

Application of Support Vector Machines on Prediction of Repeat Visitation

Rong-Chang Chen^a, Hsin-Lan Lin^b

^a Department of Logistics Engineering and Management

^b Graduate School of Business Administration
National Taichung Institute of Technology

Abstract: - Since the tourism earnings became a major part of foreign exchange in many countries, the tourism market has become much more competitive. In addition, tourism has been highly valued recently all over the world, and consequently, tourism market has been developed very fast. After quick development, it gradually becomes a saturated marketplace. Previous studies indicated that maintaining existing customers costs less than exploring new customers. Repeat visitors can not only reduce the cost but gain the long-term profits for the destinations. Thus, a method that can accurately predict repeat visitation intention is greatly needed. In this paper, we apply Support Vector Machines (SVM) to predict the repeat visitation. To evaluate the effectiveness of SVM, data are collected from tourists who visited Sun Moon Lake, which is the most famous destination in Taiwan and is the only place in Taiwan among the 50 best places to visit in China. Factor analysis (FA) is also employed to reduce variables and combined FA/SVM results are compared with those predicted from SVM only. The experimental results show that SVM can provide high accuracy rates than FA/SVM. Also, some SVM ensemble techniques can give higher true negative rates.

Key-Words: support Vector Machine, repeat visitation, tourist satisfaction, travel motivation

1 Introduction

Tourism has become a major part of foreign exchange in many countries [1]. Approximately 8 percent of foreign exchange earnings are from tourism [1]. The tourism earnings were situated the first place in all kinds of earnings. According to statistical reports in 2005, the number of outbound travelers in 2004 has reached up about 760 million people, up 10 percent and been the highest growth rate over the twenty years. In addition, World Travel and Tourism Council [2] estimated that travel and tourism total demand will reach up about USD\$ 10.7 trillion in the year of 2015 and create USD\$ 2.7 trillion of travel and tourism industry GDP. The travel and tourism economic GDP created by travel activities would reach about USD\$ 7.8 trillion, 11.3 percent of global GDP. Hence, tourism industry plays a crucial role in the development of a country in the future.

On the basis of these reports, we can evidently discover that travel and tourism total demand has become one of the most important industry in developing a country and economy. However, travel and tourism industry has the characteristic of perishability; in other word, the products can not be stored up for future use. The supply is unable to be adjusted at any time by the demands and changes of market. For this reason, it is extraordinarily essential

to predict the travel and tourism demands. Therefore, if we predict the travel and tourism demands properly, we can provide these data for the planners or strategy-makers of travel and tourism industry and hence they will make right decisions.

Previous studies indicated that enterprises which improve the retention rate of customers about 5 percent can increase about 25 to 85 percent of total profit [3]. Through public praise, enterprises can not only save the cost of advertisement but gain more customers' loyalty. Some previous studies revealed that maintaining existing customers costs much less than exploring new customers. It is costly to capture new customers from competitors because a greater degree of service improvement is needed to convince customers to switch from competitors. It is the same with tourism industry. Hence, methods that can predict repeat visitation accurately are greatly needed and by which a tourism destination can improve its market position.

Some past studies were concerned about predicting demands. Law and Au [4,5] applied the Artificial Neural Network in predicting the tourism demands. The results were better than the traditional way such as Naïve, Moving Average, Exponential Smoothing and Multiple Regression. Some scholaras suggested that the results in predicting international tourism demands which were produced by using Support Vector Regression were much more accurate

than using Artificial Neural Network [6]. In this paper, we used Support Vector Machines (SVM) to be the analytical tool and it could solve the problem effectively.

People travel due to their motivations [7]. But since each individual has different nature, the importance and degree of motivations are comparatively different. These divergences are influenced by the internal and external factors of each person. Generally, tourist motivations are able to affect the degree of satisfaction and the destination will be altered by the preference and expectation while the purpose of tourists is different [8]. In addition, providing high quality services and enhancing satisfaction of customers lead to the success of travel and tourism industry [8]. High quality services and tourist satisfaction can build good and long-term relationship and achieve the repeat visitation. Tourist satisfaction has the direct influence on the intention of repeat visitation. The costs that retain the existing customers are less than capture the new customers from competitors. Therefore, it is necessary to understand the tourist satisfaction.

Though repeat visitation has the potential to become the backbone of tourism industry, few studies have been published examining the phenomenon and motivations for repeat visitation [9]. There are two purposes in this paper: First, we analyze the relationship between traveler motivations, satisfaction and repeat visitation to provide some information to proprietors of tourism industry for future planning. Second, we compare the accuracy of two datasets, original data and data after using Factor Analysis, by using SVM. In other words, we expect to build a predictive model for repeat visitation to predict the tourists' intention of repeat visitation based on the proposed structure.

This study took the visitors of Sun Moon Lake which is one of the international tourism destinations in Taiwan as our object. We collected the information about tourist satisfaction through questionnaires. Factor Analysis was employed to find out the significant factors. Support Vector Machines was adopted to predict the respondents' intention to make a repeat visitation to Sun Moon Lake.

The rest of this paper is organized as follows. Section 2 gives a brief overview of literature review. In section 3, SVM is briefly introduced. The samples use in this study and questionnaire design are then presented in Section 4. The results are illustrated and discussed in Section 5. Finally, conclusions are drawn in Section 6.

2 Literature Review

2.1 Repeat Visitation

In the marketing literature, repeat purchase is usually referred to as customers' loyalty. The degree of loyalty is one of the significant indexes, which was used to measure the success of marketing strategy [10]. Fornell and Wernerfelt [11] indicated that discovering new customers costs four or five times than maintaining existing customers. Thus, maintaining existing customers plays a crucial role in stabilizing the long-term profits of an organization [10]. Similarly, travel destinations can be regarded as products. Hence, maintaining this group of travelers can not only gain the long-term profits for a destination but reduce the cost.

2.2 Tourist Satisfaction and Repeat Visitation

Satisfaction has been a crucial decision factor in planning tourism products and service. Satisfied tourists are important to successful destination because it influences the decision of destination, the consumption of products and service and the intention of revisiting [12].

Some scholars discovered that satisfied customers would increase the degree of loyalty; satisfied customers would rise the degree of intention of re-purchasing [13]. Customers would feel some degree of satisfaction or dissatisfaction after purchasing. If customers have been highly satisfied, they may have a higher intention of re-purchasing or recommendation; otherwise, they do not have an intention of re-purchasing or recommendation [14]. According to the literature in [15], satisfaction can directly affect the intention of re-purchasing.

2.3 Tourist Satisfaction

Generally, customers would have some expectation about the products and service before purchasing [16]. Similarly, tourists are like customers. They also have expectation about the destination or service quality. If the experience is higher/lower than expectation, they will feel satisfied/dissatisfied [17]. Thus, satisfaction basically is used to measure the quality of products, work, travel and so on. It is a useful and widely-used index.

The common ways to measure satisfaction are Overall Satisfaction, which measures the overall result after using products or experiencing services and Separated Satisfaction, which measures each characteristics after using products or experiencing services [16]. In this paper, we mainly focused on tourists' Overall Satisfaction. Hence, the simple scale was used.

2.4 Travel Motivation

Motivation has been referred to as psychological needs and wants [18]. According to Uysal and Hagan [19], motivations can drive an individual toward some types of action. On the other hand, people travel or join different kinds of activities in order to satisfy different desires and wants [20].

Thomas [21] suggested that there are eighteen travel motivations such as education and culture, rest and relaxation, tradition and others. McIntosh et al. [22] divided travel motivations into four categories: physical motivators, culture motivators, interpersonal and status and prestige motivators. Physical motivators include rest, sporting activities, beach activities and the activities related to health. Culture motivations contain music, art, folkways, dancing, painting and religion. Interpersonal motivations consist of meeting new friends, visiting relatives and escaping daily work. Status and prestige motivations contain respected by others or future development.

In tourism research, the concept of motivation can be classified into two forces: push motivation and pull motivation, which indicate that people are pushed to make travel decisions and pulled by attraction of the destination [23, 24]. Push motivations are related to internal desire such as escape, rest, relaxation, adventure, social interaction, family togetherness and excitement [23, 25, 26]. On the other hand, pull motivations are connected to external aspects, which include recreation facilities, culture and races, entertainment, natural scenery, shopping and so on [27].

In summary, Travel motivations are different according to the characteristic and object. Thus, In this paper, we try to make the research about travel motivations of travelers. We consider the characteristics of Sun Moon Lake into our travel motivation and we draw some suitable travel motivations which match on our study through the literature.

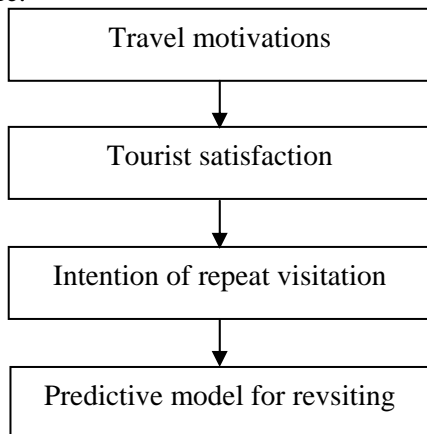


Fig. 1 Proposed Model for Revisiting

Based on the literatures above, we proposed a predictive model as Figure 4 illustrates. The model presented in Figure 4 shows how tourist motivations might impact on satisfaction factors that in turn affect repeat visitation. By constructing a predictive model for repeat visitors, we could predict if the travelers would come back.

3 Support Vector Machines (SVM)

SVM was developed by Vapnik [28]. It is a newly developed technique in recent years. It is one of the best tools to be used in classification. SVM can separate negative samples from the set in which contains both positive and negative samples with complex distribution. When the test data and the training data are similar, the result for classification is usually good. SVM is developed from statistical learning theory [29]. The main idea of SVM comes from the binary classification, namely to find a hyper-plane as a segmentation of the two types to minimize the classification error. This hyper-plane maximizes the minimum distance between the hyper-plane to the nearest negative and positive points.

In addition, SVM can also solve the problem with linear or non-linear segmentation, as shown in Figure 2 and Figure 3. SVM uses some local information to do the training, then calculates some support vectors from the training data to support the whole information while eliminates some outlier points. SVM has some attractive properties that make it a very popular technique to use. The SVM has already been successfully used for a wide variety of problems, like pattern recognition, bio-informatics, natural language learning text mining, and more [30, 31].

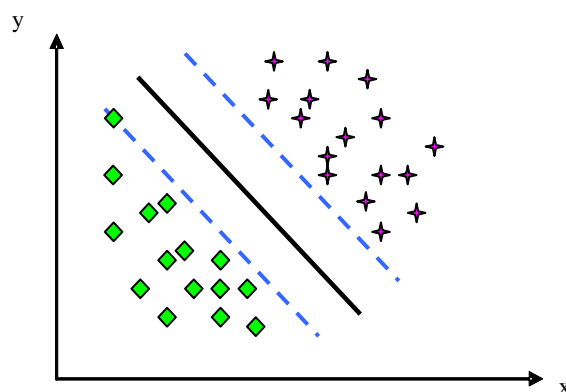


Fig. 2 Linear Classification

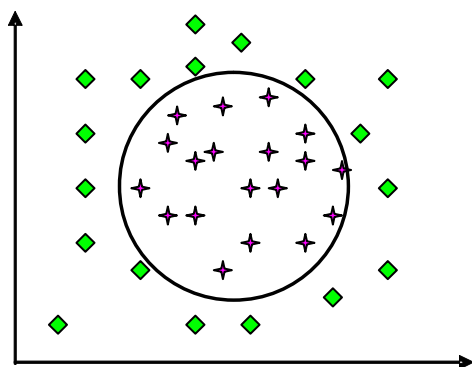


Fig. 3 Non-Linear Classification

4 Samples and Questionnaire Design

Among many ways of collecting data, two of the most popular tools are interviews and questionnaires. Both methods show their pros and cons. The biggest advantages of self-completion questionnaires over structural interviews are that they can save much time and money, as well as permit much larger samples to be gained. Thus, we employed questionnaires to collect data in this study.

A successful questionnaire depends on three elements [36]: (1) The questions must be comprehensible and definite. (2) The appearance of questions must be standardized. (3) A reliable, efficient, and preferably, cost-effective way of coding the data for following analysis must be used. With the three suggestions mentioned above, an efficient and cost-saving online questionnaire system is produced. Before a good final questionnaire is set up, a pilot test was performed to ensure clear questions in our study.

4.1 Data Collection

The samples used in this paper were collected from the tourists of Sun Moon Lake, which is the most famous destination in Taiwan. We have used a systematic sampling approach, where questionnaires were delivered to every 5th tourists that walked by. A total 400 questionnaires were distributed for this study. 380 questionnaires were found usable obtained from travelers who traveled in Sun Moon Lake, yield 95% of respondent rate.

4.2 Questionnaire Design

The questionnaire was developed by reviewing the travel motivation literatures. The questionnaire is divided into three sections. The first section separated into two parts. First, respondents were asked to rate the importance of 24 items of travel motivation to Sun Moon Lake as a travel destination

on a seven-point Likert scale, with 1 being Extremely dissatisfied to 7 being Extremely satisfied. The last question seeks to measure the level of tourists' overall satisfaction. In second section, tourists were asked to express their intention of revisiting Sun Moon Lake. The last section requires the respondents to fill in their demographical information such as gender, age, region of residence and monthly income.

The survey was pre-tested on 105 tourists in Sun Moon Lake and we employed Item Analysis to calculate the critical ratio of each item and see if we need to delete unsuitable items. (If the critical ratio item is less than 0.05, we deleted the item; otherwise, we keep the item.)

Results and Discussion

In this paper, we used Factor Analysis and Support Vector Machine for analyse the data. We employed Factor Analysis to reduce the items to a more manageable size to have a better understanding of the attributes. Five facotrs were then identified labeled Factor 1, "Mental experiment and interpersonal relationship"; Factor 2, "Climate and environment"; Factor 3, "Local food and consumption patterns"; Factor 4, "Activities and services"; Factor 5, "Local product and historical spots".

Table 1 Accuracy of cross validation before FA

Classifier	5-fold	10-fold
dot	0.842	0.821
polynomial degree 2	0.849	0.847
polynomial degree 3	0.847	0.847
radial gamma 0.0001	0.851	0.860
radial gamma 0.001	0.884	0.884

Table 2 Accuracy of cross validation after FA

Classifier	5-fold	10-fold
dot	0.840	0.847
polynomial degree 2	0.849	0.844
polynomial degree 3	0.844	0.844
radial gamma 0.0001	0.858	0.860
radial gamma 0.001	0.876	0.876

At the next stage, we used SVM to train and test the data. Since we did not know the distribution of our data, we tried three kinds of kernels, dot, polynomial and radial, to find the better kernel function and parameters. The degrees of polynomial kernel were 2 and 3. As Table 1 and Table 2 illustrate, we can see there are slight differences between 5 fold and 10 fold cross validation. We also found that the accuracy rates are lower after using Factor Analysis. We then discovered that when kernel is radial, the accuracy was the highest. Hence, the base classifier

we used was the radial kernel. Based on the confusion matrix, we found that the average accuracy is quite good. However, the true negative rates (non-repeat visitors) are quite low.

Due to the skewed data, the prediction of minority part is not well. Therefore, we adopted the majority voting technique to increase the accuracy of true negative (non-repeat visitors). We divided the data, which consists of 330 positive data (repeat visitors) and 50 negative data (non-repeat visitors; 30 samples used for training and 20 for testing) into 11 subsets as the training sets. Each subset includes 30 different positive samples and 30 same negative samples.

We compare the results of three strategies, which are summarized in Table 3. We can see that majority voting by using SVM only, more-than-1-negative rule, weighting by cross validation and voting can improve the prediction of non-repeat visitors. As Table 4 depicts, we are able to discover that majority voting, all the strategies or rules can improve the prediction of negative data. To sum up, the average accuracy was improved.

Table 3 TN rate using different method before FA

Strategy or rule	TN rate
base case	0.650
majority voting (more than 2 negative)	0.600
more than 1 negative	0.750
weighting by 5-fold cross validation & voting	0.689
weighting by TN rate & voting	0.629

Table 4 TN rate using different method after FA

Strategy or rule	TN rate
base case	0.350
majority voting (more than 2 negative)	0.400
more than 1 negative	0.600
weighting by 5-fold cross validation & voting	0.696
weighting by TN rate & voting	0.453

6 Conclusions

In an increasingly saturated marketplace, repeat visitors become a backbone of the destination. This group of tourists denotes a stabilizing influence on most destinations because of their familiarity with the destination and the stable income source they provide. This means maintaining existing tourists can reduce the cost of a destination and create long-term profits.

As a results, we identified five significant factors, labeled Factor 1, “Mental experiment and inter personal relationship”; Factor 2, “Climate and environment”; Factor 3, “Local food and consumption patterns”; Factor 4, “Activities and services”; Factor 5, “Local product and historical spots”. Based on the experimental results, the average accuracy rate of base case was quite high. However, we discovered that this model cannot provide a high prediction of negative data (non-repeat visitors) because of imbalanced data distribution. After using other strategies or rules, we found that the true negative rate was improved.

The limitation should be addressed. Travel motivations would be changed by people and destination. Tourists’ actions would also be diverse while the destinations are different. This study only investigated a local tourism destination, Sun Moon Lake. The results may not be suitable for other tourism destinations. In the future, more studies are encouraged to investigate more representative tourism destinations.

Acknowledgments

The authors wish to express their appreciation to Yin-Hua Chen, Wen-Feng Hsieh, and Hsin-Jie Liu for their help during the course of this paper.

References:

- [1] www.world-tourism.org/
- [2] www.wttc.org/
- [3] Dawkins, P., And Reichheld, F., Customer Retention a Competitive Weapon, *Directors and Broads*, Vol.14, No.4, 1990, pp. 42-47.
- [4] Law, R., and Au, N., A Neural Network Model to Forecast Japanese Demand for Travel to Hong Kong, *Tourism Management*, Vol.20, 1999, pp. 89-97.
- [5] Law, R., and Au, N., Relationship Modeling in Tourism Shopping: a Decision Rules Induction Approach, *Tourism Management*, Vol.21, 2000, pp. 241-249.
- [6] Chen, K.Y., Ho, C.H., and Shiau, H.C., Application of Support Vector Regression in Forecasting International Tourism Demand, *Tourism Management Research*, Vol.4, No.1, 2004, pp. 81-97.
- [7] Gold, S., Nouns of Neighborhood Park, *Journal of American Institute of Planners*, Vol.38, 1972, pp. 369-378.
- [8] Stevens, P., Knutson, B., and Pattin, M., Dineserv: a Tool for Measuring Service Quality in Restaurants, *The Cornell Hotel and*

- Restaurant Administration Quarterly*, Vol.36, 1995, pp. 56-60.
- [9] Lau, A.L., and Mckercher, B., Exploration Versus Acquisition: a Comparison of First Time and Repeat Visitors, *Journal of Travel Research*, Vol.42, 2004, pp. 279-285.
- [10] Flavian, C., Martinez, E., and Polo, Y., Loyalty to Grocery Stores in the Spanish Market of the 1990s, *Journal of Retailing and Consumer Service*, Vol.8, 2001, pp. 85-93.
- [11] Fornell, C., and Wernerfelt, The American Customer Complaint Management, *Marketing Science*, Vol.7, 1992, pp. 287-298.
- [12] Kozak, M., and Rimmington, M., Tourism Satisfaction with Mallorca, Spain, as an Off-season Holiday Destination, *Journal of Travel Research*, Vol.38, No.3, 2000, pp.260-269.
- [13] Reichheld, F.F., and Sasser, W.E., Zero Defections: Quality Comes to Services. *Harvard Business Review*, Vol.68, 1990, pp.105-110.
- [14] Fornell, C., Johnson, M.D., Anderson, E.W., Cha, J., and Bryant B.E., The American Customer Satisfaction Index: Nature, Purpose and Findings, *Journal of Marketing*, Vol.60, 1996, pp. 1-13.
- [15] Anderson, E.W., and Sullivan, M.W., The Antecedents and Consequences of Customer Satisfaction for Firms, *Marketing Science*,
- [16] Fornell, C., A National Customer Satisfaction Barometer: The Swedish Experience, *Journal of Marketing*, Vol.56, 1992, pp. 6-16.
- [17] Kotler, P., *Marketing Management: Analysis, Planning, Implementation and Control*, New York: Prentice Hall, 1994
- [18] Bromley, D.B., *Behavioral Gerontology: Central Issues in the Psychology of Ageing*, England Wiley, 1990
- [19] Uysal, M., and Hagan, R., *Motivation of Pleasure Travel and Tourism*, *Encyclopedia of Hospitality and Tourism*, New York Nostrand Reinhold, pp. 798-810, 1993
- [20] Manning, R.E., Diversity in a Democracy, Expanding the Recreation Opportunity Spectrum, *Leisure Science*, Vol.7, No.4, 1985, pp.377-399.
- [21] Thosmas, J.A., What Makes People Travel, *ASTA Travel News*, Vol.1, 1964, pp.169-172.
- [22] McIntosh, R., Goeldner, C., and Ritchie, B., *Tourism: Principles, Practices, Philosophies*, New York: John Wiley, 1994
- [23] Dann, G.M., Anomie Ego-enhancement and *Tourism*, *Annals of Tourism Research*, Vol.6, 1977, pp.184-194.
- [24] Dann, G.M., Tourism Motivations: An Appraisal, *Annals of Tourism Research*, Vol.4, 1981, pp. 189-219.
- [25] Crompton, K.L., Motivations of Pleasure Vacation, *Annals of Tourism Research*, Vol.6, 1979, pp. 409-424.
- [26] Klenosky, D.B., The Pull of Tourism Destination: A Means-end Investigation, *Journal of Travel Research*, Vol.40, 2002, pp.385-395.
- [27] McGehee, N.G., Loker-Murphy, L., and Uysal, M., The Australian International Pleasure Travel Market: Motivations From a Gendered Perspective, *The Journal of Tourism Studies*, Vol.7, No.1, 1996, pp. 45-57.
- [28] Vapnik, V.N., *The Nature of Statistical Learning Theory*, Springer, 1995
- [29] <http://www.kernel-machines.org/>
- [30] C.J.C, Burges, A Tutorial on Support Vector Machines for Pattern Recognition, *Data Mining and Knowledge Discovery*, Vol. 2, No.2, 1998, pp. 955-974.
- [31] R.C. Chen, J. Chen, T.S. Chen, C.H. Hsieh, T.Y. Chen, and K.Y. Wu, Building an Intrusion Detection System Based on Support Vector Machine and Genetic Algorithm, *Lecture Notes in Computer Science*, Vol. 3498, 2005, pp. 409-414.