European Journal of Operational Research 000 (2015) 1–12

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Discrete Optimization

A mean-shift algorithm for large- scale planar maximal covering location problems

Zhou He [a], Bo Fan [b,*], T. C. E. Cheng [c], Shou-Yang Wang [d], Chin-Hon Tan [a]

[a] Department of Industrial & Systems Engineering, National University of Singapore, Singapore
[b] School of International and Public Affairs, Shanghai Jiao Tong University, 1954 Huashan Road, Shanghai 200030, China
[c] Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China
[d] Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Haidian District, Beijing 100190, China

ARTICLE INFO

Article history:
Received 19 March 2015
Accepted 2 September 2015
Available online xxx

Keywords:
Location
Large scale optimization
Planar maximal covering location problem
Mean shift

ABSTRACT

The planar maximal covering location problem (PMCLP) concerns the placement of a given number of facilities anywhere on a plane to maximize coverage. Solving PMCLP requires identifying a candidate locations set (CLS) on the plane before reducing it to the relatively simple maximal covering location problem (MCLP). The techniques for identifying the CLS have been mostly dominated by the well-known circle intersect points set (CIPS) method. In this paper we first review PMCLP, and then discuss the advantages and weaknesses of the CIPS approach. We then present a mean-shift based algorithm for treating large-scale PMCLPs, i.e., MSMC. We test the performance of MSMC against the CIPS approach on randomly generated data sets that vary in size and distribution pattern. The experimental results illustrate MSMC's outstanding performance in tackling large-scale PMCLPs.

1. Introduction

Covering problems in facility location have received considerable research interest due to its applicability in the real world (Farahani, Asgari, Heidari, Hosseininia, & Goh, 2012). Each facility is able to provide services within a given critical distance, i.e., the coverage radius. A customer is considered served from a facility if the distance between them is less than or equal to the facility's coverage radius. In reality, however, budget limits often constrain the number of service facilities to be located. This gives rise to the maximal covering location problem (MCLP) (Church & Velle, 1974), which seeks to maximize the coverage of customer demands by siting a given number of new facilities. In the last 40 years, MCLP and its extensions have been widely applied to study various location issues such as planning emergency facilities (e.g., Schilling, Revelle, Cohon, & Elzinga, 1980; Eaton, Daskin, Simmons, Bulloch, & Jansma, 1985; Murray & Tong, 2007), siting telecommunications equipment (e.g., Akella, Delmelle, Batta, Rogerson, & Blatt, 2010; Oztekin, Pajouh, Delen, & Swim, 2010; Shillington & Tong, 2011), location for business (e.g., Jones & Simmons, 1993; Pastor, 1994), and public services (e.g., Hougland & Stephens, 1976; Otto & Boysen, 2014).

Most MCLP models have been built under the assumption that the candidate locations of the new facilities are known in advance. In other words, facilities can only be installed in discrete nodes. Some researchers (e.g., Mehrez, 1983; Mehrez & Stulman, 1982, 1984; Church, 1984) have relaxed this constraint and extended the discrete version of MCLP to consider facility location in a continuous space, i.e., facilities are allowed to be placed anywhere on a plane. This problem is known as the planar maximal covering location problem (PMCLP), originally defined in Church (1984). For PMCLP, it is possible to attain a greater demand coverage because many more desired locations are available for selection when making strategic facility location decisions such as infrastructure investment (Murray & Tong, 2007; Wei, 2008). Murray and Tong (2007) suggested that more general representations (points, lines or polygons) of demand can also be optimally served in a region. They introduced the extended planar maximal covering location problem-Euclidean (EPMCE) and applied it to emergency warning sirens siting in Dublin. Matisziw and Murray (2009) formulated the 1-facility continuous maximal covering problem (CMCP-1), where demand is considered as continuously distributed within a whole region (convex or non-convex). They addressed CMCP-1 by generating the medial axis, which can be viewed as a geometrical representation of the region. Later in Wei (2008), the

* Corresponding author. Tel./fax: +86 21 6293 3095.
E-mail addresses: isehz@nus.edu.sg (Z. He), fanbo@sjtu.edu.cn (B. Fan), edwin.cheng@polyu.edu.hk (T.C.E. Cheng), sywang@amss.ac.cn (S.-Y. Wang), isetch@nus.edu.sg (C.-H. Tan).

http://dx.doi.org/10.1016/j.ejor.2015.09.006
0377-2217/

multi-facility case of CMCP was solved based on Voronoi diagrams and the geometric properties of the region. Recently, several applications of PMCLP have been reported by geographical researchers (see for example Liu & Hodgson, 2013; Wei, Murray, & Batta, 2014; Wei & Murray, 2014b).

To solve PMCLP, it is natural to reduce it to MCLP by finding a finite number of potential sites on the plane. With this set of discrete sites, MCLP can be solved by either exact or heuristic approaches. In other words, PMCLP can be addressed in two phases: I – identify a candidate locations set (CLS) and II – use exact or non-exact methods to address the degraded PMCLP for coverage maximization. Therefore, it is critical to identify a *good* CLS for solving PMCLP. We can analyze a CLS from various angles: (1) Coverage. The objective of the MCLP is to find a solution with maximal coverage. Hence, the coverage of the candidates in a CLS is an important consideration. (2) Size of the CLS. MCLP is NP-hard (Downs & Camm, 1996), whose size is determined by the numbers of demand nodes, potential locations, and facilities to be located. For a very large-scale MCLP, we have to apply heuristic algorithms to search for the optimal solution. However, using such non-exact methods may cause a loss of coverage (obtaining a local optimal solution only), or even fail to produce a feasible solution. Therefore, given a large number of demand points, it is highly desirable to reduce the number of potential sites in Phase I before solving MCLP. In addition to the above two essential principles, decision-makers in reality may also be concerned with the following objectives. (3) Time to generate the CLS. Is there an efficient way to generate the CLS in Phase I? This research question has largely been neglected in the literature, yet it is a significant issue to address in real life. For example, when a severe disaster (e.g., a storm, an earthquake etc.) takes place, it is an urgent task for the government to decide where to site search and rescue (SAR) stations in order to provide medical and rehabilitation services for as many victims as possible in a large area. The scale of such a problem can be remarkably large as the numbers of SAR stations and victims may be numerous. In the context of such emergency situations, the decision-maker must consider not only the above features of the candidate locations, but also the efficiency of solving such large-scale PMCLPs in real time. (4) Average distance to demand. Given that serving faraway demands will incur a high cost, planners may also be concerned about the distances between the covered demand points to the closest facility.

A three-step procedure, often referred to as the circle intersect points set (CIPS) method in the literature, has been the dominating technique for creating the CLS since it was introduced by Church (1984). First, this method produces a demand and intersection points set (DIPS[1]) by exploiting the geometric properties of coverage. The circles are centered to cover demand locations with predefined coverage radii under the Euclidean distance measure. The second step in Phase I, which is optional if the DIPS only has few members, is to remove all the dominated points from the DIPS in order to reduce the size of the CLS. It has been shown that the reduced DIPS, i.e., the final CLS, contains at least one optimal solution to the PMCLP (Church, 1984). Given this CLS, the last step is to solve MCLP in Phase II. The CIPS method greatly facilitates coverage maximization and contributes significantly to size reduction. Consequently, it has been widely applied and extended in many studies (e.g., Younies & Wesolowsky, 2004; Murray & Tong, 2007; Canbolat & Massow, 2009; Yildiz, Akkaya, Sisikoglu, & Sir, 2011) as a standard approach to address PMCLP. However, the CIPS method has high time complexity and is generally unable to handle large-scale PMCLPs (see Section 2.2 for the details).

[1] Note that this set was also called CIPS in Church (1984) and thus the term CIPS had two meanings: the whole method and the points set. To avoid confusion, hereafter in this paper we use the term CIPS to denote the method, and DIPS to call the demand and intersection points set generated by the CIPS method at the first step. We thank an anonymous referee for this helpful suggestion.

In this study we propose a mean-shift based algorithm for treating large-scale PMCLPs, i.e., MSMC. In MSMC, we introduce a revised mean-shift procedure that is less time consuming, and hence more suitable for solving large PMCLPs, than the traditional CIPS method. The mean-shift procedure has been successfully adapted to various application domains, such as cluster analysis in computer vision and image processing (Comaniciu & Meer, 2002). To the best of our knowledge, this work is the first attempt to solve location problems using the mean-shift procedure. The advantages for choosing the mean-shift procedure to identify the CLS for PMCLP are discussed in detail in Section 3.2.

The remainder of the paper is organized as follows: in Section 2 we review PMCLP, and discuss the pros and cons of the traditional CIPS approach. In Section 3 we briefly introduce the mean-shift procedure and the core of the proposed MSMC algorithm. We present each step of MSMC in detail. In Section 4 we compare the performance of MSMC against the CIPS method on randomly generated data sets that vary in size and distribution pattern, and discuss the experimental results to reveal the various performance aspects of the MSMC and CIPS approaches. In Section 5 we conclude the paper, discuss the research limitations, and suggest topics for future research.

2. Problem formulation and the CIPS method

PMCLP seeks to maximize the demand coverage on a plane. Unlike MCLP which locates facilities on a network or discrete locations, PMCLP concerns the siting of a given number of facilities anywhere on a plane; in other words, the number of potential sites for location is infinite in PMCLP. To address this issue, there is a need to generate a set of discrete potential locations from the continuous space, essentially reducing PMCLP to MCLP. To recap, solving PMCLP can be executed in two phases, namely I – identify a CLS and II – solve the MCLP given the CLS.

In this section we first present the MCLP formulation, from which PMCLP naturally arises. We then detail Phase I of the CIPS approach to solve PMCLP proposed by Church (1984). Finally, we discuss the advantages and shortcomings of the CIPS method.

2.1. The maximal covering location problem

MCLP has been mathematically formulated by Church and Velle (1974) as follows:

$$\text{Maximize } Z = \sum_{i \in I} w_i Y_i. \tag{1}$$

subject to

$$\sum_{j \in \Theta_i} X_j \geq Y_i, \qquad \text{for all } i \in I \tag{2}$$

$$\sum_{j} X_j = p, \tag{3}$$

$$X_j = \{0, 1\}, \qquad \text{for all } j \in J \tag{4}$$

$$Y_i = \{0, 1\}, \qquad \text{for all } i \in I \tag{5}$$

where

i = index of demand points (entire set I);
j = index of facility locations (entire set J);
w_i = weight of demand node i;
$\Theta_i = \{j \in J \mid d_{ij} \leq R\}$;
d_{ij} = the shortest distance from i to j;
R = predefined coverage radius of facility;
p = number of facilities to be located;