ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940www.ijsem.org
Vol 6, Issuse.2 April 2018

A recommendation system built on a decision tree

D Navya, ,
Mallika pasham,
M Naveen Kumar,

Abstract—Choosing a tourist destination from the information that is available on the Internet and through other sources is one of the most complex tasks for tourists when planning travel, both before and during travel. Previous Travel Recommendation Systems (TRSs) have attempted to solve this problem. However, some of the technical aspects such as system accuracy and the practical aspects such as usability and satisfaction have been neglected. To address this issue, it requires a full understanding of the tourists' decision-making and novel models for their information search process. This paper proposes a novel human-centric TRS that recommends destinations to tourists in an unfamiliar city. It considers both technical and practical aspects using a real world data set we collected. The system is developed using a two-steps feature selection method to reduce number of inputs to the system and recommendations are provided by decision tree C4.5. The experimental results show that the proposed TRS can provide personalized recommendation on tourist destinations that satisfy the tourists.

***Keywords*: Recommendation System; Tourist Destination, Feature Selection; Filtering methods; Mutual information; Classification; Decision Tree**

I. INTRODUCTION

The tourism industry is an extremely important sector on a global scale and contributed 9.5% to the total world's economy in 2013. It is expected that tourism will contribute around 10.3% GDP in 2023. South East Asia is expected to be the fastest grown regions in terms of its Travel and Tourism contribution to the GDP. In particular, Thailand, Indonesia, Singapore and Myanmar were identified as the countries possessing the most attractive tourism features in 2013 [1].

International tourist arrivals in Thailand have doubled over the past nine years (See Fig 1). In 2013, Thailand is the 10th most visited destination worldwide[1]. The country attracts 26.5 million international tourists grew by 18.76% over 2012 [2]. Increasing both tourist numbers (international and domestic) and the benefits from tourism are the primary objective of the Thai government. In 2013, tourism generated 1.79 trillion BHT ($55.49 billion) in revenue for Thailand[2].

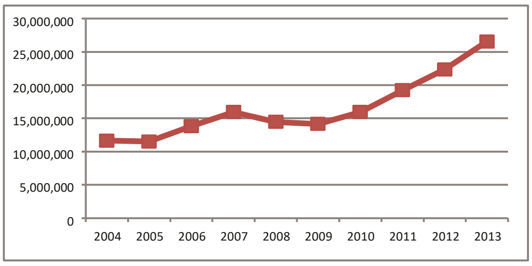


Figure 1. Number of international tourist arriving in Thailand from 2004- 2013 [1]

Assistant Professor[1]
Assistant Professor[2],
Assistant Professor[3]
Department of CSE Engineering,
Pallavi Engineering College,
Kuntloor(V),Hayathnagar(M),Hyderabad,R.R.Dist.-501505.

The Internet is now considered to be the main information source of tourists for information on products and services [3]. Due to the huge volume of heterogeneous information available on the Internet, the search for destinations, as known as travel planning can overwhelm tourists. The travel-planning task is complex and dynamic such that there are many factors involved when making a decision, for examples, the quality of the attractions, travel routes, hotels, numbers of traveler, leisure activities, weather, etc.[4].Recently, tourism has substantially benefited from ICT, and especially from Internet technology [5]. With the development of decision support tools, also known as Recommendation Systems (RS), tourists and tourism providers can search, select, compare, and make decisions more efficient than ever.

Most of the previous TRSs have focused on estimates of choosing the destination, activities, attractions, tourism services (e.g. restaurants, hotels, and transportation) based on the user's preferences and interests. With regard to technical aspects, these TRSs only provide filtering, sorting and basic matching mechanisms between the items and the user's hard constraints. However, they are lacking in technical aspects (e.g. sparsity, scalability, transparency, system accuracy, theories to improve personalization, etc.) and practical aspects (e.g. user satisfaction, usability, etc.).

One of the greatest challenges in developing a TRS that provide personalized recommendations of tourist destinations is to enhance the tourist decision-making process. In order to achieve this, it requires a deep understanding of the tourists' decision-making and develops novel models for their information search process. Also, uncertainties involved in the information search stage of a tourist decision process need to be eliminated. By reducing more parameters in the system, the model complexity could be decreased. In return, the recommendation performance and the level of user satisfaction of the system can both be increased.

This paper proposes a novel human-centric TRS that recommends destinations to tourist to solve the mentioned challenges. The proposed TRS is processed offline using the Data Mining (DM) process. This includes data acquisition, variables selection by using feature selection methods, decision making by using decision tree C4.5, and interpretation of the decision tree. The proposed TRS has three main innovations. Firstly, two feature selection methods are used to remove the unnecessary (both irrelevant and redundant) inputs into the system and to decrease the model complexity. Secondly, a decision tree C4.5 is used as a classifier to identify the tourist destination selection process. Lastly, the proposed system uses real world data that have been collected by us from Chiang Mai, Thailand.

The paper is organized into the following sections. Section 2 provides background on recommendation systems in the tourism domain. Section 3 describes the data collection process used in this paper. Section 4 presents the proposed TRS framework using the DM approach. The experiment setup for this study is demonstrated in Section 5. Section 6 shows the results and the evaluation analysis of the proposed TRS. Finally, we present some tentative conclusion and our future work in the last section.

BACKGROUND

A.Recommendation System

A recommendation system (RS), a subset of Decision Support Systems (DSS), is a tool that can recommend an item based on the aggregated information of the user's preferences [6]. It supports users by providing valuable information to assist them in their decision-making processes based on their priorities and concerns [7]. RS plays an important role and is common in many popular e-commerce websites, such as Amazon, Netflix, Pandora, etc. The e-commerce RSs suggest items to the user which involve news, articles, people, URLs, and so on [8].

B.Travel Recommendation Systems

Tourism is a leisure activity that involves complex decision processes, for example, selecting destinations, attractions, activities, and services. Thus, TRS attract the attention of many researchers from the fields of both academics and industry. Various TRS have been developed/deployed in and on many kinds of platforms (e.g. desktop, browser, mobile). TRSs recommend results to a user for the purposes of estimating user interest, choosing Points of Interests (POIs), identifying services or routes, ranking them in sequence, or as a holistic trip plan.

Most of the current TRSs aim to support an individual tourist, although there are some systems that support travel agencies as well [9]. They share similar frameworks but differ in the selection of technology, theories to improve personalization, data inputs, interaction style, and recommendation techniques. Fig 2 shows the general framework of the recent TRSs. Information from various sources (e.g. sensors, GPS Coordinates, surveys, reviews, etc.) are integrated and kept in the repository (e.g. database schema, ontology).

The recommendation engine can be composed of several subsystems such as an optimization subsystem, a statistical subsystem and an intelligent subsystem and so on. This is to suggest, rank, or predict the items (i.e. destination, attractions, activities, and services) based on user requirements, preferences, or some hard and soft constraints (e.g. user demographic information, number of travel days, travel budgets, travel type, etc.).

Generally, before or during the trip, the TRS would take some inputs from the tourist (implicit, explicit, or both) to create a user profile and calculate the recommended result which is then sent back to the tourist. Tourists can visualize the results from the system in many ways, such as by destination icons on the map interface with a route between point-to-point, agenda, and itinerary. Most TRSs present the result with the use of spatial web services such as the Google Maps API service.

Lately, some TRSs are able to adapt the results to the user by taking the user context information (e.g. location, weather) into account. Some TRSs provide a functionality to let the user modify the generated result and adapt the results based on user feedback or user rating [10], [11].

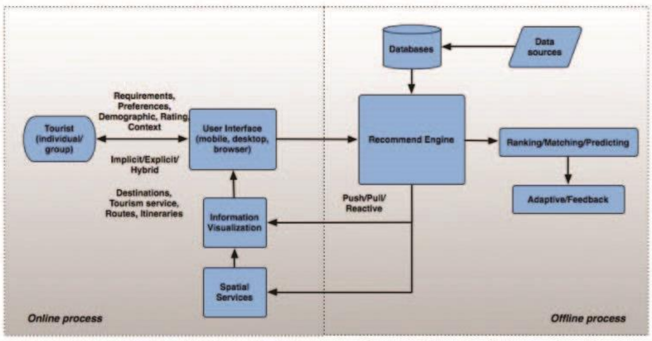


Figure 2.General framework of the travel recommendation systems

C.Recommendation techniques

According to [12], RS can be classified by the degree of personalization, including the usefulness and accuracy of the recommendations. The degree of personalization can be defined from low to high, including nonpersonalization, ephemeral personalization (short term), and persistent personalization (long term). The nonpersonalized RS is a fairly simple system that does not take the user preferences into account when making recommendations. For instance, the RS only generates a list of the most popular items based on the number of reviews or number of purchases (i.e., editor's choices or top-sellers). As a result, the recommended results would likely be of value to other generic users of the system. Due to their limited decision making power, nonpersonalized systems have not been a focus of RS research [7].

Concerning the information incorporation related to the system users (i.e. user preferences, sociodemographic information, etc.), an ephemeral and personalized RS is more advanced than a nonpersonalized RS. In other words, every user would be able to see a different list of recommendations depending on his/her preferences. For example, Trip-advisor[1] recommends a destination based on the user's sociodemographic information. In fact, there are many types of personalized RSs that have been analyzed in previous studies, and the researchers have categorized them according to the method of the information-filtering techniques [7], [13]–[15]. In the next section, we will briefly investigate the recommendation engine (Fig. 3) which is composed of several recommendation techniques based on findings in [14]. The advantages and disadvantages of each type, and the hybrid filtering approach applied (i.e. the networking of several RSs) will be discussed.

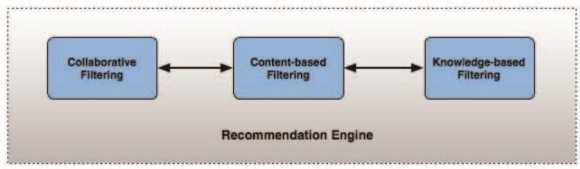


Figure 3. Recommendation Engine

a) **Collaborative filtering:** This approach is widely adopted by the most implemented recommendation systems. It recommends item(s) to the user based on the feedback of other users who share the same attributes, and suggest popular items to the user. This approach still suffers from a cold-start problem, where the new item or user would need to be rated before a recommendation can be made.

b) **Content-based filtering***:* This recommendation technique suggests items to the user based on his/her previous searches or queries for items. The main drawback is the cold-start problem for the user, in which the user needs to provide a significant amount of information before the system can generate a recommendation. Otherwise, the system needs to have archived large historical data set in order to generate quality results [13]. Another common problem is overspecialization, since the system is most likely to suggest the item that the user liked the most, with less diversity among the recommendations [7].

c) **Knowledge-based filtering***:* This technique recommends items to the user based on the knowledge of the domain. In other words, the system has some knowledge of how the particular item relates to a particular user. Predominantly, this technique can be achieved by using case-based reasoning or ontological methods. This recommendation technique can be found in [9] and [16], where the system exploits the travel agencies' and group expertise's past experiences.

d) **Hybrid filtering:** The above mentioned recommendation techniques have some strengths and weaknesses. The purpose of the hybrid recommendation technique is to achieve the best performance and to remove the weaknesses/disadvantages of one technique by complementing it with the advantages of another technique. Also, there are many hybridization methods, such as combining recommendation techniques together including weight, switching, mixed, feature combinations, cascades, feature augmentations, and metalevels [13]. The latest Information and Communications

Technology (ICT) e.g. Artificial Intelligent (AI), Semantic web, Communication network, etc. provides new opportunities for researchers to design and implement a TRS that is more intelligent, interactive, and adaptive, while being automatable, and supporting a higher degree of user satisfaction than ever before. We aim to develop a system to achieve those characteristics.

III. METHODOLOGY

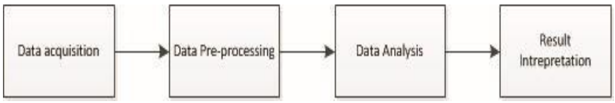


Figure 4. Data Mining Framework

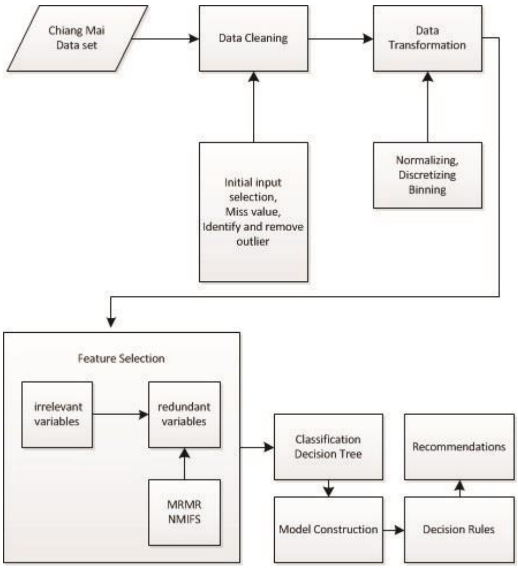


Figure 5. Methodology of the proposed destination TRS

The proposed DM framework shown in Fig. 4 consists of four phases including data acquisition, data preprocessing, data analysis, and result interpretation. (1) For data acquisition, the designed questionnaire, which has four parts, is distributed and collected from Chiang Mai, Thailand. (2) The collected data is pre-processed using several data pre-processing techniques involving data cleaning, data transformation, and feature selection methods. (3) The third phase involves the data analysis processes using a decision tree C4.5 as classifier. The aim of the third phase is to identify suitable features and find the optimal models. (4) The final phase involves the interpretation of the obtained optimal decision trees and the extracted decision rules. The flow of the processes is described in Fig 5.

A. Data acquisition

To understand tourist's search behaviour in assessing travel information and decision-making processing for destination choice, we use a questionnaire as a data collection method due to its effective mechanism for collecting information from tourists. Pre-study on variety of factors that influence tourist's preferred destinations were identified for questionnaire design. The questionnaire design contains four parts containing a set of factors related to tourist's preferred destinations as following:

Trip characteristics: These variables are the most important variables when tourists select their destinations [17]. This includes trip length, travel purpose, trip composition, and etc.

Tourist characteristics: These variables includepychological, congitive and socioeconomic status varaiblesthat influence on the tourist destination choice process[17].

Travel motivations: Travel or tour motivation is one of the important factors we have found from literature reviews when tourists are selecting their destinations. This variable describes the reason that a tourist chooses to visit a destination [18].

Tourist sociodemographic information*:* The individual demographics may influence the information seeking behaviour [19].

4,000 Questionnaires were distributed and collected at the five preferred tourist destinations in Chiang Mai, Thailand. The list of the preferred destinations was retrieved from the Trip-advisor website[1]. The survey was distributed to both international (60%) and domestic tourists (40%). The destinations included Art in Paradise (27.7%), Mae Sa Waterfall (22.06%), Huay Tung Tao Lake (19.18%), Museum of World Insect and Natural Wonders (16.97%), and Bua Thong Waterfall (14.09%). The participants took 15-30 minutes on average to complete the questionnaire. 3,695 valid questionnaires with 145 variables were imported to data pre-processing stage, while 35 samples were rejected as being incompletely filled in.

The proposed framework uses variables extracted from questionnaire as inputs for classification of the tourist's preferred destination, including travel characteristics, tourist behavior, tourist expenditure behaviour, travel motivations, and tourist demographic information as described above.

Data Pre-processing

Real world data are generally incomplete, noisy, and inconsistent. For example, with surveys like ours, respondents may intentionally submit incorrect data because they do not want to submit personal information, or there may be data entry errors. The best classification results require good quality data. To achieve this, we preprocessed the survey data through data integration, data cleaning, data transformation, and variable selection using feature selection methods.

Feature selection or variable selection is a process of selecting subsets of relevant features that describes the output classes. It is very important process for not only the utilization and usability, but also for accuracy improving. In this paper, we try to use small number of variables, which should contain the maximum information at the same time. In other words, it is to reduce the number of necessary user inputs as well as to increase performance of the classification model. In this paper, we propose a two-step filtering method based on Mutual Information (MI) to rank the features and remove irrelevant and redundant features from the dataset.

MI is used as a measurement in the feature selection process to characterize both the relevance and redundancy of the variables. If the variables were independent of each other, the MI value is zero. The greater the MI value, the more significant the dependent variable was. Given a set of *X* and *Y*, *p(x)* and *p(y)* are the marginal probability distribution functions of *X* and *Y*, and *p(x, y)* is the joint probability distribution function of *X* and Y. The MI for discrete variables is presented as:

First filtering method

The purpose of the first filtering step is to rank the variables and remove any independent variables that are unrelated to the dependent variable. We applied a MaxRelevance feature selection algorithm [20], in which we chose MI as the measurement to remove the irrelevant features. We computed the MI score between each independent and dependent variable. Then, we ranked them in descending order and used a threshold value (the threshold value is chosen manually) to remove features that contributed less or were not related to the predictive power.

Second filtering method

In the second filtering step we used two mutual information-based, feature-selection algorithms: Minimum Redundancy Maximum Relevance (MRMR) [20] and Normalized Mutual Information Feature Selection (NMIFS) [21] to remove the redundant variables. The optimal feature space was chosen using the maximum MI *G* value. Feature selection stops when *G* < 0 is reached.

MRMR algorithm

The idea of the MRMR algorithm [20] is the algorithm using the MI value to rank the features based on the minimal redundancy and maximal relevant criterion. MRMR calculates redundancy for every pair of features and calculates the relevance between the feature and the class. It is formulated as (1) below.

b)NMIFS algorithm

NMIFS [21] is a modification of the MRMR algorithm (See (2) and (3)), it normalized the original MI value by the minimum entropy (*H*(i) and *H*(j)) of both features as shown in the equation below.

□□(□;□)

C.Data Anaylsis

Decision tree is chosen as a classifier/model for the proposed TRS because it provides several benefits for decision maker such as simplicity, interpretability. Decision-making is easily understood due to its flowchart-like model. For technical aspects, it handles the TRS's technical issues in terms of sparsity and scalability.The decision tree consists of nodes and leaves. The first node is called the root node, where the instances from the test set start to navigate down to a leaf. Other nodes, referred to as internal nodes, involve testing a particular attribute; this is where the split – either binary or multi – occurs. The leaf nodes represent a class label (i.e., the output of the classification) or the final decision of the instance from the test data.[22]. To recommend a destination to tourist, we must traverse the decision tree from the root to the leaf. Many decision trees exist, such as Hunt's algorithm, Top-down Induction of Decision Tree (TDIDT), ID3, CHAID, CART and C4.5. They differ in terms of splitting criteria, pruning, type of attributes, etc.

C4.5, an extension of ID3, was devised in [23]. It was chosen for this study because C4.5 tried to solve ID3 main drawbacks. ID3 [24] is the most simple decision tree algorithm but has many drawbacks such as that the optimal solution is not guaranteed, over-fitting problem with the training data set, and it supports only nominal variables. On the other hand, C4.5 supports two types of splitting criteria, including the information gain and the entropy-based criterion. It also supports both nominal and scale variables. In order to avoid the over-fitting problem, C4.5 supports tree pruning (e.g., confidence-based and error-based pruning). Moreover, C4.5 allows attributes to be missing.

IV. EXPERIMENT DESIGN

A.Representation of data set

Table 1 describes the characteristics of the data set used in this study. The data set contains five tourist's preferred destinations. However, the decision tree model that was constructed using all five destinations archived a very low of rate classification accuracy of 36.1%. In addition, the decision tree model was too complex such that it has a large tree size and number of leafs, which makes it difficult to interpret for the decision-maker. To solve this problem, this multi-classes classification problem is divided into several sub problems by investigating the type of tourist's preferred destinations, combining the knowledge from Chiang Mai tourism domain experts and destination information from the trip advisor website.

Hence, the two categories were constructed and are presented in Table 2. The decision tree models were constructed based on these categories. The Museum data set presents a binary classification problem and the Nature data set presents a multi-classification problem. The Museum data set consists of two classes, as there both are specialized museum. The Nature data set consists of three classes, two of them represent the waterfall and one of them represents the lake.

TABLE 1.CHARACTERISTICS OF THE DATA SET USED IN THIS STUDY

TABLE 1. CHARACTERISTICS OF THE DATA SET USED IN THIS STUDY

Data set	# Features	# Classes	# Sample
Tourist destination choice	145	5	1,632

B.Data pre-processing

Initial selection is the first step for the process of cleaning the data. In this phase, knowledge acquired from tourism domains is used to select the features that are not related to output classes. Next, missing value analysis is performed for both data sets. Continuous variables were discretized using the binning method. The bin size is chosen as 10. Some of the discrete variables were normalized using tourism domain expert knowledge. After the data set had been cleaned and transformed, the proposed two-step filtering method was applied. This was done to remove the irrelevant and redundant features from the data set. For the first filtering step, different numbers of thresholds were used based on each data set to select between 17-18 relevant features. Then, MRMR and NMIFS feature selection algorithms were applied to the sub-set feature in order to remove the irrelevant features.

*C.***Classification and model construction**

After the irrelevant and redundant features were filtered out, and the designated features were selected, we then constructed a classifier using a decision tree. An investigation of C4.5 performance from the two feature selection algorithms is carried out.

The *K* repeat holdout method was applied in this experiment. In each iteration, a 60% sample from each data set was randomly selected for training, 20% was used for validating, and the rest was used for testing, with stratification (i.e. each class has the same proportion in training, validation, and testing sets). The predictive accuracy of training, validating sets on the different iterations was averaged. Different configurations on confidence levels for decision tree pruning are used to find the optimal models for the two data sets. The confidence levels ranged from 0.1 to 0.5, with a step size of 0.1. The optimal model is found when the following two conditions are met.

1. Best of mean of accuracy of validation sets.
2. Mean of accuracy of validation set must be equal or less than the mean of accuracy of training set.

V. RESULTS AND SYSTEM EVALUATION

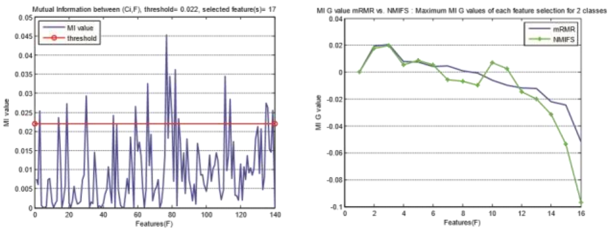
Table 2 presents result of classification rate using C4.5. For single best learner, it can be seen that the Museum data set achieved a classification rate of 80%. The Nature data set revealed a classification rate of 49.72%. Regarding the performance of the two feature selection algorithms, the NMIFS algorithm is considered superior to the MRMR algorithm for both of the data sets.

TABLE 2.ACCURACY RATE FOR EACH DATA SET

TABLE 2. ACCURACY RATE FOR EACH DATA SET

Data set	# of classes	#Sample	Confidence level	Single best learner accuracy rate
Museum	2	729	0.39	80%
Nature	3	903	0.24	49.72%

Fig 6 shows the data pre-processing result from the Museum data set. Fig 6(a) presents the MI value from the first filter method, the threshold was set as 0.022, 128 variables were removed from the data set. Fig 6(b) shows the MI G values from both of the feature selection algorithms. Feature selection stopped when negative values were reached.



(a) (b)

Fig 6. MI value (a) and MI G value (b) from the two-step feature selection method of the Museum data set

Table 3 presents the selected features from both of the feature selection algorithms of the Museum data set. The bold variables indicate that the corresponding feature belongs to the optimal subset. For the second filtering method, MRMR algorithm selected eight optimal features and NMIFS selected six optimal features for the Museum data set. It can be seen that feature ***a*** is the most important. This can be explained by the notion that one of the museums is specialized in insects. The feature *c*, *d*, and *b* were combined to help classify the data set. The optimal decision tree for the Museum data set is obtained and the decision rules are generated, combining four selected features from the NMFIS (See Fig 7 and 8). The obtained decision tree is viewed as being simple with the size of 17 and it has a number of leafs equal to 10. For the Nature data set, b2 (trip purpose) was selected as the most important feature.

TABLE 3. FEATURE RANKING BASED ON THE MRMR AND NMIFS ALGORITHMS (MUSEUM DATA SET)

Algorithm	Selected feature
MRMR	***a c d b e f g h*** *i j n k l m o p*
NMIFS	***a c d b g h*** *f j i o k e n l m p*

a:deepest impression is wildlife b:to visit place I have never been before c: The people who are companying are friend d: books and guides influences your decision to visit Chiang Mai

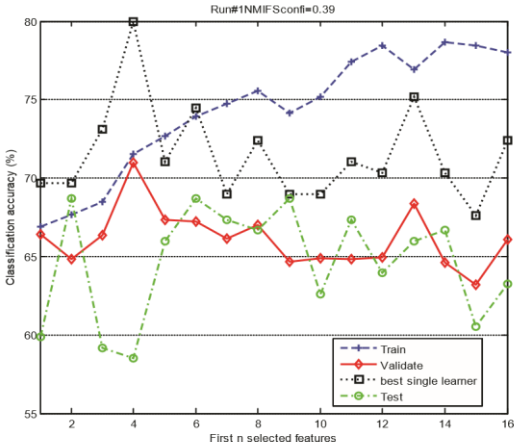


Figure 7. Accuracy rate for the Museum data set

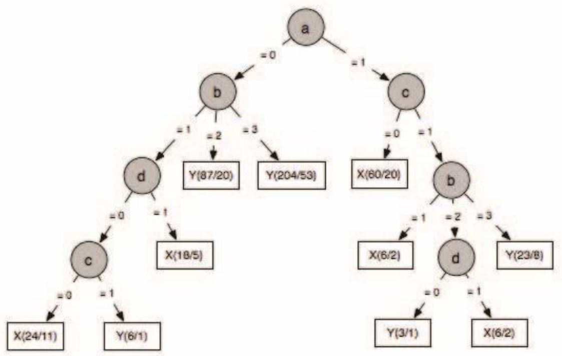


Figure 8. The optimal decision tree of the Museum data set using validation data. (X: Museum of world insects and Natural Wonders and Y: Art in Paradise, Chiang Mai 3D Art Museum)

Beside the accuracy rate, the confusion matrix is also used to evaluate the model's performance; it contains information regarding the actual and predicted classification done by the obtained optimal decision tree. According to Table 4, we can see that Museum of World Insects had a higher value of false positive (i.e. Museum of World Insects samples that were incorrectly classified as 3D Arts Museum samples).

TABLE 4.CONFUSION MATRIX OF THE MUSEUM DATA SET

TABLE 4. CONFUSION MATRIX OF THE MUSEUM DATA SET

		Predict	
		Museum of world insect	3D arts Museum
Actual	Museum of world insect	26	21
	3D arts Museum	8	90

To make it easier for a decision-maker to interpret the results, decision rules of the Museum data set are generated from the obtained optimal decision tree as shown in Table 5. There are eight rules generated for the Museum data set.

TABLE 5.THE DECISION RULES OF THE MUSEUM DATA SET

```
if a == 0, then
        if b==1 then
                if d==0
                        if c==0 then , class = X;
                        elseif c== 1 then, class = Y;
                        end
                elseif d==1 then, class = X
                end
        elseif b==2, then class = Y;
        elseif b==3, then class = Y;
        end elseif
a == 1
        if c==0 then, class = X;
        elseif c==1                     if
b==1, then class = X;
        elseif b==2
                        if d == 0 then, class = Y;
                        elseif d==1, then class = X;
                        end

        elseif b==3, then class = Y;   end
end
end
```

VI. CONCLUSION

In this paper, a decision tree based tourist recommendation system has been presented in attempt of solving the current challenge of the destination TRS. The data set has been decomposed into two sub data sets using relevant tourism domain knowledge. This was done to increase classification accuracy rate and to reduce the complexity of the decision tree. The optimal decision trees from NMIFS with the highest accuracy rate and simplicity (i.e. less number of leaf and tree size) have been constructed for destination choice. The decision rules from decision trees were extracted. It can be seen that NMIFS is the optimum method because it uses fewer number of feature than MRMR for both of the data sets. Finally, the experimental results confirm applicable of the proposed a TRS. The proposed TRS satisfies the tourists' requirements who plan to visit or during their visit the city of Chiang Mai.

For future work, different types of classifiers can be considered to increase the classification accuracy rate for the data sets. Moreover, front-end web application and an interactive and adaptive user interface will be designed and implemented.

REFERENCES

[1] "Economic Impact of Travel& Tourism 2014 Annual Update:
Summary." World travel & tourism council.

[2] "Thailand Annual Report 2013."

[3] E. Pantano and L. D. Pietro, "From e-tourism to f-tourism: emerging issues from negative tourists' online reviews," *J. Hosp. Tour. Technol.*, vol. 4, no. 3, pp. 211–227, 2013.

[4] B. Pan and D. R. Fesenmaier, "Semantics of Online Tourism and Travel Information Search on the Internet: A Preliminary Study," *Inf. Commun. Technol. Tour. 2002 Proc. Int. Conf. Innsbr. Austria 2002*, pp. 320–328, Jan. 2002.

[5] E. Pitoska, "E-Tourism: The Use of Internet and Information and Communication Technologies in Tourism: The Case of Hotel Units in Peripheral Areas," *Tour. South East Eur.*, vol. 2, pp. 335–344, Dec. 2013.

[6] G. Häubl and V. Trifts, "Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids," *Mark. Sci.*, vol. 19, no. 1, p. 4, Winter 2000.

[7] F. Ricci, L. Rokach, and B. Shapira, "Introduction to Recommender Systems Handbook," in *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds. Springer US, 2011, pp. 1–35.

[8] P. Resnick and H. R. Varian, "Recommender Systems," *Commun ACM*, vol. 40, no. 3, pp. 56–58, Mar. 1997.

[9] G. I. Alptekin and G. Buyukozkan, "An integrated case-based reasoning and MCDM system for Web based tourism
destination planning," *EXPERT Syst. Appl.*, vol. 38, pp. 2125– 2132, 2011.

[10] R. Anacleto, L. Figueiredo, A. Almeida, and P. Novais, "Mobile application to provide personalized sightseeing tours," *J. Netw.
Comput. Appl.*, vol. 41, pp. 56–64, May 2014.

[11] L. Sebastia, I. Garcia, E. Onaindia, and C. Guzman, "e-
TOURISM:: A TOURIST RECOMMENDATION AND PLANNING APPLICATION," *Int. J. Artif. Intell. Tools*, vol. 18, no. 5, pp. 717–738, Oct. 2009.

[12] J. B. Schafer, J. A. Konstan, and J. Riedl, "E-Commerce Recommendation Applications," in *Applications of Data Mining to Electronic Commerce*, R. Kohavi and F. Provost, Eds. Springer US, 2001, pp. 115–153.

[13] R. Burke, "Hybrid Recommender Systems: Survey and
Experiments," *User Model. User-Adapt. Interact.*, vol. 12, no. 4, pp. 331–370, Nov. 2002.

[14] D. Jannach, M. Zanker, A. Felfernig, and G. Friedrich, *Recommender Systems: An Introduction*. New York: Cambridge University Press, 2010.

[15] M. Montaner, B. Lopez, and J. L. de la Rosa, "A taxonomy of recommender agents on the Internet," *Artif. Intell. Rev.*, vol. 19, pp. 285–330, 2003.

[16] F. M. Santiago, F. A. López, A. Montejo-Ráez, and A. U. López, "GeOasis: A knowledge-based geo-referenced tourist assistant," *Expert Syst. Appl.*, vol. 39, no. 14, pp. 11737–11745, Oct. 2012.

[17] D. R. Fesenmaier, K. W. Wöber, and H. Werthner, *Destination UHFRPPHQGDWLRQ V\VWHPV >HOHFWURQLF UHVRXUFH@ EHKDYLRXUDO foundations and applications / edited by Daniel R. Fesenmaier, Karl W. Wöber, Hannes Werthner*. WDOOLQJIRUG 8.&DPEULGJH0$&$%,3XEF

[18] N. Leiper, "Tourist attraction systems," *Ann. Tour. Res.*, vol. 17, no. 3, pp. 367–384, 1990.

[19] K. L. Andereck and L. L. Caldwell, "The Influence of Tourists' Characteristics on Ratings of Information Sources for an Attraction," *J. Travel Tour. Mark.*, vol. 2, no. 2–3, pp. 171–190, Feb. 1994.

[20] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.

[21] P. A. Estevez, M. Tesmer, C. A. Perez, and J. M. Zurada, "Normalized Mutual Information Feature Selection," *IEEE Trans. Neural Netw.*, vol. 20, no. 2, pp. 189–201, Feb. 2009.

[22] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd Revised edition edition. $PVWHUGDP %RVWRQ 0$ 0RUJDQ .DXIPDQQ 3XEOLVKHUV ,Q 2005.

[23] J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993.

[24] J. R. Quinlan, "Induction of Decision Trees," *MACH LEARN*, vol. 1, pp. 81–106, 1986.