

텍스트 마이닝을 이용한 주제기반의 기업인 네트워크 계층 분석

Topic Based Hierarchical Network Analysis for Entrepreneur Using Text Mining

이동훈(Donghun Lee)*, 김용화(Yonghwa Kim)**, 김관호(Kwanho Kim)***

초 록

다양한 고객의 요구를 만족시키기 위한 신제품 설계 및 개발의 필요성 때문에 중소기업 간의 융합 활동의 중요성은 증대하고 있다. 특히, 최고 의사결정을 가지는 중소기업 대표는 적합한 융합 활동 파트너를 구하기 위해 인맥관리는 필수적이다. 한편 기업인들은 많은 양의 인맥을 형성하는 것이 중요할 뿐만 아니라 유사한 토픽정보를 가진 기업인과의 인맥관계를 이해하는 것이 중요하다. 그러나 중소기업의 현황 부재와 산업분야별 기업인들의 기술과 특성을 나타낼 수 있는 토픽정보를 수집하는데 어려운 한계가 존재한다. 본 논문에서는 토픽 추출 기법을 통해 이와 같은 문제점을 해결하고 3가지 측면에서 기업 네트워크를 분석한다. 구체적으로 *C, S, T-Layer* 모델이 있으며 각각의 모델은 인맥의 양, 인맥 중심성, 토픽 유사성을 분석한다. 실 데이터를 통한 실험 결과, 인맥의 양이 적은 경우 중심성이 높은 기업과 네트워크를 강화하여 인맥 네트워크를 활성화 시켜야 할 필요가 있고, 토픽 유사성이 낮은 경우 주제 기반의 네트워크를 활성화 시켜야 할 필요가 있다는 것을 실험을 통해 확인하였다.

ABSTRACT

The importance of convergence activities among business is increasing due to the necessity of designing and developing new products to satisfy various customers' needs. In particular, decision makers such as CEOs are required to participate in networks between entrepreneurs for being connected with valuable convergence partners. Moreover, it is important for entrepreneurs not only to make a large number of network connections, but also to understand the networking relationship with entrepreneurs with similar topic information. However, there is a difficult limit in collecting the topic information that can show the lack of current status of business and the technology and characteristics of entrepreneur in industry sector. In this paper, we solve these problems through the topic extraction method

이 논문은 2017년도 정부(미래창조과학부)의 재원으로 한국연구재단의 개인기초연구사업(No. NRF-2017-R1D1A1B03035639)과 2018년도 정부(과학기술정보통신부)의 재원으로 한국연구재단-현장맞춤형 이공계 인재양성 지원사업의 지원을 받아 수행된 연구임(NRF-2017H1D8A1029391).

* First Author, Dept. of Industrial and Management Engineering, Incheon National University(dhlee@inu.ac.kr)

** Co-Author, Dept. of Industrial and Management Engineering, Incheon National University (kingstwo41@gmail.com)

*** Corresponding Author, Dept. of Industrial and Management Engineering, Incheon National University (khokim@inu.ac.kr)

Received: 2018-05-24, Review completed: 2018-06-29, Accepted: 2018-07-14

and analyze the business network in three aspects. Specifically, there are *C, S, T-Layer* models, and each model analyzes amount of entrepreneurs relationship, network centrality, and topic similarity.

As a result of experiments using real data, entrepreneur need to activate network by connecting high centrality entrepreneur when the corporate relationship is low. In addition, we confirmed through experiments that there is a need to activate the topic-based network when topic similarity is low between entrepreneurs.

키워드 : 토픽 추출, 소셜 네트워크 분석, 텍스트 마이닝, Word2Vec, Color Embedding
Topic Extraction, Social Network Analysis, Text Mining, Word2Vec, Color Embedding

1. 서 론

다양한 고객의 요구를 만족시키기 위한 신제품 설계 및 개발의 필요성 때문에 중소기업간의 융합 활동의 중요성은 증대하고 있어 중소기업인들의 인맥관리는 필수적이다. 중소기업은 정보화 발전에 따른 급격한 기술혁신 확산 등의 요인으로 인하여 산업의 성장 정체를 겪고 있다. 게다가 새로운 사업 분야 진출을 위한 신기술 개발에 있어 전문 기술 부족, 인력 부족 등의 어려움을 가지고 있다. 대부분의 중소기업들은 이를 해결하기 위해 적합한 사업 파트너를 구하는데 어려움을 겪으므로 중소기업인들 간의 인맥을 미리 형성하는 것은 중요하다.

한편 중소기업인들은 인맥 형성 시 기업인들이 보유하고 있는 기술이나 유사한 토픽정보를 가지고 있는 기업인과 인맥관계를 이해하고 형성하는 것이 중요하다. 예를 들어, A기업은 반도체 설비 및 정밀 가공하는 업체이고, B기업은 사무용품 및 회화용품을 제조하는 업체, C기업은 디스플레이, 모니터 제조업체일 때 A, B 두 기업은 자사가 보유하고 있는 기술의 특성이 다르기 때문에 단위 기술이 융합하여 새로운 특성의 기술 및 가치를 창출하는데 한계가 있다. 반면에 A, C 두 기업은 보유하고 있는 기술의 특성이 유사하므로 융합 활동을 기반으로 새로운 경영

전략을 형성할 수 있고 새로운 산업 및 분야에 진출하거나 새로운 특성의 기술을 창출하는데 큰 도움이 될 수 있다. 따라서 중소기업인들은 인맥 형성 시 많은 양의 인맥을 형성하는 것이 중요할 뿐만 아니라 유사한 토픽정보를 가진 기업인과의 인맥을 형성하는 것이 중요하다.

그러나 중소기업의 현황 정보 부재로 인하여 중소기업인 간의 융합 활동을 하는데 어려움이 존재한다. 우리나라 영리기업은 537만 7천 개이며 이 중 중소기업은 537만 3천 개(99.9%)인 것으로 분석되었다[7]. 대부분의 기업이 중소기업임에도 불구하고 지역별에 따른 업종별 수, 종사자의 수와 같은 정보만을 알 수 있으며 3년 이전 제한된 자료만을 알 수 있다. 게다가 국가에서 제공하는 중소기업현황 정보시스템은 중소기업의 기술의 특성이 자주 바뀔에도 불구하고 사용자가 변경 하지 않을 시 변경된 사항을 파악하지 못하므로 중소기업이 자체적으로 중소기업의 관련 주제 및 융·복합 가능 기업 정보를 확보 할 수 없다. 그러므로 각 기업인들이 보유하고 있는 기술과 특성을 나타낼 수 있는 토픽정보를 수집하는데 어려운 한계가 존재한다.

이와 같은 문제를 해결하기 위해 융합 활동 활성화와 중요성에 대해서 분석하는 많은 연구가 진행되고 있다. 먼저 기업차원에서 타 산업과의 전략적 연계를 통한 산업간 융합 활동

중요성에 대해 연구 하였고[12] 특허정보에 있는 Nutraceuticals and Functional Foods(NFF) 기술과 IPC 공동 분류 정보를 활용하여 기술 융합 패턴을 분석하였다[9]. 또한 분석 단위와 계량서지 분석 지표를 활용하여 네트워크 분석을 통해 학문적 융합 활동에 대해 연구하거나 토픽 모델링 기법을 통하여 기술경영 분야의 연구동향을 분석하였다[13, 19]. 그러나 기존연구에서는 기업인들의 인맥 및 토픽정보를 기반으로 분석하는 연구는 존재하지 않았으며 네트워크 분석과 특허 정보를 통한 융합지표는 기업인의 인맥 관계 데이터에 국한되어 분석하는 한계점과 기업이 보유하고 있는 특허 기술이 없을 경우 토픽정보를 파악하기 어려운 한계가 존재한다.

이를 해결하기 위해 본 논문에서는 중소 기업이 활동하고 있는 웹사이트에서 토픽정보 추출 기법을 기반으로 중소 기업이 현재 보유하고 특성을 나타낼 수 있는 주제 정보를 형성한다. 따라서 현재 기업이 보유하고 있는 특성을 반영하지 못했던 점을 해결하고 각 기

업인이 보유하고 있는 토픽정보를 융합지표로 설정하여 시각화를 통해 융합 활동 활성화를 위한 주제기반의 기업인 네트워크 계층 분석 기법을 제안한다. 기업과 기업인을 같은 의미로 해석할 수 있으나 본 논문에서는 기업 활동에서 중요한 기업인이 가지고 있는 인맥과 토픽정보를 중심으로 분석한다. 따라서 포괄적인 의미로써 기업인으로 표현한다. 본 논문의 구성은 다음과 같다. 제 2장에서 관련연구에 대해 설명하고, 제 3장에서 제안한 모델에 대해 내용을 설명한다. 제 4장에서는 제안한 네트워크 분석 모델에 대한 성능을 평가하고 마지막으로 제 5장에서 결론을 기술한다.

2. 관련 연구

기존연구는 학문적 융합, 산업적 융합 등으로 융합 양상과 융합 활동에 관하여 <Table 1>과 같이 연구되었다. 먼저 학문적 융합과 관련된

<Table 1> The Overview of Previous Research on Convergence Activities

Features Considered	Analysis Methods	Research Direction	References
Academic convergence	Network analysis	Academic convergence activities	Jeong and Lee[13]
		Identification the structural characteristics of convergence activities	Heo and Yang[11]
	Latent dirichlet allocation	Technology humanities integration	Lee and Kang[19]
	Survey		Kim et al.[16]
Industrial convergence	Patent information analysis (NFF, IPC)	Technology convergence pattern analysis	Kim et al.[3], Hacklin et al.[9]
	Network analysis	Industry convergence at the enterprise level	Kim et al.[3], Jeong and Lee[12]
	Patent citation information, Keyword analysis	Technology convergence analysis	Karvonen and Kässi[14], Kim and Lee[17]
			Han and Sohn[10], Kim et al.[15]
	Patent citation information, Network analysis		
	ICT field information analysis	Technological convergence taking place in The ICT	Borés et al.[4], Choi[6]
Value chains analysis	Industrial structure convergence	Greenstein and Khanna[7], Lei[19], Wirtz[24]	

연구를 살펴보면, 벤처협회에 등록된 중소기업의 설문조사를 통해 중소기업 기술 인문융합의 정의 및 범위를 제시하였고[15], 첨단융합기술개발사업을 중심으로 네트워크 분석기법을 적용하여 융합연구의 구조적 특성을 파악하고자 하였다[11].

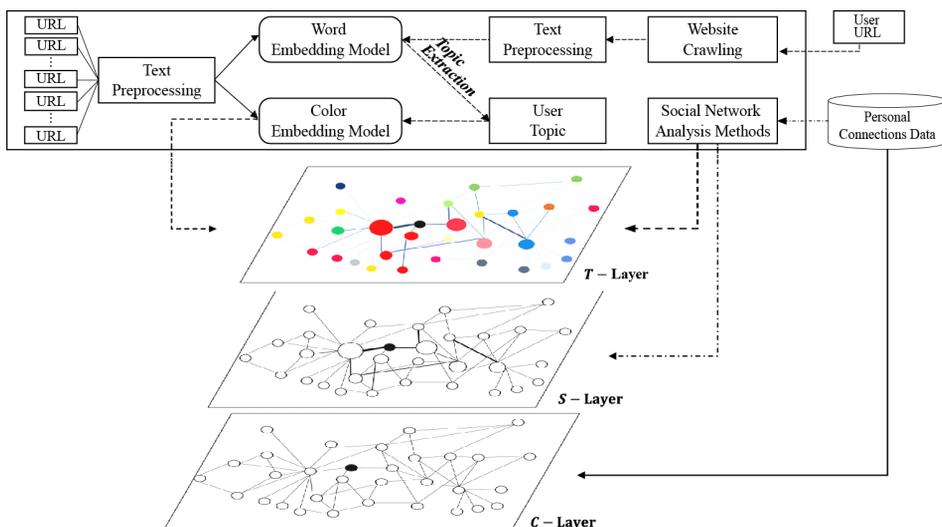
다음으로 산업적 융합 측면에서 살펴보면, IT 산업기술을 중심으로 특정 기술개발 및 기술경영을 위한 산업적 융합 분석에 대해 연구하였으며 [4, 6] 가치사슬 분석을 통한 기술과 산업구조 관점에서의 융합 분석에 대해 연구하였다[20, 24].

또한, 특허 간의 인용 정보와 키워드를 활용하여 특정 산업에 관한 융합 기술의 변화와 기술융합 현황분석 대한 연구[14, 17]와 특허 인용 정보 뿐만 아니라 네트워크 분석 기법을 활용하여 융합 기술 내에서의 핵심 기술과 융합 양상을 파악하거나 ICT 관련 표준 기술들의 융합 가능성과 연관 규칙 분석을 통해 구체적으로 융합 양상을 파악하고자 하였다[10, 15]. 하지만 위의 기존 연구들에서는 기업의 특성을 이용한 연구는 존재하지 않았으며 특허정보와 기존의

네트워크 분석기법을 활용하여 기술의 융합 양상과 융합 활동 분석에 초점을 두고 있다. 따라서 본 연구에서는 기업인의 특성을 나타낼 수 있는 정보를 기업인이 활동하는 웹사이트에서 토픽정보를 자동으로 추출하여 기업인의 특성을 반영하지 못했던 문제점을 해결하고 기업 융합 활동 활성화를 위한 주제정보 기반의 기업 네트워크 계층 분석 기법을 제안한다.

3. 제안 기법

제 3장에서는 본 연구에서 제안한 시스템을 크게 3단계로 설명한다. <Figure 1>과 같이 *C-Layer*(Corporate Connection Layer) 모델, *S-Layer*(Social Influence Layer) 모델, *T-Layer*(Topic Relevance Layer) 모델에 대해 계층적 네트워크를 분석하고 평가한다. 첫 번째, *C-Layer* 모델은 기업인의 인맥 네트워크를 분석하는 모델이다. 기업인의 인맥정보 데이터



<Figure 1> The Proposed Framework for Entrepreneur Hierarchical Network Analysis

를 이용하여 기업인의 직접적으로 알고 있는 인맥(1hop) 뿐만 1hop의 인맥(2hop), 2hop의 인맥(3hop)까지의 정보를 이용하여 네트워크를 분석한다. 두 번째, *S-Layer* 모델은 인맥 네트워크뿐만 아니라 인맥관계의 허브의 역할을 할 수 있는 영향력자를 소셜 네트워크 분석 기법을 통해 분석한다. 마지막으로 *T-Layer* 모델은 기업인의 토픽분석을 기반으로 토픽정보를 색으로 표현하고 시각화를 통해 네트워크를 분석한다.

3.1 *C-Layer* 모델 기반의 네트워크 분석

C-Layer 모델은 사용자 인맥정보 기반으로 인맥의 양적인 측면을 분석하는 모델이다. 네트워크를 분석하기 위해 임의의 기업인 u_i , $i = 1, \dots, n$ 에 대해 인맥에 해당하는 기업인 $u_{j,k}$, $j = 1, \dots, m, k = 1, \dots, K$ 의 인맥정보를 구축한다. $u_{j,k}$ 는 u_i 이웃 중 k -hop에 해당하는 기업인을 나타낸다. 따라서 기업인 $u_{1,1}, u_{2,1}, \dots, u_{j,1}$ 는 1hop으로 정의하고 기업인 $u_{1,1}, u_{2,1}, \dots, u_{j,1}$ 를 2hop으로 정의한다. 이와 같은 방법으로 인맥 네트워크를 구축하여 기업인 네트워크를 분석한다. 식 (1)은 기업인 u_i 에 대한 인맥의 양을 계산하는 식을 나타낸다.

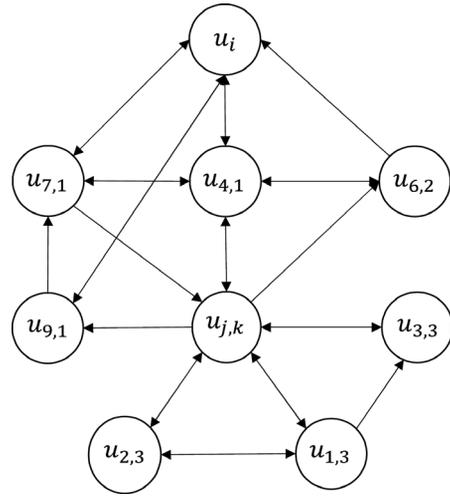
$$\sigma_C(u_i) = \frac{\sum_{j=1}^m \sum_{k=1}^K count(u_i, u_{j,k})}{\max_j \left(\sum_{j=1}^m \sum_{k=1}^K count(u_j, u_{j,k}) \right)} \quad (1)$$

여기에서 k 는 hop의 정보를 나타내고 j' 은 j 중 인맥을 가장 많이 가지고 있는 기업인을 나타낸다. $count(u_i, u_{j,k})$ 는 u_i 에 대한 K 까지의 인맥에 해당하는 기업인의 수 계산을 나타낸다.

$\sum_{j=1}^m \sum_{k=1}^K count(u_i, u_{j,k})$ 는 u_i 에 대한 K 까지의 인맥

기업인의 수의 총합, $\max_j \left(\sum_{j=1}^m \sum_{k=1}^K count(u_j, u_{j,k}) \right)$

는 전체 기업인 중 인맥을 가장 많이 가지고 있는 기업인의 인맥 수를 나타낸다.



<Figure 2> The Network Structure Obtained Using *C-Layer* Model

<Figure 2>는 *C-Layer* 모델의 예를 나타내며, 이를 기반으로 식 (1)을 통해 네트워크를 분석한다.

3.2 *S-Layer* 모델 기반의 네트워크 분석

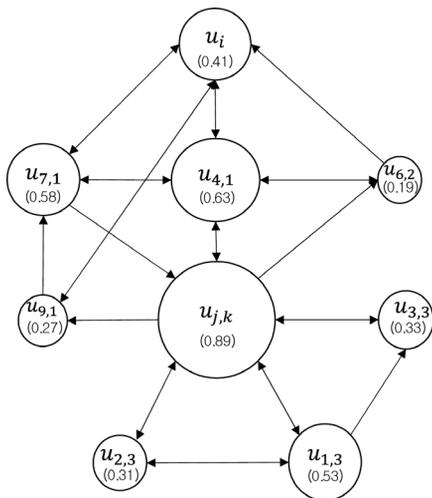
S-Layer 모델은 인맥 네트워크뿐만 아니라 인맥관계의 허브의 역할을 할 수 있는 영향력자를 분석하는 모델이다. 소셜 네트워크 기법 중 근접 중심성(Closeness Centrality) 기법을 이용하였으며 *C-Layer* 모델 기반으로 모든 개체 간의 최단거리(hops)계산을 통해 중심성을 측정한다. 중심성이 클 경우 네트워크 내에서 영향력자 또는 키 플레이어로서 기업인간의 허브의 역할을 할 수 있다[1, 3, 22].

식 (2)는 기업인 u_i 에 대한 S -Layer 모델의 중심성 측정 계산을 나타낸다.

$$\sigma_S(u_i) = \frac{\sum_{j=1}^m \sum_{k=1}^K \frac{1}{\text{hop}(u_i, u_{j,k})}}{\max_j \left(\sum_{j=1}^m \sum_{k=1}^K \frac{1}{\text{hop}(u_j, u_{j,k})} \right)} \quad (2)$$

방향성이 있는 네트워크일 경우, 거리의 합의 역수를 취하는 대신에 역수를 합한 방법을 사용한다[21]. 여기에서 $\sum_{j=1}^m \sum_{k=1}^K \frac{1}{\text{hop}(u_i, u_{j,k})}$ 는 u_i 와 u_j 의 최단 거리(hop) 역수의 총합을 나타내고, $\max_j \left(\sum_{j=1}^m \sum_{k=1}^K \frac{1}{\text{hop}(u_j, u_{j,k})} \right)$ 는 전체 기업인 중 거리(hop) 역수 총합이 가장 높은 값을 나타낸다.

<Figure 3>은 S -Layer 모델의 예시를 나타낸다. 원의 크기는 식 (2)의 결과 값이 클수록 크며 인맥의 허브의 역할을 할 수 있는 영향력 자라고 분석 할 수 있다.



<Figure 3> The Network Structure Obtained using S -Layer Model (The value in a node for each user represents that user's centrality)

3.3 T-Layer 모델 기반의 네트워크 분석

T -Layer 모델은 인맥 영향력자 뿐만 아니라 기업인의 업종별 특성 및 토픽 분석을 기반으로 시각화를 통해 토픽 유사성을 분석하는 모델이다. 먼저 기업인의 토픽 유사성을 분석하기 위해 word embedding 모델을 구축하였다. Word embedding 모델은 word2vec 모델 중 skip-gram 모델을 사용하였다. Skip-gram 모델은 명사 텍스트 데이터를 이용하여 파라미터 iteration, windows size, dimension 각각을 a, w, v 로 설정하여 학습한다[18, 21]. 다음으로 토픽정보를 시각화하여 가시화 하기위해 color embedding 모델을 구축하였다. Color embedding 모델은 word embedding 모델과 동일한 텍스트 데이터를 이용하여 구축하였으며, 파라미터 iteration, windows size, dimension 각각을 a^*, w^*, v^* 로 다르게 설정하여 학습하였다. 기업인 u_i 의 토픽정보는 $u_{i,t}, t=1, \dots, N$, 토픽 벡터는 $\vec{u}_{i,t}$, 토픽의 컬러 벡터는 $\sigma_r(u_{i,t})$, 업종별 특성 정보는 $I_d, d=1, \dots, D$, 업종의 특성 정보의 벡터는 \vec{I}_d 로 정의한다. 예를 들어, 토픽정보는 ‘블라인드, 프린트, 소프트웨어’ 등이 있고 업종별 특성 정보는 ‘기계장비, 숙박업, 교육’ 등이 있다. 식 (3)은 기업인 u_i 를 나타내는 토픽정보에 대한 토픽 컬러 벡터의 계산을 나타낸다.

$$\sigma_V(u_i) = \text{round} \left(\frac{255 \cdot \left(\sum_{t=1}^N \sigma_r(u_{i,t}) + 1 \right)}{2} \right) \quad (3)$$

여기에서 round 는 반올림을 나타내고 $\sum_{t=1}^N \sigma_r(u_{i,t})$ 의 벡터 값의 범위는 -1에서 1사이 값이다.

따라서 식 (3)의 계산은 RGB 값으로 표현하기 위한 계산을 나타낸다.

식 (4)는 기업인 간의 토픽 유사성 계산을 나타낸다.

$$\sigma_T(u_i, u_{j,k}) = \text{sim} \frac{\sum_{t=1}^N \frac{\vec{u}_{i,t} \cdot \vec{u}_{j,k,t}}{\|\vec{u}_{i,t}\| \|\vec{u}_{j,k,t}\|}}{\frac{\sum_{t=1}^N \vec{u}_{i,t} \times \vec{u}_{j,k,t}}{\sqrt{\sum_{t=1}^N (\vec{u}_{i,t})^2} \times \sqrt{\sum_{t=1}^N (\vec{u}_{j,k,t})^2}}} \quad (4)$$

여기에서 분자는 벡터의 내적을 뜻하며, 분모는 두 벡터의 크기를 각각 구해서 곱한 값을 뜻한다.

식 (5)는 식 (4)를 기반으로 u_i 에 대한 토픽 유사성 점수 계산을 나타낸다.

$$\sigma_O(u_i) = \frac{\sum_{j=1}^m \sigma_T(u_i, u_{j,k})}{\max_j \left(\sum_{j=1}^m \sigma_T(u_j, u_{j,k}) \right)} \quad (5)$$

여기에서 $\sum_{j=1}^m \sigma_T(u_i, u_j)$ 는 u_i 에 대한 토픽 유사성 점수의 총합을 나타내고, $\max_j \left(\sum_{j=1}^m \sigma_T(u_j, u_j) \right)$ 는 전체 기업인 중 가장 높은 토픽 유사성 값을 나타낸다.

식 (6)은 기업인의 업종별 특성을 파악하기 위한 계산을 나타낸다.

$$u_{i,d} = \text{Sim} \frac{\sum_{d=1}^D \sum_{t=1}^N \frac{\vec{u}_{i,t} \cdot \vec{I}_d}{\|\vec{u}_{i,t}\| \|\vec{I}_d\|}}{\frac{\sum_{d=1}^D \sum_{t=1}^N \vec{u}_{i,t} \times \vec{I}_d}{\sqrt{\sum_{t=1}^N (\vec{u}_{i,t})^2} \times \sqrt{\sum_{d=1}^D (\vec{I}_d)^2}}} \quad (6)$$

여기에서 분자는 식 (4)와 같이 벡터의 내적과 두 벡터 크기의 곱셈을 뜻 하며, 가장 높은 값을 갖는 업종정보를 기업인 u_i 의 업종정보라고 할 수 있다.

식 (3)은 토픽정보를 컬러로 표현하여 시각화를 통한 네트워크 분석에 활용되고 식 (4), 식 (5)는 기업인 간의 주제 네트워크 형성 정도를 파악하는데 활용된다. 마지막으로 식 (6)은 업종별로 특성을 파악하는데 활용된다.

4. 실험 및 평가

4.1 실험 환경

먼저 본 연구에서 제안한 *C-Layer* 모델 기반으로 네트워크를 분석하기 위해 인맥정보를 사전에 수집하였다. *T-Layer* 모델에서 기업인 토픽정보를 기반으로 기업인 간의 토픽 유사성을 분석하기 위해 Word Embedding 모델을 구축하였다. 데이터는 <Table 2>과 같이 한국 위키피디아, 나무위키, 중소기업뉴스, 중소기업 웹사이트 문서를 이용하였으며 1,392,523개의 문서에서 텍스트 데이터를 수집하였다. 수집된 텍스트 데이터들은 사용자 사전을 포함한 형태소 분석기를 통해 명사 데이터만을 추출하였다. 추출된 명사 데이터는 a, w, v 각각은 5, 10, 250으로 설정하여 학습하였다. 토픽정보를 RGB 컬러로 가시화하기 위한 Color Embedding 모델은 Word Embedding 모델을 구축하기 위해 사용 했던 텍스트 데이터와 동일한 데이터를 사용하였으며 a^*, w^*, v^* 각각은 5, 10, 3으로 설정하여 학습하였다.

<Table 2> Datasets for Learning Color and Word Embedding Models

Datasets	Numbers
Wikipedia	395,937
Namu Wiki	501,686
business news	44,900
Web documents	450,000
Total	1,392,523

기업인의 인맥 네트워크에서 hop의 수를 나타내는 K 는 3hop보다 큰 hop에 해당하는 인맥은 연관성이 적다고 판단하고 3으로 설정하였다. 기업인의 업종별 특성 정보 I_d 는 표준 한국 산업 분류표 정보를 기반으로 <Table 3>과 같이 30종의 업종정보를 사용하였다. 기업인의 토픽정보 $u_{i,t}$ 는 기업이 활동하고 있는 웹사이트 문서에서 토픽 추출 기법[18]과 Bizok 웹사이트에서 제공하는 토픽정보를 사용하였다[2]. 기업인을 나타내는 토픽정보의 개수는 토픽 추출기법에 따라 추출된 결과 값과 Bizok 웹사이트에서 제공하는 토픽정보에 따라 개수가 다를 수 있다.

<Table 3> Top10 Industrial Fields based on the Numbers of Data Collected

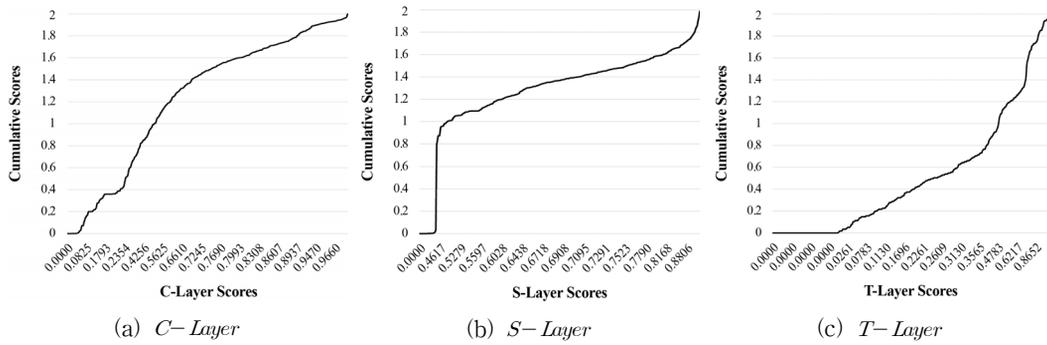
1. Precision equipment	6. Professional services
2. Machine equipment	7. Metal
3. Wholesale · Retail	8. Minerals
4. Textile · Leather	9. Business support · Facility management
5. Financial · Insurance	10. Minerals

본 연구에서 제안하는 모델들의 성능을 평가하기 위해 서울, 경기, 부천, 인천의 기업인 u_n, u_m 을 3515명 중 605명을 무작위로 선택하여 실험데이터로 설정하였다. 본 연구에서 제시한 모델들에 따라 3가지 측면에서 기업인 네트워크를 분석

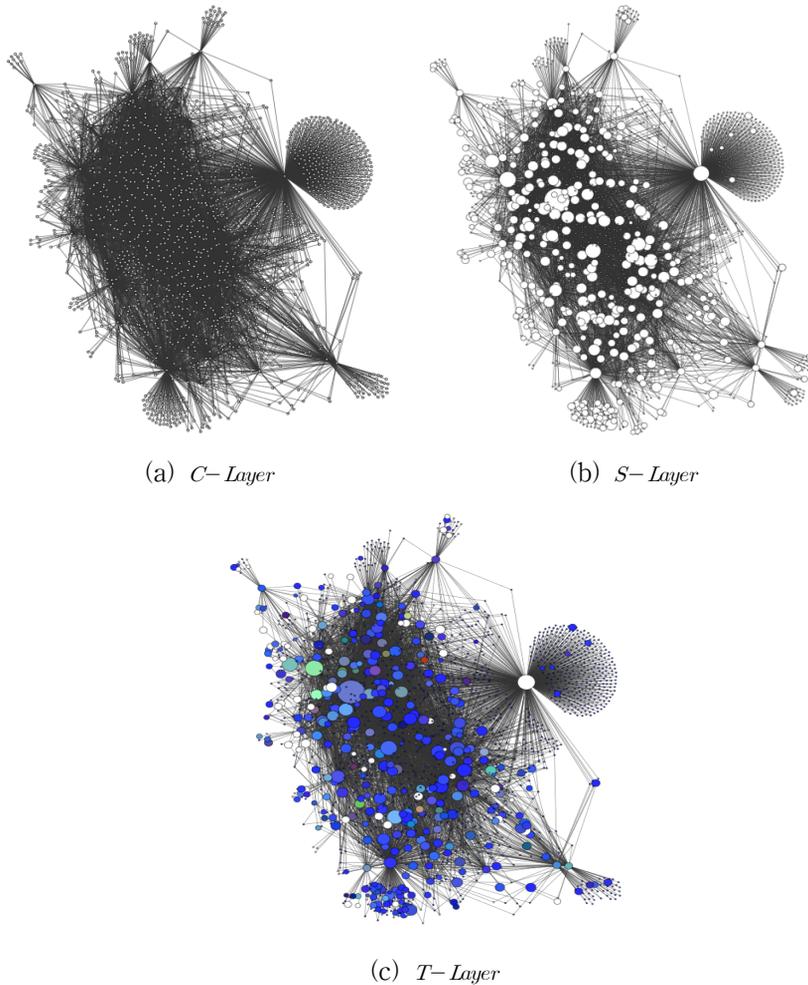
한다. 첫 번째, 제시한 모델들에 따라 전체 기업인들에 대한 네트워크 분석과 그 중 각각의 모델에 대한 차이점을 가지고 있는 기업인에 대해 네트워크를 분석한다. 두 번째, 제시한 모델을 기반으로 비율에 따른 기업인들의 네트워크 점수 변화 및 분포를 분석한다. 마지막으로 업종별 특성과 토픽분석을 기반으로 토픽정보를 RGB 컬러로 표현하여 업종에 따라 $S, T-Layer$ 모델 기반으로 분석한 결과를 2차원 공간에 시각화하여 네트워크를 분석하고 평가한다.

4.2 실험 평가 및 분석

<Figure 4>는 $C, S, T-Layer$ 각각의 모델 점수에 따른 누적분포를 나타낸다. 먼저 $C-Layer$ 모델에서는 점수가 증가할수록 누적 분포가 완만하게 증가하고 $S-Layer$ 모델에서는 낮은 점수에서는 급격하게 증가하고 점수가 증가할수록 완만하게 증가하다가 높은 점수에서 다시 급격하게 증가하였다. 반면에 $T-Layer$ 모델에 대해서는 누적 분포가 급격하게 증가하였다. 이를 통해 $C-Layer$ 모델에 대해서는 높은 점수에 형성되어 있는 기업인이 많아 인맥을 많이 형성하고 있는 많다는 것을 알 수 있었다. 그러나 그 중 $S-Layer$ 모델에 대해 높은 점수에 형성되어 있는 소수였으므로 허브의 역할을 할 수 있는 기업인은 소수라는 것을 실험을 통해 확인하였다. 또한 $C, S-Layer$ 모델 점수에 비해 $T-Layer$ 모델에 대한 점수가 낮게 형성되어 있는 기업인은 대다수였고 소수의 기업인들이 높은 점수에 형성되어 있었다. 따라서 인맥을 많이 형성하고 있는 기업인에 비해 유사한 토픽정보를 가지고 있는 기업인과의 네트워크를 잘 형성하고 있는 기업인은 소수라는 것을 실험을 통해 확인하였다.



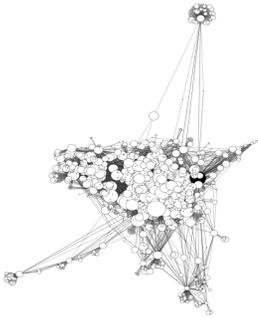
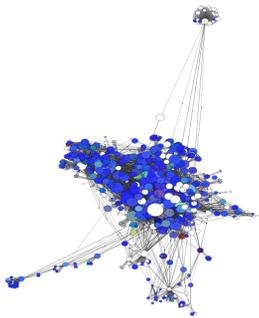
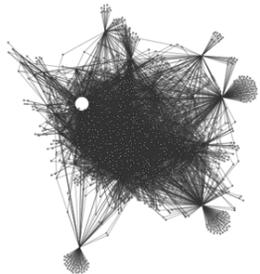
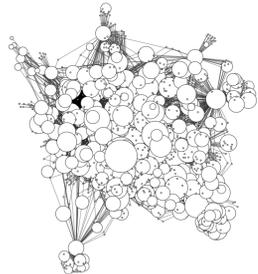
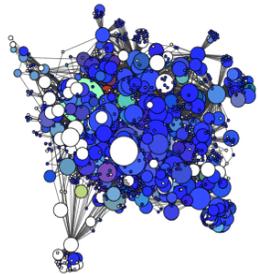
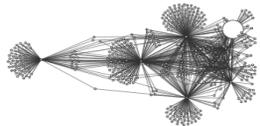
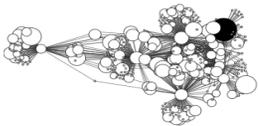
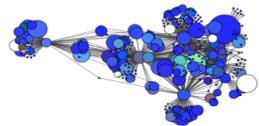
〈Figure 4〉 The Results of Cumulative Distribution According to Proposed Models



〈Figure 5〉 Network Analysis Results According to Proposed Models

<Figure 5>는 제시한 모델들에 따른 전체 네트워크 시각화 분석 결과를 나타내며, <Figure 6>은 그 중 모델에 대한 차이점을 가지고 있는 기업인에 대해 분석한 결과를 나타낸다. 먼저 <Figure 5>를 보면 *C-Layer* 모델에서는 전체 기업인 간의 hops에 따른 인맥 형성 네트워크 분석 결과를 나타낸다. *S-Layer* 모델에서는 인맥 네트워크뿐만 아니라 전체 기업인들 중 기업인 간의 많은 인맥을 형성하고 있어 키플레이어 또는 허브의 역할을 할 수 있는 기업인을 원의 크기의 가중치에 따라 네트워크를

분석한 결과를 나타낸다. 마지막으로 *T-Layer* 모델에서는 인맥의 허브의 역할을 할 수 있는 기업인을 분석할 수 있을 뿐만 아니라 기업인들의 토픽정보를 기반으로 RGB 컬러로 시각화하여 네트워크를 분석한 결과를 나타낸다. 전체 기업인에 대한 네트워크 분석에서 식 (2)를 통해 원의 크기가 클수록 기업인 간의 허브의 역할을 할 수 있는 기업인이라는 것을 알 수 있으며 식 (3)을 통해 비슷한 RGB 컬러일수록 유사한 토픽정보를 갖고 있는 기업인이라는 것을 알 수 있다.

	<i>C-Layer</i>	<i>S-Layer</i>	<i>T-Layer</i>
u_{187}	 <u>(1.000)</u>	 (0.952)	 (1.000)
u_{202}	 <u>(0.983)</u>	 (0.856)	 (0.322)
u_{257}	 (0.658)	 (0.676)	 <u>(0.813)</u>

<Figure 6> Examples of Performance Differences the Results According to Proposed Models (Bold and underline means the highest value across models)

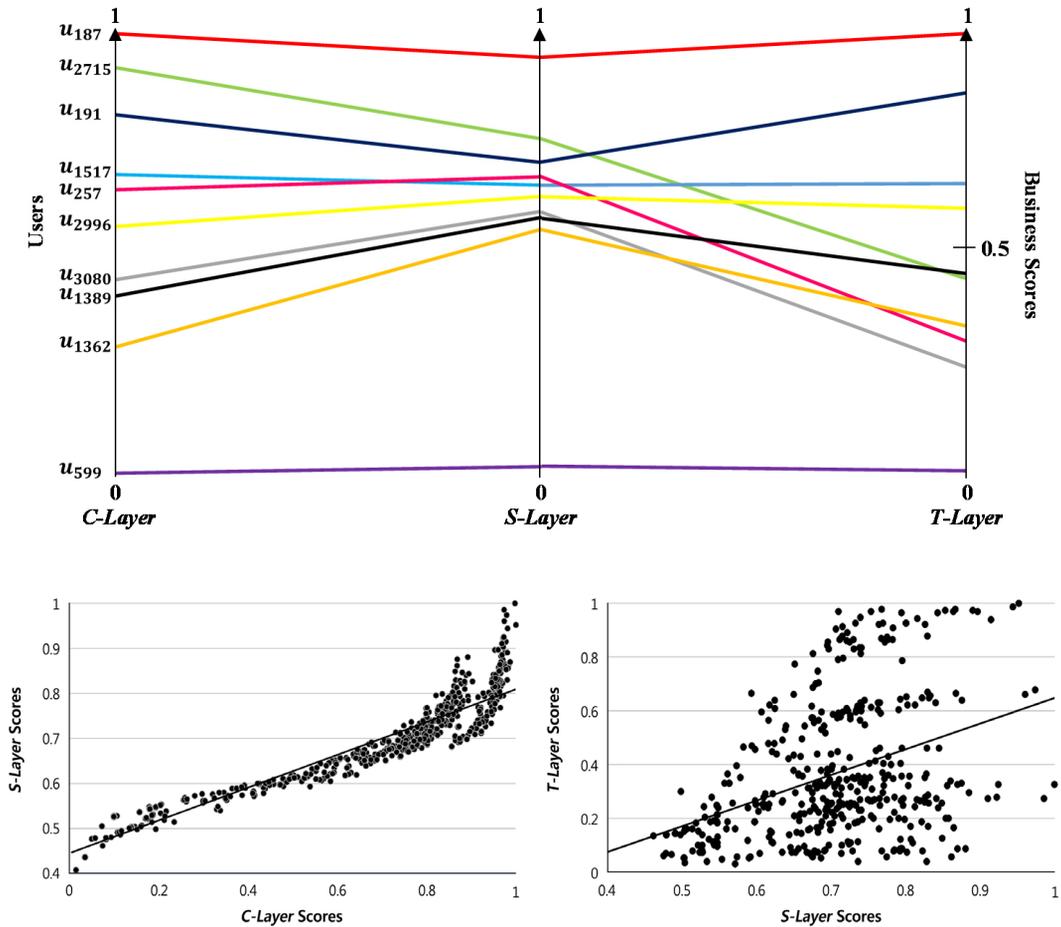
다음으로 <Figure 6>을 보면 u_{187} 은 원의 크기가 큰 기업인과 유사한 토픽클러스터를 가지고 있는 기업인과 네트워크를 잘 형성하고 있어, 제시한 모든 모델들에 대해 점수가 높았다. u_{202} 는 많은 인맥과 원의 크기가 큰 기업인과의 네트워크를 잘 형성하고 있어 *C, S-Layer*에 대해 높은 점수를 나타내고 있었으나 유사한 토픽클러스터를 가진 기업인과의 네트워크를 잘 형성하고 있지 못해 *T-Layer* 모델에 대해서는 낮은 점수를 나타내고 있었다. 반면에 기업인 u_{257} 유사한 토픽정보를 가진 기업인과의 네트워크를 잘 형성하고 있어 *T-Layer* 모델에 대해서는 높은 점수를 나타내고 있었다. 따라서 u_{187} 은 많은 인맥, 허브의 역할을 할 수 있는 기업인, 유사한 토픽정보를 가지고 있는 기업인들과의 이상적인 네트워크를 형성하고 있었다. 반면에 u_{202} 은 인맥 형성 시 유사한 토픽정보를 가지고 있는 기업인과의 네트워크를 활성화하는데 집중할 필요가 있고 u_{257} 는 허브의 역할을 할 수 있는 기업인과의 인맥을 형성하여 인맥 네트워크를 활성화 시켜야 할 필요가 있다.

다음으로 <Table 4>는 식 (1), 식 (2), 식 (5)를

기반으로 계산하여 *C-Layer* 기준으로 기업인들에 대한 점수 변화 비율을 나타내며, <Figure 7>은 제시한 모델을 기반으로 <Table 5>에 나타난 비율에 따른 기업인들의 네트워크 점수 변화와 분포를 나타낸다. <Figure 7>에서 볼 수 있듯이 비율적 측면에서는 상위 30, 50% 해당하는 기업인은 *C, S-Layer* 모델에 비해 *T-Layer* 모델의 점수가 높았다. 상위 10, 40, 60, 80% 해당하는 기업인은 모든 모델에 대한 점수가 평균화 되어 있었고, 나머지 기업인들은 *C, S-Layer* 모델에 비해 *T-Layer* 모델의 점수가 낮았다. 전체 기업인에 대한 비율적 측면에서는 *C-Layer* 모델에 대한 *S-Layer* 모델의 점수의 분포는 비례하는 반면에 *S-Layer* 모델에 비해 대부분 *T-Layer* 모델의 기업인의 수는 낮게 분포되어 있었으며 *S, T-Layer* 모델의 점수가 모든 높은 기업인의 분포는 소수였다. 이를 통해 전체적으로 많은 인맥 네트워크를 형성하고 있는 기업인들은 기업인 간의 허브의 역할을 할 수 있는 반면에 유사한 토픽정보를 가지고 있는 기업인과의 네트워크는 잘 형성하지 못하고 있다는 것을 알 수 있었다.

<Table 4> The Results of Proposed Model Score and Upper Ratios based on *C-Layer*

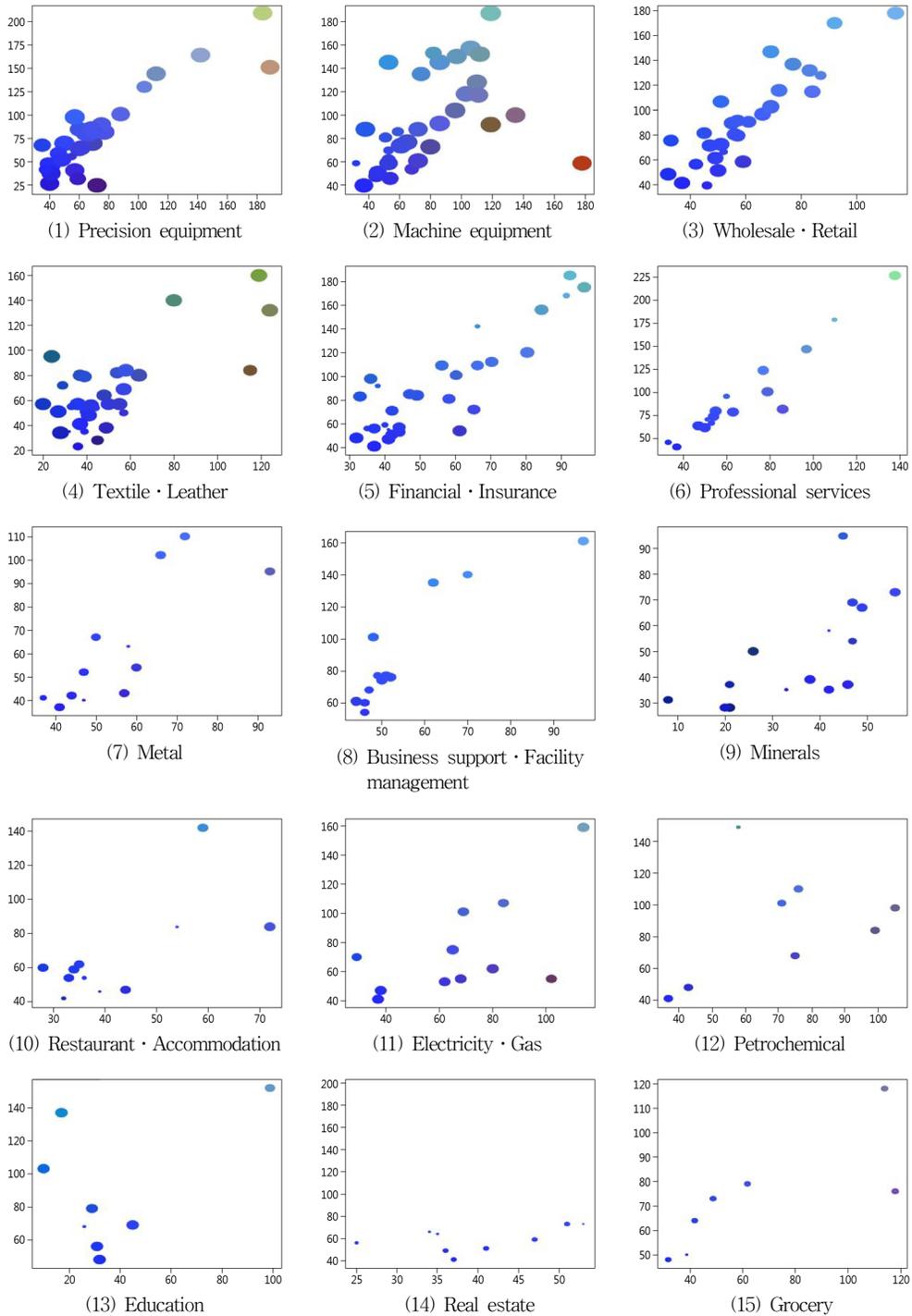
u_i	C-Layer	S-Layer	T-Layer	Ratio
u_{187}	1	0.952	1	10%
u_{2715}	0.918	0.761	0.452	20%
u_{191}	0.816	0.712	0.865	30%
u_{1517}	0.683	0.650	0.609	40%
u_{257}	0.658	0.676	0.813	50%
u_{2996}	0.571	0.625	0.689	60%
u_{3080}	0.440	0.600	0.252	70%
u_{1389}	0.413	0.582	0.465	80%
u_{1362}	0.299	0.557	0.330	90%
u_{599}	0.009	0.019	0.017	100%



<Figure 7> The Results of Proposed Model Score Change and Distribution According to Proposed Models

마지막으로 <Figure 8>은 업종에 따라 S , T -Layer 모델 기반으로 분석한 결과를 2차원 공간에 시각화한 것을 나타낸다. 원의 크기가 클수록 S -Layer 네트워크를 잘 형성하고 있는 기업인을 나타내며, 밀집의 정도가 높을수록 업종간의 T -Layer 네트워크가 잘 형성되어 있다는 것을 나타낸다. 먼저 정밀기기(Precision equipment)를 다루는 업종 분야에서는 다수의 기업인들이 분포되어 있었고 S , T -Layer 네트워크가 모두 활성화 되어 있는 것을 알 수 있었다. 다음으로

기계장비(Machine equipment), 도·소매(Wholesale·Retail), 섬유·가죽(Textile·Leather), 금융·보험(Financial·Insurance) 분야를 순서로 S , T -Layer 모델의 점수가 높아 네트워크 잘 형성되어 있었으나 순서가 뒤로 갈수록 S -Layer 모델의 점수가 T -Layer 모델의 점수에 비해 낮게 형성되어 있었다. 마지막으로 교육(Education), 부동산(Real estate), 식료품(Grocery) 등의 분야에서는 소수의 기업인들이 분포되어 있었고, S , T -Layer 모델의 점수 모두 현저히



〈Figure 8〉 The Visualization Results of Top15 Industrial Fields based on *S, T-Layer* Models

낮은 점수로 형성되어 있다는 것을 알 수 있었다. 따라서 이와 같은 실험을 통해 *S-Layer* 모델의 점수가 *T-Layer* 모델의 점수보다 낮을 경우 허브의 역할을 할 수 있는 기업인과 네트워크를 강화하여 인맥 네트워크를 활성화 시켜야 할 필요가 있으며, 정부는 이러한 기업인 간의 허브의 역할을 할 수 있는 기업인들을 파악하여 이러한 기업인들을 확산시키거나 연결시켜 줄 수 있는 모임의 장을 만들어 줄 필요가 있다. 또한 *S*, *T-Layer* 모델의 점수가 낮을 경우에는 인맥 네트워크를 강화 할 뿐만 아니라 상호 협력에 기반을 둔 융합 활동으로 기술 강화 및 고도화, 신기술 개발 및 창출 등의 효과를 내기 위해 주제 기반의 네트워크를 강화할 필요가 있고 정부는 수직적 산업(Vertical Industry)별로 모임을 강화 할 수 있는 모임 정책 강화에 집중할 필요가 있다.

5. 결 론

본 연구에서는 기업인의 인맥정보와 토픽분석 기반의 기업인 네트워크 계층 분석을 제안하였다. 먼저 *C, S, T-Layer* 각각의 모델에 따라 전체 기업인들에 대한 네트워크 분석과 각 기업인의 모델의 분석 결과에 따라 각각의 모델에 대한 차이점을 가지고 있는 기업인에 대해 네트워크를 분석하였다. 다음으로는 제시한 모델을 기반으로 기업인들의 네트워크 점수 변화 및 분포를 분석하였다. 마지막으로 업종별 특성과 토픽 분석을 기반으로 토픽정보를 RGB 컬러로 표현하여 *S, T-Layer* 모델 기반으로 분석한 결과를 2차원 공간에 시각화하여 업종에 따른 기업인의 네트워크를 분석하고 평가하였다. 실험을 통해 본 연구에서 분석한 결

과 비즈니스간의 협업을 잘하기 위해서는 인맥 네트워크가 부족한 경우 허브의 역할을 할 수 있는 기업과 네트워크를 강화하여 인맥 네트워크를 활성화 시켜야 할 필요가 있다. 또한 기업인들은 상호 협력에 기반을 둔 융합 활동으로 기술 강화 및 고도화, 신기술 개발 및 창출 등의 효과를 내기 위해 주제 기반의 네트워크를 강화할 필요가 있고, 정부는 수직적 산업별로 모임을 강화 할 수 있는 모임 정책 강화가 필요하다는 것을 확인하였다.

향 후 연구에서는 인천, 부천, 서울 등의 지역 뿐만 아니라 전국의 기업인 정보를 기반으로 네트워크를 분석할 필요가 있다. 또한 네트워크 분석을 바탕으로 융합 활동을 활성화 할 수 있는 기업인 추천에 관한 연구를 진행할 예정이다.

References

- [1] Ahn, S. M., Kim, I. W., Choi, B., Cho, Y., Kim, E., and Kim, M-K., "Understanding the Performance of Collaborative Filtering Recommendation through Social Network Analysis," *The Journal of Society for e-Business Studies*, Vol. 17, No. 2, pp. 129-147, 2012.
- [2] Company Search, Incheon dream, Bizok, "Http://bizok.incheon.go.kr/," January, 31, 2018.
- [3] Kim, H., Park, H., Lim, J., Chul C., and Kim, K., "Function-Property based Causality Network of Patents for Technology Convergence," *Proceedings of the Korean*

- Institute of Industrial Engineers, pp. 1205-1219, 2012.
- [4] Borés, C., Saurina, C., and Torres, R., "Technological Convergence: a Strategic Perspective," *Technovation*, Vol. 23, pp. 1-13, 2003.
- [5] Cho, Y. and Kim, E., "A Corporate Strategy on Technological Convergence Through Analyzing Patent Networks and Strategic Indicators," *The Journal of Intellectual Property*, Vol. 9, No. 4, pp. 191-221, 2014.
- [6] Choi, M., "IT R&D Direction for the Age of Convergence," *Journal of Communications and Networks*, Vol. 25, No. 1, pp. 25-31, 2008.
- [7] Cho, Y. and Kim, E., "A Corporate Strategy on Technological Convergence Through Analyzing Patent Networks and Strategic Indicators," *The Journal of Intellectual Property*, Vol. 9, No. 4, pp. 191-221, 2014.
- [8] Greenstein, S. and Khanna, T., "What does Industry Convergence Mean," *Competing in the age of digital convergence*, pp. 201-226, 1997.
- [9] Hacklin, F., Raurich, V., and Marxt, C., "How Incremental Innovation becomes Disruptive: The Case of Technology Convergence," *Proceedings of the IEEE International Engineering Management Conference*, Vol. 1, pp. 32-36, 2004.
- [10] Han, E. J. and Sohn, S. Y., "Technological Convergence in Standards for Information and Communication Technologies," *The Journal of Technological Forecasting and Social Change*, Vol. 106, pp. 1-10, 2016.
- [11] Heo, J. and Yang, C. H., "Applying Network Analysis in Convergent Research Relationships: The Case of High-Tech Convergence Technology Development Program," *Journal of Korea Technology Innovation Society*, Vol. 16, No. 4, pp. 883-912, 2013.
- [12] Jeong, B. K. and Lee, H. Y., "Analyzing the Domestic Collaborative Research Network in Industrial Engineering," *Journal of the Korean Institute of Industrial Engineers*, Vol. 40, No. 6, pp. 618-627, 2014.
- [13] Jeong, B. and Lee, H., "Industrial Engineering as a Multidisciplinary Field: Exploring the Structure of Academic Convergence in Industrial Engineering by Journal Citation Network Analysis," *Journal of the Korean Institute of Industrial Engineers*, Vol. 42, No. 3, pp. 182-197, 2016.
- [14] Karvonen, M. and Kässi, T., "Patent Citations as a Tool for Analyzing The Early Stages of Convergence," *The Journal of Technological Forecasting and Social Change*, Vol. 80, No. 6, pp. 1094-1107, 2013.
- [15] Kim, E., Cho, Y., and Kim, W., "Dynamic Patterns of Technological Convergence in Printed Electronics Technologies: Patent Citation Network," *The Journal of Scientometrics*, Vol. 98, No. 2, pp. 975-998, 2014.
- [16] Kim, J., Jang, Y., and Lee, S., "Issues and Efforts for Technology-Humanities Convergence: Empirical Analysis of Korean Busi-

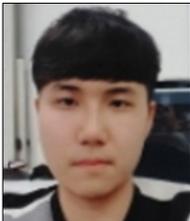
- ness,” *Journal of the Korean Institute of Industrial Engineers*, Vol. 40, No. 5, pp. 451-461, 2014.
- [17] Kim, J. and Lee, S., “Issues for Technology-Humanities Convergence: Patent Keyword Analysis,” *Journal of the Korean Institute of Industrial Engineers*, Vol. 42, No. 3, pp. 231-240, 2016.
- [18] Lee, D. and Kim, K., “WebSite Keyword Selection Method by Considering Semantic Similarity Based on Word2Vec,” *The Journal of Society for e-Business Studies*, Vol. 23, No. 2, pp. 83-96, 2018.
- [19] Lee, H. and Kang, P., “Topic Modeling and Technology Management Research Trend Analysis,” *Proceedings of the Korean Institute of Industrial Engineers Autumn Conference*, pp. 5208-5229, 2016.
- [20] Lei, D. T., “Industry Evolution and Competence Development: the Imperatives of Technological Convergence,” *International Journal of Technology Management*, Vol. 19, pp. 699-738, 2000.
- [21] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J., “Distributed Representations of Words and Phrases and Their Compositionality,” *Advances in Neural Information Processing Systems*, pp. 3111-3119, 2013.
- [22] Park, D. H., Lee, D. H., Kah J. H., and Lee, G. H., “A Study on Tourism Destination Selection of the Inbound Tourist Using Social Network Analysis: Focusing on the Seoul and Gyeongin Tourism Destination,” *International Journal of Tourism and Hospitality Research*, Vol. 29, No. 9, pp. 5-19, 2015.
- [23] Rochat, Y., “Closeness Centrality Extended to Unconnected Graphs: The Harmonic Centrality Index,” In *ASNA 2009*.
- [24] Wirtz, B. W., “Reconfiguration of Value Chains in Converging Media and Communications Markets,” *Long Range Planning*, Vol. 34, pp. 489-506, 2001.

저 자 소 개



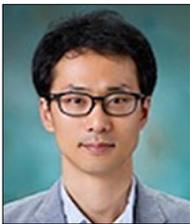
이동훈
2016년
2017년~현재
관심분야

(E-mail: dhlee@inu.ac.kr)
단국대학교 산업공학과 (학사)
인천대학교 산업경영공학과 (석사과정)
텍스트 마이닝, 통계적 기계학습



김용화
2013년~현재
관심분야

(E-mail: kingstwo41@gmail.com)
인천대학교 산업경영공학과 (학사과정)
소셜네트워크 분석, 데이터 마이닝



김관호
2006년
2012년
2013년
2014년~현재
관심분야

(E-mail: khokim@inu.ac.kr)
동국대학교 정보시스템전공 (학사)
서울대학교 산업공학과 (박사)
경희대학교 (연구박사)
인천대학교 산업경영공학과 교수
통계적 기계학습, 빅데이터, 데이터 마이닝