COBRA: A New Formulation of the Classic $p$-Median Location Problem

RICHARD L. CHURCH  
church@geog.ucsb.edu
Department of Geography and The National Center for Geographic Information and Analysis, University of California at Santa Barbara, Santa Barbara, CA 93106-4060, USA

Abstract. The $p$-median problem was first formulated as an integer-linear programming problem by ReVelle and Swain (1970) and further revised by Rosing, ReVelle and Rosing-Vogelaar (1979). These two forms have withstood the test of time, as they have been used by virtually everyone since then. We prove that a property associated with geographical proximity makes it possible to eliminate many of the model variables through a substitution process. This new substitution technique has resulted in the elimination of up to 60% of the variables needed in either of these classic model formulations.

Keywords: $p$-median problem, facility location, reformulation, integer optimization

Introduction

The $p$-median location problem was originally defined by Hakimi (1964, 1965) on a network of nodes and arcs. Hakimi assumed that each node represented a point of demand as well as a potential facility site. The level of demand at a given node was expressed as a weight. The $p$-median location problem involves the placement of $p$-facilities upon the network in such a way that the total weighted-distance of serving all demand is minimized. In calculating the total weighted-distance, it is assumed that each demand is served by their closest facility location. ReVelle and Swain (1970) demonstrated that minimizing average weighted-distance was equivalent to minimizing average distance in serving all demand. Minimizing average distance represents making a facility set as accessible as possible to a user population. Given the fact that the $p$-median model incorporates an important objective in the provision of public services (i.e., maximizing access), it has been often referred to as a public sector location model (ReVelle, 1987). From a broader perspective, the $p$-median location model is the network equivalent to the multi-facility Weber problem defined on a Cartesian plane. The objective of this paper is to present a new formulation for this classic model.

Hakimi proved that at least one optimal $p$-median solution to a given problem was comprised entirely of nodes. Given this property, the search for a “best” or optimal solution can be narrowed to a search for the best all-node solution. In 1970 ReVelle and Swain presented the first integer-linear programming formulation for the $p$-median location problem. One of the constraints in this formulation utilizes a structure originally proposed by Balinski (1965) in a plant location problem. This formulation has with-
stood the test of time, as it is the one that has been used by virtually everyone since then. In 1979 Rosing, ReVelle and Rosing-Vogelaar produced a hybrid formulation of the \( p \)-median model by using fewer Balinski constraints and adding a constraint form used by Efroymson and Ray (1966) in a formulation of a simple plant location problem. This hybrid model is a unique combination of constraints that makes it possible to effectively reduce the number of constraints needed while retaining many of the integer-friendly properties of the original problem. Rosing, ReVelle and Rosing-Vogelaar also showed how one can eliminate \( p - 1 \) assignment variables that are associated with the furthest sites from a given demand. The original model and the newer hybrid model are described in greater detail in the next two sections of this paper. We also include a discussion on different approaches that have been developed to solve the \( p \)-median problem. Given this background, we then introduce a new model element: a new variable substitution procedure. We prove that a property associated with geographical proximity with near and far sites can be invoked in a problem setting that makes it possible to eliminate a large number of model variables through a substitution process without any loss of generality. We then present a reformulation of the classic \( p \)-median model based upon the hybrid structure of Rosing, ReVelle and Rosing-Vogelaar and the variable substitution process. This new model is then tested for a small sample of problems.

1. Formulating and solving the \( p \)-median problem

The \( p \)-median model has traditionally been formulated as it originally appeared in ReVelle and Swain (1970). This model, given in its most general form, is defined on a network of nodes and arcs where each node is assumed to represent a local area of demand as well as represent a potential position for a facility. The objective is to locate exactly \( p \)-facilities so that the total weighted distance in serving all demand is minimized.

It is assumed that facilities are not limited in service by a maximum capacity, so each demand can be served by its closest located facility. We can formulate this model given the following notation:

- \( i, j \) = indices used to refer to a node or point, numbered as 1, 2, \ldots, \( n \).
- \( d_{ij} \) = shortest distance from node \( i \) to node \( j \).
- \( a_i \) = demand at node \( i \).
- \( x_{ij} = \begin{cases} 1, & \text{if demand at } i \text{ assigns to facility at } j, \\ 0, & \text{otherwise.} \end{cases} \)
- \( x_{jj} = \begin{cases} 1, & \text{if a facility is sited at site } j \text{ and demand at } j \text{ assigns to it as well,} \\ 0, & \text{otherwise.} \end{cases} \)
- \( p \) = the number of facilities that are to be located.