

Developing an intelligent trip recommender system by data mining methods

Tamer Uçar *, Software Eng. Dept., Bahçeşehir University, Istanbul 34353, Turkey

Adem Karahoca, Software Eng. Dept., Bahçeşehir University, Istanbul 34353, Turkey

Suggested Citation:

Uçar, T. & Karahoca, A. (2015). Developing an intelligent trip recommender system by data mining methods. *Global Journal of Information Technology*. 6(1), 119-127.

Received 13 January, 2016; revised 20 February, 2016; accepted 07 March, 2016.

Selection and peer review under responsibility of Prof. Dr. Adem Karahoca, Bahcesehir University, Turkey

© 2016 SciencePark Research, Organization & Counseling. All rights reserved.

Abstract

Internet has a very wide usage in almost every sector. People are continuously looking and searching for information through internet. Narrowing down relevant search results is not a very simple task. Recommender systems are being used in almost every search related area. Tourism domain is one of these sectors. This study proposes an implementation of an expert system framework which can accurately classify users and make predictions about user classifications for recommending tourism related services. Proposed approach predicts clusters for system users and according to these user clusters, trips, hotels and such services can be recommended individually or as a campaign to target user or user groups.

Keywords: Trip recommender, data mining, expert systems

1. Introduction

Internet has a very wide usage in almost every sector. People are continuously looking and searching for information through internet. Narrowing down relevant search results is not a very simple task. Recommender systems (RS) are gaining popularity for filtering and providing relevant information about a person's search on a specific topic. A recommender system tries to predict a rating value of an item for a target user. To perform an accurate prediction, such systems use their member profiles and member behaviors on the system [1].

Recommender systems are being used in almost every search related area. Tourism domain is one of these sectors. Most of the recommender system applications in tourism domain involve proposing travel destinations, trip and activity recommendations and hotel suggestions in a destination within a given set of user defined constraints. Users can define budget limits, time intervals, interests, desired locations or similar necessities. After retrieving such data, a recommender system analyzes user input and proposes a relevant output [2]. To generate accurate predictions, many approaches are tested by different researchers. Generally, most of these approaches are based on acquiring a set of parameters which can be used as constraints for the recommendation system.

This study proposes an implementation of an expert system framework which can accurately classify users and make predictions about user classifications. Proposed approach predicts clusters for system users and according to these user clusters, trips, hotels and such services can be recommended individually or as a campaign to target user or user groups.

There are many recommender systems available on tourism domain. Main purpose of our proposed approach is increasing classification accuracy. The following steps summarizes details of the proposed approach: (1) Discretizing initial travel data set. (2) Discovering user clusters for prediction model. (3) Training data mining model for prediction of user classes. (4) Predicting class for a target user. (5) Recommending services suitable for predicted user class.

The remainder of this paper is organized as follows: Section 2 includes reviews of recent studies about travel recommender systems, Section 3 describes materials and methods used in the proposed approach, Section 4 presents comparison results of the proposed approach and Section 5 contains the conclusion and summary of this study.

2. Related Works

Many studies were made about recommender systems. This section mostly focuses on previously implemented recommender systems on tourism domain in a chronological order.

In 2008, authors developed a mobile recommendation system which generates a user profile by considering users which are having similar interests on items. System tries to determine a list of activities for the target user and based on these activities, it generates trip plans. Proposed approach contains an ontology model. Recommendations were generated by the past experience of the system with similar users [3].

In other study, researchers developed a system which generates personalized recommendations of touristic attractions. Proposed system integrates heterogeneous online travel information using a tourism ontology. Travel behavior of the target user and similar users were analyzed to generate recommendations. Bayesian network technique and the analytic hierarchy process method used for recommendation engine [4].

An expert travel agent was developed for assisting tourists by suggesting package holidays and tours. The proposed method employs a hybrid approach containing both content-based and collaborative filtering methods. Demographic data was also used in recommendation system. Authors emphasized that the choice of this hybrid approach was made to cover shortcomings of each of the individual recommendation methods [5].

In another research, a semantic hotel recommender system was developed. To generate recommendations, hotel ontology was combined with a fuzzy logic approach. To involve customer experience, system contains a feedback mechanism which allows users to rate the generated recommendations. In order to generate more accurate recommendations, these ratings were used for updating fuzzy rules [6].

Another recommender engine study covers trip recommendations for both individuals and tourist groups. Group recommendation mechanism aggregates and intersects individual recommendations which were made for every member in a given group. Recommendation engine employs both demographic and content-based filtering methods [7].

In 2012, a decision support system for tourist attractions was implemented by combining Engel–Blackwell–Miniard model and Bayesian network approaches. Data which was published by the Tourism Bureau of Taiwan was used while building the proposed recommendation system. Generated recommendations were displayed on Google Maps to provide more detailed information for tourists [8].

A hybrid recommendation system which contained both content-based and collaborative filtering methods was built to propose better personalization for recommender systems in tourism. The proposed system was implemented using association based classification approach. Concepts form association and classification were combined to involve association rules in a prediction context [9].

Another mobile tourism recommendation system was implemented using a location based collaborative filtering method. The proposed system generates recommendations by considering other tourists' ratings on their visited attractions. Users exchange their rating through a mobile peer to peer connection. Three data exchange methods were proposed for effectively exchanging ratings about visited attractions [10].

Social media related recommender systems are also gaining popularity. From this perspective, [11] introduced a recommendation system which analyzes geo-tagged social media to recommend landmarks for customized travel planning. System obtains trip's spatial and temporal properties and using these properties, it computes the significance of landmarks. Specific landmark clusters are generated for similar themes and these clusters are recommended to system users [11].

A different research group proposed a travel schedule planning algorithm which generates customized recommendations based on user requirements. With a user-adapted interface, users can make changes on recommendation results and the provided feedback mechanism improves system's accuracy for later recommendations [12].

Another study which focused on social media based recommendation includes a method for city travel recommendation system. Researchers applied principals from both content based and collaborative filtering techniques. User preferences were mined from community-contributed geotagged photos archive. User similarities were taken into account for improving accuracy of the proposed model [13].

Table 1 lists all of the mentioned studies by their publication years, titles, authors and methods.

Table 1. Publications by years

Year	Title	Authors	Methods
2008	SAMAP: An user-oriented adaptive system for planning tourist visits	Castillo et al.	Multi-agent model with Ontology
2009	A Bayesian network and analytic hierarchy process based personalized recommendations for tourist attractions over the Internet	Yuxia Huang, Ling Bian	Bayesian network, analytic hierarchy process, Attraction Ontology
2009	Building an expert travel agent as a software agent	Silvia Schiaffino, Analia Amandi	Collaborative filtering with content-based recommendations
2011	Sem-Fit: A semantic based expert system to provide recommendations in the tourism domain	Crespo et al.	Fuzzy Engine, Hotel Ontology
2011	On the design of individual and group recommender systems for tourism	Garcia et al.	Demographic and content-based filtering
2012	Design and implementation of an intelligent recommendation system for tourist attractions: The integration of EBM model, Bayesian network and Google Maps	Fang-Ming et al.	Engel-Blackwell-Miniard (EBM) model with Bayesian network
2013	A hybrid recommendation approach for a tourism system	Lucas et al.	Collaborative filtering with content-based approach
2013	iTravel: A recommender system in mobile peer-to-peer environment	Wan-Shiou Yang, San-Yih Hwang	Location based collaborative filtering
2014	Adaptive landmark recommendations for travel planning: Personalizing and clustering landmarks using geo-tagged social media	Jonghyun Han, Hyunju Lee	Content-based filtering
2015	User-adapted travel planning system for personalized Schedule recommendation	Hsiu-Sen Chiang, Tien-Chi Huang	Content-based filtering
2015	Topic based context-aware travel recommendation method exploiting geotagged photos	Zhenxing Xu, Ling Chen, Gencai Chen	Collaborative filtering with content-based approach

3. Material and Methods

This section provides details about data gathering and pre-processing steps. Algorithms of the methods which were used to process the data set are also explained.

3.1. Data gathering and preprocessing

Initial data set of customer flights and details was obtained from an existing travel platform's database. Since the travel platform is working with different data sources, retrieved data was extracted from a nested XML structure. To convert nested XML structure to tabular format, XPath and T-SQL expressions were used. After converting data to tabular format, all of the identity columns were removed from data set. As a result of this data processing, 12 attributes were retrieved. Table 2 lists these attributes.

Table 2. Data set attributes

Attribute	Description
-----------	-------------

Birthdate	Passenger's birthdate.
Gender	Passenger's gender.
DepartureDate	Departure date of trip.
DepartureCity	The city in which the passenger is leaving from.
ArrivalCity	The city in which the passenger is arriving to.
DepartureAirline	Airline company for departure flight.
DepartureAirlineClass	Flight class for departure flight.
ReturningDepartureCity	The city in which the passenger is returning from.
ReturningArrivalCity	The city in which the passenger is returning to.
ReturningAirline	Airline company for returning flight.
ReturningAirlineClass	Flight class for returning flight.
ReturnDate	Returning date of trip.

Records with missing attributes and duplicate entries were also removed from data set. And a total of 3213 records were obtained for processing. To be able use the data in Xmeans and ANFIS algorithms, nominal data was converted to numeric data and normalized into same range (between 0 and 1).

Final data set was used to discover clusters. After obtaining clusters for each record, 66% of data was used for training and the remaining 34% was used for testing the prediction models.

3.2. Used methods

This section contains brief information about the methods which were used in this study.

3.2.1. Xmeans clustering

K-means clustering algorithm is a simple but popular approach for finding clusters in a given data set. But there are some important shortcomings for this method such as the necessity of providing the number of clusters and random located initial cluster centers. [14] proposed Xmeans clustering method to overcome these drawbacks. It works as extending K-means with efficient estimation of the number of clusters. The algorithm searches the space of cluster locations and number of clusters to optimize the Bayesian Information Criterion (BIC) or the Akaike Information Criterion (AIC) measure [14].

3.2.2. Fuzzy c-means clustering

In fuzzy c-means clustering, every point has a degree of belonging to clusters. This causes some points on the edge of a cluster to be shared by other clusters. For any point x , there is a set of coefficients which gives the degree of being a member for a given cluster. The weighted means of all points in a cluster is the centroid of that cluster. The degree of belonging is related inversely to the distance from x to the cluster center. It also depends how much weight is given to the closest center [15, 16].

3.2.3. Adaptive neuro fuzzy inference system (ANFIS)

ANFIS is a neural-fuzzy system which contains both neural networks and fuzzy systems. A fuzzy-logic system can be described as a non-linear mapping from the input space to the output space. This mapping is done by converting the inputs from numerical domain to fuzzy domain. To convert the inputs, firstly, fuzzy sets and fuzzifiers are used. After that process, fuzzy rules

and fuzzy inference engine is applied to fuzzy domain. The obtained result is then transformed back to arithmetical domain by using defuzzifiers. Gaussian functions are used for fuzzy sets, and linear functions are used for rule outputs on ANFIS method. The standard deviation, mean of the membership functions and the coefficients of the output linear functions are used as network parameters of the system. The summation of outputs is calculated at the last node of the system. The last node is the rightmost node of a network. In Sugeno fuzzy model, fuzzy if-then rules are used [17, 18, 19].

3.2.4. Radial basis function networks (RBFN)

Radial basis function network (RBFN) is a neural network model which uses radial basis functions (RBF) as activation functions of the system. System generates an output as a linear combination of neuron parameters and inputs' radial basis functions. These networks contain three layers. The first layer is an input layer, the next layer is a hidden layer which contains a non-linear RBF activation function and the last layer is a linear output layer. The input of the system can be a vector of real numbers [20].

3.2.5. Bayesian network

Bayesian Networks produce probability estimates as network output like logistic regression models. Prediction is not produced by the system itself. The main purpose of the system is estimating the probability of an instance for each class value whether that value suits for a class or not. If we compare plain predictions with probability estimates, we see that probability estimates are more useful than plain predictions because we can rank the predictions with probability estimates. In Bayesian Networks, the conditional probability distribution of the value of a given class attribute is predicted within other class attribute [20].

3.2.6. ID3

ID3 is an algorithm which is used for generating decision trees. Non-categorical attributes are represented by nodes. And each arc corresponds to a possible value of an attribute. A leaf of the tree specifies the expected value of the categorical attribute for the records. These records are described by the path from the root to that leaf. For measuring node informativeness, entropy is used. Each node should be associated with the most informative non-categorical attribute which is not yet considered in the path from the root [20].

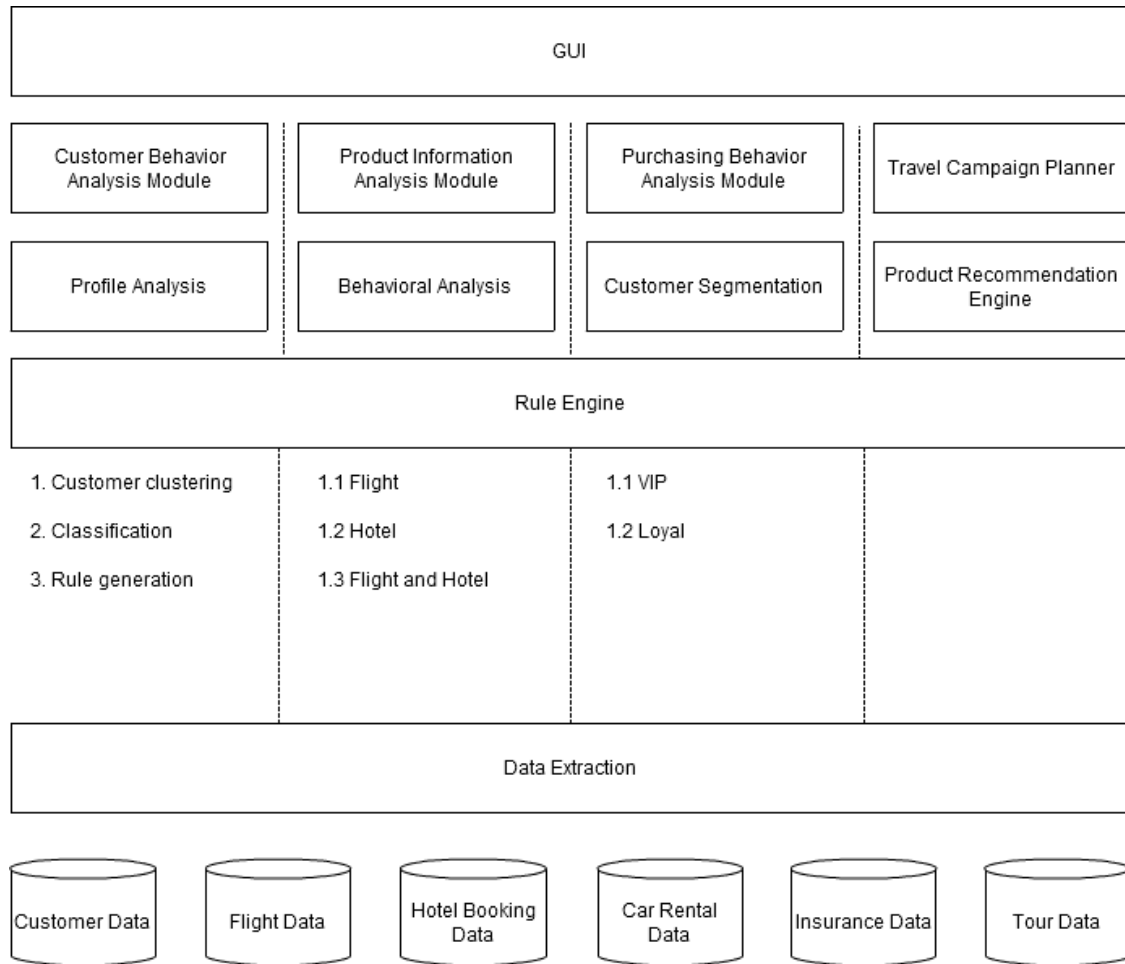


Figure 1. Recommender system architecture

4. Results and Discussions

Both Xmeans and Fuzzy C-Means clustering algorithms were applied on the same data set separately. After obtaining clusters by using these algorithms, ANFIS, Radial Basis Function Networks, ID3 and Bayesian Network algorithms were processed for each clustered data set.

Table 3 lists root mean squared error (RMSE), sensitivity, specificity, precision and correctness values for each of the methods mentioned above.

Method Name	Clustering Algorithm	RMSE	Sensitivity	Specificity	Precision	Correctness
ANFIS	Fuzzy c-means	0.04	0.97	0.99	0.97	0.99
ANFIS	Xmeans	0.07	0.88	0.98	0.85	0.97
RBFN	Fuzzy c-means	0.11	0.94	0.99	0.94	0.94
RBFN	Xmeans	0.13	0.91	0.99	0.93	0.93
Bayesian Net.	Fuzzy c-means	0.08	0.96	0.99	0.96	0.96
Bayesian Net.	Xmeans	0.11	0.92	0.99	0.92	0.93
ID3	Fuzzy c-means	0.06	0.98	0.99	0.99	0.92
ID3	Xmeans	0.08	0.97	0.99	0.97	0.89

According to the given results in Table 3, lowest RMSE and highest correctness values are obtained by combining ANFIS and Fuzzy C-means clustering algorithms whereas highest sensitivity value is obtained by combining ID3 and Fuzzy C-means clustering algorithms. If we compare Fuzzy C-means and Xmeans clustering algorithms, Fuzzy C-means algorithm generates better results which shows that the accuracy of the prediction model will be higher.

The obtained results indicate that combining Fuzzy C-means clustering algorithm with ANFIS method will be a suitable approach for implementing an expert system which can be used for classifying and proposing flight destinations for users.

According to the system output, a recommender engine can propose hotels and related services for a target user. Figure 1 shows the architecture of such a recommender system. As it is stated in the diagram, data will be extracted from various databases. Customer clustering, classification and rule generation steps will be performed using the extracted data set. And product recommendation engine can be used for proposing products or planning specific campaigns for desired user groups.

5. Conclusion

In this study, various algorithms were compared for clustering and classifying customer flight data set. And an architecture for an expert system which can be used for generating recommendations or planning specific campaigns for desired user groups was proposed. As it is stated in the previous section, best clustering and classification results are obtained by combining ANFIS and Fuzzy c-means algorithms. An expert system which can produce efficient recommendations can be implemented by using ANFIS and Fuzzy c-means algorithms together.

References

- [1] Adomavicius, G., Sankaranarayanan, R., Sen, S. & Tuzhilin, A. (2005). Incorporating contextual information in recommender systems using a multidimensional approach, *ACM Transactions on Information Systems*, 23(1), pp 103-145.
- [2] Ricci, F. Travel Recommender Systems, *IEEE Intelligent Systems*, 2002, pp 55-57.
- [3] Castillo, L., Armengol, E., Onaindía, E., Sebastiá, L., Boticario, J. G., Rodríguez, A., Fernández, S., Arias, J. D. & Borrajo, D. (2008). SAMAP: An user-oriented adaptive system for planning tourist visits, *Expert Systems with Applications*, 34(2), pp 1318-1332.
- [4] Huang, Y. & Bian, L. (2009). A Bayesian network and analytic hierarchy process based personalized recommendations for tourist attractions over the Internet, *Expert Systems with Applications*, 36(1), 933-943.
- [5] Schiaffino, S. & Amandi, A. (2009). Building an expert travel agent as a software agent, *Expert Systems with Applications*, 36(2), 1291-1299.
- [6] Crespo, Á. G., Cuadrado, J. L. L., Palacios, R. C., Carrasco, I. G. & Mezcua, B. R. (2011). Sem-Fit: A semantic based expert system to provide recommendations in the tourism domain, *Expert Systems with Applications*, 38(10), 13310-13319.
- [7] Garcia, I., Sebastia, L. & Onaindia, E. (2011). On the design of individual and group recommender systems for tourism, *Expert Systems with Applications*, 38(6), 7683-7692.
- [8] Hsua, F.-M., Lina, Y.-T. & Hoc, T.-K. (2012). Design and implementation of an intelligent recommendation system for tourist attractions: The integration of EBM model, Bayesian network and Google Maps, *Expert Systems with Applications*, 39(3), 3257-3264.
- [9] Lucas, J. P., Luz, N., Moreno, M. N., Anacleto, R. & Figueiredo, A. A. (2013). A hybrid recommendation approach for a tourism system, *Expert Systems with Applications*, 40(9), 3532-3550.
- [10] Yang, W.-S. & Hwang, S.-Y. (2013). iTravel: A recommender system in mobile peer-to-peer environment, *Journal of Systems and Software*, 86(1), pp 12-20.

- [11]Han, J. & Lee, H. Adaptive landmark recommendations for travel planning: Personalizing and clustering landmarks using geo-tagged social media, *Pervasive and Mobile Computing*, Available online 13 August 2014, doi:10.1016/j.pmcj.2014.08.002.
- [12]Chiang, H.-S. & Huang, T.-C. (2015). User-adapted travel planning system for personalized schedule recommendation, *Information Fusion*, 21, pp 3-17.
- [13]Xu, Z., Chen, L. & Chen, G. Topic based context-aware travel recommendation method exploiting geotagged photos, *Neurocomputing*, Available online 2 January 2015, doi:10.1016/j.neucom.2014.12.043
- [14]Pelleg, D., & Moore, A. (2000). X-means: Extending K-means with efficient estimation of the number of clusters. In *Proceedings of the 17th international conference on machine learning* (pp. 727–734). Morgan Kaufmann.
- [15]Dunn, J. C. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3, 32–57.
- [16]Bezdek, J. C. (1981). *Pattern recognition with fuzzy objective function algorithms*. New York: Plenum Press.
- [17]Takagi T. & Sugeno, M. (1985). Fuzzy identification of systems and its application to modelling and control. *IEEE Trans. On Systems, Man & Cybernetics*, 15(1), 116-132
- [18]Sugeno M. & Kang G. T. (1988). Structure identification of fuzzy model. *Fuzzy Sets and Systems* 28(1),15-33
- [19]Jang, J.S. (1992). Self-learning fuzzy controllers based on temporal back propagation. *IEEE Trans Neural Networks*, 3(5), 714-723
- [20]Witten, I.H. & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann Publishers, San Francisco.