한국시뮬레이션학회 논문지
Vol. 25, No. 3, pp. 107-116 (2016. 9)
http://dx.doi.org/10.9709/JKSS.2016.25.3.107
ISSN 1225-5904

토픽 모델링을 이용한 시뮬레이션 연구 동향 분석

나상태 · 김자희 † · 정민호 · 안주언

Trend Analysis using Topic Modeling for Simulation Studies

Sang-Tae Na · Ja-Hee Kim† · Min-Ho Jung · Joo-Eon Ahn

ABSTRACT

The recent diversification in terms of the scope and techniques used for simulations has highlighted the importance of analyzing state of the art trends and applying these for educational and study purposes. While qualitative methods such as literature research or experts' assessments have previously been used, such methods are in fact likely to reflect the subjective viewpoint of experts, and to involve too much time and money for the results obtained. For the purpose of an objective analysis, a quantitative analysis that included the examination of topics found in domestic academic journal articles was conducted in the present study. In this regard, simulation was found to be most actively used domestically in the electrical and electronic fields. In addition, simulation was also found to be employed for the purpose of education and entertainment in the social sciences. The results of this study are expected to help to facilitate the prediction of the direction of the development of not only the Korea Society for Simulation, but also domestic simulation studies. This study also raises the possibility of applying text mining to trend analysis, and proves that it can be a useful method for deriving future key topics and helping experts' decisions regarding quantitative data.

Keywords: *Text mining, Topic analysis, Simulation, LDA, Trend analysis*

요 약

시뮬레이션의 활용범위와 기법이 나날이 다양해지면서 시뮬레이션의 최신 연구 동향을 분석하고 이를 대학 교육과 연구에 적용하는 노력이 요구된다. 기존에는 트렌드 분석을 위해 문헌조사 또는 전문가 평가와 같은 정성적인 연구방법이 주로 사용되었으나 이런 방법들은 많은 시간과 비용이 소요될 뿐만 아니라 전문가의 주관적인 관점이 반영될 가능성이 있다. 본 연구에서는 객관적 분석을 위해 국내 학술 논문에 대하여 토픽분석을 포함한 정량적 분석을 실시하였다. 그 결과 국내에서는 시뮬레이션이 전기전자 분야에서 가장 활발하게 활용된다는 사실을 발견하였다. 또한 사회 과학에서는 교육 및 오락의 목적으로도 활용됨을 알 수 있었다. 이 연구 결과는 국내 시뮬레이션 연구와 한국 시뮬레이션 학회가 어떤 방향으로 발전할지를 예측하는 데 도움이 된다. 본 연구결과는 시뮬레이션 활용 연구 분야의 핵심 토픽을 도출하기 위하여 텍스트마이닝 기반의 트렌드분석에 대한 활용 가능성을 제시하고, 텍스트마이닝이 미래예측 키워드를 도출하는 유용한 방법임을 증명하였으며, 전문가들의 정성적인 자료를 보조하는 정량적인 자료분석 방법으로 유용할 것으로 기대된다.

주요어: 텍스트 마이닝, 토픽분석, 시뮬레이션, LDA, 트렌드 분석

1. 서론

시뮬레이션(Simulation)은 18세기의 바늘 실험 (Needle experiment)으로 시작하여 초기에는 몬테 카를로 방법 (Monte Carlo method) 위주였으나, 컴퓨터의 발달과 함께 사용범위가 확대되고 방법론도 다양해지고 있다 (Goldsman et al., 2010). 시뮬레이션(Simulation)이란 모방하다(Simulate)의 명사형으로 모의실험으로도 번역되

본 연구는 미래창조과학부의 2016년 고용계약형 SW석사과정 지원 사업을 지원받아 수행한 결과입니다.

Received: 15 July 2016, **Revised**: 22 August 2016, **Accepted**: 25 August 2016

† **Corresponding Author**: Ja-Hee Kim
E-mail: jahee@seoultech.ac.kr
Graduate school of policy and information technology, SeoulTech, Seoul, Korea

는데, 이중에서도 컴퓨터 시뮬레이션은 수학적으로 모형화하기 어려운 현실 세계의 복잡한 시스템 또는 현상을 컴퓨터로 모형화하고, 가상적으로 수행하여 의사결정을 지원하는 경영과학 기법이다(KIIE, 1992). 시뮬레이션을 사용하는 목적이 초기에는 시스템을 이해하고 성능을 향상시키기 위한 것이 대부분이었지만, 점차 교육 및 훈련 지원이나 시스템 적합도(System acceptance) 분석으로 활용분야가 확장되었다. 최근에는 적용 범위가 더욱 확대되어 가상 환경이 필요한 연구나 오락(Entertainment)분야에도 시뮬레이션 기법이 사용되고 있다(Nance and Sargent, 2002). 이와 같이 시뮬레이션은 사용범위가 확대될 뿐만 아니라 전통적 활용분야인 산업공학 분야에서 여전히 가장 많이 사용되는 기법이다(Cho et al, 2014). 시뮬레이션의 사용 분야와 목적이 다변화되고 기법들도 다양해짐에 따라 시뮬레이션 연구 흐름에 대해 이해하고, 향후 방향을 전망하는 것이 필요하다. 국제적으로는 WSC (Winter simulation conference)의 주도하에 주기적으로 시뮬레이션의 발전사를 정리하고 있지만 국가별 연구는 따로 진행되지 않고 있다(Goldsman et al. 2010). 국내에서도 시뮬레이션의 기법에 대한 조사와 발전 방향에 대한 연구가 일부 있지만 자신의 연구 분야에 기반한 정성적인 연구가 대부분이다(Song, 2012). 이에 본 연구는 우리나라에서 시뮬레이션이 어떤 분야에 사용되고 있는지 알아보고 연구 분야별 시계열 분석을 통해 미래 발전방향을 예측하고자 한다.

기존에는 트렌드 분석을 위해 문헌 조사, 전문가 평가, 델파이(Delphi) 기법과 같은 정성적(Qualitative) 연구방법이 주로 사용되었다(Noh, 2006). 하지만 이런 정성적 방법론은 대량의 정보로부터 결과를 도출하는데 많은 시간과 비용이 소요될 뿐만 아니라 전문가의 주관적인 가치가 반영될 가능성이 있다. 최근에는 정성적 연구방법의 한계를 보완하고 객관성 및 실효성 높은 트렌드 분석 결과를 확보하기 위하여 다양한 분야에서 텍스트 마이닝 등과 같은 정량적(Quantitative) 연구방법을 도입하고 있다(Oh, 2015). 본 연구에서는 빅데이터 분석방법 중 하나인 토픽분석(Topic analysis)을 사용하여 비정형 텍스트 자료에 근거한 정량적 분석을 수행한다. 토픽분석은 키워드 수준에서 단일 단어의 의미 파악이 어렵다는 한계를 보완하기 위해 텍스트 문서집합 내에 잠재되어 있는 전반적인 토픽을 도출하는 기법이다(Hornik and Grün, 2011). 본 연구는 핵심 키워드 도출 및 트렌드 분석을 위한 연구대상을 일정 수준 이상의 연구논문으로 한정하기 위하여 한국연구재단에 등재되거나 등재 후보로써 전문학술지에 게재된 논문을 대상으로 분석한다.

본 연구의 전체구성은 1장에서 본 연구의 필요성에 대한 문제를 제기하고, 2장에서 본 연구의 연구절차와 각 절차에서 이용한 방법론을 정리한다. 3장에서 연구 분류별 분석을 실시하고, 4장에서는 토픽분석을 활용한 시뮬레이션 연구동향을 제시하였다. 마지막으로 5장에서는 본 연구의 결론을 기술한다.

2. 연구방법

본 연구는 Figure 1과 같이 KCI 등재된 저널의 논문 중에서 시뮬레이션이라는 키워드가 있는 논문들을 취득하여 논문이 실린 분야와 시간적 트렌드 분석, 토픽분석의 순서로 연구를 진행하였다.

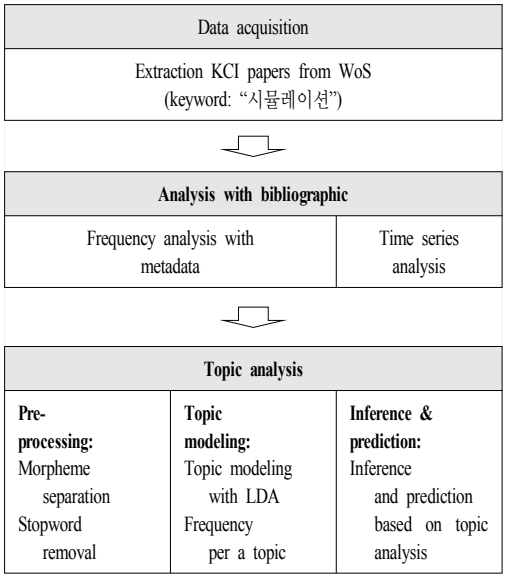


Fig. 1. Flow diagram of topic analysis

2.1 데이터의 수집 및 정리

본 연구에서는 일정 수준 이상의 연구 논문을 자료 추출 대상으로 삼기위해 KCI DB (Korea citation index database)에 저장된 논문을 자료선정 기준으로 삼았다. KCI DB에 저장된 논문들은 한국연구재단이 평가하여 선정된 등재(후보)학술지에 게재된 논문으로 학술지 수준이 일정 수준 이상으로 인정된다. 데이터 수집방법은 톰슨로이터(Thomson Reuters Co.)의 WoS (Web of Science)가 제공하는 KCI DB에서 "시뮬레이션"이라는 한글 검색어로 문헌들을 검색하여, 1992년부터 2015년

까지 11,895건의 논문을 추출하였다. 그리고 이 논문들에 대해 제목, 초록, 키워드, 게재년도를 크롤링(Crawling)하였다. 추출된 자료를 이용하여 시뮬레이션이 가장 많이 사용되는 분야와 가장 높은 연구 증가율을 보인 분야를 대상으로 대분류와 중분류 체계에서 각각 분석하고 시계열적 동향을 파악하였다.

2.2 토픽분석

2.1에서 최다 사용분야와 가장 높은 증가율의 연구분야를 선정한 후, 그 분야들에 대해 R(버전3.3.0)을 활용하여 토픽분석을 실시하였다. 먼저 영문 초록을 대상으로 데이터의 전처리를 시행하여 단어, 구, 절에 해당하는 내용을 형태소 분리하고 제외어(Stopwords)를 제거하여 단어-문서 행렬 (Term-document matrix)를 만들었다. 단어-문서 행렬의 행 합계를 구해 각 단어가 쓰인 횟수 즉, 사용빈도를 구하고 단어를 사용빈도가 많은 것부터 순서대로 정렬시켜 가장 많이 쓰이는 단어를 추출하였다. 최빈 단어로 선택하여 문서-단어 행렬 (Document-term matrix)를 생성하였다.

토픽분석은 잠재 디리클레 할당(Latent dirichlet allocation, LDA) 방법론을 활용하였다(Blei, 2003). LDA는 주어진 문서에 대하여 각 문서에 어떤 토픽들이 존재하는지에 대한 확률 모형이다. LDA는 어떤 문서에 대해 토픽 벡터인 파라미터 θ가 있는데, 앞에서부터 단어를 하나씩 채울 때마다 θ로부터 하나의 토픽을 선택하고, 다시 그 토픽으로부터 단어를 선택하는 방식으로 문서 생성 과정을 모델링한다. 두 개의 문서가 있을 때, 토픽은 비슷하더라도 각 문서에 등장하는 단어의 종류나 빈도는 다를 수 있기 때문에 키워드 기반의 모델에서는 유사도를 계산하거나 토픽을 분류하는데 한계가 있다. 그러나 이미 보유한 많은 텍스트를 기초하여 말뭉치 단위의 2차원 매트릭스 α와 β를 구하면 α와 β는 각 토픽별로 특정 단어가 생성될 확률이 담긴 테이블이 된다. 그리고 해당 문서에서 각 토픽의 가중치 θ를 계산할 수 있다면, 이 θ를 가지고 유사도 계산이나 분류 작업을 훨씬 쉽고 정확하게 해낼 수 있다. 앞에서 설명한 내용을 수식으로 적으면 아래와 같다.

$$p(z_1,\cdots,z_N)=\int p(\theta)\left(\prod_{n=1}^{N}p(z_n\mid\theta)\right)d\theta \quad (1)$$

$$p(w,z)=\int p(\theta)\left(\prod_{n=1}^{N}p(z_n\mid\theta)p(w_n\mid z_n)\right)d\theta \quad (2)$$

첫 번째 식은 문서의 토픽 생성, 두 번째 식은 문서의 토픽과 단어 생성을 나타낸다. 문서 토픽을 나타내는 z는 θ에 대한 조건부 확률이다. 말뭉치 레벨의 α와 β를 아는 상태에서 문서 w가 주어졌을 때, θ와 z에 대한 조건부 확률을 아래와 같이 나타낼 수 있다.

$$p(\theta,z\mid w,\alpha,\beta)=\frac{p(\theta,z,w\mid\alpha,\beta)}{p(w\mid\alpha,\beta)} \quad (3)$$

이어서 위의 계산이 가능하도록 모형을 단순화하면 아래와 같다.

$$q(\theta,z\mid\gamma,\phi)=q(\theta\mid\gamma)\prod_{n=1}^{N}q(z_n\mid\phi_n) \quad (4)$$

원래의 분포와 변이분포(Variational distribution)사이에서 쿨백-라이블러 발산 (Kullback-Leibler divergence, KLD)을 최소화하는 변분 매개 변수 (Variational parameter)인 γ와 φ를 찾아낸 뒤, 다시 이 값과 관찰된 문서를 사용해서 α와 β를 추정한다. γ와 φ를 구하는 과정을 E-단계, α와 β를 구하는 과정을 M-단계로 놓고, EM 알고리즘(Expectation–maximization algorithm)을 쓴다. 이런 식으로 α와 β, 그리고 개별 문서에 대해 γ와 φ를 구할 수 있으므로 LDA는 주어진 문서에 대해 토픽이 각각 어떤 비율로 섞여 있는지를 분석할 수 있고, 토픽별로 단어의 비율도 알 수가 있다. 아울러 이 토픽이 어떤 글에서 나왔는지, 단어가 어떤 토픽에서 나왔는지도 알 수 있는 것이다. 본 연구에서는 동시출현 단어 분석과 토픽 모델링 결과를 비교하여 실제 발생한 단어관계를 바탕으로 잠재된 토픽을 추출함으로써 학술 DB에서 제공하는 자료로 알 수 없었던 시사점을 찾고자 하였다.

3. 연구결과

3.1 학술 DB제공 연구 분류별 분석

본 절에서는 시뮬레이션이 어떤 분야에서 주로 사용되는지를 분석한다. WoS에서 제공하는 분류체계로 분석하면 1992년부터 2015년까지 KCI DB에 개재된 논문수는 1,190,384건이고, 이 중에서 공학분야 논문은 247,982건으로 20.8%를 점유하고 있다. 시뮬레이션을 활용한 논문을 찾기 위해 “시뮬레이션”이라는 검색어로 추출된 논문은 Table 1에서와 같이 분류된다. 시뮬레이션을 활용한 논문 중 84.7%가 공학 분야에 분포할 정도로 시뮬레이션 기법의 주 활용분야는 공학이다. 그러나, 사회과학, 경영/경제, 자연과학, 의약학, 인문학, 예술체육, 농수해양 등 다양한 분야에서도 시뮬레이션이 활용되는 것으로 집계되었다.

Table. 1. Classification of simulation studies

Classification	Number	Share
Engineering	10,073	84.7%
Social science	463	3.9%
Business&Economics	443	3.7%
Natural science	313	2.6%
Medicine/Pharmacy	270	2.3%
Humanities	209	1.8%
Art&Sports	94	0.8%
Fisheries	30	0.3%
Total	11,895	100.0%

시뮬레이션이 가장 많이 활용되는 공학을 다시 하위체계로 분류하면 Table 2와 같이 기타 공학 분야(Engineer)가 7,603건으로 75.48%의 점유율을 보이고, 컴퓨터공학, 과학기술, 건축학, 물류/운송 등에서 사용됨을 알 수 있다. 시뮬레이션이 가장 많이 활용되는 공학 분야에서 구체적으로 어떤 연구토픽에 시뮬레이션이 활용되는지를 알기 위해서는 추가적인 분석이 필요하다. 본 연구에서는 세부 연구 분야를 분석하기 위해 4절에서 토픽 분석을 실시한다.

Table. 2. Subclass of engineering

Classification	Number	Share
Engineering	7,603	75.48%
Computer science	1,023	10.16%
Science&technology	368	3.65%
Architecture	288	2.86%
Transportation	242	2.40%
Construction&building	238	2.36%
Agriculture	116	1.15%
Materials science	66	0.66%
Oceanography	59	0.59%
Instruments&Instrumentation	25	0.25%
Telecommunications	18	0.18%
Environment&ecology	7	0.07%
Film, radio&television	6	0.06%
Area studies	6	0.06%
Polymer science	4	0.04%
Metallurgy	4	0.04%
Total	10,073	100%

3.2 연도별 분석

토픽분석 이전에 미래 유망 연구분야에 대한 예측을 위해 3.1의 분류별 분석결과에 대한 연도별 추이 분석 시행한다. KCI DB (www.kci.go.kr)에 등재된 논문 중 처음 시뮬레이션이 언급된 논문은 산업경영에 등재된 “Simulation을 利用한 G.T. Layout의 效果에 관한 硏究” 이지만(Won and Won, 1989), 또 다른 국내 논문 DB인 DBPia (www.dbpia.co.kr)에 의하면 1974년 경영경제에 "경제적 설비관리 연구"에서도 이 시뮬레이션을 사용하였다(Lee, 1974). 이는 국내에서 시뮬레이션이 해외와 유사하게 의사결정을 지원하는 경영과학 기법으로부터 시작했다는 것을 보여준다. 이렇게 시작된 시뮬레이션 논문은 현재 KCI DB에 의하면 전 연구 분야에 걸쳐 매년 3,000편 이상의 논문이 발행되고 있음이 관찰된다. 본 절에서는 어떤 연구 분야에서 시뮬레이션 기법의 활용이 빠르게 성장하고 있는지 분석하기 위해 연구 분야별 시계열적 동향을 알아본다. 시뮬레이션을 활용한 대분류별 연구추이를 Table 3a가 보여주는 데 시뮬레이션을 사용한 논문이 공학 분야 학술지에 가장 많이 그리고 높은 증가를 보이고 있으며, 이외 분야에서는 사회과학분야로 분류된 학술지의 논문수가 가장 빠르게 증가한다. 특히 Table 3b에서 보이는 바와 같이 사회과학 분야에서는 교육 및 행정 분야의 연구가 가장 활발함을 알 수 있다.

Table. 3. Chronological trends

(a) Trends of All Field

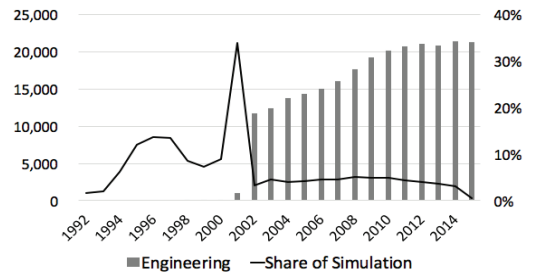
	Engineering	Social Science	Business&Economics	Natural Science	Medicine/Pharmacy	Humanities	Art&Sports	Fisheries
1993	1	0	0	0	0	0	0	0
1995	2	0	0	0	0	0	0	0
1997	13	0	0	0	0	0	0	0
1999	22	1	0	0	0	0	0	0
2001	23	1	1	0	0	0	0	0
2003	739	14	28	16	15	9	2	6
2005	1,107	38	42	28	30	24	14	2
2007	1,283	47	68	41	22	26	6	4
2009	1,658	75	80	55	42	20	13	2
2011	1,948	90	81	51	44	40	20	3
2013	1,762	110	76	61	66	44	18	2
2015	1,407	85	63	54	49	42	18	10
Total	9,965	461	439	306	268	205	91	29

(b) Trends in social science

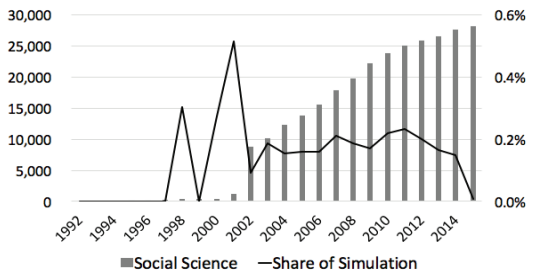
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
1999	1													
2000														
2001								1						
2002	3	2				1								
2003	2	4				1		1						
2004	8	3	4		1		1	2						
2005	8	7	3				1							
2006	9	6	3	1		2			1					
2007	8	6	8	1			1				1			
2008	16	10	5		2	2	1					1	1	
2009	14	8	10	1		1	3							
2010	12	8	9	1	2	2	2					1	1	
2011	16	14	11	3	4	2			2					
2012	7	30	14	3	2	1	1							
2013	10	28	2	4	4	1				1	1			1
2014	21	14	3	2	1		1		1	1				
2015	13	8	5	6	2	2		1		1	1	1	1	
2016	1				1									
Total	149	148	77	22	19	15	11	5	4	3	3	3	3	1

(1) Education & Educational Research
(2) Public Administration
(3) Social Sciences
(4) Government & Law
(5) Behavioral Sciences
(6) International Relations
(7) Social Work
(8) Philosophy
(9) Religion
(10) Theater
(11) Sociology
(12) Linguistics
(13) Women's Studies
(14) Information Science & Library Science

2013년 이후 시뮬레이션 관련 논문수가 감소하는 경향이 있는데, 이런 현상이 국문 논문수의 감소에 의한 것이지 확인하기 위해 대분야별 논문수에서 시뮬레이션을 활용한 논문수가 차지하는 점유율 추이를 Figure 2와 같이 확인하였다. 학진등재(후보)지에 실린 논문의 수는 80년대 100편이하였던 것이 1998년이 넘어가면서 1,000편이 넘고, 2002년에는 전년 대비 8배가 넘게 증가하게 된다. 특히 공학 분야의 논문 수는 2002년에 10배가 넘게 증가한다. 그러므로 2002년 이전은 논문수가 너무 적어 점유율이 무의미할 수 있다. 충분한 학술지의 수가 포함된 2002년부터 2015년까지의 논문 비중을 비교하면 2015년을 제외한 기간에서 시뮬레이션 점유율은 비슷하다. Figure 2a는 공학 분야에서 2010년 이후에 시뮬레이션 관련 논문수가 하락한 이유가 시뮬레이션 분야의 연구가 줄어들었다기보다는 국문으로 작성된 공학논문수가 줄어든 영향으로 보인다. Figure 2b에서 보면 사회과학분야에서도 시뮬레이션을 활용한 논문 비중이 변화하지 않음에도 불구하고 사회과학 계열 논문수가 증가하고 있는 것은 국내에서 사회과학 쪽 국문 논문수가 증가하고 있기 때문임을 보여준다.



(a) Share of simulation in engineering



(b) Share of simulation in social science

Fig. 2. Trends of study frequency depended on simulation

이상과 같이 학술 DB에서 제공하는 서지정보를 통해 시뮬레이션을 활용한 연구논문의 내역분포와 시계열적 동향을 확인할 수 있었다. 하지만 WoS에서 제공하는 분류체계만으로는 실제 산업계와 연관된 연구동향 파악에 한계가 있기 때문에 다음 장에서 토픽분석을 진행한다.

4. 토픽 분석

토픽 모델링은 수집한 논문의 초록을 대상으로 하였으며 LDA를 활용하여 논문수가 많은 공학분야는 상위 1,500개, 논문수가 적은 사회과학분야는 상위 1,000개의

단어를 추출하였다. 반복계산 횟수(num.iterations), 부정확한 내역을 버리는 양(burnin), 메인 토픽과 부토픽의 비율(alpha) 및 하나의 토픽내에서 키워드가 섞여있는 정도(eta)는 데이터량에서 차이가 많은 공학분야와 사회과학분야에 각각 다르게 설정하였다. Table 4은 본 논문에서 LDA를 시행할 때 사용한 설정값이다.

Table. 4. Settings of LDA

Setting	Engineering	Social Science
Numbers of frequent words (word.order)	1,500	1,000
Numbers of topic (K)	30	20
Numbers of iteration (num.iterations)	20,000	5,000
Numbers of discarding (burnin)	2,000	1,000
Numbers of sub topics per a main topic (alpha)	0.0001	0.001
Degree of words spreaded (eta)	0.001	0.01

4.1 공학분야

공학 분야에서 추출한 30개의 토픽과 각 토픽별 모델링에 활용된 주요단어는 Table 5와 같다. 국내에서 시뮬레이션이 가장 많이 사용되는 분야는 데이터 처리기능, 무선통신, 모바일기기, 근거리통신 등의 토픽이 도출된 통신·방송 분야(토픽번호: 1, 3, 7, 8, 11, 18, 19, 21, 23, 25, 29, 30)이다. 다음은 전기 분야 (토픽번호: 4, 5, 14, 24, 27, 28)로 스마트그리드의 연구도 전력망 개선, 전력망에 접속·발전하는 분산전원 개발, 스마트그리드 구성, 기기간 통신 및 전력에너지 운영시스템(Energy management system, EMS) 개발 등 다양한 분야에서 시뮬레이션이 사용되고 있다는 것을 파악할 수 있다. 그 외에 제어계측(6, 10, 12, 17, 24), 교통(9), 안전(20), 유체역학(13), 자동차 (22) 분야 등에서 토픽이 도출되었다. 관련 논문이 많은 통신·방송 분야의 경우 보안과 관련된 25번 토픽이나 인코딩에 관한 29번 토픽처럼 토픽이 세분화되어 도출된다. 15번과 26번 토픽은 국방 분야에 해당되는데 일반적으로 국방이 독립적인 학술 분류 체계로 존재하지 않아 논문이 게재되는 학술지를 분석하는 것으로는 알 수 없지만 시뮬레이션의 중요 활용 분야라는 것을 알 수 있다. 이처럼 토픽 분석을 통해 분류체계 분석에서는 도출할 수 없는 시사점이 도출된다.

Table. 5. Topics of engineering

	Topic	Keywords of Frequent words			
1	data traffic	simulation	time	traffic	data
2	inforgraphic	simulation	image	model	information
3	data lagging	network	mobile	wireless	time
4	energy management	analysis	performance	energy	systems
5	smart grid (electric grid)	power	circuit	current	switching
6	data signal	system	signal	frequency	data
7	mobile device	mobile	method	program	operation
8	network speed	antenna	ghz	bandwidth	microstrip
9	transportation	control	flight	nonlinear	vehicle
10	neural system	fuzzy	nonlinear	neural	parameter
11	data network	network	scheme	routing	protocol
12	sensing	spectrum	interference	sensing	signal
13	fluid control	numerical	flow	wave	fluid
14	renewable energy	inverter	converter	current	switching
15	weapon	simulation	weapon	defense	combat
16	generator	engine	power	energy	efficiency
17	wireless sensor	sensor	nodes	wireless	cluster
18	data transmission	ofdm	communication	signal	multipath
19	short-range com.	rfid	tag	slot	protocol
20	safe equipment	smoke	evacuation	safety	protection
21	antenna	antenna	band	mhz	frequency
22	hybrid car	fuel	vehicle	hybrid	electric
23	network device	routing	node	protocol	network
24	air conditioning and heating	cooling	heating	thermal	radiation
25	security	detection	attack	system	security
26	radar	radar	target	range	geometry
27	piezoelectric	piezoelectric	vibration	frequency	sensor
28	smart grid(DER)	grid	resource	scheduling	heterogeneous
29	incoding	cutting	production	decoder	analysis
30	broadcasting	resonant	station	bandwidth	broadcasting

4.2 사회과학분야

논문 수 증가율이 가장 높은 사회과학분야에서도 Table 6와 같이 세부적으로 토픽 분석을 실시하였다. 공학분야와 비교하여 데이터양이 상대적으로 적은 사회과학분야에서는 20개의 토픽을 추출하였는데 연구 및 조사 분야(1)과 교육(2, 8), 훈련(6) 등 교육 분야와 대출 (5)과 교역(11), 소비 심리(13)의 경영학 분야, 고령화(4), 지리적 접근성(19)과 같은 사회 복지 분야와 관련된 토픽이 많이 도출되었다. 그 외에 무선통신(3), 소음(7), 군사 혹은 정치적 남북문제(9), 정치(12), 주거(14), 문화(16), 법(17), 테스트 알고리즘(18), 예술(20)까지 다양한 분야에서 시뮬레이션이 사용됨을 알 수 있다. 국방분야와 무선통신은 공학과 사회과학 분야 모두에서 토픽이 도출되었다. 이와 같은 정보는 학술 DB에서 제공하는 자료만으로는 취득할 수 없는 내용이기에 텍스트 마이닝을 통한 토

픽분석의 연구에 대한 가치를 알 수 있다. 국제적으로 시뮬레이션의 활용 목적은 단순히 시스템 분석과 검증 등의 의사결정을 지원하기 위한 것이 아니라 교육 및 오락 분야에 까지 확대되고 있다(Nance and Sargent, 2002). 이런 시뮬레이션의 확장된 활용 분야는 공학 분야보다는 사회과학 분야에서 도출됨을 알 수 있다.

Table. 6. Topics of social sciences

	Topic	Keywords of Frequent words			
1	education (after school)	students	learning	education	group
2	wireless sensor	mobile	wireless	sensor	computing
3	elderly society	pension	income	elderly	retirement
4	loan	rate	mortgage	reverse	loan
5	acting training	acting	examination	cognitive	training
6	microphone	hearing	noise	microphone	sound
7	education (in school)	learning	school	elementary	grade
8	military (north Korea)	north	korea	nuclear	terrorism
9	game	game	play	genre	contents
10	world trade	fta	war	trade	international
11	politics	election	seats	political	regional
12	consumption mentality	consumers	attribute	consumer	market
13	residence	biotope	apartment	improvement	urban
14	psychology	interpersonal	selfesteem	empathy	skill
15	culture	culture	article	relation	contemporary
16	law	justice	legal	law	reform
17	testing algorithm	algorithms	multiple	explore	students
18	outdoor access	signboards	outdoor	recognition	location
19	art genre	genre	representation	art	simulation
20	researching method	simulation	model	method	research

4.3 토픽분석의 기간별 세분화

미래 유망 연구 분야를 추론하기 위해, 시기별 토픽 분석을 실시하였다. 즉, 대상 논문을 발행년도 기준으로 전반부(1995~2005)와 후반부(2006 ~2015)로 세분화하고, 비교 효과를 높이기 위해 토픽 수는 10개로 축소하여 토픽 분석을 수행하였다. 먼저, 공학 분야의 논문들을 이와 같은 방식으로 분석하면 Table 7과 같이 전기 분야의 연구 중심이 전력망에서 스마트 그리드로 변화하고, 통신 분야는 토픽 “antenna”가 별도로 분류될 정도로 무선통신 연구가 증가하였음을 알 수 있다.

Table. 7. Topic modeling in engineering

	1995～2005	2006～2015
1	system simulation	system test
2	electronic circuit	broadcasting
3	data network	wireless network
4	data transmission	air conditioning
5	neural&robot	mobile service
6	power grid	smart grid
7	network speed	wireless data
8	data traffic	data traffic
9	transportation	transportation
10	short-range communication	antenna

사회과학 분야에서의 기간별 토픽 분석을 실시한 결과 Table 8과 같이 과거보다 행정(Politics, Public administration) 및 보안(Security and safety)와 교육 분야에 대한 연구가 증가하고 있음을 알 수 있다. 교육 분야는 전반기에 “education”이라는 토픽이 도출되었지만, 후반기에는 “education (in school)” 및 “education (after school)”으로 확장 구분되며 방과 후 수업에 대한 관심이 증가하였음을 알 수 있다. 또한 “military”의 경우 전 시기에 모두 토픽으로 도출되었지만, 후반기의 최빈 단어에서 “terrorism”와 “cyber”가 추가되어 국방 분야의 관심이 전쟁뿐 아니라 상시 대응으로 확대되었음을 보인다.

Table. 8. Topic modeling in social sciences

	1995～2005	2006～2015
1	city planning	residence
2	computer	politics
3	education	education (after school)
4	residence	education (in school)
5	military	military
6	game	game
7	noise	public administration
8	eldery society	elderly society
9	outdoor access	trade
10	public assistance	security and safety

이와 같이 기간별 세분화한 토픽분석은 최근 미래 핵심이슈를 발견하고 트렌드 분석에 도움을 준다.

4.4 주요 학술지 분석

마지막으로 시뮬레이션 관련 논문이 많이 게재된 학술지를 중심으로 어떤 분야에서 시뮬레이션이 많이 활용되는지를 관찰하고 토픽분석 결과와 비교한다. 비교대상은 Table 7의 25개 학술지로 본 연구에서 추출한 총 11,895건의 논문 중 누적 점유율이 50%까지의 학술지이다. 시뮬레이션이 주로 공학 분야에서 활용되므로 학술지는 모두 공학으로 분류되어 있으며 통신·방송 분야 학술지가 12, 16, 18, 23, 24번으로 토픽분석과 동일하게 가장 비중이 높다. 전기분야는 전기, 전자, 통신, 제어의 융합기술인 스마트그리드 연구에 대해 5, 8번과 같은 전기분야 학술지 뿐만 아니라 통신 및 정보처리 관련 학술지에도 등재된다. 교통분야는 25번의 육상교통은 물론 20, 22번의 해상교통 분야에서도 시뮬레이션이 활용되고 있다. 4.1의 토픽분석 결과와 해당 분야의 논문이 등재된 학술지 성격을 비교시 통신, 방송, 전기 등 은 일치성을 확인할 수 있다. 하지만 학술 DB를 통한 분석에서 도출되지 않고 토픽분석에서만 알 수 있는 국방 토픽은 학술지 정보에서도 분석 및 추론이 불가능하다.

Table. 9. Ranking of journals posting the articles utilized simulation

No.	Journal	Num.	Share	Accu.
1	Korea Institute of Information and Communication Engineering	567	5%	5%
2	Korean Institute of Electromagnetic Engineering and Science	421	4%	9%
3	Korea Society for Simulation	354	3%	12%
4	Academia-industrial technology	343	3%	15%
5	The Transactions of Korean Institute of Power Electronics	279	2%	17%
6	The Institute of Electronics Engineers of Korea - Telecommunications	265	2%	19%
7	Korean society for Aeronautical and Space Sciences	248	2%	21%
8	Korean Institute of Illuminating and Electrical Installation Engineers	235	2%	23%
9	Korean Institute of Information Technology	231	2%	25%
10	The Institute of Electronics Engineers of Korea - Semiconductor and Devices	227	2%	27%
11	KISS : Information Networking	203	2%	29%
12	Korea Communications and Networks	195	2%	31%
13	Architectural Institute of Korea Planning Design	194	2%	33%
14	Korean Institute of Intelligent Systems	194	2%	35%
15	Korea Society of Computer and Information	188	2%	37%
16	Communications and Networks	185	2%	39%
17	Mechanical Science and Technology	182	2%	41%
18	Institute of Internet, Broadcasting and Communication	181	2%	43%
19	Korea Contents Association	173	1%	44%
20	Advanced Navigation Technology	165	1%	45%
21	The KIPS Transactions : Part C	157	1%	46%
22	Korea Navigation and Port Research	149	1%	47%
23	Korea Multimedia Society	132	1%	48%
24	Korea Institute of Electronic Communication Sciences	129	1%	49%
25	Korea Transportation Research Society	123	1%	50%

5. 결론

본 연구는 국내 시뮬레이션 연구 활동 동향을 파악하기 위해 1992~2015년까지 학술진흥재단등재(후보)학술지에 게재된 11,895건의 논문을 대상으로 토픽모델링과 시계열 분석을 실시하였다. 분석 결과 시뮬레이션은 공학 분야에서 주로 사용되며, 사회과학 분야에서의 연구 결과가 빠르게 증가하고 있다. WoS 분류 체계에서는 공학 내 어떤 분야에서 시뮬레이션이 활발히 연구되고 있는지 분석이 어려웠지만, 토픽 분석한 결과 통신 및 전기 분야에서 주로 사용된다는 것을 발견할 수 있었다. 특히 최근 10년의 연구 결과를 분석하면 그 이전에 비해 스마트 그리드를 중심으로 한 전기 분야와 무선통신 분야의 연구가 활발해졌다. 시뮬레이션을 활용한 논문 수가 빠르게 증가하고 있는 사회과학 분야에 대해 토픽 분석을 실시한 결과 시뮬레이션의 최신 활용 분야인 교육 및 오락에 관한 토픽을 도출하였다. 최근에는 교육, 국제교역 및 안전에 관한 연구가 더 활발해 졌다고 분석되었다. 추가적으로 국방 토픽 등 대량의 데이터 속에 잠재되어 학술 DB 분류에서는 인지가 어려웠던 토픽을 정량적으로 도출할 수 있었다.

향후에는 다른 세부 분야별로 토픽 분석을 실시한다면 분야별 시뮬레이션 연구 동향 분석도 가능할 것으로 기대된다. 또한 국내 논문뿐 만 아니라 뉴스 기사나 해외 논문에 대해 유사한 방식으로 트렌드를 분석하고 본 연구결과와 비교할 수 있을 것이다.

References

Blei, D. M., Ng and A. Y., &Jordan, M. I. (2003). "Latent dirichlet allocation", *Journal of machine Learning research*, 3(1), 993-1022.

Cho G.H., S.Y. Kim and S. Hur (2014) "An analysis of the research methodologies and techniques in the industrial engineering using text mining", *Journal of the Korean Institute of Industrial Engineers*, 40(1), 52-59.
(조근호, 임시영, 허선(2014), "텍스트마이닝을 이용한 산업공학 연구기법의 분석", 대한산업공학회지, 40(1), 52-59).

Goldsman,D., R. E.Nance, and J. R. Wilson (2010) "A brief history of simulation revisited", *In Proceedings of the Winter Simulation Conference* 2010, 567-574.

Hornik, K. and B. Grün (2011) "topicmodels: An R package for fitting topic models" *Journal of Statistical Software* 40(13), 1-30.

KIIE (1992) Industrial Engineering Dictionary, Chunmoongack, Seoul
(대한산업공학회 (1992) 산업공학용어사전, 청문각).

Lee, B.C. (1974) "Economic research for facilities management", *Management and Economics*, 5, 141-159.
(이병찬. (1974) "經濟的 設備管理를 위한 硏究", 경영경제, 5, 141-159).

Nance, R.E., and R.G. Sargent (2002) "Perspectives on the evolution of simulation." *Operations Research*, 50(3), 161-172.

Noh, S.Y. (2006) "Delphi Technique: Predicting future with professional insight", *Journal of National Soil Research*, 53-62.
(노승용. (2006) "델파이 기법 (Delphi Technique): 전문적 통찰로 미래예측하기", 국토 (구 국토정보), 53-62.

Oh, J.S. (2015) "Identifying Research Opportunities in the Convergence of Transportationand ICT Using Text Mining Techniques", *Journal of Transport Research* 22(4), 2015.12, 93-110.
(오준석. (2015) "텍스트마이닝방법을통한국내교통·ICT 융합분야 연구기회 발견", 교통연구 22(4), 2015.12, 93-110).

Song, H.-S. (2012) "Survey of modeling simulation methodologies and classification of dynamic systems", *Journal of Science and Culture*, 9(1), 101-119.
(송해상 (2012) "동적 시스템 분류 및 모델링 시뮬레이션 기법 조사 연구", 과학과 문화, 9(1), 101-119).

Won, Y.W. and Y.D. Won (1989) "A Study on the Effect of GT Layout by Simulation", *The Journal Of Industrial Management*, 12, 83-105
(원유원, 원유동. (1898) "Simulation을 利用한 G.T. Layout의 效果에 관한 硏究", 산업경영, 12, 83-105).

나 상 태 (amfree@gmail.com)

1998 가천대학교 전기공학 학사
2010 헬싱키경제대학 MBA(경영학) 석사
2012~ 서울과학기술대학교 IT정책전문대학원 산업정보시스템전공 박사과정
1998~ 현재 한국전력공사 계통계획처 차장

관심분야 : 스마트 그리드, 전력경제, 데이터마이닝

김 자 희 (jahee@seoultech.ac.kr)

1995 KAIST 전산학과 학사
1997 KAIST 전산학과 석사
2003 KAIST 산업공학과 박사
2004 비엔나 대학교 방문연구원
2005~ 현재 서울과학기술대학교 IT정책전문대학원 교수

관심분야 : 요구공학, 반도체 제조 스케줄, 스마트 그리드

안 주 언 (llckybbang@naver.com)

2015 서울과학기술대학교 글로벌융합산업공학 산업정보시스템 전공 학사
2013~ 현재 서울과학기술대학교 일반대학원 SW분석·설계학과 석사과정

관심분야 : 데이터마이닝, IT컨설팅

정 민 호 (odinhero@naver.com)

2015 대구대학교 전산공학 학사
2015~ 현재 서울과학기술대학교 일반대학원 SW분석·설계학과 석사과정

관심분야 : 메타데이터, 온톨로지, 데이터 품질