A multi-criteria collaborative filtering recommender system for the tourism domain using Expectation Maximization (EM) and PCA–ANFIS

Mehrbakhsh Nilashi a,⁎, Othman bin Ibrahim a,⁎, Norafida Ithnin a, Nor Haniza Sarmin b

a Faculty of Computing, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Malaysia
b Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Malaysia

ARTICLE INFO

Article history:
Received 12 November 2014
Received in revised form 10 August 2015
Accepted 10 August 2015
Available online xxxx

Keywords:
Expectation Maximization
Clustering
PCA
ANFIS
Multi-criteria CF
Tourist recommendation

ABSTRACT

In order to improve the tourist experience, recommender systems are used to offer personalized information for online users. The hotel industry is a leading stakeholder in the tourism sector, which needs to provide online facilities to their customers. Collaborative Filtering (CF) techniques, which attempt to predict what information will meet a user’s needs based on data coming from similar users, are becoming increasingly popular as ways to combat information overload. They use a single rating as input. However, the multi-criteria based CF presents a possibility to provide accurate recommendations by considering the user preferences in multiple aspects and they can be an appropriate choice for the tourist. In this paper, we propose a new hybrid method for hotel recommendation using dimensionality reduction and prediction techniques. Accordingly, we have developed the multi-criteria CF recommender systems for hotel recommendation to enhance the predictive accuracy by using Gaussian mixture model with Expectation Maximization (EM) algorithm and Adaptive Neuro-Fuzzy Inference System (ANFIS). We have also used the Principal Component Analysis (PCA) for dimensionality reduction and to address multicollinearity induced from the interdependencies among criteria in multi-criteria CF dataset. Our experiments confirmed that the proposed hybrid method achieved high accuracy for hotel recommendation for the tourism sector.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Tourism as a strategic sector has provided a significant contribution to the economies of many nations around the world (Wan Lee and Brahmasrene 2013). It has provided a remarkable impact on the global economic development in which the contribution of this sector for the employing people and economic activity have estimated around 7.6% of global employment and US$ 5474 billion of 9.4% of global GDP (World Travel and Tourism Council 2009). According to the World Tourism Organization (2006), it is predicted that by 2020, tourist arrivals around the world will increase by over 200%. Impressive changes in the Information and Communications Technologies (ICTs) and the Internet has resulted in extensive transformation of the industry. According to the Travel Industry Association of America (www.tia.org) cited by (Lucas et al. 2013), in 2003, the major United States adult population (around 30%) has used Internet as a tool to check prices and schedules and seek information regarding destinations. 66% of tourists booked travel needs using the Internet. In addition, the ICTs have considerably improved the innovations in the tourism sector in management and marketing of tourism packages and brought about new paradigm shifts in this sector as discussed in many researches (Polo Peña et al. 2013, Chiu et al. 2009, Popescu and Grefenstette 2011, Morrison et al. 2001, Singh and Kasavana 2005, Connolly and Lee 2006, Pan et al. 2007, 2011, Buhalis and Law 2008, Xiang and Pan 2010, Buhalis and O’Connor 2005).

Tourism is an activity closely linked with personal interests and preferences (Chou et al. 2008, Wang et al. 2002, Benitez et al. 2007). Recommender systems designed in the tourism domain and applications known as Travel Recommender Systems (TRSs) or destination recommendation system, are a valuable tool for customers and travel agencies (Loh et al. 2004, Werthner and Ricci 2004). That is why many tourism web applications incorporate recommender systems. With this, they try to simulate the interaction with a human travel agent. Through the introduction of tourism recommending systems, tourists can easily access information about the hotels they need, thus, resulting in shorter lead-time for bookings, making last-minute decisions and generally, tailoring their preferences. Tourism recommender systems are a class of intelligent systems that render tourism related information services in the form of

http://dx.doi.org/10.1016/j.elerap.2015.08.004
1567-4223/© 2015 Elsevier B.V. All rights reserved.

guides and suggestions to users. This class of systems can be broadly classified as web-based tourism recommender systems and mobile recommender systems. Web-based tourism recommender systems are intelligent systems that are usually embedded in e-Tourism portals in order to deliver travel information guide, travel advice and travel planning recommendations.

Collaborative Filtering (CF) can be an appropriate choice for tourism object recommendation. Recommender systems based on CF are those in which recommendations only consider the similarity of terms between users. That is, collaborative systems recommend items that other users with similar interests like. However, traditional CF use a single rating as input, usually an overall numerical ranking by a user to an item. Hence, in some applications, this kind of recommendation does not meet users’ personalized needs and multi-criteria ratings are considered.

Multi-criteria based CF presents a possibility of providing accurate recommendations by considering the user preferences in multi-aspects of items. According to Adomavicius and Kwon (2007), pure CF-based recommender systems rely solely on product ratings provided by a large user community to generate personalized recommendation lists for each individual online user. In traditional CF systems the assumption is that customers provide an overall rating for the items which they have purchased, for example, using a 5-star rating system. However, given the value of customer feedback to the business, customers in some domains are nowadays given the opportunity to provide more fine-grained feedback and to rate products and services along various dimensions (Jannach et al. 2012a, Adomavicius et al. 2011, Nilashi et al. 2014c). According to Adomavicius and Kwon (2007), multi-criteria systems provide more information about user preferences than a single-rating system. And by adopting a decision theory, multi-criteria systems can provide rich tools for system designers to build more interesting systems as well (Lakiotaki et al. 2011). In addition, nowadays, allowing online visitors to provide fine-grained multi-criteria rating feedback is common in the travel and tourism sector. TripAdvisor tourism is one of the popular platforms which has provided users to rate hotels according to different criteria such as cleanliness, service or value for money (Nilashi et al. 2015a).

Adomavicius and Kwon (2007) developed a number of basic strategies to exploit multi-criteria ratings for improving the predictive accuracy of a recommender in terms of typical information retrieval measures. Later on, a number of additional techniques to leverage the detailed ratings in the recommendation process were proposed (Liu et al. 2011, Sahoo et al. 2012, Shambour and Lu 2011a,b, Jannach et al. 2012a,b, 2014, Nilashi et al. 2014a,b, 2015a). The work presented in this paper continues on these lines of research.

Overall, our work is similar to the works of which has used, clustering, combined with methods such as methods developed by Adomavicius and Kwon (2007) and Jannach et al. (2012a,b), where we use Neuro-Fuzzy techniques to predict the overall ratings from the given multi-criteria ratings. Furthermore, our work is, in some sense similar to that of Liu et al. (2011) and Nilashi et al. (2015a) where we apply clustering.

From the literature on multi-criteria CF, at the moment there is no implementation of PCA, Neuro-Fuzzy and clustering recommenders in multi-criteria CF, and this research tries to develop a recommender system in the tourism sector based on these approaches. Thus, in order to improve predictive accuracy of multi-criteria CF, we propose a new model using fuzzy logic, neural networks and clustering techniques. To the best of our knowledge, an artificial intelligence method (ANFIS), clustering method (EM) and dimensionality reduction (PCA) is applied for the first time in this research in the context of multi-criteria CF recommendations in particular for hotel recommendation based on multi-criteria CF.

1.1. Recommendation problem

In multi-criteria CF problem, there are m users, n items and k criteria in addition to the overall rating. Users have provided a number of explicit ratings for items; a general rating $R_0$ must be predicted in addition to k additional criteria ratings ($R_1, \ldots, R_k$). It can be configured to push new items to users in two ways, either by producing a Top-N list of recommendations for a given target, or by predicting the target user’s likely utility (or rating) for a particular unseen item. We will refer to these as the recommendation task and the rating prediction task in multi-criteria CF, respectively. Fig. 1 demonstrates the multi-criteria CF problem in case of prediction and recommendation tasks for an active user $U_a$ and active item $I_j$.

Recommendation is a list of N products, $TP=\{T_{p1}, T_{p2}, \ldots, T_{pn}\}$, that the active user will like the most. The recommended list usually consists of the products not already purchased by the active customer. This output interface of multi-criteria CF algorithms is also known as Top-N recommendation. Multi-criteria CF algorithms represent the entire $m \times n \times k$ user-item-criteria data as a tensor of ratings, $A$. Each entry $a_{ij}$ in tensor $A$ as shown in Fig. 1 represents the preference score (ratings) of the $i$th user on the $j$th item (hotel) as overall preference in addition to criteria ratings in the 3rd dimension. Each overall and criteria rating is within a numerical scale and it can as well be 0, indicating that the user has not yet rated that item.

Thus, the algorithm for a multi-criteria recommender system can be extended from a single-rating recommender system. Following this approach, Adomavicius and Kwon presented two approaches to leverage multi-criteria ratings through extending single-rating CF (Adomavicius and Kwon 2007). One is computing the overall user similarity through aggregating the similarities calculated from each individual criterion (Adomavicius and Kwon 2007). The other approach is aiming for a more holistic calculation of user similarity through multidimensional distance metrics. Each rating is presented in a multivariable format, such as

$$r_{ui} = f(r_{0i}, r_{1i}, \ldots, r_{ki})$$

where $r_{ui}$ is the overall rating that user $U_i$ has rated item $I_j$ and $r_{0i}, r_{1i}, \ldots, r_{ki}$ presents the rating of criterion 1, …, k.

In this paper, we use Pearson correlation coefficient approach for users and items similarity calculation. The Pearson correlation coefficient (the most commonly used weighting approach) measures the degree to which a linear relationship exists between two variables (McLaughlin and Herlocker 2004). In this research, it is used to evaluate how a certain user is related to an active user with respect to their preferences on given items. The Pearson correlation coefficient is derived from a linear regression model (see Eq. (2)). The range of the result of the equation is from −1 to 1, inclusively. More specifically, a result of "1" means the two users are positively related (absolute agreement), "−1" denotes they are negatively related (absolute disagreement), and "0" indicates no relation at all.

$$\text{sim}_{u,v} = \begin{cases} 
0 & \text{if } \sum_{i=1}^{k} (r_{ui} - \bar{r}_i)^2 = 0 \\
0 & \text{or if } \sum_{i=1}^{k} (r_{vij} - \bar{r}_i)^2 = 0 \\
\frac{\sum_{i=1}^{k} (r_{ui} - \bar{r}_i)(r_{vij} - \bar{r}_i)}{\sqrt{\sum_{i=1}^{k} (r_{ui} - \bar{r}_i)^2 \sum_{i=1}^{k} (r_{vij} - \bar{r}_i)^2}} & \text{otherwise}
\end{cases}$$

where $I_{uv}$ in Eq. (2) is a set of items that both user $u$ and $v$ rate, $I_u$ is a set of items that user $u$ rates and $I_v$ is a set of items that user $v$ rates.
2. Related work

The fuzzy logic field has grown considerably in a number of applications across a wide variety of domains like in the semantic music recommendation system (Lesaffre and Leman 2007), movie recommendation (Nilashi et al. 2014a,b) and product recommendations (Cao and Li 2007, Stormer et al. 2006). Castellano et al. (2007) developed a Neuro-Fuzzy strategy combined with soft computing approaches for recommending URLs to the active users. They used fuzzy clustering for creating a user profile considering the similarities between the users and fuzzy set theory for presenting the vagueness in the description of users' ratings. A conceptual framework based on fuzzy logic-based was proposed by Yager (2003) to represent and then justify the recommendation rules. In the proposed framework, an internal description of the items was used that relied solely on the preferences of the active user. Carbo and Molina (2004) developed an algorithm based on CF that ratings and recommendations were considered as linguistic labels by using fuzzy sets. A model proposed by Pinto et al. (2012) that combines fuzzy numbers, product positioning (from marketing theory) and item-based CF.

In the context of web recommendation systems, traditional single-rating CF recommender systems have been highly successful however, the research area regarding of the CF with multi-criteria ratings for items has been rarely touched and fairly this issue is unexplored. Especially, few number of researches have been conducted in the tourism recommendations context using multi-criteria ratings.

According to Stock et al. (2005), tourism is a worthy area for the application of Artificial Intelligence (AI), and, especially, in Decision Support Systems (DSSs) and recommender systems (Felfernig et al. 2007). Due to the importance of tourism as a strategic sector, several attempts have been made in developing recommender systems for the tourism domain (Wallace et al. 2003, Loh et al. 2004, Castillo et al. 2008, Ricci and Nguyen 2007). These researches often have used AI techniques for their purpose. For example, Schiaffino and Amandi (2009) uses intelligent agents, Lenar and Sobecki (2007) and Ngai and Wat (2003) apply fuzzy approaches, Huang and Bian (2009) employs Bayesian networks and many researchers have incorporated semantic approaches for the tourism domain (Jakkilinki et al. 2007, Kanellopoulos 2008). Berka and Plnig (2004) developed a recommendation system for travel recommender systems (TRSs) using fuzzy association rules. Optimization techniques have been used in many researches for tourism recommendation systems (Lee et al. 2009, Vansteenwegen and Souffriau 2011, Garcia et al. 2010, 2013, Castillo et al. 2008, Meehan et al. 2013). Several researchers also have used automatic clustering in the tourism recommender systems (Fenza et al. 2011, Castillo et al. 2008, Gavalas and Kenteris 2011, Nogueira et al. 2012, Moreno et al. 2013, Batet et al. 2012, Kurata and Haru 2013, Ruotsalo et al. 2013, Lucas et al. 2013).

Collaborative and content-based approaches have been widely used in recommendation systems for tourism domain. VacationCoach’s expert advice platform and Triplehop’s TripMatcher (Ricci 2002) are good examples of content-based systems. Basically, these systems match the users characteristics, preferences and needs with the set of destinations variables and features. In case of collaborative type recommendation for tourism, there are few works which have employed only the memory-based CF approach which TripAdvisor is probably the most popular tourism recommender system of this type. The recommendations in TripAdvisor are mainly based on ratings and comments collected from users. In this way, the recommendation system follows the similarity between the users and recommends items to the target/active by comparing items of other users with similar interests like. In this type recommendation systems users are grouped based on similar items’ ratings. Hence, similar ratings will be in a one group. However, despite popularity of CF recommendation systems, they also suffer from typical recommender systems drawbacks such as sparsity, scalability and cold-start.

The use of model-based CF approaches such as case-based reasoning in recommender systems for tourism has been chosen to overcome some shortcomings associated with memory-based approaches. Accordingly, Ricci and Werthner (2002) have provided a recommender system for tourism domain based on case-based reasoning, which data is gathered from existing external tourism portals. Lucas et al. (2013) implemented a recommendation methodology in a recommender system in the tourism domain that the classification was based on association. In addition their method was composed by several data mining algorithms;
however, the key aspect of the method was the join use of classification based on association and fuzzy sets.

Several researchers have attempted to combine collaborative and content-based filtering in travel recommender systems (Pazzani 1999, Delgado and Davidson 2002). By hybrid of these two approaches, they could achieve significant accuracy improvements on the travel recommender systems.

In case of multi-criteria CF, few researches has been conducted to develop the similarity calculation of the traditional memory-based CF approach to investigate multi-criteria rating (Tang and McCalla 2009, Manouselis and Costopoulou 2007, Adomavicius and Kwon 2007) that the similarities between users are estimated through aggregating traditional similarities from individual criteria or applying multidimensional distance metrics. In order to developing the idea of Adomavicius and Kwon (2007), Sahoo et al. (2006, 2011) extended the flexible mixture model (FMM) developed by Si and Jin (2003) to multi-criteria recommender systems. Li et al. (2008) presented a multicriteria rating approach to improve personalized services in mobile commerce using multilinear singular value decomposition (MVD). The aim of their paper was to exploit context information about the user as well as multi-criteria ratings in the recommendation process. Liu et al. (2011) presented a multi-criteria recommendation approach which is based on the clustering of users. Their idea is that for each user one of the criteria is “dominant” and users are grouped according to their criteria preferences. They applied linear least squares regression, assigning each user to one cluster, and evaluated different schemes for the generation of predictions. They applied the methods on hotel domain dataset with five criteria, Value, Location, Rooms, Service and Cleanliness. Jannach et al. (2012a) further developed the accuracy of multi-criteria CF by proposing a method using SVR for automatically detecting the existing relationships between detailed item ratings and the overall ratings. In addition, the learning process of SVR models was per item and user and lastly combined the individual predictions in a weighted approach.

Shambour and Lu (2011b) implemented a hybrid Multi-Criteria Semantic enhanced CF (MC-SeCF) approach to alleviate limitations such as sparsity and cold-start of the item-based CF techniques. The experimental results on MovieLens dataset demonstrated the effectiveness of their proposed approach in alleviating the sparsity and cold-start items problems. They achieved high accuracy and more coverage in very sparse and new items datasets than the benchmark item-based CF recommendation algorithms.

In this study, we consider the proposed method for Tourism domain recommender systems. However, the method can also be adopted for e-business and e-government applications recommender systems Shambour and Lu (2011a,b) for e-business and e-government applications. In our proposed method for building a model using PCA, ANFIS and clustering, the explicit ratings provided by users without the human expert intervention. Hence, in this paper we incorporate the multi-criteria ratings to the CF for hotel recommendation in the tourism domain and propose a new approach of recommendation using Expectation Maximization (EM) clustering technique, the Principal Component Analysis (PCA) and Adaptive Neuro-Fuzzy Inference Systems (ANFIS).

Thus, in comparison with research efforts found in the literature, our work has the following differences. In this research:

- A hybrid recommendation model using EM, PCA and ANFIS is proposed for increasing the predictive accuracy of the multi-criteria CF in tourism domain.
- ANFIS is used for knowledge discovery from the multi-criteria ratings provided by users without the human expert intervention.
- PCA is used for dimensionality reduction and dealing with the multi-collinearity problem exists in the multi-criteria ratings.

The remainder of this paper is organized as follows: Section 3 provides the research methodology along with all approaches used in the proposed model. Section 4 presents the multi-criteria modeling using proposed approaches along with the evaluations. Finally, conclusions and future work is presented in the Section 5.

3. Methodology of research

This section introduces the proposed recommendation model based on multi-criteria ratings to be validated in a tourist recommender system. The general framework of proposed model is shown in the Fig. 2. As can be seen in this figure, the recommendation model is composed by several data mining techniques; however, in relation to prior researches, the key aspect of this method is the join use of supervised and unsupervised dimensionality reduction techniques. Accordingly, in this study, we use ANFIS as a supervised approach for constructing the models of prediction. However, prior to the ANFIS, we perform clustering on data using EM algorithm as a unsupervised machine learning technique. The data clustering assists ANFIS to construct effective prediction models of users and items. We also use PCA for dimensionality reduction because the greatest source of difficulties in using ANFIS method is the existence of “multi-collinearity” in many sets of data that in this research PCA will overcome this problem. We select ANFIS approach to learn prediction models for users and items in each cluster and then combine those models in a weighted approach in the online phase. The form of discovered knowledge using ANFIS is fuzzy rules that are used for predicting overall ratings. The extracted rules are employed for prediction unknown ratings and also revealing real level of user preferences on items’ features (criteria). In addition, non-stochastic uncertainty emerging from vagueness and imprecision is handled using Membership...
Functions (MFs) produced by ANFIS. They are used for representation and reasoning users' behavior in providing rating based on their perception of items' features. The MFs formed by ANFIS are continuous and are more accurate in representing the features of items and user feedbacks. Furthermore, to prevent the problem of overfitting discussed in the previous researches (Jannach et al. 2012a, Sen et al. 2009), checking data is used to minimize and solve overfitting problem in the training data. Moreover, for predicting the unknown overall ratings, we try to solve the sparsity problem in criteria ratings by the neighborhood formation in each cluster using Pearson correlation coefficient. It should be noted that tensor decomposition techniques such as Higher Order Singular Value Decomposition (HOSVD) can also be used.

Therefore, for alleviating sparsity problem in individual criteria ratings, after forming the neighborhood in each cluster, we have used the average of the other criteria ratings as an estimate of the user's rating. Two methods have been widely used in recommender systems for neighbors selection: the Top-N method and the correlation weight threshold. In Top-N, a predefined number of users with greatest correlation are selected and in the correlation weight threshold all users with similarity correlation exceeding a certain threshold are selected. We use the Top-N method as recommended by Herlocker et al. (2002). To recommend an item to a target user, first assign the target user u to one of the clusters determined in the offline phase. We thus calculate a “mean rating vector” for the target user by calculating the user’s average rating value for each dimension given the user’s past ratings. The resulting rating vector is then compared to the center of each of the clusters and the user is finally assigned to the target cluster whose center has the highest similarity with the user’s mean rating vector in terms of the Pearson correlation coefficient. It should be also noted that in the proposed recommendation model and the methods used, K-Fold Cross-Validation (CV) (k = 5) is used for estimating the models using part of the training data and then validating those models with the remaining part of the training data.

3.1. ANFIS

The Fuzzy Logic (FL) was introduced by Zadeh (1965) to provide a solution for decisions making and handle uncertainty in vague, ambiguous and imprecise situations (Prihasto et al. 2014). It represents models or knowledge using IF-THEN rules (Nilashi et al. 2015b,c). Neural Networks (NNs) is a powerful technique learn system behavior by using system input–output data. NNs have good learning and generalization capabilities. The capabilities of learning and generalization in NNs enable the system to effectively address real-world problems. They can be a good choice when sufficient data be available for the problem. Then, NNs can effectively solve problems which cannot be solved or inefficiently solved by existing techniques, including fuzzy logic.

Both fuzzy logic and NNs have been very successful in solving many real-world problems. However, both technologies have some limitations as well which have prevented them from providing efficient solutions for multi-criteria CF problems. In fuzzy logic, it is usually difficult to determine the correct set of rules and membership functions from the users’ preferences in multi-criteria CF. Moreover, fine-tuning a fuzzy solution is even more difficult and takes longer. In neural networks, it is difficult to understand the “Black Box”, i.e., it is incomplete compared to a fuzzy rule based system description.

An appropriate combination of these two technologies (NeuroFuzzy) can effectively solve the problems of fuzzy logic and neural networks and, thus, can more effectively address the multi-criteria CF problems. A Neuro-Fuzzy approach was used to take advantage of the neural network’s ability to learn, and membership degrees and functions of fuzzy logic. The weights of the neural networks are mapped to fuzzy logic rules and member functions. Expressing the weights of the neural network by fuzzy rules also provides a better understanding of the “Black Box” and thus helps the better design of the neural network itself. Thus, while the learning of neural network is parameterized by the variation in input data, the
learning of ANFIS is fixed by the rules and membership function values that we define. A Neuro-Fuzzy system is functionally equivalent to a FIS. A FIS mimics a human reasoning process by implementing fuzzy sets and approximate reasoning mechanism that uses numerical values instead of logical values. A FIS requires a domain expert to define the Membership Functions (MFs) and to determine the associated parameters in both the MFs, and the reasoning section. However, there is no standard for the knowledge acquisition process and thus the results may be different if a different knowledge engineer is at work in acquiring the knowledge from experts.

A Neuro-Fuzzy system can replace the knowledge acquisition process by humans using a training process with a set of input–output training dataset. Thus instead of dependent on human experts the Neuro-Fuzzy system will determine the parameters associated with the Neuro-Fuzzy system through a training process, by minimizing an error criterion. A popular Neuro-Fuzzy system is called an ANFIS. ANFIS is a fuzzy system that uses artificial neural network theory to determine its properties (fuzzy sets and fuzzy rules).

The main objective of ANFIS modeling is to map the inputs to outputs to find a function $f$ for a given input vector $X = (x_1, x_2, x_3, \ldots, x_n)$ in order to predict output $y$ as close as possible to its actual output $\hat{y}$. Assume $m$ observations of multi-input–single-output data pairs be available such as $X = (x_{11}, x_{21}, x_{31}, \ldots, x_{n1})$, and

$$y_{i} = f(x_{i1}, x_{i2}, x_{i3}, \ldots, x_{in}) \quad (i = 1, 2, \ldots, m)$$

(3)

It is now possible to build a model using ANFIS in prediction task for any new input vector $X = (x_{1}, x_{2}, x_{3}, \ldots, x_{n})$. This prediction, $\hat{y}$, is an approximation of $y$ that can be presented as

$$\hat{y} = f(x_{1}, x_{2}, x_{3}, \ldots, x_{n}) \quad (i = 1, 2, \ldots, m)$$

(4)

In that direction, the goal is to minimize the difference between the actual output and the predicted one by determining an ANFIS model.

$$\min \sum_{i=1}^{m} \left[ f(x_{i1}, x_{i2}, x_{i3}, \ldots, x_{in}) - y_{i} \right]^2$$

(5)

Indeed, in ANFIS the linguistic Takagi and Sugeno (TSK) type, fuzzy IF-THEN rules are used for prediction task. These rules are generated by training the model to approximate $f$ by $f$ using $m$ observations of $n$-input–single-output data pairs $(X_i, y_i)$. ANFIS has a structure that consists of nodes and directional links through which the nodes are connected. Back-propagation strategy is used to train the MFs, while the least mean squares algorithm determines the coefficients of the linear combinations in the consequent part of the model. TSK type fuzzy IF-THEN rules are used in ANFIS model, for example:

IF $x$ is $A_1$ AND $y$ is $B_1$. THEN $f_1 = p_1 x + q_1 y + r_1$

IF $x$ is $A_2$ AND $y$ is $B_2$. THEN $f_2 = p_2 x + q_2 y + r_2$

(6)

where $x$ and $y$ are the inputs, $f_i$ is the output, $p$, $q$, and $r$ are the design parameters that are determined by the users during the training process. $A_i$ and $B_i$ are the fuzzy sets according to predefined MFs. An ANFIS model with two inputs and two fuzzy rules are implemented in Fig. 3.

The first hidden layer is for fuzzification of the input variables. The outputs of layer 1 are fuzzy membership grade of the inputs, which are given by:

$$O_i^1 = \varphi_{a_i}(x), \quad i = 1, 2$$

$$O_i^1 = \varphi_{b_i}(y), \quad i = 3, 4$$

(7)

where $\varphi_{a_i}$ and $\varphi_{b_i}$ are MF.

There are fixed number of nodes in the second layer, labeled with $M$. The outputs of the second layer can be defined as:

$$O_i^2 = w_i = \varphi_{a_i}(x) \times \varphi_{b_i}(y), \quad i = 1, 2$$

(8)

where $w_i$s are so-called firing strength of the rules.

In the third layer, the number of nodes is also fixed, labeled with $N$. It normalizes the rule strengths from the second layer. The output of this layer can be defined as:

$$O_i^3 = w_i = \frac{w_i}{w_1 + w_2}, \quad i = 1, 2$$

(9)

which are the so-called normalized firing strengths. The consequent parameters of the rule are determined in the fourth layer. The output of each node in this layer is the product of the normalized firing strength and the polynomial defined in fuzzy rule, shown as:

$$O_i^4 = w_i f_i = w_i(p_i x + q_i y + r_i), \quad i = 1, 2$$

(10)

The fifth layer computes the overall output as the summation of all incoming signals. There is only one node in this layer, labeled with $S$. Hence, the output of this layer can be presented as:

$$O_i^5 = \sum_{i=1}^{2} w_i f_i = \left( \sum_{i=1}^{2} w_i f_i \right) / w_1 + w_2$$

(11)

There are two adaptive layers in the ANFIS architecture, namely the first layer and the fourth layer. There are three modifiable parameters in the first layer $a$, $b$, so-called premise parameters, which are related to the shape of the MF. In the fourth layer, there are also three modifiable parameters $p$, $q$, $r$, so-called consequent parameters, which are related to the output of the first order polynomial. When the premise parameters are fixed, the output of the ANFIS can be written as:

$$f = \frac{w_1}{w_1 + w_2} f_1 + \frac{w_2}{w_1 + w_2} f_2 = w_1 f_1 + w_2 f_2$$

(12)

Substituting the Eq. (10) into the Eq. (12):

$$f = w_1 (p_1 x + q_1 y + r_1) + w_2 (p_2 x + q_2 y + r_2)$$

(13)

which is a linear combination of the modifiable consequent parameters $p_1$, $q_1$, $r_1$, $p_2$, $q_2$, and $r_2$. The least squares method can be used to identify the optimal values of these parameters easily. In each epoch, the Least-Squares Estimator (LSE) method is used to optimize the consequent parameters, while the premise parameters are fixed. The output of the ANFIS is calculated by employing the consequent parameters found in the forward pass. Once the optimal consequent parameters are found, the Back-Propagation (BP) method will immediately start to adjust the premise parameters corresponding to the fuzzy sets in the input domain immediately, according to the output error. It has been proven that this hybrid algorithm is highly efficient comparing with a standard gradient method in training the ANFIS.

3.2. Clustering using EM algorithm

It is well known that the $k$-means algorithm is an instance of Expectation Maximization (EM) algorithm which is a general algorithm of density estimation. This algorithm is based on distance. Gaussian mixture model with EM algorithm is a power full approach for clustering. EM algorithm is model based iterative algorithm for solving the clustering problem where the data is incomplete or considered incomplete. EM algorithm is an optimization algorithm for constructing statistical models of the data (Mitra et al. 2003). In this algorithm each and every data instance belongs to each and every cluster with a certain probability. EM algorithm starts with initial estimates and iterates to find the
Fig. 3. The ANFIS architecture.

maximum likelihood estimates for the parameters. The quality of EM algorithm becomes very good when using huge dataset. It has been also demonstrated that EM is a good clustering method in terms of computation time and accuracy (Jung et al. 2014, Nathiya et al. 2010). In addition, this study EM is chosen to cluster data for the following reasons among others (Ordonez and Omiecinski 2002). (1) It has a strong statistical basis, (2) it is linear in database size, (3) it is robust to noisy data, (4) it can accept the desired number of clusters as input, (5) it can handle high dimensionality, and (6) it converges fast given a good initialization.

The mathematical background of EM algorithm is shown here in this section (Mitra et al. 2003). Given a dataset \( \{x_i\}_{i=1}^N \), the task of assigning a cluster for each instance in the dataset, is the goal that we aspire for. Let there be \( N \) data points in the dataset and let us assume that the number of clusters is \( K \). Let the index of the cluster be modeled as a random variable \( z = j \) and let its probability be given by a multinomial distribution satisfying \( \sum_{j=1}^K \pi_j = 1 \). Such that

\[
\pi_j = p(z = j), \quad \forall j, \quad j = 1, \ldots, K
\]  

It is assumed that \( p(x|z = j) \sim N(\mu_j, \sigma_j) \) is a Gaussian distribution, \( I \) denotes the identity matrix of order \( j \). The unknown parameters of the model namely the mean \( \mu_j \), variance \( \sigma_j \) and the distribution function \( \pi_j \) are estimated.

\[
\theta = \left\{ \mu_j, \sum_{j=1}^K, \pi_j \right\}
\]

where \( z \) is an unknown hidden variable. The total log likelihood of all data is given by

\[
l(\theta, D) = \log \prod_{n=1}^N \left[ \sum_{j=1}^K \pi_j \exp \left[ -\frac{||x_n - \mu_j||^2}{2\sigma_j^2} \right] \right]
\]  

The parameter values that maximize the likelihood function \( l(\theta, D) \) are the ones that are chosen. Here \( D \) denotes the data. This optimization is complicated and to solve this some of the unknowns are assumed to be known, while estimating the others and vice versa. For each class, the conditional expectation of \( z = j \) given the data and the parameters.

\[
w_{lj} = p(z = j|x, \theta) = \frac{p(x|z = j, \theta)p(z = j|\pi_j)}{p(x|\theta)} = \frac{\pi_j N(x_i|\mu_j, \sum_j)}{\sum_{j=1}^K \pi_j N(x_i|\mu_j, \sum_j)}
\]  

Since each point \( x \) contributes to \( w_j \) in some proportion, for particular \( x_i \) we have

\[
w_j = \frac{\pi_j N(x_i|\mu_j, \sum_j)}{\sum_{j=1}^K \pi_j N(x_i|\mu_j, \sum_j)}
\]

The optimization algorithm is called EM and has the following steps: Assume we have some random initial estimates of the means and variances of the model \( \mu_j^{(0)}, \sigma_j^{(0)}, \pi_j^{(0)} \). Algorithm 1, describes the EM algorithm.

**Algorithm 1 EM Algorithm.**

**Initialize:** means and variances of the model \( \mu_j^{(0)}, \sigma_j^{(0)}, \pi_j^{(0)} \).

**Step 1. Expectation:** Using the estimates of

\[
\theta^{(t)} = \left\{ \mu_j^{(t)}, \sum_j^{(t)}, \pi_j^{(t)} \right\},
\]

parameters compute the estimate of

\[
w_{lj}^{(t)} = p(z = j|x_i, \theta^{(t)}) = \frac{n_j^{(t)} p(x_i|z = j, \theta^{(t)})}{\sum_{j=1}^K n_j^{(t)} p(x_i|z = j, \theta^{(t)})}
\]

**Step 2. Maximization:** Using estimates of \( w_{lj}^{(t)} \), update the estimates of the model parameters

\[
\mu_j^{(t+1)} = \frac{\sum_{i=1}^N w_{lj}^{(t)} x_i}{\sum_{i=1}^N w_{lj}^{(t)}}
\]

\[
\sigma_j^{(t+1)} = \frac{\sum_{i=1}^N w_{lj}^{(t)} (x_i - \mu_j^{(t+1)})^2}{\sum_{i=1}^N w_{lj}^{(t)}}
\]

\[
\pi_j^{(t+1)} = \frac{1}{K} \sum_{i=1}^N w_{lj}^{(t)}
\]

**Step 3. Repeat steps expectation and maximization until the parameter change gets small enough.**

3.3. Solving multi-collinearity issue using PCA

Principal Component Analysis (PCA) is a tool for data compression and information extraction (Nilashi et al. 2015a). In some situations, there are many correlated or redundant data which must be compressed in a manner to retain the essential information. Among the widely used multivariate statistical methods, PCA is a powerful tool for analyzing such data because of its ability to handle large numbers of highly correlated, noisy and redundant variables. Using PCA, a number of related variables are transformed to a set of uncorrelated variables. It is concerned with explaining the variance–covariance structure of a set of variables through several linear combinations of these variables. Its general objectives are data reduction and interpretation.

By using initial analysis of the data in multi-criteria experimental dataset, an overlap of information can be found that affect overall rating predictions. In addition, there exists significant
correlation among the input variables which are used as inputs of multi-criteria CF to build an inferential model. However, the data can be compressed to retain the essential information and make the input variables uncorrelated.

In the field of recommender systems, PCA has for example been applied by Goldberg et al. (2001) in the Jester joke recommender. In their approach, PCA was performed in an offline phase and they applied clustering on the resulting projection of the data in a two-dimensional space. Our approach is different from their work. We applied PCA after the initial clustering process individually on each cluster and determined a suitable number of principal components to retain for each cluster. Then, as inputs in ANFIS, we used the PCs for overall ratings prediction. Following this approach, it allowed us to achieve a highly accurate and up-to-date recommendations for overall ratings prediction.

From the experimental dataset, if we consider seven variables in the matrix X, the procedure of dimensionality reduction for overcoming the multi-collinearity can be defined in two steps as follows:

- Perform PCA on matrix X that consists of user ratings on items’ criteria.
- PCs selection from PCA.

The selected number of PCs along with the desired output f (overall rating) are employed in developing the inferential models. Fig. 4 illustrates the PCA–ANFIS network structure with two PCs.

Reducing the dimensionality of a dataset which includes a large number of interrelated variables is the main objectives of PCA. It keeps as much variation as possible in the original dataset and performs this process by transforming the original variables to a new set of variables which are called Principal Components (PCs). The generated PCs are basically uncorrelated and ordered where the first of them includes most of the variation provided by the original variables. For constructing a PCA initialization model of multi-criteria CF, the multi-criteria dataset can be sufficiently described using some chosen parameters in relation to the original variables with no significant loss of information. The issue of multi-collinearity in the multi-criteria CF data is also eliminated. The number of PCs that sufficiently represents the original data set is then selected.

As in multi-criteria CF there exist interdependencies between variables of multi-criteria dataset, in such situation, the main issue is the multi-collinearity of the data that needs to be solved. If the input variables of dataset are highly collinear, using the original data for supervised learning methods such as regression, classification and also Neuro-Fuzzy methods will result an ill-conditioned problem. Thus, in this situation, by reducing the dimension of dataset, PCA can be used to address multi-collinearity problems. The compressed model generated by PCA which consists of PCs provides linear combinations of the original variables of dataset.

Hence, instead of using the original variables as inputs in the ANFIS, for the multi-criteria CF, selected PCs are used from the PCA algorithm. The developed PCA–ANFIS architecture is illustrated in Fig. 5. From this figure, it can be seen that using the PCA approach, dimensionality of multi-criteria dataset can be adequately reduced. Also, later, we will demonstrate that developing PCA–ANFIS helps the proposed multi-criteria recommendation system to overcome the issue of multi-collinearity in the data and accordingly accuracy improvement in relation to solely using ANFIS.

The inputs to the ANFIS were the selected PCs of the dataset by applying PCA. N MFs was used for each input in fuzzification process which the total number of $M^N$ rules were generated. Since PCs are used as inputs in ANFIS model, the first-order Sugeno fuzzy model provides the following rule-based structure for four PCs:

\[
\text{If PC1 is } A \text{ and PC2 is } B \text{ and PC3 is } C \text{ and PC4 is } D, \text{ then } f = p \cdot \text{PC1} + q \cdot \text{PC2} + r \cdot \text{PC3} + s \cdot \text{PC4}
\]

where $A$, $B$, $C$, and $D$ indicate fuzzy sets for the input of system, $(p, q, r, s)$ is the consequent parameter set, and $f$ denotes the output after aggregating the fuzzy rules. For the ANFIS employed in this study, the nodes in Layer 1 (premise parameters) were all generalized Gaussian MFs which have a flexible parameterization. For example, for the fuzzy set $A$, the generalized Gaussian MF takes the form:

\[
\mu_A(\text{PC}_l) = e^{-\frac{d^2}{a^2}}
\]

where $(a, b, c_i)$ is the parameter set of the MFs in the premise part of fuzzy IF-THEN rules that change the shapes of the MFs.

The next layer (Layer 2) multiplies the inputs from the nodes in Layer 1 and generates the firing strength of the rules. The output of this layer is given by:

\[
w_i = \mu_{A_1}(\text{PC}_1) \times \mu_{B_1}(\text{PC}_1) \times \mu_{C_1}(\text{PC}_1) \times \mu_{D_1}(\text{PC}_1)
\]

where $w_i$ is the firing strength of rule $i$.

Accordingly, the overall ratings ($O$) can be calculated by Eq. (13).
4. Experimental results

4.1. Analyzing and pre-processing the dataset

In order to analyze the effectiveness of the proposed method, several experiments were conducted on TripAdvisor datasets provided by TripAdvisor website (www.tripadvisor.com). TripAdvisor represents the world largest and most successful social networking and community site in tourism (O’Connor 2008). The platform facilitates the reviewing of hotels around the world and brings together individuals in discussion forums and provides users with independent travel reviews and comments. In TripAdvisor website users can rate a hotel according to 7 different aspects: Value aspect, Rooms, Location, Cleanliness, Check in/front desk, Service and Business Service. In addition, users provide overall ratings on hotels. Ratings ranges from 0 to 5 stars, and −1 indicates this aspect rating is missing in the original html file.

Generally, the information of ratings is presented in four-fold < userID; itemID; Overall rating; Criteria rating >. This data can easily be converted into a tree dimensional rating tensor, whereas the first dimension (m) spans the number of users and the second dimension (n) spans the number of items and the third dimension (k) spans the number of criteria (\(\Delta m,n,k\)).

The experimental dataset which has been used in this study includes 1264 hotels, 85,424 users and 7 criteria. Tables 1 and 2 present the sample and statistics of raw TripAdvisor dataset for seven criteria and overall rating, respectively. In Table 2, there are 28,500 tuples of ratings that the range of ratings is between 1 and 5 on criteria and overall ratings. The average evaluation grade is 2.69, 1.45, 2.70, 2.20, 2.84, 1.36, 2.72 and 0.46 for “Overall rating”, “Location aspect”, “Rooms aspect”, “Value aspect”, “Cleanliness aspect”, “Check in/front desk aspect”, “Service aspect and Business Service aspect”, respectively.

In Table 2 show that there are 28,500 tuples of rating in the original dataset, however as we can see the users have rated hotels are very few. The sparsity rate is:

\[
\text{Sparsity} = \frac{\text{numRatings}}{\text{numUsers} \times \text{numItems}} = 1 - \frac{85424}{28500 \times 1264} \approx 0.9997
\]

Fig. 5. PCA–ANFIS for multi-criteria CF.
That means, the sparsity level of the experimental dataset is 99.97%.

In order to check the collinearities between the criteria’ variables, we also compute the linear correlations between the input attributes as shown in Table 3. As can be seen high collinearities ($r^2$ value) can be found among the 7 criteria.

4.2. Clustering with EM

As indicated in Section 3, we applied the EM clustering on users’ ratings in 3-order tensor. In every clustering method, choosing the right number of clusters is important. In EM clustering, with the Gaussian mixture model, the likelihood must be optimized. Hence, for this optimization, the best cluster number is selected by evaluating various values for the number of clusters. It should be noted that according to Pelleg and Moore (2000), we used information theoretic criterion like the Akaike Information Criterion (AIC) (Akaike 1974) to choose the value optimal number of cluster. Accordingly, in the experimental dataset, we have used a resubstitution AIC estimate and evaluated a number of clusters from 1 to 20. In addition, in the clustering procedure, we applied 5-fold cross validation to obtain unbiased result. In Fig. 6, we present the various numbers of clusters to select the best cluster based on chosen criterion. This figure shows that the best criterion value (660091.311541) is obtained when 6 clusters are generated by EM. In Fig. 7, the clusters generated by EM are visualized. For visualizing the dataset clusters into the original space, a PCA is used in order to obtain a 2D representation. It was used to visualize clusters in the scatter plot using the first and second PCs. Accordingly, the cluster centers are presented in Table 4. These cluster centers are used to assign newly arriving data points to a cluster based on their Euclidian distance. From the Table 4, it can be seen that EM has obtained the 6 clusters from our experimental dataset. It should be noted that 6 clusters have been automatically generated by EM. These clusters are used in PCA and then ANFIS for prediction models.

4.3. Dimensionality reduction with PCA

We applied the PCA on the clusters obtained from the experimental dataset using EM algorithm that the results in the following are presented. It should be noted the results were obtained from dataset with seven criteria without considering the overall ratings. Tables 5–10 describe the eigenvalues associated to the factors for six clusters. We have also the percentage of the total variance (individual and cumulative). In Fig. 8, Scree plot from PCA results for six clusters are demonstrated.

Table 3 describes the eigenvalues associated to the factors obtained by PCA for first cluster. The obtained results also have the percentage of the total variance (individual and cumulative). In addition, from Table 5, it can be noticed that first two PCs provide 76.95% and 84.93% of information.

Table 6 describes the eigenvalues associated to the factors obtained by PCA for second cluster. The obtained results also have the percentage of the total variance (individual and cumulative). In addition, from Table 6, it can be noticed that first two PCs provide 65.79% and 80.04% of information.

Table 1
A sample of the multi-criteria rating of TripAdvisor dataset.

<table>
<thead>
<tr>
<th>User ID</th>
<th>Overall rating</th>
<th>Value</th>
<th>Rooms</th>
<th>Location</th>
<th>Cleanliness</th>
<th>Check in/front desk</th>
<th>Service</th>
<th>Business Service</th>
<th>Hotel ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>-1</td>
<td>5</td>
<td>-1</td>
<td>5</td>
<td>-1</td>
<td>hotel_565550</td>
</tr>
<tr>
<td>14</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>hotel_566077</td>
</tr>
<tr>
<td>20</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>hotel_566077</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>22</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>hotel_570888</td>
</tr>
<tr>
<td>45</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>hotel_570888</td>
</tr>
<tr>
<td>11</td>
<td>5</td>
<td>-1</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>hotel_572859</td>
</tr>
</tbody>
</table>

Table 2
Statistics of hotel rating on seven criteria and overall rating.

<table>
<thead>
<tr>
<th>Overall rating</th>
<th>Value</th>
<th>Rooms</th>
<th>Location</th>
<th>Cleanliness</th>
<th>Check in/front desk</th>
<th>Service</th>
<th>Business Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>28,500</td>
<td>28,500</td>
<td>28,500</td>
<td>28,500</td>
<td>28,500</td>
<td>28,500</td>
<td>28,500</td>
</tr>
<tr>
<td>Mean</td>
<td>2.69</td>
<td>2.20</td>
<td>2.70</td>
<td>1.45</td>
<td>2.84</td>
<td>1.36</td>
<td>2.72</td>
</tr>
<tr>
<td>Min.</td>
<td>0</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>Max.</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 3
Correlation among criteria.

<table>
<thead>
<tr>
<th>Variable Y</th>
<th>Variable X</th>
<th>$r^2$</th>
<th>Variable Y</th>
<th>Variable X</th>
<th>$r^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>Rooms</td>
<td>0.8372</td>
<td>Location</td>
<td>Cleanliness</td>
<td>0.7974</td>
</tr>
<tr>
<td>Value</td>
<td>Location</td>
<td>0.8422</td>
<td>Location</td>
<td>Check in/front desk</td>
<td>0.8421</td>
</tr>
<tr>
<td>Value</td>
<td>Cleanliness</td>
<td>0.7521</td>
<td>Location</td>
<td>Service</td>
<td>0.7824</td>
</tr>
<tr>
<td>Value</td>
<td>Check in/front desk</td>
<td>0.8605</td>
<td>Location</td>
<td>Business Service</td>
<td>0.8374</td>
</tr>
<tr>
<td>Value</td>
<td>Service</td>
<td>0.8514</td>
<td>Cleanliness</td>
<td>Check in/front desk</td>
<td>0.8623</td>
</tr>
<tr>
<td>Value</td>
<td>Business Service</td>
<td>0.7878</td>
<td>Cleanliness</td>
<td>Service</td>
<td>0.7826</td>
</tr>
<tr>
<td>Rooms Location</td>
<td>Cleanliness</td>
<td>0.8256</td>
<td>Check in/front desk</td>
<td>Service</td>
<td>0.8682</td>
</tr>
<tr>
<td>Rooms Location</td>
<td>Check in/front desk</td>
<td>0.7987</td>
<td>Check in/front desk</td>
<td>Business Service</td>
<td>0.7837</td>
</tr>
<tr>
<td>Rooms Location</td>
<td>Service</td>
<td>0.9123</td>
<td>Service</td>
<td>Business Service</td>
<td>0.8937</td>
</tr>
<tr>
<td>Rooms Location</td>
<td>Business Service</td>
<td>0.8125</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 7 describes the eigenvalues associated to the factors obtained by PCA for the third cluster. The obtained results also have the percentage of the total variance (individual and cumulative). In addition, from Table 7, it can be noticed that first two PCs provide 51.50% and 65.72% of information.

Table 8 describes the eigenvalues associated to the factors obtained by PCA for the fourth cluster. The obtained results also have the percentage of the total variance (individual and cumulative). In addition, from Table 8, it can be noticed that first two PCs provide 40.21% and 53.63% of information.

Fig. 6. Best cluster based on chosen criterion.

Fig. 7. Visualization of clusters on PCA axes.

### Table 7

<table>
<thead>
<tr>
<th>Number of Clusters</th>
<th>Criterion Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>704812.560326</td>
</tr>
<tr>
<td>1</td>
<td>863749.183603</td>
</tr>
<tr>
<td>2</td>
<td>836959.462740</td>
</tr>
<tr>
<td>3</td>
<td>744986.404010</td>
</tr>
<tr>
<td>4</td>
<td>716333.967586</td>
</tr>
<tr>
<td>5</td>
<td>660993.315441</td>
</tr>
<tr>
<td>6</td>
<td>821898.619479</td>
</tr>
<tr>
<td>7</td>
<td>810175.042723</td>
</tr>
<tr>
<td>8</td>
<td>1205808.277350</td>
</tr>
<tr>
<td>9</td>
<td>1195185.526033</td>
</tr>
<tr>
<td>10</td>
<td>1136947.545377</td>
</tr>
<tr>
<td>11</td>
<td>1149115.158033</td>
</tr>
<tr>
<td>12</td>
<td>1128803.253662</td>
</tr>
<tr>
<td>13</td>
<td>1119298.836241</td>
</tr>
<tr>
<td>14</td>
<td>102933.158352</td>
</tr>
<tr>
<td>15</td>
<td>899327.018714</td>
</tr>
<tr>
<td>16</td>
<td>997434.313154</td>
</tr>
<tr>
<td>17</td>
<td>1046513.971524</td>
</tr>
<tr>
<td>18</td>
<td>895661.602131</td>
</tr>
<tr>
<td>19</td>
<td>1046513.971524</td>
</tr>
<tr>
<td>20</td>
<td>1160162.357771</td>
</tr>
</tbody>
</table>

### Table 8

<table>
<thead>
<tr>
<th>Number of Clusters</th>
<th>Criterion Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>704812.560326</td>
</tr>
<tr>
<td>1</td>
<td>863749.183603</td>
</tr>
<tr>
<td>2</td>
<td>836959.462740</td>
</tr>
<tr>
<td>3</td>
<td>744986.404010</td>
</tr>
<tr>
<td>4</td>
<td>716333.967586</td>
</tr>
<tr>
<td>5</td>
<td>660993.315441</td>
</tr>
<tr>
<td>6</td>
<td>821898.619479</td>
</tr>
<tr>
<td>7</td>
<td>810175.042723</td>
</tr>
<tr>
<td>8</td>
<td>1205808.277350</td>
</tr>
<tr>
<td>9</td>
<td>1195185.526033</td>
</tr>
<tr>
<td>10</td>
<td>1136947.545377</td>
</tr>
<tr>
<td>11</td>
<td>1149115.158033</td>
</tr>
<tr>
<td>12</td>
<td>1128803.253662</td>
</tr>
<tr>
<td>13</td>
<td>1119298.836241</td>
</tr>
<tr>
<td>14</td>
<td>102933.158352</td>
</tr>
<tr>
<td>15</td>
<td>899327.018714</td>
</tr>
<tr>
<td>16</td>
<td>997434.313154</td>
</tr>
<tr>
<td>17</td>
<td>1046513.971524</td>
</tr>
<tr>
<td>18</td>
<td>895661.602131</td>
</tr>
<tr>
<td>19</td>
<td>1046513.971524</td>
</tr>
<tr>
<td>20</td>
<td>1160162.357771</td>
</tr>
</tbody>
</table>

Table 7 describes the eigenvalues associated to the factors obtained by PCA for the third cluster. The obtained results also have the percentage of the total variance (individual and cumulative). In addition, from Table 7, it can be noticed that first two PCs provide 51.50% and 65.72% of information.

Table 8 describes the eigenvalues associated to the factors obtained by PCA for the fourth cluster. The obtained results also have the percentage of the total variance (individual and cumulative). In addition, from Table 8, it can be noticed that first two PCs provide 40.21% and 53.63% of information.

---

Table 9 describes the eigenvalues associated to the factors obtained by PCA for the fifth cluster. The obtained results also have the percentage of the total variance (individual and cumulative). In addition, from Table 9, it can be noticed that first two PCs provide 40.01% and 54.31% of information.

Table 10 describes the eigenvalues associated to the factors obtained by PCA for the sixth cluster. The obtained results also have the percentage of the total variance (individual and cumulative). In addition, from Table 10, it can be noticed that first two PCs provide 69.83% and 44.75% of information.

The eigenvalues that are associated with the factors are indicators for their importance. In our work, we decided to use the rule proposed by Cattell (1966) and create “scree” plots as shown in Fig. 8 where we plot the eigenvalues of the factors to detect “elbows” that indicate possible changes in the structure of the data.
For the first plot as shown in Fig. 8a, we can include the elbow into the selection i.e. we select $k = 3$ factors. Indeed, the eigenvalue ($\lambda_3 = 0.403429$) associated with the 3rd factor is high. It corresponds to 90.69% of the variance. For the second plot as shown in Fig. 8b, we can include the elbow into the selection i.e. we select $k = 3$ factors. Indeed, the eigenvalue ($\lambda_3 = 0.421673$) associated with the 3rd factor is high. It corresponds to 90.58% of the variance. For the third plot as shown in Fig. 8c, we can include the elbow into the selection i.e. we select $k = 6$ factors. Indeed, the eigenvalue ($\lambda_6 = 0.408255$) associated with the 6th factor is high. It corresponds to 91.86% of the variance. For the fourth plot as shown in Fig. 8d, we can include the elbow into the selection i.e. we select $k = 6$ factors. Indeed, the eigenvalue ($\lambda_6 = 0.548848$) associated with the 6th factor is high. It corresponds to 93.39% of the variance. For the fifth plot as shown in Fig. 8e, we can include the elbow into the selection i.e. we select $k = 6$ factors. Indeed, the eigenvalue ($\lambda_6 = 0.449910$) associated with the 6th factor is high. It corresponds to 94.25% of the variance. For the sixth plot as shown in Fig. 8f, we can include the elbow into the selection i.e. we select $k = 4$ factors. Indeed, the eigenvalue ($\lambda_4 = 0.410137$) associated with the 4th factor is high. It corresponds to 100% of the variance.

In Table 11, we have summarized the selected PCAs for all clusters. From the Table 11, it can be found that for Cluster 1 and Cluster 2 three PCs are selected as they provides significant

<table>
<thead>
<tr>
<th>Axis</th>
<th>Eigen value</th>
<th>Difference</th>
<th>Proportion (%)</th>
<th>Histogram</th>
<th>Cumulative (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.800359</td>
<td>1.799299</td>
<td>40.01 %</td>
<td></td>
<td>40.01 %</td>
</tr>
<tr>
<td>2</td>
<td>1.001060</td>
<td>0.117355</td>
<td>14.30 %</td>
<td></td>
<td>54.31 %</td>
</tr>
<tr>
<td>3</td>
<td>0.883702</td>
<td>0.096034</td>
<td>12.62 %</td>
<td></td>
<td>66.93 %</td>
</tr>
<tr>
<td>4</td>
<td>0.787668</td>
<td>0.112935</td>
<td>11.25 %</td>
<td></td>
<td>78.18 %</td>
</tr>
<tr>
<td>5</td>
<td>0.674733</td>
<td>0.224823</td>
<td>9.64 %</td>
<td></td>
<td>87.82 %</td>
</tr>
<tr>
<td>6</td>
<td>0.449910</td>
<td>0.047341</td>
<td>6.43 %</td>
<td></td>
<td>94.25 %</td>
</tr>
<tr>
<td>7</td>
<td>0.402569</td>
<td>-</td>
<td>5.75 %</td>
<td></td>
<td>100.00 %</td>
</tr>
</tbody>
</table>
The information of MFs for second cluster.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Type</th>
<th>Linguistic values and ranges of MFs</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC1</td>
<td>Gaussian</td>
<td>[0.3693 – 0.08725]</td>
</tr>
<tr>
<td>PC2</td>
<td>Gaussian</td>
<td>[0.3136 – 2.437]</td>
</tr>
<tr>
<td>PC3</td>
<td>Gaussian</td>
<td>[0.3693 – 0.08725]</td>
</tr>
</tbody>
</table>

The information of MFs for first cluster.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Type</th>
<th>Linguistic values and ranges of MFs</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC1</td>
<td>Gaussian</td>
<td>[1.2 – 3.206]</td>
</tr>
<tr>
<td>PC2</td>
<td>Gaussian</td>
<td>[0.3525 0.9216]</td>
</tr>
<tr>
<td>PC3</td>
<td>Gaussian</td>
<td>[0.5473 – 1.65]</td>
</tr>
</tbody>
</table>
The training of the proposed network. A value closer to 1 stands for the success of learning. These estimators are determined by Eqs. (19) and (20).

$$\text{MSE} = \frac{\sum_{i=1}^{n}(\text{actual}(O) - \text{prediction}(O))^2}{n} \tag{21}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\text{actual}(O) - \text{prediction}(O))^2}{\sum_{i=1}^{n}(\text{actual}(O) - \text{actual}(O))^2} \tag{22}$$

where actual (O) indicates the real overall rating provided by user, prediction (O) implies the predicted overall rating value and n corresponds to the number of used user ratings.

For error estimation in the clusters provided by EM, it can be concluded that prediction errors for PCA–ANFIS models of EM clusters are significantly low with high values of coefficient of determination.

Comparison of performance for overall rating prediction of both Multiple Linear Regression (MLR) and ANFIS on experimental data set show that the proposed PCA–ANFIS method is more accurate than the MLR in overall rating prediction. For ANFIS models, we selected the best configurations in terms of MFs type, type of trainings and number of training. The Gaussian MF type shows the best performance in relation to the Triangular one. Note that as the overall ratings ranges are [0 5], the prediction models of PCA–ANFIS show that the proposed PCA–ANFIS method is more accurate than the MLR in overall rating prediction. For ANFIS models, we selected the best configurations in terms of MFs type, type of trainings and number of training. The Gaussian MF type shows the best performance in relation to the Triangular one. Note that as the overall ratings ranges are [0 5], the prediction models of PCA–ANFIS also obtained this range for the overall ratings. However, the prediction range for MLR was not always between [0 5], in many cases it was above than this range.

In this paper, we selected hybrid learning (training) algorithm in ANFIS. This learning algorithm combines the least squares estimator and the gradient descent method. Using the hybrid method, the ANFIS models generated rules by enumerating all possible rules generated by the hybrid learning method.

For error estimation in the clusters provided by EM, it can be concluded that prediction errors for PCA–ANFIS models of EM clusters are significantly low with high values of coefficient of determination.
combinations of MFs of all original inputs and PCs generated by PCA. Compared with the ANFIS for overall rating prediction, the models that used ANFIS with incorporating PCA had lower computation time in all models. In addition, the computation time for ANFIS is moderately large when the number of inputs (curse of dimensionality) (Brown et al. 1995). This can be a main disadvantage of using solely ANFIS for the problem of overall rating prediction in multi-criteria CF. Hence, this problem connected to the ANFIS has been eliminated with incorporating the PCA before applying ANFIS. This incorporation of PCA caused the reduction in number of inputs and accordingly hidden layers, number of MFs and rules. Evidently, the training time of prediction models was significantly reduced also as the computational overhead associated with the PCA algorithm is negligible. Note that all the prediction models by ANFIS have been developed in the offline phase and when new ratings were added to the dataset, the models needed to be retrained. An opportunity for future work is therefore to add newly arriving data points to the clusters and train the prediction models incrementally by ANFIS. The advantage of having incremental prediction models is that when incremental updates are supported, the scalability of multi-criteria CF will be improved.

4.5. Evaluating the hybrid multi-criteria CF

To evaluate the accuracy of the proposed method, we conducted a set of experiments. We determine the precision and recall of the Top-N list for recommender system. We relied on the typical evaluation protocol and accuracy measures described, e.g., in (Shani and Gunawardana 2011), to determine the quality of the recommendations and to compare our method with previous works. Specifically, we considered the data into training and test sets and tried to predict the rating or ranking of the hidden items in the test dataset. To factor out random effects, we apply a fivefold cross-validation procedure. For these measures we split the data into 80% training and 20% test data, used random subsampling and repeated the experiments appropriately to factor out effects of randomness.
The recommenders’ prediction accuracy is measured by root mean squared error (RMSE), which is a widely used metric for evaluating the statistical accuracy of recommendation algorithms, given by

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{n}}
\]  

(23)

where \(x_i\) is the rating values predicted by the recommender from the true rating \(x_i\). A lower value of RMSE indicates a higher accuracy of the recommendation system.

The accuracy result measured by RMSE is shown in Table 19. The results are based on fivefold cross-validation and 80% data in the training set with 5 test trials. The results show that the combination of the clustering and the PCA technique helps us to measurably decrease the prediction error. The coverage of methods that use Clustering with ANFIS and Clustering with PCA and ANFIS approaches was 100% compared to ANFIS and HOSVD (Nilashi et al. 2014b) with 100% coverage, Standard CF (Adomavicius and Kwon 2007) with 41% coverage, Total-Reg (Adomavicius and Kwon 2007) with 53% coverage, PCA with 74% coverage, and Clustering and Multiple Linear Regression with 85% coverage. This means that predictions for all user-item pairs in the test set could be made by the proposed method.

From Table 19, it can be seen that proposed model has outperformed other recommendation methods and obtained a better RMSE error. It should be noted that a two-tailed paired t-test has been performed and based on the results the differences between proposed model (PCA–ANFIS) and the compared methods were statistically significant \((p < 0.01)\) for the experimental dataset.

We also employ the recall and precision metrics, which are widely used in recommender systems to evaluate the quality of recommendations. Recall indicates the ability of a system to present all relevant items (Sarwar et al. 2000). In reality, it may not be possible to retrieve all the relevant items from a collection, especially when the collection is large. A system may be able to retrieve a proportion of the total relevant items. Thus, the performance of a system is often measured by the recall ratio, which

Table 18
MSE and \(R^2\) for 6 clusters in PCA–ANFIS modeling.

<table>
<thead>
<tr>
<th>Accuracy Measure</th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
<th>Cluster 5</th>
<th>Cluster 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>0.046525</td>
<td>0.04853</td>
<td>0.052925</td>
<td>0.0552</td>
<td>0.052153</td>
<td>0.045285</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.97460</td>
<td>0.96850</td>
<td>0.95260</td>
<td>0.94345</td>
<td>0.94345</td>
<td>0.971171</td>
</tr>
</tbody>
</table>

Table 19
RMSE for proposed method.

<table>
<thead>
<tr>
<th>Method</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard CF (Adomavicius and Kwon 2007)</td>
<td>0.672</td>
</tr>
<tr>
<td>Total-Reg (Adomavicius and Kwon 2007)</td>
<td>0.653</td>
</tr>
<tr>
<td>PCA</td>
<td>0.618</td>
</tr>
<tr>
<td>Clustering and Multiple Linear Regression</td>
<td>0.608</td>
</tr>
<tr>
<td>Clustering and ANFIS</td>
<td>0.513</td>
</tr>
<tr>
<td>ANFIS and HOSVD (Nilashi et al. 2014b)</td>
<td>0.489</td>
</tr>
<tr>
<td>Clustering and PCA–ANFIS</td>
<td>0.454</td>
</tr>
</tbody>
</table>
denotes the percentage of the relevant items retrieved in a given situation. Precision implies the ability of a system to present only the relevant items. This relates to its ability to not retrieve non-relevant item (Sarwar et al. 2000). This factor, that is how far the system is able to withhold unwanted items in a given situation, is measured in terms of precision ratio. Table 20 shows Contingency table for computing precision and recall. These precision and recall measures are presented by Eq. (24) and Eq. (25), respectively.

\[
\text{Precision(Reclist)} = \frac{\{\text{relevant items}\} \cap \{\text{top - Nitems}\}}{\{\text{top - Nitems}\}} = \frac{N_{SR}}{N_{ST}} \tag{24}
\]

\[
\text{Recall(Reclist)} = \frac{\{\text{relevant items}\} \cap \{\text{top - Nitems}\}}{\{\text{relevant items}\}} = \frac{N_{SR}}{N_{TR}} \tag{25}
\]

F-measure is a metric defined as the harmonic mean of the precision and recall, is also widely used to evaluate the quality of recommendations. We used the F1-metric in our evaluation, as shown in Eq. (25).

\[
F_1 = \frac{(1 + \beta^2) \cdot \text{precision} \cdot \text{recall}}{\beta^2 \cdot \text{precision} + \text{recall}} \tag{26}
\]

where parameter $\beta \in [0, 1]$ determines the relative influence of both metrics (the value $\beta = 1$ is commonly used).

In order to compare the proposed method with previous work, we evaluated our approach on YM-10-10 using MAE. The MAE is determined as the average absolute deviation between predicted ratings and true ratings that showed in Eq. (27).

\[
\text{MAE} (\text{pred. act}) = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{\text{pred}_{ui} - \text{act}_{ui}}{N} \right| \tag{27}
\]

Suppose $N$ is the number of items that user $u$ has expressed an opinion.

To experimentally show the effectiveness of clustering and PCA-ANFIS, we perform the experiments on the TripAdvisor dataset. The aim is to calculate the recommendation precision and MAE of prediction of proposed method (see Table 21). We compared the proposed method with the Standard CF and Total-Reg developed by Adomavicius and Kwon (2007), matrix factorization method using PCA (Yin and Peng 2012), and tensor factorization method (Nilashi et al. 2014b).

From the results presented in Table 21, it can be found that the Precision@5 (recommended items for a user of size 5) and Precision@7 (recommended items for a user of size 7) of the proposed method is higher than recommendation method which only uses ANFIS. According to the experiments results, the proposed hybrid method also provides better prediction accuracy with lower MAE in relation to the Standard CF and Total-Reg developed by Adomavicius and Kwon (2007). These show the effectiveness of incorporating the clustering and PCA approaches regarding the prediction accuracy of multi-criteria CF in the Tourism Domain. In addition, the results in Table 21 indicates that, compared to Standard CF and Total-Reg, our clustering and noise removal techniques help to improve the prediction accuracy by more than 20% in all tested scenarios in term of the Precision@5 and Precision@7. Furthermore, the results show that compared to the matrix factorization method using PCA, our clustering and noise removal techniques help to improve the prediction accuracy by more than 15% in all tested scenarios in term of the Precision@5 and Precision@7. It can also be observed that the matrix factorization using PCA works relatively well and is better than methods such as Standard CF and Total-Reg. This supports the findings of (Jannach et al. 2012a) with respect to the accuracy as their method based on regression approach outperformed the methods that used matrix factorization and methods developed by Adomavicius and Kwon (2007). Moreover, compared with the methods that use Clustering and Multiple Linear Regression, the proposed method improves the precision by more than 8%. Again, we can see that the clustering and PCA techniques help to increase precision and at the same time reduce the MAE compared to method that uses Clustering and ANFIS.

Over method that uses EM clustering and ANFIS, the prediction accuracy improvement is more than 6% in term of the MAE. In addition, compared with the method which uses ANFIS and Higher Order Singular Value Decomposition (HOSVD), the accuracy improvement was more than 2% in all tested scenarios in term of the Precision@5 and Precision@7 and at the same time reduced the MAE about 3%. This supports the findings of (Nilashi et al. 2014b) with respect to the accuracy as their method based on ANFIS technique outperformed the recommendation methods that only used tensor and matrix factorization approaches and recommendation methods developed by Adomavicius and Kwon (2007). In addition, it can be supported by Jannach et al. (2012a) work as they evaluated the improvements of their proposed method on Yahoo!Movies and tourism domain datasets. Their experiments showed that the suggested improvements not only lead to better results than those achieved with the techniques presented in Adomavicius and Kwon (2007), but also that the predictions are more accurate than more recent single-rating approaches based on matrix factorization.

In term of F1, the recommendation accuracy performance of the multi-criteria CF using PCA, ANFIS and clustering approach is presented in Fig. 12. For this purpose, in the experiments, we varied the number of neighbors and computed the corresponding $F_1$ ($\beta = 1$) value for proposed recommendation method. We ran the

### Table 20

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision@5</th>
<th>Precision@7</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard CF (Adomavicius and Kwon 2007)</td>
<td>62.20</td>
<td>61.45</td>
<td>1.37</td>
</tr>
<tr>
<td>Total-Reg (Adomavicius and Kwon 2007)</td>
<td>65.10</td>
<td>63.31</td>
<td>1.28</td>
</tr>
<tr>
<td>PCA</td>
<td>73.14</td>
<td>71.13</td>
<td>1.16</td>
</tr>
<tr>
<td>Clustering and Multiple Linear Regression</td>
<td>78.36</td>
<td>76.16</td>
<td>1.11</td>
</tr>
<tr>
<td>Clustering and ANFIS</td>
<td>82.45</td>
<td>81.21</td>
<td>0.92</td>
</tr>
<tr>
<td>ANFIS and HOSVD (Nilashi et al. 2014b)</td>
<td>84.86</td>
<td>83.21</td>
<td>0.89</td>
</tr>
<tr>
<td>Clustering and PCA–ANFIS</td>
<td>86.21</td>
<td>85.71</td>
<td>0.86</td>
</tr>
</tbody>
</table>

experiments on TripAdvisor dataset for $N$ equal 1, 20, 40, 60, 80, 100 and 120, where $N$ is the number of items to be recommended by the Top-$N$ recommender systems. From all F1 curves in Fig. 12, we can notice that the proposed method gives high level of accuracy when the size of neighbors is increased versus the Top-$N$ recommendation. In this figure, it can be seen that while returning Top-1 to Top-120 recommendation, the proposed method can achieve an improvement above 0.81 for all neighborhood sizes. In addition, it can be concluded that the optimal neighbor size can be obtained by considering the maximum value of F1. Thus, the neighborhood size 60 can be chosen as the optimal value in producing the best performance of the proposed method. This outcome demonstrates the significance of combining PCA–ANFIS method with EM clustering for enhancing the accuracy of multi-criteria CF in the tourism domain.

5. Conclusion and future work

In this paper, a new approach, called PCA–ANFIS, was proposed to increase the predictive accuracy and efficiency of the multi-criteria CF. The proposed method was developed for tourism domain using EM clustering, PCA and ANFIS. We selected ANFIS approach to learn the prediction models for users and items separately in each cluster and then combined two predictions in the online phase. In addition, PCA was used for dimensionality reduction and to address multi-collinearity and interdependency problems existing between criteria in the multi-criteria dataset.

We analyzed the predictive accuracy of proposed method in the domain of hotel recommendation on a real-world dataset provided by TripAdvisor. The proposed method was evaluated using RMSE, MAE, F1-measure, Precision@5 and Precision@7 using precision metric. The experimental results on TripAdvisor dataset clearly demonstrated the capability of ANFIS modeling using MFs and fuzzy rules without the human expert intervention in multi-criteria CF in the tourism domain. Our experiments confirmed that the combination of PCA–ANFIS with clustering as a hybrid method significantly leads to the improvement in predictive accuracy of tourism multi-criteria CF measured by standard accuracy metrics.

In the proposed method, ANFIS models of items and users are updated offline; however, in the multi-criteria CF recommenders, data is dramatically updated and therefore incremental learning approaches are needed to consider new ratings. Thus, future studies will focus on further improvement of the multi-criteria CF recommendation accuracy for tourism domain by incorporating fuzzy semantic technique. In addition, we will focus on developing an incremental Neuro-Fuzzy learning approach as newly arriving ratings can be crucial for the success of a practical recommendation system and have an immediate impact on the accuracy of the predictions. Recommendations for future study in this area are as follows:

- Although multi-criteria ratings can be a good choice for pure CF recommendation, the accuracy of multi-criteria CF can be improved more with incorporating other resources such as tags and content of users and items to the tensor of ratings. With this incorporation, the fuzzy semantic techniques can be applied to better alleviate sparsity problem and enhance the multi-criteria CF recommendation accuracy.
- The proposed multi-criteria recommendation model can be extended to the incremental based recommendation. Therefore, future studies will focus on further improvement of the multi-criteria CF recommendation accuracy and efficiency by incorporating incremental techniques using incremental ANFIS.

Acknowledgements

The authors would like to thank Research Management Centre (RMC) at Universiti Teknologi Malaysia (UTM) for providing adequate facilities to conduct this research and for their financial support of this research through the Post Doctorate Research Grant with Vote No. 02E50. Appreciation also goes to the editor and anonymous reviewers for their valuable comments and suggestions, which were helpful in improving the paper.

References


Fig. 12. Top-N and F-measure for different neighborhood size.


World Travel & Tourism Council. Travel & Tourism Economic Impact, 2009.


