K-means Nearest Point Search Algorithm and Heuristic Search for Transportation

Wai Mar Hlaing and Myint Myint Sein
Geographic Information System Lab.,
University of Computer Studies (UCSY)
Yangon, Myanmar
wai_mar@yahoo.com, myintucsy@gmail.com

***Abstract*— This paper aims to support the most suitable route for passengers of the taxi system using the proposed heuristic search method. Furthermore, k-Means Nearest Point Search (kMNPS) algorithm is proposed to produce the nearest road point for start and end addresses. Yangon downtown in Myanmar is selected as a case study for the transportation system. The proposed heuristic method and kMNPS algorithm reduce the distance calculations and achieve the very low time complexity for the real time transportation applications. Moreover, the proposed system can produce not only the optimal route on the map but also the popular spots near the optimal route.**

Keywords—Nearest Neighbor, K-means Clustering, Heuristic Search, Qgis, Weka

I. Introduction

Public transportation guidance is the very important role for the humans' lives. Geographic information system (GIS) and spatial datasets are widely used to build the intelligence transportation systems. Data mining techniques are applied to analyze the spatial data of the GIS. k-Nearest Neighbor is a well- known method to find the k nearest points based on a spatial query. Dr. Mohnammed Otair [1] presents the k-Nearest Neighbor spatial cluster method using kd-tree. The query can be responded during the short time by using kd-tree but it is difficult to implement. K-Means clustering is a popular unsupervised clustering algorithm to find the clusters based on the data's similar features. Xueyi Wang [2] proposed k-Means for k-Nearest Neighbors (kMkNN) with triangle inequality for searching the nearest neighbors in high dimensional space. The performance of kMkNN is faster than using the kd tree at finding the nearest neighbors.

We proposed the k-Means Nearest Point Search (kMNPS). It is similar to the kMkNN algorithm with triangular method but our proposed method does not analyze the every clusters. kMkNN calculates the distance values between every centroid of all clusters and query. This algorithm finds the maximum value among of the distance values and sets it as the initialized maximum value. Then, the algorithm compares the distance between the maximum value and objects in clusters using triangular method. However, it needs to check the objects in current cluster to move the next cluster and it also needs to compare the distance values between the current object's distance using triangular method and maximum value. The main contribution of proposed algorithm contains

- The cluster is chosen depends on the centroid and average spatial values and it removes the unnecessary clusters for the current query.
- The spatial points of all clusters are not necessary to check whether this point will contain as a nearest point.
- The algorithm needs to calculate the distance values only two small clusters although the large cluster exits.

Most of the intelligence transportation system uses the heuristic search because it is very intuitive method to find the optimal route. It is faster than the backtracking algorithm, genetic algorithm and classical linear search algorithms [3]. A* calculates the distance values between the target node and all nodes in the open list to choose a suitable node for the next level. Our proposed heuristic search method separates the large tree into small subtrees at the depth four of the tree. Each subtree works separately to find the five local optima. It considers on the neighbor nodes of current node to choose the next level. Finally, the optimal route is chosen among of the local optima. It is faster than the heuristic search A* algorithm.

II. Transportation System For Yangon Downtown Area

Yangon downtown is the main transportation township for Yangon Region. Yangon Region is composed of 33 townships and Latha, Lanmadaw, Pabedan, Kyauktada, Botahtaung and Pazundaung townships exist in Yangon downtown area. Quantum geographic information system (Qgis) software is used to create the shape file for Yangon downtown area. The Yangon downtown spatial database contains over 4000 nodes and 10, 000 edges. The shops and popular places near the roads, streets and highways are about 3000 nodes. The proposed system aims to find the optimal route for the taxi transportation system during the fast response time. Moreover, the passenger can see the places near the optimal route.

The rest of this paper is organized as follows. Section III describes the kMNPS algorithm and proposed heuristic search method. Section IV describes the experiments about the transportation system and Section V concludes the paper.

III. K-Means Nearest Point Search Algorithm and Heuristic Search

This section contains two parts. The first part is describing the k-Means Nearest Point Search (kMNPS) and the remaining part describes the proposed heuristic search method. Dividing the big data into small cluster is computed at the pre-processing stage using the data mining tool.

A. K-Means Nearest Point Search Algorithm

At the pre-processing stage, the proposed system finds the clusters for geospatial database using weka software. Weka is a popular data mining tool for data analysis. Fig. 1 describes the cluster results for Yangon downtown area. The number of cluster is decided depend on the mean square error. If the number of cluster is so many, two clusters are not enough to find the nearest point. When the number of cluster is less, mean square error may be high. Then, the suitable number of cluster for Yangon downtown area is 20. After finding the clusters, proposed kMNPS algorithm is used to find the nearest point by removing the unnecessary clusters.

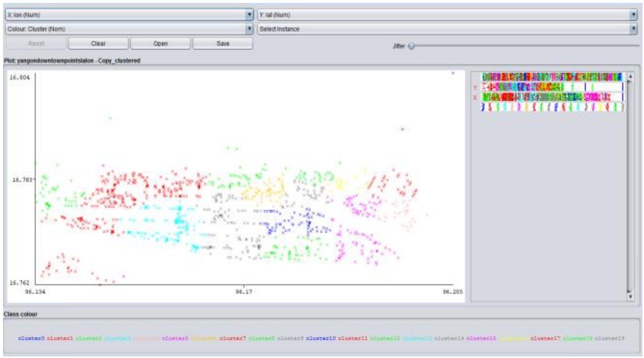

Fig. 1. The cluster results of Yangon Downtown Area

a) Detailed Steps: Detailed steps of the kMNPS algorithm are contained as follows.

1) K clusters are produced using the weka software.
2) The average values of latitude and longitude are calculated for each cluster.
3) A cluster that owns the smallest distance value is chosen as the current optimal cluster after comparing the distance values between the query and centroid of the clusters.
4) The distance value is compared between the average values of the spatial data and query to choose the next cluster. Then, it choses the cluster that owns the closet distance. If the cluster result is equal to the first chose cluster, only one cluster is needed to find the nearest point. Otherwise, the two clusters will be used for the nearest point search. At sometimes, the number of optimal clusters for query can be more than two when the proximity is less between the clusters.

b) Working Example of Nearest Road Point Search: When a customer calls a taxi, the system will get the current location l_1 of the customer. Then, the system finds the nearest road point p_1 for this location l_1 by using the proposed kMNPS algorithm. There are three main steps in proposed algorithm. The first step is to compare the distances between the location l_1 and centroids of the clusters. The second step is to compare the distances between the location l_1 and average spatial values of each cluster. At the third step, the system compares the point l_1 and all spatial values of the clusters that got from step three and four of kMNPS algorithm and finally it produces the nearest road point p_1.

c) Working Example of Popular Places Search Near the Optimal Route: Proposed heuristic search that describes in section III(B) is used to find the optimal route. After finding the optimal route, the system finds the popular places near this route. The system finds the popular places at every third point along the optimal route. For example, if the optimal route contains twelve points, the system finds the popular places at point number three, six, nine and twelve using kMNPS algorithm. The system shows the popular places around one kilometer for these points.

B. *Proposed Heuristic Search*

After finding the nearest road point using kMNPS, proposed heuristic search is used to find the optimal route. The proposed method contains three steps. The first step is dividing the large tree into small subtrees at the depth four of the tree. The second step finds the local optima of each subtree using heuristic search like A*. After finding the local optima, the final step compares the total distance values of local optima to get the optimal one.

IV. EXPERIMENTAL RESULTS

The proposed system can find the nearest point during the short time by dividing the clusters and removing the unnecessary clusters for the query. Moreover, it can find the optimal route using the proposed heuristic search. It is faster than the popular heuristic search A* algorithm because the proposed method does not need to consider all nodes in open lists to expand the next level. It bases on the neighbors of the current node. Fig. 2 describes the optimal route, the shops and interesting places near this route. When the gray triangle is pressed, the information of the shop can be viewed. The green icon is the starting point and yellow icon represents the end point.

Fig. 2. Finding the optimal route and nearest places on Yangon downtown area

V. CONCLUSION

This paper proposes two methods to find the nearest point and optimal route. kMNPS is faster than the kMkNN algorithm but it intends to find only one nearest point. Heuristic search is used to find the optimal route. It is faster than popular heuristic search A* algorithm. The system uses the divide and conquer technique at finding the optimal route and it considers only the neighbor nodes of current node. Therefore, the system can achieve the optimum route with very low time complexity. Moreover, the passenger can view the shops and popular places near the optimal route.

REFERENCES

[1] M. Otair, "A PPROXIMATE K -N EAREST N EIGHBOUR BASED S PATIAL C LUSTERING USING K- D TREE," vol. 5, no. 1, pp. 97–108, 2013.

[2] X. Wang, "A Fast Exact," pp. 1293–1299, 2011. X.Wang, "A Fast Exact k-Nearest Neighbors Algorithm for High Dimenstional Search Using k-Means Clustering and Triangle Inequality," in International Joint Conference on Neural Networks, 2011, pp. 1293–1299.

[3] A. Prof and M. Karova, "Comparative analysis of algorithms to search for the shortest path in a maze," pp. 0–3, 2016.