

토픽모델링을 활용한 '산업연구' 학술지 분석

- 과거와 미래에 대한 고찰 -

양철원* · 야마다 아키히코**

요약

본 연구는 텍스트 분석을 적용하여 『산업연구』 학술지를 분석하는 것을 목표로 한다. 산업연구의 1979년 발행된 초판 1권부터 2021년 45권까지의 논문의 제목과 국문 요약어를 추출하여 토픽모델링 LDA(Latent Dirichlet Allocation) 기법과 워드 클라우드(word cloud) 분석을 시행하였다. 이를 통해 산업연구 논문들이 지금까지 어떤 주제를 다루었는지, 또한 주제들이 어떻게 변화되어 왔는지 평가하였다. 또한 유사 학술지와 비교를 통해 산업연구가 향후 나아가할 방향을 탐색하고자 하였다. 분석 결과, LDA 기법을 통해 10개의 토픽을 특정할 수 있었으며, 이중 6개는 경영학과 4개는 경제학과 연관이 깊었다. 산업연구는 해당 시대가 요구하는 주제에 대한 연구를 적절하게 수행해오고 있었다. 마지막으로 산업연구와 유사학술지 4개와 비교하였을 때, 산업연구가 발행 논문수는 부족하지만 영향력지수(Impact Factor, IF)는 가장 우수하였다. 지속적 발전을 위해 앞으로 논문 발행수를 늘리면서도 영향력 지수를 유지하는 것이 가장 큰 과제라 하겠다.

핵심 주제어: 토픽 모델링, 산업연구, LDA, 워드 클라우드, 영향력 지수

* 제1저자, 단국대학교 경영학부 교수, yang@dankook.ac.kr

** 교신저자, 서울대학교 빅데이터 혁신공유대학 데이터 사이언티스트, ayamada5413@snu.ac.kr

<논문 투고일> 2022.6.07

<논문 수정일> 2022.8.04

<게재 확정일> 2022.8.11

I. 서론

단국대학교 부설 미래산업연구소는 한국산업의 성장과 발전에 정책적으로 기여할 목적으로 설립되었으며, 이를 위해 1979년부터 산업의 관점에서 경영·경제·무역 등 제반 분야를 포괄하는 학술지인 『산업연구』를 발간해 오고 있다. 『산업연구』의 목표는 미래 사회에서 요구되는 이론적 연구와 정책연구, 실증연구를 게재하는 것이다. 특정산업에 국한되지 않고 다양한 산업의 융·복합 주제에 대한 논문을 게재함으로써 제4차 산업혁명에서 각 산업들이 경쟁력을 갖추고 글로벌 환경의 빠른 변화에 능동적으로 대응해 나아갈 수 있는 방안을 제시하고자 한다.

산업연구는 운영 측면에서도 변화를 시도하였다. 2016년 편집위원장을 연구소장과 분리하였다. 이를 통해 논문 사전점검 강화 및 심사자의 전문성 확보 등 편집기능을 강화하였다. 이런 노력에 힘입어 산업연구는 2019년도부터 등재후보학술지로 선정되었으며, 2021년에는 계속평가를 통해 등재학술지로 승격하는 쾌거를 이루었다.

이런 중요한 시점에 서서, 본 연구는 산업연구가 지금까지 본래의 목표를 성취하였는지 평가해 보고 향후 나아가 할 방향을 탐색하는 것을 목적으로 한다. 구체적으로 다음과 같은 연구질문에 답하고자 하였다.

- (1) 산업연구는 어떤 주제를 다루었는가?
- (2) 산업연구 논문들의 주제는 어떻게 변화되어 왔는가?
- (3) 산업연구를 다른 유사 학술지와 비교했을 때 차별점과 개선점은 무엇인가?

분석자료는 산업연구의 1979년 발행된 초판 1권부터 2021년 45권까지의 논문을 대상으로 하였다. 이를 위해 산업연구 홈페이지¹⁾에서 각 논문의 제목과 국문 요약(abstract)을 추출하여 분석에 사용하였다. 주요 분석대상인 논문의 국문 요약

1) <https://cms.dankook.ac.kr/web/rifi>.

자료는 2003년 24권부터 존재하며, 그 이전은 국문 요약이 존재하지 않는다. 따라서 2002년 이전 논문에 대해서는 논문 제목만을 분석에 사용하였다.

분석을 위해 텍스트 분석기법(textual analysis)을 사용하였다. 특히 연구 목적에 적합하다고 판단한 토픽모델링(topic modeling) 방법을 적용하였다. 토픽 모델링은 문서의 주제를 추정하기 위하여 각 단어들이 통계적으로 특정 토픽에 포함될 확률을 파악하여 주제끼리 묶어주는 비지도학습 알고리즘이다. 본 연구는 토픽모델링 기법 중 가장 대표적인 LDA(Latent Dirichlet Allocation) 방법을 사용하여 국문 요약을 분석하였다. 국문 요약이 존재하는 않는 2002년 이전 자료에 대해서는 논문 제목만을 대상으로 워드 클라우드(word cloud) 분석을 시행하였다.

해외에서는 특정 분야를 설정하여 해당 분야의 논문들을 분석한 연구가 다수 존재한다. Hamelman and Mazze (1974)는 재무·금융 분야의 논문들을 대상으로 인용(citation) 분석을 처음 시행하였다. 이후 2000년대에 접어들면서 Chung, Cox, and Mitchell (2001)은 재무·금융 분야의 인용 패턴에 대한 분석을 다시 시도하였으며, Arnold, Butler, Crack, and Altintig (2003)은 논문의 영향력에 초점을 맞추어 연구를 진행하였다. 한국 문헌에 대해서는 Chan, Chang, and Chen (2011)이 한국의 기관과 저자들의 재무분야 논문을 분석하였다.

특정 학술지를 중심으로 논문들에 대해 분석한 연구들도 존재한다. Linnenluecke, Chen, Ling, Smith, and Zhu (2016)은 'Pacific-Basin Finance Journal'을 중심으로 하여 아시아, 태평양 지역의 재무 연구의 최신 경향을 연구하였다. 정재만(2021)이 '재무관리연구' 학술지에 대한 인용분석을 시도하였다. 본 연구는 '산업연구'라는 특정 학술지를 대상으로 하였다는 점에서는 정재만(2021)의 연구와 비슷하다. 하지만 기존의 연구들이 주로 논문의 인용을 분석한 것에 비해 본 연구는 논문의 텍스트 자료에 토픽 모델링을 적용하여 연구주제를 분석했다는 점에서 차이가 있다.

논문의 주요 결과는 다음과 같다. 첫째, 토픽 모델링 LDA 기법을 사용하여 산업 연구에 게재된 논문을 10개의 토픽으로 구분할 수 있었다. 10개의 주제는 회계 및 금융자산 평가, 국가의 산업, 환경, 제도, 교육 문제, 경제성장과 자본시장, 브랜드

관리 및 마케팅, 기업경영분석 및 재무관리, 정부 정책, 정보 시스템, 이론모형 및 예측, 기업 전략 및 기술, 인사조직에 해당한다. 이 10개의 토픽은 크게 경영학과 연관된 주제 6개와 경제학에 해당하는 주제 4개로 분류할 수 있었다. 이는 산업연구의 주요 대상인 ‘산업’이 경제학의 주요 관심사인 ‘국가’와 경영학의 주요 관심대상인 ‘기업’의 중간에 위치하여 둘을 다 포섭할 수 있기 때문이다.

둘째, 산업연구에 게재된 논문의 주제가 어떻게 변화되어 왔는지 살펴보았다. 2000년대 초반에는 경제성장과 자본시장, 이론 모형 및 예측 등의 토픽이 많은 비중을 차지하였다. 이는 한국의 산업발전과 연관된 주제들이다. 2008년을 중심으로 한 2000년대 후반에는 국가의 산업, 환경, 제도, 교육 문제와 재무금융 토픽의 논문이 크게 증가하였다. 이는 2008년 글로벌 금융위기를 거치면서 한국의 금융위기 대응과 연관된 연구들이 많이 진행되었음을 짐작케 한다. 2020년대 초반에 들어서면서는 정보 시스템, 기업 전략 및 기술 토픽이 많은 부분을 차지하는데, 이는 제4차 산업혁명과 관련된 ICT, 빅데이터 관련해 활발한 연구들이 많이 늘어났기 때문이다. 전반적으로 산업연구가 해당 시대의 상황에 따라 시대가 요구하는 연구들을 적절하게 수행해왔음을 보여준다.

마지막으로 산업연구가 앞으로의 나아가야 할 방향을 더 고찰해 보기 위해 유사 학술지 4개와 비교하는 연구를 수행하였다: 산업융합연구, 산업경제연구 산업혁신연구, 산업연구(JIET). 워드 클라우드를 살펴보았을 때, 산업연구의 연구주제는 전통적인 산업을 주로 다루고 있는 산업경제연구와 산업연구(JIET)와 성격이 유사하였다. 산업연구가 4차 산업혁명 등 미래 산업에 대해 다루는 방향으로 더 나아가고 싶다면 산업혁신연구의 주제들을 벤치마크할 필요가 있다. 질적인 면을 보았을 때, 산업연구의 영향력 지수(IF)가 1.13로 산업혁신연구 1.01보다 더 높으며, 나머지 유사 학술지에 비해서도 높았다. 하지만 현재 산업연구의 연간 평균 논문수 11건은 다년간 학술등재지의 위치를 유지해 온 다른 유사학술지에 비해 너무 작았다. 양적인 측면인 연간 평균 논문수를 등재 유사학술지 중 최소 수준인 30건대로 높일 필요가 있다. 논문 발행수가 많아지면 논문의 질을 관리하는 것이 힘들어지고 영향력 지수(IF)의 하락으로 이어질 수 있다. 이런 문제점이 발생하지 않도록 논문 발행수를 늘리면서도 피인용 횟수를 지속적으로 향상시킬 수 있는 방법을 강구해

야 할 것이다.

본 논문은 다음과 같이 구성된다. 제2장은 본 연구에서 사용될 토픽 모델링 LDA 기법에 대해 소개하고, 제3장은 자료에 대해 설명한다. 제4장은 본 연구의 연구질문 3개에 대한 구체적 분석결과를 제시하고 이를 해석하였다. 마지막 제5장은 논문을 요약 마무리한다.

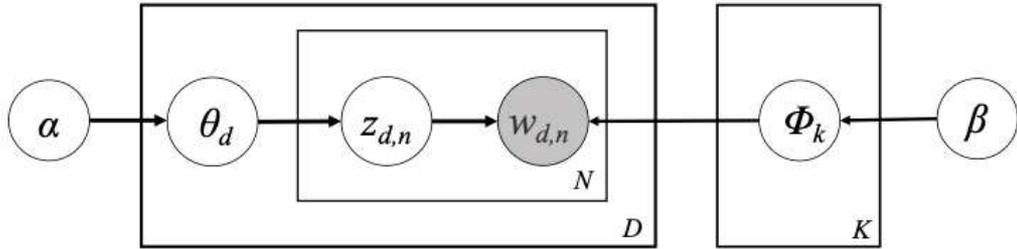
II. 토픽모델링(topic modeling)

1. 개념

본 연구에서는 토픽모델링 기법을 이용해 논문의 주제에 대한 분류를 실시하였다. 토픽모델링이란 문서를 구성하고 있는 단어를 바탕으로 각 문서의 주제를 추출하고 분류하는 기계학습모델 중 하나로 대량의 서류를 주제별로 자동으로 분류할 때 사용된다. 현재까지 다양한 토픽모델링 방법이 제안되어 왔는데 그 중에서도 일반적으로 자주 사용되는 LDA(Latent Dirichlet Allocation) 기법을 사용하였다 (Blei et al., 2003).

LDA는 문서 데이터 속에 숨어있는 토픽에 대한 확률 분포와 각 단어에 대한 토픽 확률 분포를 기반으로 한 계층적 확률분포를 학습하는 베이지안 학습모형이다. 학습은 랜덤 확률분포를 설정하는 것으로 시작되며 데이터의 특성에 맞게 확률분포가 결정되어 간다. 각 문서 안에 있는 단어를 하나하나 관찰하면서 단어에 대한 토픽 확률분포를 갱신해 나가는 학습과정을 통해 문서에 대한 토픽 확률분포와 각 단어에 대한 확률분포가 결정되어 간다. 즉, LDA는 $p(\phi_{1:K}, \theta_{1:D}, z_{1:D} | w_{1:D})$ 라는 조건확률을 구하는 모형이며, [그림 1]은 과정을 요약하고 있다. θ_d 는 문서 d 의 토픽 확률분포를 의미하며, α 에 의해 결정된다. ϕ_k 는 파라미터 β 에 의해 생성된 토픽 k 에 대한 단어의 확률분포를 나타내고 $z_{d,n}$ 는 문서 d 의 i 번째 단어에 배정되는 토픽이고 $w_{d,i}$ 는 문서 d 중에서 i 번째로 관찰된 단어이다.

[그림1] LDA 모형의 변수들과 연산과정



다음 식은 사후확률분포이며 다항분포의 형태를 가지게 된다.

$$p(\Phi_{1:K}, \theta_{1:D}, z_{1:D}, w_{1:D}) = \frac{p(\Phi_{1:K}, \theta_{1:D}, z_{1:D}, w_{1:D})}{p(w_{1:D})} \quad \text{식 (1)}$$

식(1)의 우변의 분자는 모든 확률변수의 결합분포이고 다음과 같이 계산 가능하다. 처음 두개의 항은 사전확률을 나타내고 세 번째항은 모형을 나타낸다. 관측가능한 변수는 $w_{d,i}$ 이며 문서에서 구할 수 있다. θ_d 는 파라미터 α 인 디리클레 분포를 따르고 ϕ_k 는 파라미터 β 인 디리클레 분포를 따른다. 식(1)의 우변의 분모 주변확률 $p(w_{1:D})$ 는 정확한 계산이 어렵다. 이론적으로는 숨겨진 토픽구조의 가능한 모든 인스턴스화에 대한 결합 분포를 합산하여 계산할 수 있다. 그러나 가능한 토픽 구조의 수는 기하급수적으로 많기 때문에 이 합을 계산하기는 어렵다.

이러한 상황에서 Blei et al. (2003)이 사용한 방법이 붕괴깁스샘플링(collapsed Gibbs sampling)이며, 여기서 $\Phi_{1:K}$ 와 $\theta_{1:D}$ 는 샘플링에서 제외되고 $p(z_{1:D}|w_{1:D})$ 로부터 잠재변수 $z_{1:D}$ 만 샘플링된다. 이러한 샘플링 후에 $z_{1:D}$ 가 주어지면 $\Phi_{1:K}$ 와 $\theta_{1:D}$ 의 추정치를 계산할 수 있게 된다. LDA 모델을 위해 필요한 붕괴깁스샘플링에는 아래와 같은 조건부 확률의 계산이 필요하다.

$$p(z_{d,i} = j | z_{-i}, w) \quad \text{식 (2)}$$

여기서 $z_{d,i} = j$ 는 $w_{d,i}$ 에 토픽 $z_{d,i}$ 가 j 가 되는 사건을 의미한다. z_{-i} 는 $z_{d,i}$ 외의 모든 토픽 할당을 의미하고 w 는 $w_{1:D}$ 을 의미한다. 조건부확률 계산을 적절히 하면 아래와 같은 식을 얻을 수 있다.

$$p(z_{d,i} = j | z_{-i}, w) \propto \frac{n_{d,k=j} + \alpha_j}{\sum_{k=1}^K (n_{d,k} + \alpha)} \times \frac{v_{j,c=w_{d,i}} + \beta_{w_{d,i}}}{\sum_{c=1}^C (v_{j,c} + \beta_c)} \quad \text{식 (3)}$$

여기서 $n_{d,k}$ 는 문서 d 에 나타나는 토픽 k 에 할당된 단어의 수이고 $v_{j,c}$ 는 단어 c 가 토픽 j 에서 나타나는 횟수의 합이다. α 와 β 는 디리클레 분포에 관한 하이퍼 파라미터로, K 는 토픽의 수, C 는 문서집합 전체에서 단어의 개수이다.

이상의 수식을 이용하여 붕괴깁스샘플링은 다음과 같은 단계로 이루어진다.

- (1) 각각의 단어에 랜덤으로 토픽을 할당한다.
- (2) 첫 번째 문서($d=1$)의 첫 번째 단어($i=1$)부터 시작하여 마지막 문서($d=D$)의 마지막 단어($i=N$)에 이르기까지 아래 계산 과정 (3)을 반복된다.
- (3) 토픽 j 을 변화시키면서 $p(z_{d,i} = j | z_{-i}, w)$ 를 계산하고 K 개의 토픽 중 확률 $p(z_{d,i} = j | z_{-i}, w)$ 를 가장 크게 하는 j 를 d 번째 단어의 토픽으로 결정한다.
- (4) (2)부터 (3)까지의 계산 과정을 모든 단어에 대한 토픽 할당이 안정화 될 때까지 반복한다.

본 연구에서는 LDA 분석을 위해 자료에 대해 Mecab 형태소 분석기와 python coding를 이용하여 조사나 접사 등의 내용적인 의미를 갖지 않는 불용어를 제거하였다. LDA 분석에는 LDA Mallet(Machine Learning for Language Toolkit) 모

형을 사용하였다(McCallum, 2002).

LDA를 이용한 토픽 분석에 있어 가장 어렵고 시행착오가 많은 부분이 토픽 수를 결정하는 일이다. 토픽 수의 결정의 기준 지표로 일관성 점수(Coherence Score)를 사용하였다. 일관성 점수에도 여러 종류가 있는데 본 연구에서는 UMASS를 사용하였다(Mimno et.al, 2014). UMASS은 아래와 같은 수식으로 정리된다.

$$C_{UMASS}(w_i, w_j) = \log \frac{D(w_i, w_j) + 1}{D(w_i)} \quad \text{식 (4)}$$

$D(w_i, w_j)$ 는 w_i 와 w_j 가 문서 안에 함께 나타나는 횟수이며, $D(w_i)$ 는 w_i 가 단독으로 나타나는 횟수이다. UMASS가 클수록 w_i 와 w_j 가 함께 나타날 확률이 크다는 것을 의미하며, 이를 토픽모델링 분류에 응용하면 각 토픽 안에 문서에서 함께 자주 나타나는 단어들 즉 의미적인 관계가 깊은 단어들이 모였다는 것을 의미하여 토픽 분류가 잘 됐다고 판단할 수 있다. UMASS는 대칭이 아니다 즉 $C_{UMASS}(w_i, w_j)$ 는 $C_{UMASS}(w_j, w_i)$ 와 일치하지 않는다. 본 연구에서는 정해진 토픽 개수에 대한 모형을 추정하고 추정된 모형에 대해 각 토픽을 속할 확률점수가 가장 높은 상위 20개 단어들 간의 UMASS합의 평균을 일관성점수로 사용하였다. 그리고 일관성 점수가 가장 높아지는 토픽수로 토픽수를 정하였다. UMASS는 Gensim이라는 python library가 제공하는 Coherence Model를 사용해 산출하였다.

Ⅲ. 자료

1. ‘산업연구’ 학술지

본 연구에서는 산업연구 학술지의 홈페이지²⁾에 존재하는 1권(1979년)부터 45권

(2021년)까지 산업연구에 수록된 모든 논문을 대상으로 하였다. 각 논문의 제목과 요약, 핵심어를 추출하여 분석에 사용하였다.

[표 1]은 산업연구의 주요 연혁을 보여주고 있다. 산업연구의 초판 1권은 1979년 발행되었다. 발행 초기에는 논문의 요약(abstract) 부분은 존재하지 않는다. 1997년 20권부터 논문형식 변경되어 영문 '요약'을 논문 마지막 부분에 수록하기 시작하였다. 2003년 산업연구 24권부터는 논문형식은 변경이 없었지만, 국문 '요약'을 웹사이트에 따로 게시하였다. 실제적인 논문형식 변경은 2006년 30권부터 이루어졌다. 이때부터 논문 첫 장에 목차를 없애고 대신 국문 '요약'을 수록하는 형식으로 변경하였다. 따라서 국문요약은 2003년 24권부터 존재하며, 이 국문 요약의 텍스트가 본 연구의 주요 분석대상이 된다. 2002년 이전 논문에 대해서는 논문 제목만이 분석 가능하다.

[표 1] 산업연구의 주요 연혁

연도	연혁
1963	산업관리연구소 설립(9월)
1978	산업연구소 명칭 변경
1979	산업연구 초판 발간
1997	산업연구 20권부터 논문형식 변경 (영문 '요약'을 논문 마지막에 수록함)
2003	산업연구 24권부터 논문형식 변경 (국문 '요약'을 웹사이트에 따로 수록함)
2002	연구소 운영 전반 심의 기구로 평의원회를 운영위원회로 개편
2006	산업연구 30권부터 논문형식 변경 (논문 첫 장에 목차를 없애고 대신 국문 '요약'을 수록함)
2010	미래산업연구소로 명칭 변경
2019	산업연구 한국연구재단 등재후보지 선정
2021	산업연구 한국연구재단 등재지 선정

2) <https://cms.dankook.ac.kr/web/rifi>.

[표 2] 산업연구에 게재된 논문의 현황

연도	권	호	논문편수	연도	권	호	논문편수
2021	45	3	15	2000	22	3	12
2020	44	3	15	1998	21	1	13
2019	43	2	10	1997	20	1	8
2018	42	2	10	1996	19	1	8
2017	41	2	10	1995	18	1	13
2016	40	2	8	1994	17	1	8
2015	39	2	6	1993	16	1	11
2014	38	2	8	1992	15	3	13
2013	37	2	6	1991	14	1	8
2012	36	2	10	1990	13	1	9
2011	35	2	10	1989	12	1	12
2010	34	2	11	1988	11	1	9
2009	33	2	10	1987	10	1	8
2008	32	4	17	1986	9	1	8
2007	31	2	8	1985	8	1	8
2006	30	1	14	1984	7	1	13
2005	29	1	7	1983	6	1	14
2005	28	1	10	1983	5	1	14
2004	27	1	9	1982	4	1	14
2004	26	1	11	1981	3	1	5
2003	25	1	10	1980	2	1	7
2003	24	2	8	1979	1	2	7
2002	23	1	10	총계			456

주) 2019~2020년: 등재후보학술지, 2021년 이후: 등재학술지.

[표 2]는 산업연구에 게재된 논문의 현황을 보여주고 있다. 산업연구의 초판 1권은 1979년 발행되었으며 2021년 45권까지 총 456편의 논문이 게재되었다. 보통 1년에 1권이 발간되었다. 1999년과 2001년은 발행본이 존재하지 않는다. 당시가 한국이 외환위기에 영향 받는 기간이었던 상황과 연관되어 보인다. 반면 1983년과 2003년~2005년과 같이 한 해에 2권이 발행된 연도도 존재한다. 2004년에 26, 27

권 두 번 발행되었으며, 당해 연도 발행 논문수는 20편으로 연간 발행 논문수로는 최댓값이다.

초창기의 학술지는 대부분 1권에 1호가 발행되었다. 특이하게 1992년 15권이 3호, 2000년 22권이 3호, 2003년 24권이 2호, 2007년 31권이 2호, 2008년 32권이 4호로 구성되어 있어 이 연도에 발행 호수가 많다. 이때까지도 산업연구의 편집 및 발행규정 등이 아직 정비되지 않았기 때문일 것이다. 산업연구의 발행 횟수 등이 규정된 편집규정은 2016년 처음 시행되었다.

2009년 이후로는 매년마다 1권 2호씩 발행이 규칙적으로 이루어졌다. 규정 등의 정비로 힘입어 산업연구는 안정적으로 발전하였으며, 2019년 1월부터 등재후보학술지로 선정되었다. 이후 계속평가를 준비하기 위해 2020년부터는 매년 3호를 발행하는 것으로 학술지 발행 횟수를 확대하였다. 이러한 노력의 결과로 산업연구는 계속평가 이후 2021년 1월부터 등재학술지로 승격되었다.

2. 유사 학술지 선정

산업연구와 유사분야의 대표적인 학술지를 검색하기 위해 한국학술지인용색인(Korean Citation Index: 이하 KCI) 홈페이지의 학술지 검색 시스템을 활용하였다.³⁾ 여기서 '산업'을 검색어로 하여 학술지를 검색하였다. 이후 다음과 같은 기준을 논문을 선택하였다.

- 2022년 기준 등재후보지나 등재지인 학술지
- 특정 분야를 대표하는 학술지는 제외하였다. 산업연구가 여러 분야 간의 융합연구를 주 목표로 하기 때문에 특정 분야를 초점으로 하는 학술지는 비교 대상이 아니라고 생각하였다. 이에 따라 산업조직연구, 산업노동연구, 산업경영시스템학회지, 디지털산업정보학회 논문지 등 특정 분야에 집중된 학술지는 제외하였다.⁴⁾

3) <https://www.kci.go.kr/kciportal/main.kci>.

4) 『산업경영시스템학회지』의 경우 중견 제조업체의 제조기술과 정보통신기술(ICT)의 융합을 바탕으로 제조업의 경쟁력을 강화시키는 학술연구에 주안점을 두고 있다. 『디지털산업정보학회 논문지』와 『디지털융복합연구』에서는 디지털 및 IT관련 연구주제를 중점적으로 다룬다. 『산업경제연구』는 산업경제와 기업경영관련 이론 및 정책개발을 위한 학술연구 활동에 집중하고

- 특정 산업을 대표하는 학술지도 제외하였다. 예를 들어, 한방산업연구, 산업식품공학 등은 제외되었다.

이런 기준으로 선택하였을 때, 산업융합연구, 산업경제연구, 산업혁신연구, 산업연구(JIET) 4개 학술지가 선택되었다. 산업연구원에서 발행하는 산업연구(JIET: Journal of Industrial Economics and Trade)는 산업연구와 명칭이 동일하였다. 이를 구분하기 위해 산업연구원에서 발행하는 학술지는 ‘산업연구(JIET)’로 영문 약어와 함께 표기하기로 한다. 산업연구와 가장 직접적으로 비교되는 학술지는 ‘산업융합연구(Journal of Industrial Convergence)’이다. 이 학술지도 2019년에 등재후보지로 선정되고, 2021년부터 등재지로 승격하여서 인용에 있어 산업연구와 동일한 역사를 가지고 있다. 이에 비해 산업경제연구(Review of Business & Economics)와 산업혁신연구는 이미 오래 전에 등재지로 편입되었다. 이 두 학술지는 오랜 역사를 지니며 많은 연구논문들을 출간하며 등재지의 지위를 유지하고 있기 때문에, 산업연구가 가야할 미래를 보여주는 학술지라 생각된다. 산업연구(JIET)는 아직 등재후보지이지만, 명칭이 동일하고 향후 잠재적 경쟁자가 될 수 있을 것이라 자료에 포함하였다. 위의 4개의 학술지를 중심으로 산업연구와 비교연구를 진행하였다.

IV. 분석 결과

1. 산업연구는 어떤 주제를 다루고 있는가?

토픽 모델링에 따라 산업연구의 논문들을 분류하였다. 분석은 논문의 요약(abstract)을 대상으로 하였다. 따라서 국문 요약이 존재하는 2003년부터 2021년까지 19년 동안의 논문이 분석 표본이 된다.

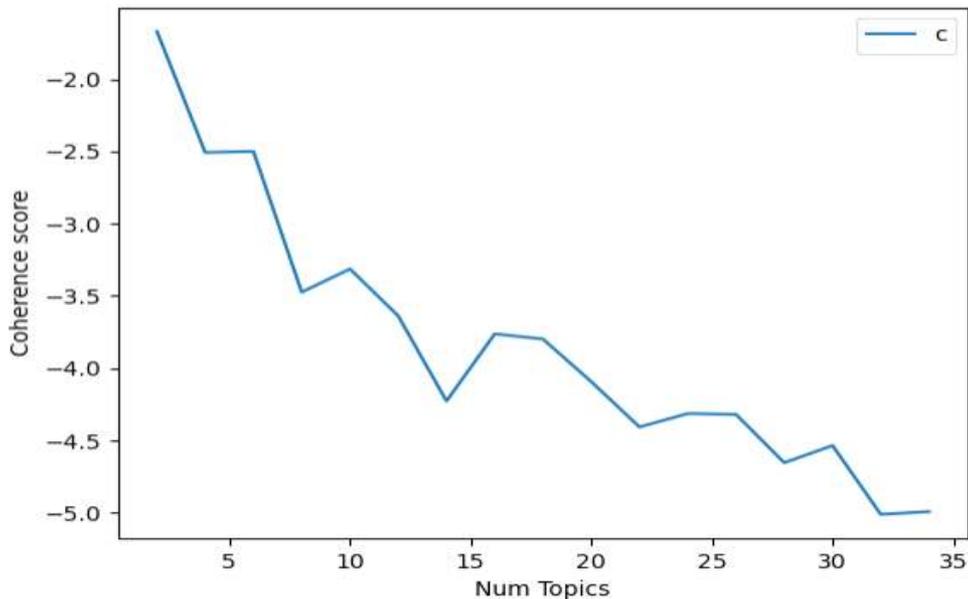
LDA를 이용한 토픽 분석에 있어 가장 어렵고 시행착오가 많은 부분이 토픽 수

있다.

를 결정하는 일이다. LDA 모델의 경우 토픽의 수를 연구자가 선정해 해당 토픽 수에 맞춰 문서를 분류한다. 토픽 수의 결정의 기준이 되는 지표는 다양하지만 그 중에서 일반적으로 사용되는 지표가 일관성 점수(Coherence Score)이다. 일관성 점수가 높을수록 서로 의미적으로 관계가 있는 단어끼리 그 그룹에 모였다는 것을 의미한다. 따라서 일관성 점수가 높을수록 토픽 분류가 잘 되어 있다고 할 수 있다. 일관성 점수에도 여러 종류가 있는데 본 연구에서는 UMass를 사용하였다 (Mimno et.al, 2014).

[그림 1]은 적절한 토픽 수를 찾기 위해 토픽 수를 변화시켜가면서 UMass 값이 어떻게 변화하는지 관찰하고 있다. 그림에서 토픽 수가 8개일 때 까지 UMass 값이 급격히 감소하다가 10개까지 상승한 후 지속적으로 감소하는 경향을 보인다. 급격한 감소는 모형의 학습이 아직 불안정하다고 판단할 수 있다. 따라서 UMass 값이 비교적 안정적이면서도 가장 높은 토픽 값 10개를 본 연구의 토픽 수로 선정하였다.

[그림 1] 토픽수의 변화에 따른 일관성 점수(UMASS)의 변화



[표 3]은 토픽모델 결과 얻어진 주제 10개를 보여주고 있다. 1번은 ‘회계 및 재무금융’ 관련 주제이다. 주제 관련 주요 단어는 회계, 금융, 경우, 평가, 자산, 위기, 위험, 문제점, 기준, 기관 등이다. 경영학의 분야인 회계와 재무금융 분야는 서로 밀접하게 연관되어 있어서 해당 논문들이 같은 토픽으로 분류되어 나온 것이라 사료된다. 2번째 주제의 주요 단어들은 환경, 제도, 나라, 문제, 산업, 국가, 교육, 방안, 측면, 해결 등이다. 이 단어들은 국가의 산업, 환경, 제도, 교육 문제에 관한 주제들에 해당하며, 경제학에서 중요하게 다루는 주제들이다. 3번째 주제도 경제성장, 자본 시장, 투자 등 경제학 관련 주제들이다. 반면 4번 주제는 브랜드 분석과 연관되어 있으며 이는 경영학의 마케팅 분야의 주요 주제이다. 따라서 4번째 주제는 ‘마케팅’이라 지정하였다. 이와 같은 방식으로 나머지 주제들도 각 주제를 명명하였다.

[표 3] LDA 모형을 통해 분류된 주제와 주요단어

토픽 번호	주제	주제 관련 주요 단어
1	회계 및 금융자산 평가	회계, 금융, 평가, 자산, 위기, 위험, 문제점, 기준, 기관
2	국가의 산업, 환경, 제도, 교육 문제	환경, 제도, 나라, 문제, 산업, 국가, 교육, 방안, 측면, 해결
3	경제성장과 자본시장	시장, 투자, 자본, 중국, 한국, 국내, 보험, 경제, 성장, 자유
4	마케팅	분석, 영향, 결과, 효과, 긍정, 유의, 실증, 브랜드, 자료
5	재무관리	기업, 영향, 가치, 결과, 비율, 성과, 경영, 분석, 대상, 규모
6	정부 정책	정책, 경제, 증가, 필요, 정부, 재정, 조세, 확대, 장기, 지출
7	정보 시스템	정보, 요인, 시스템, 개발, 도입, 사용, 활동, 성공, 제품, 제시
8	계량경제학	모형, 방법, 이용, 이론, 예측, 가격, 분석, 비교, 변화, 주식
9	기업 전략 및 기술	중요, 서비스, 효율, 기술, 감사, 제공, 전략, 비용, 구조
10	인사조직	관계, 조직, 리더십, 효과, 특성, 영향, 직무, 지향, 행동

위의 10개의 토픽을 살펴보면 크게 경영학과 연관된 주제들이 6개에 해당하며, 경제학과 관련 깊은 주제들이 4개이다. [표 4]는 이를 구분하여 보여주고 있다. 1, 4, 5, 7, 9, 10번 주제는 경영학에 해당하며, 나머지는 경제학에 해당하는 것으로 분류하였다. 이는 본 학술지의 주요 대상인 '산업'이라는 단위가 경제학의 주요 관심사인 '국가'와 경영학의 관심대상인 '기업'의 중간에 위치하여 둘을 다 포섭하기 두 학문에 걸쳐서 고르게 나타나는 것이라 사료된다. 또한 학술지의 투고자들이 경제학이나 경영학 전공자들이 대부분이기 때문에 나타나는 자연스러운 현상이라 생각된다. 결과적으로 경제학과 경영학의 다양한 분야에 해당하는 연구들이 산업연구에 게재되고 있음을 확인하였다.

[표 4] LDA 모형을 통해 분류된 토픽의 구분

토픽번호	분류	주제
1, 4, 5, 7, 9, 10	경영학	회계 및 금융자산 평가, 마케팅, 재무관리, 정보 시스템, 기업 전략 및 기술, 인사조직
2, 3, 6, 8	경제학	국가의 산업, 환경, 제도, 교육 문제, 경제성장과 자본시장, 정부정책, 계량경제학

2. 산업연구 논문의 주제는 어떻게 변화되어 왔는가?

여기서는 산업연구에 게재된 논문의 주제가 어떻게 변화되어 왔는지 살펴보고자 한다. [그림 2]는 10개의 토픽에 해당하는 논문의 수를 연도별로 추이를 보여주고 있다. 매년 학술지에 15개 이상의 논문이 게재된 2003년, 2004년, 2008년을 제외하고는 대부분 10개 내외로 일정하다가 2020년부터 15개로 증가하였다. 2019년 등재후보지로 승격되면서 2020년부터는 2회 발생에서 3회(1회 당 논문 5편) 발행으로 증가한 결과 논문의 수가 매년 10편에서 15편으로 증가하였다.

토픽별로 보면, 2000년대 초반에는 3번(경제성장과 자본시장)과 8번(계량 경제학) 토픽이 많은 비중을 차지한다. 2008년을 중심으로 한 2000년대 후반에는 2번(국가의 산업, 환경, 제도, 교육 문제)과 5번(재무관리) 토픽의 논문이 크게 증가하

였다. 특히 5번 토픽이 많이 증가하여서 재무금융관련 부분의 연구가 많이 증가하였다. 2008년 글로벌 금융위기를 거치면서 이에 대한 연구들이 많이 이루어졌다고 생각된다.

2010년대 들어서면서는 6번(정부정책)과 10번(인사 조직) 토픽의 논문들이 꾸준히 차지하고 있다. 2010년대 후반부터는 4번(마케팅)과 5번(재무관리) 토픽의 약진이 눈에 띈다. 브랜드 등 마케팅에 대한 연구와 재무관련 연구가 증가하였다. 소비자 행동, 핀테크에 대한 논문들이 다수 게재된 것과 연관되어 보인다.⁵⁾ 2020년대 초반에 들어서면서는 7번(정보 시스템), 9번(기업 전략 및 기술) 토픽이 많은 부분을 차지하는데, 이는 제4차 산업혁명과 관련된 ICT, 빅데이터 관련 연구들이 많이 늘어났기 때문이라 사료된다.⁶⁾ 전반적으로 산업연구가 해당 시대의 상황에 따라 시대가 요구하는 연구들을 적절하게 수행해왔음을 보여준다.

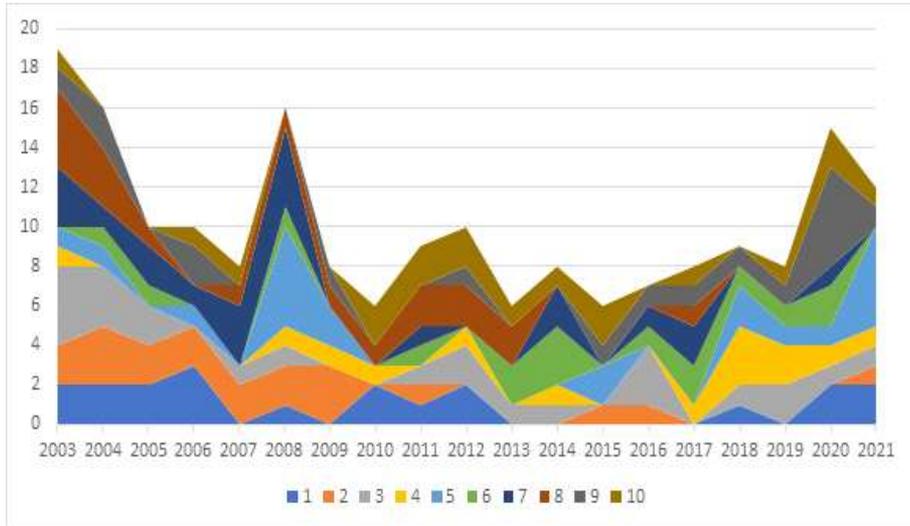
산업연구의 1권은 1979년 발생되었으나, 초기 논문 형식에는 논문에 대한 요약이 존재하지 않는다. 논문의 형식 변경으로 국문 요약이 존재하는 것은 2003년부터이다. 이런 연유로 LDA 모형 분석에는 2003년부터의 국문 요약을 사용하였다. 하지만 우리는 2002년 이전의 논문에 대해서는 제목을 가지고 있다. 따라서 산업연구에 실린 논문의 제목을 사용하여 워드 클라우드를 그려보았다. 워드 클라우드를 통해 산업연구의 주제에 대해 간단하게나마 다시 확인해 볼 수 있을 것이다. 워드 클라우드 분석에서 논문에 일반적으로 많이 들어가는 중심, 영향, 효과, 분석 등의 단어는 제외하였다.

5) 박주섭, 홍순구(2016), 박지영, 김영진, 홍사욱(2015)을 참고하십시오.

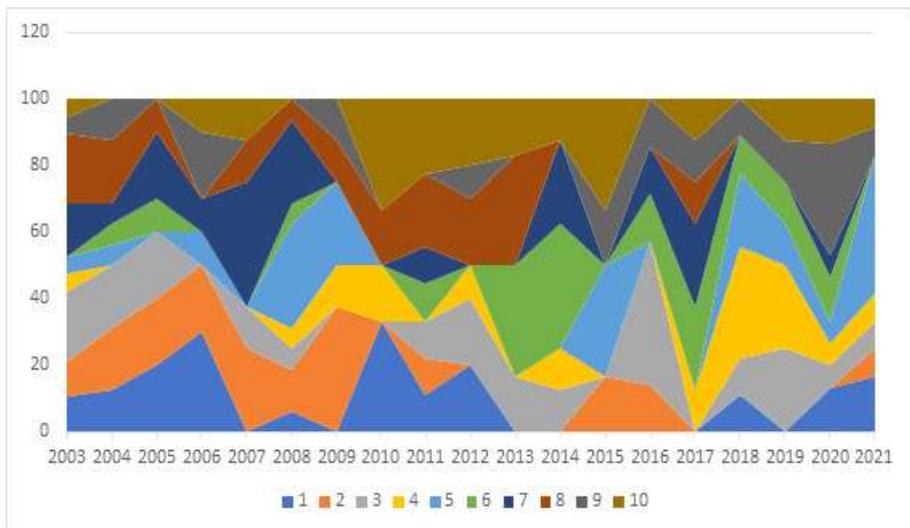
6) 권기덕 (2020), 윤상철(2019)을 참고하십시오.

[그림 2] 토픽의 연도별 추이

(a) 연도별 논문 수



(b) 연도별 논문 비중



[그림 3]은 전체기간을 1979년~2002년 기간(24년)과 2003년~2021년 기간(19년) 둘로 나누어서 분석한 워드 클라우드를 보여주고 있다. 우선 그림 (b)에서 2003년~2021년 결과를 통해 앞에서 토픽 모델링을 사용하여 분석한 결과와 비교해 볼 수 있다. 기업, 시장, 한국 등의 단어가 가장 많이 등장하여 연구 대상이 기업, 시장, 국가로 구성되어 있음을 볼 수 있다. 정책, 회계, 금융, 조직 등의 단어도 많이 등장하는데 앞의 LDA 모형을 사용하여 분류한 주제들과도 비슷하다.

[그림 3] 논문 제목의 워드클라우드

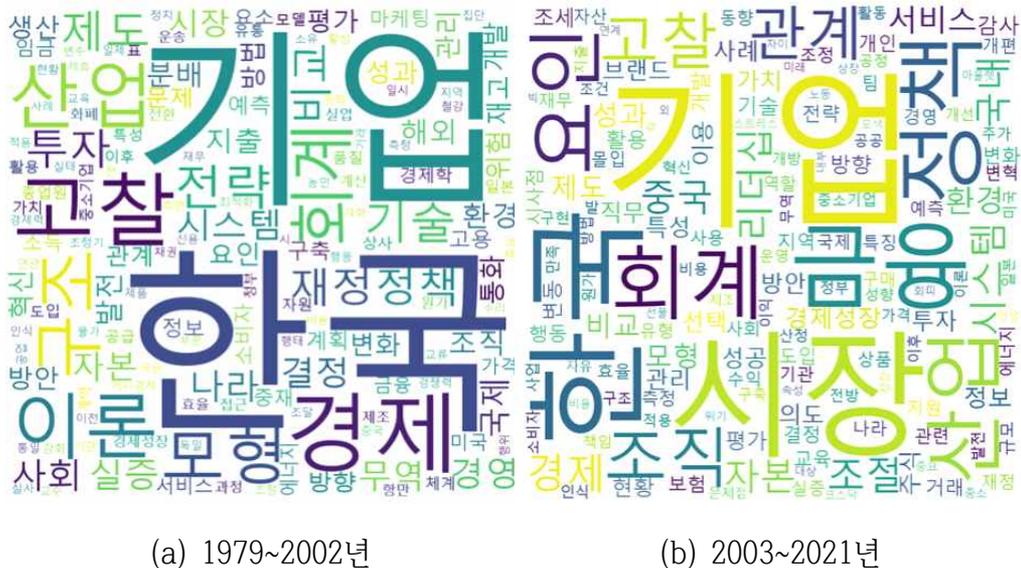


그림 (a)는 초기 1979년~2002년 논문 제목의 워드 클라우드를 보여주고 있다. 논문에서 가장 많이 등장하는 단어는 기업, 한국, 경제 등이다. 그 외에도 산업, 구조, 이론, 모형 등의 단어들이 많이 출현하다. 한국 경제나 기업, 산업 구조 등을 대상으로 한 연구들이 많았음을 보여준다. 방법론에서는 이론이나 모형을 사용한 논문이 많았다.

그림 (b)에서 2003년 이후의 워드 클라우드를 보여준다. 한국이라는 단어보다는 시장이라는 단어가 더 많이 사용된다. 한국이 초기 계획 경제에서 벗어나 시장을

중시하는 흐름으로 이동한 것과 일맥상통한다. 경제라는 작게 줄어들고 단어보다는 회계나 금융, 조직, 관계, 리더십이라는 단어의 중요도가 더 커진 것도 이러한 맥락일 것이다. 정책이라는 단어는 여전히 중요성을 지니고 있다. 한국이라는 단어가 크지만 중국이라는 단어가 등장하여 중국에 대한 관심이 커졌음도 보여준다. 이러한 결과는 산업연구가 한국 경제의 발전에 맞게 연구의 흐름도 변화되고 있음을 보여준다. 또한 시대가 요구하는 필요에 맞게 연구들을 진행하였음을 시사한다.

3. 유사 학술지와 비교 및 향후 발전방향

이 부분에서는 산업연구가 앞으로의 나아가야 할 방향을 더 고찰해 보기 위해 유사학술지들과 비교하는 연구를 수행하였다. 유사학술지에 대한 선정은 'Ⅲ.자료' 부분에서 설명하였다. 결과적으로 4개의 학술지가 선정되었다.

[표 5]는 산업연구와 4개의 유사 학술지를 비교하고 있다. 산업연구를 포함한 4개의 학술지는 등재지이며, 산업연구(JIET)만 등재후보지이다. 산업연구(JIET)는 2017년 창간되어서 가장 역사가 짧지만 발행기관인 산업연구원은 오랜 역사를 가지고 있다. 산업연구원은 1976년 1월 설립된 국무총리 산하 정부출연 연구기관으로 국내외의 산업·기술과 관련된 실물 경제의 동향과 정보를 신속히 수집·조사하고 연구하여 국가의 정책수립과 경제발전, 기업의 생산성과 국제 경쟁력 향상에 기여함을 목적으로 한다. 이런 국책 연구기관의 든든한 지원을 고려하면, 산업연구(JIET)는 앞으로 활발한 연구활동이 기대된다.

산업연구와 산업혁신연구는 대학의 연구소에 발행되고 있다는 공통점을 가지고 있으며, 산업융합연구와 산업경제연구는 학회에 의해서 발행되고 있다. 연구분야에 있어서는 산업혁신연구만이 복합학에 해당되며, 나머지 4개의 학술지는 사회과학 분야에 해당한다. 세부분야는 산업연구는 사회과학일반에, 산업경제연구와 산업연구(JIET)는 경제학에, 산업혁신연구는 경영학에 해당한다. 이런 점에서 보면, 산업연구는 경영학과 경제학의 융합적인 학술지의 성격이 강하다. 경영학은 기업의 활동에 초점을 두며, 경제학은 국가의 활동에 초점을 둔다.

[표 5] 유사학술지 개요

	산업연구	산업융합연구	산업경제연구	산업혁신연구	산업연구(JIET)
상태	등재지	등재지	등재지	등재지	등재후보지
발행기관	단국대학교 미래산업연구소	대한산업경영 학회	한국산업경제 학회	경성대학교 산업개발연구소	산업연구원
연구분야	사회과학 > 사회과학일반	복합학> 학제간연구	사회과학 > 경제학	사회과학 > 경영학	사회과학 > 경제학
창간연도	1979년 12월	2003년 6월	1988년 7월	1985년 10월	2017년 12월
연혁	2019년 등재후보지 2021년 등재지	2019년 등재후보지 2021년 등재지	2003년 등재후보지 2006년 등재지	2007년 등재후보지 2010년 등재지	2020년 등재후보지
발행간기	연3회	연6회	연6회	연4회	연2회
연간평균 논문수	11건	67건	119건	31건	9건
총 논문수	45건	266건	2,613건	462건	44건
IF	1.13	0.98	0.81	1.01	0.19
피인용 횟수	3	23	22	8	0

주) 2022년 5월 28일 기준

발행간기는 산업융합연구와 산업경제연구가 연 6회로 가장 많으며, 다음이 산업 혁신연구 연 4회, 산업연구 연 3회, 산업연구(JIET) 연 2회의 순서이다. 한국학술지 인용색인(KCI)에 등록된 연간 평균 논문 수나 총 논문의 수도 이 순서를 따라간다. 산업경제연구의 연간 평균 논문수가 119건으로 압도적이다. 다음은 산업융합연구 67건, 산업혁신연구 31건, 산업연구 11건, 산업연구(JIET) 9건 순서이다. 학회에서 발행하는 학술지의 논문 수가 연구소에서 발행하는 학술지의 논문 수보다 훨씬 많다. 총 논문수도 산업경제연구가 2,613건으로 가장 많으며, 다음이 산업혁신연구

462건, 산업융합연구 266건, 산업연구 41건, 산업연구(JIET) 44건 순서이다. 산업융합연구의 연간 평균 논문수가 산업혁신연구보다 월등히 많음에도 총 논문수에서 뒤지는 것은 역사가 짧기 때문이다. 산업융합연구는 등재후보지로 등재된 2019년부터의 논문만 KCI를 통해 공개되고 있다.

논문 피인용 횟수도 논문의 수에 비례할 것 같지만 그렇지 않다. 산업융합연구가 23회로 가장 많으며, 그 다음이 산업경제연구 22회이며, 산업혁신연구 8회, 산업연구 3회이다. 산업융합연구가 역사는 짧지만 많은 논문을 활발히 발행하고 있으며, 논문의 피인용도 우수하다. 산업경제연구는 압도적인 총논문수(2,613건)에 비해 피인용되는 횟수(22회)가 너무 작다. 산업연구의 피인용 횟수 3회는 총 논문수 45건에 비하면 꺾은 수치이다.

이런 점들은 영향력지수(IF: Impact Factor)에 그대로 반영된다. 표는 2020년 기준 영향력지수(IF)를 보고하고 있다. 산업연구의 2020년 IF가 1.13으로 5개의 학술지 중에서 가장 높으며, 그 다음이 산업혁신연구 1.01, 산업융합연구 0.98, 산업경제연구 0.81, 산업연구(JIET) 0.19 순서이다. 산업연구가 논문 수는 아직 작지만 피인용도는 높다. 이는 산업연구가 수준 높은 논문들을 생산해내고 있다는 점에서 고무적이라 할 수 있다.

각 학술지의 연구방향을 비교하기 위해 최근 5년간 게재된 논문의 제목을 사용하여 워드 클라우드를 분석하였다. 5개의 학술지 중 산업융합연구는 2019년부터의 논문만 KCI에서 제공하는 관계로 3년의 자료를 사용하였다. [그림 4]는 산업연구와 산업융합연구에 최근 3년 간 게재된 논문의 제목을 대상으로 분석한 워드 클라우드를 보여주고 있다.

[그림 4]의 (a)는 산업연구가 경영학과 경제학의 융합적인 학술지의 성격이 강함을 보여준다. 경영학은 기업의 활동에 초점을 두며, 경제학은 국가의 활동에 초점을 맞추고 있는데, 기업, 시장, 한국 등 해당되는 단어들이 가장 활발히 등장한다. 산업연구라는 학술지의 명칭에 포함된 산업에 대한 연구들은 실제적으로 많이 이루어지지 않고 있음은 아쉽다.

[그림 4]의 (b)는 산업경제연구의 워드 클라우드를 보여주고 있다. 산업경제연구에 많이 등장하는 기업, 산업, 경제, 시장, 사회 등의 단어는 산업연구와 많은 부분이 중첩된다. 산업연구보다 시장은 약한 반면, 산업은 강하게 나타난다. 산업경제연구가 경영학보다는 경제학에 더 치우친 연구경향을 보이며, 이는 산업경제연구가 표방하는 연구분야가 경제학인 것과도 상통하는 것이다. 산업연구는 브랜드, 조직, 회계 등 경영학적인 분야도 많이 취급하고 있다.

그림 (c)는 산업혁신연구의 워드 클라우드를 보여주고 있다. 산업혁신연구에는 기업, 혁신, 기술, 성과 등이 가장 많이 등장하여 학술지의 제목과 일치하는 연구들이 실리고 있음을 보여준다. 기업의 기술 혁신에 가장 많은 관심을 가지고 있음을 보여준다.

그림 (d)의 산업연구(JIET)에서 가장 많이 등장하는 단어는 산업이다. 산업에 특화된 연구들은 산업경제원에서 발행하는 산업연구(JIET)에서 더 활발히 이루어지는 경향이 있다. 제조업, 무역, 생산, 수출 등 전통적인 산업 분야를 다루는데 더 중점을 두고 있지만, 최신 산업동향을 다루는 면에서는 취약해 보인다.

[그림 5] 학술지 제목을 통해 도출한 워드클라우드 (최근 5년)



(a) 산업연구

(b) 산업경제연구



(c) 산업혁신연구



(d) 산업연구(JIET)

우리 산업연구가 제4차 산업혁명을 선도하는 융복합 연구를 게재하는 것을 목표로 한다면 1차적으로 따라가야 할 벤치마크는 산업혁신연구라고 생각한다. 산업혁신연구는 단국대 미래산업연구소와 비슷한 대학 연구소에서 발행하고 있으며, 연4회로 연간 평균 논문수가 31건으로 2010년부터 등재지의 지위를 유지하고 있다. 현재 산업연구가 연3회 발행하는 것을 고려하면, 1회의 발행 논문수를 10건으로 상향시키는 것을 목표로 할 수 있을 것이다. 이 정도의 논문 발행이 가능하다면 연4회 발행하는 것으로 개정하고 1회에 7~8건을 발행하는 것으로 변경하는 것도 고려할만 하다.

영향력 지수(IF)에 있어서는 산업연구가 1.13로 산업혁신연구 1.01보다 더 높은데, 이런 방향은 계속 유지할 수 있도록 힘써야 할 것이다. 논문 발행수가 많아지면 논문의 질을 관리하는 것이 힘들어지고 IF의 하락으로 이어질 수 있다. 이런 현상이 산업경제연구에서 발견된다. 이런 문제점이 발생하지 않도록 논문 발행수를 늘리면서도 피인용 횟수를 지속적으로 향상시킬 수 있는 방법을 강구해야 할 것이다. 이를 위해 다음과 같은 방안을 고려할 수 있다. 첫째, 논문의 질을 유지하기 위해 편집위원회의 지속적인 관심과 노력이 필요하다. 이를 위해 권위자의 논문을 유

치하는 방안을 강구할 수 있다. 또는 연구비 지원을 통해 우수한 논문을 공모하는 것도 하나의 방안일 것이다. 둘째, 산업연구에 게재된 논문에 대한 홍보를 적극적으로 하는 것이다. 편집위원들이 가장 먼저 산업연구에 관심을 가지고 자신의 연구에 산업연구를 적극 인용할 필요가 있다. 또한 산업연구에 논문을 게재한 저자들에게도 산업연구 논문을 적극 홍보할 필요가 있다. 셋째, 산업연구의 저변을 더 확대하는 것이다. 다양한 학회나 연구소, 기업 등과의 세미나나 합동연구를 통해 산업연구에 대한 관심을 고취할 수 있을 것이다.

IV. 결 론

1. 요약 및 한계점

본 연구는 토픽모델링 기법을 통해 산업연구 학술지를 분석하는 것을 목적으로 한다. 산업연구에 게재된 논문의 제목과 국문 요약을 추출하여 토픽모델링 LDA(Latent Dirichlet Allocation)기법을 통해 10개의 토픽을 찾아내었다. 또한 연도별 분석을 통해 산업연구에 게재된 논문의 주제가 어떻게 변화되어 왔는지 살펴보았다. 마지막으로 산업연구가 앞으로의 나아가야 할 방향을 더 고찰해 보기 위해 유사학술지 4개와 비교하는 연구를 수행하였다. 산업연구가 논문의 발행수는 부족하지만 영향력지수(IF)는 가장 우수하였다.

토픽모델을 사용한 연구에서는 토픽수를 정하는 것이 가장 어려운 과제이다. 토픽의 수에 따라 분석결과에서 차이가 발생할 수 있는데 토픽 수를 정하는 기준과 연관하여 연구자간에 합의된 방법이 없어 연구자의 주관에 개입될 수 있는 여지가 많다. 이러한 점이 토픽모델에 기반을 둔 분석의 한계이자 본 연구가 가지는 한계점이라고 할 수 있다.

2. 미래 산업에 대한 시사점

'산업연구' 학술지는 미래 사회에 도움이 되는 다양한 산업분야를 소개하여 한국

의 산업발전에 이바지하는 것을 목표로 한다. 이런 목표를 생각할 때, 본 연구에서 얻어진 분석 결과를 통해 산업연구 학술지의 아쉬운 점은 다음과 같다. 첫째, 산업연구의 논문들이 본래 목표로 지향하고 있는 융복합과 연관된 연구로써의 특성은 부족하였다. 각자의 분야에서 다양한 연구들이 소개되고 있는 성격이 강하였다. 둘째, 4차 산업혁명 등 미래 산업에 관련된 연구들보다는 전통적인 주제들이 많았다. 물론 최근 들어서는 많이 개선되고 있지만 이를 더 적극적으로 추진할 필요가 있어 보인다.

향후 편집위원회는 학술지의 방향에 대한 더 논의할 필요가 있다. 4차 산업혁명 등 미래산업에 초점을 두기 위해 ‘산업연구’를 ‘미래산업연구’로 개명하는 방안도 고려할 수 있지만, 연구 분야를 너무 한정시키는 자충수가 될 우려도 있다. 현재는 논문 발행수를 늘리는 것도 필요하기 때문에 일단 저변을 확대한 이후에 미래 산업과 관련된 분야로 집중할지 논할 수도 있을 것이다. 이는 장기적인 과제이며, 본 연구에서 밝혔듯이 현재 가장 시급한 과제는 지속적 발전을 위해 논문 발행수를 늘리면서도 영향력 지수(IF)를 유지 및 향상시키는 것이다. 이러한 연구방향과 개선점이 ‘산업연구’를 더욱 발전시켜 실제적으로 한국의 미래 산업에 기여할 수 있게 되길 기대한다.

참고문헌

- 권기덕, 2020, 빅데이터와 ICT융복합을 활용한 한국형 스마트팜 모델, 산업연구, 44(3), 69-88.
- 박주섭, 홍순구, 2016, 텍스트마이닝을 활용한 핀테크 동향분석, 산업연구, 40(2), 1-11.
- 박지영, 김영진, 홍사욱, 2015, 빅데이터를 활용한 이케아의 국내진출에 대한 소비자의 평판조사 연구, 산업연구, 39(2), 1-16.
- 윤상철, 2019, 4차 산업혁명시대에서의 디지털 무역과 무역보험, 산업연구, 43(2), 73-90.
- 정재만, 2021. < 재무관리연구> 인용분석 (2003 년~ 2020 년). 재무관리연구, 38(4), 315-337.
- Arnold, T., A. W. Butler, T. F. Crack, and A. Altintig, 2003. Impact: What Influences Finance Research?, Journal of Business, 76(2), 343-361.
- Chan, K. C., C. Chang, and Y. Chen, 2011. Retrospective Analysis of Financial Research Among Korean Institutions and Authors (1990-2010), Asia-Pacific Journal of Financial Studies, 40(4), 599-626
- Chung, K. H., R. A. K. Cox, and J. B. Mitchell, 2001. Citation Patterns in the Finance Literature, Financial Management, 30(3), 99-118.
- Hamelman, P. W. and E. M. Mazze, 1974. Citation Patterns in Finance Journals, Journal of Finance, 29(4), 1295-1301.
- Linnenluecke, M. K., X. Chen, X. Ling, T. Smith, and Y. Zhu, 2016. Emerging trends in Asia-Pacific finance research: A review of recent influential publications and a research agenda, Pacific-Basin Finance Journal, 36, 66-76.

- Blei, D., A. Ng, and M. Jordan, 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993-1022.
- McCallum, Andrew Kachites, 2002. "MALLET: A Machine Learning for Language Toolkit." <http://mallet.cs.umass.edu>.
- Mimno, David, Hanna Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum. 2011. Optimizing semantic coherence in topic models. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 262-272, Edinburgh, Scotland, UK. Association of Computational Linguistics.

Analysis of 'Journal of Industrial Studies' Using the Topic Modeling

- Consideration of the Past and Future -

Cheol-Won Yang* · Akihiko Yamada**

Abstract

This paper aims to analyze the 『Journal of Industrial Studies』 by applying textual analysis. The topic modeling LDA (Latent Dirichlet Allocation) technique and word cloud analysis were performed by extracting the titles and Korean summaries of the papers from the first volume 1 published in 1979 to the 45th volume in 2021. Through this, it was evaluated what topics the papers have covered so far and how the topics have changed. In addition, it was attempted to explore the future direction of the Journal of Industrial Studies through comparison with similar academic journals. As a result of the analysis, 10 topics could be identified through the LDA method, of which 6 were related to business administration and 4 to economics. Journal of Industrial Studies has been appropriately conducting research on topics required by the times. Lastly, when compared with 4 similar journals, Journal of Industrial Studies was the best at the impact factor (IF), but were not enough in published volume. For continuous development, the biggest challenge is to increase the number of papers issued in the journal while maintaining the impact factor.

Key word: Topic Modeling, Journal of Industrial Studies, LDA, Word Cloud, Impact Factor.

* First Author, Professor, School of Business Administration, Dankook University, yang@dankook.ac.kr

** Corresponding Author, Data Scientist, Bigdata Convergence and Open Sharing System, Seoul National University, ayamada5413@snu.ac.kr