Comparison of Supervised and Unsupervised Learning Classifiers for Travel Recommendations

Zohreh Bahman Isfahani[1*], Shahram Jafari[2] and Reza Akbarian[3]

[1]Faculty of E-Learning, Shiraz University, Iran
E-mail: zh_1476@yahoo.com
[2]School of Electrical and Computer Engineering, Shiraz University, Iran
E-mail: jafaris@shirazu.ac.ir
[3]AssociateProfessor of Shiraz University, Iran
E-mail: rakbarian@rose.shirazu.ac.ir

Abstract: Classification is one of the most common activities in the related fields of intelligent decision making. Neural networks are suitable approaches for solving data mining problems especially classification. Usually for solving classification problem using neural network that proper outputs are existing for them, the supervised training type is selected. In this study a comparison about classification of electronic tourism data was done by using two learning method, supervised and unsupervised and the proper output values were determined for all the input data. The output includes some travel packages appropriate for the tourists recommended to them according to the input values. The experimental result showed that despite the target output values are exist ,the neural network output with unsupervised learning and SOM architecture has more precise prediction as compared to supervised learning .The neural network proposed travel packages is in more conformation with the tourist selections for final evaluation of the results the test dataset were given to the experts and their care in predictions indicated close results to the obtained evaluations from the unsupervised learning method.

Key Words: classification, intelligent tourism, recommender model, neural network, self-organizing maps, multilayer feed-forward, supervised learning, unsupervised learning

INTRODUCTION

Tourism is one of the most successful and dynamic industries in the world. Most tourism companies have elaborate websites to present and sell their products. The tourism industry is an information intensive business, and the amount of information is increasing rapidly; however efficient access to this information is becoming a challenge. In recent years, with many countries turning to tourism to their economies, there has been a massive expansion of tourism vendor offerings. As more travel arrangements are made online, pressure is put on e-Tourism website developers to provide efficient and easy to use interfaces and intelligent services [1]. The e-tourism information service features wide-data and diverse types, such as travel agency, hotels, and tourist route as well as scenic spots. Facing with so wide tourism resources, tourists need to spend a lot of time in searching for information. On one hand, the tourist is faced with the obviously endless group of existing options, and on the other, the heterogeneity of the places to visit. Besides, tourists are not willing to spend too much time in searching online. Therefore, users need an intelligent online assistant. Artificial Neural Network is a system loosely modeled based on the human brain. It has ability to account for any functional dependency. The network discovers (learns, models) the nature of the dependency without needing to be prompted. Initially the application of data mining is not being used because of complex structure, long training time, and poor interoperability. But as neural networks is a powerful technique to solve many real world problems[2]. They have the ability to learn from experience in order to improve their performance and to adapt themselves to

changes in the environment. In addition to that they are able to deal with incomplete information or noisy data and can be very effective especially in situations where it is not possible to define the rules or steps that lead to the solution of a problem. Classification is one of the most frequently encountered decision making tasks of human activity. A classification problem occurs when an object needs to be assigned into a predefined group or class based on a number of observed attributes related to that object. Many problems in business, science, industry, and medicine can be treated as classification problems. The aim of this paper is to present the comparison of neural network techniques as data mining methods on data from tourism and also to present a model of intelligent recommender system, considering that in this model a classification problem will be solved with data mining using neural network methods.

Artificial Neural Networks (ANNs)

An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. ANNs, like people, learn by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process [3]. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons. This is true of

ANNs as well. There are many different types of ANN models rather than a single type. Each form of ANN has different characteristics for a specific set of conditions, analogous to the functional specificity associated with different regions of the brain [4]). However, all ANN models are specified in terms of three basic entities: models of the neurons themselves, models of synaptic interconnections and structures, and the training rules for updating the connecting weights[5]. An ANN consists of a number of highly connected ANs such that each AN is connected to other ANs or to itself. According to the architecture, ANNs can be roughly classified into Feed forward Neural Networks (FNNs), Recurrent Neural Networks (RNNs), and their combinations. Some popular network topologies including fully connected FNNs, RNNs, Self-Organizing Maps (SOMs), and Cellular Neural Networks (CNNs). [6]

Self-Organizing Maps (SOMs)

The SOM is a feed forward unsupervised learning network. [7] It typically contains a two-dimensional single layer of neurons in addition to an input layer of branched nodes, as illustrated in Figure 1. SOM neurons have two different types of connections. There are forward connections from the neurons in the input layer to the neurons in the output layer, and also lateral connections between neurons in the output layer. The lateral connections are used to create a competition between neurons. [8]
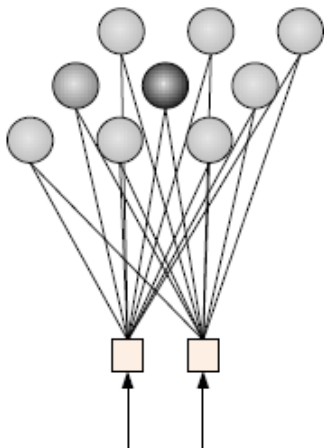

Figure 1. Self-Organizing Maps (SOMs) Topology[8]

Training algorithms for unsupervised ANNs are different as there is no desired output. SOM training is based on a competitive learning strategy: measured based on the Euclidean distance, the best neuron learns by shifting its weights from inactive connections to active ones[8]. In other words, the neuron with the largest activation level among all neurons in the output layer becomes the winner (the winner-takes-all neuron). This neuron is the only neuron that produces an output signal. The activity of all other neurons is suppressed in the competition. Neurons close to the winner are also updated according to the neighborhood relationships. In this way, SOMs effectively cluster the input vectors through a competitive learning process, while maintaining the topological structure of the input space.

Multilayer Feed forward Neural Network

Figure2 shows a typical architecture of a FNN – ANs are arranged in layers, and each AN is connected to all ANs in adjacent layers. There is no connection between the neurons within each layer. The information flows in away whereby each AN takes inputs from all the nodes in the preceding layer and sends its single output value to all the nodes in the next layer. The leftmost layer is provided with input by the user, and the output from the rightmost layer is the output which is finally used to do something useful. Popular FNNs include Multi-Layer Perceptrons (MLPs) and RBF networks, which are both fully connected layered FNNs. The MLP is the most

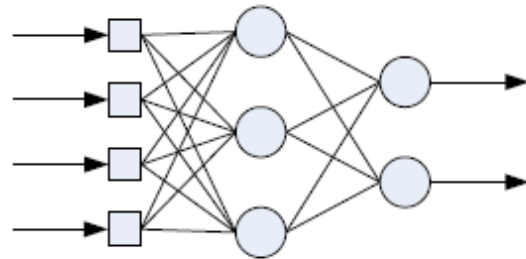Popular arrangement of ANNs.[8].

Figure 2. Multilayer Feed forward Neural Network[8]

An MLP usually consists of three layers – an input layer, a hidden layer and an output layer. The number of input neurons is typically determined to correspond to the dimension of the input vector. The number of neurons in the hidden layer is determined experimentally and the dimension of the output vector to be modeled or the number of classes to be classified generally determines the number of output neurons. Each neuron has a number of inputs (from outside the neural network or the previous layer) and a number of outputs (leading to the subsequent layer or out of the neural network[8]). A neuron computes its output response based on the weighted sum of all its inputs according to an activation function. Data flows in one direction through this kind of neural network starting from external inputs into the first layer, which are transmitted through the hidden layer(s), and then passes on to the output layer from which the external outputs are obtained.

Proposed Recommender Model for E-tourism

This is a proposed framework model for intelligent tourism guiding system and e-tourism development as a recommender system based on neural network. In this suggested model, necessary tourism data are collected not only by the companies and traveling agencies, but also by websites and tourism web sites which are including the following data groups: First data group is related to the user's profiles which are available in companies and tourism agencies in file or are obtained from membership output of tourism web sites and websites users and are stored in a special database. The Second data groups related to the travel previous records which are obtained from travel history databases by the special queries and suitable conditions. The 3rd data groups are including notes which are received as the customer's feedback or through contact forms or message panels and associations in tourism websites and web sites [9]. They are also collected through the questionnaires presented by Agencies and tourism companies. After pre-processing these data groups, enter a recommender engine implying upon neural networks via data mining

algorithms. In this engine classification on data groups Will be done based on neural network. The output of this engine is recommendation travel package for decision making tourism which is announced to the tourist via tourism companies or in a suitable division from tourism websites is introduced to him/her as a recommended travel package. The proposed recommender model is shown in figure 3.
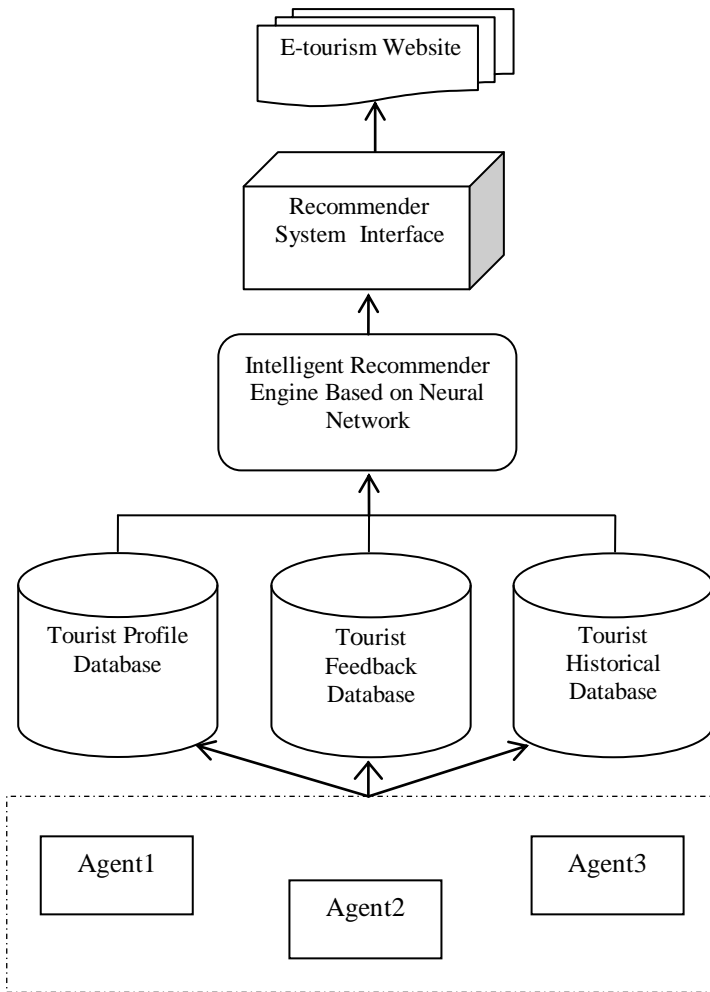

Figure3. Proposed Recommender Model for E-Tourism

Case Study:
Data Collection
Data were obtained from tourism agencies and companies in Isfahan city to simulate recommender model based on neural network. There are 400 records filtered from original databases. For tourist classification, we need to select the primary variables to identify tourist delivery and travel package, we choose 16 attributes as classification variables which are season, education level, age, etc. The classification variables is shown in Table 1.

Table I. Classification Variables

Season	Gender	Number of Nights	Interest to History
Education	Travel Type	Hotel degree	Interest to shopping
Language skill	Number of Trips	Maximum Cost	Interest to Water Sports

Age	Airline Degree	Interest to Nature	Interest to Exhibition

The experimental study for design of recommender engine according to proposed recommender model, is carried out by using Matlab software. This is a classification problem that is solved by using two methods or neural network architectures. The output of this recommender engine, is recommended travel packages for tourists includes travel

package number 1 to travel package number 12 which is shown in table 3.

Table II. Travel Packages as Target Classes

Travel Package(#1) season: Summer Destination: Thailand AirLine Degree:1 /2 Hotel Stars:5 Number of Days:7 Min Price:1200000	Travel Package(#2) season: Summer Destination: Thailand AirLine Degree:2 /3 Hotel Stars:4 Number of Days:7 Min Price:900000	Travel Package(#3) season: Summer Destination: Thailand AirLine Degree:3 /4 Hotel Stars:3 Number of Days:7 Min Price:600000
Travel Package(#4) season: Autumn Destination: Thailand AirLine Degree:1 /2 Hotel Stars:5 Number of Days:7 Min Price:1600000	Travel Package(#5) season: Autumn Destination: Thailand AirLine Degree:2/3 Hotel Stars:4 Number of Days:7 Min Price:1300000	Travel Package(#6) season: Autumn Destination: Thailand AirLine Degree:3/4 Hotel Stars:3 Number of Days:7 Min Price:1000000
Travel Package(#7) season: Autumn Destination: Thailand AirLine Degree:1 /2 Hotel Stars:5 Number of Days:7 Min Price:1500000	Travel Package(#8) season: Autumn Destination: Thailand AirLine Degree: 2/3 Hotel Stars:4 Number of Days:7 Min Price:1200000	Travel Package(#9) season: Autumn Destination: Thailand AirLine Degree: 3/4 Hotel Stars:3 Number of Days:7 Min Price:1000000
Travel Package(#10) season:Winter Destination: Dubai AirLine Degree:1 /2 Hotel Stars:5 Number of Days:7 Min Price:800000	Travel Package(#11) season: Winter Destination:Dubai AirLine Degree: 2/3 Hotel Stars:4 Number of Days:7 Min Price:600000	Travel Package(#12) season: Winter Destination: Dubai AirLine Degree: 3/4 Hotel Stars:3 Number of Days:7 Min Price:450000

Design of Recommender Engine: Determination of Network Architecture:

In the beginning, in data mining and design of recommender engine, the self-organized method of neural network is used. The reason for using the algorithm of SOM or kohonen neural network is that we expect the neural network to classify the training data which is the specifications of users' electronic tourism without the supervisor, means without a pattern or a sample. For classifying data, the different samples are given to the network and the group name of any sample is specified as the output. After training any receiving data of new samples, the network can specify which class this sample belongs to. The operational trend of this proposed model is that at first all users groups of tourism website are united and then it takes action in clustering them. This neural networks technique can assist to select the number of clusters or to select any person in any group. It is used SOM architecture in clustering, because the elements and person enter and they are gathered in a group at first. After passing clustering phase, it is classifying phase in which the information related to a person are given as the

entry and finally it is made decision who and what group must be selected. When selecting the person in the special group, the travel suggested package is introduced to him/her and put it in access which the previous generated network is used in this part.

Method1. Classification with SOM Architecture

There are 16 entries including statistical sample data with self-organized network and also there is an output which will be the same number of travel suggested package for tourist and we will have Matrix 400 in 17. In this self-organized method, network topology is including 4 clusters because In classifying method with the presented statistical sample data, there are 12 classes in which data are classified which means a vector 12 in 1 has the member 0 or 1. The classes correspond to travel recommended packages. The self-organized method is generated random centers in designing network, so that any entry is in the same weight with the other entry. After training network, the nearest vector to the cluster center is found and allocated. In the repetition of training act, the cluster centers are updated continuously. Whenever the records are presented to the network, it is an epoch which we define this repetition rate as 500 in the network. When the classification is finished, we can investigate the output dispersion. In this way and repetition of the operation, all clusters similar to each other are arranging near one another. It is determined which outputs related to the mentioned clustered data are through the entry of classification new data in which they must be placed. The output dispersion rate in this cluster is as a suggestion which is proposed to him/her as the user's suitable traveling package by the system.

Method 2.Classification with Feed Forward Architecture

In this method, the numbers of the neurons of the hidden layer are changed from 3 to 15 and it is repeated 30 times in any phase of the test. The network has the least classification error, means any data is classified correctly and this network will be selected. We train data to obtain 12 collections. When new data enter the network, the output will be a vector again, so the data belongs to the group which has the most amount in output vector. The result of solving problem in classification method was that we had the least classification error and as a result had the best network performance with 8 neurons in hidden layer. That is shown in table 4.

Table III. Network Selection with MSE Parameter

Number of Neurons	Mean Squared Error(Mse)
0.3177	3
0.2475	4
0.1338	5
0.1906	6
0.5485	7
0.1037	8
0.1271	9
0.2375	10

Result and discussion

The following figures and results show the classification of sample tourist dataset using SOM architecture. The results are defined in Figure 4, 5.
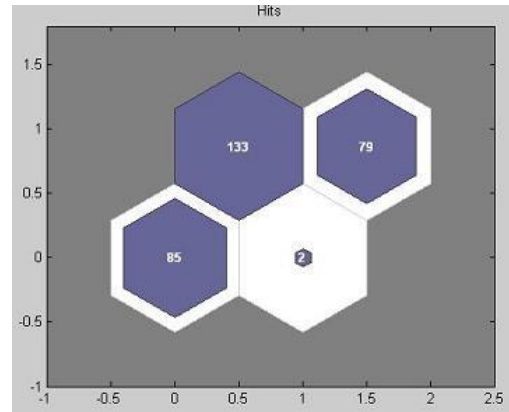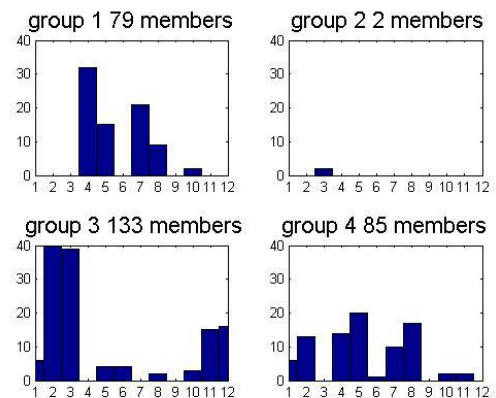

Figure4. Som Sample Hits

Figure 5. Member Grouping

Figure6 shows classification matrix of classify data with multi layer feed forward architecture. The confusion matrix gives the results obtained in train and test stage. A confusion matrix contains information about actual and predicted classifications done by a classification system. Performance of such systems is commonly evaluated using the data in the matrix. The body of the table has one row and column for each possible classification. The rows correspond to the correct classifications and the columns correspond to the predicted classifications.

Table IV. Confusion Matrix

Class	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
P1	12	0	0	0	1	0	0	0	0	0	0	0
P2	0	57	3	1	6	0	0	0	0	0	0	0
P3	0	0	37	0	0	0	0	0	0	0	0	0
P4	0	0	0	38	2	0	1	0	0	0	0	0
P5	0	0	0	7	30	0	0	0	0	0	0	0
P6	0	0	0	0	0	5	0	0	0	0	0	0
P7	0	0	0	0	0	0	28	1	0	0	0	0
P8	0	0	0	0	0	0	1	27	0	0	0	0
P9	0	0	0	0	0	0	0	0	0	0	0	0

P10	0	0	0	0	0	0	10	0	0	6	0	0
P11	0	0	0	0	0	0	0	0	0	0	17	2
P12	0	0	1	0	0	0	0	0	0	0	0	14

Network	Wrong	18	
Human Expert	Correct	68	68%
	Wrong	32	

The above results was obtained through studying the report of network function in each methods. In this process 75% of data were used in training network, It is evident that error rate in validation and test data is in higher level because it is added outside of network training cycle. The necessity for the investigation of the network with validation data is that it is prevented from over training of the network besides testing data. Here validation check has been the agent of stop in network training. It means that training of the network continues so much that the network keeps the data and it responses well just with the mentioned data.

Performance Evaluation

To evaluate the performance of recommendations according to customer's specifications, We designed a special questionnaire for receiving travel experts recommendations and professional tourists using the data records which was used in neural network testing. After study these questionnaires, we calculated percentage of correct recommended packages or prediction accuracy using our model compared with the recommendations of experts with 100 records as test dataset. The following results will show our proposed model with SOM architecture, is better. In our proposed model ,the knowledge is distributed in different agents and improve the quality of the recommendations provided[9]. A formula to calculate the correct classification rate is the model prediction accuracy formula as below:

$$P_{err} = \frac{N_{err}}{N} \quad (\text{Error Frequency}) \qquad (1)$$

$$N_{err} = \sum_{i=1}^{k} Ni \quad (\text{Number of wrongs}) \qquad (2)$$

The model prediction accuracy is calculated by subtracting the error frequency from 1. The overall accuracy of model is calculated in all three methods, including a neural network model using SOM architecture, a neural network model using FNN architecture, and Human Experts' knowledge. The results are shown in table 7.

Table V . Comparison of Classification Methods

Method	Recommendation Status	Number of instances	Classification Accuracy
MLP Network	Correct	76	76%
	Wrong	24	
SOM	Correct	82	82%

Conclusion

According to the mentioned points, neural network methods were used in this research for the tourism data mining and hence proposing travelling packages to the customers in the tourism industry. By considering the results obtained from the samples testing procedures, it can be found that designing a recommender model based on neural networks is an appropriate tool for e-tourism development and its optimization. It is to note that comparing the results from the methods of SOM architecture(unsupervised) and feed forward architecture (supervised) shows that SOM is more accurate for the offers to the customers, in the e-tourism decision making model. The other actions that could be done in future according to the proposed model in this research includes determination of the tourists' satisfaction from the offered tour package as well as defining the rate and state of the effects of interests, features and feedbacks of tourists on determination of the output package.

Acknowledgment

References

[1] N. Sharda, Intelligent visual travel recommender systems model for e-tourism websites. Gold Coast, Sustainable Tourism CRC, 2008.

[2] Du, K. L. and M. N. S. Swamy, Neural networks in a softcomputing framework, London, Springer, 2006 .

[3] P. Richard, An Introduction to Computing with Neural Nets. Lipmann, IEEE ASSP Magazine, 1987.

[4] M. Berthold, D. J. Hand , Intelligent data analysis: an introduction, Berlin, New York, Springer, 2007.

[5] K. L. Du, M. N. S. Swamy. Neural networks in a softcomputing, 2006.

[6] D. Graupe, Principles of Artificial Neural Networks, world scientific publishing, New Jersey , London, 2007.

[7] T. Kohonen, Self-organizing maps, Berlin, New York, Springer, 2001.

[8] J .Yang, Intelligent Data Mining using Artificial Neural Networks, University of Warwick , PHD Thesis., 2010.

[9] Ricci, F. , Werthner,H., "Case-Based Querying for Travel Planning Recommendation", Information Technology and Tourism, Vol. 4, No. 3, 2002.