Clustering and Multifacility Location with Constraints via Distance Function Penalty Methods and DC Programming

Tuyen Tran  
*Portland State University*

Samuel Reynolds  
*Portland State University*

Thai An Nguyen  
*Duy Tan University*

Mau Nam Nguyen  
*Portland State University*

Follow this and additional works at: [https://pdxscholar.library.pdx.edu/studentsymposium](https://pdxscholar.library.pdx.edu/studentsymposium)

Let us know how access to this document benefits you.

Tran, Tuyen; Reynolds, Samuel; Nguyen, Thai An; and Nguyen, Mau Nam, "Clustering and Multifacility Location with Constraints via Distance Function Penalty Methods and DC Programming" (2018). *Student Research Symposium*. 16. 
[https://pdxscholar.library.pdx.edu/studentsymposium/2018/Poster/16](https://pdxscholar.library.pdx.edu/studentsymposium/2018/Poster/16)

This Poster is brought to you for free and open access. It has been accepted for inclusion in Student Research Symposium by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.
Abstract

This is a continuation of our effort to use mathematical optimization involving DC programming in clustering and multifacility location. We study a penalty method based on distance functions and apply it particularly to a number of problems in clustering and multifacility location in which the centers to be found must lie in some given set constraints. We also provide numerical examples to test our method.

In the current time of “big data”, clustering is a very important problem that helps classify data in many fields such as machine learning, pattern recognition, image analysis, data compression, and computer graphics. Given a finite number of data points with a measurement distance, a centroid-based clustering problem seeks a finite number of data points with a measurement distance, which are the centers to be found.

Clustering with Constraints

We study problems of clustering with constraints in which the measurement distance is defined by the squared Euclidean norm. In the case where the measurement distance is defined by the Euclidean norm instead of the squared Euclidean distance functions to convert constrained problems to unconstrained problems. Then appropriate DC decompositions are used to solve a centroid-based clustering problem seeks a finite number of data points with a measurement distance, which are the centers to be found.

Set Clustering with Constraints

In this section, we turn our attention to a model of set clustering with constraints, i.e., for given m subsets $\Lambda_1, \ldots, \Lambda_m \subset \mathbb{R}^d$, we seek k cluster centers $x^* \in \bigcap_{i=1}^m \Omega_i^k$ for $\ell = 1, \ldots, k$, where each $\Omega_i^k$ is a subset of $\mathbb{R}^d$. The measurement distance is defined by the squared distance functions to the sets involved. The optimization modeling of the problem to be solved is given by

$$\begin{align*}
\min & \quad \psi(x^*; \Lambda^k) = \sum_{\ell=1}^k \min_{i,j \in \Lambda^k} \| x^\ell - a^j \|^2 \\
\text{subject to} & \quad x^\ell \in \bigcap_{i=1}^m \Omega_i^k \quad \forall \ell = 1, \ldots, k.
\end{align*}$$

Using the penalty method based on distance functions with a parameter $\tau > 0$, we consider the constrained set clustering model:

$$\begin{align*}
\min & \quad f(x^1, \ldots, x^k) = \frac{1}{2} \sum_{\ell=1}^k \min_{i,j \in \Lambda} \| d(x^\ell; \Lambda^j) \|^2 + \frac{\tau}{2} \sum_{i=1}^n \sum_{j=1}^m \| d(x^\ell; \Omega_i^j) \|^2, \\
\text{subject to} & \quad x^\ell \in \mathbb{R}^d, \quad \forall \ell = 1, \ldots, k.
\end{align*}$$

Multifacility Location with Constraints

Given a set of m points (nodes) $a^1, a^2, \ldots, a^n$ in $\mathbb{R}^d$, our goal is to find k centers $x^\ell$ for $\ell = 1, \ldots, k$, such that the transportation cost to the nodes is minimized. The total cost is given by

$$\psi(X) = \psi(x^1, \ldots, x^k) = \sum_{i=1}^n \min_{\ell=1}^k \| x^\ell - a^i \|.$$  

We apply Nesterov’s smoothing techniques from [4] with smoothing parameter $\mu$ and penalty parameter $\tau$ to approximate the objective function $f$, by a new DC function which is favorable for applying the DCA.

$$f_{\mu, \tau}(x^1, \ldots, x^k) = \left( \frac{\mu}{2} \sum_{i=1}^n \sum_{\ell=1}^k \frac{\| x^\ell - a^i \|^2}{\mu} \right) + \tau \sum_{\ell=1}^k \sum_{i=1}^n \| x^\ell - a^i \|^2$$

$$- \left( \frac{\mu}{2} \sum_{i=1}^n \sum_{\ell=1}^k \frac{d(x^\ell; \Omega^i)}{\mu} \right) + \sum_{\ell=1}^k \sum_{i=1}^n \max_{j \in \Lambda^k} \frac{\| x^\ell - a^j \|}{\mu} + \frac{\tau}{2} \sum_{i=1}^n \sum_{j=1}^m \| d(x^\ell; \Omega_i^j) \|^2.$$  

References