# Recommendation of Associated Tourist Attractions Based on SNS Analysis

Kaaen Kwon and Ah Cho
Department of Business Data Convergence, Chungbuk National University, Cheongju, Korea
Email: ggaaen@gmail.com, huanying20@naver.com

Wan-Sup Cho[1], Kwan-Hee Yoo[2], and Ga-Ae Ryu[3]
[1]Department of Management Information System, Chungbuk National University, Cheongju, Korea
[2]Department of Software Engineering, Chungbuk National University, Cheongju, Korea
[3]Department of Digital Informatics and Convergence, Chungbuk National, University, Cheongju, Korea
Email: wscho63@gmail.com, khyoo@chungbuk.ac.kr, relaxative@nate.com

***Abstract*—With the development of mobile devices and the supply of internet, information exchange has actively been made through SNS like blogs. In particular, blogs are widely used as a space where people share their experience after their visit to tourist attractions. Although the analysis on blog articles is expected to draw meaningful information, relevant research has yet to be conducted actively. This study proposes a method of recommending associated tourist attractions based on tourist's opinions using Association Analysis and Text Network Analysis (TNA), in order to help to develop tour products and policies**

***Index Terms*—SNS, tourist attraction, text network analysis, association rules**

## I. Introduction

Thanks to the development of IT and the emergence of various analytical techniques, it is possible to analyse unstructured text data collectable in the internet. Text data is one of the main data necessary to establish corporate marketing strategies and policies. Korea Tourism Organization has been analysed on the SNS data about local cultural festivals [1]. However, the research has the limitation in analysing the keywords associated cultural festivals of a specific region based on simple frequency analysis.

To overcome the limitation of the previous research, this study conducted Association Analysis and TNA with 55 tourist attractions in Chungcheongbukdo. This paper is comprised of as follow: in chapter 2, the studies related to the association analysis and the text network analysis are investigated; in chapter 3, a method of drawing associated tourist attractions is proposed; in chapter 4, experiment results and analysis results are presented; in chapter 5, this study comes to an conclusion.

## II. Related Research

Association rules are used to not only find meaningful patterns from transaction data, but draw relevant keywords or implement a recommended system [2], [3]. However, when a data set becomes large, it is hard to understand the results drawn.

Text network analysis (TNA) is used to find degree centrality of keywords by using nodes and links [4]. The more nodes have connections, the higher degree centrality is [5]. With the use of TNA, it is possible to find a core node.

## III. The Proposed Method

The analysis process is comprised of the following five stages: 1) data collection stage, 2) pre-processing stage, 3) data analysis stage, 4) refinement stage, and 5) result stage. Fig. 1 illustrates the flow chart of the analysis process

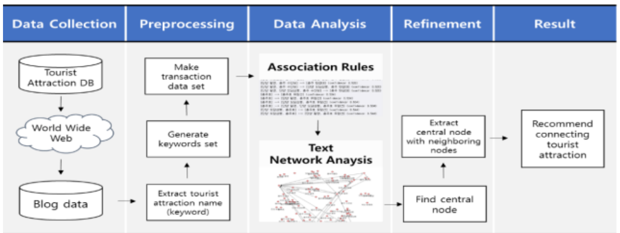


Figure 1. Analysis process of identification of proposed method

### A. Data Collection Stage

The blogs used for analysis are collected with keywords by the search engine Naver [6]. The data about a total of 381 tourist attractions of Chungcheongbukdo are drawn from Tour API [7]. Based on them, a database is established. A name of a tourist attraction is used as a search keyword.

### B. Preprocessing Stage

The names of tourist attractions are extracted from the collected blog articles. According to each blog, a set of keywords are made. With the keyword set and a blog's unique ID, transaction data are created. In other words, a blog's unique ID becomes a transaction ID, and a set of extracted tourist attraction names becomes an item set.

Manuscript received December 20, 2014; revised June 5, 2015.


doi: 10.12720/joams.4.5.393-396

### C. Data Analysis Stage

(1)Association Rules-FP-Growth algorithm is applied to find subsets of an item set. Among the created subsets, the subsets with more than 0.2 of support value are extracted to make association rules. (2) Text Network Analysis-NetMiner 4.0 is used to convert association rules into 1-mode network structure. Each node in the network structure means a tourist attraction, and a link to connect nodes represents that two keywords are mentioned in the same blog. Degree centrality analysis is conducted to draw a degree of connection between a central node and each node. In this way, the tourist attraction that becomes a central node is drawn.

### D. Refinement Stage

In-degree centrality is calculated to draw a central node. With the nodes with the highest degree centrality, a network is extracted and is visualized into Spring Network Map and Concentric Network Map. With the nodes with the next highest degree centrality, another network is extracted, and is visualized in the aforementioned way.

### E. Result

Based on the extracted networks, associated tourist attractions are recommended.

## IV. Experiment and Result

In this paper, we define associated tourist attraction as tourist attractions that show up on the same blog. The analysis period is from Sep. 1, 2013, to Aug. 31, 2014. The study subjects were the articles of Naver blogs. Among a total of 381 tourist attractions in Chungbuk Province, the first five main tourist attractions in 11 local governments were chosen. Therefore, a total of 55 tourist attractions were used as search keywords. The number of the blogs collected was 42,405. Through pre-processing, 25,620 transactions (i.e., the number of articles) and 231 items (i.e., tourist attraction name) were obtained.

To find out the relationship among tourist attractions, we did association analysis and 136 association rules were drawn. Table I presents the example of the drawn association rules.

TABLE I. Example of Association Rules

| No | Premises | Conclusion | Support |
|---|---|---|---|
| 1 | Danyang Gosu Cave | Danyang Dodamsambong Peaks | 0.088 |
| 2 | Chungjuho Lake | Danyang Eight Sceneries | 0.155 |
| 3 | Danyang Gosu Cave | Jecheon Cheongpungho Lake | 0.113 |

To intuitively and easily understand the created rules, we used NetMiner 4.0 to convert them into a network structure. The converted network has a total of 35 nodes and 136 links. However, this network cannot have a direction because we derived the association rules with the tourist attractions showed up together. So, we analyze the network data based on In-Degree Centrality because the centrality values of undirected network are same between In-Degree Centrality and Out-Degree Centrality. Table II shows the measured value of degree centrality of each node and the first letter of Tourist Attraction ID refer to the region

TABLE II. Degree Centrality of Each Node

| No | Tourist Attraction ID | Tourist Attraction Name | In-Degree Centrality |
|---|---|---|---|
| 1 | D11 | Danyang Gosu Cave | 0.040726 |
| 2 | D7 | Danyang Dodamsambong Peaks | 0.009303 |
| 3 | C2 | Chungjuho Lake | 0.008121 |
| 4 | D1 | Danyang Eight Sceneries | 0.008076 |
| 5 | C1 | Chungjuho Lake cruise ship | 0.007324 |
| 6 | D4 | Danyang Stone Gate | 0.006144 |
| 7 | Z1 | Jecheon Cheongpungho | 0.006004 |
| 8 | Z4 | Jecheon Oksunbong Peak | 0.005516 |
| 9 | D10 | Danyang Gudambong Peak | 0.004993 |
| 10 | D6 | Danyang Sainam Rock | 0.004282 |
| 11 | C7 | Chungju Namhangang | 0.003802 |
| 12 | Z3 | Jecheon Uirimji Reservoir | 0.003294 |
| 13 | U3 | Chungju Suamgol Village | 0.003071 |
| 14 | C6 | Chungju Suanbo | 0.002338 |
| 15 | D8 | Danyang Geumsusan Mountain | 0.001432 |
| 16 | C3 | Chungju Dam | 0.00132 |
| 17 | Z2 | Jecheon Cheongpung Cultural Heritage Complex | 0.001251 |
| 18 | D9 | Danyang Guinsa Temple | 0.001084 |
| 19 | U4 | Chungju Sangdangsanseong Fortress | 0.001055 |
| 20 | Z5 | Jecheon Bakdaljae Pass | 0.001025 |
| 21 | D5 | Danyang Sangseonam Rock | 0.000994 |
| 22 | C5 | Chungju Tangeumdae Terrace | 0.000924 |
| 23 | G3 | Goesan The way for the old mountain lodge | 0.000871 |
| 24 | J2 | Jincheon Nongdari Bridge | 0.000823 |
| 25 | G1 | Goesan Hwayang Valley | 0.000661 |
| 26 | U2 | Chungju Cheongnamdae Presidential Villa | 0.000643 |
| 27 | B1 | Chungbuk Alps Recreational Forest | 0.000633 |
| 28 | D3 | Danyang Ondal Tourist Site | 0.000626 |
| 29 | U1 | Chungju Zoo | 0.000608 |

According to the analysis, D11 (Danyang) node had the highest degree centrality, followed by D7 (Danyang), C2 (Chungju), and Z1 (Jecheon). High degree centrality means that the frequency of tourists' associated visits is high. Therefore, the Chungbuk Province tourist attractions with high frequency of associated visits were found to be centralized in the northern areas of Chungbuk Province, such as Danyang, Jecheon, and Chungju. The degree centrality network with 35 nodes was visualized into spring network map and concentric network map. Fig. 2 illustrated the visualized maps.

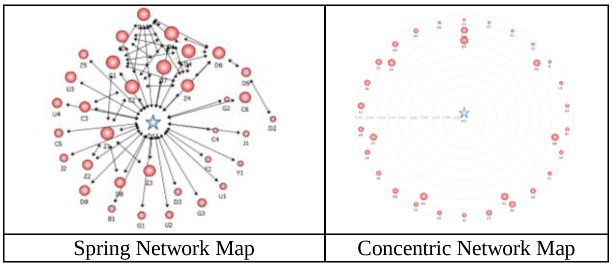


Figure 2. The result of social network analysis

A shown in Fig. 2, in the network distribution of nodes, it is possible to easily find a central node, but it is hard to make comparison according to specific node. Therefore, to overcome the limitation, this study analysed the network of D2, C2, and Z1 nodes in detail in Refinement stage.

Fig. 3 illustrates the visualized network maps of three top nodes (D7, C2, Z1) with high degree centrality

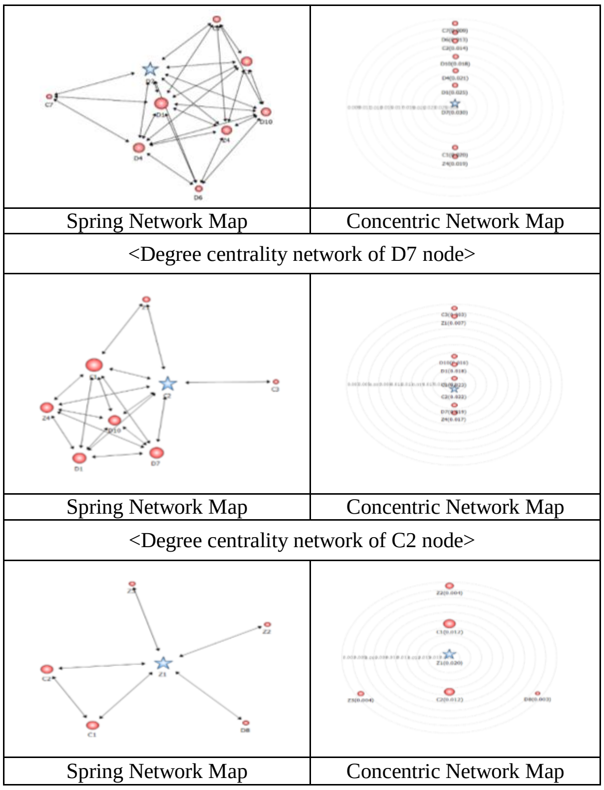


Figure 3. Visualization of degree centrality network of D7, C2, and Z1

The network analysis of the top three nodes comes to the following results.

First, the degree centrality value (0.03) of D7 node was similar to that of each D1 (0.025), D4 (0.021), and D10 (0.018) node. As a result, it was found that tourists visit associated tourist attractions in the same region.

Secondly, by drawing the degree centrality of each tourist attraction, it was possible to conjecture a core tourist attraction in a specific region. The node located at the centre of concentric network map in Fig. 3 serves as the most key role. For instance, in the degree centrality network of C2 node, the node has links with all of its neighbouring nodes, and thus it can be conjectured carefully that tourists visit the neighbouring tourist attractions of C2 node.

Thirdly, by comparing the number of links that the central node in each network has, it is possible to infer the tourist attractions vulnerable as package ones. In the case of Z1 node, it has relatively small links, which implies that the tourist attraction (Z1 node) has a smaller number of associated tourist attractions than others.

## V. Conclusion

We derived visit-associated tourist attractions using Association Analysis and TNA techniques, and then visualized them in network structure. We conducted an experimental analysis with around 42,000 blog articles. The existing method of recommending associated tourist attractions is based on neighbouring distance of tourist attractions, whereas the recommendation method proposed in this paper used the blog data reflecting tourists’ experience and comments to draw the tourist attractions actually visited by tourists. Therefore, it is expected that the proposed method would be used as a fundamental material for package tour activation policy and that since it intuitively provides the information on associated tourist attractions for users through network visualization, it is possible to offer the service of recommending associated tourist attractions.

However, this study has the limitation in the point that it failed to analyse visit route patterns of the associated tourist attractions drawn. In addition, since the tourist attractions commonly mentioned in blogs are defined as associated tourist attractions, it is hard to find the common characteristics of tourist attractions.

To overcome the limitations, it would be necessary to use the data about a floating population to analyse the patterns of visits to tourist attractions and perform the analysis on all keywords appearing in blogs to analyse relevant keywords of associated tourist attractions.

## Acknowledgment

This research was supported by the MSIP (The Ministry of Science, ICT and Future Planning), Korea, under the "SW master's course of a hiring contract" support program (NIPA-2014-HB301-14-1011)
and ITRC(Information Technology Research Center) support program (NIPA-2014-H0301-14-1022) supervised by the NIPA(National IT Industry Promotion Agency) and by Chungcheongbukdo Provincial Government Republic of Korea.

This research was financially supported by the Ministry of Education (MOE) and National Research Foundation of Korea (NRF) through the Human Resource Training Project for Regional Innovation

## References

[1] Korea Tourism Organization, Project Analysis for Big Data Applications of Tourism Business Performance, Big Data based Analysis of the Achievements of Test Tourism Project-2013 Focusing on Culture and Tour Festivals, South Korea, 2013.

[2] I. D. Cho and N. K. Kim, “Recommending core and connecting keywords of research area using social network and data mining

techniques," *Journal of Intelligence and Information Systems*, no. 17, no. 1, pp. 127-138, 2011.
[3] Y. Kim, "A study on design and implementation of personalized information recommendation system based on Apriori algorithm," *Journal of the Korean Biblia Society for Library and Information Science*, vol. 23, no. 4, pp. 283-308, 2012.
[4] D. Paranyushkin, *Text Network Analysis*, Nodus Labs, 2010, Apirl, France.
[5] Y. H. Kim, Social Network Analysis, Park Young Sa, 2003.
[6] Naver. [Online]. Available: http://www.naver.com
[7] Tour API. [Online]. Available: http://api.visitkorea.or.kr

**Kaaen Kwon** received her B.S. degrees at the Department of Information Mathematics and Industrial Engineering from Korea University, Korea in 2013. She is in Master degree at the Department of Business Data Convergence from Chungbuk National University in Korea. Her research interest include bigdata, data mining and social data analysis.

**Cho-Ah She** received her B.S. degrees at the Department of Sociology and Politics from Chungbuk National University, Korea in 2012. She is in Master degree at the Department of Business Data Convergence from Chungbuk National University in Korea. Her research interest include bigdata, Database and social data analysis.

**Wan-Sup Cho** received his M.S. and Ph.D. degrees at the Department of Computer Science from KAIST, Korea, in 1987 and 1996, respectively. He is a professor at the Department of Management Information System, Chungbuk National University in Korea. His research interests include database, bigdata, business intelligence and ERP.

**Kwan-Hee Yoo** is a professor working for the Department of Software Engineering at Chungbuk National University, Korea. He received his BS in computer science from Chungbuk National University, Korea, in 1985, and his MS and PhD degrees in computer science from Korea Advanced Institute of Science and Technology, Korea, in 1988 and 1995, respectively. His research interests include computer graphics, integral imaging system, dental/medical systems, and smart learning.

**Ga-Ae Ryu** received her B.S degrees at the Department of Cyber Security Police from Gwangju University, Korea in 2014. She is in Master degree at the Department of Digital Informatics and Convergence from Chungbuk University in Korea. Her research interest include computer graphics, computer security and smart learning.