A Study on the Utilization of Spatial Big Data For Restaurant Business Location

Myung-Hee Jo[1], Kang-Hun Lee[2]

Abstract: the business district can defined in many different ways. But the following are the indexd to be applied for analysis. The size of business district, a floating population, number of same field, access feasibility and land and lease level. This study used in the bigdata and carried out of search for “famous restaurant”, “healing”,”food”, search are conducjted in 3 words only Seoul City area is the most active business district, where Gangnam district is located and Yeoksam 1Dong which is the most active among the business district. The most active business district was located in the center of venue Yeoksam 1dong between Gangnam station and Shinnonhyon station about 50m from side to side. Even though the versatile food service business were located equally, it was shown that the western food service business was distinguished out of them as well as spatial feature as to the location was densely formed on right side of venue between Gangnam station and Shinnonhyeon station.

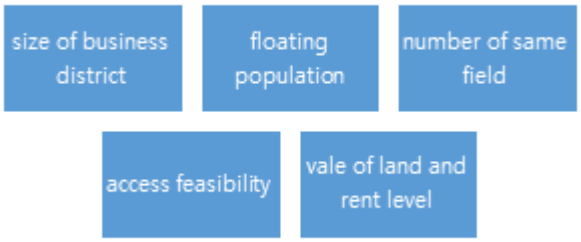
Keywords: Big data, Spatial data, SNS

1. Introduction

The business district can defined in many different ways. It can be defined as regional area where we can find a lot of customers. Business district sales can be estimated through the analysis of sale volume or number of customers in the shop as the surrogate index.

The ultimate purpose of business district analysis is to identify the number of existing customers. In spite of many of various factors for the analysis of business district, the following are the index to be applied for analysis. The size of business district, a floating population, number of same field, access feasibility and land and lease level. Both land for commercial and rental for business district with specific requirements. Even though the requirements have been completed there are other problems to be considered such as the start-up capital.

Therefore, before opening a business it is very important to search for an affordable business location base on the start-up capital.



FUGURE-1 Index of Business District Analysis

The statistic data is usually used for survey of business district analysis, however basically each index for analysis should be apprehended through the field survey. However, this kind of survey consume a lot of time and money as well as it commit errors because of personal subjective judgment.

Recently, there has been an increase in the amount of information and the new concept of the word define as "big data". In this regard, active research is being conducted with the application of this big data in various field. They attempt to combine the use of these data with the different field which is frequently

.

1. School of Convergence & Fusion System Civil engineering, Kyunpook National University, Sangju-si, Gyeongsangbuk-do, Korea
Phone : +82-10-3150-7312
E-mail : mhjo@knu.ac.kr
2. School of Convergence & Fusion System Civil engineering, Kyunpook National University, Sangju-si, Gyeongsangbuk-do, Korea
Phone : +82-10-2204-9564
E-mail : chinggu88@gmail.com

occurs. However if you use the big data for business district analysis various statistics data related to business district must be provided and subjective opinion of the user.

In this research big data was utilized to classify and analysis the business district in Seoul City using twitter data. To check the classified data of each area Arcmap Density Kerner has been used for analysis, defining size and features over the specific business district.

2. Research Scope and Related Studies

The research was conducted in Seoul City covering the whole area and the used of twitter data user between Aug 13th and Aug 20th, 2013.

The majority studies of business district analysis were based on research studies that has been made on the theoretical analysis as well as focused on the estimation of sale volume when used for specific usage. To estimate the revenue model the size of the business district and sale volume including the users ability has been considered.

Further analysis using the Customer Spotting Technique has been used to determine the costumer's district address, expenditures and distribution.

The difference between the pilot study from the previous study on business district was during on-site investigation it consume a lot of time and financial expenses and for surrogate index analysis sales volume and number of user within the business, expenditures and users in the business district area has been utilized. However, the purpose of this study is to provide specification of business district to analyze the location of user who are actually visited the store based on the user’s twitter analysis.

3. Utilization of Big Data

The establishment of big data is not for a single concept. The concept of Big Data is defined as a huge data compare to the past data that were generated in an analog environment, includes text and images as well as form of numerical data. Considering this kind of definition, the SNS, in other words the twitter data, can be considered a Big Data.

3.1 Acquisition of Twitter Data

There are necessary process to acquire the user's tweet in the twitter.

First, registration in the Twitter.

After registration in the Twitter's Homepage in the developer page the user can received consumerkey,cunsumerscretkey into the users page.

Second, authentication via of OAuth(Open Authorization)

The authentication can be acquire using 4 key after registration the user can received consumerkey and consumerscretkey, after qualification process by OAuth accesskey and accessscretkey will be granted.

Third, obtaining of data through the search API

It obtain the data through the search API 4 keys are needed. The factor used search API is query(). the factor of query() is to search for matching tweet related to the words, and then load contents of the searched tweet making use of get.Query()factor. and the setGeoCode() is only a searching factor to find a tweet within nomination position. it carried out daily search to set to the date via setSince(). and both the getLatitude() and getLongitude() are converting function over the longitude and latitude of tweet. The table below is the summary of the main parameters of API related to Twitter.

Table-1 Twitter Search API

Factor	Value	Explanation
Query()	String	Searching word either by title or contents
Get.screenname()	String	posts ID
Get.text()	String	posts contents
Set.Since()	Double	Setup the posts data
Set.Geocode()	String	Set to the spatial scope of posts
Get.latitude()	Double	Altitude coordinate
Get.Longitude()	Double	Longitude coordinate

3.2 Twitter Search

The system is designed to display the posts position, contents, ID through searching using the Twitter Open API phrase.

For searching of twitter posts use the Query s() function.

This study carried out to search for “famous

restaurant", "healing", "food", search are conducted in 3 words only.

The reason for setting 3 words because the contents to search for data was limited such as 100ea every hour, 1,000ea a day according to the API specifications.

In case of same word search per hour, it appeared to acquire the duplicated tweet. therefore the searching for tweet was proceeded to use the 3 words in sequence between 2pm and 12am every hour.

In case of searching the same word per hour the consequence is that there is a duplication. Therefore, searching for tweet was proceeded to use the 3 words in sequence between 2pm and 12am every hour.

The set.geocode() is a factor which brings only a tweet located within designation spacial. Seoul is located on middle west of Korea peninsula with an altitude 37° 34, and longitude 126° 59 according to the Wikipedia, where the distance between east and west is 36.79km and distance between south and north is 30.33km. Thus the central point set to the altitude 37° 34 longitude 126° 59, the diameter set to 19km which is half kilometer of west. east distance because its distance is more longer than south. north, and the tweet data were collected based on the central point altitude 37° 34 longitude 126° 59, and diameter 19km.

Both getLatitude() and get.Longitude(), longitude and altitude display factor, carry out the calculation of tweet's longitude and altitude such as whether or not, if not, it makes to exclude on the other if yes, it records such as ID, contents including posts longitude and altitude.

3.3 Twitter Processing

he period of acquisition of data was between Aug 13th and Aug 20th, we secured approximately 7000 tweet during 7days(1000ea/day). All data can't be used for analysis and some data need to be process.

'Delete the tweet which doesn't have a relation with this study. '

The selection of "famous restaurant", "healing", "food" out of tweets and acquisition of database. But all of tweet didn't match with this study. Result shows when analyzed tweets complied with "famous restaurant" key word, it could acquire tweets data like " cheonju hanok villiage famous restaurant- alchan galbi course", which doesn't have relation with study, accordingly it should be removed from preconditioning process.

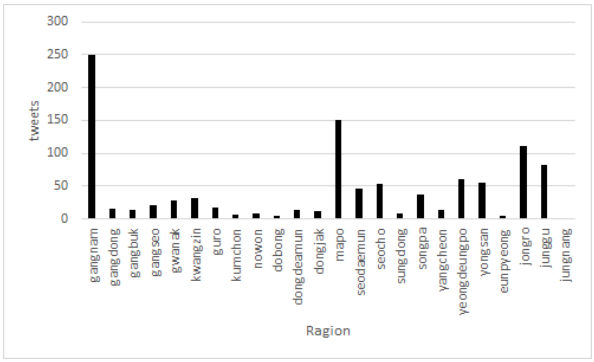
'Classification of tweet data in accordance with sort of food service business'

After removing the tweet which doesn't match with this study, the other classified such as korean, chinese, japanese, western. The korean restaurant is out of them with 3 kinds of food such as meat, tripe, and chicken and then processed because of containing versatile sorts.

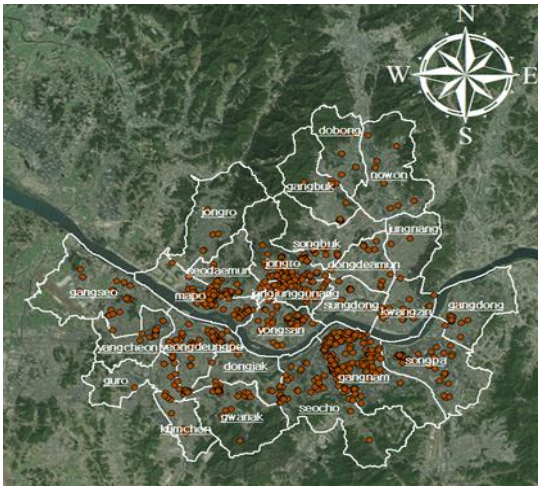
4. Data Analysis

4.1 Utilization of Twitter to check the consumer location

As a result from analysis of the acquired data through the preconditioning process using arcmap, result shows that gangnam district with 250ea(23.7%), mapo district 150ea(14.2%), and jongro district 112ea(10.6%), respectively.



FUGURE-2 tweet Distribution of seoul



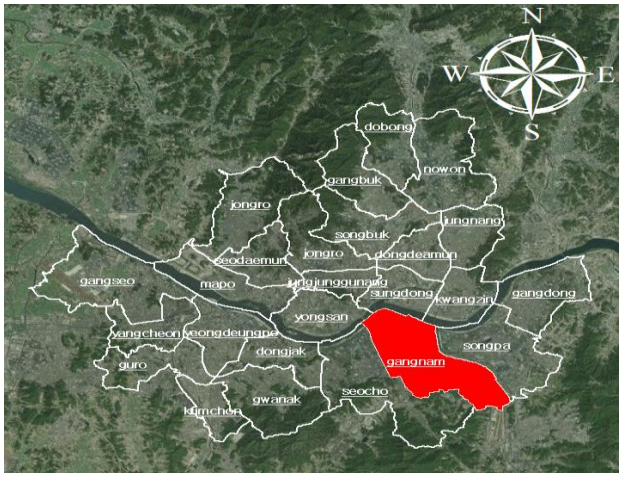
FUGURE-3 Consumption trend of Twitter users

4.2 Standard of business district definition

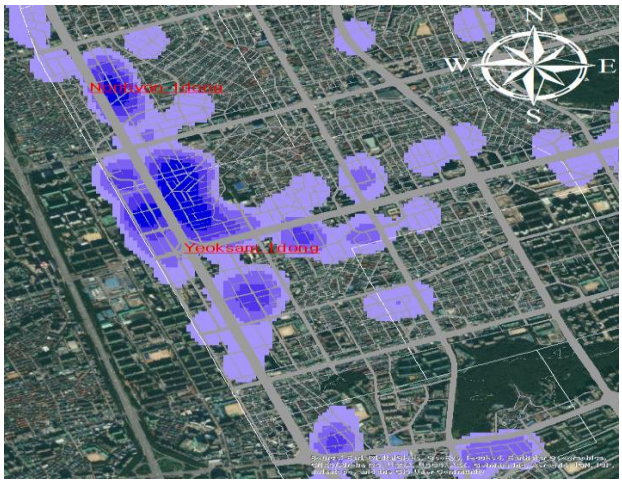
This study was conducted in Gangnam for business district using tweet to identify the location, roads and specifications of business district.

4.3 Scale of Gangnam business district

The Yeoksam dong business district of Gangnam has developed among the venues, where the administration district toward east of Gangnam venue belongs to Yeoksam dong, on the other hand the administration district toward west of Gangnam venue belongs to the Seocho district as a divided business zone. The number of searched tweet in Gangnam district were 250ea, which is the largest amount in seoul. As a result of Kernal Density analysis making use of tweet location data, most highest density area was in between Nonhyon 1 dong and Yeoksam 1 dong.



FUGURE-4 Feature of gangnam

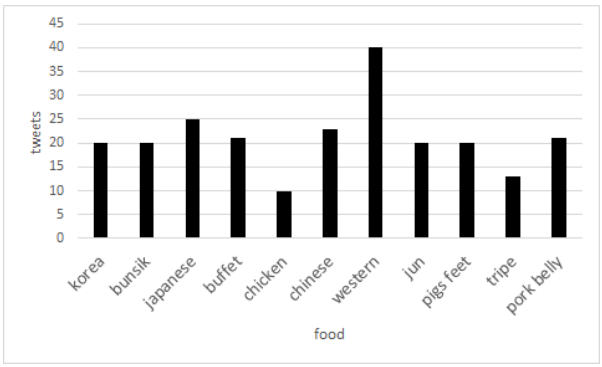


FUGURE-5 Feature of Teoksam 1dong business district

The size of business district was formed between Gangnam station and Shinnonhyeon station until 50m from side to side as a most active zone.

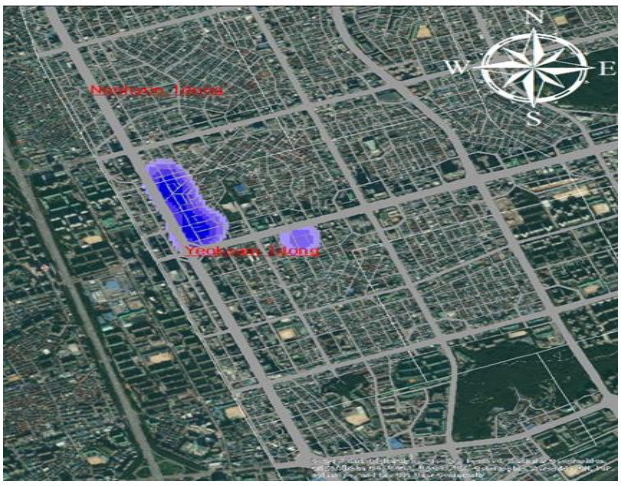
4.4 Feature of Gangnam district business zone

The sort of tweet located on Yeoksam 1dong was shown such as korean 20ea(12), snack bar 20ea(8%), Japanese 25ea(13%), chinese 23(9%), chicken 10ea(4%), pigs' feet 20ea(8%), tripe 13ea(5%), meat 21ea(8%), and western 40ea(33%), where western food is the most highest rate.



FUGURE-6 The sort of tweet in yeoksam 1dong

Also the western food rate was higher than other food service business. The spatial feature is shown that other food service business were formed surrounding on the Gangnam business district, but western food service business tend to be located on surrounding of bridged between Gangnam station and Shinnonhyeon station.



FUGURE-7 Location feature of western food service zone

Seoul City area is the most active business district, where Gangnam district is located and Yeoksam 1Dong which is the most active among the business district. The most active business district was located in the center of venue Yeoksam 1dong between Gangnam station and Shinnonhyon station about 50m from side to side. Even though the versatile food service business were located equally, it was shown that the western food service business was distinguished out of them as well as spatial feature as to the location was densely formed on right side of venue between Gangnam station and Shinnonhyeon station. The survey result shows that there is a significant impact on the store location.

5. Conclusions

The determine the existing business district size and specifications in accordance with scope, access feasibility, floating market, same type of business, and value of land and lease.

The analysis method of business district is normally carried out either by field survey or questionnaire because it consume a lot of time and manpower.

In previous research survey in order to analyze the characteristics of district areas using simple statistic data such as distribution and expenditure of user has been used as the surrogate index however in this study tweet has been used as the surrogate index and results are reliable. To identify the business district characteristics and for decision-making tweet location has been used.

The result of the analysis shows that the most active business district was Gangnam where Gangnam Station and Shinnohyeon Station located as well.

There has been a difficulty in locating the business district area in Seoul using single SNS information or twitter. In order to solve this problem, utilization of other network such as facebook, blog and other network is recommended and to reduce manual process such as preconditioning process.

References

Lee, Y. J., Choei, N. Y., 2005. A Study of Trade Area Analysis by Way of Analog Method. 75-86.

Shim, J. H., 2008. A Study on the marketing Strategy of the Trade Area Analysis in the College Town Street. 203-227

Shim, J. H., 2008. A Study on the marketing Strategy of the Trade Area Analysis in the College Town Street. 203-227

Yang, B. S., Kim, S. J., Jo, Y. W., Jo, M. H., 2009. A Study of the 5 Major Crimes based on GIS Spatial Analysis in Daegu Metropolitan City. Proceeding of 2009 KAGIS Fall Conference. 264-265