



## Toward Mobile Telecommunication Recommendation System through Intelligent Customers Categorization

RawanGhnemat, Edward Jaser

Princess Sumaya University for Technology, Computer Science Department

rawan.ghnemat@psut.edu.jo

Princess Sumaya University for Technology, Computer Graphics and Animation Department

ejaser@psut.edu.jo

### ABSTRACT

Now a day, usage of mobile devices is becoming indispensable. This is evident with current mobile penetration rates reaching 100% and even more in some countries. Customers across the world are enjoying competitive prices due to high competition among telecommunication companies. As a result of this, it is mandatory for mobile companies to provide high quality services to their customers to retain them. One aspect which will maximize customers' trust and lead to high retention rate is to offer them a suitable plan that matches their usage. Mobile customer usage categorization is therefore an essential task to develop intelligent business plans. Personalized recommendation system is needed to dynamically adapt the different customer behaviours with the most appropriate plan for them. In this paper we propose a new automatic approach for costumers' categorization. This will be the basis for the recommendation system. The proposed method is built using Fuzzy rule and aims at usage behaviour prediction. The rules was extracted from real customer data obtained from a leading provider. Comparison study with other categorization methods has been conducted and showed superior result and demonstrated the potential advantage of the proposed fuzzy based method.

### Indexing terms/Keywords

Mobile user categorization; Self-Organizing Maps; Fuzzy Systems; Machine learning.

---

# Council for Innovative Research

Peer Review Research Publishing System

**Journal: INTERNATIONAL JOURNAL OF COMPUTERS & TECHNOLOGY**

Vol 12, No. 7

editor@cirworld.com

[www.cirworld.com](http://www.cirworld.com), [www.ijctonline.com](http://www.ijctonline.com)



## 1. INTRODUCTION

Now a day, usage of mobile devices are becoming indispensable. Over the last decade we witness a booming usage of mobiles in our daily life. Beside the conventional use of phone calls and messaging, smart mobile devices are used in a wide spectrum of applications including entertainment, gaming, business follow-up, banking, reservation, Internet and many other applications. It has been also witnessing huge successes in development initiatives [1], [2] and [3]. Penetration rates stand as an evidence to the popularity of mobiles as tools. Current mobile penetration rates are approaching 100% and even more in some countries.

Customers across the world are enjoying competitive prices due to high competition among telecommunication companies. As a result of this, it is mandatory for mobile companies to provide high quality services to their customers to retain them. One issue is the nature of mobile usage: is it mainly calls, mainly data or combination of them. This issue is important to understand for users as it will affect the total spending on mobile contracts. Having inappropriate plan can lead to a customer paying higher price than she or he actually uses. One important aspect for mobile operators which will maximize customers' trust and lead to high retention rates is to offer them a suitable plan that matches their usage. Mobile customer usage categorization is therefore an essential task to develop intelligent business plans. Personalized recommendation system is needed to dynamically adapt the different customer behaviours with the most appropriate plan for them.

This paper address the issue of customers' categorization. we propose a novel automatic approach for categorizing customers based on their actual usage. This will certainly be the basis for a recommendation system which will be of great value for mobile service providers. The proposed method is built using Fuzzy rule and aims at usage behaviour prediction. The rules was extracted from real customer data obtained from a leading mobile operator. Comparison study with other categorization methods has been conducted to look at the performance of the method against widely used machine learning algorithms.

The paper is organised as follows. In Section 2, we review the topic of user categorization system and provide a summary of related research work achieved in this field. In Section 3, we describe the dataset with regards to user mobile usage in an attempt to understand what data are available for categorizing users. We describe our proposed method in details in Section 4. We describe the experiments we performed to compare our proposed method with four popular categorization algorithms in Section 5. In Section 6, we provide a concluding remarks and future work.

## 2. USER CATEGORIZATION

To build an efficient recommendation system we need a kind of differentiation for users needs; the first step to achieve this is to categorize user based on specific factor. In our research we use usage behavior as a factor to categorize the telecommunications users in order to allow enterprise understand their customers and deal with different customers demands and improve service quality [4] and improve customer satisfaction and loyalty and build reputation.

In fact the work of recommendation can be seen as segmentation of users categories to predict unknown event [5]. Recommendation systems remain an active research area that combines many different disciplines like data mining, Artificial intelligence and information retrieval [6] and [7]. Several hybrid approaches treat recommendation systems as classification task as demonstrated by Jaing et al., [8] where they propose an online recommendation system with the aim to increase and enhance customers' satisfaction based on associative classification model using customer's profile and after-sale collected feedback. the classification in [8] associate product information with customer from the needed rating data and utilization information but with minimum information about the behavior of customers, customers' demographical and psychological characteristics of customers. The recommendation system proposed by Kim et al., [9] addressed this by filtering users in recommendation system based on collaborative learning approach that model internet user interests based on usage behavior.

A general model that use collaborative filtering and match them to a groups of items and have the benefits from users pervious decisions that belongs to a specific group proposed by Bobadilla et al., in [10] which extends the regular recommendation systems by giving qualitative measures (e.g., precision and recall) while compiling recommendations. An interesting work that focused on impact is the research by Hosteler et al., [11]. They used intelligent agent technology recommendation for online shopping consumers by the classification of consumers into categories based on their "loyalty and satisfaction". They found that this method make the recommendation system more effective in increasing the perceived attractiveness of the products in online shopping and increase customers satisfaction and help retaining them.

Kim et al., [12] focus in their research on online users also using a different approach based on grouping users with similar interest by aggregating individual interest into a group behavior. They demonstrated with the aid of a prototype (online book clubs) better precision than benchmark system using this feature. Classification information considered an essential part on most of recommendation systems work; Zheng et al. in their work [13] filtered location using interesting piece of information: collaborative location and GPS data; in their system they combine geographical information with preference information.

Some work listed above approached categorization based mainly on data and feedback that is collected at individual instance of time. When addressing mobile usage it is important to continuously obtain and collecting usage date to accurately build the categorization model. We base our proposed method on 10 usage criteria as described in the next section.



### 3. DATA SET

In order to understand the categorization problem with relation to mobile usage against usage plan categories, it is important to understand what data are available for processing. In our research we relied on actual data collected from a leading mobile service provider in Jordan. A significant contribution of our work comes from the fact that it is based on usage data of actual customers. To preserve privacy, no names or sensitive information were collected. Each record of data contains the following attribute:

1. **Calls Duration:** numeric field, containing the total number of seconds used for making voice calls.
2. **Data Usage:** numeric field, containing the total number of MB used for internet usage.
3. **SMS Sent:** numeric field, containing the total number of SMS sent by the user.
4. **SMS Received:** numeric field, containing the total number of SMS received by the user.
5. **Number of Calls:** numeric field, containing the total number of calls generated by the user.
6. **Number of Received Calls:** numeric field, containing the total number of calls received by the user.
7. **On-net call Duration:** numeric field, containing the total number of seconds used for making voice calls on the same service provider.
8. **Off-net calls Duration:** numeric field, containing the total number of seconds used for making voice calls on different service provider.
9. **Fixed line calls Duration:** numeric field, containing the total number of seconds used for making voice calls on land line telephones.
10. **International calls Duration:** numeric field, containing the total number of seconds used for making international voice calls.

Each record is groundtruthed and associated a plan that best suits their needs based on expert review. We also have the actual plan a user is using but this information is not used as we are only interested in the plan that will minimize cost for users. 6,000 records were collected and groundtruth with the aid of an experts.

### 4. PROPOSED METHOD

As mentioned earlier, large data set of mobile users log files have been analysed and six main segments obtained based on the Self Organized Map (SOM) algorithm [14]. One important motivation for using the SOM comes from the fact that it is trained using unsupervised learning. Designed system based on this classification to predict user segment will save the time and the effort of the customers and the marketing sector.

Fuzzy expert systems are proving to be a powerful tool in business decision support knowledge [15]. In addition, the terms that is normally used, or understood by, customers are fuzzy in nature (e.g., high call duration or low number of SMS). The most important application of fuzzy system (fuzzy logic) is in uncertain issues. When a problem has dynamic behavior, fuzzy logic is a suitable tool that deals with this problem. Therefore, we were motivated to use this method.

First step of fuzzy expert system designing is the identification of input and output variables. There are 10 input variables and 1 output variable. Next step is to design membership functions (MF) of all variables. These MFs determine the membership of objects to fuzzy sets. In section 4.1 we describe the input variables and the output variable with their membership functions. In Section 4.2 we show the rules of system and in Section 4.3 we will explain system evaluation and tuning.

#### 4.1 Step 1: Define the linguistic variable and determine fuzzy sets

As described earlier, in our system we have 10 attributes and one output variable (listed in Table 1) with their fuzzy sets. All the input data had been taken from the log files database of usage obtained from the mobile operator. The output variable constitute a numerical range that represents the centroid of each result segments as obtained from SOM [16].

Fuzzy sets can have variety of shapes. However triangular or a trapezoid [17] can often provide an adequate representation of the expert knowledge. It also simplifies the computation process. Fig. 1 shows the membership function for the collected attributes and the fuzzy parameters defined in Table 1.

#### 4.2 Step 2: Construction of fuzzy rules

The aim of this step is to compile a set of fuzzy rules that is based on the data and that will be used to realize the categorization. We identified the clusters or user segments which are linked to plans that are available for users by the mobile operator. These segments, as explained earlier, have been obtained using SOM. Next we investigate common characteristics of customers constituting these clusters. The characteristics of each of these segments and the rules used in Fuzzy expert system is shown in Fig. 2.

**Table 1. The collected attributes, their numeric range and its associated linguistic values**

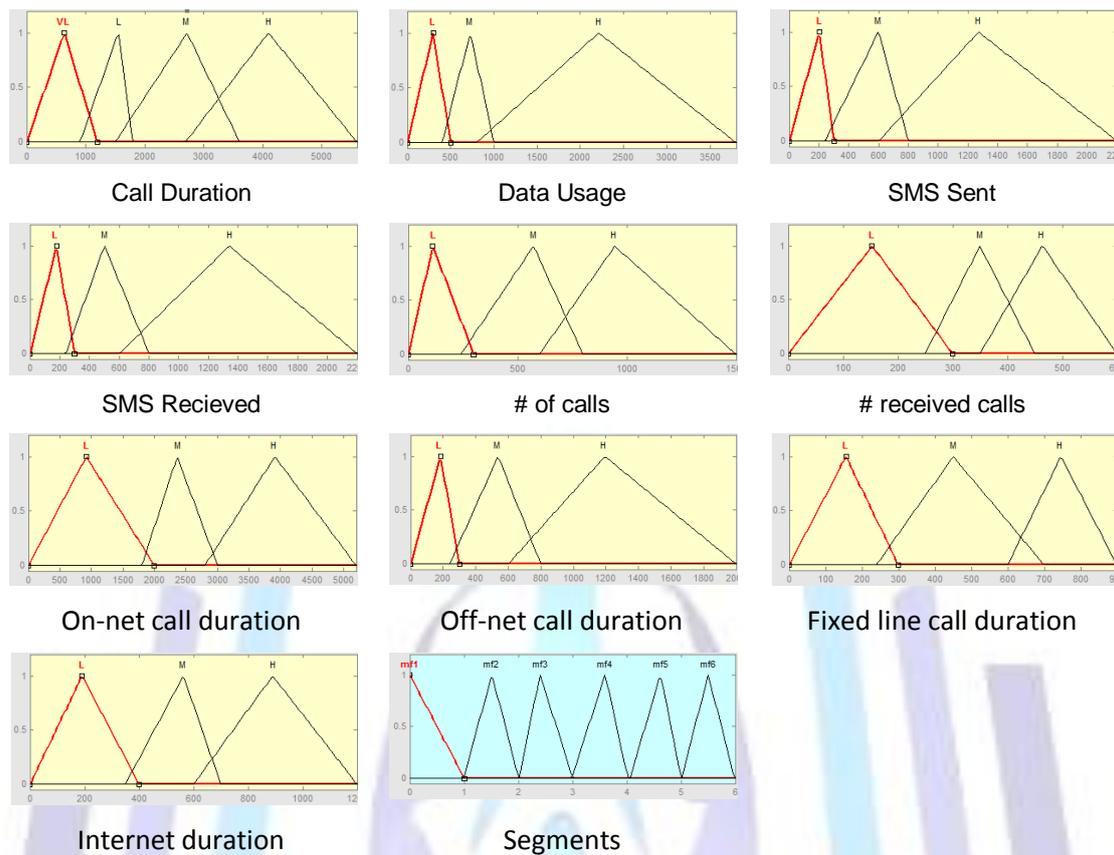
(VL: Very Low, L: Low, M: Moderate, H: High)

| Attribute                             | Linguistic Value | Numerical Range | Attribute                | Linguistic Value | Numerical Range |
|---------------------------------------|------------------|-----------------|--------------------------|------------------|-----------------|
| Duration of generated calls (minutes) | VL               | 0-1200          | Data usage (MB)          | L                | 0-500           |
|                                       | L                | 900-1800        |                          | M                | 400-1000        |
|                                       | M                | 1500-3600       |                          | H                | 800-3800        |
|                                       | H                | 2700-5600       |                          |                  |                 |
| Number of sent SMS                    | L                | 0-300           | Number of received SMS   | L                | 0-300           |
|                                       | M                | 240-800         |                          | M                | 240-800         |
|                                       | H                | 600-2200        |                          | H                | 600-2200        |
| number of generated calls             | L                | 0-300           | number of received calls | L                | 0-300           |
|                                       | M                | 240-800         |                          | M                | 250-450         |
|                                       | H                | 600-1500        |                          | H                | 350-600         |
| on-net calls (minutes)                | L                | 0-2000          | off-net calls (minutes)  | L                | 0-300           |
|                                       | M                | 1800-3000       |                          | M                | 240-800         |
|                                       | H                | 2800-5200       |                          | H                | 600-2000        |
| generated fixed line calls (minutes)  | L                | 0-300           | Internet Duration        | L                | 0-400           |
|                                       | M                | 240-700         |                          | M                | 350-700         |
|                                       | H                | 600-900         |                          | H                | 600-1200        |
| Output Variable (Segments)            | 1                | 28 -34          |                          |                  |                 |
|                                       | 2                | 21-33           |                          |                  |                 |
|                                       | 3                | 31-36           |                          |                  |                 |
|                                       | 4                | 0-23            |                          |                  |                 |
|                                       | 5                | 19-29           |                          |                  |                 |
|                                       | 6                | 9 -20           |                          |                  |                 |

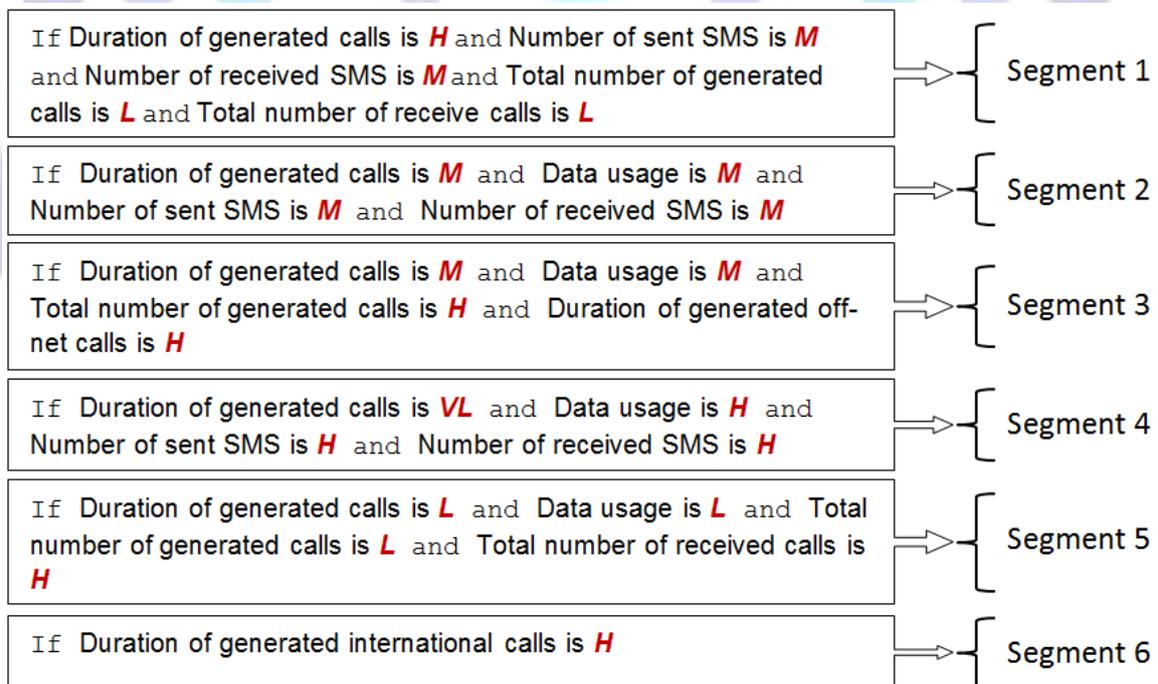
### 4.3 Step 3: Encode the fuzzy sets and perform fuzzy inference into the expert system

This study modeled and tested a fuzzy logic system to predict the segment of mobile phone users. The program's output (rules) was a text file that could be exported to Matlab for further investigation. For this purpose, Matlab's Fuzzy Logic Toolbox was used. The Fuzzy Inference System (FIS) editor of the developed program is shown in Fig 3. The summary of which is given below:

- Type = *mamdani*
- Decision method for fuzzy logic operators AND: *MIN*
- Decision method for fuzzy logic operators OR: *MAX*
- Implication method: *MIN*
- Aggregation method: *MAX*
- Defuzzification: *CENTROID* – Center of gravity



**Fig 1. Membership function for the collected attributes**



**Fig 2. Rules that govern the user segments based on the Attributes.**

*(VL: Very Low, L: Low, M: Moderate and H: High)*

### 4.4 Evaluating and tuning of the system

In order to be able to experiment with the system and properly interact with the parameter, we developed a basic UI using the Matlab IDE (Fig 3). It demonstrates how a determined value inputs affect the outputs. It can be usefully applied to give an idea about the interactions and as a tool to improve the fuzzy rules. It could also be applied as a good evaluation for performance. The figure shows and example in which duration of generated calls is 2800 and Data usage is 1900 and Total number of generated calls is 750 and Duration of generate off-net calls is 1000. The decision in this case is a user that should belong to Segment 3. Results showed that fuzzy logic provided stronger support for decisions. The shapes of membership functions may also be predicted by applying cluster or statistical analysis techniques to sub-samples of the data to be sorted. For the proposed method, we used the rule extracted from the SOM clustering to define our fuzzy expert system and we used real data collected about users to test and tune system parameters.

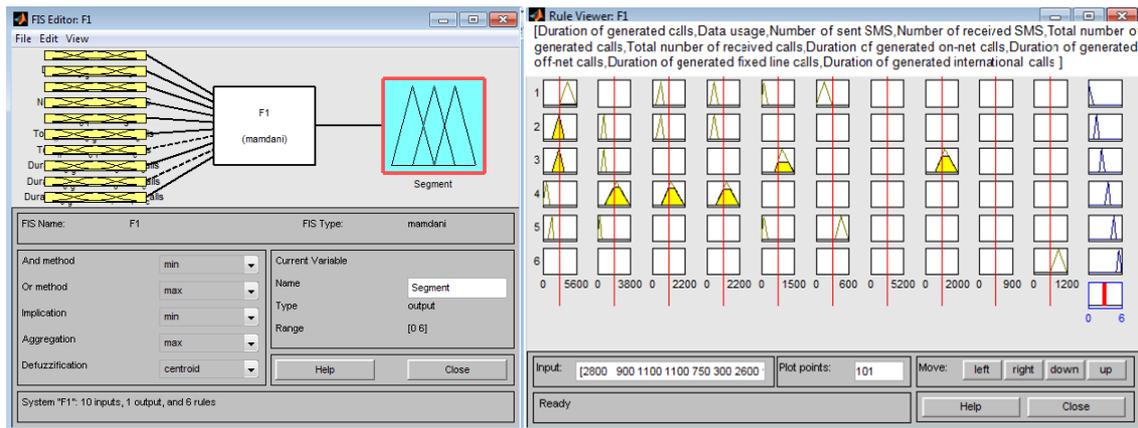


Fig 3. FIS and Rule Viewer of the system using Matlab IDE and toolboxes.

## 5. EVALUATION

To validate the proposed method we compare it against four classification methods:

- **Decision Tree:** a popular machine learning algorithm based on divide and conquer strategy. Decision trees are widely used in classification and categorization problems (e.g., [18] and [19]). Decision trees enjoys several advantages among which is its ability to process both numerical and categorical data types. It also maps efficiently to business rules.
- **Bagged Decision Tree:** Many research work (e.g., [19] and [20]) studied the adoption of machine learning ensemble to improve the result obtained from machine learning algorithms. We opted to experiment with Bootstrap aggregating (or bagging) to measure the improvement (if any) when applying it to decision trees.
- **Naive Bayes Classifier:** Naive bayes is a classical statistical classifier which estimates a set of conditional probabilities based on the training set. We attempt to estimate the conditional probability of the segment given the measurements collected about the users (i.e., attributes).
- **Neural Network:** Another classifier of different type from the ones above is artificial neural networks (ANN). It is widely popular. Maybe the attractiveness of ANN comes from the fact that it consist of sets of tunable and adjustable numerical parameters (weights) and their abilities to deal with non-linear problems.

Of course there are many more options in terms of classification methods to choose from but we opted to use the methods described briefly above because they are widely used in classification problems. We performed the experiments 10 times. We divided the data set (i.e., 6000 records) to training set (20%) and test set (80%). Each time we report the classification rate. Table 2 summarises the classification rates of the 10 runs as well as the average classification rate.



Table 2. Classification results based on 6000 records. 20% used for training and 80% for testing.

|                | Decision Tree | Bagged Decision Tree | Naïve Bayes  | Neural Network | Proposed Method (Fuzzy) |
|----------------|---------------|----------------------|--------------|----------------|-------------------------|
| Run 1          | 74.43         | 79.02                | 70.51        | <b>78.15</b>   | 88.79                   |
| Run 2          | 74.48         | 77.93                | 70.84        | 76.94          | 88.74                   |
| Run 3          | 74.47         | 78.42                | 71.06        | 76.21          | 88.76                   |
| Run 4          | 75.38         | 78.46                | 70.59        | 76.44          | 88.49                   |
| Run 5          | 74.66         | 77.95                | <b>71.58</b> | 76.17          | 88.70                   |
| Run 6          | 75.41         | <b>79.17</b>         | 70.85        | 77.07          | 88.35                   |
| Run 7          | 74.73         | 78.38                | 69.74        | 76.71          | 88.46                   |
| Run 8          | <b>75.45</b>  | 78.89                | 68.87        | 77.39          | <b>88.80</b>            |
| Run 9          | 75.06         | 78.90                | 69.04        | 77.08          | 88.43                   |
| Run 10         | 75.18         | 78.27                | 70.35        | 76.68          | 88.62                   |
| <b>Average</b> | <b>74.92</b>  | <b>78.54</b>         | <b>70.34</b> | <b>76.88</b>   | <b>88.61</b>            |

Our proposed categorization method achieved the best results based on the performed experiments. Highest classification rate achieved by the method was 88.80% and the average classification rate was 88.61%. Expectedly, Naïve Bayes Classifier reported the lowest classification rates (70% on average). This is due to its inherent naive design and probably the oversimplifying. As expected bagged decision trees outperformed single decision trees and an improvement gained from this ensemble method with classification rate reaching almost 80%.

## 6. CONCLUSION AND FUTURE WORK

In this paper we addressed the issue of the segmentation of mobile users based on their usage. This is a very important step toward the recommendation systems to suggest the best plan for users. We proposed a segmentation method that is based on Self Organizing Maps and Fuzzy System to realize the categorization. The motivation comes from the fact that users usually use fuzzy term to describe their usage behavior and understandably are not able to provide precise measurements. The system can predict the user segment based on these fuzzy terms. Experiments were designed and executed to validate the proposed method. Data has been collected from a leading mobile operator in Jordan to do the experiments. Six thousand records were collected and groundtruth. The method were compared to four widely used machine learning algorithm. Experiments demonstrated a better performance of the proposed method against the other methods. Moreover, this method will give our recommendations high adaptation capabilities to deal with new different behavior attribute and this will help also the telecommunications industry to define their customers' demands and personalize the solution based on that. Three main lines of research can be identified as a future work: (i) improve classification rates, (ii) use the proposed method as part of a recommendation system to measure properly the benefits of such system on consumers and (iii) perform longitudinal studies on the behavior of customers against recommended plan(s).

## ACKNOWLEDGMENTS

Our thanks to the experts who have contributed towards development of the template. Our thanks also to the developers and software engineers who contributed to the Weka project.

## REFERENCES

- [1] E. Jaser and I. Ahmad, "ICT Intervention to Enhance Health Services to Mothers and Children in Remote Communities in Jordan," *MobiCASE*, volume 95:303-310 (2012)
- [2] E. Jaser and I. Ahmad, "Towards the Automatic Management of Vaccination Process in Jordan," *Proceeding of the Sixth International Conference on Digital Society ICDS*, pp. 50–54, Valencia, Spain (2012)
- [3] E. Jaser, "Information and Mobile Technologies for Promoting Maternal-Child Health Care Status in Rural Areas of Jordan". *IJIM* 7(3) pp. 38-45 (2013)



- [4] D. E. Vengroff, "RecLab: a system for eCommerce recommender research with real data, context and feedback", Proceedings of the 2011 Workshop on Context-awareness in Retrieval and Recommendation, Palo Alto, California, p.31-38, February 2011,
- [5] Y. Cao and Y. Li, "An intelligent fuzzy-based recommendation system for consumer electronic products", Expert Systems with Applications: An International Journal, v.33 n.1, p.230-240 (2007)
- [6] H. Wang and H.Doong, "Online customers' cognitive differences and their impact on the success of recommendation agents", Information and Management, v.47 n.2, p.109-114, (2010)
- [7] N. Abdullah , Y. Xu and S. Geva, "Integrating collaborative filtering and matching-based search for product recommendations", Journal of Theoretical and Applied Electronic Commerce Research, v.8 n.2, p.34-48 (2013)
- [8] Y. Jiang , J. Shang and Y. Liu, "Maximizing customer satisfaction through an online recommendation system: A novel associative classification model", Decision Support Systems, v.48 n.3, p.470-479 (2010)
- [9] H. Kim , I. Ha , K. Lee , G. Jo and A. El-Saddik, "Collaborative user modeling for enhanced content filtering in recommender systems", Decision Support Systems, v.51 n.4, p.772-781 (2011)
- [10] J. Bobadilla , F. Ortega , A. Hernando and J. Bernal, "Generalization of recommender systems: Collaborative filtering extended to groups of users and restricted to groups of items", Expert Systems with Applications: An International Journal, v.39 n.1, p.172-186, (2012)
- [11] R. Eric Hostler , V. Y. Yoon and T. Guimaraes, "Recommendation agent impact on consumer online shopping: The Movie Magic case study", Expert Systems with Applications: An International Journal, v.39 n.3, p.2989-2999 (2012)
- [12] J. Kim , H. Kim , H. Oh and Y. Ryu, "A group recommendation system for online communities", International Journal of Information Management: The Journal for Information Professionals, v.30 n.3, p.212-219 (2010)
- [13] V. Zheng , Y. Zheng , X. Xie and Q. Yang, "Collaborative location and activity recommendations with GPS history data", Proceedings of the 19th international conference on World Wide Web, (2010)
- [14] Kiang, M.Y., HU, M.Y., Ficher, D.M. ,An extended self-organizing map network for market segmentation - a telecommunication example, Decision Support Systems, 42(1), 36-47, (2006)
- [15] Betty Chang , Chih-Wei Chang , Chih-Hung Wu, Fuzzy DEMATEL method for developing supplier selection criteria, Expert Systems with Applications: An International Journal, v.38 n.3, p.1850-1858, March, 2011
- [16] H. Payam and M. Mirzazadeh. "Visualizing market segmentation using self-organizing maps and Fuzzy Delphi methodADSL market of a telecommunication company" Expert Systems with Applications, 38.1 198-205, (2011)
- [17] J. O. Daramola , O. O. Oladipupo , A. G. Musa, "A fuzzy expert system (FES) tool for online personnel recruitments", International Journal of Business Information Systems, Vol.6 No.4, p.444-462, October (2010)
- [18] E. Jaser, J. Kittler and W. Christmas, "Hierarchical decision making scheme for sports video categorisation with temporal post-processing", Computer Vision and Pattern Recognition, Vol. 2 pp. 908-913. (2004)
- [19] E. Jaser, J. Kittler and W. Christmas, "Building classifier ensembles for automatic sports classification", International Workshop on Multiple Classifier Systems, pp. 366-374. (2003)
- [20] K. Machová,F. Barcák and P. Bednár, "A Bagging Method using Decision Trees in the Role of Base Classifiers", ActaPolytechnicaHungarica Vol. 3 No. 2, pp. 121-132. (2006)