

## Support Vector Machines with Manifold Learning and Probabilistic Space Projection for Tourist Expenditure Analysis

**Xin Xu**

*Institute of Automation, National University of Defense Technology,  
410073, Changsha, P.R.China (Email: xuxin\_mail@263.net)*

**Rob Law**

*School of Hotel & Tourism Management, The Hong Kong Polytechnic University,  
Hung Hom, Kowloon, Hong Kong*

**Tao Wu**

*Institute of Automation, National University of Defense Technology,  
410073, Changsha, P.R.China*

Received: 14-06-2008

Revised: 03-02-2009

### Abstract

The significant economic contributions of the tourism industry in recent years impose an unprecedented force for data mining and machine learning methods to analyze tourism data. The intrinsic problems of raw data in tourism are largely related to the complexity, noise and nonlinearity in the data that may introduce many challenges for the existing data mining techniques such as rough sets and neural networks. In this paper, a novel method using SVM-based classification with two nonlinear feature projection techniques is proposed for tourism data analysis. The first feature projection method is based on ISOMAP (Isometric Feature Mapping), which is a class of manifold learning approaches for dimension reduction. By making use of ISOMAP, part of the noisy data can be identified and the classification accuracy of SVMs can be improved by appropriately discarding the noisy training data. The second feature projection method is a probabilistic space mapping technique for scale transformation. Experimental results on expenditure data of business travelers show that the proposed method can improve prediction performance both in terms of testing accuracy and statistical coincidence. In addition, both of the feature projection methods are helpful to reduce the training time of SVMs.

*Keywords:* SVMs, Tourism data analysis, Manifold learning, ISOMAP, Feature projection, Scale transformation, Data mining

### 1. Introduction

In recent years, empirical findings of various studies have shown the economic contributions of international travelers in general, and business travelers in particular, to a destination [1, 2]. For example, the tourism figures [3, 4] of Hong Kong, a Special Administrative Region (HKSAR) of China, highlight the importance of the business traveler market and its contribution to the economy. In particular, the high-yield business traveler segment has attracted the attention of policy makers, practitioners, and, to a much lesser extent, academic

researchers in different academic disciplines including researchers in data mining. Still, few attempts, if any, have been made to understand the behavioral patterns of business travelers. The statistical data from the Hong Kong Tourism Board (HKTB) only showed the number and percentages of business visitors from different major source markets but very few efforts have been made to mine the profile of these business travelers [3]. In the existing data mining and tourism literature, the study of data mining for business travelers has been largely overlooked by academic researchers. In other words, the demographic and trip profiles of this high

yield group of visitors remain largely unknown. In a recent study [5], a model based on rough sets was developed to extract useful information and knowledge rules from business travelers to Hong Kong using primary tourist expenditure data collected in a survey. Although the work in [5] provided very promising results for data mining applications in business data analysis, the classification accuracy still needs to be improved to meet the requirements of decision makers. The main difficulties for improving the classification accuracy of tourist expenditures are in several aspects. One aspect is the complex and nonlinear relationship between attributes of tourist behaviors such as mode of travel and region of residence. Another aspect is due to unknown noises in the survey data which were collected under different situations. Therefore, the classification and prediction analysis of tourist profiles provides new challenges for data mining and machine learning techniques.

In general, there are two important performance criteria when dealing with the classification and prediction problems of business tourist profiles. The first criterion is the accuracy of classification which is popularly adopted in many data mining and pattern recognition applications. The second is the statistical coincidence between the estimated values and the actual data, which has been investigated in [5] for business traveler's profiles. Currently, existing approaches to tourism data forecasting, including rough sets, neural networks [12], etc., can not achieve good performance both in terms of the above two criteria. For example, the prediction accuracy of rough-set-based approaches is no more than 40% and the probability of statistical coincidence is no more than 0.4. As a recently developed machine learning method, the Support Vector Machine (SVM) has obtained the state-of-the-art results in many applications. Apparently, it is promising to apply SVMs to classification and prediction of business tourist profiles, which has been studied in our recent work [6]. Nevertheless, the performance of SVMs is still influenced by the selected features and kernel functions [4]. To guarantee good performance of SVMs, it is still necessary to select data features appropriately for different applications. Although, kernel functions in SVMs can be used for feature selection, it is well known that the kernel trick in SVMs is a kind of subjective technique or a kind of "art". As such, to obtain a good

result, it is necessary to choose a suitable kernel function and a suitable set of kernel parameters. However, there are few systematic and general approaches to the problem of kernel function selection.

In this paper, by incorporating a manifold learning method as well as a probability space projection technique into the SVM-based classification method, a novel data mining method for tourism data analysis is proposed. The manifold learning algorithm is called ISOMAP (Isometric Feature Mapping) [7], which was firstly used to visualize the distribution of high-dimensional data in a 2-dimensional space. By making use of ISOMAP, it was shown that some of the training data, which belong to different classes, are absolutely mixed. Therefore, it can be concluded that one of the difficulties for classifying tourist expenditure data is due to the large noises in data collections. Since the mixed training data will make it difficult to train the classifiers, it will be beneficial to discard some of these mixed training data using feature projection based on ISOMAP. Next, the probability space projection technique is presented to perform scale transformation of tourism data in order to further improve the performance of SVM-based classifiers. Thus, the proposed method improves the performance of conventional SVM classifiers in two folds. One is the manifold learning technique for data selection and noise reduction. The other is the scale transformation based on a probability space mapping. Experimental results on tourism expenditure data in Hong Kong demonstrate that the proposed approach has superior performance than previous methods.

Having introduced the research background in this section, the remaining part of this paper is organized as follows. In Section 2, a brief introduction of the classification problem for tourist expenditure data, as well as some related works, is given. The data analysis method using SVMs with manifold learning and probability space mapping is proposed in Section 3. Section 4 gives the performance comparisons among different approaches. Lastly, Section 5 concludes this research.

## 2. Problem description and related work

The data that were analyzed in this paper is a set of survey data from tourists who have recently visited

Hong Kong. The data were collected in an Omnibus Survey that was conducted by the School of Hotel & Tourism Management at the Hong Kong Polytechnic University in October 2005. During this period, a total of 1,282 non-transit visitors from seven major tourist generating regions were interviewed face-to-face in the restricted departure lounge of the Hong Kong International Airport. The regions included Mainland China, Taiwan, Singapore, Malaysia, the United States, Australia, and Western Europe. A quota sampling method was used by a team of nine interviewers. Upon the completion of an interview, a souvenir was presented to the respondent as a token of appreciation for their participation.

Following the practice of the Omnibus Survey [8], the questionnaire was developed in English and then translated into Chinese. The final version of the questionnaire was pilot-tested in September 2005 to ensure that questions were clearly understood. The

questionnaire comprised a common set of questions for demographic and trip profiles. These demographic and trip profile data were utilized in this research for classification of business travelers. Among the 1,282 respondents, 303 identified themselves as business travelers and provided usable demographic and trip profile data. The statistics of these data is shown in the following Table 1. All attributes (variables) in Table 1 were grouped as condition attributes whereas expenses per night excluding accommodation and airfares were used as the decision attributes. An equal percentile approach was adopted that split the values in the decision attribute into three categories of High (H), Medium (M), and Low (L). Based on the above data, the tourist data analysis problem becomes a pattern classification task. The input attributes include various behavior attributes of the tourists, e.g., Gender, Mode of travel, Length of stay, etc., and the decision variable is the three types of expenditure amounts, i.e., H, M, and L.

Table 1 - Demographic Information (N=303)

	Frequency	(%)
<b>Gender</b>		
Male	210	69.3
Female	93	30.7
<b>Mode of travel</b>		
Fully packaged tour	12	4.0
Partially packaged tour with transport and accommodation only	74	24.4
Non-packaged/independent tour	217	71.6
<b>Length of Stay in Hong Kong (Recode)</b>		
2 night or less	101	33.3
3 - 5 nights	137	45.2
6 - 8 nights	38	12.5
9 nights or more	27	8.9
<b>Total travel party in a group (Recode)</b>		
1 person	193	63.7
2 persons	64	21.1
3 persons or more	46	15.2
<b>First visit to Hong Kong</b>		
Yes	93	30.7
No	210	69.3
<b>Is Hong Kong the only destination you have visited during this trip?</b>		
Yes	136	44.9
No	167	55.1
<b>Annual Household Income</b>		
Less than US\$10, 000	12	4.0
US\$10,000 -29,999	40	13.2
US\$30,000 - 49,999	53	17.5

US\$50,000 - 69,999	46	15.2
US\$70,000 - 99,999	46	15.2
US\$100,000 or more	106	35.0
<b>Region of Residence</b>		
Mainland China	64	21.1
Taiwan	21	6.9
Singapore	24	7.9
Malaysia	22	7.3
United States	90	29.7
Australia	23	7.6
Western Europe	59	19.5

Due to the significant economic contributions from tourists, data mining techniques for understanding the behavioral and demographic patterns of tourists have recently received increasing research interests. However, the complexity, noise and nonlinearity in raw tourism data introduce many challenges for existing data mining techniques. Prior studies have focused on BP (back-propagation) neural network models [11-12] and rough sets [5] but their performance is far from satisfactory levels. To construct classification and regression models with high accuracy and good generalization ability, support vector machines (SVMs), which were originated from the research work on statistical learning theory and kernel learning machines [9-10], have received lots of research interests in the past decade. Nowadays, there have been many successful applications of SVMs in pattern recognition and data mining problems, where SVM-based models frequently received state-of-the-art results. However, the applications of SVMs to tourism data analysis have only received very little attention in the existing literature.

Recently, SVM-based classification and regression models were studied for tourism demand analysis [13]. However, as a class of popular kernel-based learning machines, the performance of SVM models still largely relies on the selection of feature representations, which remains as an open problem in the literature. In a recently conducted study by the authors, a multi-class SVM model with kernel principal component analysis (KPCA) was proposed to solve the problem of tourist expenditure analysis [6]. Although very promising results have been obtained in our previous study, classifying tourist expenditure data is still one of most challenging problems for SVMs as there are many

corrupted data in the training samples. It is, therefore, necessary and beneficial to study new feature projection methods for data selection and noise reduction. In this paper, SVMs will be integrated with manifold learning and a probability space projection technique so that the noisy data can be identified and discarded appropriately. In addition, the SVM-based classifier can be constructed in a more suitable feature space for better performance. In the following section, we will present the new method in detail.

### 3. ISOMAP and Probabilistic Space Projection for SVMs

This section presents the approach to solve the difficulties in tourist expenditure classification, i.e., noisy data samples and complex features. Two nonlinear feature projection methods will be integrated with multi-class SVMs for better performance. In sub-section 3.1, the ISOMAP algorithm, which is a class of manifold learning methods, is used to realize data selection and noise reduction. After that, in order to solve the problem of data features in different measurement spaces, a probability space projection method will be presented for scale transformation so that the classification performance of tourist expenditure data can be improved further.

#### 3.1. Data selection and noise reduction using manifold learning

In recent years, there have been a number of new methods studied for describing the structure of nonlinear manifolds that reside within high-dimensional data. These methods have become generally known as manifold learning algorithms. Examples of manifold

learning algorithms include the Isometric Feature Mapping (ISOMAP) algorithm [7] and the locally linear embedding (LLE) algorithms [14]. These methods have been used by different researchers in a variety of fields. By making use of the ISOMAP algorithm to perform dimension reduction, the global geometric structure of high-dimensional data can be represented in a low-dimensional manifold. Nevertheless, there has been little work on the application of manifold learning for tourist data analysis. The motivation for applying manifold learning methods to the forecasting problem of tourist expenditures includes the following two aspects. One aspect is the modeling of the complex and nonlinear relationships between the attributes of tourist behaviors. The second aspect is to reduce the unknown noises in the tourism demand data which were collected under different situations. These two aspects may be different from other data mining applications and it will

be beneficial to apply manifold learning techniques in the task of tourism demand forecasting.

In this research, to expose the intrinsic distribution of the tourist expenditure data, we propose to make use of manifold learning methods in general, and the ISOMAP algorithm in particular, to perform feature projection. As is well known, manifold learning techniques have been shown to be very efficient for dimensionality reduction in high dimensional feature spaces. Although the number of dimensions of tourist expenditure data is not very high, it is impossible to imagine the distribution of these data in the original feature space. By making use of ISOMAP, the distribution of tourism expenditure data can be visualized in a two-dimensional space and the noisy data can be easily identified.

**The ISOMAP Algorithm**

**Input:**  $X = \{x_1, x_2, \dots, x_n \in \mathbb{R}^N\}$ , neighborhood parameter  $k$ , the dimension  $d$  of the feature space that  $X$  will be embedded into;

**Output:**  $Y = \{y_1, y_2, \dots, y_n \in \mathbb{R}^d\}$

**Step 1.** Identify the  $k$  nearest neighbors of each input data point and construct a neighborhood graph where edge lengths between the points in a neighborhood are set as their Euclidean distances.

**Step 2.** Compute approximate geodesic distances,  $d_{ij}$ , containing shortest paths between all pairs of points and define the shortest paths matrix as

$$D_G = [d_{ij}^2] \in \mathbb{R}^{N \times N}$$

**Step 3.** Let  $\tau(D_G) = -\frac{1}{2} H \cdot D_G \cdot H$ , where  $H$  is the centering matrix, given

by  $H = (h_{ij}) = I - \frac{1}{N} e \cdot e^T$  and  $e = [1, \dots, 1]^T \in \mathbb{R}^N$ . Compute the top  $d$

eigenvectors vector matrix  $V \in \mathbb{R}^{N \times d}$  and the eigen-value matrix  $\Lambda \in \mathbb{R}^{d \times d}$

**Step 4.** The coordinates of the embedded data in the  $d$ -dimensional Euclidean space are given by  $Y = \Lambda^{-\frac{1}{2}} V^T$ .

Fig.1 The ISOMAP algorithm

As studied in [7], ISOMAP is one important method to embed a high-dimension manifold into a low-dimensional manifold. The basic idea of the ISOMAP algorithm is to use the Euclidean distance between two

points in the low-dimensional space to approximate the geodesic distance of that in the high-dimensional space. The ISOMAP algorithm used in our study can be described in Fig.1.

By using the ISOMAP method, the tourist expenditure data were embedded into a two-dimensional plane, which is shown in Fig.2. Since there are three classes in the data, we use three different colors and shapes to label the training data. As indicated in Fig.2, many data belonging to different classes are mixed and nearly superposed. Apparently, these superposed training data make the decision function too complex while in conventional SVM-based classification, the aim is to approximate a smooth classifier. Superposed data also make the process of solving SVM-based solutions to be unstable.

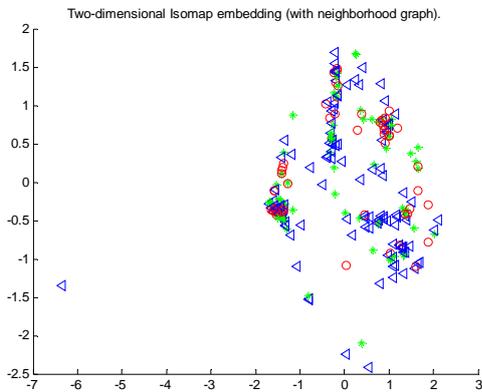


Fig. 2 Two-dimensional ISOMAP embedding of tourist expenditure data

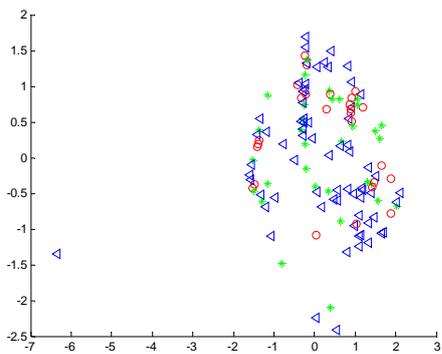


Fig. 3 Two-dimensional ISOMAP embedding of the residual training data after data selection

From Fig.2, the superposed training data can be found in the two-dimensional feature space. Since it is difficult to determine the label correctness of the superposed data, it will be beneficial to discard all the superposed data

from the training data set according to their geodesic distance in the input space or the Euclidean distance in the feature space after using the ISOMAP method. As shown in Fig.3, the data selection process based on ISOMAP can discard most of the noisy training data and the remaining training set has much better distribution of different classes. In the following experiments, it will be demonstrated that the data selection process will contribute to improve the performance of SVM-based classification.

### 3.2. Probability space projection for SVM-based classification

In addition to noise reduction and data selection, the analysis of tourist profiles presents more research challenges. Among these challenges, the problem of different measurement scales in different features is a particular one. For any pattern recognition methods, distance measure or similarity measure is a key to solve these problems. However, features with different scales will make it very hard to define an appropriate distance measure. In the tourism expenditure data, some features are discrete variables and others are continuous variables. Therefore, there will be different distance measurements among different features and it is necessary to find a uniform representation of these distance measures.

For example, one record of the training data to be analyzed is listed as follows:

$$[a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8] = [1/2, +1, 0.18, 0.10, 0.50, -1, 0.66, 0.71]$$

where feature 1, feature 2, feature 5 and feature 6 are discrete variables, and feature 3, feature 4, feature 7 and feature 8 are continuous variables. Obviously, 0.5 in feature 1 and 0.5 in feature 4 should have different meanings. However, traditional distance (similarity) measures cannot distinguish them. In this section, a scale transformation method will be presented to transform the feature values to a probability measurement, which can be called ‘Probability Transformation’ or PT.

The basic idea of PT is to use posterior probabilities to substitute the original feature values. It can formally be described as follows.

Suppose the original feature vector is  $x=(v_1, v_2, \dots, v_n)$ , then the transformed feature vector is as follows.

$$x' = [\bar{v}_1^m, \bar{v}_2^m, \dots, \bar{v}_n^m] \quad (1)$$

Here,  $n$  is the original feature number,  $m$  is the total number of classes,  $\bar{v}_i^m = [v_i^1, v_i^2, \dots, v_i^m]$ , which means that one feature value of the original feature vector will be transformed into  $m$  feature values corresponding to  $m$  posterior probabilities of  $m$  classes. The posterior probability of class  $j$  can be computed as follows:

$$v_i^j = P(Y = j | x(i) = v_i) = \frac{P(x(i) = v_i | Y = j) P(Y = j)}{P(x(i) = v_i)}, \quad i=1,2,\dots,n \quad j=1,2,\dots,m \quad (2)$$

where  $x(i)$  is the  $i$ th element of the original feature vector,  $Y$  is the class label.

Here,  $P(Y = j)$  is the prior probability which can be computed from the training data.  $P(x = v_i | Y = j)$  is the likelihood probability of the event that  $x=v_i$  under the condition of class label being  $j$ . If  $x$  is a discrete variable, one can directly compute the prior probability and the likelihood function from the training data. If  $x$  is a continuous variable, there will be two possible methods to compute  $P(x = v_i | Y = j)$  and  $P(x = v_i)$ . One is to divide the variation interval of the continuous feature into several discrete parts so that the continuous variable is transformed to a discrete variable. The other way is to use non-parameter probability estimation methods such as kernel regression to estimate the probabilities. In this paper, since there are not enough data to use non-parametric methods to estimate the prior probability, the discretization method is used.

Based on the above discussion, the original data features with different measurement scales can be transformed a uniform representation in terms of posterior probabilities of different classes. Then, the distance function computed by different classifiers will become more appropriate for classification and this will be illustrated in the following experiments when combined with SVM-based classifiers.

### 3.3. SVM-based classification for tourist data analysis

Based on the structural risk minimization (SRM) principle in statistical learning theory, SVMs are originally proposed for binary classification problems. Nevertheless, most real world pattern recognition applications are multi-class classification cases. Thus, multi-class SVM algorithms have received much attention over the last decade and several

decomposition-based approaches have been proposed in recent years [16]. The idea of decomposition-based methods is to divide a multi-class problem into multiple binary problems, which is to construct multiple standard two-class SVM classifiers and fuse their classification results. There are several strategies for the implementation of multi-class SVMs using binary algorithms, with examples of one-vs-all, one-vs-one, and error correcting output coding [16].

To solve the analysis problem of tourist expenditure, we will use the one-vs-all strategy for the multi-class SVMs, which was studied in our previous work [6] and integrated with the KPCA method for feature extraction so that better generalization ability can be obtained.

For multi-class SVMs based on the one-vs-all strategy, every binary classifier separates the data of one class from the data of all the other classes. In the training of binary SVM classifiers, a hyperplane is considered to separate two classes of samples. Based on the SRM principle, the optimal separating hyperplane can be constructed by the following optimization problem

$$\min_{\bar{w}, b} \frac{1}{2} \|\bar{w}\|^2 \quad (3)$$

subject to

$$y_i(\bar{w} \cdot \bar{x}_i + b) \geq 1, \quad i = 1, 2, \dots, N \quad (4)$$

where  $(\bar{x}_i, y_i)$  are training samples and  $y = \bar{w} \cdot \bar{x}_i + b$  is the hyperplane.

As previously discussed, in most real-world classification problems, nonlinear separating planes have to be constructed, the 'kernel trick' is thus used to transform the above linear form of support vector learning algorithm to a nonlinear one. In the kernel trick, a nonlinear feature mapping is introduced to build linear hyper-plane in the kernel-induced feature space without explicitly computing the inner products in high-dimensional spaces. Let the nonlinear feature mapping be denoted as

$$\bar{x} \rightarrow \psi(\bar{x}) \quad (5)$$

Then, the dot products  $(\bar{x}_i \cdot \bar{x}_j)$  in linear SVM algorithms can be replaced by dot products  $(\psi(\bar{x}_i) \cdot \psi(\bar{x}_j))$  in the nonlinear feature space and a

Mercer kernel function can be used to express the dot products in high-dimensional feature space

$$k(\vec{x}_i, \vec{x}_j) = (\psi(\vec{x}_i) \cdot \psi(\vec{x}_j)) \quad (6)$$

If the kernel function can be selected appropriately, the optimization problem of SVMs for two-class soft margin classifiers can be formulated as follows

$$\max_{\alpha} \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j k(\vec{x}_i \cdot \vec{x}_j) \quad (7)$$

subject to

$$0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, N \quad \text{and} \quad \sum_{i=1}^N \alpha_i y_i = 0 \quad (8)$$

To solve the above quadratic programming problem, various decomposition-based fast algorithms have been proposed, such as the SMO algorithm [15]. For details on the algorithmic implementation of SVMs, please refer to [15].

After constructing the binary SVM classifiers, the decision function of each binary SVM is

process on the tourism expenditure data was implemented in our experiments. That is, the business traveler profiles as presented in Section 2 were randomly divided into two parts independently for five times. During each division, by randomly selecting 80% (N=243) of the cases for model calibration and the remaining ones (N=60) for model testing, the accuracies of different classification methods were determined in two quality terms of percentage of correctly classified cases (the estimated value matches the actual value) and the training time on different data sets. Therefore, the performance of each data analysis method can be evaluated on five pairs of training and testing data sets. In the following paragraphs, we will denote the five pairs of training and testing data sets as HK-i, (i=1,2,...,5).

The multi-class SVMs with and without data selection using ISOMAP-based feature projections were firstly evaluated on the data sets HK-i, (i=1,2,...,5). Experimental results are presented in Table 2. In these experiments, the kernel function of SVMs was selected as Gaussian kernels and the sigma value in Gaussian

Table 2. Performance comparisons of SVMs with and without ISOMAP-based data selection

		Hk-1	HK-2	HK-3	HK-4	Hk-5
Correct Rate	SVMs	50%	50%	55%	41.67%	53.33%
	SVMs+ISOMAP	<b>50%</b>	<b>53.3%</b>	<b>55%</b>	<b>48.33%</b>	31.67%
Training time(s)	SVMs	21.39	14.7031	65.5938	12.7969	72.29
	SVMs+ISOMAP	<b>14</b>	<b>13.42</b>	<b>4.2</b>	<b>3.05</b>	<b>2.11</b>

$$f_k(\vec{x}) = \text{sgn}(\sum_{i=1}^N \alpha_{ki} y_{ki} k(\vec{x}_{ki}, \vec{x}) + b_k) \quad k = 1, 2, \dots, m \quad (9)$$

where  $f_k(\vec{x})$  is the decision function of classifier k and  $(\vec{x}_{ki}, y_{ki})$  ( $k=1, 2, \dots, m$ ) are corresponding training samples. The output of the whole multi-class SVMs can be determined by the decision outputs of every binary classifier, where various voting strategies can be used.

#### 4. Performance comparisons

To test the performance of the above data analysis method using SVMs with manifold learning and probabilistic projection, a 5-fold hold-out validation

kernel was chosen as 10, which is almost the optimal value. From the results in Table 2, it is shown that for SVMs without data selections, not only the classification precision is not satisfactory but also the training time is too long for the 250 training data. The SVM tool that was used is a Matlab toolbox designed by S. Canu (named “svm\_km”) [17]. In this toolbox, the algorithm of training SVMs was especially improved such that classifying 250 training data should not cost more than 10 seconds. But in our experiments, the average training time is 37.35 seconds. The reasons for the prolonged training speed mainly come from the noisy training data which has been shown by the ISOMAP-based feature projection. Therefore, it will be

necessary to reduce the noises in the training data by the data selection strategy using ISOMAP. In Table 2, it is shown that the classification results using SVMs with ISOMAP-based data selection are usually better than SVMs without data selection. It is illustrated that discarding some corrupted training data could not only improve the correct rate but also reduce the training time.

In Table 2, one exception is the result on HK-5, where a lower correct rate was obtained after discarding some training data. But after analyzing the testing results, it was found that almost all the errors were due to misclassifying data from class 1 as class 2 or judge 2 as 1, and only 7 errors were found to misclassify class 1 as class 3. On the contrary, for SVMs without data selection, 19 errors were found for misclassifying class 1 as 3 which will lead to a larger MAPE (Mean Absolute Percentage Error).

From the above Table 2, it can be concluded that by making use of the data selection strategy based on ISOMAP, the classification performance of multi-class SVMs can be improved. Furthermore, it will be demonstrated that the probability projection approach presented above can also contribute performance enhancement for SVM-based classifier. Table 3 makes performance comparisons of different classification methods, which include SVMs with different feature

extraction or projection techniques such as the probability transformation method presented in this paper, Principal Component Analysis (PCA) and kernel PCA. In Table 3, the testing accuracies of the five different methods are listed. It is illustrated that the proposed SVMs with probability transformation (PT) can obtain the highest test precision among all the methods. Although SVMs with KPCA can also obtain this result, it will be shown that the statistical coincidence in the results of SVMs with PT can be better than all the other approaches.

Table 3. Performance Comparison among different classifiers

Classification Methods	Test Precision
Rough Sets	33%
SVM with original features	51.4%
SVM with PCA	56.7%
SVM with KPCA	58.3%
SVM with PT features	58.3%

Table 4 lists the results of conducting Wilcoxon Signed Ranks tests on the real outputs and the estimated outputs using different classification methods. From the results of Wilcoxon Signed Rank tests, it is shown that

Table 4. Wilcoxon Signed Ranks Tests

	Estimated - Actual				
	SVM+PT	SVM+KPCA	SVM+PCA	SVM	Rough Sets
Z	-0.334 <sup>a</sup>	-0.209 <sup>a</sup>	-2.554	-1.247 <sup>a</sup>	-0.807 <sup>a</sup>
Asymp. Sig. (2-tailed)	0.738	0.835	0.011	0.213	0.420

a Based on negative ranks.

Table 5. Wilcoxon Signed Ranks Tests for different SVM parameters with PT

	Estimated - Actual			
	$\sigma = 0.18$ C=150	$\sigma = 0.175$ C=150	$\sigma = 0.17$ C=150	$\sigma = 0.15$ C=100
Z	-0.334	-0.224	-0.112	-0.071
Asymp. Sig. (2-tailed)	0.738	0.823	0.911	0.943
Correct Rate	58.3%	56.7%	55%	53.3%

the estimated values of SVMs combined with PT have one of the highest probability (0.738) for coming from the same distribution of the actual data, which is similar to that of SVMs with KPCA and much better other methods. Furthermore, Wilcoxon Signed Rank tests on different parameter settings show that SVMs with PT have good abilities to improve the statistical coincidence and it is very easy to obtain the best statistical coincidence at the little expense of accuracy degradation as indicated in Table 5.

## 5. Conclusions

The complexity, noise and nonlinearity in the tourism expenditure data introduce many research challenges for the existing data mining techniques such as rough sets and neural networks. As one of the state-of-the-art methods, the Support Vector Machines (SVMs) are also unable to consistently receive satisfactory results for real-world tourism data. In this paper, two novel feature projection methods are proposed to process the tourist expenditure data before using SVMs so that better classification accuracy and statistical coincidence can be obtained. The first method is based on ISOMAP, which is a class of manifold learning approaches for dimension reduction. By making use of ISOMAP, part of the noisy data can be identified and the classification accuracy of SVMs can be improved by appropriately discarding the noisy training data. The second feature projection method is a probabilistic space mapping technique for scale transformation. Experimental results show that the proposed method is superior to most of the existing methods and it can improve prediction performance both in terms of testing accuracy and statistical coincidence.

## Acknowledgements

The work described in this paper was fully supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. PolyU 4631/06H).

## References

1. B.M. Braun, B. Rungeling. The relative economic impact of convention and tourist on a regional economy: a case study. *International Journal Hospitality Management*, vol. 11(1), (1992) 65–71.
2. Lawson, F.R. Trends in business tourism management. *Tourism Management* vol. 3(4), (1982) 298–302.
3. Hong Kong Tourism Board, *Statistics on Conventions & Exhibitions 2004*. Accessed on February 3, 2006. (2005a) Available online at <http://partnetnet.hktourismboard.com/>
4. Hong Kong Tourism Board, *Visitor Profile Report 2004*. Accessed on February 3, 2006. (2005b) Available online at <http://partnetnet.hktourismboard.com/>
5. Law R., Bauer T., Weber K., Tse T., Towards a rough classification of business travelers. *Lecture Notes in Artificial Intelligence*, vol. 4093, (2006) 135–142
6. X. Xu, R. Law, T. Wu, Classification of business travelers using SVMs combined with kernel principal component analysis, *Advanced data mining and applications*, LNCS, vol. 4632, (2007) 524–532.
7. M. Balasubramanian, E. L. Schwartz, J. B. Tenenbaum, V. de Silva and J. C. Langford. The Isomap algorithm and topological stability. *Science*, vol. 295(5552), (2002) 7a,
8. Hui, E.L.L., McKercher. Operational issues in marketing research: an example of the Omnibus tourism survey. *Pacific Tourism Review*, vol. 5(1/2), (2001) 5–13
9. Schölkopf B., Smola A.: *Learning With Kernels*. (Cambridge, MA: MIT Press, 2002).
10. Vapnik V., *The Nature of Statistical Learning Theory*, (Springer Verlag, New York, 1995).
11. Pai P.-F., Hong W.-C., An improved neural network model in forecasting arrivals, *Annals of Tourism Research*, vol. 32(4), (2005) 1138–1141.
12. R. Law, A neural network model to forecast Japanese demand for travel to Hong Kong. *Tourism Management*, vol. 20, (1999) 80–97.
13. Chen K.-Y., Wang C.-H., Support vector regression with genetic algorithms in forecasting tourism demand, *Tourism Management*, vol. 28(1), (2007) 215–226,
14. Sam Roweis & Lawrence Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, v.290 no.5500, Dec.22, (2000) 2323–2326.
15. Platt J.: Fast Training of Support Vector Machines using Sequential Minimal Optimization. In: B.Scholkopf, C.J.C. Burges, and A.J.Smola, editors, *Advances in Kernel Methods—Support Vector Learning*, Cambridge, MIT Press. (1999) 185–208
16. Dietterich T. G., Bakiri G., Solving multiclass learning problems via error-correcting output codes. *Journal of Artificial Intelligence Research*, vol. 2, (1995) 263–286,
17. SVM and Kernel Methods Matlab Toolbox, <http://asi.insa-rouen.fr/enseignants/~arakotom/toolbox/index.html>