Computers & Industrial Engineering 63 (2012) 855–863

Contents lists available at SciVerse ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

Use of distributed data sources in facility location

Jānis Grabis [a,*], Charu Chandra [b], Jānis Kampars [a]

[a] Institute of Information Technology, Riga Technical University, Kalku 1, Riga LV-1658, Latvia
[b] College of Business University of Michigan–Dearborn, 19000 Hubbard Drive, Dearborn, MI 48126, USA

ARTICLE INFO

Article history:
Received 22 August 2011
Received in revised form 29 May 2012
Accepted 29 May 2012
Available online 6 June 2012

Keywords:
Facility location
Multi-objective optimization
Spatial data

ABSTRACT

Facility location decisions are usually determined by cost and coverage related factors although empirical studies show that such factors as infrastructure, labor conditions and competition also play an important role in practice. The objective of this paper is to develop a multi-objective facility location model accounting for a wide range of factors affecting decision-making. The proposed model selects potential facilities from a set of pre-defined alternative locations according to the number of customers, the number of competitors and real-estate cost criteria. However, that requires large amount of both spatial and non-spatial input data, which could be acquired from distributed data sources over the Internet. Therefore, a computational approach for processing input data and representation of modeling results is elaborated. It is capable of accessing and processing data from heterogeneous spatial and non-spatial data sources. Application of the elaborated data gathering approach and facility location model is demonstrated using an example of fast food restaurants location problem.

1. Introduction

The aim of facility location is to determine spatial position of different types of facilities in order to provide good customer service and to attain a competitive advantage. In the case of discrete facility location, there is a pre-defined set of alternative locations, and facilities are placed at some of these locations to optimize cost or coverage criteria. Owen and Daskin (1998) in their review of facility location problems classify facility location problems as median, covering and center problems. Melo, Nickel, and Saldanha-da-Gama (2009) analyze facility location from the supply chain management perspective and identify close relationships between facility location and supply chain design problems. They show that the majority of facility location and supply chain design models consider optimization only according to a single objective (usually cost or profit). However, the supply chain design and facility location problems are inherently multi-objective problems affected by a large number of different decision-making factors. Bhatnagar and Sohal (2005) classify factors influencing plant location in eight groups, namely, cost, infrastructure, business services, labor, government, customer/market and supplier/resources and competitor related factors. Their empirical analysis shows that the majority of these factors are important to decision makers. Nwogugu (2006) criticizes existing facility location models because of their failure to address several concerns important to practitioners. The author also discusses approaches for gathering required input data for a new facility location model designed specifically for a retail store location. Difficulties to deal with a multitude of factors relevant to facility location are also highlighted by MacCarthy and Atthirawong (2003). Julka, Baines, Tjahjono, Lendermann, and Vitanov (2007) show that the closely related facility expansion problem is also a multiple criteria problem and there are no models capturing all facets of the problem. Fernández and Puerto (2003) develop methods for obtaining the Pareto set of solutions to the multi-objective plant location problem, where different criteria are expressed in terms of their cost impact. Above mentioned studies, show that practitioners are particularly concerned with government, infrastructure and cost related factors. In order to address some of these concerns, the facility location problem should be tackled as a multiple-objective problem and a wide range of factors should be accounted for.

Additionally, many of these factors have spatial features, and spatial and cartographical information are necessary during the decision-making process (Vlachopoulou, Silleos, & Manthou, 2001). Such techniques as data warehousing (Dolk, 2000) can be used to handle non-spatial data necessary for decision-making while they are less efficient in dealing with spatial data because of data maintenance and infrastructural difficulties. Therefore, distributed data sources maintained by external service providers are an attractive option for handling spatial data. There are multiple standards for handling distributed spatial and non-spatial data. WMS (Web Map Service) and WFS (Web Feature Service) are spatial data access interfaces for requesting spatial data and features, respectively (Chen, Gong, & Chen, 2007). These standards

* Corresponding author. Tel.: +371 67089594; fax: +371 67089513.
E-mail addresses: grabis@iti.rtu.lv (J. Grabis), charu@umd.umich.edu (C. Chandra), janis.kampars@rtu.lv (J. Kampars).

0360-8352/$ - see front matter
http://dx.doi.org/10.1016/j.cie.2012.05.007

are promoted by the Open Geospatial Consortium (OGC), though they are yet to achieve a wide-spread acceptance. Another OGC standard, which can be used in spatial data exchange, is GML (Geography Markup Language) (Peng & Zhang, 2004). KML (Keyhole Markup Language) is an alternative spatial data exchange format actively promoted by Google (Du, Yu, & Liu, 2009). Recently, Google has submitted it to the OGC, allowing future harmonization of GML and KML. SOAP (Simple Object Access Protocol) based web services is the most widely used technology for exchanging non-spatial data. Standards allow reducing complexity of handling distributed data though their variety and relatively low level of approbation in decision-making applications complicates the usage of distributed data.

The objective of this paper is to develop a facility location model accounting for wide range of factors affecting decision-making and to propose a computational approach supporting the modeling process. The computational approach aims to cover the modeling life-cycle from data gathering to model-solving, and to representation of modeling results. The computational approach is necessary, because otherwise the complexity of gathering necessary data for model solving and representation of modeling results would render the proposed facility location model impractical.

The existing literature is analyzed to determine factors influencing facility location and supply chain configuration decisions. It shows that besides the traditional factors such as costs and time, additional factors such as local government, competitive situation, and labor market are highly relevant in practical decision-making. A multi-objective optimization model accounting for the identified factors is formulated. This model can be perceived as a data-driven model, which can be easily expanded if additional input factors should be accounted for. The computational approach ensures gathering data necessary for decision-making purposes from distributed spatial and non-spatial data sources. The model-solving activity is performed using a commercially available solver and a specially designed facility location heuristic. A geographical information system is used to represent the modeling results by combining multiple layers of spatial data. The main principles of the proposed computational architecture are (a) utilization of distributed data handling standards, (b) accounting for Quality of Service (QoS) characteristics of the distributed data providers, (c) utilization of local data processing facilities when possible, and (d) systematic approach to data processing process composition.

Application of the proposed facility location model and the supporting computational approach is demonstrated by providing a sample problem of locating fast food restaurants and extracting data from publicly available data sources.

Three main contributions of this paper are: (1) expanding the scope of facility location models to account for factors relevant to practitioners; (2) demonstration of utilization of spatial data in facility location and supply chain management; and (3) elaboration of the computational approach for gathering data necessary for decision-making from external, heterogeneous data sources. The proposed model differs from traditional facility location models by considering semi-continuous distribution of demand (the distribution is not continuous at the atomic level, but it is continuous if we consider the level of granularity used for defining potential facility locations). On the other hand, the demand distribution is derived empirically and does not depend upon distributional assumptions used in continuous facility location models (for instance, see Dasci & Verter, 2001). From the spatial data processing perspective, the proposed computational approach differs from the research by Vlachopoulou et al. (2001) by considering distributed spatial data sources.

The rest of the paper is organized as follows: Section 2 discusses factors influencing facility location. The multi-objective facility location model is elaborated in Section 3. Section 4 describes the proposed data gathering approach. Application of the model and the data gathering are demonstrated in Section 5 using an illustrative example of locating fast food restaurants. Section 6 concludes.

2. Factors influencing facility location

Traditional facility location models optimize facility location cost or some kind of coverage criterion. Typical factors influencing facility location are customer demand, facility location costs and distance between facilities and customers (Owen & Daskin, 1998). However, additional factors are often considered in practice. In order to identify the factors most frequently considered in facility location, two sources of information are used: (1) empirical studies on facility location factors; and (2) surveys on factors used in supply chain design.

Bhatnagar and Sohal (2005) list 41 factors influencing plant location. These factors are categorized in eight groups including cost, infrastructure, business services, labor, government, customer/market, suppliers/resources and competitors. The importance of each factor is determined using a survey. MacCarthy and Atthirawong (2003) identify thirteen groups of factors influencing facility location with focus in internationalization. They also consider such factors as proximity to parent company, quality of life and social and cultural factors. However, the average importance of these factors is lower than for other factors. Factors influencing facility location in retailing are identified by Nwogugu (2006). Position relative to competitors, impact of on-line shopping and local attractions are highlighted as important factors. He also briefly discusses methods for estimating input parameters of the proposed model. The position relative to competitors also has been considered in several mathematical facility location models referred to as competitive facility location models (e.g., Aboolian, Berman, & Krass, 2007).

Supply chain design models reviewed by Melo et al. (2009) and Chandra and Grabis (2007) deal with facility location issues as well as with product allocation, transportation planning and others. Typically these models attempt to optimize financial performance measures using various cost-related factors such as cost for opening facilities, transportation cost, and sourcing costs. Two groups of additional factors which are frequently considered in supply chain design are international factors (e.g., trade barriers, local incentives, currency exchange rates) and risk management related factors.

The discussion above shows that there is a large variety of factors influencing facility location. From the perspective of this paper, it is important to identify possible data sources for each factor. Table 1 lists groups of factors relevant to facility location as defined by Bhatnagar and Sohal (2005) and data type and sources type for each group. The data type is classified as either spatial or non-spatial data (a factor group might include both spatial and non-spatial data but the data type attribute is assigned according to the dominant data type). Spatial data identify the geographic location of features and boundaries and are usually stored

Table 1
The classification of factors relevant to facility location.

Factor	Data type	Source type
Cost	Non-spatial	Commercial
Infrastructure	Spatial	Public
Business services	Non-spatial	Commercial, field work
Labor	Non-spatial	Public, commercial
Government	Non-spatial	Public, commercial
Customer/market	Spatial	Public, commercial
Suppliers/resources	Spatial	Commercial, field work
Competitors	Spatial	Commercial, field work