

텍스트 마이닝과 소셜 네트워크 기법을 활용한 국제무역 키워드, 중심성과 토픽에 대한 빅데이터 분석

이재득
부산대학교 무역학부 교수

A Big Data Analysis on Research Keywords, Centrality, and Topics of International Trade using the Text Mining and Social Network

Chae-Deug Yi^a

^aDivision of International Trade, Pusan National University, South Korea, South Korea

Received 30 July 2022, Revised 25 August 2022, Accepted 29 August 2022

Abstract

This study aims to analyze international trade papers published in Korea during the past 2002-2022 years. Through this study, it is possible to understand the main subject and direction of research in Korea's international trade field. As the research mythologies, this study uses the big data analysis such as the text mining and Social Network Analysis such as frequency analysis, several centrality analysis, and topic analysis. After analyzing the empirical results, the frequency of key word is very high in trade, export, tariff, market, industry, and the performance of firm. However, there has been a tendency to include logistics, e-business, value and chain, and innovation over the time. The degree and closeness centrality analyses also show that the higher frequency key words also have been higher in the degree and closeness centrality. In contrast, the order of eigenvector centrality seems to be different from those of the degree and closeness centrality. The ego network shows the density of business, sale, exchange, and integration appears to be high in order unlike the frequency analysis. The topic analysis shows that the export, trade, tariff, logistics, innovation, industry, value, and chain seem to have high the probabilities of included in several topics.

Keywords: Trade, Big Data, Text Mining, Network, Frequency, Centrality, Topic

JEL Classifications: F10, C80

^a First and Corresponding Author, E-mail: givethaks@pusan.ac.kr

© 2022 The Korea Trade Research Institute. All rights reserved.

I. 서론

현재 사회는 4차 산업혁명의 시대로 데이터 기반혁신 기술인 빅데이터(Big data)와 인공지능 기법이 거의 모든 분야에서 중요하게 활용되고 있다. 이러한 데이터 기반혁신 기술과 빅데이터 분석 방법은 공학뿐만 아니라 물류, 관광, 경영 등 사회과학 분야에서도 점점 그 활동범위와 방법이 넓어지고 있다. 사회과학에서 주로 사용하는 텍스트 마이닝(Text Mining)은 비정형화 된 텍스트 데이터로부터 키워드를 분석하여 사회적 현상, 추이를 분석한다.

대부분의 경제무역 연구들은 근대 경제학의 신고전파가 주로 사용하는 기법이 구조화된 수치자료 등을 활용한 정량적 계량경제학적 방법은 치밀한 분석을 하고 있지만, 시시각각으로 변화하는 국제이슈와 경제학 모형의 모수들의 변화들을 즉각 반영하기 힘들다. 그리하여 이에 대한 보완으로, 텍스트 마이닝 기법과 소셜 네트워크 분석(Social Network Analysis: SNA) 기법은 주로 비구조화된 비정형의 자료들을 사용하여 연구하고 제공하기 때문에 텍스트 마이닝과 소셜 네트워크 분석 기법은 최근에 변화하고 있는 국제환경과 시시각각의 국제이슈들을 즉각 반영하는 관계를 이해하는데 도움이 될 것이다.

그리하여 본 연구에서도 주요 키워드를 추출하는 텍스트 마이닝과 Carter and Feld (2004)과 Poulin, Boily, and Mâsse (2000) 등의 연구에서처럼 SNA 기법을 통해 계량경제학에서 주로 사용하는 정량분석의 한계를 상호 보완함으로써 무역학 연구에서의 주요 동향과 중요한 이슈를 규명하는데 효율적으로 사용할 수 있다. 물론 기존의 무역학에 관한 논문을 분석한 연구들이 다수 있었지만, 대부분 일부 국가를 중심으로 주로 자유무역이나 경제통합 그리고 관세와 물류 등을 텍스트 마이닝 기법을 이용하여 연구하였다. 그럼에도 불구하고 아직 연구의 초기단계라 경제통합과 관세 등 무역학의 어느 특정 분야에 대한 분석에 치우쳐 있는 등 분석범위와 내용 등이 다소 국한되어 있다.

그리고 현재 무역학 연구 분야에서 빅데이터와 텍스트 마이닝과 토픽분석을 활용한 연구가

Son, Yong-Jung (2018), Suh Jeong-Meen, Han Ye-Eun, and Kim Joung-Youn (2019), 그리고 Yoon Hee-Young and Kwak Il-Youp (2020), 그리고 Lee Jee-Hoon and Kim Jung-Suk (2020) 등의 다수 있지만, 텍스트 마이닝과 소셜 네트워크 및 토픽분석을 같이 융합하여 분석한 논문은 많지 않고, 연구방법 또한 대부분 빈도분석이 많다, 더구나 무역학 연구주제가 경제통합에 주로 치우쳐 있는 등 연구주제를 보면 다소 다양하지 않으며, 경영과 관광 등 타 학문분야에 비해 연구가 상대적으로 많지 않은 실정이다.

그리하여 본 연구는 지난 20년간 무역학 분야 논문을 중심으로 우리나라의 무역학 분야의 연구내용, 방향, 동향 등을 파악하기 위하여 무역학 논문에서 키워드가 무엇인지 텍스트 마이닝과 SNA와 토픽분석을 하고자 하며, 그 정책적 시사점을 제공하고자 한다. 본 연구의 연구방법에서는 비정형 데이터인 텍스트에서 의미와 가치 있는 정보를 찾아내는 기술인 텍스트 마이닝과 단어 간의 관계를 연결망으로 추출할 수 있게 하는 SNA를 주요 분석 방법으로 활용하였다. 그리고 빈도분석과 중심성 분석 외에도 토픽을 분석하기 위해 일부 머신러닝(Machine Learning) 기법을 이용하는 Blei, Ng, and Jordan (2003)이 창안한 잠재 디리클레 할당(Latent Dirichlet Allocation, LDA) 기법을 활용하여 분석하고자 한다.

본 연구대상은 한국연구재단에서 자료를 제공하기 시작한 2002년부터 2022년 6월 현재 전국의 무역학관련 19개 학회의 14,301편의 논문들에서 주로 어떤 무역학 연구가 이루어져 있는지 분석한다. 그러나 19개 학술지들은 한국 무역학회로부터 파생되어 나왔고, 대부분 전공분류상 국제상학 분야 논문들이 많기 때문에 종합적인 무역학 연구를 분석하는데 있어 한국 무역학회 논문들을 대상으로 하는 것이 상대적으로 대표성이 더 있을 것이다.

그리하여 본 연구에서는 종합적으로 19개 무역연관 학회지도 기본적으로 분석하겠지만, 한국무역학회에서 발간하는 무역학 분야 우수등재지인 국문학회지와 SSCI 학회지인 영문 학회지 두 학회지를 묶은 2,004편 논문을 모두 활용

하였다. 먼저 텍스트 마이닝 기법과 무역학 분야 연구에서 타 분야에 비해 다소 소홀이 하였던 무역학 논문들의 주요 키워드와 네트워크를 통한 중심성(Centrality) 분석을 통한 SNA 기법 등을 이용하였고, 그리고 무역연구 네트워크에 대한 연구와 LDA 기법을 사용하여 토픽분석을 하고자 한다. 이와 같이 본 연구는 향후 무역학 분야 논문의 연구내용과 연구 키워드 그리고 향후 연구의 추세 등 시대에 따른 주요 동향을 분석하여 무역학 분야에 있어 심도 깊은 연구를 위한 기초자료로 제공할 수도 있을 것이다.

물론 무역학 분야에서 기존 연구가 주로 경제통합, 관세나 물류 등 일부 분야나 품목을 중심으로 텍스트 마이닝 기법을 이용한 빈도분석 혹은 중심성 분석과 토픽을 분석하고 있다. 그러나 본 연구는 본 연구는 2002-2022년 6월 동안의 우리나라의 모든 무역연구 논문을 대상으로, 텍스트 마이닝과 SNA, 그리고 토픽 분석 등을 종합하여 무역학 연구에서 지난 최근 20 년 동안 이루어진 주요 키워드와 연구 네트워크, 그리고 연구주제 등을 종합적으로 분석함으로써 무역연구에 있어 중요한 이슈와 연구동향 및 시사점을 제공하며, 향후 후발연구들을 유발하는데 기여하고자 한다.

II. 선행연구

기존의 선행연구에서 텍스트 마이닝과 SNA를 이용한 연구가 활발한데 비해, 무역학 분야에서는 대부분 경제통합의 효과에 대한 연구들은 대체로 키워드 빈도분석들은 많이 있는 편이지만 네트워크 중심성 분석이나 토픽분석 논문들은 별로 없는 편이다.

Park Jun-Suk, Kim Chang-Sik Kim, and Kwahk Kee-Young (2016)은 텍스트 마이닝과 간단한 SNA를 통하여 호텔영역에 있어 연구 동향을 연구하였지만, 이 연구는 텍스트 마이닝이나 아이젠벡터 중심성에 대한 연구는 없다. Ko Min-Hwan and Park Yun-Mi (2022)는 빅데이터 기반의 텍스트 마이닝을 활용한 포스트 코로나 시대의 미래관광에 대한 연구를 하였다. Ko Hyun-Jeong (2021)은 주로 공급사슬에

영향을 미치는 요인을 2015년부터 2021년까지 발표된 280편의 논문을 추출하여 상위 10개의 키워드를 분석하고 그에 대한 정책적 시사점을 밝히고 있다. 그러나 중심성에 대한 분석이 거의 없다.

무역학 분야의 기존 문헌들을 보면, Pak Jee-Moon, Kim Sung-Kuk, and Kim Han-Ho (2015)은 SNA를 이용한 1996년과 1997년 및 2011년과 2012년의 굴 수출입 무역 통계를 이용 무역네트워크를 연구하였다, Yoon Hee-Young (2015)은 FTA에 관한 논문 340편을 빈도분석을 통해 연구추세를 분석하였다. Jung sae-Yeon and Lee Chun-Su (2018)는 신발수출 산업의 중심성을 분석하였지만, 많은 수출품 가운데 비중이 작은 한 품목인 굴에 국한된 분석을 하고 있다.

무역학 분야 텍스트 마이닝을 이용한 연구들은 Kang Dong-Joon and Park Keun-Sik (2017)등의 연구에서와 같이 일부 기간의 FTA 혹은 RTA에 대한 연구들이 주를 이루고 있지만, 무역학 전반에 대한 분석은 없다. Suh Jeong-Meen (2020)은 텍스트 마이닝 접근법을 통하여 APEC에서의 지역무역협정(Regional Trade Agreement: RTA)과 자유무역협정(Free Trade Agreement; FTA)의 수렴과 발산에 대한 연구를 하였는데, APEC내의 무역협정국가들은 8-9.7%로 수렴하는 경향이 있는 것으로 나타났다. Suh Jeong-Meen, Han Ye-Eun, and Kim Joung-Youn (2019)은 텍스트 마이닝을 이용하여 한국이 체결한 RTA 간 유사도 분석결과를 Jaccard 유사성(Jaccard Similarity)을 사용하여 한국의 RTA 협정문들의 상호 수렴성과 체결 상대방의 경제적 특성에 대한 연구를 하였다.

Sohn, Sung-Pyo (2020)은 관세관련 무역학 분야 연구추이와 메타분석에 대한 연구를 한국연구재단의 한국학술지인용색인(Korea Citation Index: KCI) 자료 정보를 활용하여 총 546편의 무역학 분야 논문의 주제 동향을 연구하였다. Kang Dong-Joon and Park Keun-Sik (2017)은 1960년대부터 2016년까지 432건의 지역 무역협정을 중심으로 세계 RTA 네트워크 분석을 통한 한국 FTA의 전략과 성과에 관한 연구를 네 개의 중심성 분석을 통해 네트워크 분석은 하

였지만, 토픽분석은 하지 않고 있다.

Son, Yong-Jung (2018)은 한국의 기체결 FTA의 경제적 효과분석을 SNA를 이용한 체결 국간 네트워크를 중심으로 분석하였는데, 주로 연결 중심성(Degree Centrality)과 근접 중심성(Closeness Centrality) 및 매개 중심성(Betweenness Centrality)을 분석하였다. Chung In-Kyo, Min Wei, and Yoo Jeong-Ho (2019)는 중국의 자유 무역협정 체결과 무역네트워크의 변화를 연구하였는데, 중국은 2001년 세계무역기구에 가입하여 2018년 7월까지 24개의 국가와 16개의 FTA를 체결하였는데, 중국의 FTA 체결확대가 FTA 체결 상대국과의 무역관계를 어떻게 변화시켰는지 SNA를 통해 연결정도 중심성, 근접 중심성, 매개 중심성을 이용하여 분석하여 한국의 통상정책에 대한 시사점을 분석하고 동아시아 무역에서의 중국 역할과 아세안 국가의 중요성을 분석하였다.

Yoon Hee-Young and Kwak Il-Youp (2020)은 2002년부터 2019년까지 무역학 분야 국내학술지 논문 11,774편을 이용하여 한국무역분야 키워드 빈도를 주로 분석하였지만, SNA를 활용한 중심성은 분석하지 않았다. Lee Jee-Hoon and Kim Jung-Suk (2020)은 2002년부터 2019년까지 국내 무역학 주요 학술지인 무역학회지, 국제통상연구, Journal of Korea Trade에 수록된 논문 1,868편의 초록을 이용하여 LDA 알고리즘을 이용하여 Blei, Ng, and Jordan (2003) 연구를 바탕으로 양적인 토픽분석을 하였다. 그리고 Rha Jin-Sung (2022)은 코로나 19 팬데믹 이후 컨테이너선 운임 상승의 요인을 텍스트를 중심으로 네트워크 분석과 LDA 토픽 모델로 2020년 1월부터 2021년 7월까지 로이즈리스트에 게재된 기사들을 대상으로 분석하였다. 분석한 결과 미국과 중국의 무역마찰, 글로벌 생산 감소를 예측한 글로벌 선사들의 기항 횟수 감소와 임시결항, 수에즈 운하의 봉쇄 등을 주요 원인으로 분석하였다. Park Hyun-Hee and Cho Sung-Je (2020)는 텍스트 마이닝을 사용하여 무역보험에 대한 간단한 메타분석을 하고 지만, 연구방법면에서 정량적인 분석으로는 다소 미약하다.

그러하여 위에서 선행연구를 살펴본 결과,

무역학 분야에서도 텍스트 마이닝 방법에 의한 관세와 물류, 경제통합과 FTA에 대한 연구가 다수 있음에도 불구하고, 아직 우리나라의 무역학 분야 전반에 대한 연구동향과 연구주제에 대한 연구는 별로 없고, 더구나 연구방법면에서 무역학 연구에 대한 텍스트 마이닝과 SNA에 대한 중심성 분석을 융합한 연구들과 특히 토픽분석을 한 연구들은 거의 없는 편이다.

따라서 본 연구는 무역학 연구에 있어 주요 이슈나 동향을 분석하기 위해 2002-2022년 6월 동안 이루어진 우리나라 무역학 연관 19개 학회의 14,301편 논문들과 동시에 한국무역학회지와 Journal of Korea에 실린 논문 2,004편을 중심으로 텍스트 마이닝 분석과 SNA 분석 등을 한다. 따라서 본 연구는 연구방법 측면에서 키워드의 빈도분석뿐만 아니라 다양한 양적인 네트워크 중심성 분석과 LDA기법에 기반한 토픽 분석 등 종합적인 분석을 하고자한다. 그리하여 본 연구는 무역학 연구에서 과거와 현재 그리고 미래의 무역학 분야에 있어 중요한 연구내용과 범위, 연구주제 및 무역학 연구동향을 종합적으로 분석한다는 점에서 기존 연구와 차별성을 가진다.

Ⅲ. 연구분석의 이론적 배경

1. 텍스트 마이닝(Text Mining)과 빈도분석

Paranyushkin (2011)에 의하면, 텍스트 마이닝의 분석방법으로는 텍스트 분류, 텍스트 군집화, 정보추출, 네트워크분석, 감성분석 등이 있는데 이들은 텍스트에 나타난 주요 개념을 추출하는 것은 물론 다른 개념과의 관계를 파악하고 시각화 하는데 유용하다. Han, Jin-Sung and Ji-Hwan Yoon (2016)의 연구를 참고하면, 텍스트 마이닝은 정보검색과 정보추출 및 자연어 처리 기술을 사용하여 알려지지 않은 유용한 패턴과 지식을 텍스트로부터 발견하는 것을 말한다.

Kim, Nam-Gyu, Lee Domg-Hoon, Choi Ho-Chang, and Wong William Xiu Shun

(2017)의 연구에 나타나 있듯이, 텍스트는 인류가 정보를 표현하고 전달하는 데에 사용해 온 가장 대표적인 수단이다, 그리고 상당히 많은 양의 정보가 이미 텍스트에 포함되어 있을 뿐 아니라 최근 다양한 소셜 미디어를 통해 유통되는 비정형 데이터의 양이 급증하고 있어 텍스트 분석을 통해 새로운 지식을 창출하고자 하는 수요가 증가하고 있다, 따라서 텍스트 마이닝 분석은 급격히 성장하고 있다.

그리고 워드 클라우드(Word Cloud)는 텍스트 데이터의 대표적인 시각화 기법 중의 하나로 비정형 데이터인 텍스트는 워드 클라우드를 통해 좀 더 직관적으로 데이터의 구조를 표현할 수 있다. 이와 같이 키워드를 시각화하여 직관적으로 파악하기 위하여 워드 클라우드 기법을 사용한다.

2. 소셜 네트워크와 중심성 분석

소셜 네트워크 분석은 분석 대상 및 분석 대상들 간의 사회적 관계를 노드(node)와 라인(line)을 이용하여 연결망 구조 즉 체계적인 소시오그램(sociogram)으로 표현하고 연결 패턴을 계량적으로 제시하는 분석 방법이다(Kwahk, Kee-Young (2017)). 소셜 네트워크 분석은 주로 사람이나 조직의 사회적 관계 또는 행위자에 의해 인식되는 사회관계를 파악하는 데 사용되며 네트워크 분석의 구조적 특성을 이해하기 위해 밀도(density), 중심성(centrality), 집중성(centralization) 등의 지표가 사용된다(Park Jun-Suk, Kim Chang-Sik Kim, and Kwahk Kee-Young, 2016).

본 연구에서는 이를 위해 무역학 연구 네트워크에서 중요한 키워드로 이루어진 어떤 노드 또는 링크가 다른 노드들에 비해 더 중심적임을 나타내는 중심성 분석은 연결 중심성 혹은 연결정도 중심성(Degree Centrality), 근접 중심성(Closeness Centrality), 매개 중심성(Betweenness Centrality)을 구하여 분석하였다¹⁾.

(1) 연결(degree) 중심성:

첫 번째 중심성 지표는 연결 혹은 연결정도(degree) 중심성으로 네트워크에서 노드로 표시된 액터(actor)에 직접적으로 연결된 링크의 수를 말한다. 특정 노드의 연결정도는 네트워크 내에서 액터의 파워에 대한 가장 간단하고 효과적인 지표로 사용되기 때문에 노드 자체의 중심성을 나타내는 지표로 활용되는데, 연결 중심성은 다음과 같이 정의된다.

먼저 g개의 노드로 표시되는 액터(actor)를 가지고 있는 비방향 그래프에서 네트워크에서 액터 i 개의 연결 중심성은 액터 i가 나머지 (g-1)개의 다른 액터와 선으로 연결되어 있는 개수를 합산하여 표시된다고 할 때, 네트워크에서 액터 i의 연결 중심성($C_D(N_i)$) 다음과 같은 식으로 나타낸다.

$$C_D(N_i) = \sum_{i=1}^g x_{ij}, i \neq j.$$

여기서 g는 액터의 개수, $\sum_{i=1}^g x_{ij}$ 는 액터 I가 (g-1)개의 다른 액터와 갖는 연결관계의 계수로 x_{ij} 는 연결되어 있지 않으면 0 또는 연결되어 있으면 1의 값을 갖는다. 따라서 큰 네트워크에 속한 액터는 작은 네트워크에 속한 액터에 비해 연결된 네트워크의 크기가 커지면 연결 중심성은 더 커진다. 그리하여 네트워크 크기가 연결 중심성에 미치는 영향을 제거하기 위하여 표준화된 연결 중심성[$C_D'(N_i)$]을 Wasserman and Faust (1994)와 Knoke and Yang(2008)의 연구에서와 같이 다음과 같이 사용하기도 한다. 따라서 $C_D'(N_i)$ 값은 고립노드에서는 0, 그리고 모든 노드와 연결된 관계를 갖는 경우 1을 가진다.

$$C_D'(N_i) = \frac{\sum_{i=1}^g x_{ij}}{g-1}, i \neq j.$$

그러나 연결 중심성은 직접 링크된 이웃 노

1) Kwahk (2019), Social Network Analysis, pp.185-194 참조.

드의 링크 수만 고려하기 때문에 절대적인 링크 수로 측정된다.

(2) 근접(Closeness) 중심성

근접(Closeness) 중심성은 연결 중심성과는 달리 네트워크 내에서의 간접적 연결을 또한 고려하여 액터와 다른 액터 간의 거리를 강조하는 중심성 개념이다. 그리하여 근접 중심성은 네트워크 상에서 한 노드가 다른 노드에 얼마나 가까이 있는가를 나타내는 지표로서, 한 액터가 다른 액터들과 직접 연결 혹은 매개자를 통해 얼마나 빨리 접근하고 소통할 수 있는가를 나타낸다. 따라서 근접 중심성은 해당 네트워크의 전체적 관점에서 직접효과뿐만 아니라 간접효과도 포함하는 노드의 중심성을 고려하는 개념이다. 따라서 근접 중심성($C_C(N_i)$)은 다음과 같이 정의된다.

$$C_C(N_i) = \frac{1}{\sum_{j=1}^g d(N_i, N_j)}, i \neq j.$$

여기서 $\sum_{j=1}^g d(N_i, N_j)$ 는 액터 i 와 j 간의 최단 경로거리의 합이며, g 는 액터의 개수를 나타낸다. 한편, 액터의 근접 중심성을 네트워크의 크기를 고려하여 표준화하면, Freeman (1979)의 연구에서와 같이 다음과 같이 정의된다.

$$C'_C(N_i) = (g-1)[C_C(N_i)]$$

(3) 매개(Betweenness) 중심성

매개(Betweenness) 중심성은 네트워크 상에서 임의의 액터 혹은 노드가 직접 연결되어 있지 않는 액터 혹은 노드들 간에 매개자를 통해 매개 혹은 통제하는 정도를 나타낸다. 따라서 네트워크 상 정보교환을 통해 매개 혹은 통제력을 나타내는 중요한 영향력 지표로 Freeman (1979)은 매개 중심성($C_B(N_i)$)은 다음과 같이 정의하고 있다.

$$C_B(N_i) = \sum_{j < k} \frac{g_{jk}(N_i)}{g_{jk}}, i \neq j \neq k$$

여기서 g_{jk} 는 두 액터 j 와 k 간의 최단경로 개수, $g_{jk}(N_i)$ 는 두 액터 j 와 k 간의 최단경로 가운데 액터 i 를 포함하고 있는 경로의 개수를 말한다. 그리고 액터 i 의 최대 매개 중심성은 $\frac{(g-1)(g-2)}{2}$ 가 되므로, 표준화된 매개 중심성($C'_B(N_i)$)은 Wasserman and Faust (1994) 연구에서와 같이 다음과 같이 정의한다.

$$C'_B(N_i) = \frac{2 \times C_B(N_i)}{(g-1)(g-2)}.$$

그러므로 표준화된 매개중심성은 0과 1사이의 값을 가지며, 1에 가까울수록 액터의 중개 또는 통제능력은 증가한다.

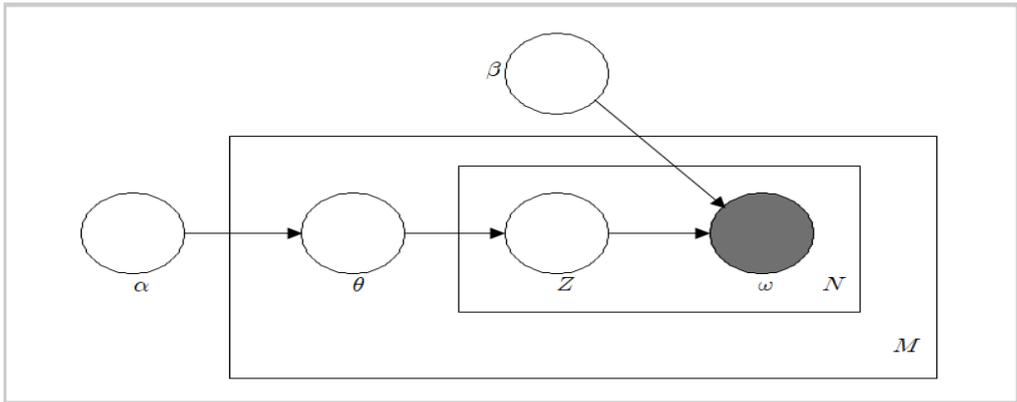
(4) 아이겐벡터(Eigenvector) 중심성

아이겐벡터(Eigenvector) 중심성은 임의의 노드와 연결되어 있는 이웃 노드가 얼마나 중요한가를 함께 고려한다. Bonacich (2007)가 주장하듯이, 연결 중심성의 개념을 확장한 것으로, 연결 관계뿐만 아니라 한 노드와 연결된 다른 노드의 중요성을 가중치를 고려한 중심성 지표이다. 액터 i 의 아이겐벡터의 중심성($C_E(N_i)$)은 다음과 같이 측정한다.

$$C_E(N_i) = \lambda \sum_{j=1}^g x_{ij} C_E(N_j), i \neq j.$$

여기서 λ 는 아이겐 값을 나타내고 x_{ij} 는 액터 i 와 j 간의 0과 1로 구성된 연결관계의 값을 나타내는 값이다. 그러므로 높은 아이겐벡터 중심성은 연결관계와 액터의 중요성을 동시에 포함하는 개념으로 연결 중심성을 확장한 개념이다. 따라서 강한 아이겐벡터 중심성을 갖고 있는 노드는 그 네트워크상에서 강한 연결 중심성을 갖고 있다.

Fig. 1. Graphical Model of LDA in Topics



source: Blei, Ng, and Jordan (2003)

(5) 에고 연결(Ego Degree)과 응집성

SNA는 분석의 초점에 따라서 자기중심의 액터와 그 액터와 연결된 그 외 다른 액터의 연결로 구성된 에고 네트워크(Ego Network)를 분석할 필요가 있다. Carter and Feld (2004)과 Poulin, Boily, and Mâsse (2000) 등에 의하면, 어떤 특정 노드를 중심으로 네트워크의 최적화 개념은, 밀도(Density)로서 표시되는데 밀도란 한 네트워크에서 노드사이의 연결이 이루어진 정도를 나타내며 네트워크 내 전체 노드가 서로 간 얼마나 많은 관계를 맺고 있는 가를 나타내는 것으로 응집성 개념이다. 그러므로 밀도가 연결관계는 적어질 수도 있다. Knoke and Yang (2008)의 연구를 따라, 본 연구에서는 비방향의 밀도(D)를 L을 라인의 개수라고 할 때, 다음과 같이 계산하여 사용한다.

$$D = \frac{L}{gC_2} = \frac{L}{\frac{g(g-1)}{2}} = \frac{2L}{g(g-1)} = \frac{\sum_{i=1}^g d(N_i)}{g(g-1)}$$

여기서 $gC_2 = \frac{g!}{(g-2)! \times 2!}$ 을 나타내고,

연결노드의 합은 $\sum_{i=1}^g d(N_i) = 2L$ 로 나타난다.

3. 토픽 모델링

토픽(Topic)은 머신러닝 기반의 텍스트 마이닝 기법의 하나로 단어나 문서의 집합에서 더 큰 주제로 묶어내는 것을 찾아내는 분석기법이다. 그리하여 토픽 모델링(Topic Modeling)은 비구조화된 문서집합에서 의미있는 주제들을 추출해 주는 확률적 모델 알고리즘이다. 토픽 분석은 기존의 분석에서 발견하지 못한 잠재된 의미구조를 파악하고 새로운 정보와 가치를 발견을 할 수 있게 해준다.

토픽모델링 중에서 사용하는 알고리즘은 주로 Blei, Ng, and Jordan (2003)이 창안한 LDA 모형을 가장 많이 사용한다. LDA는 주어진 연구문헌들의 요약이나 키워드 등 관찰된 변수를 통해 연구문헌들의 구조와 같은 보이지 않는 변수를 추론한다. 그리하여 결과적으로 전체 문헌집합의 주제와 각 연구문헌별 주제비율, 각 단어들이 각 주제에 포함될 확률들을 계산해 낸다. 따라서 각 문서에 어떤 주제들이 존재하는지에 대해 확률모형으로 주제별 단어분포나 문서별 주제의 분포를 모두 추정할 수 있다.

Blei, Ng, and Jordan (2003)은 LDA 모형을 <Fig. 1>과 같이 나타내고 있다.

<Fig. 1>에서 M은 전체 논문의 개수를 나타내고, 문서는 N개의 단어로 구성되어 있고 w는 단어를 나타낸다. N은 특정 문서에 속한 단어

Table 1. All Trade Related Journals and Number of Published Papers

Journal	Paper	Journal	Paper	Journal	Paper	Journal	Paper
International Trade & Commerce	2,345	Korea Research Society for Customs	1,124	International Trade and Industry Studies	434	Journal of International Logistics	219
Korea Trade Review	1,497	Distribution and Management Research	879	Journal of Korea Trade	476	International Journal of Trade Fairs and Exhibition Studies	215
e-Business Studies	1,533	INTERNATIONAL COMMERCIAL & LAW REVIEW	824	E-Trade Review	417	Corporation and Innovation	126
International Commerce and Information Review	1,255	ARBITRATION STUDIES	542	Korea Research Association of International Commerce	429	Global Trade & Management	68
KOREA INTERNATIONAL COMMERCE REVIEW	1,094	International Trade and Insurance	509	Asian Journal of Shipping and Logistics	317	Total	14,301

의 개수를 나타내며 포아송(Poisson) 분포를 한다. Z 는 해당 단어가 속하는 주제를 나타낸다. α 와 β 는 초모수(hyper parameter)이며, 밖의 사각형 상자는 논문을 나타내고, 안의 사각형 상자는 토픽과 단어들의 반복된 선택을 나타낸다. 말뭉치는 M 개의 문서의 수집에 의해 정해진다. θ 는 각 논문에 배당되는 토픽분포로서 모두 d 개의 디리클레(Dirichlet) 분포를 하고 있다고 상정한다. 그리하여 초모수인 α 와 β 가 주어졌다고 가정할 때, 토픽 θ , Z , 그리고 w 의 결합분포는 다음과 같이 주어진다.

$$p(\theta, Z, w|\alpha, \beta) = p(\theta|\alpha) \prod_{n=1}^N p(Z_n|\theta) p(w_n|Z_n, \beta).$$

먼저 θ 에 대해 적분하고 Z 에 대해 합하면 다음과 같은 한계(Marginal) 분포를 얻을 수 있다.

$$p(w|\alpha, \beta) = \int p(\theta|\alpha) \left(\prod_{n=1}^N \sum_{Z_n} p(Z_n|\theta) p(w_n|Z_n, \beta) \right) d\theta.$$

다음 각각의 논문들의 한계 확률의 곱을 취하면, 다음과 같은 말뭉치의 확률을 얻을 수 있다.

$$p(D|\alpha, \beta) = \prod_{d=1}^M \int p(\theta_d|\alpha) \left(\prod_{n=1}^{N_d} \sum_{Z_{dn}} p(Z_{dn}|\theta_d) p(w_n|Z_{dn}, \beta) \right) d\theta_d.$$

그리하여 LDA 분석을 통해 사전에 α 와 β , 토픽의 개수를 정한 후 Z , θ , d 를 구한다. LDA를 실시할 때 연구자는 한다. 그리고 LDA는 논문 내의 단어들, 즉 w 를 관측해 나가면서 각 단어마다 임의의 주제 Z 값을 정한다. 그리고 이 결과에 따라 θ , φ 의 디리클레분포를 업데이트하

Table 2. Annual Papers of All Trade Related and Korea Trade Research Association

Year	Number of Papers in 24 Trade Related Journal	Number of Papers in Korea Trade Review and Journal of Korea Trade
2002	126	65
2003	236	97
2004	328	88
2005	398	70
2006	515	97
2007	545	110
2008	531	86
2009	595	97
2010	704	83
2011	746	91
2012	770	95
2013	856	92
2014	967	86
2015	871	90
2016	887	79
2017	897	87
2018	944	83
2019	986	159
2020	1,004	138
2021	981	150
2022.1-2022.6	414	61
2002-2022.6	14,301	2,004

고 반복과정을 통해 가장 가능성이 높은 Z값을 찾아내고, θ , 그리고 d를 추정하기 위해 (Fig. 1)에서 나타난 것과 같이 반복적인 시뮬레이션을 통해 값을 얻는다.

따라서 특정 주제에 관한 논문에는 그 주제에 관한 단어가 다른 단어들에 비해 더 자주 등장하는 단어들을 군집화하여 잠재된 주제를 추론한다. 그리하여 주제별 단어 분포를 바탕으로 주어진 문서에서 발견된 단어의 확률분포를 분석함으로써, 앞으로 어떤 주제를 다루고 있을지를 예측한다.

IV. 실증분석

본 장에서는 무역학 분야의 빅데이터 논문자료를 기반으로 전처리를 진행하고, 텍스트 마

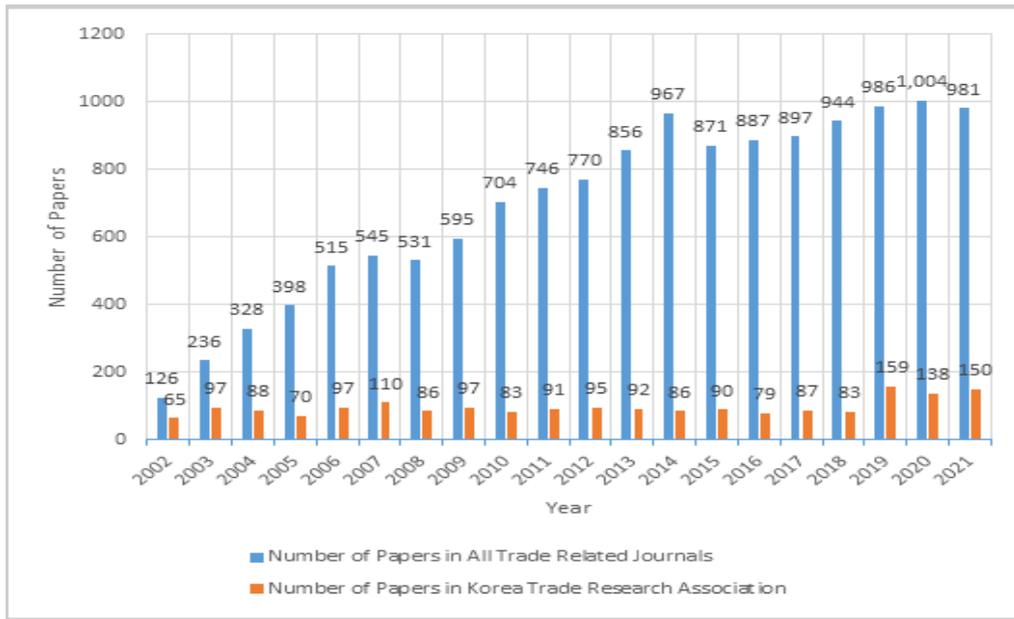
이닝 분석기법을 통해 텍스트에서 출현하는 단어의 빈도 분석과 여러 가지 다양한 중심성 분석을 통해 키워드 간의 연관관계를 도출하고, 토픽분석을 통하여 이를 기반으로 무역학 분야의 연구내용과 내용 초점, 그리고 무역연구 관계 동향을 분석한다.

1. 연구분석 자료와 전처리 방법

우리나라 무역학 연구관련 논문자료를 사용하는 본 연구는 우리나라에서 발간된 무역학 연구 관련 학회논문을 KCI 대분류-중분류-무역학 자료를 이용하면 2002년부터 2022년 6월까지 모든 연구 즉 우수등재지인 431편, KCI 등재 12,334편, KCI 후보 1,536편 등의 논문을 합하면 14,301편이 된다.

그러나 여기에 포함된 19개 학술지등은 무역

Fig. 2. Annual Papers of All Trade Related and Korea Trade Research Association



학과 다소 동떨어진 논문들도 혼재 되어 있고, 학문 영역과 범위의 경계가 불분명하고 대부분 한국무역학회에서 파생되었지만, 주제별로 다양한 차이도 있어, 부득이 하게 무역학 분야에서 우리나라에서 가장 오래되고 한국연구재단 KCI 공인 유일한 우수등재지인 한국무역학회에서 발간하는 무역학회지 1,497편과 유일한 영문 SSCI 학회지인 Journal of Korea Trade 507편을 대상으로 총 2,004편의 논문 들을 발췌하여 연구하였다.

해당 연구대상 학술논문지는 <Table 2>에 나와 있는 것과 같이 연도별 학술지는 각 지역 별로 상세히 나와 있다. 각 년도 별로 해당 논문과 논문개수가 <Table 2>와 <Fig. 2>에서 나와 있는 대로 한국무역학회에서 발간하는 논문 편수는 어느 정도 연도가 갈수록 일정하나, 기타 학회지에서 논문편수는 추세선과 동향을 보면 알 수 있듯이, 기하급수적으로 크게 증가하고 있다.

한편, 본 연구에 필요한 논문 자료는 <Table 2>에 나와 있듯이, 우리나라의 모든 한국연구재단 등재후보 혹은 등재지에 게재된 연구 간

행물을 대상으로 하기에는 논문편수의 일관성이 많이 떨어지고, 논문의 게재편수와 연간 학회지 발간 등이 학회지 별로 너무 다르고, 한글 키워드는 없는 경우가 많았다. 그리하여 연구 분석의 일관성을 위하여 한국무역학회에서 발간하는 두 개의 학회지 즉 한글 우수등재지인 무역학회지와 SSCI 영문학회지인 Journal of Korea Trade에 게재된 총 2,004편의 논문들을 대상으로 주로 영문 키워드를 수집하였다.

그리하여 본 연구에서는 영문 키워드를 중심으로 분석하였으며, 영문 키워드에서는 명사만을 대상으로 하였고, 대명사, 형용사, 부사, 동사, 전치사 접속사 등은 제외하였다. 수집된 키워드는 연구자에 의해 단어 정규화, 불용어 제거 등 데이터 전처리 작업을 진행하였다. 데이터 전처리 과정은 동일한 의미를 지닌 단어를 하나로 통합하였으며, 연구논문의 키워드에서 주로 쓰는 paper, research, 그리고 study 등과 별 의미가 없고 본 연구의 분석과 연관이 없는 낱말은 제거하였다.

그 다음에 빈도분석과 그 단어들의 네트워크 분석은 Netminer 4를 통해 분석하였고 여러 가

Table 3. Frequency of Key Word in Total Trade Related Journals

	2006		2010		2014		2018		2021		2002-2022.6	
	word	no.	word	no.	word	no.	word	no.	word	no.	word	no.
1	Trade	14	trade	24	trade	34	arbitration	31	Market	37	Trade	279
2	strategy	11	logistics	22	risk	30	trade	23	export	36	Export	171
3	firm	11	investment	21	export	29	export	23	trade	33	performance	71
4	Improvement	11	export	21	Industry	26	industry	18	Industry	33	Logistics	69
5	Relationship	8	system	20	performance	25	sale	17	performance	25	Intention	66
6	eBusiness	7	strategy	19	Law	25	ecommerce	17	efficiency	23	Market	59
7	export	6	market	19	market	25	performance	16	company	21	Agreement	54
8	Negotiation	6	quality	18	management	24	customer	16	support	20	Service	52
9	efficiency	6	service	17	industry	23	tariff	16	network	19	Digital	52
10	customer	6	port	17	service	23	service	15	chain	19	Arbitration	50
11	Price	6	customer	17	insurance	21	product	15	industry	18	Industry	43
12	Economy	5	law	16	satisfaction	21	market	15	system	16	International	41
13	plan	5	performance	15	investment	21	firm	14	Import	16	Value	39
14	measure	5	satisfaction	15	exchange	20	satisfaction	14	Convention	15	Port	38
15	industry	4	industry	14	product	20	risk	14	risk	13	System	37
16	performance	4	tariff	14	customer	19	origin	13	management	13	satisfaction	37
17	market	3	product	13	system	19	chain	13	logistics	13	FTA	37

note: The no represents the number of frequency.

지 중심성 분석과 네트워크 분석(Network Analysis)을 하였다. 그리고 토픽연구를 위해 머신러닝 기법인 LDA 알고리즘을 Python을 이용하여 토픽분석을 한다.

2. 우리나라 무역학 분야 전체 학회지 논문 키워드 분석

먼저 2002년 1월부터 2022년 6월 동안에 게재된 우리나라 전체 19개 무역연관 학술지 14,301편의 논문을 대상으로 텍스트 마이닝 기법을 이용하여 대략적으로 우리나라 무역학 연구동향을 살펴본다. 그리하여 <Table 3>에 나와 있듯이, 2002년부터 2018년 동안은 4년 마다, 그리고 최근 2021년의 게재된 논문들의 연구 키워드 빈도(frequency)를 조사하였다. 그리고 <Table 3>의 마지막 칸에는 2002년부터

2022년 6월까지 무역연관 학술지에 게재된 14,301편의 논문들 모두에서 발견된 키워드를 조사하였다.

그리하여 텍스트 마이닝에 의해 조사한 결과, 한국연구재단의 무역과 연관된 14,301편의 논문들 중에서, 적절한 논문 편수를 선정하기 위하여 키워드의 출현 빈도수가 10개 이상인 것만 선택하여 72개의 키워드를 추출하였다. 그러나 지면관계상 키워드의 출현빈도 순위가 17위 까지만 <Table 3>에서 나타내었다. 그 결과, 2006년에는 trade 14회, strategy 11회, firm 11회 등이 가장 빈번한 연구 키워드로 나타났으며, 2010년에는 trade 24회, logistics 22회, investment 21회, 그리고 export 21 회 등으로 가장 연구가 많은 키워드로 나타났다.

Fig. 3. Annual Trend of Frequency of Key Word in All Trade Related Journals

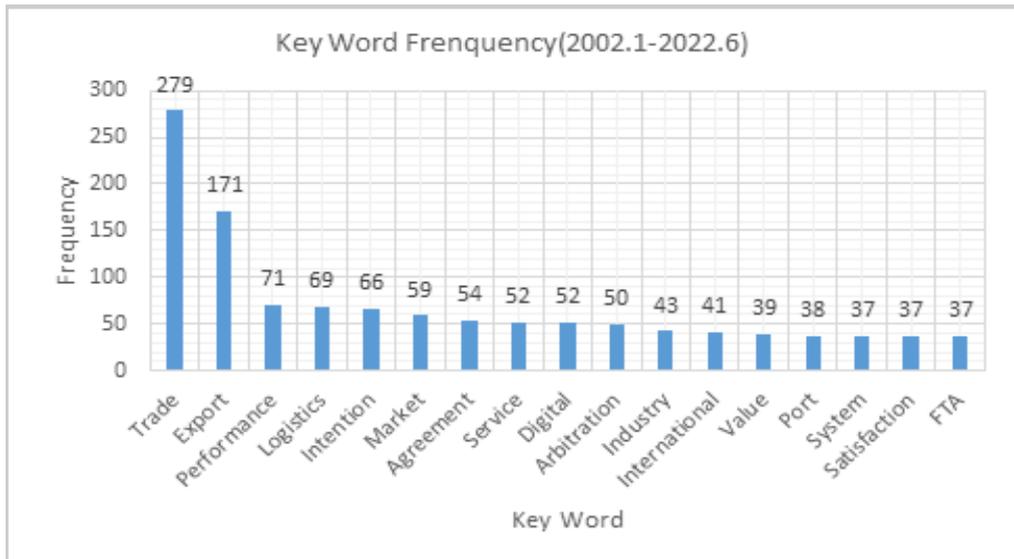
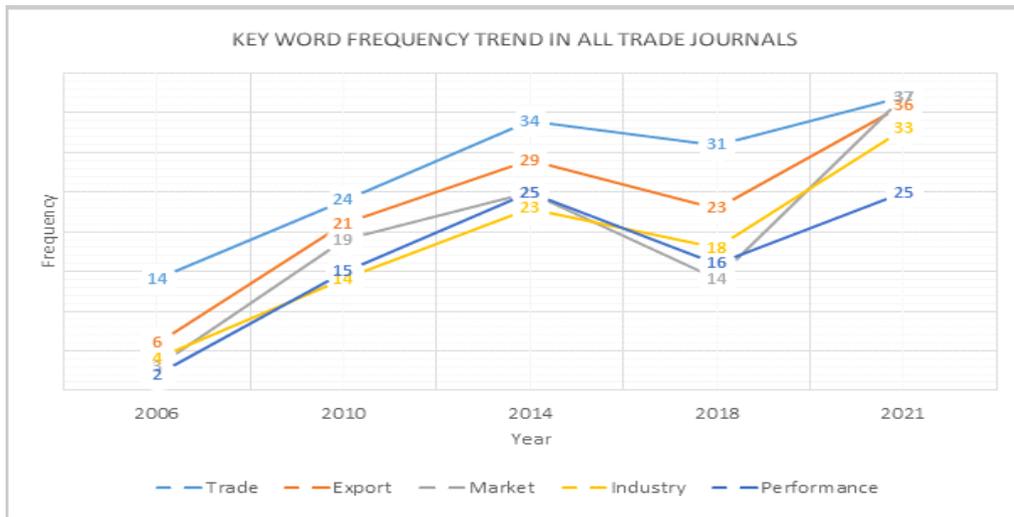


Fig. 4. Total Frequency of Key Word in All Trade Related Journals



2014년에는 trade, risk, export, 그리고 industry가 가장 많은 연구 키워드로 나타났다. 2018년에는 arbitration이 가장 높게 나타났고 trade, export, 그리고 industry 등의 순으로 높게 나타났다. 그리고 ecommerce가 6위로 빈번히 나타난 연구키워드로 부상되었다. 최근

2021년에는 market 37 회, export 36 회, trade 33 회, 그리고 industry 33회 등으로 그 빈도수가 모두 30회를 넘고 있으며, 높은 순으로 나타났다. 그리하여 우리나라 무역학에서는 사실 가장 중요한 이슈들인 무역, 수출, 시장, 물류 등이 가장 빈도가 높은 연구 키워드로 나타났다.

Table 4. Frequency of Key Word in Journals of Korea Trade Research Association

Rank	Word	Frequency	Rank	Word	Frequency	Rank	Word	Frequency	Rank	Word	Frequency
1	export	73	16	value	24	31	origin	15	46	insurance	13
2	trade	69	17	policy	23	32	strategy	15	47	knowledge	13
3	effect	41	18	efficiency	22	33	term	15	48	port	13
4	performance	39	19	measure	22	34	test	15	49	price	13
5	model	37	20	service	22	35	Volume	14	50	rule	13
6	Innovation	35	21	chain	20	36	advantage	14	51	Level	12
7	investment	35	22	logistics	20	37	competitive -ness	14	52	credit	12
8	firm	33	23	Capacity	19	38	country	14	53	negotiation	12
9	tariff	33	24	product	19	39	guarantee	14	54	network	12
10	analysis	30	25	arbitration	18	40	index	14	55	quality	12
11	rate	30	26	company	18	41	subsidiary	14	56	Breach	11
12	industry	28	27	good	18	42	system	14	57	agreement	11
13	factor	26	28	exchange	17	43	theory	14	58	capability	11
14	Consumer	24	29	management	16	44	Economy	13	59	enterprise	11
15	market	24	30	risk	16	45	eTrade	13	60	import	11

그러나 <Fig. 3>에서 나와 있듯이 2002년부터 2018년까지 매 4년 마다, 그리고 2019년부터 2021년까지는 3년간의 키워드를 조사한 결과, 이러한 키워드들의 빈도순위는 자유무역협정 등의 국제적인 경제무역 환경의 변화에 따라 연구 키워드 순위가 조금씩 변경된 것으로 나타났다. 그럼에도 불구하고 높은 순위에 있는 키워드 순위는 주로 변동하지 않고 여전히 trade, export, market, 그리고 industry 등이 상위 순위를 차지하고 있다. 그러나 2006년 eBusiness, 2018년 ecommerce 등이 각각 6위로 나타나 전자비즈니스 혹은 전자상거래 등에 대한 연구가 점점 많아지는 것으로 나타났고, 15위 보다 하위 순위에 있는 키워드들의 순위는 다소 변동이 있는 것으로 나타났다.

그 다음은 <Table 3>과 <Fig. 4>에서 나타나 있듯이, 2002년부터 2022년 6월까지 14,301편 모두의 키워드를 데이터 마이닝 기법으로 조사해본 결과, trade 279회, export 171회로 각각 나타나 무역연관 학술지에 압도적으로 많은 무역학 연구 키워드로 나타났다. 그 다음

performance 71회, logistics, 69회, market 59회 등으로 빈도수가 높게 나왔으며, 한국-EU, 한국-미국, 그리고 한국-중국 등의 자유무역협정을 비롯한 무역협정, 현재의 디지털 전환과 산업시스템 등의 영향으로 agreement, digital, industry, 그리고 FTA 등도 37회 이상의 높은 빈도를 보이고 있다.

3. 한국 무역학회 논문 빈도분석

앞에서 19개의 무역연관 학회지에 게재된 14,301편을 조사하였지만, 다소 학회지 마다 초점이 다르고, 발행연도와 게재편수, 세부적인 논문의 질 등의 편차가 심해, 무역연관 학회지에서 가장 오래되고 유일한 우수등재지인 무역학회지와 유일한 SSCI 저널인 Journal of Korea에 게재된 한국무역학회에서 발간하는 두 가지 저널을 대상으로 빈도수뿐만 아니라 키워드들 간의 네트워킹과 중심성 등 좀 더 심층적으로 분석하였다.

그리하여 2002년 1월부터 2022년 6월 동안

한국무역학회지에서 발간하였던 2,004편의 논문을 대상으로 텍스트 마이닝 기법을 이용하여 무역학 논문의 키워드의 빈도(frequency)와 중심성별로 다소 다르게 나타났다. 이를 위해 본 연구에서는 게재된 논문 중에서 저자 키워드의 출현 빈도수가 10개 이상인 72개의 키워드를 구했지만, 지면관계상 랭킹 50위 까지 빈번하게 나타난 키워드를 추출하였다.

2002년 1월부터 2022년 6월 동안 한국무역학회에서 발간된 무역과 관련 2개의 학술지에서는 <Table 4>에 나와 있는 대로, 무역과 직접적인 관련이 있는 export 73회와 trade 69회가 제일 많이 나왔고, 그 다음에는 effect 41회, performance 39회, model 37회, investment 35회, innovation 35회, tariff 33회, firm 33회 등의 순으로 많이 나타났다.

이와 같이 연구논문의 키워드는 export와 trade가 월등히 많이 나왔지만, competitiveness 14회, advantage 14회, 그리고 import는 11회만 나와 상대적으로 무역학에서 중요하게 생각하는 키워드의 빈도수는 적게 나왔다. 그러나 investment와 tariff 등도 30회 이상으로 나와 높은 순위를 이루고 있는 것으로 나타나 자유무역협정 등의 팽창 등으로 인해 한국의 연구자들은 이에 대한 관심을 더 많이 가지고 연구를 한 것으로 나타났다.

그리고 총 무역연관 19개 학회지가 발행하는 논문 14,301 편에서 무역연구 키워드상의 빈도수가 가장 높게 나타났던 export와 trade 등이 한국 무역학회에서 발간하는 무역학회지와 Journal of Korea Trade의 2,004편에서도 무역연구 키워드가 동일하게 높은 순위로 나타났다. 따라서 한국무역학회에서 발행하는 두 가지 학회지가 우리나라 전체의 무역연관 학회지의 대표성을 반영하고 있는 것으로 확인된다. 그리하여 다음 장에서는 연구 키워드 간의 한국무역학회에서 발간하는 두 가지 학회지를 대상으로 무역연구 키워드의 네트워크의 중심성을 바탕으로 심층분석을 한다.

4. 전체 네트워크 중심성 분석

네트워크 분석은 Netminer 4를 통해 분석하

였으며, 키워드 출현 빈도는 각 키워드에 대한 1-MODE 네트워크 매트릭스(Network Matrix)를 구축 후, 중심성 분석을 중심으로 네트워크 분석을 진행하였다. 그리고 각 키워드의 연결 중심성에 따라 노드크기와 텍스트크기를 설정하였으며, 이는 노드의 크기가 클수록 해당 노드가 얼마나 많은 관계에 관여하는지 파악할 수 있다. 네트워크 분석은 먼저 텍스트 파싱 및 텍스트 필터링 작업을 수행한 후 데이터를 기반으로 다음과 같이 연관분석을 수행하였다.

첫째, 2002년에서 2022년 6월까지 한국 무역학회에서 발간하는 2개의 국문 학술지와 영문 학술지 2,004편의 논문들을 대상으로 네트워크 중심성 분석을 하였다. 다음 <Table 5>를 보면, 연결 중심성의 경우 export(0.6), trade(0.5), performance(0.5), effect(0.4), investment(0.4) 등의 순으로 높은 순위를 차지하는 연구 키워드인 것으로 나타났으며, 그 외 industry, factor, value, market, logistics, chain, 그리고 company 등이 모두 0.3으로 높은 순위를 차지하고 있는 연구 키워드로 나타났다.

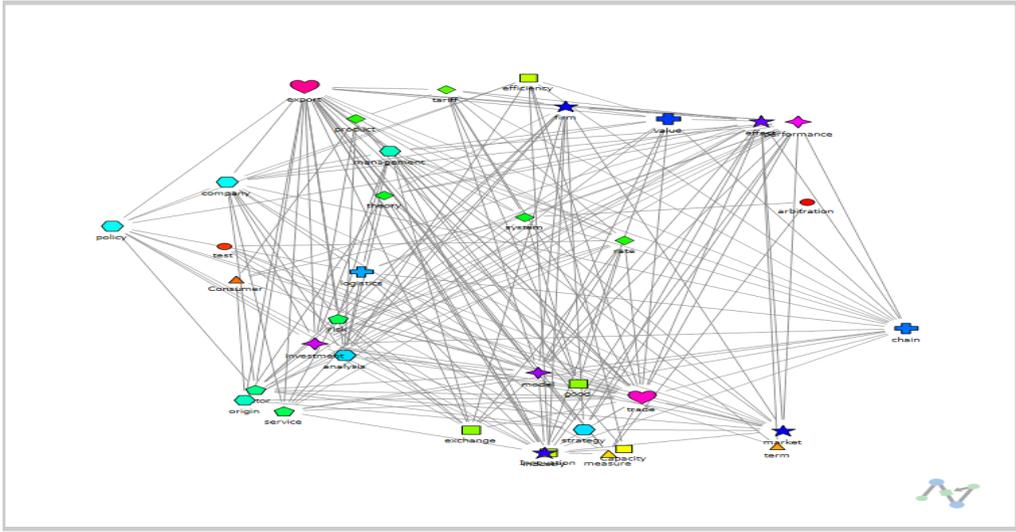
둘째, 근접 중심성의 경우, export(0.7)과 trade(0.7), performance(0.6), effect(0.6), investment(0.6), firm(0.6), industry(0.6), value(0.6), market(0.6), 그리고 chain(0.6) 등의 순으로 높은 순위를 차지하는 연구 키워드인 것으로 나타났으며, 그 외 innovation, tariff, factor, company, logistics, product 등이 모두 0.5로 높게 나타났다. 그리하여 근접 중심성의 경우는 대체로 연구 키워도 빈도나 연결 중심성, 그리고 근접 중심성과 순위는 비슷하지만 연결의 간접효과 때문에 좀 더 높게 나타났다. 특히 연결 중심성이 0.1-0.2로 비교적 낮게 나타난 경우에도 근접 중심성의 경우는 간접효과로 인해 보통 0.4-0.5로 비교적 높게 나타났다.

셋째, 매개 중심성의 경우는 순위는 연구 키워도 빈도나 연결 중심성, 그리고 근접 중심성이 가장 높게 나타난 export, trade, effect, 그리고 performance 경우는 모두 0.1로 나타나 높은 순위를 차지하고 있지만, 그 외 다른 키워드는 매개효과가 매우 작게 나타났고 거의 매개 중심성이 없는 것으로 나타났다.

Table 5. Several Centrality Analyses of the Korea Trade Research Network

Ranking	Key Word	Degree Centrality	Closeness Centrality	Betweenness Centrality	Eigenvector Centrality
1	export	0.6	0.7	0.1	0.2
2	trade	0.5	0.7	0.1	0.2
3	effect	0.4	0.6	0.1	0.1
4	performance	0.5	0.6	0.1	0.2
5	model	0.4	0.6	0.0	0.1
6	investment	0.4	0.6	0.0	0.2
7	Innovation	0.1	0.5	0.0	0.0
8	tariff	0.2	0.5	0.0	0.0
9	firm	0.3	0.6	0.0	0.1
10	rate	0.2	0.5	0.0	0.2
11	analysis	0.3	0.5	0.0	0.1
12	industry	0.3	0.6	0.0	0.1
13	factor	0.3	0.5	0.0	0.1
14	value	0.3	0.6	0.0	0.3
15	market	0.3	0.6	0.0	0.2
16	Consumer	0.1	0.4	0.0	0.0
17	policy	0.2	0.5	0.0	0.1
18	service	0.2	0.5	0.0	0.2
19	measure	0.1	0.5	0.0	0.0
20	efficiency	0.2	0.5	0.0	0.1
21	logistics	0.3	0.5	0.0	0.2
22	chain	0.3	0.6	0.0	0.3
23	product	0.2	0.5	0.0	0.0
24	Capacity	0.1	0.5	0.0	0.0
25	good	0.2	0.5	0.0	0.1
26	company	0.3	0.5	0.0	0.2
27	arbitration	0.1	0.4	0.0	0.0
28	exchange	0.2	0.5	0.0	0.2
29	risk	0.2	0.5	0.0	0.2
30	management	0.2	0.5	0.0	0.2
31	test	0.1	0.4	0.0	0.1
32	term	0.1	0.4	0.0	0.0
33	strategy	0.2	0.5	0.0	0.1
34	origin	0.2	0.5	0.0	0.1
35	theory	0.2	0.5	0.0	0.0
36	system	0.2	0.5	0.0	0.1
37	subsidiary	0.1	0.5	0.0	0.1
38	index	0.1	0.5	0.0	0.1
39	guarantee	0.1	0.4	0.0	0.0
40	country	0.2	0.5	0.0	0.1
41	competitiveness	0.2	0.5	0.0	0.1
42	advantage	0.2	0.5	0.0	0.1
43	Volume	0.0	0.4	0.0	0.0
44	rule	0.1	0.5	0.0	0.1
45	price	0.2	0.5	0.0	0.1
46	port	0.1	0.4	0.0	0.1
47	knowledge	0.1	0.5	0.0	0.0
48	insurance	0.1	0.5	0.0	0.1
49	eTrade	0.0	0.4	0.0	0.0
50	Economy	0.0	0.3	0.0	0.0

Fig. 5. International Trade Network



마지막으로 연결의 정도 외에도 액터의 중요성을 동시에 감안하는 아이겐벡터 중심성의 경우는 value와 chain 등이 0.3으로 나타나, 빈도나 연결 중심성, 그리고 근접 중심성이 가장 높게 나타난 export 0.2, trade 0.2, effect 0.1, 그리고 performance 0.2 등의 연구키워드보다 아이겐벡터 중심성이 더 높게 나타났다.

아이겐벡터 중심성은 중요성을 동시에 고려하기 때문에 value와 chain은 네트워크 망에서 노드의 빈도나 연결라인 수가 다소 상대적으로 작았지만, 생산과 무역측면에서 가격경쟁력을 나타내는 비용보다 최근 물류서비스 그리고 밸류체인 귀속여부 등이 중요하게 부각됨으로 인해 계수가 더 높게 나타났다.

5. 시각화와 네트워크

본 연구에서 무역연구 키워드의 네트워크를 나타내면 다음과 같이 나타난다. 무역연구 키워드의 빈도수 기준에 따라 복잡하고 줄이 많이 연결된 네트워크가 형성되었지만, 위의 빈도수로 선택한 무역 키워드의 중심성과 연결을 나타내는 무역연구 간 네트워크는 <Fig. 5>에서 나타나 있다. 노드의 모양은 구분하기 것이

며, 빈도수가 많이 나타나 무역연구의 키워드와 중심성이 높은 노드는 크게 나타나 있고 링크를 나타내는 줄이 많이 이어져 있다.

6. 에고 네트워크 중심성과 밀도 분석

에고 네트워크(Ego Network)는 Poulin, et al.(2000), Carter and Feld(2004) 연구를 참조하면, 네트워크의 최적화 개념으로서 네트워크의 응집성을 나타내는 것으로 본 장에서는 그것을 이웃 노드 간에 연결이 얼마나 많이 이루어져 있는가를 밀도(Density)로 측정하여 나타낸다. <Table 6>에서 나타난 키워드의 순위는 빈도수에 의한 것이며, 이러한 빈도수에 따라 에고 네트워크의 연결(degree)과 응집도를 나타내는 밀도를 나타낸다. 2002년에서 2022년 6월까지 무역학회지에서 발간하는 무역연구의 키워드의 에고 네트워크의 응집성을 나타내는 밀도의 순위는 앞의 연결 중심성이나 근접 및 매개 중심성과는 다르게 순위가 나타났다.

에고 네트워크에서 다른 노드와의 연결개수는 export 45개와 trade 39, performance 32,

Table 6. Neighbor Degree and Ego Density

Rank	Key Word	Number of Degree	Density	Rank	Key Word	Number of Degree	Density
1	export	45.0	0.2	26	origin	14.0	0.4
2	trade	39.0	0.3	27	study	13.0	0.5
3	performance	32.0	0.3	28	product	13.0	0.3
4	investment	29.0	0.4	29	price	13.0	0.4
5	model	28.0	0.3	30	orientation	13.0	0.5
6	effect	27.0	0.3	31	capital	13.0	0.4
7	industry	24.0	0.3	32	advantage	13.0	0.3
8	firm	23.0	0.3	33	sale	12.0	0.6
9	value	21.0	0.4	34	good	12.0	0.2
10	market	21.0	0.5	35	exchange	12.0	0.6
11	logistics	20.0	0.4	36	enterprise	12.0	0.5
12	chain	20.0	0.5	37	country	12.0	0.4
13	factor	19.0	0.4	38	theory	11.0	0.6
14	company	19.0	0.5	39	tariff	11.0	0.5
15	analysis	19.0	0.4	40	quality	11.0	0.5
16	policy	17.0	0.5	41	efficiency	11.0	0.3
17	management	17.0	0.5	42	agreement	11.0	0.5
18	competitiveness	17.0	0.4	43	subsidiary	10.0	0.5
19	strategy	16.0	0.5	44	rule	10.0	0.5
20	service	16.0	0.5	45	capability	10.0	0.4
21	risk	16.0	0.5	46	network	9.0	0.5
22	rate	16.0	0.4	47	knowledge	9.0	0.3
23	business	16.0	0.6	48	index	9.0	0.6
24	system	15.0	0.3	49	port	8.0	0.5
25	cost	15.0	0.3	50	integration	8.0	0.6

그리고 investment 29 등의 순으로 키워드가 이웃노드랑 많이 연결되어 있으며, 이하 연결 개수는 빈도수의 순위와 같게 나타났다. 그러나 에고 네트워크에서 응집성을 의미하는 밀도를 보면, business, sale, exchange, index, 그리고 integration 등이 모두 0.6로 나타나 제일 높게 나타났다.

그 다음은 market, chain, company, policy, management, stratrgy, service, tariff, quality, agreement, subsidiary, network, 그리고 port 등이 모두 0.5로 높게 나타나, 이러한 키워드들

의 밀도 및 응집성이 높게 나타났다. 그러나 이러한 것들의 밀도는 키워드의 빈도분석과 전체 중심성 분석에서 가장 순위가 높게 나왔던 export 0.2와 trade 0.3, performance 0.3, 그리고 investment 0.4 등보다 높게 나왔다.

7. 무역연구의 토픽 분석

본 장에서 무역학 연구동향 분석을 위해서 토픽 모델링(Topic Modeling) 방법을 활용한다. 토픽(Topic)은 머신러닝 기반의 텍스트 마

Table 7. Ten Topics by LDA in MCMC (Burn in 10)

	1st Keyword	2nd Keyword	3rd Keyword	4th Keyword	5th Keyword
Topic-1	export	rate	exchange	test	logistics
Topic-2	Capacity	tariff	good	policy	industry
Topic-3	trade	measure	analysis	effect	company
Topic-4	value	chain	service	port	export
Topic-5	tariff	industry	investment	analysis	firm
Topic-6	Innovation	performance	service	investment	effect
Topic-7	trade	analysis	investment	Innovation	industry
Topic-8	effect	performance	efficiency	model	company
Topic-9	trade	market	policy	factor	investment
Topic-10	model	product	firm	Consumer	measure

이닝 기법의 하나로 단어나 문서의 집합에서 더 큰 주제로 묶어내는 것을 찾아내는 분석기법이다. 그리하여 토픽 모델링(Topic Modeling)은 비구조화된 문서집합에서 의미있는 토픽 혹은 주제들을 추출해 주는 확률적 모델 알고리즘이다.

그리하여 본 연구에서는 무역학 연구논문의 텍스트 본문을 키워드 중심으로 숨겨진 의미구조를 발견하는 텍스트 마이닝 기법에 의해 수집된 무역학 논문에 어떤 주제들로 묶을 수 있는지 찾아낸다. 그 후 토픽을 구성하고 있는 키워드 분포를 바탕으로 무역학 연구논문들에서 발견된 키워드들의 분포를 분석하여 추출해낸 키워드들이 어떤 토픽을 다루고 있을지를 예측한다.

토픽분석에 가장 많이 사용하는 LDA는 단어의 순서는 고려하지 않고 문서 단어 행렬 즉 TF-IDF(Term Frequency-Inverse Document Frequency) 행렬을 입력으로 받는다. 그리하여 먼저 논문들의 키워드 문서를 TF-IDF 행렬로 변환한다. 그리고 TF-IDF 행렬을 입력변수로 토픽 모델링을 진행하여 LDA 모형을 구현한 함수를 구한다. 그 후 토픽 모델링에서는 모형을 평가하고 적절한 토픽의 수를 찾기 위한 과

정이 필요하다.

LDA는 무역연구 논문별 토픽별 키워드 분포의 결합분포에 의해 연구논문에 사용된 키워드를 추출하고, 그 연구논문을 나타내는 토픽들이 어떤 단어들로 구성되어 있는지 그 확률분포를 구해야 한다. 그러나 LDA 알고리즘에서는 토픽의 개수는 결정해주지 않으므로, 연구자가 선정하여야 한다. 본 연구에서는 토픽의 수를 10개로 선정하여 MCMC(Monte-Carlo Markov Chain) 방법을 활용하여 시물레이션을 100회 반복하고 10회 Burn-in하여 각 키워드가 할당받은 확률을 추정하였다. 먼저 아래와 같이 키워드 빈도수를 고려하여 키워드 빈도수 상위 50개를 기준으로 채택하여 토픽에서 키워드가 발생할 확률을 구하였다. 그리하여 먼저 <Table 7>과 <Table 8>에서 나타나 있듯이, 10개의 토픽들과 확률이 높은 키워드를 토픽별로 다섯 개씩 구하였다.

그 다섯 개 키워드가 <Table 7>에 나타나 있으며, <Table 8>에서 10개의 개별적인 토픽별로 앞에서 구한 빈도분석에서 나타난 상위 순위 키워드 50개 중에서 출현 확률이 가장 높은 키워드 다섯 개를 추출한 것이 <표 7>에 나타나 있다.

Table 8. Probability of Key Word in 10 Topics

Ranking	Word	Topic-1	Topic-2	Topic-3	Topic-4	Topic-5	Topic-6	Topic-7	Topic-8	Topic-9	Topic-10
1	export	0.274	0.007	0.003	0.051	0.003	0.046	0.035	0.033	0.033	0.022
2	trade	0.019	0.010	0.109	0.048	0.017	0.005	0.113	0.006	0.152	0.007
3	effect	0.008	0.028	0.039	0.002	0.003	0.046	0.002	0.156	0.002	0.006
4	performance	0.010	0.003	0.007	0.050	0.003	0.117	0.029	0.047	0.004	0.006
5	model	0.003	0.002	0.006	0.006	0.003	0.024	0.004	0.039	0.023	0.154
6	Innovation	0.005	0.001	0.018	0.014	0.029	0.118	0.048	0.011	0.004	0.003
7	investment	0.004	0.017	0.005	0.007	0.059	0.048	0.064	0.001	0.039	0.005
8	firm	0.018	0.020	0.021	0.001	0.044	0.007	0.036	0.020	0.018	0.052
9	tariff	0.003	0.051	0.003	0.006	0.117	0.006	0.003	0.028	0.006	0.018
10	analysis	0.003	0.004	0.058	0.006	0.056	0.003	0.080	0.005	0.002	0.002
11	rate	0.110	0.033	0.002	0.010	0.004	0.004	0.008	0.037	0.003	0.002
12	industry	0.002	0.037	0.013	0.008	0.095	0.001	0.036	0.004	0.006	0.003
13	factor	0.005	0.017	0.001	0.002	0.041	0.002	0.020	0.002	0.083	0.015
14	consumer	0.006	0.034	0.018	0.001	0.010	0.013	0.006	0.024	0.021	0.039
15	market	0.003	0.003	0.006	0.001	0.008	0.011	0.007	0.004	0.125	0.002
16	value	0.006	0.001	0.004	0.148	0.003	0.002	0.005	0.002	0.002	0.003
17	policy	0.002	0.039	0.005	0.005	0.004	0.002	0.013	0.008	0.086	0.004
18	efficiency	0.005	0.010	0.024	0.015	0.015	0.007	0.004	0.045	0.014	0.022
19	measure	0.003	0.009	0.076	0.003	0.005	0.010	0.003	0.006	0.009	0.038
20	service	0.009	0.003	0.002	0.073	0.008	0.051	0.002	0.006	0.004	0.002
21	chain	0.004	0.003	0.003	0.106	0.003	0.007	0.007	0.010	0.003	0.003
22	logistics	0.047	0.019	0.009	0.005	0.026	0.003	0.007	0.016	0.004	0.011
23	Capacity	0.002	0.072	0.013	0.005	0.004	0.003	0.002	0.031	0.003	0.006
24	product	0.004	0.021	0.005	0.002	0.011	0.002	0.009	0.012	0.006	0.071
25	arbitration	0.002	0.027	0.029	0.004	0.006	0.002	0.004	0.021	0.023	0.013
26	company	0.011	0.004	0.038	0.004	0.004	0.016	0.007	0.037	0.005	0.008
27	good	0.009	0.050	0.007	0.001	0.029	0.014	0.004	0.011	0.005	0.004
28	exchange	0.069	0.003	0.004	0.003	0.002	0.002	0.002	0.023	0.001	0.016
29	management	0.005	0.024	0.019	0.013	0.020	0.009	0.012	0.005	0.007	0.006
30	risk	0.006	0.015	0.005	0.006	0.035	0.004	0.014	0.013	0.003	0.018
31	origin	0.004	0.022	0.038	0.005	0.001	0.003	0.006	0.018	0.002	0.012
32	strategy	0.002	0.008	0.011	0.015	0.002	0.031	0.006	0.003	0.032	0.001
33	term	0.005	0.008	0.005	0.007	0.010	0.032	0.008	0.023	0.002	0.012
34	test	0.050	0.002	0.006	0.003	0.001	0.004	0.036	0.002	0.005	0.002
35	Volume	0.011	0.007	0.024	0.003	0.020	0.002	0.005	0.013	0.005	0.015
36	advantage	0.019	0.013	0.027	0.005	0.004	0.003	0.007	0.004	0.016	0.007
37	competitiveness	0.003	0.001	0.011	0.019	0.011	0.008	0.012	0.011	0.002	0.027
38	country	0.010	0.002	0.030	0.006	0.014	0.004	0.005	0.008	0.004	0.022
39	guarantee	0.009	0.008	0.002	0.020	0.031	0.008	0.005	0.003	0.014	0.005
40	index	0.002	0.003	0.021	0.003	0.002	0.002	0.028	0.006	0.005	0.035
41	subsidiary	0.023	0.007	0.006	0.003	0.005	0.005	0.016	0.017	0.015	0.009
42	system	0.003	0.020	0.014	0.004	0.006	0.043	0.002	0.004	0.006	0.002
43	theory	0.002	0.015	0.008	0.005	0.011	0.006	0.005	0.033	0.001	0.018
44	Economy	0.003	0.037	0.007	0.003	0.021	0.009	0.003	0.008	0.003	0.004
45	eTrade	0.002	0.009	0.008	0.005	0.021	0.003	0.028	0.003	0.006	0.014
46	insurance	0.007	0.014	0.006	0.004	0.002	0.016	0.007	0.021	0.003	0.017
47	knowledge	0.009	0.006	0.003	0.011	0.008	0.008	0.026	0.006	0.011	0.011
48	port	0.003	0.001	0.004	0.058	0.002	0.004	0.003	0.001	0.002	0.023
49	price	0.002	0.022	0.012	0.012	0.007	0.002	0.028	0.005	0.008	0.002
50	rule	0.002	0.019	0.030	0.007	0.008	0.009	0.005	0.008	0.003	0.008

그리고 <Table 8>의 제일 원편의 순위는 키워드 분석에서 나타난 빈도의 순위를 말하고 있다. 그리하여 <Table 8>은 특정한 무역연구 토픽에 특정 무역연구 키워드가 나타날 확률을 보여주고 있고, <표 8>에서 토픽1에서 키워드 export가 나타날 확률이 0.27을 가지고 있고, 그 다음 나타날 확률이 높은 키워드는 rate 0.11, exchange 0.07, test 0.05, logistics 0.05 등의 순으로 높게 나타나고 있다. 그리하여 토픽 1의 레이블(Label)은 '수출과 환율 및 물류'로 정할 수 있다. 토픽 2에서는 키워드 capacity 0.07, tariff 0.05, good 0.05, policy 0.04, 그리고 industry 0.04 등의 높은 확률을 가지고 있으며, 이는 '관세와 산업효과' 등으로 규정한다. 토픽 3에서는 키워드 trade가 나타날 확률이 0.11이 가장 높게 나타났고, measure 0.08로 그 다음 높게 나타나고 있으며, '무역과 기업에 대한 효과'로 토픽의 레이블을 정한다.

토픽 4에서는 키워드 value가 나타날 확률이 0.15, chain 0.11, service 0.11, port 0.06, 그리고 export 0.05 등으로 높게 나타나, 토픽 4의 레이블은 최근 중요한 이슈인 '밸류 체인(value chain)과 항만서비스와 수출'으로 특징지어 진다. 토픽 5에서는 키워드 tariff가 나타날 확률이 0.12, industry 0.09, investment 0.06, analysis 0.06, 그리고 firm 0.04 등으로 높은 확률 순으로 나타나, '관세와 산업투자'로 토픽의 레이블로 정한다.

토픽 6은 키워드 innovation가 나타날 확률이 0.12, performance 0.12, service 0.05, investment 0.05 등의 순으로 높은 확률을 가지고 있는 것으로 나타나, 토픽 6은 '혁신능력과 투자'로 토픽의 레이블로 정한다. 토픽 7은 키워드 trade가 나타날 확률이 0.11, analysis 0.08, investment 0.06, Innovation 0.05, 그리고 industry 0.04 등으로 높은 확률 순으로 나타나, 무역투자와 산업혁신 등으로 토픽의 레이블로 정한다.

토픽 8은 키워드 effect가 나타날 확률이 0.16, performance 0.05, efficiency 0.04, model 0.04 그리고 company 0.04 등으로 높은 확률 순으로 나타나, '기업의 능력과 효율성' 등으로 토픽의 레이블로 정한다. 토픽 9는 키워

드 trade가 나타날 확률이 0.15, market 0.13, policy 0.09, factor 0.08, 그리고 investment 0.04 등의 순으로 높은 확률이 나와서 '무역과 시장과 투자'으로 토픽 레이블로 정한다. 토픽 10은 키워드 model이 나타날 확률이 0.15, product 0.07, firm 0.05, consumer 0.04, 그리고 measure 0.04 등으로 나타나, 토픽 레이블을 '기업제품과 소비자'로 레이블로 정한다.

그리하여 토픽분석 결과, 무역연구에서 trade, export, 그리고 tariff가 중요한 키워드로 나타났고, 10개 토픽들에 있어 중요한 영향을 미치는 것으로 나타났다.

V. 결론

본 연구는 우리나라 무역학 연구의 토픽과 동향을 분석하기 위하여 2002년부터 2022년 6월까지 19개 무역학 연관 학회지의 14,301편의 논문들과 한국무역학회지에서 발행하는 두 가지 학술지에 게재된 2,004편을 대상으로 텍스트 마이닝과 소셜 네트워크 기법을 중심으로 분석하였다. 그리하여 본 연구의 결과를 요약하면 다음과 같다.

첫째, 우리나라 전체 19개 무역연관 학술지 14,301편의 논문을 대상으로 텍스트 마이닝 기법을 이용하여 논문들의 키워드의 빈도수를 추출해본 결과, 2002년부터 매 3-4년 마다 국제적인 무역환경의 변화에 따라 2006년부터 2021년까지 연구 키워드가 조금씩 변경되었지만, 가장 빈번한 연구 키워드는 trade, export, market, industry 등의 순으로 나타났고, 2018년에는 ecommerce 등에 관한 연구도 높아지고 있다. 2002년부터 2022년 6월까지 게재된 논문 14,301편 모두를 조사해본 결과, trade 279회, export 171회로 각각 나타나 무역연관 학술지에 압도적으로 많은 무역학 연구 키워드로 나타났다. 그 다음 logistics, 69회, market 59회 등으로 빈도수가 높게 나왔으며, 자유무역협정, 디지털 전환과 산업시스템 등의 영향으로 agreement, digital, industry, FTA 등도 37회 이상의 높은 빈도를 보이고 있다.

둘째, 2002년 1월부터 2022년 6월 동안 무역

학관련 학회들을 대표하는 가장 오래되고 회원 수도 제일 많은 한국무역학회에서 발간하는 유일한 국문 우수등재지인 한국무역학회지와 유일한 SSCI 영문학술지인 Journal of Korea Trade를 대상으로 추정된 결과, 무역과 직접적인 관련이 있는 export 73회와 trade 69회가 제일 많이 나왔고, 그 다음 무역과 투자, 혁신과 관세에 대한 초점이 높아감에 따라 investment 35회, innovation 35회, tariff 33회, 그리고 firm 33회 등의 순으로 많이 나타났다.

셋째, 한국 무역학회에서 발간하는 2개의 국문 학술지와 영문 학술지 2,004편의 논문들을 대상으로 네트워크 중심성 분석을 한 결과 먼저, 연결 중심성보다 간접적인 효과를 포함하는 근접 중심성이 조금 높게 나왔지만 대체로 순위가 유사하게 나타났다. 그리하여 export, trade, performance, 그리고 investment 등의 순으로 순위가 높게 나타났고, 그 외 industry, factor, value, market, logistics, chain, 그리고 company 등도 비교적 높게 나타났지만, 매개 중심성은 아주 낮게 나타났다.

넷째, 아이겐벡터 중심성의 경우는 연결 중심성과 근접 중심성과는 달리, value와 chain 등이 빈도나 연결 중심성, 그리고 근접 중심성이 가장 높게 나타난 export, trade, effect, 그리고 performance 등보다 더 높게 나타났다. 그리하여 value와 chain은 네트워크에서 노드의 빈도나 연결라인 수가 다소 상대적으로 작았지만, 생산과 무역측면에서의 비용보다 최근 물류 그리고 밸류체인 귀속여부 등이 중요하게 부상된 것으로 나타났다.

다섯째, 네트워크에서 예고 연결성 및 예고 밀도를 분석한 결과, 앞의 전체 네트워크 연결성 순위와 다르게 나타났다. 예고 네트워크의 키워드의 빈도수는 전체 네트워크의 빈도 순위가 대체로 비슷했지만, 그 밀도를 보면, business, sale, exchange, index, 그리고 integration 등이 모두 제일 높게 나타났는데, 이러한 키워드들은 연결 중심성에서 가장 순위가 높았던 export, trade, 그리고 investment 등보다 더 높게 나타났다.

마지막으로 본 연구는 머신러닝의 기법인 LDA 모형과 MCMC 기법을 사용하여 10개 토픽

픽모델링 분석을 한 결과, 수출과 무역, 관세, 물류 그리고 혁신과 산업, 밸류체인 등이 거의 모든 토픽에서 높은 확률을 가지고 있는 것으로 나타났다.

따라서 본 연구에서 분석한 결과, 무역학 분야에서 전통적으로 중요한 연구 키워드인 무역과 수출, 관세 외에도 투자와 산업, 그리고 기업과 산업의 효율성과 성과 등도 중요한 키워드로 나타나는 등 변화하고 있다. 특히 최근에는 밸류체인과 디지털 전환 등이 중요한 이슈로 부각되어감에 따라 value와 chain 등의 키워드가 아이겐벡터 중심성에서 높은 순위로 나타났고 그 중요성이 더해가고 있다. 그리하여 무역학 연구대상은 전통적인 수출과 무역과 관세 외에도 이러한 산업경쟁력과 기업 능력, 밸류체인과 디지털 전환에 따른 좀 더 심층적으로 많이 연구할 필요가 있다.

이와 같이 본 연구는 전통적 계량모형의 기법과는 달리 최근 사회학, 신문방송언론, 정치외교, 마케팅 등 다른 사회과학에서 많이 사용되고 있는 데이터 마이닝과 소셜 네트워크 분석을 무역학 연구에 도입하였다. 무역학 분야에서 데이터 마이닝과 소셜 네트워크를 융합한 심층적인 양적 분석이 미흡하다. 그리하여 무역학 분야에도 이러한 데이터 마이닝과 소셜 네트워크 분석 기법과 LDA 기법 등에 의한 토픽 등을 연구하여 최근 20년간 무역학 분야의 연구동향과 연구초점이 무엇이고 그 내용이 어떤가를 융합적으로 분석하는 것도 의미가 있을 것이다. 본 연구결과에 따라 향후 무역학 분야도 연구의 대상과 초점을 좀 더 넓혀 나갈 필요가 있다.

따라서 본 연구는 우리나라 무역학 연구 분야에서 아직 연구가 미흡한 데이터 마이닝과 소셜 네트워크, 그리고 토픽분석에 의해서 무역연구에 대한 주요 키워드와 네트워크 및 중심적인 토픽을 분석함으로써 향후 우리나라의 무역연구에서 기존 무역학 논문뿐만 아니라 향후 무역학 연구에 대한 내용 및 키워드 그리고 무역연구의 네트워크와 연결 관계 및 동향을 분석함으로써 많은 후속연구들을 유발할 수 있을 것이다.

그러나 본 연구는 우리나라 KCI 중분류의 무

역학 연관 학술지 등에서 국제경제학회, 금융학회, 국제금융학회 등이 무역학 분야에 누락되어 있었다. 여러 무역학 연관 학회지도 분석하였지만 국제상학 분야의 논문편수가 훨씬 많았으므로, 나중에는 무역학회에서 발행하는 학

회지를 중심으로 분석하였다. 그리하여 향후 누락되었던 무역관련 학술지를 포함시키고 좀 더 많은 자료를 수집하여 연도별로 분석한다면 좀 더 엄밀한 분석과 해석을 할 수 있을 것이다.

References

- Blei, D. M., A. Y. Ng and M. I. Jordan (2003), "Latent Dirichlet Allocation", *Journal of machine Learning research*, 3, 993-1022.
- Bonacich, P. (2007), "Some unique properties of eigenvector centrality," *Social Networks*, 29(4), 555-564.
- Carter, W. C. and S. L. Feld (2004), "Principles relating social regard to size and density of personal networks, with applications to stigma," *Social Networks*, 26(4), 323-329.
- Chung, In-Kyo, Wei Min, and Jeong-Ho Yoo (2019), "China's Free Trade Agreement and Trade Network Change", *The Journal of Humanities and Social science*, 10(2), 319-330.
- Freeman, L.C. (1978), "Centrality in social networks conceptual clarification", *Social Networks* 1, 215-239.
- Han, Jin-Sung and Ji-Hwan Yoon (2016), "Activation Strategies of the 20th BIFF using Social Big Data Text Mining Analysis", *Journal of Tourism Sciences*, 40(1), 133-145.
- Jung, sae-Yeon and Chun-Su Lee (2018), "Analysis of the Centrality of the Korean Footwear Industry through Export Network Analysis", *International Trade and Insurance*, 19(1), 151-173.
- Kang, Dong Joon and Keun-Sik Park (2017), A Study on the Development of Korea FTA strategy with the world RTA network analysis, *International Commerce and Information Review*, 19(3), 3-23.
- Kim, Nam-Gyu, Dong-Hoon Lee, Ho-Chang Choi, and William Xiu Shun Wong (2017), "Investigations on Techniques and Applications of Text Analytics", *The Journal of Korean Institute of Communications and Information Sciences*, 42(2), 471-492.
- Knoke, D. and S. Yang (2008), *Social Network Analysis*, SAGE, Thousand Oaks.
- Ko, Hyun-Jeong (2021), "Analysis on the Research Topics of Supply Chain Risk Management using Text Mining", *E-Trade Review*, 19(3), 65-83.
- Ko, Min-Hwan and Yun-Mi Park (2022), "A study on future tourism in the post-COVID-19 era using text mining based on big data: Focused on LAN cable tours", *International journal of Tourism and Hospitality*. 36(5), 79-92.
- Kwahk, Kee-Young (2017), *Social Network Analysis*, the 2nd edition, Cheongram.
- Lee Jee-Hoon and Jung-Suk Kim (2000), "A Study on the Research Trends in Int'l Trade Using Topic modeling", *Korea Trade Review*, 45(3), 55-69.
- Oh, Jin-Ho (2020), "A Study on Research Trends in Seaport Logistics in South Korea: Application of the SLNA Methods", *Korea International Commerce Review*, 35(4), 109-134.
- PaK, Jee-Moon, Sung-Kuk Kim, and Han-Ho Kim (2015), "Study on Oyster Trade Networks by Utilizing Social Network Analysis", *Korea Trade Review*, 40(2), 51-70.

- Paranyushkin Dmitry (2011), Identifying the Pathways for Meaning Circulation using Text Network Analysis, *Nodus Labs*, 26, 1-26.
- Park, Hyun-Hee and Sung-Je Cho (2020), “Meta Analysis of Trade Insurance Using Text Mining”, *Korea Trade Review*, 45(6), 157-179.
- Park, Jun-Suk, Chang-Sik Kim, and Kee-Young Kwahk (2016), “Investigation of Research Trend in Hotel Domain Using Text Mining and Social Network Analysis”, *Journal of Tourism and Leisure Research*, 28(9), 209-226.
- Poulin, R., M. C. Boily, and B. R. Mâsse (2000), “Dynamical systems to define centrality in social networks”, *Social networks*, 22(3), 187-220.
- Rha, Jin-Sung (2022), “Analysis of Factors Affecting Surge in Container Shipping Rates in the Era of Covid19 Using Text Analysis”, *Journal of the Korea Industrial Information System Research*, 27(1), 111-123.
- Sohn, Sung-Pyo (2020), “A Research Progress and Meta Analysis of International Trade Area on the Customs Study”, *The Journal of Korea Research Society for Customs*, 21(1), 3-17.
- Son, Yong-Jung (2018), “An Analysis on Economic Effects in Korea’s FTAs using SNA: A Focus on Networks between Contracting States”, *The e-Business Studies*, 19(5), 285-296.
- Suh, Jeong-Meen (2020), Measuring Convergences and Divergences in APEC RTAs/FTAs: a text-mining approach, *APEC Study Series*, 20-02, KIEP.
- Suh, Jeong-Meen, Ye-Eun Han, and Joung-Youn Kim (2019), “Comparative Analysis of Korea’s RTA using Text Mining”, *Korean journal of international economic law*, 17(3), 25-48.
- Wasserman, S., and K. Faust (1994), *Social network analysis: Methods and applications*, Cambridge University Press.
- Yoon, Hee-Young and Il-Youp Kwak (2020), “The Association Modeling on Keywords and Documents of Korea International Trade Research using Paper Abstract data”, *Korea International Commerce Review*, 35(2), 45-64.