Chapter 63
Data Mining for Multicriteria Single Facility Location Problems

Seda Tolun
Istanbul University, Turkey

Halit Alper Tayalı
Istanbul University, Turkey

ABSTRACT

This chapter focuses on available data analysis and data mining techniques to find the optimal location of the Multicriteria Single Facility Location Problem (MSFLP) at diverse business settings. Solving for the optimal of an MSFLP, there exists numerous multicriteria decision analysis techniques. Mainstream models are mentioned in this chapter, while presenting a general classification of the MSFLP and its framework. Besides, topics from machine learning with respect to decision analysis are covered: Unsupervised Principal Components Analysis ranking (PCA-rank) and supervised Support Vector Machines ranking (SVM-rank). This chapter proposes a data mining perspective for the multicriteria single facility location problem and proposes a new approach to the facility location problem with the combination of the PCA-rank and ranking SVMs.

INTRODUCTION

Businesses encounter selection problems all the time. Selecting the right machine or selecting a type of layout to increase the shop floor performance is a typical problem related to operational research. With its deterministic nature, facility location is among the most studious topics of business, operational research and related fields. Broadly, the facility location problem (FLP) can be defined as selecting the optimal location on where an operation is going to be run.

DOI: 10.4018/978-1-7998-2460-2.ch063

Location decisions are strategic by all means and they can be based on various needs such as the renewal of the facility or capacity improvement. Structural and long-term decisions tend to have immeasurable impact on the overall efficiency of the systems; meaning that while a good decision can boost the system, a poor decision may have ramifications for all parties in the end.

Decision analysis models can be used to determine the optimal selection by ranking the alternatives of a dataset that contains multiple criteria. Ranking (used interchangeably for ordination) constitutes a significant amount of research in the field of multicriteria decision making (MCDM). The background section of this chapter puts forward the mainstream facility location models used in business management and operational research.

This chapter tackles the facility location selection problem from a quantitative data analysis and data mining perspective. Data mining -a developing sub-category of data analysis that draws attention from both industry and academia- is the extraction of meaningful information and useful patterns from data using data analysis methods and machine learning algorithms. The importance of emerging techniques for the machine learning and business community is emphasized throughout the chapter. The idea of adapting contemporary methods from data analysis and data mining to facility location selection models is introduced in the background section. The aim is to equip readers with a detailed perspective on the facility location selection problem while centering on diverse business applications.

Machine learning literature classifies data mining models basically as supervised and unsupervised learning methods, according to the presence of a pre-assigned response variable within a dataset. Supervised data analysis is used to estimate an unknown dependency from a known input-output data; the learning system is modified so that the error between the model output and training data is minimized. Unsupervised data analysis, on the other hand, does not involve any supervision of an external source and hence any fine-tuning (Ahlemeyer-Stubbe & Coleman, 2014).

While analyzing multicriteria single facility location problem (MSFLP) through the unsupervised learning perspective, this chapter focuses on an exploratory data analysis technique called the principal components analysis (PCA). PCA is based on a mathematically elegant concept, namely the eigenanalysis. Through eigenanalysis, the variance of a dataset is expressed algebraically in a different way, yet the structure of the data cloud is preserved. After unfolding the connection between the variance and the optimal selection of a dataset, PCA-rank is applied to an MSFLP.

First proposed by Herbrich et al. (1999) ranking by support vector machines (SVM) is applied in this chapter for learning the ranking of facility locations. SVM-rank section provides a supervised data analysis perspective for the optimal selection in an MSFLP. Since learning to rank is a supervised learning task, and accordingly is the ranking SVM algorithm, there has to be a training set and that is considered to be the results obtained from the unsupervised PCA-rank.

This chapter approaches the optimal FLP using a combined ranking approach which includes supervised and unsupervised machine learning methods. This hybrid approach is tested on a real life dataset.

BACKGROUND

Business research literature refers to various techniques to solve facility location selection problems and related models such as factor-rating, transportation method of linear programming, center of gravity method, median method, volume-cost analysis, and Ardalan method (Chase et al., 2001; Ozcakar, 2015). An internet-based service provided by NEOS Server ("network-enabled optimization solver") offers