The trend study of academic articles on industry-academic cooperation using text mining analysis

Sun-Hee Kwon
Department of English, Busan University of Foreign Studies

텍스트 마이닝 분석을 이용한 국내 산학협력에 대한 학술 논문 동향 연구

권선희
부산외국어대학교 영어학부

Abstract This paper aims to perform a trend study of academic articles related to industry-academic cooperation in South Korea using text mining analysis. Five hundred thirty-one Korean articles on industry-academic cooperation (IAC), published in the domestic academic database (KCI) over the last ten years (2011-2021), were collected to perform the study. In particular, articles on one of the national projects for industry-university cooperation (LINC project) were separately examined to examine how industry-academic cooperation had been processed in universities. To identify the features of the collected full texts, a text mining analysis, including keyword network analysis and topic modeling analysis, was conducted. The main results of the analysis were that the top words for both IAC and the LINC articles were 'university', 'cooperation', 'industry-academic', and 'education' for TF; and 'start-up' and 'employment' for TF-IDF. The results also show that 'area' for IAC articles and 'start-up' for LINC articles were at the center of the keyword network. Lastly, it turned out that the IAC articles discussed various issues such as school, field, start-up, innovation, and market creation, while the LINC articles dealt mainly with a start-up. Based on the results of this study, this paper suggests important functions of the university as a major space for IAC, and a positive function and a developmental direction for university-business linkage.

요 약 본 논문은 텍스트 마이닝 분석을 통해 국내 산학협력 관련 학술 논문 연구에 목적이 있다. 연구 대상인 국내 산학협력 학술 논문은 대학 산학협력, 그리고 산학협력 국가재정지원사업인 사회맞춤형선도대학 육성사업(LINC)에 대해 10년간(2011~2021) 학술 데이터베이스(KCI)에 발표된 총 531개의 논문의 한글 초록이다. 수집된 전체 텍스트의 특색을 파악하기 위해 키워드 네트워크 분석과 토픽 모델링 분석을 포함한 텍스트 마이닝 분석을 실시하였다. 주요 결과는 산학협력논문과 LINC논문에 공통으로 나타난 빈도 상위 어휘는 TF기준으로 '대학', '협력', '산학', '교육'이, TF-IDF 기준으로 '창업'과 '고용'이 공통으로 발생하였다. 키워드 네트워크의 경우, 산학협력논문의 경우 '지역'이, LINC 논문의 경우 '창업'이 키워드 네트워크의 중심에 있는 것으로 나타났다. 또한 산학협력논문은 학교, 경영, 현장, 창업, 혁신, 시장 등 다양한 이슈에 대한 논의가 이루어진 반면, LINC논문은 창업이 주된 논의 분야라는 것을 알 수 있었다. 본 연구 결과를 기반으로 본 논문은 산학협력의 주요 공간으로서의 대학의 기능, 그리고 대학-기업 연계의 긍정적 효과와 발전적 방향에 대한 제언을 제시하고자 한다.

Keywords : Industry-Academic Cooperation, LINC Project, Academic Articles, Text Mining Analysis, Topic Modelling Analysis

1. Introduction

Industry-academic cooperation refers to all efforts by industry and academia to work together to pursue mutual interests. Industry-academic cooperation in South Korea is closely related to industry and government policies. In the case of government policies and support, they were centered on vocational high schools and two-year colleges before the 2000s and became targeted to four-year universities in the 2000s[1,2]. Since 2004, industry-academic cooperation activities in the universities have been actively conducted under the initiatives of relevant government ministries (Ministry of Employment and Labor, Ministry of Education, Ministry of Science and ICT, Ministry of Trade, Industry and Energy, and Ministry of SMEs and Startups) through projects such as 'industry-academic Cooperation Centered University Fostering Project (2004-2011)', 'Regional Base Research Center Project (2009-2011)', and 'Cultivation of Human Resources for Leading Industries in the Economic Area Projects (2009-2011)' and so on[3,4]. The project of 'Leaders in industry-academic cooperation (LINC)' is one of such government-led industrial-educational cooperation projects, started in 2011 and ongoing to date. The LINC project aims to strengthen national competitiveness, such as expanding youth employment and business start-ups and supporting innovation in small and medium-sized enterprises, and more importantly, it aims to resolve the difficulties in finding employment for students and job openings in companies by inducing changes to a customized university curriculum that reflects industrial demand[5]. Therefore, in order to promote the smooth transition from school to field, corporate participation in educational activities is set as a major indicator of business performance in the LINC project[6].

This study aims to examine the trends and characteristics of industry-academic cooperation in general, and in particular, to take a close look at the LINC project to see how the cooperation between the two has been going in the universities. Although the importance of industry-academic cooperation has been on the rise for a relatively long time, there is not much academic research on industry-academic cooperation[7,8]. In particular, recent research on industry-academic cooperation at universities has been very limited. Through text mining analysis, the general trend of industry-academic cooperation over the past 10 years would be examined through comparative analysis of academic articles on both industry-academic cooperation and LINC project. Based on the results, the study aims to provide suggestions and implications for industry-academic cooperation programs.

2. Research Background

Industry-academic cooperation (IAC) can be divided into several types depending on the area of cooperation. By area, there is 'education' section where students and professors participate in educational program that industries and schools run together. Next, there is a 'research' section in which industries and universities jointly pay for manpower, facilities, and expenses, and technology development is supported by universities and research institutes. The third section is 'start-up and technology transfer,' which includes support for start-ups using university human and material resources, industrial property rights such as technologies and patents, and commercialization and transfer of intellectual property rights[1,2].

Relevant government ministries that have initiated IAC programs have well recognized these three areas, 'education, research, and start-up and technology transfer,' as major areas

of activities, and have planned and supported various IAC projects in the university level. In particular, as the need for industry-academic cooperation for technological innovation increased significantly from 2010 to 2011, advanced policies for industry-academic cooperation were introduced by promoting "Industry-Academic Cooperation Advancement Plan"(2010) and "Industry-Academic Cooperation Project"(2011)[4,6]. Through such plans, various government-supported IAC projects started to get carried out by establishing and promoting implementation plans by related ministries and local governments to foster talent (education), transfer and commercialize technology, start-ups, and build research infrastructure. To examine the extent to what government-supported IAC projects were effective, a number of literature studies were conducted: For example, the correlation between the objective performance (student employment rate, number of technology transfers, etc.) and subjective performance (satisfaction survey) of IAC projects was reviewed, or various institutional and management factors were set as independent variables and analyzed their impact on each performance index [8]. In addition, the previous studies examined the status of indicators related to the personnel system (professor status, etc.), infrastructure-related indicators (number of family companies, joint equipment operating and profit, technology commercialization performance, etc.)[2].

In a similar vein, previous studies on LINC project had examined major performance indicators. For example, a result of analyzing the influencing factors on the performance of industry-academic cooperation through multiple regression analysis, it was reported that the ratio of students who completed capstone design and the input budget for the LINC project had a significant effect on the student employment rate, and the budget for the LINC project had the most significant effect on the number of industry

joint research projects per professor, research expenses, and profit from common equipment operation. Also, the effectiveness of the LINC project was reviewed through the student employment rate, the rate of on-the-job training, the number of manpower in specialized fields, and it was found that the universities that performed the LINC project showed significant increase in those ratios and the numbers after the projects[5,7].

As noted above, prior studies are meaningful in that they present reliable quantitative figures on the status, major factors, and performance level of the IAC projects. However, most of the previous studies are related to quantitative indicators, not to mention there are very few attempts of exploratory studies to see what characteristics the IAC or LINC project has had and what the sub-themes of the project activities are. To fill this gap, this study intends to explore the activities of the IAC and LINC project for about 10 years through unstructured text analysis.

3. Research Method

3.1 Data Collection

The data were collected through academic papers published in KCI(Korea Citation Index) for the past 10 years (2011~2021). There were a total of 476 articles located by the keyword for 'industry-academic cooperation', and a total of 55 articles including 'LINC'. Korean nouns were extracted in the Korean abstract of each article using Netminer BiblioCollector. For data pre-processing, proper nouns such as university names and region names were deleted. If the letter length is less than 2, it was also deleted. Since the keyword extraction was performed with only nouns in the abstract and no sufficient information on what synonym should be pre-processed, no synonym analysis is performed

during this stage. After this pre-processing, keywords were re-extracted by the two standards-TF(Term Frequency) and TF-IDF(Term-Frequency-Inverse Document Frequency). The keywords with more than TF 15, and with the higher than TF-IDF 0.5 were extracted for keyword network analysis. TF-IDF is a value for the word frequency that occurs in a particular document and is given a document-specific word weight[8]. TF-IDF can be an indicator to interpret the importance of keywords, especially when to analyze and compare the keywords of the two different texts. Through data refining process, a total of 167 keywords for IAC articles, and a total of 132 keywords for the LINC articles were set with no necessity of synonym process.

3.2 Data Analysis Method

For keyword network analysis, the relationship between keywords was quantified by the degree centrality and the betweenness centrality. Degree centrality is the degree of connection between keywords and keywords, and the more keywords

connected to neighboring keywords (nodes), the more connection centrality exists. The betweenness centrality shows how much a particular keyword acts as an intermediary to other keywords in building a network. The high mediocentric keyword is located in the shortest path that connects the sub-network[9]. Such Keyword centrality analysis ranked the size of influence of each keyword within the network, showing which keyword is influential.

In this study, topic modelling analysis was also performed using Netminer to find sub-topics of IAC and LINC articles. Topic modelling is an analysis technique that automatically finds topics from large, unstructured documents by clustering words. Topic modelling is suitable for research that tracks major issues and topic changes in a specific field over time. In particular, topic modelling is being evaluated as a suitable technique for multiple text analysis, because it is based on LDA (Latent Dirichlet Allocation) model that presumes a word can correspond to multiple topics rather than just one topic[10]. This

Table 1. Top 20 keyword by frequency

	TF					TF-IDF			
	IAC		LINC			IAC		LINC	
1	research	2016	university	266	1	area	620	start-up	71
2	cooperation	1829	cooperation	208	2	innovation	333	center	27
3	university	1702	industry-academic	204	3	start-up	262	action	25
4	university-industry	1389	business	180	4	site	235	employment	21
5	education	1341	education	151	5	professor	226	training	21
6	analysis	999	research	136	6	employment	219	satisfaction	18
7	skill	899	area	104	7	school	215	policy	18
8	corporation	844	analysis	102	8	industry-academia-research	212	environment	16
9	business	716	startup	86	9	activity	207	step	16
10	area	690	output	79	10	patent	184	economy	14
11	development	621	result	59	11	duty	181	overseas	13
12	support	563	society	58	12	market	179	focus	13
13	industry	534	problem	46	13	problem	175	field	11
14	result	529	management	43	14	factor	160	element	11
15	policy	463	training	43	15	participation	152	college student	11
16	performance	450	transfer	41	16	effect	147	measurement	11
17	design	445	organization	41	17	center	146	activity	11
18	necessity	445	step	41	18	side	144	number of cases	11
19	center	433	side	40	19	enterprise	143	completion	10
20	process	423	small business	39	20	upbringing	141	direction	10

characteristic is considered to be useful for examining possible topics of the collected texts in the study.

4. Results and Discussions

4.1 Keyword Network Analysis

Keyword network analysis was performed to identify the characteristics of the collected entire texts. Table 1 shows the results of the top 20 keywords by TF and TF-IDF.

It was shown that 'university', 'cooperation', 'industry-academic', and 'education' occur in common among the top five words by TF in both IAC and LINC articles in Table 1. However, TF-IDF had a difference between the data of the two groupss. The top 5 keywords by TF-IDF were 'area', 'innovation', 'start-up', 'site', and 'professor' for IAC articles while in the case of LINC articles, 'start-up', 'center', 'action', 'employment' and 'training' were found.

Next, keyword centrality analysis was performed. By checking the degree centrality and betweenness centrality, the vocabularies with the greater influence in the keyword network were identified. Table 2 and Table 3 shows the results of the top 10 keywords that have a major influence in the network of each text data.

Fig. 1 and Fig. 2 show the visualization of the keyword degree centrality. In the figures, the larger the node size and the more the number of connecting lines, the stronger the connection to the surrounding nodes, and therefore the greater the influence. It was seen that 'area', 'site', 'step', and 'organization' are at the center of the network in IAC articles while in the LINC articles, 'start-up', 'behavior', 'satisfaction', and 'environment' are at the center of the network with high connectivity. The results mean that 'area' in IAC, and 'start-up' in LINC articles could be most influential.
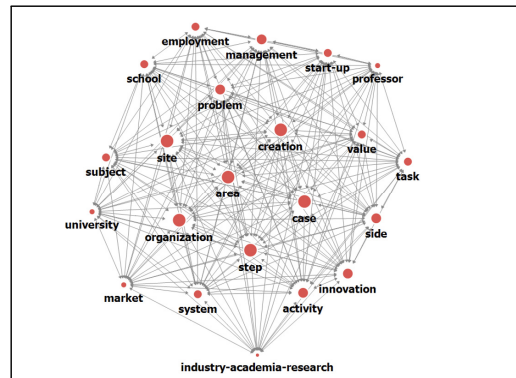
Table 2. Keyword degree centrality

	degree centrality
IAC	area(0.974), site(0.921), step(0.920), organization(0.920), case(0.898), creation(0.890), activity(0.868), problem(0.895), innovation(0.816), management(0.816)
LINC	start-up(0.344), behavior(0.237), satisfaction(0.214), environment(0.183), college student(0.178), improvement(0.175), overseas(0.160), policy(0.153), field(0.137), center(0.130)

Table 3. Keyword betweenness centrality

	betweenness centrality
IAC	area(0.012), site(0.011), organization(0.011), case(0.008), creation(0.008), step(0.007), activity(0.008), problem(0.008), innovation(0.007), management(0.007)
LINC	start-up(0.234), behavior(0.117), environment(0.079), satisfaction(0.070), field(0.068), policy(0.064), employment(0.058), overseas(0.043), element(0.035), completion(0.032)

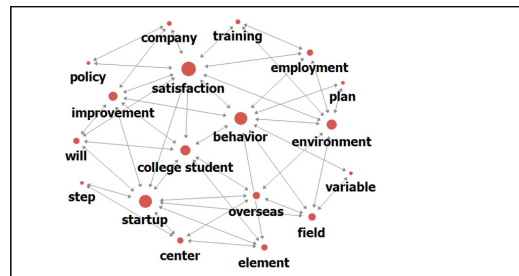Fig. 1. Visualization of IAC keyword degree centrality

Fig. 2. Visualization of LINC keyword degree centrality

4.2 Topic Modelling Analysis

In this study, Topic Modelling analysis was conducted to find out the main sub-topics of IAC and LINC articles. To set the appropriate number of topics, the analysis was repeated by setting the number of topics from 3 to 6, and the alpha value was 0.1, the beta value was 0.01, and the sampling repetition number was set to 500 repetitions. The maximum number of keywords to each topic is set to 100. As a result of the analysis, it was judged that 6 topics for IAC and 2 topics for LINC articles could be the most meaningful in interpretation. The topic issue was named considering the characteristics of the top keywords of each topic. The ratios of sub-topic of each group are presented in Fig. 3, and a list of the topics, and the top 5 keywords are presented in Table 4 and Table 5.
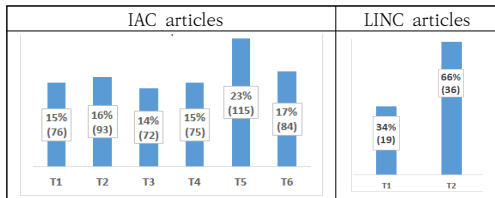

Fig. 3. Topic proportion

Table 4. Topic issues of IAC articles

	Issue	Top 5 keywords
Topic 1	employment	school, employment, duty, job, industry
Topic 2	management	industry-academic-research, management, side, efficiency, service
Topic 3	site	site, training, practice, matter, experience
Topic 4	start-up	start-up, patent, professor, transfer, organization
Topic 5	innovation	area, innovation, industry-academic-research, activity, small business
Topic 6	market	market, area, culture, value, brand

Table 5. Topic issues of LINC articles

	Issue	Top 5 keywords
Topic 1	start-up	start-up, behavior, satisfaction, focus, policy
Topic 2	start-up	start-up, center, training, job, environment

As shown in Table 4 and 5 above, the IAC articles discussed various issues such as 'employment', 'management', 'site', 'start-up', 'innovation', 'market' with the largest proportion of the 'innovation' topic, while the LINC articles dealt with students' wills and behaviors and discussed an environment of 'start-ups'. In particular, it was seen that the 'start-ups' was derived as major topics from both IAC and LINC articles. However, the difference between the two is the start-ups was related to professors, transfer and patent in IAC articles while in LINC, it seemed to be related to training and an environment of the student start-up.
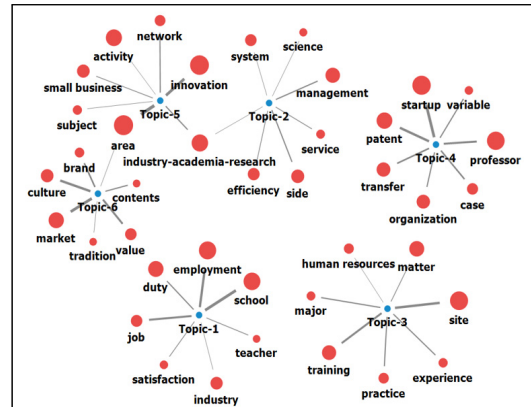

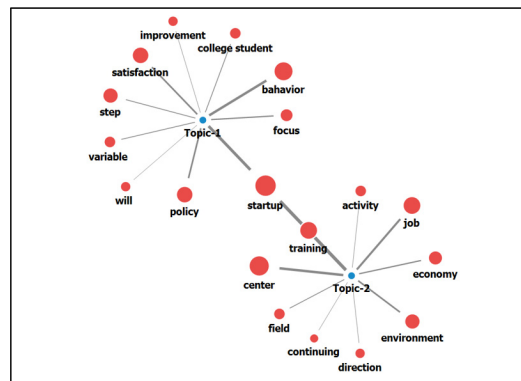Fig. 4. Visualization of topic modelling of IAC articles

Fig. 5. Visualization of topic modelling of LINC articles

Next, a 2-mode network between keywords of each topic was visualized. This is to understand the correlation between topics with a keyword network. The IAC articles showed that 'area', and 'industry-academia research' served as mediators to connect to the topic 2, 5, and 6. In the case of LINC articles, the keyword linking Topics 1 and 2 is found to be a 'start-up' and 'traning'.

5. Conclusions

This paper analyzed and compared the characteristics of industry-academic cooperation and LINC project research through keyword network analysis and topic modelling analysis. As a result, it was found that the main topics of the industry-academic cooperation research were employment, management, site, start-up, innovation, market. In the case of LINC articles, start-up behavior and start-up environment were the main topics, and related keywords were found to be influential. These results can be interpreted that industry-academic cooperation has been actively conducted in schools, and has had a great interest in how closely to work with the local communities. Also it is interpreted that university has played a major role in education program and creating environments of student start-ups over the past decade.

Implications from the results of the study are that in industry-academic cooperation, a school, a teacher, a professor have played major roles in job creation, knowledge transfer and collaborative research, thus, the more efficient system to connect the universities, companies, professors, and students should be fully considered. In order for industry-academic cooperation to have continuity, the financial support is in need, of course, but more importantly, creating an collaborative environment for innovation, start-up, or employment is ultimately in hand of members concerned. To raise the performance

level of industry-academic cooperation, the matter of how firmly cooperation and communication is established between universities and companies would be a key.

References

[1] Y. S. Noh, Y. S. Lee, "An analysis of the trends in domestic research on industry-academic cooperation," Journal of Agricultural Education and Human Resource Development, vol. 50, no. 3, pp.179-206, 2018.
DOI: http://dx.doi.org/10.23840/agehrd.2018.50.1.179

[2] J. H. Lee, S. Y. Heo, H. E. Jang, "Analysis of the government's industry-academic cooperation policy trend using the big data analysis method," Journal of The Korean Association of Regional Geographers, vol. 26, no. 3, pp.260-271, 2020.
DOI: http://dx.doi.org/10.26863/JKARG.2020.8.26.3.260

[3] S. H. Ju, Y. C. Choi, S. Y. Lee. "Exploring the future directions for the government-funded university development projects using social network analysis," Korean Governance Review, vol.25, no.1, pp.237-256, 2018.
DOI: http://dx.doi.org/10.17089/kgr.2018.25.1.009

[4] Joint Department of Relations. Amendment to the First Master Plan for Industrial Education and Industry-Academic Cooperation, 2020, https://moe.go.kr/boardCnts/view.do?boardID=337&boardSeq=83152&lev=0&searchType=null&statusYN=W&page=1&s=moe&m=0303&opType=N (accessed Aug. 20, 2021)

[5] D. J. Kim, "An empirical analysis of policy effects of the industry-academic cooperation leading university (LINC) development project," Korean Journal of Policy Analysis and Evaluation, vol. 28, no.3, pp.27-47, 2018.
DOI: http://dx.doi.org/10.23036/kapae.2018.28.3.002

[6] Ministry of Education, Development Of a Project To Foster Industry-Academic Cooperation Project, 2019, Available From: https://www.nrf.re.kr/biz/info/notice/view?menu_no=378&page=3&nts_no=113107&biz_no=356&target=&biz_not_gubn=result&search_type=NTS_TITLE&search_keyword1 (accessed Aug. 20, 2021)

[7] C. H. Kim, S. C. Yi, "A Study on the performance of industry-academic cooperations program," GRI Review, vol.18, no.2, pp.77-102, 2016.

[8] K. M. Moon, "Is government financial support for industry-academic cooperation effective?," Korean Society And Public Administration, vol. 29, no. 1, pp.53-82, 2018.

[9] Y. J. Jang, W. Son, H. J. Hwang, Unstructured Data

Analysis, p.46, KNOU press, 2020, pp. 1~120.

[10] G. C. Lee, Topic Modelling, LDA, Available From https://ratsgo.github.io/from%20frequency%20to%20semantics/2017/06/01/LDA/ (accessed August, 20, 2021)

Sun-Hee Kwon   [Regular member]

- Aug. 2006 : Univ. of Florida, Linguistics, Ph.D.
- Mar. 2007 ~ current : Busan University of Foreign Studies, Dept. of English, Professor
- Oct. 2018 ~ current : Busan University of Foreign Studies, LINC+Project, Director

⟨Research Interests⟩
Industry-Academic Cooperation, Text Analysis