Spatial Analysis on Social Media-Development of Location-based Contents System with Mobile Devices

Kazunari Ishida
Faculty of Applied Information Science
Hiroshima Institute of Technology
2-1-1 Miyake, Saeki-ku, Hiroshima, 731-5193 Japan
k.ishida.p7@it-hiroshima.ac.jp

ABSTRACT: *This paper investigates the quality of geographical information provided by a location-based service, such as Foursquare, and its geographical distribution. According to the result, geographical information described by autonomous individuals is heavily clustered in metropolitan areas. In order to reduce the geographically-biased distribution of information, location-based content systems have been developed with mobile devices for people in local communities. Members of a local community, e.g., shopping districts and tourist spots, have strong incentives to provide high quality information to their customers. Hence, the systems are provided to people in these communities so that a vast amount of geo-local content is going to be published on the Internet. To deploy the developed system, two local areas are investigated in terms of spatial analysis on social media. The areas are a rural town and a mid-size city. According to the analysis, the members in the rural town have to generate local content concerning their locality because word-of-mouth information is very scarce.*

Keywords: Spatial Analysis Social Media, Mobile Device, Twitter, Foursquare, Location based Service

Received: 12 July 2011, Revised 19 August 2011, Accepted 24 August 2011

1. Introduction

Location based services, e.g., Foursquare and Google place, provide useful functions such as writing information on a specific location and finding friends nearby a user, based on GPS embedded in a mobile device. Because of the amusement and usefulness, they are becoming popular among individuals world-wide. The posted information with location information forms social media. However, the information might be incorrect and geographically-biased because of the information posting behavior, based on an individual interest. This paper clarifies the information error and geographical bias on the social media derived from location information services. The paper also explains a geo-local contents system based on an Android device and a PC server.

2. Location Information

To collect location information posted on Foursquare [1], Twitter was employed as a sampling mechanism all over the world. The duration of the sampling was from Feb. 17 to Jun 28 in 2010. The number of tweets and that of location information were 5,324,981 and 12,361, respectively.

2.1 Geographical Distribution

To observe distribution of the information all over the world, it is divided into 24 categories based on every 15 degrees in terms of longitude. Based on the categories, frequent areas are found around America, Asia, and Europe (Figure 1).

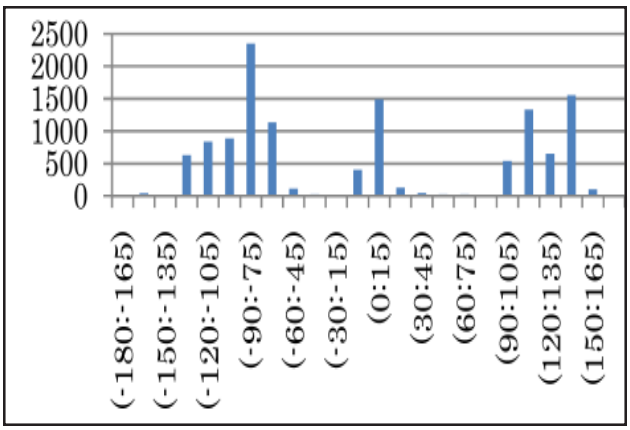


Figure 1. Geographical Distribution on Longitude

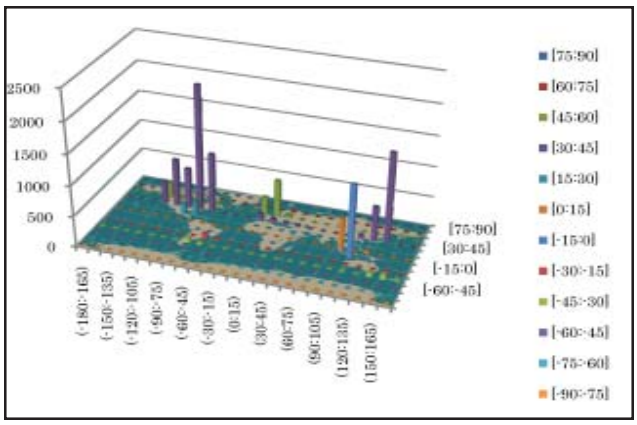


Figure 2. Geographical Distribution on the Earth

To specify a precise area of the distribution, each category is divided into 12 categories based on every 15 degrees in terms of latitude (Figure 2). Based on the categories, nine frequent areas are found: western America (lon. [-120:-105], lat. [30:45], time diff. = -8); central America (lon. [-105:-90], lat. [30:45], time diff. = -7); eastern America 1 (lon[-90:-75], lat. [30:45], time diff. = -6); eastern America 2 (lon. [-75:-60], lat. [30:45], time diff. = -5); Europe (lon. [0:15], lat. [45:60], time diff. = 0); south Asia 1 (lon. [90:105], lat. [0:15], time diff. = +6, Singapore); south Asia 2 (lon. [0:15], lat. [45:60], time diff. = +7, Indonesia); east Asia 1 (lon. [120:135], lat. [30:45], time diff. = +8, China, South Korea, and Japan);and east Asia 2 (lon. [135:150], lat. [30:45], time diff. = +9, Japan).

2.2 Time Distribution

To observe time distribution of the information throughout the world, it is divided into 24 categories based on global mean time (GMT). Figure 3 shows that there is almost a constant amount of posting.

To clarify the characteristics of hours of each area, the information is divided into 24 categories. In each category, a trajectory of the number of posting is illustrated based on a local time. With respect to the nine areas in Figure 2, there are three types of peaks on the trajectories--night, noon, and morning. The most frequent type is night peak (Figure 4). Central America, east America 1, South Asia 2 (Indonesia), east Asia 1 (China, South Korea, and Japan), and east Asia 2 (Japan) are categorized into the night peak type. The noon peak type consists of west America and south Asia 1 (Singapore) (Figure 5). Europe is categorized into morning peak type (Figure 6).

2.3 Address Distribution

In order to clarify quality and pervasiveness of location information, another dataset was extracted from the entire dataset, which has been collected continuously for this research. The sub dataset contains points of interest (POI) in a country that were

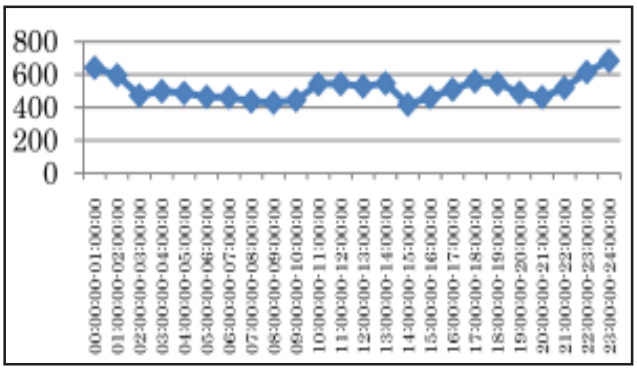


Figure 3. Time Distribution (GMT)

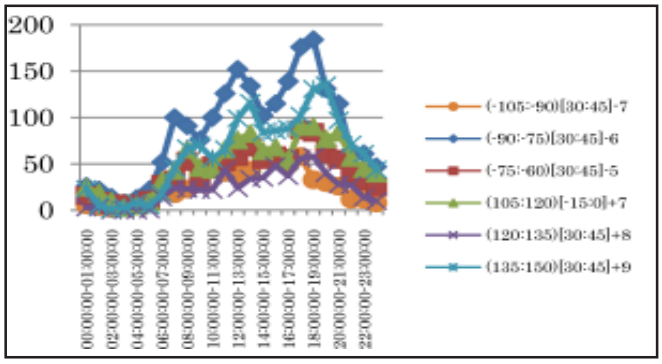


Figure 4. Peaks around Night

submitted on Twitter through Foursquare. The POI can be categorized into the administrative divisions of the country, based on the geospatial information provided by the authority of the country [2]. The dataset contains 3817 POIs and the duration is from Feb. 27, 2010 to Jan. 3, 2011.

Figure 7 shows extremely biased POI distribution among the administrative divisions. The most crowded region has 1,594 POIs, which is approximately 41.8 percent of all POIs. The region includes 444 POIs concerning the railway station, which is approximately 28 percent. The dataset tells us that the clouded region tends to have a high percentage of POIs related to public transportation.

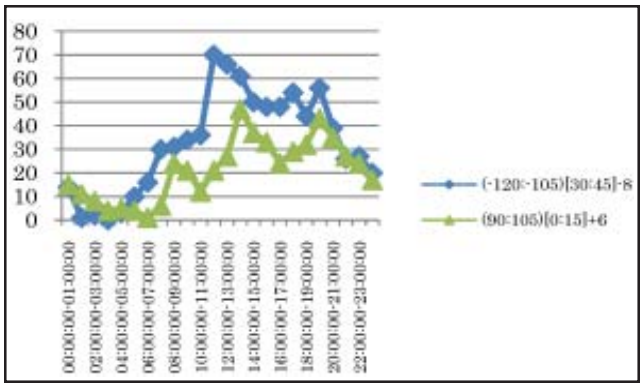


Figure 5. Peaks around Noon

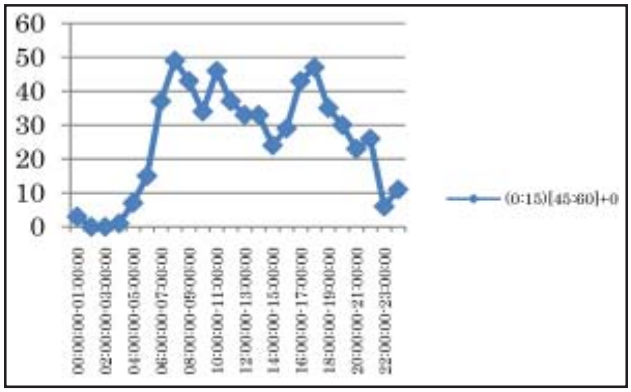


Figure 6. Peaks around Morning

The dataset is also evaluated in terms of integrity of address information written by individuals. The information is checked with the proper address information provided by the geospatial information authority of the country. Based on the observation of each POI, POIs of public places, such as a train station, have the correct address information; however, other types of POI, such as a privately owned restaurant, have incorrect or no address information. According to the result, the correct information comprises approximately 44 percent. To sum these results, POI information generated by individuals on social media has a huge regional bias and lacks correct address information.

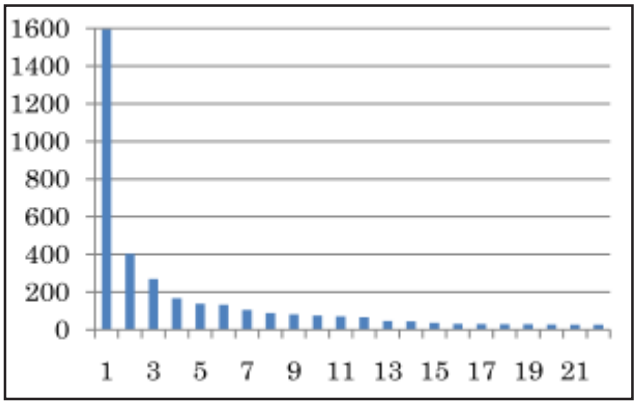


Figure 7. Address Distribution

3. GEO-Local Contents System

In order to reduce an enormous bias of location information and incorrect information, I developed a geolocal information system, which is composed of two parts--a mobile device and PC server applications.

On one hand, ordinal location information services, provided by well-known Internet companies, focus on each individual or small informal group. The result of activities on the services leads to a huge bias and incorrect information, as shown in section II C and Figure 7. On the other hand, the newly developed system in this paper will provide easy functions to collect, edit, and publish geo-local infor mation to local communities, such as shopping districts and tourist spots outside of metropolitan areas.

The members in this type of community have a strong incentive or motivation to generate geo-local information to advertise their shopping districts and tourist spots to attract customers. Hence, the systems are provided to the people in communities so that a vast amount of geo-local content is going to be published on the Internet (Figure 8).

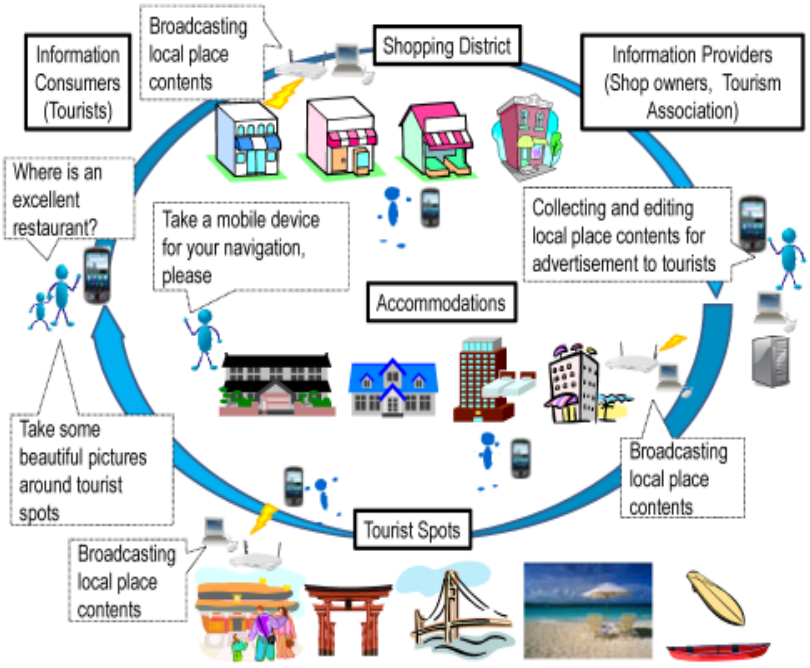


Figure 8. Information Flow in Local Community

To facilitate the geo-local information system, a smart phone and PC server applications have been developed (Figure 9). The smart phone application provides functions to take pictures, voice memos, and traces with geographical information sensed by the GPS module in the phone. It also provides local map and address browsing functions without a wired and wireless connection because it has map and address databases on the device. The databases are generated by means of a PC server application and downloaded through a Wi-Fi communication channel between a PC server and a smart phone.

A PC server provides editing and managing functions concerning map, POI, and traces. The edited geo-local information, including POIs, is also downloaded on smart phones to find places to enjoy travel. The application on a PC server has functions to publish geo-local information on the Internet to advertize shopping districts and tourist spots.

The PC servers and Wi-Fi routers will be prepared at accommodations, shopping districts, and tourist spots to provide the latest information edited by local community members. They collect and organize information with the system. The collected correct information will be published on Social media, e.g., Blog, SNS, Twitter, Foursquare, and so forth, to increase valid information on the Internet.

On one hand, ordinal mobile applications on GPS enabled devices require Internet access to browse geographical maps and server functions, e.g., online storage. On the other hand, the system of this research employs the local Wi-Fi connections around the tourist area instead of the Internet connection.

3.1 Geo-Local Information System on Mobile Device

An application software on a mobile device has been developed in order to gather geo-local contents and browse map and address information without any map application API, e.g., Google Maps API, because the software tends to be deployed outside of Internet enabled areas, such as rural beautiful beaches or mountains for sightseeing. Various maps on the Internet can be employed in order to create map databases of the software. Figure 10(a) shows browsing a map on a mobile device, generated from OpenStreetMap (OSM) [3] [4].

The software displays a map so that the current user position is centered on it. The map can be moved by dragging on the touch

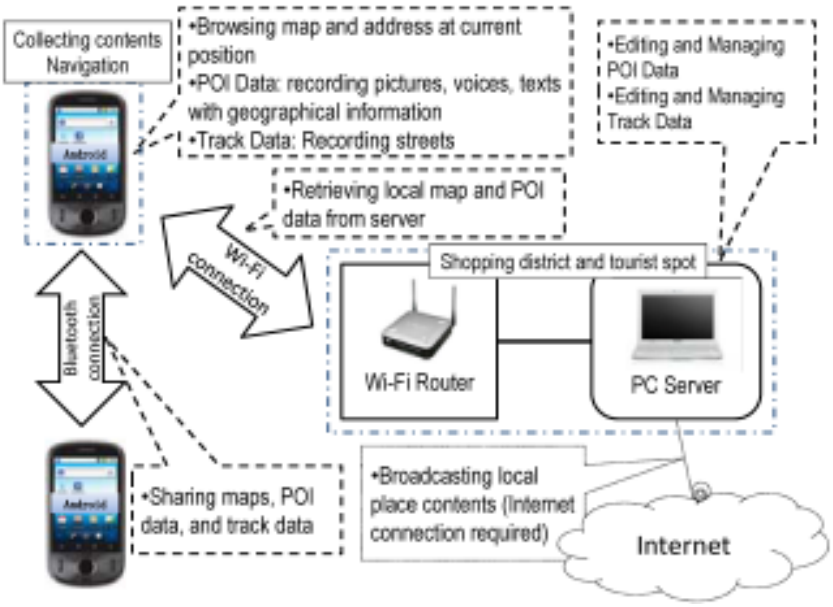


Figure 9. Overview

screen. On the top of the screen, current address information and zoom level of the map are displayed. On the bottom of the screen, there are five buttons from right to left: zoom-in and zoom-out of map, toggle of POI, enabling GPS, and taking pictures and voice memos with location information.

When toggling POI, POIs are displayed as purple dots on the map in Figure 10(a). When one of the POI is focused on the center of the map, a thumbnail of the picture and additional text information will be displayed on the screen. When we touch the thumbnail, we can browse detailed information and add voice comments on the POI. Moreover, when enabling GPS, your current position is always displayed on the center of the map. Hence, when you go to one of the POIs, the information is automatically displayed for your navigation.

When you push the camera button, a camera function is invoked in order to take a picture with current location information. After taking a picture, you can add voice memos if needed. Voice memos can be convenient to memorize information on a small mobile device without a keyboard in Figure 10(b).

(a) Map and POI

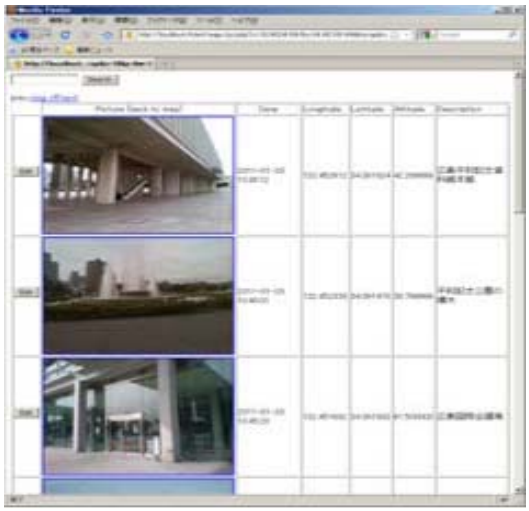
(b) Collecting Contents

Figure 10. Browsing Offline Map and POI

3.2 Editing and Publishing System on PC Server

Geo-local contents gathered by the software on mobile devices are transferred to a PC server in order to edit and manage the

(a) Map

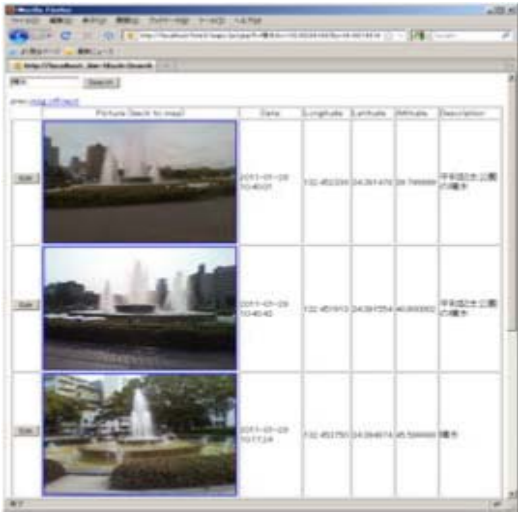


(b) Pictures

Figure 11. Browsing Offline Map and Pictures

contents based on map interface for arrangement of the contents without any map application API, e.g., Google Maps API, because the software tends to be deployed outside of Internet enabled areas. Various maps on the Internet can be employed in order to create map databases of the software. Figure 11(a) shows browsing a map with a web browser, generated from OpenStreetMap (OSM).

The system is able to be deployed on LAMP or WAMP stack on a PC server and is operated through a web browser. Figure 11(a) shows the Hiroshima Peace Memorial Museum and park. When you click the “picture” link on the left top of the map in the figure, you can browse pictures around the center of the map. Figure 11(b) shows an example of a list of pictures around the museum and park. The information concerning each picture consists of the picture itself, date, longitude, latitude, altitude, and description.



(a) Search Result with “fountain”

(b) Editing POI

Figure 12. Browsing and Editing POI

You can search pictures with a keyword based on the description of each picture. Figure 12(a) shows an example of a search result with a keyword meaning “fountain”. When you push an edit button of a picture, you can edit the description of the picture. Figure 12(b) shows an example of browsing and editing POI. The picture and a map where the picture was taken are

displayed with a text area to edit the description of the POI. You can also adjust the longitude and latitude of the POI because they might not be precise due to the spatial condition for the GPS device. The description you edit will be displayed on mobile devices as POI.

Figure 13. Local Information on Live View

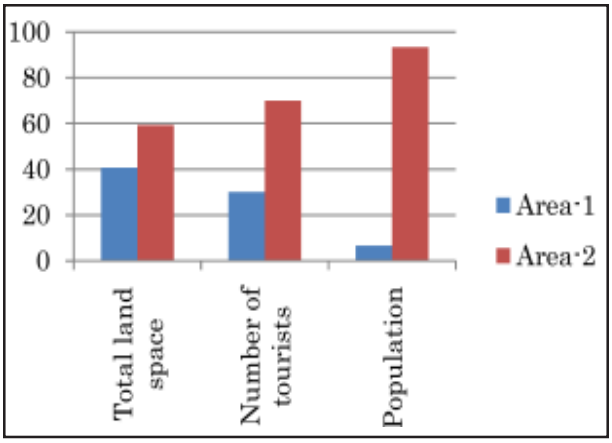


Figure 14. Comparison between Two Areas

3.3 Mobile Application for tourist

To provide local information to tourists in an intuitive way, an augmented reality (AR) application was also developed on the android because tourists are not familiar with tourist spots. Figure 13 shows an example of local information displayed on the AR application. The information was prepared by the systems in section III A and B.

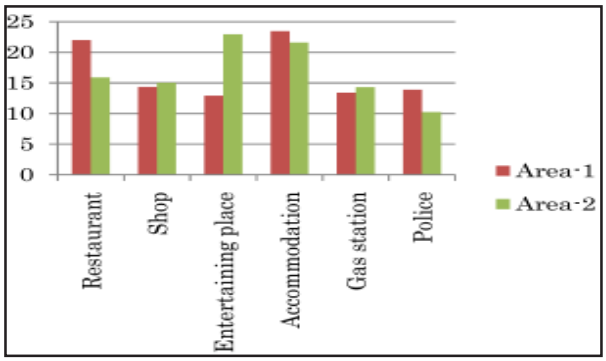


Figure 15. Distribution of Spot Categories

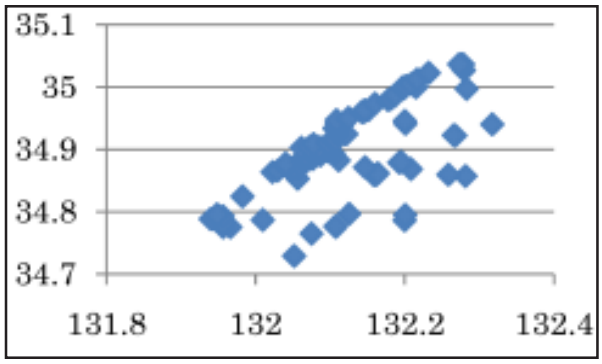


Figure 16. Tourist spots in Area-1

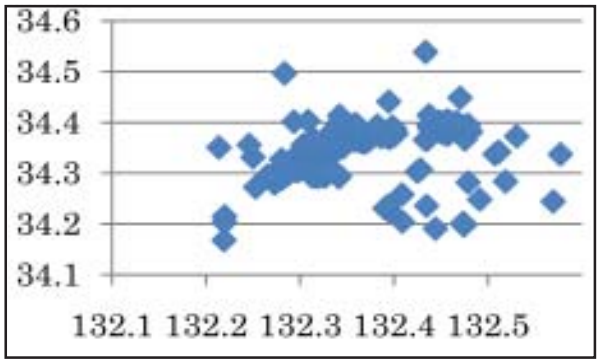


Figure 17. Tourist spots in Area-2

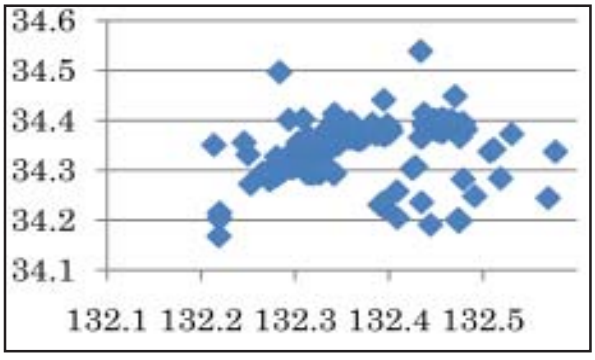


Figure 18. Restaurants in Area-1

4. Current Social Media As Information Source

Based on the systems, the circular information flow in Figure 8 will be implemented in a local community. An experiment is planned to deploy the developed systems with a local tourism association. The association maintains a tourist spot with a beautiful beach in a rural town on the west side of Japan. The members of the association cooperate by collecting and organizing local information.

To begin the data collection and organization, we have to determine what kind of information is on the web and whether enough exists concerning the rural town because we aim to add valid and useful information instead of replacing existing information.

To find out what information exists, we have to collect data concerning the rural town, called Area-1. Another tourist area is then selected for comparison. This area, called Area-2, is in a mid-size city and has two world heritage sites. The distance between the two tourist areas is approximately 60 km. In the dataset of section II-C, the numbers of POI s for Area-1 and Area-2 are 7 and 26, respectively.

4.1 Overview of Rural Town and Mid-size City

According to statistics published by local and central governments [5] [6] [7], Area-1 and Area-2 are described in terms of three aspects: total land space (958 km2 : 1,395 km ≒ 40.72 : 59.28), number of tourists (1,988,594 : 4,635,000 ≒ 30.02 : 69.98), population (90,820 : 1,269,921 ≒ 6.67 : 93.33), respectively. The differences are summarized in Fig. 14. There is a significant eifference between the two areas, especially in terms of population.

Around the two areas, six categories of tourist spot information are collected from the web. These categories are: (1) restaurants, (2) shopping, (3) entertainment, (4) accommodations, (5) gas stations, and (6) police. The number of spots in Area-1 and Area-2 are 215 and 442, respectively. Figure 15 summarizes the percentages of information among six categories for the two areas. According to the figure, Area-1 has many restaurants compared to Area-2. On the flipside, Area-2 has many entertaining places compared to Area-1. Based on the statistical test of independence, the independence between the areas and categories is rejected on a critical point of 0.01. The result of the test suggests that Area-1 has plenty of restaurants; however, it needs to facilitate entertaining places, compared to Area-2. According to the result, the local community members have to provide more information concerning entertaining places to attract tourists.

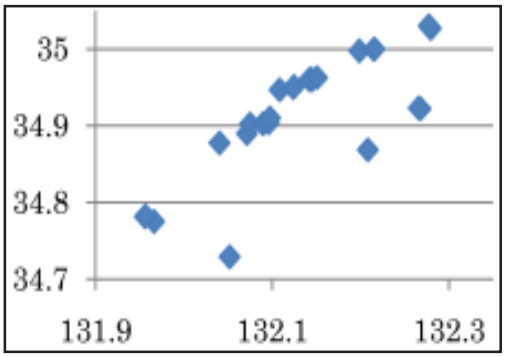


Figure 19. Shopping in Area-1

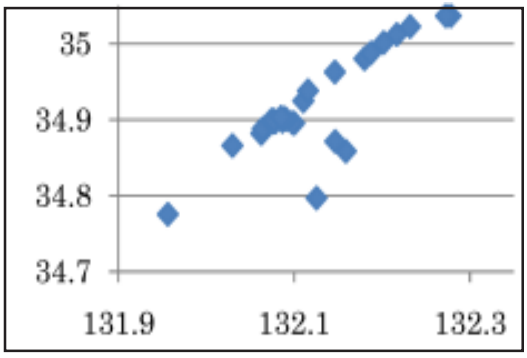


Figure 20. Entertainment

4.2 Spatial Patterns of Rural Town and Mid-size City

To understand spatial patterns of the tourist spots, Figures 16 and 17 show all tourist spots of Area-1 and Area- 2 in the data set, respectively. In both areas, the water front is the most crowded region.

To clarify characteristics in Area-1, spots in the six categories are shown in separated Figures 18, 19, 20, 21, 22, and 23. According to the figures, restaurants and gas stations are located in very narrow regions around the center of the rural city.

To clarify characteristics in Area-2, spots in the six categories are shown in separated Figures 24, 25, 26, 27, 28, and 29. According to the figures, restaurants, shops, and gas stations are located in very narrow regions around the center of the mid-size city.

To understand the spatial patterns of the tourist spots statistically, the spot data were analyzed with multi-distance spatial cluster

analysis. For statistical estimation, Ripley's K-function method was employed instead of the quadrat method and nearest neighbor method in terms of complete spatial randomness, because K-function method [8] [9] can detect multi-clusters in an area. The analysis based on K-function is described in the appendix of this paper. Tables 1 and 2 summarized the results of pairwise comparisons among the six categories in each area, in terms of density of clusters. In the tables, "S" and "W" mean relative strong and weak clusters, respectively. Each row means a relative density of clusters compared to the other categories. For example, the row pertaining to Restaurant in Table 1 means that the category has more strong density clusters compared to Shopping, Entertainment, Accommodations, and Police.

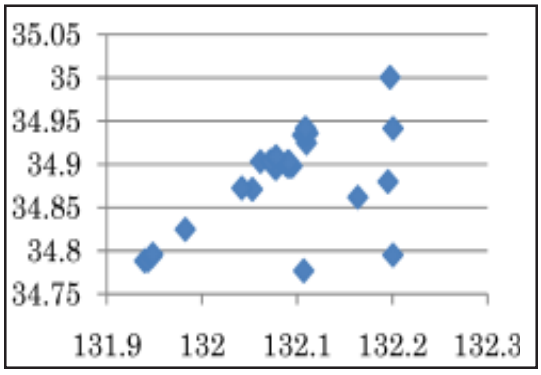


Figure 21. Accommodations in Area-1

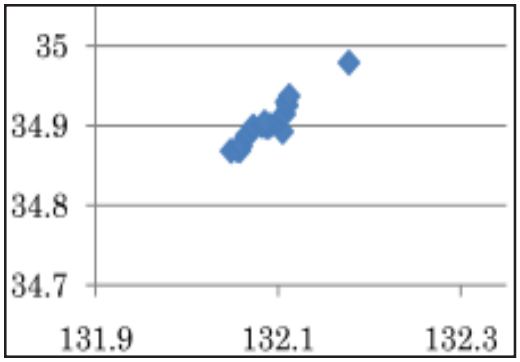


Figure 22. Gas stations in Area-1

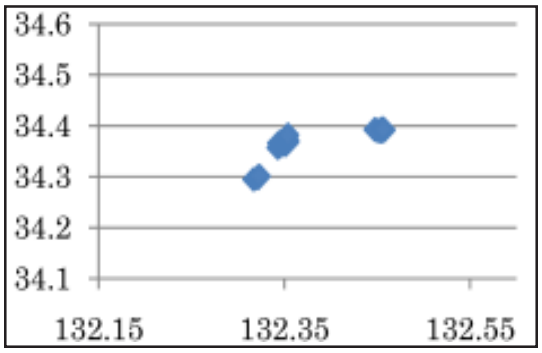


Figure 23. Police in Area-1

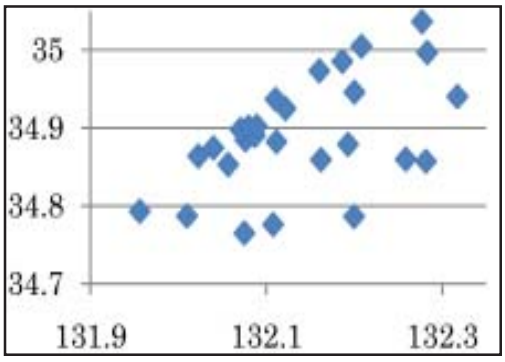


Figure 24. Restaurant in Area-2

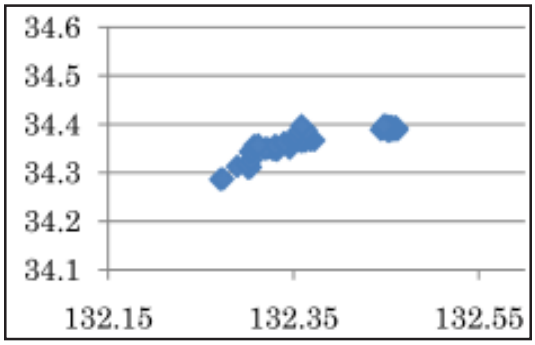


Figure 25. Shopping in Area-2

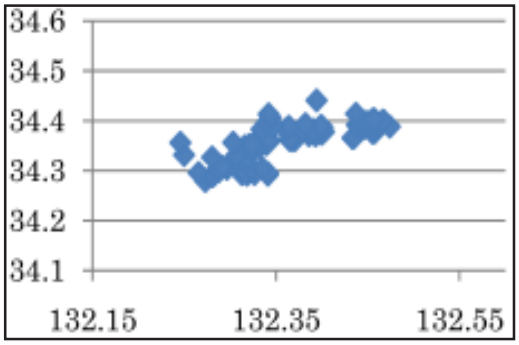


Figure 26. Entertainment in Area-2

According to the results, one of the common features in both areas is that the restaurant category has stronger density clusters than the other categories. Another common feature is that the police category does not have stronger density clusters than the other categories. On one hand, the Area-1 specific feature is that the Entertainment category does not have dense clusters. A few spots in the area might be the reason of the sparse pattern of the clusters. On the other hand, the Area-2 specific feature is that accommodation category does not have dense clusters. In that area, accommodation seems to be spread over a broad region because the mid-size city has been well developed.

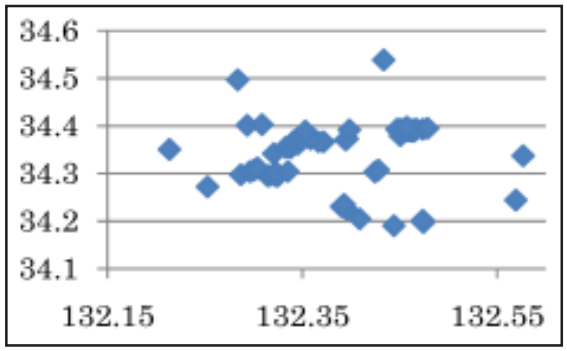


Figure 27. Accommodations in Area-2

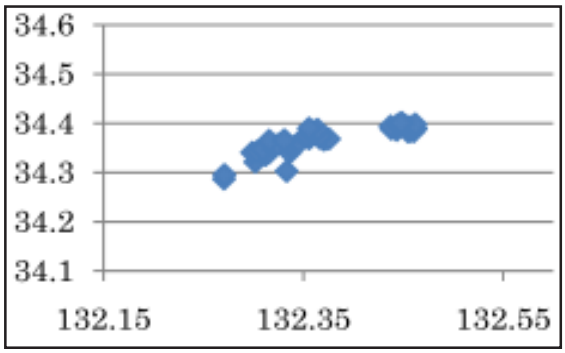


Figure 28. Gas stations in Area-2

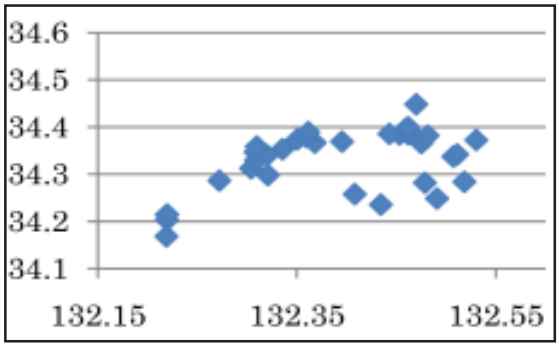


Figure 29. Police in Area-2

	R	S	E	A	G	P
Restaurants (R)		S	S	S	0	S
Shopping (S)	W	0	0	0	W	0
Entertainment (E)	W	0	0	W	W	0
Accommodations (A)	W	0	S	0	0	S
Gas stations (G)	0	S	S	0	0	S
Police (P)	W	0	0	W	W	0

Table 1. Area-1

	R	S	E	A	G	P
Restaurants (R)	0	0		S	0	S
Shopping (S)	0	0	0	S	0	0
Entertainment (E)	0	0	0	S	0	0
Accommodations (A)	W	W	W	0	W	0
Gas stations (G)	0	0	0	S	0	0
Police (P)	W	0	0	0	0	0

Table 2. Area-2

4.3 Information Patterns of Rural Area and Mid-size City

To evaluate web pages concerning each spot, the pages are also collected from the web. The number of corrected pages is 3363 and 5797 for Area-1 and Area-2, respectively. The total number of collected pages is 9103. The percentages of web pages in the 9103 pages are 36.9 and 63.1 for Area-1 and Area-2, respectively.

Tables 1 and 2 show the top 20 web sites that provide much information on Area-1 and Area-2, respectively. The total in each table means the total number of web pages collected from each web site. A1 and A2 mean the number of pages of Area-1 and Area-2 respectively. P1 and P2 mean the percentages of the pages concerning the two areas.

On one hand, the top six web sites in Table 1 only provide information concerning the rural town, i.e., specialized sites for Area-1. The other sites are mainly general portal sites in Japan. There is no Social media, e.g., blogs, in the table. On the other hand, in Table 2, the web pages of the top nine web sites are explained for Area-2, including Wikipedia, which ranks 8th. Site numbers 10, 13, 14, 15, and 16 are famous blog sites. Eleven and 17 are famous gourmet sites. These sites provide more information on Area-2 than for Area-1.

Rank	Web site	Total	A1	A2	P1	P2
1	navishimane.com	879	879	00	1.000	0.000
2	machikadoshimane.com	52	52	00	1.000	0.000
3	hamada.shimane.jp	41	41	00	1.000	0.000
4	shimane.lg.jp	27	27	00	1.000	0.000
5	www.tokusen.info	25	25	00	1.000	0.000
6	www.kankou-hamada.org	20	20	00	1.000	0.000
7	chi-ki.info	20	16	4	0.800	0.200
8	carview.co.jp	19	14	5	0.737	0.263
9	ggyao.usen.com	28	20	8	0.714	0.286
10	www.isize.com	37	25	12	0.676	0.324
11	9199.jp	103	68	35	0.660	0.340
12	qrblog.com	20	13	7	0.650	0.350
13	gnavi.co.jp	50	27	23	0.540	0.460
14	www.chizumaru.com	120	60	60	0.500	0.500
15	www.mapple.net	101	48	53	0.475	0.525
16	excite.co.jp	149	69	80	0.463	0.537
17	www.e-shops.jp	20	9	11	0.450	0.550
18	alike.jp	95	42	53	0.442	0.558
19	itp.ne.jp	114	48	66	0.421	0.579
20	www.its-mo.com	303	124	179	0.409	0.591

Table 2. Top 20 web sites for the mid size city

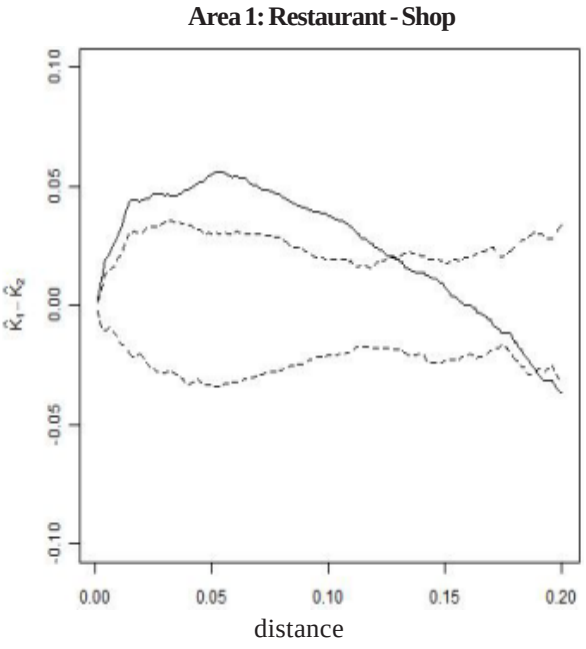


Figure 30. Restaurant and Shopping in Area-1

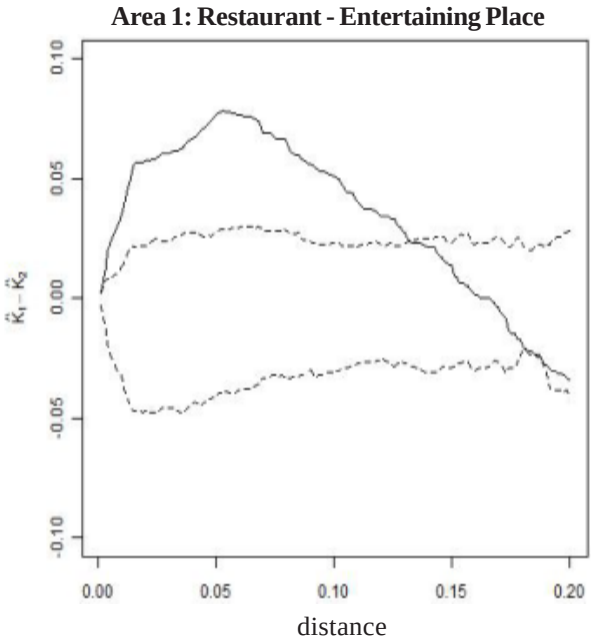


Figure 31. Restaurant and Entertanment in Area-1

Rank	Web site	Total	A1	A2	P1	P2
1	navihiroshima.com	1096	00	1096	0.000	1.000
2	www.allhotelsin.net	110	00	110	0.000	1.000
3	machikadohiroshima.com	50	00	50	0.000	1.000
4	d.katy.jp	22	00	22	0.000	1.000
5	www.tabi.tv	22	00	22	0.000	1.000
6	saeki-hiroshima.mypl.net	21	00	21	0.000	1.000
7	hiroshima.nokotonara.com	21	00	21	0.000	1.000
8	ja.wikipedia.org	19	00	19	0.000	1.000
9	chugoku-np.co.jp	18	00	18	0.000	1.000
10	exblog.jp	35	2	33	0.057	0.943
11	www.hotpepper.jp	82	6	76	0.073	0.927
12	nifty.its-mo.com	17	2	15	0.118	0.882
13	livedoor.jp	25	3	22	0.120	0.880
14	ameblo.jp	43	6	37	0.140	0.860
15	gogo.gs	62	11	51	0.177	0.823
16	livedoor.com	25	5	20	0.200	0.800
17	30min.jp	99	20	79	0.202	0.798
18	ocn.ne.jp	35	10	25	0.286	0.714
19	navitime.co.jp	50	15	35	0.300	0.700
20	tabelog.com	217	66	151	0.304	0.696

Table 2. Top 20 web sites for the mid size city

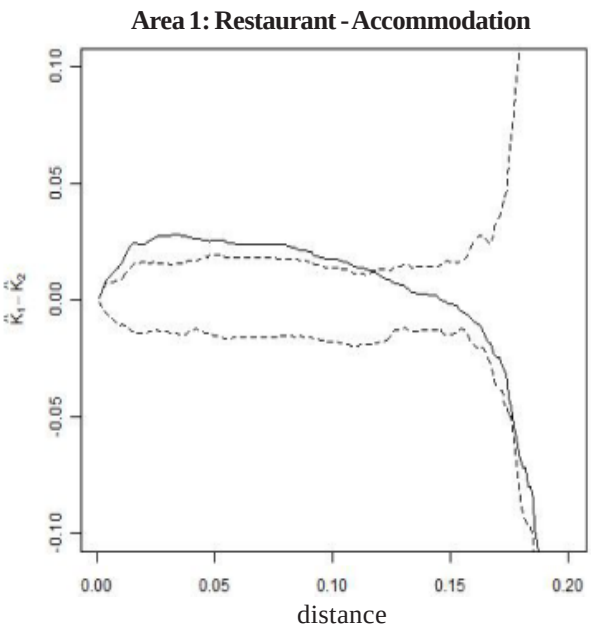


Figure 32. Restaurant and Accommodation in Area-1

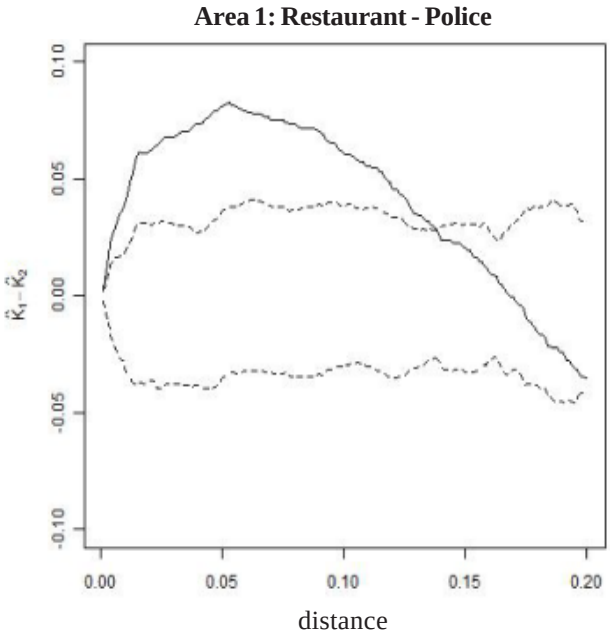


Figure 33. Restaurant and Police in Area-1

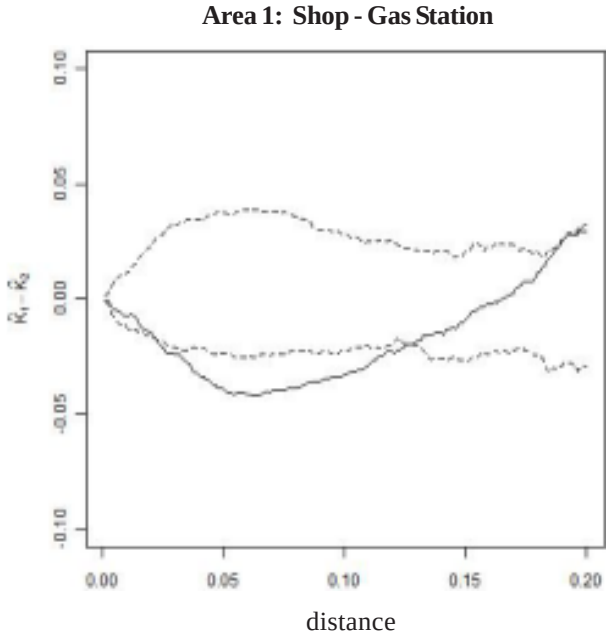


Figure 34. Shopping and Gas station in Area-1

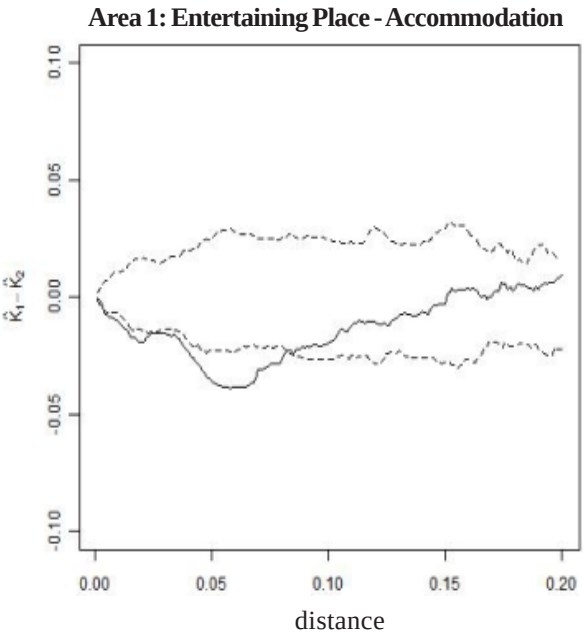


Figure 35. Entertainment and Accommodation in Area-1

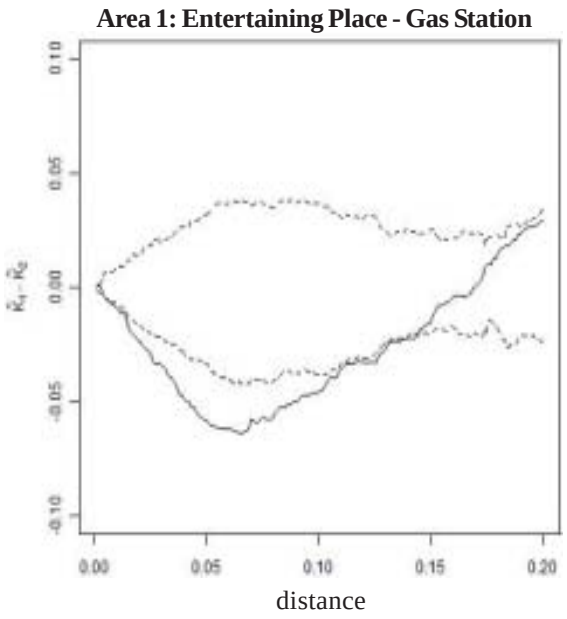


Figure 36. Entertainment and Gas station in Area-1

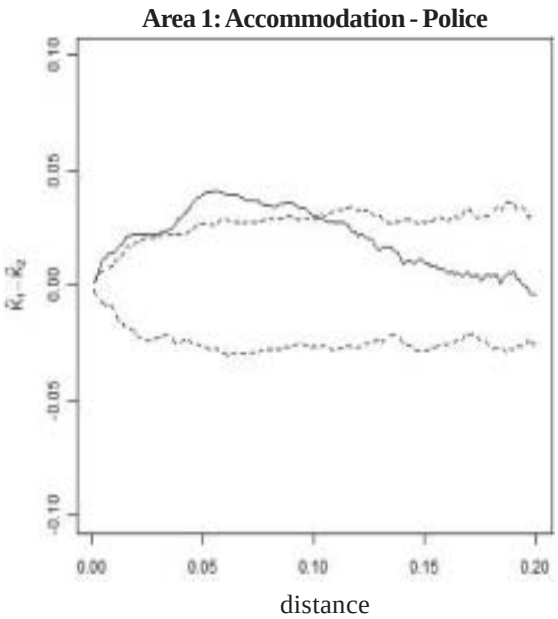


Figure 37. Accommodation and Police in Area-1

Compared to Area-1, Area-2 has many pages on social media such as famous blog sites and Wikipedia. This means that Area-2 has enough volunteers to generate word-ofmouth information on the Internet, and Area-1 does not. The difference of thenumber of volunteers seems to be related to that of populations in the areas, as mentioned in Figure 18. To promote word-of-mouth information, the local community members in Area-1 have to provide their own content on social media by themselves. Many tourists come to the rural town and become volunteer writers concerning the tourist spots.

Based on the statistics provided by government, the automatically collected information from the web, and the results of analysis, the local community members will create a strategy to attract many tourists to the tourist spots in the rural town.

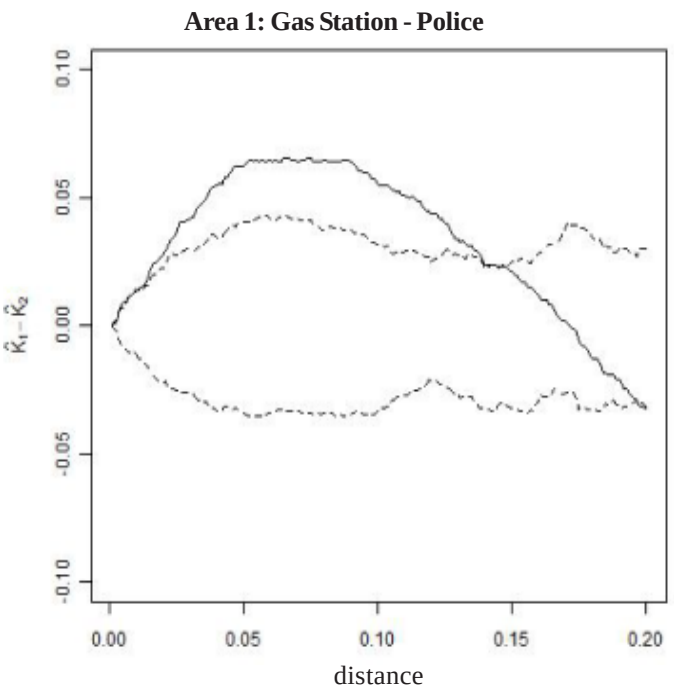


Figure 38. Gas station and Police in Area-1

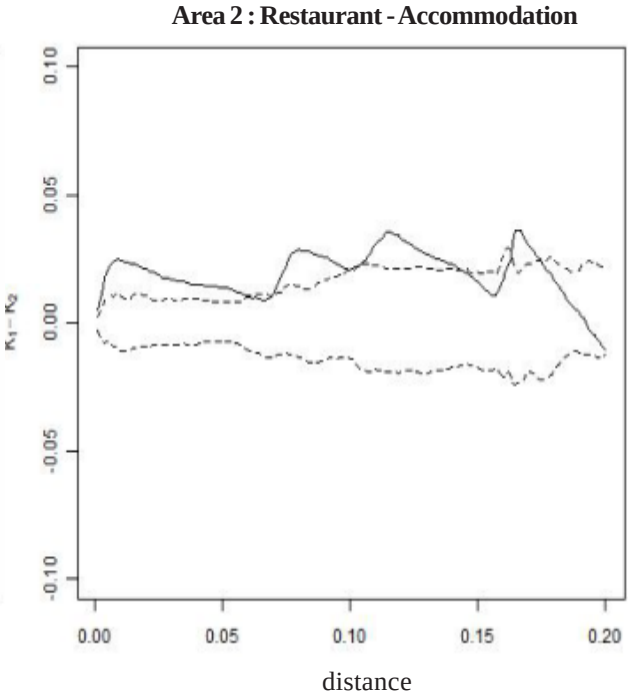


Figure 39. Restaurant and Accommodation in Area-2

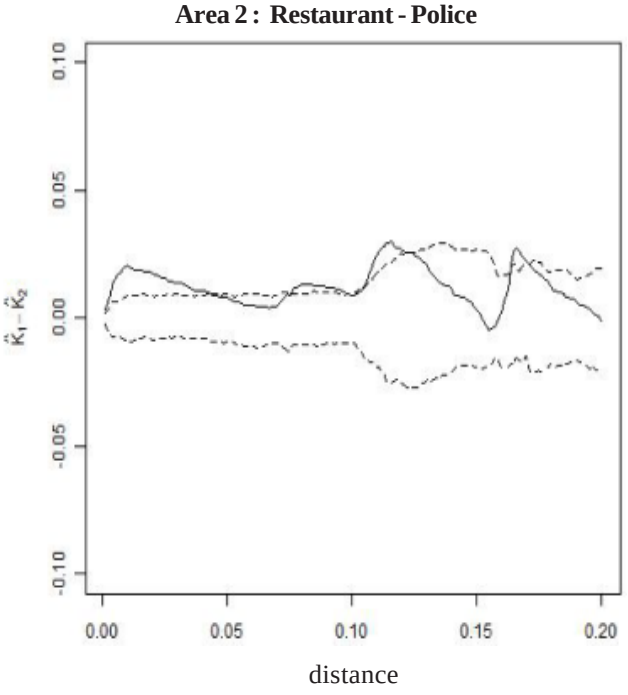


Figure 40. Restaurant and Police in Area-2

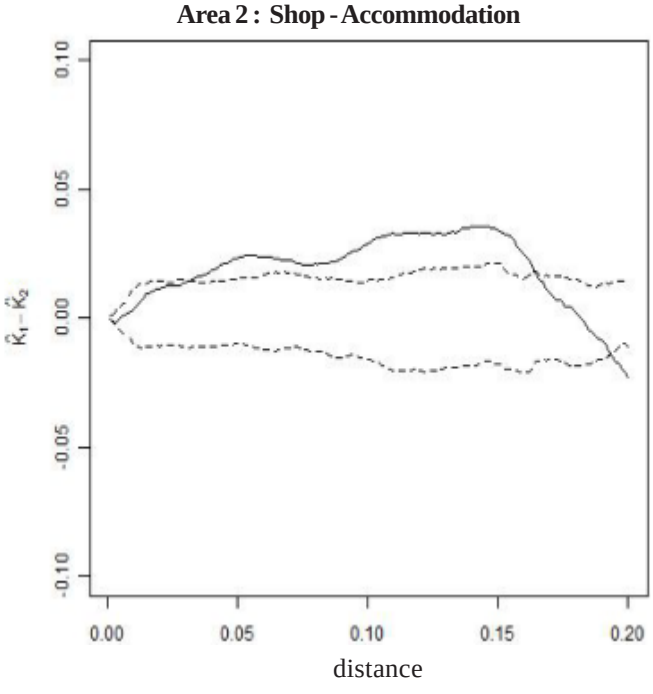


Figure 41. Shopping and Accommodation in Area-2

5. Conclusion

Geographical information, provided by social media such as Foursquare, is quite sparse in local areas, even though it is heavily clustered in metropolitan areas. In order to generate a high content in local areas, a location-based contents system was developed with mobile devices for people in local communities. Members of a local community, e.g., shopping districts and

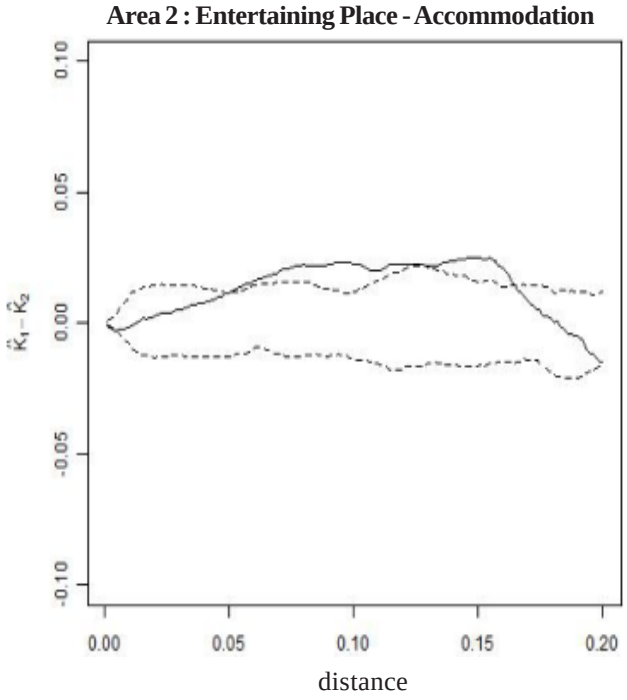


Figure 42. Entertainment place and Accommodation in Area-2

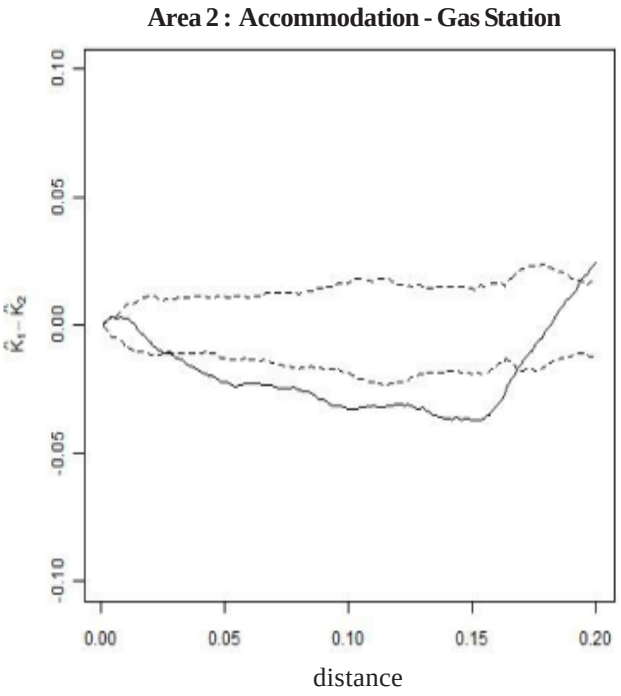


Figure 43. Accommodation and Gas station in Area-2

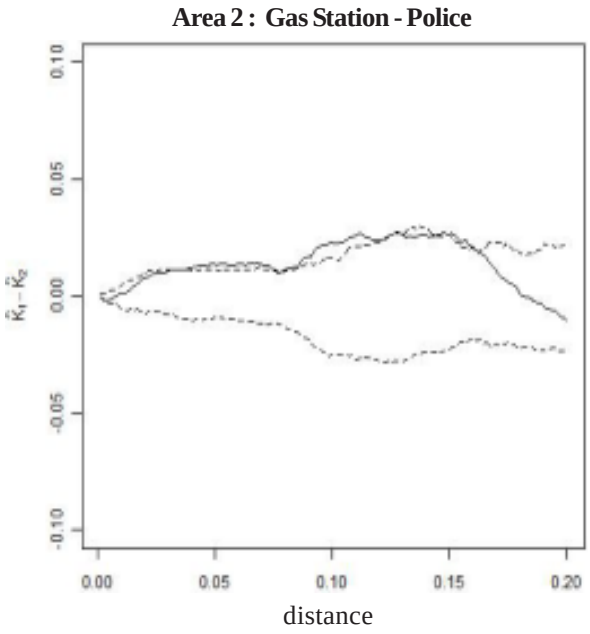


Figure 44. Gas station and Police in Area-2

tourist spots, have strong incentives to provide high quality information to their customers. Hence, the system for editing, publishing, and managing local content has been developed for these communities so that they will be published on the Internet.

To begin collecting and organizing local content, information available on the web is collected automatically concerning two areas, i.e., a rural town and a mid-size city. According to the analysis of the collected information, the members in the rural town have to generate local content concerning their locality because word-of-mouth information is very scarce.

Appendix

The figures below show the results of comparisons with respect to relative density of clusters between two categories in each area. A line in each figure shows a difference of K-function values between two categories. Two dotted lines in a figure mean boundaries of statistically significant values of the difference. When the solid line is over the upper dotted line, the first category has more dense clusters than another category. When it is below the lower dotted line, the second category has more dense clusters. For example, Figure 30 explains that restaurants form strong clusters compared to shops in Area-1.

References

[1] Foursquare0 http://foursquare.com/
[2] Geospatial Information Authority of Japan. Maps & Geospatial Information, Ministry of Land, Infrastructure, Transport and Tourism, http://www.gsi.go.jp/kiban/index.html
[3] OpenStreetMap, http://www.openstreetmap.org/
[4] Mordechai (Muki) Haklay, Patrick Weber (2008).OpenStreetMap: User- Generated Street Maps, *IEEE Pervasive Computing*, 7 (4) 12-18, October-December.
[5] National Statistics in Japan, http://www.stat.go.jp/data/ssds/5b.htm
[6] Tourism Statistics in Shimane, http://www.pref.shimane.lg.jp/kanko/report_h21.html
[7] Tourism Statistics in Hiroshima, http://www.pref.hiroshima.lg.jp/page/1276148389555/index.html
[8] Ripley, B. D. (1976). The second-order analysis of stationary point processes, *Journal of Applied Probability* 13, 255-266.
[9] Ripley, B. D. (1977). Modelling spatial patterns, *Journal of the Royal Statistical Society*, Series B 39, 172-192.