

토픽모델링을 활용한 SIAM Journal on Applied Mathematics의 연구 동향 분석

김성연
인천대학교 교육대학원

Analysis of Research Trends in SIAM Journal on Applied Mathematics Using Topic Modeling

Sung-Yeun Kim
Graduate School of Education, Incheon National University

요약 본 연구는 텍스트 마이닝 기법을 이용하여 산업수학과 관련한 논문들의 연구 현황 및 동향을 파악하는데 목적이 있다. 이를 위해 R로 1970년부터 2019년까지 SIAM Journal on Applied Mathematics 총 4910편 논문의 제목, 초록, 주제어를 수집하였으며, LDA 알고리즘 기반의 토픽모델링 분석을 수행하였다. 수집된 자료에 대한 coherence score 분석 결과, 토픽의 최적 개수는 20개로 결정하였으며, 핵심 연구 주제들은 Gibbs 샘플링 방법을 기반으로 추출하였다. 주요 분석 결과는 다음과 같다. 첫째, 해석학과 대수학을 중심으로 계산수학, 기하학, 수학적 모델링, 위상수학, 이산수학, 확률 및 통계학 등 다양한 수학 분야에서 산업수학 관련 연구가 진행되었다. 둘째, 연대별 연구 주제의 동향을 분석한 결과, 상승하는 연구 주제는 수리생물학, 비선형편미분방정식, 이산수학, 통계학, 위상수학으로, 하강하는 연구 주제는 확률론만 나타났다. 셋째, 2015개정 수학교육과정에서 반영되지 않은 분야 중 고등학교 수학교육과정에서 다루어야 할 내용으로 기수법, 행렬, 공간벡터, 복소수가 도출되었다. 마지막으로 분석 결과를 바탕으로 우리나라의 산업수학 활성화 방안과 본 연구의 제한점 및 후속 연구를 제시하였다.

Abstract The purpose of this study was to analyze the research status and trends related to the industrial mathematics based on text mining techniques with a sample of 4910 papers collected in the SIAM Journal on Applied Mathematics from 1970 to 2019. The R program was used to collect titles, abstracts, and key words from the papers and to analyze topic modeling techniques based on LDA algorithm. As a result of the coherence score on the collected papers, 20 topics were determined optimally using the Gibbs sampling methods. The main results were as follows. First, studies on industrial mathematics were conducted in a variety of mathematics fields, including computational mathematics, geometry, mathematical modeling, topology, discrete mathematics, probability and statistics, with a focus on analysis and algebra. Second, 5 hot topics (mathematical biology, nonlinear partial differential equation, discrete mathematics, statistics, topology) and 1 cold topic (probability theory) were found based on time series regression analysis. Third, among the fields that were not reflected in the 2015 revised mathematics curriculum, numeral system, matrix, vector in space, and complex numbers were extracted as the contents to be covered in the high school mathematical curriculum. Finally, this study suggested strategies to activate industrial mathematics in Korea, described the study limitations, and proposed directions for future research.

Keywords : Industrial Mathematics, LDA, Mathematics Curriculum, SIAM Journal on Applied Mathematics, Topic Modeling

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No.2019R1F1A1059437).

*Corresponding Author : Sung-Yeun Kim(Incheon National Univ.)

email: syk@inu.ac.kr

Received April 8, 2020

Revised April 28, 2020

Accepted July 3, 2020

Published July 31, 2020

1. 서론

최근 4차 산업혁명의 핵심인 인공지능과 빅데이터 분석 시장을 선점하려면 수학을 산업적으로 활용하는 산업 수학 경쟁력 확대, 관련 인재 육성, 그리고 산학 연구 기반 마련의 필요성에 대한 관심이 증대하고 있다[1]. 산업 수학은 수학적 지식 또는 방법론이 현대 산업의 문제에 적용된 것으로[2], 산업의 부가가치를 창출, 산업 현장에서 발생하는 문제를 해결, 경영상 의사결정, 자연현상 및 인간 활동을 예측, 그리고 신상품을 개발하는 등 다양한 분야에서 활용되고 있다. 특히 4차 산업혁명의 정보통신 기술이 경제 사회 전반에 걸쳐 급격한 변화를 야기함에 따라 이에 대응하기 위해 산업수학의 활용 및 역할이 강조되고 있다. 정부에서도 산업수학의 중요성을 인식하고, 2016년에 그동안 학교 안에 갇힌 수학, 사회적 고비용으로 인식되던 수학을 국가 산업경쟁력을 높이고 고용을 창출하는 수학으로 변화시키기 위한 전략과 과제를 포함하는 산업수학 육성방안을 마련하였다[3]. 2017년에는 4차 산업혁명 대응 핵심기술 분야에 특화하여 산업문제 해결과 인력양성의 거점이 될 산업수학센터를 선정하였다. 2019년에는 기초원천연구 중장기 지원방안을 통해 수학과 통계 분야 연구를 장려할 것이라고 밝히면서, 여전히 국내 산업수학 생태계 마련이 시급함을 강조했다[4]. 구체적으로 산업수학의 발전 단계는 태동기, 확산기, 도약기, 성장기, 성숙기, 그리고 정착기의 6기로 나누어 살펴볼 수 있다. 태동기는 미국, 독일 등 선진국에서 산업이 고도화되면서 수학이 산업의 기술적 문제를 해결하는 수단으로 활용되는 단계, 확산기는 산업수학 개념이 보편화되는 단계, 도약기는 수학 자체가 비즈니스가 되는 단계를 일컫는다. 이후 산업수학 육성방안에 따라 성장기는 정부 지원이 산업수학 육성을 위한 마중물이 되는 단계, 성숙기는 초, 중등 교육에 산업수학 저변을 형성하는 등 자발적 생태계가 작동되는 단계, 그리고 정착기는 민간주도하에 산업수학 생태계가 조성되는 단계이다[3].

또한 학계에서는 산업수학 활성화를 위한 노력으로 미국 등 선진국의 연구들에 대한 문헌분석이나 내용분석을 중심으로 한 정책 연구[2,5]와 특정 산업 분야에 초점을 맞춘 엔터테인먼트 산업수학 활성화 연구[6]가 수행되었다. 이러한 연구는 연구자가 가진 뛰어난 전문적 해석에 기반하여 향후 산업수학 활성화를 위한 정책의 방향과 전략을 제시해줄 수 있다는 장점이 있다. 그러나 다른 연구자들이 똑같은 방법으로 분석을 재현하는 것이 불가능하고, 분석 전에 연구자가 어떤 내용이 분석할 가치가 있

는지 사전에 알아야 하며, 연구자가 직접 수작업으로 자료를 수집 및 분류하므로 분석대상인 텍스트가 제한적일 수 있다는 한계가 지적되어 왔다[7,8].

따라서 본 연구는 연구자가 사전에 가정해야 할 내용이 없으며, 대용량의 자료를 다룰 수 있는 자동화된 텍스트 빅데이터 분석방법 중 방대한 양의 텍스트가 갖고 있는 주제를 도출하는데 효과적인 토픽모델링(topic modeling)의 잠재 디리클레 할당(Latent Dirichlet Allocation, LDA) 분석을 수행하고자 한다. LDA는 이산 자료들에 대한 확률적 생성 모형으로 단어들의 확률을 이용하여 문서 집합 내의 잠재된 토픽들을 찾아내는 방법이다[9,10]. LDA 분석은 동일한 의미를 지닌 다른 단어, 동일한 글자이지만 문맥에 따라 다른 단어를 효과적으로 변별할 수 있다. 또한 추출된 단어들은 토픽들 간의 독립성이 두드러지기 때문에 토픽을 결정짓는데 용이하다는 장점이 있다[7]. 토픽모델링의 LDA 분석을 적용해 연구 현황 및 동향을 살펴본 연구로는 경영학 분야[11], 관광 분야[12], 기후 분야[13], 보건 분야[7,14,15], 산업공학 분야[16], 수산학 분야[17], 정보시스템 분야[18-20], 커뮤니케이션 분야[21], 그리고 혁신 분야[10] 등 매우 다양하다. 그러나 현재까지 산업수학 분야의 전반적인 연구 동향을 파악하는 데 있어 토픽모델링을 적용한 연구는 거의 수행되지 않고 있는 실정이다.

본 연구에서는 현재 초, 중등 교육에 산업수학 저변을 형성해야하는 성숙기에 해당하는 우리나라 산업수학의 안정적인 정착 및 향후 나아갈 방향을 살펴보기 위해 LDA 분석을 통해 이미 산업수학이 발전해있는 미국의 대표적인 학술지의 과거와 현재를 살펴보고, 연구 현황과 학문적 추이를 분석하였다. 즉, 분석 결과를 통해 우리나라 산업수학의 향후 연구 방향과 산업수학과 관련하여 수학교육과정에서 구체적으로 다루어야 할 내용 영역을 제시함으로써 본 연구는 시대 변화에 적응할 수 있는 산업수학 활성화를 위한 전략을 수립하는 데 기초자료가 될 것이다. 구체적으로 분석대상은 전 세계 산업수학의 흐름을 주도하는 학회인 미국산업응용수학회(Society for Industrial and Applied Mathematics, SIAM)에서 출판되는 학술지 중 가장 오래되고 많은 논문들을 게재한 SIAM Journal on Applied Mathematics(SIAP)의 논문 제목, 초록, 주제어를 학회 홈페이지로부터 추출하였다.

구체적인 연구문제는 다음과 같다.

첫째, 산업수학 분야에서 연구자들이 관심을 갖는 연구 주제는 무엇인가?

둘째, 시간이 흐름에 따라 산업수학 분야의 연구 주제는 어떻게 변화하는가?

2. 연구방법

2.1 자료 수집

본 연구의 분석대상으로 선정한 SIAP은 산업수학 관련 학회 중 가장 긴 역사를 자랑하는 SIAM의 대표적인 학술지이다. SIAP의 연구 현황과 추이를 분석하기 위해 1953년부터 현재까지 게재된 6,152편의 논문 중 초록이 수록되어 있는 1970년부터 2019년까지 50년 동안 4,910편의 논문 제목, 초록, 주제어를 SIAM 홈페이지에서 R 프로그램을 작성하여 수집하였다.

Table 1. Number of abstracts published by years

Years	1970s	1980s	1990s	2000s	2010s	Total
Freq.	1165	895	939	946	965	4,910

Note. Fre.=Frequency.

수집된 논문들의 연대별 발행 빈도 분석을 살펴보면 Table 1에서 제시한 바와 같이 1970년대가 1,165편으로 가장 많았다. 반면에 1980년대가 895편으로 가장 작았으며, 그 후부터는 시간이 지남에 따라 논문 편수가 계속 증가하는 것으로 나타났다. 1970년대에 논문 편수가 많은 이유는 나머지 년대에는 1권씩 발행되던 학술지가 2권씩 발행되었기 때문이다.

2.2 자료 처리

수집한 자료에 전처리 과정은 먼저 분석에 필요한 단위로 텍스트를 나누는 토큰화(tokenization)를 수행한 뒤, 대문자는 소문자로 통일하였다. 다음으로, 구두점, 숫자, 수식, 대명사, 관사, 전치사 등 분석에 불필요한 불용어(stopwords)를 모두 제거했다. 또한 논문 제목과 초록에 빈번히 출현하는 ‘paper’, ‘study’, ‘result’ 등의 명사와 ‘present’, ‘determine’, ‘discuss’ 등과 같은 일반동사를 추가로 불용어에 포함시켜 삭제하였다. 마지막으로 문장 내에서 다양한 형태로 사용되는 단어들의 표제어를 추출하는 원형복원(lemmatization)을 수행했다. 예를 들어 ‘analyzed’, ‘analyzing’, 그리고 ‘analyzes’는 ‘analyze’로 추출하였다.

2.3 자료 분석

전처리 과정이 완료된 자료를 대상으로 LDA 분석을 실시하였다. Fig. 1은 LDA 알고리즘을 그래프로 제시한 것이다[22,23].

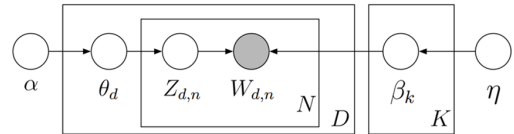


Fig. 1. LDA graphic model

Fig. 1에서 $W_{d,n}$ 은 문서 d 의 n 번째 단어로 문서에서 관측되는 변수, K 는 토픽의 수, α 는 θ_d 값을 결정하는 디리클레 분포 파라미터, η 는 β_k 값을 결정하는 파라미터이다. θ_d 는 문서별 토픽의 비율, β_k 는 토픽 내 단어들의 분포, $Z_{d,n}$ 은 문서 d 에서 n 번째 단어를 토픽에 할당하는 것이다[16]. 따라서 문서별 토픽 비율인 θ_d 에 따라 문서 내에 존재하는 단어들의 토픽인 $Z_{d,n}$ 이 결정되며, $Z_{d,n}$ 과 토픽내 단어들의 분포인 β_k 에 따라 단어 $W_{d,n}$ 이 결정된다[24].

LDA 분석을 하기 위해서는 연구자가 토픽의 수를 미리 지정해주어야 한다. 본 연구에서는 고전적인 방법으로 사용되고 있는 perplexity가 내재적으로 학습 성능 정도를 의미하고 있을 뿐 그 결과를 해석하기에 난해하다는 단점[25]을 해결한 Newman 등[26]이 제안한 coherence score를 이용하여 토픽의 수를 결정하였다. coherence score가 높을수록 LDA 분석 결과 산출되는 각각의 토픽이 의미론적으로 유사한 단어들로 구성되었다고 해석한다[14]. 본 연구에서는 최적의 토픽 수를 결정하기 위하여 토픽 수를 10부터 80까지 설정하여 coherence score를 산출하였으며, Fig. 2에 제시된 바와 같이 coherence score가 최대가 되는 20개로 정하여 연구 주제를 파악하였다.

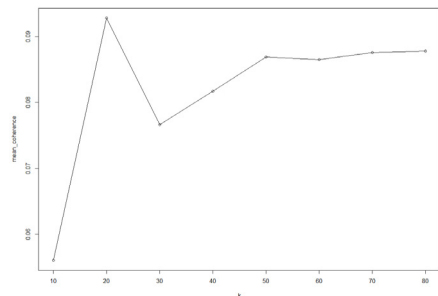


Fig. 2. Coherence score from 10 to 80 topics

연구 주제명은 연구 주제별 상위 단어들의 연관성을 바탕으로 대학수학 이상의 전공수학명으로 정하였으며, 이후 타당성 검토를 위해 산업수학 관련 경험이 있는 기하학, 해석학, 대수학, 통계학을 전공한 4인의 박사들로 구성된 전문가 협의회를 개최하였다. 또한 전문가 협의회에서 연구 주제와 관련하여 고교수학 교육과정에 필요한 수학, 그리고 연구 주제와 관련된 산업을 도출하였다.

연구 주제의 연대별 추세를 파악하기 위해서 독립변수로는 연대를, 종속변수로는 연구 주제(토픽)의 연대별 비중 평균값을 사용하여 시계열 선형회귀분석을 수행하였다. 분석 결과 유의수준 5%에서 통계적으로 유의한 회귀계수 값이 양수(+)이면 상승(hot) 주제, 그리고 음수이면 하강(cold) 주제로 구분하였다.

빈도분석과 워드 클라우드(Word cloud) 그림에는 NetMiner를, 그리고 나머지 분석에는 R 프로그램을 활용하였다. 구체적으로 LDA 분석에는 “textmineR” 패키지[28]와 “topicmodels” 패키지[29]가 사용되었다.

3. 연구결과

3.1 빈도 분석

1970년부터 2019년까지 SIAP에서 출판한 4,910편 논문의 제목, 초록, 주제어에서 전처리 과정이 완료된 자료에 빈도 분석을 실시한 결과는 Table 2와 Figure 2와 같다. Table 2에는 지면 제약 상 빈도 수 기준 상위 30개의 단어를 제시하였으며, Fig. 2에는 상위 300개의 단어를 워드 클라우드로 제시하였다. Table 2에서 보는 바

Table 2. Top 30 words

Rank	words	Freq.	Rank	words	Freq.
1	model	1860	16	application	737
2	method	1556	17	form	726
3	theory	1397	18	term	724
4	problem	1307	19	behavior	699
5	boundary	1231	20	limit	697
6	analysis	1213	21	stability	688
7	wave	1211	22	field	646
8	condition	1163	23	property	645
9	order	996	24	expansion	629
10	function	968	25	state	625
11	time	910	26	structure	619
12	parameter	883	27	perturbation	614
13	system	814	28	process	611
14	flow	761	29	effect	594
15	network	750	30	approximation	591

Note. Freq.=Frequency.

와 같이 ‘model’은 1,869번, ‘method’는 1,556번, ‘theory’는 1,397번, ‘problem’은 1,307번, boundary는 ‘1,231’번의 순으로 나타났다. 또한 Fig. 2를 통해 빈도 분석에 기반한 빈도 수가 높은 단어일수록 그 값에 비례하여 단어의 크기가 크게 표현되었음을 알 수 있다.

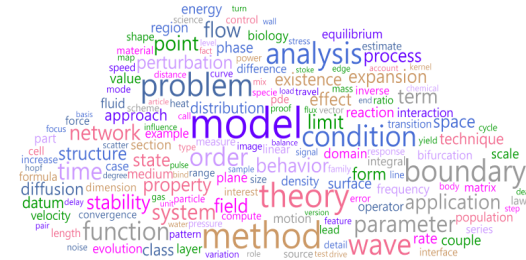


Fig. 3. Word cloud by top 300 words

3.2 토픽모델링 분석

SIAP에서 중심이 되는 주제 및 관련된 핵심 어휘를 추출하기 위하여 LDA를 분석한 결과는 Table 3과 같다. Table 3에는 각 연구 주제에 대해 빈도수 기준 상위 핵심 어휘로 10개 단어들을 제시하였다.

Table 3. Topic modeling analysis

Topic name	Top 10 words in each topic
1-optimization	condition, function, existence, linear, program, matrix, convex, vector, optimization, strategy
2-mathematical biology	model, population, biology, stability, cell, rate time, structure, disease, equilibrium.
3-differential equation	boundary, diffusion, model, order, time, limit, behavior, method, condition, expansion
4-analytic geometry	flow, fluid, surface, boundary, velocity, stoke, effect, method, layer, shape
5-nonlinear pde	boundary, inverse, datum, operator, method, domain, condition, function, measurement, field
6-calculus	model, diffusion, bifurcation, stability, limit, analysis, growth, existence, predator, prey
7-Fourier analysis	wave, perturbation, method, order, expansion, theory, boundary, approximation, point, analysis
8-discrete mathematics	network, model, couple, time, stability, biology, term, linear, analysis, neuron
9-statistics	model, datum, conservation, simulation, flow, method, transport, device, semiconductor, flux
10-nonlinear functional analysis	wave, stability, bifurcation, analysis, travel, time, theory, state, parameter, structure

11-probability theory	function, process, distribution, time, integral, application, probability, density, formula, space
12-harmonic analysis	wave, scatter, field, medium, method, frequency, theory, time, propagation, helmholtz
13-applied pde	theory, phase, limit, analysis, pattern, parameter, model, behavior, homoclinic, formation
14-topology	model, mechanic, fluid, material, structure, homogenization, surface, energy, stress, phase
15-combinatorics	function, point, class, set, theorem, existence, graph, linear, algorithm, tree
16-linear algebra	function, matrix, method, operator, property, form, theory, condition, linear, kernel
17-numerical analysis	method, algorithm, process, application, analysis, model, convergence, approach, approximation, image
18-mathematical modeling	heat, reaction, temperature, model, energy, diffusion, analysis, combustion, transfer, flow
19-stochastic process	process, time, distribution, probability, theory, application, model, state, rate, markov
20-computational mathematics	expansion, order, function, series, motion, integral, approximation, boundary, perturbation, particle

Note. pde=partial differential equation.

연구 주제 1은 ‘condition’, ‘function’, ‘existence’, ‘optimization’ 등으로 구성되어 있으며, 최적화이론과 관련되어 있다. 관련 산업으로는 운송, 수송, 물류시스템 등을 들 수 있으며, 고등학교 수학교육과정에 방정식과 부등식, 알고리즘, 행렬 등이 필요하다고 해석할 수 있다.

연구 주제 2는 ‘model’, ‘biology’, ‘cell’, ‘disease’ 등으로 구성되어 있으며, 수리생물학과 관련되어 있다. 관련 산업으로는 보건, 생명공학 등을 들 수 있으며, 고등학교 수학교육과정에 미적분, 함수와 그래프, 그리고 지수와 로그 등이 필요하다고 해석할 수 있다.

연구 주제 3은 ‘boundary’, ‘diffusion’, ‘order’, ‘expansion’ 등으로 구성되어 있으며, 미분방정식과 관련되어 있다. 관련 산업으로는 운송, 수송시스템 등을 들 수 있으며, 고등학교 수학교육과정에 미적분 등이 필요하다고 해석할 수 있다.

연구 주제 4는 ‘surface’, ‘stoke’, ‘layer’, ‘shape’ 등으로 구성되어 있으며, 해석기하학과 관련되어 있다. 관련 산업으로는 애니메이션, 철강, 항공 등을 들 수 있으며, 고등학교 수학교육과정에 공간도형과 공간좌표, 미적분 등이 필요하다고 해석할 수 있다.

연구 주제 5는 ‘inverse’, ‘datum’, ‘operator’, ‘function’ 등으로 구성되어 있으며, 비선형편미분방정식

과 관련되어 있다. 관련 산업으로는 의료영상, 정밀진단, 성능평가 등을 들 수 있으며, 고등학교 수학교육과정에 집합, 함수, 미적분, 알고리즘 등이 필요하다고 해석할 수 있다.

연구 주제 6은 ‘model’, ‘limit’, ‘analysis’, ‘predator’ 등으로 구성되어 있으며, 미적분학과 관련되어 있다. 관련 산업으로는 보건 및 개체 수 관련의 환경 등을 들 수 있으며, 고등학교 수학교육과정에 미적분, 함수의 그래프, 지수와 로그, 공간벡터, 수학적 모델링 등이 필요하다고 해석할 수 있다.

연구 주제 7은 ‘wave’, ‘perturbation’, ‘method’, ‘expansion’ 등으로 구성되어 있으며, 푸리에해석학과 관련되어 있다. 관련 산업으로는 항공, 선박, 시스템 공학 등을 들 수 있으며, 고등학교 수학교육과정에 미적분 및 수학적 모델링 등이 필요하다고 해석할 수 있다.

연구 주제 8은 ‘network’, ‘model’, ‘stability’, ‘neuron’ 등으로 구성되어 있으며, 이산수학과 관련되어 있다. 관련 산업으로는 의료 및 딥러닝과 관련한 인공지능 관련 산업 등을 들 수 있으며, 고등학교 수학교육과정에 그래프와 트리, 함수와 그래프, 삼각함수 등이 필요하다고 해석할 수 있다.

연구 주제 9는 ‘model’, ‘datum’, ‘simulation’, ‘semiconductor’ 등으로 구성되어 있으며, 통계학과 관련되어 있다. 관련 산업으로는 반도체 및 전자 등을 들 수 있으며, 고등학교 수학교육과정에 미적분 및 수학적 모델링, 확률 및 통계 등이 필요하다고 해석할 수 있다.

연구 주제 10은 ‘stability’, ‘bifurcation’, ‘travel’, ‘state’ 등으로 구성되어 있으며, 비선형함수해석학과 관련되어 있다. 관련 산업으로는 환경, 전자, 기계 등을 들 수 있으며, 고등학교 수학교육과정에 방정식, 미적분, 수학적 모델링, 복소수 등이 필요하다고 해석할 수 있다.

연구 주제 11은 ‘distribution’, ‘probability’, ‘density’, ‘formula’ 등으로 구성되어 있으며, 확률론과 관련되어 있다. 관련 산업으로는 데이터 분석, 금융 및 증권 등을 들 수 있으며, 고등학교 수학교육과정에 확률 및 통계, 알고리즘, 금융과 모델링 등이 필요하다고 해석할 수 있다.

연구 주제 12는 ‘wave’, ‘scatter’, ‘frequency’, ‘helmholtz’ 등으로 구성되어 있으며, 조화해석학과 관련되어 있다. 관련 산업으로는 데이터 분석, 전자 등을 들 수 있으며, 고등학교 수학교육과정에 미적분, 공간도형과 공간좌표 등이 필요하다고 해석할 수 있다.

연구 주제 13은 ‘phase’, ‘pattern’, ‘behavior’,

'homoclinic' 등으로 구성되어 있으며, 응용편미분방정식과 관련되어 있다. 관련 산업으로는 생명공학, 양자역학, 환경, 보건 등을 들 수 있으며, 고등학교 수학교육과정에 미적분, 수학적 모델링, 공간좌표 등이 필요하다고 해석할 수 있다.

연구 주제 14는 'mechanic', 'structure', 'homogenization', 'surface' 등으로 구성되어 있으며, 위상수학과 관련되어 있다. 관련 산업으로는 철강, 신소재, 재료공학 등을 들 수 있으며, 고등학교 수학교육과정에 미적분, 공간좌표, 수학적 모델링 등이 필요하다고 해석할 수 있다.

연구 주제 15는 'class', 'set', 'graph', 'tree' 등으로 구성되어 있으며, 조합론과 관련되어 있다. 관련 산업으로는 데이터 분석 및 소프트웨어 개발 등을 들 수 있으며, 고등학교 수학교육과정에 집합, 행렬, 그래프, 공간좌표 등이 필요하다고 해석할 수 있다.

연구 주제 16은 'matrix', 'operator', 'linear', 'kernel' 등으로 구성되어 있으며, 선형대수학과 관련되어 있다. 관련 산업으로는 수송, 물류, 시스템 개발 등을 들 수 있으며, 고등학교 수학교육과정에 행렬, 벡터, 변환, 알고리즘 등이 필요하다고 해석할 수 있다.

연구 주제 17은 'algorithm', 'model', 'approach', 'approximation' 등으로 구성되어 있으며, 이산수학과 관련되어 있다. 관련 산업으로는 데이터 분석, 컴퓨터, 소프트웨어 개발 등을 들 수 있으며, 고등학교 수학교육과정에 기수법, 집합, 행렬, 알고리즘 등이 필요하다고 해석할 수 있다.

연구 주제 18은 'heat', 'reaction', 'model', 'analysis' 등으로 구성되어 있으며, 수학적모델링과 관련되어 있다. 관련 산업으로는 환경, 에너지 등을 들 수 있으며, 고등학교 수학교육과정에 미적분, 수학적 모델링 등이 필요하다고 해석할 수 있다.

연구 주제 19는 'process', 'time', 'probability', 'markov' 등으로 구성되어 있으며, 확률과정론과 관련되어 있다. 관련 산업으로는 데이터분석, 물류, 운송 시스템 등을 들 수 있으며, 고등학교 수학교육과정에 확률 및 통계 등이 필요하다고 해석할 수 있다.

연구 주제 20은 'expansion', 'series', 'integral', 'approximation' 등으로 구성되어 있으며, 계산수학과 관련되어 있다. 관련 산업으로는 데이터분석, 동력기계 등을 들 수 있으며, 고등학교 수학교육과정에 수열, 급수, 미적분, 복소수 등이 필요하다고 해석할 수 있다.

3.3 시계열 선형회귀분석

1970년부터 2019년까지를 10년 단위로 묶은 연대별 연구 주제의 추이를 파악하는 시계열 선형회귀분석 결과는 Table 4와 같다. 또한 Fig. 4에는 이 중 상승 연구 주제와 하강 연구 주제를 나타냈다.

상승하는 연구 주제로는 주제 2인 수리생물학, 주제 5인 비선형편미분방정식, 주제 8인 이산수학, 주제 9인 수치해석학, 그리고 주제 14인 수리물리학으로 총 4개가 나타났다. 반면에 하강하는 연구 주제로는 주제 11인 확률론만 나타났다.

Table 4. Results of regression analysis for hot/cold topics

Topic name	Coefficient(SE)	p-value	Hot/Cold
1-optimization	-0.017(0.011)	0.249	-
2-mathematical biology	0.016(0.003)	0.016	Hot
3-differential equation	0.007(0.003)	0.133	-
4-analytic geometry	0.003(0.002)	0.285	-
5-nonlinear pde	0.012(0.002)	0.007	Hot
6-calculus	-0.001(0.004)	0.769	-
7-Fourier analysis	-0.003(0.004)	0.483	-
8-discrete mathematics_	0.010(0.001)	0.007	Hot
9-statistics	0.014(0.001)	0.001	Hot
10-nonlinear functional analysis	0.001(0.004)	0.976	-
11-probability theory	-0.010(0.002)	0.013	Cold
12-harmonic analysis	0.007(0.003)	0.097	-
13-applied pde	0.004(0.005)	0.449	-
14-topology	0.014(0.003)	0.011	Hot
15-combinatorics	-0.022(0.009)	0.096	-
16-linear algebra	-0.026(0.008)	0.054	-
17-numerical analysis	0.002(0.001)	0.241	-
18-mathematical modeling	-0.001(0.006)	0.900	-
19-stochastic process	-0.004(0.004)	0.469	-
20-computational mathematics	-0.006(0.003)	0.130	-

Note. pde=partial differential equation; SE=standard errors

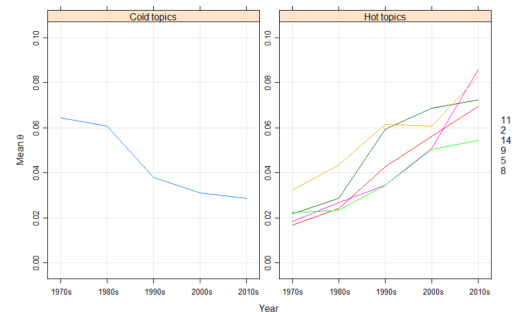


Fig. 4. Hot and cold topics

4. 결론 및 논의

본 연구에서는 LDA 기법의 토픽모델링과 시계열 선형회귀분석을 적용하여 지난 50여년간 산업수학과 관련한 대표적인 학술지인 SIAP에서 출판한 논문들의 제목, 초록, 주제어의 자료를 분석함으로써 산업수학 분야 연구의 현황과 추이를 파악하였다. 주요 연구결과를 바탕으로 본 논문의 의의와 국내에서 산업수학을 활성화시킬 수 있는 방안을 논하면 다음과 같다.

첫째, coherence score의 산출을 통해 20개 연구 주제로 분류한 결과, 해석학과 대수학을 중심으로 계산수학, 기하학, 위상수학, 이산수학, 확률 및 통계학, 수학적 모델링 등 다양한 수학 분야에서 산업수학 관련 연구가 진행되고 있음을 확인하였다. 이는 그래프 및 행렬 등 대수학을 이용해 구글의 검색엔진과 알파고의 알고리즘을 만들고, 위상수학을 이용해 암 발생 확률 예측 소프트웨어를 만들고, 수학적 모델링을 활용하여 150억 달러 헤지펀드를 만드는 등 수학으로 산업의 부가가치를 창출하는 기업의 사례를 통해서도 확인할 수 있다[2]. 이처럼 본 논문에서 제시된 방법에 의해 컴퓨터가 도출한 분석 결과가 기존의 전문가 분석 결과와 일치한다는 점에서 LDA 기법의 토픽모델링은 산업수학 분야의 현황과 추세를 파악하기에 제안된 방법이 적합함을 제시하였다는 점에서 의의가 있다. 이를 바탕으로 본 논문에서 제시한 분석방법 및 처리과정은 다른 분야에서도 적용 가능하며, 산업수학 관련 기사 및 SNS 데이터에 토픽모델링을 적용하여 실제 현장의 산업수학 동향을 파악하는 데에도 유용하게 활용될 수 있을 것이다.

둘째, 연대별 연구 주제의 추이를 분석한 결과, 상승하는 연구 주제로는 수리생물학, 비선형편미분방정식, 이산수학, 통계학, 위상수학으로 총 5개가 나타났다. 반면에 하강하는 연구 주제로는 확률론만 나타났다. 상승하는 연구 주제들은 상상이 현실이 되는 4차 산업혁명시대에 지능과 정보가 결합되면서 맞춤, 분권, 개방, 초연결, 초지능화, 자동화, 인간 생명에 관련된 산업의 인기가 많아지면서 여기에서 활용되고 있는 수학 분야로 해석할 수 있다. 반면에 하강하는 연구 주제는 데이터 분석과 관련하여 고전적인 방법론에 입각한 연구 분야에 대한 수요가 낮아지고 있기때문이라고 해석할 수 있다. 정효정[10]은 새로운 패러다임을 맞이하는 혁신 분야의 국외 연구 동향을 분석한 결과, 국외 연구 동향에 따른 연구 주제의 무분별한 도입을 지양할 것을 제안한 바 있다. 마찬가지로 현재 산업기술의 패러다임을 모방기술-추격형에서 원

천기술-선도형으로 구축[2]하기 위해서는 우리나라의 산업수학 연구가 외국의 산업수학 연구 동향에 따른 연구 주제의 무조건적인 도입은 지양할 필요가 있다. 따라서 우리나라의 산업수학 연구가 이미 20년 전부터 산업수학이 발전해있는 외국의 산업수학 연구 동향과의 격차를 줄이고, 독자적인 연구 영역을 확보하기 위해 우리나라의 상황과 정부 현안을 고려한 국내 연구자들의 지속적인 노력이 요구된다.

셋째, 연구 주제를 바탕으로 현 2015개정 교육과정 중 공통수학과 일반선택에 반영되지 않은 분야 중 고등학교 수학교육과정에서 다루어야 할 내용으로 기수법, 행렬, 공간벡터, 복소수가 도출되었다. 이러한 수학 내용은 영국, 싱가포르, 호주에서는 모두 다루고 있으며, 미국, 중국, 일본의 경우도 최소 2개 이상은 다루고 있는 현실이다. 일례로 행렬은 비교적 단순한 이론으로 수학을 현장에 적용할 때 항상 여러 개의 변수가 나오는데 이 중 가장 쉬운 일차함수와 이차함수를 다루는 수학적 방법이다. 이에 김영옥 외[30]는 행렬 없이는 미적분도 없고 아무 것도 안된다고 하였으며, 수학의 역할이 강조되면서 기하를 포함한 복합적 사고가 가능하도록 우리나라의 수학 학습 내용을 강화할 필요가 있다고 밝혔다. 또한 이해성 외[1]에 따르면 행렬의 내용 삭제는 4차 산업혁명 시대 가장 적용성이 큰 컴퓨터 알고리즘 작성과 빅데이터 처리에 가장 중요한 내용임을 감안할 때, 학생들이 이를 통해 얻을 수 있는 많은 미래의 기회를 박탈한 것과 같다고 발표한 바 있다. 따라서 고등학교 수학 교육과정에 이와 관련한 내용과 산업수학의 가치를 소개하는 내용을 함께 반영함으로써 학교 현장에서 산업수학을 활성화시킬 수 있을 것이다.

마지막으로 본 연구는 SIAP 학술지 논문으로만 분석의 범위를 한정하였기 때문에 SIAM에서 출판하는 다른 학술지들과 외국의 다른 학술지들이 포함되지 않았다는 한계가 있다. 따라서 분석 데이터에 SIAM에서 출판하는 모든 학술지들과 유럽의 산업수학 관련학과 및 연구기관들의 컨소시엄(European Consortium for Mathematics in Industry)에서 발간하는 학술지들의 논문을 분석 데이터에 추가함으로써 산업수학 연구의 국제 현황 및 동향을 면밀하게 파악하는 연구를 제안한다. 또한 본 연구에서는 약 10년을 단위로 산업수학의 연구 동향을 파악하였지만 산업수학의 최신 동향을 파악하기에는 한계가 있을 수 있다. 따라서 기간에 대한 분석 단위를 짧게 하거나, 특정 년도 이후로 한정하여 매년 분석하여 보다 세부적인 진화 과정을 파악하는 연구를 제안한다.

References

- [1] H. S. Lee, J. K. Park, O. J. Jung, AI mathematics is a global craze, but it's been deleted from high school curriculum in Korea [Internet], Hankyung, [Cited 2019 Oct. 06] Available From: https://news.naver.com/main/ranking/read.nhn?mid=etc&sid1=111&rankingType=popular_day&oid=015&aid=0004219450&date=20191006&type=1&rankingSeq=2&rankingSectionId=105 (accessed Apr. 07, 2020).
- [2] K. B. Park, Linkage between industry and academia for promoting industrial mathematics, Policy Research Report, Science and Technology Policy Institute, Korea, pp.1-122.
- [3] Ministry of Future, Planning and Science, Mathematics needed in industrial fields, the government will raise it, Press release, Sejong: Ministry of Future, Planning and Science, 2016.
- [4] T. H. Kim, Without mathematics, there is no AI. It is imperative to prepare a domestic industrial mathematics ecosystem [Internet], Chosunbiz, [Cited 2019 September 16] Available From: https://biz.chosun.com/site/data/html_dir/2019/09/16/2019091601712.html (accessed Apr. 07, 2020)
- [5] M. H. Kim, A study on strategy for vitalizing of industrial and applied mathematics as a national policy, Sejong: Ministry of Future, Planning and Science, 2015.
DOI:<https://doi.org/10.15724/jslhd.2019.28.3.081>
- [6] S. J. Kim, J. M. Sung, "Activation plan for entertainment industry mathematics", *Journal of Korea Entertainment Industry Association*, Vol.11, No.2, pp.289-297, Feb. 2017.
DOI:<https://doi.org/10.21184/jkeia.2017.02.11.2.289>
- [7] M. S. Shin, K. W. Cho, "Analysis of topic modeling and trend of journal of speech-language & hearing disorders using text mining: (2002~2018)", *Journal of Speech-Language & Hearing Disorders*, Vol.28, No.3, pp.81-91, Jul. 2019.
DOI:<https://doi.org/10.15724/jslhd.2019.28.3.081>
- [8] J. W. Mohr, P. Bogdanov, "Introduction—topic models: What they are and why they matter", *Poetics*, Vol.48, No.6, pp.545-569, Dec. 2013.
DOI:<https://doi.org/10.1016/j.poetic.2013.10.001>
- [9] T. K. Kim, H. R. Choi, H. C. Lee, "A study on the research trends in Fintech using topic modeling", *Journal of the Korean Academia-Industrial Cooperation Society*, Vol.17, No.11, pp.670-681, Nov. 2016.
DOI:<http://dx.doi.org/10.5762/KAIS.2016.17.11.670>
- [10] H. J. Jung, "Research dynamics in innovation studies using text mining", *Journal of Technology Innovation*, Vol.24, No.4, pp.249-275, Nov. 2016.
DOI:<https://doi.org/10.14383/SIME.2016.24.4.249>
- [11] A. Amado, P. Cortez, P. Rita, 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, No.1, pp.1-7, Jan. 2018.
DOI:<https://doi.org/10.1016/j.iedeen.2017.06.002>
- [12] H. J. Yoon, C. S. Kim, K. Y. Kwahk, An analysis of research trends on local tour studies: Using text mining techniques, *Proceedings of the Korean Academic Society of Tourism and Leisure*, Korea, pp.275-282, Dec. 2017.
- [13] J. H. Yoo, E. C. Jeon, H. N. Kim, "Study of research trends in climate change using text analysis-focusing on journal of climate change research-", *Journal of Climate Change Research*, Vol.10, No.3, pp.161-172, Sep. 2019.
DOI:<https://doi.org/10.15531/kscrcr.2019.10.3.161>
- [14] H. J. Yoon, J. H. Park, J. W. Yoon, "Introduction of topic modeling for extracting potential information from unstructured text data: Issue analysis on news article of dementia-related physical activity", *Korean Journal of Sport Science*, Vol.30, No.3, pp.501-512, Sep. 2019.
DOI:<https://doi.org/10.24985/kjss.2019.30.3.501>
- [15] K. W. Cho, S. K. Bae, Y. W. Woo, "Analysis on topic trends and topic modeling of KSHSM journal papers using text mining", *The Korean Journal of Health Service Management*, Vol.11, No.4, Nov. pp.213-224.
DOI:<http://dx.doi.org/10.12811/kshsm.2017.11.4.213>
- [16] S. G. Kim, S. Y. Jang, "A study on the research trends in domestic industrial and management engineering using topic modeling", *Journal of the Korea Management Engineers Society*, Vol.21, No.3, pp.71-95, Sep. 2016.
DOI:<https://doi.org/10.5762/KAIS.2016.17.11.670>
- [17] S. Syed, C. T. Weber, "Using machine learning to uncover latent research topics in fishery models", *Reviews in Fisheries Science & Aquaculture*, Vol.26, No.3, pp.319-336, Jan. 2018.
- [18] C. S. Kim, S. J. Choi, K. Y. Kwahk, "Investigation of research trends in information systems domain using topic modeling and time series regression analysis", *Journal of Digital Contents Society*, Vol.18, No.6, pp.1143-1150, Oct. 2017.
- [19] J. H. Park, M. Song, "A study on the research trends in library & information science in Korea using topic modeling", *Journal of the Korean Society for Information Management*, No. 30, No.1, pp.7-32, Mar. 2013.
DOI:<https://doi.org/10.3743/KOSIM.2013.30.1.007>
- [20] J. G. Ahn, S. D. Kim, H. W. Kim, "Exploring dynamics of information systems research trend using text mining approach", *Information Systems Review*, pp.73-96, Sep. 2016.
DOI:<https://doi.org/10.14329/isr.2016.18.3.073>
- [21] K. Lee, H. Jung, M. Song, "Subject-method topic network analysis in communication studies". *Scientometrics*, Vol.109, No.3, pp.1761-1787, Dec.

2016.
DOI:<https://doi.org/10.14329/isr.2016.18.3.073>
- [22] D. M. Blei, "Probabilistic topic models: Surveying a suite of algorithms that offer a solution to managing large document archives", *Communications of the ACM*, Vol.55, No.4, pp.77-84, Apr. 2012.
DOI:<https://doi.org/10.1007/s11192-016-2135-7>
- [23] D. M. Blei, "Probabilistic topic models", *2013 Topic Modeling Workshop at NIPS*, 2013.
- [24] J. H. Park, H. J. Oh, "Comparison of Topic Modeling Methods for Analyzing Research Trends of Archives Management in Korea: focused on LDA and HDP", *Journal of Korean Library and Information Science Society*, Vol.48, No.4, pp.235-258, Dec. 2017.
DOI:<https://doi.org/10.16981/kliss.48.4.201712.235>
- [25] J. Chang, S. Gerrish, C. Wang, J. L. Boyd-Graber, D. M. Blei, "Reading tea leaves: How humans interpret topic models", *In Advances in neural information processing systems*, pp.288-296, Jan. 2009.
- [26] D. Newman, J. H. Lau, K. Grieser, 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*, CA, USA, pp.100-108, 2010.
- [27] T. Jones, W. Doane, M. T. Jones, Functions for text mining and topic modeling [Internet], TextmineR, [Cited 2019 Jan. 01] Available From: <https://www.rtextminer.com>, (accessed Apr. 07, 2020).
- [28] K. Hornik, B. Grün, "topicmodels: An R package for fitting topic models", *Journal of statistical software*, Vol.40, No.13, pp.1-30, May. 2011.
- [29] Y. W. Kim, S. E. Ko, Y. L. Kim, H. S. Park, S. W. Lee, J. W. Jang, D. S. Cho, A study on the contents of mathematics learning suitable for future students, Korea Foundation for the Advancement of Science and Creativity, Korea, pp.110-111.

김 성 연(Sung-Yeun Kim)

[정회원]



- 1997년 2월 : 영남대학교 영남대학원 수학과 (이학석사)
- 2003년 8월 : 성균관대학교 성균관대학원 수학과 (통계학박사)
- 2011년 2월 : 연세대학교 연세대학원 교육학과 (교육학박사)
- 2016년 3월 ~ 현재 : 인천대학교 교육대학원 부교수

<관심분야>

교육측정 및 평가, 교원평가, 네트워크분석, 수학영재교육