METHODS OF APPROXIMATE SOLUTION SEARCHING FOR A SHORTEST PATH PROBLEM IN A LARGE GRAPH

A.V. Zinoviev

Omsk State Univercity, Omsk
e-mail: zaleslaw@gmail.com

As computer technology advances, modeling and solving problems based on large data sets (Big Data) becomes increasingly more relevant. Since the beginning of the 16st century, problems relating to the processing of large data sets have started to appear in many different areas. For example: astronomical observations and calculations of the planets' trajectories; weather forecasting and natural disaster prediction; data processing of trading operations; and searching optimal sea routes.

Much of the research which is conducted using big data analysis can be modeled using large graphs. Examples of the large graphs include: web graph (mathematical model of the Internet); Gmail e-mail messages graph; European road network; FriendsGraph on Facebook; Google Knowledge Graph; and citation graph (graph based on citations of academic publications).The number of vertices and edges in such graphs is estimated to be millions and billions. Usual algorithms, including shortest path algorithms – like Dijkstra and Floyd-Warshall – are ineffective and require a lot of memory to store the graph's intermediate states.

This Report presents modern approaches to the construction, processing and storage of large graphs, adopted by big companies such as Google and Microsoft. A fairly popular approach is to use two-phase algorithms ALT (Goldberg, Harrelson, 2005), TN (Geisberger, 2008), HL (Abraham, 2011). The first phase includes graph pre-processing and saving the results to a database. The second phase includes database query execution and final calculations. A basic assumption is that a road network varies only slightly over time, so it's not necessary to frequently conduct the processing phase.

Another area where it is important to be able to correctly handle large graphs is where there are problems of effective flow management in the transportation system; searching for optimal routes; making decisions about road network development; and as a network model typically used as a road graph. The process of building the road graph based on actual data about the current state of the network is described in a paper [1]. However, in the problems described above, the network changes frequently and therefore two-phase algorithms are not applicable. In this case, the bidirectional Dijkstra's algorithm was selected, with a preliminary stage of the road network transformation to a smaller size subnet.

This paper presents a comparative analysis of two-phase algorithms ALT and TN on Russia's road network and develops recommendations for using various shortest path algorithms in different cases.

REFERENCES

1. A.V. Zinoviev *Modeling of the transport network of Omsk, based on public geodata.* — Ekaterinburg, UB RAS: Proceedings of the school – conference "CSEDays 2012: Graphs theory and applications". — 2012, pp. 21-26.