Scientific
Research
Publishing

Spatial Analysis of Location-Based Social Networks in Seoul, Korea

Ilyoung Hong

Department of GIS, Namseoul University, Cheonansi, South Korea
Email: ilyoung.hong@nsu.ac.kr

Abstract

The purpose of this study is to analyze the spatial patterns of location-based social network (LBSN) data in Seoul using the spatial analysis techniques of geographic information system (GIS). The study explores the applications of LBSN data by analyzing the association between Seoul's Foursquare venues data created based on user participation and the city's characteristics. The data regarding Foursquare venues were compiled with a program we created based on Foursquare's Python API. The compiled information was converted into GIS data, which in turn was depicted as a heat map. Cluster analysis was then performed based on hotspots and the correlation with census variables was analyzed for each administrative unit using geographically weighted regression (GWR). Based on analytical results, we were able to identify venue clusters around city centers, as well as differences in hotspots for various venue categories and correlations with census variables.

Keywords

LBSN, Spatial Analysis, Foursquare

1. Introduction

With the increasing number of smartphone users, various smartphone apps have been introduced, most popular among which are social network service (SNS) applications such as Twitter and Facebook. Social network services that use global positioning system (GPS) location data are referred to as location-based social networks (LBSN), and widely used services include Foursquare and Facebook Places [1] [2]. In Foursquare, social networking takes place around venues, and various studies have been conducted on venue characteristics using statistical analysis [3] [4]. A typical approach is to create a heat map using the geographic information system (GIS) and venues' coordinates to analyze their spatial distribution properties within a city. Most studies on venue cha-

racteristics have been conducted for cities in the U.S. and Europe, where Foursquare services have primarily been focused on [5]-[7].

As of 2015, Foursquare has many users in Korea, and a large amount of data has been accumulated regarding venues. However, there has not been a survey or analysis conducted for cities in Korea and other parts of Asia. With a high level of interest in information technology (IT) and a wide distribution of smartphones, a diverse range of apps are being developed and used in Korea, producing substantial amounts of data for social media such as Foursquare. Unlike statistical information gathered through regular surveys, SNS data reflect real-time user opinions and are reported to indicate diverse social phenomena, including diseases and disasters. Similarly, LBSN data created based on voluntary participation of citizens provide the benefits of being able to analyze the changes taking place within a city [8]. At the same time, the correlations between LBSN data and the spatial distribution of census information regarding city's characteristics offer various means for future city planning and administration.

In the present study, data regarding Foursquare venues were compiled to perform statistical analysis on their spatial distributions and patterns. The study consists of Foursquare venue data compilation using Python, GIS data implementation, visualization of the collected data into maps, and spatial statistical analysis. The significance of this study is that it provides applications of LBSN data that can be quickly compiled for city administration and monitoring.

2. Related Works

According to a survey taken in 2012, 75% of smartphone users use location-based services (LBS) for finding directions and searching maps, and 18% of them use LBSN services such as Foursquare, Google Local, and Facebook Places [9]. After launching services in 2009, Foursquare has increased to 40 million users, 5 billion check-ins, and 40 million user-created tips [10].

Dodgeball in the U.S. was the first service to introduce LBSN functions in 2000. At the time, Dodgeball provided a simple LBS that allowed users to notify their friends of check-in status. To Dodgeball's existing functions, Foursquare added game components by providing badges and mayor ships to those showing significant levels of check-in activity. As a result, Foursquare was able to secure 20 million users in 2012 and achieve a rapid growth [11]. Facebook, the most popular social media network, also acquired Gowalla, an LBSN company, to provide a service called Facebook Places, making LBSN an important element in smartphones.

Gau and Liu described LBSN characteristics as a "3 + 1 framework" where 3 represents the content layer that includes audio, video, photos, and tips, the social layer that indicates social relationships, and the geographic layer that reflects users' check-in activities. The +1 framework indicates the timeline along which the three components progress [12]. The geographic layer consists of places where users' check-in data are stored. Referred to as venues in Foursquare, this type of specific information is not available in conventional social media. For each venue, information pertaining to the place, such as attributes, name, tips, photos, and category code are stored, as well as statistical data such as numbers of user check-ins and tips.

A characteristic of LBSN that differentiates it from conventional mobile phone information is that it provides information linked with GPS through Web 2.0 services [13]-[15]. Whereas mobile phone's location information has limitations in that it can only be obtained through wireless carriers due to privacy protection and security issues, access to LBNS data is permitted for researchers and developers through API, allowing liberal research and development of applications. Until now, most studies regarding GIS applications have focused on creating maps using geovisualization of venue data as well as venues' distribution characteristics, and approaches have not been made in terms of the spatial statistics of venue data.

3. Data and Method

Seoul has east-west and north-south distances of 36.78 and 30.3 km, respectively, with an area of 605.27 km^2. Although the area accounts for only 0.6% of South Korea, Seoul has a very high population density with 10.4 million inhabitants, which is more than 1/5 of country's entire population [16]. Regional analysis was performed in administrative Dong units, which is the smallest geographical area of Korean government administration.

A Python program was developed using Foursquare's developer API to compile venue data. 230,000 venue data were collected for analysis between March 15 and 21, 2015. A walking distance of 400 m was used as the search radius for Foursquare's venue search API. The collected data consisted of attributes such as venue's ID,

name, category, longitudinal and latitudinal coordinates, number of check-ins, number of users, and number of tips.

Foursquare venues are divided into ten categories: Art & Entertainment (C1), College & Education (C2), Event (C3), Food (C4), Nightlife Spot (C5), Outdoors & Recreation (C6), Professional & Other Places (C7), Residence (C8), Shop & Service (C9), and Travel & Transport (C10). In this study, the spatial distribution characteristics of each category were analyzed. For geovisualization of venues and analyzing their spatial distribution characteristics, we used hotspot analysis to verify the statistical significance and heat maps for examining the densities of point data.

For spatial statistical analysis, point data regarding the venues were converted into aerial information in administrative Dong units through spatial joining. XsDB, the spatial data from Biz GIS (biz-gis.com), was used for the census information in analyzing the correlation between venue categories and the census data.

Foursquare venues' geographic differences and spatial characteristics were analyzed by comparing the geographically weighted regressions (GWR) that estimates the relationship between independent and dependent variables based on the ordinary least squares (OLS) regression to assesses the average global association of the entire research subject region and local geographic proximities within the region [17]. For the analysis, the number of venues was set as the dependent variable and the working population in the 20 - 29 age group and land price were used as independent variables. Administrative division maps were obtained from the statistical geographic information service. The Spatial Statistics extension of ArcGIS 10.1 was used for processing data, creating maps, and performing GWR analysis. Figure 1 displays the data used in this study as well as major processes and analytical methods.

4. Results

A heat map is the most widely used means for expressing the level of density of point data's spatial distribution. Figure 2 is the heat map of venues created using 240,000 venue data gathered in Seoul. Areas with the highest venue densities in Seoul were Gangnam, Jung-gu, and Hongdae, which are representative commercial districts of Seoul. Among them, Gangnam displayed high venue densities across a vast area.

Table 1 shows the statistics of the 240,000 venue data for various categories. In addition to the venue name, stored statistical data include the numbers of visits, check-ins, and unique users, and user comments for the corresponding location. In terms of venue categories, the Food (C4) category, which provides information on restaurants, showed the highest percentage at 41.7%, followed by Professional & Other Places (C7) for offices and other facilities at 18.2% and Shop & Service (C9) for clothing stores and other shops at 17.5%. These three categories accounted for about 78% of all venues, indicating that they represent major service facilities in Seoul metropolitan area that draw the highest levels of interest from users. As for statistics involving the number of
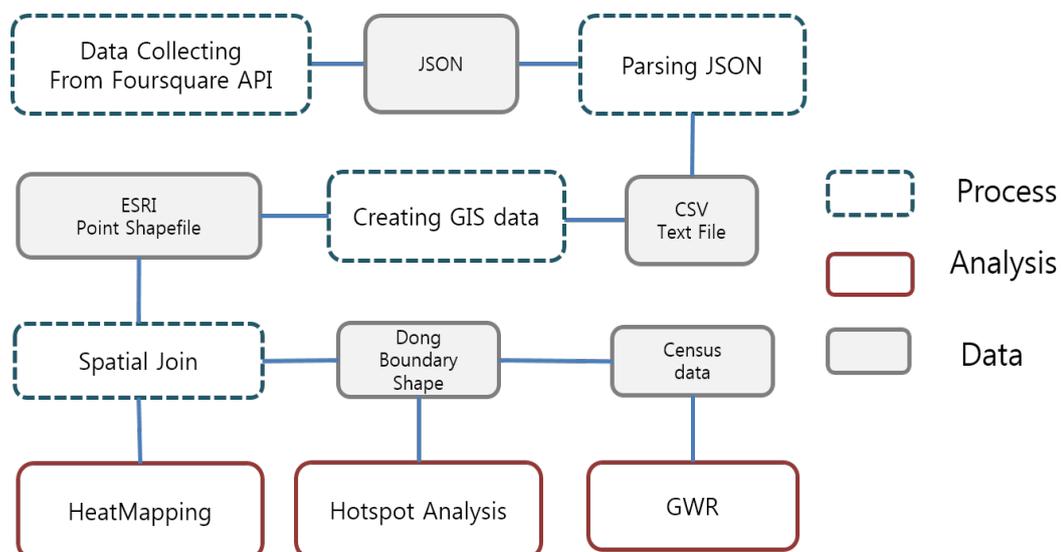

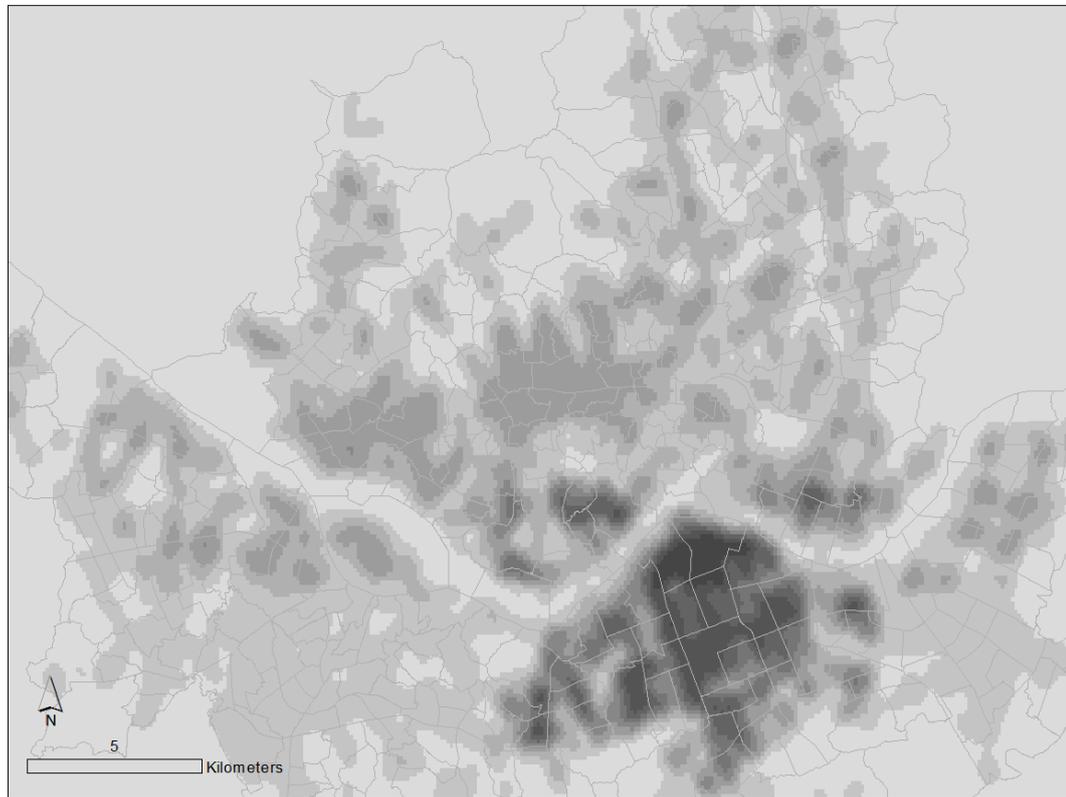**Figure 1.** Data collection and analytical process.

Figure 2. Spatial distribution heat map of venues in Seoul.

Table 1. Characteristics of 10 venue categories.

Category ID	Category Name	No. of venues	%	Average no. of check-ins	Totalno. of check-ins	Average no. of users	Total no. of users	Average no. of tips	Total no. of tips
C4	Food	90,392	41.75	79.64	7,199,075	42.78	3,867,357	1.36	122,657
C7	Professional & Other Places	39,427	18.21	70.49	2,779,278	19.36	763,287	0.41	16,002
C9	Shop & Service	38,021	17.56	92.04	3,499,332	31.22	1,186,993	0.59	22,609
C6	Outdoors & Recreation	10,762	4.97	149.46	1,608,442	57.59	619,748	0.79	8474
C5	Nightlife Spot	10,685	4.94	56.38	602,436	30.55	326,378	0.78	8375
C1	Art & Entertainment	7741	3.58	254.48	1,969,891	116.32	900,435	1.56	12,091
C10	Travel & Transport	7700	3.56	295.21	2,273,114	108.39	834,588	1.23	9445
C2	College & University	6523	3.01	148.61	969,369	31.64	206,380	0.58	3769
C8	Residence	5202	2.40	75.34	391,916	9.09	47,275	0.27	1400
C3	Event	40	0.02	24.82	993	19.80	792	0.28	11

check-ins (which is equivalent to the number of visit) and the number of unique visitors, the total figures were high for venues in C1, and C7 categories. In terms of average, however, venue categories where many people gather, such as C1, C6, and C10, showed high numbers. These characteristics indicate that although these categories contain small numbers of venues, they attract large crowds.

In order to analyze the spatial distributions of venue categories, hotspot analysis was performed using the ratios of the categories for each administrative Dong. Hotspot analysis involves identifying statistically significant clusters in a spot data distribution and provides a map of hot and cold spots that exhibit higher and lower corresponding values, respectively, than other regions. **Figure 3** displays the results of hotspot analysis performed using venue data. **Figure 3(a)** is created based on the number of venues, showing areas with the highest and lowest venue densities. **Figures 3(b)-(d)** display hotspot distributions for Food (C4), Professional & Other Places (C7) and Shop & Service (C9) categories, respectively. In terms of distribution, (c) and (d) show spatial patterns similar to that of (a), and what they have in common is that there are clusters primarily around the city center or downtown areas. On the other hand, the Food (C4) category, which has the highest number of venues, is showing a relatively low percentage in the city center with more hotspots in outskirt residential areas. It can be confirmed that whereas the absolute number of venues on the Food (C4) category is greater in the case of city centers with large daytime populations and high land prices than in suburban areas, the percentage is lower compared to venues in other categories.

GWR analysis was performed to analyze the census data, which is related to the venue distribution. One of the general characteristics of the census data is that young adults (in the 20 - 29 age group) produce the highest amount of content for social media. Furthermore, it can be expected that a venue characteristic, which is a type of a point of interests (POI), is that there would be more venues in areas with large daytime population and therefore high land prices. Therefore, GWR analysis was performed with the working population in the 20 - 29 age group and the land price as independent variables, and the number of venues as a dependent variable. GWR provides a map that expresses Zscore values for the goodness of fit of the regression analysis performed with each of the two variables. Furthermore, the coefficients indicate the characteristics of the regression model for each region. **Figure 4** displays the Zscore values of the GWR model. The Zscore values in city centers such as Gangnam and Yeouido showed high levels of goodness of fit. In terms of age group characteristics, there was a high correlation with the distribution of working population in the 20 - 29 age group that accounted for the large daytime population.

5. Conclusions

In this study, LBSN data were compiled and their spatial patterns were analyzed using the GIS spatial analysis
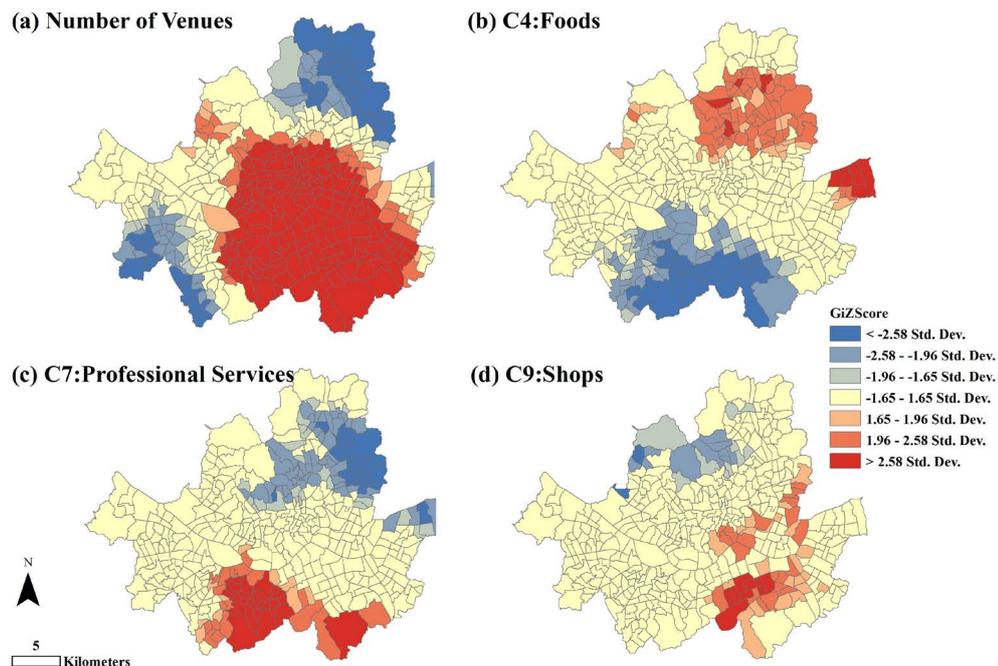

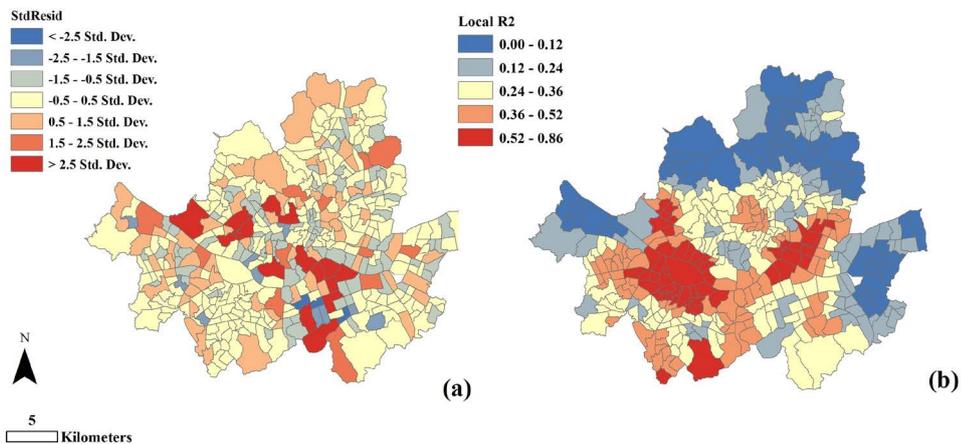**Figure 3.** Hotspot analysis with foursquare venues.

Figure 4. Zscore values of the GWR model.

technique. Using Foursquare venues in Seoul, the distribution characteristics of venues were surveyed. In terms of spatial distribution, the characteristics of venues' geographical concentration were assessed and high densities around major city centers were identified.

The spatial statistics analysis was performed based on the hotspot analysis of the distribution characteristics for each category. The hotspot analysis was conducted for venues categories with highest percentages: Food (C4), Professional & Other Places (C7), and Shop & Service (C9). As a result, we were able to identify different venue clusters according to their characteristics. Although Food (C4) was the most prevalent venue category, its percentage was relative lower in areas with high venue densities. As for venues' correlation with the census data, land price and the working population in the 20 - 29 age group were used as variables. From the results of analysis performed using the variables, it can be deduced that venue distribution is closely associated with the size of daytime population.

Among various types of LBSN data, this study focused on venue location data for analysis. In the future, additional research is required on social networks created around venues and user tips, which are comments made regarding venues. In addition, detailed study should be conducted on the association with pedestrian flow, which is anticipated to be closely related to venue distribution.

References

[1] Allamanis, M., Scellato, S. and Mascolo, C. (2012) Evolution of a Location-Based Online Social Network: Analysis and models. *Proceedings of the* 2012 *ACM Conference on Internet Measurement Conference*, ACM, New York, 145-158. http://dx.doi.org/10.1145/2398776.2398793

[2] Bao, J., Zheng, Y. and Mokbel, M. (2012) Location-Based and Preference-Aware Recommendation Using Sparse Geo-Social Networking Data. *Proceedings of the* 20*th International Conference on Advances in Geographic Information Systems* (*SIGSPATIAL*'12), ACM, New York, 199-208. http://dx.doi.org/10.1145/2424321.2424348

[3] Cheng, Z., Caverlee, J., Lee, K. and Su, D. (2011) Exploring Millions of Footprints in Location Sharing Services. *Proceedings of the Fifth International Conference on Weblogs and Social Media*, Barcelona, 17-21 July 2011, 81-88.

[4] Noulas, A., Scellato, S., Mascolo, C. and Pontil, M. (2011) An Empirical Study of Geographic User Activity Patterns in Foursquare. *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media*, Barcelona, 17-21 July 2011, 570-573.

[5] Humphreys, L. (2010) Mobile Social Networks and Urban Public Space. *New Media & Society*, **12**, 763-778. http://dx.doi.org/10.1177/1461444809349578

[6] Ferrari, L.,Rosi, A., Mamei, M. and Zambonelli, F. (2011) Extracting Urban Patterns from Location-Based Social Networks. *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks* (*LBSN*'11), ACM, New York, 9-16. http://dx.doi.org/10.1145/2063212.2063226

[7] Bawa-Cavia, A. (2011) Sensing the Urban: Using Location-Based Social Network Data in Urban Analysis. In: *Workshop on Pervasive and Urban Applications* (*PURBA*).

[8] Goodchild, M. (2007) Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0. *International Journal of Spatial Data Infrastructures Research*, **2**, 24-32.

[9] Zickuhr, K. (2012) Three-Quarters of Smartphone Owners Use Location-Based Services. Pew Internet & American Life Project.

[10] Smith, C. (2014) By the Numbers: 12 Important Foursquare User Stats. http://expandedramblings.com/index.php/by-the-numbers-interesting-foursquare-user-stats/

[11] Tsotsis, A., (2011) Foursquare Now Officially at 10 Million Users. http://techcrunch.com/2011/06/20/foursquare-now-officially-at-10-million-users/

[12] Gao, H. and Liu, H. (2014) Data Analysis on Location-Based Social Networks. In: *Mobile Social Networking*, Springer, New York, 165-194.

[13] Haklay, M., Singleton, A. and Parker, C. (2008) Web Mapping 2.0: The Neogeography of the Geoweb. *Geography Compass*, **2**, 2011-2039. http://dx.doi.org/10.1111/j.1749-8198.2008.00167.x

[14] Ratti, C., Sevtsuk, A., Huang, S. and Pailer, R. (2007) Mobile Landscapes—Graz in Real Time. In: *Location Based Services and Telecartography*, Springer, Berlin, 434-444.

[15] Scharl, A. and Tochtermann, K. (2009) The Geospatial Web: How Geobrowsers, Social Software and the Web 2.0 Are Shaping the Network Society. Springer Science & Business Media.

[16] Wikipedia (2015). http://en.wikipedia.org/wiki/Seoul#Geography

[17] Fotheringham, A.S., Brunsdon, C. and Charlton, M. (2002) Geographically Weighted Regression: The Analysis of Spatially Varying Relationships, Wiley. Chichester.