

HDP를 통해 살펴본 컴퓨터 네트워크 및 통신 연구 동향 분석

권태훈*

*한국과학기술정보연구원

*kth78@kisti.re.kr

Analysis of research trends in computer networks and communications through hierarchical dirichlet process

Taehoon Kwon*

*Korea Institute of Science and Technology Information

요약

본 논문에서는 computer networks and communications분야의 연구 동향을 살펴보기 위해 토픽 모델링 분석을 수행하였다. 이를 위해 2017년 1월 1일부터 2021년 5월 12일까지의 SCOPUS DB중 해당 ASJC(all science journal classification)의 영문 논문을 정제를 거쳐 총 466,756건을 수집하였다. 이 중 출판연도별로 분석을 위한 DB를 정제 및 재구성후 title, abstract을 대상으로 키워드 추출을 수행하였다. 토픽 모델링 분석은 주제와 함께 적절한 주제수를 같이 찾아줄 수 있는 계층적 디리클레 프로세스(hierarchical dirichlet process, HDP) 기법을 활용하였다.

I. 서론

논문은 학문적 의미나 결과의 공표를 위해 저자가 일정한 양식에 따라 작성한 연구 성과물이다. 따라서, 특정분야의 연구 동향을 살펴보기 위해서는 텍스트 마이닝을 활용하여 논문을 살펴보는 것이 많은 도움이 된다 [1]. 특히, 토픽 모델링은 텍스트로부터 주제를 분석하는데 사용되는 기법으로 논문의 연구 동향을 분석하는데 효과적인 텍스트 마이닝 기법중 하나이다[2].

토픽 모델링 기법중 가장 대표적 기법은 잠재 디리클레 할당 (Latent Dirichlet Allocation, LDA)기법이다. 하지만, LDA는 정확한 주제수를 미리 결정하지 않으면 결과가 달라진다. 이를 보완하기 기법 중 하나가 계층적 디리클레 프로세스 (Hierarchical Dirichlet Process, HDP)로서 주제와 함께 적절한 주제수를 같이 찾아주도록 LDA를 개선한 기법이다[3].

본 논문에서는 2017~2021년동안의 SCOPUS DB의 컴퓨터 네트워크 및 통신 분야 저널 및 학회 논문들에 HDP 토픽모델링을 적용하여 최근 5년간의 기술동향을 분석하고자 한다.

II. 본론

SCOPUS DB중 2017년 1월 1일부터 2021년 5월 12일까지의 컴퓨터 네트워크 및 통신 분야 영문 저널 및 학회 논문들을 수집하였다. 수집은 한국과학기술정보연구원(KISTI)의 내부 논문 DB 검색 시스템인 SCOPUS 시스템을 활용하였다[4]. 과학기술분류 코드인 ASJC(all science journal classification)중 computer networks and communications(분류코드 1705)로 분류되는 논문들을 정제를 거쳐 총 466,756건을 수집하였다. 수집 및 정제된 데이터는 출판연도별로 분석을 위한 DB로 재구성하였다.

DB 논문들의 title, abstract을 대상으로 uni-gram과 bi-gram를 고려하여 키워드 추출을 수행하였다. 해당 키워드들을 대상으로 HDP를 활용하여 토픽 모델링을 수행하였다. 가장 비중이 높은 토픽별로 정리한 결과는 다음과 같다.

	주요 키워드
1	service, security, cloud, iot, device, attack, resource

2	communication, channel, scheme, transmission, node, signal, wireless
3	student, learning, development, service, use, social, knowledge
4	feature, classification, dataset, number, accuracy, task, set
5	detection, feature, accuracy, classification, recognition, sensor, human
6	control, energy, load, voltage, controller, current, simulation
7	antenna, frequency, structure, band, low, ghz, measurement
8	water, sensor, temperature, parameter, area, effect, structure
9	jamming, ppm, japan, non, swap, file, engine, ransomware
10	state, color, bug, cipher, teaching, email, word
11	land, use, lbw, pattern, part, ii, identification, work
12	day, nvm, medicine, feature, selection, owp, cta, mal, activity
13	group, lms, content, pre, eclampsia, p4, gif, image, ash
14	round, cca, clock, ulcer, rc5, hdr, image, confidante
15	preimage, attack, stereotype, threat, don, chd, sr8, effect, ethical, board, secretary

III. 결론

본 논문에서는 HDP기법을 활용하여 ASJC가 computer networks and communications로 분류되는 논문들에 대해서 연구동향을 살펴보았다. 분석결과 최근 5년동안 service, security, cloud, IOT, device, attack, resource관련 연구가 가장 많은 것으로 판단되었다.

ACKNOWLEDGMENT

본 논문은 한국과학기술정보연구원(KISTI)의 지원으로 작성되었습니다.

참고 문헌

- [1] 권태훈, “자연어 처리 기법을 통해 살펴본 스마트 팩토리 연구동향,” 2020년도 한국인터넷정보학회 추계학술발표대회, 2020. 10. 31
- [2] 권태훈, “특허성과물 HDP 토픽 모델링 분석을 통한 기술동향 분석,” 2021년도 통신정보 합동학술대회, 2021. 4. 29
- [3] <https://bab2min.tistory.com/>
- [4] <https://scopus.kisti.re.kr/>