INTERNATIONAL JOURNAL

of

COMPUTERS, COMMUNICATIONS & CONTROL

ISSN 1841-9836

ISSN-L 1841-9836

A Bimonthly Journal
With Emphasis on the Integration of Three Technologies

Year: 2014 Volume: 9 Issue: 3 (June)

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

Agora University Editing House

CCC Publications

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

International Journal of Computers, Communications & Control

EDITORIAL ADDRESS:
Agora University/ R&D Agora Ltd. / S.C. Cercetare Dezvoltare Agora S.R.L.
Piaţa Tineretului 8, Oradea, jud. Bihor, Romania, Zip Code 410526
Tel./ Fax: +40 359101032
E-mail: ijccc@univagora.ro, rd.agora@univagora.ro, ccc.journal@gmail.com
Journal website: http://univagora.ro/jour/index.php/ijccc/

International Journal of Computers, Communications & Control

DATA FOR SUBSCRIBERS

Supplier: Cercetare Dezvoltare Agora Srl (Research & Development Agora Ltd.)
Fiscal code: 24747462
Headquarter: Oradea, Piata Tineretului Nr.8, Bihor, Romania, Zip code 410526
Bank: MILLENNIUM BANK, Bank address: Piata Unirii, str. Primariei, 2, Oradea, Romania
IBAN Account for EURO: RO73MILB0000000000932235
SWIFT CODE (eq.BIC): MILBROBU

International Journal of Computers, Communications & Control

Brief Description of IJCCC

Title of journal: International Journal of Computers, Communications & Control
Acronym: IJCCC
Abbreviated Journal Title: INT J COMPUT COMMUN
International Standard Serial Number: ISSN 1841-9836, ISSN-L 1841-9836
Publisher: CCC Publications - Agora University
Starting year of IJCCC: 2006
Founders of IJCCC: Ioan Dzitac, Florin Gheorghe Filip and Mişu-Jan Manolescu
Logo:

Publication frequency: Bimonthly: Issue 1 (February); Issue 2 (April); Issue 3 (June); Issue 4 (August); Issue 5 (October); Issue 6 (December).

Coverage:

- Beginning with Vol. 1 (2006), Supplementary issue: S, IJCCC is covered by Thomson Reuters - SCI Expanded and is indexed in ISI Web of Science.
- Journal Citation Reports(JCR)/Science Edition:
  - Impact factor (IF): JCR2009, IF=0.373; JCR2010, IF=0.650; JCR2011, IF=0.438; JCR2012, IF=0.441.
- Beginning with Vol. 2 (2007), No.1, IJCCC is covered in EBSCO.
- Beginning with Vol. 3 (2008), No.1, IJCCC, is covered in Scopus.

Scope: International Journal of Computers Communications & Control is directed to the international communities of scientific researchers in computer and control from the universities, research units and industry.

To differentiate from other similar journals, the editorial policy of IJCCC encourages the submission of scientific papers that focus on the integration of the 3 "C" (Computing, Communication, Control).

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

- Integrated solutions in computer-based control and communications;
- Computational intelligence methods (with particular emphasis on fuzzy logic-based methods, ANN, evolutionary computing, collective/swarm intelligence);
- Advanced decision support systems (with particular emphasis on the usage of combined solvers and/or web technologies).

5th International Conference on Computers, Communications & Control, ICCCC 2014

Hotel President, Baile Felix, Oradea, Romania, May 6-10, 2014

Brief Description of ICCCC 2014

International Conference on Computers Communications and Control (ICCCC) and International Journal of Computers, Communications & Control (IJCCC) was founded by Ioan Dzitac, Florin Gheorghe Filip and Misu-Jan Manolescu in 2006. ICCCC is organized in every even year by Agora University of Oradea under the aegis of Romanian Academy - Information Science and Technology Chapter, in order to exchange ideas, problems, solutions, and to work together in a friendly environment.

Honorary Chair of ICCCC 2014: Prof. Yong Shi, Executive Deputy Director of Research Center on Fictitious Economy & Data Science - Chinese Academy of Sciences & A. Dean of School of Management, University of Chinese Academy of Science.

Program Committee Chair: Prof. Florin Gheorghe Filip, President of Information Science and Technology Chapter -Romanian Academy.

Organizing Committee Chair: Prof. Misu-Jan Manolescu, President of Agora University.

ICCCC 2014, 5th edition of ICCCC, is focused to four special sessions:

Special Session 01: *Decision Support Systems for Supply Chain Management and their Dynamics. New Proposals and Directions*, Chair: Angel Ortiz, Universidad Politecnica de Valencia, Spain.

Special Session 02: *Network Optimization and Security*, Chair: Prof. Yezyd Donoso, University de los Andes, Colombia.

Special Session 03: *Data Mining and Intelligent Knowledge Management*, Chairs: Prof. Gang Kou & Prof. Yi Peng, Southwestern University of Finance and Economics / University of Electronic Science and Technology of China, Chengdu, China.

Special Session 04: *Computational Intelligence Methods in Decision Support Systems*, Chairs: Prof. Dan Benta & Prof. Adriana Manolescu, Agora University of Oradea, Romania.

With every edition of the conference the evaluation process is more demanding. Due to the certain number of submissions that can be published in IJCCC we could accept just a small number of submitted papers. Therefore many good works had to be rejected. The abstracts of all accepted papers for ICCCC2014 will be published in the volume *Abstracts of ICCCC Papers*, 4(2014), ISSN 1844-4334. All presented papers at ICCCC 2014, after strict peer-reviewing process, will be included in the *Proceedings of the International Conference on Computers, Communications and Control*, which will be published as a supplementary issue of IJCCC (4-6 pages/paper).

After presentation and discussions in conference sessions, extended versions of some selected high-quality papers will be published in IJCCC. The title of extended paper must be different of proceedings paper and the old title must be cited. Is mandatory also provide a brief description of the differences between the extended manuscript and the preliminary version published in conference proceedings. Paper previously published in conference proceedings are eligible for consideration provided that the papers have undergone substantial revision.

Prof. Ioan Dzitac,
Rector of Agora University,
General Chair of ICCCC2014 & A. Editor in Chief of IJCCC,
E-mail: rector@univagora.ro, Tel. +40 359 101 032.

International Journal of Computers, Communications & Control
ISSN 1841-9836, ISSN-L 1841-9836, Volume 9, Issue 3, June, 2014.

Contents

INT J COMPUT COMMUN, ISSN 1841-9836
9(3):253-260, June, 2014.

Solving Continuous Trajectory and Forward Kinematics Simultaneously Based on ANN

C.K. Ang, S.H. Tang, S. Mashohor, M.K.A.M. Arrifin

Chun Kit Ang*, Sai Hong Tang,
Syamsiah Mashohor, Mohd Khairol Anuar Mohd Arrifin
Universiti Putra Malaysia
Department of Mechanical and Manufacturing Engineering,
Faculty of Engineering, 43400 UPM, Serdang, Selangor Darul Ehsan.
Email: ack_kit@hotmail.com, saihong@upm.edu.my,
syamsiah@upm.edu.my, khairol@upm.edu.my
*Corresponding author

Abstract: Robot movement can be predicted by incorporating Forward Kinematics (FK) and trajectory planning techniques. However, the calculations will become complicated and hard to be solved if the number of specific via points is increased. Thus, back-propagation artificial neural network is proposed in this paper to overcome this drawback due to its ability in learning pattern solutions. A virtual 4-degree of freedom manipulator is exploited as an example and the theoretical results are compared with the proposed method.
Keywords: Artificial Neural Network (ANN), back propagation neural network, forward kinematics, continuous trajectory.

1 Introduction

Trajectory planning can be considered as an important issue in robot industry. Without an appropriate trajectory planning, a robot's motion will become unpredictable and it may collide with obstacles or go through undesirable points. Thus, an appropriate trajectory planning enables us to determine the required workspace and provides us an opportunity to avoid obstacles.

Basically, there are two types of trajectory planning, point to point and continuous trajectory (with via points). However, continuous trajectory planning will be more practical compare to point to point trajectory because it allows robot to pass through various specific points and it is useful for real time robot.

Conventionally, there are a few methods that can be used for continuous trajectory planning such as Bezier function, Linear Segments with Parabolic Blends (LSPB), and high order polynomials. Unfortunately, those methods require extensive calculations when the number of specific points is increased and thus leading to the emergence of a large number of formulas ([1]- [5]). Hence, it is advisable to develop a method which is able to overcome this drawback.

2 Preliminaries

Forward kinematics (FK) can be used to determined the position and orientation of a robot's hand if all the configurations are known and the equations can be derived by using Denavit-Hartenberg (DH) method because it is the standard way of modelling the robot motion due to its simplicity of modelling robot links and joints that is applicable for any robot configuration s regardless of its complexity. The total transformation for a robot manipulator from base to hand will be

$$ {}^{R}T_{H} = {}^{R}T_{1}^{1}T_{2}{}^{2}T_{3}...{}^{n-1}T_{n} = A_{1}A_{2}A_{3}...A_{n} = \begin{pmatrix} n_x & o_x & a_x & p_x \\ n_y & o_y & a_y & p_y \\ n_z & o_z & a_z & p_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (1) $$

and the end effector can be defined as:

$$ r_e(t) = f_e(\theta(t)) \quad (2) $$

where $r_e \in R^m$ in Cartesian space is related to the joint space vector, m is dimensional position vector of end effector.

In fact, what we are actually interested in is the inverse kinematics (IK) because it allows us to determine the value of each joint in order to place the arm at a desired position and orientation. The inverse kinematic problem is to find the joint variable θ for any given$r_e(t)$. Unfortunately, it is impossible to find an analytic solution due to the nonlinearity of$f(\theta(t))$in reality. The inverse kinematics problem of manipulators is thus usually solved at the velocity level. Differentiating $r_e(t) = f_e(\theta(t))$ with respect to time yields a linear relation between the Cartesian velocity $\dot{r}_e$and the joint velocity$\dot{\theta}$:

$$ J_e(\theta)\dot{\theta} = \dot{r}_e \quad (3) $$

where $J_e(\theta) \in R^{m \times n}$ is the end effector Jacobian matrix which is defined as $J_e(\theta) = \frac{\partial f(\theta)}{\partial \theta}$ [6-13]. However, the derived IK equations are not necessarily available when the degeneracy and singularity problems ([6], [9], [12]- [14])] occurred and thus we need to use some new methods to solve it such as artificial neural network, genetic algorithm, neural fuzzy and etceteras ([7]- [15]).

Generally, high order polynomial is used to plan a trajectory with the intention to let the robot's hand to pass through all the specific points. However, solving the high order polynomial equations requires extensive calculations. Instead, it is advisable to use combinations of lower order polynomials for different segments of the trajectory and blend them together to satisfy all required boundary conditions ([1]- [5]).

$$ \theta\left(t\right) = c_0 + c_1 t + c_2 t^2 + ... + c_{n-1}t^{n-1} + c_n t^n \quad (4) $$

For an example, for a 4-3-4 trajectory, a fourth order polynomial is used to plan a trajectory between the initial point and the first via point, followed by a third order polynomial to plan a trajectory between two via points, lastly another fourth order polynomial is implemented to plan the trajectory between the last via point and final angle.

$$ \theta\left(t\right)_1 = a_0 + a_1 t + a_2 t^2 + a_3 t^3 + a_4 t^4 $$

$$ \theta\left(t\right)_2 = b_0 + b_1 t + b_2 t^2 + b_3 t^3 $$

$$ \theta\left(t\right)_3 = c_0 + c_1 t + c_2 t^2 + c_3 t^3 + c_4 t^4 $$

In short, IK allows us to allocate the robot's hand at desired location and the motion of robot's hand can be determined by incorporating the FK and trajectory techniques. Sadly, the calculations will become tedious and hard to be solved when the number of specific points is increased.

3 Theoretical and Simulation Results

Experiments were performed on a simulated 4-degree of freedom (DOF) robot manipulator which shown in Figue 1, IK was used to determine the starting and ending configurations for the robot's hand while the high order polynomials and FK were used to track the motion of the robot's hand from beginning to the end in Cartesian space. It was specified that the robot need to pass through two specific points. At the same time, velocities and accelerations constraints need to be taken into consideration.

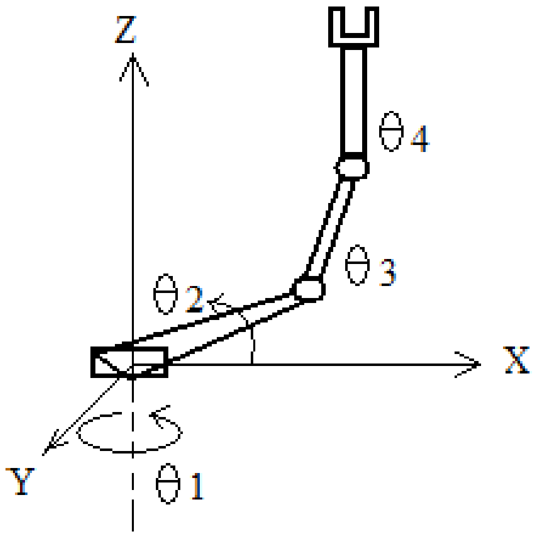


Figure 1: Simulated 4 DOF manipulator

Assuming that a manipulator needs to spend $t_f$to complete the moving path and a manipulator joint needs to reach $\theta_a$at interval time t_a then $\theta_b$at interval time t_b where $t_a < t_b$and $t_a, t_b \in [0, t_f]$. One may need to match the position, velocity, and accelerations of the two segments at each point to plan a continuous trajectory. Hence, higher order polynomials in the form of $\theta\left(t\right) = C_0 + C_1 t + C_2 t^2 + ... + C_{n-1} t^{n-1} + C_n t^n$ are required. However, it is an extensive calculation process to solve high order polynomial and it is advisable to use the combinations of lower order polynomials for different segment of the trajectory such as 4-3-4 trajectory or 3-5-3 trajectory.

The velocity and acceleration for initial and final point are zero. In order to fulfill the entire initial, final, and via points requirement, a matrix form can be generated where

$t_{ab} = t_b - t_a$and $t_c = t_f - t_a - t_b$.

$$\begin{bmatrix} \theta_1 \\ \dot{\theta}_1 \\ \ddot{\theta}_1 \\ \theta_2 \\ \theta_2 \\ 0 \\ 0 \\ \theta_3 \\ \theta_3 \\ 0 \\ 0 \\ \theta_4 \\ \dot{\theta}_4 \\ \ddot{\theta}_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & t_a & t_a^2 & t_a^3 & t_a^4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2t_a & 3t_a^2 & 4t_a^3 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 6t_a & 12t_a^2 & 0 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & t_{ab} & t_{ab}^2 & t_{ab}^3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2t_{ab} & 3t_{ab}^2 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 6t_{ab} & 0 & 0 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & t_c & t_c^2 & t_c^3 & t_c^4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2t_c & 3t_c^2 & 4t_c^3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 6t_c & 12t_c^2 \end{bmatrix} \times \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ b_0 \\ b_1 \\ b_2 \\ b_3 \\ c_0 \\ c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix}$$

Or

$$[\theta] = [M]\,[C] \tag{5}$$

By solving those constant values in matrix$[C]$, the joint turning angle at any interval time can be obtained where

$$\theta(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3 + a_4 t^4 \qquad 0 < t \le t_a$$

$$\theta(t) = b_0 + b_1 t + b_2 t^2 + b_3 t^3 \qquad t \in [t_a, t_b]\ , t = t - t_a$$

$$\theta(t) = c_0 + c_1 t + c_2 t^2 + c_3 t^3 + c_4 t^4 \qquad t \in [t_b, t_f]\ , t = t - t_a - t_b$$

The equation which is used to represent the robot's hand at any interval time can be denoted as:

$$r_E(t_i) = \begin{pmatrix} \cos(\theta_1(t_i)) \left[l_1 \times \cos(\theta_2(t_i)) + l_2 \times \cos\left(\sum_{k=2}^{3} \theta_k(t_i)\right) + l_3 \times \cos\left(\sum_{k=2}^{4} \theta_k(t_i)\right)\right] \\ \sin(\theta_1(t_i)) \left[l_1 \times \cos(\theta_2(t_i)) + l_2 \times \cos\left(\sum_{k=2}^{3} \theta_k(t_i)\right) + l_3 \times \cos\left(\sum_{k=2}^{4} \theta_k(t_i)\right)\right] \\ l_1 \times \sin(\theta_2(t_i)) + l_2 \times \sin\left(\sum_{k=2}^{3} \theta_k(t_i)\right) + l_3 \times \sin\left(\sum_{k=2}^{4} \theta_k(t_i)\right) \end{pmatrix} \tag{6}$$

In fact, all these calculations can be solved simultaneously by back-propagation artificial neural network. The back-propagation neural network can be activated by presented a training set data with inputs $x_1(p), x_2(p)...x_n(p)$and desired outputs$y_{d,1}(p), y_{d,2}(p), ..., y_{d,n}(p)$. Assume that the back-propagation network consists of three layers which are input layer, hidden layer, and output layer. The actual outputs of the neurons in the hidden layer will be $y_j(p) = sigmoid\left[\sum_{i=i}^{n} x_i(p) \times w_{ij}(p) - \theta_j\right]$ where n is the number of inputs for neuron j in the hidden layer, and sigmoid is the sigmoid activation function$Y^{sigmoid} = 1/(1+e^{-X})$. For the output layers, the actual outputs of the neurons will be $y_k(p) = sigmoid\left[\sum_{j=1}^{m} x_{jk}(p) \times w_{jk}(p) - \theta_k\right]$where m is the number of inputs for neuron k in the output layer. An error signal $e_k(p) = y_{d,k}(p) - y_k(p)$will be generated from neuron $k$at output layer and propagated backward for updating the weights.

For updating the weights at the output neurons with learning rate, α:

Error gradient,$\delta_k(p) = y_k(p) \times [1 - y_k(p)] \times e_k(p)$

$e_k(p) = y_{d,k}(p) - y_k(p)$

Weight correction, $\Delta w_{jk}(p) = \alpha \times y_j(p) \times \delta_k(p)$

Update weight, $w_{jk}(p+1) = w_{jk}(p) + \Delta w_{jk}(p)$

Similar to output layer, for updating the weights at the hidden neurons with learning rate, α:

Error gradient,$\delta_j(p) = y_j(p) \times [1 - y_j(p)] \times \sum_{k=1}^{1} \delta_k(p) \times w_{jk}(p)$

Weight correction, $\Delta w_{ij}(p) = \alpha \times x_i(p) \times \delta_j(p)$

Update weight,$w_{ij}(p+1) = w_{ij}(p) + \Delta w_{ij}(p)$

The same procedures can be repeated by increasing the iteration $p$until the output error criterion is satisfied.

A training data set which consists of inputs and outputs is presented to the networks for learning purpose. Inputs consists of the final angle of each joint $\theta_{f,n}$ and interval time variable t_i where $t_i \in [0, t_f]$ and $t_f$is the time for the manipulator to complete the moving path. The outputs will be the coordinates of robot's hand at any interval time with respect to t_i,$r_E(t_i) = [p_x\ p_y\ p_z]^T$.

There are two set of results in table 1. The theoretical results which obtained by using conventional FK and high order polynomial trajectory were compared to the artificial neural network (ANN) results. The robot started moving from its initial position ($\theta 1$=$\theta 2$=$\theta 3$=$\theta 4$=0) and passed through two via points before it reached its destinations. It shows that the robot still able to pass through those specific points even though the destination point was changed and the moving path could be tracked by using ANN. The theoretical and ANN results were almost the same and the errors were less than ±0.3unit (less than 1%). This can be used as an evidence to proof that artificial neural network is able to solve FK and continuous trajectory simultaneously due to its adaptive learning ability.

Since back-propagation neural network is able to track the motion of end effector, it can be used to track the position of each joint in Cartesian space as well. Referring to Figure 1, Joint 1 and joint 2 are located at the origin of reference frame, (0,0,0). For tracking the position of joint 3, the training data set inputs and outputs will be $[t_i\ \ \theta_1(t_i)\ \ \theta_2(t_i)]^T$and $r_{E,3}(t_i)$. Correspondingly, for joint 4, the training data set inputs and outputs will be $[t_i\ \ \theta_1(t_i)\ \ \theta_2(t_i)\ \ \theta_3(t_i)]^T$and $r_{E,4}(t_i)$. Once the joins location of all joints are computed, the manipulator link equations can be derived by using parametric equations.

Assuming that the location and vector for joint m and joint n at any interval time t_i: $r_{E,n}(t_i) = [p_{x,n}\ p_{y,n}\ p_{z,n}]^T$, $\overrightarrow{OP_n} = [a_n\ b_n\ c_n]^T$ and $r_{E,m}(t_i) = [p_{x,m}\ p_{y,m}\ p_{z,m}]^T$, $\overrightarrow{OP_m} = [a_m\ b_m\ c_m]^T$.

The link equation that connecting joint n and m will be

$$\frac{x-p_{x,n}}{a_n-a_m} = \frac{y-p_{y,n}}{b_n-b_m} = \frac{z-p_{z,n}}{c_n-c_m} = \zeta$$
$$or$$
$$[x\ y\ z]^T = [p_{x,n}\ p_{y,n}\ p_{z,n}]^T + [a_n - a_m\ b_n - b_m\ c_n - c_m]^T\,\zeta$$

For the robot manipulator which is mentioned in this paper, the robot link equations can be denoted as:

Link 1: $\frac{x-p_{x,3}}{a_3-a_2} = \frac{y-p_{y,3}}{b_3-b_2} = \frac{z-p_{z,3}}{c_3-c_2} = \zeta_2 \qquad 0 \le z \le p_{z,3}$

Link 2: $\frac{x-p_{x,4}}{a_4-a_3} = \frac{y-p_{y,4}}{b_4-b_3} = \frac{z-p_{z,4}}{c_4-c_3} = \zeta_3 \quad p_{z,3} \le z \le p_{z,4}$

Link 3: $\frac{x-p_{x,5}}{a_5-a_4} = \frac{y-p_{y,5}}{b_5-b_4} = \frac{z-p_{z,5}}{c_5-c_4} = \zeta_4 \quad p_{z,4} \le z \le p_{z,5}$

Conclusively, the link equation can be derived as long as the coordinates of two successive joints are provided by using back-propagation neural network and the links parametric equations can be derived based on the vector of the two successive joints. Once the equation of each manipulator link is known, a person is able to predict or compute the entire motion of a robot manipulator.

Besides, obstacle collision can be detected if the link equations are known. Assuming that there is a rigid body obstacle enclosed in a sphere which possesses a r radius and center locates at (x_o, y_o, z_o), the sphere equation will be

$$(x - x_o)^2 + (y - y_o)^2 + (z - z_o)^2 = r^2$$

The intersection points between the link and obstacle can be obtained by solving the link and sphere surface equation. For an example, a line that pass through the point (e, f, g)possessing a same direction as vector$[a\ b\ c]^T$, the equation for the link will be:

$$\frac{x-e}{a} = \frac{y-f}{b} = \frac{z-g}{c} \quad = \quad \zeta \text{or} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \\ \mathbf{z} \end{pmatrix} = \begin{pmatrix} \mathbf{e} \\ \mathbf{f} \\ \mathbf{g} \end{pmatrix} + \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \\ \mathbf{c} \end{pmatrix} \zeta$$

Substituting the link equation into sphere surface equation yields,

$$(a\zeta + e - x_o)^2 + (b\zeta + f - y_o)^2 + (c\zeta + g - z_o)^2 = r^2$$

simplified and yields,

$$A\zeta^2 + 2B\zeta + C = 0$$

where,

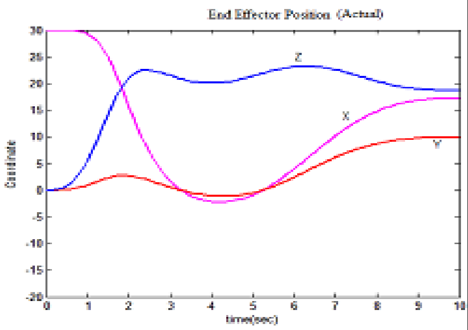
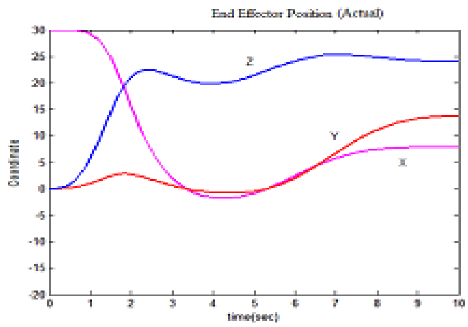
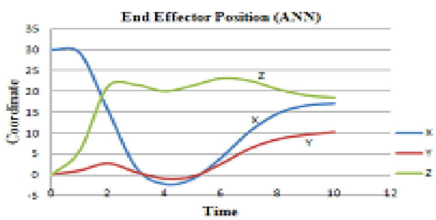
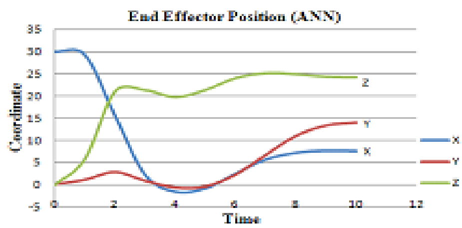
$$A = a^2 + b^2 + c^2$$
$$B = a(e - x_o) + b(f - y_o) + c(g - z_o)$$
$$C = (e - x_o)^2 + (f - y_o)^2 + (g - z_o)^2 - r^2$$

By examining the discrimination roots, the intersection points between link and sphere surface exists when$\Delta \geq 0$,

$$[a(e-x_o)+b(f-y_o)+c(g-z_o)]^2 \geq 4\left[a^2+b^2+c^2\right]\left[(e-x_o)^2+(f-y_o)^2+(g-z_o)^2-r^2\right]$$

The intersection point between the manipulator links and obstacle can be determined by solving the constant ζ.

Table 1: Comparison of Conventional and ANN Techniques in Tracking Robot End Effector Moving Path

	Destination 1 Coordinate: (x=17.24; y=9.95; z=18.68) Joint values: (θ1=30; θ2=15; θ3=25; θ4=35)					Destination 2 Coordinate: (x=7.89; y=16.67; z=24.06) Joint values: (θ1=60; θ2=35; θ3=25; θ4=15)				
	Time	θ1	θ2	θ3	θ4	Time	θ1	θ2	θ3	θ4
Specific point 1 (x=15.58; y=2.75; z=21.08)	2	10	20	30	40	2	10	20	30	40
Specific point 2 (x=-0.87; y=-0.5; z=21.43)	5	30	40	50	60	5	30	40	50	60
Actual										
ANN										

4 Conclusions

This paper shows that back-propagation neural network is able to solve trajectory and kinematic problems simultaneously. The moving path of the robot end effector still able to be tracked even though the ending point is changed. Besides, this method can be used to track the position of robot joints which can be used to derive the robot link equations. In comparison to conventional continuous trajectory planning methods, ANN is much easier to be applied and provide high accuracy results.

Bibliography

[1] Saeed B. N. (2001). *Introduction to Robotics Analysis, System, Application.* Prentice Hall, 29-172.

[2] Ata, A., Myo, T. (2008). Optimal trajectory planning and obstacle avoidance for flexible manipulators using generalized pattern research. *World J. Modeling and Simulation*, 4:163–171.

[3] Guan, Y., Yokoi, K. , Stasse, O. , Kheddar, A. (2005). A. On robotic trajectory planning using polynomial interpolations. *Proc. IEEE Int. Conf. on Robotics and Biomimetics*, 111-116.

[4] Reichenbach, T, Kovacic, Z.(2005). Collision-Free Path Planning in Robot Cells Using Virtual 3D Collision Sensors. *Cutting Edge Robotics*, 683–697.

[5] Campos, J., Flores, J.A.R., Montufar, C.P.(2008). Robot Trajectory Planning for Multiple 3D Moving Objects Interception: A Polynomial Interpolation Approach. *Proc. IEEE Int. Conf. on Electronics, Robotics and Automotive Mechanics*, 478–483.

[6] Zarkandi S., Vafadar A., Esmaili M. R.(2011). PRRRRRP redundant planar parallel manipulator: Kinematics, workspace and singularity analysis. *IEEE 5th Int. Conf. on Robotics, Automation and Mechatronics (RAM)*, DOI: 10.1109/RAMECH.2011.6070457, 61–66.

[7] Al-Mashhadany, Y. I. (2010). Inverse Kinematics Problem (IKP) of 6-DOF Manipulator by LRNNs. *Int. Conf. on Management and Service Science (MASS)*, 1 – 5.

[8] Zhang, D., Lei, J.(2011). Kinematic analysis of a novel 3-DOF actuation redundant parallel manipulator using artificial intelligence approach. *Robotics and Computer-Intergrated Manufacturing*, 27: 157–163.

[9] Firmani, F., Podhoprodeski, R. P.(2009). Singularity analysis of planar parallel manipulatos based on forward kinematic solutions. *Mechanism and Machine Theory*, 44: 1386–1399.

[10] Vesselenyi, T., Dzitac, S., Dzitac, I., Manolescu, M.J. (2007). Fuzzy and Neural Controllers for a Pneumatic Actuator. *Int J Comput Commun*, ISSN 1841-9836, 2(4):375-387.

[11] Alvandar, S., Nigam, M. J. (2008). Neuro-Fuzzy based approach for inverse kinematics solution of industrial robot manipulators. *Int J Comput Commun*, ISSN 1841-9836, 3(3):224–234.

[12] Mahidzal, D., and Jian-Ding, T. (2011). Forward and Inverse Kinematics Model for Robotic Welding Process Using KR-16KS KUKA Robot. *Proc. IEEE Int. Conf. on Modeling, Simulation and Applied Optimization*, 1–6.

[13] Shah, J., Rattan S. S., Nakra, B.C.(2011). Kinematic Analysis of 2-DOF planer robot using artificial neural network. *World Academy of Science, Engineering and Technology*, 81: 282–285.

[14] Parikh, J.P., Lam, S.S. (2009). Solving the forward kinematics problem in parallel manipulators using an iterative artificial neural network strategy. *Int. J. Adv. Manuf. Technol.*, 40: 595–606.

[15] Her, M.-G., Chen, C.Y., Karkoub, M. (2002). Approximating a Robot Inverse Kinematics Solution Using Fuzzy Logic Tuned by Genetic Algorithms. *Int. J. Adv. Manuf. Technol.*, 20: 375–380.

INT J COMPUT COMMUN, ISSN 1841-9836
9(3):261-275, June, 2014.

Advantages of Using an Ontological Model of the State Development Funds

S. Arsovski, B. Markoski, P. Pecev, N. Petrovački, D. Lacmanović

Saša Arsovski
Guarantee Fund of Autonomous Province of Vojvodina
Serbia, 21000 Novi Sad, Hajduk Veljkova 11
E-mail: garfondapv@neobee.net

Branko Markoski*,
Predrag Pecev, Dejan Lacmanović
University of Novi Sad
Technical faculty "Mihajlo Pupin" Zrenjanin
Serbia, 23000 Zrenjanin, Djure Dakovica, bbi
E-mail: redrag.pecev@tfzr.rs, dlacman@tfzr.uns.ac.rs
*Corresponding author: markoni@uns.ac.rs

Neboša Petrovački
University of Novi Sad
Faculty of Technical Sciences
Computing and Control Department
Serbia, 21000 Novi Sad,Trg Dositeja Obradovića 6
E-mail: petrovackin@uns.ac.rs

Abstract: Ontologies generated from the workflow of administrative procedures, can provide significant improvements and reduce the time of modeling, testing and integration in the process of building an information system of the public administration. In this paper we has analyzed the ontology annotation aimed to provide the groundwork for automatic generation administrative act. The proposed semantic representation of the administrative procedures enables use of the document templates, which are the framework for the automatic generation of an administrative act. This would be a result of the administrative procedures execution. The proposed approach is verified by a case study of building an ontological model of the administrative procedures for the Guarantee Fund of Autonomous Province of Vojvodina (APV).
Keywords: e-Government, ontology, services, information system.

1 Introduction

State development funds are an interventionist mechanism by which the state affects the development of small and medium enterprises. These are institutions set up by authorities of different levels (state, provincial, local government) and whose activity is aimed at stimulating the development of small and medium-sized enterprises while reducing risk and transaction costs related to the implementation of stimulating instruments (e.g. loans) of Small and Medium Enterprises. The following development funds are established in AP Vojvodina: Vojvodina Development Fund, Agricultural Development Fund, Capital Investment Fund and the Guarantee Fund of APV. There are two scenarios of use of ontologies in the development of information systems, depending on the type of information system that is being modeled [1]: Traditional information systems: semantic content described in the ontology is transformed into a standard component of the information system Ontology-driven information systems: the ontology is an individual component of the information system.

The ontological concept, which defines the administrative procedures and tasks, will be described and analyzed in detail. The course of generating ontology will be resting on the role of defined entity domains and the activities performed by those entities which are observed during the process of the state administration.

The model should meet the following conditions:

1. To describe structural aspect of public administration, administrative units (Funds) and their hierarchical relationship.

2. To provide an explicit representation of knowledge of administrative processes in Guarantee fund of APV.

3. To provide ontological representation of the participants and documents in the development funds.

4. To describe electronic services profiles those are providing the groundwork for automatic generation administrative act.

Our main intention is to reduce development time and effort in order to meet the demands of the information system based on knowledge of the administrative processes and documents. The core idea in our method is ontology annotations with user interface components in order to automatic generate user interface and modeling ontology of the electronic services that will provide generation of the final administrative act.

Main advantages of the model which will be described hereinafter, is a semantic representation of the administrative procedures business logic which, in the use of traditional techniques for developing information systems, are firmly integrated into the user interface code. [4] [8]

The paper is organized as follows: the second section provides an overview of the related work. The third section presents an ontological model of the Guarantee Fund of APV. The fourth section shows the detailed administrative tasks semantic representation and shows practical use of designed ontology. In the final, fifth section, concluding remarks and directions for further research are given.

2 Related work

This section will present research results related to two aspects of the problem investigated in this paper. The first aspect relates to the use of ontologies in modeling business processes of public administration, and the second aspect relates to the use of ontologies in the generation of the administrative act. A ontology is an information structure, which helps to acquire knowledge, share it, and check consistency within knowledge. Authors in [19] conclude that the ontology provides a better communication, reusability and organization of knowledge by decreasing language ambiguity and structuring data. Author [18] proposed a formal definition of ontology as a 5-tuple (N, R, D, F, T) where each element is defined as follows: N, a set of nodes. R,N is a set of relation Types. D, is a set of description logic sentences. Each sentence can use the elements in N and two variables subject and object. Indicating respectively the first and third element in 5-tuple in T. F is a function that maps each element from R onto one element in D. T is a set of relations which is defined as a set of 3-tuples where for each element consists of (s, r, o) where: s is the subject, an element of N, r is the relation, an element of R and o is the object, an element of N.

The authors in [2], propose a model of the system in which the procedures performed by each administrative unit in the decision-making process and the creation of administrative documents are described. The main component of the proposed system is ontology of the public administration. The proposed ontology comprises following aspects of the observed domain. Structural aspects of public administration represent administrative units and their hierarchical relationships. Textual aspect of ontology describes the documents that appear as a product of the administrative unit described in the structural aspects of ontology. Procedural aspects of public administration ontology given in [2] are represented as an extension (specialization) to OWL-S. Service aspect is also represented by OWL-S. An ontology modeled in this way allows hierarchical control of the administrative units, control of administrative acts anticipated by legislation, communication between administrative units involved in the creation of administrative documents and calling of the procedures that trigger electronic services. The authors [5], [6], define three main reasons for the use of ontologies in generating user interfaces: 1. Improved visualization of UI; 2. Improvement of interactions between the system and users, 3.Improvement of the development process of user interface. In [3] the authors propose the creation of an ontology, which represents different aspects of service including administrative documents and legislation. This ontological model is focused on describing the electronic services of public administration. Basic concepts of the proposed ontology are Services, Service users, Organization, Administrative, Service implementation, Legislation, Form, Document, Event.

The ontological model created in this paper relies on the results of [2] and [3].

In a study of applying ontologies in the development of e-Government the authors presented a method of modeling ontologies in the domain of e-Government [15]. According to the authors of [9], each public service of eGovernment is semantically modeled and contains references that point to the required input data. Predefined values of input data or preconditions can be expressed with semantic rules. In this way, this allows automatic creation of (web) forms and interactive validation of input data. [13] According to the authors, the development of new applications and projects begins with modeling the ontology. When modeling the ontology, the authors used two types of classes and subclasses, by introducing the following assumptions:

Each class that contains a subclass is seen as an abstract class;

Each class that does not contain a subclass is considered a realistic class (the basis for the creation of form);

Electronic services accept only the instances of realistic classes as input data.

The presented rules and ways of modeling ontologies, according to the authors of [9], allow the unambiguous identification of e-Government services. The authors of [10] proposed the use of algorithm for the direct transformation from OWL to relational data structures. The authors of [11] also dealt with these problems. The basic idea is to transform the created ontology with the help of the transformation tools in the DLL script and thus preserve all relations, constraints, and information about the domain. DLL script is used to generate a relational database. Authors of [14] proposes a feasible implementation of a multi-agent environment which makes use of ontologies and ontology mapping to achieve semantic interoperability. Authors use an ontology model to facilitate semantic interoperability in a simulated multi-agent environment. The Authors in [16] exploits different semantic web technologies and builds a prototype of semantic web mash up functionality based on combination of RDF/OWL with SQWRL. The main scope is to improve decision-making processes. Sugumaran and Storey present a heuristics-based method for developing and creating ontologies [17]. They identify all the basic terms; this is done by using use cases and then revising synonyms and related terms manually or by an online thesaurus. In the next step they identify the relationships among these terms. They define three types of relationships: generalization, synonyms and associations. Generalization corresponds to an is-a-relationship. This paper will present the methodology of ontology construction and annotation of state development funds.

3 Creating an ontology

This section presents a method aimed at building an ontological model of state development funds. The basic administrative activities that are common to all development funds in APV are as follows: Registration of participants of the competition, Analysis of the submitted documentation, Risk Assessment of funds placement, The decision on funds placement, and Type of placement. It is noticeable that the administrative activities are almost identical for all state development funds, regardless of the type of fund and type of placement. The authors in [12] propose the following four-level typology of administrative activities of public administration: Identification-identification of types of services offered to end users:

1. Specifications specifying administrative procedures and documents for the identified type of service;

2. Interaction specifying a communication protocol for the identified type of service (for example, signing a contract);

3. Transaction the realization of services (for example, issuing guarantees for loan).

This typology can be applied to state development funds. An important role in the development of ontology represents the conceptualization and organization of knowledge. The task of conceptualization is to transform informal knowledge into an ontological concept with the help of professionals in their field of ontology modeling. Figure 1 shows a simplified model of the activities of state development funds. [7]

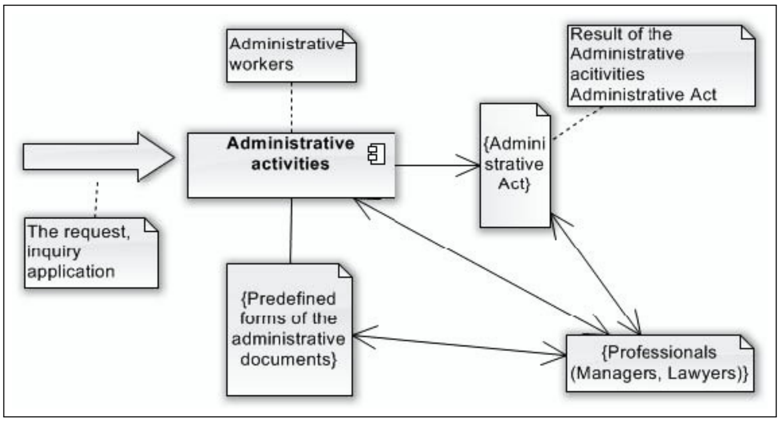


Figure 1: A simplified model of the activities of state development funds

Ontological descriptions of the administrative activities are generally applicable to all government development funds. Below is a detailed description of the ontological model of the Guarantee Fund of APV (GFAPV OM), which defines the administrative procedures and tasks.

3.1 Ontological model of the GFAPV

The public administration aspects given in [2] were taken as the basis for the creation of concepts in OM GFAPV. Ontology of public administration in [2] consists of two parts. In the first part of the ontology, two aspects are defined:

1. A structural aspect of public administration in which administrative units and their hierarchical relationship are described where hierarchical relationships between the administrative units were presented with properties (belongs to and supervised by).

2. The textual aspect of ontology describes the documents that appear as a product of the administrative units described in the structural aspect of the ontology. The textual aspect is presented in the ontology with four main classes as follows:

1. Administrative documents -presented as a product of administrative units;

2. Civil documents -all types of documents that people create and fill out in the course of communication with public administration;

3. The legal texts -the laws and secondary legislation;

4. Court decisions -decisions related to administrative actions and decisions of the Supreme Court.

In the second part of the ontology given in [2], both the procedural and service aspects are presented.

The procedural aspect of public administration ontology is presented as an extension (specialization) of OWL-S. The basic concepts of procedural aspects are tasks, procedures and full procedures. Tasks are atomic actions that cannot be further broken down and are executed by an administrative unit. Each task has input data that needs to be filled by the system administrator or worker. The result of executing the task is an administrative act. A procedure contains at least one informative task (i.e. it tries to find or notify some information from/to another unit) and only one executive task (i.e. it produces a single act). Thus, the procedures are composed of one or more tasks. Each task is executed within a framework of procedure. Full procedure represents a number of procedures intertwined. A full procedure may reflect to the provision of a service to one or several entities (property providedTo). Procedures may be sequential or in an acyclic graph. In this ontology, the control constructs of OWL-S are adopted. When defining the ontological model of OM GFAPV, the following elements of the domain were analyzed:

1. Participants. This aspect corresponds to the structural aspect of the model [2]. It represents internal funds structure and its position within the public administration, as well as actors who perform tasks defined in the business process and external participants to the process.

2. Documents. This part, which corresponds to the text aspect of the model [2], represents administrative acts appearing during the execution of tasks within business processes.

3. The business process logic, which corresponds to the procedural aspect of the model [2], is defined by business rules and operating procedures of the business system (state credit guarantee funds).

4. Electronic services are the services invoked in order to execute procedures implementing business process logic.

In the case of OM GFAPV, administrative procedures are defined as ontology concepts. Communication between administrative procedures and services is presented as atomic process. In the case of OM GFAPV each service belongs to only one procedure that is defined in the ontology. Services are described by service profile properties. The generalization (superclass-of) and specialization (subclass-of) of ontological concepts is represented by the taxonomy of the main concepts of the ontology of the Guarantee Fund of AP Vojvodina.

3.2 The taxonomy of the concept Participants

By this concept, all participants in procedures are represented. The concept Participants is shown in Figure 2.

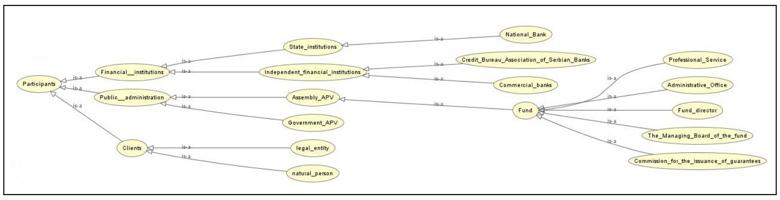


Figure 2: Participants-ontological concept

Three subclasses of the Participants class are defined in this way: Financial-institutions, Public- administration, Clients.

The Financial-institutions concept

This ontological concept describes financial institutions that cooperate with the Guarantee Fund as well as external participants in carrying out work procedures of the fund. Two groups of financial institutions were identified: 1. State institutions (The National Bank), 2. The Independent financial institutions (Commercial banks, Credit Bureau Association of Serbian Banks). The properties of these concepts are SubClassof and haveService. The property SubClass of describes the structural position of the concept, while the property haveService indicates the existence of external services that are invoked during the execution of procedures.

The Public-administration concept

The classes described by the concept Public-administration represent the organizational structure of state administration in AP Vojvodina and the place and role of the Guarantee Fund of APV in this organizational structure.

The Fund concept The classes identified in the taxonomy of the Fund are, as follows: Managing Board, Fund director, Professional Service, Administrative Office, Commission for the issuance of guarantees.

The Clients concept

The classes described by the concept Clients represent participants who are allowed to apply to Fund open competition. The classes identified in the taxonomy are: Legal entity and Natural person

3.3 The Documents concept

The concept Documents is created by analyzing all of the documents identified in the Guarantee Fund of APV. Figure 3 shows the taxonomy of the Documents concept.

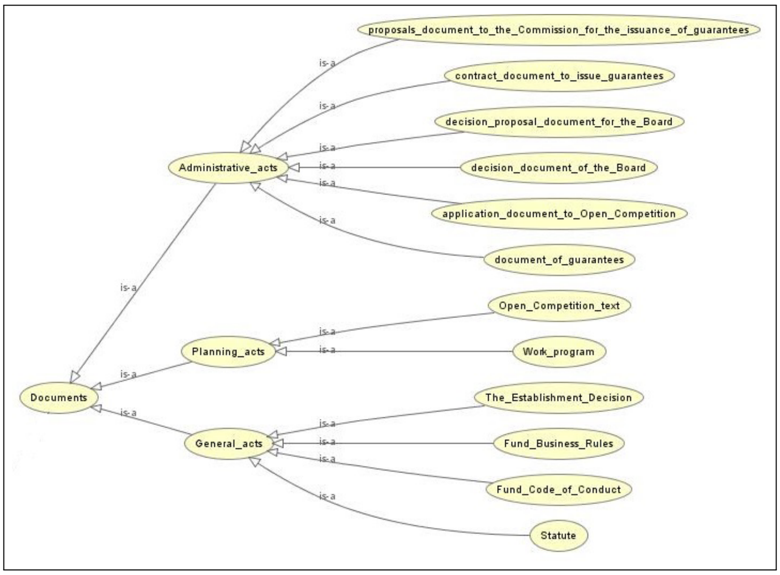


Figure 3: Taxonomy of the Documents ontological concept

Three groups of documents were identified: Administrative acts, Planning acts, and General acts.

The Administrative-act concept

As defined above, administrative documents are the product of administrative tasks within business procedures. Semantic descriptions of documents that have been identified as a result of the execution of administrative tasks are given below. If we look at the documents as a product of administrative tasks in the case of the Guarantee Fund of APV, the following documents that appear as a result of the execution of administrative tasks in the business process of issuing a guarantee can be identified:

The application document to Open Competition;

The proposals document to the Commission for the issuance of guarantees;

The decision proposal document for the Board;

The decision document of the Board;

The contract document to issue guarantees.

The document of guarantees Administrative-act concept is a subclass of the Documents concept. In addition, it is associated with the property Producedby that describes it in terms of administrative procedure which produces the document of Administrative-act type, and the property FillBy which determines the entity from the Participants taxonomy who fills out a pre-defined form of the document of Administrative-act type.

The Planning-act concept

Documents belonging to this group are created each year. This group includes the following documents: Work program, and Open Competition text. Apart from their property SubClass of which defines hierarchical relations among documents, they are associated with properties that describe them in terms of creation and approval:

Createdby defines the entity from the Participants taxonomy that created the document.

Acceptedby defines the entity from the Participants taxonomy that approved the document.

The General-acts concept

The documents that determine the legal framework of the fund belong to this group. In the case of the Guarantee Fund of APV, these are the following documents: The Establishment Decision, The Statute, The Fund Business Rules, and The Fund Code of Conduct. Apart from their property SubClass of which defines hierarchical relations among documents, they are associated with properties that describe them in terms of creation and approval:

Createdby defines the entity from the Participants taxonomy that created the document.

Acceptedby defines the entity from the Participants taxonomy that approved the document.

3.4 Procedures concept

The Procedures concept represents taxonomy of the administrative procedures of the fund. This taxonomy is created based on operational procedures for issuing guarantees and procedures relating to the creation of planning acts of the fund. In addition, the taxonomy contains procedures related to utilization of the documents that define legal and regulative framework of funds within the state administration. The taxonomy is shown on Figure 4.

The taxonomy of the Procedures concept consists of following classes: Planning-procedures, Operative-procedures.

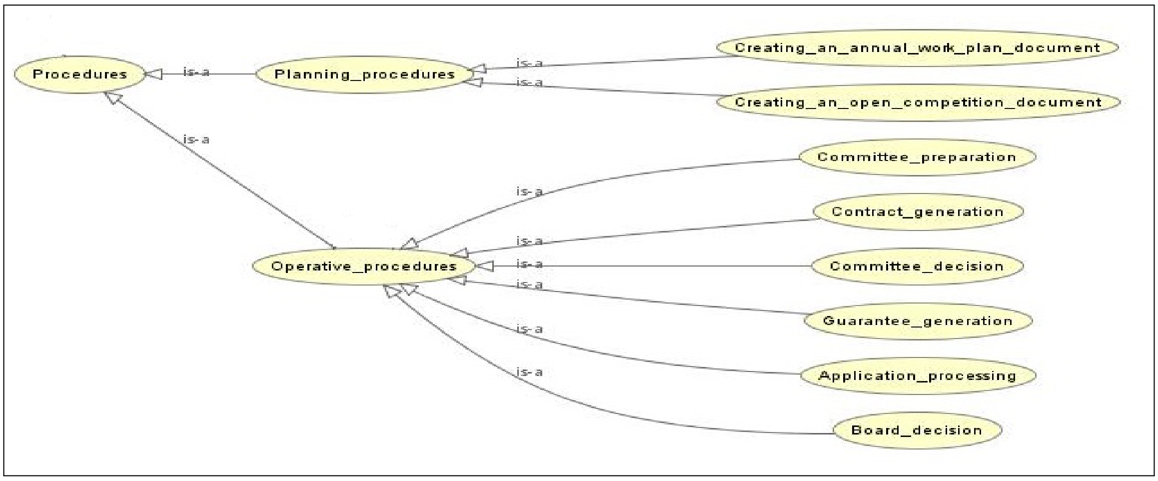


Figure 4: The Procedures taxonomy

Planning-procedures concept

This concept describes the procedures aimed at planning annual activities of the Fund. The outputs of these procedures are various documents that contain activities to be performed in a calendar year and related financial resources. The following procedures have been identified: Creating an annual work plan document, and Creating an open competition document. The properties of classes describe these procedures:

SubClass of identifies the structural position of the concept within the Procedures taxonomy.

HaveOutputDoc -identifies the document that results from performing general administrative procedures.

PerformedBy -identifies the perpetrators of the procedure.

GeneralActsReferenced identifies all General-Act concepts, which are legal/regulative basis for the created planning document.

Operative-procedures concept

Basis for creating the Operative-procedures ontological concept are: Task, Procedure, and Full Procedure as defined in [2]. Each task is executed within a Procedure. Each task has input data that an administrative worker or system should fill out and the result of the execution of the task is an administrative act. The Procedure contains only one task with the corresponding input and output data, which is filled out by an administrative worker or system. Full-Procedure is composed of one or more procedures, while Operative- procedure is composed of one or more tasks. In the case of the Guarantee Fund of APV, the Full-Procedure is the procedure of issuing the guarantee. Analyzing Operative-procedures concept the following classes are identified: Application-Processing, Committee-Preparation, Committee-Decision, Board-Decision, Contract-Generation, and Guarantee-Generation. Figure 5. depicts Application-Processing class, and properties and relations of the Operative-procedures concept.

Services that call the specified Operative-procedures during their execution are presented by their profiles (Pservice-A-P, etc.). Communication between these procedures and the service can be defined as an atomic process since the action of the service can be performed in a single interaction with the service.

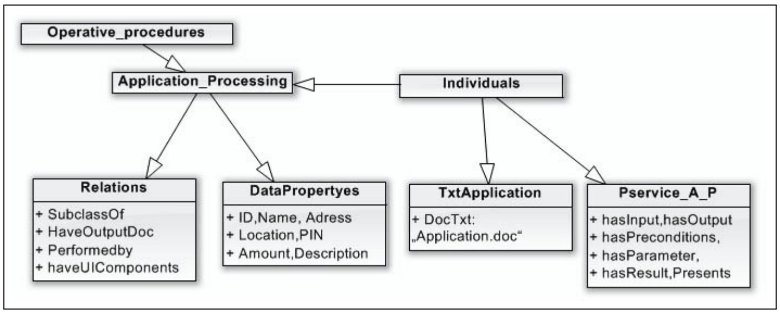


Figure 5: Properties and relations of the Operative-procedures concept

3.5 The Services concept

Classes of internal and external services that are invoked for executing the work procedures of issuing guarantees are described in this taxonomy. The following classes are defined: InternalServices and ExternalServices. Internal Services are the services that are provided by the Fund, while External services are those that are provided by third parties like banks, other administrative bodies, etc. Both have same properties SubClass of and availableOn. The property SubClass identifies the structural position of the concept within the Services taxonomy, while the property availableOn indicates service provider URI. Members of the class InternalServices (Service-A-P, Service-C-P, etc.) are presented by the PresentedBy property. This property points to the service profile that presents the service and the operational procedure that uses this electronic service.

4 The semantics of the administrative task

As described above, in our model, an Operative-procedure is composed of one or more tasks. We call these tasks Administrative tasks. Figure 6 shows the ontological representation of an Administrative task.

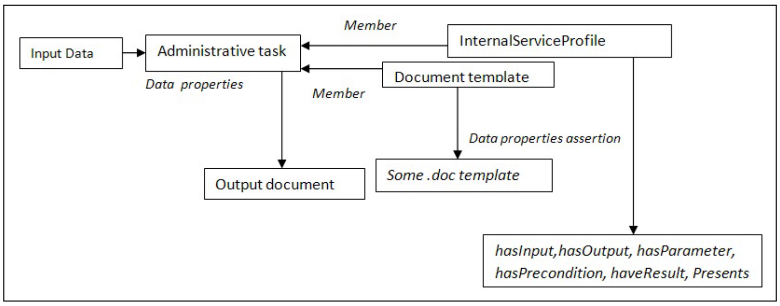


Figure 6: Ontological representation of an Administrative task

Each class and subclass of the ontology within the concept of the Operative-procedures has input data. Properties dataProperties describe input data of each Operative-procedures class. An administrative worker or service fills input data into pre-defined Document template(s) corresponding to an Administrative task. Document template is an active document that contains a code aimed at invoking a service presented by InternalServices profile. It serves as a starting point for a new document creation. Individuals of each Operative-procedures class represent corresponding service and corresponding document template that are filled with input data during the execution of an administrative task. Result of an Administrative task execution is a custom administrative act Output document (one or more in general, because the ontology allows it). Output documents of an administrative task have their corresponding Object Properties. These properties describe all relations between an Administrative task and concepts Documents (Producedby, FillBy) and Participants (Performedby). Figure 7 shows the ontological model of the administrative task aimed at processing applications on open competition.

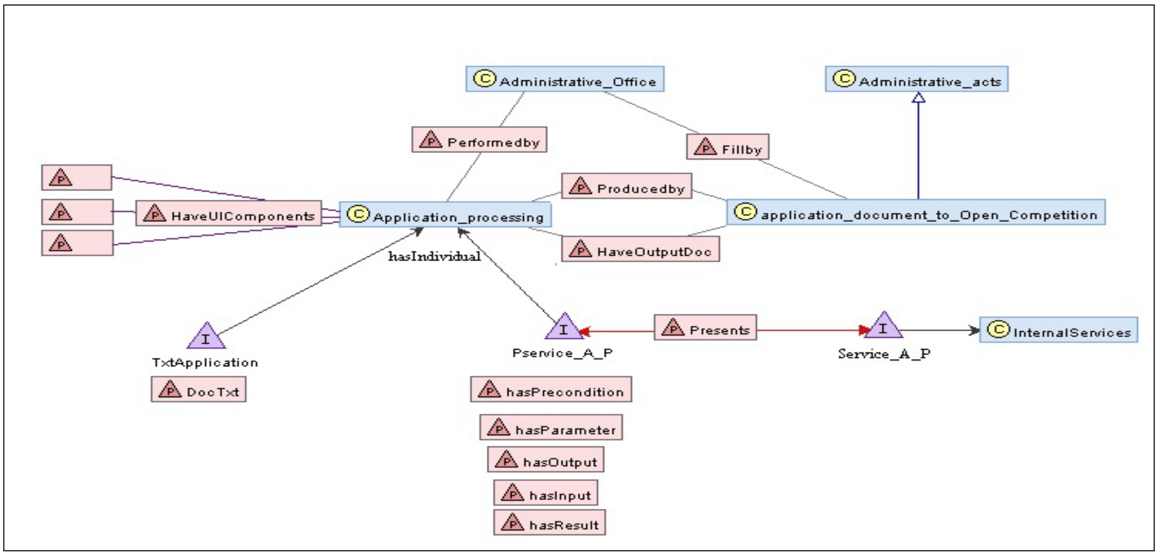


Figure 7: Ontological model of the Administrative task

Semantics of an administrative task can be described as follows: The procedure Application-processing is executed by the Administrative-Office. The document application-document-to-Open-Competition is the result of the execution of Application-processing procedure. The procedure for application processing has a unique document template TxtApplication. The service profile Pservice-A-P describes the electronic service invoked by the procedure Application-processing.

4.1 Application of the ontology of administrative task

The annotation of the ontology with elements of user interface is based on the rule that administrative tasks have input data. As shown in Figure 7, input data is a series of variables that are filled with the resulting document template of the administrative task. The names of the input data are represented with dataProperties of the observed task. The descriptions of input data are defined in dataProperties isDefinedBy property in the following form:

```
<Control>Component_type</Control>...(Type of the UI /database component )
<Data>data_type</Data>.(Data Type (string), (integer))
<Order>X</Order>.(Component index order on the form)
<Label>Component_label</Label>.(Label of the Component)
<Id>unique_name</Id>(Unique name of the defined Component)
```

Individuals of each class denote the template of the document that is filled with input data The feature Data-Properties-assertion references a document that represents the template.

Data-Properties-ID defines the order of execution of the administrative procedures and/or tasks.

An algorithm for a direct transformation of the semantic description of the administrative task in the user interface components is shown in Figure 8.

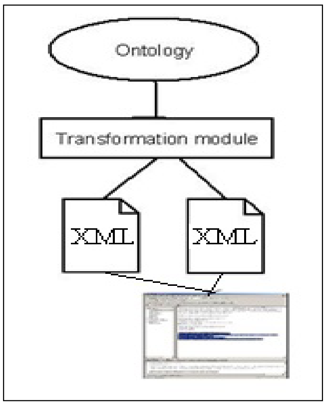


Figure 8: Transformation ontology into user interface

Module for semantic content transformation: The requirement set in the introductory section of this paper is that the created ontology should enable automatic generation of information system components. The basis of this procedure is the transformation of the created Administrative Task ontology represented in OWL format in two XML documents witch enable creation of the user interface. The module for transformation of semantic content represents the application that loads the created ontology and executes SPARQL queries. Parsing the query results generates two XML documents (OntoClass, OntoForm).

Administrative tasks sorted in the order of execution within the administrative process are presented by a Tree View component. For each administrative task, it is necessary to load the descriptions of the components of the information system from the property DataProperty - isDefinedBy.

The first XML document is a representation of classes and subclasses of the ontological concept Procedure that represents an ontological description of the administrative business processes. OntoClass.xml is an XML document, which represents a workflow of the defined administrative procedures. The second XML document represents components that are extracted from the semantic description of the administrative task. As shown in Figure 8, our model proposes the application of ontologies for the representation of administrative processes defined in workflows and extraction of information system components from the generated ontology. Figure 9 shows the user interface of the module for the transformation of semantic content.

The module for semantic content transformation provides the following functions:

Selection and loading of the desired ontology;

Entering the names of basic classes of the ontology that describe the tasks of the business process;

Generating the first XML document;

Entering the name of an administrative procedure for which it is necessary to generate user interface components;

Generating the second XML document;

Generating user interface for the selected administrative task.

As described, annotating administrative task ontology with user interface components, enable automatic creation of the User Interface for each procedure defined in the ontology. Depending on the application, the appropriate XSL file provides mechanisms of transformation and formatting OntoForm.xml and OntoClass.xml documents into user interface components.

Ontological representation of the administrative task contains the resulting document template of a selected task. Listing 1 displays the SPARQL queries by which we extract the name of the resulting document template of the chosen procedure.

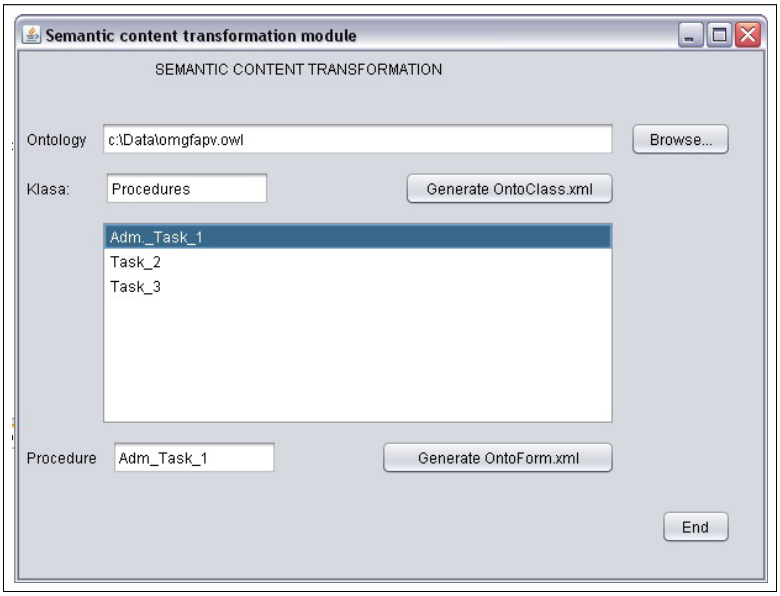


Figure 9: The module for semantic content transformation

```
Listing 1. Query code for the template extraction
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX gf: <http://localhost/GFontologies/2011/8/19/Ontology1316412802796.owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT ?x ?y WHERE {
  ?x rdf:type gf:Adm_Task_1
     ?x gf:DocTxt ?y
}
```

4.2 Using document templates

Each administrative task has exactly two ontological individuals: a corresponding service and a corresponding document template to be filled with input data during the execution of the administrative task thus generating the final document of this administrative task. IT and legal experts are responsible for creating document templates and corresponding electronic services based on the semantic annotation of the administrative task. An example of executing Application-processing procedure, which illustrates the flexibility of the administrative task representation achieved by the proposed administrative task semantic description and use of document templates, is shown in sequel:

When a user select desired administrative task (on the generated user interface Application-processing Figure 10) and fill out all the required fields, corresponding service generate document template shown on Figure 10. Document template acquires application data from the generated User Interface and enable generation of the final administrative act.

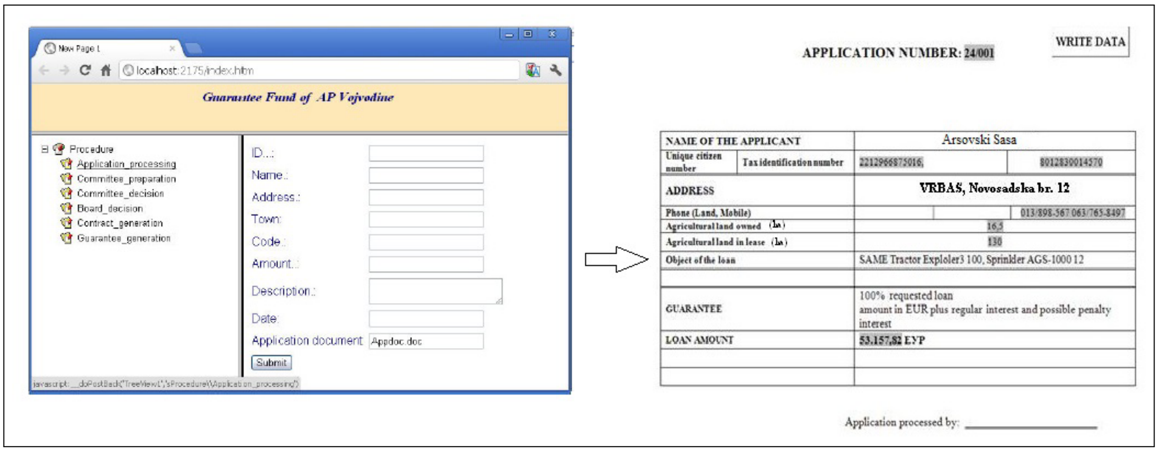


Figure 10: Functionality of the created document template

5 Conclusion

eGovernment solutions represent a major challenge in redefining the role of public administration agencies and organizations. Creating and using ontologies of knowledge of administrative processes and modeling and controling systems that would speed up and automate the work of state administration, constitute prerequisites for technical and organizational interoperability of different government agencies.

The proposed method reduces development time and effort in order to meet the demands of the information system based on knowledge of the administrative processes and documents. In this paper, we presented a methodology for ontology creation and annotation utilized to describe knowledge of the administrative task.

In this paper, we have proposed an ontological model of state development funds. The proposed model enables document templates creation, which are the framework for the automatic generation of an administrative act as a result of the administrative procedures execution. The proposed approach is verified by a case study of building an ontological model for the Guarantee Fund of APV. When defining the ontological model of OM GFAPV, the following elements of the domain were analyzed:

1 Participants, represents internal funds structure and its position within the public administration, as well as actors who perform tasks defined in the business process and external participants to the process;

2 Documents, represents administrative acts appearing during the execution of tasks within business processes;

3 The business process logic, is defined by business rules and operating procedures of the business system (state credit guarantee funds);

4 Electronic services are services that are invoked in order to execute procedures implementing business process logic.

The generalization (superclass-of) and specialization (subclass-of) of ontological concepts are represented by the taxonomy of the main concepts of the ontology of the Guarantee Fund of AP Vojvodina. An atomic process presents communication between administrative procedures and services. In the case of OM GFAPV, each service belongs to only one procedure that is defined in the ontology. Services are described by service profile properties. Following the proposed ontology, semantics of an administrative task is described in details. Finally, an example illustrating usage of the proposed ontology for implementation of Application-processing procedure is presented. Further research should focus on the development of ontologies of administrative processes within the domain of state bodies. Such ontologies could serve for fast and efficient creation of interoperable eGovernment applications.

Acknowledgement

This work was partially supported by the Serbian Ministry of Education and Sciences (Grant No: 171039).

Bibliography

[1] Guarino, N. (1998). Formal Ontology and Information Systems, *In Proc. of the First Int. Conf. on Formal Ontologies in Information Systems* (FOIS), 3-15.

[2] Savvas I, Basiliades N. (2009). A Process-Oriented Ontology-Based Knowledge Management System for Facilitating Operational procedures in Public Administration. *Expert Systems with Applications, An International Journal*, 36(3):4467-4478.

[3] Vassilakis C., Lepouras G. (2006). An Ontology for e-Government Public Services, Encyclopedia of E-Commerce, E-Government, and Mobile Commerce 721-728.

[4] Paulheim H., Probst F. (2010). Ontology-Enhanced User Interfaces, A Survey. *International Journal on Semantic Web and Information Systems* (IJSWIS), DOI: 10.4018/jswis.2010040103, 6(2), 24 p.

[5] Kleshchev A.,Gribova V. (2003). From an ontology-oriented approach conception to user interface development, *International Journal Information Theories and Applications*, 10:87-93.

[6] Paulheim H., Probst F. (2010). Improving UI Integration with Formal Semantics, SAP Research CEC Darmstadt Bleichstrasse 8 64283 Darmstadt, Germany, February 7th, 2010 SEMAIS Workshop @ IUI.

[7] Furtado, E., Furtado, J.J.V., Silva, W.B., Rodrigues, D.W.T., Taddeo, L.S., Limbourg, Q., Vanderdonckt, J. (2002). An Ontology-Based Method for Universal Design of User Interfaces. In: *Seffah, A., Radhakrishnan, T., Canals, G. (eds.) Proc. of Workshop on Multiple User Interfaces over the Internet: Engineering and Applications Trends* MUI 2001 (Lille, September 10, 2001), In Task Models and Diagrams For User Interface Design (TAMODIA 2002).

[8] Liu B., Chen H., He W. (2005). Deriving User Interface from Ontologies: A Model-Based Approach. In *ICTAI '05, Proceedings of the 17th IEEE International Conference on Tools with Artificial Intelligence*, Washington DC, USA; IEEE Computer Society, 254-259.

[9] Salhofer P., Stadlhofer B., Tretter G.(2009). Ontology Driven e-Government. *The Electronic Journal of e-Government*, 7(4):87-93.

[10] Vysniauskas E., Nemuraite L. (2006). Transforming ontology representation from owl to relational database, *Information technology and control*, 35(3a):333-343.

[11] Gali A., Chen C.X., Claypool K.T., Uceda-Sosa R. (2005). From Ontology to Relational Databases. Shan Wang et all (Eds.), Conceptual Modeling for Advanced Application Domains, *LNCS*, 3289:278-289.

[12] Todorovski L., Leben A., Kunstelj, M., Cukjati, D., Vintar, M. (2006). Methodology for Building Models of Life Events for Active Portals. In *Gronlund, A. et al. (Eds.):Communication proceedings of 5th EGOV International Conference, EGOV 06*, Trauner Verlag, 61-68.

[13] Berners-Lee, T. (1999). *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor*, Harper San Francisco.

[14] Toma I.F. (2010). Contributions to the Study of Semantic Interoperability in Multi-Agent Environments -An Ontology Based Approach, *Int J Comput Commun*, ISSN 1841-9836, 5(5):946-952.

[15] Klischewski R. (2003). Semantic Web for E-Government. R. Traunmuller (Ed.): EGOV 2003, *LNCS*, 2739: 288-295.

[16] Necula S.C. (2012). Implementing the Main Functionalities Required by Semantic Search in Decision-Support Systems, *Int J Comput Commun*, ISSN 1841-9836, 7(5):907-915.

[17] Sugumaran V., V. Storey C. (2002). Ontologies for conceptual modeling: their creation, use, and management, *Data Knowl. Eng.* 42(3): 251-271.

[18] Gruber T. (2007). Ontologies, Web 2.0 and Beyond. Keynote presentation at the Ontology Summit 2007 Ontology, Taxonomy, Folksonomy: Understanding the Distinctions, March 1, 2007.

[19] Ramanauskaite S., Olifer D., Goranin N., Cenys A. (2013). Security Ontology for Adaptive Mapping of Security Standards, *Int J Comput Commun*, ISSN 1841-9836, 8(6):878-890.

INT J COMPUT COMMUN, ISSN 1841-9836
9(3):276-283, June, 2014.

Enhancing Nodes Lifetime Optimum Protocol for Dissemination of Information in WSN

M. Gholami, A. Panahi

Mohammadreza Gholami
Department of Electrical Engineering,
Saveh Branch, Islamic Azad University,
Saveh, Iran
E-mail: Reza5762@ yahoo.com

Abdorreza Panahi*
Department of Mathematics,
Saveh Branch, Islamic Azad University,
Saveh, Iran
*Corresponding author: panahi53@ gmail.com

Abstract: Challenging issue in Wireless Sensor Networks (WSNs) is assessment of energy and lifetime at different nodes within the networks. Various methods may be employed to improve lifetime (reduce energy consumption). One such method involves balancing loads on nodes when data is being transmitted from source to destination nodes. Multi-path routing techniques can be used for this purpose. In these techniques no global information is available regard the path, making it difficult to create multi-path routes from sources nodes to destination nodes. Another problem with these networks is a routing applied to source nodes independently from that applied used for destination nodes. This creates energy loss and reduces lifetime. To overcome this problem, the present paper makes use of clustering by selecting virtual nodes to gather information from sources and sending it to destination nodes. The New Protocol for Enhancing Nods Lifetime (PENL) is implemented through NS-2.
Keywords: Wireless Sensor Networks (WSN), optimum routing, dissemination of information, enhancing nodes lifetime.

1 Introduction

Sensor networks have experienced a considerable growth over the recent years [4]. These networks are composed of a large number of nodes of very small sensors used to collect and process environmental information [5]. Nodes in sensor networks usually do not have unique addresses, and the information collected through nodes is of greater importance [6]. In addition, the nodes become inaccessible once they are distributed in the environment. They become useless (reach the end of their lifetime) once they consume the available energy [8]. Therefore, energy and its optimization is a major challenge is sensor network and received attention from a large body of research over the past years [1, 3]. To this date, sensor networks have found increasing applications in different areas including military, environmental monitoring, medicine, agriculture, and so on. One data-centric method proposed for routing data in sensor networks is directed diffusion [7], in which nodes use only local data in routing packets. In this method, interest packets are disseminated over the network and to all nodes by basic nodes. Then the nodes containing information of interest (information sources) receive these packets and direct collected information to the destination node. The present paper attempts to overcome problems (e.g. late aggregation, extra explanatory data, and high levels of power consumption) experienced in previous protocols. Figure 1 illustrates the process.

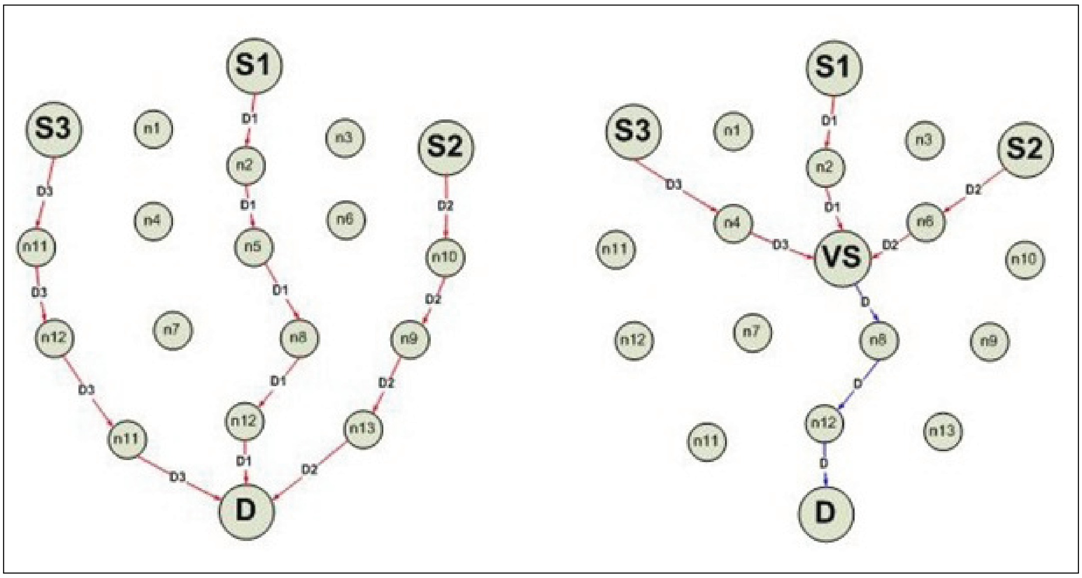


Figure 1: Comparison between routing in MECH (left one) and PENL (right one)

The method uses a virtual node in vicinity of source nodes to collect information and send it to destination nodes. This paper presents PENL and uses the routing method described above to improve lifetime and reduce overload compared to MECH [2], as we shall see in simulation results presented in the final section of this paper.

2 Maximum energy cluster head

In MECH [2], source and destination nodes use characteristics of the graph to determine information that needed to be disseminated and to find a multi-direction efficient path connecting source and destination nodes. To send data, an interest message is disseminated over an area of interest in the network. Each node remembers the node through which it has received information and assigns a gradient to that node. The gradient represents both the direction of information flow and the status of query (which can be active or inactive). If the node is able to predict the next path using the gradient, then it delivers the query to an adjacent node related to that query; otherwise, the query will be sent to all adjacent nodes. The sending node will be recognized as a source. When being send to destination, data is stored in intermediate nodes in order to prevent repeated sending. If one node stops working, other nodes will try to locally recover the path. Once initial exploratory data are sent, the next data will be sent only through reinforced paths. Source nodes alternatively send exploratory data from time to time to update gradients based on dynamic changes in network.

Properties of MECH:

- MECH uses neighbor-to-neighbor or step-by-step in which each node can interpret data.
- Information diffusion is a data-centric method and all connections in a WSN use interests to determine named data for dissemination.
- Nodes are not assigned globally unique addresses and since each node can individually interpret data, it is possible to reduce data load and send data in concise form.

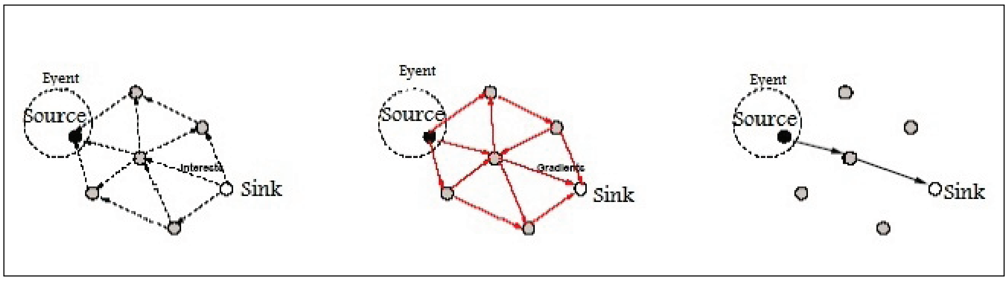


Figure 2: MECH protocol

A drawback of MECH is the increased number of data steps which, in turn, shortens lifetime of nodes and the overall network. NPDI, described below, overcomes these issues.

3 The proposed protocol (PENL)

As seen in C, an appropriate node close to source nodes is selected as virtual node. In D, the virtual node creates a path to destination node. In E, another node adjacent to the first virtual node is selected as the second virtual node. And finally in F, in cases where source nodes do not receive local interest packets for a while, they overlook the virtual node and send collected data directly to the destination node.

3.1 Selecting virtual node

A major and one of the most difficult steps in PENL is selecting virtual node which has to:

- Be spatially close to sources and have the largest number of nodes adjacent to it in order to be able to collect data as quickly as possible.
- Maintain a minimum level of energy above some threshold (e) in order to be able to handle a large amount of data.

Since each node reports its location, it is easy to select a virtual node with the largest number of adjacent nodes close to sources (the goal is to prevent reduction in lifetime as a result of information dissemination). Selection of virtual node in this manner meets the above mentioned properties to a large extent. The minimum distance for the virtual node (denoted by D) can be determined based on node density in the network. In simulations through NS-2, D is equal to 3 steps. As seen in Fig. 3, VS1 meets the above mentioned conditions and therefore is selected as virtual node.

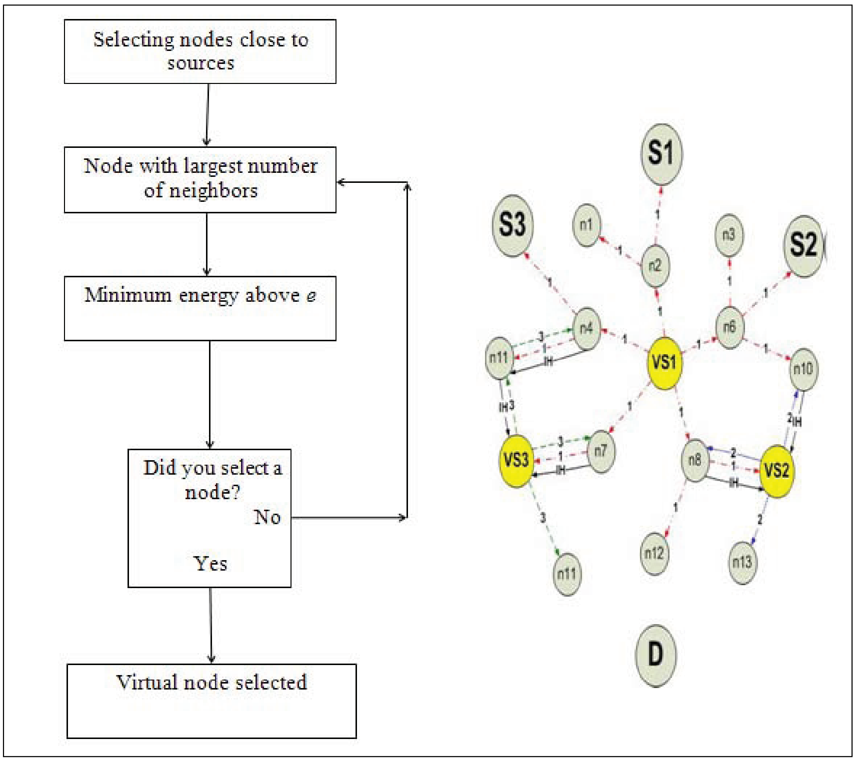


Figure 3: Flowchart for selecting virtual node

3.2 The shortest path

The shortest path problem can be formulated into a linear programming model:

$$
\begin{aligned}
&Min\ z = \sum_{i}\sum_{i\neq j} c_{ij}x_{ij} \\
&st:\ \sum_{k\neq j} x_{jk} - \sum_{k\neq j} x_{kj} = 1 \qquad if\ j\ is\ source \\
&\quad \sum_{k\neq j} x_{jk} - \sum_{k\neq j} x_{kj} = 0 \qquad if\ j\ is\ neither\ source\ nor\ destination \\
&\quad \sum_{k\neq j} x_{jk} - \sum_{k\neq j} x_{kj} = -1 \qquad if\ j\ is\ destination
\end{aligned}
\tag{1}
$$

The model does not account for the direction of edges (links). Each link may transmit data from j to k or in opposite direction. The following graph shows a part of a network where this technique is applied to find the shortest path. Here, the goal is to find the shortest path between Vs and S. Fig. 4 presents a typical graph of a WSN.

$$
\begin{aligned}
&x_{01} + x_{02} + x_{03} - (x_{10} + x_{20} + x_{30}) = 1 \\
&x_{10} + x_{13} + x_{14} - (x_{01} + x_{31} + x_{41}) = 0 \\
&x_{20} + x_{23} + x_{25} - (x_{02} + x_{32} + x_{52}) = 0 \\
&x_{30} + x_{31} + x_{32} + x_{34} + x_{35} + x_{36} - (x_{03} + x_{13} + x_{23} + x_{43} + x_{53} + x_{63}) = 0 \\
&x_{41} + x_{43} + x_{46} + x_{48} - (x_{14} + x_{34} + x_{64} + x_{84}) = 0 \\
&x_{52} + x_{53} + x_{56} + x_{57} - (x_{25} + x_{35} + x_{65} + x_{75}) = 0 \\
&x_{63} + x_{64} + x_{65} + x_{67} + x_{68} + x_{69} - (x_{36} + x_{46} + x_{56} + x_{76} + x_{86} + x_{96}) = 0 \\
&x_{75} + x_{76} + x_{79} - (x_{57} + x_{67} + x_{97}) = 0 \\
&x_{84} + x_{86} + x_{89} - (x_{48} + x_{68} + x_{98}) = 0 \\
&x_{96} + x_{97} + x_{98} - (x_{69} + x_{79} + x_{89}) = -1 \\
&a \leq x_{ij} \leq a + \epsilon \qquad (i, j = 1, 2, \ldots, 9) \\
&Min\ z = a_{01}x_{01} + a_{02}x_{02} + \ldots + a_{89}x_{89}
\end{aligned} \tag{2}
$$

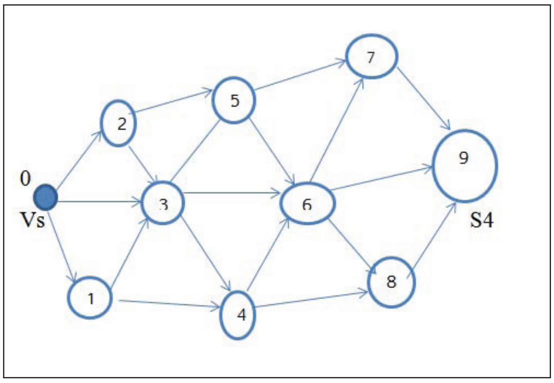


Figure 4: Connecting paths in a typical graph

3.3 Selecting a new virtual node

Since virtual node is required to handle a large amount of data, it will be eliminated once its energy is used up. To prevent this, a new virtual node will be selected after a certain period of time is passed. In PENL, this time interval is denoted by Pe. When Pe is reached, the virtual node sends an NR message to its neighbors requesting them for sending an Na message in response indicating the remaining amount of energy for the virtual node. The virtual node allows for a delay in responses and then selects an adjacent node with the highest level of energy as the new virtual node. An SN message is sent to this node. The new node will disseminate an interest message to update paths to the new node. After a certain period, the new virtual node floods exploratory data globally over the network to find a path to the destination node. The process is illustrated in Fig. 5.

3.4 Virtual node expiration

In some cases virtual node may become inoperative. This can be caused by different factors such as hardware problems, expiry of working period, used up energy, failure in finding a node with required level of energy, etc. In this case, if virtual node is still working, a message will be sent over the network to request source nodes for overlooking the virtual node and sending individual packets of exploratory data over the network to find a path to the destination node. The process is illustrated in Fig. 5.

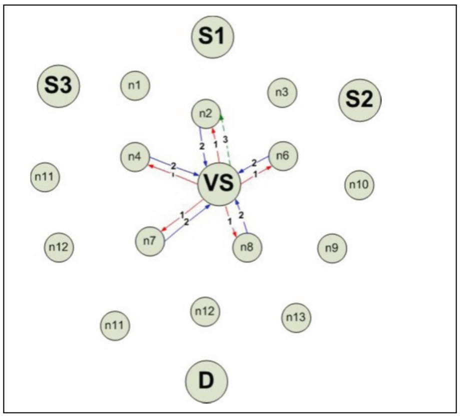


Figure 5: Selecting a new virtual node using the original virtual node

4 Analysis of simulation results

This section presents simulation results obtained through NS-2.

4.1 Routing overload

Overloads contain additional bits used to identify and correct errors. This increases the level of disseminated unwanted information and redundant processing at intermediate nodes as well as end stations. Fig. 6 shows routing overload. As seen in this figure, routing overhead of PENL when the number of sources is greater than 2, is smaller compared to that of MECH.

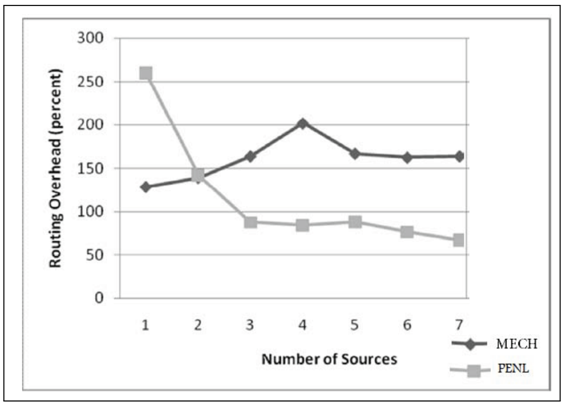


Figure 6: Comparison of routing overload in PENL and MECH

4.2 Packet loss rate

Fig. 7 compares number of lost data packets for different number of sources when connection is established between source and destination nodes. As seen in this figure, packet loss rate does not change considerably with the increase in the number of sources for PENL while increase in the number of sources for MECH, particularly from 6 to 7 sources, significantly raises packet loss rate.

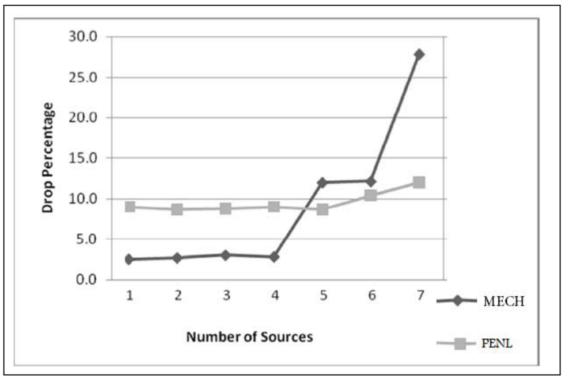


Figure 7: A comparison of packet loss rate of nodes in PENL and MECH

4.3 Energy consumption in networks

Fig. 8 shows overall energy used by all nodes of the network. As seen in this figure, PENL is much more energy-efficient compared to MECH mainly because of reduction in the number of transmission paths. In this simulation 7 sources were used over a 25 × 15 grid containing 320 nodes.

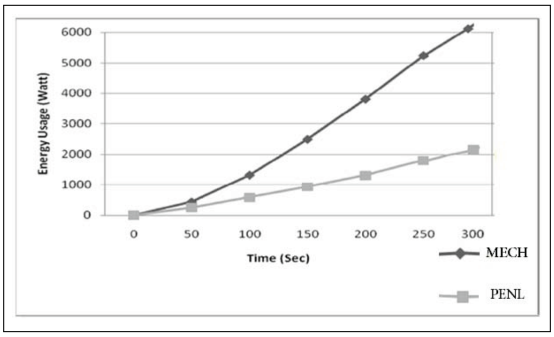


Figure 8: A comparison of energy consumption of nodes in PENL and MECH

5 Conclusions

In this paper, PENL was proposed as a protocol to improve routing in WSNs. The protocol is often used to increase efficiency of the previously used protocols in terms of energy consumption by reducing routing overload and balancing loads on nodes. PENL outperforms previous protocols such as MECH in many aspects including packet loss rate, energy consumption, and routing overload.

Bibliography

[1] Allahviranloo, T. et al (2009); A computational method to find an approximate analytical solution for fuzzy differential equations, *Analele Stiintific Ale Universitatii Ovidius Costanta-Seria Matematica* ISSN 1224-1784, 17(1): 5-14.

[2] Chang, R.S.; Kuo, C.J. (2006); An energy efficient routing mechanism for wireless sensor networks, *Proceedings of the 20th International Conference on Advanced Information Networking and Applications*, 2(6): 308-312.

[3] Chong, C.Y.; Pumar, S.; (2003); Sensor networks: Evolution, opportunities, and challenges *Proceedings of the IEEE*, 1247-1256.

[4] Dai, L., Xu, H.K. Chen,T., Qian, C. Xie, L.J. (2014). A Multi-objective Optimization Algorithm of Task Scheduling in WSN. *Int J Comput Commun*, ISSN 1841-9836, 9(2):160-171.

[5] Gaomez, M.; Perez, G.M. (2008); Providing trust in wireless sensor networks using a bio inspired technique, *Proc. of the Networking and Electronic Commerce Research Conference*, 312-321.

[6] Heinzelman, W.R. et al (2000); Energy-efficient communication protocols for wireless microsensor networks, *Proceedings of the Hawaii International Conference on Systems Sciences.*

[7] Levis, P. et al (2004); The emergence of networking abstractions and techniques in tinyOS, *Proceedings of the First USENIX/ACM Symposium on Networked Systems Design and Implementation.*

[8] Manjeshwar, A.; Agrawal, D. (2001); A routing protocol for enhanced efficient in wireless sensor networks, *The 15th International Parallel and Distributed Processing Symposium*, DOI: 10.1109/IPDPS.2001.925197, 2009-2015.

INT J COMPUT COMMUN, ISSN 1841-9836
9(3):284-291, June, 2014.

LAR-1: Affirmative Influences on Energy-Conservation and Network Lifetime in MANET

N. Gupta, R. Gupta

Neelesh Gupta
Department of ECE,
TRUBA Institute of Engineering and Information Technology,
Bhopal (M.P.)-INDIA
E-mail: neelesh.9826@gmail.com

Roopam Gupta
Department of Information Technology, UIT, RGPV, Bhopal (M.P.)-INDIA
E-mail: roopamgupta@rgtu.net

Abstract: Nowadays, reduction of energy-consumption in Mobile Ad-Hoc Networks (MANETs) has been a herculean task. Appropriate location-based routing protocols like Location-Aided Routing-1 (LAR-1) can be incorporated for reducing the energy consumption as it extends the network lifetime in dynamic network. The research work proposed integrating energy-conservation along with LAR-1 route-discovery mechanism, named as New-Location-Aided Routing-1 (N-LAR-1). In this paper, developments of performance-metrics using this N-LAR-1 approach have been reported. Affirmative results have been achieved through N-LAR-1 by utilizing the sleep or inactive mode condition of mobile nodes in Ad-Hoc Networks. It has been observed that performance of N-LAR-1 is better than N-DSR approach.
Keywords: MANETs, LAR-1, Energy Conservation, Network lifetime.

1 Introduction

MANET is a self-configuring, self-organizing infrastructure less network. For instance in military application, enemy tank can be tracked within the network at the time of battle. Similarly, local community can use an Ad-Hoc network to detect moving car and its speed along with direction. Presently limited energy, limited bandwidth, multi-hop routing, dynamic topology and security are the leading technical challenges in MANETs. Numerous routing protocols have been proposed to overcome these challenges in Ad-Hoc networks [1]. Routing protocols play an important role in dynamic network communication. Commonly, routing protocols are divided into topological and position-based protocols. Further topological-based routing protocols may be classified mainly into proactive, reactive and hybrid types. On the other hand, position-based routing protocols diminish the limitations of topological-based routing protocols using location information via GPS [2]. There are three position-based routing approaches such as, Greedy, Restricted Directional-Flooding and Hierarchical type techniques. Earlier several position-based protocols had been projected such as LAR-1, Distance Routing Effect Algorithm for Mobility (DREAM), Most Forward within Distance R (MFR) and so on.

LAR-1 is widely used source routing protocol, like a DSR. It is a type of restricted directional flooding based position-based routing protocol. Initially it starts flooding in network by the source but after expecting destination, the routing will be only in the direction of the destination. Since nodes in MANET are battery dependent. To extend network lifetime, routing protocol should concentrate for achieving energy conservation in the network [3] [4]. In this paper, N-LAR-1 technique has been developed with association of Power-Aware Dynamic Source Routing (PADSR) energy model in MANET. This proposed work is an endeavour to address challenging the issue of developing an effective energy-conserving technique for MANET. This paper also shows optimization of Ad-Hoc network performance metrics using proposed N-LAR-1 technique. In second section, literature survey has been presented. The short description of conventional LAR-1 protocol is shown in next section.

Afterward N-LAR-1 is elaborated in next portion. In fifth section, simulation results and its effect on proposed technique have been displayed. Last part concludes the paper.

2 Preliminaries

MANET is the group of wireless mobile nodes without any pre-defined infrastructure. Each mobile device has a limited energy and additional energy is required to forward the packets during route discovery mechanism in Ad-Hoc Network. Therefore, energy-conservation is required in network. Several researchers have been contributed in the era of energy-conservation. This section shows literature survey of work on LAR-1, energy-conservation on LAR-1 along with comparison amongst various routing protocols and energy-conserving routing techniques in MANET.

At first optimization in route discovery overhead was given by Ko Y.B. et al. [5] in the form of Location-Aided Routing (LAR) protocols in MANETs. They suggested an approach to utilize the location information via GPS to improve performance of routing protocols in the network. They developed two location-aided routing algorithms LAR-1 and LAR-2. Both use directional forwarding flooding, in which the source node floods data packets in the direction of the destination node. Both limit the search for a route in restricted area. They had concluded that routing overhead is reduced using LAR-1 scheme.

Ahvar E. et al. [1] analysed the performances of LAR-1, Dynamic Source Routing (DSR) and Ad-Hoc On-demand Distance Vector (AODV) routing protocols on the basis of energy consumption in Ad-Hoc networks. This analysis was accomplished by varying the network load, size and node mobility of the network. They concluded that LAR-1 is much better than others and affords energy-conservation in high node density networks. Xu Y. et al. [6] proposed an energy saving Geographic Adaptive Fidelity (GAF) algorithm. It concentrates on turning the radio off as much as possible. GAF nodes utilize geographic location information to segregate into fixed square grids. Nodes within a grid, switch between sleeping and listening state with the guarantee that one node in each grid stays up to route packets. Their results show that it can save 40-60% more energy than conventional protocol. Patel A. et al. [7] recognized an unreachable corner of GAF scheme and its effect in the network. GAF-h (Hexagonal) algorithm has been derived and checked with different traffics by them. It replaces virtual square grid into hexagonal grid. They concluded that it optimizes the Packet Delivery Fraction (PDF) and throughput of the network.

Chen B. et al. [8] proposed a power saving protocol SPAN that reduces energy consumption without shrinking the capacity or connectivity of an Ad-Hoc network. It adaptively elects coordinators and to form a fundamental structure that has set defined rules that are primarily based on the residual energy at nodes plus the number of nodal neighbours for coordinator announcement and withdrawal. Further Joshi N. et al. [3] modified LAR-1 scheme in terms of energy-constrained operations is known as Variable Range Energy aware Location-Aided Routing (ELAR-1VAR). It controls the transmission power of a node according to the distance between the nodes. They evaluated ELAR-1-VAR with LAR-1 in the terms of PDF, End-to-End (E2E) Delay and average energy consumption. In conclusion, it improves the network lifetime by reducing energy consumption in the network. Ramkrishnan S. et al. [9] exposed PADSR protocol and implemented on small to medium size network. It is concluded by them that PADSR protocol outperforms than DSR protocol by power saving of 30% in MANET.

3 Conventional LAR-1

The dynamic nature of MANET results in frequent and unpredictable changes of network topology, adding complexity to routing among the mobile nodes. The chief aim of routing is to find and maintain route between nodes in a dynamic topology with possibly unidirectional links using minimum resources in Ad-Hoc Network.

Several routing protocols have been proposed for MANETs to accomplish effectual routing. One such example is LAR-1 scheme that was utilized for reducing the transfer range of broadcasting during route discovery in networks. The objective of LAR-1 is to perform more efficient route discovery and limit the flooding of route request packets. At first Ko Y.B. et al [5] presented the idea of utilizing the location information for mobile nodes in terms of route discovery optimization, is called LAR-1. LAR is one of the most popular amongst location-based routing protocol. Conventional LAR-1 is an on-demand restricted direction flooding location-based routing protocol. It uses source routing like a DSR, but after expecting the destination packets have flooded in the direction of destination only. Source routing has been used with statically and dynamically configured routes for routing in MANET. Concept of LAR-1 routing depends upon request and expected region. Request region comprises of source node and expected region. If location is available, a request region (including the expected region) is formed that is defined as the region containing set of nodes that should forward the route discovery packet. Request region is also defined by location information of destination node. Size of this region depends upon movement of destination, elapsed time in which prior location of destination is included. As a result, the routing overhead reduces and better performance of LAR-1 protocol is achieved in the network [5] [10]. LAR-1 can be effectively used when an infinite queue of data packets is to be sent in the network.

4 Proposed Technique

Proposed technique integrates the method of energy-conservation of mobile nodes and route-discovery mechanism via LAR-1, is named as, N-LAR-1 technique.

4.1 General Criterion of the technique

1. At first, Source node (S) broadcasts (B) RREQ packets to nodes within their individual radio range (Rj).
2. Whole area in request region from S to D (Destination) is divided into numerous hexagonal grids within the radio range of mobile nodes in the network.
3. At the time of flooding, only having higher energy (non-survival condition of energy) of nodes are selected, called Coordinators.
4. At the same time, other remaining nodes will go into the sleep mode correspondingly and save their energy in the network.
5. Through these coordinators D is achieved and after some time Acknowledgment(ACK)goes to the S from D that contains the information of destination. Otherwise, the process is restarted.
6. Now, destination is achieved in flooding as a DSR. Hence, the computing the new route occurs through LAR-1 schemes again only in the direction of an expected region of destination node.
7. At a moment, the route is established in the direction of destination through the selected coordinators under LAR-1 scheme within the request region of an Ad-Hoc network.

4.2 Algorithm

STAGE-1

1. Create mobile node mj;
2. Set Source Node S, $S\xi mj$;
3. Set Destination Node D, $D\xi mj$;
4. Set routing protocol, DSR & LAR-1;
5. Set radio range Rj // Rj=550 m;
6. Compute _ route (RREQ_B, S, D);
7. If (next_neighbor node == true, energy≥10 J & next_neighbor≤Rj);
8. Next_neighbour_node_table (A, B, C, D - - - -);
9. Check_Eng = Max (A_Eng, B_Eng, C_Eng, — - - - - - -) ;
10. Set C = = Max_Eng_Node // for coordinators selection;
11. If (next_neighbour_node = = D);
12. find destination_node;
13. send ACK to S via RREP through selected route ;
14. Call to LAR-1;
15. Else Go to Step-7

STAGE-2

1. Call to LAR-1;
2. Information_D (elapsed time, speed, radius)_ expected region, send to S;
3. S Broadcast (B), RREQ only in D_expected-region;
4. Compute_route (RREQ_B_expected_region, S, D);
5. If (next_neighbour node = = true, energy≥ 10 J & next_neighbour≤ Rj;
6. Next_neighbour_node_table (A, B, C, D - - - -);
7. Check_Eng = max (A_Eng, B_Eng, C_Eng, — - - - - - -) ;
8. Set C = = Max_Eng_Node //for coordinator selection;
9. If (next_destination_node = = D);
10. Route is established;
11. send ACK to S via RREP from D;
12. Else Go to Step-5.

In figure 1, the expected region shows circular area radius of R = V $(t_1 - t_0)$. The request region is divided into number of hexagonal grids 1,2,3,4,5,6,7,8,9,10,11,12 (shown above in figure 1). A, e, F, w, q, I, K, a, v, m, O, s are the route forwarding nodes having the higher energy values than other nodes in their respective grids. Other remaining nodes will be in sleep or inactive mode at the same time and save their energy in MANET. Finally, dashed line in figure 1 shows the route from S to D using proposed N-LAR-1 energy-conserving routing technique. Thus, number of hexagonal grids is to be monitored and forward route tracing is more reduced with decreased routing overhead, E2E Delay and energy consumption in the network. In such a way, energy conserves and network lifetime is enhanced in Ad-Hoc networks.

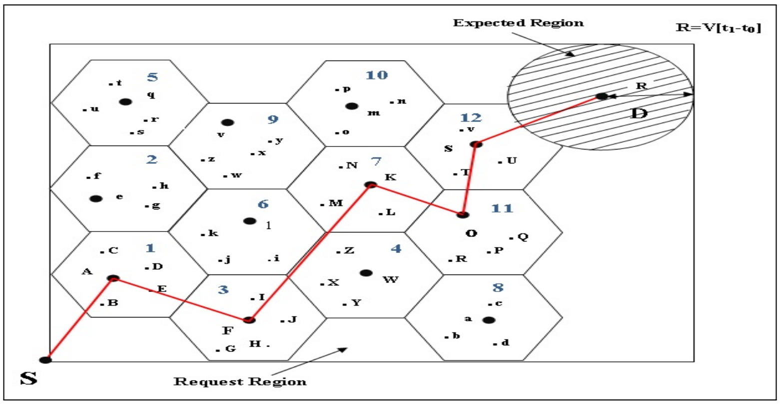


Figure 1: N-LAR-1 Routing Technique

5 Simulation and Results

The simulation was conducted in the Network Simulator (NS2) environment. All simulations were performed on Intel (R) core i3 CPU, 2.3 GHZ, 3072 MB of RAM running on Inspiron N5010 configuration. Simulation includes physical, data-link and Medium Access Control (MAC) Layer models. The Distributed Coordination Function (DCF) of IEEE 802.11 is used to model the contention of nodes for the wireless medium. The radio model uses characteristics similar to Lucents Wave LAN direct sequence spread spectrum radio [11]. The simulations were run for 100 seconds. The number of nodes (n) in network is 25, 50, 75, 100 and situated in a network area of 800 x 600 square meter region having a transmission range of 550m. In implementation, simulator setup (Radio Propagation Model, Random Waypoint Mobility Model) and all simulation parameters with simulation environment has been shown in table1 separately.

Table 1: Simulation Environment

Network Parameter	**Value**
Simulator	NS-2.31
Simulation time	100 Seconds
Transmission range	550 m
Node movement model	Random way point
Routing Protocols	DSR and LAR-1
Simulation area	800 x 600m^2
Bandwidth	2 Mb/s
Traffic type	CBR
Transport Agents	TCP
Node Speed	0-20 m/s
Energy value	18-100 J

5.1 Simulation Results and Discussion

If the concept of proposed technique is applied separately on DSR protocol, is named as N-DSR approach. In this simulation, implementation has accomplished between N-LAR-1 and N-DSR approaches for 25, 50, 75 and 100 nodes. Comparison of performance metrics like PDF, throughput, Normalized Routing Load (NRL), E2E delay and average node energy consumption is displayed in figures 2, 3, 4, 5 and 6 respectively.

5.2 Results

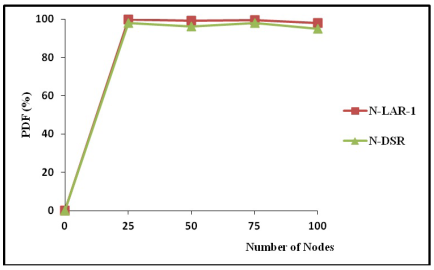


Figure 2: PDF Comparison

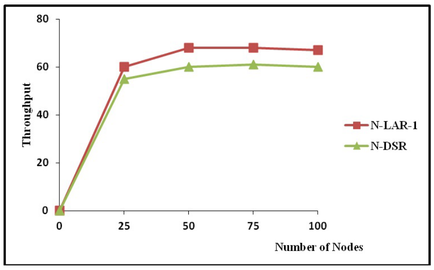


Figure 3:Throughput Comparison

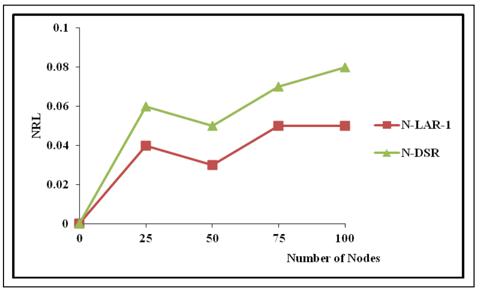


Figure 4:NRL Comparison

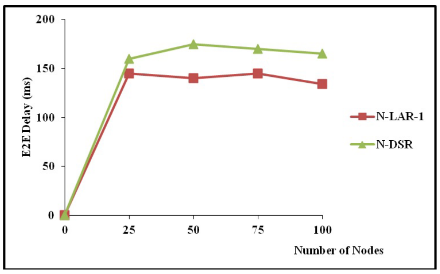


Figure 5:E2E Delay Comparison

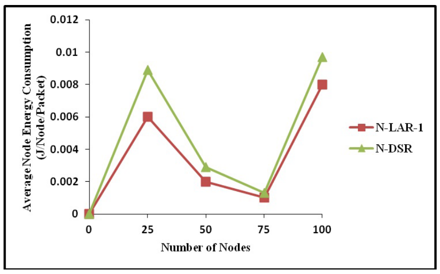


Figure 6:Average node energy consumption comparison

5.3 Analysis of Performance Metrics

1. **PDF:** PDF is consistent in N-LAR-1 and N-DSR techniques with increasing the number of nodes. It is clear from figure 2 that PDF ratio is higher using N-LAR-1 than N-DSR approach.
2. **Throughput:** N-LAR-1 technique provides better throughput than N-DSR scheme. This analysis presents consistent behaviour in both techniques on 50, 75 and 100 nodes.
3. **NRL:** NRL ratio increases in both techniques till 25 nodes. Further, it varies with increasing the number of nodes in the network. NRL ratio is reduced using N-LAR-1 in comparison to N-DSR.
4. **E2E Delay:** E2E Delay analysis illustrates linear variation in both approaches till 25 nodes. E2E delay is less in N-LAR-1 than N-DSR and can be optimized when the number of nodes is 75 in the network.

5. **Average Node Energy Consumption:** Figure 6 depicts the average node energy consumption analysis for N-LAR-1 and N-DSR techniques. This parameter varies with increasing the number of nodes in network. Average node energy consumption using N-LAR-1 is less than N-DSR and can be optimized on 75 nodes.

It is concluded from the analysis, N-LAR-1 technique achieves energy conservation along with sending successfully higher data packets to the destination. All above-mentioned performance metrics are optimized and overall network lifetime of network is improved using N-LAR-1 in comparison to N-DSR.

6 Conclusion and Future Work

LAR-1 extends the on-demand routing approach in MANETs and packets have flooded towards the destination only instead of entire Ad-Hoc network. This research work shows development of an effective energy-conserving technique N-LAR-1 for MANET. Proposed N-LAR-1 technique has arrived at a conclusion that performance metrics such as average node energy consumption, PDF, E2E Delay, NRL, throughput and Network lifetime have been improved when the sleep or inactive mode of the nodes in the networks are utilized. It has been analysed that performance of N-LAR-1 is better than N-DSR approach. These improvements in the performance of N-LAR-1 are possible due to the flooding of data in the direction of expected destination only. Research can be extended to integrate such type of work like proposed N-LAR-1 technique using other reactive routing protocols like AODV, OLSR, TORA and position-based routing protocols like GLS, DREAM in future. Further this proposed technique can be implemented and analyzed for wireless sensor networks, Vehicular Ad-Hoc networks and cognitive radio networks.

Bibliography

[1] E. Ahvar, and M. Fathy (2007), Performance Evaluation of Routing Protocols for High Density Ad-Hoc Networks based on Energy Consumption by GlomoSim Simulator, *Proceedings of World Academy of Science, Engineering and Technology*, 97-100.

[2] Mauve M. Widmer A. and Hartenstein H (2001), A Survey on Position-Based Routing in Mobile Ad-Hoc Networks, *IEEE Network*, 15 (6):30-39.

[3] Joshi N., Joshi D (2011), Energy Conservation in MANET Using Variable Range Location-Aided Routing Protocol, *International Journal of Wireless and Mobile Networks* (IJWMN), ISSN: 0975-3834, 3(5): 265-280.

[4] Qabajeh L., Kiah L and Qabajeh M. (2009), A Qualitative Comparison of Position-Based Routing Protocols for Ad-Hoc Networks, *International Journal of Computer Science and Network Security* (IJCSNS), ISSN:1738-7906, 9(2):131-140.

[5] Y. Ko and N. Vaidya (1998), Location-aided routing (LAR) in Mobile Ad-Hoc Networks, *Proceedings of the ACM/IEEE International Conference on Mobile Computing and Networking* (MOBICOM98):66-75.

[6] Y. Xu, J. Heidemann, and D. Estrin (2001), Geography-informed Energy Conservation for Ad-Hoc Routing, *Proceedings of 7th Annual International Conference on Mobile Computing and Networking*:70-84.

[7] Patel A., Joshi R. (2009), Energy Conservation for Wireless Mobile Ad-Hoc Networks using Hexagonal GAF Protocol, *Proceedings of the 12th International Conference on networking, VLSI and signal processing*: 29-34.

[8] Chen B., Jamieson K., Balkrishnan H. and Morris R (2002), Span: An Energy-Efficient Coordination Algorithm for Topology Maintenance in Ad-Hoc Wireless Networks, *ACM Wireless Networks Journal*, 8(5):481-494.

[9] Ramkrishnan S., T. Thyagarajan, P. K. Ram and R.Vinodh (2005), Design and Analysis of PADSR Protocol for Routing in MANET, *IEEE Indicon-2005 Conference*, Chennai, India: 193-197.

[10] Gupta N., Gupta R (2010), Coordinator Selection Technique for Energy Conservation in MANET, *CiiT International Journal of Wireless Communication*, CiiT Published by the Coimbatore Institute of Information Technology, ISSN:0974-9756, 2(3):75-78.

[11] Fall Kevin (2011), NS Manual (Formerly ns Notes and Documentation), VINT Project, www.isi.edu/nsnam/ns/doc.

INT J COMPUT COMMUN, ISSN 1841-9836
9(3):292-304, June, 2014.

Formal Specification and Verification of Mobile Agent Systems

L. Kahloul, M. Grira

Laid Kahloul*
LINFI Laboratory, Computer Science Department,
University of Biskra, Biskra, 07000, Algeria
*Corresponding author: kahloul2006@yahoo.fr

Messaouda Grira
Computer Science Department, University of Biskra
Biskra, 07000, Algeria

Abstract: Mobile agent systems offer efficiency and flexibility as a design paradigm. These two characteristics allow to these systems to be an adequate solution for many problems. These systems are used in many critical domains. This expansion, in use, obliges designers to insure the reliability and correctness of such systems. Formal methods can be used to verify the correctness of these systems. This paper presents a formal specification and verification of mobile agent systems using the High Order π-calculus. The verification exploits the two tools UPPAAL and SPIN.
Keywords: Mobile Agent, Formal Verification, π-calculus, Promela, SPIN, UPPAAL.

1 Introduction

Formal methods are methods with mathematical background. Unlike informal methods, descriptions using formal methods are more precise, understandable, and unambiguous. There are many formal languages and methods, among these languages we recall: TL (Temporal Logics) [2], Processes Algebras (CCS [1], π-calculus [4], HOπ-calculus [19]) and State Transition Systems [3, 4]. To be formal, a language expressing a specification must comprise three components: a syntax defining the rules for forming expressions, a semantics with rules for the interpretation of formed sentences and a proof theory governing rules for inferring useful information from the specification [3].

π-calculus is a mathematical model of processes whose interconnections change as they interact. The basic computational step is the transfer of the communication link between two processes; the recipient can then use this link for further interaction with other parties. This makes the calculus suitable for modelling systems where the accessible resources vary over time [1]. Higher-Order-π-calculus (HOπ) treats another kind of mobility, where processes (called agents) themselves move. The π-calculus and its extensions are processes-algebras that focus on mobility. In these calculi, processes communicate using channels. A process sends a channel's name in the monadic π-calculus, tuples of channels names in the polyadic π-calculus, and tuples of processes and channels names in Higher-Order-π-calculus (HOπ) [4].

In software engineering, formal specification is the expression in a formal language, and at an abstract level, a set of proprieties; that a system is designed to satisfy [3]. It is recognised that good specifications should be adequate, internally consistent, unambiguous, complete, minimal and satisfied by lower-level ones [2].

System Verification is a domain appeared after the crisis witness of enterprises specialised in the development of systems, where the verification is a viable solution to these problems. During this period, verification tools were designed to verify systems. Some of the most important verification tools are:

1. UPPAAL: is the acronym of the university of Uppsala (Sweden) and Aalborg (Denmark). It is an educational tool for formal specifications and verification of systems [5];

2. SPIN: is the acronym of Simple Promela Interpreter. It is a tool for verifying the logical consistency of concurrent systems, specifically of data communication protocol. The verified systems must be described in the Promela [6] modelling language;

3. LOTOS: is the acronym of (Language of Temporal Ordering Specifications). LOTOS has been applied to describe complex systems, formally. A number of tools have been developed for LOTOS, covering user needs in the areas of simulation, compilation, test generation, and formal verification [7].

The formal specification and verification of mobile agent systems contribute to the best formalisation of these systems. Specification can be used to analyse some known MAS (mobile agent system) properties, which are: safety, accessibility, boundedness, and liveness. The objective of our work is to specify and verify an example of a MAS. This system manages service locations in particular networks. This system is called Service Location Protocol (SLP) [26].

To present this work, we organize this paper as follows: In the next section, we present briefly mobile agent systems, the section three will present HOπ-calculus, Promela and UPPAAL languages. Section four will present the modelling and verification using SPIN and UPPAAL tools. Before concluding this paper, we will show some related works in section five. Finally, a conclusion will summarize this paper and will discuss possible prospects.

2 Mobile Agent Systems (MAS)

2.1 Agent Concept

Agent's term comes from two distinct domains: *the distributed systems domain* and the *multi-agent systems domain*. These last belongs in the base to intelligence artificial domain, where the programmers tent to imitate humane intelligence [8]. Russel and Norvig define the agent term as entity that broach its environment and interact on it [9].

2.2 Mobile Agents

Mobile agents have been introduced initially in 1994 with the Telescript environment [16] that permitted to processes to choose themselves to move on the sites of a network in order to work locally onto resources. A mobile agent [12, 13, 15] is a process that can move from a site to another site in order to achieve a task. In general, the mobility is provided using some primitive like: **move(site)** that permits the agent to move toward the site designated by the parameter. A mobile agent is composed of his corresponding code, as well as of a context including some data. This context can evolve under execution, for example while collecting some data when an agent achieves a research of information on a set of servers. The code and the agent's context are displaced with the agent when this one visits different servers. In general, a mobile agent system provides the primitive of communication allowing to the agents to interact between them, but also to the agents to interact with the services that they visit. These primitives of communications take the form of sending messages or calling procedures or methods [17].

2.3 Mobility mechanisms

Mobile agents can move between network hosts, transporting their code, data and state information to continue their execution on different environments. In literature [5], we can find many mechanisms. In the case of remote execution, the agent is sent before it starts to be executed. When it arrives on the destination, it is executed until it finishes. In this case the agent is transferred once. When it is executing it can use the same remote execution mechanism to start the execution of other agents. In the remote execution the destination of the agent is determined by the execution starter. The mobile agent can do a weak migration by sending its data with its code. Usually, the implementations of this scheme allow choosing which part of data will be transferred to the new location of the agent. In this case, the agent programmer might design some mechanism based on the value of agent's data to resume the execution from some point. Strong migration is the highest degree of mobility. Using this scheme not only agent code and data is sent, but also the state of execution. When the agent arrives to the destination, it is fully restored and its execution is resumed from the same execution point it was just before migration. Strong migration turns to be complex, since it involves low level internal mechanisms for execution restoring that must be standard to provide migration transparency in heterogeneous environments.

2.4 Applications and limits of mobile agents

Mobile agents [1] are software abstractions that can migrate across the network representing users in various tasks. This is a contentious topic [14, 18] that attracts some researchers. Mobile agents provide a very appealing, intuitive, and apparently simple abstraction. The authors in [1] give some applications of mobile agents:

1. Distributed research of information;
2. Active Documents: Active e-mail, web page (hypertext);
3. Advanced telecommunications services: video-conferencing, mobile users (with the potential disconnections);
4. Monitoring and remote configuration of devices: industrial processes, network management;
5. Management and cooperation in the work-flow: the work-flow defines activities, sites, relationships, time for their implementation, to achieve an industrial product. Mobile agents are responsible for conveying information between co-workers in a work-flow;
6. Active networks: flexible and dynamic networks according to application needs. Two approaches are proposed: (i) Programmable switches: dynamically extend the networks. This approach is based on Code on Demand paradigm; and (ii) Capsule approach: attach codes to the transferred packets. The node that receives the packet performs the associated code to treat the data in the packet;
7. E-commerce: an agent looks in a market for catalogues, and then it returns to the laptop of a customer with the best rates available;
8. Applications deployment and maintenance of components in distributed environments.
9. Parallel processing: the agents dispatch several computing units to make parallel certain tasks.

3 Formal Languages for MAS

We are interested to present two formal languages in this section HOπ-calculus, Promela language and two tools which are SPIN and UPPAAL.

3.1 HOπ-calculus

In this higher order paradigm, mobility is achieved by allowing agent to be passed as values in a communication. HOπ-calculus [19] is an extension of the first order π-calculus introduced by D.Sangiorgi [4]. This calculus enriches the π-calculus with an explicit higher order communication. In the HOπ-calculus, not only names, but also agents of arbitrarily high order, can be transmitted. The syntax of the HOπ-calculus is an extension of the syntax of the first order π-calculus.

Let be $\{a, b, ..., x, y, ...\}$ a set of names and $\{P, Q, ...\}$ a set of processes [19]. Processes of HOπ-calculus are defined by the following grammar:

$P ::= \tilde{x}(K).P|x(U).P|P||Q|\tau.P|\upsilon xP|P + Q|[x == y]P|!P|0$

Where:

1. The output prefix $\tilde{x}(K).P$ can send the name K via the name x and continue as P;
2. The input prefix $x(U).P$ can receive any name via x and continue as P with the received name substituted for all free occurrences of U in P;
3. The Parallel Composition $P||Q$ represents the combined behaviour of P and Q executed in parallel, where P and Q can proceed independently and interact via shared names;
4. The Silent Prefix $\tau.P$ represents an agent that can evolve to P without interaction with the environment;
5. The Restriction υxP behaves as P but the scope of the name x is restricted to P. P can not interact with other process through x;
6. The Sum $P + Q$ represents an agent that can enact either P or Q;
7. The Match $[x == y]P$ can evolve as P if x and y are the same name, and will do nothing otherwise;
8. The Replication $!P$ can be seen as an infinite composition $!P = P|P|$ or, equivalently, an agent satisfying the equation $!P = P|!P =!P|P$;
9. The empty agent 0 cannot perform any action.

3.2 Promela

Promela [6] is the acronym of Process (or Protocol) Meta Language; it is the modelling language for the SPIN (Simple Promela Interpreter) [27]. Promela programs consist of processes, message channels, and variables. Processes are global objects that represent the concurrent entities of the distributed system. Message channels and variables can be declared either globally or locally in a process. It supports rendezvous and asynchronous communication between processes via channels. Processes specify behaviour, while channels and global variables define the environment in which the processes run.

3.3 UPPAAL

UPPAAL [6] is a tool to model, to validate and to verify real time systems; it is appropriated for the systems that can be modelled by timed automata or linear hybrid automata. The model-checker UPPAAL is based on the theory of timed automata [29,30] and its modelling language offers additional features such as bounded integer variables and urgency (priority between the same actions defined in two automata). The query language of UPPAAL, used to specify properties to be checked, is a subset of CTL (computation tree logic) [31].

4 Specification and Verification of the SLP Protocol

4.1 The SLP Protocol

The Service Location Protocol is an IETF standard track protocol [26]. The IETF is the Internet Engineering Task Force committee that is part of the IAB (Internet Activities Board) and determines internet standards. SLP provides a framework to allow networking applications to discover the existence, location, and configuration of networked services in enterprise networks. SLP can eliminate the need for the user to know the technical features of network hosts. Using the SLP, the user needs only to know the description of the service he is interested in. Based on this description, SLP is then able to return the URL of the desired service. SLP is a language independent protocol. Thus the protocol specification can be implemented in any language. The SLP infrastructure (Figure 1) consists of four types of agents [26]:

1. User Agent (UA), which is a software entity that is looking for the location of one or more services;

2. Service Agent (SA), which is a software entity that provides the location of one or more services;

3. Directory Agent (DA), which is a software entity that acts as a centralised repository for service location information;

4. Memory Directory Agent (MDA), which is a software entity that memorizes the service information.

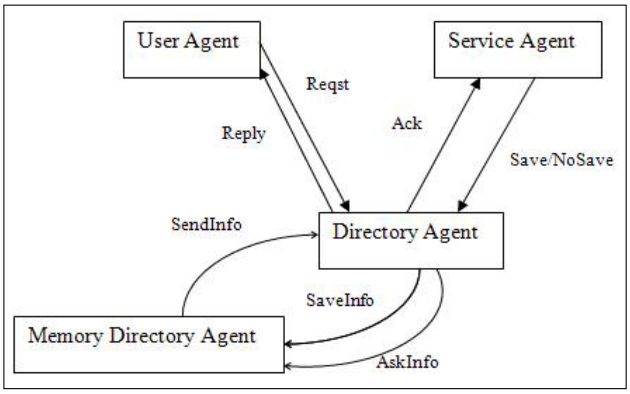


Figure 1: The SLP infrastructure

In next section, we present a formal specification of SLP in HOπ-calculus. This specification will be transformed firstly into the Promela language, then into the UPPAAL modelling language. The section six will present the verification phase. We achieve the verification using the two tools: SPIN and UPPAAL, then we discuss a comparison between the results obtained from the two tools.

5 SLP Specification

In this section, we will specify the SLP protocol by three formal specification languages: the HOπ-calculus, Promela, and UPPAAL modeling language. The formal specification of the SLP system using the HOπ-calculus is given as the following:

$$\begin{array}{l} system = (Reqst,\ Reply,\ Save,\ NoSave,\ Connect,\ Ack,\ Ask_{info},\ Send_{info})| \\ UA(Reqst,\ Reply)|SA(Connect,\ Save,\ NoSave,\ Ack)| \\ DA(Reqst,\ Ack,\ Connect,\ Save,\ NoSave,\ Ask_{info},\ Send_{info})| \\ MDA(Ask_{info}, Send_{info}) \end{array}$$

In this specification, the system is composed of one User Agent (UA), one Directory Agent (DA), one Memory Directory Agent (MDA), and one Services Agent (SA). The mobility is described through the parameters of the channel. For each new service registered by the SA within the DA, there are MDA belonging to this specific service. In our case the service will be **service**. The parameters for each agent are used for the communication between agents.
In the UA specification, we use **Request** and **Reply** as two channels for the communication between the UA and DA. The term **service** specifies the service that will be asked. In this specification, the UA sends a request to ask a service, and receives a reply that contains the service information. The User Agent (UA) Specification is as follows:

$$\begin{array}{l} UA(Reqst,\ Reply) = \upsilon(service) \\ \overline{Reqst}(service, Reply).Reply(service, info) \end{array}$$

The SA can register a service within the DA using the channel *Save* and also remove a registered service within the DA using the channel *NoSave*, and it receives an *Ack* as an ok to *Save* or *NoSave*. The Service Agent (SA) Specification is as follows:

$$\begin{array}{l} SA(Connect, Save, NoSave, Ack) = \upsilon(service) \\ \overline{Connect}(adrs).\overline{Save}(service).Ack|\ SA(Connect,\ Save,\ NoSave,\ Ack) \\ +\overline{Connect}(adrs).\overline{NoSave}(service).Ack|\ SA(Connect,\ Save,\ NoSave,\ Ack) \end{array}$$

The DA can communicate with UA and SA, it generates the service location. This DA saves the service information if it receives a $Save(service)$, and does not save it otherwise. If the DA receives a $Reqst(service)$, it asks the Memory Directory Agent (MDA) for the service information via $Ask_{info}(service)$, and it waits the reply $(Send_{info}(service))$ from this agent, before sending this service information to the UA via $\tilde{Reply}(info, service)$. The specification of the DA is as follows:

$$\begin{array}{l} DA(Reqst,\ Ack,\ Connect,\ Save,\ NoSave,\ Save_{service},\ NoSave_{service},\ Ask_{info}, \\ Send_{info}) = Connect(adrs)|Save(Service).\overline{Ack}.DA_{Mem}(Ask_{info}, send_{info})| \\ \overline{Save}_{service}(service)|\ Reqst(service).\overline{Ask_{info}}(service)| \\ Send_{info}(service, info).\overline{Reply}(service, info)| \\ DA(Reqst, Ack, Connect, Save, NoSave) \\ +Connect(adrs)|NoSave(service).\overline{Ack}| \\ DA_{Mem}(Ask_{info}, Send_{info}).\overline{NoSave}_{service}(service)| \\ DA(Reqst, Ack, Connect, Save, NoSave) \end{array}$$

The MDA is the Memory Directory Agent; it saves information about services. It communicates only with the DA. The specification of the MDA is as follows:

$$\begin{array}{l} MDA(Ask_{info}, Send_{info}, Save_{service}, NoSaveservice) = \\ Save_{service}(service)|Ask_{info}(service).\overline{send}_{info}(service)|MDA(Ask_{info}, Send_{info}) \\ +NoSave_{service}(service).MDA(Ask_{info}, Send_{info}) \end{array}$$

Formal methods may be used to specify the system behaviour and to verify that the designed and implemented system satisfies the expected properties. The higher-order π-calculus is able to describe the mobile systems behaviour and, because it possesses formal semantics, it is capable of verification as well. To insure verification of the HOπ-calculus, one is obliged to use some automated tools. We have found two tools that can be used: SPIN and UPPAAL. To realise verification using this two tools, we were obliged to apply some transformation on the former specification. The two next sections will present the specification in SPIN and in UPPAAL.

5.1 Modelling SLP Using Promela

To allow the verification of the SLP protocol, we propose (firstly) to use the SPIN tool. The SPIN tool uses Promela as a specification language. Requirements and properties are specified as LTL (Linear Temporal Logic) [2]. These LTL formulae can be entered as assertions in the Promela specification, and so be used in simulation, or they can be verified using the model-checker of SPIN. The following paragraphs present the Promela specification of the SLP protocol.

The messages declarations are formalised as follows:

$\#define\, msgtype\, 10$
$mtype = service, name, ok, info$

The channels between the entities (agents) are formalised as follows:

$chan\, Reqst = [1]\, of\, mtype, byte$
$chan\, Reply = [1]\, of\, mtype$
$chan\, Save = [1]\, of\, mtype$
$chan\, NoSave = [1]\, of\, mtype$
$chan\, Ack = [1]\, of\, mtype$
$chan\, Connect = [1]\, of\, byte$
$chan\, Ask_info = [1]\, of\, mtype$
$chan\, Send_info = [1]\, of\, mtype$
$chan\, Save_service = [1]\, of\, mtype$
$chan\, wait = [1]\, of\, byte$
$bool free = true,\ souscribe = false;$
$byteadrs;$

We declare four processes. The first one is a process named *User Agent*, specified as follows:

$active\, proctype\, UA()\{$
$xsReqst;\ xrReply;$
$again1:$
$Reqst!service,\ Reply;$
$if\ \ ::\ Reply?service \rightarrow free = false;\ wait!msgtype(10)$
$fi;$
$goto\, again1\}$

The second one is a process named *Service Agent*, specified as follows:

```
active proctype SA(){
xs Save, NoSave; xr Ack;
again2;
Connect!adrs;
if  :: wait?(10) → free = false
fi
if
:: free == true → Save!service
:: else → NoSave!service
fi;
Ack?ok;
goto again2}
```

The third one is a process named *Directory Agent*, specified as follows:

```
active proctype DA(){
xs Ack, Ask_info, Reply;
xr Save, NoSave, Send_info, Reqst;
if  :: Connect?adrs → subscribe = true
fi
assert((free == true) & &(subscribe == true));
if
:: Save?service → Ack!ok; Save_service!service
:: NoSave?service → Ack!ok
:: Reqst?service → Ask_info!service
fi;
if Send_info?service → Reply!service, info
fi}
```

The four one is a process named *Memory Directory Agents* and it is specified as follows:

```
active proctype MDA(){
xs Send_info; xr Ask_info, Save_service;
if
:: Save_service?service;
:: Ask_info?service → Send_info!service
fi}
```

5.2 Modelling SLP by UPPAAL

Figure 2 presents the graphic modelling of the whole system specification as a timed automaton. The agents SA, DA, and MDA are presented respectively in the Figures: 3, 4, and 5.

6 SLP Verification

This section is dedicated to present the verification phase in this work. We consider three important properties (i.e. requirements), that we will verify:

- **Safety**: Every agent SA which registers a service within DA is recognised by the DA and the service becomes available;

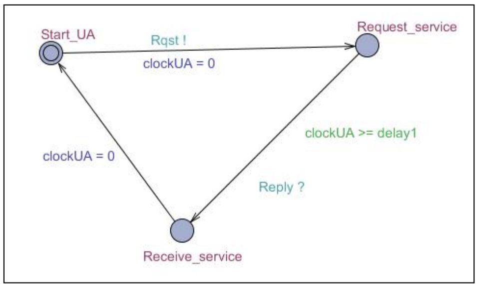


Figure 2: Diagram of the system

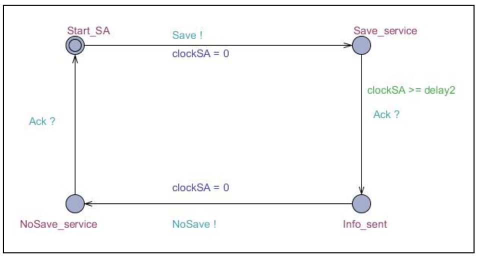


Figure 3: Service agent diagram

- **Reachability**: If UA asks for the service "service" and if SA subscribes this service at the DA, then UA obtains the service;

- **Bounded Liveness**: During the registration period within the DA, a service must be accessible.

6.1 SLP Verification by the SPIN Tool

We use the SPIN tool to verify the Safety, Reachability and Liveness properties. This tool allows verifying properties during the simulation using the assert function (assertion): $assert((free == true)\,\&\,\&(subscribe == true))$, or using the following never claims:

$never\{$
$TO\_init:$
if
$::\ (!(ask)\ \&\ \&(frr)\ \&\ \&(subscribe)) \rightarrow goto\, accept\_all$
$::\ (!ask) \rightarrow goto\, TO\_S3$
$fi;$
TO_S3
if
$((free)\ \&\ \&(subscribe) \rightarrow goto\, accept\_all$
$(1) \rightarrow goto\, TO\_S3$
$fi;$
$accept\_all$
$skip\}$

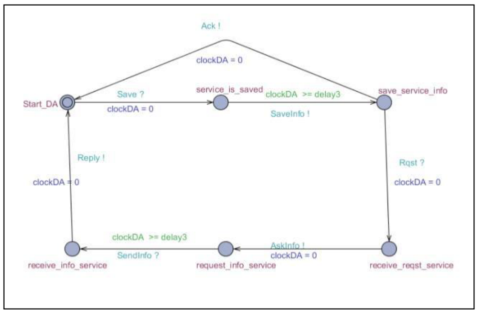


Figure 4: Directory agent diagram

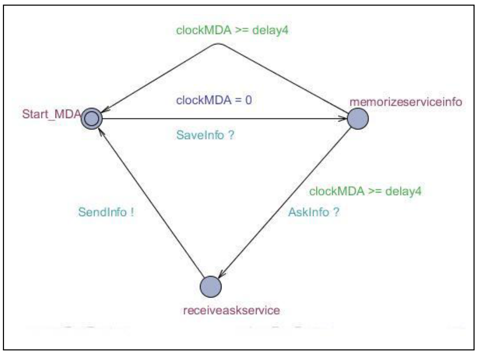


Figure 5: Memory directory agent diagram

6.2 SLP Verification by the UPPAAL Tool

We use the SPIN tool to verify the Safety and Liveness properties using the CTL (Computational Tree Logic) [31] as following:

$E{<>}Directory.receive\_reqst\_service$: This expression allows us to verify the reachability of some sate in the directory agent behaviour.

$A[]notdeadlock$: This expression allows us to verify that the system has no deadlock.

7 Related Works

Mobile agent paradigm attracts researchers in the domain of software engineering and artificial intelligence. Many interesting researches on formal specification and verification of mobile code system can be found. In the present work, we have presented firstly a specification of the service location protocol system in the HOΠ-calculus. This specification was then verified using the two tools: SPIN and UPPAAL. This verification validates the safety propriety. In the literature, we found some similar works that tended to specify and verify formally mobile code systems using other specification languages and other verification tools, as examples:

- In [1], the SLP is specified in HOPi-calculus and verified by UPPAAL tool. In this work, the HOPi-calculus specification is then translated to an automaton model, where the mobile aspect of data is not respected. In our work, we have used a specification with only four agents. Another difference is that we have used the SPIN tool and Promela.

- In [3], a formal specification and verification of secure routing protocols for mobile ad hoc networks (MANETs) is presented. In this work, authors used the Promela as specification languages and SPIN as verification tool.

- Formal specification of standards for distance victor routing protocol is presented in [10]. This project presents how to use an interactive, HOL (High Order Logic), together with a model checker, SPIN, to prove key proprieties of distance victor routing protocols. The formal verification techniques in this project are suited to routing protocol generally.

8 Conclusion

A mobile agent is an entity that can change its location during its execution. It can migrate from one host to another, in a network. The use of mobile agents knows an expansion in several domains. This inclusion of mobile agents in critical domains obliges the designer and developer to make attention during the elaboration of these agents. The reliability and the correctness of these programs (mobile agents) are important. Formal methods can help the designer to specify systems then to prove the correctness of these systems. Some of these methods are: π-calculus, HOπ-calculus.

In this paper, we have presented the use of the HOπ-calculus and Promela language as specification languages for mobile agent systems. We have presented a specification of the SLP protocol (Services Location Protocol), and we have done verification using the two tools: UPPAAL and SPIN.

The use of formal methods meets some difficulties in software development. These difficulties are argued by the lack in experiments of these methods and the requirement of some mathematical background by the developer. In this work and some other previous works [32–34], we have tried to show the power of formal methods to insure the reliability of a mobile agent systems. In our future work, we will be interested to apply formal methods on more examples of mobile agent systems. These experiments will establish an approach on the use of formal methods, in software engineering process.

Bibliography

[1] Milner, R. (1989). *Communication and Concurrency*, Prentice Hall, International Series in Computer Science, ISBN 0-13-115007-3.

[2] Rescher,N., Garson, J. (1968). Topological Logic, *Journal of Symbolic Logic*, 33(4):537-548.

[3] Wagner, F. (2006). *Modeling Software with Finite State Machines: A Practical Approach*, Auerbach Publications, ISBN 0-8493-8086-3.

[4] Murata, T. (1989). Petri Nets: Properties, Analysis and Applications, *Proceedings of the IEEE*, 77(4): 541-580.

[5] Fuggetta, A., Picco, G.P., and Vigna, G. (1998). Understanding Code Mobility, *IEEE transactions on software engineering*, 24(5): 342-361.

[6] Bengtsson, J., Larsen, k.G., Larsson,F.,Pettersson, P., Yi W. (1995). Uppaal - a Tool Suite for Automatic Verification of Real-Time Systems, in *Proceedings of the 4th DIMACS Workshop on Verification and Control of Hybrid Systems*, New Brunswick, New Jersey, 22-24 October, 1995.

[7] Barbu, A. (2005). *Developing Mobiles Agents Through a Formal Approach*, Thesis, Paris XII, 12 September 2005.

[8] Lamsweerd, A. V. (2000). Formal Specification: a Roadmap, *ICSE '00 Proceedings of the Conference on The Future of Software Engineering*, 147-159.

[9] Gurdag, A.B., Caglayan, M.U. (2008). A Formal Security Analysis of SAODV using Model Checking, *International Symposium on Computer Networks (ISCN)*, June 2008.

[10] Sangiorgi, D., Walker, D.(2003). *The Pi-calculus: A Theory of Mobile Process*, Cambridge University Press.

[11] Faez CHARFI. (2003). *Une approche dinterfaage de CoD UPPAAL pour la spcification et la vrification des sysmes temps el*, Thesis, September 2003.

[12] Ruys, T. (2006). *SPIN and Promela*, January 18, 2006.

[13] Sighireanu, M., *LOTOS NT User Manual*, February 21, 2008.

[14] Gray,R.S., Cybenko,G., Kotz, D., Peterson, R.A., Rus, D. (2002) *D'agents: Application and performance of a mobile-agent system*, 2002.

[15] Akhil Sahai, A., Morin, C. (1998). Mobile agents for enabling mobile user aware application, *Proceedings of the 2nd International Conference on Autonomous agents*, 205-211.

[16] Bhargavan,K., Obradovic, D., Gunter, C.A. (2002). Formal Verification of Standards for Distance Vector, *Journal of the ACM*, Vol. 49(4):538-576.

[17] Picco, G.P. (1998). *Understanding, Evaluating, Formalizing, and Exploiting Code Mobility*, Ph.D. Thesis, Politecnico di Torino, Italy, February 1998.

[18] Satoh, I. (2002). Physical mobility and Logical mobility in ubiquitous computing environments, *Proceeding MA '02 Proceedings of the 6th International Conference on Mobile Agents*, 186-202.

[19] Johansen, D. (2004). Mobile agents: Right concept, wrong approach, *Proc. IEEE, Mobile Data Management-MDM*, 300-301.

[20] Robles S. (2002). *Mobile Agent Systems and Trust, a Combined View Toward Secure Sea-of-Data Applications*, PhD Thesis, Barcelona, July 2002.

[21] Kotz, D., Gray, R. S. (1999). Mobile agents and the Future of internet, *ACM Operating Systems Review*, 7-13.

[22] Roth, V. (2004). Obstacles to adoption of mobile agents, *IEEE International Conference on Mobile Data Management*, 296-297.

[23] Milojicic, D. (1999). Trend Wars: Mobile agent applications. *IEEE Concurrency*, 80-90.

[24] Sangiorgi, D. (1993). From π-calculus to Higher-Order-π-calculus- and back. *TAPSOFT '93 Proceedings of the International Joint Conference CAAP/FASE on Theory and Practice of Software Development.*

[25] Milner, R., Parrow, J., Walker, D. (1992). *A calculus of mobile processes*, part I/II. 1992.

[26] Perkins, C., Guttman, E., *Service Location Protocol (SLP)*, online: http://www.ietf.org/rfc/rfc2608.txt.

[27] Kaliappan, P.S., *Simple Promela Interpretor (SPIN)- Model Checker.*

[28] Guillaume, S. (2008). *A Promela front-end for Spot*, 20 p., May 2008.

[29] Alur R., Dill, D.L. (1990). Automata for modeling real-time systems, *Colloquium on Algorithms, Languages, and Programming*, 443:322-335.

[30] Hopcroft, J.E., Ullman, J.D. (2001). *Introduction of Automata Theory, Languages*, Addison Wesley.

[31] Huth, M., Ryan, M. (2004). *Logic in Computer Science*, (Second Edition), Cambridge University Press. p. 207. ISBN 0-521-54310-X. 2004.

[32] Kahloul, L., Chaoui, A. (2008). Coloured Reconfigurable Nets for Code Mobility Modeling, *Int J Comput Commun*, ISSN 1841-9836, Suppl. issue, 3(S):358-363.

[33] Kahloul, L., Chaoui, A., Djouani, K. (2009). *Code Mobility Modelling: A Formal Study*, *International Review on Computer and Software*, http://www.praiseworthyprize.com/IRECOS.htm.

[34] Kahloul, L., Chaoui, A., Djouani, K. (2010). Modeling Reconfirgurable Systems Using Flexible Petri Nets, *4th IEEE International Symposium on Theoretical Aspects of Software Engineering*, August 24 - 27, 2010, Taipei, Taiwan.

INT J COMPUT COMMUN, ISSN 1841-9836
9(3):305-312, June, 2014.

Development of MCDM Methods – in Honour of Professor Edmundas Kazimieras Zavadskas on the Occasion of His 70th Birthday

O. Kaplinski, F. Peldschus, L. Tupėnaitė

Oleg Kaplinski
Poznań University of Technology
Piotrowo 5, 60-965 Poznań, Poland
oleg.kaplinski@put.poznan.pl

Friedel Peldschus
Leipzig University of Applied Science
132 Karl Liebknecht St, 04227 Leipzig, Germany
friedel@peldschus.net

Laura Tupėnaitė*
Vilnius Gediminas Technical University
Sauletekio Ave. 11, Vilnius, LT-10223, Lithuania
E-mail: laura.tupenaite@vgtu.lt
*Corresponding author

Abstract: Multiple Criteria Decision Making (MCDM) substantially evolved during the past decades and became one of the most important areas in Operational Research/Management Science. The article presents a review of extensive scientific work of Professor Edmundas Kazimieras Zavadskas on development of MCDM methods on the occasion of his 70th birthday. The article also highlights his research carrier, and lists some of his publications.
Keywords: decision support, MCDM, methods, development, Edmundas Kazimieras Zavadskas.

1 Introduction

Multiple Criteria Decision Making (MCDM) methods have substantially evolved since 1970s, and had various types of real world applications. New MCDM methods have been developed, and existing methods improved, showing that research in the decision-making is still critical and valuable. One of the early and exceptional authors, continuously working on the development and improvement of MCDM methods since 1976, is Professor Edmundas Kazimieras Zavadskas (Fig. 1). This article is an attempt to summarise his research and achievements in development of the MCDM methods on the occasion of Professor˙s 70th birthday.

2 Achievements in Development of MCDM Methods

Edmundas Kazimieras Zavadskas was born on the 12th of May 1944 in Vilnius. He presented his PhD in 1973 – he researched the applications of polymer resins in reinforced concrete. This was the time when he took interest in optimising constructions, technologies and organisations. A selection of decision-making solutions dominated his research. This is how he significantly developed some elements of rational decisions theory. As a synthesis of research results, in 1987 E. K. Zavadskas defended his Post-Doctoral (Habilitation) Thesis where MCDA/MADM methods (TOPSIS, SAW, ELECTRE, ENTROPY, Game Theory, Utility Theory, Permutation

Method, Judgement Methods) were applied for construction tasks solutions. Furthermore, these methods were used for development of decision support systems [1].

There is a monograph summarising his achievements from that period [2]. This monograph has had a strong influence on research conducted by young academics working towards their PhD theses in number of countries, from Uzbekistan through Russia, Poland, Germany, Denmark, and Cuba to Syria. Another monograph [3] strengthened the Professor˙s position as a leader in this part of Europe, conquering the area of multi criteria decision aiding methods, and operational research applications in construction industry.

Synthetic works on multiple criteria decision support systems in construction have been published in individual monographs or in collaboration with his colleagues [4] – [6]. New methods of performing multiple criteria analysis in a project have been developed by Professor and his team, including:

- A method of COmplex PRoportional ASsessment (COPRAS) [5], COmplex PRoportional ASsessment of Alternatives with Grey Relations (COPRAS-G) – presented in publication [7] which was nominatd the New Hot Paper for January 2010 by Science Watch (Thomson Reuters) in the field of Engineering (see http://archive.sciencewatch.com/dr/nhp/2010/10jannhp/10jannhpZavaET/) and COmplex PRoportional ASsessment of Alternatives Applying Fuzzy Sets (COPRAS-F) [8];
- Additive Ratio Assessment (ARAS) method [9], Additive Ratio Assessment Applying Attributes Values Determined in Intervals (ARAS-G) method [10] and Applying Fuzzy Sets (ARAS-F) [11];
- A Selection of Rational Dispute Resolution Method by Applying New Step-Wise Weight Assessment Ratio Analysis (SWARA) [12];
- TOPSIS Method Applying Mahalanobis Distance Measure (TOPSIS-M) [13];
- A new Normalization method in Games Theory [14];
- A method of Weighted Aggregated Sum Product Assessment [15];
- Algorithm of Maximising the Set of Common Solutions for Several MCDM problems [16].

Furthermore, variety of the new MCDM methods and software was developed in collaboration with academic colleagues from abroad, i.e.:

- Software for Multiple Criteria Evaluation [17];
- A method of Multi-Objective Optimisation on the Basis of Ration Analysis (MOORA) [18];
- A method MULTIMOORA (MOORA plus Full Multiplicative Form) [19];
- COPRAS method for Group Decision Making in an Interval-Values Intuitionistic Fuzzy Environment [20];
- Extensions of LINAMP Model for Multi Criteria Decision Making with Grey Numbers [21];
- Fuzzy DEA Approach Based on Parametric Programming [22];
- Intuitionistic Fuzzy DEA for Efficiency Evaluation under Uncertainty [23];
- Stepwise DEA Analysis and Grey Incidence Analysis [24].

All the above listed methods had wide real world applications in such areas as: sustainable development in civil engineering, building life cycle, modelling of construction and real estate sector, quality control of construction projects, etc. Professor E. K. Zavadskas continuously develops new and researches existing MCDM methods for further improvements. Researches results, among many others, as illustration of Professor˙s works can be distinguished:

- Measuring Congruence of Ranking Results Applying Particular MCDM methods [25];
- Evaluation of Ranking Accuracy in Multi-Criteria Decisions, presented in paper [26] which was titled as Fast Breaking Paper for June 2009 by Science Watch (Thomson Reuters) in

the field of Mathematics (see
http://archive.sciencewatch.com/dr/fbp/2009/pdf/09junfbpVysh.pdf);
- Verification of Robustness of Methods when Assessing Alternative Solutions [27] – [31];
- Proposal of Multi-Criteria Assessment Model of Technologies [32] – [39] .

Professor E. K. Zavadskas, in collaboration with his colleagues, has also developed hybrid decision making methods by combining MCDM methods TOPSIS, SAW, ELECTRE, AHP and the methods proposed by the Professor (see, i.e. [40] – [45]). About 20 papers were nominated as Hot Papers in Thomson Reuters database. In one of these papers there is a published interview with the author (see http://archive.sciencewatch.com/dr/nhp/2011/11maynhp/11maynhpZavaET/).

Professor E. K. Zavadskas was granted a Lithuanian award for research achievements for a cycle of works ´Multiple Criteria Assessment of Construction Projects and Technological SolutionsĄ (1980–1996) in 1996, and for a cycle of works ´Modelling in construction (methods, simulation, decision support and information systems, web-based technologies, practical application)Ą (1996–2003) – in 2004. In 1996 he also was awarded the 4th class medal of the Lithuanian Grand Duke Gediminas.

3 Biography and General Data

Edmundas Kazimieras Zavadskas (Fig. 1) was born in 1944 Vilnius, Lithuania. He graduated from an elementary school in Vorkuta, USSR and secondary school in 1962, Dūkštas, Lithuania.

E. K. Zavadskas studied at Faculty of Construction Economics in Vilnius branch of Kaunas Polytechnic Institute (VISI) (1962–1967) (now – Vilnius Gediminas Technical University, VGTU). In 1968 he became an assistant, 1969–1972 he was a PhD student at VISI, lecturer – in 1974, associate professor in 1977 and became a professor in 1988.

The Professor became a Rector of Vilnius Civil Engineering Institute (VISI) in 1990. During the period between February-October 1990, he successfully reorganized the Institute, which became the Vilnius Technical University, and was nominated a Rector for period of 1990–1996. Later the University was renamed as Vilnius Gediminas Technical University (VGTU) and Prof. E. K. Zavadskas became a Rector for the period of 1996–2002. In 2002–2011, he was a Vice-Rector of VGTU. In this time, he worked towards making the University one of the largest universities in Lithuania, taking a leading position in technical and engineering education and research.

Since 1986 till now Professor Zavadskas has been a Head of the Department of Construction Technology and Management, Civil Engineering Faculty, VGTU.

His impressive academic carrier started in 1973 when he became a PhD student at the Vilnius Civil Engineering Institute (VISI, now VGTU) and defended his Post-Doctoral (Habilitation) Thesis at Moscow Civil Engineering Institute in 1987 (on Multi Attribute Decision Making in Construction). In 1993, he won laurea doctorali ad habilitationem at Technical Sciences at VGTU.

Professor E. K. Zavadskas was an Expert Member (1991–1993), Corresponding Member (1993–2011) and in 2011, became a Full Member of the Lithuanian Academy of Sciences. Furthermore, he has been granted a title of Honorary Doctor of three universities: Poznań, Kiev, and St. Petersburg (2001–2003) and the Honorary International Professor of the National Taipei University of Technology. He presides and participates in a number of scientific corporations and editorial boards of scientific publishing houses. He is a member of two Russian Academies, the Ukrainian Academy of Cybernetics, and many (17) scientific and research organisations, from Melbourne to Brussels. He also represents the Baltic States in international organisations. The Professor was a President of Operational Researchers Society in Lithuania and the Baltic States

Figure 1: R. J. Slowiński, B. Roy, E. K. Zavadskas at the 52nd Meeting of the EURO Working Group Multicriteria Aid for Decisions (MCDA), 6–7 October, 2000, Vilnius, Lithuania (organised by H. Pranevičius, L. Sakalauskas, E. K. Zavadskas, A. Kaklauskas)

in 2001–2012.

The Professor˙s research and teaching includes fields of construction materials, materials resistance, construction technology and management, operational research methods, decision support systems, life cycle analysis, etc.

Prof. E. K. Zavadskas has published over 50 books, including 5 textbooks and 16 monographs as single author, or in collaboration with other authors, 10 popular science books, over 400 research articles as well as several hundreds of articles on various social and cultural topics. He has edited over 20 collective volumes.

Professor E. K. Zavadskas has set up three famous international scientific journals: ´Technological and Economic Development of EconomyĄ (Editor-in-chief since 1994), ´Journal of Civil Engineering and ManagementĄ (Editor-in-chief since 1995) and ´International Journal of Strategic Property ManagementĄ (Editor-in-chief since 1997 till 2011). Since 2008 all three journals have been referred in Thomson Reuters Web of Science database, and since 2010 – have impact factor (IF). Furthermore, since 2010 these journals are published by VGTU publishing house ´TechnikaĄ in collaboration with a famous publishing house – Taylor & Francis.

The Professor is also a member of editorial boards of 16 international journals referred in Thomson Reuters Web of Science database and 17 other journals.

On various occasions special issues of journals and collective volumes were dedicated to Professor˙s works, i.e. Journal of Management and Decision Making: ´Normalisation in Decision Making methodsĄ (2007), International Journal of Environment and Pollution (2007, 2008), Automation in Construction (2010), Informatica (2001), Ecology (2007).

Professor E. K. Zavadskas was a chairman and member of organizing committees of numerous international conferences, as well as editor of conference proceedings, including:

- Modelling and Simulation of Business Systems (Vilnius, Lithuania 2003);
- 33rd Symposium International FESF Strasbourg: Recent developments in Environmental Protection (Vilnius, Lithuania, 2003);
- Simulation and optimization in Business and Industry: International Conference on Oper-

ational Research (Tallinn, Estonia, 2006);
- The 20th International Conference EURO mini Conference ĽConstruction Optimization and Knowledge-Based Technologies´ (EurOPT'2008) (Neringa, Lithuania, 2008);
- The 25th International Symposium on Automation and Robotics in Construction (ISARC 2008) (Vilnius, Lithuania, 2008);
- International Conference on Modelling of Business, Industrial and Transport Systems (Riga, Latvia, 2008);
- International Conference ´Modern Buildings, Materials and StructuresĄ, (Vilnius, Lithuania, 2004, 2007, 2010, 2013), etc.

Professor E. K. Zavadskas was one of the main initiators of international German – Lithuanian – Polish colloquium dedicated to Operational Research (OR) in Civil Engineering. The first one was held in 1986, Leipzig, Germany. The Colloquia are organized every two years. Since the first one, 14 colloquia have already been organized.

On the basis of collaboration during 11th and the 12th colloquia, the idea of setting up of a new EURO working group ´OR in Sustainable Development and Civil Engineering (EW-GORSDCE)Ą was presented. On the initiative of professor Zavadskas, the Working Group was established during the 23rd European Conference on Operational Research ´OR creating competitive advantageĄ, which took place in Bonn, Germany 5–8 July 2009 (http://www.euro-online.org/web/ewg/32/ewg-orsdce-or-in-sustainable-development-and-civil-engineering). Prof. E. K. Zavadskas is a Chairman of this working group.

Under E. K. Zavadskas supervision, 33 PhD dissertations were presented (four of his former students were awarded the title of Full Professor).

On the day of his Jubilee, we would like to congratulate Professor Edmundas Kazimieras Zavadskas – an exceptional scientist of greatest format. We wish the Professor good health and creativity in further contributions to MCDM methods.

4 Conclusions

Professor Edmundas Kazimieras Zavadskas has greatly contributed to development and practical applications of MCDM methods. This is why a part of the article was devoted to his achievements. His numerous articles and, most of all, authorship or co-authorship of books contribute to the MCDM theory and practice, as well as encourage continuous innovations in this field.

Bibliography

[1] Zavadskas, E.K. (1987); *Multiattribute Decision Making in Construction.* Dr.Sc. Dissertation. Moscow: Institute of Civil Engineering. 720 p. (in Russian).

[2] Zavadskas, E. (1987); *Complex Estimation and Choice of Resource-Saving Decisions in Construction.* Vilnius: Mokslas. 210 p. (in Russian).

[3] Zavadskas, E.K. (1991); *System of Estimation of Technological Solutions in Building Construction.* Leningrad: Stroizdat. 256 p. (in Russian).

[4] Zavadskas, E.; Peldschus, F.; Kaklauskas, A. (1994); *Multiple Criteria Evaluation of Projects in Construction.* Vilnius: Technika. 226 p.

[5] Zavadskas, E.K.; Kaklauskas, A. (2007); *Mehrzielselektion für Entscheidungen im Bauwesen.* Stuttgart: IRB Verlag. 276 p. (in German).

[6] Zavadskas, E.; Kaplinski, O.; Kaklauskas, A.; Brzezinski, J. (1995); *Expert Systems in Construction Industry. Trends, Potencial & Applications.* Vilnius: Technika. 180 p.

[7] Zavadskas, E.K.; Kaklauskas, A.; Turskis, Z.; Tamosaitiene, J. (2008); Selection of the Effective Dwelling House Walls by Applying Attributes Values Determined at Intervals, *Journal of Civil Engineering and Management*, 14(2): 85-93.

[8] Zavadskas, E.K.; Antucheviciene, J.; (2007); Multiple Criteria Evaluation of Rural Building's Regeneration Alternatives, *Building and Environment*, 42(1): 436-451.

[9] Zavadskas, E.K.; Turskis, Z. (2010); A New Additive Ratio Assessment (ARAS) Method in Multicriteria Decision-Making, *Technological and Economic Development of Economy*, 16(2): 159-172.

[10] Turskis, Z.; Zavadskas, E.K.; (2010); A Novel Method for Multiple Criteria Analysis: Grey Additive Ratio Assessment (ARAS-G) Method, *Informatica*, 21(4): 597-610.

[11] Turskis, Z.; Zavadskas, E.K. (2010); A New Fuzzy Additive Ratio Assessment Method (ARAS-F). Case Study: the Analysis of Fuzzy Multiple Criteria in Order to Select the Logistic Centers Location, *Transport*, 25(4): 423-432.

[12] Keršulienė, V.; Zavadskas, E.K.; Turskis, Z. (2010); Selection of Rational Dispute Resolution Method by Applying New Step-Wise Weight Assessment Ratio Analysis (SWARA), *Journal of Business Economics and Management*, 11(2): 243-258.

[13] Antuchevičienė, J.; Zavadskas, E.K.; Zakarevičius, A. (2010); Multiple Criteria Construction Management Decisions Considering Relations between Criteria, *Technological and Economic Development of Economy*, 16(1): 109-125.

[14] Zavadskas, E.K.; Turskis, Z. (2008); A New Logarithmic Normalization Method in Games Theory, *Informatica*, 19(2): 303-314.

[15] Zavadskas, E.K.; Turskis, Z.; Antucheviciene, J.; Zakarevicius, A. (2012); Optimization of Weighted Aggregated Sum Product Assessment, *Electronics and Electrical Engineering = Elektronika ir elektrotechnika*, 122(6): 3-6.

[16] Dadelo, S.; Krylovas, A.; Kosareva, N.; Zavadskas, E.K.; Dadeliene, R. (2014); Algorithm of Maximizing the Set of Common Solutions for Several MCDM Problems and it˙s Application for Security Personnel Scheduling, *International Journal of Computers, Communications & Control (IJCCC)*, 9(2): 140-148.

[17] Zavadskas, E.K.; Ustinovičius, L.; Peldschus, F. (2003); Development of Software for Multiple Criteria Evaluation, *Informatica*, 14(2): 259-272.

[18] Brauers, W.K.M.; Zavadskas, E.K. (2006); The MOORA Method and its Application to Privatization in a Transition Economy, *Control and Cybernetics*, 35(2): 445-469.

[19] Brauers, W.K.M.; Zavadskas, E.K. (2010); Project Management by MULTIMOORA as an Instrument for Transition Economies, *Technological and Economic Development of Economy*, 16(1): 5-24.

[20] Razavi Hajiagha, S.H.; Hashemi, S.S.; Zavadskas, E.K. (2013); A Complex Proportional Assessment Method for Group Decision Making in an Interval-Valued Intuitionistic Fuzzy Environment, *Technological and Economic Development of Economy*, 19(1): 22-37.

[21] Razavi Hajiagha, S.H., Hashemi, S.S.; Zavadskas, E.K.; Akrami, H. (2012); Extensions of LINMAP Model for Multi Criteria Decision Making with Grey Numbers, *Technological and Economic Development of Economy*, 18(4): 636-650.

[22] Razavi Hajiagha, S.H.; Mahdiraji, H.A.; Zavadskas, E.K.; Hashemi, S.S. (2013); A Fuzzy Data Envelopment Analysis Approach Based on Parametric Programming, *International Journal of Computers, Communications & Control (IJCCC)*, 8(4): 594-607.

[23] Razavi Hajiagha, S.H.; Akrami, H.; Zavadskas, E.K.; Hashemi, S.S. (2013); An Intuitionistic Fuzzy Data Envelopment Analysis for Efficiency Evaluation under Ucertainty: Case of a Finance and Credit Institution, *E&M Economics and Management = E&M Economie a Management*, 1: 128-137.

[24] Razavi Hajiagha, S.H.; Zavadskas, E.K.; Hashemi, S.S. (2013); Application of Stepwise Data Envelopment Analysis and Grey Incidence Analysis to Evaluate the Effectiveness of Export Promotion Programs, *Journal of Business Economics and Management*, 14(3): 638-650.

[25] Antuchevičienė, J.; Zakarevičius, A.; Zavadskas, E.K. (2011); Measuring Congruence of Ranking Results Applying Particular MCDM methods, *Informatica*, 22(3): 319-338.

[26] Zavadskas, E.K.; Zakarevicius, A.; Antucheviciene, J. (2006); Evaluation of Ranking Accuracy in Multi-Criteria Decisions, *Informatica*, 17(4): 601-618.

[27] Zavadskas, E.K.; Antucheviciene, J.; Saparauskas, J.; Turskis, Z. (2013); MCDM Methods WASPAS and MULTIMOORA: Verification of Robustness of Methods when Assessing Alternative Solutions, *Journal of Economic Computation and Economic Cybernetics Studies and Research* (ECECSR), 47(2): 5-20.

[28] Brauers, W. K. M.; Zavadskas, E. K. (2012); Robustness of MULTIMOORA: a method for multi-objective optimization, *Informatica*, 23(1): 1-25.

[29] Brauers, W. K. M.; Kildienė, S.; Zavadskas, E. K.; Kaklauskas, A. (2013); The construction sector in twenty European countries during the recession 2008-2009 – country ranking by MULTIMOORA, International *Journal of Strategic Property Management*, 17(1): 58-78.

[30] Yazdani-Chamzini, A.; Yakhchali, S. H.; Zavadskas, E. K. (2012); Using a integrated MCDM model for mining method selection in presence of uncertainty, *Ekonomska Istraživanja - Economic Research*, 25(4): 869-904.

[31] Brauers, W. K. M.; Zavadskas, E. K. (2012); A multi-objective decision support system for project selection with an application for the Tunisian textile industry, *E&M Economie a Management*, 15(1): 28-43.

[32] Zavadskas, E. K.; Turskis, Z.; Volvačiovas, R.; Kildienė, S. (2013); Multi-criteria assessment model of technologies, *Studies in Informatics and Control*, 22(4): 249-258.

[33] Hashemkhani Zolfani, S.; Sedaghat, M.; Zavadskas, E. K. (2012); Performance evaluating of rural ICT centers (telecenters), applying Fuzzy AHP, SAW-G and TOPSIS Grey, a case study in Iran, *Technological and Economic Development of Economy*, 18(2): 364-387.

[34] Hashemkhani Zolfani, S.; Zavadskas, E. K.; Turskis, Z. (2013); Design of products with both international and local perspectives based on Yin-Yang balance theory and SWARA method, *Ekonomska Istraživanja - Economic Research*, 26(2): 153-166.

[35] Yazdani-Chamzini, A.; Fouladgar, M. M.; Zavadskas, E. K.; Moini, S. H. H. (2013); Selecting the optimal renewable energy using multi criteria decision making, *Journal of Business Economics and Management*, 14(5): 957-978.

[36] Hashemkhani Zolfani, S.; Esfahani, M. H.; Bitarafan, M.; Zavadskas, E. K.; Arefi, S. L. (2013); Developing a new hybrid MCDM method for selection of the optimal alternative of mechanical longitudinal ventilation of tunnel pollutants during automobile accidents, *Transport*, 28(1): 89-96.

[37] Staniunas, M.; Medineckiene, M.; Zavadskas, E. K.; Kalibatas, D. (2013); To modernize or not: ecological-economical assessment of multi-dwelling houses modernization, *Archives of Civil and Mechanical Engineering*, 13(1): 88-98.

[38] Kracka, M.; Zavadskas, E. K. (2013); Panel building refurbishment elements effective selection by applying multiple-criteria methods, *International Journal of Strategic Property Management*, 17(2): 210-219.

[39] Brauers, W. K. M.; Kracka, M.; Zavadskas, E. K. (2012); Lithuanian case study of masonry buildings from the Soviet period, *Journal of Civil Engineering and Management*, 18(3): 444-456.

[40] Zavadskas, E.K.; Vainiūnas, P.; Turskis, Z.; Tamošaitienė, J. (2012); Multiple Criteria Decision Support System for Assessment of Projects Managers in Construction, *International Journal of Information Technology & Decision Making*, 11(2): 501-520.

[41] Hashemkhani Zolfani, S.H.; Aghdaie, M.H.; Derakhti, A.; Zavadskas, E.K.; Varzandeh, M.H.M. (2013); Decision Making on Business Issues with Foresight Perspective; An Application of New Hybrid MCDM Model in Shopping Mall Locating, *Expert Systems with Applications*, 40(17): 7111-7121.

[42] Fouladgar, M. M.; Yazdani-Chamzini, A.; Zavadskas, E. K.; Haji M., Hamzeh, S. (2012); A new hybrid model for evaluating the working strategies: case study of construction company, *Technological and Economic Development of Economy*, 18(1): 164-188.

[43] Rezaeiniya, N.; Hashemkhani Zolfani, S.; Zavadskas, E. K. (2012); Greenhouse locating based on Anp-Copras-G Methods – an empirical study based on Iran, *International Journal of Strategic Property Management*, 16(2): 188-200.

[44] Hashemkhani Zolfani, S.; Rezaeiniya, N.; Pourhossein, M.; Zavadskas, E. K. (2012); Decision making on advertisement strategy selection based on life cycle of products by applying FAHP and TOPSIS GREY: growth stage perspective; a case about food industry in IRAN, *Inžinerinė Ekonomika - Engineering Economics*, 23(5): 471-484.

[45] Zavadskas, E. K.; Sušinskas, S.; Daniūnas, A.; Turskis, Z.; Sivilevičius, H. (2012); Multiple criteria selection of pile-column construction technology, *Journal of Civil Engineering and Management*, 18(6): 834-842.

INT J COMPUT COMMUN, ISSN 1841-9836
9(3):313-324, June, 2014.

Numerical Prediction of Time Series Based on FCMs with Information Granules

W. Lu, J. Yang, X. Liu

Wei Lu*, Jianhua Yang, Xiaodong Liu
School of Control Science and Engineering
Dalian University of Technology
Dalian 116023, China
*Corresponding author: luwei@dlut.edu.cn

Abstract: The prediction of time series has been widely applied to many fields such as enrollments, stocks, weather and so on. In this paper, a new prediction method based on fuzzy cognitive map with information granules is proposed, in which fuzzy c-means clustering algorithm is used to automatically abstract information granules and transform the original time series into granular time series, and subsequently fuzzy cognitive map is used to describe these granular time series and perform prediction. two benchmark time series are used to validate feasibility and effectiveness of proposed method. The experimental results show that the proposed prediction method can reach better prediction accuracy. Additionally, the proposed method is also able to precess the modeling and prediction of large-scale time series.
Keywords: Fuzzy Cognitive Maps (FCMs), time series, prediction, , information granules.

1 Introduction

Time series is a sequence of real-data, with each element in this sequence representing a value recorded at some time moment. As a classic issue, time series prediction has been used in diverse fields, which utilizes prediction model describing some useful temporal relationship that is developed by observing a past certain variable or a past family of variables to extrapolate future values. How to construct prediction model of time series is core for prediction of time series. Many researcher early focus on how to construct predictive model of time series with the aid of the linear system theory [1], the stochastic process theory [2] and the black-box methodology [3], and dynamical system analysis [4]. However, the constructed prediction model by using these methods cannot solve prediction problem in which the historical data are missing or uncertain.

Fuzzy sets theory can be used to make semantics and represent the data themselves and fuzzy reasoning offers a viable alternative to ensure robustness to the inherent uncertainty, which has involved into modeling and prediction of time series. Song and Chissom [5–7] defined the concept of fuzzy time series and developed two fuzzy time series prediction models — the time-invariant model [6] and the time-variant model [7]. Following the work of Song and Chissom, many the improved prediction models associated with fuzzy time series are emerged such as the Markov model [8], Chens model [9], Hwangs model [10], the heuristic model [11], the high order model [12], the local trend model [13] and so on. These models have been used to predict enrollment of university, stock index, temperature etc., which also shows better prediction performance. As data, time series is inherently associated with large size, high dimensionality and a stream-like nature. Whereas construction of the fuzzy logic relationships which has important impact on performance of models is a tedious work in the development process of existing fuzzy time series models, which is difficult and complicated for modeling of the large-scale fuzzy time series. To overcome the deficiency, fuzzy cognitive map (FCM for short) with information granules seem to can become an alternative.

Fuzzy cognitive map as a soft computing technology for modeling complex systems was proposed by Kosko in 1986 [14], which are treated as an alternative way for knowledge-based representation and inference process for its easy of usage and numeric matrix operation for complex system. Based on the nature, FCM has capable to handle prediction problem of time series [15, 16]. Further, information granules and information granulation [17, 25] play key roles when dealing with large data. As expression of meaningful the abstract entities, information granules come with information granulation. The role of information granulation is to organize detailed numerical data into some meaningful, semantically sound entities (information granules). In particular, granulate information are able to achieve a high level of interpretability and manage phenomena which are complex and consequently the data are overwhelming. The formalism of information granules includes intervals, fuzzy sets, rough sets, shadow sets or alike. There are many methods supporting design of information granules, one of which is fuzzy c-means clustering. It can provide an ability abstracting fuzzy information granules from data with multivariate attribute.

In this paper, starting with global view of information granules, we attempt to construct fuzzy cognitive map with information granules to realize prediction of time series. Our proposed prediction method of time series includes two stage: the first stage is to construct fuzzy cognitive map prediction model on basis of information granules, and the second stage is to perform prediction by exploiting the fuzzy cognitive map with information granules based model which is constructed by the first stage.

The remainder of this paper is organized as follows. Section 2 briefly introduces some concepts related to FCM, and then focus on learning method of FCM. Section 3 presents a method of constructing FCM with information granules by using fuzzy c-means clustering algorithm.The proposed prediction method of time series based on FCM with information granules is detailed in section 4. In section 5, two benchmark time series are used to validate the feasibility and effectiveness of proposed method, and experimental results is also discussed. Finally, some conclusions is provided in section 6.

2 Fuzzy cognitive map and its learning method

In this section, some concept associated with fuzzy cognitive map is first recalled, and then the learning method based on particle swarms optimization (PSO) technology, an effective learning method of FCM, is presented to learn weights of FCM.

2.1 Fuzzy cognitive map

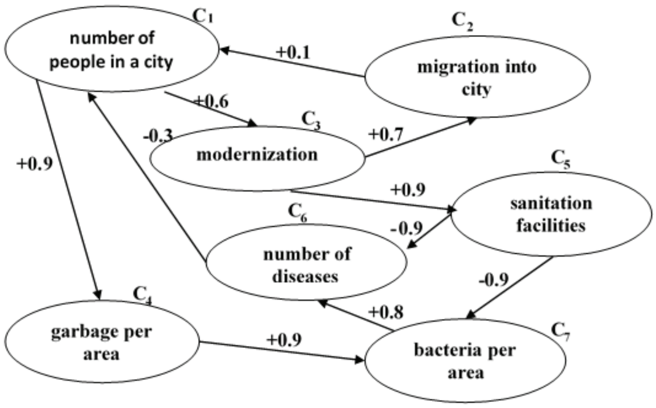
FCM is simple, very powerful tool for representation of human knowledge and performing reasoning. FCM can describe a given system by concepts and mutual relationships among them, which play an important role for time series modeling and prediction in this paper.

FCM is composed of a collection of nodes and directed links (edges) between nodes. In the FCM, nodes is used to represent concepts, say $C_1, C_2, \cdots, C_n$. These concepts can be envisioned as status, variables etc. which is used to describe main dynamic characteristic of problem/process/system. Values of nodes (concepts) are fuzzy and change with time. The directed edges represent casuality between nodes or more precisely a way in which one node affects another one. The connections between nodes could be asymmetric. The strength of connections (it is also called weights of FCM) from node C_j to node C_i, denoted by ω_{ij}, is quantified to be range from -1 to 1. The value of ω_{ij} reflects different casuality between C_i and C_j, viz.,

- $\omega_{ij} > 0$, which indicates positive causality between C_j and C_i, i.e., an increase of value of C_j leads to an increase of value of C_i (and vice versa).

- $\omega_{ij} = 0$, which indicates neutral causality between C_j and C_i, i.e., no relationship between C_j and C_i. In this case, the connection from C_i to C_j can be removed.
- $\omega_{ij} < 0$, which indicates negative causality between C_j and C_i, i.e., an increase of value of C_j leads to an decrease of value of C_i (and vice versa).

The fuzzy cognitive map can be represented as not only the directed graphic but also a square matrix. The square matrix, which is also called relationship matrix, stores all values of weights of FCM. Fig.1 shows an example of FCM model and its relationship matrix that concerns public city health issues.



(a) FCM model of public city health issues

	C_1	C_2	C_3	C_4	C_5	C_6	C_7
C_1	0	0	0.6	0.9	0	0	0
C_2	0.1	0	0	0	0	0	0
C_3	0	0.7	0	0	0.9	0	0
C_4	0	0	0	0	0	0	0.9
C_5	0	0	0	0	0	0.9	0.9
C_6	−0.3	0	0	0	0	0	0
C_7	0	0	0	0	0	0.8	0

(b) The relationship matrix of FCM

Figure 1: FCM model of public city health issues and its relationship matrix.

Behind FCM, there is mathematical mechanism which is described in forms of (1):

$$C_i(t+1) = f(\sum_{j=1}^{n} \omega_{ij} C_j(t) + \omega_{0i}) \tag{1}$$

where $C_j(t)$ is the the active level (value) of jth node at the tth time moment, $\omega_{ij} \in [1, 1]$ is the value of weight from the concept C_j to the concept C_i and $\omega_{0i} \in [0, 1]$ is the bias associated with the ith node. Besides, n is the number of nodes of FCM, and f is the transformation function that is generally selected as sigmoid function with steepness parameter σ — $f(u) = 1/(1 + e^{-\sigma u}))(u \in R, \sigma > 0)$, where the steepness parameter σ is associated with the individual node of the FCM. The role of this parameter is to provide some additional calibration of the value of the node [16].

Let $\mathbf{w}_i = [\omega_{i1}\ \omega_{i2}\ \cdots\ \omega_{in}]^T$ be the weight vector which includes all values of weights from all other nodes to ith node. Likewise the vector of active level of nodes is described as $\mathbf{C} = [C_1\ C_2\ \cdots\ C_n]^T$, which includes all values of nodes of FCM. Thus we can rewrite (1) to obtain a more concise vector notation as follows.

$$C_i(t+1) = f(\mathbf{w}_i^T \mathbf{C}(t) + \omega_{0i}) \tag{2}$$

Once nodes of FCM and weights between these nodes are determined, FCM starts with a given initial state vector $\mathbf{C(t)} = [C_1(t), C_2(t), \cdots, C_n(t)]^T$ to perform iteration computation according to (1). After finite iteration, FCM can reach an equilibrium point or a limit cycle. It is worthy noting that at $t+1$ time moment the active level of ith node depends on all the value of node of FCM at t time moment.

2.2 The learning of fuzzy cognitive map

The sound values of weights can ensure FCM accurately describing dynamic behavior of system to research. Several methods such as the Hebbian learning [14] on basis of the scheme of unsupervised learning, the numeric data-based PSO algorithm [16] and evolutionary optimization [19] have been proposed. As an optimization method, PSO algorithm offers possibilities of global, population-based optimization yet not imposing a very heavy computational overload. The method allows for determining parameters of FCM from original time series data without human input. In this paper, we take PSO algorithm to learn all parameters of FCM.

In brief, PSO involves a population of particles whose dynamics is guided by the mechanisms of social interactions and personal experience. The details of PSO and its enhancement version are documented in the literature [20] and [21].

The objective of utilizing PSO algorithm to learn parameters of FCM is to develop candidate FCM and enable it to mimic the given input data. This optimization problem requires to establish $n \times (n+2)$ parameters. Consequently, the particles structure is defined as

$$\begin{aligned}\hat{\mathbf{W}} = \ [\ & \omega_{11}, \omega_{12}, \cdots, \omega_{1n}, \omega_{21}, \omega_{22}, \cdots, \omega_{2n}, \cdots, \omega_{n1}, \omega_{n2}, \cdots, \omega_{nn}, \\ & \omega_{01}, \omega_{02}, \cdots, \omega_{0n}, \sigma_1, \sigma_2, \cdots, \sigma_n] \end{aligned} \tag{3}$$

where $w_{ij} \in [1, 1]$ is the weight from node j to node it, $w_{0i} \in [0, 1]$ is the bias associated with the ith node, and $\sigma_i > 0$ is the steepness parameters of transformation function f.

Objective function (4) is used to evaluate quality of particles in population, and is defined by taking advantage of an inherent property of FCM model.

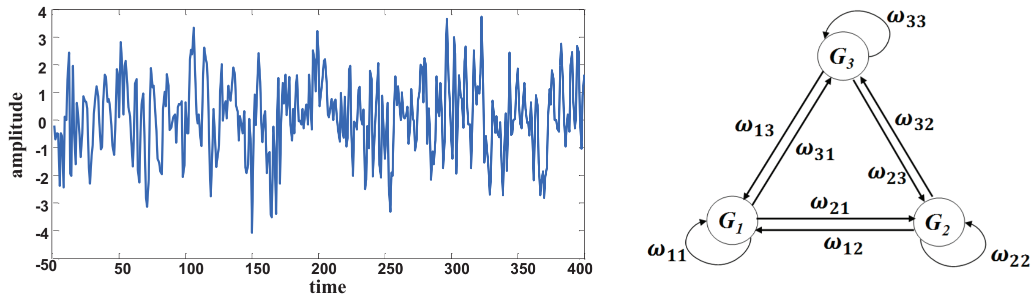
$$min : f = \frac{1}{(l-1)\,n} \sum_{t=1}^{l-1} \sum_{i=1}^{n} \parallel \hat{C}_i(t+1) - C_i(t+1) \parallel^2 \tag{4}$$

where $\mathbf{C}(t+1) = [C_1(t+1)\ C_2(t+1)\ \cdots\ C_n(t+1)]^T$ is a actual response for initial state vector $\mathbf{C(t)} = [C_1(t)\ C_2(t)\ \cdots\ C_n(t)]^T$, $\hat{\mathbf{C}}(t) = [\hat{C}_1(t+1)\ \hat{C}_2(t+1)\ \cdots\ \hat{C}_n(t+1)]$ is a response of the condidate FCM for initial state vector $\mathbf{C(t)}$, l is a number of input data points, n is a number of nodes of FCM. The objection function can realize comparison between the single-step response of the candidate FCM and the actual response for the same initial state vector.

3 The design of fuzzy cognitive map with information granules

One fundamental problem when FCM is used in time series modeling is how to design structure of FCM for a given time series, i.e., how to map the given time series onto the structure of FCM, viz. expressing a meaning of the nodes of the graph and specifying possible causal linkages between pairs of nodes. In general, if the concepts of system to research are easily identifiable, these concepts can be directly cast into FCM. However, the clearly and understandable concepts are hardly obtained from time series including a series of numeric data, which need to be discovered or mined. Here the idea of information granulation is adopted. On a basis of numeric data formed are information granules and they can be treated as nodes of FCM. Fuzzy c-means clustering [22] can serve as a convenient vehicle to construct information granules from data. In what follows, an illustrative example is presented how to design structure of FCM for time series.

Let us consider a certain time series as shown in Fig.2 (a). The numeric data included in the time series is clustered by using the standard version of fuzzy c-means clustering algorithm with Euclid distance, where the number of clusters c is set into 3 and the fuzzification coefficient m is set into 2 as well. The results of clustering is reported in Table 1 in where G_1, G_2, G_3



(a) The original time series plot.

(b) The expression of FCM with information granules

Figure 2: FCM expression of a certain time series.

are prototypes. These prototypes can be viewed as information granulations represented by the following semantics:

- G_1 — the amplitude of time series is negative high where locates nearby -1.9019.
- G_2 — the amplitude of time series is zero where locates nearby 0.
- G_3 — the amplitude of time series is positive high where locates nearby 1.7832.

From perspective of information granules, the dynamic behaviour of timer series can be described through establishing relationships among these granules.

Table 1: The prototypes and granular description of time series showed in Fig.2 (a)

Cluster No.	1	2	3
prototypes	-1.9019	0.025	1.7832
Description of prototypes information granules	G_1	G_2	G_3

From Table 1, if the number of clusters is envisioned as the total number of node of FCM, and the fuzzy semantics associated with prototypes, say G_1, G_2 and G_3, are regarded as the concepts of corresponding nodes of FCM, FCM is constructed to express time series showed in Fig.2 (a), which is shown in Fig.2 (b). FCM with sound weights can represent dynamic characters of the time series, which can become realization by PSO algorithm to minimize objective function (4) (refer to section 2.2).

4 The proposed prediction method of time series based on fuzzy cognitive map with information granules

In this section, the proposed prediction method of time series based on FCM with information granules is detailed. FCM is exploited to realize prediction of time series, which is based on the idea that the structure of it can store the values of weights between nodes that describe dynamic behaviour of time series at each iteration step.

Suppose that $X = \{x_1, x_2, \cdots, x_l\}$ is a time series. The outline of proposed method is presented in Fig.3, which consists of two stages — the first stage (from step 1) to step 5)) is to construct FCM model to express time series and the second stage (from step 6) to step 8)) is to use the constructed FCM model to perform prediction. In what follows, the two stages are detailed respectively.

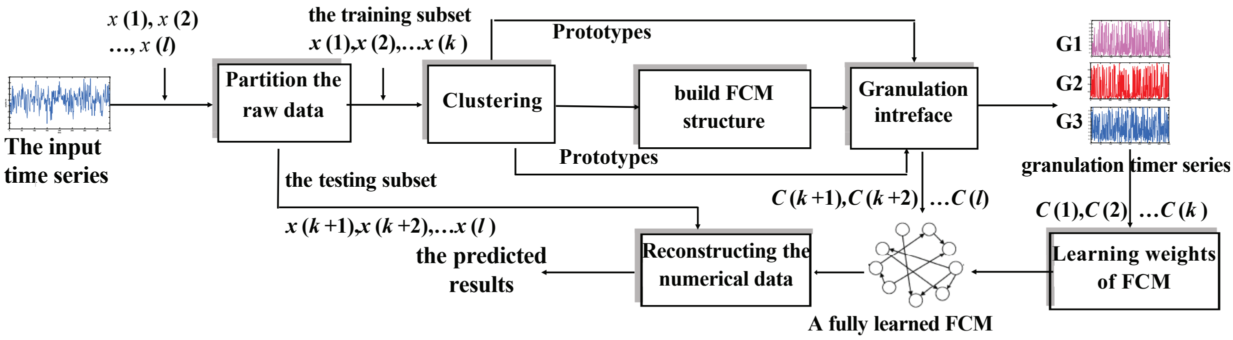


Figure 3: The outline of proposed prediction method.

Step 1) Partition the raw time series $\mathbf{X}$ into the training set consisted of k observations and the testing set consisted of lk observations. The former is used to develop FCM model, whereas the latter one is separate and is used to carry out prediction of the non-observed data.

Step 2) Generate information granules by fuzzy c-means clustering algorithm. Its object is to capture the linguistic and numerical characters of the training set. The stand version of fuzzy c-means clustering algorithm is adopted in this paper, which accepts the all observations from the training set and the predefined number of clustering c, and by the end of clustering algorithm, a prototype vectors $\mathbf{P} = [p_1 \ p_2 \ \cdots \ p_c]$ can be obtained. Each prototype p_i $(i = 1, 2, \cdots, c)$ is assigned into semantics, which result in the formation of information granules, say $G_1, G_2, \cdots, G_c$ (see section 3), i.e., prototypes can be make semantics, which form information granules.

Step 3)) Build the structure of candidate FCM via the information granules generated by step 2). The step plays an important role. As mentioned previously in section 3, the predefined number of clusters c is regarded as the total number of nodes of FCM and information granules are directly taken as the concept of corresponding nodes of FCM. Subsequently, the structure of fuzzy cognitive maps is formed.

Step 4) Granulate each observation in the training set by granular interface, which creates ordered data points by the well-known (5) according to prototypes $\mathbf{P}$ obtained previously and the each observation in the training set. Note that u_i can become the active level of the ith node of FCM related to ith granules G_i which is formed by assigning semantics into prototype p_i.

$$u_i(p_i, x(t)) = \frac{1}{\sum_{j=1}^{c} \| \frac{x(t)-p_i}{x(t)-p_j} \|^{\frac{2}{m-1}}}, i = 1, 2, \cdots, c; t = 1, 2, \cdots, k \tag{5}$$

where $x(t)$ is an observation at t time moment from the training set, c is the number of cluster, p_j $(j = 1, 2, \cdots, c)$ is the jth prototype obtained previously, m is fuzzification coefficient, $u_i(p_i, x(t))$ is the membership value that $x(t)$ belonged to prototype p_i and t is time label.

$$\mathbf{F} = \begin{bmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \\ \vdots \\ \mathbf{G}_c \end{bmatrix} = \begin{bmatrix} \mathbf{C}(1) \\ \mathbf{C}(2) \\ \vdots \\ \mathbf{C}(l) \end{bmatrix}^T = \begin{bmatrix} c_1(1) & c_1(2) & \cdots & c_1(l) \\ c_2(1) & c_2(2) & \cdots & c_2(l) \\ \vdots & \vdots & \vdots & \vdots \\ c_c(1) & c_c(2) & \cdots & c_c(l) \end{bmatrix} = \begin{bmatrix} u_1(p_1, x(1)) & u_1(p_1, x(2)) & \cdots & u_1(p_1, x(l)) \\ u_2(p_2, x(1)) & u_2(p_2, x(2)) & \cdots & u_2(p_2, x(l)) \\ \vdots & \vdots & \vdots & \vdots \\ u_c(p_c, x(1)) & u_c(p_c, x(2)) & \cdots & u_c(p_c, x(l)) \end{bmatrix} \tag{6}$$

According to (5), each observations $x_i(t)$ $(i = 1, 2, \cdots, l-k)$ from the training subset can be transformed along with time variable t, and become available in the form of their membership values corresponding to each granules G_j, which means original time series is converted to granular time series. We unfold (5) along with time label t, (6) can be obtained. Let us observe (6), $G_1, G_2, \cdots, G_c$ are defined as granular time series which includes c subsequence. Note that each row vector of $\mathbf{F}$, say $\mathbf{G}_1, \mathbf{G}_2, \cdots, \mathbf{G}_c$, expresses the level that the given time series can be characterized by corresponding fuzzy semantics p_i, whereas each column vector of $\mathbf{F}$, say $\mathbf{C}(1), \mathbf{C}(2), \cdots, \mathbf{C}(l)$, expresses the level that an observation of time series $x(t)$ at t time moment can be characterized by all semantics.

Step 5) Learn weights of FCM by PSO algorithm according to the transformed granular time series data. Its role is to establish a fully learned FCM model. The PSO algorithm described in Section 2.2 is used to learn all parameters vector $\hat{\mathbf{W}}$ of FCM constructed by step 3) on basis of the granulation time series data $\mathbf{C}(1), \mathbf{C}(2), \cdots, \mathbf{C}(l)$ which is formed by step 4).

Step 6) Granulate datum from the testing set by granulation interface, which is regarded as input of the fully learned FCM formed by step 5). A point $x(k+s)$ at $k+s$ time moment within the testing set, which is transformed into a tuples initial state vector $\mathbf{C}(k+s)$ according to (5), which indicate the activation level of all nodes in FCM model.

Step 7) Perform iteration computation on basis of (1). FCM model perform one step iteration from initial state vector $\mathbf{C}(k+s)$ and generates response vector $\hat{\mathbf{C}}(k+s+1)$. Results of iteration computation of FCM indicate the state of granular time series in the next time moment, in other words, the iteration process of FCM is actually the prediction process on the level of information granules.

Step 8) Reconstruction the numerical values according to the activation level of all nodes of FCM at a certain time moment. The numerical prediction is carried out according to (7) which reconstructs numerical values on basis of the activation value among all nodes of FCM at a certain time moment and prototypes formed by fuzzy c-means clustering algorithm prior.

$$\hat{x}(t) = \frac{\sum_{j=1}^{c} \hat{c}_j(t)^m p_j}{\sum_{j=1}^{c} \hat{c}_j(t)^m} \tag{7}$$

where c is the number of cluster, p_j $(j = 1, 2, \cdots, c)$ is prototypes, m is fuzzification coefficient, $\hat{c}_j(t)$ is an activation value of jth node of FCM model at t time moment, and $\hat{x}(t)$ is the predicted numerical value at t time moment.

5 Experimental study

In this section, two benchmark time series — the enrollment of university and Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX) time series, are used to carry out profound experiment for validating feasibility and effectiveness of proposed predicted method. The goal is to assess quality of the proposed method and compare with other method based on fuzzy sets theory.

5.1 Experimental setup

The proposed prediction method concern only one adjustable parameter,say the predefined number of clustering c. The all experiments for the two time series include the two tasks. One is that the quantification of the impact on the prediction accuracy of the proposed method being brought by the number of clusters c which implies the level of granularity. The other is that the predicted results are obtained by using our proposed method are compared with other prediction methods.

Considering a given time series and a predefined value of c, we first divide the original time series into the training set and the testing set. The training subset is used to develop HFCM model and the testing subset is used to perform prediction. In the sequence, the experiments are carried out by systematically sweeping through the values of the number of clusters c which start with 3 and increment it until the value of 8 is reached. To assure high confidence in the produced prediction results, for each of c, the experiments were repeated 50 times, the average root mean squared error (RMSE) are reported whose goal is to evaluate prediction accuracy and compare with other methods. Besides, the other one parameters of fuzzy c-means clustering algorithm — fuzzification coefficient m is set into 2 for each of experiments. The standard version of PSO algorithm is used to learn parameters of all candidate FCM models of the two time series. Parameters of PSO algorithm used in all experiments is shown in Table 2.

Table 2: Parameters of PSO algorithm used in all experiments

Description	Value
Population size	50
Acceleration constant ϕ_1	2
Acceleration constant ϕ_2	2
Inertial weight ζ	0.9
Initial positions	Random number
maximum number of iterations	1000
minimum objection function value	10^5

5.2 Experimental results and analysis

There are two benchmark time series are applied to validate and analyze the proposed prediction method. The first is the enrollment of university of Alabama during 1971–1992, which includes 2 time series which includes 22 observations. The time series is often used by many researchers [6]– [13]. The second time series involves the daily values of TAIEX from January 1, 2000 to December 30, 2000, which consists of 242 observations. The time series has been used in the related literature [9, 15].

In order to directly compare with other methods, for the first time series, its all observations are taken as not only the training data for constructing FCM model but also the testing data for validating prediction accuracy. Whereas for the second time series, data from 2000/1/4 to 2000/10/31 are used to construct FCM model and data from 2000/11/1 to 2000/12/30 are used to perform prediction. In this case, the comprehensive tests that involve the two data sets are carried out respectively according to the description in section 5.1, whose results are reported in Fig.4, Fig.5, Fig.6, Table 3 and Table 4.

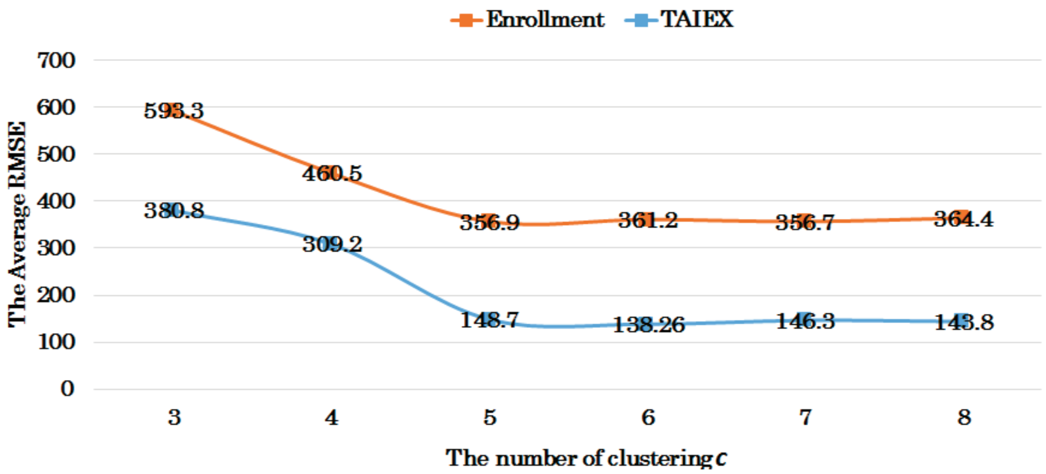


Figure 4: the plot of the average RMSE versus the number of clustering for the enrollment and TAIEX time series.

Table 3: The comparison result for enrollment time series.

Method	RMSE
Song's time-invariant method [6]	677
Chen's method [9]	663
Markov method [8]	638
Hwang's method [10]	567
Huarng's method [11]	489
Dan's method [13]	438
Proposed method (c=5)	346

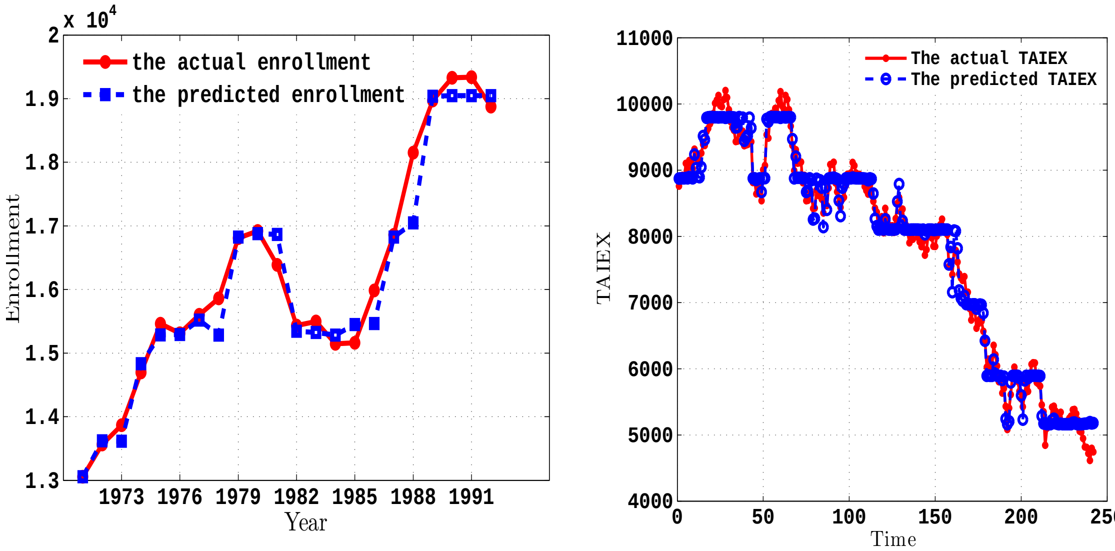
1) Experiments with respective to parameter c of proposed prediction method. Fig.4 represents a plot of the average RMSE versus the number of clustering c. Several interesting conclusions can be drawn when analyzing Fig.4. First, the prediction accuracy is highly sensitive to the choice of the number of clusters. The value of average RMSE get lower dramatically with the increasing number of clusters however the reduction of the average RMSE becomes less visible when going beyond a certain number of clusters i.e. 5 or 6. In other words, the prediction accuracy is not continuously increase with the increasing number of clusters. The significant increase in the prediction accuracy happens when moving from very low values of clusters to some higher values. The prediction accuracy changes unobviously with the change of clusters number when the value of cluster number c greater than 5 or 6. For example, as to enrollment time series, the average RMSE is 593.3 for c=3, whereas it is 356.9 for c=5 and 364.4 for c=8.

Table 4: The comparison result for TAIEX time series.

Method	RMSE
Chen's method [9]	176
Yu's method [23]	170
Huarng's method [24]	139
Proposed method (c=6)	132

There is smaller difference in RMSE at the higher number of cluster just 7.5, but there is larger difference in RMSE at the lower number of clusters reaching 236.4. For TAIEX time series, the situation is also similar and optimal values of c are located at the value of 6. The change of the average RMSE is slightly when the value of c over 6. The explanation for this is that when the number of clusters is low, very few granules are generated which is not sufficient to capture the character of time series and gives rise to the larger prediction error. But the number of clusters is larger, the granules may be more minute, which drastically reduces their interpretability.

2) Comparison with other prediction methods based on fuzzy sets theory. Table 3 and Table 4 reports comparison with other methods based on fuzzy sets theory respectively. The results in Table 3 show that the proposed prediction method with the best parameters (c=5) can obtain the better accuracy for enrollment time series. The RMSE of the proposed methods can achieve 346, while the best result Dans method [13] scored 438. At the same time, results for the TAIEX time series also show superiority of Huarngs method [24] . The raw time series and the selected best predicted results for enrollment and TAIEX time series are illustrated in Fig.5 and Fig.6 respectively.



(a) The plot of the actual and predicted enrollment. (b) The plot of the actual and predicted TAIEX.

Figure 5: The experimental results of using our proposal prediction method

6 Conclusions

In this paper, a time series prediction method based on fuzzy cognitive with information granules was proposed. It includes two important components — fuzzy clustering algorithm and fuzzy cognitive maps. The former extracts information granules from original time series data and then transform the original time series into the granular time series, while the latter is used to describe these granular time series and perform prediction. Comprehensive experiments have been carried out for two benchmark time series to validate feasibility and effectiveness of our proposed method. The results of experiment show predicted accuracy is related to the number of cluster. The proposed prediction method can obtain satisfying prediction accuracy in the case of the sound number of cluster. Additional, our proposed prediction method can automatically perform modeling and prediction of time series without more human intervention, its potential advantage is capability of handling modeling and prediction of large-scale time series.

Acknowledgements

The research is supported by the Natural Science Foundation of China under Grant 61175041 and The Major National Science and Technology Programs of China in the "Twelfth Five Year" Plan Period (No. 2011ZX05039-3-3).

Bibliography

[1] Kailath, T. (1980); *Linear Systems*, Prentice Hall.

[2] Papoulis, A. (1991); *Probability, Random Variables and Stochastic Processes*, Mcgraw-Hill College.

[3] Juditsky, A. et al (1994); *Wavelet in identification: wavelets, splines, neurons, fuzzies: how good for identification?*, INRIA reports No.2135.

[4] Kaplan, D.; Glass L. (1995); *Understanding Nonlinear Dynamics*, Springer Verlag.

[5] Song, Q.; Chissom, B.S. (1993); Fuzzy time series and its models, *Fuzzy Sets Systems*, 54(3): 269-277.

[6] Song, Q.; Chissom, B.S. (1993); Forecasting enrollments with fuzzy time series — Part I, *Fuzzy Sets Systems*, 54(1): 1-9.

[7] Song, Q.; Chissom, B.S. (1994); Forecasting enrollments with fuzzy time series — Part II, *Fuzzy Sets Systems*, 62(1): 1-8.

[8] Sullivan, J.; Woodall, W.H. (1994); A comparison of fuzzy forecasting and Markov modeling, *Fuzzy Sets Systems*, 64(3): 279-293.

[9] Chen, S.M. (1996); Forecasting enrollments based on fuzzy time series, *Fuzzy Sets Systems*, 81(3): 311-319.

[10] Hwang, J.R.; Chen, S.M.; Lee, C.H. (1998); Handling forecasting problems using fuzzy time series, *Fuzzy Sets Systems*, 100(1-3): 217-228.

[11] Huarng, K. (2001); Heuristic models of fuzzy time series for forecasting, *Fuzzy Sets Systems*, 123(3): 137-154.

[12] Chen, S.M. (2002); Forecasting enrollments based on high-order fuzzy time series, *Cybernetics and Systems: An International Journal*, 33(1): 1-16.

[13] Dan, J.; Dong F.; Hirota, K. (2011); Fuzzy Local Trend Transform based Fuzzy time series Forecasting Model, *International Journal of Computers, Communications* & *Control* , VI(4): 603-614.

[14] Kosko, B. (1986); Fuzzy Cognitive Maps, *International Journal of Man-Machine Studies*, 7: 65-75.

[15] Stach, W.; Kurgan, L.; Pedrycz, W. (2008); Numerical and Linguistic Prediction of Time Series With the Use of Fuzzy Cognitive Maps, *IEEE Transactions on Fuzzy system*, 16(1): 61-72.

[16] Pedrycz, W. (2010); The design of cognitive maps: A study in synergy of granular computing and evolutionary optimization, *Expert system with applications*, 37(10): 7288-7294.

[17] Pedrycz, W.; Vukovich, G. (2010); Abstraction and specialization of information granules, *IEEE Transactions on Systems Man and Cybernetics, Part B: Cybernetics*, 31(1): 106-111.

[18] Axelrod, R. (1976); *Structure of Decision: The Cognitive Maps of Political Elites*, Princeton University Press.

[19] Stach, W.; Kurgan, L.; Pedrycz, W.; Reformat M. (2005); Genetic learning of fuzzy cognitive maps, *Fuzzy Sets Systems*, 153(3): 371-401.

[20] Kennedy, J.; Eberhart, R. (1995); Particle Swarm Optimization, *Proceeding of IEEE International Conference on Neural network*, 4: 1942-1948.

[21] Shi, Y.H.; Eberhart, R. (1998); A modified particle swarm optimizer, *Proceeding of IEEE International Conference on Evolutionary Computation*, 69-73.

[22] Bezdek, J.C. (1981); *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press.

[23] Yu, H.K. (2005); A rened fuzzy time-series model for forecasting, *Physica A: Statistical Mechanics and its Applications*, 346(3-4): 657-681.

[24] Huarng, K.; Yu, H.K. (2005); A Type 2 fuzzy time series model for stock index forecasting, *Physica A: Statistical Mechanics and its Applications*, 353(1): 445-462.

[25] Lu, W.; Pedrycz, W; Liu, X.; Yang, J.; Li, P. (2014); The modeling of time series based on fuzzy information granules, *Expert Systems with Applications*, 41: 3799-3808.

INT J COMPUT COMMUN, ISSN 1841-9836
9(3):325-339, June, 2014.

An Adaptive GA in Partitioned Search Space

F. Nadi, A.T. Khader

Farhad Nadi*
School of computer schiences
Universiti Sains Malaysia, 11800 Penang, Malaysia
*Corresponding author: nadi@cs.usm.my

Ahamad Tajudin Khader
School of computer schiences
Universiti Sains Malaysia, 11800 Penang, Malaysia
E-mail: tajudin@cs.usm.my

Abstract: Evolutionary algorithms are population based meta-heuristics inspired from natural survival of fittest phenomena. Despite their reasonable performance, these algorithms suffer from some weaknesses including the need for finding the values of their parameters that affect their performance. A new algorithm is proposed that divide the search space into equal sized partitions. Each partition is assigned with two parameters that determine the intensification and diversification rates. The partitions will be intensified or diversified adaptively with regards to the corresponding parameters. Traditional crossover and mutation operators are replaced with two new parameter-free operators. The experiments conducted on a wide range of multi-modal and epistatic problems showed the superiority of the proposed method in comparison to other algorithms in literature.
Keywords: Genetic Algorithms,Adaptive Parameter Control, Crossover Rate, Mutation Rate

1 Introduction

Evolutionary algorithm (EA) is a search meta-heuristics that improvise the quality of a set of candidate solutions iteratively using variation and selection operators [8, 13, 27].

Two major issues in the design of global search methods are intensification and diversification [3]. The ability to visit different regions of the search space is generally referred to as diversification. The ability to obtain high quality solutions within those locations is generally referred to as intensification [3]. The balance between these two and the way it is conducted is the main factor that differentiate these algorithms from each other [19]. EAs consist of different components (operators), each with different role for intensification or diversification or both. Usually, each operator in EAs come with some parameters that could be used to change the role of that operator, and consequently the so called balance in the search. This way, the parameters are found to have important effects on performance of the algorithms [9, 24, 27].

The literature on parameter adjusting could be divided into two main approaches [23] as follows. The first approach is known as *parameter assignment*, where values for the parameters will be provided using different methodologies. Whereas, in the second approach, which will be referred to as *parameter-reduced*, removal of the parameters is of interest.

Adaptive GA in a Partitioned Search Space (AGAPSS) is a new EA that assign the parameters to portions of the search space. It divides the search space into a predefined number of regions and tries to search the regions with carefully designed intensification and diversification operators. Two parameters are assigned to each partition of the search space holding the intensification and diversification rates. The probabilities for intensification or diversification of each region will be determined from the behaviour of the regions. The region's behaviour is a function of overall fitness of the region and the number of visits from that region. The population will be distributed among the regions depending on their behaviour. The designed operators for intensification and diversification are working based on the probability vectors that holds the probability of the alleles on each loci for each individual. The probability vectors will be updated based on the changes in fitness of an individual. The AGAPSS iteratively evolves until it reaches a stopping criteria. Algorithm 1 shows a pseudo code of the proposed method.

Algorithm 1 Pseudo code for the main loop of the proposed method.

while stopping criteria is not met **do**
  for $i = 0$ to m //m is the number of regions in the search space See § 3.2 **do**
    if $U(0,1) < G_i.\mu_D$ **then**
      //$U(0,1)$ returns a random generated number between $[0,1]$ based on the uniform distribution
      Diversify(I_1, I_2) $I_1, I_2 \in G_i$
    end if
    if $U(0,1) < G_i.\mu_I$ **then**
      Intensify(I_1) $I_1 \in G_i$
    end if
    Update related regions informations //See sub-section 3.2
    Update: $G_i.F, G_i.V, G_i.I_B, G_i.I_W$ See § 3.2
    Update: $G_i.\mu_D$. See § 3.2
    Update: $G_i.\mu_I$. See § 3.2
  end for
  Adjust region population sizes. See § 3.2
end while

The remainder of this paper is organised as follows. Section 2 briefly looks into the background and related works. Explanation of the proposed method will be introduced in section 3. How the experiment has been conducted is mentioned in section 4. Experimental results will be reported in section 5. Finally, the last section, that is section 6, will be the conclusion and future works.

2 Background

This section will briefly review the literature on the research focusing on parameters.

2.1 Parameter Calibration Approaches

Categorizing approaches that are dealing with parameters could be done in different ways [8]. To this end, the literature could be divided into two main branches [23], which are parameter calibration methods and parameter-reduced methods.

In parameter calibration approaches, the parameters are actually within the algorithm and different methodologies are used for determining their values. Whereas in parameter-reduced approaches, the effort is towards removal of the parameters of the algorithms. The ideal form would have no parameters within the algorithm.

Parameter Control

This category covers all of the approaches on which the optimal values for the parameters will be provided by different methods. This covers those that try to find the optimal parameter values prior to the run, or the methods that are trying to find the values of the parameters during the run.

REVAC [6], tries to find the parameters of a given EA by refining the possible parameter vectors iteratively. Here, an Estimation of Distribution Algorithm (EDA) [25], is used for finding the best parameter set for a given EA.

In an adaptive GA [20], a combination of static rules, inference engine of fuzzy logic controller and feedback from the algorithm are used for determining the values of the parameters.

Frequency of the best individuals within the population, number of duplicate individuals, and number of expected optimal values are used [4] for determining the parameter's values.

In a self-adaptive method [28], each individual is extended with a part that holds the mutation rate in itself. The global mutation rate will be calculated based on the individual mutation rate and a global value. Similarly, in self-Adaptive GA (SAGA) [1], every individual is extended with an extra bit (μ) holding the mutation rate. On each update, the new mutation rate (μ') will be derived based on the previous rate (μ).

Hybrid Self-adaptive GA (HSGA) [7], add an extra gene to the end of the chromosomes, representing the size of tournament. In this method, the less fit individuals get less selection pressure while fitter individuals get higher selection pressure.

Lobo in his Ph.D. thesis [18] addressed all of the parameters of the GA, either automatically (population size) or rationally (selection rate and crossover rate) based on theoretical foundations. An extension of this work [17] integrates local search with the previous work. It has been shown that the use of local search for exploitation of the search space is beneficial.

Parameter-reduced Approaches

This approach covers all of those methods where even one parameter is removed from the algorithm. Generally, in parameter-reduced GAs, the population and variation operators will be replaced with probabilistic model representation and generation models [2, 21, 26].

Compact GA (cGA) was proposed as a variation of GA where population is represented as a probability distribution over the set of solutions [12]. It processes each gene independently while it tries to create the same distribution like the previous population.

In a canonical GA referred to as SSRGA [32] crossover operator is removed from the algorithm, and mutation is replaced with another customised operator. In this method, firstly, the population will be divided into a set of sub-populations, secondly, a site-specific rate vector will be calculated from each sub-populations, and finally each sub-population will be mutated using its corresponding site-specific vector.

The extended version of the SSRGA, referred to as SSRGA-II [31], explores and exploits the search space simultaneously using a different strategy used in SSRGA. The algorithms will select top m individuals from the population and a position frequency matrix will be constructed based on them. Each component of this matrix refers to the probability of occurrence of allele i in site j. The probability for each site will be calculated proportionate to the fitness of the individuals.

3 Proposed Method

In the proposed method, the search space will be divided into regions. Each region is given two probability rates, one for diversification and the other one for intensification. In the beginning of the search, these probabilities will be initialised randomly. In other words, some of the regions will have more chance for either diversification or intensification. However, these probabilities will be updated as the search progresses. The probabilities of the regions change over time depending on the behaviour of the regions, which in turn will be determined by a rank given to each region. The rank of each region is with regards to its overall fitness and the number of visits that have been done from that region. Population of each region will also change over time, though in the beginning of the search, the same number of individuals will be assigned to each region. The rank of the region will also be used to determine the eligibility of a given region for attracting more individuals. Regions will be prioritised based on their eligibility and individuals will be assigned to regions proportional to their eligibility. Each individual is assigned with a counterpart probability vector, which holds the probability of having allele 1 for each loci of a given individual. The occurrence probabilities of the alleles will get updated depending on the changes in the fitness and the number of changes in each individual.

The AGAPSS is mainly inspired from PGA [23] however, these two methods differ in some aspects including the variation operators. Besides that, the AGAPSS is an adaptive method due to utilization of parameter for intensification and diversification.

The remaining of this section is organized as follows. First, a formal definition of the concepts that have been used in this research will be introduced in section 3.1. Details on dividing the search space and management of the divisions will be explained in section 3.2.

3.1 Terminology

The formal definitions of the concepts used in the algorithm will be introduce in this section. For an individual (I),

$$I = i_1, ..., i_l \quad | \quad i_k \in A, 1 \leq k \leq l,$$

where l is the length of individual, and $A = \{0, 1\}$ is the alphabet of each locus.

There would be a probability vector (Z) for every individual,

$$\forall \;\; I_i \quad \exists \quad Z_i \;, \;\; 1 \leq i \leq N,$$

where N is the population size, and

$$Z = z_1, ..., z_l \quad | \quad z_j \in [0, 1], 1 \leq j \leq l,$$

where z_j holds the probability of 1 in j^{th} locus of individual (I).

Depending on the number of changes (τ) that have occurred on a given individual, its respective probability vector (Z) will get updated [23] using,

$$\acute{Z} = Z + R.$$

Where $R \subseteq Z$ is defined as in Eq. 1 and $\acute{Z}$ is the updated probability vector. Initially all of the Z vectors will be initialized, in all of their loci, with 0.5. Each locus in Z vector will be bounded between 0 and 1 This way, equal chance will be given to each loci for being either 0 or 1. As the search progresses, the Z vectors will get updated based on Eq. 1, which is with regards to Table 1. According to Eq. 1 only those loci in Z vector will be updated that their corresponding loci in individuals have been changed. The locations of the changed loci will be determined using another vector, U, (see eq. 2) which will be constructed by applying logical *exclusive OR* operator on a given individual before and after changes.

In Eq. 1, $\epsilon = 1$ is the learning factor, $f : I \rightarrow \mathbb{R}$ is assumed to be the objective function of a maximization problem,

$$\tau = \sum_{i=1}^{l} u_i, \quad U = u_1, ..., u_l | u_j = i_j \oplus \acute{i_j}, 1 \leq j \leq l \tag{2}$$

where $\oplus$ is bitwise *exclusive OR* operator.

$$R = r_1, ..., r_l | 1 \leq j \leq l, \quad r_j = \begin{cases} \frac{\epsilon}{\tau} \cdot u_j & (f(I) > f(\acute{I}) \wedge i_j = 0) \vee (f(I) < f(\acute{I}) \wedge i_j = 1) \\ -\frac{\epsilon}{\tau} \cdot u_j & (f(I) < f(\acute{I}) \wedge i_j = 0) \vee (f(I) > f(\acute{I}) \wedge i_j = 1) \\ 0 & otherwise \end{cases} \quad (1)$$

Table 1: Probability vector updates is based on the change in fitness prior and after the changes on individual. Depending on the fitness, the probability vector will get positive or negative credit.

Change in locus		Fitness	Credit
From	To		
0	1	Improved	+
0	1	Worsen	-
1	0	Improved	-
1	0	Worsen	+

As an example, assuming the OneMax function as optimization problem on which the number of ones within the individual will be the fitness. For an individual $I = \{1, 1, 0, 0, 1, 0, 0, 1\}$, the fitness will be 4. Let assume this individual is changed to $I' = \{1, 1, \mathit{1}, \mathit{1}, \mathit{0}, 0, 0, 1\}$, the new fitness will be 5. Here the fitness improved after the changes. Accordingly, $U = \{0, 0, 1, 1, 1, 0, 0, 0\}$ and $\tau = 3$. Consequently, $R = \{0, 0, 1/3, 1/3, -1/3, 0, 0, 0\}$. Therefore, if we assume before the update $Z = \{0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5\}$, after the changes it will be updated to $Z' = \{0.5, 0.5, \mathit{0.83}, \mathit{0.83}, \mathit{0.17}, 0.5, 0.5, 0.5\}$.

3.2 Dividing the Search Space

Initially, the search space will be divided into a predefined number (m) of regions (G) with the same size [16, 18, 22].

Representation

A set of loci in each individual will be reserved as indicators of regions [22]. It has to be noted that no extra genes will be added to the individuals. Each individual in the population will be member of a region, formally,

$$I_i \in G_j | 1 \leq i \leq N, 0 \leq j \leq m - 1. \quad (3)$$

where N is population size, and m is the number of regions in the population,

$$m = |A|^x, x \geq 1 \quad (4)$$

where x is the number of loci in the individual that is needed to be reserved for the region indicator, and $|A|$ is the cardinality of loci.

Binary representation of the region index (i), in the first x loci of an individual, will be used for indication of the members of the i^{th} region (G_i). This way the search space could be divided into equal sized partitions.

As an example, assume the case of dividing the search space into $m = 4$ regions. Referring to equation 4, the number of loci needed to be reserved is $\log_{|A|} m$ which in this case will be $x = 2$. Consequently, values of the first two loci of each individual, will be set to either of $00, 01, 10, or 11$, that refers to the regions 0, 1, 2, and 3 respectively. A sample population of this example is shown in Figure 1. Here, three individuals will be assigned to each region assuming $N = 12$.

Region	I					f	F	V	δ	γ	P_I	P_D	r_I	r_D	χ	N
G0	**0**	**0**	1	1	0	2	1.67	3	$\overline{H}$	$\overline{H}$	3	1	4	3	1	2
	0	**0**	0	1	1	2										
	0	**0**	0	0	1	1										
G1	**0**	**1**	0	0	1	2	3	3	$\overline{H}$	$\overline{H}$	3	1	2	1	3	3
	0	**1**	1	0	1	3										
	0	**1**	1	1	1	4										
G2	**1**	**0**	0	1	1	3	2.33	3	$\overline{H}$	$\overline{H}$	3	1	3	2	2	3
	1	**0**	0	1	0	2										
	1	**0**	1	0	0	2										
G3	**1**	**1**	0	0	0	2	2.67	3	$\overline{H}$	H	1	2	1	4	4	4
	1	**1**	1	1	0	4										
	1	**1**	0	0	0	2										

Figure 1: The very first locus on each individual (highlighted in bold) will be used to indicate the regions. The onemax function is assumed as optimization problem in this example.

Region Parameters

As mentioned earlier, each region is assigned with two rates, one for intensification (μ_D) and the other for diversification (μ_I). In the beginning of the search, these probabilities will be set randomly for all of the regions. However, these rates change adaptively over time (iteration), depending on the ranks of the regions. The diversification rate will be adjusted using,

$$\acute{\mu_D}(i) = \mu_D(i) + 0.1(r_D(i) - \acute{r_D}(i)), \tag{5}$$

where, $\acute{r_D}(i)$ and $r_D(i)$ are diversification ranks of region G_i at previous and current iterations respectively. On the other side, changes in intensification rates will be based on,

$$\acute{\mu_I}(i) = \mu_I(i) + 0.1(r_I(i) - \acute{r_I}(i)), \tag{6}$$

where $\acute{r_I}(i)$ and $r_I(i)$ are intensification ranks of region G_i at previous and current iterations respectively. In both of the cases, the rates are bounded between 0.1 and 0.9 to prevent either neglect or over looking into a region.

For example, assuming that in the beginning of the search, the diversification rank ($r_D(1)$) for the first region is 3, and diversification rate ($\mu'_D(1)$) is 0.10. Also assume that in the subsequent generation, the diversification rank ($r_D(1)$) will be changed to 2. Based on Eq. (5) the diversification rate for the first region will change to 0.20.

Rank of the Regions

As mentioned, the regions' parameters will be updated based on the ranks of the regions. The rank of the regions will be determined with regards to the priority and the fitness of the regions.

The priorities of the regions will be determined based on the fitness of the regions and the number of visits that have been done from them. To this end, two behavioural statuses will be given to any given region, one for fitness and the other for the number of visits from a region.

If the fitness value of a given region (See Eq. (7)) was larger than mid-range value on that region, its status will be considered as high (H), otherwise it will be determined as not high ($\overline{H}$).

$$\gamma(i) = \begin{cases} H & G_i.F > \frac{f(G_i.I_B)+f(G_i.I_W)}{2} \\ \overline{H} & \text{otherwise.} \end{cases} \tag{7}$$

The number of visits from a given region (see Eq. 8), will be determined as high (H) if the mid-range value of the number of visits among all of the regions is lower than the number of the visit from the given region.

$$\delta(i) = \begin{cases} H & G_i.V > \frac{\underset{0\le i\le m-1}{\text{argmax}}\ \text{G}_\text{i}.\text{V}+\underset{0\le i\le m-1}{\text{argmin}}\ \text{G}_\text{i}.\text{V}}{2} \\ \overline{H} & \text{otherwise} \end{cases}. \tag{8}$$

As the first step, priorities of the regions will be retrieved using Eq. (11) or Eq. (12) for diversification or intensification respectively. A region with a low fitness and a relatively high number of visits would get the lowest priority (lowest value in Eq. (11)) for either diversification or intensification. It is because that the expectation of finding a good solution within that region is low. This way, lower attention will be on this kind of regions.

In the next step, ranks of the regions ($r_D(i)$ or $r_I(i)$) will be determined. A list will be constructed where the regions are sorted first based on their priorities and then their fitnesses. The row index of each region in the sorted list will be returned as the ranks for the regions. In other words,

$$\begin{aligned} r_D(i) > r_D(j) &\Leftrightarrow (p_D(i) > p_D(j)) \\ \vee(p_D(i) = p_D(j) \wedge\ & G_i.F > G_j.F) \\ &\mid 0 \le i, j \le m-1. \end{aligned} \tag{9}$$

$$\begin{aligned} r_I(i) > r_I(j) &\Leftrightarrow (p_I(i) > p_I(j)) \\ \vee(p_I(i) = p_I(j) \wedge\ & G_i.F > G_j.F) \\ &\mid 0 \le i, j \le m-1. \end{aligned} \tag{10}$$

where in both Eq. 9 and Eq. 10, $\vee$ is logical OR operator, and $\wedge$ is logical AND operator.

$$p_D(i) = \begin{cases} 2 & (\gamma(i) = H) \wedge (\delta(i) = H) \\ 3 & (\gamma(i) = H) \wedge (\delta(i) = \overline{H}) \\ 1 & (\gamma(i) = \overline{H}) \wedge (\delta(i) = H) \\ 4 & (\gamma(i) = \overline{H}) \wedge (\delta(i) = \overline{H}) \end{cases} \tag{11}$$

$$p_I(i) = \begin{cases} 3 & (\gamma(i) = H) \wedge (\delta(i) = H) \\ 4 & (\gamma(i) = H) \wedge (\delta(i) = \overline{H}) \\ 1 & (\gamma(i) = \overline{H}) \wedge (\delta(i) = H) \\ 2 & (\gamma(i) = \overline{H}) \wedge (\delta(i) = \overline{H}) \end{cases} \tag{12}$$

As another example, the overall fitness (F) of the regions of the example population, and the number of visits from each population are also reported in Figure 1 on page 330. Consequently, δ, and γ are calculated for all of the regions. Using the required information, the diversification and intensification priorities and also the ranks for the regions (according to Eq. (8) and Eq. (7)) are also calculated and shown in Figure 1.

Region Population

As mentioned earlier, each region will be instantiated with the same number of individuals. However, during the progress of the search, the number of individuals within each region varies. It has to be noted that the population size is fixed and it is the membership of individuals that changes.

The population of regions changes with proportion to the diversification rank of the regions. The lower is the diversification rank of a given region, the higher is the eligibility (χ) of that region for attracting new individuals. Formally,

$$\chi_i = m - r_D(i) + 1. \tag{13}$$

In each iteration of the run, the new population size of each region will be calculated using,

$$G_i.N = G_i.N_{min} + \left\lfloor \frac{\chi_i}{\sum_{1 \leq i \leq m} \chi_i} G_i.N_{max} \right\rfloor ,$$

where $N_{max} = N - N_{min}(m-1)$.

The extra individuals of each region will be distributed among the other regions with proportionate to their eligibility.

The eligibility rank and population size of the regions for the sample population in the previous example are shown in Figure 1.

4 Experimental setup

Three different test functions are chosen from two different family of problems. Selection of the compared algorithms was based on their relevance to the proposed method. The test functions are the same as those that have been used by the chosen algorithms for comparison [1,23,30–32]. This facilitate, the comparison of the results with other methods in literature.

To come up with a fair experiment, the algorithms are tuned as it was reported in literature, which is reported in Table 2.

Table 2: Attributes of the comparing algorithms

Feature	epistatic	multimodal	MPG
Population model	steady state		
Parent selection	Tournament selection		
Survival selection	delete oldest		
Selection pressure	8	8	100
Population size	50	50	100
Max. no. of generations	500	100	10000
Chromosome length	based on the epistasis level	30	100
Number of regions (m)	8		
Region population size	8		

The performance of each algorithm is measured over 50 independent runs for each of the problems. The average (*avg.*) of the best results are derived for all of the experiments. However, standard deviation (*stdev.*) of the results are reported when the compared methods have reported them. Best average is highlighted in **bold**, while the lowest standard deviation is <u>underlined</u> in each case.

The following will be the explanation on the utilized test problems for the course of experiments.

4.1 MAX-SAT problems

MAX-SAT is a generalized version of satisfiability (SAT) decision problem that belongs to the family of NP-hard optimization problems.

The problem in SAT is to find if there exists any assignment for the variables that satisfies the following formula, which is in Conjunctive Normal From (CNF),$C_1 \wedge C_2 \wedge \ldots \wedge C_n$.. However, in MAX-SAT the problem is to find an assignment which maximises the number of satisfied clauses [5, 10, 11].

Multi-modal boolean satisfiability and epistatic problems are the two classes of benchmarks used in the comparisons. These problems will be constructed based on the MAX-SAT problems. Degree of multi-modality in a problems could be defined as a measure of difficulty of the problem. In these problems, the number of false peaks grow with the number of modality. While, the degree of epistasis of a problem expresses the relationship between the genes in a chromosome. Dependency of a large number of alleles at other loci is a sign of high epistasis in a system [29]. The degree of epistasis has direct relation with difficulty level of the problem.

The compared algorithms are canonical GA (CGA) [32] with a randomly chosen constant mutation rate, SSRGA [32], *self-adaptive (SAGA)* [1], *adaptive (AGA)* [30] parameter control methods, Compact GA (cGA) [12] and a proposed method, [23] which will be referred to as $PGACMCO$.

Epistatic problems

Spears [29] has introduced a method for the creation of epistatic problems using the boolean expressions. His proposed method could be tuned for different levels of epistasis. The method is based on conversion of the Hamiltonian Circuit (HC) problems into equivalent SAT expressions. In a directed HC problem, the aim is to find if a given graph has a Hamiltonian cycle. By definitions, a Hamiltonian cycle is a cycle in a graph that meet all of the vertex exactly once. The definition of HC constrains the nodes of feasible solutions to have only one input edge and one output edge. As such, any tour that does not satisfies this constraint cannot be a solution [14].

MAX-SAT Multi-modal problems

The multi-modal problems used for the course of experiment are created using a mechanism proposed by Spears [29].

A uni-modal problem of length 30 could be created as follows,

$$1Peak \equiv (x_1 \wedge x_2 \wedge \ldots \wedge x_{30}).$$

Using the above uni-modal problem, a bimodal problem will be,

$$2Peak \equiv 1Peak \vee (x_1 \wedge \bar{x}_1 \wedge \bar{x}_2 \wedge \ldots \wedge \bar{x}_{30}).$$

For the course of this experiments, multi-modal problems up to 5 peaks have been used.

4.2 Multi-modal Problem Generator

Multi-modal Problem Generator (MPG) is another benchmark, proposed by Spears [15, 29]. The benchmark has the ability to create problems with multiple peaks where the number of peaks defines level of problem difficulty. For all of the peaks we have used, the heights are linearly distributed and lowest height is 0.5. Fitness value of a given individual will be calculated according to the Hamming Distance (HD) between the individual and the nearest peak [33].

The compared algorithms are Compact GA (cGA), [12] the PGA , [23] GASAT [9], GAHSAT [9], hand-tuned [9], meta-GA [6] and REVAC [6]. It has to be noted that the hand-tuned algorithm here is a canonical GA on which the values of the parameters are carefully tuned.

5 Results and discussion

The Mean of Best Fitness (MBF) [8] for all of the algorithms are reported. The MBFs of the MPG function reported in Table 3 show that the AGAPSS has obtained better performance in all of the instances of this function in comparison to the other compared methods.

Table 3: Comparison of the AGAPSS with compared methods over MPG benchmark in [6, 7]

Peaks	AGAPSS	GASAT	GAHSAT	GA	meta-GA	REVAC	cGA	PGA	SSRGA-II
1	1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.959
2	1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.958
5	1	1.0	1.0	1.0	0.988	1.0	1.0	1.0	0.962
10	**1**	0.9956	0.9939	0.9961	0.993	0.996	0.9989	0.9994	0.960
25	**0.9992**	0.9893	0.9879	0.9885	0.994	0.991	0.9863	0.9933	0.963
50	**0.9988**	0.9897	0.9891	0.9876	0.994	0.995	0.9794	0.9896	0.962
100	**0.9976**	0.9853	0.9847	0.9853	0.983	0.989	0.9847	0.9909	0.965
250	**0.9969**	0.9867	0.9850	0.9847	0.992	0.966	0.9739	0.9879	0.969
500	**0.9949**	0.9834	0.9876	0.9865	0.989	0.970	0.9673	0.9885	0.967
1000	**0.9935**	0.9838	0.9862	0.9891	0.987	0.985	0.9649	0.9884	0.959

The results of the MAX-SAT epistatic problem are reported in Table 4. The rustles are very comparable with the results of the cGA even though the AGAPSS has obtained higher MBFs in five instances of the epistasis problems.

Table 4: Comparison of the AGAPSS with benchmark methods [23, 32] over epistatic problem with different levels of epistasis

Deg. of epistasis		N=6	N=11	N=16	N=21	N=26	N=31	N=36	N=41
AGAPSS	avg.	0.986	0.986	0.982	0.965	**0.946**	**0.926**	**0.91**	**0.896**
	stdev.	0.014	0.003	0.005	0.008	0.011	0.0081	0.011	0.012
PGA	avg.	0.991	**0.990**	**0.994**	0.967	0.909	0.869	0.842	0.827
	stdev.	0.019	0.004	0.001	0.005	0.007	0.007	0.006	0.006
SSRGA-II	avg.	0.953	0.926	0.885	0.858	0.839	0.823	0.811	0.805
	stdev.	0.013	0.007	0.013	0.014	0.01	0.007	0.007	0.007
AGA	avg.	**1**	0.96	0.922	0.888	0.865	0.847	0.836	0.826
	stdev.	0	0.007	0.008	0.007	0.006	0.005	0.004	0.004
SAGA	avg.	0.980	0.943	0.904	0.873	0.853	0.837	0.827	0.817
	stdev.	0.019	0.007	0.01	0.006	0.007	0.005	0.004	0.004
CGA	avg.	0.989	0.948	0.906	0.876	0.856	0.840	0.827	0.819
	stdev.	0.017	0.011	0.009	0.007	0.008	0.005	0.005	0.004
Compact GA	avg.	0.985	0.983	0.983	**0.979**	0.945	0.926	0.909	0.894
	stdev.	0.014	0.002	0.002	0.003	0.005	0.006	0.006	0.005

The results of the MAX-SAT multi-modal problem are reported in Table 5. In this benchmark the AGAPSS has performed better in four out of the six algorithms. However, the results of the AGAPSS and the PGA are comparable for all the instances. It has to be noted that this test function as like the MPG function is a multi-modal problem. However, the obtained results of the AGAPSS was far different from what that have been obtained over MPG function.

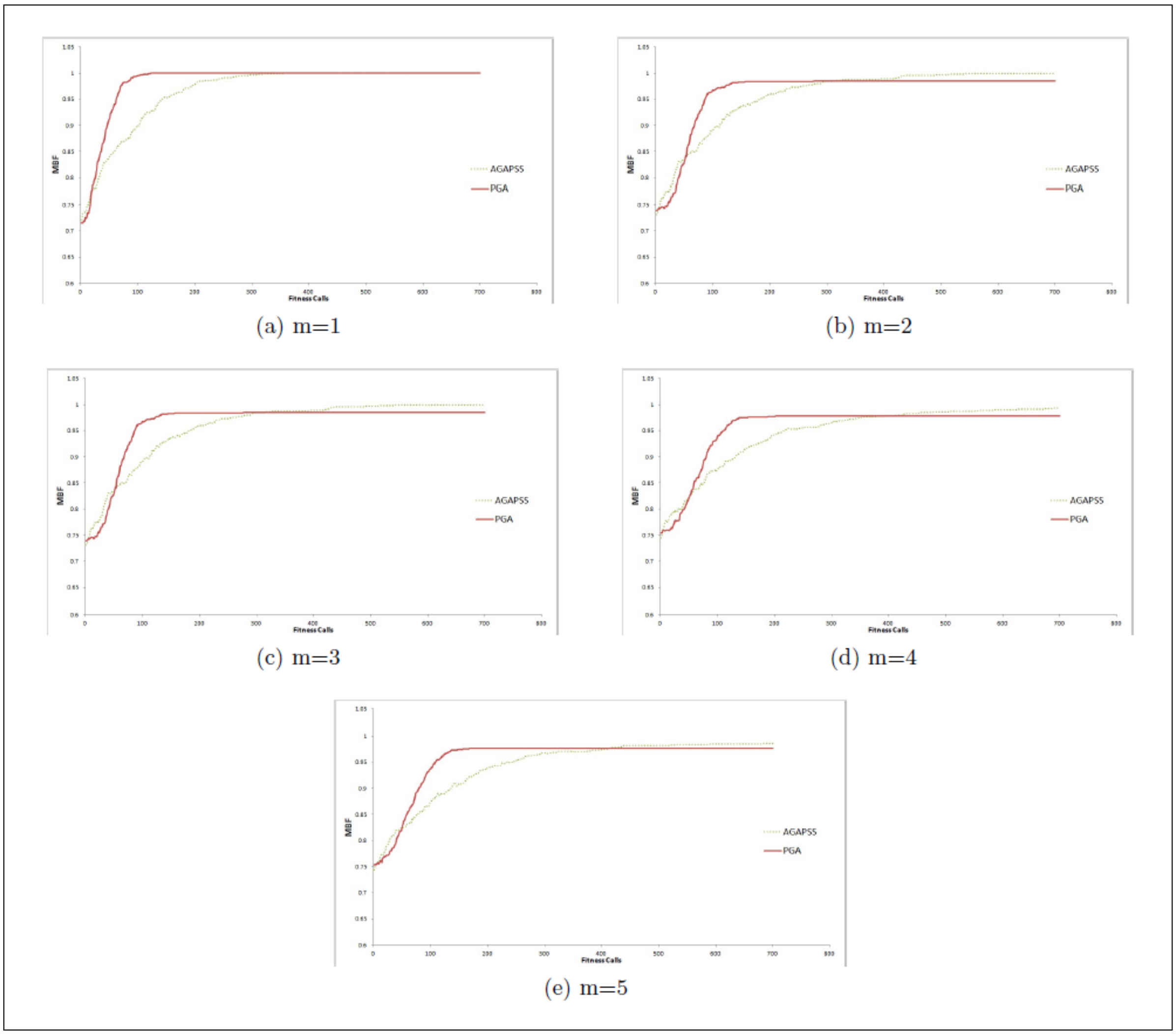
The statistical test reported in table 6 allow systematic comparison of the AGAPSS with the other compared methods. The proposed method is compared pairwise with other compared methods and the results of the statistical are reported for each pair.

The statistical tests confirms that the results of the AGAPSS have been significantly different from all of the other compared methods over MPG test function. According to the obtained ranks of the AGAPSS it could be inferred that the proposed methods has obtained significantly better results in comparison to the other compared methods over this test function.

Considering the obtained rankings and the p-values, the statistical tests confirms the superiority of the proposed method over epistasis test function in four out of the six compared algorithms. Although the AGAPSS has performed better than the PGA and cGA, regarding to the obtained ranks (R^-), however the statistical difference between them was not significant.

The statistical results have confirmed significant difference in favour of the proposed method for five out of six compared algorithm over MAX-SAT multi modal problem. However, the proposed algorithm has performed significantly worse than the PGA method. It would be more interesting to note that two different results have been obtained for the same family of problems. Both of the MAX-SAT multi-modal and MPG functions are multi-modal functions. While the proposed method has performed significantly better than the PGA over MPG method, it has performed worst over MAT-SAT multi-modal problem.

The main difference between the MPG and MAX-SAT multi-modal function is that the MPG function is ran for 10000 fitness calls, while the MAX-SAT problem stopped after 100 fitness calls. It suggest that the AGAPSS might be able to outperform the PGA over higher number of iterations. In order to check the validity of the aforementioned hypothesis, a new experiment has been conducted. Both of the PGA and AGAPSS methods have been ran over MAX-SAT multi-modal problem while the maximum number of fitness calls was set to 10000. Figure 2 depicts the progress of improvement in fitness as the search progress over different instances of MAX-SAT multi-modal function.



(a) m=1 (b) m=2 (c) m=3 (d) m=4 (e) m=5

Figure 2: Comparison of the progress in the performance of the AGAPSS and PGA over the MAX-SAT multi-modal function in.

As it could be seen in all of the instances of the compared algorithms, the AGAPSS have been able to obtain better results in comparison with the PGA method. However, the AGAPSS has obtained the better performance after some iterations. In other words, the performance of the PGA method was better in the early stages of the run. The statistical analyses, on the results of the last four instances of the MAX-SAT multi-modal function, confirmed that the AGAPSS has performed significantly better than the PGA . The performance of the AGAPSS has also been significantly better than PGA in longer iterations over the epistasis problems.

As a result it has been shown that the assignment of the parameters to the parts of the search space could be beneficial. However, it take some generations, in the early stages of the run, for the AGAPSS to conduct the proper balance between the partitions of the search space.

In other words, it takes time for the proposed method to find the promising partition and focus on it. This time would slow down the algorithm from finding the promising area in the search space, however, the accuracy of the proposed method in comparison to the other methods has shown that the proposed method is able in finding significantly better solutions.

Table 5: The AGAPSS in comparison to some of the relevent methods from literature over multi-modal MAX-SAT problem with different levels of multi-modality.

Modality		$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 5$
AGAPSS	avg.	0.936	0.9155	0.9015	0.8996	0.9117
	stdev.	0.023	0.033	0.031	0.031	0.037
PGA	avg.	**0.999**	**0.976**	**0.970**	**0.964**	**0.961**
	stdev.	0.007	0.023	0.025	0.029	0.026
SSRGA-II	avg.	0.871	0.835	0.842	0.835	0.846
	stdev.	0.030	0.040	0.034	0.034	0.029
AGA	avg.	0.874	0.870	0.866	0.868	0.876
	stdev.	0.022	0.029	0.026	0.022	0.025
SAGA	avg.	0.827	0.842	0.840	0.853	0.846
	stdev.	0.029	0.029	0.022	0.027	0.024
CGA	avg.	0.834	0.843	0.848	0.850	0.842
	stdev.	0.029	0.029	0.022	0.027	0.024
cGA	avg.	0.8360	0.7642	0.7909	0.7729	0.7794
	stdev.	0.0386	0.0274	0.0347	0.0292	0.0280

6 Conclusions and future works

A new adaptive algorithm is proposed that divides the search space into predefined number of regions. Different search strategies will be applied on different regions based on ranks that are given to each region based on their behaviour. The behaviour of a regions will be determined in relation to the overall fitness of the regions and the number of visits that have been done from them.

The proposed method has been examined using different instances of three benchmark problems: MAX-SAT multi-modal problems, MAX-SAT epistatic problems, and MPG problems.

The performance of the proposed method shown to be superior over the MPG in comparison to all of the compared methods. However, the results are negotiable with the MAX-SAT multi-modal test function. The proposed method shown to be able to perform better than the PGA when the maximum number of fitness calls increased. Ignoring the first instance of this benchmark, the statistical analysis shown significant difference The results over the epistasis problem has also shown that the AGAPSS was significantly better in compare to the most of the compared methods. Overall, based on the results of the benchmarks, the performance of the proposed method is superior compared to the compared algorithms.

Table 6: The Wilcoxon signed-rank test between the AGAPSS and other algorithms. p-values below 0.05 are shown in bold.

AGAPSS vs.		multi-modal	epistatic	MPG
PGA	R^-	0	26	53.5
	R^+	15	10	1.5
	p-value	0.043*	0.263**	**0.012**
SSRGA-II	R^-	15	36	55
	R^+	0	0	0
	p-value	**0.043**	**0.012**	**0.005**
AGA	R^-	15	35	
	R^+	0	1	
	p-value	**0.043**	**0.017**	
SAGA	R^-	15	36	
	R^+	0	0	
	p-value	**0.043**	**0.012**	
Hand-tuned	R^-	15	35	52
	R^+	0	1	3
	p-value	**0.043**	**0.017**	**0.018**
Compact GA (cGA)	R^-	15	25	52
	R^+	0	11	3
	p-value	**0.043**	0.327**	**0.018**
GASAT	R^-			52
	R^+			3
	p-value			**0.018**
GAHSAT	R^-			52
	R^+			3
	p-value			**0.018**
meta-GA	R^-			53.5
	R^+			1.5
	p-value			**0.012**
REVAC	R^-			52
	R^+			3
	p-value			**0.018**

** : No significant difference between the AGAPSS and the compared algorithm.

* : Significant difference between the AGAPSS and the compared algorithm in favour of the compared algorithm.

Bibliography

[1] Bäck, T., Schutz, M. (1996). Intelligent mutation rate control in canonical genetic algorithms. *In Foundations of Intelligent Systems*, 158-167.

[2] Baluja, S. (1994). Population-based incremental learning: A method for integrating genetic search based function optimization and competitive learning. Science, *CMU-CS-94-(CMUCS- 94-163)*:1-41.

[3] Blum, C., Roli, A. (2003). Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM Comput. Surv.*, 35(3):268-308.

[4] Houat de Brito,F., Noura Teixeira, A., Noura Teixeira, O., Oliveira R.C.L. (2007). A fuzzy approach to control genetic algorithm parameters. *SADIO Electronic Journal of Informatics and Operations Research*, 1(1):12-23.

[5] Cook, S.A. (1971). The complexity of theorem-proving procedures. *Proceedings of the third annual ACM symposium on Theory of computing, STOC 71*, 151-158.

[6] de Landgraaf, W.A., Eiben, A.E., Nannen, V. (2007). Parameter calibration using metaalgorithms. *In Evolutionary Computation*, 71-78.

[7] Eiben, A.E., Schut,M.C., de Wilde, A.R.(2006). Boosting genetic algorithms with (self-) adaptive selection. *Proceedings of the IEEE Conference on Evolutionary Computation (CEC 2006)*, 1584-1589.

[8] Eiben, A.E., Smith, J.E. (2007). *Introduction to Evolutionary Computing.* Natural Computing Series. Springer, 2 edition, 2007.

[9] Eiben, G., Martijn C. Schut, M.C. (2008). New ways to calibrate evolutionary algorithms. In *Patrick Siarry and Zbigniew Michalewicz, editors, Advances in Metaheuristics for Hard Optimization, Natural Computing Series*, 153-177.

[10] Garey, M.R., Johnson, D.S. (1990). *Computers and Intractability; A Guide to the Theory of NP-Completeness.* W. H. Freeman & Co., New York, NY, USA, 1990.

[11] Gu, J. (1993). Local search for satisfiability (sat) problem. Systems, Man and Cybernetics, *IEEE Transactions on*, 23(4):1108-1129, July/August 1993.

[12] Harik, G.R., Lobo, F.G., Goldberg, D.E. (1999). The compact genetic algorithm. *Evolutionary Computation, IEEE Transactions on*, 3(4):287-297.

[13] De Jong, K.E. (2006). *Evolutionary Computation. A Unified Approach.* MIT Press, 2006.

[14] De Jong, K.E., Spears, W.M. (1989). *Using genetic algorithms to solve np-complete problems.* In J. David Schaffer, editor, ICGA, 124-132. Morgan Kaufmann, 1989.

[15] De Jong, K.E., Spears, W.M. (1992). *A formal analysis of the role of multi-point crossover in genetic algorithms*, 1992.

[16] Kuo, T. Hwang, S.-Y. (1996). Why DGAs work well on ga-hard functions? *New Gen. Comput.*, 14:459-479.

[17] Lobo F.G., Goldberg, D.E. (2004). The parameter-less genetic algorithm in practice. *Information Sciences*, 167(1-4):217-232.

[18] Lobo F.G., (2000). *The Parameter-Less Genetic Algorithm: Rational and Automated Parameter Selection for Simplified Genetic Algorithm Operation.* PhD thesis, 2000.

[19] Lozano, M., Garcia-Martinez, C, (2010). Hybrid metaheuristics with evolutionary algorithms specializing in intensification and diversification: Overview and progress report. *Comput. Oper. Res.*, 37:481-497, March 2010.

[20] McClintock,S., Lunney, T., Hashim, A. (1997). A fuzzy logic controlled genetic algorithm environment. *Systems, Man, and Cybernetics*, 3:2181-2186.

[21] Muhlenbein, H., Mahnig, T., Ochoa Rodriguez, A. (1999). Schemata, distributions and graphical models in evolutionary optimization. *Journal of Heuristics*, 5:215-247.

[22] Nadi, F., Khader, A.T. (2010). Managing search in a partitioned search space in GA. *Cybernetics and Intelligent Systems (CIS), 2010 IEEE Conference on*, 114-119.

[23] Nadi, F., Tajudin, A., Khader, A.T. (2011). A parameter-less genetic algorithm with customized crossover and mutation operators. In *Natalio Krasnogor and Pier Luca Lanzi, editors, ACM, GECCO*, 901-908.

[24] Nannen, V., Smit,S.K., Eiben, A.E. (2008). Costs and benefits of tuning parameters of evolutionary algorithms. *Proceedings of the 10th international conference on Parallel Problem Solving from Nature*, 528-538.

[25] Pelikan, M., Goldberg,D.E., Lobo F. (1999). A survey of optimization by building and using probabilistic models. *IlliGAL Report No. 99018, Illinois Genetic Algorithms Laboratory*, University of Illinois at Urbana-Champaign, Urbana, IL, 1999.

[26] Sastry, K., Goldberg, D.E. (2000). On extended compact genetic algorithm. Technical report, *GECCO-2000*, late breaking papers, Genetic And Evolutionary Computation Conference.

[27] Smit, S.K., Eiben, A.E. (2009). Comparing Parameter Tuning Methods for Evolutionary Algorithms. *IEEE Congress on Evolutionary Computation (CEC)*, 399-406.

[28] Smith, J., Fogarty, T.C. (1996). Adaptively parameterised evolutionary systems: Selfadaptive recombination and mutation in a genetic algorithm. *of Lecture Notes in Computer Science*, 1141:441-450.

[29] Spears, W.M. (2000). *Evolutionary Algorithms: The Role of Mutation and Recombination (Natural Computing Series)*. Springer, June 2000.

[30] Srinivas, M., Patnaik, L.M. (1994). Adaptive probabilities of crossover and mutation in genetic algorithms. *IEEE Transactions on Systems, Man, and Cybernetics*, 24(4):656-667.

[31] Vafaee, F. (2011). *Controlling Genetic Operator Rates in Evolutionary Algorithms*. PhD thesis, Dept. Computer Science, 2011.

[32] Vafaee, F., Nelson P.C. (2010). An explorative and exploitative mutation scheme. *IEEE Congress on Evolutionary Computation*, 1-8.

[33] Vajda, P., Eiben, A.E., Hordijk, W. (2008). Parameter control methods for selection operators in genetic algorithms. *Proceedings of the 10th international conference on Parallel Problem Solving from Nature*620-630.

INT J COMPUT COMMUN, ISSN 1841-9836
9(3):340-347, June, 2014.

An Efficient Solution for the VRP by Using a Hybrid Elite Ant System

M. Yousefikhoshbakht, F. Didehvar, F. Rahmati

Majid Yousefikhoshbakht, Farzad Didehvar*, Farhad Rahmati
Department of Mathematics and Computer Science
Amirkabir University of Technology
No.424, Hafez Avenue, Tehran 15914, Iran
e-mail: khoshbakht@aut.ac.ir, frahmati@aut.ac.ir
*Corresponding author: didehvar@aut.ac.ir

Abstract: The vehicle routing problem (VRP) is a well-known NP-Hard problem in operation research which has drawn enormous interest from many researchers during the last decades because of its vital role in planning of distribution systems and logistics. This article presents a modified version of the elite ant system (EAS) algorithm called HEAS for solving the VRP. The new version mixed with insert and swap algorithms utilizes an effective criterion for escaping from the local optimum points. In contrast to the classical EAS, the proposed algorithm uses only a global updating which will increase pheromone on the edges of the best (i.e. the shortest) route and will at the same time decrease the amount of pheromone on the edges of the worst (i.e. the longest) route. The proposed algorithm was tested using fourteen instances available from the literature and their results were compared with other well-known meta-heuristic algorithms. Results show that the suggested approach is quite effective as it provides solutions which are competitive with the best known algorithms in the literature.
Keywords: Vehicle Routing Problem (VRP), elite ant system, , global and local updating, NP-hard problems.

1 Introduction

The vehicle routing problem (VRP) is one of the most important combinational optimization problems that was first defined by Dantzig and Ramser more than 50 years ago [1]. The VRP involves designing a set of vehicle routes each of which starts and ends at a depot. These routes are used for a fleet of vehicles which provide services for a set of customers with known demands. Each customer is visited by exactly one vehicle only once and the total demand of a route does not exceed the capacity of the vehicle type assigned to it. The objective is to minimize the total distance traveled by all the vehicles. To make VRP models more realistic and applicable, there are various forms of the VRP obtained by adding constraints to the basic model. Examples of such extensions are VRP with pickup and delivery (if the vehicles need to pick up and delivery) [2], VRP with time windows (if the services have time constraint) [3], heterogeneous fleet OVRP (if the capacity of vehicles in OVRP are different) [4], VRP with backhauls (if the customers with delivery demand have to be visited before by the customers with a pickup demand) [5] and generalized VRP (if the customers are partitioned into clusters with given demands such that exactly one customer from each cluster should be visited) [6].
The VRP solution methods fall into three main categories: exact methods, heuristic and meta-heuristic algorithms. Exact approaches for solving the VRP are successfully used only for relatively small problem sizes but they can guarantee optimality based on different techniques. These techniques use algorithms that generate both a lower and an upper bound on the true minimum value of the problem instance. If the upper and lower bound coincide, a proof of optimality is achieved. Branch-and-bound [7] and branch-and-cut [8] are two of exact methods which have

been proposed for VRP by researchers.
Since this problem is known to be NP-hard (Non-deterministic Polynomial-time Hard) problem [1], exact algorithms are not often suitable for real instances because of the computational time required to obtain an optimal solution. Therefore, in the past forty years, researchers have developed the heuristic algorithms using a permissible solution instead of the optimal solution. There are many celebrated algorithms in this class such as savings heuristic of Clarke and Wright [9] that gains a single route instead of two routes according to the savings obtained by this merger. The sweep algorithm is another famous construction heuristic that is proposed by Gillett and Miller [10].
A new kind of algorithm which basically tries to combine basic heuristic methods in higher level frameworks aimed at efficiently exploring a search space is meta-heuristics. Since the meta-heuristic approaches are very efficient for escaping from local optimum, they are one of the best group algorithms for solving combinatorial optimization problems. Some of the most popular metaheuristics applied to the VRP are simulated annealing (SA) [11], genetic algorithm (GA) [12], tabu search (TS) [11], Computational Intelligence Approach [13], large neighborhood search [3], ant colony optimization (ACO) [14], hybrid ant colony optimization [15] and particle swarm optimization [16].
The VRP is intrinsically a multiple objective optimization problem (MOP) in nature that has received much attention because of its practical application in industrial and service problems [17]. On the other hand, there are few research studies which have used the ACO in order to solve VRP. Furthermore, the elite ant system (EAS) is one of the most important and powerful versions of ACO that nowadays is applied on a lot of combinatorial optimization problems [18]. Therefore, this paper proposes a hybrid EAS (HEAS) mixed with insert and swap algorithms for solving the VRP.
The remaining parts of the paper are organized as follows. Section 2 describes the EAS, and the proposed algorithm. Section 3 describes computational experiments carried out to investigate the performance of the proposed algorithm. Finally, section 4 presents the results of the conclusions and future works.

2 Our Algorithm

In this section, first, the EAS are presented and then the proposed algorithm will be analyzed in more detail.

2.1 Elite Ant System

The first modification to ant system (AS) which was conducted by Dorigo and his colleagues in 1996 was using elite strategy in EAS [19]. The decision for choosing the unvisited N_i node by ant k located in node i is made based on formula (1) where τ_{ij} indicates the amount of pheromone on (i, j) edge while η_{ij} shows inverse distance between i and j. However, both are powered by α and β which can be changed by the user. Therefore, their relative importance can be altered.

$$P_{ij}^k = \begin{cases} \frac{{\tau_{ij}}^{\alpha}{\eta_{ij}}^{\beta}}{\sum\limits_{j\in N_i} {\tau_{ij}}^{\alpha}{\eta_{ij}}^{\beta}} & if \quad j \in N_i \\ 0 & if \quad j \notin N_i \end{cases} \tag{1}$$

In the EAS, in addition to depositing pheromone on all local edges, in each iteration pheromone is released on the edges of the best path. Imagine T^{gb} is the best path gained after the algorithm is completed, T_k is total length of edges traversed by ant k and ρ is the rate of the pheromone evaporation in order to prevent rapid convergence of ants to a sub-optimal path.

While pheromone is being updated, the edges marched by the ant which have constructed the T^{gb} path absorb an additional amount of pheromone which is equal to $e/L^{gb}(t)$. The formula for pheromone updating can be written in (2). In this way, the edges of the shortest path up to the current iteration become more attractive and are updated based on the value of the best $L^{gb}(t)$ tour. As it is shown, $1/L^k(t)$ is the formula for the local trail updating. While traversing between nodes i and j, ants release pheromone on the respective edge which is equal to the inverse of the cost of the tour $L^k(t)$ taken by ants.

$$\tau_{ij}(t+1) = (1-\rho).\tau_{ij}(t) + \sum_{k=1}^{m} 1/L^k(t) + e/L^{gb}(t) \tag{2}$$

2.2 The Proposed Algorithm

In EAS, local pheromone release is not a good guide for finding the best route because there might be some edges which belong to no best route in none of the iterations but pheromone is still deposited on them in each iteration. Therefore, local pheromone release must be abandoned so that the ants resort to global pheromone release in finding new solutions. This attracts the attention of ants to the edges which belong to the best route ever found. Furthermore, the accuracy of solutions in EAS is low at the beginning but it increases with the iterations of the algorithm and pheromone release. Therefore, constant e coefficient cannot be a suitable formula for encouraging the best path found ever because it is not important when and with what accuracy the best solution was found. To improve the mentioned shortcomings, the HEAS has made some changes to the EAS as follows:

Route Construction

In the original EAS, an ant, say k, moves from the present node i to the next node j according to the state transition rule given by formula (3). Here we define μ_{ij} as the savings of combining two nodes on one tour as opposed to serving them on two different tours. The savings of combining any two customers i and j are computed as $\mu_{ij} = d_{i0} + d_{0j} - d_{ij}$ where d_{ij} denotes the distance between nodes i and j, and node 0 is the depot. Furthermore, λ like α and β is control parameter.

$$P_{ij}^k = \begin{cases} \dfrac{{\tau_{ij}}^{\alpha}{\eta_{ij}}^{\beta}\mu_{ij}^{\lambda}}{\sum\limits_{j \in N_i} {\tau_{ij}}^{\alpha}{\eta_{ij}}^{\beta}\mu_{ij}^{\lambda}} & if \quad j \in N_i \\ 0 & if \quad j \notin N_i \end{cases} \tag{3}$$

Pheromone Updating

In EAS, the pheromone of all edges belonging to the routes obtained by ants will be updated. The pheromone updating includes local and global updating rules. The pheromone updating formula was meant to simulate the change in the amount of pheromone due to both the addition of new pheromone deposited by ants on the visited edges and the pheromone evaporation. In HEAS, not only does the pheromone release increases pheromone on the edges of the best solution in each iteration but also the pheromone release reduction is used in order to distract ants from the edges of the worst solution. The idea of the elitist strategy in the context of the HEAS is to give extra emphasis to the best and the worst paths found so far after every iteration. The value of coefficient e in pheromone increase and pheromone decrease which is shown as -e is a function. It was found out that using polynomial k^2 can generate better solutions for the algorithm. In k^2, k is an integer number whose value is 1 at the beginning and increases one unit. Moreover, this polynomial is considered an appropriate function since it is not only an ascending function but also has variable slope as well. In other words, the function increases and the best routes are encouraged and strengthened compared with the previous paths. It should be noted that the low value of the function at the beginning causes the pheromone released on the best route to have less impact on the route selection in the following iterations and ants demonstrate a kind of forgetting in their search. In other words, this helps the ants to forget the weak solutions found at the start of the algorithm. However, as the algorithm is iterated and the accuracy of the solutions increases, the value of the polynomial rises rapidly and the edges belonging to the best solution absorb more pheromone and the edges of the worst solution lose more pheromone. Finally, in order to prevent the edges of the worst solution from gaining a negative value, a minimum value which is equal to the half of the initial pheromone is considered for the edges so that when the amount of pheromone on the edges falls below this minimum, it is replaced with the minimum value.

Local Search

A local search approach starts with an initial solution and searches within neighborhoods for better solutions. Abundant literature on ACO indicates that a promising approach for obtaining high-quality solutions can only be obtained through coupling ACO with a local search algorithm. Therefore, in the HEAS, after the ants have constructed their solutions and before the pheromone is locally updated, each ant's solution is improved by applying a local search. Because local search is a time-consuming procedure, only a local search is applied to the iteration's best solution. The idea here is that better solutions may have better chance to find a global optimum. In the proposed algorithm, at first a local search based on insert move and then the swap move is applied to the ant. In insert algorithm a customer is moved to another route. However, in swap algorithm a customer in a certain route is swapped with another customer from a different route. It should be noted that the new solution will be accepted only if first, VRP constraints are not violated especially about each vehicle's capacity and second, a novel solution will gain a better value for a problem than the previous solution.

3 Results

The HEAS was coded in Matlab 7. All the experiments were implemented on a PC with Pentium 4 at 2.4GHZ and 2GB RAM and Windows XP Home Basic Operating system. Because the proposed HEAS approach is a meta-heuristic algorithm, the results are reported for ten independent runs and in each run the algorithm was iterated n times. Furthermore, the pack of optional parameters obtained through several tests is $\alpha = 1, \beta = 4, \lambda = 3, \rho = 0.2$.
The performance of the proposed algorithm was tested on a set of 14 benchmark instances designed by Christofides et al. which have been widely used as benchmarks in order to compare the ability of proposed HEAS to the results of six meta-heuristic algorithms including SA and TS [11], genetic algorithm (GA) [12], scatter search algorithm combined by ACO (SS-ACO) [14], particle swarm intelligent (PSO) [16] and genetic algorithm combined with particle swarm intelligent (GAPSO) [20]. The information of the 14 instances is shown in Table 1. In this table, n, m, HEAS, mHEAS, BKS, and PD are the number of customers, the number of vehicles of best known solution (BKS), the best solutions found by HEAS algorithm, the used vehicles of HEAS, the best known solution, and the percentage deviation of HEAS compared to the best know solutions (BKSs) respectively. The PD is computed by formula (4) where $c(s^{**})$ is the best solution found by our algorithm for a given instance, and $c(s^{*})$ is the overall BKS for the same

instance on the Web. A zero PD indicates that the BKS is found by the algorithm.

$$PD = \frac{c(s^*) - c(s^{**})}{c(s^*)} \times 100 \quad (4)$$

Table 1: Computational results for standard VRP problems

Instance	n	m	SA	TS	GA	SS-ACO	PSO	GAPSO	HEAS	mHEAS	PD	BKS
C1	50	5	528	524	524.61	524.61	524.61	524.61	524.61	5	0	524.61
C2	75	10	838	844	849.77	835.26	844.42	835.26	847.14	10	-1.42	835.26
C3	100	8	829	835	840.72	830.14	829.40	826.14	712.36	8	13.77	826.14
C4	150	12	1058	1052	1055.85	1038.20	1048.89	1028.42	1066.89	12	-3.74	1028.42
C5	199	17	1376	1354	1378.73	1307.18	1323.89	1294.21	1311.35	17	-1.54	1291.45
C6	50	6	555	555	560.29	559.12	555.43	555.43	555.43	6	0	555.43
C7	75	11	909	913	914.13	912.68	917.68	909.68	909.68	11	0	909.68
C8	100	9	866	866	872.82	869.34	867.01	865.94	865.94	9	0	865.94
C9	150	14	1164	1188	1193.05	1179.4	1181.14	1163.41	1162.89	14	-0.03	1162.55
C10	199	18	1418	1422	1483.06	1410.26	1428.46	1397.51	1404.75	18	-0.64	1395.85
C11	120	7	1176	1042	1060.24	1044.12	1051.87	1042.11	1042.11	7	0	1042.11
C12	100	10	826	819	877.8	824.31	819.56	819.56	840.64	10	-2.57	819.56
C13	120	11	1545	1547	1562.25	1556.52	1546.20	1544.57	1545.93	11	-0.31	1541.14
C14	100	11	890	866	872.34	870.26	866.37	866.37	866.37	11	0	866.37

As can be seen from this table, the proposed algorithm finds the optimal solution for 6 out of 14 problems that are published in the literature. The results indicate that HEAS is a competitive approach compared to the BKSs. Furthermore, one new best known solution of the benchmark problem including C3 is also improved by the proposed method. For instances C4 and C12, the gap is about as high as 3%. However, in most of the instances, the proposed algorithm finds nearly the BKSs and for overall the average difference is 0.25%.

As the results in table 1 indicate, the GA has not been able to find the best solutions in thirteen of the fourteen examples. Therefore, it is considered to be the weakest algorithm among the seven presented algorithms. However, SS-ACO has been able to find better solutions than the GA and has come up with the best solutions in 12 examples. Among remaining 5 algorithms, in 11 examples SA has not been almost capable of finding the BKS. PSO has failed in improving the solutions in 10 examples and has come up with solutions similar to the ones found by SA. Hence, it can be concluded that TS is more efficient than GA, SS-ACO, PSO and SA in finding better solutions. From the comparison between GAPSO and HEAS, it can be seen that GAPSO in four examples has been able to find solutions with a gap of less than 1 percent. Despite being able to find the best solution ever found for ten examples, it has failed to achieve these results in the remaining four examples. However, HEAS has found better solutions than GAPSO for the two examples.

A simple criterion to measure the efficiency and the quality of an algorithm is to compute the average of solutions on specific benchmark instances. In figure 1, the average of each algorithm's solution is reported. From this table we conclude that the HEAS method has the best average with 975.44 and has been able to escape local optimum points. The algorithms in terms of their performance from the worst to the best can be listed as: GA, SA, TS, PSO, SS-ACO, GAPSO and HEAS. In addition, in order to demonstrate the efficiency of the proposed algorithm, two of the solutions found for the examples in table 1 are presented in Figure 2. It should be noted

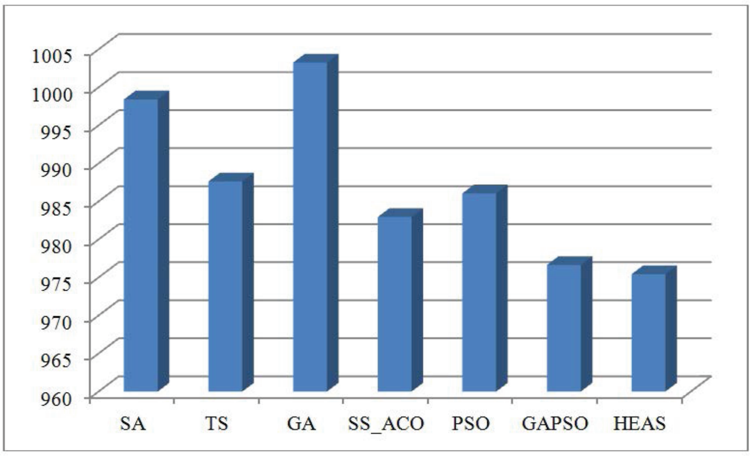


Figure 1: Comparison of mean for 14 instances between meta-heuristic algorithms in Table 1

that in 1 out of 2 examples presented in this figure, the HEAS has been able to improve the best solution ever found.

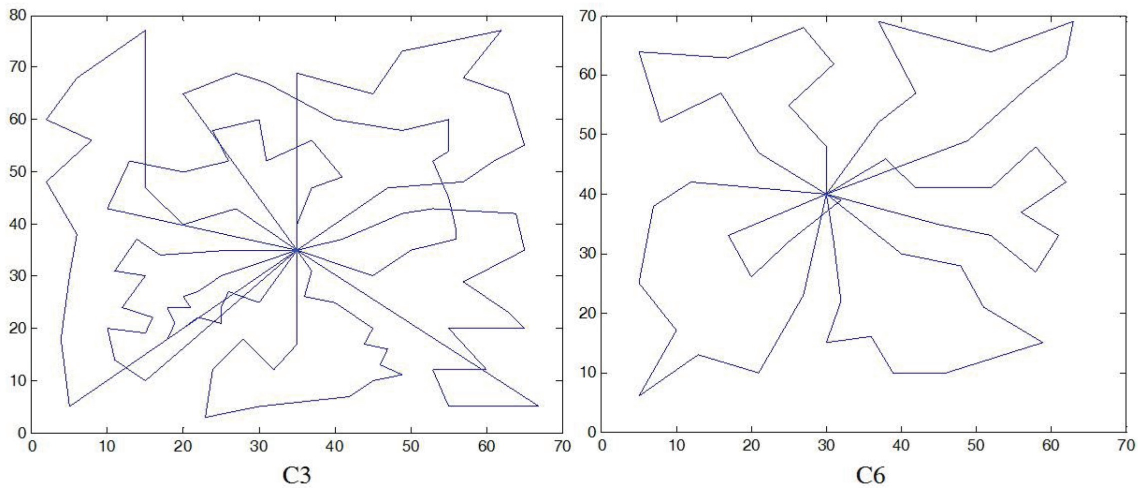


Figure 2: Some of the Solutions to the VRP Found by the proposed algorithm

4 Conclusion and Future Works

In this paper, a modified version of EAS which employs several effective modifications was presented. The modifications improved the performance of the classic EAS algorithm in escaping from local optimum points and finding better solutions in comparison with the other metaheuristic algorithms. It seems that combining the proposed algorithm with other metaheuristic algorithms like tabu search and making use of strong local algorithms like lin-kernigan algorithm can bring better results for the HEAS. Furthermore, HEAS can be used for other versions of VRP like OVRP and heterogeneous fixed fleet OVRP. Future projects will focus on working on such ideas and making them operational.

Bibliography

[1] Yousefikhoshbakhtm, M. and Khorram, E. (2012). Solving the Vehicle Routing Problem by a Hybrid Meta-Heuristic Algorithm, *Journal of Industrial Engineering International*, 8(12):1-9.

[2] Yousefikhoshbakht, M., Didehvar, F. and Rahmati, F. (2014). A Combination of Modified Tabu Search and Elite Ant System to Solve the Vehicle Routing Problem with Simultaneous Pickup and Delivery, *Journal of Industrial and Production Engineering*, 31(2): 65-75.

[3] Hong, L. (2012). An improved LNS algorithm for real-time vehicle routing problem with time windows, *Computers and Operations Research*, 39(2):151-163.

[4] Yousefikhoshbakht, M., F. Didehvar, and F. Rahmati, (2013). Solving the heterogeneous fixed fleet open vehicle routing problem by a combined metaheuristic algorithm, *International Journal of Production Research*, http://dx.doi.org/10.1080/00207543.2013.855337.

[5] Toth, P. and Vigo, D. (1997). An exact algorithm for the vehicle routing problem with backhauls, *Transportation Science*, 31:372-385.

[6] Pop P.C., Sitar C.P., Zelina I., Lupse V. and Chira C. (2011). ,Heuristic Algorithms for Solving the Generalized Vehicle Routing Problem, *International Journal of Computers Communications & Control*, ISSN 1841-9836, 6(1): 158-165.

[7] Valle, C. A., Martinez, L. C. and Cunha, A. S., Mateus, G. R. (2011). Heuristic and exact algorithms for a min-max selective vehicle routing problem, *Computers and Operations Research*, 38(7):1054-1065.

[8] Brad, J., Kontoravdis, G. and Yu, G. (2002). A Branch-and-Cut Procedure for the Vehicle Routing Problem with Time Windows. *Transportation Science*, 36(2):250-269.

[9] Clarke, G. and Wright, J. W. (1964). Scheduling of vehicles from a central depot to a number of delivery points. *Operations Research*, 12: 568-581.

[10] Gillett, B. E. and Miller, L. R. (1974). A heuristic algorithm for the vehicle dispatch problem. *Operations Research*, 22: 340-349.

[11] Osman, I. (1993). Metastrategy simulated annealing and Tabu search algorithms for the vehicle routing problem, *Operations Research*, 41: 421-451.

[12] Baker, B., and Ayechew, M. (2003). A genetic algorithm for the vehicle routing problem, *Computers and Operations Research*, 30: 787-800.

[13] Hirota, K., Dong, F. and Chen, K. (2006), A Computational Intelligence Approach to VRSDP (Vehicle Routing, Scheduling, and Dispatching Problems), *International Journal of Computers Communications & Control*, ISSN 1841-9836, Suppl. issue, 1(S): 53-60,

[14] Zhang, X. and Tang, L. (2009). A new hybrid ant colony optimization algorithm for the vehicle routing problem, *Pattern Recognition Letters*, 30(152): 848-855.

[15] Negulescu S.C., Kifor C.V. and Oprean C. (2008). Ant Colony Solving Multiple Constraints Problem: Vehicle Route Allocation , *International Journal of Computers Communications & Control*, ISSN 1841-9836, 3(4):366-373.

[16] Ai, J., Kachitvichyanukul, V. (2009). Particle swarm optimization and two solution representations for solving the capacitated vehicle routing problem, *Computers and Industrial Engineering*, 56:380-387.

[17] Pintea C.-M., Chira C., Dumitrescu D. and Pop P.C. (2011). Sensitive Ants in Solving the Generalized Vehicle Routing Problem, *International Journal of Computers Communications & Control*, ISSN 1841-9836, 6(4):734-741.

[18] Yousefikhoshbakht, M., Didehvar, F. and Rahmati, F. (2013). Modification of the Ant Colony Optimization for Solving the Multiple Traveling Salesman Problem, *Romanian Journal of Information Science and Technology*, 16(1):65-80.

[19] Yousefikhoshbakht, M. and Sedighpour, M. (2012). A Combination of Sweep Algorithm and Elite Ant Colony Optimization for Solving the Multiple Traveling Salesman Problem, *Proceedings of the Romanian academy, A*, 13(4):295-302.

[20] Marinakis, Y. and Marinaki, M. (2010). A hybrid genetic-particle swarm optimization algorithm for the vehicle routing problem, *Expert Systems with Applications*, 37(33):1146-1455.

INT J COMPUT COMMUN, ISSN 1841-9836
9(3):348-355, June, 2014.

Backstepping-based Robust Control for WMR with A Boundary in Prior for the Uncertain Rolling Resistance

M. Yue, S. Wang, X.L. Yang

Ming Yue*, Shuang Wang, Xiaoli Yang
School of Automotive Engineering
Dalian University of Technology, Dalian 116024, China
*Corresponding author: yueming@dlut.edu.cn

Abstract: In this study, we focus on the trajectory tracking control problem of a wheeled mobile robot (WMR) in an uncertain dynamic environment. Concerning the fact that the upper boundary may be usually achieved in prior according to the physical properties of the terrain, this crucial message is utilized to construct the controllers. Firstly, a dynamic model for WMR including the rolling resistance is presented, whose state variables are longitudinal and rotational velocities, as well as the rotational angle of the mobile platform. Secondly, with the aid of backstepping technique, the robust controllers based on the upper boundary are proposed and the globally asymptotic stability of the closed-loop system is proven by the Lyapunov theory in the following. Lastly, a saturation function is applied to replace the signum function, by which the inherent chattering can be suppressed greatly. Numerical simulation results demonstrate that the proposed controllers with upper bound in prior possess robustness characteristics which yields potentially valuable applications for the mobile robot, especially in the unstructured environment.
Keywords: Wheeled Mobile Robot (WMR), boundary estimation, uncertain rolling resistance.

1 Introduction

In practice, the physical properties of the wheeled mobile robot (WMR) within an unstructured environment are seldom considered in lots of trajectory tracking control issues. Among these physical properties, the rolling resistance acting on the robot wheels is such a crucial physical behavior that can effect the control performances greatly. Particulary, in some rough terrain, such as loose soil and sand conditions, it is difficult to obtain high-performance of trajectory tracking without considering the influences of the rolling resistance. Therefore, when the WMR explores in an outdoor environment and even planetary surface, the rolling resistance may not be neglected for designing the system controllers [1].

Many efforts are devoted to the rolling resistance acting on the WMR's wheels recently. Most References (see [2,3]) had investigated the interactions of soil and wheels and then some new mechanical structures had also been reported to enhance the contact between the wheels and terrain. However, there are only a few investigations to deal with rolling resistance problem by control methods. After analysing of the terrain properties, we notice that the boundary (especially for the upper bound) of off-road rolling resistance can be calculated in advance according to the geologic parameters when the terrain of the exploration is determined. For practical implementation, the rolling resistance appears drastically changing, which is similar to the disturbances to the mobile benchmark, but with the aid of the boundary message in prior, the robust controller for WMR can be designed to make it possible to against the uncertain dynamics directly. It should be mentioned that, compared with modifying the system mechanical structure, the adopted control approach yields much more flexibility for the mobile robot. Certainly, we are also aware of adaptive scheme to update the upper bound, but the control algorithm with adaptive scheme may make the regulation process more elastic and cause further time-delay usually [4, 5]. To overcome this drawback, a powerful robust control approach is determine to employ to deal with the uncertain problem of the unavoidable rolling resistance.

On another hand, if the rolling resistance is concerned, the dynamics of the WMR must be considered. For the past decades, most efforts have been done to the WMR with nonholonomic constraint on kinematic model, all of which the system velocities were assumed to be the control inputs [6, 7]. But unfortunately, there are only a few researches address the dynamic behaviors of nonholonomic system, where control inputs are transformed to be the system actuators, i.e., in most cases the driving torques of motors [8–10]. Since the rolling resistance is described in torque form when it takes place in practice, the dynamic properties must be considered when designing the system controllers to overcome this kind of uncertain dynamics.

With a widely survey of the current dynamic models for the WMR, we observe that a dynamic model proposed by Watanabe et.al [11] is so succinct and effective to describe the relationship between the rolling resistance and robot posture on the level of dynamics. Then, we utilize this dynamic model to establish the control system in this investigation. Particularly, it should be noted that the rolling resistance is somewhat difference from the external disturbances and unmodeled dynamics, because the rolling resistance lies in the movement process around a constant value; while the external disturbances and unmodeled dynamics possess white noise property which is around zero. The objective of this study is to address the rolling resistance behaviors of the WMR in detail, and derives the dynamic model to discuss the influence of rolling resistance. Based on addressed dynamic model, a simple and effective approach to stabilize the trajectory tracking system via backstepping techniques is developed. In order to illustrate the efficacy of the control approach, numerical simulations for a practical WMR have been performed.

2 WMR dynamics

The typical WMR can be described in Fig.1. It is assumed that the mobile robot is driven by two independent wheels as well as a passive auxiliary wheel is adopted to support the workbench. To begin with, some notations is introduced to help the controlling system design, as follows: l is the distance between the driving wheels and the symmetry axis; d is the distance between point $C(x, y)$ and the mass center of the robot $P(x_c, y_c)$, which is assumed to be on the symmetry axis; r is the radius of the driving wheels; m_p and m_w are the masses of vehicle body and wheel; I_c is the inertia moment of vehicle body w.r.t. vertical axis through C; I_w and I_m are the inertia moments for the wheel w.r.t. wheel axis and diameter, respectively. According to the parallel axis theorem, the equivalent mass $m = m_p + 2m_w$ and rotation inertia $I = 2I_m + 2m_w l^2 + m_c d^2 + I_c$ can be introduced for simply.

The robot position can be described by the coordinates (x, y), which is the midpoint C of the axis of two robot wheels, and the orientation angle ϕ, which is heading angle of body coordinate with respect to fixed frame (see Fig.1). With the hypothesis of *pure rolling and non slipping* condition, the nonholonomic constraint, i.e., $\dot{x}\sin\phi - \dot{y}\cos\phi = 0$, hold throughout the movements.

Let v and w be represent the linear and angular velocities of the mobile robot, and τ_r and τ_l express the driving torque of the right and left wheels, respectively. Through the analysis of the forces acting on the mobile robot, we can obtain that

$$I\ddot{\phi} = \frac{\tau_r}{r}l - \frac{\tau_l}{r}l \qquad (1)$$

$$m\dot{v} = \frac{\tau_r}{r} + \frac{\tau_l}{r} \qquad (2)$$

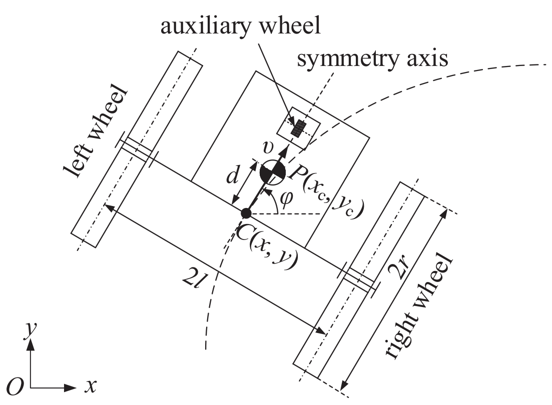


Figure 1: A wheeled mobile robot

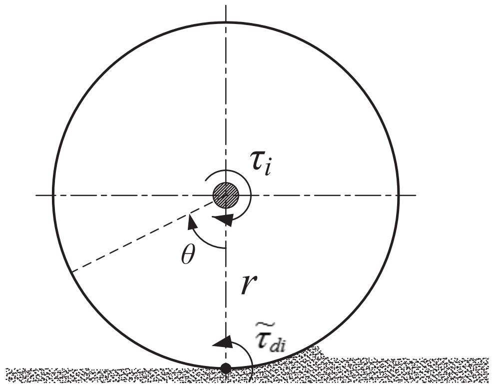


Figure 2: Rolling resistance generation

While rolling on the ground, especially operating on the soft soil or sandy terrain, the robot will suffer unavoidable rolling resistance, as shown in Fig.2. Let the notation $\tilde{\tau}_d$ represent this rolling resistance, and the dynamic behaviors of the wheel can be governed by

$$I_w\ddot{\theta}_i + c\dot{\theta}_i = ku_i - \tau_i - \tilde{\tau}_{di} \tag{3}$$

where τ_i is the driving input of the wheel (i expresses r or l for left or right wheels, the same hereinafter); c is the viscous friction coefficient and θ is the rolling angle of the wheel; k is driving gain between the motor's voltage and its output torque; u_i is the excited voltage signals for the motor installed in the relative wheel.

Let v_r, v_l denote the right and left linear velocity of the wheel center, and we can further have $v_r = r\dot{\theta}_r = v + l\dot{\phi}$ and $v_l = r\dot{\theta}_l = v - l\dot{\phi}$. Then, the relationship between linear velocity of the robot's platform and angular velocity of the robot wheels can be formulated by $r(\dot{\theta}_r + \dot{\theta}_l) = 2v$ and $r(\dot{\theta}_r - \dot{\theta}_l) = 2l\dot{\phi}$.

With the analysis, we can ultimately obtain the dynamics described by the linear velocity v and orientation angle ϕ as follows:

$$\dot{v} = -\frac{2cv}{mr^2 + 2I_w} + \frac{kr}{mr^2 + 2I_w}(u_r + u_l) - \frac{r}{mr^2 + 2I_w}(\tilde{\tau}_{dr} + \tilde{\tau}_{dl}) \tag{4}$$

$$\ddot{\phi} = -\frac{2cl^2}{Ir^2 + 2I_w l^2}\dot{\phi} + \frac{krl}{Ir^2 + 2I_w l^2}(\tilde{\tau}_{dr} - \tilde{\tau}_{dl}) \tag{5}$$

If the system state variables are chosen as $x = [v\ \phi\ \dot{\phi}]^T$, the driving input is selected as $u = [u_r\ u_l]^T$, the disturbance vector is defined as $\tilde{\tau}_d = [\tilde{\tau}_{dr}\ \tilde{\tau}_{dl}]^T$, and the output variables are $y = [v\ \phi]^T$, a new dynamic model for control system design can be rewritten by:

$$\begin{cases} \dot{x} = Ax + Bu + D\tilde{\tau}_d \\ y = Cx \end{cases} \tag{6}$$

with

$$A = \begin{bmatrix} a_1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & a_2 \end{bmatrix},\ B = \begin{bmatrix} b_1 & b_1 \\ 0 & 0 \\ b_2 & -b_2 \end{bmatrix},\ D = \begin{bmatrix} d_1 & d_1 \\ 0 & 0 \\ d_2 & -d_2 \end{bmatrix},\ C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix},$$

where $a_1 = -\frac{2c}{mr^2 + 2I_w}$, $a_2 = -\frac{2cl^2}{Ir^2 + 2I_w l^2}$, $b_1 = \frac{kr}{mr^2 + 2I_w}$, $b_2 = \frac{krl}{Ir^2 + 2I_w l^2}$, $d_1 = -\frac{r}{mr^2 + 2I_w}$, $d_2 = -\frac{rl}{Ir^2 + 2I_w l^2}$.

3 Control system design

3.1 System decoupling

Noticing that the system (6) is a coupled system, for the convenience of designing the system controller, the proposed dynamic model should be decoupled in the first place. To achieve this objective, a new control input vector should be introduced as follow:

$$\begin{bmatrix} u_r \\ u_l \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \tag{7}$$

With these modified inputs u_1 and u_2, the system (6) is transformed into two independent subsystems:

$$\dot{v} = a_1 v + b_1 u_1 + \tilde{\tau}_{dv} \tag{8}$$

$$\dot{w} = a_2 w + b_2 u_1 - 2b_2 u_2 + \tilde{\tau}_{dw} \tag{9}$$

where $\tilde{\tau}_{dv} = d_1(\tilde{\tau}_{dr} + \tilde{\tau}_{dl})$ and $\tilde{\tau}_{dw} = d_2(\tilde{\tau}_{dr} + \tilde{\tau}_{dl})$, which represent disturbances for the two independent subsystems.

Supposing the robot moves on a special terrain ground, although the rolling resistance is in a changeable state, the maximum amplitude for $\tilde{\tau}_{dr}$ and $\tilde{\tau}_{dl}$ could be estimated in advance from the terrain property. Let the notation $\bar{\tau}_{dr}$ and $\bar{\tau}_{dl}$ represent the upper bound values of $\tilde{\tau}_{dr}$ and $\tilde{\tau}_{dl}$ respectively, and then it can be achieved that

$$\bar{\tau}_{dv} = \max |d_1(\tilde{\tau}_{dr} + \tilde{\tau}_{dl})| \tag{10}$$

$$\bar{\tau}_{dw} = \max |d_2(\tilde{\tau}_{dr} + \tilde{\tau}_{dl})| \tag{11}$$

3.2 Control scheme design

Considering the decoupled form of the system (8) and (9), a backstepping technique can be applied to derive the robot controller due to the high dimension feature of the system. The process of control design procedures may be divided into two steps which is given as bellows.

Step 1. Linear velocity control

After the trajectory tracking system has been decoupled, the control algorithms for linear and angular velocities can be derived stage by stage. Here, the control input u_1 is used to control the linear velocity, such that the robot can track an desired trajectory from an arbitrary original point with the required performances. Suppose that the desired linear velocity v_d, and the tracking error may be expressed by $v_e = v_d - v$. Then, a theorem can be given as follow:

Theorem 1. *For the linear velocity system (4) with the bounded linear velocity, the tracking error v_e will globally converge to zero, i.e., $\lim_{t\to\infty} v_e(t) = 0$, if the control law is given by*

$$u_1 = \frac{1}{b_1}[\dot{v}_d - a_1 v + c_1 v_e + sgn(v_e)\bar{\tau}_{dv}] \tag{12}$$

where c_1 is a positive constant and $sgn(\cdot)$ is a signum function.

Proof: Consider a candidate Lyapunov function as

$$V_1 = \frac{1}{2} v_e^2 \tag{13}$$

Differentiating V_1 with respect to time yields

$$\dot{V}_1 = v_e \dot{v}_e = v_e(\dot{v}_d - a_1 v - b_1 u_1 - \tilde{\tau}_{dv}) \tag{14}$$

Substituting (12) into (14), it can be ultimately obtained $\dot{V}_1 = -c_1 v_e^2 \leq 0$. Obviously, by Lyapunov stability theorem, the propose Theorem 1 is proved. □

Step 2. Angular velocity control

Let ϕ_d represents the desired orientation angle of the robot to be tracked. Then, the tracking error of the orientation angle can be defined as $\phi_e = \phi_d - \phi$ and $\dot{\phi}_e = \dot{\phi}_d - \dot{\phi}$. In terms of angular velocity tracking control, it is to design the control input u_2 such that the ϕ_e and $\dot{\phi}_e$ converge to zero as $t \to \infty$. In order to realize this purpose, two virtual control inputs are introduced, which are defined as $z_1 = \phi_e$ and $z_2 = \dot{\phi}_e + c_2 \phi_e$ where c_2 is an arbitrary positive constant. Similar to the linear velocity control, based on the above definitions, a Theorem could be given as

Theorem 2. *For the angular velocity system (5) with the bounded angular velocity, the tracking error ϕ_e and $\dot{\phi}_e$ will both globally converge to zero simultaneously , i.e., $\lim_{t\to\infty} ||\phi_e \; \dot{\phi}_e|| = 0$, if the control law is given by*

$$u_2 = \frac{1}{2b_2}[a_2 w + b_2 u_1 - \ddot{\phi}_d - c_2 z_1 + c_3 z_2 + sgn(z_2)\bar{\tau}_{dw}] \tag{15}$$

where c_3 is a positive constant and $sgn(\cdot)$ is a signum function.

Proof: Consider a candidate Lyapunov function as

$$V_2 = \frac{1}{2} z_1^2 \tag{16}$$

Differentiating V_2 with respect to time achieves

$$\dot{V}_2 = z_1 \dot{z}_1 = z_1(\dot{\phi} - \dot{\phi}_d) = z_1 z_2 - c_2 z_1^2 \tag{17}$$

Notice that only Lyapunov function V_2 can not guarantee the global stability of the system. In order to stabilize the system, another virtual input z_2 is designed such that the entire tracking errors converge to zero. Hence, another candidate Lyapunov function constructed by z_2 is established for this purpose. The new candidate Lyapunov function is put forward as

$$V = \frac{1}{2} z_1^2 + \frac{1}{2} z_2^2 \tag{18}$$

Differentiating V with respect to time obtains

$$\dot{V} = \dot{V}_2 + z_2 \dot{z}_2 = z_1 z_2 - c_2 z_1^2 + z_2(a_2 w + b_2 u_1 - 2b_2 u_2 + c_2 \dot{z}_1 - \ddot{\phi}_d) \tag{19}$$

Substituting (12) and (15) into (19), one can obtain that $\dot{V} = -c_2 z_1^2 - c_3 z_2^2 \leq 0$. Obviously, according to Lyapunov stability theorem, the Theorem is proved. □

Since the systems (8) and (9) are decoupled, the proposed controller (12) and (15) may guarantee the stabilization of the linear and angular velocity subsystems overall. To sum up, we can summarize the investigation results and ultimately give a theorem as follow:

Theorem 3. *For the trajectory tracking system (4) and (5), the tracking errors v_e, ϕ_e and $\dot{\phi}_e$ will all globally converge to zero, i.e., $\lim_{t\to\infty} ||v_e \; \phi_e \; \dot{\phi}_e|| = 0$, with the controller (12) and (15).*

The proof can be easily achieved according to the above analysis.

From the controller designing process, it can be observed that the candidate Lyapunov functions are established step by step, which make the control design process be simple and convenience. This is a backstepping approach which is usually applied for controller design of a sophisticated nonlinear systems. Meanwhile, it should be emphasized that there are inherent chattering caused by signum function in the proposed controllers. To overcome this drawback, one can replace signum function sgn(·) by saturation functions sat(·) to suppress this behaviors, one of which can be given by:

$$sat(\chi) = \frac{\chi}{|\chi| + \epsilon} \tag{20}$$

where ϵ is a positive constant and χ represents arbitrary variable.

4 Simulation results

In this section we perform numerical simulations to verify the effectiveness of the proposed controllers. The simulated parameters of the mechanical structure of the WMR are supposed to be m_c = 14kg, m_w = 3kg, I_c = 0.9kgm^2, I_m = 0.05kgm^2, I_w = 0.015kPgm^2, l = 0.3m, d = 0.05m, r = 0.1m, k = 0.95, and c = 0.01. From the required performance of the trajectory tracking, the control parameters as selected as follows: $c_1 = c_2 = c_3 = 10$ and $\epsilon = 0.002$. Suppose the mobile robot starts from initial point $[x\ y\ \phi]^T = [0\ 0\ 0]^T$; meanwhile, the desired trajectory of linear and angular velocity are hypothesized to be v_d = 1m/s, $w_d = \sin t$ with the initial state vector $x_0 = [v_0\ \phi_0\ w_0]^T = [0.5\ 0.2\ 0]^T$. Moreover, the rolling resistance acting on the left and right wheels are supposed to be a worse situation in simulation than that in practice to verify the effectiveness of the control algorithms. So a synthesis signal consisted of two sinusoid signals with different frequency is introduced to imitate the rolling resistance. Here, assume that the rolling resistances are same on the left and right wheels, i.e., $\tilde{\tau}_{dr} = \tilde{\tau}_{dl} = \sin w_d t + \sin 2w_d t$ where w_d denotes the resistance frequency. According to practical experiment, the $\bar{\tau}_{dv}$ and $\bar{\tau}_{dw}$ can be set as the value of 0.75Nm and 3Nm respectively in this simulation circumstance. In addition, the saturation should be considered for actual driving motor, so a maximum value 10Nm is introduced to restrict the control outputs, i.e., $|\tau_r| \leq$ 10Nm and $|\tau_l| \leq$ 10Nm.

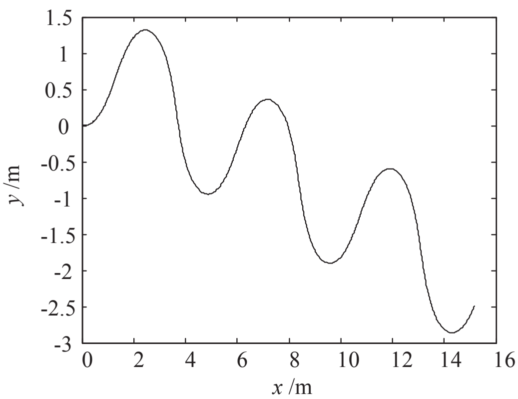


Figure 3: Actual trajectory on the plane

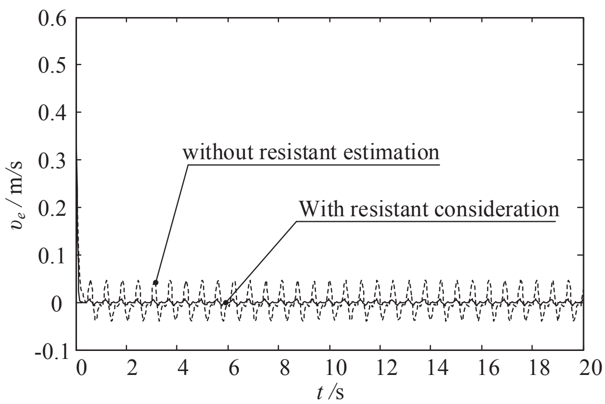


Figure 4: Tracking error for v_e

Under the proposed simulation circumstance, the actual tracking process for the desired trajectory is plotted in Fig.3. It indicates the proposed control method can make the mobile robot to track the desired trajectory with satisfactory tracking performances. Besides, we simulate two cases, that is, with and without estimation, to illustrate the signification of control with and without upper bound information. As shown in Fig.4, the tracking error relative to the linear velocity without boundary is described as the dashed line, as well as the case with boundary is plotted by the dashed line. The results state the tracking errors become much smaller (reduced to 20% of the amplitude) when $\bar{\tau}_{dv}$ and $\bar{\tau}_{dw}$ are adopted. Particularly, owing to the sat(·), the inherent chattering is also suppressed. Lower chattering may be achieved with smaller ϵ, but it will cause more violent changes for the control inputs.

Furthermore, to exhibit the better tracking performance with the upper bound, the tracking errors of $\dot{\phi}_e$ and ϕ_e are also addressed in Fig.5 and Fig.6, respectively. It is observed that the amplitudes of ϕ_e and w_e are reduced to 21% and 18% nearly compared with the case without upper bound estimation, which yields that the derived controller can provide better performances for tracking system and the robustness is enhanced in the following.

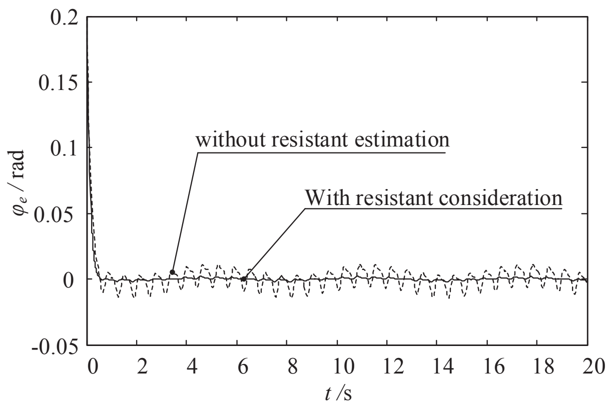


Figure 5: Tracking errors for ϕ_e

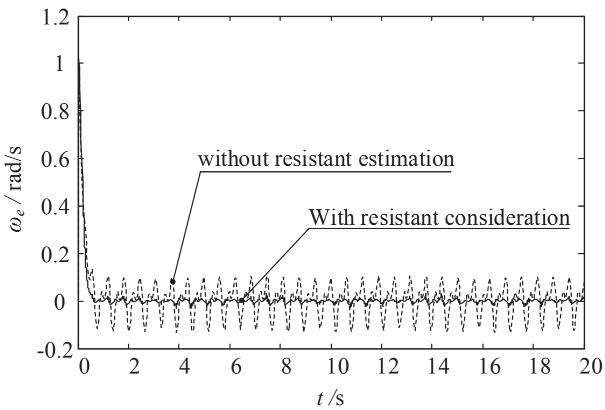


Figure 6: Tracking errors for ω_e

5 Conclusions

The trajectory tracking issues were discussed in this study, and the main contribution can be summarized as follows: 1) a decoupled dynamic model was used to describe the complicated nonlinear behaviors of WMR where the rolling resistances are concerned. The decoupled model consisted of two parts: a first-order model for linear velocity and a second-order model for orientation angle of the robot, which not only kinematic postures but also dynamics behaviors were all included; 2) based on the fact that the upper bound of rolling resistance which can be achieved in prior, a robust controller was developed to obtain the better tracking performances. The backstepping methodology was employed to derive the system controller considering the higher dimensional property of the tracking system; and 3) a saturation function was introduced to alleviate the inherent chattering caused by signum function. Such treatment was simply but effective for a practical WMR system.

Acknowledgement

This research was supported by grants from the National Natural Science Foundation of China (No. 61175101), and the Fundamental Research Funds for the Central Universities (No. DUT12LK45).

Bibliography

[1] L. Ding, H. B. Gao, Z. Q. Deng, K. Nagatani, and K. Yoshida, Experimental study and analysis on driving wheels performance for planetary exploration rovers moving in deformable soil, *Journal of Terramechanics*, 48(1): 27-45, 2011.

[2] C. C. Ward and K. Iagnemma,A dynamic-model-based wheel slip detector for mobile robots on outdoor terrain, *IEEE Transactions on Robotics*, 24(4): 821-831, 2008.

[3] P Lamon and R. Siegwart, Wheel torque control in rough terrain-modeling and simulation, *IEEE Int. Conf. on Robotics and Automation,* Barcelona, Spain, 867-872, 2005.

[4] X. M. Sun, G. P. Liu, W. Wang and D. Rees, L_2-gain of systems with input delays and controller temporary failure: Zero-order hold model, *IEEE Transactions on Control Systems Technology*, 19(3): 699-706, 2011.

[5] R. Wang, G. P. Liu, W. Wang, D. Rees and Y. B. Zhao, H_∞ control for networked predictive control systems based on the switched Lyapunov function method, *IEEE Transactions on Industrial Electronics,* 57(10): 3565-3571, 2010.

[6] A. Chatterjee, O. Ray, A. Chatterjee and A. Rakshit, Development of a real-life EKF based SLAM system for mobile robots employing vision sensing, *Expert Systems with Applications*, 38(7): 8266-8274, 2011.

[7] D. Buccieri, D. Perritaz, P. Mullhaupt, Z. P. Jiang and D. Bonvin, Velocity-scheduling control for a unicycle mobile robot: Theory and experiments, *IEEE Transactions on Robotics*, 25(2): 451-458, 2009.

[8] Z. J. Li and S. Z. S. Ge and A. G. Ming, Adaptive robust motion/force control of holonomic-constrained nonholonomic mobile manipulators, *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 37(3): 607-616, 2007.

[9] M. Yue, P. Hu and W. Sun, Path following of a class of non-holonomic mobile robot with underactuated vehicle body, *IET Control Theory and Applications*, 4(10): 1898-1904, 2009.

[10] M. Yue, W. Sun and P. Hu, Sliding mode robust control for two-wheeled mobile robot with lower center of gravity, *International Journal of Innovative Computing Information and Control*, 7(2): 637-646, 2011.

[11] K. Watanabe, J. Tang, M. Nakamura, S. Koga and T. Fukuda, Fuzzy-Gaussian neural network and its application to mobile robot control, *IEEE Transactions on Control Systems Technology*, 4(2): 193-199, 1996.

INT J COMPUT COMMUN, ISSN 1841-9836
9(3):356-369, June, 2014.

A Feedback-corrected Collaborative Filtering for Personalized Real-world Service Recommendation

S. Zhao, Y. Zhang, B. Cheng, J.-L. Chen

Shuai Zhao*, Yang Zhang, Bo Cheng, Jun-liang Chen
State Key Laboratory of Networking and Switching Technology
Beijing University of Posts and Telecommunications
Beijing, 100876, China
*Corresponding author: zhaoshuaiby@bupt.edu.cn

Abstract: The emergence of Internet of Things (IoT) integrates the cyberspace with the physical space. With the rapid development of IoT, large amounts of IoT services are provided by various IoT middleware solutions. So, discovery and selecting the adequate services becomes a time-consuming and challenging task. This paper proposes a novel similarity-measurement for computing the similarity between services and introduces a new personalized recommendation approach for real-world service based on collaborative filtering. In order to evaluate the performance of proposed recommendation approach, large-scale of experiments are conducted, which involves the QoS-records of 339 users and 5825 real web-services. The experiments results indicate that the proposed approach outperforms other compared approaches in terms of accuracy and stability.
Keywords: Internet of Things, service recommendation, similarity measurement, collaborative filtering.

1 Introduction

1.1 Background

Internet of Things (IoT) seamlessly integrates user requirement, cyberspace and physical space, which enables the dynamic cooperation of Ą°human-machine-thingĄą. As the adopting of SOA (service oriented architecture) paradigm in IoT environment [1], real-world devices will be able to offer their functionality via service interfaces, which enable other components to interact with them dynamically. The functionalities provided by these devices (e.g. the provisioning of online sensing data) are referred to as *real-world services* because they are provided by embedded systems that are directly related to the physical world [2]. With the developments of IoT, lots of middleware solutions like OpenIoT [3], GSN [4], COSM [5] are proposed which act as service provision platforms. These platforms access real-world resources around the world and provide their capability in form of millions of *real-world services*, which enable sharing, monitoring and controlling environmental data on the web. However, these existing platforms only provide limited functions for service selecting and recommendation. As the rapid increase of available services, selecting the appropriate services becomes challenging and time-consuming [6]. Therefore, service discovery becomes the critical issues of IoT development. Other than web-service discovery based on functional property which has been deeply studied [7], the studies of discovery based on non-functional property are far from mature, i.e. it is difficult to differentiate services with similar or identical functions. As the user-received performance of service is tightly related to the personalized information of specific users, identifying the optimal one for service users is difficult and costly in the case of many services with equivalent functions. Effective personalized service selection and recommendation based on non-functional property become more and more important [8]. Quality of Service (QoS) (e.g. observation-accuracy, round-trip time (RTT), etc.)

is served as key non-functional property, which acts as important element considered when discovering and selecting services [9, 10]. The values of QoS are usually influenced by the specific environment (such as network quality) of users and tend to vary with each user. Because conducting actual service invocation is time and resource consuming, it is unpractical to evaluate the QoS of all candidate services for every user [11]. So, the idea of making personalized QoS prediction for users using a small amount of available QoS value is extremely useful. Based on the predicted QoS values, personalized service recommendation is available for service-users. It enables users to select the service with optimal QoS from a number of services which are function-equivalent.

1.2 Motivation

Collaborative filtering (CF) are widely used in recommender systems [12]. The algorithms of CF can be divided into two categories: memory based and model based. Cosine-based approach (COS) [14, 15] and Pearson Correlation Coefficient (PCC) [12, 13] are two of the most popular memory-based approaches [16] to calculate the similarity between items. A number of works that employ COS-CF (Cosine based Collaborative-filtering) or PCC-CF (Pearson Correlation Coefficient based Collaborative filtering) for QoS based service recommendation and selection have been proposed recently [11] [17–19]. However, the performance that using PCC and COS to measure similarity leaves much to be desired and the prediction accuracy of these works cannot satisfy the requirement of practical application. Moreover, the experiments of these existing works are not convincing enough. The existing service recommendation [17, 19] approaches are short of sufficient-scale and systematic evaluation to verify their recommendation results. Some of them employ item dataset (such as MovieLens [20]) instead of real service dataset to evaluate their approaches.

In order to address these issues, this paper proposes a novel similarity-measurement for computing the similarity between service users and introduces a new personalized recommendation approach for real-world service based on collaborative filtering, which named feedback-corrected Tan-NED (Tanimoto Normalized Euclidean Distance). The contributions of this paper are summarized as the following aspects:

- This paper proposes a novel similarity measurement for memory based CF, which avoids the shortcomings of existing approaches for service recommendation and takes the characteristics of real-world service QoS into account. Therefore, it finds similar users more accurately and obtains more accurate QoS-prediction.
- A feedback-corrected CF approach is proposed, which significantly improves the performance of service recommendation in comparison with existing approaches.
- To evaluate the performance of proposed approach, we conduct comprehensive evaluative experiments based on a large-size real-service QoS dataset which includes 5825 real web-services and 339 users.

The following parts of this paper are organized as follows. In section 2, we review existing similarity measurement approaches for memory-based CF. Section 3 presents the proposed feedback-corrected Tan-NED CF approach. The experiment results of proposed approach are discussed in section 4. Section 5 concludes the work and discusses the future work.

2 Existing Similarity Measurement

The service similarity measurements are divided into two types, namely, functional-measurement and non-functional-measurement. Our previous work [21] focuses on measuring the functional similarity between real-world services based on semantic model. Other than functional similarity, this paper focuses on non-functional similarity, i.e. QoS similarity. The idea of memory-based CF is inspired by the fact that users trust the recommendations from the one who have similar context and preference. These methods predict the QoS of a particular service for a user based on the QoS obtained from users who have similar context and preference. In memory-based CF, Cosine-based approach (COS) [14,15] and Pearson Correlation Coefficient (PCC) [12,13] are two of the most popular algorithms to measure the similarity.

Assuming that a service recommender system includes N services and M users, then we obtain an $M \times N$ user-service matrix, in which the entry $r_{m,n}$ denotes the QoS value of service n observed by user m. If the entry $r_{m,n} = \emptyset$, it indicates that user m has never invoked service n. PCC-CF can be used to calculate the similarity between user u and user v by following formula:

$$Sim(u,v) = \frac{\sum_{i \in I}(r_{u,i} - \overline{r_u})(r_{v,i} - \overline{r_v})}{\sqrt{\sum_{i \in I}(r_{u,i} - \overline{r_u})^2}\sqrt{\sum_{i \in I}(r_{v,i} - \overline{r_v})^2}} \quad (1)$$

Where $I = I_u \cap I_v$ is the set of services that are co-invoked by users u and v, $r_{u,i}$ denotes the quantized QoS value of service i observed from the view of user u, and $\overline{r_u}$ is the average value on the QoS of services in I observed by user u. The values of PCC range from -1 to 1 according to the definition of equation (1).

In COS CF, the similarity between users can be measured by calculating the cosine similarity of the vectors between them:

$$sim(u,v) = \frac{\sum_{i \in I} r_{u,i} r_{v,i}}{\sqrt{\sum_{i \in I} r_{u,i}^2}\sqrt{\sum_{i \in I} r_{v,i}^2}} \quad (2)$$

Table 1: An example of user-service QoS matrix

	$service_1$	$service_2$	$service_3$	$service_4$	$service_5$
$user_1$	2	4	2	4	5
$user_2$	1	2	1	2	5
$user_3$	2	4	2	$\emptyset$	$\emptyset$
$user_4$	2	2	2	1	4
$user_5$	5	5	5	4	$\emptyset$
$user_6$	3	2	3	3	1
$user_7$	3	2	3	4	$\emptyset$
$user_8$	3	1	$\emptyset$	$\emptyset$	$\emptyset$

Table 1 is an example of QoS matrix which contains 5 services ($service_1$ to $service_5$) and 8 users ($user_1$ to $user_8$). The values from 1 to 5 are the minimum to the maximum QoS-values of the user-service matrix. $\emptyset$ denotes the user has never invoked the corresponding service before.

We calculate similarity value adopting COS approach (Eq. (2)), and get the following arithmetic expression:

$$Sim(user_3, user_1) = Sim(user_3, user_2)$$

It indicates that $user_1$ is similar to $user_3$ as much as $user_2$ is. Actually, $user_1$ is more similar to $user_3$ than $user_2$, which can be easily observed according to the values in Table 1. Hence, calculation results and facts are contradictory. We can also compute the similarity between $user_4$ and $user_5$:

$$Sim(user_4, user_5) = 0.9885$$

We can draw conclusion that $user_4$ and $user_5$ are very similar according to this computation result. However, this is in conflict with the fact shown in Table 1, since $user_4$ and $user_5$ almost have the opposite QoS-values. $user_5$'s QoS-values are approximated to the maximum QoS-value 5 in the user-service matrix while $user_4$'s values are approximated to the minimum QoS-value 1 of the matrix. This contradiction arises from that when measuring the similarity between two vectors, COS only considers the angle between two vectors and does not consider the length of vector.

LetĄŻs consider another example. If we employ PCC defined in Eq.(1) to measure the similarity between users, we can get the following result:

$$Sim(user_7, user_6) < Sim(user_7, user_8)$$

It indicates that $user_6$ is less similar to $user_7$ than $user_8$ is. Actually, $user_6$ is more similar to $user_7$ than $user_8$, because there are the same value in three dimensions and a difference of 1 in one dimension for $user_7$ and $user_6$, and there are the same value in one dimension and a difference of 1 in one dimension for $user_7$ and $user_8$. Therefore, the calculation results are in conflict with the facts. This contradiction arises from that PCC does not consider the number of services co-invoked which implies the similarity of selection preference and style between users.

In addition to above mentioned approaches, some other similarity measurements are also proposed such as rated-item pools (RIPs) user similarity [22], proximity impact popularity (PIP) measure [13], and mean squared difference [14], which are either for special purposes, or for special situations, or not used widely. Among the traditional similarity measurement approaches, the approaches that we elaborated above are strongly representative.

3 Feedback-corrected Tan-NED CF

3.1 Tan-NED Similarity Measurement

In order to address the issues of existing similarity measurement, this paper proposes a novel similarity measurement named Tanimoto [23] Normalized Euclidean Distance (Tan-NED). Compared with the traditional similarity measurement approaches, our approach measures the similarity based on normalized Euclidean distances of difference multidimensional vector spaces.

Although Euclidean-distance approach can also be employed to measure similarity, Tan-NED is completely different from it. Since the number co-invoked services is different deal with different couples of users in a recommender system, the Euclidean distances of different couples of users tend to compute in the different dimensions of vector spaces. Moreover, the maximal values of Euclidean distances in different vector spaces are usually very different. A value which denotes the maximal value in one vector space may be a very small Euclidean distance in another vector space. Therefore, putting them together to measure similarity is meaningless. For instance, if users a, b both invoked the same 10 services while users c, d both invoked the same 300 services, $dist(a, b)$ is the Euclidean distance between the two users a and b, $dist(c, d)$ is the Euclidean distance between the two users c and d, then it will be meaningless to mention $dist(a, b)$ and $dist(c, d)$ in the same breath, because $dist(a, b)$ is calculated in 10-dimension vector space, and

$dist(c, d)$ is computed in 300-dimension vector space. Therefore, to measure the similarity of vectors we should consider the difference of dimension-number.

Our NED normalizes the QoS values of different users to the same range in order to address the different maximal-value issue. Then, it unifies the similarity metrics of different vector spaces. Before measuring the similarity between users, it uses the maximal and minimal QoS-value of each row to normalize every value of the same row in the original matrix M. After that, the QoS-values of each row are normalized to $[0, 1]$. In consequence, the original QoS value matrix M is mapped into a row-normal matrix M_{nu}. Assuming that the number of co-invoked services by user u, v is num, and user u, v have the observed QoS-value vectors $\vec{u}, \vec{v}$ respectively in matrix M_{nu}. Then, the Euclidean distance between vector $\vec{u}$ and $\vec{v}$ in the num dimensions can be calculated by $dist(u, v)$. Besides, the maximal Euclidean distance of the num dimensions are calculated by $dist_{max}$. Since the matrix M_{nu} has been normalized, each dimension ranges from 0 to 1, and the maximal distance of each dimension is 1. In M_{nu}, $nr_{(u,i)}$, $nr_{(v,i)}$ are the QoS-value of service i towards user u, v respectively. The similarity between user u and v can be calculated by NED as follows:

$$nr_{u,i} = \frac{r_{u,i} - r_{umin}}{r_{umax} - r_{umin}}, nr_{v,i} = \frac{r_{v,i} - r_{vmin}}{r_{vmax} - r_{vmin}}$$

$$dist(\vec{u}, \vec{v}) = \sqrt{\sum_{i \in I}(nr_{u,i} - nr_{v,i})^2}$$

$$Sim_{ned}(u, v) = 1 - \frac{dist(\vec{u}, \vec{v})}{dist_{max}} = 1 - \frac{\sqrt{\sum_{i \in I}(\frac{r_{u,i}-r_{umin}}{r_{umax}-r_{umin}} - \frac{r_{v,i}-r_{vmin}}{r_{vmax}-r_{vmin}})^2}}{\sqrt{\sum_{k=1}^{|I|}(1-0)^2}}$$

i.e.,

$$Sim_{ned}(u, v) = 1 - \frac{\sqrt{\sum_{i \in I}(\frac{r_{u,i}-r_{umin}}{r_{umax}-r_{umin}} - \frac{r_{v,i}-r_{vmin}}{r_{vmax}-r_{vmin}})^2}}{\sqrt{|I|}} \quad (3)$$

Where $I = I_u \cap I_v$ is the set of services which is co-invoked by user u and v; $|I|$ is the number of I; r_{umin} and r_{umax} are the minimal and the maximal QoS-values of user u in the original matrix M, $r_{(u,i)}$ denotes the QoS-value of service i towards user u in M. The results of Eq.(3) range from 0 to 1, i.e., $Sim_{ned}(u, v) \in [0, 1]$, where $Sim_{ned}(u, v) = 0$ represents that two users are dissimilar and $Sim_{ned}(u, v) = 1$ indicates that these two user are exactly similar even the same.

Further, in order to use more information of the two users, we also take the number of invoked services by each user and that by both users which implies the QoS preference and style of users into account. We propose Tan-NED which combines Tanimoto similarity coefficient [23] with NED. The formula of Tan-NED is as follow:

$$Sim(u, v) = \frac{|I|}{|I_u| + |I_v| - |I|} \times Sim_{ned}(u, v)$$

i.e.,

$$Sim(u, v) = \frac{|I|}{|I_u| + |I_v| - |I|} \times (1 - \frac{\sqrt{\sum_{i \in I}(\frac{r_{u,i}-r_{umin}}{r_{umax}-r_{umin}} - \frac{r_{v,i}-r_{vmin}}{r_{vmax}-r_{vmin}})^2}}{\sqrt{|I|}}) \quad (4)$$

All the contractions mentioned in section 2 can be eliminated by our Tan-NED. Adopting Eq.(4) we get the following results which are consistent with the facts:

$$Sim(user_3, user_1) > Sim(user_3, user_2)$$

$$Sim(user_4, user_5) = 0.3381$$

$$Sim(user_7, user_6) > Sim(user_7, user_8)$$

3.2 Feedback-corrected Tan-NED Collaborative Filtering

Tan-NED can calculate the similarity between two users, based on it, a novel memory-based CF approach named Tan-NED CF is proposed. Tan-NED CF predicts the unknown QoS-value $r_{(u,i)}\prime$ of service i towards user u based on the already available QoS-values of service i towards other users that are similar with user u. The more similar user v to user u is, the greater user v's QoS-value influences on $r_{(u,i)}\prime$. The normalized predicting value $\hat{r}_{(u,i)}\prime$ can be calculated by Eq.(5), and then we recovers the normalized value to the original scale of user u by the maximal and minimal values of user u. The QoS-value predicted by Tan-NED CF is defined as follow:

$$\hat{r}_{(u,i)}\prime = \frac{\sum_{v \in U} Sim(u,v) \times nr_{v,i}}{\sum_{v \in U} Sim(u,v)} \tag{5}$$

$$r_{(u,i)}\prime = r_{umin} + (r_{umax} - r_{umin})\hat{r}_{(u,i)}\prime \tag{6}$$

U is the set that contains the similar users to user u. Each element $v \in U$ has also invoked service i. $nr_{v,i}$ denotes the normalized QoS-value of user v on service i in matrix M_{nu} which is row-normal. $Sim(u,v)$ can be calculated by Eq.(4), and r_{umax} and r_{umin} are the maximal and minimal QoS-values of user u in the original matrix M. Then we employ the feedback (i.e., the difference value between prediction QoS value and real QoS value of invoked services) to correct the following prediction. The correction is a continuous process, so uncertain abnormal real QoS values may appear which cause "dirty" feedback to the correction process. For example, if a users router is congested then the RTT of service invoked by this user will become very long, which will not response the real QoS of service. In order to reduce the impact of "dirty" feedback, we adopt a Gaussian distribution coefficient to control the correction efforts of feedback. The QoS value prediction of feedback corrected Tan-NED is as follow:

$$r_{u,i} = \Delta r \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(\Delta r - \frac{\Delta r_{min} + \Delta r_{max}}{2})^2}{2\sigma^2}} + r_{u,i}\prime \tag{7}$$

Where Δr denotes the difference value between real QoS and prediction QoS of the previous service invoked by user u, i.e., $\Delta r = r_{real} - r_{pred}$. Δr_{min} is the minimum value of historical Δr and Δr_{max} is the maximum value of historical Δr. σ is the standard deviation of Gaussian distribution which determines the amplitude of distribution, and an appropriate σ value will be obtained by experiments. The feedback correction is a continuous recursive process along with every service invoking.

3.3 Feedback-corrected Tan-NED CF for Service Recommendation

In the case of candidate services having equivalent functions, the predicted QoS-values which are calculated by Feedback-corrected Tan-NED can be used for service recommendation, and the service that has the best predicted QoS performance will be recommended to the corresponding user. Then, our approach enables personalized service recommendation using a small amount of available QoS value without the time-consuming and costly actual service invocation. In service recommender system, either services or users may be remotely distributed in different location. Besides, the network performances which influence the QoS of services are highly dynamic. Hence, the QoS styles and preferences of user are quite different from each other. Since the proposed approach considers the diversity of QoS style and preference and adopts the information of similar users to make the prediction, it is applicable to a variety of environments.

4 Evaluative Experiments

4.1 QoS Dataset

In order to has sufficient data to evaluate our approach, we use the web-service QoS set [24] which contains 1873838 real RTT (route trip time) records on 5825 real web-services from 339 distributed service users. To our knowledge, this is the largest dataset in the domain of service-computing. In order to collect the data, Zheng et al. monitor 5825 web-services using 339 distributed planet-lab computers. Assuming that a $M \times N$ user-service matrix contains N services, M users and T non-null records, then the density of this matrix can be defined as $density = \frac{T}{(M \times N)}$. According to this concept, the density of user-service matrix used for evaluation is 94.9%.

4.2 Evaluation Metric

We adopt the MAE (Mean-Absolute-Error) metric to evaluate the prediction accuracy of proposed approach. MAE denotes the average-absolute-deviation between the ground-truth values and the predictions values. It is defined as follow:

$$MAE = \frac{\sum_{u,i} |\bar{r}_{u,i} - r_{u,i}|}{N} \tag{8}$$

$r_{u,i}$ is the prediction RTT of web-service i towards user u, $\bar{r}_{u,i}$ is the real RTT of i observed by u, N is the number of predicted RTTs. The less the value of MAE is, the better the accuracy of prediction is.

4.3 Experimental Setup

The experiments are divided into three parts, namely, performance of similarity measures, impact parameters of prediction, and performance of prediction approaches. First, we compare the performance of our Tan-NED with other similarity measures. In this experiment, we use the original user-service matrix. The RTT records of matrix are divided into two parts: 80% of the records as the training set and 20% of the records as the test set.

Then, we measure the impact of σ (controls the correction effort of feedback), neighbour size k (top-k similar users to calculate the prediction QoS value), and density of matrix. In the third part, we compare the performance of our feedback-corrected Tan-NED with other prediction approaches. In order to evaluate the accuracy of RTT value prediction by different prediction algorithms, the user-service RTT records in the original matrix is removed randomly to generate ten sparse matrices. As Section 4.1 defined, the densities of these ten matrices are incremental with the step-size of 2%, their densities range from 2% to 20%. We adopt these small density matrices in order to get closer to the practical situation that a user may only invokes limited number of services in large amount of available services. So, the real user-service matrix is generally very sparse. We divide each of the ten matrices into three parts: 70% of the RTT records as the training set, 10% as the feedback set, 20% as the test set.

4.4 Performance Evaluation of Similarity Measurements

In order to validate the validity of proposed Tan-NED similarity measurement (Eq.(4)), we compare its performance with other two well-known similarity measurements: COS (Eq.(2)) and PCC (Eq.(1)). Then, we combine Tan-NED, COS, and PCC with formula Eq.(6) to predict the missing RTT values respectively. Figure 1 presents the accuracy of predictions by different similarity measurements. As Figure 1 shows, the proposed Tan-NED consistently outperforms other compared approaches under different k-values; even the worst-case of Tan-NED still over-match the best cases of the compared approaches. Therefore, compared with COS and PCC, the proposed Tan-NED significantly improves the accuracy of prediction.

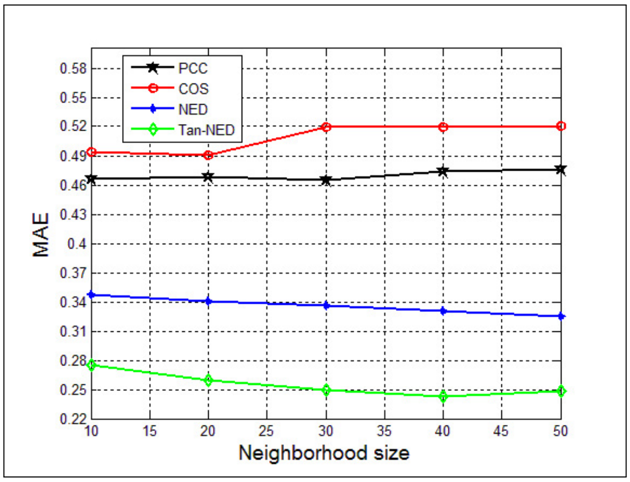


Figure 1: Performance comparison of similarity measures

4.5 Impact of Parameters

Impact of σ

The parameter σ determines the amplitude of distribution which controls the correction efforts of feedback. The higher the value, the more flat the distribution, which means that the probability distribution of Δr ($r_{real} - r_{pred}$) is dispersive. Then, the correction efforts of different Δr value are approximate. Whereas, the lower the value, the steeper the distribution. It means that a few values which close to the median of Δr have strong efforts, however, other values have relatively weak efforts. In this experiment, we increase the parameter σ from 0.1 to 1.0 with the step-size of 0.1 in order to study the impact of ŚŇ on the prediction result. The original user-service matrix is adopted, and the neighbour size k is set to 30. The influence of parameter σ on prediction is presented in Figure 2. As it shows, the values of MAE first slightly decline and then sharply rise. When $\sigma = 0.4$, it hits the bottom. This experiment result indicates that the prediction accuracy can be improved by adjusting the amplitude of correction-effort distribution.

Impact of neighbour size k

In the proposed Tan-NED approach, the neighbour size k determines the number of similar users used for missing value prediction. It acts an important role in the prediction performance. If the value of k is too low, many similar users will be filtered out. If the value of k is too high, dissimilar usersĄŹ records will be considered to calculate the prediction QoS value. The neighbor size k is increased from 10 to 100 with the step-size of 10 in order to study the impact

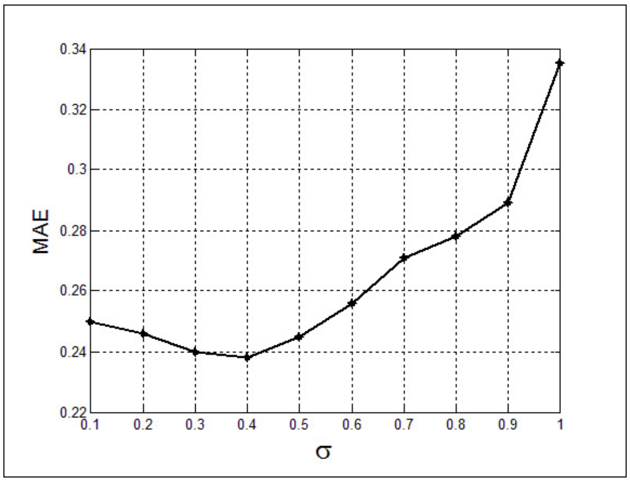


Figure 2: Impact of Sigma

of k. The original user-service matrix is employed, and the parameter σ is set to 0.4. The impact of k on the prediction of Tan-NED is presented in Figure 3. As it shows, the value of MAE first slightly declines and then slightly rises, indicating that our Tan-NED achieves best performance for this dataset when $k = 50$. The deviation between the highest value and the lowest value is only 0.041, which means that our approach is not sensitive to the neighbour size. It is because that our approach uses the value of similarity degree to restrict the effects of each similar user when calculating the prediction QoS value (shown in Eq.(5)).

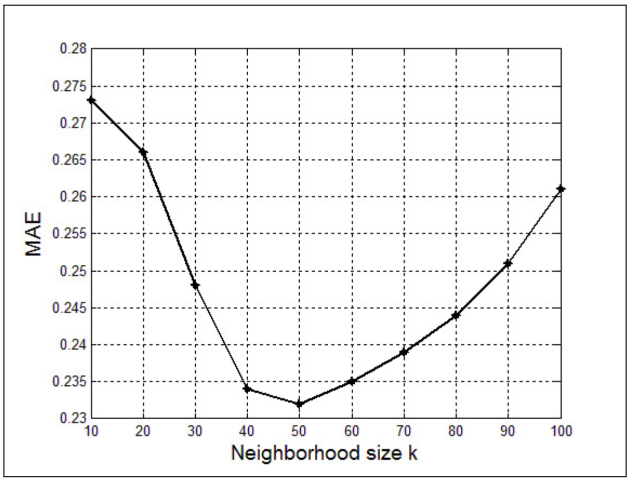


Figure 3: Impact of neighbour size

Impact of User-service Matrix Density

Matrix density denotes the proportion of non-null records that can be used for missing value prediction in the matrix. This section studies the impact of the matrix density on the accuracy of Tan-NED. In this experiment, the density is increased from 0.04 to 0.2 with the step-size of 0.02. The parameter σ is set to 0.4 and k is set to 50. The impact of matrix density on the accuracy of Tan-NED prediction is presented in Figure 4. As it shows, the value of MAE declines as the density increase. The result of this experiment denotes that the prediction of Tan-NED becomes more accuracy as the matrix density increase. The reason is that a more intensive matrix provides more reference information for Tan-NED to predict missing value.

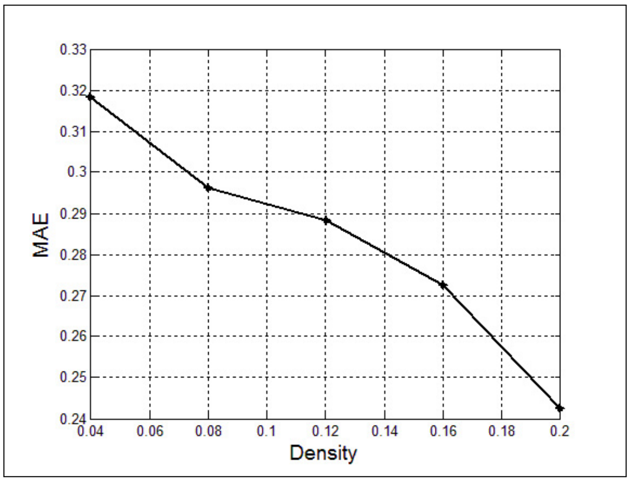


Figure 4: Impact of user-service matrix density

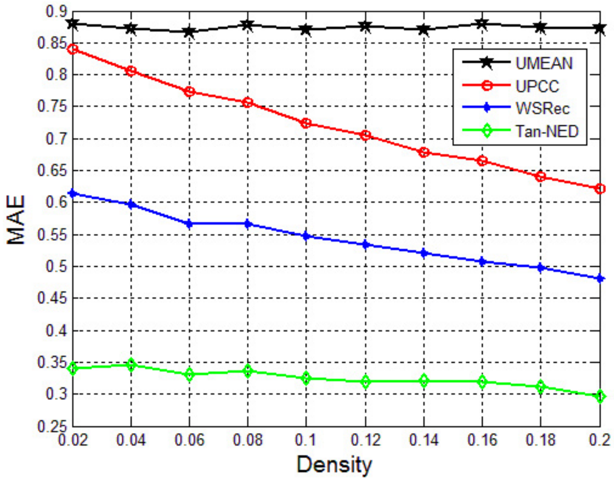
4.6 Performance Evaluation of Prediction Approaches

We compare the proposed feedback corrected Tan-NED with other three prediction approaches: UPCC (User-based CF adopting PCC), UMEAN (User-Mean) and WSRec (Web-services recommender) in order to validate its effectiveness. UPCC refers the information of similar users to predict the missing value [7, 19]. WSRec [11] is a novel memory-based CF for web-service recommendation, which achieves a relatively good performance. UMEAN uses the average RTT values of other web-services from the same user to make missing value prediction. The predictions of these four approaches are influenced by the neighbour size k. The neighbour size is increased from 10 to 50 with the step-size of 20 in this experiment. The parameter σ of Tan-NED is set to 0.4 and the confidence weight of WSRec is set to 0.1[1].

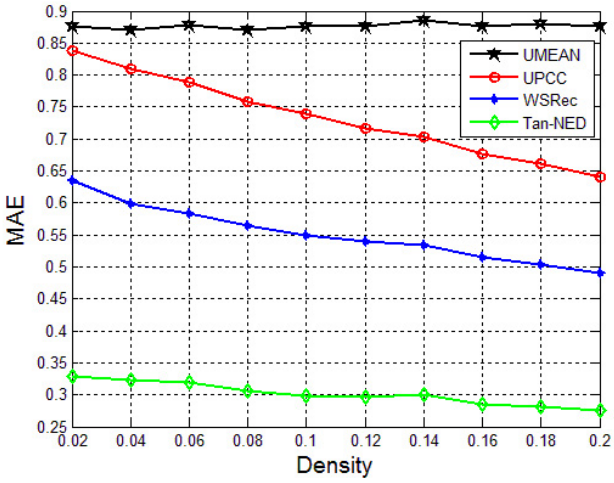
Figure 5 presents the prediction accuracy measured by MAE of the evaluated approaches. The three subfigures of Figure 5 correspond to the neighbour-size of 10, 30, and 50 respectively. We increase the density of matrix from 0.02 to 0.2 with the step-size of 0.02. Each subfigure shows the value of MAE with the matrix density changes. As shown in Figure 5, our feedback corrected Tan-NED is significantly superior to other compared approaches. When the density of user-service matrix becomes sparser, the co-valued dimensions between vectors decrease. It means that the number of available values used for missing value prediction is limited, which expands the gap of performance between Tan-NED and compared approaches. As the increase of density, the improvement rate of Tan-NED declines due to each approach has enough available values for prediction. However, in practical situation, the user-service matrix is usually very sparse. Moreover, as each subfigure shows, the deviation between the maximum MAE and the minimum MAE of tan-NED is small. It means that Tan-NED keeps a stable MAE performance under the different density of user-service matrix.

The results of this experiment indicate two features of Tan-NED: 1) compared with other existing approaches, the sparser the user-service matrix is, the more superior of Tan-NED is; 2) the performance of Tan-NED is insensitive to the decrease of matrix density, namely, even with limited available QoS records Tan-NED can also make relatively accurate prediction. These two features of Tan-NED are suitable for the actual situation of real-world service environment. In

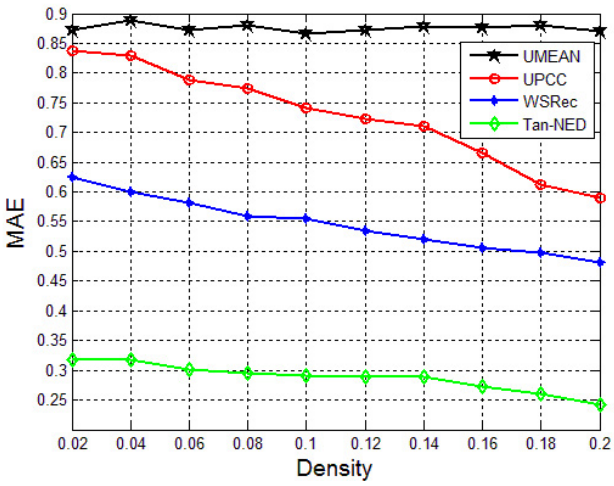
[1] The confidence weight in WSRec denotes the impact of user-based method on the final prediction result. As [11] discussed, when confidence weight is set to 0.1, it achieves the best performance.



(a) Set the number of neighbour to 10



(b) Set the number of neighbour to 30



(c) Set the number of neighbour to 50

Figure 5: The Performance of compared prediction approaches

real-world service recommendation, the density of matrix is generally very sparse. Therefore, compared with existing approaches, our approach can make more accurate and stable prediction for QoS-value.

5 Conclusion

This paper proposes a feedback corrected Tan-NED approach to solve the issue of real-world service personalized recommendation. It studies the features of the QoS-values of real-world services, and proposes a novel similarity measurement which seeks similar users more accurately and provides a basis for accurate QoS-value prediction. Then, the proposed approach can use a small number of available QoS-values from similar users to predict the service QoS-value for the user according to his personalization. In the service recommender system, the proposed approach assists service users to select the service with optimal QoS from a number of function-equivalent services instead of conducting the costly actual service invocation. In order to evaluate the performance of feedback corrected Tan-NED, this paper conducted comprehensive evaluative experiments using a real-world web-service dataset which has sufficient QoS records. Experiment results indicate that compared with existing approaches the proposed approach improves the accuracy of QoS-value prediction significantly.

Since dynamic is a new feature of IoT environment, the QoS of real-world service changes with time frequently. In the future work, we will focus on enhancing the efficiency of our approach to handle the dynamic QoS issue. It may be addressed by using the latest advanced technologies of machine learning.

Acknowledgement

This study is supported by 973 program of National Basic Research Program of China (Grant No. 2011CB302704 and 2012CB315802). National Natural Science Foundation of China (Grant No. 61001118, 61171102); Program for New Century Excellent Talents in University (Grant No. NECT-11-0592); Project of New Generation Broadband Wireless Network under Grant (Grant No.2010ZX03004-001).

Bibliography

[1] Gustafaason, J. (2011); Integration of wireless sensor and actuator nodes with IT infrastructure using service-oriented architecture, *IEEE Trans Industrial Informatics*, ISSN 1551-3203, 6(1): 1-10.

[2] Guinard, D.; Trifa, V. (2010); Interacting with the SOA-Based Internet of Things: Discovery, Query, Selection, and On-Demand Provisioning of Web Services, *IEEE Trans Services Computing*, ISSN 1939-1374, 3(3): 223-235.

[3] ICT FP7 OPEN IoT Project. Open source solution for the internet of things into the cloud, (2011). http://vmusm03.deri.ie/.

[4] EPFL GSN project (2009). http://sourceforge.net/apps/trac/gsn/.

[5] Cosm. Cosm platform, (2007). https://cosm.com/.

[6] Perera, C.; Zaslavsky, A.; Christen, P. (2013). Context-aware sensor search, selection and ranking model for internet of things middleware. *14th IEEE International Conference on Mobile Data Management*, 314–322.

[7] Sreenath, R.M.; Singh, M.P. (2003); Agent-based service selection, *J Web Semantics*, ISSN 1570-8268, 1(3): 261-279.

[8] Zhang, L.J.; Zhang, J.; Cai, H. (2007) Services computing, *Springer and Tsinghua University Press*, ISSN 0895-4852.

[9] Moser, O.; Rosenberg, F.; Dustdar, S. (2008). Non-intrusive monitoring and service adaptation for ws-bpel, *17th Intl Conf. on World Wide Web*, 815–824.

[10] Papazoglou, M; Georgakopoulos, D. (2003). Service-oriented computing, *Communications of the ACM*, ISSN 0001-0782, 46(10): 25–28.

[11] Zheng, Z.; Ma, H.; Lyu, M.R.; King, I. (2009). Wsrec: A collaborative filtering based web service recommender system, *7th Intl Conf. Web Services*, 437–444.

[12] Resnick, P.; Iacovou, N.; Suchak, M.; Bergstrom, P.; Riedl, J. (1994). Grouplens: An open architecture for collaborative filtering of net news, *ACM Conf. Computer Supported Cooperative Work*, 175–186.

[13] Shardanand, U.; Maes, P. (1995). Social information filtering: Algorithms for automating word of mouth, *SIGCHI Conf. Human Factors in Computing Systems*, 210–217.

[14] Sarwar, B.; Karypis, G.; Konstan, J.; Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms, *10th Intl Conf. World Wide Web*, 285–295.

[15] Breese, J.; Heckerman, D.; Kadie, C. (1998). Empirical analysis of predictive algorithms for collaborative filtering, *14th Intl Conf. on Uncertainty in artificial intelligence*, 43–52.

[16] Adomavicius, G.; Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions, *IEEE Trans on Knowledge and Data Engineering*, ISSN: 1041-4347, 17: 734–749.

[17] Chen, X.; Zheng, Z.; Liu, X.; Huang, Z.; Sun, H. (2013). Personalized qos-aware web service recommendation and visualization, *IEEE Trans on Service Computing*, ISSN: 1939-1374, 6(1):35-47.

[18] Karta, K. (2005). An investigation on personalized collaborative filtering for web service selection. *Honours Programme thesis, University of Western Australia.*

[19] Shao, L.S.; et al. (2007). Personalized qos prediction forweb services via collaborative filtering. *Intl Conf. on Web Services*, 439–446.

[20] Miller, B.; Albert, I.; Lam, S.; Konstan, J.; Riedl, J. (2003). MovieLens unplugged: Experiences with an occasionally connected recommender system. *8th International Conference on Intelligent User Interfaces*, 263–266.

[21] Zhao, S.; Zhang, Y.; et al. (2013). A multidimensional resource model for dynamic resource matching in internet of things. *Concurrency and Computation: Practice Experience.*

[22] Thio, N.; Karunasekera, S. (2005). Automatic measurement of a qos metric for web service recommendation, *Australian Software Engineering Conference*, 202–211.

[23] Lipkus, A.H. (1999). A proof of the triangle inequality for the Tanimoto distance, *Journal of Mathematical Chemistry*, ISSN: 0259-9791, 263-265.

[24] Zheng, Z.; Ma, H.; Lyu, M.R.; King, I. (2011). QoS-aware Web service recommendation by collaborative filtering, *IEEE Trans on Service Computing*, ISSN: 1939-1374, 4(2): 140-152.

INT J COMPUT COMMUN, ISSN 1841-9836
9(3):370-378, June, 2014.

Comparison and Weighted Summation Type of Fuzzy Cluster Validity Indices

K.L. Zhou, S. Ding, C. Fu, S.L. Yang

Kaile Zhou*, Shuai Ding, Chao Fu, Shanlin Yang
School of Management
Hefei University of Technology
Hefei 230009, China
*Corresponding author: zhoukaile@mail.hfut.edu.cn
E-mail: dingshuai@hfut.edu.cn,
wls_fuchao@163.com, yangsl@hfut.edu.cn

Abstract: Finding the optimal cluster number and validating the partition results of a data set are difficult tasks since clustering is an unsupervised learning process. Cluster validity index (CVI) is a kind of criterion function for evaluating the clustering results and determining the optimal number of clusters. In this paper, we present an extensive comparison of ten well-known CVIs for fuzzy clustering. Then we extend traditional single CVIs by introducing the weighted method and propose a weighted summation type of CVI (WSCVI). Experiments on nine synthetic data sets and four real-world UCI data sets demonstrate that no one CVI performs better on all data sets than others. Nevertheless, the proposed WSCVI is more effective by properly setting the weights.
Keywords: fuzzy clustering, fuzzy c-means (FCM), cluster validity indices (CVIs), WSCVI.

1 Introduction

Clustering [1] is an unsupervised learning process to discover significant patterns in a given data set by partitioning a data set into groups (i.e., clusters) such that the elements assigned to the same group are as similar as possible while those in different groups are dissimilar in some sense. Clustering is an unsupervised process, the data objects in a data set are typically unlabeled and no structural knowledge about the data set is available [2]. Therefore, evaluating the quality of clustering results and determining the optimal number of clusters are difficult tasks. Also, the number of clusters is a prerequisite input parameter for many clustering algorithms [3].

Cluster validity index (CVI) is a kind of criterion function to determine the optimal number of clusters [3]. Currently, a large number of CVIs have been proposed [4,5]. So it is necessary to evaluate and compare the performances of these CVIs. Extensive comparisons of crisp CVIs have been presented [6,7], while few studies have focused on the performance comparison of CVIs for fuzzy clustering. Most comparison studies of fuzzy CVIs are presented when a new fuzzy CVI was proposed [8,9], but the extent of these comparisons was limited. In this paper, we present an extensive comparative study of ten well-known fuzzy CVIs.

Previous studies on CVIs have demonstrated that there is no single CVI that can deal with any data sets and always perform better than the others [10,11]. The idea of weighted CVIs has been mentioned in literature [12]. However, few studies have focused on the weighted summation type of CVIs for fuzzy clustering. Hence, in this paper we propose a weighted form of fuzzy clustering CVIs which is the weighted sum of ten well-known fuzzy CVIs.

The remainder of this paper is organized as follows. Section 2 reviews the fuzzy c-means clustering algorithm and ten well-known CVIs for fuzzy clustering. Then, in Section 3, we introduce the weighted summation type of fuzzy clustering CVI. Finally, experimental results are presented in Section 4. The conclusions are drawn in Section 5.

2 Fuzzy C-means and Fuzzy CVIs

2.1 Fuzzy c-means algorithm

Fuzzy c-means (FCM) algorithm [13] starts with determining the number of clusters followed by guessing the initial cluster centers. Each cluster center and corresponding membership degrees are updated iteratively by minimizing the objective functions until the termination criterion is met. The objective function of FCM is defined as:

$$J_m(U,V) = \sum_{i=1}^{c}\sum_{j=1}^{n} \mu_{ij}^{m} d_{ij}^{2} \tag{1}$$

where U denotes the membership, V is the cluster center matrix, n is the number of data objects, c is the number of clusters, m is the fuzzifier in FCM, v_i is the center of cluster i, μ_{ij} is the membership degree of the jth data object x_j to v_i, d_{ij}^2 is the Euclidean distance of x_j to v_i, and $d_{ij}^2 = ||x_j - v_i||^2$.

See Ref. [13] for the iterative formulas of membership degree μ_{ij} and cluster centers v_i.

2.2 CVIs for fuzzy clustering

Proposed by Bezdek [14] in 1974, PC is the first CVI used for FCM clustering, which was defined as

$$PC = \frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n} \mu_{ij}^{m} \tag{2}$$

The optimal cluster number c^* is obtained with the maximum value of PC.

To reduce the monotonic trend of PC index with the increase of the cluster numbers, a normalized form of PC, called NPC [15], was defined as

$$NPC = 1 - \frac{c}{c-1}(1 - PC) \tag{3}$$

The optimal number of cluster c^* is also found when NPC reach the maximum value.

The concept of entropy was introduced in PE index by Bezdek [16]. Much like PC index, PE index is defined as

$$PE = -\frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n} \mu_{ij} \log_{\alpha} \mu_{ij} \tag{4}$$

where α is the base of the logarithm. The optimal cluster number c^* is determined by the minimum value of PE index.

Like NPC index, NPE [17] was the normalized form of PE index to reduce the monotonic tendency of PE index and was defined as

$$NPE = \frac{n}{n-c} PE \tag{5}$$

Like PE index, the optimal cluster number is corresponding to the minimum value of NPE index.

Different from PC, NPC, PE and NPE index which only considered the membership degree elements in U, the XB index [18] consists of both the membership degree values and the information about data set itself. XB index is defined as

$$XB = \frac{\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^2 d_{ij}^2}{n \times \min_{i \neq j} ||v_i - v_j||^2} \tag{6}$$

The optimal cluster number is found at the minimum value point of XB index.

Kwon [19] extended XB index and proposed a new CVI, VK, which is defined as

$$VK = \frac{\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^2 d_{ij}^2 + (1/c)\sum_{i=1}^{c}||v_i - \overline{v}||^2}{\min_{i \neq j} ||v_i - v_j||^2} \tag{7}$$

In order to determine the best clustering results and the optimal cluster number c^*, we should find the most compact and separate partition, that is, the minimum value of VK index.

Pakhira et al. [20] proposed a CVI, known as PBM index, for crisp clustering, and a corresponding form for fuzzy clustering, called PBMF index. The definition of PBMF index is

$$PBMF = (\frac{1}{c} \times \frac{E_1}{E_c} \times D_c)^2 \tag{8}$$

where $E_c = \sum_{i=1}^{c} E_i$, $E_i = \sum_{j=1}^{n} \mu_{ij} d_{ij}$, $D_c = \max_{i,j=1}^{c} ||v_i - v_j||^2$

The optimal cluster number is found when the maximum value of PBMF index is achieved.

Different from the ratio type of CVIs, Fukuyama and Sugeno [21] proposed a summation type of CVI called FS index. Its definition is

$$FS = \sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^m (d_{ij}^2 - ||v_i - \overline{v}||^2) \tag{9}$$

The optimal cluster number c^* is achieved at the minimum value of FS index.

A new CVI, referred to VT index, was proposed by Tang et al. [22] based on the idea of penalty function of Kwon's index VK. VT index is defined as

$$VT = \frac{\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^2 d_{ij}^2 + \{1/[c(c-1)]\}\sum_{i=1}^{c}\sum_{k=1;k \neq i}^{c}||v_i - v_k||^2}{\min_{i \neq k} ||v_i - v_k||^2 + 1/c} \tag{10}$$

The optimal cluster number c^* is also found at the minimum value of VT index.

A CVI proposed by Bensaid et al. (SC) [23] is defined as

$$SC = \sum_{i=1}^{c} \frac{\sum_{j=1}^{n}\mu_{ij}^2 d_{ij}^2}{\sum_{j=1}^{n}\mu_{ij}\sum_{k=1}^{c}||v_i - v_k||^2} \tag{11}$$

The optimal cluster number c^* is determined by the minimum value of SC index.

3 Weighted Summation Type of Fuzzy CVIs

In order to take advantage of each fuzzy CVI and weaken its limitation, we proposed a weighted summation type of CVI (WSCVI), which is the weighted sum of the above ten fuzzy CVIs. WSCVI is defined as

$$WSCVI = \sum_{i=1}^{N} \omega_i \cdot CVI_i \tag{12}$$

where N is the number of CVIs. ω_i is the weight of index CVI_i, which represents the relative importance of the ith CVI. ω satisfies $0 \le \omega_i \le 1$, $\sum_{i=1}^{N} \omega_i = 1$.

CVI_i is one of the ten above CVIs for fuzzy clustering. Among them, PC, NPC and PBMF index are maximum type indices, i.e., the optimal cluster number c^* is achieved at the maximum value of these CVIs, while other seven indices are minimum type. In order to obtain c^* when WSCVI is minimum, we convert the three maximum type indices into their corresponding reciprocal types, which are presented as $PCr = 1/PC$, $NPCr = 1/NPC$, $PBMFr = 1/PBMF$. The values of different CVIs change in different range. To overcome the dominate influence of CVIs in large values, all the CVIs are normalized so that all of their values range in [0, 1].

The corresponding cluster number is optimal cluster number c^* when the values of the normalized CVIs equal to 0, and the minimum value of WSCVI is achieved.

4 Experimental Results

Nine synthetic data sets (six 2-D data sets and three 3-D data sets) and four real-world data sets were used in the experiments. The value of fuzzifier in FCM is set m=2. We suggest $\omega_1 = \omega_2 = \cdots = \omega_N = 1/N$ when there is no prior knowledge available. In the experiments, we also set some other weights to obtain the optimal cluster numbers.

The synthetic data sets are expressed as Data_d_c_n, in which d is the dimension of the data set, c is the number of clusters in the data set, and n is the total number data objects in the data set.

4.1 Data sets

Three well-known 2-D synthetic data sets, Butterfly, Example_01 and Example_02 presented in [19] and the other three 2-D synthetic data sets, Data_2_3_60, Data_2_3_70, and Data_2_4_110 [24], are shown in Figure 1 (a) to (f), respectively. Figure 2 shows the three 3-D synthetic data sets, Data_3_3_200, Data_3_3_300, and Data_3_4_320, respectively.

We also use four real-world data sets, *bupa*, *wdbc*, *iris* and *glass* data set, from UCI Machine Learning Repository [25] to test the performance of WSCVI and the ten single CVIs.

4.2 Results

The optimal cluster numbers found by the ten single CVIs and the proposed WSCVI with equal weights of each CVI are shown in Table 1. It can be seen from Table 1 that there is no one CVI that can find the optimal cluster numbers for all of the data sets. WSCVI with equal weights, $1/N$, found the correct optimal cluster numbers except Data_3_3_300, Data_3_4_320, Iris, and Glass. In order to find the optimal cluster numbers of Data_3_3_300, Data_3_4_320, Iris, and Glass using WSCVI, we adjust the weights of each CVI in WSCVI.

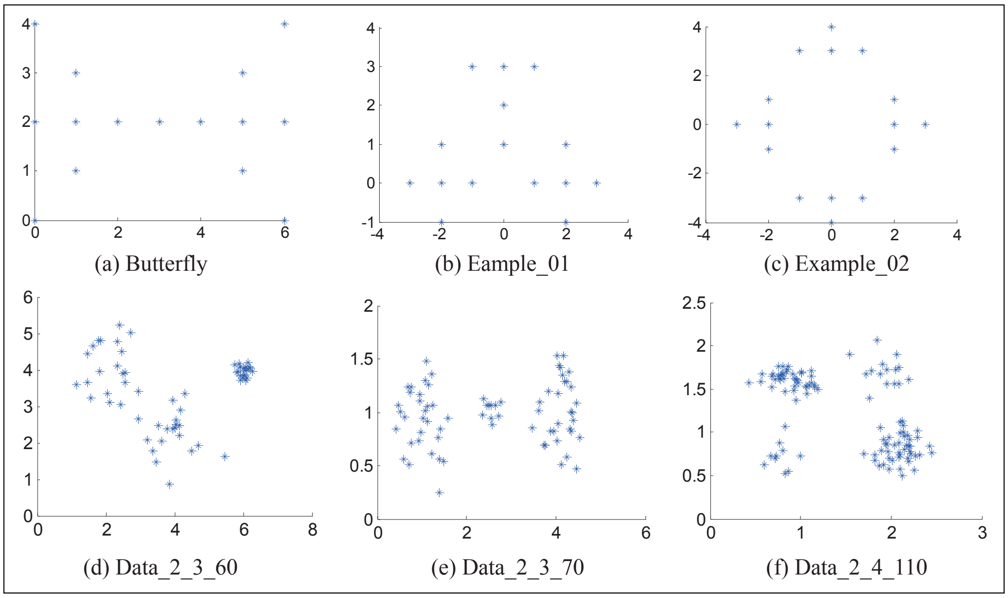


Figure 1: Six synthetic 2-D data sets

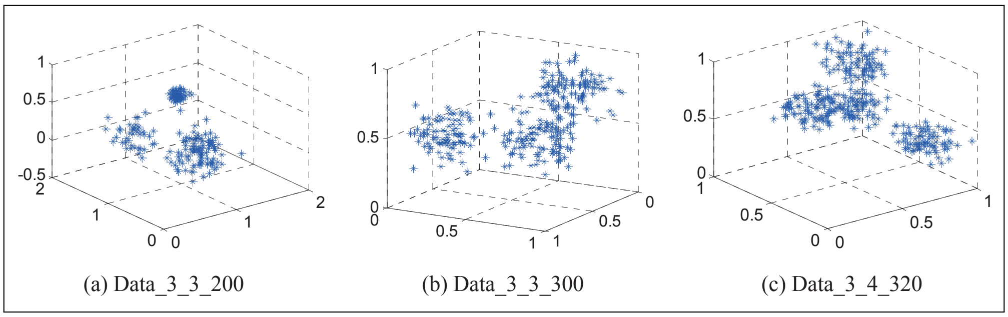


Figure 2: Three synthetic 3-D data sets

For Data_3_3_300 and Data_3_4_320, we set the weight of SC index ω_{SC}=0.5, the weights of the other CVIs are all equal to 0.056. The changes of WSCVI values with equal weights and with adjusted weights are shown in Figure 3.

As Figure 3 shows, the minimum values of WSCVI with adjusted weights are achieved at c=3 for Data_3_3_300 and c=4 for Data_3_4_320. Therefore, the optimal cluster number c^*= 3 for Data_3_3_300 and c^*= 4 of Data_3_4_320 are both found.

For iris data set, the optimal cluster number c^*= 2 considering the geometric structure is found by WSCVI when all the weights are equal to 0.1. Now we consider an adjusted weights case, in which the weight of PBMF index ω_{PBMF}=0.7, and the weights of other indices are equal to 0.025. The changes of WSCVI values with equal weights and adjusted weights on iris data are shown in Figure 4 (a).

As shown in Figure 4 (a), the optimal cluster number c^*= 3 for iris can be found when one CVI dominate other CVIs. Since six classes in glass data set are heavily overlapped, it is difficult to find six clusters. The changes of WSCVI values with equal weights and adjusted weights for glass data are shown in Figure 4 (b).

Table 1: Optimal cluster numbers preferred by each CVI

	c^*	PC	NPC	PE	NPE	XB	VK	PBMF	FS	VT	SC	WSCVI
Butterfly	2	2	2	2	2	2	2	2	2	2	2	2
Example_01	3	3	3	3	2	3	3	2	3	3	3	3
Example_02	4	4	4	4	2	4	4	2	4	4	4	4
Data_2_3_60	3	3	3	3	3	3	3	2	7	3	3	3
Data_2_3_70	3	2	3	2	2	2	2	2	4	2	4	3
Data_2_4_110	4	2	4	2	2	2	2	2	4	2	10	4
Data_3_3_200	3	3	3	2	2	3	3	2	10	3	3	3
Data_3_3_300	3	2	2	2	2	2	2	2	15	2	3	2
Data_3_4_320	4	3	4	2	2	3	3	2	5	3	4	3
Bupa	2	2	2	2	2	2	2	3	4	2	2	2
Wdbc	2	2	2	2	2	2	2	5	4	2	2	2
Iris	3	2	2	2	2	2	2	3	5	2	2	2
Glass	6	2	2	2	2	2	2	6	6	2	5	2

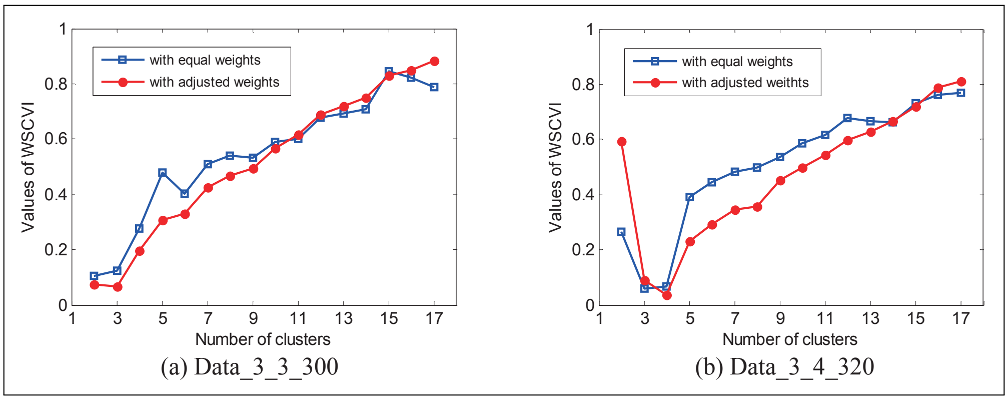
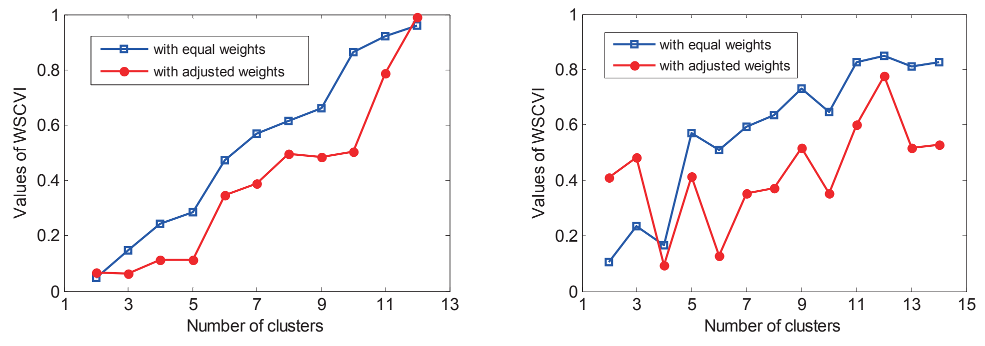


Figure 3: Values of WSCVI

From Figure 4 (b), the minimum value of WSCVI with adjusted weights achieved at c=4, and its value is the second smallest when c=6. Also, there is a large increase when c is greater than 6. Therefore, WSCVI index with adjusted weights offers the information that c^*=6 is a good cluster number estimate for glass data set.



(a) iris (b) glass

Figure 4: Values of WSCVI

With equal weights for nine data sets and adjusted weights for four data sets of each CVI, WSCVI finally find all the optimal cluster numbers for the above thirteen data sets.

5 Conclusion

We investigate ten well-known CVIs for fuzzy clustering and present an extensive comparison of the ten single CVIs and the proposed WSCVI on nine synthetic data sets and four real-world data sets. Experimental results demonstrate that most single fuzzy CVIs are effective in finding optimal cluster numbers for data sets which are low-dimensional and well-separated. But some CVIs fail to find the optimal cluster numbers for some high-dimensional or heavily overlapped data sets. The experimental results indicate that, by properly setting the weights of each CVI, the proposed WSCVI is more effective in finding the optimal cluster numbers than single CVI.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (71131002, 71201042).

Bibliography

[1] A.K. Jain, M.N. Murty, P.J. Flynn (1999). Data Clustering: A Review, *ACM Computer Surveys*, 31(3):264-323.

[2] P.A. Devijver, J. Kittler (1982). Pattern Recognition: A Statistical Approach, *Prentice-Hall*, London.

[3] F. Hoppner, F. Klawon, R. Kruse, T. Runkler (1999). Fuzzy Cluster Analysis: Methods for Classifications Data Analysis and Image Recognition, *Wiley*, New York.

[4] M. Kim, R.S. Ramakrishna (2005). New Indices for Cluster Validity Assessment, *Pattern Recognition Letters*, 26 (15):2353-2363.

[5] W. Wang, Y. Zhang (2007). On Fuzzy Cluster Validity Indices, *Fuzzy Sets and Systems*, 158(19):2095-2117.

[6] E. Dimitriadou, S. Dolnicar, A. Weingessel (2002). An Examination of Indexes for Determining the Number of Clusters in Binary Data Sets, *Psychometrika*, 67(1):137-159.

[7] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J.M. P¨Śrez, I. Perona (2013). An Extensive Comparative Study of Cluster Validity Indices, *Pattern Recognition*, 46(1):243-256.

[8] K.L. Wu, M.S. Yang (2005). A Cluster Validity Index for Fuzzy Clustering, *Pattern Recognition Letters*, 26 (9):1275-1291.

[9] H. Le Capitaine, C. Frelicot (2011). A Cluster-validity Index Combining an Overlap Measure and a Separation Measure based on Fuzzy-aggregation Operators, *IEEE Transactions on Fuzzy Systems* , 19(3):580-588.

[10] U. Maulik, S. Bandyopadhyay (2002). Performance Evaluation of Some Clustering Algorithms and Validity Indices, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:1650-1654.

[11] K.R. Zalik (2010). Cluster Validity Index for Estimation of Fuzzy Clusters of Different Sizes and Densities, *Pattern Recognition*, 43(10):3374-3390.

[12] W. Sheng, S. Swift, L. Zhang, X. Liu (2005). A Weighted Sum Validity Function for Clustering with a Hybrid Niching Genetic Algorithm, *IEEE Transactions on Systems, Man, and Cybernetics - Part B, Cybernetics*, 35(6):1156-1167.

[13] J.C. Bezdek, R. Ehrlish, W. Full (1984). FCM: The Fuzzy C-means Clustering Algorithm, *Computers & Geosciences*, 10(2-3):191-203.

[14] J.C. Bezdek (1974). Numerical Taxonomy with Fuzzy Sets, *Journal of Mathematical Biology*, 7(1):57-71.

[15] M. Roubens (1978). Pattern Classification Problems and Fuzzy Sets, *Fuzzy Sets and Systems*, 1(4):239-253.

[16] J.C. Bezdek (1974). Cluster Validity with Fuzzy Sets, *Journal of Cybernetics*, 3(3):58-72.

[17] J.C. Dunn (1977). Fuzzy Automata and Decision Processes, *Elsevier*, New York.

[18] X.L. Xie, G. Beni (1991). A Validity Measure for Fuzzy Clustering, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(8):841-847.

[19] S.H. Kwon (1998). Cluster Validity Index for Fuzzy Clustering, *Electronics Letters*, 34(22):2176-2177.

[20] M.K. Pakhira, S. Bandyopadhyay, U. Maulik (2004). Validity Index for Crisp and Fuzzy Clusters, *Pattern Recognition*, 37(3):487-501.

[21] Y. Fukuyama, M. Sugeno (1989). A New Method of Choosing the Number of Cluster for the Fuzzy C-means Method, *Proceedings of the 5th Fuzzy Systems Symposium*, 247-250.

[22] Y.G. Tang, F.C. Sun, Z.Q. Sun (2005). Improved Validation Index for Fuzzy Clustering, *American Control Conference*, 1120-1125.

[23] A.M. Bensaid, L.O. Hall, J.C. Bezdek, L.P. Clarke, M.L. Silbiger, J.A. Arrington, R.F. Murtagh (1996). Validity-guided (Re) Clustering with Applications to Image Segmentation, *IEEE Transactions on Fuzzy Systems*, 4(2):112-123.

[24] K.L. Zhou, S.L. Yang (2013). A Fuzzy Cluster Validity Index in Consideration of Different Size and Density of Data Set, *Journal of the China Society for Scientific and Technical Information*, 32(3):306-313.

[25] A. Asuncion, D.J. Newman (2007). UCI Machine Learning Repository, *University of California, School of Information and Computer Science*, Irvine, CA, http://www.ics.uci.edu/mlearn/MLRepositor-y.html.

Author index