JDCS
디지털콘텐츠학회논문지
Journal of Digital Contents Society
Vol. 20, No. 9, pp. 1839-1846, Sep. 2019

구조적 토픽 모델링 기반 스마트 시티 연구 동향 분석

박 한 샘[1] · 김 동 현[2] · 장 성 주[1*]
[1]한국과학기술원 건설 및 환경공학과
[2]공주대학교 컴퓨터 공학과

Research Trend Analysis on Smart City based on Structural Topic Modeling(STM)

Hansaem Park[1] · Dong-Hyun Kim[2] · Seongju Chang[1*]

[1]Dept. of Civil and Environmental Engineering, Korea Advanced Institute of Science and Technology, Daejon 34051, Korea
[2]Dept. of Computer Engineering, Kongju National University, Cheonan, 31080, Korea

[요 약]

본 연구에서는 구조적 토픽 모델링을 활용하여 Scopus에 게재된 스마트 시티 관련 연구논문 12,400들의 데이터들을 수집하여 동향 분석을 수행하였다. 총 15개의 스마트 시티 주요 연구 토픽들은 “Machine Learning”, “Network Performance”, “Waste Disposal”, “Air Quality”, “Energy Management”, “Intelligent Context Recognition”, “Big Data Analytics”, “Cloud Computing”, “IoT & Security”, “Social Media”,“Sustainable Urban Planning”, “Intelligent Traffic System”, “Healthcare”, “GIS”, “Disaster Management”로 나타났으며, 추가적으로 토픽 발현률에 따라 Hot/Cold 토픽으로 분류한 결과로 기계학습 및 IoT와 같은 연구 분야들이 핫토픽으로 분류되었고, 이에 반해 소셜미디어 및 GIS와 같은 연구 분야는 상대적으로 감소추세를 보이므로 콜드토픽으로 분류되었다. 본 연구의 결과를 통해 현재까지의 스마트시티 관련 연구 동향을 파악하고, 향후 연구 및 정책에 대한 방향성을 제시하고자 한다.

[Abstract]

In this paper, the 12,400 datasets of smart city-related research papers published in SCOPUS were collected and analyzed based on Structural Topic Modeling (STM). As a result, 15 topics (“Machine Learning”, “Network Performance”, “Waste Disposal”, “Air Quality”, “Energy Management”, “Intelligent Context Recognition”, “Big Data Analytics”, “Cloud Computing”, “IoT & Security”, “Social Media”, “Sustainable Urban Planning”, “Intelligent Traffic System”, “Healthcare”, “GIS”, “Disaster Management”) were derived. In order for analysis of research trends of each topic, we used the topic proportion of topics to classify hot/cold topics. Research fields such as machine learning and IoT are represented by hot topic. On the other hands, social media and GIS related topics are included in a cold topic. The result of this study is to grasp the current research trends related to smart city and to suggest directions for future researches and policy makings.

색인어 : 데이터 분석, 스마트시티, 구조적 토픽 모델링, 동향 분석, 텍스트 마이닝
Key word : Data analysis, Smart city, Structural Topic Model(STM), Trend analysis, Text mining

http://dx.doi.org/10.9728/dcs.2019.20.9.1839

Received 30 July 2019; **Revised** 10 August 2019
Accepted 20 September 2019

***Corresponding Author; Seongju Chang**

Tel: +82-42-350-5667
E-mail: saem@kaist.ac.kr

Ⅰ. 서 론

도시는 시민들의 정치·경제·문화 활동을 할 수 있는 공간이며, 이에 따라 인구 집약적 특성을 가지고 있다. 이러한 특성으로 인해 도시는 교통혼잡, 환경오염, 지속가능성, 에너지부족 등 다양한 문제들을 가지고 있으며 점진적으로 더욱 복잡해지고 있다. 또한, 기존 도시 사회문제를 해결하기 위한 수단으로 투입 자원을 증대시키는 전통적인 방법은 한계를 보이고 있다[1].

4차 산업혁명 개념의 출현 이후 다양한 산업 분야에서는 현실 문제를 보다 지능적·효율적으로 해결하기 위해 정보통신기술(ICT)과의 융·복합 기술 개발을 추진 중에 있으며, 이러한 추세에 따라 기존 도시 개념 또한 지속 가능한 혹은 스마트 도시(Smart City)로써의 도약을 준비하고 있다[1, 3, 4].

<그림 1>에서 보는 바와 같이, Research and Market에서 분석한 스마트 도시의 시장 규모에 따르면 2020년까지 14조달러에 육박할 정도로 스마트 도시에 대한 관심이 전 세계적으로 고조되고 있으며, 각 나라별 실정에 따라 다양하게 정의되고 있다[2]. 국제전기통신연합(ITU)에서 정의한 스마트 도시의 개념은 "시민의 삶의 질, 도시운영 및 서비스 효율성, 경쟁력을 향상시키기 위해 ICT 등의 신기술을 활용한 혁신적인 도시"로 축약시켜 정의하였다. 하지만 스마트 도시가 갖는 사회적 요구와 현실적인 문제를 해결하기 위한 기술 및 시민들이 직접적으로 체감할 수 있는 서비스 적용 가능성에 대한 실질적인 연구가 필요함에도 불구하고 각 국은 현재 스마트 시티 대한 사전적 개념 정의에 초점을 맞추고 있는 실정이다. 이에 따라, 본 논문에서는 스마트 도시와 연관된 융·복합 기술들의 관한 최근 연구 동향을 분석하는데 목적을 두며, 이를 위해 Scopus 데이터베이스로부터 "Smart City" 키워드를 통해 검색된 2010년부터 2018년까지 학술문헌의 다양한 정보(제목, 키워드, 초록, 저자 등)을 수집하여 토픽 모델링 분석 방법 중 하나인 구조적 토픽 모델링(STM; Structural Topic Modeling)을 기반으로 스마트 시티 관련 연구 토픽들 도출하고 각 연구 토픽들의 향후 동향에 관해 분석하였다. 이를 통해, 학술적 관점에서는 스마트 도시 구축에 대한 향후 연구 방향 수립에 기여할 것으로 기대된다.

Ⅱ. 토픽 모델링

토픽 모델링은 텍스트 마이닝의 한 분야로 비구조화(Unstructured) 혹은 반구조화(Semi-unstructured) 된 대량의 텍스트 데이터로부터 숨겨진 잠재 구조들을 추출해내는 비지도 학습 방법 중 하나이며[5, 7], 잠재디리클레할당(LDA; Latent Dirichlet Allocation) 방법은 토픽 모델링 방법 중 가장 널리 알려져 있는 방법으로 동향 분석 뿐만 아니라 관광·문헌정보학·소셜미디어 및 뉴스 등 다양한 분야에서 가치있는 정보를 추출하기 위해 활용되어지고 있다[5, 6, 7, 8]. LDA에서 문서는 잠재

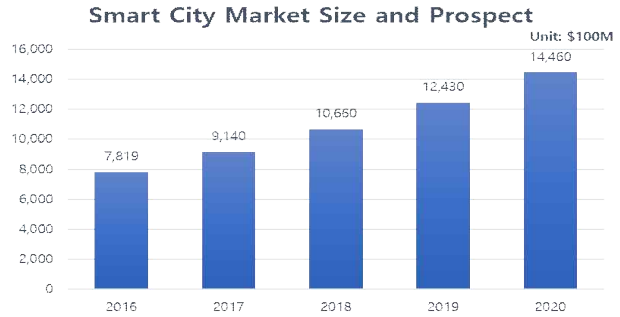

그림 1. 스마트시티 시장 현황[2]
Fig. 1. Smart City Market Size[2]

적인 토픽들의 집합으로 모델링되고 주어진 문서 내에 다양한 확률 가진 단어들의 집합으로 구성된다. 하지만 기본적인 LDA 방법은 문서에 대한 메타데이터(예: 작성 시간, 저자 정보, 종류 등)를 활용하지 않고 각 문서안에 존재하는 단어들의 빈도수를 기반으로 토픽들을 추출하기 때문에 부정확한 결과를 도출할 수 있는 단점이 존재한다. 이러한 제한점을 해결하기 위해, Robert ME는 STM 방법을 제안하였다. STM은 디리클레 할당 방식을 사용하는 LDA의 확장된 프레임워크로써 문서 안에 존재하는 단어의 빈도수 뿐만 아니라 문서에 메타데이터를 활용하여 토픽을 구성하는 단어들의 분포를 결정한다. 이에 따라, STM은 메타 데이터와 문서 내 존재하는 토픽들의 상관관계에 대해 추정할 수 있고, 각 토픽들의 관계를 구분하여 해석할 수 있는 장점을 가지고 있다[8, 9].

<그림 2>는 LDA와 STM의 토픽 추출 과정에 대한 차이를 나타내는 다이어그램이다. 각 노드들은 데이터 생성 과정에서 활용되는 파라미터 값을 표현한다. 음영처리 되지 않는 노드와 음영 처리된 노드는 각각 잠재 변수(Latent variable)와 관측 변

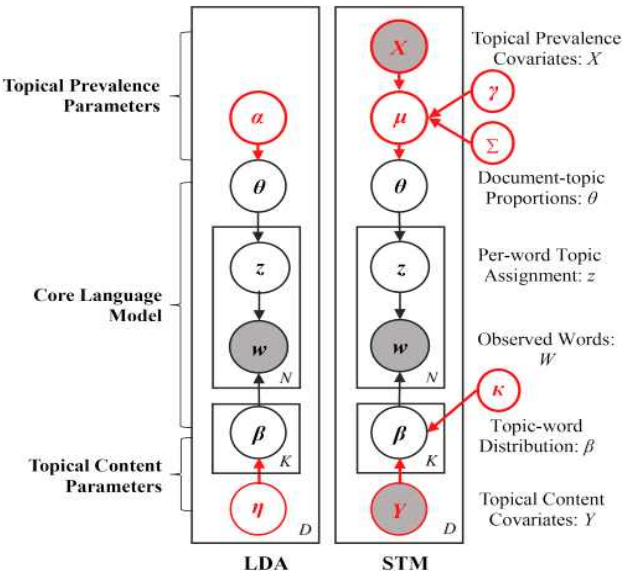

그림 2. LDA와 STM 비교 플레이트 다이어그램[6]
Fig. 2. Comparative Plate Diagram of LDA and STM[6]

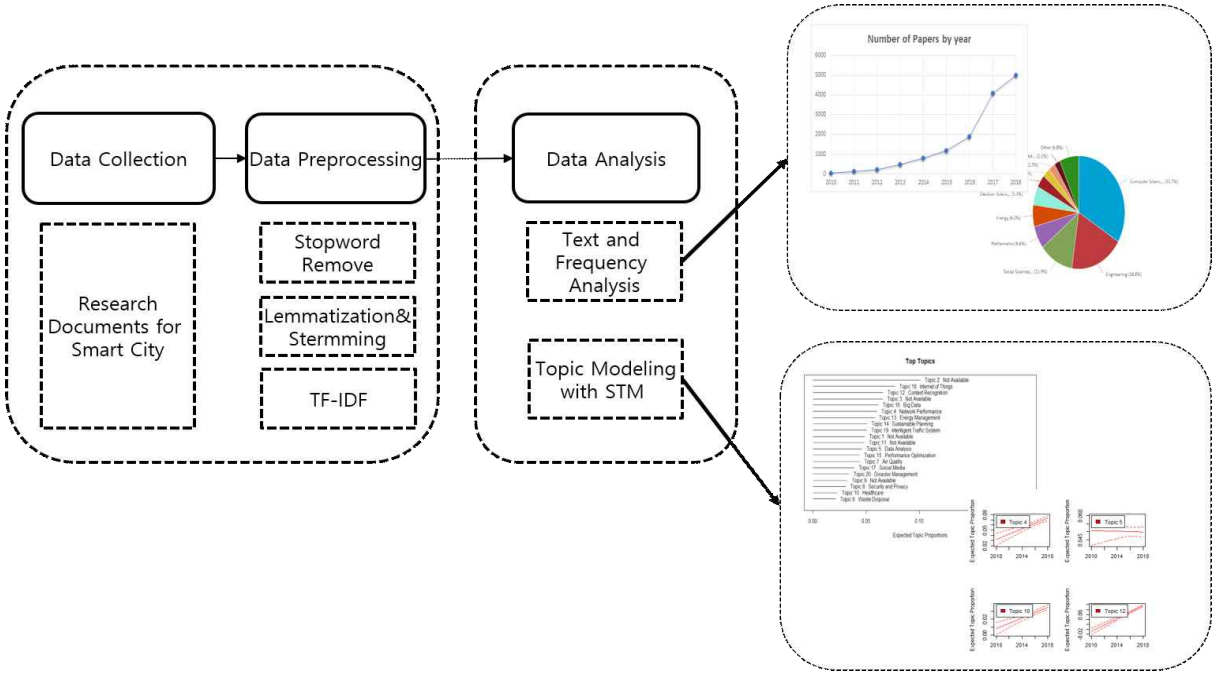

그림 3. 개괄적 연구 프레임워크
Fig. 3. Overall Research Framework

수(Observable variable)를 나타낸다. N은 문서내에 있는 단어들을 나타내며 K는 사용자가 정의한 토픽의 개수를 나타낸다. D는 전체 문서들의 집합을 나타낸다. 추가적으로 각 파라미터에 대한 설명은 아래와 같다.

α : 주제들에 대한 분포
β : 단어들에 대한 분포
θ : 문서에서의 각 토픽별 분포
Z: 해당 토픽에 대한 각 단어 확률
W: 해당 토픽에서의 실제 관측 단어

위 그림에서 보는 바와 같이, LDA와 STM 모형은 상당히 유사한 면을 보이지만 두 토픽 모델의 가장 큰 차이는 Topical Prevalence(Content) 파라미터에서 다른 모습을 보인다. LDA에서는 디리클레 분포를 사용하는 반면 STM에서는 문서와 문서의 메타데이터 사이의 공변량에 대한 선형 모델을 활용하여 각 토픽 및 토픽들의 단어들의 분포를 결정하는 모수로써 사용한다. 또한, STM은 기댓값-최대화 알고리즘(EM: Expectation-Maximization algorithm)을 기반으로 모수를 추정한다. EM 알고리즘은 잠재 변수에 의존되는 확률론적 모델에서 모수들의 최대우도추정(MLE: Maximum Likelihood Estimate) 값을 찾는 방법으로 다양한 확률론적 모델을 풀기 위해 널리 사용되어지는 알고리즘이며, E와 M 단계로 나뉘는데 E는 잠재 변수의 기대치를 계산하고, M에서는 주어진 데이터와 기대치가 포함된 잠재 변수를 이용하여 모수들의 최대우도 추정치를 산출한다[5, 6, 7].

토픽 모델링에서 중요하게 고려되어져야 하는 부분은 사용자가 직접 토픽의 개수를 결정 해줘야 하는 것이다. 앞서 언급했듯이, 토픽 모델링은 비지도 학습을 통해 토픽들을 추출하고 군집화하는데 다른 비지도 학습과 다르게 훈련 및 시험 데이터 셋이 존재하지 않아 모형의 성능을 비교평가하기 어려운 면이 있어 적절한 토픽의 개수를 산출하기 것이 중요하다[9, 10]. 이를 위해, 최대우도 추정치(Held-out likelihood)와 잔차(Residual)을 활용하여 토픽의 개수를 최적화 시키는 것이 일반적이다. 하지만 토픽의 개수가 점점 증가함에 따라 추정치와 잔차의 성능은 개선되는 반면 토픽들의 의미론적 일관성을 나타내는 평가지표(Semantic coherence)가 낮아지는 문제가 발생한다. 이와같은 이유로 인해 토픽의 개수 및 각 토픽들이 갖는 의미들에 대한 부분들은 사용자가 직접 판별하여 선택해야 한다[11, 12, 13].

본 논문에서는 스마트 시티와 연관된 연구 논문들의 초록들과 메타데이터(출판 년월, 문서 타입, 인용횟수 등)을 수집하여 스마트 시티 관련 연구 토픽들을 추출하고 각 연구토픽들의 출현 분포를 활용하여 향후 어떤 연구에 대한 관심이 증가하거나 감소하는 연구 동향 분석을 수행하고자 한다.

Ⅲ. 연구 방법

3-1 연구 프레임워크(Research Framework)

본 논문은 스마트 시티 연구에 대한 학술문헌 데이터를 기반으로 하여 다양하게 변화하고 있는 연구 주제 파악 및 동향을 분석하기 위한 연구 절차를 수립하였으며 <그림 3>과 같다.

연구 문헌 데이터 수집은 스마트 시티 관련 연구 주제에 대

표 1. 수집 데이터 현황
Table 1. Description of Collected Dataset

Date	2010 ~ 2018
Taerget	Scopus Database (https://www.scopus.com/search/form.url?display=basic)
Keyword	Smart city
Document Type	Article(Journal), Conference Paper
Feature(s)	Title, Keyword, Abstract, Year, Citation, Etc,.
Total	12,400

한 분석을 위해 Scopus 데이터베이스로부터 연관 키워드를 활용하여 데이터를 수집하였다. 2010년부터 2018년까지의 총 9년간 게재된 총 12,400편의 학술문헌들을 수집하였지만 연구 주제 도출을 위해 다양한 문헌 종류들 중 논문(Article, Conference Paper) 데이터를 제외하였다. 이에 따라, 총 10,955편을 대상으로 토픽모델링을 적용하여 분석하였다.

텍스트 전처리(Text preprocessing)는 자연어처리에서 가장 중요한 부분이라고 할 수 있을 정도로 분석 결과의 신뢰도에 큰 영향을 끼친다. 텍스트 전처리 과정은 일반적으로 토큰화, 불용어 처리, 어간 및 표제어 추출 등으로 나뉘게 되며 수집된 데이터 중 문헌 초록에 각 단계별 전처리 과정을 수행하였다.

- 토큰화(Tokenization) : 토큰화는 주어진 말뭉치(Corpus)에서 토큰이라 불리는 단위(단어, 단어구 등)로 분리하는 전처리 과정을 말하며 이는 텍스트 데이터에 존재하는 구두점(Puctuation)과 같은 문자를 제외시키기 위해 활용된다. 구두점이란 온점, 컴마, 물음표 등과 같은 기호를 지칭하며 이러한 기호들은 텍스트 분석에 무의미하기 때문에 제거해야 한다.
- 불용어처리(Stopword Remove) : 구두점과 마찬가지로 텍스트 데이터 에 존재하는 의미없는 단어들을 제거하는 방법을 불용어처리라고 하며, 문장 내에 자주 출현하지만 분석하는데 불필요한 단어들을 지칭한다. 대표적인 불용어로써는 관사(a, an, the)가 존재하며 이외에도 사용자가 분석 목적에 맞게 추가적으로 정의하여 제거 할 수도 있다.
- 어간(Stemming) 및 표제어(Lemmatization) 추출 : 이 전처리 과정은 다양한 형태의 단어들을 그 뿌리 단어로 변환하는 과정을 뜻한다. 예를 들어, “compute”, “computation”,“computing”이란 서로 다른 형태의 단어이지만 같은 뜻을 지니고 있다면 이 형태를 단일 형태 “compute”로 치환하는 과정이다. 이 과정을 통해 처리되어지는 단어의 개수를 줄일 수 있으므로 연산 속도를 증가시킬 수 있다.

또한, 논문 초록 텍스트 데이터들에 대해 TF-IDF를 산출하여 기간별에 따른 핵심 단어들을 도출하여 워드클라우드 혹은 네트워크를 통해 표현하였다. 마지막으로는 STM을 기반으로 과거부터 현재까지의 연구 토픽들을 도출하여 각 토픽들이 가지고 있는 출현 분포를 활용하여 연도별 연구 동향을 분석하였다.

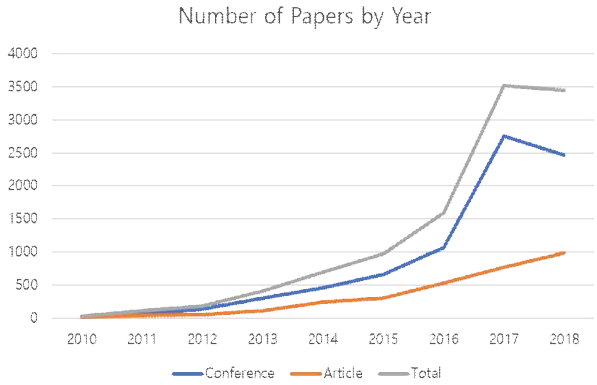

그림 4. 연도별 스마트 시티 관련 연구 문헌의 수
Fig. 4. Number of Research Papers by Years

Ⅳ. 연구 결과

4-1 빈도 분석 및 텍스트 분석

본 연구에서는 각 문헌 초록 정보들에 대한 토픽 모델링을 도출하기 전 문헌 정보에 기반을 둔 빈도 분석을 수행하였다. <그림 4>에서 보는 바와 같이 스마트시티와 관련된 연구 분야는 2010년 이전까지 아주 미미하게 진행되어왔지만 빅데이터, IoT와 같은 ICT 기반 핵심 기술들에 대한 패러다임의 출현 이후인 2016년 이후부터 가파른 증가 추세를 보이는 것을 알 수 있었다. 추가적으로 <그림 5>와 같은 경우에는 국가별 스마트시티 관련 연구 현황 분석을 위해 2010년부터 2018까지 전체 문헌에 대해 조사한 결과를 보여준다. 이를 통해, 중국이 스마트 시티 연구에 가장 활발한 것을 알 수 가 있었다. 이같은 결과는 중국이 현재 겪고 있는 심각한 도시 문제(공기질 오염, 에너지 부족 등)들에 대한 심각성을 인식하고 해결하고자 꾸준히

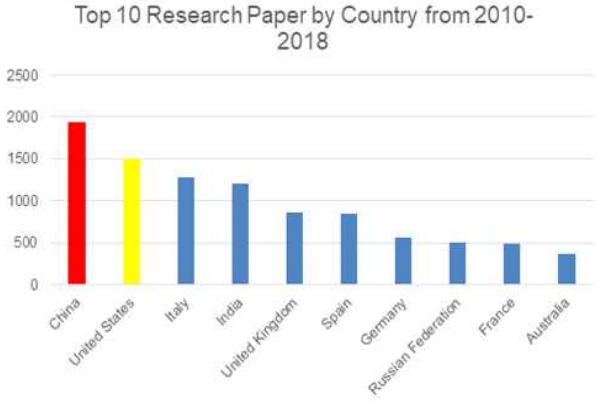

그림 5. 국가별 연구 문헌 출판 현황
Fig. 5. Number of Research Papers by Country

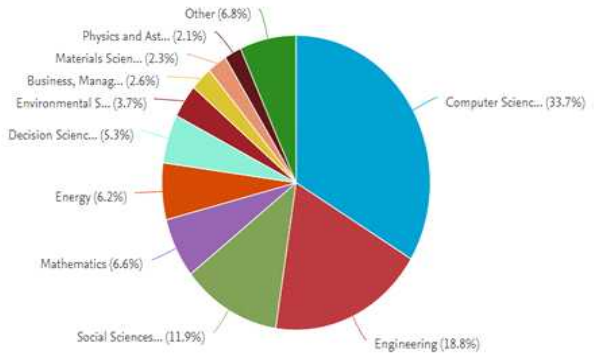

그림 6. 분야별 연구 문헌 출판 현황
Fig. 6. Number of Research Papers by Fields

노력 중인 것을 알 수 있었다. 또한, <그림 6>는 스마트시티 연구에 대한 각 분야별 현황을 나타낸다. 다양한 학문 분야에서 스마트시티 관련 연구를 진행해오고 있지만 그 중 가장 많은 비율을 차지하는 분야는 바로 컴퓨터 공학 분야이다. 스마트 시티 구축을 위한 노력으로 기존 도시 문제를 ICT 패러다임과의 융복합을 통해 해결하고자 하는 연구가 활발히 이루어지고 있다는 사실을 간접적으로 알 수 있었다.

또한, 각 연구 문헌에서 나타나는 단어 중 빈도가 높은 단어일수록 연구 주제와 아주 밀접한 관계를 가지고 있을 가능성이 높기 때문에 기간별 텍스트 분석을 통해 어떠한 단어들이 새롭게 변화했는지 추출하였다. 기간은 4차 산업혁명의 개념이 나온 이전과 이후 즉 2016년을 기점으로 분리하였다. 수집된 데이터 셋은 2010년부터 2018년까지이기 때문에 2010년-2015년, 2016년-2018년으로 나누어 기간별 상위 10개의 빈도수가 높은 단어들을 도출하였고 그 결과는 <표 2>와 같다. 분리된 두 기간별 빈도 단어들은 상당히 유사하게 나타났지만, 2016년-2018년의 단어에서 iot가 새롭게 등장한 것을 알 수 있었다. <그림 7>은 2016년~2018년 데이터를 통한 스마트시티에 대한 워드클라우드이다.

표 2. 기간별 빈도수에 따른 상위 10개 단어 비교
Table 2. Comparison of Top 10 Words based on Frequency by Year

Date			2010 ~ 2015		2016 ~ 2018	
words	frequency	1	smart	5037	smart	15057
		2	city	3652	data	10772
		3	data	2741	city	9069
		4	urban	1451	system	6185
		5	system	1417	energy	4328
		6	information	1403	information	4298
		7	energy	1190	urban	4119
		8	development	950	**iot**	**3821**
		9	management	861	network	3569
		10	use	839	time	3526

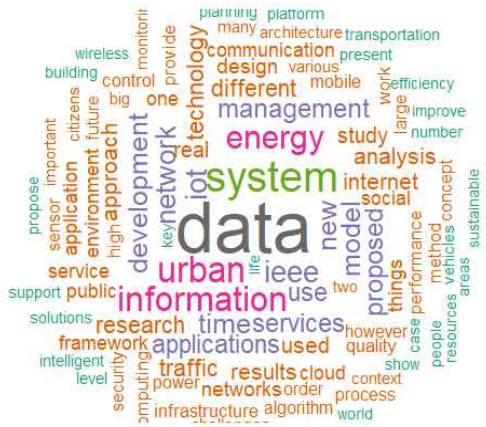

그림 7. 스마트 시티 워드클라우드 (2016~2018)
Fig. 7. Wordcloud for Smart City (2016~2018)

4-2 토픽 모델링

1) 최적 토픽(K) 개수 도출

앞서 언급했듯이, 문서로부터 주요한 토픽들을 추출하기 위한 가장 중요한 단계는 문서 내에 잠재되어 있다고 가정하는 토픽의 개수를 결정하는 것이다. 이 과정은 모든 토픽 모델링 방법에서 필수적으로 거쳐야 하는 과정이며, 본 연구에서 활용되어진 STM에서도 동일하다. 잠재 토픽 개수 설정은 사용자가 직접적으로 설정해야하며, 본 논문에서 최적 토픽 개수를 결정하기 위해 STM 토픽의 개수를 5~20개까지 점진적으로 증가시키면서 추정치와 잔차 그리고 의미론적 일관성이 가장 적합한 모델을 추출하였다. <그림 8>에서 보는 바와 같이 토픽의 개수가 낮을 경우 추정치와 잔차는 높지만 반대적으로 의미론적 일관성을 높은 것을 볼 수 있으며 토픽의 개수가 증가함에따라 감

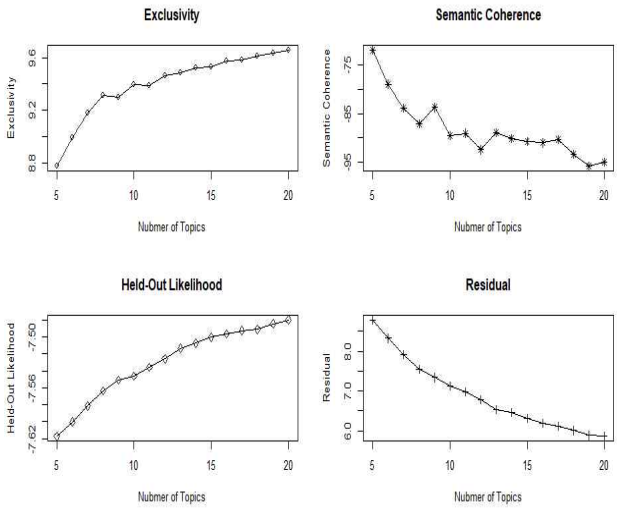

그림 8. 최적 토픽 개수
Fig. 8. Number of Optimal Topics

표 3. STM 기반 토픽 추출 및 키워드
Table 2. STM based Topic Extraction and Keywords

Topic		Major Temrs
1	Machine Learning	Model, Method, Learning, Performance, Prediction, Accuracy, Neural
2	Network Performance	Network, Communication, Wireless, Performance, Routing, Transmission, Node
3	Waste Disposal	Water, Waste, Collection, Temperature, Study, Sludge, Garbage
4	N/A	Technology, Information, Development, Business, System, Reserved, Right
5	Air Quality	Monitoring, Sensor, Quality, Air, Pollution, System, Environmental
6	Energy Management	Energy, Power, Consumption, System, Grid, Building, Demand
7	Intelligent Context Recognition	Algorithm, Detection, Image, Video, Method, Recognition, Surveillance,
8	Big Data Analytics	Data, Big, Information, Processing, Analytics, Web, Real, Time
9	Cloud Computing	Cloud, Service, Computing, Architecture, Platform, Application, Approach
10	IoT & Security	Iot, Internet, Things, Security, Privacy, Access, Authentication
11	Social Media	Mobile, Social, Media, User, Crowdsourcing, Information, Public
12	Sustainable Urban Planning	Urban, Research, Sustainable, Development, Innovation, Social, Sustainability
13	Intelligent Traffic System	Traffic, System, Congestion, Road, Parking, Control, System
14	Healthcare	Human, People, Health, Healthcare, Activity, Medical, Elderly
15	N/A	Project, Development, Design, Article, Concept, Engineering, Implementation
16	GIS	Urban, Planning, Areas, Spatial, GIS, Growth, Development
17	Disaster Management	Infrastructure, Safety, Management, Disaster, Improve, Communication, Information

소하는 것을 볼 수 있다. 본 논문에서는 추정치, 잔차, 의미론적 일관성에 대한 지표를 통해 추정치와 잔차가 극소점에서 가깝고 의미론적 일관성이 최소화 되지 않는 지점을 기준으로 최적 토픽 개수를 결정하였으며, 최종적으로 17개의 토픽 개수를 도출하였다.

2) 토픽 추출 및 가시화

본 연구의 목적은 스마트 시티에 대한 과거부터 현재까지의 연구분야 및 동향을 분석하여 향후 연구방향 수립에 대한 가이드를 제시하고자 하는 것이다. 이에 따라, 앞서 도출한 17개의 최종 토픽의 개수를 설정하고 STM을 적용하였으며, 각 토픽 및 각 토픽에 속하는 단어들을 추출하였다. 하지만 토픽모델링 방법에서는 각 군집화된 단어들의 집합인 토픽들에 대한 의미가 없기 때문에 도출된 토픽들에 대한 현실적 의미에 기반한 라벨을 부여해야하며, 그 결과는 <표 3>과 같다. 또한, STM 기반으로 추출된 17개의 토픽 중 본 연구의 목적을 달성하기 위해 두가지 기준을 만족시키는 적절 수준의 분석 대상 토픽들을 선별하였다. 사용자 주관적 판단을 최소화 하기 위해 두가지 기준에 대한 기준을 세웠다. 본 연구에서 포함하지 않는 토픽을 선별하는 첫 번째 기준은 '토픽의 의미가 연구와 무관한 경우', 두 번째는 "토픽을 표현하는 단어만으로 토픽의 의미가 명확하지 않는 경우"로 설정하였으며, 이에 해당되는 토픽(Topic4, Topic15)들은 본 연구에서 제외하였다.

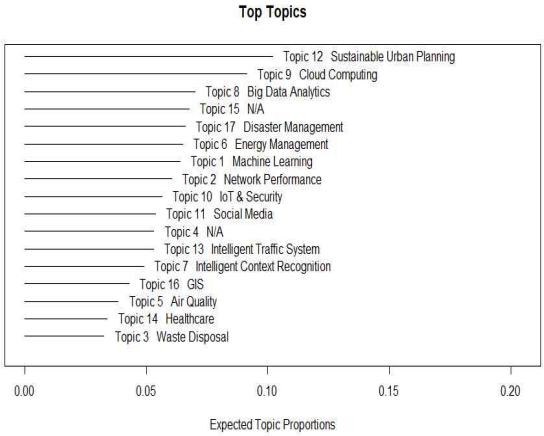

그림 9. 토픽 발현에 따른 상위 토픽
Fig. 9. Top Topics based on Topic Propotions

제외된 토픽들 외에 도출된 상위 토픽들로는 "토픽12", "토픽9", "토픽8", "토픽17", "토픽6"이 있으며, 이는 데이터들을 분석하는 ICT 기반의 빅데이터 분석 방법과 관련된 연구동향이 많은 것을 알 수 있었다. 뿐만아니라, 실내외 환경 요소에 다른 에너지 관리도 상당히 높은 비율로 스마트시티 구축에 관련하여 관심있는 주제라고 생각할 수 있으며, 추가적으로 재난재해에 대한 연구들도 많이 수행되어 온 주제라고 생각할 수 있다.

4-3 토픽들에 대한 연구 동향 분석

마지막으로 본 절에서는 시간이 지남에 따라 스마트시티 관련 토픽들에 대한 동향을 분석하였다. 토픽들에 대한 관심의 증가/감소를 도출하기 위해 토픽에 대한 공변량으로 시간를 설정하여 토픽들을 추정하였다. 이를 통해 각 토픽들 중에서 증가 추세를 나타내는 토픽들을 핫토픽(Hot-Topic)으로 분류하고, 감소 추세를 나타내면 콜드(Cold-Topic)으로 분류하였다. <그림 x>에서 보는 바와 같이, "machine learning", "network performance", "waste disposal", "intelligent context recognition", "IoT & security", "intelligent traffic system", "healthcare" 들은 핫토픽으로 분류되었다. 이는 현재 스마트시티 구축과 관련된 주요 연구들임을 알 수 있으며, 특히 기계학습은 이중 제일 가파른 증가 추세선을 보인다. 이를 통해, 현재 스마트 시티 구축과 관련된 연구들 중 제일 주요 토픽으로 센싱을 통한 데이터 분석과 관련된 연구임을 알 수 있고 향후 이러한 연구들이 더욱 성장할 것으로 추정할 수 있다. 이에 반해, "Social Media",

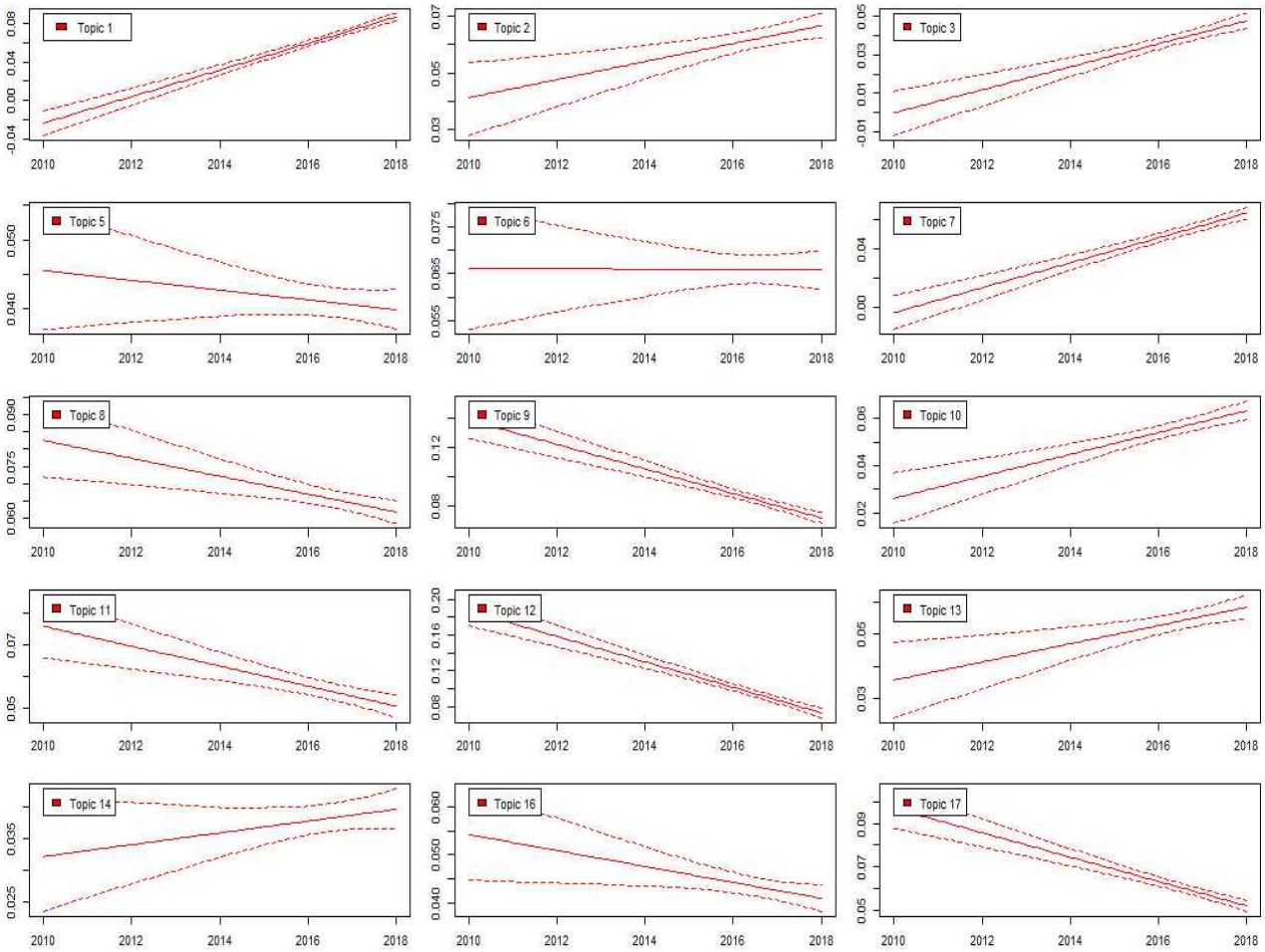

그림 10. 토픽 발현에 따른 토픽별 Hot/Cold 토픽 분류
Fig. 10. Classification between Hot and Cold Topics on Topic Propotions

"GIS"와 같은 경우에는 현재 감소추세인 연구들인 것을 알 수 있으며, 이 분야들은 IoT나 상황인지 그리고 지능형 교통시스템과 같은 연구 주제보다 현재 많이 수행되고 있지 않는 연구들임을 확인하였다.

Ⅴ. 결 론

본 연구는 STM 기반 토픽 모델링을 활용하여 스마트 시티 관련 토픽 도출 및 각 토픽별 동향을 추출하기 위해 Scopus로부터 연구 문헌 데이터를 대상으로 분석하였다. STM 토픽모델링 알고리즘을 이용하여 스마트시트의 세부 연구 토픽들을 도출하였으며, 토픽 발현 확률을 기준으로 연구 동향을 파악하였다. 지능형 교통 시스템, IoT, 기계학습 등 센싱 데이터셋을 기반으로한 데이터 분석 관련 연구 및 이를 적용한 분야의 연구들이 핫토픽들로 분류되었으며, 이런 기술들은 최근 4차산업혁명 이후 다양한 산업에서 활발하게 연구되는 핵심 기술들임을 알 수 있다. 반면, GIS, 소셜미디어 등과 같은 주제들은 상대적으로 핫토픽들에 비해 감소추세를 보이며 콜드토픽들로 분류되었다.

본 연구는 스마트시티 관련 연구 동향 및 관련 정책 수립을 위해 연구 문헌들에 대한 빅데이터 분석을 수행하였으며, 향후 후속 연구를 위한 프레임워크를 도출하였다는 점에서 학문적 기여도가 있으며, 추가적으로 현재 뿐만아니라 향후 스마트 시티 구성을 위한 가이드 및 실무적 시사점을 제시하기 위한 초기의 학술 연구로서 의의가 있을 것으로 판단된다.

감사의 글

본 연구는 국토교통부의 스마트시티 혁신인재육성사업으로 지원되었으며, 2019년도 국토교통부 주거환경연구사업의 연구비 지원(19RERP-B090228-06)에 의해 수행되었습니다.

참고문헌

[1] T. Nam., "Smart city as urban innovation: Focusing on management, policym and context." *The Proceedings of the 5th international conference on theory and practive of electronic governance,* pp. 185-194, Sep 2011.

[2] Research and markets. Smart City market [Internet]. Available: https://www.researchandmarkets.com/.

[3] Preuveneers D., "The intelligent industry of the future: A survey on emerging trends, research challenges and opportunities in industry 4.0", *Journal of Ambient Intelligence and Smart Environment,* Vol. 9, No. 3, pp. 287-298, April 2017.

[4] Sengers, F., "Smart city construction: Towards an analytical framework for smart urban living labs," In *Urban Living Labs*, Routledge, ch. 5, pp. 74-88, May 2018.

[5] Roberts, M.E., "stm: R package for structural topic models," *Journal of Statistical Software,* Vol. 10, No. 2, pp. 1-40, 2014.

[6] Kuhn, K. D., "Using structural topic modeling to identify latent topics and trends in aviation incident reports," *Transportation Research Part C: Emerging Technologies,* Vol. 87, pp. 105-122, Feb 2018.

[7] Hu, N., "What do hotel customers complain about? Text analysis using structural topic model," *Tourism Management,* Vol. 72, pp. 417-426, June 2019.

[8] Bagozzi, B.E., "The politics of scrutiny in human right monitoring: Evidence from structural topic models of US State Department human rights report," *Political Science Research and Methods*, Vol. 6, No. 4, pp. 661-677, Oct 2018.

[9] Roberts, M.E., "Structural topic models for open-ended survey respnses" *American Journal of Political Science*, Vol. 58, No. 4, pp. 1064-1082, Mar 2014.

[10] Moro, S., "Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent dirichlet allocation" *Expert Systems with Application*, Vol. 42, No. 3, pp. 1314-1324, Feb 2015.

[11] Liu, L., "An overview of topic modeling and its current applications in bioinformatics" *Expert Systems with Application*, Vol. 42, No. 3, pp. 1314-1324, Feb 2015.

[12] Krestel, R., "Latent dirichlet allocation for tag recommendation" *In Proceeding of the third ACM conference on Recommender systems*, pp. 61-68, Oct 2009.

[13] Griffiths, T., "Finding scientific topics" *Proceeding of the National Academy of Sciences*, 101(suppl 1), pp. 5228-5235, Apr 2004.

박한샘(Hansaem Park)

2014년~2016년 : 고려대학교 컴퓨터정보학과 대학원 (공학석사)

2016년~현 재: 한국과학기술원 건설환경공학과 박사과정

※관심분야 : 지능형 건물 시스템, 기계학습, 건물 공조 시스템 예측 제어

장성주(Seongju Chang)

1984: 서울대학교 건축학 학사
1986: 서울 대학교 건축학 석사
1999년: 카네기멜런 대학 건축학 박사
2002년~2006년: 한국정보통신대학교 연구교수
2002년- 2003년: MIT Media Lab 객원 연구원
2007년~2012: 한국과학기술원 전문교수
20013년~현 재: 한국과학기술원 초빙교수
2011-현 재: 미래도시 포럼 의장

※관심분야 : 스마트그린 오브젝트/빌딩/시티, 인터랙티브 멀티미디어 인터페이스, 의사결정지원 시스템

김동현(Dong-Hyun Kim)

1986년 : 중앙대학교 전기공학과 (공학사)
2005년 : 공주대학교 대학원 컴퓨터멀티미디어공학과(공학석사)
2010년 : 공주대학교 대학원 컴퓨터공학과(공학박사)
2010년~2015년: 우송대학교 IT경영학부 초빙교수
2016년~현 재: ㈜정보소프트 기술이사

※관심분야 : 인공지능(AI), 지식관리(Knowledge Management), 데이터 분석(Data Analysis)