Data mining framework based on rough set theory to improve location selection decisions: A case study of a restaurant chain

Li-Fei Chen [a],[*], Chih-Tsung Tsai [b]

[a] Department of Business Administration, Fu Jen Catholic University, New Taipei City, Taiwan
[b] Jie Kune Precision Technologies Co., Ltd., Taoyuan City, Taiwan

HIGHLIGHTS

- A data mining framework was established to support location selection decisions.
- Rough set theory was applied to predict store performance with location factors.
- A case of restaurant chain was studied to demonstrate the proposed approach.

ARTICLE INFO

ABSTRACT

Location selection plays a crucial role in the retail and service industries. A comprehensive location selection model and appropriate analytical technique can improve the quality of location decisions, attracting more customers and substantially impacting market share and profitability. This study developed a data mining framework based on rough set theory (RST) to support location selection decisions. The proposed framework consists of four stages: (1) problem definition and data collection; (2) RST analysis; (3) rule validation; and (4) knowledge extraction and usage. An empirical study focused on a restaurant chain to demonstrate the validity of the proposed approach. Twenty location variables relevant to five location aspects were examined, and the results indicated that latent knowledge can be identified to support location selection decisions.

1. Introduction

Location selection is one of the most critical factors to the success of long-term strategic decisions taken in the restaurant industry. A suitable restaurant location can attract more customers, provide convenient service to customers, and enhance customer loyalty. Moreover, it can decrease the length of the period required to pay for fixed capital investments and increase market share and profitability (Chou, Hsu, & Chen, 2008; Prayag, Landré, & Ryan, 2012; Tzeng, Teng, Chen, & Opricovic, 2002). Therefore, academics and practitioners have focused on examining location decisions.

Because of the importance of retail location decisions, a number of models have been developed to address this decision. Central place theory, spatial interaction theory, and the principle of minimum differentiation are the most discussed retail location models in the literature (Brown, 1993; Litz & Rajaguru, 2008; Prayag et al., 2012). Although they are normative and require unrealistic assumptions, these models continue to attract considerable academic attention (Brown, 1993; Chou et al., 2008; Prayag et al., 2012). The models provide in-depth information on certain dimensions, however, more extensive perspectives must be considered when modeling a location selection problem (Kuo, Chi, & Kao, 2002). Although simple analytical techniques, such as checklists and analogs, have been available for at least 60 years to support location decisions, most managers still favor their personal experience and instincts (Hernández & Bennison, 2000). Many studies have explored the location selection problem by using statistical methods, such as regression, cluster, and factor analysis (Davies, 1973; Rogers & Green, 1979). However, conventional statistical methods require a high level of specialist knowledge regarding model building in general, and they assume that data are normally distributed and exhibit linear relationships to provide meaningful

* Corresponding author. Department of Business Administration, Fu Jen Catholic University, No. 510, Zhongzheng Rd., Xinzhung Dist., New Taipei City, 24205, Taiwan.
E-mail address: 075033@mail.fju.edu.tw (L.-F. Chen).

inferences (Chen, 2014; Coates, Doherty, French, & Kirkup, 1995; Hernández & Bennison, 2000). Some studies have applied mathematical programming methods to location selection problems. However, these methods can only use quantitative data and the modeling process is relatively time consuming (Hernández & Bennison, 2000; Ho, Chang, & Ku, 2013).

Data mining is a powerful tool that can be used to analyze large quantities of data and discover potentially helpful patterns or hidden rules. Data mining is widely used in many fields, but few researchers have applied it to location selection (Hernández & Bennison, 2000). Rough set theory (RST) is an effective data mining method that can be used to explain and explore how a decision was made using simple, understandable, and useful rules in the presence of uncertainty and vagueness without requiring the assumptions that are made during regression analysis (Chien & Chen, 2007). The RST method may be more effective than a regression-based approach to capture the relationship between location factors and store performance.

This study aims to develop a data mining framework based on RST to explore store location data for predicting store sales performance. The proposed framework consists of four stages: (1) problem definition and data collection; (2) RST analysis; (3) rule validation; and (4) knowledge extraction and usage. An empirical study focused on a restaurant chain to demonstrate the validity of the proposed approach. The results indicated that latent knowledge can be revealed to identify critical location selection factors, allowing specific location strategies to be derived for selecting optimal locations.

2. Related location selection studies

A number of studies have examined the location selection problem. Central place theory was proposed by Christaller (1933), who suggested that a retail location is determined by the range and threshold of a good. The range refers to the maximal distance consumers are willing to travel to obtain a good and the outer limit of a store's market area is determined accordingly. Actual range differences may vary based on individual mobility, price, and customer preferences. The threshold is defined as the minimal amount of demand that must exist in an area for a store to be economically viable; that is, an area should be sufficiently populated to support a store (Craig, Ghosh, & McLafferty, 1984). Craig et al. (1984) indicated that central place theory is the most well-developed normative theory related to retail location. However, the main limitation of this theory is that it cannot provide different retail location patterns based on product offerings, store image, and competition levels (Litz & Rajaguru, 2008).

Central place theory assumes that consumers shop at the nearest source that provides a required good or service. By contrast, spatial interaction theory (Reilly, 1929; 1931) assumes that consumers trade off the attractiveness of alternative shopping areas against the obstacle effect of distance (Brown, 1993). Using the gravity model, spatial interaction theory treats a number of variables that could influence competing locations—including site related factors, such as store size, distance, price, service levels, and image features, and other attractiveness factors, such as atmosphere and consumer cognition—as determinants of store location (Teller & Reutterer, 2008). Although it is widely used in practical situations, the main shortcoming of the gravity model is its mathematical complexity (Burnaz & Topcu, 2006; Prayag et al., 2012).

Principle minimum differentiation theory—developed by Hotelling in 1929—emphasizes the concept of the clustering effect. It suggests that proximity to competitors is an indicator of attractiveness and competitiveness (Chou et al., 2008). Prayag et al. (2012) summarized the reasons why restaurants cluster together.

For example, clustering increases the attractiveness of individual restaurants and an area as a whole; it facilitates comparison of restaurants based on cuisine type; and it enables the sharing of facility and promotion costs. Restaurant clustering also benefits customers. For example, it decreases the search time and costs required to find a suitable restaurant and it provides various choices for customers in a certain area.

A wide variety of analytical techniques have been applied to support location decisions. Although simple methods, such as checklists and analogs, have been available for at least 60 years, most managers still solely rely on their personal experiences and instincts (Hernández & Bennison, 2000). Despite of multivariate statistical techniques, certain studies have managed the location selection problem by using mathematical programming methods (e.g., Jovanovic, 2003; Kolli & Evans, 1999). However, the modeling process is computer- and data-intensive and relatively time consuming. Moreover, this method can only be used to process quantitative data. It cannot incorporate qualitative location criteria; therefore, its applications are limited (Ho, Chang, & Ku, 2013). An increasing number of studies have used multi-criteria decision making models, such as the analytic hierarchy process (AHP) (Chou et al., 2008; Ho et al., 2013; Tzeng et al., 2002), analytic network process (ANP) (Burnaz & Topcu, 2006; Tuzkaya, Önüt, Tuzkaya, & Gülsün, 2008), PROMETHEE, and the technique for order of preference by similarity to ideal solution (Ishizaka, Nemery, & Lidouh, 2013) to evaluate location selection problems. Researchers have used the AHP and ANP because they can process both qualitative and quantitative criteria. The AHP was developed by Saaty (1980) to determine the relative importance of a set of alternatives in a complex, unstructured and multi-criteria decision problem. There are three basic steps in using AHP: (1) the design of the hierarchy to describe the decision problem; (2) the prioritization of various attributes in each level of the hierarchy by pairwise comparisons; and (3) the integration of the pairwise comparisons to develop the overall evaluation of these alternatives (Partovi, 2001). The ANP technique does not require a strict hierarchical structure as AHP and it can deal with more complicate interdependencies among and between levels of attributes and alternatives (Partovi, 2007).

However, it is difficult and time consuming for decision makers to evaluate too many criteria because more pairwise comparisons are needed. Moreover, the accuracy of the results largely depends on the user's experience and knowledge in the area concerned, and this may result in an unreliable analysis (Ravi, Shankar, & Tiwari, 2005; Yurdakul, 2003).

Because of advances in information technology, data mining techniques have been applied to the location selection decision. For example, Coates et al. (1995) used an artificial neural network (ANN) and Wang, Chen, and Su (2015) proposed a fuzzy-connective-based aggregation networks method to support location selection decisions. Kuo et al. (2002) proposed an integrated fuzzy ANP and ANN method to select convenience store locations by considering competition, commercial area, convenience, availability, store characteristics and population characteristics. They performed a comparison analysis and found that their method provided more accurate results than a regression model.

3. Rough set theory

RST was developed by Pawlak (1982; 1997; 2002). It is a data mining approach used for various purposes, such as feature selection, feature extraction, feature reduction, and extraction of decision rules from data, especially in the presence of uncertainty and vagueness (Chien & Chen, 2007). RST has been applied in various domains, such as quality engineering (Su & Hsu, 2006), human resource management (Chien & Chen, 2007), health care