Using of Clustering Algorithm CWSP-PAM for Rural Network Planning

Lamiaa Fattouh Ibrahim
Department of Computer Sciences and Information
Institute of Statistical Studies and Research, Cairo University
Giza, EGYPT
Currently Department of Computer Sciences
Faculty of Sciences, King Abd El Aziz university
Jeddah, Saudi Arabia
lfattouh@mailer.eun.eg

Abstract

Network planning is of key importance during the construction of new communities and cities, in which telephone and data services have to be introduced as a component of the overall master plan of the city. The system accepts the map of a city in the form of its streets and intersection nodes coordinates, the specifications of the available cable sizes, and the cost information for the cables. The system determines the minimum cost network that satisfies the demand and constraints. The problem is treated as a clustering around medoids problem where the distances are represented by weighted shortest paths. In this paper, the Partitioning Around Medoids (PAM) original algorithm have been modified. Results demonstrate the effectiveness and flexibility of the modifying algorithm in tackling the important problem of rural network planning. Comparisons with related work are presented showing the advantages of the CWSP-PAM (Clustering with Shortest Path-PAM) algorithm introduced in this paper.

1. Introduction

The network planning process has to consider a variety of constraints including: policy of administrations, planning objective, etc, there is no universal method that is applicable to all network planning problems. Due to the complexity of this process artificial intelligence (AI) [1], [2] and clustering techniques [3], [4], [5] has been successfully deployed in a number of areas. The process of network planning is divided into two sub problems: Determining the location of the exchanges and Determining the layout of the subscribers' network lines paths from the exchange to the subscribers premises while satisfying both cost optimization criteria and design constraints[1].

The goal of a clustering algorithm is to partition a given data set into clusters or groups, which are not predefined, such that the data points in a cluster are similar to each other more than points in different clusters [6]. These groups are formed according to some measures of goodness that differ according to application. Clustering Problem is defined as follows [7]:

- Given a database $D=\{t_1,t_2,\ldots,t_n\}$ of tuples and an integer value k, the *Clustering Problem* is to define a mapping $f : D\rightarrow\{1, .., k\}$ where each t_i is assigned to one cluster Kj, $1<= j <= k$.
- A *Cluster*, Kj, contains precisely those tuples mapped to it.

The PAM (Partioning Around Medoids) algorithm, also called K-medoids algorithm, represents a cluster by medoid. Initially, the number of desired clusters is input and a random set of K items is taken to be the set of medoids. Then at each step, all items from the input dataset that are not currently medoids are examined one by one to see if they should be medoids. That is, the algorithm determines whether there is an item that should replace one of the existing medoids. By looking at all pairs of medoids, non-medoids objects, the algorithm chooses the pair that improves the overall quality of the clustering the best and exchages them. Quality here is measured by the sum of all distances from a non-medoids object to the medoids for the cluster it is in. A item is assigned to the cluster represented by the medoid to which it is closest (minimum distance or direct Euclidean distance between the customers and the center of the cluster they belong to). We assign the location of medoids to the location of switchs in the problem of netwok planning.

In network planning, the direct Euclidean distance ignores the presence of streets and paths that must be taken into consideration during clustering. In this paper, The Partitioning Around Medoids (PAM) original algorithm have been modified depending on using the weighted physical shortest available path. The weights used are assigned to represent subscribers' loads. The constraints of cable length are used to determine the number of clusters. The CWSP-PAM algorithm is developed in the spirit of PAM algorithm.

Proceedings of the Third International Conference on Information Technology and Applications (ICITA'05)
0-7695-2316-1/05 $20.00