Available online at www.sciencedirect.com
SCIENCE DIRECT®
PERGAMON
Tourism Management 24 (2003) 511–517
TOURISM MANAGEMENT
www.elsevier.com/locate/tourman

Incorporating the rough sets theory into travel demand analysis

Carey Goh*, Rob Law

School of Hotel and Tourism Management, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

Received 16 June 2002; accepted 9 December 2002

Abstract

This research integrates the rough sets theory into tourism demand analysis. Originated from the area of Artificial Intelligence, the rough sets theory was introduced to reveal important structures and to classify objects. A rough sets approach can capture useful information from a set of raw hybrid data and discover knowledge from the data in a form of decision rules. This makes the rough sets approach a useful classification and pattern recognition technique. Because of its ability to accommodate hybrid data and its algorithms without rigorous theoretical and statistical assumptions, the theory could complement the orthodox demand framework. This paper introduces a new rough sets approach for deriving rules from an Information Table of tourist arrivals. The induced rules were able to forecast change in demand with 87% accuracy.

Keywords: Rough sets; Tourism demand; Volume of trade; Continuous-value discretization

1. Introduction

As tourism becomes the largest industry in the world, its impact on the global economy is widely recognized. To keep pace with the rapid flow of tourists, a destination needs to plan the utilization of its resources. One of the key and preliminary elements in the planning process is to study the demand for tourist arrivals in terms of both volume and determinants. To understand the relationship between tourist arrivals and their determining factors, most of the existing causal studies of tourism demand apply economic models that use mathematical functions. These models are established on the basis of many statistical assumptions and limitations. To a certain extent, the traditional models serve the purpose of acquiring new knowledge. These models, however, do not provide sufficient predictive capability when it comes to problems involving interactions among many interdependent variables with unknown probability distribution. In other words, these models are unable to perform consistently well in situations where the exogenous variables correlate with each other, and when distributions of the samples of variables do not meet the required independent and identical distribution (iid) condition. Moreover, these forecasting models have very much been exploited and only marginal improvements might be expected from their continued use. This paper presents a new approach that applies the rough sets theory to form a model for tourism demand. The rough sets theory was originally introduced by Pawlak (1982) to reveal important structures in a data set and to classify objects. Unlike a conventional data analysis, which uses statistical inferential technique, the rough sets approach is based on data-mining techniques to discover knowledge. Originating from Artificial Intelligence, a subdivision of Computer Science, the rough sets approach has been found successful in pattern recognition and object classification in medical and financial fields (Slowinski & Zopounidis, 1995; Tanaka & Maeda, 1998). The theory has been incorporated into tourism and hospitality research by Law and Au (1998, 2000), and Au and Law (2000). Through the rough sets approach, the relationship of a decision (dependent) variable and a set of condition (independent) variables in terms of decision rules could be modeled. Most importantly, these decision rules can represent data in both numeric and non-numeric forms. This makes the rough sets approach a very useful classification and pattern recognition

*Corresponding author. Tel.: +852-2766-4034; fax: +852-2362-9362.

E-mail addresses: hmcarey@polyu.edu.hk (C. Goh), hmroblaw@polyu.edu.hk (R. Law).

0261-5177/03/$ - see front matter
doi:10.1016/S0261-5177(03)00009-8

technique. However, to date, no published work has ever linked the rough sets theory with relationship modeling and forecasting in tourism demand analysis. This study makes an attempt to bridge this gap.

The objective of this research is to induce patterns in a form of decision rules, which are able to distinguish between the classes of arrival volume (in terms of percentage change) based upon differences in the factors that affect tourist arrivals. The aim is to understand how demand is affected by changes in demand determinants.

The remaining part of this paper is organized as follows: Section 2 presents the methodology of the research and concepts of the rough sets theory. Section 3 depicts the empirical results and model performance and Section 4 is concluded with discussions on implications and suggestions for future research within the tourism industry.

2. Research methodology

2.1. Data collection and data preprocessing

This study pooled data on tourism demand determining factors of the ten main tourist generating countries/region to Hong Kong over a period of 16 years from 1985 to 2000. These ten countries/region were China, Taiwan, Japan, Korea, the United States, United Kingdom, Australia, Singapore, Thailand, and the Philippines. Six determining factors were used to forecast whether arrivals increase or decrease in number, these being country of origin, Gross Domestic Product (GDP), relative consumer price index (CPI), population, volume of trade and foreign exchange rate. With the exception for trade volume, these factors were commonly used in many published articles on tourism demand analysis. Volume of trade was included in this analysis because arrivals on business purposes consistently made up a large percentage of about 32% on average of total arrivals to Hong Kong (Hong Kong Tourist Association, 1993–2000). For this reason, volume of trade was expected to influence the demand for travel to Hong Kong and it was therefore included in the model to help explain demand. The initiative of including this variable in tourism demand analysis is inline with that of Turner, Reisinger, and Witt (1998), Turner and Witt (2001a, b) in which tourist flows are disaggregated into different purpose of visit, including business visit. Volume of trade was measured as the total value of import and export of merchandize and services between Hong Kong and the tourist origin countries/region. Data on arrivals were collected from the Hong Kong Tourism Board (Hong Kong Tourist Association, 1985–2000) and other secondary data were collected from the online database, Datastream. Other than country of origin, secondary data for the other determining factors were recorded in continuous values and were transformed into percentage-change values to avoid biased estimation and spurious relationship modeling. For the purpose of rough sets-based data analysis, these percentages were then discretized into categorical values for further analysis. In this study, these percentages were automatically discretized using an algorithm based on entropy measures introduced by Fayyad and Irani (1992). The algorithm represents supervised, local and dynamic approaches to discretization. A supervised approach means that information about classes of objects is used during discretization, local means that only one attribute is considered per class, and dynamic-algorithm determines the number of intervals, into which the domain of attribute is divided (Obersteiner & Wilk, 1999). For a set S of N examples in which data are partitioned into k number of classes, the class entropy, $Ent(S)$ is defined as

$$Ent(S) = -\sum_{i=1}^{k} P(C_i, S)\log(P(C_i, S)),$$

where $P(C_i, S)$ is the proportion of examples in a system S that have class C_i.

To evaluate the robustness of the rough sets model, six samples, each of which contains 131 observations were randomly chosen and a leave-ten-out experiment was conducted by using those 131 observations for training and the remaining 10 observations for testing.

2.2. The rough sets concept

The rough sets theory provides a relatively new technique of reasoning from vague and imprecise data. This section presents the main concepts of the theory. More introductions to the theory can be found in Pawlak, Wong, and Ziarko (1988) and thorough theoretical foundation is illustrated in Pawlak (1991).

2.2.1. Concept 1: Indiscernibility of objects

Indiscernibility means similarity and it is the mathematical basis of rough sets theory. In rough sets theory, an information system, S, is represented as follows:

$$S = \langle U, Q, V, f \rangle,$$

where U represents the closed universe with a finite set of N objects $\{x_1, x_2, x_2, \dots, x_N\}$ in the system, Q is a finite set of n attributes $\{q_1, q_2, \dots, q_n\}$ in that information system, V is the domain (value) of attributes q. f is the total decision function called the information function. Some objects (let them be x and y, where $x, y \in U$) in S cannot be distinguished in terms of an available set of attributes (let it be A, where $A \subseteq Q$). Then it is said that x and y are 'indiscernible' by a set of attributes A, denoted by

$$IND(A) \cdot \text{(indiscernibility relation in terms of } A).$$