International Journal of Geospatial and Environmental Research

Volume 1 | Number 2 Article 5

March 2015

An Analysis on the Spatial Characteristics of Satisfaction on the Residential Environment Using Tweets

Youngok Kang
Ewha Womans University, ykang@ewha.ac.kr

Jaehee Park
Korea Research Institute for Human Settlement, parkjh@krihs.re.kr

Aetti Kang
Ewha Womans University, aetti@naver.com

An Analysis on the Spatial Characteristics of Satisfaction on the Residential Environment Using Tweets

Abstract

The purpose of this study is to analyze the regional difference of spatial distribution of residential satisfaction by extracting the elements of residential satisfaction in the text of tweet data. We determined three themes such as "safety", "amenity" and "convenience", base search terms by theme. And we detailed the search terms by base search term in order to retrieve the tweets related to the satisfaction of residential environments. We analyzed the selected tweets and visualized the results of analysis on the map and then investigated the satisfaction of residential environments through the index analysis which was a proportion of tweet ratio of theme to whole tweet ratio by region This study shows that it may replace the offline survey method by the analysis of tweets on SNS in investigating the satisfaction of residential environments by regions in South Korea.

Keywords

SNS analysis, Tweets, Big data, Satisfaction on residential environment, Search terms, Mapping

1. BACKGROUND AND PURPOSE OF STUDY

With the popularization of internet and smart-phone, Social network service (SNS) such as Twitter and Facebook is becoming the means of communication that people give and take their interests. Many researchers have realized the importance of the big data, stored in SNS platform, which can be a gold mine to identify social signal. So they have performed lots of researches to predict and detect various social phenomena by analyzing the big data stored in SNS platform (Achrekar et al. 2011; Bollen et al. 2011b; Tumasjan et al. 2010).

Twitter (http://www.twitter.com) among many SNS tools is designed that the users upload their short text messages on their timeline and the following users who read that post can retweet or reply it. Tweet data is appropriate to the study in the field of social science because people record their interests and sentiment in the text of Twitter with short but compelling messages. In addition, tweet data also contains user information such as user location, follower, followee, language, attention information such as the number of re-tweets and location information that tweet occurred. As a result, it is possible for us to analyze the distribution of social environment such as residential satisfaction by combining the text information which record people's interests and the location information (Frank et al. 2013a; Mitchell et al. 2013).

Up to now subjective indicator of residential environment such as living satisfaction has been generated through the method of qualitative measurement by survey questionnaires. The survey has a weakness in that it uses a limited amount of sample data and one cannot verify whether the survey represents people's true thoughts or not. On the other hand, one can overcome survey's drawbacks if he/she makes full use of the tweet data which contains people's frank opinions.

We think that it is possible for us to analyze the regional distribution of residential satisfaction by utilizing the text and the location information stored in Twitter platform. The purpose of this study is to analyze the regional difference of spatial distribution of residential satisfaction by extracting the elements of residential satisfaction in the text of tweet data. In addition, this study has significance in that one can attain the implications on the process of study because this kind of research methodology has not been handled much in Korea up to now.

2. LITERATURE REVIEW

Big data has been defined differently in various fields since its appearance. Mayer-Schönberger and Cukier (2013) tried to define big data from the side of social scientist whom big data can be used to analyze the change of society, politics, economy and culture. They defined it as follows: "big data is to extract the insights and/or values from a very large amount of data, which cannot be obtained in small amount of data." The reason that new insights and/or values such as catching the emotion of people and predicting the trend can be extracted from SNS data, which is impossible in existing data, is because today one can utilize a typical data such as social data and web pages which has been accumulated in Twitter and Facebook. Researches that are trying to find new meanings by analyzing these kinds of a typical data are getting increased more and more (Achrekar et al. 2011; Aramaki, Lampos and Cristianini 2010; Paul and Dredze 2011). Google analyzed the distribution of flu sufferers by investigating the frequency of key words related to the flu that users input on the Google search site and then by collecting the locational information through the IP address that the key word occurred. Google confirmed that the number of flu suffers in its analysis is proportional to the real number of flu suffers shown in prevention center for statistics of disease in America. Google's study shows that it is possible for us to find social phenomenon by analyzing the record of users' web searches (Ginsberg et al. 2009).

There are millions of texts which can grasp people's sentiment and circumstances in tweet data because people exchange them with others easily using Twitter (Mitchell et al. 2013). Researchers have tried to perform the analysis of sentiment to take intrinsic emotions out of text in tweet data which can be obtained easily comparing with other SNS tools (Bollen et al. 2011a; Bollen et al. 2011b; Tumasjan et al. 2010). The analysis of sentiment is a method which analyzes people's opinions and sentiment using natural language processing and statistical analysis after detecting key words that express people's opinions and sentiment in various texts on SNS. Bollen J. et al. (2011b) performed the analysis of emotions in order to predict stock price after collecting six different kinds of key words related to emotions using two tools for analysis of emotions. Their prediction results showed 87.6% accuracy comparing with the closing price of DJIA (Dow Jones Industrial Average). This study shows that it is possible for us to detect social phenomena using big data stored in SNS.

Tweet data is coupled with location information such as (1) GPS, (2) Geo-IP and (3) Home location. Home location is input by users themselves and included in Twitter user profile. There are many research that grasp the distribution of social phenomena using location information coupled with tweet data (Ardon et al. 2011; Chandra, Caverlee and Cheng 2011; Roick and Heuser 2013; Xu, Wong and Yang 2013). Li, Goodchild and Xu (2013) explored the socio-economic characteristics by analyzing the coordinated tweets and pictures provided by regional residents among the people who use Twitter. They utilized upload time and frequency of tweets and pictures in order to find where regional residents are dwelling. They developed the model of using tweets and the model of using Flicker picture. And then they found out that the distribution of people who used the Twitter and Flicker were affected by their age and education level. Also, other researches show that location information coupled with tweet data can be utilized to detect the disaster and emergency situations. Earle, Bowden and Guy (2012) performed a research to predict the area of earthquake occurrence as developing a long- and short-term model by inputting the frequency of key words related to the earthquake in tweet data. They compared their results with the records of 5,175 earthquake occurrences during five months that USGS (United States Geological Survey) has measured. It shows that one can predict the area of earthquake occurrences with 75% accuracy through their models. Kent and Capello Jr. (2013) collected and analyzed tweet data related to the forest fire in which occurred in the Horsethief canyon in reality in 2013 in America in order to find the elements which take an effect on the distribution of human beings who inform regional events such as forest fires. They tried to (1) find the index pattern of community that contributes to situational awareness during emergency period by applying the spatial analysis technique and (2) prove the community profile that contributes to the action during forest fire by applying spatial regression analysis with demographic variable. Unfortunately, their research has some weakness in that (1) the ratio of tweet with location information to whole tweet forms around one percent (Davis Jr. et al. 2011) and (2) the tweets that occurs out of residential area of tweet users are not reflected in the process of event awareness. In order to resolve these kinds of drawbacks, Hwang (2013) tried to extract the event based on the district that the tweet occurred instead of GPS coordinates of the tweet. He confirmed that it is possible for us to detect the event effectively as detecting the forest fire which occurred in Pohang in March 9 in 2013 using the tweet data of Korean users that occurred from March 4 to 29 in 2013.

In addition, new studies, which aim at grasping social atmosphere and trend using SNS that people freely exchange their interests and opinions each other, are also emerging (Paul and Dredze 2011; Frank et al. 2013b; Mitchell et al. 2013; Paul and Dredze 2011). Usually, it is difficult detect social atmosphere and trend because they include people's subjective elements. Especially, it is more difficult to visualize the distribution of social atmosphere and trend because people who form the distribution of social atmosphere and trend don't want their home location. However, Ghosh and Guha (2013) showed it is possible to predict the distribution by analyzing tweet text coupled with the geographical coordinates. They predicted the distribution of obesity and analyzed its causes. They collected tweet data that contain the topic components related to obesity through API and transformed tweet data with geographical coordinates into spatial data in the process of removing the noise of tweet data. They analyzed the general pattern of obesity by using point density tool. As a result of analyzing the correlation with the percentage of adult obesity from CDCP (Center for Disease Control and Prevention), they found that obesity is inversely proportional to the level of education. In addition, they extracted the themes related to obesity from tweet text by applying text mining technique. They selected three themes by applying topic modeling and analyzed the spatial distribution by theme. As a result, they confirmed that both government policy and tweet occurrence location take an effect on the obesity distribution.

In the past, survey questionnaire and spatial analysis which utilized the census data or survey data have been used to analyze the people's preference such as happiness and residential satisfaction. But recently, there is a new study to visualize the people's sentiments or emotions by geocoding tweet text. Mitchell et al. (2013) performed the study to draw a map of happiness by using the sentiment analysis, location of tweet text and objective characteristics of places. They gathered 800,000 of geographically coordinated tweet data produced in 2011 and compared their result with the result of survey questionnaire. And then they developed the taxonomies of states and cities based on the similarity of terminology usage and measured the level of happiness by states and cities. They concluded that happiness is strongly correlated with wealth by showing large positive correlation with high household income. Unfortunately, in case of Korea we can find few any researches to analyze subjective index such as happiness and residential satisfaction using geographically coordinated tweet data.

3. RESEARCH PROCEDURE

The research procedure, which shows the work flow of this study, has six steps as shown in Figure 1. First, we collected the tweet data with GPS information through streaming API by defining the rectangular area which covers whole South Korea. Total 1,305,173 tweet data was collected during the period from Nov.30 in 2012 to Jan.10 in 2013. The distribution of the tweet data is shown on the map, as shown in Figure 2. Figure 2 shows that tweet data has occurred intensively in major large cities such as Seoul and Busan and others along the highway.

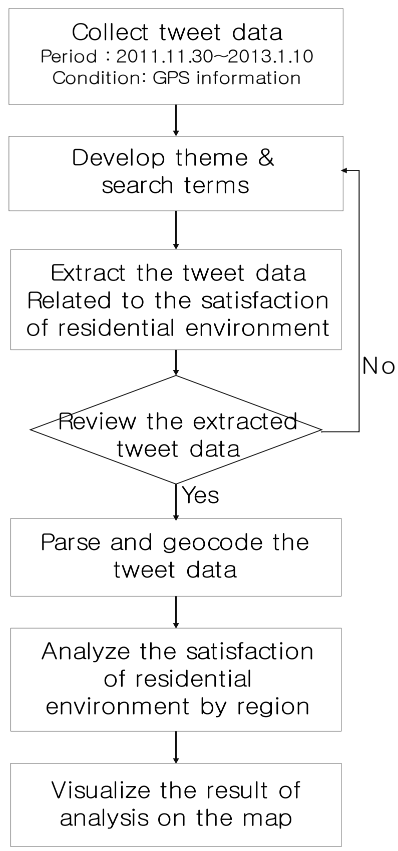


Figure 1. Research Procedure

Second, we developed themes and search terms to retrieve the tweet data related to the satisfaction of residential environments. Three themes such as “safety”, “amenity” and “convenience” and search terms by theme were developed. Third, we extracted total of 516 tweet data related to the satisfaction of residential environments by using the search terms developed above. Fourth, we parsed and encoded the tweet data. That was, we had parsed the values of tweet’s time, text and location and then geocoded the tweet data through point layer function in ArcGIS Desktop 10.1 in order to draw a map.

Fifth, we analyzed the satisfaction of residential environments by regions in South Korea using the satisfaction index which represented the relative satisfaction of residential environments by theme. Finally, we visualized the results of analysis on the map. The satisfaction of residential environments in 16 large cities and states in South Korea was compared each other by visualizing the values of satisfaction index on the map. That was, we divided the satisfaction indexes into five classes which had different colors and then visualized them on the map in order to compare the distribution of residential satisfaction by region.

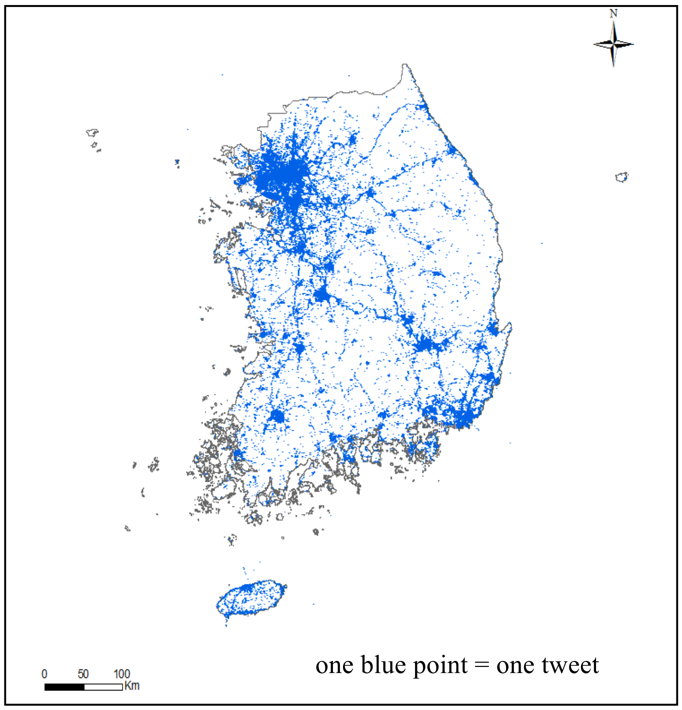


Figure 2. Distribution of geolocated tweets collected

4. SELECTION OF THEMES AND SEARCH TERMS

Generally, the elements of residential environments are classified into 4 categories such as safety, amenity, convenience and health based on the basic living requirements of human beings that WHO (World Health Organization) proposed in 1961 (Kim and Ahn 2004). Some researchers add sociality and economy in addition to four categories above. (Jang 1992, Kim 2003, Kim and Ahn 2004).

The safety of residential environment is defined to be the state of being "safe" in view of life, body, property, activity and function of human beings (Lee 2012). The health of residential environment is defined as maintaining healthy life when considering one's own house and/or its macro-environment (Lee 2012). The convenience of residential environment means whether human being can perform any kinds of activities without receiving any stress when he intends to take any action in his life (Lee 2012). The convenience is usually evaluated according to the daily living convenience, accessibility and availability of all sorts of facilities and organizations (Lee 2012). The amenity of residential environment is defined to be physiologically comfortable in view of safety and sanitary (Lee 2012). Lee (2012) performed a study of satisfaction of residential environment by defining survey terms as shown in Table 1, based on previous researches (Yasushi 2003, Kim and Ahn 2004).

Table 1. Themes and survey terms of residential environment satisfaction

Theme	Survey Terms
Safety	Fear on night journey, Nervousness of door locking, Car accident, Fire, Flooding, Landslide
Health	Less noise, Less factories, Enough space for walking and exercising, Emotional stability
Convenience	Water, Electricity, Gas, Parking, Garbage disposal, Hospital, House remodeling, Public facilities, Telephone, Market, Public transportation, Commuting, Road accessibility
Amenity	Good for ventilation, sunlight, grass, tree and park, view, No sewage and garbage

Source: Lee 2012

In order to determine themes and search terms related to the residential satisfaction we utilized the survey terms shown in Table 1 on the tweet and excluded the theme and the search term that are not retrieved on the tweet from survey themes and terms. As a result, the themes are classified into three categories such as “safety”, “amenity” and “convenience” by considering the characteristics of information obtained from the tweet and the search terms by each theme are illustrated in Table 2.

Table 2. The list of themes and search terms

Theme	Search term
Safety	Fear of night journey
	Car accident
	Risk of disaster and accident
Amenity	Noise
	Parking inconvenience
Convenience	Inconvenience of public and owner-driven transportation
	Inconvenience of super-market, traditional market, town and house itself

We selected search terms by referring previous study, but we need to revise search terms a little bit because colloquial or slang expressions are more frequently used in the tweet. In addition, it is important to choose appropriate scope of search terms. If we define the scope of search terms too broadly, many unrelated tweets can be retrieved. On the other hand, if we define the scope of search terms too narrowly, some related tweets cannot be retrieved. Therefore, for the purpose of our research, we extracted detailed search terms from base search terms by referring both their characteristics and the expressions frequently used in the tweet.

In order to define detailed search terms, we referred to the ‘search word map’ and ‘related positive – negative word’ provided by Social Metrics (www.socialmetrics.co.kr) which is a website of Daum Soft, a provider of social analysis service. The website provides a result of analysis by collecting the documents which contain search terms in Twitter and blog for one month from the starting point. The procedure of defining detailed search terms is as follows: First, we selected a natural language by filtering the documents and then removing both spam and noise data within the documents. Second, we selected related search terms based on the frequency of the words that mentioned in those search terms. Finally, we defined related ‘positive–negative’ terms after judging positive, negative, and neutral of the search terms based on the sentiment dictionary which was a collection of expressions about subjective evaluation of things and situations. As a result, we could grasp the rough tendency of which detailed key terms are used together with any kinds of positive-negative terms even though there was a time difference between the data collection for this study and the one for the analysis.

Table 3. Sentiment Dictionary of search terms by theme

Theme	Base search term	Detailed search term
Safety	Fear of night journey	night journey + fear, fear_xx (various suffix in Hangeul)
	Car accident	car accident, transportation accident, bus accident
	Risk of disaster and accident	accident + risk, accident + risk
Amenity	Noise	floor noise, neighbors noise, side/upper/lower house + noise_xx (various suffix in Hangeul) + call
	Parking inconvenience	parking inconvenience, illegal parking, parking violation
Convenience	Inconvenience of public and owner-driven transportation	transportation, stop + inconvenience, station + inconvenience, commuting + inconvenience
	Inconvenience of Supermarket, traditional market and town	super/mart/market + inconvenience, town/house + inconvenience

As we mentioned before, the terms generally used in the tweet are colloquial or slang, not literary. Most people do not mind grammar or spelling when they give and take the messages on SNS. Besides, in Korean Hangeul we need to include the stem words[1] in

[1] It means the indeclinable part of word. In Korean Hangeul, word is composed of stem word of indeclinable part and ending word of declinable part. The sentence of “This is love” in English is translated into “사랑한다(Saranghanda)”. And the sentence of “Is this love?” in English is translated into

detailed search terms because the change of suffixation in Korean Hangeul is dynamic. If we composed detailed search terms using the term only, the homonymic terms, which had totally different meanings, were also retrieved. But if we included the stem terms into detailed search terms, we could extract the tweet which was more adequate for our purpose. In addition, since there were much more negative expressions than positive expressions on the tweet, we needed to define the detailed search terms that express the dissatisfaction of residential environments. Table 3 shows the dictionary of themes, base search terms and detailed search terms we defined.

If any irrelevant tweets were searched due to the use of homonymic word, we made them not be selected by using the exceptive word. During the period of tweet data collection, there was a presidential election in South Korea. Therefore, a lot of tweets related to the noise of election campaign had been retrieved when we extract 'noise' related tweets. In this case, we excluded those tweets related to the 'noise' of election campaign from 7.3% of entire tweets. Figure 3 is the distribution of tweets related theme.

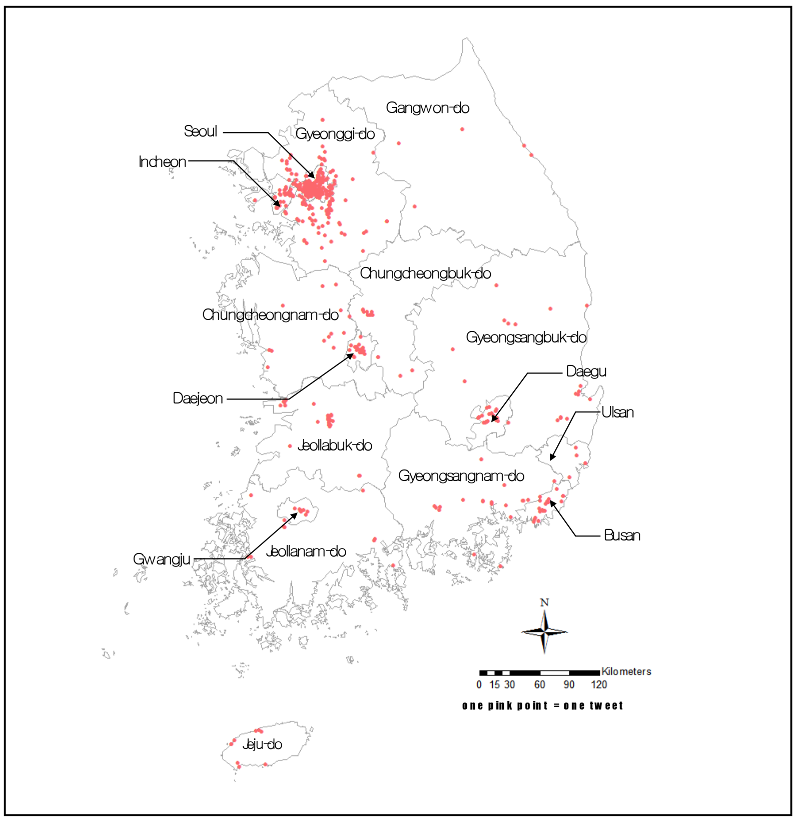


Figure 3. Distribution of geolocated tweets that were used in analysis

"사랑하니?(Saranghani)" in Korean. In this point, "사랑(Sarang)" is the stem word and "한다(handa)" or "하니(hani)" is the ending word.

5. ANALYSIS AND VISUALIZATION OF SPATIAL CHARACTERISTICS

We extracted the tweets by using JavaScript and then arrange those tweets in Excel. Then we transformed the tweet data into the point data by using Make XY Event Layer tool of ArcGIS Desktop 10.1. Finally, we depicted the distribution of the tweet as a map. Total 516 tweets, which included three themes such as "safety", "amenity" and "convenience", were extracted from the detailed search terms we defined. Table 4 shows the ratio of tweet distribution by each theme and by each city or state.

Table 4. Tweet ratio of each region by theme

Region[2]	Whole tweet		Safety tweet		Amenity tweet		Convenience	
	ratio (%)	number	ratio (%)	number	ratio (%)	number	ratio (%)	number
Seoul	35.85	185	31.09	83	38.51	56	44.55	44
Busan	4.26	22	4.12	11	3.38	5	5.94	5
Daegu	4.26	22	3.75	10	4.73	7	4.95	4
Incheon	5.04	26	7.49	19	2.03	3	2.97	2
Gwangju	1.55	8	1.12	2	2.03	3	1.98	1
Daejeon	3.49	18	2.62	6	5.41	8	2.97	2
Ulsan	0.97	5	0.37	0	2.7	3	0	0
Gyeonggi-do	22.48	116	22.47	59	20.27	29	25.74	25
Gangwon-do	0.97	5	1.5	4	0.68	1	0	0
Chungcheongbuk-do	2.52	13	2.62	6	3.38	5	0.99	0
Chungcheongnam-do	2.71	14	3.37	8	3.38	5	0	0
Jeollabuk-do	4.84	25	6.37	17	3.38	5	2.97	2
Jeollanam-do	2.13	11	3	8	1.35	1	0.99	0
Gyeongsangbuk-do	3.49	18	4.49	11	2.7	3	1.98	1
Gyeongsangnam-do	3.68	19	3.37	8	5.41	8	1.98	1
Jeju-do	1.74	9	2.25	6	0.68	1	1.98	1

[2] We say the region for your information, Seoul is the largest city in South Korea and then Busan, Incheon, Daejeon , Daegu, Gwangju, Ulsan in that order. These cities are called as metropolitan city. Do is the equal to metropolitan city level such as Seoul and Busan. Gyeonggi-do is the largest province in South Korea and the others have very similar populations.

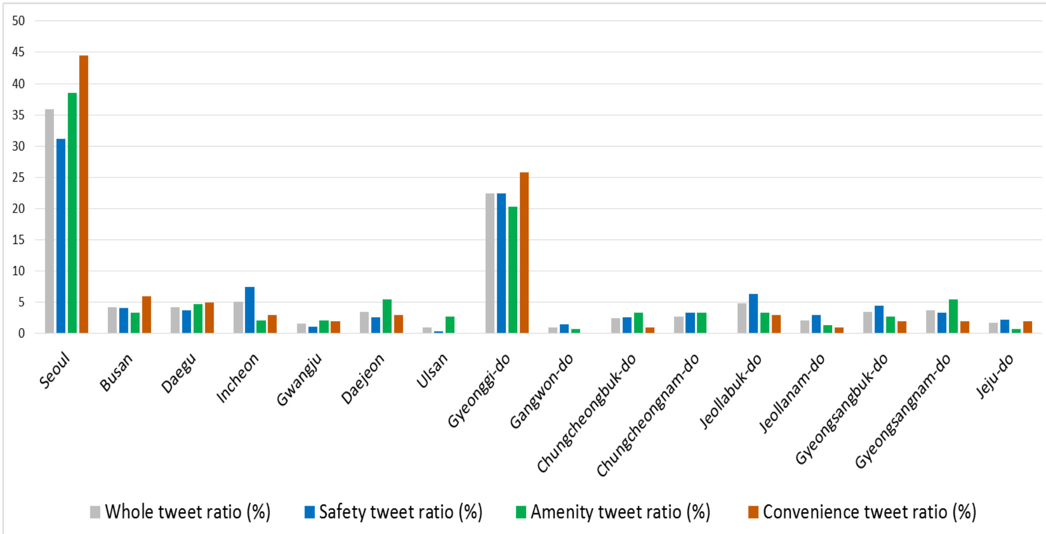


Figure 4. Tweet ratio by theme (%) (Refer to Figure 3 for *si-do* names.)

This study, which analyzed the spatial characteristics with satisfaction of residential environments with tweet data have significance in two points of view. First, we analyze the people in which region feel that its residential environments are relatively satisfactory comparing with other regions. Second, we distinguish whether there are any differences in the satisfaction of residential environments by detailed search term of the region.

It is not appropriate for us to analyze the satisfaction of residential environments with the tweet ratio of each city or state. So we think that it is more appropriate for us to analyze the relative index in the situation that there is a huge difference in the number of tweet users by region. The relative index is calculated as follows: (1) Calculate the whole tweet ratio by region - dividing the number of tweet by region by the total number of tweet. (2) Calculate the tweet ratio of each theme by region - dividing the number of tweet of each theme by region by the total number of tweet of each theme. (3) Calculate the index – dividing the tweet ratio of each theme by region by the whole tweet ratio by region. For example, as shown in Table 4, if the whole tweet ratio of Seoul is 35.85% and tweet ratio upon "safety" of Seoul is 31.09%, then the index of Seoul is 0.86. The index represents the relative satisfaction of residential environments by theme. For example, in case of "safety" factor, if the index of one region is higher than the one of other region, it means that the region is relatively less safe than other region.

5.1 Safety

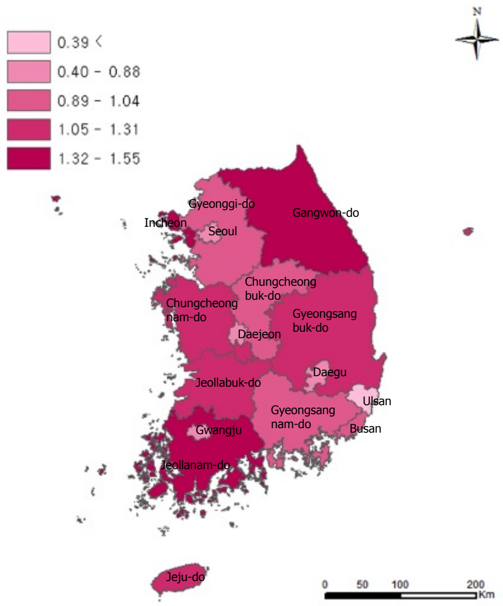
As a matter of fact, the "safety" represents the "unsafety" of residential environments in this study because the detailed key words upon "safety" belong to negative words in Korean, not positive words. Total of 267 tweets, which represent the "unsafety" of residential environments, have been extracted from whole 516 tweets. Gangwondo, Incheon and Jeonranamdo are relatively unsafe in the order of high index (Figure 5). On the contrary, Ulsan, Kwangju, Daejeon and Seoul show relatively higher values of safety in the order of low index. High index represents that people in the region express more active opinion upon unsafety of residential environments. In three regions with high density, that is, where people used more expressions of unsafety, we analyzed in which base search term of "safety" takes an effect on the dissatisfaction of residential environments. In all three regions, the 'fear of night journey' largely takes an effect on "safety", comparing with car accident and other accidents (Figure 6).

5.2 AMENITY

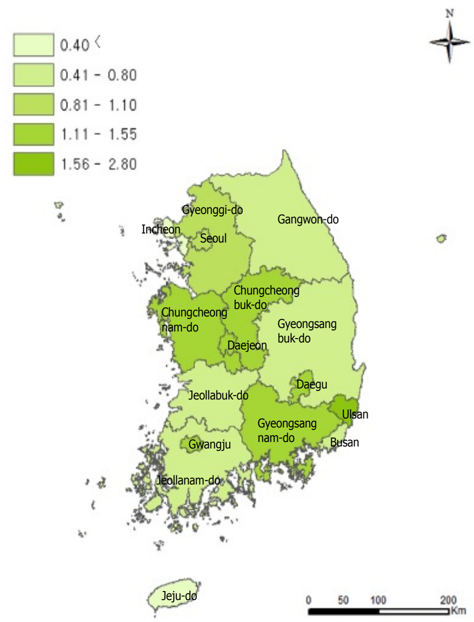
In the same way, the "amenity" represents the "dis-amenity" of residential environments because the detailed search terms upon "amenity" belong to negative words, not positive words. Total of 148 tweets, which represent the "dis-amenity" of residential environments, have been extracted from whole 516 tweets. Ulsan, Daejeon, Gyungsannamdo and Chungcheongbukdo have higher index upon "amenity", comparing with other regions (Figure 5). High index represents that people in the region express more active opinion upon dis-amenity of residential environments. In three regions with high density, that is, where majority of Twitter users expressed the feeling of dis-amenity, we analyzed in which base search term of "amenity" takes an effect on the dissatisfaction of residential environments. In all three regions, the 'noise' largely takes an effect on "amenity", comparing with parking problem (Figure 6).

5.3 CONVENIENCE

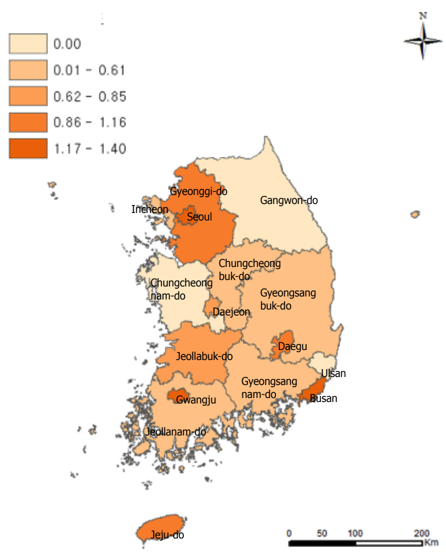
In same way, the "convenience" represents the "in-convenience" of residential environments because the detailed search terms upon "convenience" in Korean belong to negative words, not positive words. Total of 101 tweets, which represent the "in-convenience" of residential environments, have been extracted from whole 516 tweets. The region which has more tweets upon "in-convenience" means that the region show relatively higher values of inconvenience in view of the availability of transportation and the use of supermarket and store. Even though Busan, Kwangju and Seoul have more transportation and convenience facilities, people in those regions express more "inconvenience", comparing with the people in other regions (Figure 5). Figure 6 shows the ratios by detailed search terms in three regions in which part of "inconvenience" people express "inconvenience" in their tweets. As shown in Figure 5, in the satisfaction of residential environments people put more tweets upon "inconvenience" of mart or store than those upon "inconvenience" of commuting.



Safety Index



Amenity Index



Convenience Index

Figure 5. Satisfaction index of residential environment by theme

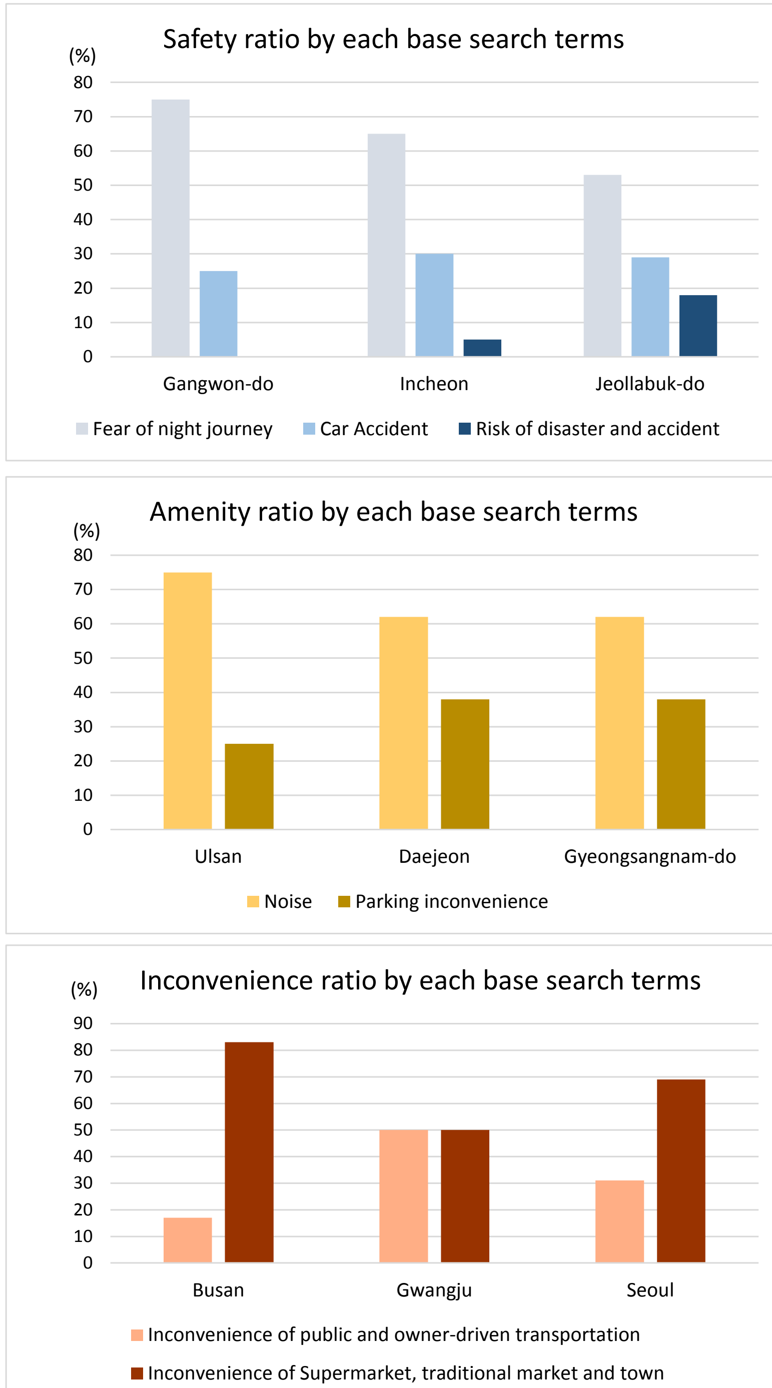


Figure 6. The ratios of base search terms in selected three regions with high index by theme

6. CONCLUSION

We collected a total of 1,305,173 tweets with GPS information for three months from November, 2012 using Twitter API within the bounding box of covering South Korea. And then we remove the tweets uploaded in Japan. Then we determined three themes such as “safety”, “amenity” and “convenience”, base search terms by theme and detailed search terms by base search term in order to retrieve the tweets related to the satisfaction of residential environments. Next we analyzed the selected tweets and visualized the results of analysis on the map. We investigated the satisfaction of residential environments through the index analysis which is a proportion of tweet ratio of theme to whole tweet ratio by region. According to our analysis, the index of “safety” is relatively high in Gangwondo and Jeonranamdo, the index of “amenity” is relatively high in Ulsan, Daejeon, Gyungsangnamdo and Chungcheongbukdo, and the index of “convenience” is relatively high in Busan, Kwangju and Seoul, South Korea.

This study has significance in three folds: First, we defined themes, base search terms and detailed search terms to retrieve tweets related to the satisfaction of residential environments. Second, we analyzed the difference in the satisfaction of residential environments by regions in South Korea and we visualized the satisfaction of residential environments on the map. Finally, this study shows that it may replace the offline survey method, which is very time consuming and expensive, by the analysis of tweets on SNS, which is very fast and cheap to collect, in investigating the satisfaction of residential environments by regions in South Korea.

But this study has some limitations as follows; first, the number of tweets related to search terms is not sufficient due to the limited period of data collection time. Second, only text analysis based on count technique has been applied in retrieving the tweets related to the search term. Therefore, the methodology such as the sentiment analysis which finds out the feeling of satisfaction more sophisticatedly would be applied in further research. Third, we do not consider the synchronization between the uploaded tweet location and the home location. And the last, as the problem of survey terms it’s the problem that the elements of the satisfactory of residential environments was missing. For example terms about “health” is not provided in Table 1. In the future, it is necessary for us to (1) increase the number of tweets by securing enough period of time and by utilizing the sentiment analysis and relationship information of tweets, (2) collect extensive opinions of residential satisfaction by applying the technique which predicts home location and (3) utilize the census data to make prediction in the future extension.

REFERENCES

Achrekar, H., A. Gandhe, R. Lazarus, S.-H. Yu & B. Liu (2011) Predicting flu trends using twitter data. In *Computer Communications Workshops* (INFOCOM WKSHPS), 2011 IEEE Conference on, 702-707. IEEE.

Aramaki, E., S. Maskawa & M. Morita (2011) Twitter catches the flu: detecting influenza epidemics using Twitter. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 1568-1576. Association for Computational Linguistics.

Ardon, S., A. Bagchi, A. Mahanti, A. Ruhela, A. Seth, R. M. Tripathy & S. Triukose (2011) Spatio-temporal analysis of topic popularity in twitter. *arXiv preprint arXiv:1111.2904.*

Bollen, J., H. Mao & A. Pepe (2011a) Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In *ICWSM.*

Bollen, J., H. Mao & X. Zeng (2011b) Twitter mood predicts the stock market. *Journal of Computational Science*, 2, 1-8.

Chandra, S., L. Khan & F. B. Muhaya (2011) Estimating twitter user location using social interactions--a content based approach. *In Privacy, security, risk and trust (passat), 2011 IEEE third international conference on and 2011 IEEE third international conference on social computing (socialcom)*, 838-843. IEEE.

Davis Jr, C. A., G. L. Pappa, D. R. R. de Oliveira & F. de L Arcanjo (2011) Inferring the location of twitter messages based on user relationships. *Transactions in GIS*, 15, 735-751.

Earle, P. S., D. C. Bowden & M. Guy (2012) Twitter earthquake detection: earthquake monitoring in a social world. *Annals of Geophysics*, 54.

Frank, M. R., L. Mitchell, P. S. Dodds & C. M. Danforth (2013) Happiness and the Patterns of Life: A Study of Geolocated Tweets. *arXiv preprint arXiv:1304.1296.*

Ghosh, D. & R. Guha (2013) What are we 'tweeting'about obesity? Mapping tweets with topic modeling and Geographic Information System. *Cartography and Geographic Information Science*, 40, 90-102.

Ginsberg, J., M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski & L. Brilliant (2009) Detecting influenza epidemics using search engine query data. *Nature*, 457, 1012-1014.

Jang, H. Y. (1992) *A Study on Housing Satisfaction according to Housing Values style*, Dong-A University, M.A. Thesis.

Kent, J. D. & H. T. Capello Jr (2013) Spatial patterns and demographic indicators of effective social media content during the Horsethief Canyon fire of 2012. *Cartography and Geographic Information Science*, 40, 78-89.

Kim, C. S. (2003) *Effects of Apartment Residing Environment Factors on Satisfaction of Residing Environment*, Changwon National University, M.A.thesis.

Kim, S. H. & K. O. Ahn (2004) Needs of Local Housing Plans according to the Evaluation of Residential Environment, *Journal of The Architectural Institute of Korea*, 20(1), 159-169.

Lampos, V. & N. Cristianini (2010) *Tracking the flu pandemic by monitoring the social web.*

Lee, H. N. (2012) *A Study on the Housing Type Choice through the Residential Environment Evaluation*, University of Seoul, M.A.Thesis.

Li, L., M. F. Goodchild & B. Xu (2013) Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. *Cartography and Geographic Information Science*, 40, 61-77.

Mayer-Schönberger, V. & K. Cukier (2013) *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

McGee, J., J. A. Caverlee & Z. Cheng (2011) A geographic study of tie strength in social media. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, 2333-2336. ACM.

Mitchell, L., M. R. Frank, K. D. Harris, P. S. Dodds & C. M. Danforth (2013) The Geography of Happiness: Connecting Twitter sentiment and expression, demographics, and objective characteristics of place. *PloS one*, 8, e64417.

Paul, M. J. & M. Dredze (2011) You Are What You Tweet: Analyzing Twitter for Public Health. In *ICWSM*.

Roick, O. & S. Heuser (2013) Location Based Social Networks–Definition, Current State of the Art and Research Agenda. *Transactions in GIS*.

Tumasjan, A., T. O. Sprenger, P. G. Sandner & I. M. Welpe (2010) Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *ICWSM*, 10, 178-185.

Xu, C., D. W. Wong & C. Yang (2013) Evaluating the “geographical awareness” of individuals: an exploratory analysis of twitter data. *Cartography and Geographic Information Science*, 40, 103-115.

Yasushi, A. (2003) *Residential Environments: Method of Evaluation and Theory*, L. Kang, I. Park, K. Lee, and J. Choi (ed), Seoul: SpaceTime.

Yim, J., S. Kim, J. Yoon, P. Oh, B. Lee & B. Y. Hwang (2013) Detecting Local Events Using Twitter. *Korean Institute of Information Scientists and Engineers, Korea Computer Congress 2013, 248~ 250.*