

## 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



















