2Q-05

Characterizing Areas on Road Networks by Distances from Facilities

Maulana ARIEF　　Kazumi SAITO　　Tetsuo IKEDA　　Hiroaki YUZE
Graduate School of Management and Information of Innovation, University of Shizuoka

1 Introduction

Studies of the structure and functions of large complex networks have attracted a great deal of attention in many different fields such as sociology, biology, physics and computer science [1]. As a particular class, we focus on spatial networks embedded in the real space, like road networks, whose nodes occupy precise positions in a two- or three-dimensional Euclidean space, and whose links are real physical connections between them [2]. For such spatial networks, we explore techniques for extracting areas with some coherent characteristics. Although this task is somehow relating to community extraction ones widely studied in complex network analysis [1], unlike these general techniques only using network structure, we focus on exploiting the distances from positions of some pivotal points, like facilities, on the spatial network.

In this paper, we propose a new method for automatically classifying and characterizing some areas on a given road network, based on distances from facility positions. The proposed method first computes a feature vector for each node, consisting of the minimum distances from several types of facilities, then classifies these vectors into some areas by the use of the greedy K-medoids clustering algorithm, and finally provide each of the classified areas with the urbanness degree as the inverse harmonic average of the feature value for the corresponding medoid vector. In our experiments, using the three road networks of three cities in Shizuoka prefecture collected from OpenStreetMap (OSM), we show that our method can produce some promising results.

2 Proposed Method

Let $G = (\mathcal{V}, \mathcal{E})$ be a given road network constructed by mapping the intersections of roads into nodes and the roads between the nodes into links. where $\mathcal{V} = \{u, v, \cdots\}$ and $\mathcal{E} = \{(u, v), \cdots\}$ mean the sets of nodes and links, respectively. Here, we consider M types of facilities, and let $\mathcal{A}(m)$ be the set of facilities of type $m \in \{1, \cdots, M\}$, where the position of each facility $a \in \mathcal{A}(m)$ is identified with the position of the nearest node $u \in \mathcal{V}$. Then, for each node $v \in \mathcal{V}$, we can define the m-th feature by the minimum distance from the facilities of type m as

$$x_v(m) = \min_{a \in \mathcal{A}(m)} d(v, a),$$

where $d(v, a)$ stands for the geodesic distance between the positions of nodes v and a over the road network G. Thus, for each node $v \in \mathcal{V}$, we can compute the M-dimensional feature vector expressed as $\mathbf{x}_v^T = (x_v(1), \cdots, x_v(M))$, where $\mathbf{x}_v^T$ means the transposed vector of $\mathbf{x}_v$.

Let $\mathcal{X} = \{\mathbf{x}_u, \mathbf{x}_v, \cdots\}$ be the set of feature vectors computed by the above procedure. Then, we consider classifying $\mathcal{V}$ into K areas, $\mathcal{V}_1, \cdots, \mathcal{V}_K$, by the K-medoids algorithm. More specifically, we select the set of medoids (representative vectors) $\mathcal{R} \subset \mathcal{X}$ whose number of elements is K, i.e., $|\mathcal{R}| = K$, so as to maximize the following objective function based on the cosine similarity:

$$f(\mathcal{R}) = \sum_{\mathbf{x} \in \mathcal{X}} \max_{\mathbf{r} \in \mathcal{R}} \left\{ \frac{\mathbf{x}^T \mathbf{r}}{\sqrt{\mathbf{x}^T \mathbf{x}} \sqrt{\mathbf{r}^T \mathbf{r}}} \right\}.$$

Then, the greedy K-medoids algorithm can be summarized as follows: after initializing $k \leftarrow 1$ and $\mathcal{R} \leftarrow \emptyset$, we repeatedly select and add each medoid by

$$\mathbf{r}_k \leftarrow \max_{\mathbf{x} \in \mathcal{X} \setminus \mathcal{R}} \{ f(\mathcal{R} \cup \{\mathbf{x}\}) - f(\mathcal{R}) \}, \ \mathcal{R} \leftarrow \mathcal{R} \cup \{\mathbf{r}_k\},$$

during $|\mathcal{R}| \leq K$ together with increment $k \leftarrow k + 1$. From the obtained set of the K medoids, $\mathcal{R} = \{\mathbf{r}_1, \cdots, \mathbf{r}_K\}$, we can compute each classified area as

$$\mathcal{V}_k = \left\{ v \in \mathcal{V}; \mathbf{r}_k = \max_{\mathbf{r} \in \mathcal{R}} \{\rho(\mathbf{x}_v, \mathbf{r})\} \right\}.$$

For the classified areas, the closeness to various types of facilities can be regarded as one reasonable measure of the urbanness. Thus, we provide each classified area $\mathcal{V}_k$ with the urbanness degree U_k as the inverse harmonic average of the feature values for the corresponding medoid vector $\mathbf{r}_k$ defined by

$$U_k = \frac{1}{M} \sum_{m=1}^{M} \frac{1}{r_k(m)},$$

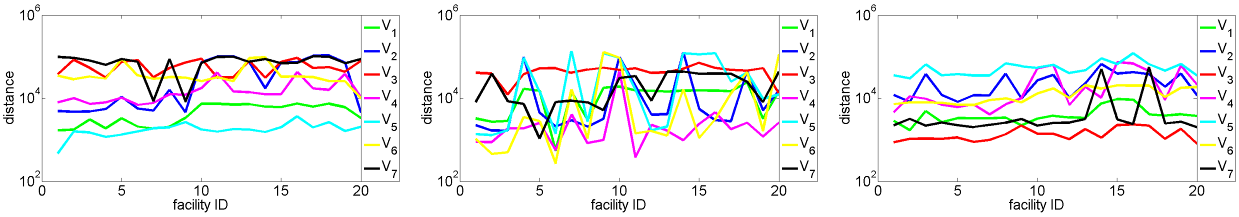
for each $k \in \{1, \cdots, K\}$. Note that we employ the above inverse hamonic average version as our initial study, although it might be possible to device more elaborate ones.

3 Experiments

We used OSM data of three cities, Hamamatsu, Numazu and Shizuoka, in Shizuoka prefecture, where we extracted the largest connected components from all highways and all nodes appearing in them, and deleted continuous curve-fitting-points of roads. Then, the numbers of nodes and links became $104,813$ and $127,648$ for Hamamatsu, $15,477$ and $19,046$ for Numazu, and $53,903$ and $66,444$ for Shizuoka. From OSM data, we also extracted all facilities located in Shizuoka prefecture. Table 1 shows the twenty types of facilities sorted according to the number of registered facilities in Shizuoka prefecture.

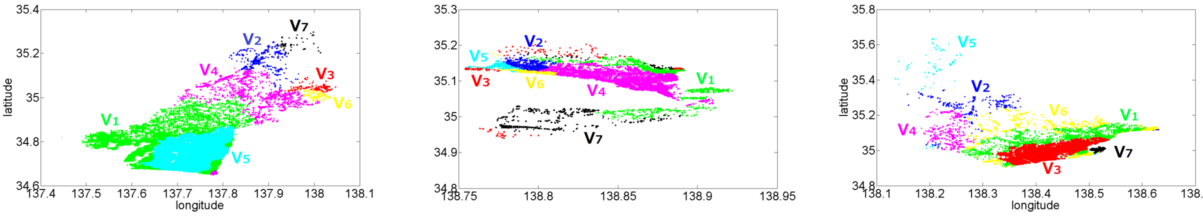
We applied our proposed method to road networks of three cities by setting the number of areas to seven, i.e.,

Copyright ©2017 Information Processing Society of Japan.
All Rights Reserved.



(a) Hamamatsu (b) Numazu (c) Shizuoka

Figure 1: Feature values of medoid vectors



(a) Hamamatsu (b) Numazu (c) Shizuoka

Figure 2: Classified areas by our method

Table 1: 20 types of facilities in our experiments.

i	A	A_i	i	A	A_i
1	kindergarten	999	11	bank	250
2	social facility	929	12	post box	244
3	restaurant	922	13	public building	233
4	school	903	14	hospital	217
5	place of worship	693	15	cafe	197
6	post office	597	16	pub	186
7	police	398	17	fast food	182
8	fuel	349	18	telephone	179
9	toilets	342	19	pharmacy	169
10	doctors	260	20	fire station	166

$K = 7$. Figure 1 shows the feature values of each medoid vector $\mathbf{r}_k \in \mathcal{R}$, where the horizontal and vertical axes stand for the types of facilities and the minimum distances from each type of facilities. Figure 2 plots the nodes at their positions with colors of their classified area as shown in Fig. 1. We can see that these visualization results are reasonably interpretable in the sense that several areas such urban, mountainous and intermediate regions were clearly classified by using different colors. Table 2 shows the urbanness degree of each area sorted according to its rank, From these experimental results, we can see that the feature values for the corresponding medoid vector can be used as a reasonable measure of the urbanness.

Table 2: Urbanness degree U_k of each city

Rank	Hamamatsu		Numazu		Shizuoka	
1	$\mathcal{V}_5$	6.392×10^{-04}	$\mathcal{V}_4$	7.773×10^{-04}	$\mathcal{V}_3$	8.046×10^{-04}
2	$\mathcal{V}_1$	2.942×10^{-04}	$\mathcal{V}_6$	7.327×10^{-04}	$\mathcal{V}_7$	3.645×10^{-04}
3	$\mathcal{V}_2$	9.706×10^{-05}	$\mathcal{V}_2$	2.608×10^{-04}	$\mathcal{V}_1$	2.847×10^{-04}
4	$\mathcal{V}_4$	8.331×10^{-05}	$\mathcal{V}_5$	2.343×10^{-04}	$\mathcal{V}_4$	9.491×10^{-05}
5	$\mathcal{V}_6$	3.307×10^{-05}	$\mathcal{V}_1$	2.201×10^{-04}	$\mathcal{V}_6$	9.176×10^{-05}
6	$\mathcal{V}_7$	2.285×10^{-05}	$\mathcal{V}_7$	1.185×10^{-04}	$\mathcal{V}_2$	5.982×10^{-05}
7	$\mathcal{V}_3$	1.959×10^{-05}	$\mathcal{V}_3$	2.683×10^{-05}	$\mathcal{V}_5$	2.070×10^{-05}

Below we explain two examples for clarifying characteristics of the classified areas. First, from Figs. 1 (c) and 2 (c), we can see that as for the area 7 including Miho in Shizuoka city ($\mathcal{V}_7$ colored in black), the distances from the medoid $\mathbf{r}_7$ to the nearest hospital (facility ID 14) and fast food restaurant (facility ID 17) are remarkably longer than those in the area 3 interpreted as the central Shizuoka city ($\mathcal{V}_3$ colored in red). Second, from Figs. 1 (b) and 2 (b), we can see that as for the area 7 including Heda in Numazu city ($\mathcal{V}_7$ colored in black), the distance from the medoid $\mathbf{r}_7$ to the nearest place of worship (facility ID 5) is remarkably shorter in comparison to the area 4 interpreted as the central Numazu city ($\mathcal{V}_4$ colored in magenta). These examples also show that the proposed method is promising for characterizing areas.

4 Conclusion

We proposed a method for classifying and characterizing the areas on road networks by distances from facilities. In our experiments, we confirmed that our method can produce promising results. In future, we plan to evaluate our method using various road networks.

Acknowledgements

This work was supported by Grants-in-Aid for Scientific Research (No.15K00311).

References

[1] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, Vol. 45, pp. 167–256, 2003.

[2] P. Crucitti, V. Latora, and S. Porta. Centrality measures in spatial networks of urban streets. *Physical Review E*, Vol. 73, No. 3, 036125, 2006.

Copyright ©2017 Information Processing Society of Japan. All Rights Reserved.