European Journal of Operational Research 237 (2014) 590–605

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Decision Support

An adaptive multiphase approach for large unconditional and conditional p-median problems

Chandra Ade Irawan *, Said Salhi, Maria Paola Scaparra

Centre for Logistics & Heuristic Optimization (CLHO), Kent Business School, University of Kent, Canterbury, Kent CT2 7PE, UK

A R T I C L E I N F O

Article history:
Received 26 July 2013
Accepted 23 January 2014
Available online 31 January 2014

Keywords:
Variable neighbourhood search
Exact method
Aggregation
Large p-median problems
Adaptive learning

A B S T R A C T

A multiphase approach that incorporates demand points aggregation, Variable Neighbourhood Search (VNS) and an exact method is proposed for the solution of large-scale unconditional and conditional p-median problems. The method consists of four phases. In the first phase several aggregated problems are solved with a "Local Search with Shaking" procedure to generate promising facility sites which are then used to solve a reduced problem in Phase 2 using VNS or an exact method. The new solution is then fed into an iterative learning process which tackles the aggregated problem (Phase 3). Phase 4 is a post optimisation phase applied to the original (disaggregated) problem. For the p-median problem, the method is tested on three types of datasets which consist of up to 89,600 demand points. The first two datasets are the BIRCH and the TSP datasets whereas the third is our newly geometrically constructed dataset that has guaranteed optimal solutions. The computational experiments show that the proposed approach produces very competitive results. The proposed approach is also adapted to cater for the conditional p-median problem with interesting results.

1. Introduction

The p-median problem is a discrete location problem where the objective is to find the location of p facilities among n discrete potential sites in such a way to minimise the sum of the weighted distances between customers and their nearest facilities. The p-median problem becomes the conditional problem when some (say q) facilities already exist in the study area and the aim is to locate p new facilities given the existing q facilities. This problem is also known as the (p,q) median problem. A customer can be served by one of the existing or the new open facilities whichever is the closest to the customer. When $q = 0$, the problem reduces to the unconditional problem (the p-median problem for short). A further but brief description related to the conditional p-median problem will be presented in Section 6 where some results are also given.

The p-median problem is categorised as NP-hard (Kariv & Hakimi, 1979). For relatively large problems, optimal solutions may not be found and hence heuristic or metaheuristic methods are usually considered to be the best way forward for solving such problems. Mladenovic, Brimberg, Hansen, and Moreno-Perez (2007) provided an excellent review on the p-median problem focusing on metaheuristic methods. The p-median problem was originally formulated by ReVelle and Swain (1970). However, Rosing, ReVelle, and Rosing-Vogelaar (1979) enhanced the p-median problem formulation to reduce its solution time. In their model, the furthest $p-1$ assignments associated with each demand point are ignored. This reduction scheme is based on the observation that in the worst case, a demand point i is served by its $(n-p+1)$th closest site. The enhanced p-median formulation is formulated as follows:

$$\text{Minimise} \quad \sum_{i \in I} \sum_{j \in F_i} w_i d(i,j) Y_{ij} \tag{1}$$

$$\text{s.t.} \quad \sum_{j \in F_i} Y_{ij} = 1 \quad \forall i \in I \tag{2}$$

$$\sum_{j \in J} X_j = p \tag{3}$$

$$Y_{ij} - X_j \leqslant 0, \quad \forall i,j \in F_i \tag{4}$$

$$X_j \in \{0,1\} \quad \forall j,j \in J \tag{5}$$

$$Y_{ij} \in \{0,1\} \quad \forall i,j \in F_i \tag{6}$$

where (I,J) is the set of customers ($i \in I = \{1,\ldots,n\}$) and set of potential sites ($j \in J = \{1,\ldots,M\}$) (i.e.:$n = |I|$ and $M = |J|$) respectively; w_i the demand or weight of customer i; $d(i,j)$ the distance between customer i and potential site j (Euclidian distance is used here); p the required number of facilities to locate; $Y_{ij} = 1$, if customer i is fully served by a facility at site j and = 0 otherwise; $X_j = 1$, if a facility is

* Corresponding author. Tel.: +44 (0)1227 82 4672.
E-mail address: ca259@kent.ac.uk (C.A. Irawan).

http://dx.doi.org/10.1016/j.ejor.2014.01.050
0377-2217/

opened at potential site j and = 0 otherwise; F_i is the set of all sites except the $p-1$ furthest sites from demand point i.

The objective function (1) minimises the total demand-weighted distance. Constraints (2) guarantee that each customer i is assigned to one facility only. Constraint (3) states that the number of facilities to be located is p. Constraints (4) ensure that customer i can only be allocated to facility j (i.e., $Y_{ij} = 1$) if a facility is opened at site j (i.e., $X_j = 1$). The use of the sets F_i in constraints (2), (4), and (6) yields a more compact formulation, requiring a fewer number of variables and constraints than the classical formulation.

In some applications, p-median problems may involve a large number of demand points and potential facility sites. These problems arise, for example, in urban or regional areas where the demand points are individual private residences. Francis, Lowe, Rayco, and Tamir (2009) stated that it may be impossible and time consuming to solve location problems consisting of a large number of demand points. To simplify the problem, it is quite common to aggregate demand points (and/or potential facility sites) when solving large-scale location problems. In other words, the number of demand points (and/or potential facility sites) can be reduced from n to m points ($m \ll n$) so that the approximated problem can be solved within a reasonable amount of computing time. However, aggregation introduces errors in the data as well as in the models output, thus resulting in less accurate results.

The main contributions of this paper include: (i) a novel multiphase approach that incorporates aggregation, Variable Neighbourhood Search (VNS) and an exact method for solving large p-median problems, (ii) new best solutions for some benchmark problems, (iii) the construction of a new large dataset for p-median problems with guaranteed optimality, and (iv) an adaptation of the proposed approach for the conditional p-median problem.

The paper is organised as follows. Section 2 presents a brief review of the past efforts at solving large p-median problems. Section 3 describes the ingredients that make up our method as well as the overall algorithm. Detailed explanations of the main steps and the "Local Search with Shaking" procedure are described in Section 4. Computational results are presented in Section 5 using large datasets including the one with guaranteed optimal solutions which we constructed. Section 6 presents a brief review on the conditional p-median problem followed by the adaptation and the implementation of our approach for this related problem. The last section provides a summary of our findings and highlights some avenues for future research.

2. Past efforts at solving large p-median problems

This section presents an overview of past efforts at solving large p-median problems (see Francis et al., 2009, for an excellent review). Hillsman and Rhoda (1978) introduced a classification of aggregation errors using three types, namely source A, B, and C errors. Source A error occurs when the distance between an Aggregate Spatial Unit (ASU) and a facility is utilised in the model, instead of the true distance between a Basic Spatial Unit (BSU) and a facility. Source B error exists in the special case when a facility is located at an ASU whereas source C error appears when a BSU is assigned to the wrong facility.

Goodchild (1979) stated that aggregation tends to produce more dramatic effects on location than on the values of the objective function while also noting that there is no aggregation scheme without a possible resulting error. Bach (1981) mentioned that "the level of aggregation exerts a strong influence on the optimal locational patterns as well as on the values of the locational criteria". Mirchandani and Reilly (1986) examined the effect of replacing distances to demand points (BSUs) in a region by the distance to a single point (ASU) representing that region.

Current and Schilling (1987) proposed a method for eliminating source A and source B errors. They introduced a novel way of measuring aggregated weighted travel distances for p-median problems. Let $d(i,j)$ denote the distance between the ith and the jth BSUs and $\tilde{d}(k,j)$ the distance between the representative point of the kth ASU and the jth BSU. The distance between the kth ASU and the jth facility is traditionally defined as:

$$\hat{d}(k,j) = W_k \tilde{d}(k,j) \tag{7}$$

where $W_k = \sum_{i \in A_k} w_i$ with A_k being the set of aggregated BSUs at the kth ASU.

To eliminate source A and B errors, the distance proposed in Current and Schilling (1987) is set as:

$$\hat{d}(k,j) = \sum_{i \in A_k} w_i d(i,j) \tag{8}$$

However, this method is not able to eliminate source C errors.

Casillas (1987) introduced two measures to assess the accuracy of aggregated models. These include the cost error ($ce = f(F':C) - f(F':C')$) and the optimality error ($oe = f(F:C) - f(F':C)$) where F and F' represent the optimal locations of the p facilities found with the original and the aggregated models respectively, while C and C' denote the list of BSUs and ASUs. The objective functions $f(F:C)$, $f(F':C)$ and $f(F':C')$ represent the objective function evaluated using F and C, F' and C, and F' and C' respectively.

Oshawa, Koshizuka, and Kurita (1991) studied the location error and the cost error due to "rounding" in the unweighted 1-median and 1-centre problems in the one-dimensional continuous space. Aggregation error bounds for the median and the centre problems were developed by Francis and Lowe (1992). A Geographical Information System (GIS) method for eliminating source C error was proposed by Hodgson and Neuman (1993). Transport costing errors for the median problems were investigated by Ballou (1994) who demonstrated that cost errors increase with p but decrease with m. An investigation by Fotheringham, Densham, and Curtis (1995) suggested that the level of aggregation affects the location error more significantly than the objective function value. Francis, Lowe, and Rayco (1996) introduced a median row-column aggregation method to find an aggregation which gives a small error bound. In addition to the A, B, and C errors, Hodgson, Shmulevitz, and Körkel (1997) introduced source D error which arises when the BSU locations act as potential sites.

Murray and Gottsegen (1997) investigated the influence of data aggregation on the stability of facility locations and the objective function for the planar p-median model. Demand point aggregation procedures for the p-median and the p-centre network location models were studied by Andersson, Francis, Normark, and Rayco (1998). Hodgson and Salhi (1998) proposed a quadtree-based technique to eliminate source A, B, and C errors in the allocation process. Bowerman, Calamai, and Brent Hall (1999) investigated the demand partitioning method for reducing source A, B, and C aggregation errors in p-median problems. Erkut and Bozkaya (1999) provided a review of aggregation errors for the p-median problem. Francis, Lowe, and Rayco (2000) computed error bounds for several location models. Plastria (2001) investigated how to minimise aggregation errors when selecting the ASUs location at which to aggregate given groups of BSUs. Hodgson (2002) introduced data surrogation error in the p-median problem which appears when an original population's demand is substituted by inappropriate values.

To solve large p-median problems without aggregation, Church (2003) put forward an enhanced Mixed Integer Linear Programming formulation called COBRA. He also proved that there are redundant assignment variables that can be consolidated if they satisfy some equivalent assignment conditions. These conditions