
기술평가를 위한 평가 참조정보 생성 방안에 관한 연구*

—기술적 속성 유사성 관점에서—

최성철** · 서원철***

I. 서 론

II. 관련 연구

1. 기술평가를 위한 peer group 기반의
참조정보

2. Doc2vec 기반의 기술적 속성 유사도

III. 기술평가를 위한 참조정보 생성 방 법론

1. 기존 평가결과를 보유하고 있는 기업
들의 기술적 속성 정량화

2. 기술적 속성 간 유사도 수준 측정

3. Peer group 정의 및 평가 참조정보
생성

IV. 사례분석

1. 평가 대상기업 및 관련기업들의 기술
적 속성 정량화

2. 평가 대상기업 및 관련기업들의 기술
적 속성 간 유사도 수준 측정

3. 평가 대상기업의 peer group 정의
및 평가 참조정보 생성

V. 논의 사항

VI. 결 론

* 이 논문은 부경대학교 자율창의학술연구비(2017년)에 의하여 연구되었음.

** 가천대학교 산업경영공학과 조교수(제1저자).

*** 부경대학교 기술경영전문대학원 기술경영학과 부교수(교신저자).

초 록

기술금융은 기업의 기술혁신 및 사업화 역량에 대한 평가를 바탕으로 적정 수준의 자금을 공급하는 것을 의미하며, 기술평가는 이를 실현하기 위한 기반으로 인식된다. 기술평가는 계량화된 데이터에 의한 정량적 평가와 해당 분야의 전문가에 의한 정성적 평가의 혼합으로 이루어지며, 계량화된 모형은 정성적 평가를 지원하는 역할을 수행하기 때문에 정성적 평가 결과가 최종 기술평가 결과에 직접적으로 영향을 미치는 경향이 강하다. 하지만 정성적 평가는 평가자의 개별적 역량 및 성향 차이에 따라 일관적이지 못한 결과를 도출할 수 있기 때문에, 다양한 기술평가 관련 기관들은 평가 대상기업과 속성이 유사한 기업들을 peer group으로 정의하고 이들의 평가 관련 정보를 평가 참조정보로서 제공하고 있다. 하지만 기존의 방안들은 주로 산업분류 또는 기술분류 간 유사성만을 고려할 뿐, 기술 자체의 속성을 고려하지 못한다는 한계를 지닌다. 이에 본 연구는 딥러닝에서 사용되는 텍스트 문서의 벡터 정보 모델링기법인 doc2vec 모델을 사용하여 기술내용 간 유사도 수준을 측정하여 평가 대상기업과 기술적 속성 관점에서 유사성을 지닌 기업들을 발굴하고 이들로부터 평가 참조정보를 생성하기 위한 방법론을 제시한다. 또한 본 방법론에 의해 산출된 결과와 기존의 peer group 정의 방식에 의해 산출된 결과를 비교하고, 본 방법론의 타당성에 대한 논의를 수행한다. 본 방법론은 실용적 평가 참조정보를 정량적 방식에 따라 제공함으로써 정성적 기술평가 작업의 효율성을 향상시키는 데 기여할 수 있다. 또한 평가자의 개인적 역량 및 성향에 따른 기술평가 결과의 편차를 줄여 줌으로써 기술평가 작업의 신뢰성도 증진시킬 수 있을 것으로 기대된다.

주제어

기술평가, 평가 참조정보, Peer group, 기술적 속성 정량화, 기술내용 유사도

I. 서론

지식기반경제 시대로의 전환이 가속화됨에 따라 기술, 지식, 정보로 대표되는 무형자산이 기업의 지속가능한 경쟁우위를 창출하기 위한 핵심 원천으로서 인식되고 있다.¹⁾²⁾ 이러한 무형자산을 창출, 관리, 활용하기 위한 선순환 체계의 조성은 기술혁신 역량을 기반으로 기업 경쟁력을 제고하기 위한 필수적 요소가 되기 때문에 이러한 체계의 구축을 지원할 수 있는 다양한 활동들이 요구된다. 이 중, 기술혁신의 전 주기에 걸쳐 발생하는 혁신 활동에 금융자원을 배분하기 위한 기술금융은 가장 중요한 요구사항으로 볼 수 있다.³⁾ 기술금융은 기술평가 기반으로 이루어지는 기업금융의 한 분야로서 개별 기업이 보유하고 있는 혁신역량과 기술사업화 역량에 대한 정량적, 정성적 평가를 바탕으로 적정 수준의 자금을 공급하는 것을 의미한다.⁴⁾ 하지만 자금 공급자와 수요자 사이에 높은 정보 비대칭성이 존재하고⁵⁾ 대상 기술이 실제 시장에서 영향력 있는 결과물을 창출해 낼 수 있을지에 대한 불확실성 및 위험성이 높다는 점은 기술에 대한 명확한 평가를 어렵게 함으로써 기술금융의 실현 및 안정적 정착을 방해하는 요인이 된다.⁶⁾ 따라서 성공적인 기술금융 생태계의 조성을 위해서는 기술평가를 효율적으로 지원하기 위한 체계가 필수적이다.

기술금융 실현을 위한 가장 기초적 기반인 기술평가는 일반적으로 계량화된 데이터에 의한 정량적 평가와 해당 분야의 전문가에 의한 정성적 평가의 혼합으로 이루어지며,⁷⁾⁸⁾⁹⁾¹⁰⁾¹¹⁾¹²⁾ 기술 자체에 대한 평가뿐만 아니라 기술사

1) 박선영, “특허기술 평가요인에 관한 연구”, 『지식재산연구』, 제2권 제1호(2007), 30-56면.

2) 이상현 외 4인, “DEA와 MPI를 활용한 OECD 국가간 지식기반산업의 효율성 분석”, 『Entrue Journal of Information Technology』, 제13권 제3호(2014), 7-20면.

3) 김광희, “기술금융의 현황과 과제”, 『과학기술정책』, 제21권 제3호(2011), 51-71면.

4) 이성상, “지식재산 금융 활성화 방안에 관한 연구 — 미국과 이스라엘의 사례를 중심으로”, 『지식재산연구』, 제8권 제3호(2013), 149-169면.

5) 서경 · 권영관, “IP금융의 새로운 패러다임 — 선진 IP금융 모델사례로부터의 시사점”, 『지식재산연구』, 제8권 제3호(2013), 121-148면.

6) 김광희, 앞의 글(주 3), 51-71면.

업화를 위한 주체인 기업에 대한 평가까지 포괄한다.¹³⁾¹⁴⁾¹⁵⁾ 정량적 평가를 위한 계량화된 모형은 전문가에 의한 정성적 평가를 지원하는 역할을 수행하기 때문에 정성적 평가에 의한 결과가 최종 평가결과에 직접적으로 영향을 미치는 경향이 강하다. 하지만 정성적 평가는 평가자 개인의 역량 및 성향에 따라 매우 다른 결과를 도출해 내게 되는데, 이는 기술평가 결과에 대한 신뢰성을 저하시키는 요인이 된다.¹⁶⁾ 이와 같은 평가자 개인의 차이에 의한 평가 결과의 편차를 보정하기 위하여 기술평가 기관들은 평가 대상기업과 속성이 유사한 기업들을 peer group으로 정의하고 이들의 평가 관련 정보를 평가 참조정보로서 제공하기 위한 노력을 시행하고 있다. 대표적 기술금융기관인 기술보증기금 또한 비즈니스 및 재무적 관점에서 기술의 잠재적 미래 가치를 측정하기 위한 평가모형으로서 KTRS(Kibo Technology Rating System)을 갖추고 있다.¹⁷⁾ KTRS는 기술 또는 기술 보유 주체인 기업에 대해 기술성, 시장성,

7) 여인국, 『기술사업화 이론과 실제』, 학현사, 2013, 15-24면.

8) 김성민 외 8인, 『기술사업화의 이해와 적용』, 경문사, 2015, 447-464면.

9) Jun, Sunghae et al., "A technology valuation model using quantitative patent analysis: A case study of technology transfer in Big Data marketing", *Emerging Markets Finance and Trade*, Vol.51 No.5(2015), pp.963-974.

10) Noh, Heeyong et al., "How to improve a technology evaluation model: A data-driven approach", *Technovation*, Vol.72-73(2018), pp.1-12.

11) Shen, Yung-Chi et al., "A hybrid selection model for emerging technology", *Technological Forecasting and Social Change*, Vol.77 No.1(2010), pp.151-166.

12) Yu, Peng & Lee, Jang Hee, "A hybrid approach using two-level SOM and combined AHP rating and AHP/DEA-AR method for selecting optimal promising emerging technology", *Expert Systems with Applications*, Vol.40 No.1(2013), pp.300-314.

13) 김태완 · 윤재홍, "기술가치평가 방법론을 활용한 사례 분석", 『경영논총』, 제33권 (2012), 29-56면.

14) Chiu, Yu-Jing & Chen, Yuh-Wen, "Using AHP in patent valuation", *Mathematical and Computer Modelling*, Vol.46 No.7-8(2007), pp.1054-1062.

15) Kim, Seo Kyun et al., "The effect of R&D, technology commercialization capabilities and innovation performance", *Technological and Economic Development of Economy*, Vol.17 No.4(2011), pp.563-578.

16) Kim, Eungchan et al., "An approach to generating reference information for technology evaluation", *Sustainability*, Vol.10 No.9(2018), p.3200.

17) Liu, Yao-Jen & Liu, Shang-Jyh, "The intellectual property policy of Taiwan: A strategic viewpoint", In *Proceedings of IEEE International Engineering Management*

사업성 및 경영환경의 관점에서 기술사업화 가능성을 평가하고 평가결과를 등급화하여 산출하는 모형이다.¹⁸⁾ KTRS 계열 3개 종류 및 정책목적용 9개 종류 등 총 12개 종류, 61개 세부모형이 운영 중에 있으며, KTRS 계열모형은 기술 및 기업에 대해 미래 성장가능성뿐만 아니라 부실가능성의 측면에서 기술사업화 가능성을 평가한다. 성장가능성은 기술사업 역량과 사업성장 가능성에 따라 측정하고, 부실가능성은 사업위험과 환경위험에 따라 측정한다. 총 33개의 세부항목에 대한 평가 후, 평가자는 평가등급을 A~E 중 하나로 지정한다. 각 세부항목은 평가하고자 하는 목적에 따라 평가대상을 다르게 가져가도록 정의된다. 경영주의 역량을 살펴보는 항목들은 경영주 및 기업 차원을 대상으로 하고, 기술성 항목들은 주로 기술 및 제품을 대상으로 하며, 시장성 항목들은 산업 자체를, 마지막으로 사업성 항목들은 기술사업화 활동의 주체인 기업을 대상으로 한다. KTRS 계열모형에서는 비즈니스 관점의 평가에 있어, 평가 대상기업과의 유사성에 기인하여 peer group을 정의하고 이들의 관련정보를 평가 참조정보로서 제공하고 있다.¹⁹⁾²⁰⁾²¹⁾ 하지만 현재의 peer group은 산업분류 또는 기술분류 간 유사성이라는 단순한 기준에 의해 정의된다. 즉, 평가 대상기업의 한국표준산업분류(Korean Standard Industrial Classification: KSIC)와 동일한 분류에 속한 기업들을 도출하거나 해당 기업의 평가 대상기술과 기술분류가 동일한 기술을 보유한 기업들을 도출하는 형태로 peer group이 정의되는 것이다. Peer group 기반의 참조정보는 근본적으로 기술평가를 지원하는 역할을 수행하는데, 이러한 현재의 방안은 기술 자체가 지니는 속성을 고려하지 못한다는 한계를 지닌다.²²⁾ 따라서 본 연구에

Conference, IEEE, 2004, pp.42-46.

18) 김대철 외 2인, “거시경제변수를 고려한 기술평가모형의 개선: 기술보증기금의 사례”, 『경영과학』, 제30권 제2호(2013), 117-132면.

19) Sohn, So Young et al., “Improved technology scoring model for credit guarantee fund”, *Expert Systems with Applications*, Vol.28 No.2(2005), pp.327-331.

20) Ruskov, Petko et al., “Model for technology adoption supporting system”, In *Proceedings of International Scientific Conference*, UNITECH, 2016.

21) Ju, Yonghan & Sohn, So Young, “Technology credit scoring based on a quantification method”, *Sustainability*, Vol.9 No.6(2017), p.1057.

서는 기술내용 간 유사도 수준을 측정하여 평가 대상기업과 기술적 속성 관점에서 유사성을 지닌 기업들을 발굴하고 이들을 peer group으로 정의하기 위한 방안을 제안함으로써 참조정보의 활용성을 증진시키고자 한다.

본 연구의 방법론은 1) 기존 평가결과를 보유하고 있는 기업들의 기술적 속성 정량화, 2) 이들과 평가 대상기업의 기술적 속성 간 유사도 수준 측정, 3) 유사도 수준에 기반한 peer group 정의 및 평가 참조정보 생성의 3단계로 구성된다. 기술적 속성에 대한 정량화는 유사한 기술내용을 보유하고 있는 기업들을 발굴하기 위한 목적으로 이루어지며, 이를 위해 본 연구에서는 doc2vec을 이용한다. Doc2vec은 비정형 텍스트 기반의 문서를 벡터 공간에 표현하는 알고리즘으로,²³⁾²⁴⁾²⁵⁾ 가변 길이의 특징을 지닌 텍스트 기반의 문서를 고정 길이의 벡터로 표현하는 문서 임베딩을 위한 비지도 학습 기법이다.²⁶⁾²⁷⁾ Doc2vec에는 Distributed Memory(DM) 모델과 Distributed Bag-of-Words(DBOW) 모델의 두 가지 모델이 존재하는데, 일반적으로 DM 모델의 성능이 우수한 것으로 알려져 있어²⁸⁾ 본 연구에서도 DM 모델을 이용한다. 이러한 문서 임베딩을 이용하여 저널 논문의 시각화 방안 연구,²⁹⁾ 재무정보에 대한 감성 분석 연구,³⁰⁾ 특허문서 군집화 연구³¹⁾ 등 다양한 연구들이 수

22) Kim, Eungchan et al., *op. cit.*, p.3200.

23) Le, Quoc & Mikolov, Tomas, "Distributed Representations of Sentences and Documents", In *Proceedings of the 31st International Conference on Machine Learning*, ICML, 2014, pp.1188-1196.

24) Lee, Younghoon et al., "Applying convolution filter to matrix of word-clustering based document representation", *Neurocomputing*, Vol.315(2018), pp.210-220.

25) 육지희·송민, "토픽모델링과 딥 러닝을 활용한 생의학 문헌 자동 분류 기법 연구", 『정보관리학회지』, 제35권 제2호(2018), 63-88면.

26) 김도우·구명완, "Doc2Vec과 Word2Vec을 활용한 Convolutional Neural Network 기반 한국어 신문 기사 분류", 『정보과학회논문지』, 제44권 제7호(2017), 742-747면.

27) 김병식·신주현, "DBSCAN을 활용한 유의어 변환 문서 유사도 측정 방법", 『멀티미디어학회논문지』, 제21권 제9호(2018), 1035-1043면.

28) 김선혜·윤병운, "SAO2vec: SAO(Subject-Action-Object) 구조와 Doc2vec을 이용한 SAO 임베딩 알고리즘 개발", 『대한산업공학회 학술대회논문집』, 대한산업공학회, 2018, 387-398면.

29) Koyamada, Koji et al., "Visualization of JOV abstracts", *Journal of Visualization*, Vol. 21 No. 2(2018), pp.309-319.

행되어 왔다. 본 연구에서 제시하는 방법론의 실제적 활용성을 탐색하기 위하여 하나의 기업을 평가 대상기업으로 선정하고 이와 유사성을 지니는 peer group을 발굴한 후, 평가 참조정보를 생성하는 사례분석을 시행한다. 또한 본 방법론에 의한 결과가 실제 평가결과를 설명하는 수준을 산출한 후, 기존의 peer group 정의를 위해 활용되는 산업분류 및 기술분류 기준에 의한 결과와 비교하고 논의를 수행한다. 본 연구의 방법론은 실용적인 평가 참조정보를 정량적 방안에 따라 제공함으로써 정성적으로 이루어지는 기술평가 작업의 효율성을 향상시키는 데 기여할 수 있다. 또한 평가자의 개인적 역량 및 성향에 따른 기술평가 결과의 편차를 줄여 줌으로써 기술평가 작업의 신뢰성도 증진시킬 수 있을 것으로 기대된다.

II. 관련 연구

1. 기술평가를 위한 peer group 기반의 참조정보

기술평가는 산업계 및 학계 모두에서 오랜 시간동안 그 중요성이 높게 인식되어 왔다.³²⁾³³⁾ 이는 R&D 및 사업화투자, 기술이전 및 거래 등과 같은 다양한 의사결정 영역에 기술평가가 미치는 영향이 매우 크기 때문이다. 따라서 기술평가가 원활하게 이루어질 수 있도록 지원하기 위한 다양한 체계가

30) Sohangir, Sahar et al., "Big Data: deep learning for financial sentiment analysis", *Journal of Big Data*, Vol.5 No.3(2018), pp.1-25.

31) 김재영 외 2인, "Deep Embedding Clustering 기술을 이용한 특허 문서의 비지도 군집화", 『대한산업공학회 학술대회논문집』, 대한산업공학회, 2018, 73-80면.

32) Perkmann, Markus et al., "How should firms evaluate success in university-industry alliances? A performance measurement system", *R&D Management*, Vol.41 No.2 (2011), pp.202-216.

33) Hsu, David et al., "Toward successful commercialization of university technology: Performance drivers of university technology transfer in Taiwan", *Technological Forecasting and Social Change*, Vol.92(2015), pp.25-39.

요구되었고, 이를 위한 다양한 연구들이 수행되어 왔다. 기술평가와 관련된 기존연구들은 크게 평가대상과 평가방법이라는 2가지 측면을 중심으로 논의 수행한다.³⁴⁾ 평가대상 측면에서는 기술 그 자체뿐만 아니라 기술을 활용하는 주체의 역량 또한 평가의 주요 대상이 되어야 함을 지적하고,³⁵⁾ 평가 방법 측면에서는 기술평가의 유형이 정량적, 정성적 그리고 이들의 혼합으로 구분될 수 있음을 지적한다.³⁶⁾³⁷⁾ 정성적 기술평가는 주로 관련 분야 전문가의 판단에 의해 수행되는 반면,³⁸⁾³⁹⁾ 정량적 평가는 계량화된 데이터를 이용한 자동화된 방법론에 기인하여 이루어진다. 정량적 평가결과는 일반적으로 정성적 평가 수행을 위한 보조적인 수단으로 이용되고 있다. 그럼에도 불구하고 정성적 평가는 평가를 수행하는 주체인 평가자의 개별적 역량 또는 성향 차이에 따라 일관되지 못한 결과를 산출하는 문제를 지니고 있다.⁴⁰⁾

따라서 본 연구에서는 평가자에 의해 수행되는 정성적 평가 활동을 지원하기 위한 평가 참조정보를 생성하는 방안을 제안하고자 한다. 참조정보는 다양한 분야에서 다양한 목적으로 구축 및 활용되어 오고 있다.⁴¹⁾ 예를 들어, 과학기술정보 콘텐츠의 가치평가 결과에 대한 참조정보를 통해 무형자산 콘텐츠 시장에서 활용 가능한 객관적 모형을 제안하는 연구,⁴²⁾ 기술의 특

34) Noh, Heeyong et al., *op. cit.*, pp.1-12.

35) Roure, Juan & Keeley, Robert, "Predictors of success in new technology based ventures", *Journal of Business Venturing*, Vol.5 No.4(1990), pp.201-220.

36) Shen, Yung-Chi et al., *op. cit.*, pp.151-166.

37) Yu, Peng & Lee, Jang Hee, *op. cit.*, pp.300-314.

38) Azzone, Giovanni & Manzini, Raffaella, "Quick and dirty technology assessment: The case of an Italian Research Centre", *Technological Forecasting and Social Change*, Vol.75 No.8(2008), pp.1324-1338.

39) Facey, Karen et al., "Patients' perspectives in health technology assessment: A route to robust evidence and fair deliberation", *International Journal of Technology Assessment in Health Care*, Vol.26 No.3(2010), pp.334-340.

40) Choi, Sungchul et al., "Analyzing technological spillover effects between technology classes: The case of Korea technology finance corporation", *IEEE Access*, Vol.6(2018), pp.3573-3584.

41) Hernández Pardo, Ricardo et al., "Sustainable product service systems in small and medium enterprises (SMEs): Opportunities in the leather manufacturing industry", *Sustainability*, Vol.4 No.2(2012), pp.175-192.

성을 반영한 기술가치 산출 모델에서 활용 가능한 참조정보 생성에 관한 연구,⁴³⁾ 과학기술 연구의 지속성 및 재현성 확보를 위한 참조표준데이터 생성 및 관리 방안에 관한 연구⁴⁴⁾ 등이 수행되어 왔다. 이러한 참조정보는 관련 연구에 도움이 될 수 있는 실용적 패턴을 확립하는 역할뿐만 아니라 효율적 검색 활동을 위한 유용한 도구로서의 역할도 수행할 수 있다. 따라서 기술평가를 위한 참조정보를 생성하고자 하는 본 연구는 기술평가 활동에 유용한 정보를 정성평가를 수행하는 평가자에게 제공함으로써 평가 활동의 효율성을 증진시키는 데 기여할 수 있다. 또한 평가자의 개인적 역량 및 성향의 차이에 기인하여 발생하는 평가결과의 편차를 줄여 줌으로써 정성적 기술평가의 신뢰성을 증진시키는 데에도 기여할 수 있을 것으로 기대된다.

2. Doc2vec 기반의 기술적 속성 유사도

Word2vec은 기존의 자연어처리 기법에서 고려되지 않았던 유사도(similarity) 개념을 바탕으로 연속적인 단어를 벡터 공간상에 표현하기 위한 목적으로 제시된 딥러닝 기반의 도구이다.⁴⁵⁾ 단어 또는 단어들의 집합을 고정된 크기의 벡터로 변환하는 가장 일반적인 방법은 원-핫 인코딩(one-hot encoding)이다. 하지만 이는 코퍼스에 존재하는 중복을 제외한 모든 단어 수만큼의 벡터 차원을 요구하기 때문에 메모리 낭비가 큰 문제가 있다.⁴⁶⁾ 게다가 단어들 간

42) 성태웅 외 3인, “과학기술정보 콘텐츠의 가치평가모형 프레임워크 연구”, 『한국콘텐츠학회논문지』, 제16권 제11호(2016), 421-433면.

43) 류승미 · 성태웅, “수자원 분야 공익형 기술가치평가 시스템에 대한 연구”, 『지능정보연구』, 제24권 제3호(2018), 177-198면.

44) 채균식 · 이웅봉, “연구중에 생산된 과학기술 참조데이터 관리에 관한 연구”, 『한국문헌정보학회지』, 제37권 제4호(2003), 131-149면.

45) Zhang, Dongwen et al., “Chinese comments sentiment classification based on word2vec and SVMperf”, *Expert Systems with Applications*, Vol.42 No.4(2015), pp.1857-1863.

46) Kim, Donghwa et al., “Multi-co-training for document classification using various document representations: TF-IDF, LDA, and Doc2Vec”, *Information Sciences*, Vol.477(2019), pp.15-29.

관계가 서로 독립임을 가정하고 있어 이들 사이에 실제로 존재하는 의미론적 관련성을 무시하는 문제 또한 지니고 있다.⁴⁷⁾ 이러한 문제점을 극복하기 위하여 word2vec은 유사한 의미를 지니고 있는 단어는 유사한 문맥에서 나타날 것이라는 분산적 가정⁴⁸⁾⁴⁹⁾⁵⁰⁾에 기반하여 비슷한 의미를 지니고 있는 단어들을 비슷한 벡터 값 또는 수치로 표현한다.⁵¹⁾ 이와 같은 벡터 기반의 수치화를 통해 단어와 단어 간의 의미상 거리를 계량화하거나 벡터 간 연산을 수행하는 것이 가능해질 뿐만 아니라, 단어의 문맥상 의미에 기반한 추론 또한 가능해지게 된다.⁵²⁾⁵³⁾ 이러한 추론 역량에 기인하여 word2vec은 장문의 문서에 대한 특성별 분류,⁵⁴⁾⁵⁵⁾ 웹사이트 콘텐츠를 대표할 수 있는 의미론적 키워드 추출,⁵⁶⁾ 신문기사 내용에 따른 특성 분석,⁵⁷⁾ 영화 리뷰 콘텐츠에 대한 감성 분석⁵⁸⁾ 등 다양한 텍스트마이닝 관련 연구 분야에서 널리 활용되고 있다.

단어만이 아닌, 문장, 단락, 문서 등 가변 길이의 텍스트를 벡터 공간상에 표현하기 위하여 word2vec의 확장 형태인 doc2vec이 제시되었다.⁵⁹⁾⁶⁰⁾⁶¹⁾

47) Kim, Donghwa et al., *op. cit.*, pp.15-29.

48) Harris, Zellig, "Distributional structure", *Word*, Vol.10 No.2-3(1954), pp.146-162.

49) Turney, Peter & Pantel, Patrick, "From frequency to meaning: Vector space models of semantics", *Journal of artificial intelligence research*, Vol.37(2010), pp.141-188.

50) 기홍도 외 5인, "스마트폰 어플리케이션 설치 목록을 이용한 사용자 특성 추론", 『정보과학회논문지』, 제45권 제12호(2018), 1240-1249면.

51) 육지희·송민, 앞의 글(주 25), 63-88면.

52) 최가람·최성필, "단어 임베딩(Word Embedding) 기법을 적용한 키워드 중심의 사회적 이슈 도출 연구", 『정보관리학회지』, 제35권 제1호(2018), 231-250면.

53) 정지수 외 6인, "문서 유사도를 통한 관련 문서 분류 시스템 연구", 『방송공학회논문지』, 제24권 제1호(2019), 77-86면.

54) 김정미·이주홍, "Word2vec을 활용한 RNN기반의 문서 분류에 관한 연구", 『한국지능시스템학회 논문지』, 제27권 제6호(2017), 560-565면.

55) 김도우·구명완, 앞의 글(주 26), 742-747면.

56) 이동훈·김관호, "Word2Vec 기반의 의미적 유사도를 고려한 웹사이트 키워드 선택 기법", 『한국전자거래학회지』, 제23권 제2호(2018), 83-96면.

57) 한은진 외 3인, "Word2vec algorithm 기반 뉴스 기사에 나타난 성별 관련 어휘 규명 연구", 『대한산업공학회지』, 제44권 제4호(2018), 272-282면.

58) 박성수·이건창, "Word2vec과 앙상블 분류기를 사용한 효율적 한국어 감성 분류 방안", 『한국디지털콘텐츠학회 논문지』, 제19권 제1호(2018), 133-140면.

59) Le, Quoc & Mikolov, Tomas, *op. cit.*, pp.1188-1196.

이는 다양한 문서들 간에 존재하는 의미론적 관련성을 유지할 수 있도록 적절한 벡터 공간에 문서를 임베딩하는 것을 목적으로 한다.⁶²⁾ Doc2vec은 다른 문서 임베딩 도구들에 비해 정확도 등 문서 분류 성능이 매우 우수한 것으로 입증되고 있으며, 이에 다양한 분야에서 문서 콘텐츠에 대한 효과적인 벡터화에 기반한 문서 분류를 위한 목적으로 널리 활용되고 있다.⁶³⁾

Doc2vec 모델은 기존 word2vec 모델과 유사한 방식으로 문서에 대한 벡터를 학습한다. Word2vec은 연속적인 단어들이 출현한 후 다음 단어를 예측하도록 모델을 구성하고, 이를 가장 잘 예측할 수 있는 단어들의 임베딩 매트릭스(embedding matrix)를 생성하는 방식으로 학습을 진행한다. Doc2vec은 기존 word2vec에서 연속적인 단어들의 정보 앞에 해당 문서에 대한 정보를 추가함으로써 문서에 대한 벡터와 함께 쓰이는 단어들을 동시에 학습할 수 있도록 한다. 이를 통해 최종적으로 생성된 벡터 공간에서 비슷한 내용을 가진 문서들의 벡터와 함께 사용되는 단어들의 벡터와 유사한 위치에 놓일 수 있도록 학습이 진행된다.

본 연구는 새로운 기업에 대한 평가를 수행하는 과정에 있어, 해당 평가 대상기업과 기술적 속성 관점에서 유사성을 지니고 있는 기업들을 발굴하여 peer group으로 정의하고, 이들의 기존 평가결과 정보를 평가 참조정보로서 제공하는 것을 목적으로 한다. 기술평가에 있어, peer group은 평가 대상기업과 기술분류 또는 산업분류 관점의 공통성을 바탕으로 정의되고 있지만 이는 평가 대상기업이 보유한 기술 자체를 고려하지 못한다는 한계를 지니고 있다.⁶⁴⁾ 따라서 본 연구에서는 기술적 속성 관점에서의 유사성을 바탕으로 평가 대상기업에 대한 peer group을 정의하기 위한 방안을 제시함으로써

60) Kim, Donghwa et al., *op. cit.*, pp.15-29.

61) 정지수 외 6인, 앞의 글(주 53), 77-86면.

62) Le, Quoc & Mikolov, Tomas, *op. cit.*, pp.1188-1196.

63) Xing, Chao et al., "Document classification with distributions of word vectors", In *Proceedings of Signal and Information Processing Association Annual Summit and Conference*, 2014 APSIPA, 2014.

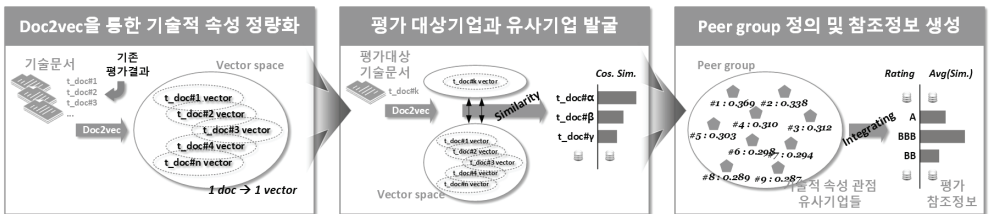
64) Kim, Eungchan et al., *op. cit.*, p.3200.

참조정보의 활용성을 증진시키고자 한다.

Ⅲ. 기술평가를 위한 참조정보 생성 방법론

기술평가 활동에 유용한 평가 참조정보를 정성평가를 수행하는 평가자에게 제공함으로써 평가 활동의 효율성을 증진시키고자 하는 본 연구는 <그림 1>과 같이 1) 기존 평가결과를 보유하고 있는 기업들의 기술적 속성 정량화, 2) 이들과 평가 대상기업의 기술적 속성 간 유사도 수준 측정, 3) 유사도 수준에 기반한 peer group 정의 및 평가 참조정보 생성의 3단계로 이루어진다. 그리고 본 방법론의 실제적 활용성을 탐색할 수 있도록 하나의 기업을 평가 대상기업으로 선정하고 사례분석을 시행하여 참조정보를 산출한 후, 해당 정보의 유용성 및 타당성에 대한 논의를 수행한다.

<그림 1> 연구 프레임워크



1. 기존 평가결과를 보유하고 있는 기업들의 기술적 속성 정량화

평가 참조정보는 평가 대상기업과 기술적 속성 관점에서 유사성을 지니고 있는 기업들의 집합인 peer group의 기존 평가결과를 활용하여 정의된다. 따라서 첫 번째 단계에서는 참조정보 생성을 위해 근간이 되는 기존 평가결과를 가진 기업들에 대한 평가 관련 정보를 수집하고 텍스트로 표현되어 있는 이들의 기술내용으로부터 doc2vec을 통해 기술적 속성을 정량화한다. 이러한 기술적 속성의 정량화를 위해서는 정형화된 형태로 기술내용을 서술하

고 있는 기술문서가 요구된다. 특허는 무형의 기술지식을 체계적 구조에 따라 함축하여 나타내기 때문에⁶⁵⁾ 본 연구에서는 정형화된 기술내용 기반의 기술적 속성 정보로서 특허문서를 활용한다.

평가 대상기업과 기술적 속성 관점에서 유사성을 지니고 있는 기업들을 발굴하여 peer group을 정의할 수 있도록 기존 평가결과를 지니고 있는 기업들의 특허문서로부터 doc2vec 모델을 학습한다. 특허문서는 기본적으로 한글을 기반으로 기술되어 있기 때문에 한글에 대한 처리가 가능하도록 형태소 추출, 품사 태깅, 조사 등 불필요한 품사와 불용어 처리 등의 전처리 과정을 거친다. 또한 doc2vec 모델의 학습을 위해서는 다양한 파라미터 튜닝이 요구되는데, 본 연구에서는 반복실행을 통해 다양한 튜닝 값 중 최적의 값을 선택하는 과정을 시행한다. 이와 같은 과정을 통해 구축된 기존 평가결과를 지니고 있는 기업들의 doc2vec 모델은 이후 평가 대상기업 doc2vec 모델과의 연계를 통해 평가 대상기업과의 유사도 수준을 측정하고 peer group을 생성하기 위한 목적으로 활용된다.

2. 기술적 속성 간 유사도 수준 측정

평가 참조정보는 peer group의 기존 평가결과를 결합하여 생성되는데, 이때 평가 대상기업과 peer group 간 유사도 수준이 가중치로서 활용된다. 따라서 본 단계에서는 평가 대상기업과 기존 평가결과를 지닌 기업들 간 기술적 속성 관점의 유사도 수준을 측정한다. 이미 앞 단계에서 기존 평가결과를 지닌 기업들에 대해 특허문서를 기반으로 doc2vec 모델을 구축하였기 때문에, 본 단계에서는 평가 대상기업에 대해 기술관련 문서를 활용하여 doc2vec 모델을 생성한다. 이러한 기업 간 기술적 속성 관점의 유사성은 해당 기업들이 보유하고 있는 기술내용 간 유사성에 기인한다. 따라서 다양한 관련 기업들의 특허문서 및 평가 대상기업의 기술관련 문서를 활용하여 구축한 doc2vec 모

65) 정오뜸 외 2인, “기술적 파급효과 도출을 통한 기술융합 분석 연구: 나노바이오 융합기술 사례를 중심으로”, 『지식재산연구』, 제10권 제4호(2015), 255-290면.

델을 통해 이들 간 유사도 수준을 계량화할 수 있다.

본 연구에서는 특허문서의 doc2vec 변환에 있어 특허문서 내 다양한 구성 요소 중, 제목과 한글 요약문을 활용한다. 변환과정의 첫 번째 단계는 수집된 한글 정보에 대해 전처리를 수행하는 것이다. 전처리를 위해 우선 불용어(stopword) 리스트를 구성하여 의미 없는 단어들을 제거한다. 다음으로 대표적인 한국어 품사분석 도구(PoSTagger)인 KoNLPy⁶⁶⁾를 사용하여 문장 내에서 주어, 동사, 형용사, 외래어만 추출한다. 그 외의 문장 구성 요소인 조사, 대명사 등은 문서정보에 노이즈를 추가하는 경향이 강하기 때문에 제외하도록 한다. 외래어의 경우, 특허문서의 기술 키워드가 외래어인 경우가 매우 많기 때문에 제외되지 않도록 처리한다. 이와 같은 전처리 과정을 거친 후, 대표적 자연어처리 도구인 gensim을 활용하여 수집된 정보 중 복합단어로 생성 가능한 trigram 단어를 생성하고, 기존 텍스트와 생성된 trigram 모두를 포함하여 doc2vec 모델을 생성한다. 기술용어의 경우 한 개 이상의 단어로 구성된 복합단어들이 많기 때문에 trigram의 처리가 필수적이다. 특허문서에서 gensim을 활용하여 doc2vec을 생성할 경우, 기술용어에 대한 벡터 추출이 가능하며, 이는 기술용어 유사도의 산출을 가능하게 한다. Doc2vec 모델 생성 시 하이퍼 파라미터의 설정이 요구되는데, 본 연구에서는 생성되는 문서벡터 크기 300, 윈도우 크기 6, 최소 출현횟수 10을 활용한다.

본 연구에서는 코사인 유사도를 이용하여 기술적 속성 간 유사도 수준을 측정한다. 코사인 유사도는 두 벡터 간 유사도를 방향성의 관점에서 계량화하며,⁶⁷⁾ 완벽한 반대의 방향성을 의미하는 -1에서 완벽히 일치하는 방향성을 의미하는 1 사이의 값을 갖는다. 두 벡터가 갖는 차원 및 크기에 무관하게 방향성에 대한 비교를 가능하게 하는 장점을 가지고 있어, 다차원 공간상에서의 유사도 비교를 목적으로 널리 활용되고 있다. 코사인 유사도는 다음 식

66) KoNLPy, "KoNLPy: Korean NLP in Python", <<http://konlpy.org/en/latest/>>, 작성일: 2019.5.23.

67) Garcia, Noa & Vogiatzis, George, "Learning non-metric visual similarity for image retrieval", *Image and Vision Computing*, Vol.82(2019), pp.18-25.

을 통해 측정된다.

$$\cosine(A, B) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n (A_i \times B_i)}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (1)$$

여기서, A_i 와 B_i 는 각각 벡터 A 와 B 의 i 번째 성분을 의미한다.

3. Peer group 정의 및 평가 참조정보 생성

본 단계에서는 평가 대상기업과의 유사도 수준에 기반하여 peer group을 정의하고 peer group의 기존 평가결과에 유사도를 결합하여 최종적으로 평가 참조정보를 생성한다. Doc2vec 및 코사인 유사도를 이용하여 유사도 수준을 측정하게 되면, 모든 기술문서에 대해 언제나 유사도 값이 존재한다고 나타나는 것이 일반적이기 때문에 유사도 값을 가지는 모든 기업들을 peer group에 포함시키는 것은 비합리적이다. 따라서 너무 미미한 유사성을 지닌 기업들은 배제하면서도 참조정보 생성에 충분할 만큼 적정 수의 기업들이 peer group에 포함될 수 있도록 유사도 값에 대한 임계치를 설정하는 과정이 요구된다. Peer group의 기존 평가결과와 유사도 수준을 결합할 수 있는 가장 직관적인 방법은 기존 평가등급별 유사도 수준에 대한 가중평균을 산출하는 것이다. 기존 평가결과가 가능한 모든 평가등급에 대해 균일한 분포를 이루고 있다면 이 방법은 매우 합리적이라 볼 수 있을 것이다. 하지만 특정 소수의 등급에 대부분의 평가결과가 집중되어 있다면, 가중평균은 언제나 비슷한 결과만을 산출해 내게 된다. 이는 참조정보의 활용성 및 실효성을 저조하게 만드는 요인이 된다.⁽⁶⁸⁾ 따라서 본 연구에서는 peer group 내 기업

(68) Kim, Eungchan et al., *op. cit.*, p.3200.

들에 대해 기존 평가등급별 유사도 수준의 평균값을 산출하고 이를 이용하여 참조정보를 생성하고자 한다. 이는 기존 평가결과가 등급에 따라 큰 편차를 보이더라도 이에 무관하게 유사도 수준에 대한 트렌드를 확인할 수 있도록 하는 장점을 지닌다. 또한 각 등급에 포함된 기업의 수가 현저히 적을 경우에도, 그들이 지닌 유사도 수준만을 그대로 나타낼 수 있도록 하기 때문에 결과의 왜곡을 유발하지 않는 장점도 지닌다. 등급별 평균 유사도 수준은 다음 식을 통해 측정된다.⁶⁹⁾

$$density_{i,r} = \frac{\sum_{j \in r} cosine(i, j)}{count_{i,r}} \quad (2)$$

여기서, $\sum_{j \in r} cosine(i, j)$ 는 평가 대상기업 i 와 peer group 내 유사기업 j 중, 기존 평가결과 등급이 r 인 기업들 간 유사도 수준의 합을 나타내며, $count_{i,r}$ 은 평가 대상기업 i 의 유사기업들 중, 기존 평가결과 등급이 r 인 기업들의 수를 의미한다. 이러한 정의에 따라 $density_{i,r}$ 는 평가 대상기업 i 에 대해 평가등급 r 별 평균 유사도 수준을 나타내는 density의 의미를 갖는다. 본 연구에서는 이러한 density에 대한 기술통계량 및 분포도 형태로서 평가 참조정보를 생성하고자 한다. 이와 같은 기술통계 및 그래픽 기반의 참조정보를 기술평가를 수행하는 평가자들에게 제공함으로써 평가 활동의 효율성을 증진시키는 데 기여할 수 있다.

69) Kim, Eungchan et al., *op. cit.*, p.3200.

IV. 사례분석

1. 평가 대상기업 및 관련기업들의 기술적 속성 정량화

본 연구에서 제안한 방법론의 활용성을 탐색하기 위하여 하나의 평가 대상기업을 선정하고 사례분석을 시행한다. 본 단계에서는 평가 대상기업 관련 정보와 함께 peer group 생성을 위해 필요한 특허데이터 및 기존 평가결과를 지닌 기업들의 정보를 수집한다. <표 1>은 본 사례분석을 위한 평가 대상기업의 기본 정보를 나타낸다. 해당 기업은 KSIC상 자동차 및 트레일러 제조업인 30에 속한 기업으로서 자동변속기용 유성기어 관련 기술을 보유하고 있으며, 해당 기술은 기술보증기금의 KTRS를 통해 BB 등급을 획득하였다. 평가 대상기업과 유사한 내용의 기술을 보유하고 있는 기업들을 도출하기 위하여 한국특허청에 출원된 특허를 수집하고 doc2vec을 이용하여 유사도 수준을 측정한 후 peer group을 생성한다. Peer group에 포함된 기업들의 기존 평가결과를 그 유사도 수준과 결합하여 최종적으로 평가 참조정보를 생성하기 때문에 peer group에 포함되기 위해서는 기존 평가결과를 보유하고 있어야만 한다. 따라서 본 연구에서는 기술보증기금의 KTRS 결과를 기존 평가결과로서 활용하며, 특허데이터 또한 이러한 기존 평가가 이루어진 시기에 부합하도록 추출한다. 이러한 과정을 통해 총 824,710건의 특허데이터를 수집하였고, 이를 이후 분석에 활용하고자 한다.

<표 1> 사례분석을 위한 평가 대상기업의 기본 정보

KSIC	기술명	기술분류코드	기존 KTRS 평가 등급
30	자동변속기용 유성기어	F251	BB
기술내용			
유성기어장치는 자동변속기 내 동력 전달용 기어로서 주행속도에 맞추어 동력을 전달하는 기능을 수행하는 핵심 부품이며 유성기어장치는 그 중심에 위치하는 선기어(변속기 주축의 구동축에 고정되어 회전력을 전달), 외주에 위치하는 링기어(변속기 주축의 피동축에 연결되어 구동력을 추진축으로 전달), 그리고 선기어와 링기어 사이에 위치하는 유성기어 및 유성기어를 지지하는 캐리어로 구성되고, 3개의 기어 요소를 고정하거나 구동 또는 자유로 하여 직결, 감속, 가속, 역전 및 중립 등으로 변속이 가능토록 하는 장치임.			

2. 평가 대상기업 및 관련기업들의 기술적 속성 간 유사도 수준 측정

본 단계에서는 평가 대상기업과 기술적 속성 관점에서 높은 유사성을 지닌 기술을 보유하고 있는 기업들을 발굴한다. <표 1>의 기술내용과 수집된

<표 2> 평가 대상기업의 보유기술과 유사성을 지닌 특허 정보(유사도 기준 상위 10개)

출원인코드	출원번호	특허명	기술적 속성 유사도
119980045675	1020060104990	자동변속기의 8속 파워트레인	0.4041
119980045675	1020120144869	하이브리드 자동차의 동력전달장치	0.3671
119980045675	1020120116826	차량용 자동변속기의 유성기어트레인	0.3601
119980003169	1020060083112	4륜구동차량용 부변속기	0.3581
119980018093	1020110058075	선박용 추진장치 및 이를 포함하는 선박	0.3577
119980045675	1020060118483	차량용 10속 자동 변속기의 기어 트레인	0.3557
119980045675	1020060069220	자동변속기의 7속 파워트레인	0.3533
120050288508	1020130060801	자동변속기의 6속 파워트레인	0.3423
319999029382	1020100064016	복수 개의 동력전달 경로를 가지는 유성기어 어셈블리	0.3355
519980783324	1020097008817	다단 감속 기어	0.3327

<표 3> 평가 대상기업과 기술적 속성 관점의 유사성을 지닌 유사기업(유사도 기준 상위 10개)

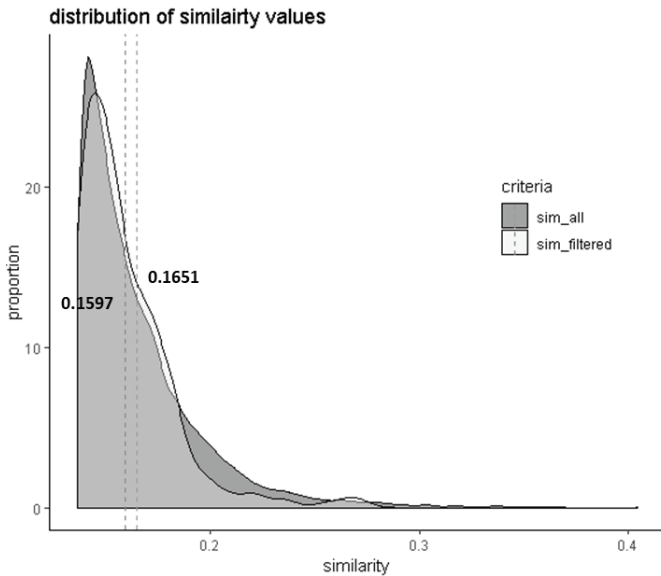
기업ID	기술명	기술분류 코드	기술적 속성 유사도
294650	Magnet Coupling을 이용한 양방향 모터제어	F159	0.2729
268542	3D 라스트콘텐츠를 이용한 맞춤형 제작	ED9999	0.2683
322969	산업차량 및 건설장비용 자동변속기	EA0901	0.2636
279081	AMOLED기관 표면 처리 장비	F482	0.2548
287340	감속장치(Reduction Gear) 제조기술	F251	0.2374
243172	탈부착 및 정비가 용이한 유량계가 구비된 지열 보호관	N240	0.2339
94250	비상용 전동 액추에이터	F147	0.2230
256143	무인 이송 시스템용 서플 킬팅 장치	F100	0.2211
236925	AG/CU세정액조성물을 활용한 CATHODE MASK	H481	0.2205
322969	산업차량용 자동변속기 구동장치	EA0901	0.2072

특허문서의 요약문에 doc2vec을 적용하여 벡터화를 실행한 후, 코사인 유사도를 측정하였으며, <표 2>는 그중 유사도 수준이 높은 상위 10개 특허를 나타낸다. 모두 변속기, 유성기어, 추진장치 등 평가 대상기업의 보유기술과 밀접한 연관성을 지니고 있는 것들이 도출된 것으로 판단된다. 3D 라스트콘텐츠 활용, 기관 표면 처리, 마스크 등 직접적 연관관계가 표면상 잘 나타나지 않는 기술적 속성들도 발굴되었지만, 이들 또한 평가 대상기술과 연관성을 지니고 있는 것으로 볼 수 있다. 3D 라스트콘텐츠의 경우, 개별 사람의 특성을 반영한 맞춤형 보조기구와 관련된 것으로 기구의 효과적 구동을 위한 동력 전달용 기어에 대한 기술적 내용을 포함하고 있다. 기관 표면 처리의 경우, 회전 프레임을 기반으로 표면에 대한 처리 과정을 기계화하는 것과 관련된 것으로 회전 프레임을 회전 구동하는 구동부에 대한 기술적 속성 또한 포괄한다. 마스크의 경우, 반도체 기관 또는 평판 디스플레이장치 제조 공정에 필요한 척에 대한 것으로 이는 제조대상인 기관의 이송 및 거치에 있어 중요한 역할을 담당한다. 척 또한 제조대상의 효과적 이송 및 거치를 위하여 구동 관련 기능을 함께 포함한다. 평가 참조정보는 관련 기술들의 기존 평가결과를 이용하여 정의되기 때문에, 이러한 유사성을 지니고 있는 기술들 중, 기존 평가결과가 존재하는 것들만 추출하여 최종적으로 peer group을 생성한다. 이는 KTRS 기반의 기존 평가결과와 연계함으로써 수행되며, <표 3>은 이 중 유사도 수준이 높은 상위 10개 기업 및 보유 기술을 나타낸다. 충분한 유사성이 있다고 판단되는 기업들을 묶어 peer group으로 정의하고 이들의 기존 평가결과와 평가 대상기업과의 유사도 수준을 결합함으로써 평가 참조정보를 산출한다.

3. 평가 대상기업의 peer group 정의 및 평가 참조정보 생성

본 단계에서는 앞 단계에서 측정한 유사도 수준에 기반하여 평가 대상기업과 기술적 속성 관점의 유사성을 지니고 있는 유사기업들을 발굴하고 이들을 묶어 peer group으로 정의한다. 그리고 peer group의 기존 평가결과를 활용하여 최종적으로 평가 참조정보를 생성한다. 기본적으로 유사도 수준이

〈그림 2〉 전체 및 기존 평가결과를 지닌 기업들과 평가 대상기업 간 유사도 분포



0보다 큰 경우, 유사성을 지니고 있다는 판단을 할 수 있지만, 너무 미미한 수준까지 포괄하여 모든 기업들을 peer group에 포함시키는 것은 바람직하지 않다. 따라서 평가 참조정보를 생성하기에 충분한 수의 기업들을 확보하면서도 매우 미미한 수준의 유사도 값을 가지는 기업들을 배제해 낼 수 있도록 적절한 임계치를 설정하는 것이 필요하다. 평가 대상기업의 기술내용과 유사성이 있다고 판단된 특허문서들이 약 16,000건 추출되었고, 이들의 유사도 값 평균은 0.1651이었다. 이들 중, 기존 평가결과를 가지고 있는 기업들의 특허문서는 280건이었고, 이들의 유사도 값 평균은 0.1597이었다. 〈그림 2〉는 전체 특허문서 및 기존 평가결과를 가지고 있는 기업들과 평가 대상기업 간 기술적 속성 관점의 유사성 수준에 대한 분포를 나타내는데, 전체적으로 평균보다 낮은 수준에 밀집되어 있는 양상을 확인할 수 있다. 따라서 본 연구에서는 peer group 정의를 위한 유사도 수준 임계치로서 평균 유사도 값보다 약간 낮은 수준인 0.140을 활용한다. 이를 통해 전체 대비 약 75% 수준의 유사기업들이 peer group에 포함되게 된다.

〈표 4〉 Peer group 내 유사기업들의 기존 평가결과에 대한 기술통계량

유사기업 수	min.	max.	Q1	Q2	Q3	Avg.	Stdev.
208	5	9	7	7	8	7.3846	1.1104

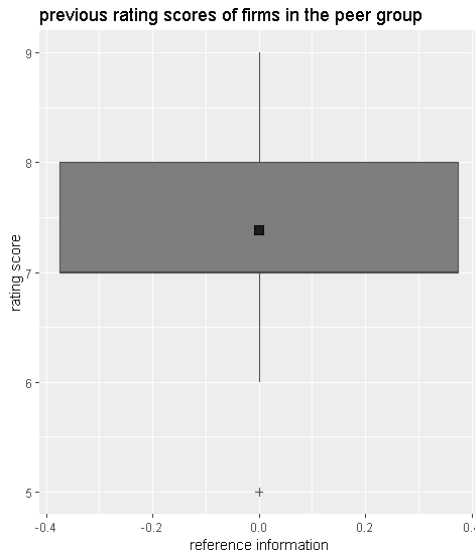
〈표 4〉는 이러한 과정을 통해 산출된 평가 대상그룹의 peer group에 속한 208개 기업들의 기존 평가결과를 정리한 것이다. 가장 높은 등급인 AAA를 10점으로 하고, 가장 낮은 등급인 D를 1점으로 할당함으로써 각 기업들의 기존 평가결과를 수치화하고 이들에 대한 기술통계량을 산출하면, 최소값 5, 최대값 9, 중앙값 7, 평균 7.3846, 표준편차 1.1104와 같다. 이러한 peer group의 기존 평가결과에 대한 수치정보를 이용하여 box plot을 나타내면 〈그림 3〉과 같다. 〈그림 4〉는 평가등급별 평균 유사도 수준인 density 분포를 의미하는데, BB 등급에서 가장 높은 값을 보여 준다. 본 사례분석에서 대상으로 삼고 있는 기업의 기존 KTRS 평가결과가 〈표 1〉에서 제시된 것처럼 BB 등급이었다는 사실에 비추어 볼 때, 본 연구에서 제안한 density 기반의 평가 참조정보가 합리성을 지니고 있다고 판단된다.

기술평가 활동은 주로 정성적 작업에 기반하여 수행되며, 이는 평가자 개인의 역량 및 성향에 따라 매우 다른 결과를 산출해 내는 경향이 있다. 따라서 정성적 평가활동의 일관성과 신뢰성을 향상시킬 수 있도록 평가자의 기술평가 활동을 지원하기 위한 가이드를 갖추는 것이 필요하다. 본 연구에서 제안하는 density 기반의 평가 참조정보는 평가 대상기업과 기술적 속성 관점에서 유사성을 지닌 peer group이 기존에 획득한 평가결과가 어떠한지를 종합적 체계에 따라 산출하며, 평가결과에 대한 추이를 쉽게 확인할 수 있도록 함으로써 이러한 가이드로서의 역할을 충실히 수행해 낼 수 있다.

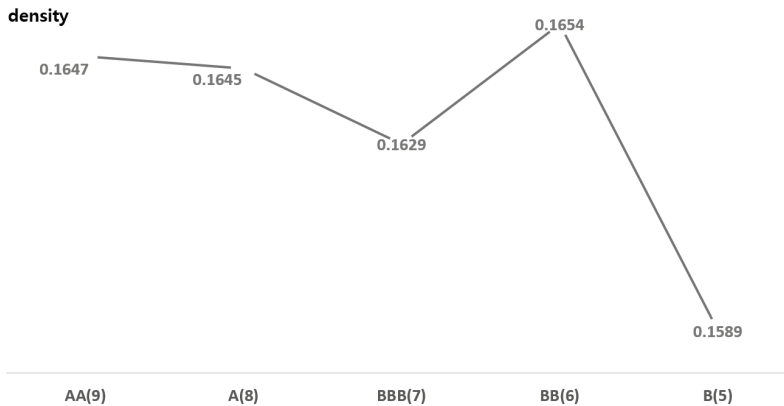
평가자들은 본 연구에서 제안하는 참조정보를 평가 기준선으로서 활용할 수 있다. 평가대상 기술과 속성이 유사한 기술들의 기존 평가결과를 기반으로 생성된 density 정보를 통해 새로운 평가 수행을 위한 기준선의 경계를 설정하고 이를 다른 평가자들과 공유함으로써 평가자의 개인적 성향 및 역량에 따른 편차를 최소화할 수 있다. Density는 평가자에게 제공하는 정보의

객관성 및 시인성 증진을 목적으로 기술통계량과 분포와 같이 단순한 형태로 정의되었다. 따라서 실제적인 활용성을 높일 수 있도록 보다 구체적인 개별 기업 정보 또한 참조정보의 범주에 포함시키고자 한다. 한 예로서, 〈그림 5〉는 유사도 수준 상위 5개 기업에 대한 KTRS 세부 항목별 평가결과를 나타

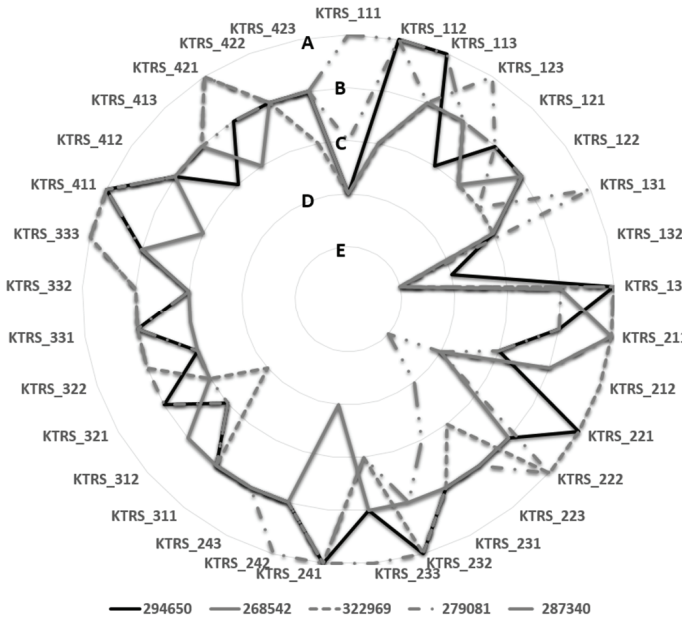
〈그림 3〉 Peer group 내 유사기업들의 기존 평가결과에 대한 box plot



〈그림 4〉 평가 참조정보로서 생성된 density 분포



〈그림 5〉 유사도 수준 상위 5개 기업에 대한 KTRS 세부 항목별 평가결과

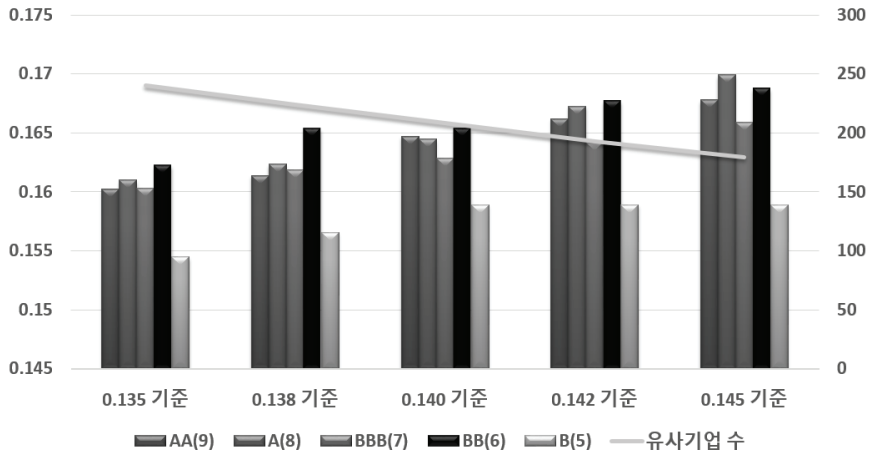


낸다. 이와 같은 형태로 해당 기업들의 정성적 평가사유 등 기존 평가관련 세부 이력정보를 참조정보로서 함께 제공할 수 있을 것이다. 평가자들은 이러한 복합적 참조정보 체계를 활용하여 새로운 평가 수행을 위한 합리적 평가 기준선을 구축할 수 있게 된다.

V. 논의 사항

기술평가 활동의 신뢰성 및 효율성을 향상하기 위하여 기술 외적인 요소와 관련된 다양한 정보들을 폭넓게 활용하기 위한 노력들이 활발하게 시도되고 있다. 기술보증기금 또한 평가 대상기업과 기술분류 또는 산업분류 관점에서 유사성을 지닌 기업들을 peer group으로 정의하고 이들의 기존 평가 결과를 평가 참조정보로서 활용해 오고 있다.⁷⁰⁾ 본 연구에서도 기술적 속성

〈그림 6〉 임계치 수준의 변화에 따른 유사기업 수 및 density 값의 변화 양상



관점의 유사도 수준을 정량적 방법으로 측정하고 이에 기인하여 peer group을 생성하는 방안을 제안하였다. 그런데 유사도 수준의 값이 존재하는 모든 기업들을 peer group에 포함시키는 것은 불합리하기 때문에 peer group으로 판단하기 위한 기준으로서 유사도 수준에 대한 임계치를 설정하는 것이 필요하였다. 이를 위하여 본 연구에서는 매우 미미한 수준의 유사성을 지닌 기업들을 배제하면서도 평가 참조정보 생성에 충분한 수의 기업들이 포함될 수 있도록 한다는 매우 단순한 기준을 정의하고, 이 기준하에서 자의적 판단에 따라 임계치를 설정하였다. 따라서 이러한 임계치 수준의 변화에 따른 density 기반 평가 참조정보의 변화를 살펴보는 것은 본 연구에서 제안한 방안의 타당성 및 강건성을 보다 깊이 있게 탐색하기 위해 반드시 필요한 활동으로 볼 수 있다. 〈그림 6〉은 임계치 수준의 변화에 따른 평가등급별 density 값의 변화를 나타낸다. 임계치 값을 0.135로 설정할 경우, 기존 평가 결과를 가진 기업들 중 86%가 peer group으로 포함되는 반면, 0.145로 설정할 경우, 그 비율이 64%까지 떨어지는 양상을 보였기 때문에, 참조정보 생성에 필요한 충분한 수의 기업들을 포함시킬 수 있도록 0.135부터 0.145 사이

의 값들만을 임계치로 설정하여 탐색을 수행하였다. 평가등급별 평균 유사도 수준인 density가 BB 등급에서 대체적으로 높게 나타나는데, 이는 4.3절에서 제시한 결과와 일관된 양상임을 확인할 수 있다. 따라서 본 연구에서 제안한 방안은 작의적 임계치 설정에 기반하고 있음에도 불구하고 충분히 안정적인 결과의 참조정보를 생성해 낼 수 있을 것으로 여겨진다.

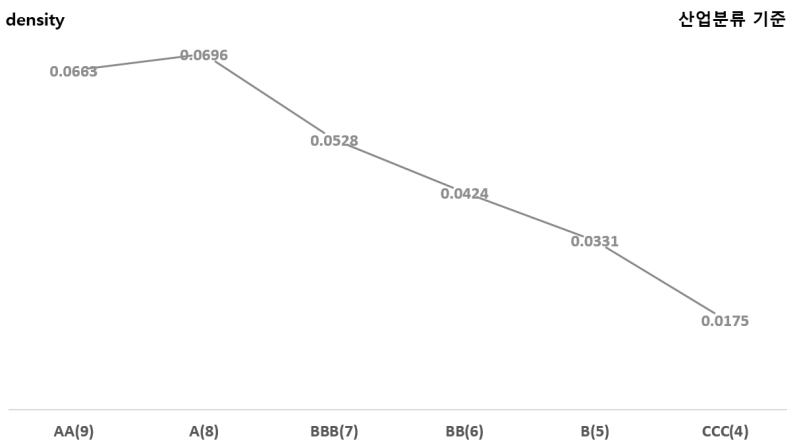
이미 앞에서 언급한 것처럼, peer group은 다양한 목적으로 다양한 방안에 따라 생성되고 이용되어 오고 있다. 기술평가의 영역에서도, 기술분류 또는 산업분류 간 유사성이라는 기준에 따라 peer group을 정의하고 이를 바탕으로 평가 참조정보를 생성하기 위한 방안이 활용되고 있다. 하지만 peer group 기반의 참조정보가 기술평가 지원을 목적으로 함에도 불구하고, 현재의 방안은 기술 자체가 지니는 속성을 전혀 고려하지 못한다는 한계를 지닌다.⁷¹⁾ 이러한 한계를 극복하고 평가 참조정보의 활용성을 증진시키기 위하여 본 연구에서는 기술적 속성 관점의 유사성을 통해 peer group을 정의하기 위한 방안을 제안하였다. 본 연구에서 제안한 방안의 타당성을 확인하기 위해서는 peer group을 정의하기 위해 활용되어 오고 있는 기존의 방안과 본 연구의 방안을 비교 분석하는 것이 필요하다. 이를 위해 기존의 방안에 따라 사례 분석의 평가 대상기업에 대한 peer group을 생성하고 평가등급별 density를 산출한다. <그림 7>과 <그림 8>에 나타난 것처럼 모두 AA 및 A 등급에서 가장 높은 수준의 density를 나타내고 있는데, 평가 대상기업의 평가결과가 BB 등급이었다는 점을 감안할 때, 이는 합리적인 결과라고 볼 수 없다. 따라서 하나의 기업을 대상으로 하는 사례분석에 비추어 볼 때, 본 연구에서 제안한 방법론은 기존 방안에 비해 높은 타당성을 지니고 있다고 판단된다.

본 연구에서 제안한 참조정보 생성방안 및 기존방안 모두 기술평가 결과의 예측을 목적으로 하지 않는다. 참조정보는 오직 평가 대상기업과의 유사도 수준 및 peer group 내 유사기업들의 기존 평가결과만을 활용하여 생성되며, 이는 새로운 평가를 위한 기준선 설정의 역할을 수행한다. 기술평가에 대한 최

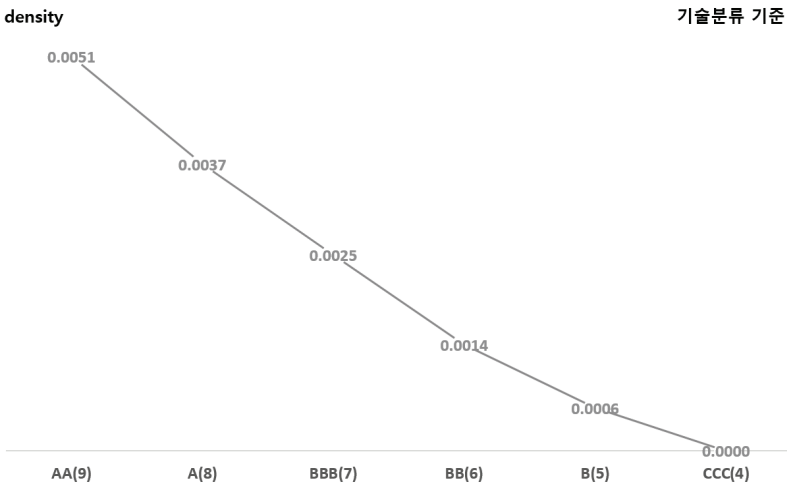
71) Kim, Eungchan et al., *op. cit.*, p.3200.

종결과를 획득하기 위해서는 다양한 요인들을 고려해야만 하기 때문에 본 연구의 결과물이 새로운 평가결과를 예측하기 위한 목적으로 활용될 수 없다. 다만, 본 연구에서는 제시한 방안에 따라 생성된 참조정보의 타당성 및 합리성을 살펴보기 위한 목적으로 최종 기술평가 결과와의 비교를 수행하고 있다.

〈그림 7〉 기존 방안인 산업분류 유사성에 따라 생성된 density 분포



〈그림 8〉 기존 방안인 기술분류 유사성에 따라 생성된 density 분포



VI. 결 론

성공적인 기술금융 생태계의 조성을 위해서는 명확한 기술평가 시스템을 갖추는 것이 매우 중요하다. 기술금융 실현을 위한 기술평가는 일반적으로 계량화된 데이터에 의한 정량적 평가와 해당 분야의 전문가에 의한 정성적 평가의 혼합으로 이루어지며, 정량적 평가를 위한 계량화된 모형은 전문가에 의한 정성적 평가를 지원하는 역할을 수행하기 때문에 정성적 평가 결과가 최종 기술평가 결과에 직접적으로 영향을 미치는 경향이 강하다. 하지만 정성적 평가는 평가를 수행하는 주체인 평가자의 개별적 역량 및 성향 차이에 따라 일관되지 못한 결과를 도출하게 되는데, 이는 기술평가 결과에 대한 신뢰성을 저하시키는 요인이 된다. 이와 같은 평가자 개인의 차이에 의한 평가결과의 편차를 보정하기 위하여 기술보증기금을 포함한 기술평가 기관들은 평가 대상기업과 속성이 유사한 기업들을 peer group으로 정의하고 이들의 평가 관련 정보를 평가 참조정보로서 제공하기 위한 노력을 시행하고 있다. 하지만 기존의 방안들은 주로 산업분류 또는 기술분류 간 유사성이라는 단순한 기준을 활용하고 있어, 기술 자체가 지니는 속성을 고려하지 못한다는 한계를 지니고 있으며, 이는 궁극적으로 평가 참조정보의 활용성 및 실효성을 저조하게 만드는 요인이 된다. 이러한 문제를 해결하기 위하여 본 연구는 기술내용 간 유사도 수준을 측정하여 평가 대상기업과 기술적 속성 관점에서 유사성을 지닌 기업들을 발굴하고 이들로부터 평가 참조정보를 생성하기 위한 방안을 제시하였다. 그리고 본 연구에서 제시하는 방법론의 실제적 활용성을 탐색하기 위하여 하나의 기업을 평가 대상기업으로 선정하고 이와 유사성을 지니는 peer group을 발굴한 후, 평가 참조정보를 생성하는 사례 분석을 시행하였다. 또한 본 방법론에 의해 산출된 결과와 기존의 peer group 정의를 위해 활용되는 방안에 의해 산출된 결과를 비교하고 본 방법론의 타당성에 대한 논의를 수행하였다. 본 연구의 방법론은 실용적인 평가 참조정보를 정량적 방안에 따라 제공함으로써 정성적으로 이루어지는 기술

평가 작업의 효율성을 향상시키는 데 기여할 수 있다. 또한 평가자의 개인적 역량 및 성향에 따른 기술평가 결과의 편차를 줄여 줌으로써 기술평가 작업의 신뢰성도 증진시킬 수 있을 것으로 기대된다.

이러한 기여점에도 불구하고 본 연구는 여전히 한계점을 지니고 있기 때문에 이에 따른 추후 연구방향을 제시하고자 한다. 우선, 본 연구는 기술내용만을 고려하여 peer group을 생성하였기 때문에 여전히 단편적이라는 한계를 지닌다. 보다 다양한 관점에서 평가 대상기업과 속성이 유사한 기업들을 발굴할 수 있도록 기술을 보유 및 활용하는 주체인 기업의 특성을 고려하여 유사성을 판단하기 위한 시도가 이루어져야만 한다. 이를 위해서는 기업 규모 등의 구조적 관점과 기업 간 거래 관계 등의 연계적 관점을 반영할 수 있는 계량적 기준을 설계하는 과정이 요구된다. 두 번째로 본 연구에서 제시하는 평가 참조정보는 peer group 내 기업들이 지닌 유사도 수준과 기존 평가결과의 결합만을 통해 산출되기 때문에 다양성이 부족한 한계를 지닌다. 기업 재무정보 등 보다 폭넓은 정보를 복합적으로 활용하여 평가 참조정보의 유용성을 증진시키기 위한 연구가 수행되어야 한다. 마지막으로 본 연구에서 제안하는 방법론의 활용성 및 타당성을 확인하기 위하여 하나의 기업을 대상으로 사례분석을 시행한 후, 기존의 peer group 생성 방안과 비교 분석하였다. 하지만 보다 엄밀한 의미의 타당성 탐색을 위해서는 보다 많은 사례들을 발굴하고 이들의 결과를 종합하여 설명력을 비교하기 위한 과정이 이루어져야만 한다. 많은 수의 사례들에 대해 본 연구의 방안과 기존 방안 간 설명력의 차이를 도출하고 이에 대한 유의성을 탐색하는 후속연구가 수행되어야 한다.

참고문헌

〈단행본(국내 및 동양)〉

김성민 외 8인, 『기술사업화의 이해와 적용』, 경문사, 2015.

여인국, 『기술사업화 이론과 실제』, 학현사, 2013.

〈학술지(국내 및 동양)〉

기흥도 외 5인, “스마트폰 어플리케이션 설치 목록을 이용한 사용자 특성 추론”, 『정보과학회논문지』, 제45권 제12호(2018).

김광희, “기술금융의 현황과 과제”, 『과학기술정책』, 제21권 제3호(2011).

김대철 외 2인, “거시경제변수를 고려한 기술평가모형의 개선: 기술보증기금의 사례”, 『경영과학』, 제30권 제2호(2013).

김도우 · 구명완, “Doc2Vec과 Word2Vec을 활용한 Convolutional Neural Network 기반 한국어 신문 기사 분류”, 『정보과학회논문지』, 제44권 제7호(2017).

김병식 · 신주현, “DBSCAN을 활용한 유의어 변환 문서 유사도 측정 방법”, 『멀티미디어학회논문지』, 제21권 제9호(2018).

김정미 · 이주홍, “Word2vec을 활용한 RNN기반의 문서 분류에 관한 연구”, 『한국지능시스템학회 논문지』, 제27권 제6호(2017).

김태완 · 윤재홍, “기술가치평가 방법론을 활용한 사례 분석”, 『경영논총』, 제33권(2012).

류승미 · 성태웅, “수자원 분야 공익형 기술가치평가 시스템에 대한 연구”, 『지능정보연구』, 제24권 제3호(2018).

박선영, “특허기술 평가요인에 관한 연구”, 『지식재산연구』, 제2권 제1호(2007).

박성수 · 이건창, “Word2vec과 앙상블 분류기를 사용한 효율적 한국어 감성 분류 방안”, 『한국디지털콘텐츠학회 논문지』, 제19권 제1호(2018).

서경 · 권영관, “IP금융의 새로운 패러다임 — 선진 IP금융 모델사례로부터의 시사점”, 『지식재산연구』, 제8권 제3호(2013).

성태웅 외 3인, “과학기술정보 콘텐츠의 가치평가모형 프레임워크 연구”, 『한국콘텐츠학회논문지』, 제16권 제11호(2016).

육지희 · 송민, “토픽모델링과 딥 러닝을 활용한 생의학 문헌 자동 분류 기법 연구”, 『정보관리학회지』, 제35권 제2호(2018).

이동훈 · 김관호, “Word2Vec 기반의 의미적 유사도를 고려한 웹사이트 키워드 선택 기법”, 『한국전자거래학회지』, 제23권 제2호(2018).

- 이상현 외 4인, “DEA와 MPI를 활용한 OECD 국가간 지식기반산업의 효율성 분석”, 『Entrue Journal of Information Technology』, 제13권 제3호(2014).
- 이성상, “지식재산 금융 활성화 방안에 관한 연구 — 미국과 이스라엘의 사례를 중심으로”, 『지식재산연구』, 제8권 제3호(2013).
- 정으뜸 외 2인, “기술적 파급효과 도출을 통한 기술융합 분석 연구: 나노바이오 융합 기술 사례를 중심으로”, 『지식재산연구』, 제10권 제4호(2015).
- 정지수 외 6인, “문서 유사도를 통한 관련 문서 분류 시스템 연구”, 『방송공학회논문지』, 제24권 제1호(2019).
- 채균식 · 이용봉, “연구중에 생산된 과학기술 참조데이터 관리에 관한 연구”, 『한국문헌정보학회지』, 제37권 제4호(2003).
- 최가람 · 최성필, “단어 임베딩(Word Embedding) 기법을 적용한 키워드 중심의 사회적 이슈 도출 연구”, 『정보관리학회지』, 제35권 제1호(2018).
- 한은진 외 3인, “Word2vec algorithm 기반 뉴스 기사에 나타난 성별 관련 어휘 규명 연구”, 『대한산업공학회지』, 제44권 제4호(2018).

〈학술지(서양)〉

- Azzone, Giovanni & Manzini, Raffaella, “Quick and dirty technology assessment: The case of an Italian Research Centre”, *Technological Forecasting and Social Change*, Vol.75 No.8(2008).
- Chiu, Yu-Jing & Chen, Yuh-Wen, “Using AHP in patent valuation”, *Mathematical and Computer Modelling*, Vol.46 No.7-8(2007).
- Choi, Sungchul et al., “Analyzing technological spillover effects between technology classes: The case of Korea technology finance corporation”, *IEEE Access*, Vol.6(2018).
- Facey, Karen et al., “Patients’ perspectives in health technology assessment: A route to robust evidence and fair deliberation”, *International Journal of Technology Assessment in Health Care*, Vol.26 No.3(2010).
- Garcia, Noa & Vogiatzis, George, “Learning non-metric visual similarity for image retrieval”, *Image and Vision Computing*, Vol.82(2019).
- Harris, Zellig, “Distributional structure”, *Word*, Vol.10 No.2-3(1954).
- Hernández Pardo, Ricardo et al., “Sustainable product service systems in small and medium enterprises (SMEs): Opportunities in the leather manufacturing industry”, *Sustainability*, Vol.4 No.2(2012).
- Hsu, David et al., “Toward successful commercialization of university technology:

- Performance drivers of university technology transfer in Taiwan”, *Technological Forecasting and Social Change*, Vol.92(2015).
- Ju, Yonghan & Sohn, So Young, “Technology credit scoring based on a quantification method”, *Sustainability*, Vol.9 No.6(2017).
- Jun, Sunghae et al., “A technology valuation model using quantitative patent analysis: A case study of technology transfer in Big Data marketing”, *Emerging Markets Finance and Trade*, Vol.51 No.5(2015).
- Kim, Donghwa et al., “Multi-co-training for document classification using various document representations: TF-IDF, LDA, and Doc2Vec”, *Information Sciences*, Vol.477(2019).
- Kim, Eungchan et al., “An approach to generating reference information for technology evaluation”, *Sustainability*, Vol.10 No.9(2018).
- Kim, Seo Kyun et al., “The effect of R&D, technology commercialization capabilities and innovation performance”, *Technological and Economic Development of Economy*, Vol.17 No.4(2011).
- Koyamada, Koji et al., “Visualization of JOV abstracts”, *Journal of Visualization*, Vol.21 No.2(2018).
- Lee, Younghoon et al., “Applying convolution filter to matrix of word-clustering based document representation”, *Neurocomputing*, Vol.315(2018).
- Noh, Heeyong et al., “How to improve a technology evaluation model: A data-driven approach”, *Technovation*, Vol.72-73(2018).
- Perkmann, Markus et al., “How should firms evaluate success in university-industry alliances? A performance measurement system”, *R&D Management*, Vol.41 No.2(2011).
- Roure, Juan & Keeley, Robert, “Predictors of success in new technology based ventures”, *Journal of Business Venturing*, Vol.5 No.4(1990).
- Shen, Yung-Chi et al., “A hybrid selection model for emerging technology”, *Technological Forecasting and Social Change*, Vol.77 No.1(2010).
- Sohangir, Sahar et al., “Big Data: deep learning for financial sentiment analysis”, *Journal of Big Data*, Vol.5 No.3(2018).
- Sohn, So Young et al., “Improved technology scoring model for credit guarantee fund”, *Expert Systems with Applications*, Vol.28 No.2(2005).
- Turney, Peter & Pantel, Patrick, “From frequency to meaning: Vector space models

- of semantics”, *Journal of Artificial Intelligence Research*, Vol.37(2010).
- Yu, Peng & Lee, Jang Hee, “A hybrid approach using two-level SOM and combined AHP rating and AHP/DEA-AR method for selecting optimal promising emerging technology”, *Expert Systems with Applications*, Vol.40 No.1(2013).
- Zhang, Dongwen et al., “Chinese comments sentiment classification based on word2vec and SVMperf”, *Expert Systems with Applications*, Vol.42 No.4(2015).

〈인터넷 자료〉

- KoNLPy, “KoNLPy: Korean NLP in Python”, (<http://konlpy.org/en/latest/>), 작성일: 2019. 5. 23.

〈기타 자료〉

- 김선혜 · 윤병운, “SAO2vec: SAO(Subject-Action-Object) 구조와 Doc2vec을 이용한 SAO 임베딩 알고리즘 개발”, 『대한산업공학회 학술대회논문집』, 경주, 대한민국, 4월 4-7일(2018).
- 김재영 외 2인, “Deep Embedding Clustering 기술을 이용한 특허 문서의 비지도 군집화”, 『대한산업공학회 학술대회논문집』, 경주, 대한민국, 4월 4-7일(2018).
- Le, Quoc & Mikolov, Tomas, “Distributed Representations of Sentences and Documents”, In *Proceedings of the 31st International Conference on Machine Learning*, ICML, 2014.
- Liu, Yao-Jen & Liu, Shang-Jyh, “The intellectual property policy of Taiwan: A strategic viewpoint”, In *Proceedings of IEEE International Engineering Management Conference*, IEEE, 2004.
- Ruskov, Petko et al., “Model for technology adoption supporting system”, In *Proceedings of International Scientific Conference*, UNITECH, 2016.
- Xing, Chao et al., “Document classification with distributions of word vectors”, In *Proceedings of Signal and Information Processing Association Annual Summit and Conference*, 2014 APSIPA, 2014.

A reference information generation method for technology evaluation from the perspective of similarity of technological features

Choi Sungchul*, Seo Wonchul**

Technology evaluation is generally recognized as a basis for realizing technology finance which refers to providing financial support based on the evaluation of technology innovation and commercialization capability of firms. It is naturally performed in a combination of quantitative evaluation by quantified data and qualitative evaluation by experts in relevant field. Among the both, the results from qualitative evaluation tend to directly affect the ultimate evaluation results. To facilitate the qualitative work, various institutes related to technology evaluation have tried to identify several firms that are similar to the target firm and generate reference information for the evaluation of target firm's technology. However, the existing approaches have limitations in that they cannot reflect the features of technology itself, only considering the similarity in terms of the industrial or technological classification. This study proposes a methodology for identifying similar firms to the target firm in terms of technological features by measuring the similarity their

* Assistant Professor, Department of Industrial and Management Engineering, Gachon University(First Author).

** Associate Professor, Graduate School of Management of Technology, Pukyong National University(Corresponding Author).

technological contents and generating reference information from the information of the similar firms. To measure the similarity between technological context, we apply a doc2vec method which is one of the most representative deep-learning based models to represent textual document information to vector. This study can contribute to improving the efficiency and reliability of qualitative technology evaluation work by providing practical reference information in a quantitative way.

Keyword

Technology evaluation, Reference information, Peer group, Quantification of technological features, Similarity in technological contents