

FARS: Fuzzy Ant based Recommender System for Web Users

Shiva Nadi¹, Mohammad. H. Sarace², Mohammad Davarpanah Jazi³ and Ayoub Bagheri⁴

¹ Islamic Azad University, Najafabad Branch, Isfahan, Iran

² Department of Electrical and Computer Engineering, Isfahan University of Technology
Isfahan, Iran

³ Department of Computer Engineering, Foulad Institute of Technology
Fouladshahr, Isfahan, Iran

³ Department of Electrical and Computer Engineering, Isfahan University of Technology
Isfahan, Iran

Abstract

Recommender systems are useful tools which provide an adaptive web environment for web users. Nowadays, having a user friendly website is a big challenge in e-commerce technology. In this paper, applying the benefits of both collaborative and content based filtering techniques is proposed by presenting a fuzzy recommender system based on collaborative behavior of ants (FARS). FARS works in two phases: modeling and recommendation. First, user's behaviors are modeled offline and the results are used in second phase for online recommendation. Fuzzy techniques provide the possibility of capturing uncertainty among user interests and ant based algorithms provides us with optimal solutions. The performance of FARS is evaluated using log files of "Information and Communication Technology Center" of Isfahan municipality in Iran and compared with ant based recommender system (ARS). The results shown are promising and proved that integrating fuzzy Ant approach provides us with more functional and robust recommendations.

Keywords: *Web personalization, Recommender Systems, Ant colony optimization, Fuzzy set.*

1. Introduction

Recommender systems (RS) are useful tools which guaranties that right information are accessible for right users at right time [1]. RSs are useful in different domains, such as web personalization, information filtering, e-commerce, providing recommendations of books, movies and music. One of the most popular applications of recommender systems is web environment personalizing by providing a list of items related to user's interests. This paper proposes a fuzzy-ant based Recommender system (FARS) which provides a list of recommendations for currently online user by comparing active user's navigational behavior with data collected from other users.

Ant colony optimization (ACO) is a computational algorithm that mimics the behavior of ants and is proposed by Dorigo in 1996. As the uncertainty is the nature of user's behavior the dynamic intelligent FARS proposed in this paper, uses ACO and fuzzy logic to prepare the high potential and suitable promoting recommendations for active user.

Usually demographic, content based and collaborative based filtering techniques are employed to generate recommendations. In demographic filtering technique (DMF), users are categorized based on their personal attributes such as their age range and provide recommendation based on these demographic categories. Recommendations produced by these systems are too general and not adaptive with changes in user preferences over time. Content based filtering (CBF) provides users with similar conceptual recommendations based on previously evaluated items. Content-based filtering methods usually utilize text extraction techniques for building user profiles. These methods have some disadvantages such as mismatch between user profile terms and item profile terms, that leads to decreasing the performance [2]. Collaborative based filtering technique (CF) suggests items based on similar users preferences. This method provides poor prediction when the number of similar users is small. Overcoming disadvantages of these models, in this paper we employ an integrated CF and CBF based hybrid model. Applying this hybrid model on the web user's behaviors obtained from the analysis of log files, we propose a fuzzy-ant based recommender system.

The rest of the paper is organized as following. In section 2 the background of this research is explained. The purposed model is described theoretically in section 3. Experimental results about the implemented system and a

brief discussion about model is represented in section 4 and finally in section 5 the conclusion and future research directions are presented.

2. Research Background

During the recent years, many researches are done for personalizing websites using Variety of techniques. In the field of swarm intelligence algorithms, Ujjin and Bently have presented a recommender system based on particle swarm optimization algorithm [3]. They proved that the results obtained from their PSO recommender system are more accurate than the genetic and Pearson algorithm. In another work a fuzzy genetic recommender system with the accuracy of memory based CF and the scalability of model based CF. their novel user model helps achieving complexity and sparsity reduction in system. The performance of their method is proved by comparing the results with Pearson and fuzzy recommender system [4]. Sobecki used ant colony metaphor for selecting optimal solutions in his hybrid recommendation method and Bedi also, presented a recommender system based on collaborative behavior of ants. He used collaborative filtering approach and generated recommendations for Jester dataset [5]. Clustering is an important step in all recommendation systems, choosing an appropriate clustering algorithm leads to producing more qualified recommendations. The c-means and k-means algorithms are most well known clustering algorithms [6]. Fuzzy c-means is a method of clustering which allows one piece of data to belong to two or more clusters. This method developed by Dunn in 1973 and improved by Bezdek in 1981. Swarm algorithms are also used for clustering items. Ant based clustering has been introduced by Deneuborg, in this algorithm ants discriminate between different kinds of items and spatially arrange them according to their properties. This algorithm is modeled of the real behavior of ants in nature. The proposed approach by Kanade and Hall (2003), presents the combination of ant based clustering and FCM [7]. Their model is employed in this study for clustering web users based on their accesses to web pages.

3. Proposed fuzzy-ant recommender system

In this section, a dynamic intelligent recommender system for providing high potential recommendations is proposed, which uses fuzzy logic in accordance with ant colony optimization for increasing the accuracy and relevancy of predictions. The system analyzes a user's navigational behavior during a period of time and find out the most ideal recommendations for him. Figure 1 show the schema of the proposed system, which consists of modeling and recommendation phases. In first step, user preferences are

identified using web access log data called web usage data. In the next step, the knowledge which is achieved through the previous step is used to identify the possibly interested URLs and provide recommendations to the users. This recommendation can be done in different manners such as adding related hyperlinks to the web page requested by the user.

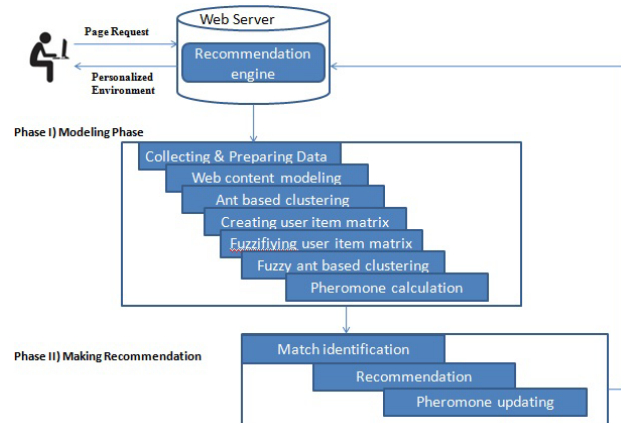


Fig. 1 Proposed work schema

In this paper, Represented method dynamically recommends the highest match score URLs to the users with similar interests, the recommended URLs are also related conceptually to each other. In this way, we integrated fuzzy logic to ant colony semaphore to provide a recommender system which produce optimum recommendations for users, considering the uncertainty among user's interests.

3.1 Data Modeling Phase

In this phase through several sub phases, pure data extracted from logs files become prepare for grouping users in appropriate clusters based on their interests' similarity.

Step1) Data Collection

Web servers provide log files which have useful information about access of all users to a specific website. Extracting these information, some preprocesses should be done on log files. The result will be a reformatted log file which contains useful information about the accessed URLs and IP addresses.

Step2) Web Content Mining

Various numbers of clustering techniques are used for clustering documents. In this work, following algorithm is used for document clustering. Assume $P = \{p_1, p_2, \dots, p_k\}$ is the set of k website's pages that will be grouped in content based clusters using following steps:

Step 1. Assign each page to a single cluster

Step 2. Merge clusters based on Jaccard coefficient similarity measure by Eq. 1.

$$sim(p_x, p_y) = \frac{|p_x \cap p_y|}{|p_x \cup p_y|} \quad (1)$$

Where $|p_x \cap p_y|$ is the number of common words between two basic clusters and $|p_x \cup p_y|$ is total number of words in both clusters.

Step 3. Repeat step 2 until all documents being clustered.

The result is the set, $DC = \{DC1, DC2, \dots, DC_n\}$ and DC_i represents a set of URLs with similar content.

Step3) Creating User-Item Matrix

Most of recommender systems which are CF based use user-item matrix for identifying similar users and recommending items which are highly ranked by those users. User-item matrix is a matrix of size $m \times n$ where each entry represents the interest degree of i -th user to j -th document cluster. The interest degree is the value of each document cluster for each user. We use the interest degree which proposed by Castellano (2007), they defined it as the ratio of the number of accesses to each document cluster to the total number of accesses to all document clusters for each user [8].

$$Val(DC_i, U_i) = \frac{|\{DC_j | DC_j \in A_i\}|}{|A_i|} \quad (2)$$

$A = \{A_1, A_2, \dots, A_m\}$ is a set of user's accesses to document clusters. For example A_k indicates the access list of K -th user to a subset of document clusters ($A_k \in DC$, for $K=1$ to n)

Step4) Normalizing and fuzzifying user-item matrix

Fuzzy sets theory which model vagueness was introduced by Lotfi A. Zadeh in 1965. Fuzzy sets support a flexible membership of elements to the sets. While in crisp set theory, an element absolutely belongs to a set or doesn't but in fuzzy sets theory variety of membership degrees in the range of 0 to 1 can be allocated to items in classifying interest domains. We use triangular fuzzy numbers to characterize user interests to document features. A triangular number has a triangle shaped membership function, which can be viewed as possibility distribution. To express user interests, the linguistic terms are used to linguistically evaluate the importance of user interests. Five linguistic sets can be used to describe the diversity of user's interests: VS (Very Small), S (Small), M (Medium), H (High), VH (Very High)

The membership function of user's interests is shown in Figure 2.

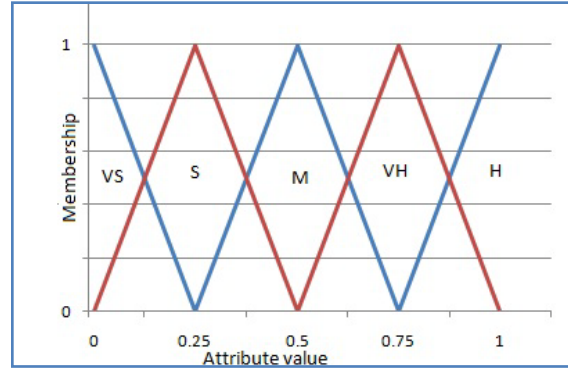


Fig. 2 membership function

The distance between two users, fuzzy sets illustrates the degree in which these users are the same. The global distance is an aggregation of distances between any two partial feature distances. The aggregation operator may be the average of the local fuzzy distances [9].

$$fd(x, y) = \frac{1}{N} \sum_{i=1}^N fd(x_i, y_i) \quad (3)$$

Step 5) Fuzzy ant based clustering

The ant based algorithm provides acceptable clusters of data without any knowledge of the initial clusters. In the ant based algorithm if an object is assigned to an inappropriate heap then it takes long time to be transported to a better heap. Overcoming this problem, a synthetic algorithm proposed by Kanade and Hall (2003) that uses the integrated FCM and ant based clustering algorithms [7]. In the web recommendation field we supposed that objects are users associated with a vector of fuzzy numbers which indicates their interests to document clusters.

The fuzzy C-means algorithm requires good initialization. These initial values are provided by the ant based algorithm. The result will be small homogenous heaps that will be merged by repeating the steps. By increasing the number of iterations the number of heaps decreases. The following algorithm is used in this study to cluster users in appropriate groups: [7]

- Scatter the users randomly on the board
- Initialize the ants with random position, and random direction
- For N iterations do
 - For each ant do
 - Move the ant
 - If the ant is carrying an user U then possibly drop the user U else
 - Possibly pick up a user

- Use the cluster centers obtained in step 3 to initialize cluster centers for the fuzzy C-means algorithm
- Cluster the data using the fuzzy C-means algorithm
- Harden the data obtained from the Fuzzy C-means algorithm, using the maximum membership criterion, to form new heaps
- Repeat step 1-6 by considering each heap as a single object

The advantage of this algorithm lies within the calculation of initial clusters' centers which is based on Ant colony algorithm. Table 1 illustrates a cluster containing 3 users. The center of the cluster is a vector (table 2), which is computed as the mean of user preferences in the user cluster.

Table 1: k-th user cluster

UC _k	DC1	DC2	DC3 ...	DC8
U1	0.3	0.0	0.5	0.2
U6	0.2	0.1	0.7	0.0
U9	0.0	0.06	0.64	0.3

Table 2: Center of cluster

UC _k	DC1	DC2	DC3 ...	DC8
U1	0.080200	0.0000	0.01100	0.208000
U6	0.208000	0.604000	0.002080	0.0000
U9	0.0000	0.76024000	0.00440560	0.080200
Center	0.060500600	0.45021000	0.00540450	0.0605300600

Step 6) Calculating pheromone for each cluster

The pheromone associated with each cluster is calculated based on the density of each cluster using Eq. 4.

$$P_{heromone}(t) = \frac{\text{Number of users in cluster } i}{\text{total Number of users}} \quad (4)$$

3.2. Recommendation process for active user

When a new user starts a transaction, our model matches the new user with the most similar user clusters and provides suitable recommendations for him/her. It will be done by computing the similarity of the active user profile with center of the user cluster. By combining the similarity and pheromone value allocated to each cluster the active user can be matched with existing clusters.

Step1) Matching active user with best user clusters

Similarity of fuzzy sets is based on their distance which is calculated using Eq. 5 [10].

$$SM(U_{new}, U_{center}) = \frac{1}{1 + DM(U_{new}, U_{center})} \quad (5)$$

Where SM is the similarity measure and DM is the distance measure of two U_{new} and U_{center} fuzzy sets. U_{new} is the active user fuzzified profile and U_{center} is center of i -th cluster.

The distance between two fuzzy sets of the active user profile and the center of the cluster can be computed using Eq.6 [10].

$$DM(U_{new}, U_{center}) = \sum_{i=1}^l w_i \cdot DM_k(U_{new}, U_{center}) \quad (6)$$

Where l is the total number of fuzzy sets, w_i is the weighting for i -th linguistic variable, here we got it as 1, and DM_k is the distance measure along the fuzzified features variables.

$$DM = \sqrt{\sum_{i=1}^l (U_{new_i} - U_{center_i})^2} \quad (7)$$

Where, l is the total number of fuzzy sets.

Using E.q. 5 the similarity of active user is calculated with other clusters and clusters which have users with similar preferences are distinguished.

The clusters which have more pheromone and more similarity to active user are chosen as clusters with most match score with active user. The match score at time t is calculated using Eq. 8 [11].

$$MatchScore_i(t) = \frac{\text{pheromone}_i(t) \cdot SM(U_{new}, U_{center})}{\sum_{i=1}^k \text{pheromone}_i(t) \cdot SM(U_{new}, U_{center})} \quad (8)$$

The clusters in the range of [highest MatchScore – α , highest MatchScore] are chosen for recommendation.

Table 3 illustrates an active user profile applying fuzzy ant based clustering algorithm described in section 3.1.5, 4 user clusters are obtained. The match score for the active user and these 4 clusters are illustrated in table 4.

Table 3: Normalized active user profile

	DC1	DC2	DC3	DC4	DC5	DC6	DC7	DC8
active user profile	0.12	0	0.24	0	0.3	0	0	0.64

Table 4: match score between active user and user clusters

	Pheromone	Similarity	MatchScore
UC1	0.4	0.237	0.364
UC2	0.2	0.153	0.117
UC3	0.3	0.370	0.426
UC4	0.1	0.240	0.092

By choosing $\alpha = 0.1$, the clusters between the range of

$0.426 - 0.1 \leq \text{Match Score} \leq 0.426$ are chooses for extracting suitable recommendations. Here UC1 and UC3 have more compatibility with active user.

Step 2) Recommendation

The values of documents which have been accessed in chosen clusters and not by active user are candidates to be presented for active user as recommended items. Table 5 illustrates the centers of clusters Uc_1 and Uc_3 which has the most match score with the active user. As presented in active user profile DC₁, DC₃, DC₅ and DC₈ are accessed before and are not considered for recommendation.

Table 5: the value of candidate document clusters in chosen user clusters

	DC2	DC4	DC6	DC7
Center of UC ₁	0.222	0.208	0.358	0.095
Center of UC ₃	0.193	0.064	0.194	0.111

DC2, DC4 and DC6 are chosen from user clusters 1, and DC7 is chosen from user cluster 3.

Step 3) Pheromone Updating

The pheromone associated with good solutions must be increased and the pheromone values associated with bad ones must be decreased. This is possible by decreasing the value of pheromone through pheromone evaporation and increasing the pheromone levels associated with a suitable solution. The amount of pheromone associated with each cluster is decreased by small value and pheromone associated to the cluster from which recommendation is generated is increased according to formula. 9 [11].

$$pheromone_i(t) = (1 - \rho) \times pheromone_i(t - 1) + \Delta Q \times pheromone_i(t - 1) \quad (9)$$

Where, $\rho = 0.01$, is pheromone evaporation rate. ΔQ is calculated using Eq. 10.

$$\Delta Q = \frac{Q_{cc}}{Q_{cc} + 1} \quad (10)$$

Where, Q_{cc} is the value of item in selected user cluster. The result of pheromone updating process is shown in Table 6.

Table 6: Pheromone updating

	UC1	UC2	UC3	UC4
Initial Pheromone	0.400	0.200	0.300	0.100
Pheromone after recommending DC2	0.468	0.190	0.290	0.090
Pheromone after recommending DC4	0.543	0.180	0.280	0.080
Pheromone after recommending DC6	0.680	0.170	0.270	0.070
Pheromone after recommending DC7	0.670	0.160	0.260	0.076

The pheromone assigned to each cluster is updated based on recommendations made for active user and will be used for future recommendations.

4. Discussion and experimental results

This paper improved web user recommendation quality by considering the fuzzy behavior of users for modeling them in suitable clusters and supports them by a range of relevant recommendations. The fuzzy ant based recommender system proposed in this paper has been

applied on the access log file of “Information and Communication Technology Center” of Isfahan municipality in Iran for IP address 80.191.136.6. We collected log file during a period of one week. After data cleaning in preprocessing step, the number of requests was 5232. The number of accessed URLs in this website was 200 pages. Employing content based document clustering algorithm the URLs were grouped to 15 clusters. In next step, user’s behaviors have been modeled as access matrix. Using equation 2, interest degrees are calculated as a 5232×15 matrix. A set with 4000 rows and 5 columns is considered as the training set and is used as the input of user clustering algorithm.

In clustering step, the combination of ant based clustering and fuzzy C-means was used for clustering web site’s users. The number of algorithm iterations was set to 1000 and the values of T_{create} , P_{drop} , $P_{destroy}$, P_{load} , T_{remove} was set to 0.5,0.2,0.3,0.3,1.5, respectively and the number of ants was set to $n/3$ where n is the total number of objects to be clustered [10]. Here we set the number of ants to be 1000 and the value of 0.01 is taken for pheromone evaporation rate. Following the steps described in section 3 a set of documents are derived to be recommended for active user.

The proposed model predicts user preferences not only based on the interests of other users but also the conceptual similarity between website pages is intended. The combination of ant clustering and FCM used in this model locates users in appropriate classes and provide perfect clusters of users with similar interests. As providing recommendations is based on the interest of users in the same clusters, improving quality of clusters increases the completeness and exactness of recommendations. In clustering process, the fuzzy C-means algorithm requires some initializations that feed by ant based algorithm without any initial knowledge of the center of clusters. For calculating the exactness and completeness of produced recommendations, precision and recall measures are computed. In this way, a user profile is chosen and divided into a training and test set. According to the knowledge obtained from analyzing the training set, some predictions are produced. The ratio of the number of correctly predicted items over the number of predicted items indicates the recall measure. Precision is defined as the ratio of the number of correctly predicted items over the size of top n set. Recall and precision are computed for the proposed method and were compared with the results of ant based recommendation method in which user clustering method was singly ant based. The results are shown in Figure 3 and 4. N is the number of topmost recommendations.

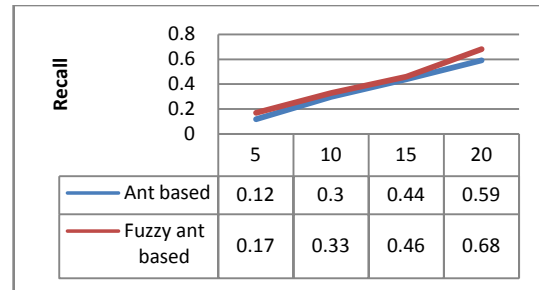


Fig. 3 Recall measures for N top recommendations

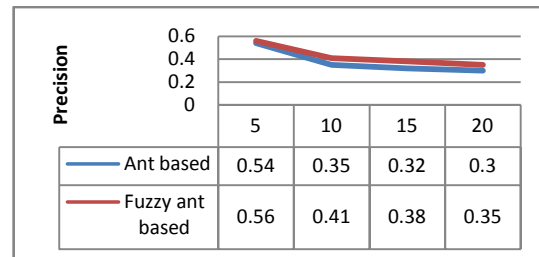


Fig. 4 Precision measures for N top recommendations

The Figure 3 and 4 indicates the proposed approach improves the performance of Ant based method and modifies the results with respect to the user demand and she/he interests and therefore our approach shows more qualified recommendations.

5. Conclusions

Recommender systems are tools which provide a personalized environment for the users of a web site by investigating their navigational behavior in a period of time. In this paper a fuzzy ant based recommender system (FARS) was proposed. In this method, the user’s interests to the web pages are extracted from web server log files. By applying FCM and ant based clustering algorithms users are grouped in appropriate clusters. Ant based algorithm helps in providing optimal solutions and FCM considers the uncertainty in user’s interests. Then the match score of currently online user with existing clusters is calculated based on their similarity and the amount of pheromone on each cluster. The pheromone assigned to each cluster is updated at end of recommendation process for future uses. Finally the precision and recall are computed to expressing the exactness and completeness of produced recommendations. The results indicate that using the proposed method for clustering users will lead to offering more qualified recommendations. Our work can be extended by assigning an appropriate weight to the documents in each document cluster for increasing their effectiveness in match score identification.

References

- [1] Andomavicius, G. and A. Tuzhilin, "Toward the next generation of recommender system: A survey of the state-of-the-art and possible extensions". *IEEE Trans, Knowledge Data Eng.*, 17: 734-749, 2005.
- [2] pertz shoval, Veronica Maidel, Brancha shapira, "international journal of information theories and applications", Vol.15.pp. 303-314,2008.
- [3] Ujjin, S. and P.J. Bentley. "Particle swarm optimization recommender system". *Proceedings of the 2003 IEEE Swarm Intelligence Sysposium-SIS '03*, April 24-26, London, UK. PP: 124-131, 2003.
- [4] Mohammad Yahya H., Al-Shamiri, Kamal K., Bharadwaj, "Fuzzy Genetic Approach to Recommender Systems based on a Novel Hybrid User Model", *Expert systems with applications*, Elsevier, pp.1386-1399, 2007.
- [5] Sobecki, J., "Web-Based System User Interface Hybrid Recommendation Using Ant Colony Metaphor. In: Knowledge-Based Intelligent Information and Engineering Systems", Apolloni, B. et al. (Eds). LNCS 4694, Springer-Verlag, Berlin, Heidelberg, ISBN: 978-3-540-74828-1, pp: 1033-1040, 2008.
- [6] Vathatis, M.N., B. Boutsinas, P. Alevizos and G. Pavlides, "The new K-windows algorithm for improving the K-means clustering algorithm". *J.Complexity*, 18: 375-391, 2002.
- [7] Kanade P. M. and L. O. Hall, "Fuzzy ants as a clustering concept". *North American Fuzzy Information Processing Society, NAFIPS 2003, 22nd International Conference of the*, pp. 227-232, 2003.
- [8] Castellano, G. and Fanelli, A. M. and Mencar, C. and Alessandra Torsello, M., "Similarity-based Fuzzy clustering for user profiling", *IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, pp. 75-78, 2007.
- [9] Gadi, T., Daoudi, R. B. M., & Matusiak, S., "Fuzzy similarity measure for shape retrieval", In *Vision Interface '99, Trois-Rivie` res, Canada* (pp. 19-21), 1999.
- [10] Koczy T. Laszlo, *Tikk Domonkos: Fuzzy rendszerek*, Typotex, 2000.
- [11] Bedi P., Sharma R. and kaur H. , "Recommender System based on collaborative behavior of ants", *Journal of artificial intelligence*, ISSN 1994-5450, PP. 40-55.

Shiva Nadi received her BS and MS degrees in computer engineering from Islamic Azad University of Najafabad, Iran, in 2010. Her research interests are in several areas of artificial intelligence, data mining, web mining, evolutionary algorithms and software engineering.

Mohamad H Saraei received his PhD from University of Manchester in Computation, MSc from University of Wyoming, USA in Software Engineering and BSc from Shahid Beheshti University, Iran. His main areas of research are intelligent databases, Mining advanced and complex data including medical and Bio, Text Mining and E-Commerce. He has published extensively in each of these areas and served on scientific and organizing committee on number of journals and conferences.

Mohamad Davarpanah Jazi received his PhD from University of Manchester in Information Systems, MSc from University of Caltech, USA in Software Engineering and BSc from Shahid Beheshti University, Iran. He is now a member of computer department of Foulad Institue of Technology. His main areas of

research are Information Systems, DataBases, Programming Languages and Operating Systems.

Ayoub Bagheri received his BS degree in computer engineering from Ferdowsi University of Mashhad, Iran, in 2007, and the MS degree in computer engineering from Isfahan University of Technology, Iran, in 2009. Now he is PhD student in computer engineering in Isfahan University of Technology. His research interests are in several areas of artificial intelligence, data mining, machine learning and evolutionary algorithms.