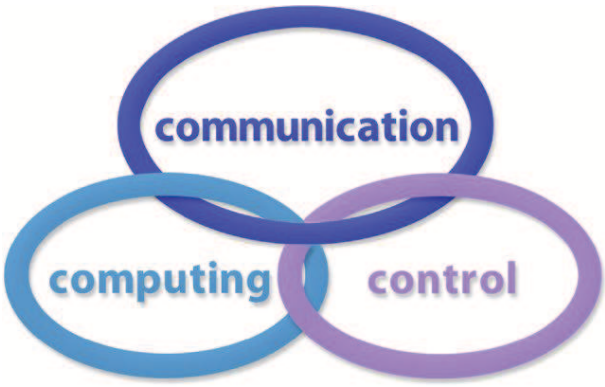
INTERNATIONAL JOURNAL

of

COMPUTERS COMMUNICATIONS & CONTROL

ISSN 1841-9836



A Bimonthly Journal
With Emphasis on the Integration of Three Technologies

Year: 2018 Volume: 13 Issue: 1 Month: February

This journal is a member of, and subscribes to the principles of, the Committee on Publication Ethics (COPE).

http://univagora.ro/jour/index.php/ijccc/

BRIEF DESCRIPTION OF JOURNAL

Publication Name: International Journal of Computers Communications & Control.
Acronym: IJCCC; **Starting year of IJCCC:** 2006.
ISO: Int. J. Comput. Commun. Control; **JCR Abbrev:** INT J COMPUT COMMUN.
International Standard Serial Number: ISSN 1841-9836.
Publisher: CCC Publications - Agora University of Oradea.
Publication frequency: Bimonthly: Issue 1 (February); Issue 2 (April); Issue 3 (June); Issue 4 (August); Issue 5 (October); Issue 6 (December).
Founders of IJCCC: Ioan DZITAC, Florin Gheorghe FILIP and Misu-Jan MANOLESCU.

Indexing/Coverage:

- Since 2006, Vol. 1 (S), IJCCC is covered by Thomson Reuters/Clarivate Analytics and is indexed in ISI Web of Science/Knowledge: Science Citation Index Expanded.

  2017 Journal Citation Reports® Science Edition (Thomson Reuters, 2016):
  Subject Category: (1) Automation & Control Systems: Q4(2009, 2011, 2012, 2013, 2014, 2015), Q3(2010, 2016); (2) Computer Science, Information Systems: Q4(2009, 2010, 2011, 2012, 2015), Q3(2013, 2014, 2016).

  Impact Factor/3 years in JCR: 0.373(2009), 0.650 (2010), 0.438(2011); 0.441(2012), 0.694(2013), 0.746(2014), 0.627(2015), **1.374(2016)**.
  Impact Factor/5 years in JCR: 0.436(2012), 0.622(2013), 0.739(2014), 0.635(2015), **1.193(2016)**.

- Since 2008 IJCCC is indexed by Scopus: **CiteScore 2016 = 1.06**.
  Subject Category:
  (1) Computational Theory and Mathematics: Q4(2009, 2010, 2012, 2015), Q3(2011, 2013, 2014, 2016);
  (2) Computer Networks and Communications: Q4(2009), Q3(2010, 2012, 2013, 2015), Q2(2011, 2014, 2016);
  (3) Computer Science Applications: Q4(2009), Q3(2010, 2011, 2012, 2013, 2014, 2015, 2016).

  SJR: 0.178(2009), 0.339(2010), 0.369(2011), 0.292(2012), 0.378(2013), 0.420(2014), 0.263(2015), 0.319(2016).

- Since 2007, 2(1), IJCCC is indexed in EBSCO.

Focus & **Scope:** International Journal of Computers Communications & Control is directed to the international communities of scientific researchers in computers, communications and control, from the universities, research units and industry. To differentiate from other similar journals, the editorial policy of IJCCC encourages the submission of original scientific papers that focus on the integration of the 3 "C" (Computing, Communications, Control).

In particular, the following topics are expected to be addressed by authors:

(1) Integrated solutions in computer-based control and communications;

(2) Computational intelligence methods & Soft computing (with particular emphasis on fuzzy logic-based methods, computing with words, ANN, evolutionary computing, collective/swarm intelligence);

(3) Advanced decision support systems (with particular emphasis on the usage of combined solvers and/or web technologies).

FAMOUS FORMER MEMBER IN THE EDITORIAL BOARD OF IJCCC

Lotfi A. Zadeh (Feb. 4, 1921 - Sept. 6, 2017)
The inventor of Fuzzy Sets, Fuzzy Logic, and Soft Computing
Former member in the Editorial Board of IJCCC between 2008-2017

FAMOUS EXCELLENCE AWARD PROPOSED FOR IJCCC

Journals Excellence Awards proposed by Scopus (2015)

EDITORIAL STAFF OF IJCCC (2018)

EDITORIAL ADDRESS:
Agora University, Cercetare Dezvoltare Agora, Tineretului 8, 410526 Oradea, Bihor, Romania,
Tel./ Fax: +40 359101032, E-mail: ijccc@univagora.ro, rd.agora@univagora.ro
URL: http://univagora.ro/jour/index.php/ijccc/

EDITORIAL BOARD OF IJCCC (MEMBERS, 2018):

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, Volume 13, Issue 1, February, 2018.

Contents

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 13(1), 8-23, February 2018.

A Comparative Study of the PSO and GA for the m-MDPDPTW

E. Ben Alaïa, I. Harbaoui, P. Borne, H. Bouchriha

Essia Ben Alaïa*, Imen Harbaoui Dridi, Hanen Bouchriha
LACCS : Laboratoire d'Analyse, Conception et Commande des Systèmes
École Nationale des Ingénieurs de Tunis, LR11ES20, Tunisie
Université de Tunis El Manar
*Corresponding author: essia.benalaia@enit.rnu.tn
imen.harbaoui@issatkr.rnu.tn
hanen.bouchriha@enit.rnu.tn

Pierre Borne
CRIStAL : Centre de Recherche en Informatique Signal et Automatique de Lille
Ecole Centrale de Lille
Villeneuve d'Ascq, France
pierre.borne@centralelille.fr

Abstract: The m-MDPDPTW is the multi-vehicles, multi-depots pick-up and delivery problem with time windows. It is an optimization vehicles routing problem which must meet requests for transport between suppliers and customers for the purpose of satisfying precedence, capacity and time constraints. This problem is a very important class of operational research, which is part of the category of NP-hard problems. Its resolution therefore requires the use of evolutionary algorithms such as Genetic Algorithms (GA) or Particle Swarm Optimization (PSO). We present, in this sense, a comparative study between two approaches based respectively on the GA and the PSO for the optimization of m-MDPDPTW. We propose, in this paper, a literature review of the Vehicle Routing Problem (VRP) and the Pick-up and Delivery Problem with Time Windows (PDPTW), present our approaches, whose objective is to give a satisfying solution to the m-MDPDPTW minimizing the total distance travelled. The performance of both approaches is evaluated using various sets instances from [10] PDPTW benchmark data problems. From our study, in the case of m-MDPDPTW problem, the proposed GA reached to better results compared with the PSO algorithm and can be considered the most appropriate model to solve our m-MDPDPTW problem.
Keywords: : Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Vehicle Routing Problem (VRP), Pick-up and Delivery Problem with Time Windows (PDPTW), m-MDPDPTW, optimization.

1 Introduction

The multi-vehicles, multi-depots pick-up and delivery problem with time windows (m-MDPDPTW) combine several variant of the well-known Vehicle Routing Problem (VRP), which is part of the category of NP-hard problems. Even for a small problem size, the resolution of such complex problem requires to use heuristic and meta-heuristic methods, since these allow finding feasible solutions, in a reasonable calculation time.

The m-MDPDPTW principle is to construct a set of routes in order to pick up and to deliver goods between a set of suppliers (pickup nodes) and a set of customers (delivery nodes). We consider several depots which does not contain any goods and where is based a homogeneous fleet of vehicles. Every single vehicle has a limited capacity and must leave and return to its starting depot. And each load must be transported by one vehicle without any transshipment at

other locations. A time window is associated with each pick-up and delivery node, thus defining for each vehicle the earliest time to visit and the latest permitted time to leave each node.

We aims to minimize the sum of travel distance without violating the different problem constraints, which are: (1) the capacity constraint which ensures that at any time in the route, the load on the vehicle must not exceed its maximum capacity. And all the vehicles leave and return to depot unloaded. (2) The precedence constraint which ensures that for each request, the origin (suppliers) precedes the destination (customers). (3) The soft time constraint: we consider that if the vehicle arrives before the earliest permitted time to visit the node, then it must waits the beginning of the time window to serve it. But, if the end of service time, in the node, is after the latest permitted time to leave it, then a tardiness time is calculated and the solution is accepted but penalized.

In this context, we develop and compare two improving optimization approach based on population search methods, which are Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) for solving our m-MDPDPTW.

This paper is organized as follows. The literature review of the Vehicle Routing Problem and the Pick-up and Delivery Problem is presented in section 2. The problem formulation and the mathematical model for the m-MDPDPTW is described in Section 3. Section 4, proposes our developed approach, based on GA and PSO, for solving a m-MDPDPTW to minimize the total travel distance. Section 5 validates and compares the proposed approaches by numerical example. Finally, the concluding remarks and further research are included in Section 6.

2 Related work

2.1 Genetic algorithm for the vehicle routing problem and the pick-up and delivery problem

Genetic algorithms are the most popular and most used evolutionary algorithms. They evolve a set of coded solutions called individuals populations. For each individual, the degree of adaptation to their local environment is measured by a predefined objective function to optimize, called fitness. From one generation to another, the best adapted individuals are selected for the reproduction by applying of genetic operations of crossover and mutation in order to produce better populations with better performing individuals. This process is repeated until a stop criterion is reached.

A literature review was proposed by [19], in which the authors detail 15 variants of the VRP and give a synthesis of the 48 heuristics for these problems. In [20], the authors propose an approach based on metaheuristic method, which combines a hybrid genetic algorithm for research and adaptive control for diversification to solve the multi-depot and periodic vehicle routing problems. In the work of [18], the authors present two metaheuristics for the resolution of the Multi depot Vehicle Routing Problem (MDVRP) which are based on a local research and a hybrid genetic algorithm.

A genetic algorithm was developed by [6] for the resolution of the multi-criteria PDPTW with multiple vehicles, based on the aggregation method and minimizing the compromise between the total travel cost, the total tardiness time and the vehicles number. This algorithm has been treated in the dynamic case in [5].

The authors in [1] proposed a multi-criteria approach based on GA for the optimization of the m-MDPDPTW. The aim is to discover a set of satisfying solutions (routes) minimizing total travel distance, total tardiness time and the total number of vehicles.

A tabu method, a genetic algorithm based on chromosome permutation, a "split" procedure and a local search are proposed by [9] to solve a particular problem of PDPTW. This Problem

considers several requests: the delivery of medicines and medical care pharmacy at home patients and the pick-up of biological samples and unused drugs at the patients.

The authors in [13] present a memetic algorithm for the resolution of PDPTW with an efficient crossover operator. Memetic algorithm is obtained by combining GA with local search.

A new genetic algorithm has been introduced in [14] for optimization of the MDVRP with capacity constraints and restrictions on the traveled distance. These authors use indirect coding and adaptive mutation operator inter-depots for the affectation of customers.

For the resolution of the MDVRP, [7] use a stochastic approach based on the GA and the fuzzy logic to adapt the crossover and mutation rates. They consider the total traveled cost and the time spent in the objective function. [2] propose an approach which is based on the combination of Genetic Algorithm (GA) with the clustering algorithm for the optimization of the m-MDPDP. The main contribution in this work is to find new depot locations in order to obtain feasible solution (routes) for the m-MDPDP.

The disadvantage of the GA is that it requires a high number of iterations. Particle Swarm Optimization (PSO) is a relatively recent heuristic algorithm which is based on the behavior of swarming characteristics of living organisms. These approaches are similar, nevertheless they each have its own particularities, its own research strategy and two different ways to develop the set of solutions.

2.2 PSO for the vehicle routing problem and the pick-up and delivery problem

The principle of the PSO is to start from an initial swarm (population) and apply a research strategy based on the cooperation of its Ne particles (individuals). In the PSO algorithm, the speed is the basic mechanism which drives the search in the promising areas of the solution space. This speed allows to put update the particles positions.

In [4], the authors present a solution approach based on particle swarm optimization (PSO) in which a local search is performed by variable neighborhood descent algorithm (VND). This approach was developed to solve the vehicle routing problem with simultaneous pickup and delivery (VRPSPD).

A hybrid particle swarm optimization (HPSO) is proposed by [16] to solve the multi-objective PDPTW. This algorithm adds particles neighbor information to diversify the particle swarm and use the variable neighborhood search (VNS) to enhance the convergence speed.

The PSO is also used to solve the vehicle routing problem with multi-depot. [8] propose an improved PSO for the multi-depot vehicle routing problem with time windows. Another PSO algorithm is proposed for solving the practical case of multi-depot vehicle routing problem with simultaneous pickup and delivery and time window [17].

A GA which evolves the VRP solutions using a PSO is proposed in [12]. This algorithm improves the performance of each individual of the population. We find other metaheuristic methods that have been developed for the resolution of the VRP who is one of the most famous combinatorial optimization problems [15], [21].

We choice to study two evolutionary algorithms GA and PSO. These are two optimization techniques based on the populations and have been widely compared in the literature. These two approaches provide a coding that perfectly represents the data of our problem. The difference is in their research strategy: The Genetic algorithm is categorizing as global research heuristics that uses crossover and mutation operators and a competition between individuals to find desired solutions, while the PSO has no evolutionary operator and in its research strategy, it gives more importance to cooperation between individuals.

Many researches in the literature showed that the PSO gives better results for the vehicle

routing problems. For our problem studied case, GAs gave better results compared to the PSO [11].

3 Mathematical model

Like the PDPTW problem, our m-MDPDPTW takes into account the following variables and parameters:
L: Set of depots;
H: Set of nodes (pick-up and delivery) $\{1, 2, ..., n\}$;
H^+: Sets of pick-up nodes;
H^-: Sets of delivery nodes;
H_c: Set of couples: delivery and pickup;
C_i: The couple (c_i, f_i) : the pick-up node (f_i) with its corresponding delivery node (c_i), $\forall i \in \{1, ..., (n/2)\}$;
V_m: Set of available vehicles from depot m,$\{V_1, ..., V_{dep}\}$;
d_{ij}: Euclidean distance between node i and j;
K: The total number of vehicles available for all the depots;
q_i: Goods quantity request of the node i, (if $q_i < 0$ it is a delivery node else if $q_i > 00$ it is a pick-up node);
t_{ij}^k: Time taken by the vehicle k to travel from node i to node j;
Q: The maximum capacity of a vehicle;
y_i^k: The load of vehicle k after leaving the node i;
ET_i: The earliest time that node i can be serviced by a vehicle;
LT_i: The latest permitted time to leave node i;
S_i: Service time at node i;
A_i: Arrival time of the assigned vehicle at the node i;
D_i: Departure time of the vehicle from the node i;
W_i: Waiting time of the vehicle at node i;
T_i: Tardiness time of the vehicle at node i.
We add the above sets
L: Set of depots $1, ..., dep$;
H_c: Set of couples: delivery and pickup; C_i: The couple c_i, f_i: the pick-up node f_i with its corresponding delivery node c_i, $\forall i \in \{1, ..., (n/2)\}$ V_m: Set of available vehicles from depot m $V_1, ..., V_{dep}$.

Our decision variable is defined as follows:

$$x_{ij}^{mk} = \begin{cases} 1 \text{ if vehicle } k \text{ originates from depot } m \text{ travel along arc } (i,j) \\ 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{otherwise} \end{cases}$$

The m-MDPDPTW considered in this study aims to minimize the total distance travelled (f_1). The objective function is formulated as:

$$f_1 = \sum_{m \in L} \sum_{k \in V_m} \sum_{i \in (H \cup m)} \sum_{j \in (H \cup m)} d_{ij} x_{ij}^{mk} \tag{1}$$

subject to:

$$\sum_{m \in L} \sum_{i \in H \cup L} \sum_{k \in V_m} x_{ij}^{mk} = 1 \qquad (\forall j \in H \cup L) \tag{2}$$

$$\sum_{m \in L} \sum_{j \in H \cup L} \sum_{k \in V_m} x_{ij}^{mk} = 1 \qquad (\forall i \in H \cup L) \tag{3}$$

$$\sum_{j\in H} x_{ij}^{mk} = \sum_{j\in H} x_{ji}^{mk} \qquad (\forall i = m \in L \, and \, k \in V_m) \tag{4}$$

$$x_{ij}^{mk} = 1 \quad \Rightarrow \quad y_i^k = 0 \qquad (\forall i \in \ L, \ j \in H \, and \, k \in V_m) \tag{5}$$

$$x_{ji}^{mk} = 1 \quad \Rightarrow \quad y_i^k = 0 \qquad (\forall i \in \ L, \ j \in H \, and \, k \in V_m) \tag{6}$$

$$x_{ij}^{mk} = 1 \quad \Rightarrow \quad y_j^k = y_i^k + q_j (\forall i, j \in H \, and \, k \in V_m \,) \tag{7}$$

$$0 < y_i^k \leq Q \qquad (\forall i \in H \text{and} k \in V_m) \tag{8}$$

$$x_{ij}^{mk} = 1 \quad \Rightarrow A_j = D_i + t_{ij}^k \qquad (\forall k \in V_m) \tag{9}$$

$$D_i = A_i + S_i \qquad (\forall i \in H) \tag{10}$$

$$D_i = S_i = 0 \qquad (\forall i \in L) \tag{11}$$

$$ET_i > A_i \quad \Rightarrow W_i = ET_i - A_i \qquad (\forall i \in H) \tag{12}$$

$$T_i = \max(0, D_i - LT_i) \qquad (\forall i \in H) \tag{13}$$

$$D_{f_i} < D_{c_i} \quad (\forall i \in \ H_c \, , f_i \in H^+ and \, c_i \in H^-) \tag{14}$$

In this formulation, constraints (2) and (3) ensure that each request is served once by the same vehicle, while constraint (4) guarantee that each vehicle starts and ends its route at the same depot. Constraints (5) and (6) impose that all the vehicles which leave and return to depot are unloaded. For each vehicle of each depot, the load of vehicle k leaving node i to j is defined in (7), while capacity constraint (8) guarantee that at any time the load, on the vehicle k, must not exceed the vehicle capacity. Each node i have time interval $[ET_i, \ LT_i]$ in which service at location i must take place. This time windows define in constraints (9) to (11) the arrival time, the departure time, service times at every depot, respectively. If the vehicle arrives at customer before the beginning of the applicable service time, a waiting time is calculated according the equation (12). And if the departure time from a node is later than its latest time of service, we calculate a tardiness time by equation (13). Finally, the precedence constraints (14) ensure that the pick-up node (f_i) of every couple i must be visited before the corresponding delivery node (c_i).

4 Optimization approach for solving the m-MDPDPTW

4.1 Solution representation

We have adopted a coding that is easy to use and to program, which fits well with the needs of our problem. Our choice was based on direct-type permutation list encoding to represent the solutions (individuals for the GA or particles for PSO) of the m-MDPDPTW. Our solution is a

sequence of genes encoded in integers. Each gene identifies a node and the order of the genes gives for each vehicle the order in which these nodes will be visited. Each coded individual contains both customers and suppliers. We chose to indicate the start and the end of each path by the depot indices. The index \0\ is not used throughout the work. Figure 1 shows a solution encoded as a direct type permutation list. The example consists of twenty nodes numbered 1 to 20, which is ten pairs (customers/suppliers), three depots (index 21, 22 and 23) and two vehicles located in each depot. In the first depot the two vehicles are used, so we will have two routes that start at the depot (index 21) and ends at the same depot. The first vehicle visit two nodes in this order (13 before 4) and the second one visit three couple that is to say six nodes. On the other hand in the third depot just one vehicle is used to visits four nodes.

Depot1	**21**	13	4	**21**	8	17	10	5	12	20	**21**
Depot2	**22**	7	18	3	6	**22**	19	2	14	16	**22**
Depot 3	**23**	1	15	11	9	**23**					

Figure 1: Solution representation

4.2 Genetic algorithm optimization for the m-MDPDPTW

Our developed approach based on GA, manipulates several types of populations and the solutions of the m-MDPDPTW are constructed using different heuristics which breaks down the problem into two major parts: regrouping then routing.

Structure of the initial population

The choice of the initial population is very important because it has an influence on the convergence speed of the genetic algorithm used. The initial population is constructed into two steps:

Step 1: The depot-grouping phase: creation of the population couple/depot

The m-MDPDPTW considers several depots in which a fixed number of vehicles are located. It is therefore necessary to determine from which depot the nodes will be served. The process rules of depot grouping phase is explained in details in the following reference [1]. At the end of this strategy, each couple (pick-up node and its associated delivery node) are assigned to the nearest depot.

Step 2: Generating the initial solution

A group of initials solutions is randomly generated, constructing the first population named Pcouple/depot containing N individuals. The chromosomes of the solution are encoded using path representation in which, for each depot, the couples are listed in the order in which they are visited [21].

The routing phase

For each depot, the number of vehicles used and the order of delivery and pick-up within each route are specified by the population named $P_{node/vehicle/depot}$. To construct this population type we should follow three steps:

Step1: Creation of population $P_{vehicle/depot}$

Knowing the number of vehicles available in each depot, we start by creating $2N$ individuals of a new population named $P_{vehicle/depot}$. This population is generated at each iteration to indicate the new number of vehicles used and the new number of couples to be visited by each vehicle.

Step 2: Use of GA operators

The genetic operators are used to create new population Pcouple/depot containing 2N individuals. Its first part represents an exact copy of the N individuals of Pcouple/depot created in phase 2, while the remaining 50 percent are created by applying GA operators on the first part. In our case, we select two parent chromosomes from population of step 2 by using tournament selection. For recombination, it is difficult to determine the most effective crossover method in advance. Therefore, we have designed our recombination operator based on the one point crossover and uniform crossover. Our crossover algorithm is adapted for permutation coding with individuals containing multiple depots. The principle is to start by applying a binary mask of the same length as the number of depots. This mask will determine: (1) the depot that will be copied: the genes of the first parent contained in this depot will be copied in the second child and those of the second parent will be copied into the first child). (2) The depot that will be crossed: the same crossing point is chosen in each depot, for the two parent chromosomes. The first part of each child is copied, gene by gene, from its parent. The first child will be completed with genes that were not inherited from the first parent but rearranged according to their order of appearance in the second parent. And we apply the same procedure to complete the second child using the order of appearance in the first parent.This procedure is repeated until the crossing rate ($T_c = 0.8$) is reached (80% of children population size is reached). For diversification, what remains of the population (2 individuals) will be mutated with applying swap mutation according to a fixed probability ($T_m = 0.2$).

Step 3: Population $P_{node/vehicle/depot}$

For each depot, the number of vehicles used and the order of delivery and pick-up within each route are specified by the population named $P_{node/vehicle/depot}$. To construct this population type we should follow three steps: Knowing the number of vehicles available in each depot, we start by creating $2N$ individuals of a new population that indicates the number of nodes visited by each vehicle, named $_{P}vehicle/depot$. After, we affect visited couples to vehicles by coding each individual of the population Pcouple/depot created in phase 3 by an individual of the population $P_{vehicle/depot}$. This new population, named Pcouple/vehicle/depot, verifies that each couple belongs exactly to same route. To find, for each depot the order in which all pick-up and delivery nodes are visited, we develop a heuristic algorithm. Its principle is to choose randomly, in each route, a starting node from the assigned couples. Then, we follow it by the closest node. Our heuristic minimizes the total distance traveled for each individual of $P_{node/vehicle/depot}$.

Figure 2 and Figure 3 show an example of the individual of $P_{vehicle/depot}$ and $P_{node/vehicle/depot}$.

	k_1	k_2	k_3
Depot1	6	2	0
Depot2	4	4	
Depot3	4	0	

Figure 2: Example of individual of $P_{vehicle/depot}$

For this example we have a vehicles total number equal to 7 and 10 couples customer/supplier which are defined as follows:

Depot 1	**21**	8	13	4	17	10	5	**21**	12	20	**21**
Depot 2	**22**	19	7	2	18	**22**	3	6	14	16	**22**
Depot 3	**23**	1	15	11	9	**23**					

Figure 3: Example of individual of $P_{node/vehicle/depot}$

$C_1(13,8), C_2(10,4), C_3(20,12), C_4(5,17), C_5(7,19), C_6(18,2), C_7(14,3), C_8(16,6), C_9(9,1), C_10(11,15)$.

Heuristics for creation of feasible solutions

Each individual of population Pnode/vehicle/depot must respect different constraints. The precedence constraint ensures that each delivery point, on the same route, is not visited before its supplier. The capacity correction constraint ensures that the total load of the vehicle must be smaller than or equal to the maximum capacity of the vehicle. The heuristics algorithms for precedence and capacity corrections procedures, to transform each individual into feasible solution, are explained in details in [1].

The structure of the genetic algorithm proposed for the m-MDPDPTW optimization is illustrated in Table 1. In our genetic algorithm we use the elitism strategy for each generation, the best N solutions in the current population are copied in the new initial population. The best solution found throughout the search is returned when a fixed number of iterations is reached. This number is determined after several experiments.

Table 1: Structure of the genetic algorithm proposed for the m-MDPDPTW optimization

Begin **Step 1:** Apply Depot-Grouping phase
Step 2: Generate randomly an initial population $P_{couple/depot}$ containing N individuals. **Repeat until maximum of generation reached**.
Step 3: Create a new $P_{couple/depot}$ containing $2 \times N$ individuals. The first part of this population represents one copy of the N individual $P_{Best-couple}$, while the remaining 50 percentage of this population are created by applying GA operators on population $P_{Best-couple}$. We select two parent chromosomes from population of step 2 by using tournament selection. For recombination, we apply our designed crossover and for diversification, we apply swap mutation according to a fixed probability.
Step 4: Generate $P_{vehicle/depot}$ containing $2 \times N$ individuals and respecting constraint vehicle numbers.
Step 5: Apply routing phase to create $P_{node/vehicle/depot}$. This population with size $[2 \times N \times m]$ specifies, for each depot, the number of routes (that are vehicles) and the order of delivery and pick-up within each route.
Step 6: Apply the precedence then the capacity correction procedure, to transform each individual into feasible solution.
Step 7: Calculate for every individual of $P_{node/vehicle/depot}$ fitness values (f_1, f_2).
Step 8 : Elitism: Copy the N best $P_{couple/depot/vehicle}$ solution into the new initial population.
Increment the generation number
End

4.3 Particle Swarm Optimization for the m-MDPDPTW

General principle of the algorithm PSO proposed

PSO shares many similarities with GAs. All two techniques begin with a group of a randomly generated population; both utilize a fitness value to evaluate the population.

The principle of the PSO is to start from an initial swarm and to apply a research strategy based on the cooperation of its Ne particles. The search for optimums is done by generating several generations. In every generation, a potential solution to the problem is generated, and then evaluated to record the best solutions found. The solution to our m-MDPDPTW represents the best solution found on all generations.

The main difference between the PSO approaches compared to GA, is that PSO does not have genetic operators such as crossover and mutation. Particles update themselves with the internal velocity. In PSO, only the best particle gives out the information to others. It is an on way information sharing mechanism, the evolution only looks for the best solution.

In most applications, the particles positions represent the solutions of the problem studied, but in our case the solution to m-MDPDPTW is decoded from the new particle position.

The travel strategy of a particle i is illustrated in Figure 4.

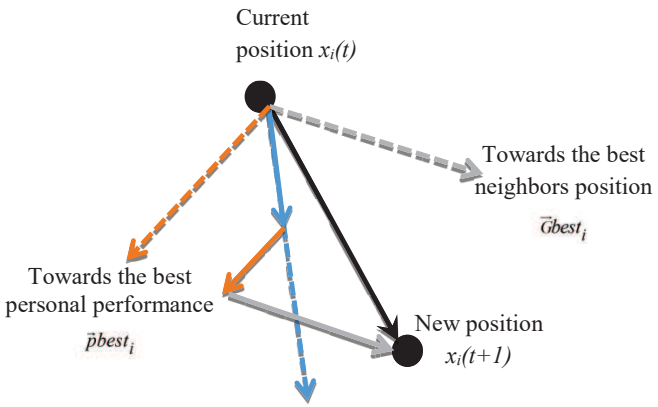


Figure 4: Analysis of the particle travel

For the creation of the initial swarm $P_{node/depot}$ and the Decoding of the initial solutions: routing phase for the creation of the swarm Pnode/vehicle/depot we follow the same steps as in the genetic algorithm.

Initializing and updating speed and position

We consider that the particles movement is controlled by the limitation of their traveled maximum distance during iteration. Thus, in order to escape the deviation problem of the particles from the search space, we use a technique of interval confinement. The speed of each particle is initialized by values between 0 and n (n is the nodes number to be visited).

The update of the speed and the best position is obtained by the particle ($\vec{P}best_i$: personal best) and by the swarm ($\vec{G}best_i$: global best), using equations (15), (16) and (17). The new position of the particle is calculated from equation (18).

$$\vec{P}best_i(t+1) = \begin{cases} \vec{x}_i(t+1), & if\ f(\vec{x}_i(t+1))\ is\text{better than}\ f(\vec{P}best_i(t)) \\ \vec{P}best_i(t), & else \end{cases} \tag{15}$$

$$\vec{G}best(t+1) = \arg\min_{\vec{P}best_i} f\left(\vec{P}best_i(t+1)\right), 1 \leq i \leq N \tag{16}$$

$$v_{i,j}(t+1) = \begin{cases} w\ v_{i,j}(t) + \\ c_1 r1_{i,j}(t)(pbest_{i,j}(t) - x_{ij}(t)) + \\ c_2 r2_{i,j}(t)(gbest_j(t) - x_{i,j}(t)) \end{cases} \tag{17}$$

$$x_{i,j}(t+1) = x_{i,j}(t) + v_{i,j}(t+1) \tag{18}$$

with :
$x_{i,j}(t)$: The position of the particle i in dimension j at time t;
$v_{i,j}(t)$: The speed of the particle i in dimension j at time t;
$pbest_{i,j}(t)$: The best position obtained by the particle i in dimension j at times t;
$gbest_{i,j}(t)$: The best position known by the particle i at time t;
c_1, c_2: Acceleration coefficients;
r_1, r_2: Random numbers drawn uniformly in [0, 1], at each iteration t and for each dimension j;
$w\,\vec{v_i}$: The inertia component of the movement;
$c_1 r_1(pbest_i - \vec{x_i})$: The cognitive component of the particle movement (moving to its best known position);
$c_2 r_2(gbest_i - \vec{x_i})$: The social component of the particle movement (closer to the best position of its neighbors).

Decoding the new particle position

The elements of the new particle position after its updating do not reveal directly the nodes index in the route. A decoding phase of the new visit order is therefore necessary in order to find the solutions adapted to m-MDPDPTW.

We consider that the speed is defined as a probability vector, where the value of each element corresponds to the probability of its permutation in the route.

The decoding phase of the new particles gives us the new permutation of the nodes order in the route.

In the same route, nodes will be visited from the nearest to the furthest. If there are several nodes that are in the same position, then they will be visited according to the speed of their movement. The node with the highest speed will be visited first.

We therefore reorder in ascending order of the values of $\vec{x_i}(t+1)$.

These values are subsequently replaced by the corresponding node index to build the new visit order of the nodes.

The structure of the *adaptation algorithm of the PSO for the m-MDPDPTW optimization* is illustrated in the following.

n: nodes number to visit; f : function to minimize ; T: maximum number of iterations; *dep*: number of depot; $t = 0$; $x_{max} = n$; $x_{min} = 0$;

1. Initialization : N_e : Swarm size ; c_1; c_2; r_1; r_2;
2. Generating N_e initial particles of $P_{node/depot}$;
3. Application of the routing phase for the creation of initial solutions $P_{node/vehicle/depot}$;

4. Application of the corrections heuristics of precedence, capacity and belonging of each couple to the same route;
5. Evaluate the Ne initial particles of $P_{node/vehicle/depot}$ (1);
6. Initialize the speed of each particle by random values between [0; n];
7. Initialize the position of each particle

$$\vec{x}(0) = P_{node/depot}$$

8. Initialize

$$\vec{P}best_i(0) = P_{node/vehicle/depot}[i]$$

9. Initialize the best solution found by the swarm

$$\vec{G}best(0) = \arg \min_{\vec{P}\,best_i} f(\vec{P}best_i(0))$$

10. While $(t < T)$ do;
11. For $i = 1$ to N_e do;
12. Speed Update (18);
13. Update of the new position (18);
14. Verify the new particle does not leave the search space:

$$\begin{aligned} &\text{if}(\vec{x}_i(t+1) > x_{\max})then\vec{x}_i(t+1) = x_{\max};\\ &else\text{if}(\vec{x}_i(t+1) < x_{\min})then\vec{x}_i(t+1) = x_{\min}; \end{aligned}$$

15. Decoding the new particles position;
16. Generating new solutions: Routing phase;
17. Apply corrections heuristics;
18. Evaluate the N_e particles of $P_{node/vehicle/depot}$ (1);
19. Save the best result found by the particle;
    For $i = 1$ to N_e do;

$$\begin{aligned} &\text{if}f(P_{node/vehicle/depot}(i)) \leq f(pbest_i(t))\text{then}\\ &\text{Pbest}_\text{i}(t+1) = f(x(t+1))\\ &else\text{Pbest}_\text{i}(t+1) = Pbest_i(t) \end{aligned}$$

    End For;
20. Update the best solution found by the swarm

$$\vec{G}best(t+1) = \arg \min_{\vec{P}\,best_i} f(\vec{P}best_i(t+1))$$

21. End For;

22. End While;
    Save the best solution found on all generation;

The same, to find the optimal solution, we browse the swarm $P_{node/vehicle/depot}$ particle by particle, to determine that which minimizes our objective function (1).

5 Computational results

This section describes computational results to assess and compare the performance of the two proposed algorithms. We programmed the algorithms in C language using Microsoft Visual Studio2010 and ran them on Mobile Intel Core i7, CPU 2.50 GHz and 6.00 Go memory (RAM) under the operating system Windows 8 Professional.

In literature, the existing benchmark instances do not combine all the characteristics of the m- MDPDPTW. So our simulation results are obtained using the problem instances generated by Li and Lim [10] which are created for the single depot pickup and delivery problem with time windows. We only added the additional information that can accommodate the m-MDPDPTW by generating multi depot locations using the algorithm described in our work [2]. The locations of the different depots considered in our simulations for each type of Li and Lim's problems are summarized in Table 2.

Table 2: Location of the different depots

Instances	Coordinates of the 1^{st} depot	Coordinates of the 2^{nd} depot
LC	(40 , 50)	(34 , 32)
LR	(35 , 35)	(60 , 40)
LRC	(40 , 50)	(65 , 30)

The parameters of the two approaches are selected after many experiments.

The parameters values for PSO are: Number of swarms=1, swarm size=500 particles, the maximum number of generation=1000 iterations. The inertia weight linearly decreasing from 0.8 to 0.5, the acceleration constants, random numbers for personal best and global best are 0.2, 0.2, 0.3 and 0.5.

The parameters values for GA are: Population size equal to number of particles=500, maximum number of generations=1000. The crossover rate and mutation rate are 0.8 and 0.2, respectively.

The best results obtained from our two proposed GA and PSO approaches solving our m-MDPDPTW were compared as shows in Table 3.

We consider the depot locations already given in the Table 1 and we evaluate the total distance travelled, considered in equation 1, for some selected categories of problem instances with different sizes: clustered locations with short schedule horizon (LC101 and LC102), clustered locations with long schedule horizon (LC201 and LC202), randomly distributed locations with short schedule horizon (LR101 and LR102), randomly distributed locations with long schedule horizon (LR201 and LR202), and finally, problems category that have partially random and

partially clustered locations with a tight time window width (LRC101 and LRC102) and with a large time window width (LRC201 and LRC202).

Table 3: Comparison of the results of our GA and PSO approaches to solve our m-MDPDPTW problem

Instances	Best distance of GA approach	Best distance of PSO approach
LC101	916.078	1839.3962
LC102	910.598	1606.801
LC201	1081.395	947.914
LC202	1204.900	1472.292
LR101	1938.980	2137.17
LR102	1898.109	2128.639
LR201	1979.990	2060.405
LR202	1241.264	2497.547
LRC101	1079.137	2513.376
LRC102	1129.326	2567.438
LRC201	1453.306	2937.202
LRC202	1421.923	2829.999

From Table 3, we can observe that the values of the total distance traveled given by our approach based on the GA are better than those given by the PSO. This last only gave a low improvement of the fitness for clustered locations instance LC201.

However, these results show the strength of GAs that is in the parallel nature of their research and their ability to manage multiple large data sets.

Compared with GA, the advantages of PSO are that it is easy to implement and there are few parameters to adjust. A source of the AG's power is their used genetic operators: crossover and mutation. The crossover attempts to preserve the beneficial aspects of candidate solutions and to eliminate undesirable components, while the random nature of mutation is probably more likely to degrade a strong candidate solution than to improve it.

Through its genetic operators, even weak solutions may continue to be part of the makeup of future candidate solutions and thus allows the creation of new solutions that have, a higher probability of exhibiting a good performance. This tends to make the algorithm likely to converge towards high quality solutions within a few generations.

To validate and evaluate our proposed AG approach, we solve the single depot PDPTW

Table 4: Comparison of the results of our AG approaches using the Li and Lim instances of single depot problem

Case	Best known solutions of Li and Lim	Best known solutions of our GA approach
LC101	828.94	861.24
LC102	828.94	926.941
LC103	1035.35	959.303
LC104	860.01	940.05
LC105	828.94	867.609
LC106	828.94	955.339
LC107	828.94	922.66
LC108	826.44	914.211
LC109	1000.60	869.537

problem. The comparison with the best results published by Li and Lim with respect to the total traveled distance minimization is shown in Table 4.

The proposed GA approach shows the promising result in the clustered problem with short schedule horizon. Compared with the Lim and Lim's best solutions approach for the PDPTW problem with single depot, this algorithm produced good quality results that are sometimes even better than the results obtained, making it even more suitable for population-based solution algorithms.

Furthermore, the total traveled distances for the benchmark data sets are near the best result and are a little better than the best known distance in LC103 and LC109 category.

6 Conclusion

In this paper we have developed two meta-heuristic approaches based on the Genetic Algorithm (GA) and Particle swarm optimization (PSO), in order to compare and identify the best approach that can be used for solving a multi depots multi vehicles pickup and delivery problems with time windows (m-MPDPTW). We proposed a brief literature review on the VRP and the PDPTW. The mathematical formulation of our problem was subsequently presented. Then, we have detailed the use of GA and PSO algorithm to determine the solution which minimizes our

objective function. Simulation was presented in a last part by using benchmark's data. The experimental results on a large number of benchmark instances indicate that the use of GA seems to be the most favorable method to reach final best solutions for our m-MDPDPTW. For our future work, we propose to study the multi-objective optimization by comparing the genetic algorithms, the Particle Swarm Optimization and other metaheuristic methods.

Bibliography

[1] Ben Alaia, E.; Harbaoui, D.I.; Bouchriha, H.; Borne, P. (2015); Genetic Algorithm for Multi-Criteria Optimization of Multi-Depots Pickup and Delivery Problem with Time Windows and multi vehicles, *Acta Polytechnica Hungarica*, 12(8), 155–174, 2015.

[2] Ben Alaia, E.; Harbaoui, D.I.; Bouchriha, H.; Borne, P. (2015); Insertion of new depot locations for the optimization of multi-vehicles Multi-Depots Pickup and Delivery Problems using Genetic Algorithm, *IEEE International Conference on Industrial Engineering and Systems Management*, Seville, Spain, 695-701, 2015.

[3] Ben Alaia, E.; Harbaoui, D.I.; Bouchriha, H.; Borne, P. (2015); Genetic Algorithm with Pareto Front selection for Multi-Criteria Optimization of Multi-Depots and multi- vehicle Pickup and Delivery Problems with Time Windows, *IEEE International Conference on Sciences and Techniques of Automatic control and computer engineering*, 21-23 December, Hammamet, Tunisie, 488-493, 2014.

[4] Fatma, P.G.; Fulya, A.; Ismail, K. (2012); A Hybrid Particle Swarm Optimization for Vehicle Routing Problem with Simultaneous Pickup and Delivery, *Computers and Industrial Engineering*, 1-6, 2012.

[5] Harbaoui, D.I., Ben Alaia, E.; Borne, P. (2015); Heuristic Approach for The Optimization of The Dynamic Multi-Vehicle Pickup and Delivery Problem with Time Windows, *IEEE International Conference on Industrial Engineering and Systems Management*, 21–23 October, Seville, Spain, 488-493, 2015.

[6] Harbaoui, D.I.; Kammarti, R., Ksouri, M.; Borne, P. (2011); Multi-objective optimization for the mpdptw : Aggregation methode with use of genetic algorithm and lower bounds, *International Journal of Computers Communications & Control*, 6(2), 246–257, 2011.

[7] Lau, H.C.W.; Chan, T.M.; Tsui, W.T.; Pang, W.K. (2010); Application of genetic algorithms to solve the multidepot vehicle routing problem, *IEEE Transactions on Automation Science and Engineering*, 7(2), 383–392, 2010.

[8] Lei, W.; Fanhua, M. (2008); An Improved PSO for the Multi-Depot Vehicle Routing Problem with Time Windows, *Pacific-Asia Workshop on Computational Intelligence and Industrial Application*, 852-856, 2008.

[9] Liu, R.; Xi, X.; Augusto, V.; Rodriguez, C. (2013); Heuristic algorithms fora vehicle routing problem with simultaneous delivery and pickup andtime windows in home health care, *European Journal of Operational Research*, 230(3), 475-486, 2013.

[10] Li, H.; Lim, A. (2001); A metaheuristic for the pickup and delivery problem with time windows, *In IEEE International Conference on Tools with Artificial Intelligence*, 13, 160–167, 2001.

[11] Marinakis, Y.; Iordanidou, G.R.; Marinaki, M (2013); Particle Swarm Optimization for the Vehicle Routing Problem with Stochastic Demands, *Applied Soft Computing*, 13 (4), 1693–1704, 2013.

[12] Marinakis, Y.; Marinaki, M. (2010); Ahybrid genetic - particle swarm optimization algorithm for the vehicle routing problem, *Expert Systems with Applications*, 37(2), 1446–1455, 2010.

[13] Nagata, Y.; Kobayashi, S. (2011); A memetic algorithm for the pickup and delivery problem with time windows using selective route exchange crossover, *In: Schaefer R., Cotta C., Kolodziej J., Rudolph G. (eds), Parallel Problem Solving from Nature, PPSN XI. PPSN 2010. Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, 6238, 536–545, 2011.

[14] Ombuki, B. B.; Hansharin, F. (2009); Using genetic algorithms for multi depot vehicle routing, *Studies in Computational Intelligence*, 161, 77–99, Springer Berlin Heidelberg, 2009.

[15] Pop, P.C.; Pop Sitar, C.; Zelina, I.; Lupse, V., Chira, C. (2011); Heuristic Algorithms for Solving the Generalized Vehicle Routing Problem, *International Journal of Computers Communications & Control*, 2011(1), 158–165, 2011.

[16] Shuilong, Z.; Jin, L.; Xueqian, L. (2013); A Hybrid Particle Swarm Optimization Algorithm for Multi-Objective Pickup and Delivery Problem with Time Windows, *Journal of Computers*, 8(10),2583-2589, 2013.

[17] Sombuntham, P., Kachitvichayanukul, V. (2010); A Particle Swarm Optimization Algorithm for Multi-depot Vehicle Routing problem with Pickup and Delivery Requests, *The International MultiConference of Engineers and Computer Scientists*, 1-6, 2010.

[18] Vidal, T.; Crainic, T.G.; Gendreau, M., Prins, C. (2014); Implicit depot assignments and rotations in vehicle routing heuristics, *European Journal of Operational Research*, 237(1), 15–28, 2014.

[19] Vidal, T.; Crainic, T G.; Gendreau, M.; Prins, C. (2013); Heuristics for multiattribute vehicle routing problems : a survey and synthesis, *European Journal of Operational Research*, 231(1), 1–21, 2013.

[20] Vidal, T.; Crainic, T.G.; Gendreau, M.; Lahrichi, N.; Rei, W. (2012); A hybrid genetic algorithm for multi-depot and periodic vehicle routing problems, *Operations Research*, 60(3), 611–624, 2012.

[21] Yousefikhoshbakht, M.; Didehvar, F.; Rahmati, F. (2014); An Efficient Solution for the VRP by Using a Hybrid Elite Ant System, *International Journal of Computers Communications & Control*, 9(3), 340-347, 2014.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 13(1), 24-38, February 2018.

Selective Feature Generation Method for Classification of Low-dimensional Data

S.-I. Choi, S.T. Choi, H. Yoo

Sang-Il Choi, Haanju Yoo
Department of Computer Science and Engineering
Dankook University
152, Jukjeon-ro, Suji-gu, Yongin-si
Gyeonggi-do, 16890, Korea
choisi@dankook.ac.kr, haanju.yoo@gmail.com

Sang Tae Choi*
Department of Internal Medicine
Chung-Ang University College of Medicine
102 Heukseok-ro, Dongjak-gu
Seoul, 06974, Korea.
*Corresponding author: beconst@cau.ac.kr

Abstract: We propose a method that generates input features to effectively classify low-dimensional data. To do this, we first generate high-order terms for the input features of the original low-dimensional data to form a candidate set of new input features. Then, the discrimination power of the candidate input features is quantitatively evaluated by calculating the 'discrimination distance' for each candidate feature. As a result, only candidates with a large amount of discriminative information are selected to create a new input feature vector, and the discriminant features that are to be used as input to the classifier are extracted from the new input feature vectors by using a subspace discriminant analysis. Experiments on low-dimensional data sets in the UCI machine learning repository and several kinds of low-resolution facial image data show that the proposed method improves the classification performance of low-dimensional data by generating features.
Keywords: feature generation, input feature selection, feature extraction, discriminant distance, low-dimensional data, data classification.

1 Introduction

Advances in information technology have resulted in a rapid increase in the amount of digital data that is available, and a significant amount of research has been carried out to develop tools to extract useful and necessary information from vast amounts of data. Such tools are currently being applied in various fields, including biometrics (e.g., iris, fingerprint and face recognition), data mining, diagnosis systems and pattern classification [22, 26].

When working with data samples, which are represented as 'input features', feature extraction methods can effectively improve classification performance by extracting useful information. When there are input features in a data sample, feature extraction methods find projection vectors to get new features containing the maximal information for problem solving [4, 14, 15, 27, 31]. Then, an input data sample is represented by a set of new features (feature vector), each of which is a linear combination of the input features.

The different feature extraction methods have different properties, and the appropriate method must be used corresponding to the characteristics of the data and the problem that is to be solved, e.g., data representation, classification, restoration, etc. Common feature extraction methods such as Principal Component Analysis (PCA) [27] and Linear Discriminant Analysis

(LDA) [15] have been the basis to develop other methods, including Null space LDA (NLDA) [4], Biased Discriminant Analysis (BDA) [31], etc. In these methods, data is stored in vector form, and the appropriate features are extracted using a covariance matrix which is appropriately defined depending on the problem to be solved. Methods such as MatFLDA [5], Two-Dimensional LDA (2DFLD) [29], Composite LDA (C-LDA) [18] and Composite BDA (C-BDA) [17] use an image covariance matrix instead of the covariance matrix. These image covariance-based methods can be used effectively for data in which input features are strongly correlated [17]. C-LDA can be viewed as a generalized image covariance-based method because C-LDA becomes identical to the 2DLDA or MatFLDA form when the composite vector is defined as a row or column vector.

In classification problems, an object is described as an array of attributes to search for the underlying patterns in the object. These attributes are represented as numerical values, which are stored in a vector form (input feature vector) [8]. For example, for blood test data for a person in a hospital, the dimension of the data is the number of test items. Even when using the same object, the attributes can be defined in different ways depending on the problem that is to be classified. For example, when classifying a dog, attributes such as food or skeletal structure can be used to classify species of mammals, amphibians, and the like, and when distinguishing individual objects belonging to the same group of animals, attributes such as hair color, size, age, etc. can be used. However, when expressing an object with attributes in this manner, the number of attributes is limited, and it is usually represented using low-dimensional data. On the other hand, temporal sensing data such as speech, or spatial data such as images is usually stored as high-dimensional data. Even such data is often reduced and stored as low-dimensional data such as a thumbnail image in order to effectively use the data in a small device, which has a relatively small computing power.

Most feature extraction methods mentioned above use a statistical correlation of input features and extract features from the shape information of the pixels constituting the image, so their classification performance is limited when the number of input features is too small and is affected by the resolution of the image. In the case of the DCV method, which offers a high performance for generic high-dimensional data, the dimension of the null space may decrease or disappear when the dimension of the data decreases. Therefore, it is necessary to generate meaningful features from the input features to effectively utilize the existing data classification techniques with low-dimensional data.

In this paper, we propose an input feature generation method for classification of low-dimensional data. According to the Theorem of Cover [10], if data samples are not distributed linearly and separably, they can be made into a linearly separated distribution through conversion into higher dimensions. Many methods use kernel functions to convert low-dimensional data into higher dimensions [9, 24, 28, 30]. These methods use a kernel matrix instead of directly computing kernel functions because doing so would require extensive computation. However, in this case, since the value of the high-dimensional data that is created can not be confirmed, even if the feature corresponding to the individual dimension of the high-dimensional data includes unnecessary information that do not help in classification, they can not be removed or separately used. In the proposed method, new input features are generated by adding a higher order term of individual input features, and the separability power for the original input features and the generated input features is measured using the discriminant distance scale [21]. Then, only features with high discrimination information are selectively used during data classification. The new input features improves the performance of existing discriminant feature extraction methods especially when classifying low-dimensional data. We recently investigated the feature generation method for face recognition and presented preliminary results in [6]. In this paper, we provide a more detailed analysis of the method, as well as an extensive discussion, and we apply the method to other classification problems other than face recognition. Through experiments on

various low-dimensional data sets, we confirmed that the classification performance is improved when using the proposed input feature generation method. The results of the experiment for low-resolution facial images show that the proposed method offers a higher recognition rate than when the resolution of such images is increased via interpolation.

This paper is organized as follows. In the next section, we examine the effect of the data dimension on the classification performance. Then, we describe the feature generation method and the optimal input feature selection method. Finally, the experimental results are described and the conclusion follows.

2 Effect of data dimensionality on classification performance

2.1 Subspace discriminant analysis

Subspace discriminant analysis methods represent a data sample as an n-dimensional vector $\mathbf{x}$. LDA, NLDA and BDA are representative methods of these subspace discriminant analysis methods. When there are N data samples with C classes and N_i samples for each class $c_i (i = 1, ..., C)$, the within class scatter matrix S_W and the between class scatter matrix S_B can be defined as follows:

$$\begin{aligned} S_W &= \sum_{i=1}^{C} \sum_{\mathbf{x_k} \in c_i} (\mathbf{x_k} - \boldsymbol{\mu}_i)(\mathbf{x}_k - \boldsymbol{\mu}_i)^T, \\ S_B &= \sum_{i=1}^{C} N_i (\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T \end{aligned} \tag{1}$$

where $\boldsymbol{\mu}_i$ is the average of the samples in class c_i and $\boldsymbol{\mu}$ is the average of all N samples.

LDA finds a projection matrix $W_{Fisher} = [\boldsymbol{\mu}_1, \cdots, \mathbf{w}_{C-1}]$ consisting of projection vectors $\mathbf{w}_l (l = 1, \cdots, C-1)$ that satisfies the following objective function. This means that the LDA constructs a feature space that maximizes the covariance between the other classes while minimizing the covariance between the same classes in the range space of S_W [15]. W_{LDA} can be obtained by calculating the eigenvectors of $S_W^{-1} S_B$.

$$W_{Fisher} = argmax_W \frac{\left|W^T S_B W\right|}{|W^T S_W W|} \tag{2}$$

Unlike the LDA, which uses the range space of S_W, the NLDA uses the null space of S_W containing more discriminating information [4]. That is, a projection matrix W_{DCV} satisfying the following objective function is obtained in a space of $\left|W^T S_W W\right| = 0$ and $\left|W^T S_B W\right| \neq 0$.

$$W_{DCV} = argmax^T_{|W^T S_W W|=0} \left|W^T S_B W\right| \tag{3}$$

NLDA shows good performance especially when the number of input features of the data samples and the null space of S_W are large.

BDA is a modified form of LDA. Unlike LDA, which maximizes the distance of the mean values of classes in a multi-class classification problem, the BDA aims to classify one class of interest and the rest [31]. The BDA constructs a positive sample in the form of a normal distribution, and negative samples constitute a feature space that is distributed away from the mean of positive samples, and has the following objective function. Assuming that (i) the data samples $\mathbf{x}^P$ and $\mathbf{x}^N$ are positive and negative samples, respectively, (ii) their numbers are N_P

and N_N, respectively, and (iii) the average of the positive samples is $\boldsymbol{\mu}^P$, the scatter matrix of the positive samples S_P and the scattering matrix for the negative samples are defined as shown in Eq. (4). The objective function of BDA is defined as shown in Eq. (5).

$$\begin{aligned} S_P &= \sum_{k=1}^{N_P} (\mathbf{x}_k^P - \boldsymbol{\mu}^P)(\mathbf{x}_k^P - \boldsymbol{\mu}^P)^T \\ S_N &= \sum_{k=1}^{N_N} (\mathbf{x}_k^N - \boldsymbol{\mu}^P)(\mathbf{x}_k^N - \boldsymbol{\mu}^P)^T \end{aligned} \tag{4}$$

$$W_{BDA} = argmax_W \frac{\left|W^T S_N W\right|}{|W^T S_P W|} \tag{5}$$

To avoid the small sample size problem [11], we use ν and γ instead of $S_N^R = (1-\nu)S_N + \frac{\nu}{n} tr[S_N I]$ and $S_P^R = (1-\gamma)S_P + \frac{\gamma}{n} tr[S_P I]$ by using a regularization factor S_N and S_P for each scattering matrix [31]. After investigating classification rates for various values of ν and γ, we set ν and γ to 0 and 0.1, respectively.

In the subspace-based analyses, after finding W in the training phase, the feature vector ($\mathbf{y} \in R^{m\times 1}$, $m < n$) for a given sample $\mathbf{x}$ can be obtained through a linear transformation as $\mathbf{y} = W^T\mathbf{x}$. Also, the problem is effectively solved by defining the covariance matrices and objective function according to the particular type of problem. However, the number of input features should be secured for the covariance analysis of the input features to be successful. Besides, some methods, such as NLDA, may not be able to conduct an analysis if the number of input features is less than the number of samples. Therefore, to more efficiently use subspace discriminant analysis, it is necessary to ensure a certain number of input features.

2.2 Classification performance over data dimensionality

To confirm the effect of the dimension of the data sample on the classification performance in the subspace discriminant analysis, it is necessary to examine how the classification rate changes with respect to the data representing the same object with vectors of different dimensions. As an example, we performed recognition experiments on facial images with various resolutions [6]. We have experimented on images with 120×100, 60×50, 30×25, 24×20 and 15×12 resolution for the FERET database [25], CMU-PIE database [1], Yale B [12] and Yale database [32] database (Fig. 1). The NLDA method was used for 120×100, 60×50, 30×25, and 24×20 images, and the LDA method [2] was used for 15×12 images because there is no null space of S_W.

As can be seen in Fig. 2, the recognition rate decreases as the resolution decreases in all databases. The recognition rate of the 15×12 images, which can not use the NLDA method, is significantly lower than that for the 120×100 to 24×20 images because the applicable classification methods are limited when the dimension of the data is low. As a result, when data is a dimension higher than a certain level, it is possible to attempt effective classification using various methods. On the other hand, the variations in illumination and facial expression in facial images can be regarded as a kind of noise. In this sense, the FERET database, which has less variation in images than the CMU-PIE, Yale B and Yale databases, can be regarded as relatively noiseless.

The results of the experiment for the FERET database show that the recognition rates for 120×100, 60×50 and 24×20 images are almost the same. This indicates that, when the influence of the noise is not large, if the dimension of the data becomes larger than a certain level, there is no further advantage in classification accuracy, and the amount of unnecessary calculation

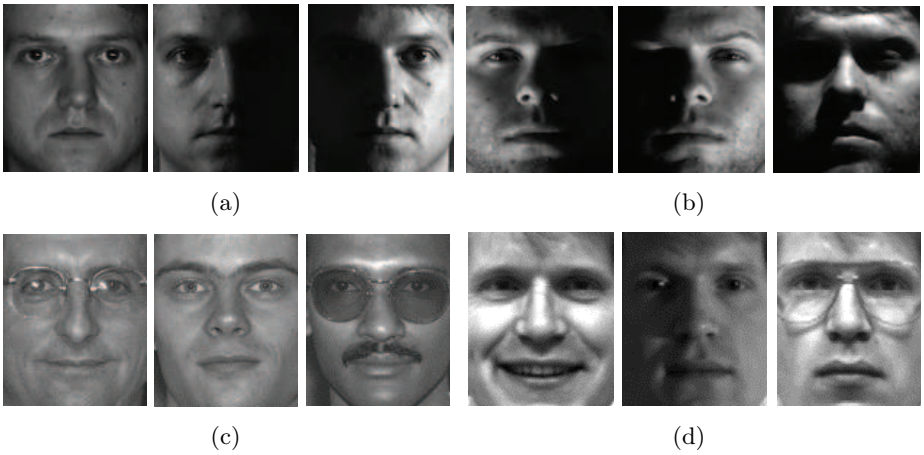


Figure 1: Examples from (a) CMU-PIE database. (b) Yale B database. (c) FERET database. (d) Yale database.

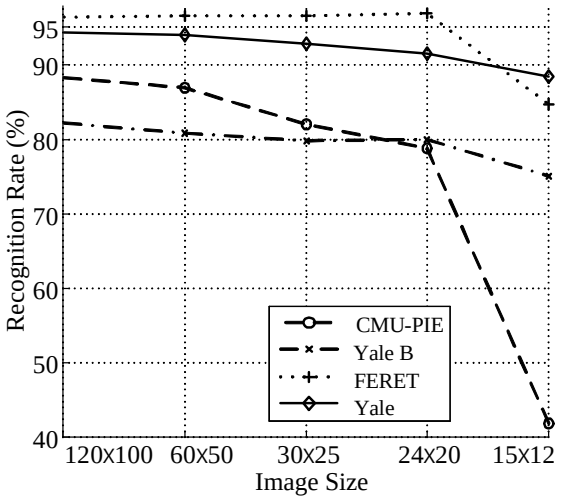


Figure 2: Face recognition rates for various face image resolutions.

increases due to data redundancy. Therefore, to efficiently classify the data, it is necessary to construct appropriately sized data.

3 Feature generation and construction of optimal features

As noted above, low-dimensional data samples may have limitations when classified only with the original input features. Therefore, to improve the performance of the data classification, it is desirable to increase the separability of the samples by converting the dataset with the samples into a high-dimensional space through a non-linear transformation $\varphi(\cdot)$ (Cover' theorem [10], Fig. 3). One simple way to increase the dimension of the input feature space is to create and add a higher order term from the input features of the data sample.
In this paper, we use the correlation between the input features as a new feature by adding the quadratic term $(x_i x_j, (i,j = 1,...,n))$ of the input features (pixels) of the data sample $(\mathbf{x} = [x_1,...,x_n]^T)$. The dimension of the data increases through the addition of a higher order n_{new} as follows.

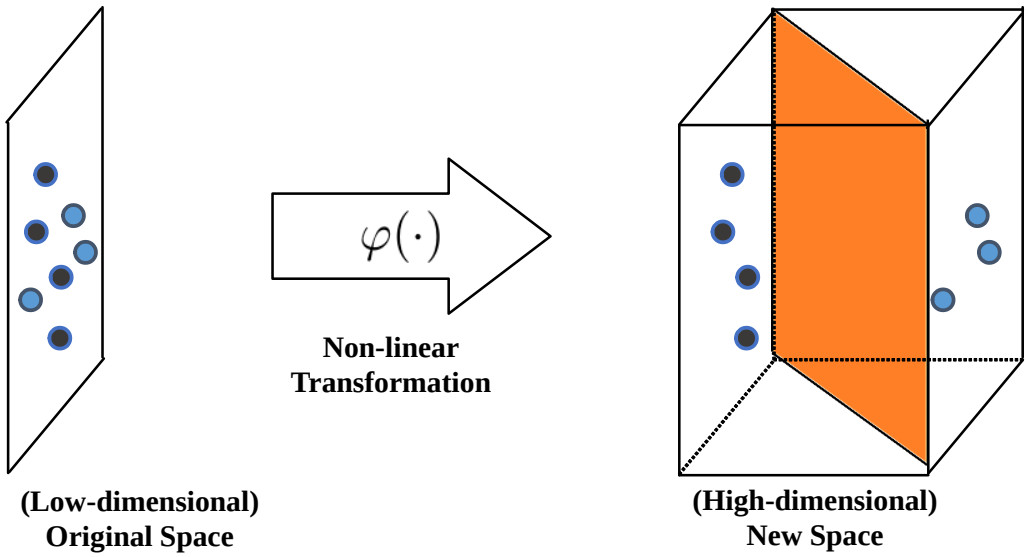


Figure 3: Cover's theorem.

$$n_{new} = \binom{n+2-1}{2} + n = \frac{(n+2-1)\,!}{2\,!(n-1)\,!} + n = \frac{(n^2+3n)}{2} \tag{6}$$

Since the dimension of the input feature space increases through feature generation, the accuracy of the whole classification can be improved. However, at the same time, the amount of computation needed in the classifier increases due to the increase in dimensionality. Furthermore, if the dimension of the feature space increases beyond a certain level, the classification accuracy would be rather reduced due to overfitting or the like. This phenomenon is called the "curse of dimensionality" [23]. This happens because as the dimension of the feature vector increases, the volume of the feature space increases exponentially, so the number of data samples required to effectively utilize the huge feature space also increases. However, there is a limit to collecting the necessary data samples in reality.

Since all generated input features do not have a positive effect on the classification performance, creating a feature is not itself a solution to the problem. For example, for an image with a size of 100×120, according to Eq. (6), 16290 input features can be created by adding a quadratic term, and some of these features are useful for classification, while others have little effect in solving the classification problems. Therefore, to obtain the optimal classification performance, it is necessary to generate only useful input features to construct a new input feature space of the appropriate dimensions.

Using the proposed method, the amount of discriminative information of individual features is quantitatively measured before using the original input features and the generated input features in the classification process. Then, based on the results of the measurement, only features with a large amount of discriminative information are selected to construct a new input feature vector, and the discriminant features that are to be used for classification are extracted using subspace discriminant analysis on the new input feature vectors $\mathbf{x}^{SFG}$.

The separability of the individual features is measured using the discriminant distance scale [21]. The distance between the different classes and the class can be defined as follows for a j-th component (feature) of $\mathbf{x}^{FG}$, where $\mathbf{x}^{FG} \in R^{n_{new}\times 1}$ is a data sample including newly generated input features.

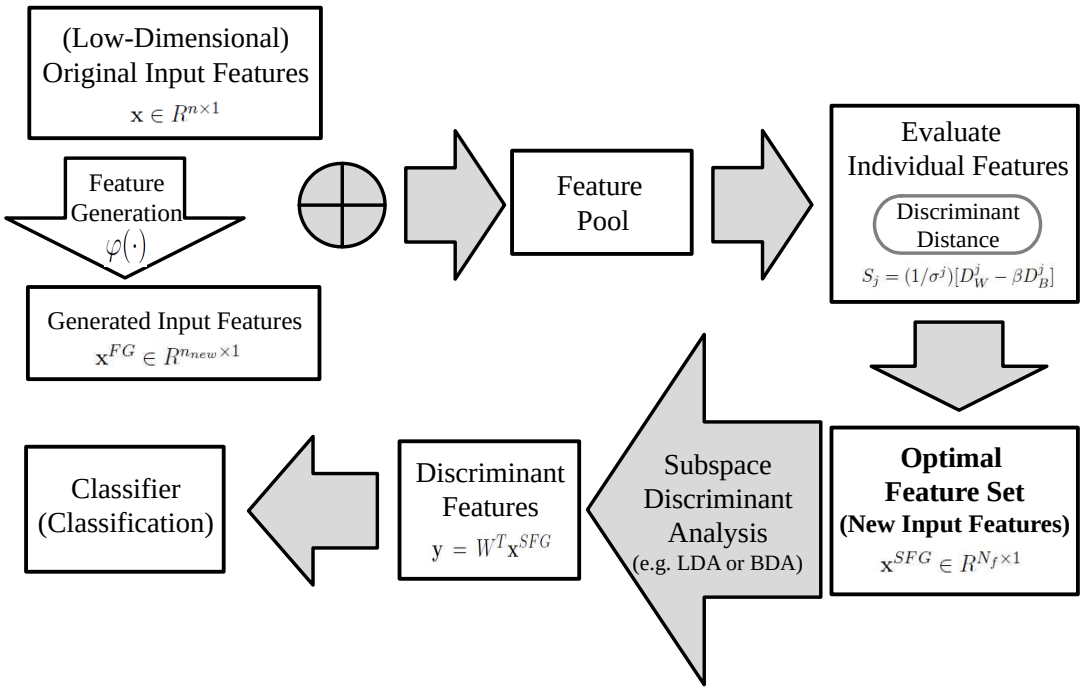


Figure 4: Overall procedure of the proposed method.

$$
\begin{aligned}
D_W^j &= \sum_{i=1}^{C} \frac{N_i}{N(N_i-1)} \sum_{x_{kj}^{FG}\in c_i} \left(x_{kj}^{FG} - \mu_i^j\right)^2 \\
D_B^j &= \sum_{i=1}^{C} \frac{N_i}{N} (\mu_i^j - \mu^j)^2
\end{aligned}
\tag{7}
$$

, where $\boldsymbol{\mu}_i^j$ and $\boldsymbol{\mu}^j$ are the j-th component of the mean of class c_i and all training data samples, respectively. The discriminant distance of the j-th feature from Eq. (7) can be defined as $(1/\sigma^j)[D_B^j - \beta D_W^j]$, $\sigma^j = (1/(N-1))\sum_{k=1}^{N}(x_{kj}^{FG} - m^j)^2$ [21], which can be used as a measure of the amount of discriminative information possessed by the j-th feature. β can be determined according to the distribution of the data samples as a user coefficient. In the case where the distribution within a class is large but the class separability is relatively good, it is preferable to reduce the value of β, which means a penalty of D_W^j. We set the value of β to 2 in this paper. Then, a measurement vector $\mathbf{S} = [S_1, S_1, .., S_{n_{new}}]^T$, $S_j = (1/\sigma^j)[D_W^j - \beta D_B^j]$ of the same size as n_{new} is defined and the new input feature vector $\mathbf{x}^{SFG}$ is constructed with the features corresponding to the large S_j. The entire process of the proposed method is shown in Fig. 4.

4 Experimental results and discussion

To show the effectiveness of the proposed method, we applied the proposed method to various real world problems. Through experiments on face databases and the UCI machine learning repository [3], we show that the proposed method works effectively for various kinds of low-dimensional data sets and for low-resolution images.

Table 1: Datasets from UCI machine learning repository used in the experiments

Dataset	**No. of classes**	**No. of instances**	**No. of original input f.**	**No. of new input f.** (for LDA/BDA)
Breast cancer	2	683	9	40/6
Pima	2	768	8	27/9
Bupa	2	345	6	20/15
Monk3	2	432	6	4/4
Balance	3	625	4	6/2
Wine	3	178	13	28/61
Glass	6	214	9	6/21
Car	4	1728	6	13/20

Table 2: Classification rates for UCI data sets

Feature extraction	**LDA**					**BDA**				
DatasetInput features	$\mathbf{x}^{ori}$	$\mathbf{x}^{IVS}$	$\mathbf{x}^{FG}$	$\mathbf{x}^{com}$	$\mathbf{x}^{SFG}$	$\mathbf{x}^{ori}$	$\mathbf{x}^{IVS}$	$\mathbf{x}^{FG}$	$\mathbf{x}^{com}$	$\mathbf{x}^{SFG}$
Breast.	95.9	96.0	95.7	96.5	96.0	95.1	95.8	95.3	96.8	95.8
Pima	68.9	69.1	69.8	68.6	**70.7**	69.3	70.0	69.5	68.7	**70.4**
Bupa	59.8	59.8	63.7	57.7	**64.1**	64.1	65.5	62.7	63.7	**64.8**
Monk3	87.4	100.0	91.2	99.6	**99.9**	68.6	100	68.4	99.4	**99.8**
Balance	87.7	87.7	94.1	88.9	**99.2**	84.3	84.3	85.3	96.3	**99.8**
Wine	98.7	98.7	96.4	98.7	98.6	98.0	98.8	99.3	98.6	**99.7**
Glass	61.8	71.2	64.5	71.7	71.5	71.2	77.6	70.0	72.3	70.6
Car	83.5	90.7	91.9	87.3	**94.9**	95.3	95.3	95.3	87.0	**95.5**
aver.	80.4	84.1	83.4	83.6	**86.8**	80.7	85.9	80.7	85.3	**87.0**

4.1 UCI Machine learning repository

We applied the propose method to several data sets in UCI machine learning repositories. Brief summaries of eight data sets that have been used in many other studies are given in Table 1. For each data set, we performed 10-fold cross validation 10 times and computed the average classification rate. Each input feature in the training set was normalized to have zero mean and unit variance, and the input features in the test set were also normalized using the means and variances of the training set. The one nearest neighbor rule was used as a classifier and the l_2 norm was used to measure the distance between two samples.

LDA and BDA were used to extract the discriminant features from the input feature vectors. LDA is a supervised learning method that is extensively used in data classification. In addition, as shown in Table 1, most of the data sets used in the experiments have binary classes, so we evaluated the classification performance using the BDA developed for one-class problems as well. We should find ways to extend BDA to multi-class problems in order to apply it to a few data sets having more than two classes, such as an iris data set, balance data set, glass data set and car data set. One of the simplest ways [20] to extend the BDA to D-class classification problems is to construct D data sets with only two classes (positive and negative). In constructing the i-th data set, the samples from the i-th class are regarded as positive samples, and the rest are regarded as negative samples. Then, we obtain D feature spaces by applying BDA to each of these data sets. During the test of a sample, a combined feature vector, which is concatenated with D resulting feature vectors from D feature spaces as in [19] is used with the classifier. The necessary parameters for CLDA and CBDA, i.e., the length of a composite vector and the number of composite features, were set to the values with which each classification method exhibited the best performance, as in [17].

Table 2 shows the classification performance using LDA and BDA for new input feature vectors obtained by applying various methods to input features. The values in the column corresponding to $\mathbf{x}^{ori}$ are the classification rates obtained by applying LDA or BDA to the original data. $\mathbf{x}^{IVS}$ are the data samples containing only some input features selected by the IVS method [8] among the original input features, and $\mathbf{x}^{FG}$ are data samples with quadratic terms added to original input features using Eq. (6). Columns corresponding to $\mathbf{x}^{Com}$ are the results of CLDA and CBDA using a composite vector, which is a subset of input features. For the last row, the average classification rate of nine data sets was reported for each method.

From the results in the table, the proposed method that selectively generated new input features ($\mathbf{x}^{SFG}$) provided the best classification performance in most data sets, showing that the average classification rates were 6.% and 6.3% higher in the LDA and BDA, respectively, than when using the original input features. The effects of the proposed method are prominent in the monk3 and balance data sets. In particular, for the balance data set, both the LDA and BDA classification results showed that when new features were selectively generated using the proposed method, the classification rate increased by more than 10% when using the original input features intact. The common characteristic of these two data sets is that the input features have fewer types of values. In both the monk3 and balance data sets, input feature values can only be four and five kinds of integers, respectively. In this case, when new features are generated using the proposed method, not only the dimension of the data but also the kinds of values that the input feature can have increases, so the data samples can be distributed more effectively in the feature space. On the other hand, in the case of the monk3 data set, LDA and BDA showed 87.4% and 68.6% of the original input features, respectively. However, when some input features were removed using the IVS method, both LDA and BDA showed 100%, respectively. This means that among the original input features, unnecessary features were included that would disturb the classification. As a result, the performance of $\mathbf{x}^{FG}$, including all quadratic terms generated by these unnecessary input features, increased slightly (in the case of LDA) or was even lower than the for $\mathbf{x}^{ori}$ (in the case of BDA). However, in the case of the selective feature generation using the proposed method ($\mathbf{x}^{SFG}$), the classification rate can be seen to have increased to nearly 100% because the unnecessary input features were effectively filtered.

4.2 Face database and preprocessing

We also applied the proposed method to a face recognition problem. The FERET, CMU-PIE database, Yale B, and Yale databases, which are used in the experiments, are widely used in face recognition research (Table 3, Fig. 2). In order to represent each database's degree of variation, we selected an image taken under normal conditions (no illumination and expression variations) for each subject as a reference image and computed the PSNR of the subject's other images. As shown in Table 3, the PSNR of the FERET database is higher compared to the other databases; thus, the images in the FERET database exhibited a relatively small variation.

For the FERET database, images for 792 subjects were used, and two images ('fa', 'fb') taken from the front of each person were used, that is, a total of 1584 images [7]. Of 792 participants, 200 images for 100 subjects were used as training images to evaluate the recognition performance, and the remaining images for 692 subjects were used as test images. For the test, the 'fa' image was used as a gallery image and the 'fb' image was used as the probe image.

Among the frontal pose images of the CMU-PIE database, the 'illum' category includes 21 images with different lighting conditions for a total of 68 subjects. In this experiment, we used 21 images for 65 subjects, that is, 1365 images in total, except for images of people who have some shooting defects or do not include all 21 kinds of illumination variations. We used three images ('27_06', '27_07', '27_08') for each subject, i.e., 195 total images that have a relatively

Table 3: Characteristics of each face database used for the experiments

Database	FERET	CMU-PIE	Yale B	Yale
No. of subjects	992	65	10	15
No. of images per subject	2	21	45	11
Illumination variation	none	large	large	small
Expression variation	small	none	none	large
Occlusion	none	none	none	glasses
No. training / test	200 / 1384	195 / 1170	70 / 380	10-fold CV
Degree of variations (avr.PSNR)	16.9	12.6	12.4	14.1

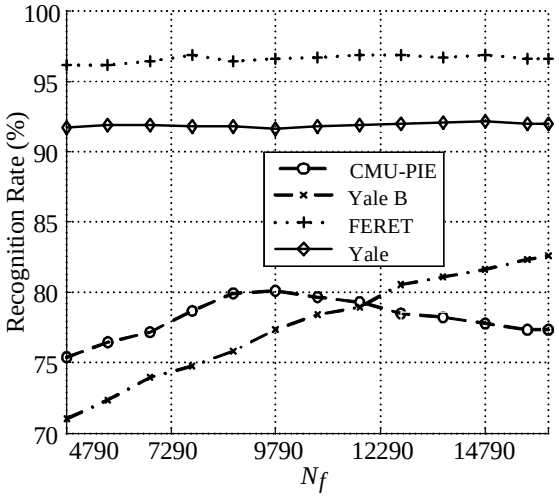


Figure 5: Recognition performance for various N_f.

small variation in illumination as training images, and the '27_20' image from the front lighting was used as a gallery image. The remaining images for each subject (total 65 pieces x17 = 1105 pieces) were used as proof images.

The Yale B database contains images for 10 subjects, and each subject's image consists of 45 kinds of images with illumination variations. The images are divided into subsets 1, 2, 3, and 4 according to the degree of variation in the illumination. In this experiment, the images for the subset 1 with less variation in illumination were used as training images and gallery images, and the images for remaining subset 2, 3 and 4 were used as probe images.

The Yale database contains 165 gray images of 15 subjects, with different facial expressions, with or without glasses, and under different illumination variations. In order to evaluate the recognition rates, we performed 10-fold cross validation 10 times and computed the average classification rate.

For face recognition experiments, facial images should be aligned to have the same size. For this, the whole face image is cropped based on the distance between the two eyes using manually detected eye coordinates and is then down scaled to a size of 120×100 [8], and the 60×50, 30×25, 24×20, and 15×12 images are downscaled versions of the 120×100 image again. All images were pre-processed for histogram equalization [13] and all pixels were normalized to have zero mean and unit standard deviation [7, 8]. The face recognition rates were evaluated from the 15×12 image (I_{180}), for which the recognition rate decreased sharply in Fig. 2, to I^{IP}_{12000} which is resized from the I_{180} to the 120x100 size via the bicubic interpolation [16], I^{FG}, to which the features

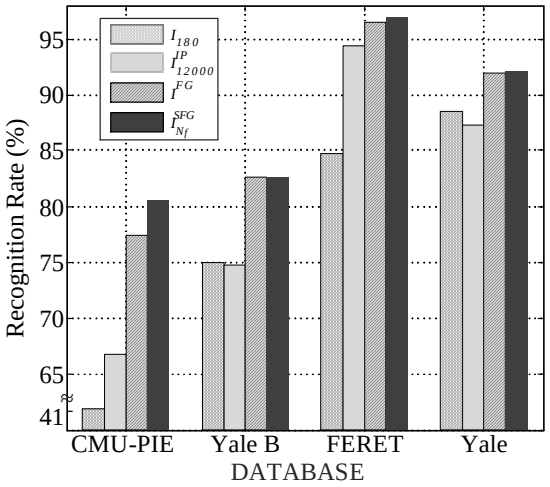


Figure 6: Comparison of recognition rates for I_{180}, I^{IP}_{12000}, I^{FG} and $I^{SFG}_{N_f}$(proposed method).

generated using Eq. (6) from I_{180} are added. $I^{SFG}_{N_f}$ (proposed) consists of the optimal features selected by the discriminant distance scale from 16,290 features of I^{FG}. The optimal number of features (N_f) is experimentally determined because it depends on the nature of the database [8]. As in Fig. 5, after we investigated the recognition rates by changing N_f, we set N_f to 12290, 10290, 16290, and 14290 for the FERET, CMU-PIE, Yale B, and Yale databases, respectively. Among the appearance-based face recognition methods, the DCV method was used for feature extraction and the Fisherface method was used only for I_{180} where the SSS problem occurred. The NN (Nearest Neighborhood) method was used as a classifier, and the Euclidean distance was used as the distance measurement.

Fig. 6 shows the recognition rates for I_{180}, I^{IP}_{12000}, I^{FG}, $I^{SFG}_{N_f}$ for various databases. Fig. 6 shows that the recognition rate for the CMU-PIE database and the FERET database was improved when the image size (i.e., the number of pixels) was increased using the interpolation method (I^{IP}_{12000}), but in the case of the Yale database, the recognition rate for I^{IP}_{12000} is less than I_{180} because the pixels (input features) generated via the interpolation method have brightness values estimated from the spatial relationship of adjacent pixels in the existing image, and thus the generated pixels do not help extract features using linear discriminant analysis. On the other hand, $I^{SFG}_{N_f}$, which is composed of the selected features by the discriminant distance scale among features generated in a non-linear way, showed a higher recognition rate than I_{180} for all databases.

Compared to I^{IP}_{12000}, the recognition rates of $I^{SFG}_{N_f}$ were significantly improved in the CMU-PIE, Yale B, and Yale databases than in the FERET database. The images of the FERET database, which have a relatively small variation compared to the CMU-PIE, Yale B and Yale databases, are less likely to suffer a loss of identity information due to image reduction. Since the images of the CMU-PIE, Yale B and Yale databases have already lost much of the identity information in the original image due to the variations such as in illumination and facial expressions, the reduced image (I_{180}) includes many pieces of face identification information as well as distortion information. Bicubic interpolation uses 16 adjacent pixels in I_{180} to determine the brightness value of a new pixel when expanded from I_{180} to I^{IP}_{12000}, so if any one of the 16 pixels contains distorted information (variation), the distortion is also reflected in the generated pixels. Consequently, in the case of the CMU-PIE, Yale B, and Yale databases, the improvement in the recognition rate through the use of I^{IP}_{12000} is not large or is rather worse than using I_{180}. In

contrast, the features generated by using the high order terms of the input features are relatively low in the distortion ratio of the identity information, and as a result, the recognition rate of I^{FG} is higher than I_{180} in all databases. In addition, even if distorted information is included in the generated features, all features are evaluated using the discriminant distance scale. Using only features with a high separability based on this ($I_{N_f}^{SFG}$), an additional improvement in the recognition rate can be obtained.

5 Conclusions

In pattern recognition problems, data for an object is represented vector composed of input features. The dimensions of data sample are determined by the attributes of the object samples are often stored as low-dimensional vectors according to the nature of the problem. Several discriminant feature extraction methods developed for data classification use statistical correlation of input features, but their performance is limited when the dimension of data is small or the range of values input features is small. Also, in the case of high-dimensional data such as image data, the image taken from a high-resolution camera converted into low-resolution image to reduce the calculation for data processing and effectively use the storage space. However, the performance may when a low-resolution image is used for recognition due to loss of information occur reducing the dimension of data. In this paper, we propose an input feature generation method effectively low-dimensional data to solve these problems. First, by generating high-order terms of the input features of the low-dimensional data samples, information on the correlation between the input features used as a new feature candidate group. Then, using the discriminant distance scale, new data samples were constructed with only input features with high separability by removing the features that are not helpful or obstructive to classification among the original input features and newly generated features. The experimental results on various low-dimensional data sets of UCI machine learning repository and several kinds of low-resolution facial images showed that the classification performance improved by selectively generating input features using the proposed method.

Acknowledgments

This work was supported by the Human Resources Program in Energy Technology of the Korea Institute of Energy Technology Evaluation and Planning (KETEP) granted financial resource from the Ministry of Trade, Industry and Energy, Republic of Korea (no. 20174030201740), and also supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2017-2015-0-00363) supervised by the IITP (Institute for Information and communications Technology Promotion).

Author's contributions

Conceived and designed the experiments: S.-I. Choi (SIC), H. Yoo (HY), S.T. Choi (STC). Performed the experiments: SIC, STC. Analyzed the data: SIC, HY, STC. Contributed reagents/-materials/analysis tools: SIC, HY. Wrote the paper: SIC, HY. Revised the manuscript critically for important intellectual content: SIC, HY, STC.

Bibliography

[1] Baker, S.; Sim, T.; Bsat, M. (2003); The CMU pose, illumination, and expression database, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, DOI: 10.1109/TPAMI.2003.1251154, 25(12), 1615-1618, 2003.

[2] Belhumeur, P. N.; Hespanha, J. P.; Kriegman, D. J. (1997); Eigenfaces vs. fisherfaces: Recognition using class specific linear projection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, DOI: 10.1109/34.598228, 19(7), 711-720, 1997.

[3] Blake, C.; Merz, C. J. (1998); UCI Repository of machine learning databases, https://www.nist.gov/, 1998.

[4] Cevikalp, H.; Neamtu, M.; Wilkes, M.; Barkana, A. (2005); Discriminative common vectors for face recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, DOI: 10.1109/TPAMI.2005.9, 27(1), 4-13, 2005.

[5] Chen, S.; Zhu, Y.; Zhang, D.; Yang, J.-Y. (2005); Feature extraction approaches based on matrix pattern: MatPCA and MatFLDA, *Pattern Recognition Letters*, DOI: 10.1016/j.patrec.2004.10.009, 26(8), 1157-1167, 2005.

[6] Choi, S.-I. (2015); Feature generation method for low-resolution face recognition, *Journal of Korea Multimedia Society*, 18(9):1039-1046, 2015.

[7] Choi, S.-I.; Choi, C.-H.; Jeong, G.-M.; Kwak, N. (2012); Pixel selection based on discriminant features with application to face recognition, *Pattern Recognition Letters*, DOI: 10.1016/j.patrec.2012.01.005, 33(9), 1083-1092, 2012.

[8] Choi, S.-I.; Oh, J.; Choi, C.-H.; Kim, C. (2012); Input variable selection for feature extraction in classification problems, *Signal Processing*, ISSN: 01651684, DOI: 10.1016/j.sigpro.2011.08.023, 92(3), 636-648, 2012.

[9] Cortes, C.; Vapnik, V. (1995); Support-vector networks, *Machine Learning*, DOI: 10.1023/A:1022627411411, 20(3), 273-297, 1995.

[10] Cover, T. M. (1965); Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition, *IEEE Transactions on Electronic Computers*, ISSN: 03677508, DOI: 10.1109/PGEC.1965.264137, (3):326-334, 1965.

[11] Duda, R. O.; Hart, P. E.; Stork, D. G. (2001); *Pattern classification. 2nd*, New York, 55, 2001.

[12] Georghiades, A. S.; Belhumeur, P. N.; Kriegman, D. J. (2001); From few to many: Illumination cone models for face recognition under variable lighting and pose, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, DOI: 10.1109/34.927464, 23(6), 643-660, 2001.

[13] Gonzalez, R.; Woods, R. (2002); *Digital image processing*, A. Dowrkin, Ed. Upper Saddle River, New Jersey 07458, Prentice Hall, 2002.

[14] Jain, A. K.; Duin, R. P. W.; Mao, J. (2000); Statistical pattern recognition: A review, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, I, DOI: 10.1109/34.824819, 22(1), 4-37, 2000.

[15] Keinosuke, F. (1990); *Introduction to statistical pattern recognition*, Academic Press Inc., 1990.

[16] Keys, R. (1981); Cubic convolution interpolation for digital image processing, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, I, DOI: 10.1109/TASSP.1981.1163711, 29(6), 1153-1160, 1981.

[17] Kim, C. (2007); *Pattern recognition using composite features*, Ph. D. Thesis, Seoul National University, 2007.

[18] Kim, C.; Choi, C.-H. (2007); A discriminant analysis using composite features for classification problems, *Pattern Recognition*, DOI: 10.1016/j.patcog.2007.02.008, 40(11), 2958-2966, 2007.

[19] Kim, C.; Oh, J. Y.; Choi, C.-H. (2005); Combined subspace method using global and local features for face recognition, *Neural Networks, 2005. IJCNN'05. Proceedings. 2005 IEEE International Joint Conference on*, DOI: 10.1109/IJCNN.2005.1556212, 4, 2030-2035, 2005.

[20] Kwak, N.; Oh, J. (2009); Feature extraction for one-class classification problems: Enhancements to biased discriminant analysis, *Pattern Recognition*, I DOI: 10.1016/j.patcog.2008.07.002, 42(1), 17-26, 2009.

[21] Liang, J.; Yang, S.; Winstanley, A. (2008); Invariant optimal feature selection: A distance discriminant and feature ranking based solution, *Pattern Recognition*, DOI: 10.1016/j.patcog.2007.10.018, 41(5), 1429-1439, 2008.

[22] Lin, F.; Zhou, X.; Zeng, W. (2016); Sparse online learning for collaborative filtering, *International Journal of Computers Communications & Control*, 11(2), 248-258, 2016.

[23] Marimont, R.; Shapiro, M. (1979); Nearest neighbour searches and the curse of dimensionality, *IMA Journal of Applied Mathematics*, DOI: 10.1093/imamat/24.1.59, 24(1), 59-70, 1979.

[24] Mika, S.; Ratsch, G.; Weston, J.; Scholkopf, B.; Mullers, K.-R. (1999); Fisher discriminant analysis with kernels, *Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop*, 41-48, 1999.

[25] Phillips, P. J.; Wechsler, H.; Huang, J.; Rauss, P. J. (1998); The FERET database and evaluation procedure for face-recognition algorithms, *Image and vision computing*, 16(5), 295-306, 1998.

[26] Suto, J.; Oniga, S.; Pop Sitar, P. (2016); Feature analysis to human activity recognition, *International Journal of Computers Communications & Control*, ISSN: 18419836, 12(1), 116-130, 20106.

[27] Turk, M.; Pentland, A. (1991); Eigenfaces for recognition, *Journal of Cognitive Neuroscience*, DOI: 10.1162/jocn.1991.3.1.71, 3(1), 71-86, 1991.

[28] Viriri, S.; Lagerwall, B. (2016); Increasing face recognition rates using novel classification algorithms, *International Journal of Computers Communications & Control*, 11(3), 381-393, 2016.

[29] Xiong, H.; Swamy, M.; Ahmad, M.O. (2005); Two-dimensional FLD for face recognition, *Pattern Recognition*, ISSN: 00313203, DOI: 10.1016/j.patcog.2004.12.003, 38(7), 1121-1124, 2005.

[30] Yang, J.; Frangi, A. F.; Yang, J.-y.; Zhang, D.; Jin, Z. (2005); KPCA plus LDA: a complete kernel fisher discriminant framework for feature extraction and recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, DOI: 10.1109/TPAMI.2005.33, 27(2), 230-244, 2005.

[31] Zhou, X. S.; Huang, T. S. (2001); Small sample learning during multimedia retrieval using biasmap, *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, DOI: 10.1109/CVPR.2001.990450, 1, 111-117, 2001.

[32] Center for Computational Vision and Control, Yale University, The Yale FaceDatabase, http://cvc.cs.yale.edu/cvc/projects/yalefaces/yalefaces.html.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 13(1), 39-49, February 2018.

The Integrated Environment for Learning Objects Design and Storing in Semantic Web

V. Dagiene, D. Gudoniene, R. Bartkute

Valentina Dagiene, Daina Gudoniene*
Institute of Mathematics and Informatics
Vilnius University
Akademijos 4, LT-08663 Vilnius, Lithuania
*Corresponding author: daina.gudoniene@mii.stud.vu.lt

Reda Bartkute
Informatics Faculty
Kaunas University of Technology
Studentu 50, LT-51367 Kaunas, Lithuania

Abstract: There is a variety of tools and environments for Learning Objects (LOs) design and delivery as well as learning object repositories (LOR) but the researchers could not find a repository that includes both functions: creation and storing of LOs. A number of different integrated learning systems are suggested for users that demonstrate the variety of e-learning methods and semantic capabilities. LO repository oer.ndma.lt/lor, that we are going to present, is very friendly and interoperable to use and assure LO design, search in semantic web, adaptation of the re-used objects and storing. There are no more existing LO repositories with the functionality presented by researchers. Transformation of closed education into open one without existence of well-structured, multifunctional and integrated environment becomes problematic. Authors will present an integrated environment for the LO design, search in semantic web, adaptation and storing of newly designed or re-designed LO. Measures will support the transformation of closed education into open and will assure effective design, re-usability and adaptation of LO in the integrated environment.
Keywords: learning objects (LO), models, semantic web, semantic technologies.

1 Introduction

The authors of the paper explore the existing learning objects repositories throw the open education and transformation of education into openness where it is a need to have an integrated environment for LO design, search and storing in the Semantic Web [4, 5]. Bellow presented different scientific papers and authors have analyzed technological challenges for LOR and IS implementation into learning design process. In this paper, Learning Objects (LOs) are referred as "small, modular, discrete units of learning design for electronic delivery and use" [27]. All Learning Objects can be identified by these features: interoperability, reusability, manageability, flexibility, accessibility, durability and scalability [22]. Analyzing Learning Objects from the technological perspective, they are based on the "paradigm of object orientation" [14] which means that the parts of a learning object can be used, changed or created repeatedly. This feature allows the user to create new and unique Learning Objects that can be used multiple times. In this process, the semantic web plays an important role.

The Semantic Web is "the extension of the World Wide Web that enables people to share content beyond the boundaries of applications and websites" [12]. In other words, the Semantic Web can be understood as a web of related meanings. The authors of the paper analyze the opportunities of the Semantic Web in the perspective of the search of a Learning object. In this

case, semantic web broadens the search and provides more and better fit with the search requests on the various aspects.

The context of semantic web, most of the practical implementations and use of standards are related to the marking of learning objects, which creates a lot of additional requirements for the successful use of standards [16]. When searching for ways to improve the courses using semantic web technologies, it is very important to develop simple methods and tools for LO labeling, separation of the objective and subjective metadata to create metadata sets and schemes from a variety of sources, to integrate production and labeling, to include formal semantics into existing standards and dynamically associate metadata with different LO. These are the challenges that are expected to be solved.

For sharing reusable learning objects, repositories are required so that these learning objects could be stored and delivered . The role and functions of learning objects' repository are described in the IMS digital storage compatibility specification [11].

A variety of learning object repositories are existing, however metadata studies show that the majority of metadata of currently existing Learning Objects Repositories (LORs) is only a general description of the content and settings [21]. Such data is difficult to use for program agents. Therefore, it is important to create semantic relations in the repositories so that learning objects would be fully integrated and linked [10]. A Semantic learning object repository is a system containing the educational resources and metadata (or metadata only) which provides search interface to people or other systems [2].

The analysis shows that the existing LORs and IS analysed and presented by the authors do not assure effective e-learning objects (ELO) design, search in semantic web and adaptation, as well as there is no suggested model, which will assure ELO adaptation in semantic web and automatically will integrate ELO to be re-used.

2 A research methodology

The research methodology was prepared in the frames of constructive method. The constructive research method takes 5 phases (see fig.1): (1) literature review; (2) problem identification; (3) theoretical framework; (4) practical implementation and (5) experiment. (Figure 1).

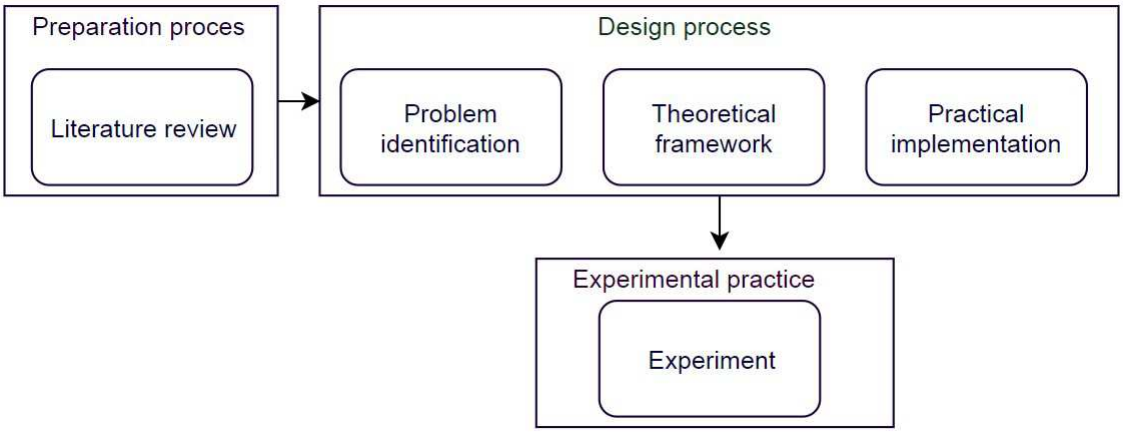


Figure 1: Constructive research method

A systematic review of related research works and analytical research methods were used for revealing the advantages of the use of Semantic Web technologies for integrated environment and for raising issues related to the semantic learning objects' use in semantic education as well

as for exploring existing LOs design approaches and models and for extracting initial data from our model linked to a theoretical framework.

The theoretical framework is designed to explain the integrations in the environment and the need of the new technological solutions after problem identification. Practical implementation phase evaluated during the experiment phase.

3 Related works on learning object repositories implementation

The comparative analysis of smart e-learning systems and architectures [26] [25] highlights the challenges of e-learning: (1) development of mobile agent's architecture in semantic web-based e-learning systems, while the agent knowledge is improved; (2) increased number of visualized educational resources; (3) new cooperation and communication with the learner to apply artificial intelligence achievements and the updating of internal and online databases, wikis etc.

3.1 Learning object repositories

Learning object repositories are operating online by providing the users a lot of different LOs by covering different educational levels and topics and also developed by variety of technologies and having different metadata described with the aim to classify LOs [19].

There are a lot of different well known learning object repositories like: CLOE; MIT Open Courseware (OCW); VCILT; CAREO; OLI; Commonwealth of Learning Object Repository; Edclicks; Encore; GEM; LOLA; MERLOT (specialized in microworlds); NDMA (www.oer.ndma.lt/lor/) repository for LO and different educational activities design and store.

Most of these LORs follow IEEE-LOM metadata standard and metadata annotation is done manually. However, the learning object repositories are used not only for storing but also for sharing, reusing and LO design [18].

Mohan and Brooks [23] discussed a situation where learning objects are embedded within learning systems. Today we have an analogous situation with mainframe systems. Learning systems today are essentially centralized, with learning objects (data) managed at a single place by a single system. However, with the growth of learning object repositories on the Semantic Web, it is necessary to find a model for the LO search, exchange and design where a learning object may also contain links to other courses or content packages, where it has been used before to support searches by software and human agents on the Semantic Web [17].

Goncales and Pimenta [13] analyzed the LO's integration into virtual learning environment and made a conclusion that in a dynamic environment of content sharing and educational practices the LOR has to integrate, thus creating dynamic learning environments to allow the interoperable user to retrieve them by searching through federated repositories, with the ability to modify those objects and compose lessons out of them.

The integration aspects were discussed by Shen, Ullrich and Borau [28] who outlined the importance of metadata in the LOR implementation process. LOR uses metadata to describe LOs and helps to improve the search for users. Learning Object Metadata lower barriers for repository growths as the server or learning environment does not contain the whole object but the metadata. This feature allows to have big Learning Object Repositories with less expenses. The model based on metadata of Learning Objects were implemented in proposed LOR as well.

3.2 Learning object design models

The researchers worked on the integrated environment for LO design and search in the semantic web.

Table 1: Comparative analysis of the LO models

Models	Features of the model
Verbert and Duval model	The model components are: content parts, learning objects, content objects. The model explains how the authors distinguish the fragments of content, content and learning objects (text, audio and video), presents individual resources [29].
Meyer model	LO components are designed on the basis of a clearly defined concept. LO components are well defined and focused on reusing. LO has one or more evaluation criteria and their usage areas includes several lessons [3].
Boyle model	Structural generative learning model (GLO) of an object contains inner and outer structures. GLO behavioral model consists of development tool, an XML file, and player program. As the main characteristic of the model, Boyle distinguishes the ability to change the flexibility of the XML file, using authoring tools to create instances of GLO layouts [8].
Santiago and Raabe generative model	Generative learning objects of higher levels (AGLO) come from the advanced options of heterogeneous meta-programming technology that allows to convey many aspects of learning (such as content, didactic, social and technological aspects) clearly through parameterization [7].
"Learnativity" content model	The model defines five levels of content hierarchy [29]: 1. The raw data and media items belong to the lowest level. 2. Information object includes raw data as well as media items and focuses on one piece of information. 3. On one task basis, information objects are connected to the third level - software objects. Learning objects are sets of information objects. 4. The fourth level - complex sets for larger (end) tasks. 5. Lessons or sections can be combined into larger sets.
NETg LO model	This model defines the rate as a matrix, divided into three main components: subjects (vertical), lessons (horizontal) and topics (fields) [1].
BNTOPM model	Widely used set of specifications for creating learning content independently from specific content delivery platform. BNTOPM includes content model compounds, consisting of shared content objects and containing compound [14].
"Navy" content model	"Navy" content model is an improved BNTOPM content model that offers more specific content definitions for detail levels that are necessary for the "Navy Interactive Learning Environment". The content of "Navy" is compatible with BNTOPM. "Navy" distinguishes LO compounds, final LOs, enabling LOs [29].
"Cisco" DNMO/DNIO model	Cisco classifies each DNIO. Possible content element classifications are: definition, example, overview, further steps, analogy, topology illustration, block scheme, additional resources, pie charts, teacher's notes, introduction, a key wording, illustration, importance, plan, facts, list, objectives, contrast, table, working scenario, conditions, guidelines, procedures' table, decisions' table, demonstration, table of prepared or combined [9].

Dynamic learning content management system model	The aim is to increase learning content reusability providing a modular design strategy together with a structured description [20]. The system includes a component model that defines three circuit levels: 1. The asset is media elements: pictures, video clips, animations, and simulation. 2. The content elements are defined as small, modular learning content units, which: (1) form the basis of learning content, (2) can be combined into larger, didactically based learning units (3) are independent, (4) are based on a single didactic content type (5) can be reused in different teaching contexts and (6) can be made of assets. Learning unit is defined as a set of elements.
ALOCOM model	It defines different levels of detail of contemporary content models as well as their mutual relations. It is based on OWL language, which uses ontologies for relating the content models [24].
Integrated modeling method - conceptual, educational and didactical model	It is a sustainable model developed to model the educational content. It consists of several models: conceptual, educational and didactic. Each of them defines specific learning object development aspects. The conceptual model is the basis for the field of knowledge, because it provides teaching concepts and determine their mutual relations [6].

The analysis shows that currently there is no model where LO functionality would be directly semantically tied to LO repository and LO development environment, where different learning objects are created.

4 The integrated environment for learning object's design and storing

Traditional teaching must be re-evaluated and adapted in a way to of open education. We cannot forget the educational aspects of the LO, which have a direct impact on a successful organization of learning process, and for efficient LO development. For the creation of such learning object, the educational model of LO planning, development and delivery is applicable. The essential functionality is related with LO design, delivery and search for additional material by using semantic web [15] i.e. search in other repositories of open educational resources for improving the designed LO.

Semantic network software agents can use contractual language service that allows agents to work together and actively introduce the learning material in the context of current problems. The aim to create links between learning objects and semantic network is very important in LO adaptation or improvement (Figure 2). The framework for the user profile is close related with the learning program and a different types of LO distributed in the integrated environment having also functions of learning object repository (LOR) (Figure 3).

The context is complex and dynamic, composed of a variety of problematic situations. According to the LOM specifications, context is the environment in which the LO may be created.

All contextual components are directly related to the tasks, content structure, available resources and contextual model, bringing together all these resources into the whole with the help of LO repository. The aim is to provide the necessary information to the consumer, and to carry out a search of electronic sources using keywords.

The new framework links semantic web technologies and learning objects: the technologies that facilitate search and re-use of learning objects in LO repositories are developed. The idea of search in the semantic web, is to find similar content LO not only in national but also internation-

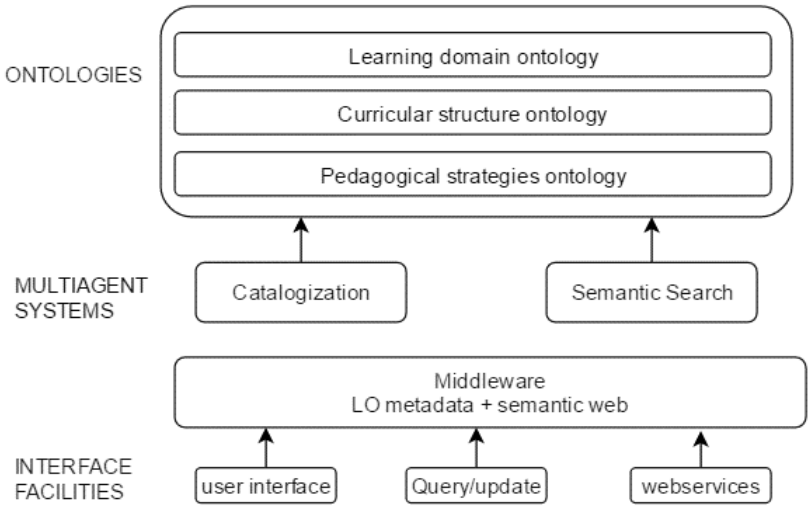


Figure 2: Integrated environment for learning objects design and storing in semantic web

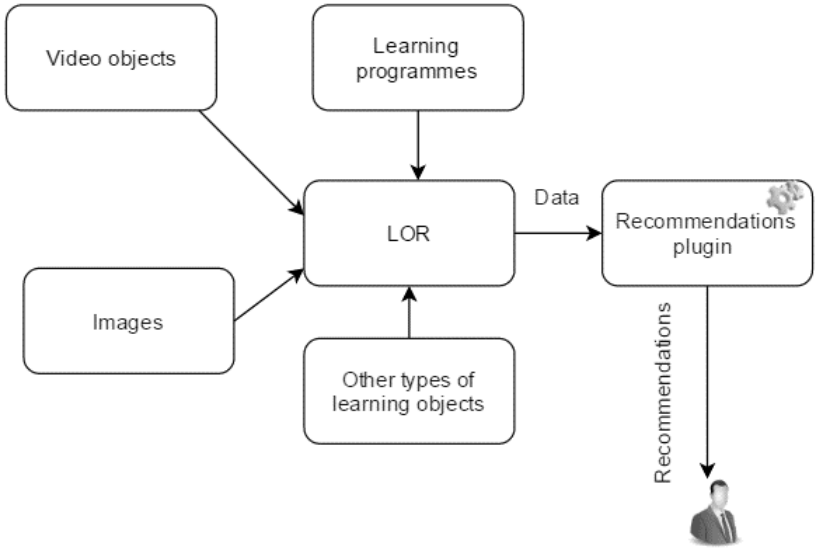


Figure 3: Framework for the user profile in the LOR

al/external learning object repositories, directing into storage that support certain educational topics (Figure 4).

LO search is performed from the user development environment and directly channeled into educational learning object repositories. Figure 4 shows the LO search algorithm in semantic web. Such a search cannot be carried out without using, i.e. each LO must be described and assigned to a specific storage area containing the various technologies and content objects (see Table 2).

For example, when modeling the LO repository, we distinguish the following entities: book - abstract publication in a repository; catalogue - systematic LO catalogue, in which the knowledge area includes a part of books or LO of the repository; teacher - LO developer, etc. The designed environment's functionality allows to users to develop different types of LO (see Figure 5) and

to upload it directly to the repository www.oer.ndma.lt.

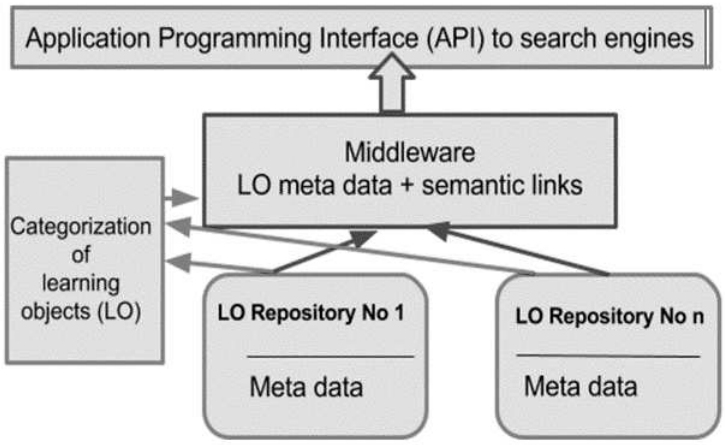


Figure 4: The framework of integrated environment for LO design and delivery

Table 2: Meta data for LO description

Type of file	Format
Text	(plain, richtext, html, xml)
Images	(bmp, gif, jpeg, png, tiff)
Video records	(avi, mpeg, quicktime)
Other formats	(pdf, doc, xls, ppt, java, x-shockwave-flash, zip, scorm)

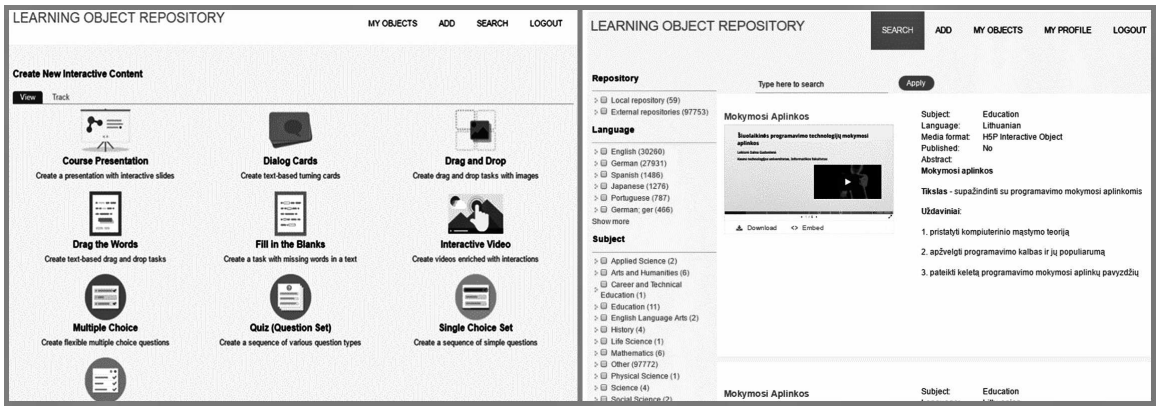


Figure 5: Learning object design tools in the environment (oer.ndma.lt/lor)

Search LO in the semantic web is assured by relations with external repositories of open educational resources, giving open access to it.

In a search in the semantic web the meaning of the words used in the query is considered in the search process what involves, for instance, the understanding of the intention of the user and the context of the search term, either on the newly designed LOR (Figure 5).

5 Experiment on effectiveness of the integrated environment for LO design and delivery

The distributed nature of the semantic web enables continuous improvement of learning materials. It enables the use of distributed knowledge provided in various forms, enabled by semantic annotation of content. For example, if user is in a video presentation system, he gets similar videos based on analysis of all data on systems together.

The created environment enables personalized LO development and delivery as well as its integration into different learning environments, regardless of their origin and type. This means that it assures minimal time consumption and gives an opportunity for teachers to create learning objects, complement them semantically with other educational content, and integrate them into personalized learning environments, integrated learning environments, repositories and teaching management environments. It allows improve educational module's fitness for the learner, and to increase learning efficiency and quality as well as the effectiveness of different learning environments.

The proposed framework and its practical application is useful for course developers, when designing and developing distance learning courses, massive open online courses, and other ICT-based content.

Table 3: Comparison of existing IS for LO design and the newly designed

Features of the systems/environments	Integrated environment	Virtual learning Environment	Information System
LOs form the basis of learning content	X	X	X
LOs can be combined into larger	X	X	
LOs are independent	X		X
LOs are based on a single didactic content type	X	X	X
LOs can be reused in different teaching contexts	X		X
LOs can be searched in the semantic web	X		X
LOs described in metadata	X	X	X
Learning domain ontology	X	X	
Curricular structure ontology	X	X	
Pedagogical strategies ontology	X	X	
Environment have categorization functions	X		X
Environment have a semantic search for other external LO	X		
Existing user interface	X	X	X
Existing Query / update	X	X	
Existing Web Services Facilities	X		X

The newly designed framework identifies user's need to develop and organize the learning

content. Content modeling activities can be carried out at different levels of abstraction - from coordination to instructional and educational level. LO models defined a formal framework, where learning objects can be modelled defining their formats, functions, participants and activity sequences.

High-resolution digital objects are stored in the repository (oer.ndma.lt/lor). However, using them for further development of the learning process and creating educational programs and courses, requires these objects to be transformed into different formats and to ensure their reusability in other content or courses.

Information systems and their components designed to work in an environment that is adapted to communicate with standard IT platforms, operating systems and computer networks.

6 Conclusions and future studies

The authors have proposed a friendly and interoperable integrated environment framework, which is able to use and assure LO design, search in semantic web, adaptation of the re-used objects and storing.

There are no more existing LO repositories with the functionality presented by researchers.

The integrated environment will contribute to the transformation of education into openness.

The integrated environment for the LO design, search in semantic web, adaptation and storing of newly designed or re-designed LO is identified as an effective to use.

The created environment enables personalized LO development and delivery as well as its integration into different learning environments, regardless of their origin and type.

The authors of the paper are planning to continue the research related with smart learning objects for smart education. Following the new strategies of the HE to be more active in MOOCs design and transformation of closed education into open one, the smart learning objects and integration with the wide functionality will serve for the new learning methods implementation and new smart learning objects design.

Bibliography

[1] Allen, C.; Mugisa, E. (2010); Improving Learning Object Reuse through OOD: A Theory of Learning Objects, *Journal of Object Technology*, 9(6), 51-75, 2010.

[2] Aroyo, L.; Dicheva, D. (2013); The New Challenges for E-learning, *The Educational Semantic Web Educational Technology & Society*, 7 (4), 59-69, 2013.

[3] Arreola, R. 1998; *Writing Learning Objectives*, The University of Tennessee, Memphis, 1998.

[4] Berners-Lee, T.; Hall, W.; Hendler, J.A. et al (2006); A Framework for Web Science, *Foundations and Trends® in Web Science*, 1(1), 1-130, 2006.

[5] Brasoveanu, A.M.P.; Dzitac I. (2012); The Role of Visual Rhetoric in Semantic Multimedia: Strategies for Decision Making in Times of Crisis, *International Journal of Computers Communications & Control*, 7(4), 605-615, 2012.

[6] Bridges, D.; Davison, R.; Odegard, P.S.; Maki, I.; Tomkowiak, J. (2011); Inter-professional collaboration: three best practice models of inter-professional education, *Medical Education Online*, DOI: 10.3402/meo.v16i0.6035, 16(10), 2011.

[7] Burbaite, R. (2014); *Advanced Generative Learning Objects in Informatics Education: The concept, models and Implementation*, Kaunas University of Technology, 2014.

[8] Cheng, H.M.; Yen, Y.N.; Chen, M.B.; Yang, W.B. (2010); A Process for Digitizing Historical Architecture, *Euro-Mediterranean Conference (EuroMed'10)*, 1-12, 2010.

[9] Cisco (2006); *Product Bulletin No. 1545*, Cisco, 1-12, 2006.

[10] Dhuria, S.; Chawla, S. (2014); Ontologies for Personalized E-Learning in the Semantic Web, *International Journal of Advanced Engineering and Nano Technology (IJAENT)*, 1(4), 13-18, 2014.

[11] Ermalai, I.; Mocofan, M.; Onita, M.; Vasiu, R. (2009); Adding semantics to online learning environments. Computational Intelligence and Informatics, *5th International Symposium on Applied Computional Intelligence and Informatics (SACI'09)*, 569-573, 2009.

[12] Gladun, A. (2009); An application of intelligent techniques and semantic web technologies in e-learning environments Expert Systems with Applications, *Expert Systems with Applications*, 36(2-1), 1922-1931, 2009.

[13] Goncalves, M.J.A.; Perez Cota, M.; Pimenta, P. (2013); A Study to determine what Kind of Learning Objects are Used in Higher Education Institutions, *Education*, 3(1), 30-36, 2013.

[14] Gutierrez, I.; Alvarez, V.; Paule, P.; Perez-Perez, J.R.; Freitas, S. (2016); Adaptation in E-Learning Content Specifications with Dynamic Sharable Objects, *Systems*, 4(2), 24, 2016.

[15] Jovanovic, J.; Gasevic, D.; Knight, C.; Richards, G. (2007); Ontologies for Effective Use of Context in e-Learning Settings, *Journal of Educational Technology & Society*, 10(3), 47-59, 2007.

[16] Klasnja-Milicevic, A.; Vesin, B.; Ivanovic, M.; Budimac, Z. (2010); E-Learning personalization based on hybrid recommendation strategy and learning style identification, *Computers & Education*, 56(3), 885-899, 2010.

[17] Kontopoulos, E. (2008); An ontology-based planning system for e-course generation, *Expert Systems with Applications*, 35, 398-406, 2008.

[18] Lupeikiene, A. (2007); *Theoretical and technological aspects of information systems: a textbook for doctoral and postgraduate students in computer science*, Institute for Mathematics and Informatics, Vilnius University, 200, 2007.

[19] Magnisalis, I.; Demetriadis, S.; Karakostas, A. (2011); Adaptive and Intelligent Systems for Collaborative Learning Support: A Review of the Field, *IEEE Transactions on Learning Technologies*, 4(1), 5-20, 2011.

[20] McGreal, R. (2004); Learning Objects: A Practical Definition, *International Journal of Instructional Technology and Distance Learning*, 1(9), 2004.

[21] McIlraith, S.; Son, T.; Zeng, H. (2001); Semantic Web Services, *Intelligent Systems*, 16(2), 46-53, 2001.

[22] Menolli, A., Reinehr, S., Malucelli, A. (2013); Improving Organizational Learning: Defining Units of Learning from Social Tools, *Informatics in Education*, 12(2), 273-290, 2013.

[23] Mohan, P.; Brooks, C. (2003); Learning Objects on the Semantic Web, *Proceedings of 3rd IEEE International Conference on Advanced Learning Technologies (ICALT03)*, 2003.

[24] Psyllidis, A. (2015); Ontology-Based Data Integration from Heterogenous Urban Systems: A Knowledge Representation Framework for Smart Cities, *Proceedings of the 14th International Conference on Computers in Urban Planning and Urban Management (CUPUM'14)*, 2015.

[25] Raju, P.; Ahmed, V. (2012); Enabling technologies for developing next-generation learning object repository for construction, *Automation in Construction*, 22, 247-257, 2012.

[26] Sakarkar, G.; Deshpande, S.; Thakare, V. (2012); Intelligent Online e-Learning Systems: A Comparative Study, *International Journal of Computer Applications*, 56(4), 21-25, 2012.

[27] Salas, K.; Ellis, L. (2006); The Development and Implementation of Learning Objects in a Higher Education Setting, *Interdisciplinary Journal of Knowledge and Learning Objects*, https://doi.org/10.28945/398, 2, 1-22, 2006.

[28] Shen, R.; Ullrich, C.; Borau, K. (2013); Learning from Learning Objects and their Repositories to Create Sustainable Educational App Environments, *IEEE 13th International Conference on Advanced Learning Technologies*, 285-287, 2013.

[29] Verberts, K.; Duval, E. (2008); ALOCOM: a generic content model for learning objects, *International Journal on Digital Libraries*, 9(1), 41-63, 2008.

[30] The Semantic Web (2012); *The Semantic Web*, 2012. [Online] www.semanticweb.org

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 13(1), 50-61, February 2018.

Text Classification Research with Attention-based Recurrent Neural Networks

C. Du, L. Huang

Changshun Du*, Lei Huang
School of Economics and Management
Beijing Jiaotong University
Beijing 100044, China
*Corresponding author: summer2015@bjtu.edu.cn

Abstract: Text classification is one of the principal tasks of machine learning. It aims to design proper algorithms to enable computers to extract features and classify texts automatically. In the past, this has been mainly based on the classification of keywords and neural network semantic synthesis classification. The former emphasizes the role of keywords, while the latter focuses on the combination of words between roles. The method proposed in this paper considers the advantages of both methods. It uses an attention mechanism to learn weighting for each word. Under the setting, key words will have a higher weight, and common words will have lower weight. Therefore, the representation of texts not only considers all words, but also pays more attention to key words. Then we feed the feature vector to a softmax classifier. At last, we conduct experiments on two news classification datasets published by NLPCC2014 and Reuters, respectively. The proposed model achieves F-values by 88.5% and 51.8% on the two datasets. The experimental results show that our method outperforms all the traditional baseline systems.
Keywords: machine learning, text classification, attention mechanism, bidirectional RNN, word vector.

1 Introduction

Text classification refers to the process of determining text categories based on the text content under a given classification system. On the Internet, major news sites, forums, blogs and so on are used as text for the main information subject and studying their automatic classification has a wide range of uses. In the field of journalism, press and publication need to be classified according to the columns in order to organize different news in different columns; intelligent recommendation system needs to be calibrated according to the user's different personality characteristics and preferences for the corresponding category of news; in mail processing tasks, the Mail system needs to be governed by the contents of the messages to determine whether the message is spam and decide whether to show to the user. Therefore, the main goal of this paper is to study the text automatic classification algorithm and meet the current mass of automatic classification of text requirements.

As early as the 1960s, people began to study text classification. At that time, it was artificial to write classification rules according to language phenomena and rules. By the 1990s, people began to study computer-based automatic classification technology. This method is first trained by pre-tagging data, learning discrimination rules or classifiers, and then starting to automatically classify new samples of unknown categories. The results show that in the context of large data volumes, its classification accuracy is much better than the expert definition of the rules. Therefore, the current research focuses on automatic text categorization of computer algorithms. Mingzhu Yao [16] et al. used the Latent Dirichlet Allocation (LDA) model to automatically classify text. The LDA model is expressed as a fixed probability distribution, the Gibss sampling

in Markov chain Monte Carlo (MCMC) is used to reason and the parameters of the model are calculated indirectly. The probability distribution of the text on the fixed subject is obtained, the large probability is for the text of the category. Aili Zhang et al. [18] used the support vector machine (SVM) algorithm for multi-class text classification. The method mainly uses the vector space model as a feature, which transforms the document into a high dimension sparse vector per the features of the text and then enters it into the SVM classifier. Liu Hua [4] uses the key phrases of the text to classify, and he thinks that the key words or phrases of the response text category information are more important, so the vector features of the key phrases are first extracted by statistical methods, and then the cosine similarity is calculated to determine the category. With the rise of in-depth learning methods in recent years, the limited Boltzmann machine has also been widely applied to text classification methods. Hilton et al. [10] used the depth Boltzmann machine simulation document to automatically learn the classification characteristics of the document, and in current document classifications has achieved good results. Note that except of document classification, deep learning has been also drawn attention in other pattern classification tasks, such as images [15] [6].

The above methods have achieved some success in text classification tasks, but they each have their own shortcomings. In the literature [4], it is pointed out that the categories of texts are usually associated with some key phrases and words, so they are modeled using the method of extracting keywords. These keywords are important, but other words that link these keywords together also contain a lot of information about the document, and the direct abandonment of these words can seriously damage the information that the document represents. In [10], the neural network is used to study the document, taking into account the interrelations and sequences of words, has the ability to extract text features automatically, and has the strongest performance on current classical data sets. However, the whole model does not take the role of key words into account, but rather treats all the words as a network of input, not giving the key words any special treatment. Therefore, we believe that if we can combine the advantages of the two methods, redesign the neural network model, and increase the weight keywords in the network, that the final text classification results should see a significant improvement. In order to verify this hypothesis, we designed the recursive neural network model to learn the representation of the text, and added the attention mechanism to the neural network [1]. The function of the attention mechanism is to learn a weight for each word of the input document, expecting key words to have a heavier weight, and the non-critical words to have a lighter weight, and the weight of the word reflects its contribution to the subject of the document. In the attention mechanism, the values of these weights are obtained through network learning, which is different from the previous subject model. Here, we will assign a vector to each category. The weight of the word is calculated according to the similarity between the word vector and the category vector. All the vectors in the model, including the word vector and the class vector, are learned through the optimization algorithm.

In the lab section, we collected data sets for news categories in Chinese and English, and experimented with different settings on both datasets. The Chinese data came from The 3rd Natural Language Processing and Chinese Computing Conference (NLPCC2014) public evaluation of news classification data, provided by the Xinhua News Agency and marked category tags into a total of 347 categories; English news data was selected from RCV1-v2 released by Reuters [10], which contains a total of 103 categories. We first use the text preprocessing technique to remove low frequency words and stop words, then use the recursive neural network with the attention mechanism to extract the feature vector of the article, and finally pass the feature vector to the softmax classifier. The experimental results show that the attention mechanism is very effective in assigning a higher weight to the key words, and can effectively improve the accuracy of the classification.

The method proposed in this paper makes full use of the advantages of the depth learning model which can employ self - learning characteristics, and embeds the traditional method of using keywords to classify the text in the neural network by way of the attention mechanism. The advantages of the two are organically combined in this paper. In Part 2, the structure of the model is described in detail. Part 3 gives the optimization objective function of the model and the parameter settings of the experiment. The fourth part shows the experimental results of the model on the above two datasets. Part 5 is the Model and experimental summary.

The development of workable assessment systems is difficult largely due to the fact that the value of assessment is often controversial:

2 Recurrent neural network model based on attention mechanism

The model consists of two parts, the first part is the feature extraction operation which mainly utilizes the recursive neural network to gradually synthesize the vector characteristics of the text. We first introduce the recurrent neural network model, and then describe it in detail on the basis of this model to increase the structure of the attention mechanism; the second part is the classifier, the classifier has a dropout [14] layer and softmax layer composition. The biggest advantage of this model is that only simple preprocessing of text is required, you can use the attention mechanism to select keywords and learn the text of the feature representation. The various parts of the model are described in detail below.

2.1 The representation of the word

In the text, we represent the word as a distributed word vector, and there are already many work studies in the vector field to learn word representations [11] [5] [8] [9] [7]. We use [8] the proposed language model to learn the representation of words. First, we collect an unsupervised text corpus and the New York Times corpus (NYT), pre-training the word vector with the Baidu Encyclopedia. [13] [12] [17] and other work points out that in a large-scale unsupervised corpus, the learning the word vectors can improve the effectiveness of the model, and the model can also provide a better initial value. In this paper, we use $\mathbf{E}$ to represent the matrix of word vectors, each of which represents a vector of words, and the dimension of the column vector is d . The k th word is expressed as a one-hot vector v_k (the k th position is 1 and the remaining position is 0), then the vector of the k th word can be denoted as $\mathbf{E}v_k$.

2.2 The input layer of the network

In the past, the input of the network was the word vector itself. Here, we use this input method as a stepping stone to improve. We believe that the word definition cannot be determined by only the word itself, but also should look at the specific environment where the word is placed, and furthermore see how it works in different environments, where the meaning of the same word differs. This feature is especially important for Chinese context. Therefore, we get the word's pre-training initial value, and the word in each text section, we use it before and after each word as background content, to calculate the given word as the center of the window for the average of the word vector 3, As a vector representation of the current word. Here is an example.

</s> Shanghai Forest Coverage Year by Year Increase </s>

In this sentence, the word "coverage" of the vector is calculated as:

$$V'_{Coverage} = \frac{V_{Forest} + V_{Coverage} + V_{Year}}{3},$$

For the words in the beginning and ending of a text, we use the symbol </s> to fill it, which is also given a vector representation in the model, so the beginning and ending words can also take a similar calculation. It is worth noting that this method of calculation constitutes the input layer of the model, which is included in the objective function expression of the optimization model, not just the initial calculation.

2.3 Recurrent neural network (RNN)

The RNN (Recurrent Neural Network) model has demonstrated a strong learning ability in many natural language processing tasks. It is characterized by good modeling of sequence data and full utilization of sequence information. Since the RNN is to semantically synthesize each word in the text in turn, the RNN can adapt to the variable sentence, that is, the uniformity of the text length is not required, and both long and short texts can be learned.Fig. 1 shows a traditional recurrent neural network structure.

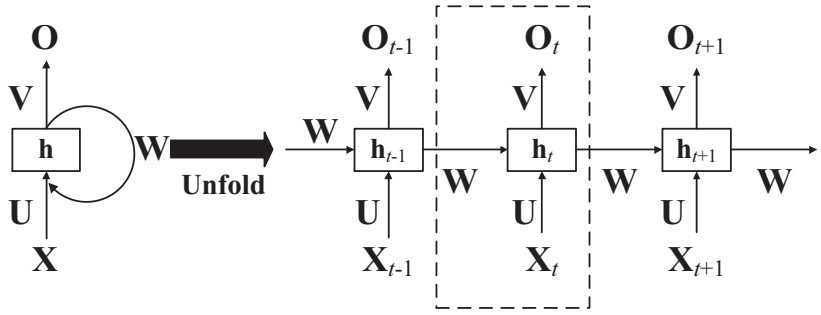


Figure 1: Traditional structure of RNN

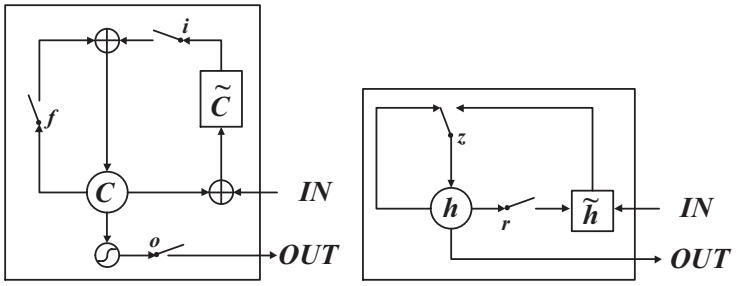
In Fig.1, $\mathbf{x}_t$ is the input unit of step t , which represents the word vector of the t th word in the text; $\mathbf{h}_t$ is the hidden state of step t ; $\mathbf{o}_t$ represents the output of step t , the output of this step is a softmax classifier, the output is selected according to the needs of the model; $\mathbf{U}$,$\mathbf{V}$, and $\mathbf{W}$ are the weight parameters of the network that all need to be learned in the model. As shown in Fig. 1, the dotted line box is the calculation of the t th unit, as follows:

$$\begin{cases} \mathbf{h}_t = f(\mathbf{W}\mathbf{h}_{t-1} + \mathbf{U}\mathbf{x}_t + \mathbf{b}_h) \\ \mathbf{o}_t = softmax(\mathbf{V}\mathbf{h}_t + \mathbf{b}_0) \end{cases} \tag{1}$$

Where variables $\mathbf{b}_h$ and $\mathbf{b}_0$ represent biased terms. As can be seen from Eq.1, each hidden state of the recursive neural network is determined by the current input word and the hidden state of the previous step. If you do not need to add a classifier to each synthetic step in a particular task, $\mathbf{o}_t$ can not be output. The disadvantage of the traditional recurrent neural network is that with the increase of the length of the text, the number of layers of the network is gradually deepened. The loss of the network in the process of information synthesis is relatively large, which tends to focus on the learning of the final stage of memory. Therefore, the efficiency of long text learning is not good.

In this paper, Long Short-term Memory (LSTM) [3] and Gated Recurrent Unit (GRU) [2] are used because of the shortcomings of traditional RNN in dealing with long text. The advantage of the LSTM and GRU nodes is that they can be set up in the process of synthesizing to control how much information should be received in the current synthesis step, how much is forgotten, and how much information is passed back. Through these gate controls, RNN has a proficient learning ability for long text. The difference between LSTM and GRU is that LSTM has more parameters. The GRU has fewer parameters and thus has a faster calculation speed.

LSTM and GRU are two kinds of calculate nodes of RNN. In the method of calculating hidden state, it is different from traditional method, and it is consistent with RNN structure of main body. The LSTM and GRU nodes are calculated as shown in Fig.2.



(a) LSTM structure (b) GRU structure

Figure 2: Structures of LSTM and GRU unit

LSTM node:

$$\begin{cases} \mathbf{i} = \sigma(\mathbf{U}^i\mathbf{x}_t + \mathbf{W}^i\mathbf{h}_{t-1}) \\ \mathbf{f} = \sigma(\mathbf{U}^f\mathbf{x}_t + \mathbf{W}^f\mathbf{h}_{t-1}) \\ \mathbf{o} = \sigma(\mathbf{U}^o\mathbf{x}_t + \mathbf{W}^o\mathbf{h}_{t-1}) \\ \mathbf{g} = tanh(\mathbf{U}^g\mathbf{x}_t + \mathbf{W}^g\mathbf{h}_{t-1}) \\ \mathbf{c}_t = \mathbf{c}_{t-1}o\mathbf{f} + \mathbf{g}o\mathbf{i} \\ \mathbf{h}_t = tanh(\mathbf{c}_t)o\mathbf{o} \end{cases} \tag{2}$$

GRU node:

$$\begin{cases} \mathbf{z} = \sigma(\mathbf{U}^z\mathbf{x}_t + \mathbf{W}^z\mathbf{h}_{t-1}) \\ \mathbf{r} = \sigma(\mathbf{U}^r\mathbf{x}_t + \mathbf{W}^r\mathbf{h}_{t-1}) \\ s = tanh(\mathbf{U}^s\mathbf{x}_t + \mathbf{W}^h(\mathbf{h}_{t-1}o\mathbf{r})) \\ \mathbf{h}_t = (\mathbf{1\text{-}z})o\mathbf{s} + \mathbf{z}o\mathbf{h}_{t-1} \end{cases} \tag{3}$$

σ represents the sigmoid function, and the symbol o represents the operation of the corresponding vector element multiplication. In the LSTM node, $\mathbf{i}$,$\mathbf{f}$, $\mathbf{o}$ represent the input gate, the memory gate, and the output gate respectively, which control the proportion of the information throughput; $\mathbf{g}$ is the hidden state of the candidate, similar to the way the traditional RNN calculates the hidden state; $\mathbf{c}_t$ is the internal memory, by the $t-1$ step of the memory $\mathbf{c}_{t-1}$ and $\mathbf{g}$ through the memory gate and input gate weight to form; $\mathbf{h}_t$ is the true output state, which is the amount of information that the internal memory $\mathbf{c}_t$ outputs at the output gate. In the GRU node, $\mathbf{z}$ is the update gate, $\mathbf{r}$ is the reset gate, s is the hidden state of the current candidate. It can be seen by $\mathbf{s}$ calculation that the reset node controls the amount of the previous node information $\mathbf{h}_{t-1}$, and the final output state $\mathbf{h}_t$ is weighted by the current candidate's hidden state s and the previous node output state $\mathbf{h}_{t-1}$ by updating the gate $\mathbf{z}$.

2.4 Bidirectional RNN model of attention mechanism

In this paper, we use the bidirectional RNN to learn the characteristics of the text, because the meaning of a word is not only related to the text content in front of it, but also related to the text content behind it. We use the bidirectional RNN method to implement the text representated from the learning, and then the two directions to learn the feature vector spliced together, this as a text vector, so that relative to the unidirectional RNN, the eigenvector of the semantics is more comprehensive and rich. At the same time, we add a mechanism of attention to the network model, for each word to learn a weight, making the key words have a heavier

weight, and non-key words have a lighter weight, which can make important features become more prominent. Fig.3 shows the overall architecture of the model.

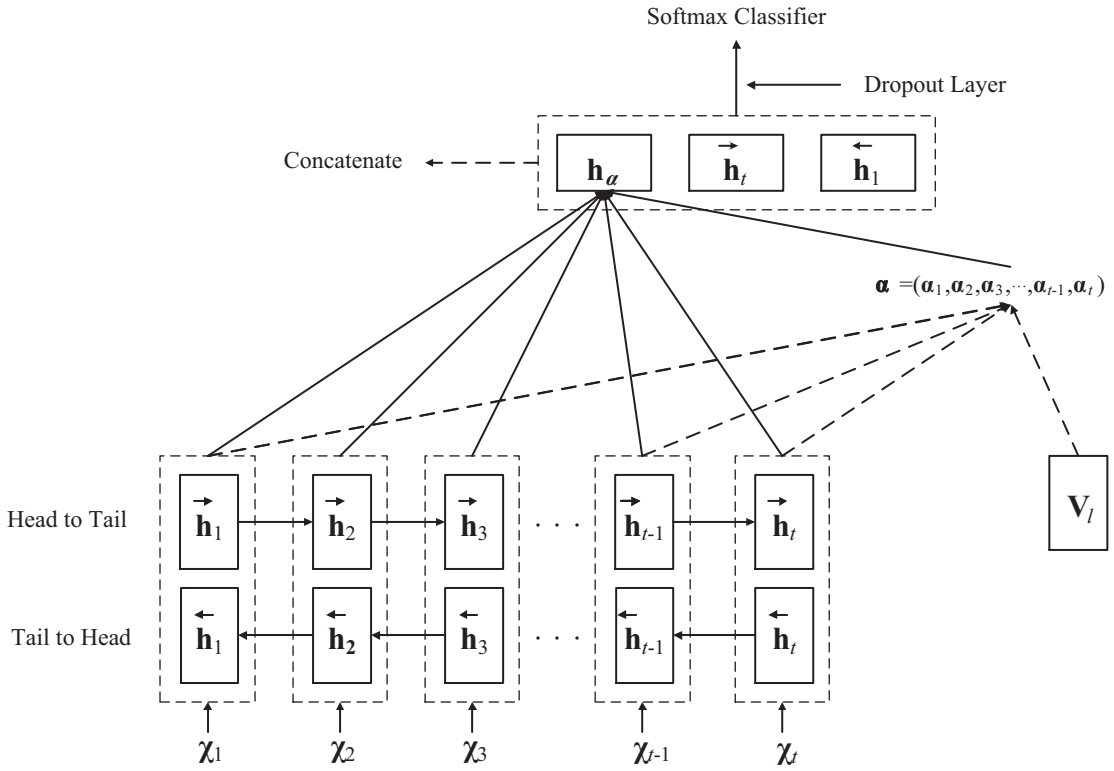


Figure 3: Attention-based bidirectional RNN structure

As shown in Fig.3, the input sentence is x_1 , x_2 , x_3 , L , x_{t-1} , x_t. The recursive neural network RNN has both forward and backward directions, and in the forward RNN, the hidden state is $\overrightarrow{\mathbf{h}_1}, \overrightarrow{\mathbf{h}_2}, \overrightarrow{\mathbf{h}_3}, L, \overrightarrow{\mathbf{h}_{t-1}}, \overrightarrow{\mathbf{h}_t}$ (" $\leftarrow$ " indicates that the direction of the RNN is forward); In this setting, the hidden state of the word x_i corresponds to $\mathbf{h}_i = \left[\overrightarrow{\mathbf{h}_i}; \overleftarrow{\mathbf{h}_i}\right]$ $(i = 1, 2, L, t)$, that is, the hidden state of the two directions together, such as the original hidden state of the k dimensional vector, then the $2k$ vector after the stitching.

In the traditional bidirectional recurrent neural network, the vector of $\overrightarrow{\mathbf{h}_t}$ and $\overleftarrow{\mathbf{h}_1}$ is usually concatenated as a text representation. Since the words that reflect the subject in a text are primarily a few keywords, the importance of each word is different. Here we introduce the attention mechanism. We calculate the corresponding weights for each word by the attention mechanism, and then weighted the sum of the hidden states of all the words according to the weight, and the result of the summation $\mathbf{h}_a$ is also taken as part of the textual feature.

As shown in Fig.3, $\mathbf{v}_l$ represents the category vector, $\alpha_i (i = 1, 2, L, t)$ represents the similarity between the hidden state of the i word and the category vector, that is, the weight of the i word, the similarity formula is [1]

$$\alpha_i = \frac{e^{\mathbf{h}_i^T \mathbf{M} \mathbf{v}_l}}{\sum_{j=1}^{t} e^{\mathbf{h}_i^T \mathbf{M} \mathbf{v}_l}} \tag{4}$$

And $i = 1, 2, L, t$. S is a parameter matrix that calculate generalized similarity, When it is a unit array, the similarity of $\mathbf{h}_i$ and $\mathbf{v}_l$ is degenerated into the inner product. $\alpha = (\alpha_1, \alpha_2, L, \alpha_t)$ represents the weight vector. According to Eq.4 we can see that it has been normalized. The resulting weighted feature vector $\mathbf{h}_a$ is

$$\mathbf{h}_a = \sum_{i=1}^{t} \alpha_i \mathbf{h}_i \tag{5}$$

And then $\mathbf{h}_a$ and RNN forward and backward results together, that is the character representation of the text.

$$\mathbf{s} = \left[\mathbf{h}_a, \overrightarrow{\mathbf{h}_t}, \overleftarrow{\mathbf{h}_1}\right]$$

2.5 Softmax classifier

As shown in Fig.3, after extracting the features of the text, the feature vectors are entered into the softmax classifier for classification. Here we use dropout [14] to connect the feature vector with the softmax classifier. To illustrate the dropout method, we consider the eigenvector as an input to the classifier. The traditional neural network connection method is comprised of the whole connection mode, the dropout algorithm is connected to the random input data(the feature vector after splicing in this paper) according to a certain proportion of 0, and the only other elements that are not set to 0 are ones participating in the operation and connection. For the sake of convenience, suppose that a learning sample is updated once, and the specific process is as follows: First, the input vector is placed in proportion to a portion of the elements, and the element not set to 0 is involved in the operation and optimization of the classifier; then the second sample enters the vector, this time in accordance with the random set way to select the elements of training, until all the samples have been studied once. Since for each input of a sample, the way to set 0 is randomized, each network update will have differing weight parameters. In the final prediction process, the entire network parameters are multiplied by $1 - \rho$ to get the final classifier network parameters. Because the parameters of each update are not the same, the dropout algorithm can be seen as the neural network being cast into a combination of multiple models, and can effectively prevent over-fitting and improve the model's prediction rate [14]. According to the literature [14], the dropout algorithm is similar to evolution, and the genes of the offspring are made up of half of the genes of the parents. This combination has a tendency to produce more vigorous genes. Similarly, in the final network of the dropout algorithm the parameter is a combination of the parameters of multiple models, which is a process of choosing and keeping the advantages over the shortcomings, and thus yielding a better generalization ability.

Suppose that the vector obtained by a bi-directional recurrent neural network is $\mathbf{c}$. The way the dropout algorithm sets its element to 0 can be represented by the Bernoulli distribution. The Bernoulli distribution is used to produce a binary vector $\mathbf{r}$ (it only contains 0 or 1) equal to the latitude of $\mathbf{c}$:

$$\mathbf{r} \sim \mathbf{Bernoulli}\,(\rho)$$

he vector entered into the softmax classifier is recorded as:

$$\mathbf{c}_d = \mathbf{c} \cdot \mathbf{r}$$

Where softmax classifier's network parameter is $\mathbf{W}_c$ and the offset term is $\mathbf{b}_c$, the output of the network is :

$$\mathbf{o} = f\left(\mathbf{W}_c \mathbf{c}_d + \mathbf{b}_c\right)$$

f is a sigmoid function or a tanh function. The probability that the current text belongs to category i is:

$$p\left(i|S\right) = e^{o_i} / \sum_{j=1}^{N} e^{o_j}$$

o_i represents the i th element of the vector $\mathbf{o}$, and N represents the number of classes.

2.6 Objective function

In this paper, we mainly study the classification problem. The parameters that need to be optimized include the following parts: word vectors, parameters of bi-directional recurrent neural networks, generalized similarity matrices of attention mechanisms, class vectors and classifier parameters. A word vector is denoted by $\mathbf{E}$, and the parameters of the bi-directional RNN are denoted by $\hat{\mathbf{W}}$ and $\hat{\mathbf{U}}$; the parameters of the attention mechanism layers are denoted as $\mathbf{M}$, the vector of all the categories is represented by the matrix $\mathbf{V}$, the i line vector $\mathbf{v}_i$ represents the i th category; the parameters of the classifier are denoted by $\mathbf{W}_c$. The sample set of the training set is $\Omega = \left\{\left(T_1, y_1\right), \left(T_2, y_2\right), L, \left(T_{|\Omega|}, y_{|\Omega|}\right)\right\}$, where T_i is the i th text, y_i is its category label, and $|\Omega|$ is the number of training set samples. $\theta = \left\{\mathbf{E}, \hat{\mathbf{W}}, \hat{\mathbf{U}}, \mathbf{M}, \mathbf{V}, \mathbf{W}_c\right\}, p\left(y_i|T_i, \theta\right)$ represents the probability that the category of the text T_i is divided into y_i when the parameter θ is known, so the optimized objective function is:

$$L = \sum_{i=1}^{|\Omega|} logp\left(y_i|T_i, \theta\right) + \lambda \left\|\theta\right\|_2^2$$

λ is the parameter of the regular term. In the actual experiment, we use the random gradient descent method to optimize θ update method is:

$$\theta = \theta - \alpha \frac{\partial L}{\partial \theta},$$

α is the learning rate.

3 Experiments and results

3.1 Experimental data

Two data sets are used in the experiment. The first is the Chinese data set, it is the 2014 Chinese Computer Society organized by the natural language processing conference published by the news classification evaluation data set. It is responsible for organizing and annotating Xinhua, which is a large-scale news classification corpus. It can be downloaded directly from the official website of NLPCC2014. The corpus training set has a size of 30,000 news articles, the test set contains 11,577 articles, and it's test set and training set have good consistency in the distribution of each category. The data set has two categories, the first layer comprised of 24 categories, and the second layer with a total of 367 categories. In this paper, the category of text is unified as a single-level category, and for a multi-level hierarchy of trees, we consider the final small category as a single category and report the classification of the final category. The second data set is an English data set and is an REV1-v2 dataset published by Reuters. This data set contains 804,414 news articles, comprising a total of 103 topics (103 categories, here is the hierarchical classification, processing with the previous data set). Following the literature [10], we randomly divided the data set into a training set and test set, of which the training set contains

794,414 news articles, and the test set contains 1,000 news articles. All of the experiments in this paper are performed on both datasets.

3.2 Data preprocessing

For the Chinese data set, we first use the Chinese word segmentation package NLPIR developed by the Chinese Academy of Sciences for Chinese word segmentation. The functions of NLPIR include Chinese word segmentation, partnered annotation, named entity recognition, user dictionary function, support of a variety of Chinese coding formats, and the ability to discover new words as well as facilitate keyword extraction. As the experiment in this article has a Chinese data set, we need to call the software packet word. English data itself is a separate word, so the operation of word segmentation is unnecessary. After the word segmentation operation is completed, the word frequencies of the two data sets are calculated, and the low frequency words and stop words are deleted. Because these words are not helpful in judging the subject, it may be helpful to use them when classifying them, which is not conducive to the prediction of the classification model.

Since we used the minibatch training model during the training process, we needed to perform a fixed-length operation on the length of the text. Since the sentence lengths of the natural language text are inconsistent, we first calculate the longest sentence length l_{max} . For any sentence length less than l_{max} , the uniform use of the text </s> symbol to l_{max} (</s> vector is always set to $\mathbf{0}$). The purpose of unifying the text length is to improve the efficiency of computing. When the length of the data is unified, you can use a matrix calculation, which when compared with circular computing is time-saving.

3.3 Pre - training of word vector

Before the model training, we need to pre-train the training vector on an unregulated large-scale corpus. A word vector is a distributed representation of a word that expresses an input suitable for a neural network. Many of the current studies have shown that word vectors without oversight learning in a large corpus are more conducive to the convergence of the neural network model leading to a good local optimal solution. In this paper, we use the Skip-gram model to pre-train the training vector. The vector of this model has a strong performance in many natural language processing tasks. The Skip-gram algorithm has been integrated in the word2vec package which we use directly to train Chinese and English word vectors. We use the text content read on the Baidu Encyclopedia to carry out the pre-training of the Chinese word vector, we pre-train the English word vector with the New York Times corpus.

3.4 The setting of the experimental parameters

In this paper, the model has the following super parameters: the dimension of the vector d , the dimension of the class vector, the dimension n of the hidden state in the recurrent neural network, the ratio ρ of the dropout algorithm, and the learning rate α of the SGD optimization algorithm. We use the grid search method to determine these parameters. The dimension of the word vector d is taken in $\{50, 100, 200, 300\}$; The dimension of the class vector l is in $\{50, 100\}$, The dimension n of the hidden state is taken in $\{500, 1000, 2000\}$; According to experience, the dropout algorithm ratio P is 0.6; the SGD algorithm learning rate α is in $\{1, 0.1, 0.01, 0.001\}$. For the Xinhua Newsroom data set, the best parameter value is: $d = 100, l = 100, n = 1000, \alpha = 0.05$. For Reuters RCV1-v2 datasets, the best parameter values are: $d = 300, l = 100, n = 1000, \alpha = 0.01$. The range of these parameters is based on experience, generally within the scope of the value can be achieve better experimental results. In

this experiment, we use these parameters for multiple experiments, and then obtain the average of the results.

3.5 Data experiment and comparative analysis

In this paper, we design a bi-directional recurrent neural network based on an attention mechanism to deal with the problem of text classification. The attention mechanism can be used to learn a weight for each word in the text based on the information of the category, where words closely related to the category receive a relatively heavy weighting, and words that are relatively weak in relation to the category receive lighter weighting. In the experiment, we vectorize 20,000 words of high frequency occurrence in Chinese, and vectorize 100,000 words of high frequency occurrence in an English data set. Tab.1 lists some of the baseline models and the results presented in this paper.

Table 1: Test results of document classification Chinese and English data sets

Model	**Accuracy**		**Recall**		**F-Value**	
	Xinhua News Agency	Reuters corpus RCV1-v2	Xinhua News Agency	Reuters corpus RCV1-v2	Xinhua News Agency	Reuters corpus RCV1-v2
TF-IDF+SVM	72.1	31.8	88.7	45.8	79.5	37.5
AveVec+SVM	70.8	29.3	92.3	33.2	80.1	31.1
TRNN	74.4	40.5	90.7	51.7	81.7	45.4
LDA	77.3	35.1	93.3	44.3	84.5	39.2
DocNADE	76.5	41.7	84.5	42.5	80.3	42.1
Replicated Softmax	82.3	42.1	94.0	47.2	87.8	44.5
Over-Rep. Softmax	82.7	45.3	89.3	51.4	85.9	48.2
Bi-TRNN	82.4	44.8	91.1	52.5	86.5	48.3
LSTM Bi-RNN	81.9	46.2	92.8	55.8	87.0	50.6
GRU Bi-RNN	83.3	45.8	93.9	51.9	88.3	48.7
Attention LSTM Bi-RNN	81.7	46.4	92.8	56.1	87.0	51.8
AttentionGRU Bi-RNN	83.9	46.0	93.5	52.8	88.5	49.2

In order to verify the validity of the model, we compared it with the methods of some baseline systems. The first comparison method is to calculate the TF-IDF feature of the text, form a set of vectors, and then use the support vector machine (SVM) to classify the eigenvectors; The second method is to use the average of the text word vector (the text is preprocessed to calculate the mean of the vector of all words) and then use the SVM classifier for classification; The third method TRNN (Traditional RNN) is to achieve the Traditional RNN model, The fourth method is to call the matlab LDA algorithm package for text classification; The fifth method is the neural autoregressive density estimation method; The sixth and seventh methods are the use of the depth of the Boltzmann built RMB model [10], and the softmax classifier is transformed accordingly; Bi-TRNN is based on the traditional RNN, using the results of a two-way network; Finally, the LSTM Bi-RNN and the GRU Bi-RNN represent the bi-directional recurrent neural networks based on the LSTM and GRU compute nodes, respectively. The Bi-RNN is a bi-directional recurrent neural network. Finally, the Attention LSTM Bi-RNN and Attention GRU

bidirectional Recurrent Neural Network of Attention Mechanism. From the results in Table 1, it can be seen that the neural network model with an attention mechanism achieves the strongest performance on most indicators, reaching 83.9% accuracy and 88.5% F value when tested on the news corpus of Xinhua News Agency. When tested on the Reuters corpus it reached a precision of 46.4% and F value of 51.8% effect. It can be seen that in the task of text learning, the advantages of the two methods of neural network representation learning and traditional keyword classification are taken into account, which has positive significance for the task of text classification.

In addition, in regard to the LSTM and GRU contrast, regardless of whether there is no attention mechanism used in the network structure, because LSTM has more fitting parameters than GRU, it is more suitable for large data learning and prediction. So, for the Reuters news corpus, the predictive effect of LSTM is significantly better than GRU, while for the Xinhua news data, the GRU results are better than the LSTM node.

4 Conclusion

Based on the task of text classification, this paper proposes a bi-directional recurrent neural network algorithm based on the neural network attention mechanism. After extracting the vector features of the text, the feature is input into the softmax classifier by dropout. Previous methods either based on keywords, or the use of neural networks, each have their own shortcomings. The former is too concerned about the keyword and ignores the role of other words. The latter treats all words equally, regardless of the particularity and importance of the keyword. The attention mechanism described in this paper can be a good combination of the advantages of both.

Bibliography

[1] Bahdanau, D.; Kyunghyun Cho, K.; Bengio Y. (2014); Neural machine translation by jointly learning to align and translate, ICLR 2015, *arXiv preprint arXiv*, 1409.0473, 2014.

[2] Chung, J.; Gulcehre, C.; Cho, K. et al. (2015); Gated feedback recurrent neural networks, *International Conference on Machine Learning*, 37, 2067-2075, 2015.

[3] Graves, A.; Schmidhuber, J. (2005); Framewise phoneme classification with bidirectional LSTM and other neural network architectures, *Neural Networks*, 18(5), 602–610, 2005.

[4] Hua, L. (2007); Text Categorization Base on Key Phrases, *Journal of Chinese Information Processing*, 21(4), 34–41, 2007. (in Chinese)

[5] Huang, E.H.; Socher, R.; Manning, C.D.; et al. (2012); Improving word representations via global context and multiple word prototypes, *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*, Association for Computational Linguistics, 873–882, 2012.

[6] Li, W.; Wu, G.; Zhang, F.; Du, Q. (2017); Hyperspectral Image Classification Using Deep Pixel-Pair Features, *IEEE Transactions on Geoscience and Remote Sensing*, 55(2), 844-853, 2017.

[7] Luong, T.; Socher, R.; Manning, C.D. (2013); Better Word Representations with Recursive Neural Networks for Morphology, *CoNLL*, 104–113, 2013.

[8] Mikolov, T.; Sutskever, I.; Chen, K.; et al. (2013); Distributed representations of words and phrases and their compositionality, *Advances in neural information processing systems*, 3111–3119, 2013.

[9] Mikolov, T.; Yih, W.T.; Zweig, G. (2013); Linguistic regularities in continuous space word representations, *Proceedings of NAACL HLT 2013*, Atlanta, USA, 746-751, 2013.

[10] Nitish, S.; Salakhutdinov, R.R.; Hinton G.E. (2013); Modeling documents with deep boltzmann machines, *Uncertainty in Artificial Intelligence - Proceedings of the 29th Conference*, 616-624, 2013.

[11] Pennington, J.; Socher, R.; Manning, C.D. (2014); GloVe: Global vectors for word representation, *Conference on Empirical Methods in Natural Language Processing*, Doha, Qatar, 1532–1543, 2014.

[12] Socher, R.; Huval, B.; Manning, C.D.; et al. (2012); Semantic compositionality through recursive matrix-vector spaces, *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, Association for Computational Linguistics, 1201–1211, 2012.

[13] Socher, R.; Perelygin, A.; Wu, J.Y.; et al. (2013); Recursive deep models for semantic compositionality over a sentiment treebank, *Proceedings of the conference on empirical methods in natural language processing (EMNLP)*, 1631–1642, 2013.

[14] Srivastava, N.; Hinton, G.; Krizhevsky, A.; et al. (2014); Dropout: A simple way to prevent neural networks from overfitting, *The Journal of Machine Learning Research*, 15(1), 1929–1958, 2014.

[15] Xu, X. ; Li, W.; Ran, Q.; et al. (2018); Multisource Remote Sensing Data Classification Based on Convolutional Neural Network, *IEEE Transactions on Geoscience and Remote Sensing*, 56(2), 937-949, 2018.

[16] Yao, Q.Z.; Song, Z.L.; Peng, C. (2011); Research on text categorization based on LDA, *Computer Engineering and Applications*, 47(13), 150–153, 2011. (in Chinese)

[17] Zeng, D.; Liu, K.; Lai, S.; et al. (2014); Relation Classification via Convolutional Deep Neural Network, *COLING*, 2335–2344, 2014.

[18] Zhang, A.-L., Liu, G.-L., Liu C.-Y. (2004); Research on multiple classes text categorization based o SVM, *Journal of Information*, 9, 6–10, 2004. (in Chinese)

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 13(1), 62-70, February 2018.

Elder Monitoring Workflow System for Independent Living

S. Jecan, D. Benta, L. Rusu, R. Arba

Sergiu Jecan
Computer Science for Economics Department
Babes-Bolyai University of Cluj-Napoca
Romania, 400591 Cluj-Napoca, M. Kogălniceanu, 1
sergiu.jecan@econ.ubbcluj.ro

Dan Benta
1. Beck et al. Services Cluj-Napoca
dan.benta@bea-services.com
2. Agora University of Oradea
Romania, 410526 Oradea, Piata Tineretului, 8

Lucia Rusu*
Computer Science for Economics Department
Babes-Bolyai University of Cluj-Napoca
Romania, 400591 Cluj-Napoca, M. Kogălniceanu, 1
*Corresponding author: lucia.rusu@econ.ubbcluj.ro

Raluca Arba
DSEGA-German
Babes-Bolyai University of Cluj-Napoca
Romania, 400591 Cluj-Napoca, M. Kogălniceanu, 1
raluca.arba@econ.ubbcluj.ro

Abstract: This paper presents an automatic workflow framed in a gerontechnology solution, as part of the Active and Assisted Living (AAL) platform in Mobile@Old project. Our solution aims to increase or preserve cognitive functions, to track medication and coordinate physical activity through an exercising game (exergame). The exergame is customized according to each elderly person's reactions and specificities. The workflow involves doctors, physiotherapists, the elderly person and their caregivers, in an ecosystem designed to ensure well-being and independence.
Keywords: workflow management systems, gerontechnology, personalized medicine, control.

1 Introduction

Elderly population is increasing around the world and technologies for this area should be more and more common. EU statistics show that by 2025 more than 20% of population will be over 65 years old, and a significant percent of them over 80. The European Commission has committed to offer healthy and dignified ageing for older people, in order to allow them to enjoy a good quality of life and their independence, as well as to remain active in society and their families. Healthy Life Years (HLY) or "disability-free life expectancy" indicator is the number of years a person of a certain age can expect to live without disability [11, 12]. HLY indicator is part of the core set of European Structural Indicators, recognized in the Lisbon Strategy [11, 12].

In this area we can mention several prominent research projects for elders' well-being and independent living: ASPA (Activating senior potential in ageing Europe); Demhow (Demographic change and housing wealth); LEPAS (Long-run economic perspectives of an ageing society); Maggie (Major ageing and gender issues in Europe); Multilinks - How demographic changes

shape intergenerational solidarity, well-being and social integration: a Multilinks framework ; Sharelife - Employment and health at 50+: a life history approach to European welfare state interventions; SPReW - Generational approach to the social patterns of relation to work; Recwowe - reconciling work and welfare in Europe [6].

Following standard assumptions, we take an elderly person to be defined as being at least 65 years old. Most elderly people are retired and are affected by several chronical conditions, most commonly related heart problems, mobility and mental deficiencies. Clinical studies show that the elderly's cognitive abilities are affected by various degrees, ranging from simple forms of amnesia, mild cognitive impairment (MCI) or mild to severe age-associated memory impairment (AAMI), to mild or severe dementia and Alzheimer [1, 7]. For these reasons, the Active and Assisted Living Joint Program (AAL JP) encourages several research programs for assistive technology and ubiquitous computing in order to offer several e-heath or m-heath solutions for independent living [6]. In recent years, gerontechnology has developed a symbiosis between gerontology and technology, an interdisciplinary research into technology for an ageing society, for improving the quality of life of elderly individuals [1].

Elderly monitoring activities can be assimilated with a workflow due to the repetitive character of the steps to be followed as well as the daily activities specific to the elderly. For this reason we offer a gerontechnology solutions based on an automatic, as a module of AAL platform Mobile@Old. Workflow based on to elderly personalized treatment, track medication and feedback for increasing or preserving cognitive function, physical activity focused on personalized exercises and wellbeing. Section 2 describes software and AAL solutions for elderly people, used as assistive technologies in the rehabilitation domain. Workflow description and business process analysis is given in Section 3. Section 4 presents a usability elders' test of the proposed workflow. Conclusions and future work are discussed in the last section.

2 Related work

Ambient assistive living technology (ALT) improves the elderly's lives, especially those with numerous conditions as a barrier to quality of life. The two major barriers are cognitive and physical in nature. Physical barriers consist of loss of physical function (lack of mobility, hemiparesis, paresis or failing eyesight and hearing) and cognitive barriers consist of loss of cognitive function, starting with amnesia, AAMI, MCI, evolving to severe MCI, dementia or Alzheimer [7, 8].

For these reasons, for most elderly people, drug therapy is part of their everyday life. Many important mobile applications are focused on medication management and reminders of daily activities or medical tests and appointments in order to offer personalized medicine based on companion diagnostic, monitoring and disease surveillance [9].

MyMeds is an application available for Android and iOS devices, as well as on the Web. It manages medication by sending daily reminders by text, email or push notification. The application shows medication that should be taken and its use, tracks the elderly's medication, saves and analyzes personal history. MyMeds has notifications on refilling prescriptions and suggests the best price for each upcoming prescription.

Care4today is another mobile health manager for seniors in both assisted living and independent living. It manages medication with reminders, tracking and connectivity. Through a user-friendly interface, it offers medication reminder tiles with information on timing, dosage, adherence tracking and dates for prescription refills. Each drug can be selected from a medication database built into the application. The tracking of medication schedule provides reports for every type of medication. Care4today can share these records with doctors directly from the screen and can connect with healthcare providers and caregivers [10].

There are also several applications for failing eyesight and hearing. *Dragon Dictation* is a free application that offers the possibility to dictate text and then send it as an email message. In the same manner, caregivers or seniors can dictate reminders to them and post on *Facebook* and *Twitter*. *VizWiz* is designed for partially sighted elderly people, to help them use their phone for taking photos, ask questions and get spoken answers. *Read2Go* provides a choice of font size and settings and offers several functions as an e-book reader: browse, search and download books [10].

If the elderly accept a physical activity based on exergames they get a relaxing daily fun and self-motivation [4]. Exergames use remote hand held controllers, motion sensors for capturing and monitoring body movements. Some of the exergames that we could mention: Nintendo Wii, Playstation Move, Dance Dance Revolution and Xbox Kinect [14].

3 Workflow automations for elders monitoring

3.1 Business process in exergame elders monitoring

Physiotherapy for the elderly is an essential component of their well-being. Its main goal is to preserve balance while walking and maintain core physical gestures that allow them to perform basic daily activities on their own - eating, drinking, getting (un)dressed, washing, combing, shaving, walking etc. The loss of any one of these abilities leads to loss of independence. Most common causes for loss of physical abilities are neurological conditions (paraplegia, hemiplegia, Parkinson), psychopathy (AAMI, MDI, Alzheimer, dementia), conditions affecting the locomotors system. In many of these cases, as well as in the case of seniors who do not suffer from any serious conditions, it is recommended to have occupational therapy, which focuses on daily activities to increase mobility and muscle tonus, as well as improve state of mind and reduce any form of psychopathy [5].

The main participants involved as users are: the *Elder* - defined as a person who has over 65 years - and sometimes Carers as supervisors that receive notifications from application. Carers cooperate with elderly and will be alerted when health conditions decrease or new health worrying symptoms appear. Carers can be husband, wife, children, other relatives or friends. Another participant is the Doctor, with a key role in the elderly's examination and monitoring. All messages, alerts, health monitoring parameters, daily habits, and time schedule for medication are set by the Doctor and all physical exercises or exergames are set by the Physiotherapist (*Kinetotherapist*). Both are main actors in Mobile@Old platform and coordinate automations of the elderly person's communication, in order to ensure their well-being [2, 3].

Since most elderly suffer from various diseases and health conditions, a *Preliminary evaluation* is necessary, carried out by one or several medical specialists (Step 1, Figure 1). The physicians fill the module *Recommendation* with information on the patient's health, chronic conditions, necessary medication, recommended behavior and lifestyle, as well as the module *Restrictions*, if applicable. They also determine the frequency of *Periodical Evaluations*. In this case the senior person will repeat Step 1 through Step 4 whenever necessary (Figure 1). All the data is handled in *Senior Database*. The data concerning the patient's health is updated daily/weekly/periodically according to the relevant parameters. As a result of the reservations expressed by our participants, the module Physical Activity Trainer (PAT) was conceived in two different versions. For each exercise in *Exercise Database* we store several specific fields like: exercise category, difficulty level, exercise name, relevant joints for the exercise, repeating time, exercise description [2, 3]. If the elders refuse to join Exergame program, they must follow medical recommendations. In this case, the person is not a user in Physical Activity Trainer (PAT), but if she changes her mind later all medical examinations can be used by the physiotherapist

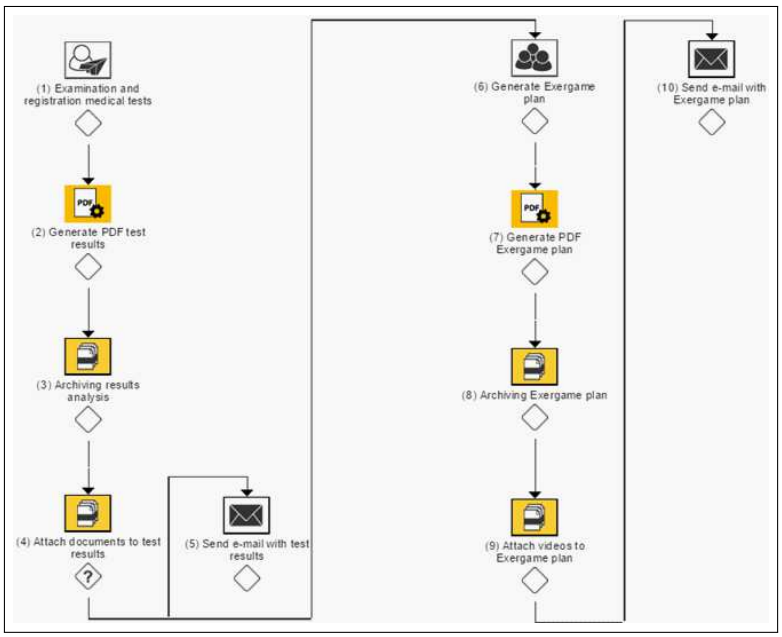


Figure 1: Elder monitoring workflow

to prepare a personalized exergame.

If the elderly person does not use Kinect and interaction with Physiotherapist is done only by tablet and smart mobile, the Physiotherapist records personalized exercises as a Videogame in *Reference Exercises* [2, 5].

For elderly person who accepts Kinect, PAT implementation offers several exergames as video serious game, designed in a customized manner, depending on the restrictions imposed by the chronic elderly. Every exergame has two avatars: one for user and one for trainer. Physiotherapist's recommendation takes into account personal elder profile, which includes diseases information, medical recommendation, sex, age, aso. All type of physical exercises is personalized, depending on the restrictions imposed by the chronic disease of the elderly. Physiotherapists will register a set of exercises with gradually levels of difficulty. All monitoring processes and elders' performance are directed by procedures associated to Kinect parameters which are recorded in *Exercise Database* [3].

3.2 Elders monitoring workflow

This workflow ensures automatic monitoring for PAT and Vital Sign Monitoring (VSM) module as a part of Mobile@Old platform. Daily analyzes and health parameters (systolic and diastolic blood pressure, pulse, glucose value, breathing rate) are repetitive activities that can be easily coordinated by a workflow. Caregiver intervention is justified only when they exceed normal limits, in which case the designed workflow alerts the physician and caregiver. To monitor physical exercise for the elderly who does not accept Kinect as an automated solution, the proposed workflow automates this process. Daily analyzes and health parameters (systolic and diastolic blood pressure, pulse, glucose value, breathing rate) are repetitive activities that can be easily coordinated by a workflow. Caregiver intervention is warranted only when they exceed normal limits, in which case the projected workflow alerts the doctor and caregiver.

Technical tools for workflow development was provided by JobRouter AG, consist on cloud storage with JobTable, JobSub, JobPDF, JobSelect and JobArchive modules. The designed workflow was tested in several browsers like: Firefox 47.0.1, Google Chrome Version 51.0.2704.106

m (64-bit), Safari 5.1.7 (7534.57.2), Opera 38.0 Version 38.0.2220.41 and Internet Explorer 11 Version 11.0.9600.17416. All tests passed with success. Designed workflow involves the following steps (Figure 1):

(1) Examination and registration medical tests are a Start Step. The designated doctor (*Doctor*) fills the form with basic information on the patient, common medical measures (blood pressure, heart rate), dates for various medical tests, information on whether these have been taken before or after meals, and optionally, details on other conditions or restrictions due to any chronic disease. Files documenting additional tests can be added. The patient can decide whether he wants to start Exergame or not. Doctor, a specialized professional, has the rights to determine the senior patient's examination, to indicate or change medication, to access for add, modify or delete the elderly's records and/or fields in several tables, such as: *Daily_ Habit*, *Elder_ Disease*, *Disease_ Type*, *Generic_ Drug*, *Drug*, *Normal_ Analysis*, and *Schedule for LAB Analyse*. *Doctor* decides *time management* for the medication: before or after eating, how long before lunch, 3/2 times per day (morning, lunch, evening), weekdays or even afternoon table, indication and contraindication for exercise and daily activities at home, and also follows and decides prescriptions based on *Analysis/ Medication Administration History* table. This information serves as a guide for *Physical Therapist* in the exergame individual plan [3, 13].

(2) Generate PDF test results is a System Step processed by JobPDF module. Based on a previously designed template, a PDF version of the form is generated.

(3) Archiving results analysis is a System Step processed by JobArchive module, archive action. The previously mentioned PDF is archived. It can be accessed by doctors or elderly patients for monitoring and disease surveillance.

(4) Attach documents to test results is a System Step processed by JobArchive module, clip action. It allows the addition of test results to the form archived in step 3.

(5) Send email with test results is next System Step processed by Send E-mail automatic procedure, with a document that contains the results of medical tests. If the patient does not wish to continue the Exergame course of treatment, he receives an email confirmation of this decision.

(6) Generate Exergame plan is a User Step directed by the *Physical Therapist*. If the patient accepts the Exergame program, the designated physiotherapist receives the test results and a request to design an appropriate course of treatment. He must then fill the corresponding form with the following information: name of the exercise, number of repetitions (5/10/20 ...), and frequency (once, twice or three times a day). Optionally, a video demonstrating the exercises can be added. It should be mentioned that one and the same video may come with different indications (for example, different number of repetitions), depending on the elderly's patients physical abilities and health condition [2, 3, 5].

(7) Generate PDF Exergame plan is a System Step processed by JobPDF module. Based on a previously designed template, a PDF corresponding to the course of treatment is generated.

(8) Archiving Exergame plan is a System Step processed by JobArchive module, archive action. The PDF is archived.

(9) Attach videos to Exergame plan is a System Step processed by JobArchive module, clip action. Videos corresponding to each exercise are attached to the form archived in step 3.

(10) Send e-mail with Exergame plan is a System Step processed by Send E-mail automatic procedure. After the completion of the treatment plan, the patient and/or the caregiver receive an e-mail with the treatment plan and corresponding test results.

System Steps of all types (JobPDF, JobArchive or Send E-mail) are automatically processed by the system. If further needs will arise and multiple Doctors and Physiotherapist will be involved for a single elder, we can implement parallelization steps as a beginning and ending sub-process. For this case we used only User steps (directed by doctor, physiotherapist and elder) and System steps (7-10).

The physician or physical therapist (as workflow users) can access periodic reports detailing the evolution of the medical condition of the elderly included in the AAL program, the accepted or refused personalized exercise program as well as the performance and evolution of each elderly person. Monitoring exergame program, tests, performance and evolution are directed by the PAT module which involves Kinetics and other wearable technologies and is detailed in other papers [5, 13]. Moreover, if a patient is transferred from one doctor to another, the last one will have access to a full patient medical history (Figure 2).

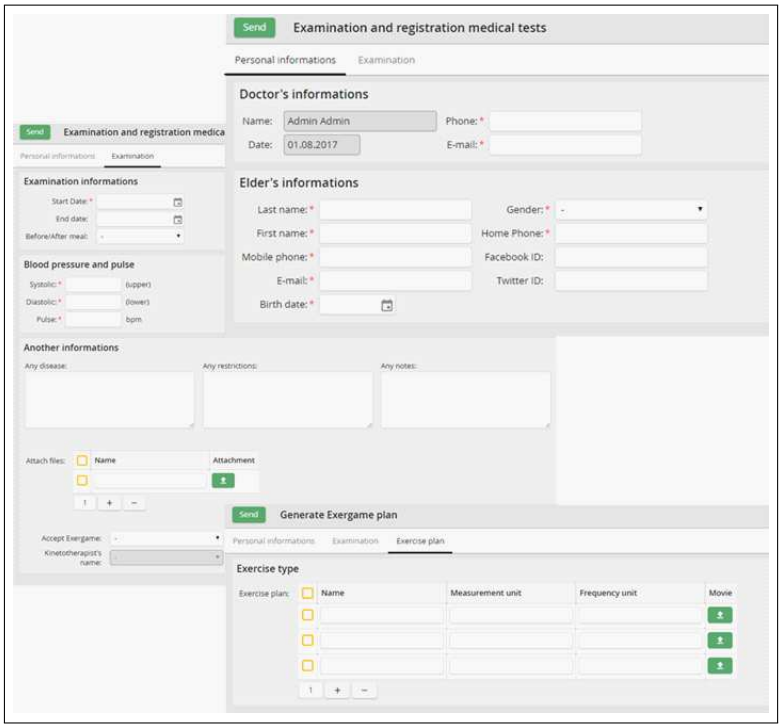


Figure 2: GUI for personal information, examination and exercise plan

4 Experimental results

We offer the doctor same level of importance as the patient on the platform, based on the analysis of response related to health issues. Needs analysis started with a questionnaire with 61 items, that was applied to 69 persons, elderly people between 60 and 87 years (median 67,4), with chronic or severe diseases, and several of them (12% - 8) with disabilities. These results (Figure 3) showed that most of the elders suffer from at least one chronic disease (95%), or two

chronic illnesses (62%) even 3 chronic diseases (47%) and some of them have impairments or disabilities [3, 5].

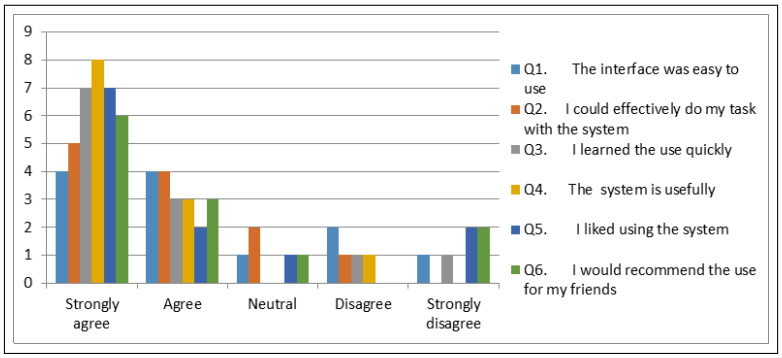


Figure 3: Feedback from elders

Out of the elderly that participated in our study, only 8 out of 69 (12%) accepted to use the proposed workflow as an automatic solution to monitor their health parameters and exergames, and accepted to use the mobile application made available to them. Only 4 out of 69 (6%) accepted monitoring exergames via Kinect. The target audience has coronary heart disease (7), Myopia (6), Hypertension (5), Hypermetropia (4), AAMI (4), Osteoporosis (4), Diabetes (3), MCI (3), Thyroid disorders (2), overweight (3) or Angina (2). Only two of them have disabilities: Hemiparesis. All of them are computer literate users (CLUs) and have a healthy life style. Elders accepted our proposal about workflow and usability evaluation [3, 5, 13].

On the other site, we invited several CLUs: doctors (5) physiotherapists (3) and caregivers (6) to use our workflow and evaluate usability based on a questionnaire with 15 items. They interacted with workflow as users and with elders as patients. In fact, everyone has a user role in our proposed workflow. We presented several relevant results related to the elderly because all the others (doctors, physiotherapists and caregivers) gave only answers that were strongly agree or agree. Seniors find the system useful (Q4), learn to use it quickly (Q3) and like it (Q5). Only few of them dislike the interface (Q1), not recommended it to others and are neutral to performing tasks with the system (Q2). Figure 3 shows details about 6 of 15 answers, the rest included open questions (4) related to CLU (3) and wearable technologies (2).

5 Conclusions

This paper presents an automatic workflow for monitoring of the elderly. We focused on two major problems of third age: monitoring health parameters and chronically disease and monitoring exercise plans. We discussed three major participants in the workflow system: (i) the doctor, who coordinates periodic medical examinations and personalized treatment, (ii) the physiotherapist, who follow doctor's indications and contraindications and coordinate exercise plans and daily activities at home and (iii) the elder person, who is the core beneficiary.

All steps are only User steps conducted by the doctor, physiotherapist and the senior user or System steps. System steps use several database tables managed by Mobile@Old platform and generate several documents, such as medical examination reports, medication schedule or exergame plan.

All actors involved in the workflow testing are computer literate users (CLUs) and have accepted our questionnaire related to system usability. Moreover, the consulted seniors have a healthy life style and find our solution useful and easy to learn. A real benefit seems to be

for doctors, physiotherapists and caregivers, who appreciated asynchronous communication and message automation.

As future improvement, we intend to develop new functionalities and to adapt the application to new requirements in order to best fit elders' needs in this area.

Acknowledgment

This research was supported by the Executive Unit for Financing Higher Education, Research and Development and Innovation through the Partnership Program, the project "Mobility pattern assistant for elderly people", project number PN-II-PT-PCCA-2013-4-2241. We gratefully acknowledge the contribution of colleagues from Beck et al. Services SRL Cluj-Napoca (RO) and JobRouterŽ AG Mannheim (DE).

Bibliography

[1] Hyry, J. (2015); Designing Projected User Interfaces as Assistive Technology for the Elderly, *Acta Universitatis Ouluensis, A Scientiae Rerum Naturalium*, 664, 2015.

[2] Lohan, E.S.; Cramariuc, O.; Malicki, L.; Brencic, N.S.; Cramariuc, B. (2015); Analytic Hierarchy Process for assessing e-health technologies for elderly indoor mobility analysis, *MOBIHEALTH'15 Proceedings of the 5th EAI International Conference on Wireless Mobile Communication and Healthcare*, Great Britain - October 14 - 16, 2015, pp. 54-57.

[3] Rusu, L.; Mocanu, I.; Jecanm S., Sitar, D. (2016); Monitoring Adaptive Exergame for Seniors, *Journal of Information Systems & Operations Management*, 10(2), 2016.

[4] Ying-Yu, C. (2015); Exergaming: therapeutic benefits in older adults, *http://lermagazine.com/article/exergaming-therapeutic-benefits-in-older-adults*, accessed January 2016.

[5] Zdrenghea, D.; Ile, M.; Zdrenghea, M.; Sitar-Taut, A-V., Pop, D. (2014); The Effects of Maximal and Submaximal Exercise Testing on NT-proBNP Levels in Patients with Systolic Heart Failure, *Romanian Review of Laboratory Medicine*, 22(1), 25-33, DOI: 10.2478/rrlm-2014-0008, March 2014, *http://www.degruyter.com/view/j/rrlm.2014.22.issue-1/rrlm-2014-0008/rrlm-2014-0008.xml?format=INT*.

[6] Ambient Assisted Living Programme; *http://www.aal-europe.eu*, accessed August 2017.

[7] Alzheimer's & Dementia Association; *http://www.alz.org/ dementia /mild-cognitive-impairment-mci.asp*, accessed August 2017.

[8] The Active and Assisted Living Joint Programme (AAL JP); *https://ec.europa.eu/digital-single-market/en/active-and-assisted-living-joint-programme-aal-jp*, accessed August 2017.

[9] President's Council of Advisors on Science and Technology (2008); Priorities for Personalized Medicine, 2008.

[10] Top app for the elderly (2014); *https://myageingparent.com/technology/communication/top-ipad-apps-for-the-elderly/*, accessed August 2017.

[11] European Commission (2014); Population ageing in Europe, Facts, implications and policies, Research and Innovation, *http://europa.eu*, Luxembourg: Publications Office of the European Union, doi:10.2777/60452, 2014, accessed August 2017.

[12] European Commission; *https://ec.europa.eu/health/population_groups/elderly_en*, accessed August 2017.

[13] Life Sciences-Healthcare and the Institute of Bio-Sensing Technology for the Microelectronics and Biomedical iNets (2012); Assisted Living Technology, A market and technology review ;*www.inets-sw.co.uk*, accessed March 2016.

[14] Xbox.com; Xbox kinect ;*www.xbox.com/kinect*, accessed March 2016.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 13(1), 71-82, February 2018.

A Knowledge Base Completion Model Based on Path Feature Learning

X. Lin, Y. Liang, L. Wang, X. Wang, M. Yang, R. Guan

Xixun Lin, Xu Wang
Key Laboratory for Symbol Computation and
Knowledge Engineering of National Education Ministry,
College of Computer Science and Technology,
Jilin University, Changchun 130012, China

Limin Wang
School of Management Science and Information Engineering,
Jilin Province Key Laboratory of Internet Finance,
Jilin University of Finance and Economics, Changchun 130117, China

Mary Qu Yang
MidSouth Bioinformatics Center and Joint Bioinformatics Ph.D. Program,
University of Arkansas at Little Rock and
University of Arkansas for Medical Sciences, 2801 S.
University Avenue, Little Rock, Arkansas 72204, USA

Yanchun Liang, Renchu Guan*
Key Laboratory for Symbol Computation and
Knowledge Engineering of National Education Ministry,
College of Computer Science and Technology,
Jilin University, Changchun 130012, China
Zhuhai Laboratory of Key Laboratory of Symbolic Computation and
Knowledge Engineering of Ministry of Education,
Zhuhai College of Jilin University, Zhuhai 519041, China
*Corresponding author: guanrenchu@jlu.edu.cn

Abstract: Large-scale knowledge bases, as the foundations for promoting the development of artificial intelligence, have attracted increasing attention in recent years. These knowledge bases contain billions of facts in triple format; yet, they suffer from sparse relations between entities. Researchers proposed the path ranking algorithm (PRA) to solve this fatal problem. To improve the scalability of knowledge inference, PRA exploits random walks to find Horn clauses with chain structures to predict new relations given existing facts. This method can be regarded as a statistical classification issue for statistical relational learning (SRL). However, large-scale knowledge base completion demands superior accuracy and scalability. In this paper, we propose the path feature learning model (PFLM) to achieve this urgent task. More precisely, we define a two-stage model: the first stage aims to learn path features from the existing knowledge base and extra parsed corpus; the second stage uses these path features to predict new relations. The experimental results demonstrate that the PFLM can learn meaningful features and can achieve significant and consistent improvements compared with previous work.
Keywords: knowledge base completion, random walks, path features, extreme learning machine.

1 Introduction

Large-scale knowledge bases (KBs), such as Never-Ending Language Learning (NELL) [6], Freebase [3], Yago [35], and DBpedia [11], construct their own ontologies derived from facts that

are manually or automatically extracted from databases, such as Wikipedia or other unannotated web pages. These KBs usually contain billions of facts, and each fact can be organized as a triple $R_k(a_i, b_j)$, such as AthletePlaysForTeam (Messi, Barcelona) and Professions (van Gogh, Artist). The variables a and b represent entities or attributes in the real world, and R represents the binary relationship between them. Billions of these facts constitute a large complicated knowledge network. The construction of such large-scale KBs is significant for many types of natural language processing research, e.g., question answering [2], semantic analysis [1], and information retrieval [13].

However, existing facts stored in KBs are not comprehensive compared with real-world knowledge. Many important entity-relationships are missing due to improper manual operations or the drawbacks of fact-extraction models [38]. Fortunately, most of the information can be inferred from existing facts in KBs. Thus, the task of knowledge base completion (KBC) has aroused intense attention in both academic and industry research [27, 36].

Graph-based approaches for KBC is an important subfield of statistical relational learning (SRL) [10]. A classic learning method in SRL is Markov logic network (MLN) [32], which combines the probabilistic graphical model with first-order logic for knowledge inference. Although MLN is very powerful, it needs to explore all derivation trees and combine them to calculate even a single ground fact; therefore, MLN does not perform well when the amount of data is large. Inductive logic programming (ILP), such as the first-order inductive learner (FOIL) and its variants, is another type of SRL [18, 23, 31]. FOIL deploys the separate-and-conquer strategy to learn the first-order logic rule set; however, similar to MLN, FOIL cannot handle large-scale KBs. ProPPR [37] is a new direction that attempts to improve the scalability of FOIL.

The path ranking algorithm (PRA) [19] aims to reason new facts directly from facts observed in the KBs and achieves state-of-the-art performance compared with previous work. PRA attempts to encode the entire KB as a large edge-labeled directed graph and exploits random walks to extract informative path features. Each path feature can be viewed as the most frequent sequence of relations in the graph. However, PRA is deficient in terms of addressing long-tail data distributions because some rare entities lack common path features. Another direction emphasizes the extraction of relations from a large text corpus and combination with current KBs to enhance the performance of PRA [7–9, 20, 21]. The extracted relations are added to the original directed graph as new edges to compensate for the lack of sufficient facts. However, directly adding relations with similar semantics to the KB results in severe path explosion and feature sparsity.

In addition to the aforementioned models, many approaches focus on latent feature models [27], which aim to learn latent representations of entities and relations by minimizing a reconstruction loss or a margin-based ranking loss [4, 12, 25, 28, 29]. Latent feature models are effective to encode knowledge representations, but when the KB tensor constructed from data has a higher rank, it is more difficult to obtain meaningful embeddings.

In this paper, motivated by PRA, we propose a more general framework called the path feature learning model (PFLM). The PFLM is a two-stage model: In the first stage, two types of path features (directed relation paths (DRPs) and supplement relation paths (SRPs)) are generated from different target entity pairs by random walks; in the second stage, we incorporate the learned features into the kernel extreme learning machine (KELM) [14, 15] for KBC. In addition to the advantages of good generalization and fast learning speed [16], KELM is robust in terms of triple classification, which is illustrated in our experiments. The main highlights of this paper are summarized as follows:

1. With the stage of path feature learning, plenty of classification algorithms (in our case, we use ELM and KELM) can be easily incorporated into our framework. Using the new path

features and the single-hidden layer feedforward neural network, the PFLM can perform KBC effectively and efficiently.

2. The PFLM is an extensible scheme. In addition to the DRPs introduced by PRA, the proposed SRPs and other expected path information (e.g., immediate nodes) can be collected in the path feature learning stage.

3. The results of experiments show that our model achieves significant and consistent improvements compared with baseline models, such as PRA [19] and CPRA [7].

The rest of our paper is organized as follows: Section 2 introduces the background of basic model setting. In Section 3, we introduce the detailed implementation strategies for our model. Experimental details and discussions are provided in Section 4. The last section draws our conclusions and identifies future work.

2 Background

In this section, we first give a brief overview of PRA and KELM. PRA is a classic algorithm of KBC. KELM is an efficient neural network architecture that we employ to implement triple classification. Triple classification [33] is a standard way to evaluate KBC.

2.1 Path ranking algorithm (PRA)

PRA leverages multiple random walks to reach tail entity b from head entity a for each entity pair (a,b). The filtered paths that connect entity pair (a,b) serve as different path features. Then, the random probability of head entity a reaching tail entity b through path-constrained random walks is calculated as the feature value. Finally, PRA adopts logistic regression based on the limited-memory Broyden-Fletcher-Goldfarb-Shanno(L-BFGS) to perform triple classification. The effectiveness of PRA depends on the power of the Horn clause rules to obtain the entity constraints. A relation path is generated by the conjunction of a sequence of triples, for example, given WriterCreatedRole (Hemingway, Santiago) → RoleDescribledInBook (Santiago, The Old Man and the Sea), the predicative information WriterWroteBook (Hemingway, The Old Man and the Sea) can be inferred. By exploiting the implicit relation paths, novel facts that do not originally exist in the KBs are produced.

The more detailed procedures are as follows: PRA encodes the whole KB as a directed edge-labeled graph $G(N, T, E)$. N is the set of entities in KB, E is the set of edges connecting entity pairs, and T is the collection of edge types representing first-order logic rules. A path P is the ordered sequence of edges P=$\{R_1,\ldots,R_n\}$. PRA applies multiple random walks, starting from the head entity a, to obtain the common path set $S(P) = \{P_1, \ldots, P_k\}$ for the common tail entity b. $S(P)$ is relevant to a specific relation R. The most frequent paths from $S(P)$ are selected as path features. Each path feature value $V_{a,P}(b)$ is calculated using the recursive formulas defined as follows.

If P is an empty path:

$$V_{a,P}(b) = 1 \quad \text{if} \quad a = b \tag{1}$$

$$V_{a,P}(b) = 0 \quad \text{otherwise.} \tag{2}$$

If P is not an empty path, $P' = R_1, \ldots, R_{n-1}$:

$$V_{a,P}(b) = \sum_{b' \in range(P')} V_{a,P'}(b') \cdot P(b|b'; R_n), \tag{3}$$

where $P(b|b';R_n)$ is the probability of reaching target node b from b' with one edge labeled R_n, and $range(P')$ is the set of target nodes where the path P' ends.

2.2 Kernel extreme learning machine (KELM)

Extreme learning machine is a learning algorithm aims to train single-hidden layer feedforward neural networks. Huang provided strict theoretical proof that the standard single-hidden layer feedforward neural networks training process can be considered as finding a least-squares solution β' for the linear system $H\beta = T$ that allows the hidden node parameters w_i and b_i to be randomly generated. H is the hidden layer's output matrix, β is the output weights and T represents the training sample outputs. In most cases, when the number of hidden nodes is much less than the number of distinct training samples, $\beta' = H^{\dagger}T$ can be calculated with zero error to approximate these training examples, where $H^{\dagger}$ is the Moore-Penrose generalized inverse of H. ELM shows good generalizability and learning speed; it performs better than the conventional learning algorithm in many areas, such as face recognition [40], text classification [39], image classification [5], and medical diagnosis [24].

Kernel extreme learning machine (KELM), in contrast to traditional ELM, conducts classification and regression with kernel functions and kernel parameters (C, γ) instead of the number of hidden-layer nodes and path feature mappings $h(x)$. The reason we choose KELM as our classifier is its robustness, which will be discussed in our experiments. The kernel matrix is defined as

$$\Omega_{ELM} = HH^T \tag{4}$$

$$\Omega_{ELM}(i,j) = h(x_i) \cdot h(x_j) = K(x_i, x_j), \tag{5}$$

where $K(x_i, x_j)$ is the kernel function.

The output weights and output function are expressed as follows:

$$\beta = H^T(\frac{I}{\lambda} + HH^T)^{-1}T \tag{6}$$

$$f(x) = h(x)\beta = \begin{bmatrix} K(x, x_1) \\ . \\ . \\ . \\ K(x, x_N) \end{bmatrix}^T (\frac{I}{\lambda} + \Omega_{ELM})^{-1}T. \tag{7}$$

3 Proposed method

Motivated by PRA, we propose a novel model called the path feature learning model (PFLM), which combines path feature learning with KELM to achieve triple classification. Fig. 1 illustrates the framework for our model. The inputs of our model are the knowledge base and the large parsed corpus. The path feature learning aims at providing a large real-valued matrix for KELM. The final output of PFLM is predicting whether the entity pair contains the specific relation, which is a binary classification problem. The main reason for combining these operations is that the first stage of our model assumes that all background knowledge and samples are ground facts, which makes it more suitable for machine learning models. On the premise of representing the KB as a large directed heterogeneous graph with encodable Horn clause rules, random walk inference outperforms traditional searching methods, such as the breadth-first search algorithm, in the process of the feature searching. The retrieved first-order Horn clause rule sets, which are represented as paths, can be recognized as candidate path features. We first introduce the

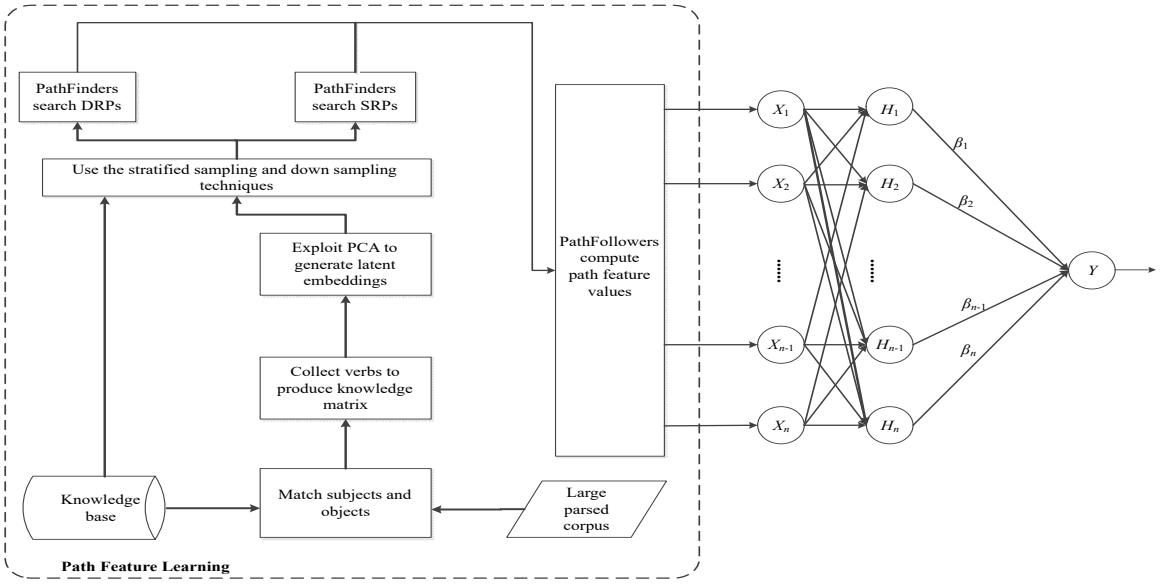


Figure 1: Flowchart of the path feature learning model.

concept of path feature learning, which includes DRPs and SRPs. Then, we introduce two practical techniques in subsection 3.2 for efficient implementation.

3.1 Path feature learning

In the path feature learning stage, we define two path types according to their components. DRPs are the existing chains of the Horn clause rules that are directly searched by PRA. SRPs are the sequences of relations with some latent embeddings that are learned from factorization of the KB matrix M. Specifically, each row of the matrix is a tuple $t = (subject, object)$, while *subject* and *object* are entities that exist in the KB and also occur in 500 million dependency-parsed Web documents [7, 21]. Each column of the matrix is the verb in t. The elements of the matrix are the frequencies of (t, *verb*) in the extra corpus. After row normalization and centering, we apply PCA to M to obtain the latent embeddings of the *verb*s. As shown in the Fig. 1, both DRPs and SRPs can be searched and computed by PRA according to the equations 1, 2 and 3.

We consider a specific example to explicitly describe the two path types. An important reasoning path for the triple AthletePlaysInLeague (LeBron James, NBA) can be expressed by DRP as AthletePlaysForTeam (LeBron James, Cleveland Cavaliers) → TeamBelongedToLeague (Cleveland Cavaliers, NBA). Unfortunately, the knowledge repository may miss the valuable relational edge AthletePlaysForTeam (LeBron James, Cleveland Cavaliers), so it seems impossible to infer the fact AthletePlaysInLeague (LeBron James, NBA), which is known to us but disappears in the KB. This feature sparsity problem leads to severe over-fitting for many isolated entity pairs, which limits the performance of PRA.

In our method, the proposed SRPs combine the extracted subject-verb-object (SVO) information that expresses similar semantics of AthletePlaysForTeam from the extra text corpus. The method exploits verb clustering techniques to map the lexicalized edge labels (e.g., 'works for' , 'plays for', 'leads to') to some latent embeddings, forming a new edge type: LatentEM1 (LeBorn James, Cleveland Cavaliers). The probability of extracting important path features is increased by using both DRPs and SRPs compared with that of PRA, which only considers DRPs in

KBs. The path features are quite logical for our understanding and effective for the large feature space. After the entity pair path information has been taken into account for all datasets, the most frequent m candidate path features are selected as the final path features. Each path feature is composed of relations or latent embeddings or mixtures of both. It is reasonable to regard the random probability values as path feature values, even if the computation requires enormous running time. The larger the probability is, the more likely a specific path feature will be selected for our target relation. This attribute cannot be reflected by binary features. Finally, the abstract graph information is mapped to relevant feature vectors, and the KELM completes the task of triple classification.

It is noted that the PFLM is a general scheme. We chose KELM as our second stage classification model, in contrast to the logistic regression model adopted by PRA. Different kernels and parameters can be selected for different relation decisions. In general, in the stage of path feature learning, in addition to DRPs, we propose SRPs to explictly identify path features from the KBs and extra corpus; in the second stage, the feature matrices are transferred to the kernel extreme learning machine classifier to complete the task of triple classification. Furthermore, each feature matrix is computed by random walks following the concrete relations indicated by DRPs or SRPs. Missing relations can be added to the KBs based on accurate classification results.

3.2 Two practical sampling techniques

Due to the knowledge bias in KBs, sampling techniques must be employed to balance the datasets. When adopting random walks to obtain positive and negative samples, two points must be considered. First, the relation types contained in KBs are probably uneven, and this phenomenon may affect the overall model's feature distribution. Therefore, we consider [21], which employs stratified sampling [32] to obtain identical sample numbers for different types of relations when possible. Secondly, PRA holds the closed-world assumption. In addition to the small portion of positive samples already existing in knowledge graph G, most of the samples are negative ones produced by random walks, which may lead to a serious distribution imbalance. We use a downward sampling technique to solve this problem. To control the positive and negative samples within a reasonable proportion (in our case, the ratio is 1:10), the PFLM selects the common relation paths in accordance with the sampling numbers. For example, in a large-scale KB, we present thousands of entity pairs for AthletePlaysInLeague. In addition to the path AthletePlaysForTeam $\rightarrow$ TeamBelongedToLeague, many other related paths, such as AthletePlaysSport $\rightarrow$ AthletePlaysSport^{-1} $\rightarrow$ LeaguePlayers^{-1}, AthletePlaysSport$\rightarrow$StadiumHomeToSport$^{-1}\rightarrow$ LeagueStaduim^{-1}, and AthletePlaysForTeam $\rightarrow$ AthletePlaysForTeam^{-1} $\rightarrow$LeaguePlayers^{-1} are contained in the set. We denote R^{-1} as the inverse of relation R (i.e., WriterWroteBook^{-1} is equivalent to BookWritenByWriter). After sampling and feature computing, the PFLM provides the final feature matrices composed of multiple entity pairs with their feature vectors for the target relation AthletePlaysInLeague to the KELM classifier.

4 Experiments and discussion

We use the Never-Ending Language Learning (NELL) dataset to evaluate our model. NELL is a large-scale KB whose contents are learned by reading from the web over time. The dataset we employ is a benchmark and can be downloaded from `http://rtw.ml.cmu.edu/emnlp2013_pra/`. This dataset contains 15 relations, and each relation is split into two parts: 10% test data and 90% training data. To rigorously compare and decrease the extensive feature computing, we follow the same rules as CPRA [7] and set the number of each sample's path features as 750.

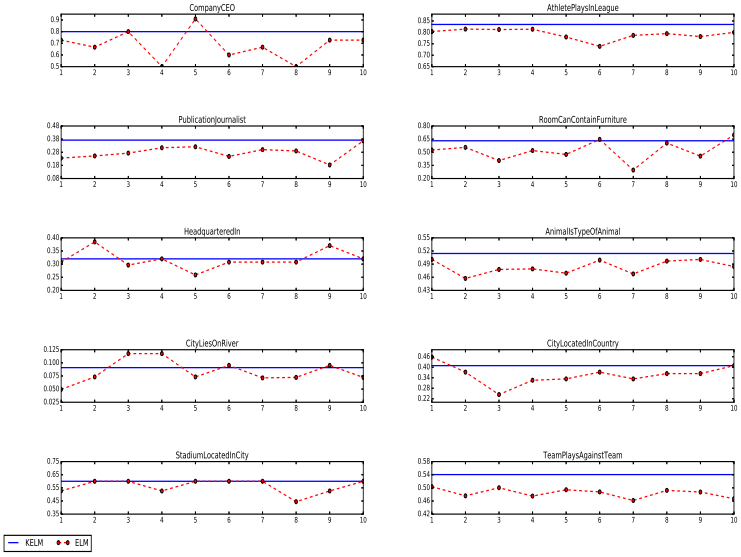


Figure 2: Comparison of 10 relations' data with ELM and KELM based on RBF kernel function. The horizontal axis represents the experiment times, and the vertical axis represents the F1-measure.

Table 1: The comparisons of four different kernels for KELM

	Macro-precision	**Macro-recall**	**Macro-accuracy**	**Macro-F1**
Linear kernel	0.5249	0.3798	0.8939	0.4130
Wavelet kernel	0.5974	0.3862	0.8960	0.4332
RBF kernel	**0.8825**	**0.4172**	**0.9488**	**0.5279**
Polynomial kernel	0.8021	0.4034	0.9446	0.4986

We first compare the robustness of ELM and KELM for triple classification on ten diverse relation datasets; the corresponding path features are the same and are both provided by path feature learning. The results are shown in Fig. 2. From Fig. 2, it can be seen that when the number of samples is small, the ELM's F-measure oscillates between several fixed values, and the vibration is drastic in some relation datasets, such as StadiumLocatedInCity. When the number of samples is larger, we intend to implement more hidden layer nodes to improve the network generation, and ELM fluctuates more severely, thus causing a decline in accuracy and stability, such as the relation RoomCanContainFurniture. By contrast, when we choose KELM and fix the corresponding kernel parameters, KELM is more robust and performs better than ELM. Therefore, we chose KELM as our classifier instead of ELM based on the experimental results.

Moreover, we compare the performance of different kernel functions on the same task. Table 1 presents the experimental results. The best performances are indicated in bold for all tables in this section. We employ the grid-search strategy to select the best parameters C and γ for different kernel functions. The parameters are both tuned in $\{2^{-5}, 2^{-3}, \ldots, 2^{15}\}$. We evaluate the linear, wavelet, RBF and polynomial kernels on four measurements among 15 relation datasets. From the results, we can observe that the effects of different kernels vary considerably, and the RBF kernel achieves the best performance. The most notable result is that the RBF kernel is 35.8%, 3.8%, 5.5%, and 11.5% higher than the Linear kernel, which indicates that the dot

Table 2: Running time comparison for PRA, CPRA and PFLM

	PRA	**CPRA**	**PFLM**
Time(min)	1313	1160	1191

Table 3: Detailed comparison of the F1-measure for PRA, CPRA, and PFLM

	PRA	**CPRA**	**PFLM**
AnimalIsTypeOfAnimal	0.5214	0.5270	**0.6069**
AthletePlaysForTeam	0.2156	0.6387	**0.6667**
AthletePlaysInLeague	0.8099	0.7402	**0.8370**
CityLiesOnRiver	0.0493	0.3076	**0.5484**
CityLocatedInCountry	0.1538	0.5454	**0.5778**
CompanyCEO	0.2857	0.3529	**0.7273**
CountryHasCompanyOffice	0.0000	0.0000	**0.1481**
DrugHasSideEffect	**0.9629**	0.9427	0.9474
HeadquarteredIn	0.3076	**0.6382**	0.6047
locationLocatedwithinLocation	0.3950	0.4147	**0.4286**
PublicationJournalist	0.0967	0.1594	**0.5444**
RoomCanContainFurniture	0.7206	0.7320	**0.7985**
StadiumLocatedInCity	0.5263	0.6666	**0.7143**
TeamPlaysAgainstTeam	0.4736	0.2086	**0.5800**
WriterWroteBook	0.5911	0.8000	**0.8218**

product in an infinite-dimensional space is more suitable for our problem. Therefore, we choose the RBF kernel as our kernel function in the following experiments. In Table 2 we compare the running times of PRA, CPRA, and PFLM. From Table 2 we can conclude the following. 1) PRA adopts logistic regression with L2 regularization optimized by L-BFGS to complete the triple classification. It calculates the path feature values with dozens of iterations, which means that its convergence is slow. 2) CPRA reduces the number of relation paths that random walkers need to search, therefore it is the fastest method. 3) Although the computation and selection for the path feature learning of the DRPs and SRPs consume substantial running time, the PFLM is not as time-consuming as PRA.

In Table 4 we compare PRA, CPRA and PFLM on 15 relations. The experimental results show that the PFLM achieves significant and consistent improvement. For example, the PFLM is 12% and 23% higher than CPRA and PRA on the macro F1 criterion, respectively. We believe the PFLM outperforms CPRA because the PFLM can better incorporate expressive path information to address the problem of feature sparsity, which limits the performance of CPRA and PRA. We reproduce the experiments in [30], and Table 3 shows a more detailed comparison of the F1 measurement for the three models on 15 NELL relations. From Table 3, we can conclude that the PFLM shows significant improvement in the triple classification of 13

Table 4: Macro measurement comparison for PRA, CPRA and PFLM (%)

	Macro-precision	**Macro-recall**	**Macro-F1**	**Macro-accuracy**
PRA	0.7458	0.3443	0.4073	0.9373
CPRA	0.8094	0.4241	0.5116	0.8708
PFLM	**0.9009**	**0.5152**	**0.6367**	**0.9439**

Table 5: Most impressive path features in the three relations

	Path Type	Element of path
AtheletePlaysForTeam	DRPs	-athleteledsportsteam- -athelteledssportsteam-teamhomestadium-teamhomestadium^{-1}-
	SRPs	-LE1^{-1}-LE2- -LE1^{-1}-LE2-teamhomestadium-teamhomestadium^{-1}-
CityLiesOnRiver	DRPs	-proximityfor-subpartof^{-1}-riverflowsthroughcity^{-1}- -statecontainscity^{-1}-atlocation-riverflowsthroughcity^{-1}-
	SRPs	-LE1LE5^{-1}- -LE1LE5^{-1}-riverflowsthroughcity-riverflowsthroughcity^{-1}-
CompanyCEO	DRPs	-organizationleadbyperson- -worksfor-
	SRPs	-LE1LE5- LE1LE5^{-1}-agentcollaborateswithagent- -LE1^{-1}-LE2-LE1LE2-agentcollaboratesswithagent-

relations, with the largest increases of 463.0 % and 241.5 % compared with PRA and CPRA on the relation PublicationJournalist. Table 5 displays the impressive path features (DRPs and SRPs) that are searched by path feature learning on three different relations. These paths are common to all entity pairs, and their importance in triple classification are reflected by their own path feature values. For each path type, we present two examples, and we can observe that the highest-weighted DRPs and SRPs have similar semantics with the target relations.

5 Conclusions and future work

In this paper, we propose the PFLM to solve the problem of large-scale KBC. The PFLM extracts DRPs and SRPs during the path feature learning stage and sends these features to the kernel extreme learning machine. The PFLM shows superior classification ability compared with the original algorithm and its variants on the benchmark datasets. For our future work, we plan to 1) incorporate more path information into the path feature learning schemes, e.g., the path-type limitation and immediate nodes; 2) enhance the efficiency for computing path feature values by exploiting parallel or distributed computing.

Acknowledgements

The authors are grateful for the support of the National Natural Science Foundation of China (No. 61572228, No. 61472158, No. 61300147, No. 61602207 and No. 61572225), the United States National Institutes of Health (NIH) Academic Research Enhancement Award (No.1R15GM114739), the National Institute of General Medical Sciences (NIH/NIGMS) (No.5P20GM103429), the Science Technology Development Project from Jilin Province (No. 20160101247JC), the Premier-Discipline Enhancement Scheme supported by the Zhuhai Government and the Premier Key-Discipline Enhancement Scheme supported by Guangdong Government Funds.

Bibliography

[1] Agirre, E.; Lacalle, O.; Soroa, A. (2014); Random walks for knowledge-based word sense disambiguation, *Computational Linguistics*, 40, 57–84, 2014.

[2] Berant, J.; Chou, A.; Frostig, R.; Liang, P. (2013); Semantic parsing on Freebase from question-answer pairs, *Proceedings of EMNLP*, 1533–1544, 2013.

[3] Bollacker, K.; Evans C.; Paritosh, P.; Sturge, T.; Taylor, J. (2008); Freebase: a collaboratively created graph database for structuring human knowledge, *Proceedings of KDD*, 1247–1250, 2008.

[4] Bordes, A.; Usunier, N.; García-Durán, A.; Weston, J.; Yakhnenko O. (2013); Translating embeddings for modeling multi-relational data, *Proceedings of NIPS*, 2787–2795, 2013.

[5] Cao, F.; Liu, B.; Park, D. (2013); Image classification based on effective extreme learning machine, *Neurocomputing*, 102, 90–97, 2013.

[6] Carlson, A.; Betteridge, J.; Kisiel, B.; Settles, B.; Hruschka, E.; Mitchell T. (2010); Toward an architecture for never-ending language learning, *Proceedings of AAAI*, 1306–1313, 2010.

[7] Gardner, M.; Talukdar, P.; Kisiel, B.; Mitchell, T. (2013); Improving learning and inference in a large knowledge-base using latent syntactic cues, *Proceedings of EMNLP*, 833–838, 2013.

[8] Gardner, M.; Talukdar, P.; Krishnamurthy, J.; Mitchell, T. (2014); Incorporating vector space similarity in random walk inference over knowledge bases, *Proceedings of EMNLP*, 833–838, 2014.

[9] Gardner, M.; Mitchell, T. (2015); Efficient and expressive knowledge base completion using subgraph feature extraction, *Proceedings of EMNLP*, 1488–1498, 2015.

[10] Getoor, L.; Taskar, B. (2007); Introduction to statistical relational learning, MIT press, 2007.

[11] Glassick, C.E.; Huber, M.T.; Maeroff, G.I. (2015); DBpedia - A large-scale, multilingual knowledge base extracted from Wikipedia, *Semantic Web*, 6, 167–195, 2015.

[12] Guo, S.; Wang, Q.; Wang, B.; Wang, L.; Guo, L. (2015); Semantically smooth knowledge graph embedding, *Proceedings of ACL*, 84–94, 2015.

[13] Hoffmann, R.; Zhang, C.; Ling, X.; Zettlemoyer, L.; Weld, D. (2011); Knowledge-based weak supervision for information extraction of overlapping relations, *Proceedings of ACL*, 541–550, 2011.

[14] Huang, G.; Wang, D.; Lan, Y. (2011); Extreme learning machines: a survey, *International Journal of Machine Learning and Cybernetics*, 2, 107–122, 2011.

[15] Huang, G.; Zhou, H.; Ding, X.; Zhang, R. (2012); Extreme learning machine for regression and multiclass classification, *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42, 513–529, 2012.

[16] Huang, G.; Zhu, Q.; Siew, C. (2006); Extreme learning machine: theory and applications, *Neurocomputing*, 70, 489–501, 2006.

[17] Lanckriet, G.; Cristianini, N.; Bartlett, P.; Ghaoui, L.; Jordan, M. (2004); Learning the kernel matrix with semidefinite programming, *Journal of Machine Learning Research*, 5, 27–72, 2004.

[18] Landwehr, N.; Kersting, K.; Raedt, L. (2005); nFOIL: Integrating naıve bayes and FOIL, *Proceedings of AAAI*, 795–800, 2005.

[19] Lao, N.; Mitchell, T.; Cohen, W. (2011); Random walk inference and learning in a large scale knowledge base, *Proceedings of EMNLP*, 529–539, 2011.

[20] Lao, N.; Minkov, E.; Cohen, W. (2015); Learning relational features with backward random walks, *Proceedings of ACL*, 666–675, 2015.

[21] Lao, N.; Subramanya, A.; Pereira, F.; Cohen, W. (2012); Reading the web with learned syntactic-semantic inference rules, *Proceedings of EMNLP*, 1017–1026, 2012.

[22] Lao, N.; Mitamura, T.; Mitchell, T.; Zuo, W. (2012); Efficient random walk inference with knowledge bases, *PhD Thesis*, 2012.

[23] Lavrac, N.; Dzeroski, S. (1994), Inductive logic programming, *Proceedings of Workshop on Logic Programming*, 146–160, 1994.

[24] Lee, K.; Man, Z.; Wang, D.; Cao, Z. (2013); Classification of bioinformatics dataset using finite impulse response extreme learning machine for cancer diagnosis, *Neural Computing and Applications*, 22, 457–468, 2013.

[25] Lin, Y.; Liu, Z.; Sun, M.; Liu, Y.; Zhu, X. (2015); Learning entity and relation embeddings for knowledge graph completion, *Proceedings of AAAI*, 2181–2187, 2015.

[26] Ma, C.; OuYang J.; Chen, H.; Ji, J. (2016); A novel kernel extreme learning machine algorithm based on self-adaptive artificial bee colony optimisation strategy, *International Journal of Systems Science*, 47, 1342–1357, 2016.

[27] Nickel, M.; Murphy, K.; Tresp, V.; Gabrilovich, E. (2015); A review of relational machine learning for knowledge graphs, *Proceedings of IEEE*, 104, 11-33, 2015.

[28] Nickel, M.; Tresp, V.; Kriegel, H. (2011); A three-way model for collective learning on multi-relational data, *Proceedings of ICML*, 809–816, 2011.

[29] Nickel, M.; Rosasco, L.; Poggio, T. (2016); Holographic embeddings of knowledge graphs, *Proceedings of AAAI*, 1955–1961, 2016.

[30] Niu, F.; Ré C.; Doan, A.; Shavlik, J. (2011); Tuffy: Scaling up statistical inference in markov logic networks using an rdbms, *Proceedings of the VLDB Endowment*, 4, 373–384, 2011.

[31] Quinlan, J. (1990); Learning logical definitions from relations, *Machine Learning*, 5, 239–266, 1990.

[32] Richardson, M.; Domingos, P. (2006); Markov logic networks, *Machine Learning*, 62, 107–136, 2006.

[33] Socher, R.; Chen, D.; Manning, C.; Ng, A. (2013); Reasoning with neural tensor networks for knowledge base completion, *Proceedings of NIPS*, 926–934, 2013.

[34] Su, L.; Yao, M. (2013); Extreme learning machine with multiple kernels, *Proceedings of ICCA*, 424–429, 2013.

[35] Suchanek, F.; Kasneci, G.; Weikum, G. (2007); Yago: a core of semantic knowledge, *Proceedings of WWW*, 697–706, 2007.

[36] Wang, Q.; Mao, Z. Wang, B.; Guo, L. (2017); Knowledge graph embedding: a Survey of approaches and applications, *IEEE Transactions on Knowledge and Data Engineering*, 2724–2743, 2017.

[37] Wang, W.; Mazaitis, K.; Cohen, W. (2013); Programming with personalized pagerank: a locally groundable first-order probabilistic logic, *Proceedings of CIKM*, 2129–2138, 2013.

[38] West, R.; Gabrilovich, E.; Murphy, K.; Sun, S.; Gupta, R.; Lin, D. (2014); Knowledge base completion via search-based question answering, *Proceedings of WWW*, 515–526, 2014.

[39] Zheng, W.; Qian, Y.; Lu, H. (2013); Text categorization based on regularization extreme learning machine, *Neural Computing and Applications*, 22, 447–456, 2013.

[40] Zong, W.; Huang, G. (2011); Face recognition based on extreme learning machine, *Neurocomputing*, 74, 2541–2551, 2011.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 13(1), 83-98, February 2018.

Factors Space and its Relationship with Formal Conceptual Analysis: A General View

H. Liu, I. Dzitac, S. Guo

Haitao Liu*
1. Institute of Intelligence Engineering and Mathematics
Liaoning Technical University, Fuxin 123000, China
2. College of Science
Liaoning Technical University, Fuxin 123000, China
*Corresponding author: liuhaitao@lntu.edu.cn

Ioan Dzitac
1. Aurel Vlaicu University of Arad
310330 Arad, Elena Dragoi, 2, Romania
ioan.dzitac@uav.ro
2. Agora University of Oradea
410526 Oradea, P-ta Tineretului 8, Romania,
idzitac@univagora.ro

Sicong Guo
1. Institute of Intelligence Engineering and Mathematics
Liaoning Technical University, Fuxin 123000, China
2. College of Science
Liaoning Technical University, Fuxin 123000, China
guosizong@tom.com

Abstract: Conceptual generation is a key point and basic problem in artificial intelligence, which has been probed in the Formal Concept Analysis (FCA) established by G. Wille. Factors Space (FS) is also a branch of cognition math initiated by P.Z. Wang at the end of last century, which has been applied in information processing with fuzzy concepts effectively. This paper briefly introduces the historic background of FS and its relationship with FCA. FS can be seen as a good partner of FCA on conceptual description and structure extraction; combining FCA with FS, we can get more clear and simple statements and more fast algorithms on conceptual generation.
Keywords: Factors Space (FS), Formal Concept Analysis (FCA), rough sets, conceptual generation, basic concept semi-lattice, fuzzy logic.

1 Introduction

There were three mathematical branches describing cognition and thought of human beings simultaneously: Formal Concept Analysis [32] (Wille 1982), Rough Sets [14] (Pawlak, 1982) and Factors space [29] (Wang, 1982). We call them the branches of Cognition Math since there may be not any mathematical branch clearly declaring its object is to describe cognition process before. Even though many mathematical branches, such as mathematical logic, probability, optimization have been applied widely in artificial intelligence, those branches just do AI jobs spontaneously by their own natural characteristics, while cognition math conscientiously does AI jobs, fitting the needs of the era.

Relying on the involution between intension and extension, R. Wille gave a mathematical definition to the concept and gave algorithms for the generation of conceptual lattice, and since then, the computer has done conceptual generation automatically. He had opened the door of

cognition math. Similarly, Z. Pawlak brought in the rough sets to do knowledge discovering for databases, and P. Z. Wang brought in the factors space to represent things and thinking.

The articles of Wille are very serious. Unfortunately, he focuses on the attribute value rather than the attribute name. The number of columns in his formal background table exponentially increases, and the algorithms on concept lattice in FCA encounters the trap of N-hard complexity. Z. Pawlak emphasizes attribute name rather than attribute value, making the number of columns of formal background table greatly reduced and turned into the table of information system in rough sets, which plays the theoretical foundation for relational database. Unfortunately, his theory is not deep enough to avoid N-hard trap in attribute induction. Pawlak School emphasizes attribute name, but is not really aware of the significance of the attribute name. What is the difference between attribute value and attribute name? Attribute values, such as tall, middle and short, are a group of varying states under the name Height. From the viewpoint of mathematics, an attribute value is a constant, while the name is a variable that represents each one of its states. Height is an important factor for basketball player selection since the selection limits the varying of height. For the same reason, it is important for weightlifter selection, while it is not important for student examination since there is no limit to its varying. To manifest the influence of height with respect to the talents selection, we would like to say that height is a factor. In this paper, the word factor stands for a name, which influences or causes something.

People engaged in database know FCA and RS very well, but they are not familiar with factors space when FS was presented by P. Z. Wang at 1982, which was used to do ontology and cognition rather than database. To make mathematical discussion on randomness vs fuzziness, Wang presented FS to represent the basic space in a probability field and represent the discussion universe in fuzzy field respectively, and discovered a dual relationship: A fuzzy set defined on universe U can be represented by a probability field in the power of U; fuzzy measures can be determined by the falling shadow of a random set. This is the core of falling shadow theory [18]. Wang has proved the existence and uniqueness theorem around the probability distribution in the power $P(U)$, which lays the foundation for the applications of fuzzy sets and other subjective non-additive measures. FS plays a very important role in the development of fuzzy mathematics in China. Since 2012, Wang has turned his target to data science and finds that by means of the supports of FS, we can unify and deepen the theories of FCA and RS. All subjects in FCA and RS can be briefly and clearly stated, and the mentioned N-hard traps can be resolved by faster algorithms. FS provides natural base for cognation, information and data science. In this paper, we will introduce and develop the theory of FS briefly. We will introduce the historic background of factors space theory in Section 2. What is factors and factors space? We will answer the questions in Section 3. The core idea of background relation of FS will be given in Section 4. In section 5, we will deepen the conceptual generation and its AI applications based on FS. Section 6 will draw a brief conclusion.

2 The historic background of factors space theory

In 1982, Wang presented the first paper of factors space to explore and compare two kinds of uncertainty: fuzziness and randomness. He used factors space to represent basic space in probability and the discussion universe in fuzzy sets respectively. He clearly answered questions about the sources, and distinguished and relationship of the two kinds of uncertainties. We will introduce the historic background of factors space in following sub-sections.

2.1 Basic space in probability

What is randomness? Why and when does it occur? Tossing a coin, we can get two outcomes, Head or Back randomly. To know why randomness occurs, we need to find out those factors that influence the tossing process, such as Shape of coin, Actions of fingers, Condition of desktop, Environmental influence and so on. They will impact the outcomes. We may make such a determinism hypothesis: When the states of all factors are fixed, the outcome will be determinate. If not, there must be some influential factors that have not been considered, and then the hypothesis should be held considering all missing factors. Each factor f has a state space $X(f)$ as a dimension; let F_o be the set of all factors, and they form the high-dimensional space $X(F_o)$, which is what we call the factors space. This hypothesis tells us that, if the considered set F_o of factors is sufficient and if the state of factors can be observed and controlled precisely, then there is no randomness; randomness occurs if the considered factors are not complete or even if complete, they cannot be observed and controlled precisely. Randomness is the uncertainty on the occurrence of event caused by the lack of conditional factors with their observation and controlling.

If we cannot observe a state $\mathbf{x}$ precisely into a single point but a subset C in $X = X(F_o)$, C is called the condition of experiment. Under the condition, we only know $\mathbf{x}$ is in C but don't know where it is. When C is not small enough and crosses the border of the opposite sides with respect to an event, the randomness occurs. Even though conditional factors are not sufficient, there exists causality rule between condition and result. Probability is generalized causality, which is the essential scale of the frequency of an event. The basic model of probability statistics is: Fix the circle, move the point, where circle stands for the condition C, and point stands for the point $\mathbf{x}$ in X.

Kolmogorov gave an axiomatic definition on probability field $(\Omega, \mathcal{A}, p)$, where Ω is the basic space, $\mathcal{A}$ is an σ-algebra on Ω, and the probability p is an additive measure defined on $\mathcal{A}$. Wang admires him so much that he defines the random variable as a mapping $\xi : \Omega \to R$; based on the definition, classical probabilities are carried by ξ and become probability distribution functions or distribution densities on real line R, and then probability becomes a modern mathematical branch. However, how can we define a random variable ξ, a mapping from Ω to R? Only factors space can describe a random variable by a determinate mapping, and the basic space Ω must be a factors space [17]. As a great mathematician, Kolmogorov has had the idea of factors space already.

Viewing the basic space Ω as a factors space, we can study probability from a new aspect: Focusing on the transformation between randomness and certainty. The core problem is how to decrease the randomness so that it is transferred into certainty?

We can divide the set F_o of factors definition into three subset $F_o = G_1 + G_2 + G^c$, where $G_1 + G_2$ is the set of considered factors, and G^c is the set of unconsidered factors, G_1 is the set of considered factors whose states can be observed and controlled with accuracy C, which is small enough so that it does not cross the borders of opposite side of an events. G_2 is the factors in G but not in G_1. Then we can have a formula

$$\xi = g_1(\mathbf{x}) + g_2(\mathbf{x}) + g^c \tag{1}$$

where g_1 is a function, g_2 is a probability density under the condition C, and g^c is a Gaussian distribution of white noise when all unconsidered factors are neglect-able.

Let ξ be the hitting score of a shooter. We need to calculate its mean $a = M\xi$ and the mean square error $\sigma = (\xi - M\xi)^{1/2}$. Passing through special trains, we can reduce the error σ, and then move the parameter a to be the real target. An important task of probability is how

to promote the hitting probability of a shooter throughout special trains on improving the two parameters a and σ respectively.

2.2 Discussion universe of a fuzzy set

L. A. Zadeh (1921-2017) was a great world-renowned computer scientist in the history [4]. To open a path to develop artificial intelligence from the mathematical base, he presented fuzzy sets in 1965 [36], deeply influencing the progress of AI applications in the world. A fuzzy subset of the universe U is defined by a mapping $\mu_A : U \to [0,1]$, called membership function of A. He encouraged Wang to make a study on the discussion universe U; while the selection of U is crucial for the fuzzy description. For example, "young" is a fuzzy concept. To see whether a man is young, it is very difficult to measure the membership under the factor Age, but is much easier if we consider more factors such as "face", "action", "vigor". P.Z. Wang treated the discussion universe U as a factors space and found that fuzziness is the uncertainty on the conceptual extension caused by the lack of cognitive factors [18]. Even though cognitive factors are not sufficient, there exists the law of excluded middle in recognition still. Membership degree reflects the generalized law of excluded middle, which essentially scales the frequency of covering.

Lacked factors are essentially the objective factors in randomness but essentially the subjective factors in fuzziness. The basic model of probability statistics is: Fix the circle, move the point; the basic model in fuzzy statistics is: Fix the point, move the circle. Where the point stands for a state in factors space X, the circle stands for a subset in X. This is a duality. A circle of U is a point in the power $P(U)$; while a point $\mathbf{x}$ in U can be transferred a circle $\mathbf{x}^* = \{C | C \ni \mathbf{x}\}$ in $P(U)$. Therefore, we get a dual principle between fuzziness and randomness: A fuzzy experience model in the ground U can be transferred to a probability experience model in the sky $P(U)$. According to the principle of duality, Wang presented the falling shadow theory, which defines the membership degree of A as the covering probability of the related random set ξ:

$$\mu_A(\mathbf{x}) = p(\xi \ni x) \tag{2}$$

Falling shadow theory is the population theory for fuzzy statistics (interval statistics, set-valued statistics), which is a special contribution of factors space for fuzzy sets theory praised by the founder L. A. Zadeh [36].

Factors space theory gave subjective measures, including fuzzy measure, a united statement and gave a deep theorem: For any subset A of U, set $A^* = \{C | C \supseteq A\}$, ${}^*A = \{C | C \subseteq A\}$ and called the idea and filter of A respectively. Set $(A^*)^c = \{C | C^c \cap A \neq \emptyset\}$, $({}^*A)^c = \{C | C \cap A^c \neq \emptyset\}$. Let (X, T) be a topological space, denote $T_1 = \{A^* | A \in T\}$, $T_2 = \{{}^*A | A \in T\}$, $T_3 = \{(A^*)^c | A \in T\}$, $T_4 = \{({}^*A)^c | A \in T\}$, and let $\mathbf{T}_i (i = 1, 2, 3, 4)$ be the hyper-topologies generated by T_1, T_2, T_3 and T_4 respectively. $(P(T), \mathbf{T}_i)(i = 1, 2, 3, 4)$ are hyper-topological spaces on the power of $i = 1, 2, 3, 4$, let $\mathbf{A}_i$ be the σ-algebra generated from $\mathbf{T}_i$ respectively, $(P(T), \mathbf{A}_i)(i = 1, 2, 3, 4)$ are called four basic hyper-measurable spaces on the power $P(T)$ respectively. Let $\mathbf{p}$ be a probability defined on $(P(T), \mathbf{A}_i)$, for any $A \in T$ set

$$b(A) = \mathbf{p}(A^*), n(A) = \mathbf{p}({}^*A), an(A) = \mathbf{p}(({}^*A)^c), ab(A) = \mathbf{p}((A^*)^c) \tag{3}$$

where b, n, an and ab are known as the belief, plausibility, anti-plausibility and anti-belief measures respectively. They are non-additive measures.

Theorem 1. (Extension and Uniqueness Theorem) *Given any one of the four non-additive measure on A, there exists one and only one probability $\mathbf{p}$ on $(P(T), \mathbf{A}_i)$ such that Eq.(2) is held.*

Proof is given in [18].

Why did Wang use factors space to study on randomness vs randomness? He wants to develop cognition math. Fuzzy sets theory brings fuzzy concepts into math so that the programs can do smart recognition and control like brain of human being; fuzzy sets and fuzzy logic open a new door of mathematics toward AI applications. Factors space can deepen the researches of fuzzy sets. Treating the discussion universe U as a factors space, we can select factors to make a concept clearer and decrease its fuzziness, and we can calculate membership degree more reasonably.

Fuzzy sets represent cognition and thinking from the aspect of extension only, but lack the aspect of intension. Factors space is a plate to combine extension and intension, which can represent things and thinking. H. X Li and others use factors space to do fuzzy computer [25] and fuzzy controller with 4 stages inverse pendulum and factor space has been successfully applied to AI in China.

3 Factors and factors spaces

Based on the ideological background from history, we introduce and develop the factors theory. A lot of theoretical papers before 2012 can be found in the reference [3, 6–11, 13, 16, 19, 20, 24, 26, 31, 34, 35]. Since 2012, FS has turned to the data science [1, 2, 12, 21–23, 27, 28, 30, 37, 38].

We have mentioned that a factor causes something. But according to the interpretation of dictionary, factor is the reason causing something or the essence of thing, which tells us that a factor essences the very thing. What is the thing? Human being has accumulated a lot of knowledge on the generation of things, and the perfect model occurs in biology. Gene is the key to opening the door of biology, which induces the discovery of DNA, the mystery of life. Gene was called the factor by G. Mendel; factor is the generalization of gene, which is the key to opening the door of ontology and cognition, said Wang, who gives special emphasize from philosophy: Anything is united in the opposite of quality and quantity; factor is the root of quality, as gene is the root of biological attributes. Without the string of factors, attributes are like broken strings of pearls sprinkled over a floor [28]. By means of factors, things are organized and concepts are generated.

The group of states bunched by factor f forms a dimension $X(f)$ called the state space of f. Any factor must be defined on a discussion universe U. A factor f can be mathematically defined as a mapping $f:\ U \to X(f)$.

Factor is the aspect we observe and focus on; factors can be operated from rough to fine or from fine to rough, reflecting the analysis process and the synthesis process in thinking respectively. Consider a group of simple factors $f_1, \cdots, f_n$ defined on U, and they combine to a complex factor denoted as $F_o = \{f_1, \cdots, f_n\} = f_1 \vee \cdots \vee f_n$. The state space of F_o can be defined as the Cartesian product space $X(F_o) = \prod_i X(f_i)$. Each simple factor f_i performs analysis, which maps an object u in U to $f_i(u)$ in $X(f_i)$; the complex factor F_o performs synthesis and get the coordinate of u in $X(F_o)$. For example, consider Height(f_1), Weight(f_2), Age(f_3) and Sex(f_4) defined on a group of people U. Suppose that u=John in U, f_1(John)=1.75m, f_2(John)=75kg, f_3(John)=25years, f_4(John)=Male, then, we get that F_o(John)=(1.75m, 75kg, 25years, Male). John can be represented as a point in the coordinate system. Factors space is a generalized coordinate system of things.

Definition 2. Let $F = (F; \vee \wedge, \mathbf{0})$ be a lattice with minimum $\mathbf{0}$, $X_F = \{X(f)\}_{(f \in F)}$ is a family of sets satisfying
(1) $X(0) = \{\emptyset\}$;
(2) For any irreducible subset T of F ($s, t \in T$ implies $s \wedge t = 0$), we have that

$$X(T) = \prod_{f \in T} X(f) \tag{4}$$

where $\prod$ stands for Cartesian product. $\psi = (U, X_F)$ is called a factors space on U, if for any $u \in U$ and $f \in F$, there exists an unique $f(u) \in X(f)$. (i.e., each $f \in F$ is a mapping $f : U \to X(f)$). A member f in F is called a factor, and $\mathbf{0}$ is called empty factor. The operations $\vee$ and $\wedge$ are called the synthesis and decomposition of factors respectively.

We often limit the definition that $F_o = \{f_1, \cdots, f_n\} = f_1 \vee \cdots \vee f_n$, and $F = P(F_o)$, F is the power of F_o. In this case, the operations $\vee$ and $\wedge$ become to $\cup$ and $\cap$ respectively.

This definition is a relaxed vision given here. Originally, the index set was defined as a Boolean algebra $F = (F; \vee \wedge, \neg, \mathbf{0}, \mathbf{1})$ [26], which is too strong for our study. Of course, we often treat F as a Boolean algebra if it is needed.

Remark 3. About the name "Factors Space", we need to emphasize here: our definition is different from the factor space $X_{/\sim}$ representing the set of classes classified by the equivalent relation $\sim$. From now on, we use the word Factors, not Factor, in writings; But, we keep using the word in factor analysis initiated by L. L. Thurston [15]. As the great psychologist in 1931, even though he did not take high mathematical description into cognitive psychological measurement, he raised the banner of factors more than half a century earlier as our pioneer. Today, factors space inherits his target forward.

Factors space is a generalization of Cartesian coordinate space. Any quantitative Cartesian space can be viewed as a factors space since every variable is a factor; factors space extends those coordinate spaces from fixed dimension to variable dimension. The main breakthrough of factors space is that the states of factor can be changed from quantitative to qualitative. There are two types of states of a factor; one is quantitative; for example, the state of height can be measured as a real number in the interval [10, 250] (cm); the other type is qualitative; the state is taken as a word in natural language, tall, middle or short. In most cases, the two types of states can be employed simultaneously, and form a pair of state space $X(f)$ and $\underline{X}(f)$, the former one stands for qualitative and the latter one the quantitative State space. In the AI applications, we employ qualitative on semantic description and quantitative on formulae calculation. The most important point is how to transfer each other. Fuzzy sets theory provides technique to do such a kind of transformations, which enables FS to build its coordinate system by available axes.

4 Background set and correlation distribution

In the Section, we need to introduce the core concept of factors space.

Definition 4. Given a factors space $\psi = (U, X_F)$ with $F = P(F_o)$ and $F_o = f_1 \vee \cdots \vee f_n$, denote

$$R = F_o(U) = \{(x_1, \cdots, x_n) | \exists u \in U; x_1 = f_1(u), \cdots, x_n = f_n(u)\}. \tag{5}$$

Which is called the background set of ψ.

The background set R is the image of universe U in $X(F_o)$, which can be understood the real Cartesian product space of $f_1, \cdots, f_n$. Outside of R, the state-configuration of the n factors is not possible to be realized, so that R can be seen as the state space of the combined factor F_o. When $f_1, \cdots, f_n$ are independent, we have that $R = X(f_1) \times \cdots \times X(f_n)$; otherwise, there are correlations existing between those factors. The background relation R is also called the correlation of factors $f_1, \cdots, f_n$.

Correlation of factors is the source of factors space theory, which generates causality, logic, concept generation and decision making. Correlation of factors is the foundation of cognition math. A basic theorem was a proposal in paper [23].

We give an intuitive proof as follows:

Theorem 5. *Let R be the correlation of f and g, set $X = X(f)$ and $Y = X(g)$. For any $E \subseteq X$ and $E' \subseteq Y$, the inference "If x is in E then y is in E'" is identically true if and only*

$$(E \times Y) \cap R \subseteq (X \times E') \cap R. \tag{6}$$

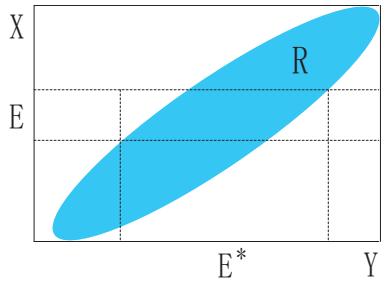


Figure 1: Background relation determines inference

Proof: The proposition "x is in E" is equivalent to "x is in E and y is in Y" and "(x, y) is in $E \times Y$". Since (x, y) is in R always, The proposition "x is in E" is equivalent to "(x, y) is in $(E \times Y) \cap R$". Similarly, the proposition "y is in E'" is equivalent to "(x, y) is in $(X \times E') \cap R$". The inference "If x is in E then y is in E'" is true if and only $(E \times Y) \cap R \subseteq (X \times E') \cap R$ □

Since the factors f and g can be complex factors in high dimension space, theorem 5 can be used in general cases, which tells us that the correlation of factors determines all inferences between those factors.

It is obvious that the background set is indeed the formal background in FCA. FS theory further underscoring the significance of the formal background emphasized by Wille.

So far, we get background set as a theoretical base with determinacy. But, it has to be extended to handle uncertainty from the following two reasons:

1) An object u in U is not a point but a granule, which may be a complex system. According to the viewpoint of granular computing [33], the size of granule can be varying in hierarchical model. For example, u was a man, the factor $f = height$ maps u to a determinate state $f(u) = Tall$ in $X(f)$, but, when the size of granule u is changed from a man to a group of men, what is the state $f(u)$? Here comes the uncertainty.

2) U is a theoretical term, which is not available in practice. For example, consider that U consists of all people. Where to identify a person? In the East or the West? When to identify a person? Include those who have been died or will be born? In practice, the universe is a sampling of U with randomness.

From the two reasons, we need to treat u as random variable and extend background set to background distribution relatedly.

Definition 6. Given a factors space $\psi = (U, X_F)$ with $F = P(F_o)$ and $F_o = f_1 \vee \cdots \vee f_n$, and suppose that $X(f_i) = \{a_{i1}, \cdots, a_{ik(i)}\}\,(i = 1, \cdots, n)$ are qualitative state spaces both. Let p be the probability carried by factors from random variable u. Denote

$$\begin{array}{l} r(x_1, \cdots, x_n) = p(f_1(u) = x_1, \cdots, f_n(u) = x_n) \\ (\sum \{b(x_1, \cdots, x_n) | x_i = a_{i1}, \cdots, a_{ik(i)} \;\; (i = 1, \cdots, n)\} = 1) \end{array} . \tag{7}$$

where r is called the background distribution of ψ, or the correlation distribution between factors $f_1, \cdots, f_n$.

The definition can be extended to quantitative state spaces.

5 Factors space in concept generation

Concept generation is a key point and a basic topic in cognation math. We only introduce the topic on concept generation in the paper.

5.1 General target of AI

Suppose that there are some intelligent problems about bodies of water treated by hydrologists. There are 21 objects as follows: 1. Tarn, 2. Trickle, 3. Rill, 4. Beck, 5. Rivulet, 6. Runnel, 7. Brook, 8. Bum, 9. Stream, 10. Torrent, 11. River, 12. Canal, 13. Lagoon, 14. Lake, 15. Mere, 16. Plash, 17. Pond, 18. Pool, 19. Puddle, 20. Reservoir, 21. Sea. We take those terms as objects to form a set U. The tasks of AI are treating objects according to a general target: Picking up information and transferring to the knowledge. Even though there are only 21 objects, each of them is not a piece of body in the land, but has been classified as a word in hydrology, and they are rambling too. The first task of AI is organizing objects into a conceptual system: From tousle to intellect.

5.2 Using factor to distinguish and classify objects

Concepts come from comparing and distinguish objects; it could not be done without factors.

Definition 7. We call object u and v in U are different if there is a factor f defined on U such that

$$f(u) \neq f(v). \tag{8}$$

Otherwise, they are called the same with respect to factor f.

Any two things in the world have the same points and different points simultaneously, without factors, we couldn't judge if two things are same or different.

A factor f is a mapping defined on the universe U, which determines an equivalent relation $\sim$ on U:

$$u \sim v \text{ if and only if } f(u) = f(v).$$

The equivalent relation determines a classification on U; if we consider only qualitative state space here, then the classified factorial space can be denoted as

$$U_{/f} = \{C_k = (u_{k1}, \cdots, u_{kn(k)})\}_{k=1,\cdots,K} \tag{9}$$

where the number of states of $X(f)$ is $m = n(1) + \cdots + n(K)$, $n(j)$ is the number of objects in k^{th} subclass.

Definition 8. [38] Denote

$$c_f = n(1)(n(1)-1) + \cdots + n(K)(n(K)-1)/m(m-1) \tag{10}$$

Which is called the distinguish degree of factor f with respect to U.

The bigger the distinguish degree, the finer the division taken by f on U.

Table 1: Factorial table for bodies of water

U	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
f_1	H	N	N	N	N	N	N	N	N	N	N	A	N	N	A	N	A	N	N	A	N
f_2	M	R	R	R	R	R	R	R	R	R	R	R	M	M	M	M	M	M	M	M	M
f_3	S	S	S	S	S	S	S	S	S	S	S	S	I	S	S	S	S	S	S	S	I
f_4	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	T	C	C	T	C	C

5.3 Atomic concepts

As for the bodies of water, to pick up information and organize objects into a conceptual system, we need to consider some factors: Is the body natural or artificial? How is its manner, running or muddy? Is the water body in sea or in land? Is it constant or temporary? And so on. Those factors are often the names of attributes, but factors do not play the role of passive descriptors but initiative leaders. In the bodies of water, we can define that $f_1 = Natural?$, $X(f_1) = \{Natural, Artificial\}$; $f_2 = Running?$, $X(f_2) = \{Running, Muddy\}$; $f_3 = Sea - inland$, $X(f_3) = \{Sea, Inland\}$; $f_4 = Timeliness$, $X(f_4) = \{Constant, Temporary\}$. Then we can get a factor space $\psi = (U, X_F)$ with $F_o = f_1 \vee f_2 \vee f_3 \vee f_4$, The factor tableau can be taken as the sampling in $X = X(f_1) \times X(f_2) \times X(f_3) \times X(f_4)$. Citing the data from paper [5], we get Table 1.

How to analyze on the tableau? Given a factors space $\psi = (U, X_F)$ with $F_o = f_1 \vee \cdots \vee f_n$, for any $[\mathbf{a}] \in X(F_o)$, denote

$$[\mathbf{a}] = F_o^{-1}(\mathbf{a}) = \{u | F_o(u) = (f_1(u), \cdots, f_n(u)) \in \mathbf{a}\} \tag{11}$$

If $[\mathbf{a}] \in R$, i.e., $[\mathbf{a}] \neq \emptyset$, we say that $\mathbf{a}$ is an atomic intension, or say a is a real configuration. The collection of all the atom intensions form the division $U_{/F_o} = \{[\mathbf{a}]\}$, which includes all subclasses divided by factors $f_1, \cdots, f_n$. It is clear that F_o is an isomorphic mapping from $U_{/F_o}$ to R.

Definition 9. Given factors space $\psi = (U, X_F)$ with $F = P(F_o)$ and $F_o = f_1 \vee \cdots \vee f_n$, for any $\mathbf{a} \in R$, the pair $\alpha = (\mathbf{a}, [\mathbf{a}])$ is called an atomic concept and $[\mathbf{a}]$, $\mathbf{a}$ are called the extension and intension of α respectively.

Return to the bodies of water, the Cartesian product space $X(f_1) \times X(f_2) \times X(f_3) \times X(f_4)$ concludes $2 \times 2 \times 2 \times 2 = 16$ elements, but, the real product X_F, i.e., the background set R includes 6 elements only

R={NMSC, NRSC, ARSC, NMIC, AMSC, NMST}

According to our definition, there are six atomic concepts:

α_1=(NMSC, {1,14,18}), α_2=(NRSC, {2,3,4,5,6,7,8,9,10,11}), α_3=(ARSC, {12}),
α_4=(NMIC, {13,21}), α_5=(AMSC, {15,17,20}), α_6=(NMST, {16,19}).

5.4 How important are the atomic concepts?

Atomic concepts have 4 basic functions in AI:
1. Atomic concepts provides the finest classification of U under factors $f_1, \cdots, f_n$.

Theorem 10. *Given factors space $\psi = (U, X_F)$ with $F = P(F_o)$ and $F_o = f_1 \vee \cdots \vee f_n$, each atomic concept $\alpha = (\mathbf{a}, [\mathbf{a}])$ is not able to be refined (i.e.,$[\mathbf{a}]$ is not able to be divided) by the factors $f_1 \vee \cdots \vee f_n$.*

Proof: For any atomic extension [**a**] and any $u, v \in [\mathbf{a}]$, we have that $f_i(u) = [\mathbf{a}] = f_i(v)$ for $i = 1, \ldots, n$, so that $u \sim v$ for $i = 1, \ldots, n$. □

The smaller the number of atomic concepts with respect to the number of objects in U, the simpler and clearer the structure of knowledge is but the rougher the classification is. To break the limit of finest division, we need to find out more factors, and this is a deepening process and will be discussed in the future.

2. The finest classification provides the universal code for knowledge organization, retrieval and inquiring. Of course, the code is edited under the guidance of factors.

3. An answer system can answer following questions based on atomic concepts:

Is Rill α_1? No, since α_1=(NMSC, {1,14,18}) does not includes Rill (3).

Does α_2 include Rill? Yes, since α_2=(NRSC, {2,3,4,5,6,7,8,9,10,11}) includes Rill.

Does Rill have the attribute A? No, since $fi(Rill) = A \neq N$.

What is the concept an object with attributes N and S belonging to? It may α_1, α_2 or α_4 be.

4. All atomic concepts form the background relation or the correlation R. What is the shape of R? What is the causality generated from R? There comes a lot of functions applied in AI, we will discuss in the future.

5.5 Basic concepts semi-lattice

Definition 11. Given a factors space $\psi = (U, X_F)$ with $F = P(F_o)$ and $F_o = f_1 \vee \cdots \vee f_n$, for any $A \subseteq R$, denote $[A] = \cup\{[\mathbf{a}] | \mathbf{a} \in A\}$, For any $A \subseteq R$, $\gamma = (A, [A])$ is called a concept with the intension A and extension $[A]$. Set $\Gamma = \{\gamma = (A, [A]) | A \subseteq R\}$, $\Gamma = (\Gamma, \mathrm{OR}, \mathrm{AND})$ is called the conceptual lattice on $\psi = (U, X_F)$, Where $\gamma \mathrm{OR} \gamma' = (A \mathrm{OR} A', [A] \cup [A'])$ and $\gamma \mathrm{AND} \gamma' = (A \mathrm{AND} A', [A] \cap [A'])$; even $\Gamma = (\Gamma, \mathrm{OR}, \mathrm{AND}, \mathrm{NOT})$ is called the conceptual Boolean algebra on $\psi = (U, X_F)$, Where NOT $\gamma = (\mathrm{NOT}\ A, [A]^c)$.

Do not be afraid that the generated concepts are too little; we are afraid that too much. If there are k atomic concepts, then 2^k new concepts will be generated. To avoid the information overflow, we generate basic concepts only.

Definition 12. A concept is called basic if its intension can be stated in a conjunctive normal form.

Intuitively speaking, the extension of a basic concept can be mapped to a hyper-rectangle within R in $X(F_o)$.

In bodies of water, each atomic concept is a basic concept. The concept γ=(NMC, {1,14,18,13,21}) is a basic concept since its intension can be written to the conjunctive normal form $\mathrm{N} \wedge \mathrm{M} \wedge (\mathrm{S} \vee \mathrm{I}) \wedge \mathrm{C} = \mathrm{N} \wedge \mathrm{M} \wedge X(f_3) \wedge \mathrm{C} = \mathrm{N} \wedge \mathrm{M} \wedge \mathrm{C}$.

How to get the basic concept semi-lattice? We introduce Wang's algorithm [21] by the example of bodies of water.

1) Calculating distinguish degree c_f of each factor and then changing the order of columns to do f-classification on U, where f holds the maximal distinguish degree c_f.

$m = 22$

f_1: $n(\mathrm{N}) = 16, n(\mathrm{A}) = 5, c_{f_1} = 1 - {}^{(16 \times 15 + 5 \times 4)}/_{21 \times 20} = 0.4$;

f_2: $n(\mathrm{R}) = 11, n(\mathrm{M}) = 10, c_{f_2} = 1 - {}^{(11 \times 10 + 10 \times 9)}/_{21 \times 20} = 0.5$;

f_3: $n(\mathrm{S}) = 19, n(\mathrm{I}) = 2, c_{f_3} = 1 - {}^{(19 \times 18 + 2 \times 1)}/_{21 \times 20} = 0.2$;

f_4: $n(\mathrm{C}) = 19, n(\mathrm{T}) = 2, c_{f_4} = 0.2$;

$c_{f_2} > c_{f_1} > c_{f_3} = c_{f_4}$.

U	12	2	3	4	5	6	7	8	9	10	11	1	13	14	15	16	17	18	19	20	21
f_1	A	N	N	N	N	N	N	N	N	N	N	H	N	N	A	N	A	N	N	A	N
f_2	R	R	R	R	R	R	R	R	R	R	R	M	M	M	M	M	M	M	M	M	M
f_3	S	S	S	S	S	S	S	S	S	S	S	S	I	S	S	S	S	S	S	S	I
f_4	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	T	C	C	T	C	C

We get the f_2-classification $U = U_1\{2,3,4,5,6,7,8,9,10,11,12\}+U_2\{1,13,14,15,16,17,18,19,20,21\}$.

Repeatedly, do classification on U_1 and return to step 1):

U	12	2	3	4	5	6	7	8	9	10	11
f_1	A	N	N	N	N	N	N	N	N	N	N
f_2	R	R	R	R	R	R	R	R	R	R	R
f_3	S	S	S	S	S	S	S	S	S	S	S
f_4	C	C	C	C	C	C	C	C	C	C	C

m=11

$f_1 : n(\mathrm{N}) = 10, n(\mathrm{A}) = 1, c_{f_1} = 1 - {}^{(10\times 9+1\times 0)}/_{11\times 10} = 9/11;$

$f_2 : n(\mathrm{R}) = 11, n(\mathrm{M}) = 0, c_{f_2} = 1 - {}^{(11\times 10)}/_{11\times 10} = 0;$

$f_3 : n(\mathrm{S}) = 11, n(\mathrm{I}) = 0, c_{f_3} = 1 - {}^{(11\times 10)}/_{11\times 10} = 0;$

$f_4 : n(\mathrm{C}) = 11, n(\mathrm{T}) = 0, c_{f_4} = 1 - {}^{(11\times 10)}/_{11\times 10} = 0.$

$c_{f_1} > c_{f_2} = c_{f_3} = c_{f_4}$

We get the f_1-classification $U_1 = U_{11}\{12\} + U_{12}\{2,3,4,5,6,7,8,9,10,11\}$.

Similarly, we can repeat the step 1) on other sub-classes. Finally, we get a tree with following branches

$U_2 = U_{21}\{1,13,14,16,18,19,21\} + U_{22}\{15,17,20\};$

$U_{21} = U_{211}\{13,21\} + U_{212}\{1,14,16,18,19\};$

$U_{212} = U_{2121}\{1,14,18\} + U_{2122}\{16,19\}.$

The figure of the tree is drawn in Fig.2.

The complexity of basic concept semi-lattice is $\mathrm{O}(m^2n)$, where m and n are the number of rows and columns of factorial tableau respectively.

In Fig.2, each node is a basic concept, and the node at the top is the original concept with extension U. When a baby born, the origin is zero concept (Empty, U), No intension statement, the extension is the chaotic universe. From top to bottom, the intension increases and the extension shrinks. The undermost node is not a concept but an ideal extreme with empty extension. We add it in the Fig.2 to match the original figure drawn by Ganter and Wille.

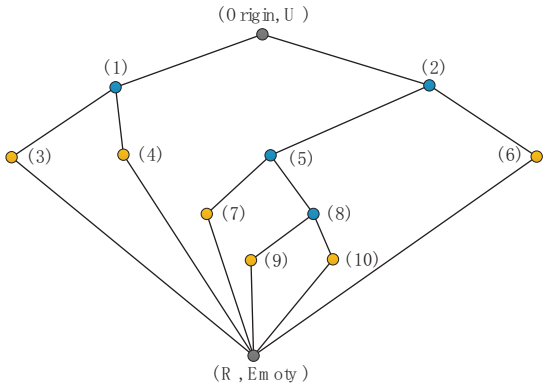


Figure 2: Semi-lattice of basic concepts

Theorem 13. *The operation OR may not be closed for two basic concepts except that the intension of two concepts are different at only one factor.*

Proof is obvious, we only hint that γ_1 OR $\gamma_4 = (\text{NMSCORNMIC}, \{1, 14, 18\} \cup \{13, 21\}) = (\text{NMC}, \{1, 14, 18, 13, 21\})$, which is the basic concept γ mentioned above. Performing the operation OR on two basic concepts, we get a basic concept. What is the condition of the result? The intensions of the two basic concepts are NMSC and NMIC, and their difference is on the factor f_3 only. Without the condition, the operation of OR is not closed in basic concepts. Be careful, all basic concepts form a semi-lattice, but may not be a lattice.

5.6 The consistence with formal concept analysis

Concept generation was established by Wille, who defined the formal background as $K = (G, M, I)$, where G is a group of objects, M a group of attribute values, I a relation from G to M : $(g, m) \in I$ implies that object g holds attribute value m. For any $A \subseteq G$, define $s(A) = \{m \in M | \forall g \in A; (g, m) \in I\}$, and for any $B \subseteq M$, define $t(B) = \{g \in G | \forall m \in B; (g, m) \in I\}$. $\alpha = (B, A)$ is called a concept if the fallowing convolution principle is holds:

$$s(A) = B, \quad t(B) = A \tag{12}$$

Theorem 14. *Given a factors space $\psi = (U, X_F)$ with $F = P(F_o)$ and $F_o = f_1 \vee \cdots \vee f_n$, for any basic concept $\alpha = (A, [A])$, $[A]$ and A satisfy the convolution principle.*

Proof: Let $\alpha = (\mathbf{a}, [\mathbf{a}])$ be an atomic concept. Taking $G = U$, $M = X(F_o)$ and $I = \{(u, \mathbf{a}) | \mathbf{a} = F_o(u)\}$, we have

$$s([\mathbf{a}]) = \{m \in M | \forall g \in [\mathbf{a}]; (g, m) \in I\} = \{F_o(u) | u \in [\mathbf{a}]\} = \mathbf{a};$$
$$t([\mathbf{a}]) = \{u \in U | (u, \mathbf{a}) \in I\} = \{u \in U | F_o(u) = \mathbf{a}\} = [\mathbf{a}].$$

Hence $[\mathbf{a}]$ and $\mathbf{a}$ satisfy convolution Eq.(12).

Let $\alpha = (A, [A])$ be a basic concept. We have

$$s([A]) = \{A \subseteq X(F_o) | \forall u \in [A]; \ F_o(u) \in A = A\};$$
$$t(A) = \{u \in U | F_o(u) \in A\} = [A].$$

Hence $[A]$ and A satisfy convolution Eq.(12). □

The proposition proves that the definition of basic concepts in FS is consistent with the original definition in FCA. It does also hint that general concepts may not satisfy the convolution principle.

What is the relationship between Table 1 and the original Table of Wille? The original table shows a formal background with 22 objects and 8 attribute values. Table 1 replaces the 8 attribute values by 4 factors, and since the object *channel* has no attribute value in the original table, it has been taken out of the table.

As shown in Fig.2, apart from the top and bottom nodes, there are 8 red atomic and 4 blue non-atomic basic concepts. To be easy to match to Fig.3, we number the nodes as follows:

Node	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$[A]$	U_1	U_2	U_{11}	U_{12}	U_{21}	U_{22}	U_{211}	U_{212}	U_{2121}	U_{2122}

It is obvious that nodes (3),(4),(6),(7),(9),(10) are atomic, color red, nodes (1),(2),(5),(8) are non-atomic basic concepts, color blue. We can see that Fig.2 is consistent with Fig.3. Indeed, the semi-lattice drawn in Fig.2 is a sub-lattice drawn in Fig.3.

	natural	artificial	stagnant	running	inland	maritime	constant	temporary
tarn	×		×		×		×	
trickle	×			×	×		×	
rill	×			×	×		×	
beck	×			×	×		×	
rivulet	×			×	×		×	
runnel	×			×	×		×	
brook	×			×	×		×	
bum	×			×	×		×	
stream	×			×	×		×	
torrent	×			×	×		×	
river	×			×	×		×	
channel				×	×		×	
canal		×		×	×		×	
lagoon	×		×			×	×	
lake	×		×		×		×	
mere		×	×		×		×	
plash	×		×		×			×
pond		×	×		×		×	
pool	×		×		×		×	
puddle	×		×		×			×
reservoir		×	×		×		×	
sea	×		×			×	×	

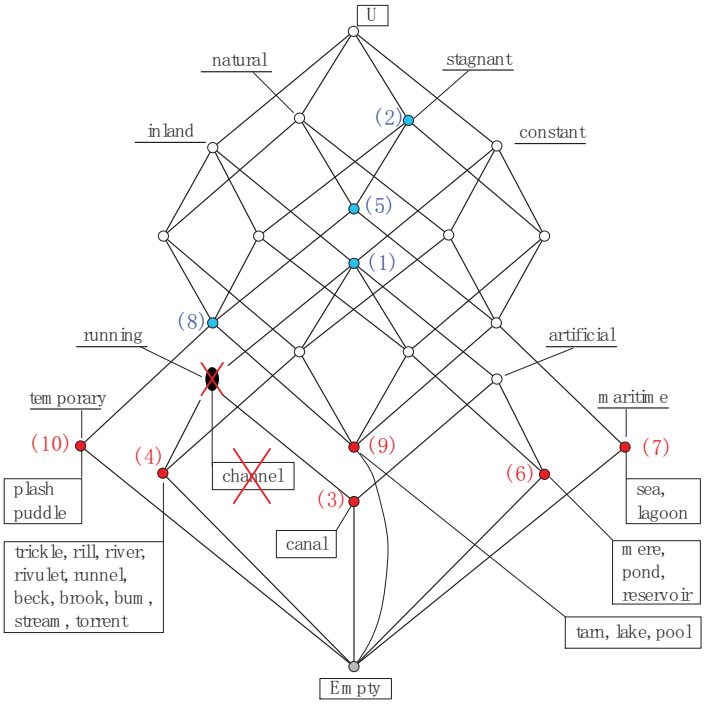


Figure 3: Conceptual lattice(Ganter & Wille [5]).

In practical applications, the generated basic concepts are too much, we need to review those concepts by experts, rarely select those concepts who can be understood by human being, and then rename them into a special database. Fig.2 decreases the number of non-atomic basic concepts from 15 in origin to 4 in Fig.2.

However, we can change the semi-lattice by changing the order of factor-divisions, and the union of that semi lattice can completely full the original 'concept lattice':

Theorem 15. *Given a factors space $\psi = (U, X_F)$ with $F = P(F_o)$ and $F_o = f_1 \vee \cdots \vee f_n$, for any basic concept $\gamma = (A, [A])$, there is a semi-lattice S generated by Wang's algorithm taking γ as a member in it.*

Proof: Let the intension A is written in the conjunction normal form: $A = A_{i(1)} \wedge \cdots \wedge A_{i(k)}$ $(k \leq n)$, where $A_{i(j)} = a_{i(j),j(1)} \vee \cdots \vee a_{i(j),j(r)}$ is a disjunction of atomic intensions with respect to the factor $f_{i(j)}(j = 1, , k)$. According to Wang's algorithm, doing classifications by factors $f_i(1), \cdots, f_i(k)$ successively, the basic concept must be occurred in the semi-lattice of basic concept. □

Basic concepts have 2 basic functions in AI:

1. A set of inference rules can be shown in the graph of semi-lattice.

Let S be the graph of a semi=lattice. From bottom to up, any node $\gamma = (A, [A])$ links at least one node $\gamma' = (A', [A'])$. It is obvious that the inference arrow $\gamma \Rightarrow \gamma'$ is constant true.

2. An answer system can answer the following questions based on basic concepts:

Is Pool in U_1? No, since Pool is α_1 numbered red 9, which links to blue 8, then links to blue 5, then links to blue 2, no way to link to blue 1. It does not belong to U_1.

Is Drill in U_1? Yes, since Drill is α_2 numbered red 4, which links to the node numbered blue 1, having extension U_1.

6 Conclusion

Factor space provides a new view point toward formal concept analysis, which promotes the signification of formal background. By means of the support of FS, all problems in FCA or RS can get more clear and simple statement, and the N-hard trap FCA and RS encountered will be overcome by faster algorithms. Concept generation is a key point and basic problem in artificial intelligence and factor space has got good progress on concept generation, which will be an important tool for cognition, information and data science.

Acknowledgement

The authors specially thank Professor P.Z. Wang, was a good friend and collaborator with the father of fuzzy sets: Lotfi A. Zadeh (1921-2017), for his guidance and valuable suggestions. This study was partially supported by the grants (Grant Nos. 61350003, 11401284, 70621001, 70531040) from the Natural Science Foundation of China, and the grant (Grant Nos. L2014133) from the department of education of Liaoning Province.

Bibliography

[1] Cheng, Q.F.; Wang, T.T.; Guo, S.C.; Zhang, D.Y.; Jing, K.; Feng, L.; Wang, P.Z. (2017); The Logistic Regression from the Viewpoint of the Factor Space Theory, *International Journal of Computers Communications & Control*, 12(4), 492–502, 2017.

[2] Cui, T.J.; Wang, P.Z.; Ma, Y.D. (2016); Structured representation methods for 01 space fault tree, *J Dalian Jiaotong Univ*, 37(1), 82–87, 2016.

[3] Dzitac, I. (2015), The Fuzzification of Classical Structures: A General View, *International Journal of Computers Communications & Control*, 10(6), 772-788, 2015.

[4] Dzitac, I.;Filip, F.G. ; Manolescu, M.J. (2017); Fuzzy Logic Is Not Fuzzy: World-renowned Computer Scientist Lotfi A. Zadeh, *International Journal of Computers Communications & Control*, 12(6), 748-789, 2017.

[5] Ganter, B.; Wille, R. (1996); Formal concept analysis, *Wissenschaftliche Zeitschrift-Technischen Universitat Dresden*, 45, 8–13, 1999.

[6] Kandel, A.; Peng, X.T.; Cao, Z.Q.; Wang, P.Z. (1990); Representation of concepts by factor spaces, *Cybernetics and Systems: An International Journal*, 21(1), 43–57, 1990.

[7] Li, H.X.; Wang P.Z.; Yen, V.C. (1998); Factor spaces theory and its applications to fuzzy information processing.(I). The basics of factor spaces, *Fuzzy Sets and Systems*, 95(2), 147–160, 1998.

[8] Li, H.X.; Yen, V.C.; Lee, E.S. (2000); Factor space theory in fuzzy information processing-Composition of states of factors and multifactorial decision making, *Computers & Mathematics with Applications*, 39(1), 245–265, 2000.

[9] Li, H.X.; Yen, V.C.; Lee, E.S. (2000); Models of neurons based on factor space, *Computers & Mathematics with Applications*, 39(12), 91–100, 2000.

[10] Li, H.X.; Chen, C.P.; Yen, V.C.; Lee, E.S. (2000); Factor spaces theory and its applications to fuzzy information processing: Two kinds of factor space canes, *Computers & Mathematics with Applications*, 40(6-7), 835–843, 2000.

[11] Li, H.X.; Chen, C.P.; Lee, E.S. (2000); Factor space theory and fuzzy information processing-Fuzzy decision making based on the concepts of feedback extension, *Computers & Mathematics with Applications*, 40(6-7), 845–864, 2000.

[12] Liu, H.T.; Guo, S.C. (2015); Inference model of causality analysis, *Journal of Liaoning Technical University(Natural Science)*, 2015, 34(1), 124–128.

[13] Liu, Z.L. (1990); *Factorial Neural Networks*, Beijing Normal University Press, 1990.

[14] Pawlak, Z. (1982); Rough sets, *International Journal of Parallel Programming*, 11(5), 341–356, 1982.

[15] Thurstone L. L. (1931); Multiple factor analysis, *Psychological Review*, 38(5), 406–427, 1931.

[16] Vesselenyi, T.; Dzitac, I.; Dzitac, S.; Vaida, V. (2008); Surface roughness image analysis using quasi-fractal characteristics and fuzzy clustering methods, *International Journal of Computers Communications & Control*, 3(3), 304–316, 2008.

[17] Wang, P.Z. (1981); Randomness, *Advance of Statistical Physics*, Science and Technology Press, 1981.

[18] Wang, P.Z. (1985); *Fuzzy sets and falling shadows of random set*, Beijing Normal University Press, 1985.

[19] Wang, P.Z. (1990); A factor spaces approach to knowledge representation, *Fuzzy Sets and Systems*, 36(1), 113–124, 1990.

[20] Wang, P.Z. (1992); Factor space and concept description, *Journal of Software*, 1, 30–40, 1992.

[21] Wang, P.Z. (2013); Factor spaces and factor data-bases, *Journal of Liaoning Technical University (Natural Science)*, 32(10), 1–8, 2013.

[22] Wang, P.Z. (2015); Factors space and data science, *Journal of Liaoning Technical University (Natural Science)*, 34(2), 273–280, 2015.

[23] Wang, P.Z.; Guo, S.C.; Bao, Y.K.; Liu, H.T. (2014); Causality analysis in factor spaces, *Journal of Liaoning Technical University (Natural Science)*, 33(7), 1–6, 2015.

[24] Wang, P.Z.; Jiang, A. (2002); Rules detecting and rules-data mutual enhancement based on factors space theory, *International Journal of Information Technology & Decision Making*, 1(01), 73–90, 2002.

[25] Wang, P.Z.; Li, H.X. (1995); *Fuzzy system theory and fuzzy computer*, Publishing Company of Science, 1995.

[26] Wang, P.Z.; Li, H.X. (1994); *A mathematical theory on knowledge representation*, Tianjin Scientific and Technical Press, 1994.

[27] Wang, P.Z.; Liu, Z.L.; Shi, Y.; Guo, S.C. (2014); Factor space, the theoretical base of data science, *Annals of Data Science*, 1(2), 233–251, 2014.

[28] Wang, P.Z.; Ouyang, H.; Zhong, Y.X.; He, H.C. (2016); Cognition math based on factor space, *Annals of Data Science*, 3(3), 281–303, 2016.

[29] Wang, P.Z., Sugeno, M. (1982); The factor fields and background structure for fuzzy subsets, *Fuzzy Mathematics*, 2(2), 45–54, 1982.

[30] Wang, H.D.; Wang, P.Z.; Shi, Y.; Liu, H.T. (2014); Improved factorial analysis algorithm in factor spaces, *International Conference on Informatics*, 201–204, 2014.

[31] Wang, P.Z.; Zhang, X.H.; Lui, H.C.; Zhang, H.M., Xu, W. (1995); Mathematical theory of truth-valued flow inference, *Fuzzy Sets and Systems*, 72(2), 221–238, 1995.

[32] Wille, R. (1982); Restructuring lattice theory: an approach based on hierarchies of concepts, *Ordered sets*, Springer Netherlands, 445–470, 1982.

[33] Yao, Y. (2009); Three-Way Decision: An Interpretation of Rules in Rough Set Theory, *RSKT*, 9, 642–649, 2009.

[34] Yuan, X.H.; Wang, P.Z.; Lee, E.S. (1992); Factor space and its algebraic representation theory, *J Math Anal Appl.*, 171(1), 256–276, 1992.

[35] Yuan, X.H.; Wang, P.Z.; Lee, E.S. (1994); Factor Rattans, Category FR (Y), and Factor Space, *Journal of Mathematical Analysis and Applications*, 186(1), 254–264, 1994.

[36] Zadeh, L.A. (1965); Fuzzy sets, *Information and control*, 8(3), 338–353, 1965.

[37] Zeng, F.H.; Zheng, L. (2017); Sample cultivation in Factorial analysis, *Journal of Liaoning Technical University (Natural Science)*, 36(3), 320–323, 2017.

[38] Zeng, F.H.; Li, Y. (2017); An improved decision tree algorithm based on factor space theory, *Journal of Liaoning Technical University (Natural Science)*, 36(3), 109–112, 2017.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 13(1), 99-116, February 2018.

Dynamic Multi-hop Routing Protocol Based on Fuzzy-Firefly Algorithm for Data Similarity Aware Node Clustering in WSNs

M. Misbahuddin, A.A. Putri Ratna, R.F. Sari

Misbahuddin Misbahuddin*
Department of Electrical Engineering, Faculty of Engineering,
Universitas Indonesia, Depok, 16424, Indonesia
*Corresponding author: misbahuddin@ui.ac.id

Anak Agung Putri Ratna
Department of Electrical Engineering, Faculty of Engineering,
Universitas Indonesia, Depok, 16424, Indonesia
ratna@eng.ui.ac.id

Riri Fitri Sari
Department of Electrical Engineering, Faculty of Engineering,
Universitas Indonesia, Depok, 16424, Indonesia
riri@ui.ac.id

Abstract: In multi-hop routing, cluster heads close to the base station functionaries as intermediate nodes for father cluster heads to relay the data packet from regular nodes to base station. The cluster heads that act as relays will experience energy depletion quicker that causes hot spot problem. This paper proposes a dynamic multi-hop routing algorithm named Data Similarity Aware for Dynamic Multi-hop Routing Protocol (DSA-DMRP) to improve the network lifetime, and satisfy the requirement of multi-hop routing protocol for the dynamic node clustering that consider the data similarity of adjacent nodes. The DSA-DMRP uses fuzzy aggregation technique to measure their data similarity degree in order to partition the network into unequal size clusters. In this mechanism, each node can recognize and note its similar neighbor nodes. Next, K-hop Clustering Algorithm (KHOPCA) that is modified by adding a priority factor that considers residual energy and distance to the base station is used to select cluster heads and create the best routes for intra-cluster and inter-cluster transmission. The DSA-DMRP was compared against the KHOPCA to justify the performance. Simulation results show that, the DSA DMRP can improve the network lifetime longer than the KHOPCA and can satisfy the requirement of the dynamic multi-hop routing protocol.
Keywords: clustering, data similarity, multi-hop routing, fuzzy system, firefly algorithm, Wireless Sensor Networks (WSNs).

1 Introduction

Wireless Sensor Networks (WSNs) rapidly grow in various applications for many domains. Besides, WSNs is also an integral part of the Internet of Things (IoT) that can share data for improving the human capability in monitoring a local environmental condition and process automation. It consists of a set of sensor nodes that are deployed in an appropriate environment in the ad hoc model to observer and interact with the physical world or biological system remotely. Therefore, WSNs should be able to adapt dynamically with the charged environment. Recently, WSNs play important role in various necessaries of human such as flood monitoring [14], weather monitoring, earthquake detection [22], tracking [19], volcanic eruption, military necessity [10], healthcare observation [3] agriculture automation [26], and manufacturing automation [17].

Each node is composed of sensor, low ability processor, limited capacity storage and power supply, and transceiver. Therefore, the efficient usage of energy that is supplied by battery is

important issue in order to prolong the network lifetime. Topologically, the clustering techniques have been commonly used to improve the network performance such as prolonging the network lifetime, enhancing the network scalability, increasing the bandwidth efficiency, and increasing fault tolerance [1]. Clustering technique divides the nodes on a network into many logical or physical groups termed as clusters. Each cluster is composed of a node selected as Cluster Head (CH) and many regular nodes called cluster members. Each regular node senses data of the environmental condition and forwards to its CH. Meanwhile, the CH functions to sense data, aggregate data, and relay them to the other CH or Base Station (BS).

Clustering techniques consist of two fashions, equal sized clustering and unequal sized clustering. In equal sized clustering, all clusters have the same size number of cluster members. The CHs closer to BS have an additional function, not only sensing data, aggregating data, and sending the aggregated data to BS but also forwarding data from the other CHs to BS. These CHs have a heavier load than the CHs farther from BS, so that they consume more energy and deplete energy more quickly than the other CHs. Thus, the network connectivity is disrupted in relaying data to BS. This event is termed as a hot spot problem.

To overcome the hot spot issue in the network, the topology of unequal sized clustering can be used to organize the load balancing among the CHs [**?**]. Architecture of the unequal sized clustering is to reduce the clusters size closer to BS and increase the clusters size as the distance between CH and BS. In our work, load of clusters can not be arranged through such way because the cluster size is determined according to the clustering technique based on the data similarity referred spatial and temporal correlation. Therefore, such clustering technique requires a specific routing protocol to increase the energy efficiency in transmitting the sensed data by the regular nodes to BS via either a CH with a single-hop or some CH with multi-hop. Furthermore, this technique is also a dynamically changed clustering in each round. The topology of the network changes in each round because each cluster is established based on the data similarity of the adjacent nodes.

Because such clustering technique generates unequal sized clusters, the selection of some nodes as CHs is a crucial problem for improving the energy saving in order to prolong the network lifetime. This paper proposed a dynamic multi-hop routing protocol designed specially for a data similarity aware node clustering that is a topology of the unequal sized cluster to improve the network lifetime, and satisfy the requirement of multi-hop routing protocol for the dynamic node clustering that consider the data similarity of adjacent nodes.

Generally, this protocol runs in two main steps. The first step is the dynamic node clustering based on the data similarity using fuzzy aggregation technique. The second step is a routing algorithm using the modified K-HOP Clustering Algorithm (KHOPCA) [6] by adding a priority factor that is obtained by a hybrid approach of fuzzy system and firefly algorithm. There are two variables that are considered to obtain the priority factor, i.e., the residual energy and distances between CHs and BS.

The remainder of this paper is organized as follows: Section 2 presents literature review related with the routing protocols in wireless sensor networks. Section 3 describes our approach used for dynamic multi-hop routing protocol. Section 4 presents the simulation results to show the performance evaluation. Finally, Section 5 concludes this paper.

2 Routing protocol in WSNs

Routing is the best path to transmit a data packet from a source node to a destination node. The clustering-based routing in WSNs, there are two types of path, i.e. data traffic within a cluster termed as intra-cluster, and data traffic between clusters called inter-cluster. In the intra-cluster, each regular node senses a local environmental condition and transmits it to

corresponding CH. Meanwhile, the CH senses, receives, and aggregates data. Then it transmits the aggregated data packet to either BS directly or via intermediate CHs.

One of highly important issues in many literatures of the clustering-based routing techniques is the use of more energy efficient methods in order to prolong the network lifetime. There are two main steps that needs the appropriate technique in order to achieve better network performance in term of lifetime. Both steps are the clustering technique and the CHs election method. The clustering techniques are classified into two major categories, i.e. unequal sized clustering and equal sized clustering techniques. Similarly, the CHs election approaches are categorized into three major groups, i.e. preset, random, and attribute based method [1]. In preset approach, all nodes that are selected as CHs are adjusted before they were deployed in the environment. In random methods, the CHs are selected among of the nodes randomly in the field. On the other hand, attribute-based approaches select the CHs among of the nodes based on some of their characteristics, such as the residual energy and distance to the BS.

There are several equal clustering based routing protocols that select CHs randomly. Among other is the Low Energy Adaptive Clustering Hierarchy (LEACH) [12] that is a hierarchical clustering-based routing protocol which has been used widely as a benchmark. There are many LEACH-based routing protocols that have been proposed to improve the energy efficiency in order to prolong the network lifetime such as LEACH-Centralized (LEACH-C) [13], LEACH-based Energy (LEACH-E) [15], and LEACH with Distance-based Threshold (LEACH-DT) [16]. The disadvantage of the LEACH-based routing protocol is that CHs are close to the BS that consumes more energy than the CHs farther from BS. Consequently, the CHs near the BS died earlier. This causes a disruption of the network connectivity termed as a hot spot problem.

Single-hop routing protocol can overcome the hot spot problem because it does not require the intermediate BS to relay the data packet to BS. However, this approach has a limitation of transmission coverage, so that the scalability of the network cannot be achieved. In order to overcome the hot spot problem on the network, some unequal sized clustering approaches using single-hop routing have been proposed. However, these approaches are a waste of energy and the transmission coverage are limited [18]; [4]; [7]. Therefore, several multi-hop routing protocols using unequal sized clustering technique [2]; [21]; [24] have been proposed to overcome the hot spot problem, improve the network scalability and optimize the energy-saving in order to prolong the network lifetime.

In fact, some of specific applications in WSNs not only satisfy the three purposes, but also require a clustering technique based the data similarity readings of adjacent nodes to obtain the data pattern of the observed environment to make decision or prediction. Therefore, to fulfill the requirements of the applications, this paper proposes a new dynamic multi-hop routing protocol using the unequal size clustering technique based on the data similarity. This protocol proposes an incorporation the K-Hop Clustering Algorithm (KHOPCA) rules [6] and fuzzy system-firefly algorithm [25]. Our proposed routing protocol is called Data Similarity Aware for Dynamic Multi-hop Routing Protocol (DSA-DMRP). The DSA-DMRP starts to establish the node clustering based on data similarity among all nodes in the network using the Fuzzy Aggregation Technique as a dynamic unequal sized clustering mechanism. Finally, the CHs election and establishment of the best route path uses the KHOPCA rules and a priority factor in the CHs election. The priority factor considers the residual energy and distance to the BS using integration of Fuzzy System (FS) and Firefly Algorithm (FA) to improve the network lifetime.

3 The proposed model of routing protocol

3.1 Network model

Data similarity aware node clustering in WSNs is an unequal sized clustering-based WSNs. A node will merge or leave its cluster depend on its data similarity degree to those of neighbor nodes. Therefore, perhaps there are some nodes that are an intersection in some clusters. Due to an unequal sized clustering, there are some clusters that have member nodes fewer than those of other clusters. Moreover, there are some individual nodes which does not belong to any cluster. The characteristic of such a clustering-based model requires a properly specific routing algorithm. Hence, a multi-hop clustering-based routing protocol was developed to overcome the problem. All sensor nodes that are randomly deployed on the network are stationary. They are homogeneous in their capabilities of sensing, processing, and communicating. The BS is assumed as a stationary site, and it has no energy constraint. Not all nodes can communicate directly with other nodes. Only nodes that can communicate directly to the sink will be considered to become the cluster heads. The node clustering and the data gathering are run into rounds. In each round, there are four steps, (i) all nodes sensed the local data, (ii) the nodes divided themselves into several clusters, (iii) the nodes created a multi-hop routing, (iv) the member nodes sent their data to the corresponding CHs, and the CHs aggregates the data in order to forward them to the BS.

3.2 Energy consumption model

Energy consumption in WSNs is an urgent issue to be considered in order to increase the longer network lifetime. The energy consumption models have been developed in several literatures. Figure 1 shows the consumption model of radio energy.

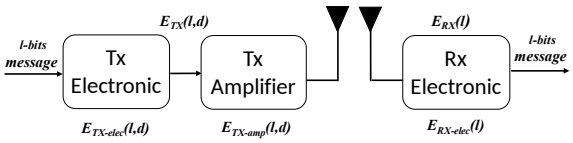


Figure 1: Transmission energy dissipation model [24]

The dissipated energy to transmit *l-bits* message with distance d and receive *l-bits* can be computed respectively using the following equations [7].

$$\begin{aligned} E_{TX}(l,d) &= E_{TX-elec}(l) + E_{TX-amp}(l,d) \\ &= l(E_{elec} + E_{amp}d^p) \\ &= \begin{cases} l(E_{elec} + \varepsilon_{fs}d^2, & if \quad d \leq d_0 \\ l(E_{elec} + \varepsilon_{mp}d^4, & if \quad d > d_0 \end{cases} \end{aligned} \tag{1}$$

$$E_{RX}(l) = E_{RX-elec}(l) = l \times E_{elec} \tag{2}$$

where E_{elec} is the consumed energy in either transmitter or receiver circuit, is for amplifier radio model, and p is the path loss. The path loss is adjusted to $p = 2$ and the free space model is used by assigning $E_{amp} = \varepsilon_{fs}$ if the distance d is less than or equal to the threshold distance. On the contrary, the path loss is assigned $p = 4$ and the multi-path fading model is employed by setting $E_{amp} = \varepsilon_{mp}$ if distance d more than the distance threshold d_0. Besides, the sensing data by sensor also consumes amount of energy significantly as follows [24].

$$E_{Sensing}(l) = l \times E_{sens} \tag{3}$$

The clustering model based on data similarity has a high correlation between the data read by each node in any cluster. Therefore, CH can use data aggregation ways to bundle all highly related data into a single length-fixed packet. The consumed energy to aggregate n packet of *l-bits* by the cluster head is calculated by [24]:

$$E_{aggre}(l) = n \times l \times E_{DA} \tag{4}$$

In the network, there are K clusters in which each cluster contains of s_i (i = 1, 2, ..., K) regular node. In each round, a node senses *l-bit* packet data and transmits it to the corresponding cluster head once a frame. Thus, the consumed energy by a regular node in the intra-cluster communication between the regular node i and its cluster head j is calculated by [24]:

$$E_{reg-j}(i, l, d) = E_{Sensing}(l) + E_{TX}(l, d) \tag{5}$$

where *d(i, j)* is the distance between regular node i and CH j. On the other hand, the consumed energy by the CH in both intra-cluster and inter-cluster communication consist of five activities i.e. CH senses data, CH receives data packets, CH aggregate data packets, CH receives and transmits data packets from the other CH, and CH transmits the aggregated packets and the received packets of other CH to the BS [24]. Thus, the consumed energy by a CH can be obtained using the following equation [24].

$$\begin{aligned} E_{CH}(i, l, d) = l(&E_{Sensing}(l) + E_{elec}s_i + E_{DA} \\ &(s_i + 1) + E_{elec}relay(j) + (relay(j) + 1) \\ &(E_{elec} + E_{amp}d(i, NH))_p) \end{aligned} \tag{6}$$

where *d(i, NH)* is the distance between CH j and next hop (NH) in which *NH* may be CH or base station. Moreover, *relay(j)* is the number of forwarded packets from other CHs.

3.3 Dynamic node clustering based on data similarity

In WSNs, there are several applications that require a data similarity based node clustering approach. Such approach is highly related to two important issues, i.e. the spatial and temporal correlation. In the spatial correlation, the data that is sensed by adjacent nodes tend to have a high data similarity degree. Meanwhile, in the temporal correlation, the data that is sensed by each node consecutively at an observed location tend to have a high correlation of its data readings.

The fuzzy aggregation technique in equation 7 [9] can measure the data similarity degree considering the spatial correlation of two data a and b that is sensed by two adjacent nodes.

$$Sim(a, b) = exp\frac{-||a - b||^2}{2 \times \eta^2}, \quad Sim(a, b) =\in [0, 1] \tag{7}$$

where η is the constant Gaussian Kernel by setting $\eta = 1.74$. The data similarity degree of 1 is a highest level and 0 is a lowest level. To establish the unbalanced clusters based on the data similarity, the DSA-DMRP use two main algorithms to identify the neighbor nodes that satisfy the data similarity degree. Both algorithms are embedded in each node in order to broadcast and receive the beacon message as shown in Algorithm 1 and 2 [20] :

1. In each node, two simple data structures are required: (i) the *SpatNeighbor* is a vector data structure to store the information of a spatially closed neighbor nodes; and (ii) *simNeighbor* is a vector data structure to store the information of the similar neighbor nodes. The information contains all of the three local variables, i.e., address (*add*), current data readings (*cdR*), and the number of similar neighbors (*nsN*).

2. Algorithm 1: Beacon message, the **Broadcasting** (line 1-3) is a main function to broadcast a beacon message periodically. This function consists of three sub functions: (i) the Send sub function (line 1) transmits the information about the identifier address (*add*), the current data reading (*cdR*), and the number of similar neighbors (*nsN*). (ii) the Delay sub function prevents the transmissions simultaneously. (iii) the TimeExpire sub function limits the broadcasting time of the nodes.

3. Algorithm 2: The **Receiving** main function is employed by the node to receive a beacon message. The message is used by any node receiver to identify whether the transmitter node includes as a similar neighbor node.

4. The beacon message of *add*, *cdR*, and *nsN* is utilized to identify the similar neighbor candidates (line 1-2). The data similarity degree is measured by the fuzzy aggregation technique (line 7). If its similarity degree is more than or equal to *SiDegree* and also the similar neighbor candidate is not existed within the data structure, it is added into the data structure as one of the members of the similar neighbor (line 10). The number of similar neighbor (*nsN*) is incremented (line 11). In contrast, if the similar neighbor candidate existed within the data structure, it is removed from the data structure (line 13), and the number of similar neighbor (*nsN*) is decremented (line 14).

Algorithm 1 Beacon Message

Broadcasting()

```
1: Send(add, cdR, nsN)
2: Delay(interval + rand())
3: TimeExpire()
```

3.4 Cluster head election and multi-hop routing

DSA-DMRP forms routes that are started through selecting several cluster heads. Establishing routes and selecting CHs utilize K-Hop Clustering Algorithm (KHOPCA) [6] that consists of four rules. However, the rules proposed by the KHOPCA do not consider how to prolong the network lifetime and overcome the hot spot problem. To address both problems, our DSA-DMRP proposed modified KHOPCA's rules by adding a Priority Factor (PF) to select the prospective common nodes as the CHs. The PF is calculated via an incorporation between fuzzy model and Firefly Algorithm (FA) that consider two input variables, i.e. residual energy and distance to the base station.

The KHOPCA constructs the network routes using a set rules that were inspired by Game of Life [8]. The KHOPCA is implemented over the networks that have been clustered. A route from a regular node to the base station is established based on a set of rules that define transition of previous state to current state of a node depending on the previous state of its similar neighbors. The state of a node is represented by a weight $w \in [wMin, wMax]$. Minimum distance to a cluster head is defined as the minimum weight *wMin*, whilst the maximum weight *wMax* is the

Algorithm 2 Receiving data for node clustering

Receiving*(add,cdR,nsN)*

```
spatNeighbor
simNeighbor
simNeighbor ← CreateSimNeighbor(add, cdR, nsN)
sigma ← 1.74
n ← cdR
m ← DataReading()
s ← exp( - || n-m || ^2/(2 * sigma^2)) Eq. 7
if (s ≥ SiDegree) then
    if (!IsExistSimNeighbor(simNeighbor) then
        AddSimilarNeighbor(simNeighbor)
        IncreNumSimilarNeighbor()
    else if ( thenIsExistSimNeighbor(simNeighbor)
        RemoveSimNeighbor(simNeighbor)
        DecrNumSimNeighbor()
    end if
end if
```

maximum distance to a cluster head. The KHOPCA has four rules to establish the network route as follows [6]:

1. If node i with wight w_i has a neighbor node that has the highest weight w_n where $w_n > w_i$, $\forall$n $\in$ *LN(i)* of its all neighbor nodes, and *LN(i)* is the list of i similar neighbors, the node i changes its weight to $w_i = w_n$ - 1.

2. If node i has no a similar neighbor with a higher weight, where $w_i \neq wMax$, $w_i \geq w_n$ and $\forall w_n \in$ *W(LS(i))*, the node i decreases its weight to $w_i = w_i$ - 1

3. If weight of the node i is $w_i = wMin$ and also $w_i = w_n$ where $w_n \in$ *W(LS(i))*, the node i adjusts its weight to $w_i \leftarrow wMax$ and states itself as CH. In this case, none of its similar neighbors has a higher weight than *wMin*.

4. If weight of node i is $w_i = wMax$ and weight of one of its neighbor nodes has also weight $w_n = wMax$, where $\exists$ $w_n \in$ *W(LS(i))*, the node n decreases its weight to $w_n = w_n$ - 1. In this case, there are two CHs in the same cluster.

Although those four rules are simple, they can construct all nodes in the network to create a multi-hop routing. The first rule aims to form a top-down hierarchical structure through adjusting its weight with a difference-one of the highest weight node existing in the list of similar neighbors. The second rule intends to avoid the less weighted nodes that are most likely to quit from a cluster. Thus, the higher weighted nodes that the CH attract at the surrounding nodes will merge into its cluster in order to a fragmented cluster. The third rule describes that a node declares itself as a CH if all similar neighbors have a minimum weight. This situation shows that the isolated nodes are chosen as CH on the minimum weight level. The fourth rule overcomes a situation where there are two CHs in the same cluster. Therefore, one of them must survive as a CH, while other node must be a follower node of the CH.

The weakness of the rules is that it has not considered how to prolong the network lifetime. Therefore, the third rule determines a node to be a CH. The fourth rule defined itself as a CH.

Both rules are modified by adding a priority factor that considers the residual energy and the distance to the BS as follows.

3. If weight of the node i is $w_i = wMin$, $w_i = w_n$ where $w_n \in W(LS(i))$, and also its PH is highest in the same cluster, the node i adjusts its weight to $w_i \leftarrow wMax$ and states itself as CH. In this case, none of its similar neighbors has a higher weight than *wMin*.

4. If weight of node i is $w_i = wMax$, and weight of one of its neighbor nodes has also weight $w_n = wMax$, where $\exists\ w_n \in W(LS(i))$, as well as its PH is highest in the same cluster, the node n decreases its weight to $w_n = w_n$ - 1. In this case, there are two CHs in the same cluster.

Priority factor of cluster head selection

One of criteria to select prospective regular node as CH is the priority factor as shown in third and fourth rule of modified KHOPCA's rules. The Priority Factor (PF) is calculated using a incorporation between Fuzzy Logic and Firefly Algorithm (FA). The PF is obtained through a procedure that consists of four steps: normalization, fuzzification, inference engine, and defuzzification as shown in Figure 2. To obtain the proper selection of fuzzy rule in inference process, firefly algorithm is used to optimize the the Tsukamoto fuzzy rule base table. As input variable of the fuzzy logic, our proposed DSA-DMRP considers two variables, i.e. the residual energy $E_r(n)$ and the distance to the BS $d_{BS}(n)$.

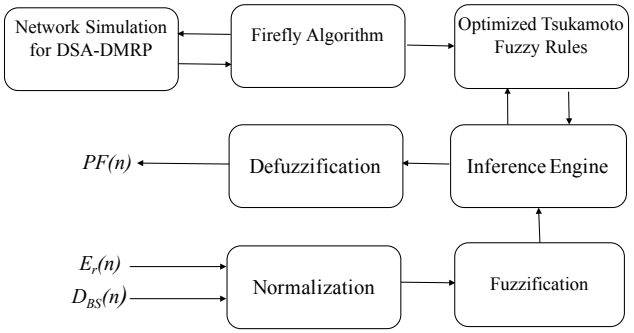


Figure 2: Procedure of priority factor to select cluster head

Before fuzzification that maps crisp input to membership degree via membership function, the first step normalizes both input variables $E_r(n)$ and D_{BS} (n) into range of [0, 1]. This step is performed to avoid the difference of range value in each cluster. Normalization of both input variables use a general formula as follows:

$$d^n(n) = \frac{d(n) - min(d)}{max(d) - min(d)} \tag{8}$$

where $d^n(n)$ is normalized value and $d(n)$ is real values of input variable d for node n. The input variable d can be applied for both input variables E_r and D_{BS}. The maximum real value of d is defined as $max(d)$ and the minimum real value of d is represented as $min(d)$. In the second step, the fuzzifier maps normalized crisp values d to the membership functions to convert to be the linguistic fuzzy values. In this study, our proposed fuzzy model uses five membership functions that are symbolized with Very Low *(VLow)*, *Low*, *Medium*, and *High* for both input variables as shown in equation 9 up to 13 respectively.

$$y_{VLow} = \begin{cases} 1 & if \quad x \leq 0.1 \\ \frac{0.3-x}{0.3-0.1} & if \quad 0.1 \geq x \leq 0.3 \\ 0 & if \quad x \geq 0.3 \end{cases} \tag{9}$$

$$y_{Low} = \begin{cases} 0 & if \quad x \leq 0.1 \; or \; x \geq 0.5 \\ \frac{x-0.3}{0.3-0.1} & if \quad 0.1 \geq x \leq 0.3 \\ \frac{0.5-x}{0.5-0.3} & if \quad 0.3 \geq x \leq 0.5 \end{cases} \tag{10}$$

$$y_{Madium} = \begin{cases} 0 & if \quad x \leq 0.3 \; or \; x \geq 0.7 \\ \frac{x-0.5}{0.5-0.3} & if \quad 0.3 \geq x \leq 0.5 \\ \frac{0.7-x}{0.7-0.5} & if \quad 0.5 \geq x \leq 0.7 \end{cases} \tag{11}$$

$$y_{High} = \begin{cases} 0 & if \quad x \leq 0.5 \; or \; x \geq 0.9 \\ \frac{x-0.7}{0.7-0.5} & if \quad 0.5 \geq x \leq 0.7 \\ \frac{0.9-x}{0.9-0.7} & if \quad 0.7 \geq x \leq 0.9 \end{cases} \tag{12}$$

$$y_{High} = \begin{cases} 0 & if \quad x \leq 0.7 \\ \frac{0.9-x}{0.9-0.7} & if \quad 0.7 \geq x \leq 0.9 \\ 1 & if \quad x \geq 0.9 \end{cases} \tag{13}$$

Likewise, the defuzzifier use seven membership functions i.e. Very Small (*VSmall*), *Small*, Rather Small (*RSmall*), *Medium*, Rather Large (*RLarge*), *Large*, and Very Large (*VLarge*) as shown in equation 14 up to 20 respectively to obtain the fuzzy output.

$$PF_{VSmall} = \begin{cases} 1 & if x \leq 0.05 \\ \frac{0.2-x}{0.2-0.05} & if \quad 0.05 \geq x \leq 0.2 \\ 0 & if \quad x \geq 0.2 \end{cases} \tag{14}$$

$$PF_{Small} = \begin{cases} 0 & if \quad x \leq 0.05 \; or \; x \geq 0.35 \\ \frac{x-0.2}{0.2-0.05} & if \quad 0.05 \geq x \leq 0.2 \\ \frac{0.35-x}{0.35-0.2} & if \quad 0.2 \geq x \leq 0.35 \end{cases} \tag{15}$$

$$PF_{RSmall} = \begin{cases} 0 & if \quad x \leq 0.05 \; or \; x \geq 0.35 \\ \frac{x-0.35}{0.35-0.2} & if \quad 0.2 \geq x \leq 0.35 \\ \frac{0.5-x}{0.5-0.35} & if \quad 0.35 \geq x \leq 0.5 \end{cases} \tag{16}$$

$$PF_{Madium} = \begin{cases} 0 & if \quad x \leq 0.35 \; or \; x \geq 0.65 \\ \frac{x-0.5}{0.5-0.35} & if \quad 0.35 \geq x \leq 0.5 \\ \frac{0.65-x}{0.65-0.5} & if \quad 0.5 \geq x \leq 0.65 \end{cases} \tag{17}$$

$$PF_{RLarge} = \begin{cases} 0 & if \quad x \leq 0.5 \; or \; x \geq 0.8 \\ \frac{x-0.65}{0.65-0.5} & if \quad 0.5 \geq x \leq 0.65 \\ \frac{0.8-x}{0.8-0.65} & if \quad 0.65 \geq x \leq 0.8 \end{cases} \tag{18}$$

$$PF_{Large} = \begin{cases} 0 & if \quad x \leq 0.65 \; or \; x \geq 0.95 \\ \frac{x-0.8}{0.8-0.65} & if \quad 0.65 \geq x \leq 0.8 \\ \frac{0.95-x}{0.95-0.8} & if \quad 0.8 \geq x \leq 0.95 \end{cases} \tag{19}$$

$$PF_{VLarge} = \begin{cases} 0 & if \quad x \leq 0.8 \\ \frac{0.95-x}{0.95-0.8} & if \quad 0.8 \geq x \leq 0.95 \\ 1 & if \quad x \geq 0.95 \end{cases} \tag{20}$$

In third step, the inference engine performs a fuzzy reasoning against the crisp input in the fuzzy rule base table containing n rules. In our study, we use Tsukamoto Fuzzy system [23] with two inputs and an output. The typical rules of Tsukamoto uses AND-based fuzzy rule base table that are represented as IF-THEN as shown in Equation (21) as follows:

$$\textbf{IF} \quad in_1 = A \ \ \textbf{AND} \ \ in_2 = B \ \ \textbf{THEN} \ \ out = C \tag{21}$$

where A and B are the membership degree of the corresponding input membership functions and C is min(A, B).

In final step, all outputs of the fired rules are aggregated and converted to be a single-crisp output value. To obtain the single-scrip output value as value of the priority factor, our fuzzy model uses the average-weighted Tsukamoto defuzzification model [23]. The Priority Factor *PF(n)* can be formulated in Equation (22) as follows:

$$PF(n) = \frac{\Sigma_{i=1}^{25} \mu_{i \times C_i}}{\Sigma_{i=1}^{25} \mu_i} \tag{22}$$

where μ_i is min(μ_{Er}, $\mu_{D_{BS}}$) to corresponding membership functions within the fuzzy rule i. Also, C_i is the output of corresponding membership function in rule i.

Optimization of AND-based fuzzy rule via firefly algorithm

The Inference engine of fuzzy system has usually many rules. The selection of fuzzy rule requires a proper method to obtain the best performance of the fuzzy system. In our fuzzy model, there are two input variabels in which each input variable has five membership functions. These mean that the number of rules is 5 x 5 = 25. Because the output fuzzy has seven member functions, the number of output alternatives of the 25 rules is 7^{25}. The output alternatives of 7^{25} become an NP-hard problem to find the best solution in turning of the fuzzy rule base table. The NP-hard problem can be addressed a fuzzy system that uses Firefly Algorithm (FA) to optimize the Tsukamoto fuzzy rule base table in order to prolong the network lifetime based data similarity aware node clustering.

FA is a population-based swarm intelligent search algorithm [11]. Each individual firefly in population has a role as a candidate solution in the search space. Each firefly moves toward a new position. The new position represents a better candidate solution. Finally, they find the best solution. The movement is represented by their attractiveness. The attractiveness is proportional to the intensity of the emitted light by adjacent fireflies. The better solution is usually measured by the fitness value.

Let x_i be the ith firefly in the population, where i = 1, 2, ..., N and N is the population size. The attractiveness β with the Euclidian distance r_{ij} between two adjacent fireflies x_i and x_j can be computed using the Equation (23) as follows [27]:

$$\beta(r_{ij}) = \beta_{min} + (\beta_0 - \beta_{min}) e^{-\gamma r_{ij}^2} \tag{23}$$

$$r_{ij} = ||x_i - x_j|| = \sqrt{\sum_{k=1}^{D} (r_{ik} - r_{jk})^2} \tag{24}$$

where D is the problem dimension with $k = 1, 2, \ldots, D$. The parameter β_0 indicates the fireflies' attractiveness at the distance $r = 0$, and γ is the light absorption coefficient. β_{min} is the minimum value of β as shown in Equation 23. The attractiveness of β is limited in the range of $[\beta_{min}, \beta_0]$. Each firefly X_i is compared with all other fireflies x_j , where $j = 1 , 2 , \ldots , N$ and . If x_j is brighter than x_i, x_i will be attracted to and move toward x_j. The movement of the firefly x_i that is attracted toward the firefly x_j can be calculated by [27]:

$$x_{ik}(t+1) = x_{ik}(t) + \beta^{-\gamma r_{ij}^2} + (x_{ik} - x_{jk} + \alpha(t) s_i \varepsilon_i \tag{25}$$

$$\alpha(t+1) = \alpha(t)(1/9000)^{\frac{1}{t}} \tag{26}$$

where ε_i is a uniformly distributed random value in the range of [-0.5, 0.5] and parameter α is the dynamically updated step factor using Equation (26) and s_i is the length of scale of each designed variable.

In Algorithm 3 [28], the FA starts to optimize the AND-based fuzzy rules through generating randomly the population of N fireflies. Since there are 25 controllable parameters in fuzzy rules as mentioned previously, the length of feasible solution is a string of 25. Each value of feasible solution contains number 1 to 7 representing seven output member functions.

Algorithm 3 FA to optimize the AND-based fuzzy rules

Input: Population: N; Dimension D; Iterator time: T Output: Global best firefly's brightness x(t)

```
t ← 1 (initialization)
initialize all fireflies brightness x_i^k
while t ≤ T do
    Update the parameter α according to Eq. 26
    for i=1:N do
        for j=1:N do
            if j ≠ i then
                Compute the attractiveness of β
                according to Eq. 23
                if f(x_j(t) < f(x_i(t)) then
                    Move x_i(t) toward x_j(t)
                    according to Eq. 25
                    f(x_i(t)) ← Evaluate Fitness of Firefly
                    according to Eq. 27
                    t ← t + 1
                end if
            end if
        end for
    end for
end while
```

Before conducting the iteration process, the fitness value of each feasible solution is computed using *f(X_i)* fitness function. The optimal solution will be obtained when the iteration reaches the maximum iteration time M. In each iteration, the parameter is updated firstly using Equation (26). Next, the attractiveness β between two fireflies x_i and x_j is calculated according to Equation (23), where $j \neq i$. The movement of the x_i firefly towards the x_j firefly using Equation (25) is processed if the fitness value of x_j firefly is better than that of x_i firefly.

To evaluate each feasible solution, the FA requires a fitness function $f(x_i)$. The feasible solutions are taken from the corresponding Tsukamoto fuzzy rules to be extracted and simulated in the network using Network Simulation of DSA-DMRP. The best solution will be obatined if all fitness within same approaching value. The fitness is computed using three parameters, i.e. the number of rounds when the first node dies (FND), the number of rounds in which half of nodes are dead (HND), and the number of rounds until the last node dies (LND). The network lifetime is measured using the three parameters. The Equation (27) is the fitness function and its constrains, which are used to maximized the network lifetime formulated as follows [29]:

Maximize:

$$Fitness = w_1 FND + w_2 HND + w_3 LND \tag{27}$$

Address to

$$0 \leq w_j \leq 1, \qquad \sum_{j=1}^{3} w_j = 1 \tag{28}$$

where w_j with $j = (1, 2, 3)$ are the weights to determine the important objective of parameters of *FND*, *HND*, and *LND*.

4 Performance evaluation

This study is a experimental research using a a NS-3 network simulator version 3.25. The performance of our proposed DSA-DMRP was evaluated using term of the network lifetime of *FND*, *HND*, and *LND*, as well as the round history of alive nodes in various data similarity degrees of the cluster. Furthermore, the DSA-DMRP was compared agianst the KHOPCA with exactly the same scenario.

4.1 Simulation scenario

The experimental data used to simulate the both protocols utilized the humidity readings gathered by the Intel Berkeley Research Lab [5]. The data was collected from 54 sensor nodes deployed in a 640m x 480m sized network. Before the simulation is executed, there are some the network parameters and setting parameters of firefly algorithm that need to be set as shown in Table 1 and Table 2 respectively.

Table 1: Simulation Parameters of Networks

Parameter	**Value**
Data similarity degree (siDegree)	0.7 to 0.9
Gaussian Kernel constant η	1.74
Initial energy	0.5 Joule
E_{elec}	50 nJ/bit
ε_{fs}	100 pJ/bitt/m^2
ε_{amp}	0.03 pJ/bit/m^2
Data packet size	4000 bit

Table 3: Comparison between KHOPCA and proposed DSA-DMRP in term of the network lifetime and in various data similarity degrees (siDegree)

	KHOPCA			DSA-DMRP		
SiDegree	FND	HND	LND	FND	HND	LND
0.7	206	1393	1421	543	1489	1575
0.8	134	1368	1391	316	1420	1477
0.9	240	1380	1452	316	1368	1470

Table 2: Setting parameters of firefly algorithm

Parameter	**Value**
Maximum iterations time (T)	100
The number firefly in population (N)	30
β_0	1
β_{min}	0.2
ε_{amp}	0.005
Dimension of population (D)	25

4.2 Experimental results

Table 3 shows the results obtained in the comparison between KHOPCA and our proposed DSA-DMRP in term of *FND*, *HND*, and *LND*. In this experiment, there are three weights to set the fitness function i.e. w_1=0.8, w_2=0.2 and w_3=0. Moreover, the node clustering based on data similarity was established through three experiments with the data similarity degree (*SiDegree*) 0.7, 0.8, and 0.9.

The DSA-DMRP was compared against the KHOPCA to justify the performance. It shows that the DSA-DMRP can reach a longer network lifetime than the KHOPCA in all conditions due to an addition of the priority factor in the KHOPCA's rules. However, the difference between *FND* and *HND* is highly significant both in the KHOPCA and the DSA-DMRP because the unequal sized cluster is not designed specially to overcome the hot spot problem for CHs close to BS but actually, it is designed in other to satisfy the requirement of the multi-hop routing protocol for the dynamic node clustering based on the data similarity of their neighbors.

Figure 3 shows three round histories of alive nodes versus rounds for each routing protocol with *SiDegree* 0.7, 0.8 and 0.9 respectively. It clearly shows that the DSA-DMRP and the KHOPCA have an approximately same stability of alive nodes. Both protocols show that most of nodes died simultaneously in term of LND. However, the DSA-DMRP is longer in all terms FND, HND, and LND. Therefore, the DSA-DMRP can extend the network lifetime in a relatively significant manner. The stability of alive nodes in both protocols are caused by the stability of KHOPCA's rules in selecting the CHs. Meanwhile, the network lifetime in DSA-DMRP that longer than the KHOPCA is caused by adding the factor priority in selecting CHs.

In order to justify the stability of alive nodes in each protocol for three data similarity degrees, we compare them in variable the data similarity degree for each protocol. Figure 4 shows the round history of alive nodes for the KHOPCA and the DSA-DMRP. The KHOPCA clearly shows the data similarity degree of 0.7 and 0.8 are more stable than at of 0.9. Whereas, the DSA-DMRP indicates that the data similarity degree of 0.8 and 0.9 is more stable than those of 0.7.

However, Figure 4 shows an inverse phenomenon between both the alive node stability graphs. The phenomenon of the KHOPCA protocol shows the graphs of similarity degrees of 0.7 and 0.8 that are more stabile than the similarity degree of 0.9. On the other hand, in the phenomenon

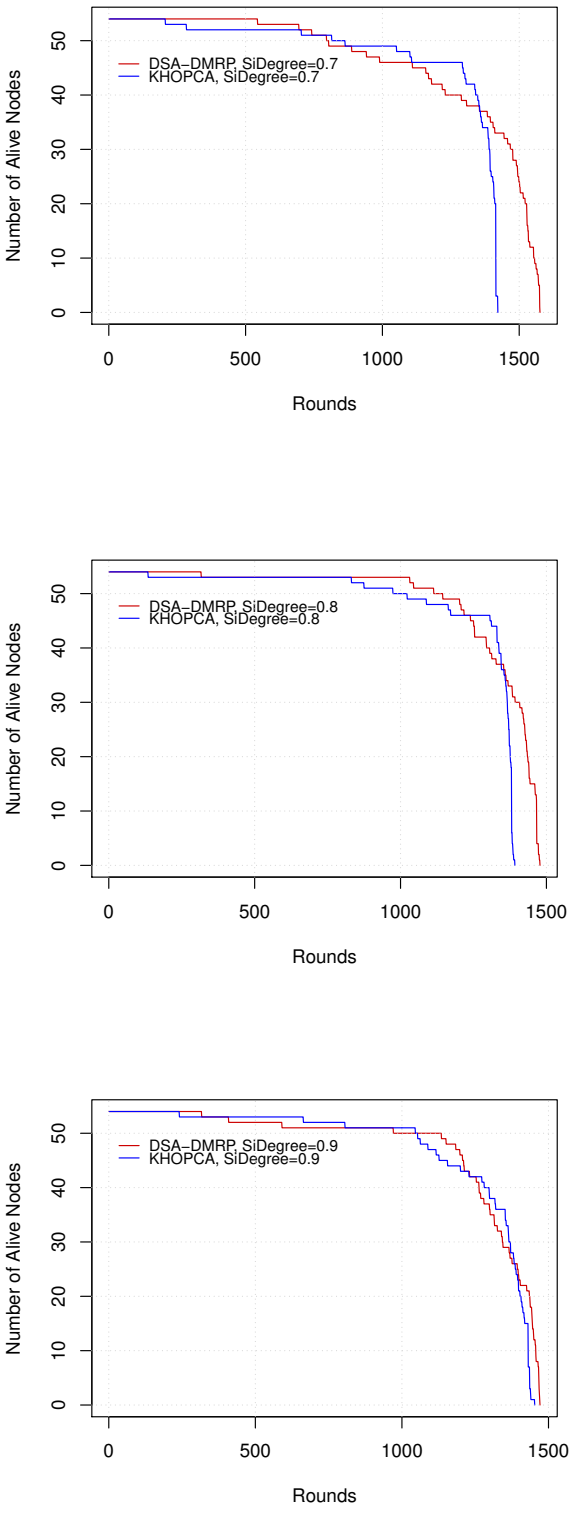


Figure 3: Relation between the number of alive nodes and rounds in *SiDegree* = 0.7, 0.8, and 0.9

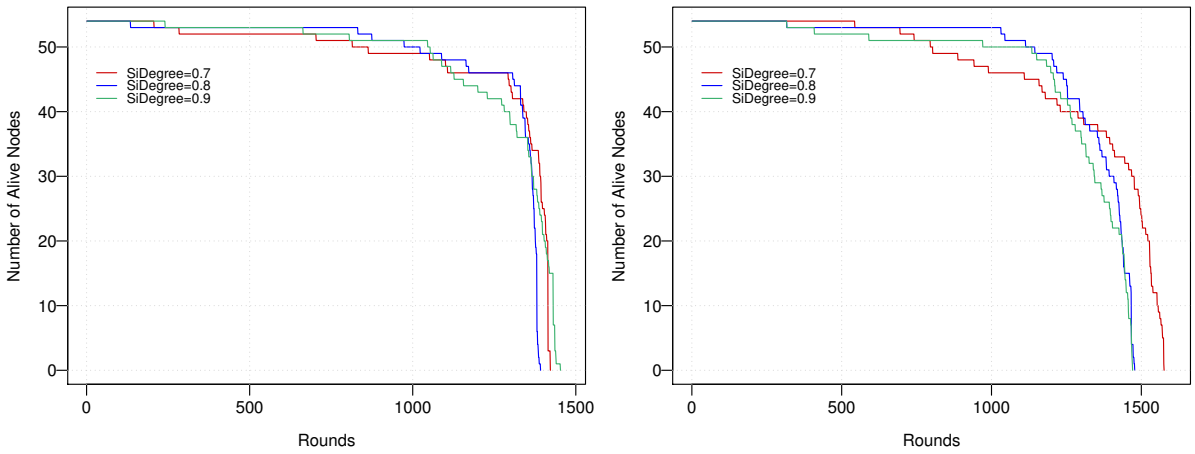


Figure 4: Comparation between the data similarity degree in the alive nodes vs the rounds for the KHOPCA and the DSA-DMRP

of DSA-DMRP protocol, the similarity degrees of 0.8 and 0.9 are more stable than the similarity degree of 0.7. Moreover, in the rounds before reaching 1000 of the KHOPCA protocol, the number of formed clusters in the similarity degrees of 0.7, 0.8, and 0.9 are almost same. Thus, more and more cluster heads in the similarity degree of 0.9, more and more the probabilities of the nodes will be dead because the cluster heads more consume the energy than those of the member nodes. On the contrary, the rounds before reaching 1000 of the DSA-DMRP, the number of formed clusters is almost same the number of formed clusters in the KHOPCA protocol. However, before reaching 1000 rounds, the DSA-DMRP has less the number of cluster heads in the similarity degrees of 0.8 and 0.9 than those of the similarity degree of 0.7. Finally, the problems in this case are caused by the elected cluster heads that are not only effected by the similarity degree but also affected by the residual energy and the distance to the base station.

5 Conclusions

Our proposed DSA-DMRP is a dynamic multi-hop routing protocol using unequal sized clustering approach. This protocol is based on the modified KHOPCA rules by adding a priority factor. The priority factor is a parameter for selecting the CHs in the network that consider the residual energy and distance to the BS. The fuzzy aggregation technique is used to measure the data similarity degree of adjacent nodes.

The DSA-DMRP was compared against the KHOPCA to justify the performance. The DSA-DMRP and the KHOPCA have an approximately same stability of alive nodes. However, The DSA-DMRP can reach a longer the network lifetime than the KHOPCA in all terms of FND, HND, and LND. Therefore, the DSA-DMRP can extend the network lifetime in a relatively significant manner and can satisfy the requirement of multi-hop routing protocol for dynamic node clustering based on the data similarity of their neighbors.

Acknowledgement

The authors gratefully acknowledge the Ministry of Research, Technology and Higher Education of Indonesia for supporting this work through DIKTI's research grant under grant No. 1173/UN2.R12/HKP.05.00/2016

Bibliography

[1] Afsar, M. M.; Tayarani, M. H. (2014); Clustering in sensor networks: A literature survey, *Journal of Network and Computer Applications*, 46, 198-226, 2014.

[2] Ahmed, N.; Kanhere, S.; Jha, S. (2010); Experimental Evaluation of Multi-hop Routing Protocols for Wireless Sensor Networks, *In Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks*, 416–417, New York, NY, USA, 2010.

[3] Amin, R.; Islam, S. H.; Biswas, G. P.; Khan, M. K.; Kumar, N. (2016); A robust and anonymous patient monitoring system using wireless medical sensor networks, *Future Generation Computer Systems*, 2016.

[4] Bagci, H.; Yazici, A. (2013); An energy aware fuzzy approach to unequal clustering in wireless sensor networks, *Applied Soft Computing*, 13(4), 1741-1749, 2013.

[5] Bodik, P., Hong, W., Guestrin, C., Madden, S., Paskin, M., and Thibaux, R., Intel Lab Data. Retrieved from http://db.csail.mit.edu/labdata/labdata.html.

[6] Brust, M. R.; Frey, H.; Rothkugel, S. (2008); Dynamic Multi-hop Clustering for Mobile Hybrid Wireless Networks, *Proceedings of the 2Nd International Conference on Ubiquitous Information Management and Communication*, 130-135. New York, NY, USA: ACM, 2008

[7] Gajjar, S.; Talati, A.; Sarkar, M.; Dasgupta, K. (2015); FUCP: Fuzzy based unequal clustering protocol for wireless sensor networks, *39th National Systems Conference (NSC)*, 2015.

[8] Gardner, M. (1970); Mathematical Games: The Fantastic Combinations of Jhon Conway's New Solitaire Game "Life", City, 1970.

[9] Ghaddar, A.; Razafindralambo, T.; Simplot-Ryl, I.; Tawbi, S.; Hijazi, A. (2010); Algorithm for data similarity measurements to reduce data redundancy in wireless sensor networks, *World of Wireless Mobile and Multimedia Networks (WoWMoM), 2010 IEEE International Symposium on*, 1-6, 2010.

[10] Gupta, R.; Sultania, K.; Singh, P.; Gupta, A. (2013); Security for wireless sensor networks in military operations, *Computing, Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference on*, 1-6, 2013.

[11] Haibin, D.; Qinan, L. (2015); New progresses in swarm intelligence-based computation, *Int. J. Bio-Inspired Computation*, 7(1), 26-35, 2015.

[12] Heinzelman, W. R.; Chandrakasan, A.; Balakrishnan, H. (2000); Energy-efficient communication protocol for wireless microsensor networks, *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, 3005-3014, 2000.

[13] Heinzelman, W. B.; Chandrakasan, A. P.; Balakrishnan, H. (2002); An application-specific protocol architecture for wireless microsensor networks, *IEEE Transactions on Wireless Communications*, 2002.

[14] Horita, F. E. A.; Albuquerque, J. P. de; Degrossi, L. C.; Mendiondo, E. M.; Ueyama, J. (2015); Development of a spatial decision support system for flood risk management in Brazil that combines volunteered geographic information with wireless sensor networks, *Computers and Geosciences*, 80, 2015.

[15] Jia, J.G.; He, Z.W.; Kuang, J.M;, Mu, Y.H. (2010); An Energy Consumption Balanced Clustering Algorithm for Wireless Sensor Network, *2010 6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM)*, 2010.

[16] Kang, S. H.; Nguyen, T. (2012); Distance Based Thresholds for Cluster Head Selection in Wireless Sensor Networks, *IEEE Communications Letters*, 16(9), 1396-1399, 2012.

[17] Kollam, M.; Shree, S. R. B. S. (2011); Zigbee Wireless Sensor Network for better Interactive Industrial Automation, *Advanced Computing (ICoAC), 2011 Third International Conference on*, 304-308, 2011.

[18] Li, C.; Ye, M.; Chen, G.; Wu, J. (2005); An energy-efficient unequal clustering mechanism for wireless sensor networks, *IEEE International Conference on Mobile Adhoc and Sensor Systems Conference*, 597-608, 2005.

[19] Lu, K.; Zhou, R.; Li, H. (2016); Event-triggered cooperative target tracking in wireless sensor networks, *Chinese Journal of Aeronautics*, 29(5), 1326-1334, 2016.

[20] Misbahuddin, M.; Sari, R. F. (2016); Data Similarity Based Dynamic Node Clustering Using Bio-Inspired Algorithm for Self-Organized Wireless Sensor Networks, In *P. Novais and S. Konomi (Eds.), Intelligent Environments*, 318-327, London, UK: IOS Press, 2016.

[21] Purkait, R.; Tripathi, S. (2015); Fuzzy based unequal energy aware clustering with multi-hop routing in wireless sensor network, *2015 IEEE Workshop on Computational Intelligence: Theories, Applications and Future Directions (WCI)*, 2015.

[22] Rahman, M.; Rahman, S.; Mansoor, S.; Deep, V.; Aashkaar, M. (2016); Implementation of ICT and Wireless Sensor Networks for Earthquake Alert and Disaster Management in Earthquake Prone Areas, *Procedia Computer Science*, 85, 92-99, 2016.

[23] Ross, T. J. (2010); *Fuzzy Logic with Engineering Applications*, Mexico, USA: Jhon Wiley & Sons, 2010.

[24] Sabor, N.; Abo-Zahhad, M.; Sasaki, S.; Ahmed, S. M. (2016); An Unequal Multi-hop Balanced Immune Clustering protocol for wireless sensor networks, *Applied Soft Computing*, 43, 372-389, 2016.

[25] Saleem, M.; Di Caro, G. A.; Farooq, M. (2011); Swarm intelligence based routing protocol for wireless sensor networks: Survey and future directions, *Information Sciences*, 181(20), 4597-4624, 2011.

[26] Tuan Dinh, L.; Dat Ho, T. (2015); Design and deploy a wireless sensor network for precision agriculture. In Information and Computer Science (NICS), *2015 2nd National Foundation for Science and Technology Development Conference on*, 294-299, 2015.

[27] Xin-She, Y. (2008); Cuckoo Search and Firefly Algorithm Xin-She Yang Editor Theory and Applications. (X.-S. Yang, Ed.). London, UK: Springer. http://doi.org/10.1007/978-3-319-02141-6, 2008.

[28] Wang, H.; Wang, W.; Zhou, X.; Sun, H.; Zhao, J.; Yu, X.; Cui, Z.(2017); Firefly algorithm with neighborhood attraction, *Information Sciences*, 382-383, 374-387, 2017.

[29] Zahedi, Z.M.; Akbari, R.; Shokouhifar, M.; Safaei, F.; Jalali, A. (2016); Swarm intelligence based fuzzy routing protocol for clustered wireless sensor networks, *Expert Systems with Applications*, 55, 313-328, 2016.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 13(1), 117-128, February 2018.

Modern Interfaces for Knowledge Representation and Processing Systems Based on Markup Technologies

A. A. Mohammed Saeed, D. Dănciulescu

Ali Amer Mohammed Saeed*
Doctoral School of Informatics
University of Piteşti
110040 Piteşti, 1 Târgul din Vale Str., Romania
*Corresponding author: ali.amer81@gmail.com

Daniela Dănciulescu
Computer Science Department
University of Craiova
200585 Craiova, 13 A. I. Cuza Str., Romania
danadanciulescu@gmail.com

Abstract: The usage of markup technologies to specify knowledge to be processed according to a specific field of application is a common technique. Representation techniques based on markup language paradigm to describe various types of knowledge including graph based models is considered and details on using Knowledge Representation and Processing (KRP) Systems in education are presented. XML, and VoiceXML were selected to implement smart interface for KRP systems.
Keywords: KRP systems, markup technologies, intelligent interfaces, VXML

1 Introduction

This article deals with Markup mechanisms for knowledge, but also for voice interfaces. It is based on [13] being an extended version of the previous work of the first author.

The coverage of the subject follows. The next section deals with the usability and efficiency of the following approaches to be used in KRP context: SGML / XML, RDF extensions, state-based modeling - SCXML, and Voice XML.

From the RDF (Resource Description Framework) category, in the context of KRP systems, CKML (Conceptual Knowledge Markup Language), Ontology Markup Language (OML) and DLML (Descriptive Logic Markup Language) are useful.

Other approaches are based on the Ontology Interface Layer (OIL) and the DARPA Agent Markup Language (DAML). Of the ontological development tools, the most commonly used are: DUET (UML Enhanced Tool), UBOT, Protege, and Ontolingua. An example of processing using descriptions in natural language is illustrated using SCXML.

SCXML and VoiceXML are covered in the third section.

Interaction of knowledge bases using JAVA technologies is demonstrated in the fourth section. For this purpose, the legacy knowledge model is modeled by a graph that indicates the inheritance relationship of object attributes.

The fifth section is dedicated to the usage of KRP systems in education. It is shown that, for visually impaired users, the usage of VoiceXML based technologies to translate various educational resources is feasible.

2 Markup models and knowledge representation

By models and markup technologies, in the context of this paper, we understand such models and technologies obtained from SGML (Standard Generalized Markup Language; ISO 8879:1986

[28]). SGML is a meta-language, i.e. an artificial language which allow us to describe other languages, in general for the formatting of documents [13].

SGML was used initial by the Association of American Publishers. Then it has become a powerful model with applications and multiple influences. For example, Coleman and Willis (1997) proposed the usage of SGML in the conservation of the publications of the libraries [4]. In the same year, already appeared HTML (HyperText Markup Language, 1990 [23]) useful for WWW, and Extensible Markup Language (XML, 1996) as the language of the description of the structured information [31]. Therefore, SGML is known as being the father of both HTML and XML [13]. However HTML is a court specifies its DTD of SGML (with markers predefined), and XML is a subset of SGML where users can define their own tags and attributes.

An XML document is composed of *markers* (tags) and *data* "character" (char, character). A *marking* is a string of characters bounded by the symbols "<" and ">". An XML file contains three sections: a *header* (<?xml version="1.0" encoding="UTF-8?">), the *definition* of document type internal or external (example: <!DOCTYPE document SYSTEM "location of its DTD">) and the *root* (XML Information in this part may be set as a tree structure).

A XML schema to describe the set of rules used by Knowledge Representation and Processing (KRP) Systems can be given as below (regula.xsd).

```
<?xml version="1.0"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
<xs:element name="KRPRule">
  <xs:complexType>
   <xs:sequence>
     <xs:element name="RLabel" type="xs:string"/>
     <xs:element name="RLeft" type="xs:string"/>
     <xs:element name="RRight" type="xs:string"/>
   </xs:sequence>
  </xs:complexType>
</xs:element>
</xs:schema>
```

XML Processors are used to verify whether the XML documents are well formed or not. To access and editing an XML document, initially is loading the XML document in associated task (example with JavaScript) [13]:

```
parser = new DOMParser();
xmlDoc = parser.parseFromString(text,"text/xml");
```

Then extract the elements of the XML document for processing.

In the context of the knowledge representation of the rules used by KRP systems we can consider <KR> as a root element which may include one or more elements of the rule type. Each item rule has a unique identifier *rid*. A rule of association (A $\rightarrow$ B) is formed of "The hypothesis of A" and "B - the consequent part". Therefore, each hypothesis must have one or more items. Each item hypothesis has a name that is represented by a sequence. This model is described below.

XML document:

```
<KR><Rule rid="1">
      <Hyp>
          <ItemHyp>
              <Name>A</Name>
          </itemHyp>
      </Hyp>
      <Infer>
          <ItemInfer>
              <Name>B</Name>
          </ItemInfer>
       </Infer>
</Rule> </KR>
```

The DTD of the model:

```
<!ELEMENT KR (Rule+)>
<!ELEMENT Rule (Hyp Infer)>
<!ATTLIST Rule rid CDATA>
<!ELEMENT Hyp (ItemHyp+)>
<!ELEMENT ItemHyp (Name)>
<!ELEMENT Infer (ItemInfer+)>
<!ELEMENT ItemInfer (Name)>
<!ELEMENT Name (#PCDATA)>
```

To describe structures useful to outline knowledge in the field of science of the soil for agriculture, the authors of [12] have converted XML declarations in a format useful for application, called KBML (Knowledge Based Markup Language). All meta-information is stored in a file KBML, while the actual data may be available in any data source (distributed, etc.). According to [13], KBML is not a markup language, but merely an application of XML.

In the context of modeling and knowledge processing many specialized Markup notations have been developed, such as: RDF/XML (model supertitles for expressing graphs as RDF documents that XML [27]), CKML (The Conceptual Knowledge Markup Language, 2000 [10]), OML (Ontology Markup Language [26]), DLML (Logical Description Markup Language [22]). OML is an extension of the SHOE and supports the lambda expressions. OML and CKML are based on the conceptual graphs introduced by Sowa (2008) in [17].

Querying RDF data is possible by specific languages, some in the lines of traditional database query languages, others based on logic and rule language [1]. Stratified graphs can be used to automatic generation of queries in formal or natural language [5, 6, 14].

The kernel of a RDF model is made up of *nodes* and *pairs of attached attributes/values*. A description of the RDF syntax is presented in [3] and can be understood on the basis of the following example that describes the creator of the file tey.rdf located in a folder on a Windows Server:

```
<rdf:RDF>
  <rdf:Description
    rdf:about="file:///home/ali/tey.rdf">
     <xf:Creator>
       <rdf:Description
          rdf:about="http://www.upit.ro/">
           <xf:Name>A L I</xf:Name>
       </rdf:Description>
```

```
      </xf:Creator>
  </rdf:Description>
</rdf:RDF>
```

For RDF diagrams one shall specify the space of rdfs *names*. The fundamental RDF *classes* are: rdfs:Resource (class resources), rdf:Property (describes the properties of the resources) and rdfs:Class (for specifying the type or category). To define a new class of RDF diagram, the corresponding resource class has the property rdfs:type whose value resource is rdfs:Class. The resources which belong to the defined class are called *courts*. An example that describes a collection of resources is:

```
<rdf:Bag rdf:ID="docs-apply">
  <rdf:li
      rdf:resource="file:///lucru/teza.docx" />
  <rdf:li
      rdf:resource="https://www.upit.ro/\_doc/8806/a_27_c_taxe.pdf" />
  <rdf:li
     rdf:resource=" https://www.upit.ro/\_doc/11836/proc_mencs.pdf" />
</rdf:Bag>
```

New versions CKML have included the ideas and techniques on the informational flow (IF - *Information Flow*) and the design of the logic of the distributed systems. The final version CKML is both a language based on the logic of the information document and a language based on frames. In accordance with Kent(2000), "in CKML the specification requires the use of the concept of mathematical lattice or the most practical notion of conceptual space" [10]. The basis of the theoretical portion of the practice based on CKML is the CKP Theorem which states *the equivalence between data structures of type conceptual lattice and formal context (classification).*

OML provides three levels of further specify the restrictions [26]: top - sequences (corresponding informational flow); the intermediate pipe - calculation of binary relations; Lower logical expressions (corresponding to concept graphs).

Expressing an ontology is possible using the languages of specification such as [13]: KIF (Knowledge Interchange Format), CL (Common Logic), OIL, DAML+OIL AND ALLURE.

KIF is based on the logic of the predicates [25], but provides a LISP oriented syntax for this. From the point of view of the semantic, there are four categories of constant in KIF: constant of type object, constant of function type, constant of relation and logical constant.

OIL (Ontology Inference Layer [7]) extends RDF diagram to provide an intuitive syntax and a great power of expression and a semantics more clearly defined with easy to use descriptive logic within the framework of the schemes of reasoning. Such OIL brings together and unifies three directions: descriptive logic, modeling based on frames and modeling RDF/XML.

(DAML DARPA Agent Markup Language) + OIL has a syntax diagram type RDF, that inherits the primitives of RDF (subclass, domain, range) and primitive added extras like transitivity, cardinality etc. Schematic DAML+OIL is oriented on the objects in which the concepts are abstracted by grades and roles through the properties of the objects. Thus, the ontological model DAML+OIL is based on a lot of the axioms about the classes and properties, as well as a set of builders very useful from the perspective of the RPC systems [13]: intersectionOf; unionOf; complementOf; oneOf; toClass; hasClass; hasValue; minCardinalityQ; maxCardinalityQ; cardinalityQ.

The result of the foregoing the evolutionary process is [13]: 1) OIL extends RDF; 2) DAML extend RDF; 3. DAML+OIL DAML integrates and OIL and extends the RDF; 4) ALLURE extends DAML+OIL and RDF.

The final result of the research on ontological modeling using RDF/XML has led to the specification of the ALLURE, in three versions [13]: ALLURE LITE (simple hierarchy, hierarchy of classes with simple constraints), ALLURE DL (maximum expressiveness) and ALLURE FULL (very expressive). For the processing of meta-data described using specific Markup ontologies have been developed a variety of tools for annotation, navigation, utilities (API), edit, view graphics, marking, pan, validation, import, export, compilation, query, search etc. A list of them would be too long. We will be limited to the most important tools, the rest being described in the references indicated: DUET (DAML UML Enhanced Tool), UBOT, The Platform Protégé, and Ontolingua. Ontolingua Editor allows for the creation of ontologies, exploration and collaborative editing. Using Ontolingua, it is possible to export and import formats like: KIF, DAML + OIL, OKBC, Prologue, the LOOM, Ontolingua and CLIP. Can import data in the *protégé* format.

3 SCXML and Voice XML

SCXML provides a generic state-machine, an execution environment based on CCXML and Harel State Tables, according to W3C(2015) in [30]. Also in [11] it is mentioned that: "using SCXML as the representation of the state machine is seen as a benefit". The mentioned authors found that "large portions of the SCXML standard are not necessary for it to be useful to our customers and us." CCXML is designed to upgrade VXML dialog systems with advanced telephony functions. An example of the SCXML representation is for speech recognition in the natural language. For the implementation of the KRP systems, the role of the SCXML is active in the framework of the failures, through voice and natural language.

According to the above considerations, it was our choice to propose the usage of VXML to create voice-enabled applications [29]. VoiceXML (VXML) is a markup language for specifying the vocal dialog between a man and a software application, for example a KRP system. Thus, using VoiceXML 2.0 one can develop KRP applications which provides automatic recognition of speech (ASR - Automated Speech Recognition) and interactive vocal response (IVR - Interactive Voice Response).

The main elements of voiceXML are:

- <vxml> - start/close any vxml document;
- <var>, <assign>, <clear> used to declare, assign and delete variables;
- <grammar> to specify the grammar of the text under recognition;
- <catch>, <throw>, <error>, <noinput>, <nomatch> to manage exceptions;
- <menu>, <choice>, <enumerate> to deal with menu;
- <if>, <else>, <elseif> to describe conditional aspects;
- <initial>, <form>, <field>, <filled>, <option> to process forms;
- <prompt>, <reprompt>, <value> for input operations;
- <prompt>, <audio>, <record>, <reprompt> to deal with multimedia entities;
- <block> to describe the code to be executed;
- <disconnect>, <exit> for the management of the sessions;
- <meta>, <metadata> for metadata management;

- <noinput>, <nomatch>, <help> to manage events and actions;
- <subdialog>, <goto>, <return>, <link> for dialog control;
- <object>, <property>, <param>, <script>, <submit>, <transfer>, <log> for server oriented processing of parameterized queries.

Vxml applications may be of the type uni - or many - document. An application many - document allows us to define a root document which defines all the entities visible in and recovered by the documents son. VXML applications are oriented to the following categories:

- *Queries* - to retrieve information from Web-based infrastructures (like voice portals, web call centers);
- *Transactions* - to execute specific transactions with a Web-based back-end database.

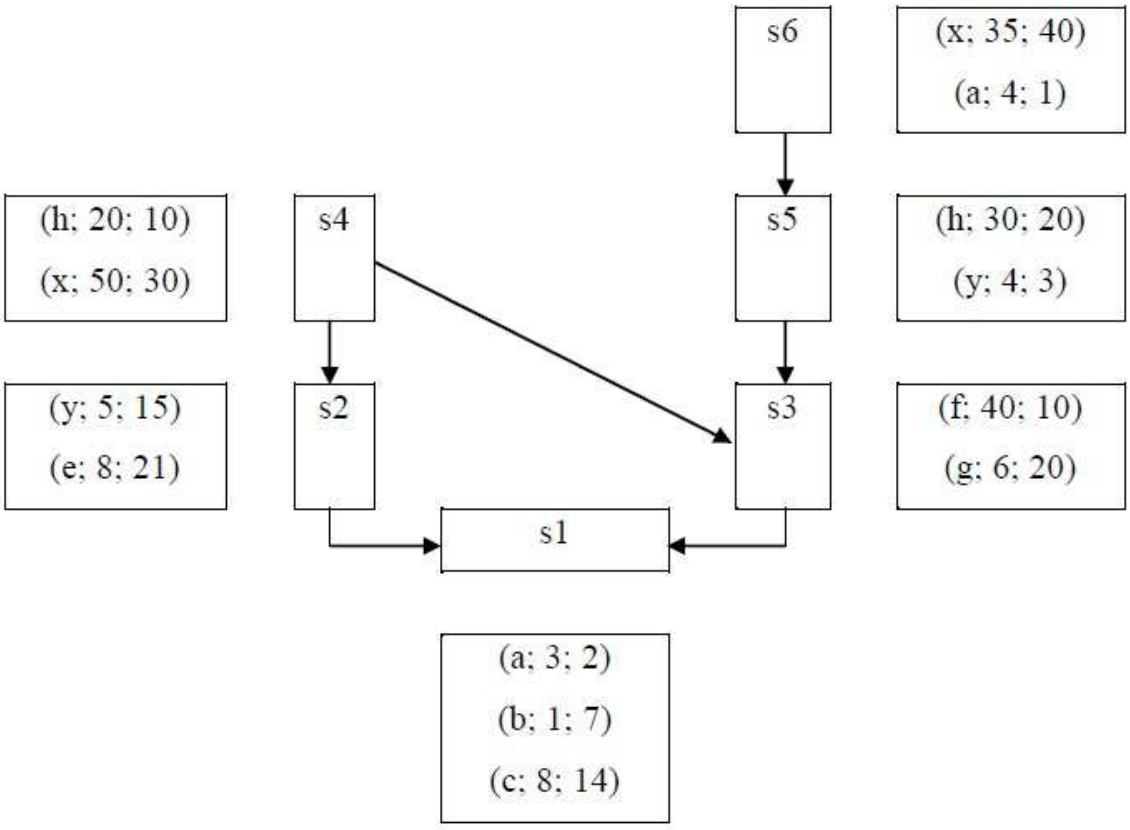


Figure 1: A Knowledge database - inheritance and its graph representation

To optimize the development of voice based user interfaces, the following facts should be understood:

1. A VXML application is a single VoiceXML document, or a set of documents which forms a *conversational finite state machine* (CFSM). The root document is loaded firstly and remains loaded while transitioning over documents belonging to the same application.
2. A *session* is opened by the user to start the interaction with the VXML interpreter, and is closed by a request from the user, a document, or the interpreter itself.

3. VXML has two types of dialogs: *forms* and *menus*. Each dialog has associated one or more speech and/or DTMF (Dual-Tone Multi-Frequency) grammars.
4. A *form* defines an interaction that obtains values for a set of variables.
5. A *menu* provides the user with some alternative options and follows other dialogs depending on the selection.
6. A *subdialog* is like a function call; after interaction returns to the original form.
7. There exist two types of grammars in VXML: machine directed (the form items are executed in the sequential order) and mixed initiative (the flow has to be directed both by the user and by the application).

The example described below is a skeleton query type using the voice-based interface when querying inheritance graphs of knowledge is considered, as [18, 19], proposed for text interfaces. This model can be extended for the implementation of the interfaces based on voice within web-based KRP systems.

To demonstrate the basic principles of voice-based interfaces, a simple knowledge database (described in Fig. 1) is considered. Based on a client-server implementation in Java, a dialog for the computation of the certitude factor is shown:

```
private Object[][] b3 = {
{"s1", "s2 33","atr(a,3,2) atr(b,1,7) atr(c,8,14)"},
{"s2", "s4","atr(y,5,15) atr(e,8,21)"},
{"s3","s4 s5","atr(f,40,10) atr(g,6,20)"},
{"s4","","atr(h,20,10) atr(x,50,30)"},
{"s5","s6","atr(h,30,20) atr(y,4,3)"},
{"s6","","atr(x,35,40) atr(a,4,1)"}
};

Dialogue: The first step for client1
>>> Welcome. Ready for you!
>>> Select your knowledge base: b3
---  The knowledge base b3 was selected for processing.
>>> Select the object to investigate: s1
--- The object s1 was selected.
>>> Select the attribute under investigation: h
--- Your choice was the attribute h.
Answer: The value of attribute h of object s1\\
 is 30 and the certitude factor is fc=20.

VXML code:
<prompt>Welcome. Ready for you!</prompt>
<prompt count="1">
Select your knowledge base: <value expr="KnowledgeBase"/>?
</prompt>
// Javascript Code
<prompt> The knowledge base </prompt>
// Code to print --KnowledgeBase--
<prompt> was selected for processing. </promt>
<prompt count="2">
```

```
Select the object to investigate: <value expr="Object"/>?
</prompt>
// Javascript Code
<prompt> The object </prompt>
// Code to print -- Object Name--
<prompt> was selected. </promt>
<prompt count="3">
Select the attribute under investigation: <value expr="Attribute"/>?
</prompt>
// Javascript Code
<prompt> Your choice was the attribute </prompt>
// Code to print --Attribute--
<prompt>. </promt>
```

4 Markup technologies

From the point of view of the process of XML tying - JAVA implemented by JAXB, it is noticed the existence of two major components [24]: a generator of diagrams and compiler of diagrams and the process actually involves tying seven actions: the generation of classes; the compilation of classes; unmarshal (XML documents which satisfy the restrictions in the diagram source are processed by the JAXB. Also, JAXB lets you transfer XML data from sources other than the files and XML documents such as the nodes DOM, paintings rows of characters, SAX sources and so on and so forth); the generation of the shaft into which describes the contents of an XML document; validation (Unmarshal process involves the validation of the source before generating shaft into which describes the contents. Where there is a change in the shaft in the next step can be used the operation JAXB validation to confirm the changes before to transform the contents into a document XML); the client application may change the XML data represented by JAXB shaft using the interfaces generated by the compiler JAXB; marshal (the shaft that describes the contents is converted into the XML document. The content can be validated before the conversion. A process called "Marshalling" offers a client applications the ability to convert a Java derived from JAXB in data XML.)

With a force greater than the programming, can be used JAXP technology (Java API for Processing XML) based on SAX (Simple API for Parsing XML) and DOM (Document Object Model). During the operation of the "parsing" based on SAX it generates events that notify the components identified, and Java application must deal with the events of the callback methods (for the construction of the structure of the database). The operation of parsing DOM build in the memory a representation tree diagram of the data from the XML document. JAXP technology allows the transformation of XML documents using XSLT technology (Extensible Stylesheet Language Transformation).

XMLBeans technology is used to compile the XML layout with obtaining in memory, the classes, and has been developed in the period 2003-2014 by the Apache Software Foundation to enable the processing of large structures.

Therefore, for the processing of databases structured knowledge which comply with a diagram and are stored in XML files may be used JAXB technologies facing on the diagram XML and JAXP facing on the direct rendering of documents XML. JAXP is a good choice for large knowledge database to be processed in terms of low computational capacity.

The following example shows how a XSLT stylesheet is used to transform a sample data set into VoiceXML 2.0 format.

```
<?xml version="1.0"?>
<phdstud>
   <stud>
      <pid>06</pid>
      <uni>University of Pitesti</uni>
      <phds>Ali Amer Mohameed Saeed</phds>
      <year>2017</year>
   </stud>
   <stud>
      <pid>107</pid>
      <uni>University of Bucharest</uni>
      <phds>Radu Mihai</phds>
      <year>2017</year>
   </stud>
</phdstud>

<vxml version="2.0">
<form id="start">
 <audio>Some PhD students </audio>
 <xsl:for-each select="phd">
       <audio>PhD student id is <xsl:value-of select="pid" /></audio>
         <break time="100ms"/>
         <audio>Comes from PhDsch.<xsl:value-of select="uni"/></audio>
         <break time="100ms"/>
         <audio>The PhD name is <xsl:value-of select="phds" /></audio>
         <break time="100ms"/>
     <audio>Year of defence is <xsl:value-of select="year"/></audio>
         <break time="100ms"/>
  </xsl:for-each>
  </form>
</vxml>
```

5 KRP systems in education

According to [16], a KRP system for Artificial Education (AE) should take into consideration four elements. In AE, the first element, "knowledge would include knowledge of pedagogy (teaching practices and beliefs), curriculum, and knowledge regarding the individual student's needs, assessments, evaluating, and more". The second element is connected to problem solving. In this context, the KRP system should "look at past successful and unsuccessful pedagogies used with individual student, and it would be able to present instructional material to that specific student in a way the benefited him or her individually". The last two elements are connected to developers and administrators, but the mentioned authors did not conclude on smart interfaces for KRP educational systems. However, they emphasize on Intelligent Tutoring Systems (ITS), but ITS are "emphasizing those aspects which have relevance to user support, rather than detailed consideration of the merits of pedagogical or student knowledge modelling strategies" as shown by Hefley & Murray (1993) in [8].

Following Horvitz(1999), an intelligent user interface should consider imprecision and uncertainty aspects during run-time [9]. This is more important in AE, due to the nature of queries

formulated by learners. As Salih(2014) mentioned [15], the Natural User Interfaces (NUI) will be the next generation of user interfaces to improve user experiences. Our proposal is based both on Artificial Intelligence Techniques to deal with imprecision/uncertainty and natural language aspects with speech understanding and knowledge restructuring for fast answering systems.

Therefore, any KRP system for education should consider preliminary requirements to understand the learner's behaviour, markup models and technologies to implement solutions to queries given in "approximate" natural language by learners. One KRP system for education should be able to represent not only pedagogical aspects, but also, different variants of content, and appropriate behaviour according to the learn initiatives.

Smart interfaces of KRP systems for education are based on Voice-enabled applications to support e-learning in many ways, making possible the usage of e-learning systems by visually impaired users. In Web-based e-Learning systems, the output is generated in HTML format. In order to support Voice type output, one step more is required to translate HTML to VXML. In the following, only a short example for translating a table is given. Only the VXML code is shown.

```
<?xml version=''1.0''?>
<vxml version=''2.0''>
..
<block> The next structure is table 1 </block>
<block> The table name is </block>
// Code ...
<block> Row 1 </block>
<block> Column 1 </block>
<block> E[1][1] </block>
<block> Column 2 </block>
<block> E[1][2] </block>
<block> Row Ending 1 </block>
<block> Row 2 </block>
<block> Column 1 </block>
<block> E[2][1] </block>
<block> Column 2 </block>
<block> E[2][2] </block>
<block> Row Ending 2 </block>
<block> This the ending of table 1 </block>
</vxml>
```

6 Conclusions

This work has analyzed the detailed rules for the description of the information structured used in context of KRP systems, using markup languages. Some markup technologies based on Java are considered.

The best choice is a model of the XML, and from the point of view of the java technologies for the processing of XML documents, it is found that for practical application JAXB (object interrogation, processing in memory) and JAXP (linear, facing processing on the fragments identifying and dealing with events) are more appropriate to be used. The effort of JAXB programming is less and object processing is promoted.

In addition, by Voice XML can be describes the smart interfaces of the KRP systems based on voice.

Bibliography

[1] Angles, R.; Gutierrez, C. (2005); Querying RDF Data from a Graph Database Perspective, In: *Gómez-Pérez A., Euzenat J. (eds) The Semantic Web: Research and Applications. ESWC 2005, Lecture Notes in Computer Science, vol 3532*, Springer, Berlin, Heidelberg, 2005.

[2] Berners-Lee, T. (1989); *Information Management: A Proposal*, CERN, [Online] Available: https://www.w3.org/History/199/proposal.html, Accessed on December 20, 2016.

[3] Buraga, S. (2004); *Semantic Web. Fundamente şi aplicaţii*, Matrix Rom, Bucureşti, 2004.

[4] Coleman, J.; Willis, D. (1997); *SGML as a Framework for Digital Preservation and Access, Commission on Preservation and Access*, Washington DC, 1997.

[5] Dănciulescu, D. (2015); Formal Languages Generation in Systems of Knowledge Representation Based on Stratified Graphs, *Informatica*, 26(3), 407-417, 2015.

[6] Dănciulescu, D.; Colhon, M.; Grigoraş, G. (2017); Right-Linear Languages Generated in Systems of Knowledge Representation based on LSG, *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, 8(1), 42-51, 2017.

[7] Fensel, D.; van Harmelen, F.; Horrocks, I.; McGuinness, D.L.; Patel-Schneider, P.F. (2001); OIL: An Ontology Infrastructure for the Semantic Web, http://www.cs.man.ac.uk/~horrocks/Publications/download/2001/IEEE-IS01.pdf (Last visited on 9/09/2017).

[8] Hefley, W.E.; Murray, D. (1993); Intelligent User Interfaces, In Proceedings of IUI'93, Orlando, Florida, ACM Press, NY, 3-10, 1993.

[9] Horvitz, E. (1999); Principles of Mixed-Initiative User Interfaces, In Proc. of CHI, 159-166.

[10] Kent, R.E. (2000); Conceptual Knowledge Markup Language (CKML): An introduction, Netnomics 2, 139-169.

[11] Kistner, G.; Nuernberger, Ch. (2014); Developing User Interfaces using SCXML State charts, *NVIDIA*, Publication Rights Licensed to ACM, http://phrogz.net, 1-7, 2014.

[12] Meenakshi, A.; Aghila, R.; Suganthi, P.; Kavya, S.(2016); A Knowledge Representation Technique for Intelligent Storage and Efficient Retrieval using Knowledge based Markup Language, *Indian Journal of Science and Technology*, 9(8), 1-8, 2016.

[13] Mohameed Saeed, A.A. (2017); Intelligent Interfaces for Knowledge Representation and Processing Systems, *Proceedings of ICVL 2017*, 370-375. 2017.

[14] Negru, V.; Grigoraş, G/; Dănciulescu, D. (2015); Natural Language Agreement in the Generation Mechanism based on Stratified Graphs, *Proceedings of the 7th Balkan Conference on Informatics Conference (BCI '15)*, ACM, New York, NY, USA, Article 36, 1-8, 2015.

[15] Salih, D. (2014); Natural User Interfaces, LM Research Topics in HC, http://www.cs.bham.ac.uk/, 2014.

[16] Sora, J.C.; Sora, S.A. (2012); Artificial Education: Expert systems used to assist and support 21st century education, *GSTF Journal on Computing (JoC)*, 2(3), 2012.

[17] Sowa, J.F. (2008); Conceptual Graphs, In *F. van Harmelen, V. Lifschitz, and B. Porter (Eds.): Handbook of Knowledge Representation*, Elsevier, 213-237, 2008.

[18] Ţăndăreanu, N. (2000); Knowledge Bases with Output, Knowl. Inf. Syst., 2(4), 438-460.

[19] Tăndăreanu, N.(2007); Communication by Voice to Interrogate an Inheritance Based Knowledge System. Research Notes in Artificial Intelligence and Digital Communications, *7th International Conference on Artificial Intelligence and Digital Communications*, 107, 1-15, 2007.

[20] Vohra, A.; Vohra, D. (2006); Pro XML Development with Java Technology, Apress.

[21] [Online] DAML tools; Available: http://www.daml.org/tools/, Accessed on December 15, 2016.

[22] [Online] DLML; Available: http://co4.inrialpes.fr/xml/dlml/, Accessed on December 15, 2016.

[23] [Online] HTML; Available: https://en.wikipedia.org/wiki/HTML, Accessed on September 15, 2016.

[24] [Online] JAVA XML parsers; Available: http://docs.oracle.com/javase/8/docs/api/index.html, Accessed on December 15, 2016.

[25] [Online] KIF; Available: http://www-ksl.stanford.edu/knowledge-sharing/kif/, Accessed on September 15, 2016.

[26] [Online] Ontology Markup Language; Available: http://www.ontologos.org/OML/OML%200.3.htm, Accessed on December 15, 2016.

[27] [Online] RDF/XML; Available: https://www.w3.org/TR/rdf-syntax-grammar/, Accessed on December 15, 2016.

[28] [Online] SGML - ISO 8879:1986; Available: https://www.iso.org/obp/ui/#iso:std:iso:8879:ed-1:v1:en, Accessed on December 15, 2016.

[29] [Online] VXML 2.0; Available: https://www.w3.org/tR/voicexml20/#dml1.4, Accessed on December 15, 2016.

[30] [Online] W3C (2015); State Chart XML (SCXML): State Machine Notation for Control Abstraction, Available: https://www.w3.org/TR/scxml/, Accessed on December 15, 2016.

[31] [Online] XML COVER PAGES; Available: http://xml.coverpages.org, Accessed on December 15, 2016.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 13(1), 129-142, February 2018.

Tracing Public Opinion Propagation and Emotional Evolution Based on Public Emergencies in Social Networks

H. Wei-dong, W. Qian, C. Jie

Huang Wei-dong*
Nanjing University of Posts and Telecommunications
Nanjing, Jiangsu, Peoples R China
*Corresponding author:huangwd@njupt.edu.cn

Wang Qian
Nanjing University of Posts and Telecommunications
Nanjing, Jiangsu, Peoples R China
18262621679@163.com

Cao Jie
Nanjing University of Information Science and Technology
Nanjing, Jiangsu, Peoples R China.
cj@amss.ac.cn

Abstract: Social network has become the main communication platform for public emergencies, and it has also made the public opinion influence spread more widely. How to effectively obtain public opinions from it to guide the healthy development of the society is an important issue that the government and other functional departments are concerned about. However, the interaction and evolution mechanism between the subject and the environment in the public opinion propagation is complicated, and the public and media attention and reaction to the incident are closely linked with the progress of the incident disposal. And public mining corpus has some shortcomings in the distribution of emotional classification. Only the timely update of artificial rules and emotional dictionary resources, it can handle new text data well. In fact, from the perspective of public opinion propagation, this paper built the network matrix between Internet users through the forwarding relationship, and used the social network analysis method and the emotion mining analysis technology to study the interaction and evolution mechanism between the subject and the environment in the public opinion propagation, and it studied the role of users in the emotional propagation of social networks. This paper proposed a sentiment analysis method on the micro-blog platform, which expanded the emotional dictionary and took sentence and emoticon and sentence patterns into account, which improved the accuracy of positive and negative classifications and emotional polarity analysis of the micro-blog.
Keywords: opinion mining, the public opinion propagation, semantic classes, machine learning.

1 Introduction

People gather on social networking sitesdue to the similar values and habits, and they become the realistic social digital projection. Based on modern information dissemination of digital social network, on the one hand, information dissemination speed and efficiency of social network have been greatly enhanced; on the other, information sharing based on social relations provides an effective filtering mechanism to spread a lot of information on the Internet. Thus, social networking sites become the most important platform for information sharing and dissemination [3].

Public opinion phenomenon on social network and its development laws are very rich and complicated. Especially, the subject of social network is the people who are with highly intelligence and adaptive capacity, and its cognition and decision-making behavior is the adaptive process of continuous evolutionary learning, and they changed their behavior through interaction with other subjects as well as the environment [19]. This complex adaptive and learning mechanism is projected to digital social network, and its propagation mechanism has the dynamic characteristics of the time-varying dynamic evolution, complexity and adaptability in the subject created more profound complexity of the system [18]. Therefore, it is necessary to explore an effective method to elaborate its run and propagation mechanism. And netizens are always generating massive amounts of data. With the characteristics of rapid growth, structural diversity [21], dynamic updating and wide ranges, these data contains public emotional information of all levels of society. Mining the public emotional information makes sense for the research of information retrieval, electronic commerce and public opinion supervision.

In recent years, scholars have carried out relevant research on the public opinion propagation mechanism. FeiXiong [16] studied the diffusion of micro-blog information based on forwarding mechanism and proposed the information diffusion model and verified the correctness of the model by micro-blog hot event. Lee [5] applied density-based online clustering method to mine text flow to assess the influence of public opinion events to achieve the goal of situational awareness and risk management; Zhao Haiqing [22] analyzed the nature of the complex network of public opinion propagation, the study found that: the key node in the core position has a high degree of activity and participation, which should be highly valued by network public opinion monitors and leaders; Su Xiaoping [9] put forward a local central measurement method, and the measurement results show that in the social network, the importance evaluation of nodes depends on both the centrality of the nodes in the community and "bridging" nature between the community.

At present, the researches on emotional analysis mainly focus on the emotion excavation and calculation of the information released by Internet users in the Internet [12], [13], [14], as well as the analysis the related factors of emotion, such as the stock market volatility the political election results and movie box office, which are predicted by Internet emotions. Riloff [8] proposed to construct a emotion dictionary based on a large number of corpus statistical analysis in order to classify emotion; Wiebe [15] added part-of-speech analysis based on the emotion dictionary and completed the subjective and objective classification of corpus based on Bootstrapping algorithm; Xu Lin-hong [17], based on primary school textbooks, screenplays, literary periodicals and fairy tales, annotated a large amount of Chinese corpus, and formed a "emotional vocabulary ontology library". Although it has been widely used in the field of emotion analysis of Chinese texts, its ontology has some shortcomings in the distribution of corpus and genre Only the timely update of artificial rules and emotional dictionary resources, it can handle new text data well.

So, this paper extended the emotion dictionary and considered the impact of sentence patterns on sentence sentiment in order to achieve more accurate results.

This paper attempted to focus on specific social hot spots and measured the network structure of the public opinion, mined the characteristics of the network structure, and researched the interaction and evolvement mechanism between the main body and the environment in the public opinion propagation. And it introduced the emotion analysis into the evolution of public opinion events, explored the reasons for the sharp fluctuations in emotion throughout the evolutionary life cycle of public opinion, and then provided the decision-making basis for the control of Internet public opinion and the dynamic grasp of public opinion.

2 Research on the propagation mechanism of public opinion in social network

The emergence of social networks provides people with a new information media, but also greatly increases the possibility of the outbreak of public opinion. Public opinion propagation on social networks is influenced by many factors. This paper mainly studied the three aspects of the subject elements of public opinion propagation in social networks, propagation networks and emotional tendencies.

2.1 The subject elements of public opinion propagation in social network

Netizens group is the main body of public opinion in the Internet. They have also provided most of the impetus, and they are the main generating force and affected objects of network public opinion. Representative of netizens are mainly divided into the following four groups: grassroots, net-a-porter, network opinion leaders and web hyper.

Grassroots: The grassroots is the most active basic crowd of netizens in china. When grassroots gathered together due to a topic or event that causes a resonance or mood fluctuations, the group leaders assert, repeat, and infect constantly to act on the individual in the way of implication and cross-infected, and they will turn to a common direction, and will immediately turn this concept or belief into action tendencies, so that emotions are greatly vented.

Net-a-porter: Net-a-porter is the netizen who is engaged in forwarding and spreading network information. The existence of the net-a-porter makes network voices amplified infinitely, and the information gets fully widely spread and exchange [10].

Network opinion leaders: The behaviors of network opinion leaders are very active and they have excellent ability of expression. For the majority of netizens, they have a strong influence and tractive effort, and the network has decided to focus on the focus and direction of opinion to a certain extent.

Web hyper: The web hyper is a network planner who popularizes enterprise products, customer brands, events and individuals with the help of the network medium. This purely commercial purpose makes the Internet more complex, and they render it through the secondary reports of the traditional media, so that more netizens who are unknown truth are generally involved, and the topic or the event is very easy to be hype, and even become the network killers that hit businesses and slander the individuals [23].

2.2 Individual contact process of propagation

Because of the initiative and interaction of the individual participation, network propagation has special advantages in the aspects of information acquisition and propagation. Individuals in the interpersonal communication are not always passive acceptance of information, and it is not like the real life of the traction of the mass media "agenda setting", and they also play multiple roles, such as publishers and communicators. Because of the strong concern, in fact, some individuals have developed into similar network Medias that have the extraordinary influence. The gap between the network propagation media and the network space is very weak, and they together constitute the diverse individuals in the network space. The interactive propagation among diverse individuals has become the main way to spread the information of network public opinion. In fact, the macro level of diffusion process of network public opinion information is the diverse individuals contact process through the micro level. Specifically, in the emerging social media that represented by social network and micro-blog, emergency network public opinion information always released firstly by a few individuals in the network, and then they were

forwarded or shared by other individuals in order to communicate one relatively stable group of receiving information. Through a virtual social network, "infected group" scale of network public opinion has been expanded, and then they have the social power that cannot be ignored [1].

2.3 The design of analytic framework

To analyze the evolution of public opinion of the social network, it needs to analyze the characteristics of the whole network structure and the characteristics of the key nodes from the two levels. Considering the whole development evolutionary characteristics of public opinion events, we need the evolutionary analysis of time series that is aimed at network event scale, density, average degree, clustering coefficient and so on [2], [4], and we focused on the evolutionary process of public opinion of rapid growth stage and the mature stage and calculated the network structure and the network properties of the important nodes in order to explore the evolution rules of public opinion of the whole network. To understand the emotional state held by users in the process of event information, the user's emotions in the propagation network need to be judged. Based on the above research, this paper first crawled and texted the specific events studied, and divided the text into clauses. In the second step, to segment the text, we needed to create a dictionary of special-purpose emotion words that belongs to the event, based on the existing definition of thesaurus, and marked the sentiment classification of the emotion words, polarity value and its intensity. The third step, combined with the sentence, expression and emotional words of this paper gave the emotional tendency of the text and emotional polarity values.

3 Construction and analysis of public opinion propagation network in social network

The development of public opinion events is always carried out in the dimension of time. After the formation, proliferation, explosion and termination of public opinion, the social network of public opinion presents the communication of user emotion, attitude, intention and view in turn. On the basis of the research on the overall development of public opinion event of the "Shandong illegal vaccine", in this paper, the social network analysis method based on time series was used to segment the daily relational data in the process of public opinion social network and form the public opinion social network of daily data. Then we analyzed the overall network structure, the evolution of network characteristics and the evolution of key nodes in these different stages.

3.1 Data source collection and overall structure analysis

On March 18, 2016, an article entitled "Billions of yuan of vaccine without refrigeration flows into 18 provinces" caused widespread comments and forwarding from users. Micro-blog users expressed their opinions, including celebrities, media, government and other users. The Shandong illegal vaccine case caused public opinion to continue to pay attention. In this incident, what was the attitude of different types of users to events? What was the direction of public opinion? What role did the different users play? Based on this, a study was conducted on the "illegal vaccine". In this paper, the "illegal vaccine" as the keywords for micro-blog search, we collected data from March 18, 2016 to April 4, 2016, and used web crawler software for micro-blog data crawl, and obtained the data through the preparation of crawling rules, a total of 10 035 data collection.

Data were cleaned and pre-processed by Excel. Duplicate data, garbled data, incomplete data, and data without records were deleted. After cleaning, 8,222 data were obtained, and we

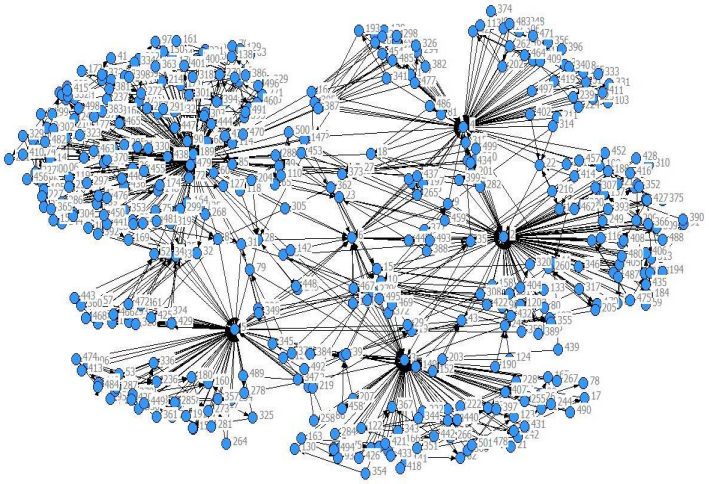


Figure 1: The forwarding network visualization

teased out the micro-blog's forwarding relationship through Sina micro-blog's forwarding rules: if A forwards B, the value of A <-> B is 1; if the two are not forwarded, the value is 0; and the main diagonal value of the matrix is defined as 1. We constructed the public opinion network matrix of "Illegal Vaccine in Shandong". Based on the constructed matrix, visualization software [20] was used to process the forwarding network of the event, and the visualization result was shown in Figure 1.

3.2 Whole characteristics evolution analysis

To analyze the evolution of the overall characteristics of the selected cases, the index of public opinion network of daily data of the event needed to be analyzed, according to the trend and regularity of daily public opinion network, and the law of the overall network topological index with event changes was explored.

(1) The evolution of network density

The network density is used to describe the close degree of node interaction. The greater the density of the whole network is, the more frequent the interaction between nodes is. The blue curve in fig.2 showed the close degree of interaction of each user daily posting and replies in the public opinion event, and the green curve indicated the change of delta value of each variable in two days. It can be found from the curve trend: from March 19 to the end of March, network density had remained flat changing trend. But in April, this topic again appeared fluctuations.

(2) The evolution of network average degree

The blue curve in Figure 3 showed daily users' network average degree of the public opinion event, and the green curve indicated the change of delta value of each variable in two days. From the point of curve trend, the development trend of the network average degree was fluctuating. From March 19 to the end of March, the public opinion event was gradually come to an end from the outbreak of public opinion, the value of network average degree appeared larger fluctuation, especially in the period of hot topic.

The average degree of the whole network became larger and each node had more surrounding nodes for forwarding the interaction, and the communication was more extensive. Namely between March 22-24, the network average degree was higher, it indicated that during this period,

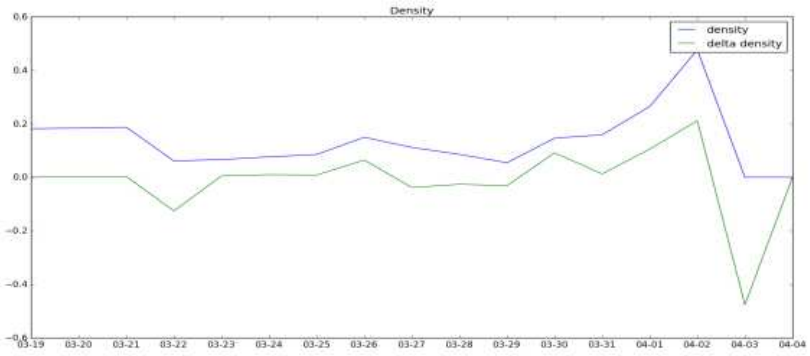


Figure 2: The diagram of network density evolution

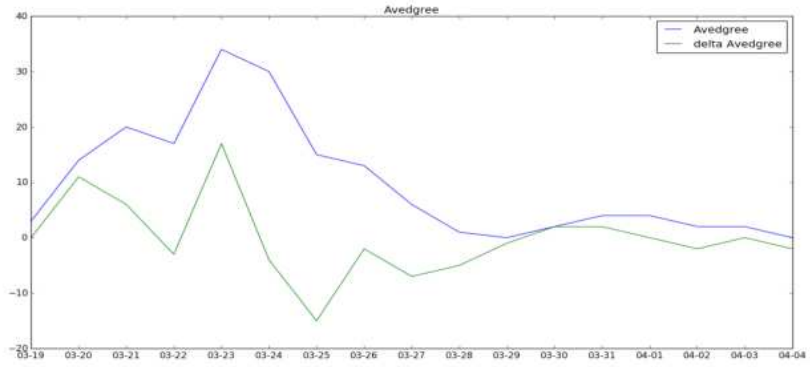


Figure 3: The diagram of the evolution of network average degree

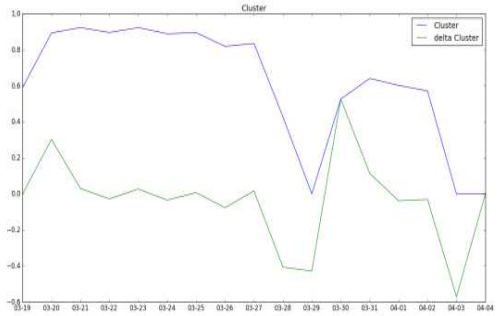


Figure 4: The diagram of the evolution of network clustering coefficient

many users on the network were repeatedly posting and replying, it may be for a topic or a post and caused a lot of participants to discuss, we can judge, there may be potential opinion leaders.

(3) The evolution of network clustering coefficient

If a node j in the network has nj neighbor nodes, then there may be a maximum of $nj(nj-1)/2$ edges between the nj nodes. We define the ratio of actual number of edges $E(j)$ between nodes nj and the possible number of edges $nj(nj-1)/2$ as the clustering coefficient of nodes $CC(j)$, that is:

$$CC(j) = 2E(j)/[nj(nj-1)]$$

The clustering coefficient CC of the whole network is the average of the clustering coefficients of all nodes$j$:

$$CC = \frac{\sum_{j=1}^{N} CC(j)}{N}$$

Network clustering coefficient describes the degree of network collectivization [6], that is, it is used to identify whether there is a central tendency in the network. Whole network clustering coefficient is bigger, the network of centralized trend is higher, and the users of public opinion tend to the interaction with small groups [7]. The blue curve in fig.4 showed the daily change of clustering coefficient of the network public opinion, and the green curve indicated the change of delta value of each variable in two days. From the point of curve trend, after the outbreak of public opinion from March 19, it was the first time to achieve the highest value of network clustering coefficient on March 20, and then it entered into a state of smooth fluctuation and maintained between [0.6, 1]. After March 27, the development of public opinion was into a smooth state, the number of people involved in public opinion and the interaction frequency dropped. In the small batch of node communication, the network clustering coefficient was the same as the network density, which will show a large fluctuation phenomenon.

3.3 Key nodes characteristics evolution analysis

In the research of the key nodes in the social public opinion network, scholars often used the static structure characteristics of social network, such as the degree centrality and the structural holes, to analyze the differences in grades and advantages between the key nodes and other nodes in the network. Therefore, based on the public opinion event of "Shandong illegal vaccine",combined with the life cycle theory, we analyzed the characteristics of key nodes in different stages, so as to study the characteristics evolution of network public opinion.

According to the event of "Shandong illegal vaccine", we focused on the characteristics of the key nodes in the rapid growth stage and the mature stage. Statistics showed that the number of people involved existed obvious periodic: after the outbreak of public opinion from March 19, it reached the climax of the public opinion firstly on March 22, and then the growth rate of public opinion had been reduced. It reached the biggest outbreak of the whole public opinion event on March 24, and then it began to subside slowly. Based on this, we set the stage of rapid growth from March 19th to March 22nd, the mature stage, from March 23rd to March 24th.

Degree centrality

It can be seen from the diagram of the propagation network, the propagation of public opinion presented divergent trend around the individual points, in the propagation network of public opinion, and it showed that there was a strong interaction space around some topics or some

users. Through measuring degree centrality of this nodes, it can be seen that "The concept of tide", "Friendly 88", "Swineherd is officer","bing-lv" and other users' centrality degrees are bigger, and their participation was relatively high, and it showed some characteristics of opinion leaders. In the two stages of the topic development, different nodes played an important role in the network. In other words, the node was only active before a certain stage, which was not coherent. Therefore, the evolution of a single node cannot be analyzed. Therefore, this section focused on the study of the network structure changes in different stages and the performance of nodes in this stage.

Structural holes

Structural hole refers to a non-repeating relationship between two nodes, which is equivalent to a buffer in a network.The value of the contribution to the network of the two related people who exist structural holes between nodes can be accumulated. By using UCINET software, we analyzed structural holes in public opinion propagation network of "Shandong illegal vaccine".

It can be seen from the table 1, the public opinion propagation network of "Shandong illegal vaccine" existeda large number of structural holes. We sorted out the effective scale, and "Golden Orchard fruit", "qfy180" and "black feeling" ranked the top three. The value of effective scale reflected the position of nodes in the propagation network, and the bigger the value is, the more core the position becomes. The global constrain of these three nodes was relatively small, they were less than 0.2, and it also reflected that the three nodes were not easy to be controlled by other nodes and easier to access and spread public opinion. As can be seen from the results of the structural hole in the mature stage, there were still a large number of structural holes in the network of public opinion of "Shandong illegal vaccine". We sorted out the effective scale, and "Happy stock886644", "The Lord of Peach Blossom Island-6" and "Dawn is now tomorrow" ranked the top three. By comparing with the rapid growth stage, it reflected the similar characteristics of the network in different stages of the topic.

Table 1: The structural holes of rapid growth stage

Node	**Degree**	**Constrain**	**Density**
Golden Orchard fruit	25.000	0.178	0.850
qfy180	25.000	0.178	0.850
black feeling	37.000	0.117	0.857
Open the treasure to eat	31.000	0.131	0.677
Friendly 88	36.000	0.122	0.943
I don't know at all	22.000	0.184	1.390
haohuanlea	11.000	0.350	0.764
pour out words	37.000	0.117	0.857

The analysis showed that in the evolution of network density and network clustering coefficient, the events of rapid growth and maturity showed a relatively steady state, and the trend of its delta curve was basically the same as the original curve. By comparing the network attributes of rapid growth stage and mature stage, the density of networks in mature stage was larger, which indicated that in the mature stage, the overall participation and interaction of network users were more frequent, the propagation distance was shorter and the communication ability was the strongest. In the two stages of topic development, different nodes played an important role in the network. Nodes were more active before a certain stage. Such activities are not coherent. Therefore, it was harder to dismiss the opinion leaders as the whole process of the incident, it needed to identify potential opinion leaders at a finer level of granularity and

Table 2: The Structural holes of mature stage

Node	Degree	Constrain	Density
Happy stock 886644	111.000	0.039	1.175
The Lord of Peach Blossom Island_6	21.000	0.217	0.543
Dawn is now tomorrow	102.000	0.042	1.351
Love the dog	100.000	0.043	1.404
Beijing United Hospital of China and America	28.000	0.158	0.860
put quality before quantity	40.000	0.111	0.801
Love Wolf	11.000	0.358	0.655
ren398540882	107.000	0.040	1.256

conducted early warning of incident developments and public opinion regulation.

4 The incident text emotional calculation

As the most popular social network platform, micro-blog users express their opinions and feelings through words and expressions. These messages obviously include emotional information. These sections will statistics emotions and intensity in the whole text from three aspects: emoticons, emotion words and sentence patterns. We analyzed the data of "Shandong illegal vaccine" from March 18, 2016 to March 28, 2016.

4.1 Emotional dictionary construction

However, due to the colloquy of micro-blog, a large number of emoticons and related terms for specific events, we constructed an event-specific emotion dictionary. On the basis of comprehensive comparison of existing emotion dictionary, this paper referred to Dalian polytechnic Chinese emotional vocabulary ontology database, which divides the words into two categories: positive emotion ("happy", "good") and negative emotion ("angry", "sad", "evil", "fear", "shock"). Use the word segmentation software to segment the collected text, according to the word frequency from high to low, and then we manually added words that did not come from Dalian Polytechnic dictionary, but also gave emotional classification and emotional strength to these words. The words included not only adjectives, nouns, but also modal word. Part of the expansion of the emotional dictionary is shown.

Table 3: Expansion of emotional dictionary (part)

Words	Emotional classification	Emotional strength	Emotional tendency
Bully	evil	5	negative
Affording general satisfaction	good	9	positive
Bend the law for the benefit of relatives or friends	evil	7	negative
Beyond description	angry	7	negative
Get out	angry	5	negative
Barbarian	evil	5	negative

In order to reduce the influence ofthe collocation of positive emotion words or negative emo-

tion words and negative words, in general, the emotional tendency of positive emotion and negative word tends to be negative emotion, and the emotional tendency of negative emotion and negative word tends to be weak emotion or not Emotions [11]. Therefore, we first determined whether there are negative words adjacent to or similar to the emotional words, in this paper, negative emotions that are negative words modified by negative words, positive emotion words modified by negative words are regarded as negative emotions, negative emotion words modified by negative words are considered as neutral emotions. The negative words include "no", "not", "none", "never", and so on. Emoticons are also commonly used to express emotions in micro-blog. This paper still classified emoticons appearing in texts based on the seven emotions categories in Dalian Polysemous Emotional Dictionary, and the corresponding relationship between emotion classification and expressions is shown in the Table 4.

Table 4: The corresponding table of emotions and emotional categories (part)

Emoticonal text	**Emotional classification**	**Polarity value**	**Emotional tendency**
[Too happy]	happy	7	1
[Haha]	happy	9	1
[Humph]	angry	3	2
[Angry]	angry	7	2
[Retched]	sad	3	2
[Disappointed]	sad	7	2
[Surprised]	shock	7	**2**
[Smiles]	happy	5	1

4.2 Micro-blog emotional polarity calculation

The polarity value of each microblog is determined by the emotion word, the sentence pattern and the emoticons in the microblog text. Specific model algorithm steps are as follows:

(1) For each word in the established emotional dictionary, calculate its TF-IDF value with all the text as the total corpus, and mark it as then the emotional polarity of the ith word in this event is:$PWi = Fi * PWi$;

(2) The emotional polarity value of each sentence is determined by the emotion words, emoticons and sentence pattern The process of calculating the emotional polarity of a sentence is to calculate the sum of the emotional polarity values of all the emotional words and emoticons in the sentence firstly, and then adding the sentence rules to determine the polarity of the entire sentence. Let $P(Si)$ be the emotional polarity of the sentence before considering the sentence pattern, that is: $P(Si) = \sum SE + \sum P(W)$, where $\sum SE$ is the sum of the emotional intensities of all the emoticons in the sentence, and $\sum P(W)$ is the sum of all the emotion words in the sentence.

(3) Let $P'(Si)$ be the sentence emotion value after considering the sentence pattern characteristics, and Table 5 is the sentence emotional polarity weighted rule of six sentence patterns.

(4) The emotional polarity value of a single micro-blog Sij is the sum of the emotional polarity values of each sentence, that is, $P(Sij) = \sum_{k=i}^{j} P'(Sk)$

Table 5: Special sentence calculation instructions

Sentence pattern	$P'(Si)$
Rhetorical questions	$P'(Si) * (-0.6) + (-0.05)$
Interrogative sentence	$P'(Si) * (-0.2) + (-0.05)$
Exclamation sentence	$P'(Si) * (1.5)$
Assumptions sentence	$P'(Si) * (-0.2)$
Declarative sentence	$P(Si)$
Ironically	$-P(Si)$

4.3 Statistical results

Emotional calculation was carried out on the collected micro-blog texts of this event, and the emotion word, the sentence pattern and the emoticons were taken into account in the calculation. Emotional values of the event micro-blog texts were calculated through a calculation rule of a single micro-blog text. Statistics on the intensity of micro-blog texts for a single micro-blog text as a Unit were conducted.

Table 6: Positive and negative polarity text statistics

Negative texts number	Positive texts number
6434(85%)	1113(14.7%)

As can be seen from Table 6, negative micro-blog texts occupied most of share, about 0.3% of the texts did not have emotional values, and 85% of the texts were negative texts. Negative texts were very unfavorable to the development of the event and the subject of the event, so it was especially important to pay attention to the negative micro-blog in the event. This paper drew the negative polarity intensity figure according to the time (Figure 5). Combined with the release of time-critical information of vaccine events, the dynamic process of emotion trend can be initially determined. The highest value of first-pass negative emotional polarity was reported on March 18 after the outbreak of illegal vaccine, culminating on the 19^{th}. The second round of the highest negative emotions appeared at 23rd-24th, which was due to the spread of emotions caused by the widespread spread of "The sorrow of vaccine".

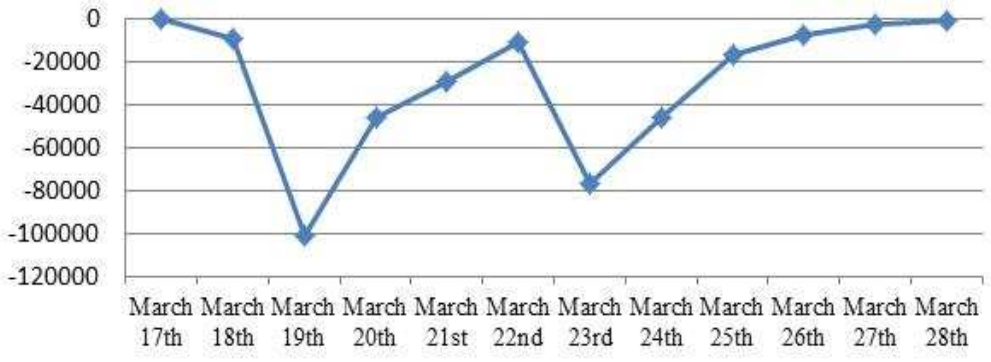


Figure 5: Emotional polarity of Shandong illegal vaccine

The public had a strong emotional response as the incident had many risk characteristics such as affecting children and influencing the population. State Food and Drug Administration, the State Council, the National Health and Family Planning Commission quickly intervened,

and issued and responded several times, and communicated with the public. From the above-mentioned incident, the public opinion and emotion tendency of internet public opinion showed that the public and the media paid close attention to the incident and responded to the progress of the incident handling. The key nodes of incident handling coincided with the peaks of the network public opinion. It was suggested that government departments plaid an important role in the handling of incidents and the release of key information. Timely release of authoritative and transparent objective information was conducive to dredge network public opinion, and it can reduce the negative influence brought by internet public opinion. In this incident, the government sector will definitely not make a clear statement of its impatience and will effectively cut the responsibility so that public emotions will be released and it will be easy to establish and restore trust.

5 Conclusion

Based on the theory of network public opinion evolution, this paper took the example analysis of microblog public opinion events, and used the ways of the data crawler, segmentation technology, the construction of emotional dictionary and annotation patterns to get the emotional tendency and emotional polarity of events Then, we explored the causes of the peaks and emotional evolutions that had occurred throughout the evolutionary life cycle of public opinion. The analysis of vaccine events showed that risk perception of the public and society in the media environment tended to deviate easily, and it was difficult to form a self-correcting mechanism. To standardize the processing of media information and to monitor and guide public sentiment is the key to risk communication in crisis And it is a need to speed up the formulation and construction of a system of rules governing the expression of public opinion. The investigation of government and relevant departments' intervened investigation and its result has played an important role in the development of public opinion. The timely announcement of the government and the factual release of the truth in real time were conducive to resolving the contradictions and quelling the public opinion. In the information era, the government and relevant departments should not expect to reduce the spread by covering up with negative news And openly and fairly dealing with them according to law, which is the basic requirement of a society ruled by law.

Although this paper is a single case study of recent typical risk events, the illegal vaccine event is a microcosm of the outbreak of public opinion in recent years. Lack of rigorous media information and the extensive emotional evolution in the network have a direct impact on risk perception. The changes of network structure and emotional polarity exhibited during the amplification of the risk of illegal vaccines have empirical value in coping with other similar social focus events.

Acknowledgment

Work described in this paper was funded by the National Natural Science Foundation of China under Grant No. 71671093 and the National Natural Social Foundation of China under Grant No. 16ZDA054. The authors would like to thank other researchers at Nanjing University of Posts and Telecommunications.

Bibliography

[1] Chen Y.J. (2010); An Effective Way to Explore the Internet, *Computer and Information Science (ICIS)*, 529-534, 2010.

[2] Cho, Y.; Hwang, J.; Lee, D. (2012); Identification of effective opinion leaders in the diffusion of technological innovation: a social network approach, *Technological Forecasting and SocialChange*, 79(1), 97-106, 2012.

[3] Han, S.C.; Liu, Y.; Chen, H.L.; Zhang, Z.J. (2016); Influence Model of User Behavior Characteristics on Information Dissemination, *International Journal of Computers Communications & Control*, 11(2), 209–223, 2016.

[4] Hong, X.J.; Jiang, N.; Xia, J.J. (2014); Study on Micro-blog Rumor Based on Social Network Analysis - A Case Study on the Micro-blog about Food Security, *Intelligence Journal*, 8, 161-167, 2014.

[5] Lee, C. H. (2012); Mining Spatio-temporal Information on Micro-blogging Streams Using a Density-based Online Clustering Method, *Expert Systems with Applications*, 39(10), 9623-9641, 2012.

[6] Nair, H.S.; Manchanda, P.; Bhatia, T. (2010); Asymmetric Social Interactions in Physician Prescription Behavior: The Role of Opinion Leader, *Journal of Marketing Research*, 47, 883-895, 2010.

[7] Paltoglou, G. (2016), Sentiment-based Event Detection in Twitter, *Journal of the Association for Information, Science and Technology*, 67(7), 1576-1587, 2016.

[8] Riloff, E.; Patwardhan, S.; Wiebe, J. (2006); Feature Subsumption for Opinion Analysis. *Conference on Empirical Methods in Natural Language Processing*, 440-448, 2006.

[9] Su, X.-P.; Song, Y.-R. (2015); Leveraging neighborhood "structural holes" toidentifying key spreaders in social networks, *Acta Phys. Sin*, 2, 5-15, 2015.

[10] Sznajd, W.K. (2006); Opinion Evolution in Closed Community, *Physics C*, 11, 1157-1165, 2000.

[11] Wang, L.; Zhao, Y.; Liang, S.H. (2013), Micro-blog Social NetworkAnalysis Based on Network Group Behavior, *Advanced Materials Research*, 79(12), 435-438, 2013.

[12] Wen, F.; He, Z.; Dai, Z.; Yang, X. (2014); Characteristics of Investors' Risk Preference for Stock Markets, *Economic Computation and Economic Cybernetics Studies and Research*, 48(3), 235-254, 2014.

[13] Wen, F.; Gong, X.; Cai, S. (2016); Forecasting the volatility of crude oil futures using HAR-type models with structural breaks, *Energy Economics*, 59, 400-413, 2016.

[14] Wen, F.; Xiao, J.; Huang, C.; Xia, X. (2018); Interaction between oil and US dollar exchange rate: nonlinear causality, time-varying influence and structural breaks in volatility. *Applied Economics*, 50(3), 319-334, 2018.

[15] Wiebe, J.; Riloff, E. (2012); Finding Mutual Benefit between Subjectivity Analysis and Information Extraction, *IEEE Transactions on Affective Computing*, 2(4), 175-191, 2012.

[16] Xiong, F.,; Liu, Y.; Zhang, Z.J. (2012); An information diffusion model based on retweeting mechanism for online social media, *Physics Letters A*, 6, 2103-2108, 2012.

[17] Xu, L.; Lin, H.; Zhao, J. (2008); Construction and Analysis of Emotional Corpus, *Journal of Chinese Information Processing*, 1, 116-122, 2008.

[18] Zhang, D. (2017): High-speed Train Control System Big Data Analysis Based on Fuzzy RDF Model and Uncertain Reasoning, *International Journal of Computers Communications & Control*, 12(4), 577-591, 2017.

[19] Zhang, H.; Small, M.; Fu, X. (2011); Staged Progression Model for Epidemic Spread on Homogeneous and Heterogeneous Networks, *Journal of Systems Science and Complexity*, 24, 619-630, 2011.

[20] Zhang, D.; Sui, J.; Gong, Y. (2017); Large scale software test data generation based on collective constraint and weighted combination method, *Tehniki vjesnik*, 24(4), 1041-1049, 2017.

[21] Zhao, J.L.; Cheng, J.H. (2015); Analysis of Micro-blog Public Opinion Diffusion Based on SNA: AN Empirical Study on April 20 Ya'an Earthquake in Sichuan, *Management Review*, 1, 148-157, 2015.

[22] Zhao, H.Q. (2015); Analysis of Complex Network Public Opinion Communication, *Journal of Qinghai University (Natural Science)*, 4, 29-37, 2015.

[23] Zhou, M.; Liu, X.; Pan, B. (2017); Effect of Tourism Building Investments on Tourist Revenues in China: A Spatial Panel Econometric Analysis, *Emerging Markets Finance and Trade*, 53(9), 1973-1987, 2017.

Author index