

과학외교를 위한 데이터기반의 연구주제선정 방법

Data-Driven Approach to Identify Research Topics for Science and Technology Diplomacy

여운동, 김선호, 이방래, 노경란
한국과학기술정보연구원 RnD투자분석센터

Woon-Dong Yeo(wdyeo@kisti.re.kr), Seonho Kim(haebang@kisti.re.kr)
BangRae Lee(brlee@kisti.re.kr), Kyung-Ran Noh(infor@kisti.re.kr)

요약

두 국가가 본격적으로 외교적 협약을 진행하기 전 우호적인 분위기를 만들기 위해서나, 국가간 정치적 우호 관계를 지속하기 위한 목적 등으로 과학외교를 사용한다. 최근에는 과학기술이 국가 발전에 미치는 영향이 커짐에 따라서 과학외교에 대한 관심이 더욱 집중되고 있다. 과학외교를 수행하기 위해 두 국가가 서로 흥미를 가질 수 있는 협동연구주제를 찾는 것은 전문가 집단에 의해 추천에 의해 이뤄진다. 그러나 이 방법은 전문가의 주관적 판단에 의지하기 때문에 편향성과 이에 따른 문제가 존재한다. 개인적 및 조직적 편향, 유명한 연구자의 후광효과, 전문가마다 다른 추천기준 등이 있을 수 있다. 본 논문에서는 전문가 기반의 방식이 가지는 문제점을 극복하기 위해 한국에서 시도된 빅데이터 기반의 외교를 위한 연구주제 추천방법을 소개한다. 빅데이터를 분석하기 위한 알고리즘은 전통적인 연구분야인 계량서지학 뿐만 아니라 최신 딥러닝 기술을 사용한다. 제안된 방식은 한국과 헝가리 간의 과학외교에 사용되었으며, 데이터기반 주제선정 방식의 가능성을 확인할 수 있었다.

■ 중심어 : | 과학계량학 | 추천시스템 | 계량서지학 | 딥러닝 | 과학외교 |

Abstract

In science and technology diplomacy, major countries actively utilize their capabilities in science and technology for public diplomacy, especially for promoting diplomatic relations with politically sensitive regions and countries. Recently, with an increase in the influence of science and technology on national development, interest in science and technology diplomacy has increased. So far, science and technology diplomacy has relied on experts to find research topics that are of common interest to both the countries. However, this method has various problems such as the bias arising from the subjective judgment of experts, the attribution of the halo effect to famous researchers, and the use of different criteria for different experts. This paper presents an objective data-based approach to identify and recommend research topics to support science and technology diplomacy without relying on the expert-based approach. The proposed approach is based on big data analysis that uses deep-learning techniques and bibliometric methods. The Scopus database is used to find proper topics for collaborative research between two countries. This approach has been used to support science and technology diplomacy between Korea and Hungary and has raised expectations of policy makers. This paper finally discusses aspects that should be focused on to improve the system in the future.

■ keyword : | Scientometrics | Recommender System | Bibliometrics | Deep Learning | Science for Diplomacy |

I. Introduction

The use of science and technology to achieve international diplomatic goals is called science and technology diplomacy. It is generally used to create a positive atmosphere between two countries before they enter into a main diplomatic agreement or decide to maintain long-term political relations[1]. In science and technology diplomacy, major countries actively utilize their capabilities in science and technology for public diplomacy, especially for promoting diplomatic relations with politically sensitive regions and countries. Small- and medium-sized countries use science and technology diplomacy as a means of raising their potential in science and technology to international standards[2]. Recently, with an increase in the influence of science and technology on national development, interest in science and technology diplomacy has increased. To promote science and technology diplomacy, cooperation is required between the government departments involved in diplomacy and science and technology. Such cooperation includes the identification of the strengths and weaknesses of the partner countries and the exploration of international joint research topics (RTs) in close relationship with the scientific community[1].

So far, RTs for science and technology diplomacy have been selected by a bottom-up approach driven by scientists and researchers or a top-down approach defined by government agencies in pursuit of top-down strategic priorities[3][4]. Both these approaches are qualitative selection methods. The approach involving the selection of RTs by scientists can have similar problems to the approach

involving expert-driven research evaluation. In the case of the bottom-up approach, a lack of purity in the intent of the researchers participating in science diplomacy may lead to a lack of objectivity and credibility in the RT selection process. The problems include the bias arising from the subjective judgment of scientists, the attribution of the halo effect to famous researchers, and the use of different criteria for different experts[5][6]. It may be difficult for an expert in a particular field to judge the suitability of an interdisciplinary topic beyond his or her domain or compare topics from different fields. The process of selecting an RT requires knowledge of a wide range of subjects, time, and effort. Moreover, unlike scientists, who are expected to be devoted to their country, researchers many participate in science diplomacy programs with the primary purpose of securing financial aid and enhancing their reputation[7]. If there is a lack of purity in the intent of the scientists participating in science diplomacy according to the bottom-up approach, the objectivity and credibility of the RT selection process may be questionable.

In the case of a top-down approach to science diplomacy, there may be complaints arising from scientists being asked to conduct collaborative research. According to a study conducted in Germany, scientists believe that there is no transparency in the decisions of relevant government agencies and that these decisions are difficult to understand and are too self-oriented[7].

One solution to the problems of the bottom-up and top-down approaches is to use a data-driven approach to identify potential RTs. The data-driven approach involves making

decisions based on data analysis results rather than the judgment of scientists, researchers, and policy makers. Although data-driven approaches are being used in international collaborative research, most of them focus on finding suitable researchers who can collaborate[8-11]. Some prior data-based studies on finding collaborative RTs used a method of recommending similar RTs after learning past collaborative research patterns. This is a machine-learning method that predicts the future by learning the past. This method assumes that the joint RTs selected in the past were desirable. However, it is not appropriate to use past cases as training data because the selection results based on the past cases may contain the abovementioned problems. There is no evidence that past choices are the best. In addition, unlike general science and technology cooperation, science diplomacy is carried out for promoting mutual reciprocal relations with foreign countries, so it mainly uses science for easily achieving the economic and political objectives of countries [12]. Because the economic and political objectives of countries can change from time to time, it is not appropriate to make selections by learning past patterns. Hence, finding a solution to this problem is the motivation of this study.

Science diplomacy is broadly defined as a form of cooperation among countries or regions to solve complex problems through scientific research. The Royal Society and the American Association for the Advancement of Science (publisher of the magazine *Science & Diplomacy*) divide science diplomacy into three types: science in diplomacy, science for diplomacy, and diplomacy for science. This paper is associated with science for diplomacy

[13]. Science for diplomacy involves the use of science to help build and improve international relations, especially in cases where there may be strain or tension in the official relationship. Science for diplomacy primarily draws on the “soft power” of science, i.e., its attractiveness and influence both as a national asset and as a universal activity that transcends national or partisan interests.

This paper suggests a new data-driven method to overcome the previously mentioned problems of the top-down and bottom-up approaches. This method of selecting RTs uses big data and has recently been tried as part of the science for diplomacy initiatives in Korea. Algorithms for analyzing big data use deep-learning techniques as well as bibliometric methods, which are quantitative data measurement methods. The present study makes the following two contributions. First, the paper suggests a new and practical data-driven approach to recommend RTs for science and technology diplomacy. This approach has actually been used in the science for diplomacy program between Korea and Hungary and has raised expectations of policy makers. The second contribution is that the paper suggests a new method to find technologies based on relationships with hypernyms. We call this method as the “usability method” because it explores the different usages of a technology. The usability method helps science and technology nonspecialists to understand a specific technology more easily. It is not a new algorithm but a new application of the well-known word2vec model[14].

This paper is organized as follows. Section II outlines the techniques used in the data-based

approach. A detailed description of each technique is beyond the scope of this paper and is therefore omitted. Section III describes the principle of operation of the implemented system, that is, an RT recommendation algorithm. Section 4 presents the RTs recommended by the system for science diplomacy between Korea and Hungary. Finally, Section V summarizes the study, describes problems that have been found in practical applications, and outlines future directions of development.

II. Technical background

This study uses bibliometrics and deep learning to identify RTs based on data for science and technology diplomacy. Bibliometrics is used to find RT candidates, and deep learning is used to select the final RTs considering the relationship between the target country and candidate technology (accessibility, growth rate, and usability). In this paper, a new combination of deep-learning algorithms is suggested to measure the usability of a technology.

Number of papers, number of citations, and citations per paper

Bibliometrics is a data-based statistical evaluation method that is commonly used for measuring and evaluating science and technology research[15]. When bibliometrics is used for assessing the level of technology, it is often called as scientometrics. Bibliometrics mainly evaluates science and technology research using the bibliographic data of

publications. The number of papers, number of citations, and citations per paper are the most basic and important indicators used by bibliometrics. The existence of a large number of papers indicates that the topic is being actively researched by many researchers. The fact that there are a large number of citations to a specific topic means that qualitatively, the research result of this topic has more impact than that of other topics. However, as the number of papers increases, the sum of the number of citations received by other papers is more likely to be greater. This problem can be alleviated by using the parameter called citations per paper, which is a value that is normalized by the number of papers.

Activity and attractiveness

Activity and attractiveness are indicators of the number of papers and number of citations normalized to the world average, respectively [16][17]. The activity index is defined as a country's share in the world's publication output in a given field divided by the country's share in the world's publication output in all science fields. The attractiveness index is defined as a country's share in the citations attracted by publications in a given field divided by the country's share in the citations attracted by publications in all science fields. Activity and attractiveness are used to compare interdisciplinary performance using a multidimensional scaling chart, as shown in [Figure 1]. For example, the RTs located at the top right can be considered to be more active in terms of research activity and more qualitative than the other RTs. The RTs at the bottom right can be understood to be topics

that show good performance in terms of quantity but not in terms of quality. In contrast, the RTs at the upper left are poor in quantity but more impactful in quality, and those at the bottom left are poor in terms of both quantity and quality.

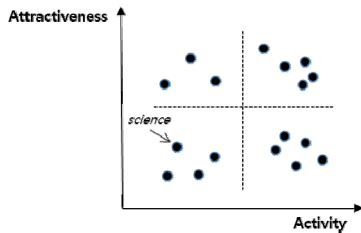


Fig. 1 Activity and Attractiveness index

Fractional citation count

Generally, a paper cites other papers that are related to its own topic. Not many papers from other topics are cited, thereby forming a dense citation network within the same topic[18]. Consequently, topics such as mathematics that display a small number of references in a paper have a smaller number of citations than topics such as biology that display a large number of references. A fractional citation count is a parameter that can be used to solve this problem. It counts the citations in inverse proportion to the reference number of the source paper. That is, the number of citations available from the source article is not considered to be “1” but “1/(the reference number of the source).” In this study, we calculate the number of citations according to the fractional citation count.

Accessibility

The computer science technology that has

been the most popular recently is deep learning. It is a machine-learning algorithm and is used as a core algorithm for realizing artificial intelligence because it shows considerably better performance than other machine-learning algorithms. One of the deep-learning models is word2vec[14], which is a representative algorithm that shows excellent performance in word embedding and converts each word into a unique number. This model maps all the words we use to a multidimensional vector space. An extension of word2vec to the document level is an algorithm called doc2vec[19]. This algorithm is designed to map documents and words to the same vector space. When words are input to word2vec, the document that includes the words is trained simultaneously. The underlying principle of doc2vec is identical to that of word2vec. In addition, item2vec, which is a modified algorithm of word2vec, is used to analyze a shopping cart[20]. Note that word2vec uses the position of the words that occur in sentences, whereas item2vec calculates the relationship of the items in a shopping cart.

Once a word has an absolute position in space, the similarity between two words can be measured. This principle is applied to word2vec, doc2vec, and item2vec. In this study, the author keyword of a research paper is regarded as the commodity contained in the shopping cart (item2vec), and learning is performed by recognizing the country of the first author of the paper as a document (doc2vec). Accessibility is achieved by measuring the similarity between the vectors of a country and keyword.

Usability

By using word2vec’s word embedding and performing a vector operation, we can infer a new word that has the same vector relation that we are interested in [Figure 2]. A well-known example of such an inference is finding the word "Queen" using the vector operations of King, Queen, Man, and Woman ("King" - "Man" + "Woman" = ?) [14]. This study uses vector-oriented reasoning to select RTs with high usability. The vector operation is <"Automobile" - "Diesel engine" + "Specific technology" = ?>. That is, this study uses the relationship between a diesel engine and an automobile to identify RTs with higher usability.

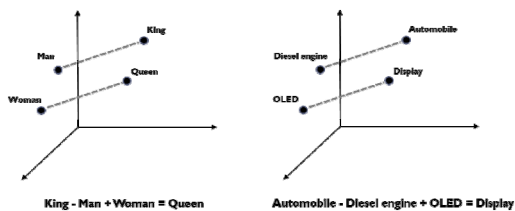


Fig. 2 Vector operation using word2vec’s word embedding

III. Proposed Method

The proposed method consists of two stages. The first stage identifies candidate RTs using bibliometric data such as the author keywords, publication year, number of citations, and author’s country from a scientific database. The second stage measures the accessibility, growth rate, and usability and determines the final RTs for diplomacy.

Stage 1. Identification of candidate RTs

Science and technology diplomacy is

expected to be a form of cooperative research between two countries. In some fields of science and technology, one country may be stronger than other countries. In this case, the delivery of technical assistance may be required from one country to another. The matching criteria for the candidate RTs depend on the field of science and technology. [Table 1] and [Figure 3] show the matching criteria and identification method for the candidate RTs used for diplomacy in relation to the activity and attractiveness index.

For the fields of technology for which the quality level is low in Korea, the research level of the target country is hoped to be similar to or higher than that of Korea. In this regard, the target country should be higher than Korea in terms of at least one of the following three indicators: the number of papers, number of citations, and citations per paper. A disjunctive strategy [21] is used to increase the recall of the algorithm. The three indicators are likely to be influenced by non-research factors such as the population and language of a country. In particular, because the number of references varies according to the topic, there is a difference in the number of citations on average[18]. In addition, the RTs with similar levels are ideal subjects in terms of mutual reciprocity, so there is no reason to exclude them.

Not all RTs need to be candidates for diplomacy just because the target country’s level of research is higher than that of Korea. When diplomatic RTs are to be selected for two very developed countries, more stringent criteria should be applied. Even if the target country has a better research performance on most RTs, the topics that are not of interest to

Korea will not be suitable. The topics that are actively researched in Korea and are of at least some interest in the target country will be suitable. In terms of the quality of the research results, it is highly likely that the topics that perform well in the target country and are not excellent in Korea will help the development of science and technology in Korea. Here, the activity is used to assess the amount of interest in the two countries, and the attractiveness is used to determine the qualitative performance of the research related to the topic. Consequently, a conjunctive strategy [21] is used, which indicates that the desirable topics are the ones for which the activity is very high in the target country and somewhat high in Korea and the attractiveness is very high in the target country but low in Korea. The bold lines in [Figure 3] correspond to the topics, and the default values are empirical values, which are 20% and 50% of the critical points.

For the research fields in which Korea has technical advantages, Korea looks for RTs that can be helpful to the target country. In such cases, Korea becomes a helping country. Accordingly, the algorithm is reversed. At least one of the three indicators (number of papers, number of citations, and citations per paper) should be higher in Korea than in the other country. The activity index should be very high in the target country. The activity index and attractiveness index differ from the previous case. In the RT selection for diplomacy, the needs of the target country with regard to technical support are of considerable importance. The activity index in Korea does not have to be very high if it is high in the target country. The attractiveness index is not a major indicator in this case because

attractiveness is only a relative comparison between domestic technologies and it can be far below the world level. In [Figure 3], the dashed lines correspond to these RTs, and the default values are 20% and 50% of the critical points.

Table 1. Index criteria for candidate topic selection

| Index | For research field in which Korea has technical disadvantages | For research field in which Korea has technical advantages |
|----------------------------------------------------------------|--------------------------------------------------------------------------|---------------------------------------------------------------------------|
| Number of papers, number of citations, and citations per paper | At least one index in target country should be higher than that in Korea | At least one index in Korea should be higher than that in target country. |
| Activity | Very high in target country and somewhat high in Korea | Very high in target country and somewhat high in Korea |
| Attractiveness | Very high in target country and slightly low in Korea | - |

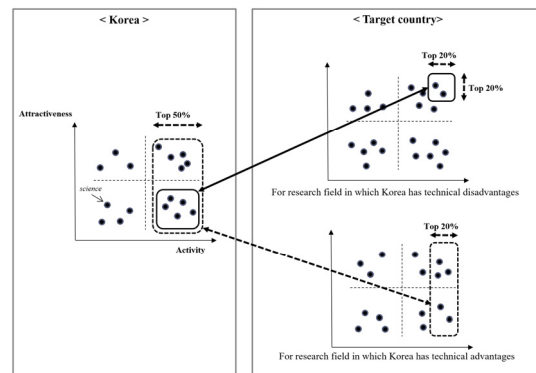


Fig. 3 Identification method for candidate RTs used for diplomacy

Stage 2. Determination of RTs

The candidate topics identified in stage 1 do not have any particular priority. In particular, there can be dozens of candidate topics depending on the threshold setting of various index values. Hence, it is important to determine the priority among the candidate topics and recommend topics with higher

rankings to help policy makers reduce the time required to select the final RTs for diplomacy. In this stage, the values of accessibility and growth rate are calculated for each topic to prioritize the RTs [Figure 4]. The values are used as a tool for decision making. Thus, the values need not have thresholds, and the tool can be used just by sorting topics in terms of the accessibility or growth rate. RTs with high accessibility are easily accessible based on existing research experience in the countries. To gain competitiveness for a specific RT, it is often necessary to acquire research capability in both the periphery and core of the topic. This is because a technology is not used independently by itself. In addition, it would be preferable to choose highly accessible topics to achieve rapid visibility in diplomatic support. Topics with a high growth rate have recently attracted attention in the scholarly community. Papers on these topics have been published in related fields, and these topics are fast growing. In recent years, disruptive technologies that have surpassed the performances of previous technologies are emerging, so it is desirable to seek cooperation with other countries for utilizing the emerging technologies. The fast-growing topics correspond to the emerging technologies. To calculate the growth rate, the compound annual growth rate (CAGR) formula is used. The linear regression slope can be used as an alternative to the CAGR because it shows robust performance against various data types [22].

Usability can be obtained by performing a vector operation based on the word embedding from word2vec. The topic recommendation based on author keywords has an advantage in that various topics can be identified. However,

it has a disadvantage in that the author keywords consist of very detailed technical terminologies, so it is difficult to understand the concept, function, and usages of technologies. Usability information is provided to compensate for this disadvantage. The accessibility and growth rate are numeric, whereas the usability lists author keywords. For example, the usability for “RFID” is represented by “radio frequency identification,” “bluetooth,” “nfc,” “lot,” and “wireless.” Usability is used to understand the usages, but it is also used to help nonexperts such as policy makers to understand the topic.

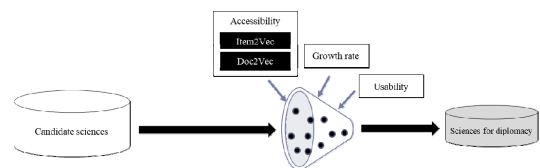


Fig. 4 Determination of RTs for diplomacy

IV. Case Study

Korea and Hungary have strengthened their cooperative relations by focusing on science and technology and the economy since they established formal diplomatic relations in 1989. In March 2018, Korea and Hungary agreed to enhance cooperation in the field of high technology. Korea occupies a higher position than Hungary in engineering fields such as electronics and information and communication technology. We identified RTs from different fields of engineering and used matching criteria and the proposed method for the research fields in which Korea has technical advantages.

To identify potential RTs for science and technology diplomacy between Korea and

Hungary, we extracted author keywords from the Scopus database, which has been widely used by worldwide researchers since its launch in 2004. The keywords were limited to those included in the papers published in 20 categories related to science and technology within the last 5 years. Accordingly, we retrieved 8,401 author keywords from papers published between 2012 and 2016. [Figure 5] shows the process of data extraction. An author keyword can be replaced with an indexer keyword. However, our experimental results showed that the author keyword is more appropriate for topic diversity. The indexer keyword has recently been mainly extracted automatically by natural language processing, and it contains a lot of noise and has the tendency to select words with high frequency as keywords.

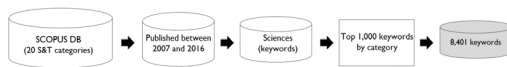


Fig. 5 Data extraction from Scopus

[Figure 6] represents the system implemented using the proposed method. The system has a text field to set the minimum number of papers per keyword because relatively big science is preferred in diplomacy. For the convenience of users, the minimum value of activity and attractiveness can be set to the same value for each country and the target country. In the present case, the minimum number of related papers is 1500 and the minimum value of activity and attractiveness is 50 for Korea and 20 for Hungary.

[Table 2] shows results in the Scopus engineering field. It is sorted by the growth rate. The first recommended RT is UAV (unmanned aerial vehicle), followed by

MICROFLUIDICS, ELECTRIC FIELD, DELAMINATION, etc. After discussion between the two countries with regard to the recommended topics, UAV was selected as the main RT in the field of engineering. The accessibility of UAV in Hungary is relatively low; that is, Hungary does not have many research results for UAV. This means that Hungary may find it difficult to achieve the technical goals for UAV based on research experience, but the reward for success will be greater than that when the accessibility is high. The agreement between Korea and Hungary was the first case in which a data-driven approach was used for science and technology diplomacy with a foreign country by Korea.

Table 2. Results in engineering

| Rank | Science | Usability | Accessibility | Growth Rate |
|------|---------------------|----------------------------------------------------------------------|---------------|-------------|
| 1 | UAE | unmanned aerial vehicle, uas, multirotor | 1.418 | 12.170 |
| 2 | MICROFLUIDICS | microfluidic, lab on a chip, dielectrophoresis | 2.128 | 5.839 |
| 3 | ELECTRIC FIELD | electric field strength, electric field distribution, electric field | 7.801 | 2.535 |
| 4 | DELAMINATION | laminar, cfrp, debonding | 9.220 | 2.039 |
| 5 | FLEXIBILITY | agility, planning, manufacturing | 2.837 | 2.002 |
| 6 | EVAPORATION | wick, thin film evaporation, condensation | 9.929 | 0.258 |
| 7 | PERMEABILITY | porosity, sandstone, tortuosity | 8.511 | -0.205 |
| 8 | SURFACE ROUGHNESS | surface finish, roughness, machining | 4.965 | -1.220 |
| 9 | ROBUST CONTROL | nonlinear control, disturbance observer, tracking control | 5.674 | -1.508 |
| 10 | MICROWAVE | microwaves, ultrasonic, microwave heating | 10.683 | -1.577 |
| 11 | IMAGE ANALYSIS | image processing, imagej, segmentation | 0.709 | -2.865 |
| 12 | ENERGY EFFICIENCY | energy consumption, energy saving, energy efficient | 3.546 | -3.581 |
| 13 | COMPUTER SIMULATION | mathematical models, calculations, flowcharting | 7.092 | -4.945 |
| 14 | GPS | gnss, global positioning system, gzs | 6.383 | -5.546 |
| 15 | PROTECTION | cable, substation, protection system | 4.255 | -5.983 |

V. Discussion

As the amount of scholarly data grows and powerful analytical tools become more available, scholarly data are used more easily and widely[23]. Literature recommendation, venue recommendation, and expert recommendation are representative examples of the results of the use of publication data. Deep learning is further accelerating the use of publication data and increasing the performance of analysis.

This paper introduced a method that recommends RTs for science and technology diplomacy using scientific data. The need for such a data-driven approach was raised by a government agency in charge of science and technology diplomacy. The biggest problem faced by the agency was the limitations of the expert-based method, i.e., lack of subject diversity and limited time and personnel. The government agency saw the possibility of data-driven matchmaking through the first application of our method to science and technology diplomacy between Korea and Hungary.

Despite this success, the proposed approach has some limitations. The approach selects an excessive number of biology-related topics. This may be attributed to the fact that publication data contains a large number of papers on biology. Fractional citation counting can mitigate this problem to some extent. We can overcome this problem by recommending more topics from the field of science and technology. Another limitation is that the deep-learning algorithm employed for usability does not show perfect performance. The results of the algorithm included acronyms, full names,

and synonyms. However, these results have helped the government agency to understand RTs extensively. This paper empirically used 20% and 50% of the critical values related to activity and attractiveness. However, further studies on the imposition of these values is required. Finally yet importantly, environmental factors other than science and technology such as politics, society, and industry are not reflected in the selection of the RTs. The reflection of these environmental factors is important from the perspective of diplomatic science. The reflection of political, social, and industrial information requires the use of other types of data such as news and a more sophisticated algorithm.

In conclusion, we need to overcome the limitations of the data-driven approach through close relationships with scientific communities.

참고 문헌

- [1] P. Boekholt, J. Edler, P. Cunningham, and K. Flanagan, *European Commission: Drivers of International collaboration in research*, Luxembourg: Publications Office of the European Union, 2009.
- [2] C. Vaughan, M. Sarah, C. Daryl, S. Lloyd, G. Robert, and P. Maria, "The Emergence of Science Diplomacy," *Science Diplomacy*, pp.3-24, 2015.
- [3] H. Ceballos, J. Fangmeyer, N. Galeano, E. Juarez, and F. Cantu-Ortiz, "Impelling research productivity and impact through collaboration: A scientometric case study of knowledge management," *Knowledge Management Research and Practice*, Vol.15, No.3, pp.346-355, 2017.
- [4] The Royal Society, "New Frontiers in Science Diplomacy: Navigating the changing balance

- of power,” 2010.
- [5] D. E. Chubin and E. J. Hackett, *Peer review and the printed word*, In: Chubin DE, Hackett E.J. Peerless Science: Peer Review and U.S. Science Policy. Albany, NY: SUNY Press. 1990.
- [6] R. N. Kostoff, “Assessing research impact: US government retrospective and quantitative approaches,” *Science and Public Policy*, Vol.2, No.1, 1994.
- [7] B. FÄHNRIK, “STD: Investigating the perspective of scholars on politics-science collaboration in international affairs,” *Public Understanding of Science*, 2015.
- [8] G. R. Lopes, M. M. Moro, L. K. Wives, and J. P. M. D. Oliveira, “Collaboration recommendation on academic social networks,” in *Advances in Conceptual Modeling-Applications and Challenges*. Springer, pp.190-199, 2010.
- [9] F. Xia, Z. Chen, W. Wang, J. Li, and L. T. Yang, “Mvwalker: Random walk-based most valuable collaborators recommendation exploiting academic factors,” *Emerging Topics in Computing*, IEEE Transactions on, Vol.2, No.3, pp.364-375, 2014.
- [10] P. Chaiwanarom and C. Lursinsap, “Collaborator recommendation in interdisciplinary computer science using degrees of collaborative forces, temporal evolution of research interest, and comparative seniority status,” *Knowledge-Based Systems*, Vol.75, pp.161-172, 2015.
- [11] X. Kong, H. Jiang, Z. Yang, Z. Xu, F. Xia, and A. Tolba, “Exploiting publication contents and collaboration networks for collaborator recommendation,” *PloS ONE*, Vol.11, No.2, e0148492. 2016.
- [12] M. M. Zolfaghazadeh, A. A. Sadabadi, M. Sanaei, F. L. Toosi, and M. Hajari, “Science and technology diplomacy: a framework at the national level,” *Journal of Science and Technology Policy Management*, Vol.8, No.2, pp.98-128, 2017.
- [13] L. M. Frehill and K. Seely-Gant, “International Research Collaborations: Scientists Speak about Leveraging Science for Diplomacy,” *Science & Diplomacy*, Vol.5, No.3, 2016. [Online] Available: <https://www.sciencediplomacy.org/article/2016/international-research-collaborations>
- [14] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean, Efficient Estimation of Word Representations in Vector Space, In *ICLR Workshop Papers*, 2013.
- [15] OECD, *Bibliometrics*, OECD Glossary of Statistical Terms, 2015.
- [16] J. D. FRAME, “Mainstream research in Latin America and the Caribbean,” *Interciencia*, Vol.2, No.143, pp.143-148, 1977.
- [17] A. Schubert and T. Braun, “Relative indicators and relational charts for comparative assessment of publication output and citation impact,” *Scientometrics*, Vol.9, 1986.
- [18] F. Radicchi and C. Castellano, “Testing the fairness of citation indicators for comparison across scientific domains: the case of fractional citation counts,” *J Informetr*, Vol.6, No.1, pp.121-130, 2012.
- [19] Q. Le and T. Mikolov, *Distributed Representations of Sentences and Documents*, In *Proceedings of ICML 2014*.
- [20] O. Barkan and N. Koenigstein, Item2vec: neural item embedding for collaborative filtering. In *MLSP Workshop*, 2016.
- [21] I. Linkov, A. Varghese, S. Jamil, T. P. Seager, G. Kiker, and T. Bridges, *Multi-criteria decision analysis: a framework for structuring remedial decisions at contaminated sites*, Comparative risk assessment and environmental decision making, 15-54, 2004.
- [22] Y. H. Tseng, Y. I. Lin, Y. Y. Lee, W. C. Hung, and C. H. Lee, “A comparison of methods for detecting hot topics,” *Scientometrics*, Vol.8, No.1, pp.73-90, 2009.
- [23] F. Xia, W. Wang, T. M. Bekele, and H. Liu, “Big Scholarly Data: A Survey,” *IEEE Transactions on Big Data*, Vol.3, pp.18-35, 2017.

저 자 소 개

여 운 동(Woon-Dong Yeo)

정회원



- 2014 : 고려대학교 컴퓨터학과 (Ph.D.)
- 2002~현재 : 한국과학기술정보연구원 책임연구원

〈관심분야〉 : 데이터마이닝, 과학계량학

김 선 호(Seonho Kim)

정회원



- 2008년 : 버지니아공대 전산학과 (Ph.D.)
- 1997년 ~ 현재 : 한국과학기술정보연구원 책임연구원

〈관심분야〉 : 기계학습, 자연언어처리, 인공지능, HCI

이 방 래(BangRae Lee)

정회원



- 2000년 : 부산대학교 전기공학과(공학사)
- 2002년 : 한국과학기술원 전기및전자공학과(공학석사)
- 2017년 : 서울시립대학교 전자전기컴퓨터공학과(공학박사)
- 2002년 ~ 현재 : 한국과학기술정보연구원 책임연구원

〈관심분야〉 : 데이터마이닝, 과학기술정보분석

노 경 란(Kyung-Ran Noh)

정회원



- 2006년 : 연세대학교 대학원 문헌정보학과 (Ph.D.)
- 1994년 ~ 현재 : 한국과학기술정보연구원 책임연구원

〈관심분야〉 : 과학계량학, 변혁적 연구영역 탐지