On Transforming a Road Network Database to a Graph for Localization Purpose

Xiangli Meng, School of Technology and Business Studies, Dalarna University, Falun, Sweden & School of Statistics, Shandong University of Finance and Economics, Jinan, China
Pascal Rebreyend, School of Technology and Business Studies, Dalarna University, Falun, Sweden

ABSTRACT

The problems of finding best facility locations require complete and accurate road networks with the corresponding population data in a specific area. However the data obtained from road network databases usually do not fit in this usage. In this paper the authors propose a procedure of converting the road network database to a road graph which could be used for localization problems. Several challenging problems exist in the transformation process which are commonly met also in other data bases. The procedure of dealing with those challenges are proposed. The data come from the National road data base in Sweden. The graph derived is cleaned, and reduced to a suitable level for localization problems. The residential points are also processed in ordered to match the graph. The reduction of the graph is done maintaining the accuracy of distance measures in the network.

KEYWORDS
GIS, Graph, Population, Road Network

1. INTRODUCTION

Consider the common location-allocation model (the $p$-median problem) which allocate $P$ facilities to a population geographically distributed in $Q$ demand points such that the population’s average or total distance to its nearest service facility is minimized. Solving such problems is based on computing the distance between candidate locations and all or a subset of points representing the habitat of people. The candidate locations should be connected in a graph and the corresponding distance matrix should be suitable for further calculations. A popular approach is to avoid the trouble caused by transforming the original data into a connected road network and instead use the Euclidian distance between the candidate locations and residential points. Thus, all points are assumed to be connected pairwise via straight lines. However, the Euclidian distance may be inaccurate, especially if two points are not directly connected and require detours, in which case the Euclidian distance would underestimate the real distance. Thus the quality of the distance is deteriorated, particularly in a area with a lot natural barriers. Consequently, the usage of Euclidian distance would lead to poor locations of the model.

Han & al (2013) shows that using the road network to compute distances instead gives a good treatment of the problem. The road network distance gives accurate estimate of distance between 2 points. It reflects the road connections and natural barriers. In their analysis, although the complexity of the problem leads to sub-optimal solutions, the road network distance solutions still outperforms those of Euclidian ones.

However, even though the usage of the road network distance is essential in solving the models, the original network data sets researchers obtained usually do not fit and are unable to be used directly. Those data sets need to be cleaned and transformed to a connected road network graph before being
used for finding best locations of facilities. Many challenging problems could pop out in this process such as incomplete information or isolated sub graph. Dealing with these problems inappropriately would produce inaccurate graph and then lead to bad solutions. For data sets in a large scale, the necessity of giving proper treatment of the problems in the data set and the transformation process is even more crucial.

Despite the nontrivial role an accurate road network graph plays in solving the location models, few research papers give proper illustrations on how to deal with the troubles in the process of deriving it. Especially for data at a large scale for the \( p \)-median models, there is no research investigation on how to deal with the challenges in the transformation process, which is the focus of this paper. In our case, we are interested in locating public facilities such as hospitals, public services at the scale of the country Sweden. The objective function we want to minimize is the average distance between residence’s habitants and the closest facilities nationwide. The scale of the problem is very large, making the quality for graph and distance matrix vital. The original data we get is from the National road data base (NVDB), provided by the National Swedish Road Agency (Trafikverket). It is a very detailed data set consisting of over 34 million data points with auxiliary information like speed limits and direction. Many commonly met challenges appear in the transformation process. Proper methods need to be used for handling those challenges before getting the graph for solving the \( p \)-median models. The main challenges we encountered are as follows.

A. Filling in the missing crossing and connectivity information properly.
B. Reducing a super large graph to a manageable scale and keep the distance information as accurate as possible.
C. Calculating the distance between different nodes in an affordable time.

These 3 problems are commonly met in deriving large graphs for location-allocation models. Inappropriate filling of missing information would result in incorrect graph, giving wrong distances. In our data set, cross and connection information are missing, which need to be filled properly. Wrong fill of the cross would connect 2 separated roads ‘together’. Adding connections to isolated sub graph incorrectly usually underestimates distances. Reducing the large graph is necessary since our data set is based on national level, therefore the scale itself to this problem would cause trouble for further analysis. Due to the fact that \( p \)-median problems call for the distance between any pair of nodes, the distance matrix corresponding with the graph would have very large dimensions, and it would be beyond the ability of the computers to process with heuristic algorithms or other methods. Thus it is crucial and tricky to reduce the data appropriately to an affordable level without losing too much accuracy. The commonly used Dijkstra algorithm could give the distance between any pair of nodes, but it consumes a very long time, making it inconvenient for researchers to make adjustment to it and difficult to derive large distance matrix. Thus the algorithm needs to be refined to be able to get the distance matrix in an affordable time.

Sweden is about 450,000 km\(^2\) in area with 10 million population. It is fairly large in Europe and the world. The Swedish road networks are quite advanced and connect almost all the parts of the country. Deriving a graph in such a large scale is quite innovative and could provide helpful techniques to researchers. The methods we proposed to deal with challenges are indicative for data at similar scales. We have tried our approach on different regions separately in Sweden, and it worked well. It shows our methods have good adaptability also for data with smaller scales. The region sizes are quite comparable to many other areas. For example, the smallest regions in Sweden are of the size of Luxembourg, the middle sized are comparable with Belgium and German states, and the largest
Advanced Data Compression Techniques for SOAP Web Services
www.igi-global.com/chapter/advanced-data-compression-techniques-soap/26914?camid=4v1a

The Full Knowledge of Big Data in the Integration of Inter-Organizational Information: An Approach Focused on Decision Making
www.igi-global.com/chapter/the-full-knowledge-of-big-data-in-the-integration-of-inter-organizational-information/217864?camid=4v1a