Analysis of Research status based on Citation Context

Seon-heui Choi*, Byungkyu Kim**
*Department of Domestic Information, Korea Institute of Science and Technology Information
E-mail : sunny.choi@kisti.re.kr*, yourovin@kisti.re.kr**

1. Introduction

Citation analysis, which utilizes the relations among citations, is the most popular method of bibliometric methods. This analysis is based on 1) the evaluation by paper, journal and researcher of research output, 2) the identification of emerging research topics, 3) the production of the map for the intellectual structure analysis by research domain and 4) various services for academic information. However, this approach has a limitation since a citation is treated very simply, even though the purposes of citation are various. To complement this problem, new approaches based on citation context have been studied. This research separates citations by citation functions and tries to analyze based on the newly classified citations. Furthermore, research on citation summarization and visualization based on both citation context and citation function of citations has been tried. On the other hand, there are very few studies related to citation context in South Korea. Therefore, more research and development for this area is needed. This study analyzed the status of research for citation context. For this, we utilized the method of the social network analysis.

2. Research status based on citation context

2.1. Analysis of author keyword

In the library and information science, studies on both purpose of citations and utilization has been accomplished. Since year 2000, an advanced researches have increased in computer science such as research group for natural language processing and text mining. The studies based on citation context can be separated as 1) the research on citation function as both identification of citation context and classification scheme design for citation function, 2) the research for the summarization of papers utilizing citation contexts 3) the research for performance improvement of information retrieval used the keywords from citation contexts and 4) the research of new service models based on citation context and citation function. We analized the articles related to the research for citation context since year 2000, and specifically selected the article which Simone Teufle published in 2006 as the key paper. This paper suggests the method for the automatic classification of citation function. Over 90 papers cited this key paper of their research. These studies usually utilize the computer based technology such as natural language processing, text mining and machine learning. Also, as a result of analysis of author information such as author occupation and research paper's source, there is a lot of research output in mainly computer science, information science and partly medical science. We produced a proximity matrix using author keywords or the manually extracted keywords in case of paper without author keywords by ourself of related papers and keywords networks based on this such as below Figure1. Netminer of Cyram inc., which is the software for the social network analysis, is used to produce matrixs and the networks for this study. Figure1.A is the matrix produced with the Pearson correlation coefficient and Figure1.B is the network of keywords based on the matrix. Through these figures, we can see how key words such as citation context, citation sentence, citation function, sentiment analysis, citation intention analysis, citation classification and citation summarization are related.

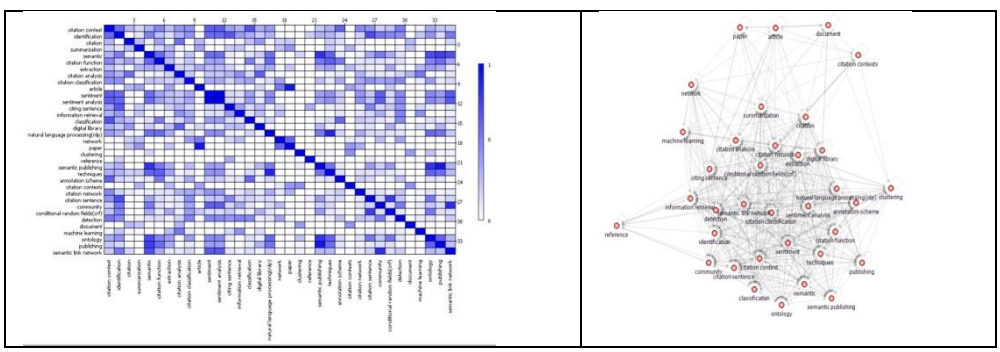

Figure 1. analysis result of author keyword (A: proximity matrix, B: keyword network)

2.2. Analysis of researchers

Based on author information such as the author's name, occupation and co-authorship, we identified the central author and groups. Figure2.A shows the co-occurance matrix and Fighre2.B is the network of co-authors produced by the matrix.

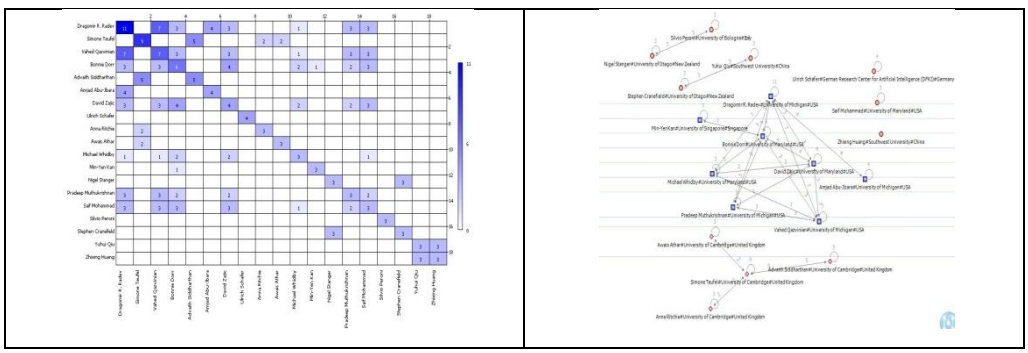

Figure 2. Process of Automatic Subject Classification

Through the analysis of co-authorship among authors, authors and institutes, which play a role mainly in this research area, was revealed such as University of Michigan(Department of EECS and School of Information: Dragomir R. Radev, Vahed Qazvinian) and University of Maryland(Human Language Technology Center of Excellence : Bonnie Dorr,David Zajic) in USA and University of Cambridge (Natural Language and Information Processing Group: Simone Teufel, Anna Ritchie) in UK. Specifically, the research partnership between University of Michigan and University of Maryland was strong as a result.

3. Conclusion

This study analyzed the status of research related citation context using the social network analysis. Through this, the central research topics and research groups in this research area was revealed. In the next study, it needs the next study for applicable researches in this research area such as citation function and citation summarization based on citation contexts using KSCD(Korean Science Citation Database) which is representative the citation database in Korea.

4. References

[1] Nanba, H., Kando, N., Okumura, M.: Classification of Research Papers Using Citation Links and Citation Types: Towards Automatic Review Article Generation. In: Proceedings of the 11th SIG Classification Research Workshop, pp. 117–134 (2000)
[2] Teufel, Simone, Advaith Siddharthan, and Dan Tidhar. "Automatic classification of citation function." Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, pp. 103–110 (2006)
[3] KSCI: Korea Science Citation Index (http://ksci.kisti.re.kr)