

# 텍스트마이닝 기법을 활용한 미국산업응용수학 학회지의 연구 현황 및 동향 분석

## Analysis on Status and Trends of SIAM Journal Papers using Text Mining

김성연

인천대학교 교육대학원

Sung-Yeun Kim(syk@inu.ac.kr)

### 요약

본 연구의 목적은 전 세계 산업수학의 흐름을 주도하는 미국산업응용수학 학회에서 출판하는 논문들의 연구 현황 및 동향을 거시적으로 파악하는 데 있다. 이를 위해 2016년부터 2019년까지 6,255편의 논문 제목 및 초록을 수집하였으며, LDA 기법을 활용한 토픽모델링과 시계열회귀모형 분석을 수행하였다. 분석 결과 첫째, 산업수학 분야는 해석학을 중심으로 기하학, 대수학, 위상수학, 이산수학, 확률 및 통계 등 다양한 분야에서 연구가 진행되었다. 둘째, 시간이 흐름에 따라 상승하는 연구 주제는 수리유체역학, 그래피론, 확률미분방정식이었으며, 하강하는 연구 주제는 계산이론과 고전기하로 나타났다. 연구 결과는 산업수학 분야의 지적 구조에 대한 전체적인 흐름 및 변화에 대한 이해를 바탕으로 연구자들에게는 향후 연구 방향에 대해서, 그리고 교육 현장에는 시대 변화를 반영한 산업수학 교육과정을 수립하는데 시사점을 제공할 것이다.

■ 중심어 : | 미국산업응용수학학회 | 산업수학 | 잠재 디리클레 할당 모형 | 토픽 모델링 |

### Abstract

The purpose of this study is to understand the current status and trends of the research studies published by the Society for Industrial and Applied Mathematics which is a leader in the field of industrial mathematics around the world. To perform this purpose, titles and abstracts were collected from 6,255 research articles between 2016 and 2019, and the R program was used to analyze the topic modeling model with LDA techniques and a regression model. As the results of analyses, first, a variety of studies have been studied in the fields of industrial mathematics, such as algebra, discrete mathematics, geometry, topological mathematics, probability and statistics. Second, it was found that the ascending research subjects were fluid mechanics, graph theory, and stochastic differential equations, and the descending research subjects were computational theory and classical geometry. The results of the study, based on the understanding of the overall flows and changes of the intellectual structure in the fields of industrial mathematics, are expected to provide researchers in the field with implications of the future direction of research and how to build an industrial mathematics curriculum that reflects the zeitgeist in the field of education.

■ keyword : | Industrial Mathematics | LDA | SIAM | Topic Modeling |

\* 본 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No.2019R1F1A1059437).

접수일자 : 2020년 05월 08일

심사완료일 : 2020년 06월 08일

수정일자 : 2020년 06월 08일

교신저자 : 김성연, e-mail : syk@inu.ac.kr

## I. 서론

4차 산업혁명에는 18세기 후반에 영국에서 시작된 증기기관의 사용과 기계화에 따른 1차 산업혁명, 19세기 후반에 증기기관에서 전기와 석유를 활용한 2차 산업혁명, 20세기 후반에 컴퓨터와 인터넷, 인공지능을 이용한 정보혁명을 나타내는 3차 산업혁명에 이어 인공지능, 사물인터넷, 빅데이터, 모바일 등 첨단정보통신기술이 사회, 경제 전반에 융합되어 혁신적인 변화가 나타나는 차세대 산업혁명을 일컫는다[1][2].

국내에서는 이러한 4차 산업혁명 ‘문제해결사’로 산업수학을 꼽고 있으며, 이세돌 9단과 알파고의 바둑대결로 인공지능에 대한 관심이 고조된 2016년부터 정부 주도하에 4차 산업혁명의 핵심 기초학문을 산업수학으로 하고 이를 육성하는 방안을 발표하였다[3]. 같은 해 8월에는 과학기술전략회의에서 9개 국가전략 프로젝트로 성장동력 확보를 위한 자율주행자동차, 경량소재, 스마트시티, 인공지능, 가상증강현실의 5개 과제와 삶의 질 향상을 위한 정밀의료, 탄소자원화, (초)미세먼지, 바이오의약의 4개 주제를 선정하였다[4]. 이러한 국가전략을 달성하기 위해서는 과학, 공학 및 기술의 발전이 이루어져야 가능하겠지만 산업수학의 발전으로 인하여 달성 가능성을 높이고 효과 및 효율의 향상을 거둘 수 있다[5]. 또한 2017년에는 산업문제 해결과 인력양성의 거점이 될 산업수학센터를 선정하였으며, 최근에는 기초원천연구 중장기 지원방안을 통해 수학과 통계 분야 연구를 장려할 것이라고 발표하는 등 산업수학에 대한 관심과 중요성이 더욱 부각되고 있는 실정이다[6]. 이처럼 산업수학은 학문 중심의 순수수학이 주류를 이루었던 우리나라의 수학계 풍토를 정부 주도하에 순수수학 영역을 넘어 다양한 학문과 융·복합을 이룰 수 있도록 정부, 학계, 교육계가 모두 바뀌도록 하고 있다.

융·복합 분야에서 새로운 연구 주제를 탐색하고, 기존의 연구 흐름을 파악하는 데 있어 연구 현황 및 동향에 대한 포괄적인 이해는 필수적이다. 연구 현황은 거시적인 관점에서 해당 학문 분야의 연구 주제를 요약해서 보여주며, 연구 동향은 시간이 흐름에 따라 연구 주제가 어떻게 변화했는지의 양상을 보여준다. 이처럼 기존 연구의 현황 및 동향을 파악하는 것은 한 학문의 새로

운 패러다임의 등장에 대응하고 향후 연구의 방향성을 예측하는 데 있어 기반의 역할을 한다. 특히 국외 연구 동향에 대한 분석 결과는 글로벌 이슈를 고려한 연구 주제 도출에 기여할 수 있다[7].

그동안 산업수학과 관련한 국내 연구들로는 이미 산업이 고도화되어 산업수학 역사가 20년 이상 된 선진국 사례에 대한 문헌분석이나 내용분석을 중심으로 한 정책연구들과[8][9] 특정 산업에 초점을 맞춘 엔터테인먼트 분야에서 산업수학을 활성화하는 연구가[10] 수행되었다. 이러한 기존 연구들은 연구자가 가진 뛰어난 전문성에 기반하여 향후 산업수학 활성화를 위한 정책 방향과 전략을 제시해 줄 수 있으며, 산업수학 분야 내 세부 분야의 연구 동향을 미시적 관점에서 이해하는 데 적합하다고 할 수 있다. 그러나 다른 연구자들이 같은 분석 방법으로 결과를 재현하는 것이 어려우며, 분석 전에 연구자가 분석할 가치가 있는 내용을 사전에 알아야 하며, 연구자가 직접 수작업으로 자료를 수집 및 분류하여야 하므로 분석대상이 되는 텍스트의 용량이 제한적일 수 밖에 없다는 한계가 존재한다[11][12].

따라서 본 연구는 4차 산업혁명이란 용어가 처음으로 언급된 2016년부터 최근 2019년까지 전 세계 산업수학의 흐름을 주도하는 학회인 미국산업응용수학회(Society for Industrial and Applied Mathematics, SIAM)에 게재된 논문 제목과 초록을 대상으로 토픽모델링 분석을 이용해 산업수학 분야의 연구 현황 및 동향을 분석하고자 한다. 토픽모델링은 비정형 대규모 텍스트자료에서 반복적으로 제시되는 주제들을 찾기 위해 사용되는 빅데이터 분석 방법 중 하나이다. 토픽모델링의 기법 중 잠재 디리클레 할당(Latent Dirichlet Allocation, LDA) 기법은 이산 자료들에 대한 확률적 생성 모형으로 단어들의 확률을 활용하여 문서 집합 내의 잠재된 토픽들을 찾아내는 방법으로서[13][14], 동일한 의미를 갖지만 철자가 다른 단어, 철자는 같지만 문맥에 따라 의미가 달라지는 단어를 효과적으로 구별할 수 있다. 또한 추출된 단어들은 토픽들 간의 독립성이 두드러지기 때문에 토픽을 결정하는데 용이하다는 장점이 있다[11].

특정 학문 분야의 현황 및 동향을 살펴보기 위해 LDA 기법의 토픽모델링을 적용한 연구들을 살펴보면

[표 1]과 같다. 먼저 국외 연구로 Shiryayev et al.[15]은 웹 페이지에서 과학 기술 분야의 연구와 관련하여 논문들을 수집하고 분석한 결과, LDA 기법이 전문가의 평가 결과와 일치함으로써 과학기술 분야의 연구 경향을 파악하기에 적합함을 제시하였다. Choi et al.[16]은 개인 정보 보호에 대한 학술 연구의 동향을 1972년부터 2015년까지 총 2,356편의 논문 초록을 대상으로 LDA 기법을 활용하였으며, Amado et al.[17]은 마케팅 분야에서의 연구 동향을 2010년부터 2015년까지 출판된 1,560개 논문에 LDA 기법을, 그리고 Sun & Yin[18]은 1990년부터 2015년까지 총 17,163편의 논문 초록에 LDA 기법을 적용하여 분석한 결과로 추출된 토픽들은 전문가 평가에 의해서도 대표적인 연구 주제들을 확인하였다.

국내에서 박자현과 송민[19]은 문헌정보학 분야의 논문 초록을 대상으로 LDA 기법을 적용하여 연구동향을 파악하였으며, 안정국 외[20]는 정보시스템 분야의 연구 동향을 파악하고자 1980년부터 2015년까지 48,102개의 논문 제목, 초록, 키워드 분석을 통해 연구 토픽 추출 결과를 연대별로 분석화하고 시각화하였다. 조경원 외[21]는 보건의료산업학회 학술지의 영문초록과 키워드에 대해 LDA 기법을 적용하여 10년간의 연구 주제와 연구 주제 추이를 파악하였으며, 토픽모델링이 개별 논문의 주제에 대한 탐색이라기보다는 보건의료산업의 학문 분야에 대해 전체적인 이해와 통찰을 가능케 했다고 밝혔다. 또한 장익영과 이원미[22]는 한국스포츠사회학회지에서 출판한 777편 논문의 제목을 대상으로 LDA 기법으로 분석하여 연구 주제 동향의 특성과 시간의 흐름에 따라 연구 주제 동향의 변화를 파악하였다.

이상에서 살펴본 것처럼 공학, 과학, 사회과학, 체육 등 다양한 학문 분야에서 짧게는 6년, 길게는 44년 동안 출판된 516편에서 48,102편에 이르는 논문들에 자동화된 텍스트 빅데이터 분석방법을 활용하여 각 학문 분야의 연구 내용을 총괄적으로 이해하고 분석하는 연구들이 수행되었다. 특히 이 중에서도 방대한 양의 텍스트가 갖고 있는 주제를 도출하는데 효과적인 LDA 기법이 사용되고 있다.

표 1. 특정 학문분야의 현황 및 동향을 파악한 연구들

| 연구                     | 수집 기간     | 논문 편수  | 주요 결과  |
|------------------------|-----------|--------|--|
| Shiryayev et al.(2017) | -         | 1,070  | IT 기술 분야에서 LDA 기법 분석 결과 추출된 토픽은 전문가 평가와 일치함  |
| Choi et al. (2017)     | 1972-2015 | 2,356  | 개인 정보 보호 분야에서 알고리즘, 페이스북 개인정보, 온라인 소셜네트워크 추출함  |
| Amado et al. (2017)    | 2010-2015 | 1,560  | 경영학 분야에서 빅데이터, 마케팅, 저자들이 소속된 기관의 지리적 위치, 제품, 분야를 추출함                                       |
| Sun & Yin (2017)       | 1990-2015 | 17,163 | 운송 분야에서 상품 주제로 지속가능성, 여행 행동, 자전거의 이동성이 나타남   |
| 박자현, 송민 (2013)         | 1970-2012 | 3,834  | 국내 문헌정보 분야에서 상송토픽은 도서관 유형별 서비스 및 평가, 인터넷, 메타데이터, 하강토픽은 도서, 분류, 편목, 고서지로 나타남                |
| 안정국 외. (2016)          | 1980-2015 | 48,102 | 정보시스템 분야에서 시간이 흐름에 따라 의사결정 지원시스템, 인터넷 상용화, 데이터마이닝, 클라우드 컴퓨팅 추출함                            |
| 조경원 외. (2017)          | 2000-2017 | 516    | 보건의료산업학회지에서 의료서비스, 보건관리, 병원조직 및 인적자원관리, 병원재무관리, 노인보건 추출함                                   |
| 장익영, 이원미 (2019)        | 1993-2018 | 777    | 한국스포츠사회학회지에서 스포츠 참가요인, 주제, 미디어 및 성역할, 거시이론 및 방안, 하위문화, 조직문화, 사회문제, 사회구조와 행위, 미시이론 관련개념 추출함 |

주. "-"는 Habrahabr.ur에서 정보 보안과 관련된 모든 논문을 수집함으로써 수집 기간이 따로 명시되어 있지 않음

본 연구에서는 이러한 시대적 변화에 따라 LDA 기법을 활용하여 이미 산업수학이 발전해있는 대표적인 SIAM에서 출판한 17종 학술지의 6,255개 논문의 제목과 초록을 분석함으로써 산업수학 분야의 연구 현황 및 동향을 거시적으로 파악하고자 한다. 본 연구는 산업수학 분야의 지적 구조에 대한 전체적인 흐름 및 변화에 대한 이해를 바탕으로 연구자들에게는 향후 연구 방향에 대해, 그리고 교육 현장에는 시대 변화를 반영한 산업수학 교육과정을 수립하는데 기초 자료로 활용될 수 있을 것이다. 구체적인 연구문제는 다음과 같다.

첫째, SIAM 학술지의 연구 주제는 어떠한 특성을 갖고 있는가?

둘째, 시간이 흐름에 따라 SIAM 학술지의 연구 주제는 어떻게 변화되는가?

## II. 연구방법

### 1. 자료수집

본 연구는 전 세계의 산업수학 흐름을 주도하는 학회인 SIAM에서 출판하는 17종 학술지의 2016년부터 2019년까지 게재된 총 6,255개의 논문 제목 및 초록을 분석대상으로 정하여 SIAM 홈페이지에서 R로 웹 스크래핑 프로그램을 작성하여 수집하였다. 또한 연구동향을 분석하기 위해 출간 연도도 함께 수집하였다. 수집된 논문들의 연도별 빈도수는 [표 2]에 제시된 바와 같이 2018년도의 논문 편수가 약간 많은 것을 제외하면 거의 비슷한 것으로 나타났다.

표 2. 분석 변수

| 학술지명\연도                                | 2016 | 2017 | 2018 | 2019 |
|--|------|------|------|------|
| Applied Algebra and Geometry           |      | 33   | 25   | 28   |
| Applied Dynamical Systems              | 77   | 74   | 87   | 69   |
| Applied Mathematics                    | 106  | 99   | 146  | 104  |
| Computing                              | 70   | 62   | 81   | 55   |
| Control and Optimization               | 135  | 160  | 176  | 162  |
| Discrete Mathematics                   | 123  | 132  | 150  | 116  |
| Financial Mathematics                  | 33   | 31   | 42   | 30   |
| Mathematical Analysis                  | 147  | 157  | 202  | 159  |
| Mathematics of Data Science            |      |      |      | 33   |
| Matrix Analysis and Applications       | 78   | 70   | 77   | 65   |
| Multiscale Modeling & Simulation       | 53   | 64   | 69   | 45   |
| Numerical Analysis                     | 147  | 130  | 147  | 114  |
| Optimization                           | 114  | 105  | 131  | 115  |
| Scientific Computing                   | 262  | 232  | 264  | 257  |
| SIAM Review                            | 44   | 45   | 46   | 44   |
| Theory of Probability&Its Applications | 52   | 49   | 66   | 54   |
| Uncertainty Quantification             | 53   | 48   | 63   | 48   |
| 연도별 합계                                 | 1494 | 1491 | 1772 | 1498 |
| 총 합계                                   | 6255 |      |      |      |

### 2. 자료 처리

수집한 자료에 전처리 과정을 수행하기 위해 먼저 제목과 초록의 두 열로 나누어져 있는 자료를 한 단위로 처리할 수 있도록 하나의 열로 병합하였다. 이후 텍스트를 분석에 필요한 단위로 나누는 토큰화(tokenization) 과정을 수행하였으며, 대문자를 소문자로 통일하였다. 또한 n개의 단어가 연쇄적인 확률로 나타나는 단어의 동시 출현 빈도수를 측정할 수 있는

n-gram에서 n의 최소값은 1로, 최대값은 2로 적용하였다. 다음으로 분석에 필요하지 않은 구두점들과 숫자, 수식과 함께 R 프로그램의 tm 패키지에서 제공하는 'english'와 'SMART'를 통해 대명사, 관사, 전치사 등 불용어(stopwords)를 삭제하였다. 또한 추출된 단어들 중에서 논문 제목이나 초록에 빈번히 출현하면서 연구 목적과 무관한 'paper', 'purpose', 'result', 'study' 등을 불용어에 포함시켜 삭제하였다. 마지막으로 문장 내에서 다양한 형태로 나타나는 단어들의 표제어를 추출하는 원형복원(lemmatization)을 수행하였다. 예를 들어 'imagined', 'imagining', 'imagines'는 'imagine'으로 추출되었다.

### 3. 자료 분석

전처리 과정이 모두 완료된 자료를 대상으로 LDA 기법의 토픽모델링 분석을 실시하였다. [그림 1]은 LDA 알고리즘을 그림으로 제시한 것이다[14].

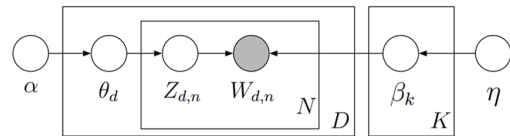


그림 1. LDA 그래픽 모형

[그림 1]에서  $W_{d,n}$ 은 문서  $d$ 의  $n$ 번째 단어를 나타내며 문서에서 관측되는 변수이고,  $K$ 는 토픽의 수를,  $\alpha$ 는  $\theta_d$ 값을 결정하는 디리클레(Dirichlet) 분포 파라미터를,  $\eta$ 는  $\beta$ 값을 결정하는 파라미터를 나타낸다.  $\theta_d$ 는 문서별 토픽의 비율을,  $\beta_k$ 는 토픽 내 단어들의 분포를,  $Z_{d,n}$ 은 문서  $d$ 에서  $n$ 번째 단어를 토픽에 할당하는 것을 나타낸다[23]. 따라서 문서별 토픽 비율인  $\theta_d$ 에 따라 문서 내에 존재하는 단어들의 토픽인  $Z_{d,n}$ 이 결정되며,  $Z_{d,n}$ 과 토픽 내 단어들의 분포인  $\beta_k$ 에 따라 단어  $W_{d,n}$ 이 결정된다[24].

LDA 분석을 하기 위해서는 먼저 연구자가 토픽의 수를 지정해야 한다. 본 연구에서는 고전적인 방법으로 자주 사용되고 있는 perplexity는 내재적으로 학습 성

능 정도를 의미하고 있을 뿐 그 결과를 해석하기에 난해하다는 단점[25]을 해결한 coherence score를 이용하여 토픽의 수를 결정하였다. Coherence score는 Newman et al.[26]이 제안하였으며, 이 값이 클수록 LDA 분석 결과 산출되는 각각의 토픽이 의미론적으로 유사한 단어들로 구성되었다고 해석한다[27]. 본 연구에서는 최적의 토픽 수를 coherence score가 최대가 되는 22개로 정하고 연구 주제로 해석하였다. 또한 22개의 주제가 잘 구분되는지를 토픽 간 거리 지도(intertopic distance map, IDM)로 확인하였다. IDM은 주제의 중요도와 주제 간 거리를 나타내는 그림으로 왼쪽은 추출된 주제를 원의 형태로 보여주며, 원의 면적은 분석 데이터 내에서 토픽의 상대적 중요도를 나타낸다[28]. 또한 각 주제의 위치는 다차원 척도법(multidimensional scaling)에 따라 유사성 거리에 의해 표현된 것으로 이를 통해 주제 간의 유사성 또는 차이를 확인할 수 있다. 즉, 각 원이 겹치는 영역 없이 잘 구분될수록 연구 주제들도 명확하게 구분된다고 해석한다. 왼쪽 그래프에서 원하는 주제를 선택하면 주제의 색이 붉은색으로 변하고, 오른쪽에는 해당 주제를 설명하는 핵심 단어가 나타난다[29].

연구 주제명은 주제별 상위 단어들의 연관성을 바탕으로 산업수학 관련 경험이 있는 기하학, 대수학, 통계학, 해석학을 전공한 4인의 박사들로 구성된 전문가 협의회를 개최하여 수학 세부 분야로 정하였다. 또한 연구 주제명과 관련하여 고교수학 교육과정에 필요한 수학, 그리고 관련된 산업명을 부여하도록 하였다.

연구 주제의 연도 변화에 따른 동향을 파악하기 위해서는 독립변수로는 연도를, 종속변수로는 개별 주제(토픽)의 연도별 비중 평균값을 사용하여 시계열 선형회귀 분석을 수행하였다. 분석 결과 중 유의수준 5%에서 통계적으로 유의한 회귀계수 값이 양수(+)이면 상승(hot) 주제, 그리고 음수이면 하강(cold) 주제로 구분하였다. 모든 분석은 R 프로그램으로 수행하였다. 구체적으로 coherence score 산출에는 “textmineR” 패키지[30], LDA 분석에는 “topicmodels” 패키지[31], 그리고 시계열 회귀모형 분석에는 “stats” 패키지를 활용하였다.

### III. 연구결과

#### 1. 토픽모델링 분석

LDA 기법의 토픽모델링 분석 결과, 2016년부터 2019년까지 SIAM에 게재된 총 6,255개의 논문 제목 및 초록에서 중심이 되는 22개의 연구 주제 및 관련 핵심 단어는 [표 3]과 같다. [표 3]에는 지면 제약 상 각 연구 주제에 대해 빈도수 기준으로 상위 10개의 단어만을 제시하였다. 그러나 2-gram 결과 나타나는 주제가 각각 단일 주제어 상위 10위안에 드는 경우는 중복을 피하고 독립적인 정보를 제공하기 위하여 다음 순위에 드는 주제어들을 제시하였다.

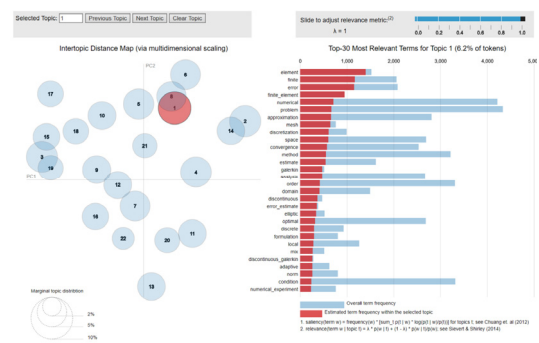


그림 2. IDM을 활용한 LDA 시각화

또한 [그림 2]에는 22개의 토픽이 잘 구분되는지를 IDM으로 제시하였으며, 왼쪽에서 선택된 원은 연구 주제 1로 유한요소론을 나타내며, 오른쪽의 30개 단어들은 이 주제를 설명하는 핵심 단어를 나타낸다. [그림 2]에서 알 수 있듯이, 각 주제들의 비중은 전체적으로 유사하게 나타나 편차가 크지 않으며, 각 주제들 간의 겹치는 원의 영역이 매우 작아 각 주제들 간의 연관성이 낮은 것으로 나타났다. 이를 통해 22개의 연구 주제들이 비교적 명확하게 구분되어 있음을 확인할 수 있다.

연구 주제 1은 유한요소론과 관련되어 있으며 ‘element’, ‘finite’, ‘error’ 등으로 구성되어 있다. 이는 고등학교 수학교육과정 중 구분구적법, 확률 및 통계 등과 관련되어 있으며, 건축 및 신소재 산업 등과 관련되어 있다고 해석할 수 있다.

연구 주제 2는 푸리에해석과 관련되어 있으며 'wave', 'periodic', 'schrodinger' 등으로 구성되어 있다. 이는 고등학교 수학교육과정 중 미적분과, 벡터, 삼각함수 등과 연관되어 있으며, 통신, 제조, 및 시스템 산업 등과 관련되어 있다고 해석할 수 있다.

연구 주제 3은 수리유체역학과 관련되어 있으며 'boundary', 'fluid', 'navierstokes' 등으로 구성되어 있다. 이는 고등학교 수학교육과정 중 미적분 및 공간도형 등과 연관되어 있으며, 양자역학, 제조 및 항공 산업 등과 관련되어 있다고 해석할 수 있다.

연구 주제 4는 확률과정론과 관련되어 있으며 'random', 'process', 'markov' 등으로 구성되어 있다. 이는 고등학교 수학교육과정 중 확률 및 통계 등과 연관되어 있으며, 데이터 분석 산업 등과 관련되어 있다고 해석할 수 있다.

연구 주제 5는 역문제응용과 관련되어 있으며 'datum', 'image', 'inverse' 등으로 구성되어 있다. 이는 고등학교 수학교육과정 중 미적분 및 행렬 등과 연관되어 있으며, 정밀진단, 딥러닝, 의료 영상산업 등과 관련되어 있다고 해석할 수 있다.

연구 주제 6은 수치해석학과 관련되어 있으며 'numerical', 'order', 'rungekutta' 등으로 구성되어 있다. 이는 고등학교 수학교육과정 중 그래프, 미적분, 알고리즘 및 행렬 등과 연관되어 있으며, 제조 및 반도체 산업 등과 관련되어 있다고 해석할 수 있다.

연구 주제 7은 선형대수와 관련되어 있으며 'matrix', 'rank', 'algebraic' 등으로 구성되어 있다. 이는 고등학교 수학교육과정 중 행렬, 벡터 및 공간좌표 등과 연관되어 있으며, 시스템 개발 및 인공지능 산업 등과 관련되어 있다고 해석할 수 있다.

연구 주제 8은 최적화이론과 관련되어 있으며 'optimization', 'gradient', 'minimization' 등으로 구성되어 있다. 이는 고등학교 수학교육과정 중 부등식의 영역 및 미분 등과 연관되어 있으며, 정밀진단, 성능평가, 유통 및 물류 산업 등과 관련되어 있다고 해석할 수 있다.

연구 주제 9는 위상수학과 관련되어 있으며 'network', 'dynamic', 'node' 등으로 구성되어 있다. 이는 고등학교 수학교육과정 중 그래프, 집합 및 확률

등과 연관되어 있으며, 정보시스템공학 및 의료 산업 등과 관련되어 있다고 해석할 수 있다.

연구 주제 10은 전산통계와 관련되어 있으며 'approximation', 'monte', 'carlo' 등으로 구성되어 있다. 이는 고등학교 수학교육과정 중 행렬, 미적분, 알고리즘, 확률 및 통계 등과 연관되어 있으며, 금융 및 증권 산업 등과 관련되어 있다고 해석할 수 있다.

연구 주제 11은 그래프이론과 관련되어 있으며 'graph', 'vertex', 'edge' 등으로 구성되어 있다. 이는 고등학교 수학교육과정 중 그래프 등과 연관되어 있으며, 인공지능, 통신 네트워크, 유통 및 물류 산업 등과 관련되어 있다고 해석할 수 있다.

연구 주제 12는 행렬분해와 관련되어 있으며 'matrix', 'factorization', 'krylov' 등으로 구성되어 있다. 이는 고등학교 수학교육과정 중 알고리즘 및 행렬 등과 연관되어 있으며, 시스템 개발 산업 등과 관련되어 있다고 해석할 수 있다.

연구 주제 13은 계산이론과 관련되어 있으며 'computation', 'reduction', 'method' 등으로 구성되어 있다. 이는 고등학교 수학교육과정 중 미분 및 알고리즘 등과 연관되어 있으며, 전산 및 소프트웨어 산업 등과 관련되어 있다고 해석할 수 있다.

연구 주제 14는 확률미분방정식과 관련되어 있으며 'system', 'stochastic', 'differential' 등으로 구성되어 있다. 이는 고등학교 수학교육과정 중 미분, 확률 및 통계 등과 연관되어 있으며, 금융공학, 보건, 정책 및 빅데이터 산업 등과 관련되어 있다고 해석할 수 있다.

연구 주제 15는 수리생물학과 관련되어 있으며 'flow', 'cell', 'membrane' 등으로 구성되어 있다. 이는 고등학교 수학교육과정 중 미적분, 행렬 및 벡터 등과 연관되어 있으며, 열유체공학 및 의료 산업 등과 관련되어 있다고 해석할 수 있다.

연구 주제 16은 함수해석과 관련되어 있으며 'dynamic', 'bifurcation', 'system' 등으로 구성되어 있다. 이는 고등학교 수학교육과정 중 미적분, 공간도형 및 공간좌표 등과 연관되어 있으며, 동역학계와 보건 산업 등과 관련되어 있다고 해석할 수 있다.

표 3. 토픽모델링 분석 결과

| 연구 주제       | 상위 10개 단어   |
|-------------|---|
| 1-유한요소론     | element, finite, error, numerical, problem, approximation, mesh, discretization, space, convergence               |
| 2-푸리에해석     | wave, stability, nonlinear, pattern, periodic, propagation, perturbation, schrodinger, localize, travel           |
| 3-수리유체역학    | boundary, condition, energy, model, flow, initial, fluid, dirichlet, navierstokes, equilibrium                    |
| 4-확률과정론     | random, distribution, process, probability, measure, variable, density, markov, walk, independent                 |
| 5-역문제응용     | datum, image, inverse, regularization, model, measurement, problem, error, weight, signal                         |
| 6-수치해석      | scheme, numerical, convergence, stability, diffusion, order, step, differential, accuracy, rungekutta             |
| 7-선형대수      | matrix, tensor, polynomial, rank, algebraic, set, symmetric, vector, norm, lattice                                |
| 8-최적화이론     | optimization, problem, convex, convergence, constraint, gradient, condition, algorithm, minimization, iteration   |
| 9-위상수학      | model, network, system, dynamic, process, node, state, sensitivity, chemical, neural                              |
| 10-전산통계     | stochastic, approximation, sample, parameter, uncertainty, monte, carlo, numerical, approach, pdes                |
| 11-그래프이론    | graph, vertex, edge, path, bound, degree, connect, tree, cycle, subgraph  |
| 12-행렬분해     | matrix, system, parallel, linear, preconditioner, factorization, sparse, multigrid, krylov, algorithm             |
| 13-계산이론     | approach, computation, numerical, application, simulation, reduction, technique, efficiency, method, optimization |
| 14-확률미분방정식  | control, system, stochastic, state, stability, equilibrium, differential, nonlinear, delay, dynamic               |
| 15-수리생물학    | model, flow, simulation, transport, cell, fluid, numerical, diffusion, traffic, membrane                          |
| 16-함수해석     | model, dynamic, bifurcation, system, parameter, equilibrium, manifold, oscillator, orbit, synchronization         |
| 17-수치선형대수   | algorithm, matrix, numerical, approximation, compute, method, experiment, iteration, subsapce, error              |
| 18-고전기하     | tree, point, distance, planar, embed, graph, permutation, line, sequence, euclidean                               |
| 19-금융수학     | optimal, control, strategy, cost, risk, time, price, inequality, market, portfolio                                |
| 20-미분기하     | function, operator, eigenvalue, space, spectral, matrix, linear, manifold, vector, interpolation                  |
| 21-응용편미분방정식 | field, scatter, boundary, inverse, potential, asymptotic, numerical, magnetic, homogenization, maxwell            |
| 22- 알고리즘    | algorithm, time, complexity, bound, approximation, polynomial, query, circuit, run, input                         |

주. Burnin 5,000과 iterations 5,000으로 Gibbs Sampling으로 추정했음

연구 주제 17은 수치선형대수와 관련되어 있으며 ‘algorithm’, ‘matrix’, ‘numerical’ 등으로 구성되어 있다. 이는 고등학교 수학교육과정 중 근사값, 알고리즘 및 행렬등과 연관되어 있으며, 데이터 분석 산업 등과

관련되어 있다고 해석할 수 있다.

연구 주제 18은 고전기하와 관련되어 있으며 ‘distance’, ‘line’, ‘euclidean’ 등으로 구성되어 있다. 이는 고등학교 수학교육과정 중 그래프, 조합, 공간도형 및 공간좌표 등과 연관되어 있으며, 물류 및 유통 산업 등과 관련되어 있다고 해석할 수 있다.

연구 주제 19는 금융수학과 관련되어 있으며 ‘cost’, ‘risk’, ‘portfolio’ 등으로 구성되어 있다. 이는 고등학교 수학교육과정 중 방정식 및 함수의 미분, 확률 및 통계와 연관되어 있으며, 금융, 보험 및 증권 산업 등과 관련되어 있다고 해석할 수 있다.

연구 주제 20은 미분기하와 관련되어 있으며 ‘operator’, ‘space’, ‘manifold’ 등으로 구성되어 있다. 이는 고등학교 수학교육과정 중 복소수, 공간도형 및 공간좌표 등과 연관되어 있으며, 동력기계공학 및 디자인 산업 등과 관련되어 있다고 해석할 수 있다.

연구 주제 21은 응용편미분방정식과 관련되어 있으며 ‘scatter’, ‘homogenization’, ‘maxwell’ 등으로 구성되어 있다. 이는 고등학교 수학교육과정 중 미적분, 행렬 및 벡터 등과 연관되어 있으며, 재료, 전자 및 제조 산업 등과 관련되어 있다고 해석할 수 있다.

연구 주제 22는 알고리즘과 관련되어 있으며 ‘algorithm’, ‘approximation’, ‘input’ 등으로 구성되어 있다. 이는 고등학교 수학교육과정 중 근사값, 알고리즘 및 미분 등과 연관되어 있으며, 컴퓨터, 전자, 소프트웨어 개발 산업 등과 관련되어 있다고 해석할 수 있다.

## 2. 시계열회귀모형 분석 결과

2016년부터 2019년까지 연도별 연구 주제 동향을 파악하기 위해 독립변수는 연도를, 종속변수는 개별 주제의 연도별 비중 평균값을 사용하여 시계열회귀모형으로 분석한 결과는 [표 4]와 같다. 또한 [그림 3]에는 이 중 상승하는 연구 주제와 하강하는 연구 주제의 비중을 연도별로 제시하였다.

유의수준 5%에서 통계적으로 유의하게 상승하는 연구 주제는 수리유체역학, 그래프이론, 확률미분방정식으로 나타났다. 반면에 유의수준 5%에서 통계적으로 유의하게 하강하는 연구 주제는 계산이론과 고전기하

로 나타났다.

표 4. 시계열회귀모형 분석 결과

| 연구 주제       | 회귀계수(SE)        | 유의확률  | 상승/하강 |
|-------------|-----------------|-------|-------|
| 1-유한요소론     | -0.0008(0.0016) | 0.683 |       |
| 2-푸리에해석     | 0.0006(0.0027)  | 0.852 |       |
| 3-수리유체역학    | 0.0027(0.0008)  | 0.043 | 상승    |
| 4-확률과정론     | 0.0029(0.0013)  | 0.159 |       |
| 5-역문제응용     | 0.0003(0.0010)  | 0.784 |       |
| 6-수치해석      | 0.0005(0.0025)  | 0.859 |       |
| 7-선형대수      | -0.0016(0.0017) | 0.446 |       |
| 8-최적화이론     | -0.0003(0.0018) | 0.866 |       |
| 9-위상수학      | 0.0020(0.0009)  | 0.160 |       |
| 10-전산통계     | 0.0004(0.0019)  | 0.867 |       |
| 11-그래프이론    | 0.0031(0.0009)  | 0.081 | 상승    |
| 12-행렬분해     | -0.0025(0.0014) | 0.206 |       |
| 13-계산이론     | -0.0012(0.0003) | 0.047 | 하강    |
| 14-확률미분방정식  | 0.0037(0.0009)  | 0.048 | 상승    |
| 15-수리생물학    | -0.0018(0.0011) | 0.231 |       |
| 16-함수해석     | -0.0010(0.0019) | 0.662 |       |
| 17-수치선형대수   | -0.0003(0.0003) | 0.409 |       |
| 18-고전기하     | -0.0015(0.0004) | 0.042 | 하강    |
| 19-금융수학     | -0.0014(0.0002) | 0.583 |       |
| 20-미분기하     | -0.0021(0.0008) | 0.111 |       |
| 21-응용편미분방정식 | 0.0001(0.0003)  | 0.763 |       |
| 22- 알고리즘    | -0.0017(0.0016) | 0.381 |       |

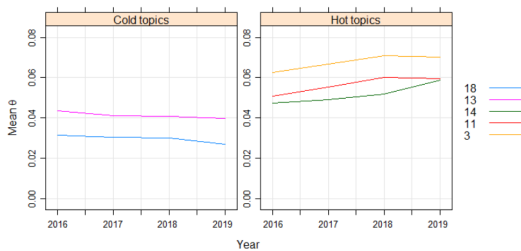


그림 3. 상승 주제 및 하강 주제

### V. 결론 및 제언

본 연구는 산업수학의 흐름을 주도하는 SIAM 학회에서 출판한 논문들의 제목과 초록을 대상으로 LDA 기법의 토픽모델링 분석을 수행함으로써 산업수학 분야 연

구의 현황과 동향을 거시적으로 파악하였다. 주요 연구 결과를 바탕으로 본 논문의 의의와 시사점을 제시하면 다음과 같다.

첫째, coherence score의 산출을 통해 토픽의 최적 개수는 22개로 결정되었으며, 산업수학 연구는 해석학을 중심으로 기하학, 대수학, 위상수학, 이산수학, 확률 및 통계 등 다양한 분야에서 진행되는 것으로 나타났다. 이는 기존의 산업수학 관련 연구에서 제시하고 있는 것처럼 21세기 지식정보화사회에서 수학이 담당해야 할 역할이 재정립되면서 세상의 문제를 해결하고 산업의 부가가치를 창출하는 데 다양한 수학 분야가 활용되고 있다는 연구 결과와 일치한다[5][8-10]. 구체적으로 기하학은 구제역과 조류 독감 등의 전염 경로를 파악하고 차단하는 데, 대수학과 이산수학은 구글의 검색 엔진과 알파고의 알고리즘을 만드는데, 그리고 위상수학은 암 발생 확률 예측 소프트웨어를 만드는데 활용되었다.

이처럼 본 논문에서 제시한 LDA 기법의 토픽모델링 분석을 통해 컴퓨터가 도출한 분석 결과는 기존의 전문가가 직접 내용을 판단하여 분석한 결과와 유사하게 도출됨을 확인하였다. 즉, 산업수학 분야의 현황과 동향을 파악하는 데 있어 본 연구에서 사용한 분석 방법이 적합함을 제시하였다는 점에서 본 논문의 의의가 있다. 이를 바탕으로 본 논문에서 활용한 데이터 처리과정 및 분석 방법은 다른 학문 분야에도 적용 가능하며, 산업수학 관련 뉴스, 신문기사, 페이스북, 트위터 및 블로그 데이터에 적용하여 실제 산업현장의 수요를 반영한 산업수학 동향을 파악하는 데 활용될 수 있을 것이다.

둘째, 연도별 연구 주제 동향을 분석한 결과, 상승하는 연구 주제는 수리유체역학, 그래프이론, 확률미분방정식, 그리고 하강하는 연구 주제는 계산이론과 고전기하로 나타났다. 상승하는 연구 주제는 김명환[8]이 제시한 신산업 적용 분야 중 금융IT 및 금융정보 산업, 뇌과학, 문화사업, 의료사업 등과 관련한 수학 분야라고 할 수 있다. 구체적으로 수리유체역학은 영상학, 생물학, 의학 등을 융합하여 수리영상과학 기반의 문화사업과 생물수학 기반의 의료사업에, 그래프이론은 생물학 및 뇌과학을 융합하여 뇌에서 이루어지는 정보처리 방식을 이해하는 계산수리 뇌과학에, 그리고 확률미분방정



식은 경영학과 공학을 융합하여 금융수학의 금융IT 및 금융정보 산업에 활용될 수 있다. 예컨대 수리유체역학은 물, 불, 눈, 공기 같은 특정한 형태가 없이 움직이는 유체의 움직임을 모사하는데 활용됨으로써 애니메이션에 등장하는 눈, 비, 먼지 등의 자연현상을 현실 세계와 거의 유사하게 재현할 수 있으며, 혈액 속 암세포를 신속하게 찾음으로써 개인 맞춤의학을 실현할 수 있으며, 그리고 전염병 전달의 메커니즘을 파악할 수 있다. 따라서 최근 전 세계가 코로나19가 확산되면서 새로운 위기에 처해있지만, 오히려 산업수학을 기반으로 적절한 대응 및 후속 조치를 취한다면, 이러한 위기를 극복 및 대처하고 4차 산업발전을 가속화시킬 수 있을 것이다. 반면에 하강하는 연구 주제는 응용수학 분야의 방법론적인 발전에 따라 이전의 연구 방법에 대한 수요가 감소하기 때문이라고 해석할 수 있다.

이를 바탕으로 4차 산업혁명을 실행하는 동안, 그리고 그 후에도 수학이 국내 산업현장에서 혁신의 도구 역할을 계속 해나가기 위해서 국내 연구자들은 상승하는 산업수학 연구 주제에 대한 지속적인 연구와 하강하는 산업수학 연구 주제에 대한 면밀한 탐색과 검토가 필요하다. 정효정[7]에 따르면 다양한 학문적 배경을 가진 융합 학문으로 혁신 연구의 국외 연구 동향을 분석하고, 그 결과로 국내 혁신 연구는 정부 현안을 고려한 Selective Follower로서 국외 연구 동향에 따른 연구 주제의 무분별한 도입을 지양할 필요가 있다고 강조하였다. 마찬가지로 국내 산업수학 연구의 경우도 추격형에서 창조형으로 전환하기 위해서는 외국의 연구 동향을 무조건 반영하는 것은 지양할 필요가 있다. 특히 하강하는 연구 주제인 계산이론과 고전기하처럼 순수수학은 현재 우리가 당면한 문제 중 가장 큰 이슈가 되고 있는 고차원 데이터를 수학적으로 분석하거나, 인공지능이 분석한 내용을 수학적으로 재구성하는데 활용되고 있으며, 이를 능가하는 새로운 이론의 출현을 기대하고 있는 실정이다. 또한 4차 산업혁명 이후의 정보와 관련된 모든 문제들은 대부분 수학문제로 바뀌어져서 많은 계산과 그 해석에 의존할 수 밖에 없다. 따라서 20년 이상 산업수학이 발전되어 있는 외국 산업수학 연구 동향과의 시간 격차를 줄이고, 향후 산업수학 연구에서 선도적인 역할을 하기 위해서는 순수수학에 대

한 연구가 지속적으로 수행될 필요가 있다.

셋째, 도출된 연구 주제를 바탕으로 산업수학 인재양성 및 산업화 방안으로 개설된 실용수학 교육과정에 반영되지 않은 내용 영역은 그래프, 근사값, 알고리즘 및 행렬로 나타났다. 우리나라 고등학교 수학과 교육과정에서는 7차 시기의 이산수학 과목에서 독자적으로 그래프와 알고리즘을, 그리고 공통수학 수준에서 근사값은 2차와 3차 시기에 다루었다. 행렬은 9차 시기에서는 필수과목에서 제외되었으며, 10차 시기부터는 인문계와 자연계 구분 없이 완전히 삭제된 상태이다. 그러나 외국의 고교 수학교육과정을 살펴보면 미국, 중국, 영국, 싱가포르, 호주 등 대부분의 국가에서 행렬을 다루고 있는 실정이다. 윤강준[32]에 따르면 수학의 기능은 단순히 계산능력을 배양하는 것이나 이론의 적용이 아니라 주어진 상황이나 조건에서 정답 등 최적의 답을 구하는 능력을 기르는 것이라고 하였다. 그리고 이것이 현재 인공지능 시대에서 수학이 핵심 기술로 부각되는 이유라고 하였다. 왜냐하면 인공지능 기술의 개발은 존재하는 기술의 적용보다는 주어진 상황에서 최적의 알고리즘을 찾고, 이 알고리즘을 구현하는 방법을 제시하는 데 있기 때문이다. 이러한 알고리즘 작성이 가장 중요한 수학 영역으로 행렬과 그래프를 들 수 있다. 또한 실세계와 수학의 관련성을 보여주는 영역으로, 그래프, 알고리즘 및 행렬이 포함되며, 그래프는 행렬에 의해 대수적으로 표현되므로 행렬은 그래프와 함께 다루는 수학적 도구가 된다[5].

이를 바탕으로 그래프, 근사값, 알고리즘 및 행렬을 실용수학 교육과정에 차기부터 포함시키는 것이 필요하다. 실용수학은 산업수학 인재양성 및 산업화 방안으로 개설된 과목으로 학생들에게 실생활과 첨단산업에 활용되는 수학의 원리를 소개하고 있다. 따라서 그래프 단원에서는 인공지능, 통신 네트워크, 유통 및 물류 산업 등을, 행렬 단원에서는 정밀진단, 딥러닝, 의료 영상 산업, 시스템 개발 산업 등을 예시로, 그리고 알고리즘 단원에서는 다양한 산업에서의 예시뿐만 아니라 일상 생활에서 부딪히는 단순한 문제에서부터 컴퓨터 프로그래밍까지 포함시킬 필요가 있다. 구체적으로 어떤 상황을 그래프로 표현하고, 행렬로 계산하며, 알고리즘으로 구현하고, 근사값으로 정확도를 평가하는 등의 수학

적 경험을 통해 학생들이 산업수학의 가치를 발견하도록 원리와 실용성 중심의 교육과정을 개발할 필요가 있다.

마지막으로 본 연구의 한계점과 향후 연구방안을 제시하면 다음과 같다. 본 연구에서는 SIAM에 게재된 논문들로만 분석의 범위를 한정하였기 때문에 산업수학과 관련한 국내 학회와 다른 외국의 학회에서 출판한 논문들이 포함되지 않았다. 따라서 분석의 범위를 확장하여 국내의 산업수학 관련 학회와 외국의 다양한 학회에서 출판한 논문들을 함께 분석하고 비교함으로써 산업수학 연구의 국내 및 국외 현황을 면밀히 파악하는 연구를 수행할 필요가 있다. 또한 본 연구에서는 SIAM의 연구 동향을 파악하기 위해 2016년도 이후로 논문이 게재된 연도를 한정하여 분석하였다. 그러나 산업수학 분야와 같이 융합 분야에서 새로운 연구 주제를 탐색하고, 기존 연구 동향에 대한 포괄적인 이해와 예측을 위해서는 논문이 게재된 연도를 변화시키면서 일년 주기의 단위를 확장해서 분석하는 연구를 수행할 필요가 있다.

**참 고 문 헌**

[1] 윤호영, "기술-사회 패러다임 용어는 어떻게 소비되는가? 탐색적 접근: 4차 산업혁명에 관한 26만건 뉴스(제목) 분석," 정보사회와 미디어, 제19권, 제2호, pp.1-28, 2018.

[2] 조경원, 우영운, "텍스트 마이닝을 이용한 4차 산업 연구 동향 토픽 모델링," 한국정보통신학회논문지, 제23권, 제7호, pp.764-770, 2019.

[3] 미래창조과학부, "산업현장에서 필요한 수학, 정부가 나서서 키운다," 미래창조과학부 보도자료, 2016.5.2.

[4] 미래창조과학부, "대한민국 미래 책임질 9대 국가전략 프로젝트 선정," 미래창조과학부 보도자료, 2016.8.10.

[5] 김영욱, 고성은, 김영록, 박혜숙, 이상욱, 장정욱, 조도상, 미래 인재상에 적합한 수학학습 내용 연구, 한국과학창의재단, 2017.

[6] 김태환, "수학 없으면 AI도 없다 국내 산업수학 생태계 마련 시급," 조선비즈, 2019.9.16.

[7] 정효정, "텍스트 마이닝을 이용한 혁신 분야의 국외 연

구 분석," 기술혁신연구, 제24권, 제4호, pp.249-275, 2016.

[8] 김명환, 국가 정책으로서의 산업응용수학 활성화 전략 수립, 미래창조과학부, 2015.

[9] 박기범, 산업수학 활성화를 위한 국내 산업수학 생태계 분석, 정책연구, 2015.

[10] 김성주, 정문성, "엔터테인먼트산업수학의 활성화 방안," 한국엔터테인먼트산업학회논문지, 제11권, 제2호, pp.289-297, 2017.

[11] 신명선, 조경원, "텍스트마이닝을 활용한 한국어어치료 학회지의 토픽 모델링 및 트렌드 분석:(2002~2018)," 언어치료연구, 제28권, 제3호, pp.81-91, 2019.

[12] J. W. Mohr and P. Bogdanov, "Introduction—topic models: What they are and why they matter," Poetics, Vol.48, No.6, pp.545-569, 2013.

[13] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," Journal of Machine Learning Research. Vol.3, pp.993-1022, 2003.

[14] D. M. Blei, "Probabilistic topic models," Communications of the ACM, Vol.55, No.4, pp.77-84, 2012.

[15] A. P. Shiryayev, A. V. Dorofeev, A. R. Fedorov, L. G. Gagarina, and V. V. Zaycev, "LDA models for finding trends in technical knowledge domain," IEEE, pp.551-554, 2017.

[16] H. S. Choi, W. S. Lee, and S. Y. Sohn, "Analyzing Research Trends in Personal Information Privacy Using Topic Modeling," Computers & Security, Vol.67, pp.244-253, 2017.

[17] A. Amado, P. Cortez, P. Rita, and S. Moro, "Research trends on Big Data in Marketing: A text mining and topic modeling based literature analysis," European Research on Management and Business Economics, Vol.24, pp.1-7, 2017.

[18] L. Sun and Y. Yin, "Discovering themes and trends in transportation research using topic modeling," Transportation Research Part C: Emerging Technologies, Vol.77, pp.49-66, 2017.

[19] 박자현, 송민, "토픽모델링을 활용한 국내 문헌정보

학 연구동향 분석,” 정보관리학회지, 제30권, 제1호, pp.7-32, 2013.

[20] 안정국, 김소담, 김희용, “텍스트 마이닝 기법을 이용한 정보시스템 분야 연구 동향 분석,” Information Systems Review, 제18권, 제3호, pp.73-96, 2016.

[21] 조경원, 배성권, 우영운, “텍스트마이닝을 활용한 보건의료산업학회지의 토픽 모델링 및 토픽트렌드 분석,” 보건의료산업학회지, 제11권, 제4호, pp.213-224, 2017.

[22] 장익영, 이원미, “토픽모델링을 적용한 한국스포츠회학회지의 연구주제 분석,” 한국스포츠회학회지, 제32권, 제2호, pp.20-35, 2019.

[23] 김상겸, 장성용, “토픽모델링을 이용한 국내 산업경영공학 연구동향 분석,” 한국경영공학학회지, 제21권, 제3호, pp.71-95, 2016.

[24] 박준형, 오효정, “국내 기록관리학 연구동향 분석을 위한 토픽모델링 기법 비교- LDA와 HDP를 중심으로-,” 제48권, 제4호, pp.235-258, 2017.

[25] J. Chang, S. Gerrish, C. Wang, J. L. Boyd-Graber, and D. M. Blei, “Reading tea leaves: How humans interpret topic models,” In Advances in neural information processing systems, pp.288-296, 2009.

[26] D. Newman, J. H. Lau, K. Grieser, and T. Baldwin, “Automatic evaluation of topic coherence,” In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp.100-108, 2010.

[27] 윤효준, 박재현, 윤지운, “비정형 텍스트 자료에서 잠재정보 추출을 위한 토픽모델링 소개: 치매관련 신체활동 뉴스 기사의 이슈 분석,” 체육과학연구, 제30권, 제3호, pp.501-512, 2019.

[28] C. Sievert and K. E. Shirley, “LDAvis: A method for visualizing and interpreting topics,” In Proceedings of the workshop on interactive language learning, visualization, and interfaces, pp.63-70, 2014.

[29] 유예림, 2015개정 교육과정 정책에 대한 언론보도 분석, 서울대학교, 박사학위논문, 2017.

[30] T. W. Jones, “textmineR,” <https://www.rtextminer.com>, 2019.11.16.

[31] B. Grün and K. Hornik, “topicmodels: An R Package for Fitting Topic Models,” Journal of Statistical Software, Vol.40, No.13, pp.1-30, 2011.

[32] 윤강준, “수학, 산업수학에 길을 묻다”, 중도일보, 2019.10.24.

저 자 소 개

김 성 연(Sung-Yeun Kim)

정희원



- 1997년 2월 : 영남대학교 수학과 (이학석사)
- 2003년 6월 : 성균관대학교 수학과 (이학박사)
- 2011년 2월 : 연세대학교 교육학과 (교육학박사)
- 2016년 3월 ~ 현재 : 인천대학교

교육대학원 부교수

〈관심분야〉 : 교육측정 및 평가, 교원평가, 네트워크 분석, 수학영재교육