

토픽 모델링을 활용한 물류 로봇 연구 동향 분석

박현성*, 최승일*

*국립공주대학교 산업공학과

e-mail:sichoi@kongju.ac.kr

Analyzing of Research Trends in Logistics Robots Using Topic Modeling

Hyeon-Seong Park*, Seungil Choi*

*Dept. of Industrial Engineering, Kongju National University

요약

본 논문에서는 국내 물류 로봇 연구 동향 분석을 위해 LDA 토픽 모델링을 활용하여 핵심 주제가 무엇인지 살펴보았다. 하이퍼 파라미터를 튜닝하기 위해 복잡도와 일관성 지표를 활용하였고, 주제 범주화가 잘되는지 중복은 없는지 확인하였다. 또한, 토픽 모델링 결과로 토픽별 분류된 논문을 확인하여 토픽별 주제를 네이밍 하였다. 그리고 토픽별 분류된 논문 비율을 확인하고, 코로나 전과 후의 토픽 추이를 살펴보았다.

1. 서론

오늘날의 물류 산업은 물류 4.0 기술을 통해 스마트하고 유연한 물류 시스템을 구축하여, 처리 시간과 비용을 줄이고, 서비스의 정확성과 속도를 높여 고객 맞춤형 서비스를 제공한다[1]. 다가오는 물류 5.0 시대에는 모든 작업을 무인 자동화 시스템이 대체하여 인간이 참여하지 않아도 되는 비부가 가치 활동을 지능화하고 자동화할 것이다[2].

저출산 고령화 시대의 도래는 물류 산업에 커다란 도전 과제를 안겨주고 있다. 통계청(2023)에서 집계한 출생 통계 결과에 따르면, 매년 감소 추세를 보이며, 2018년 합계출산율 0.98명을 시작으로 2023년 0.72명으로 합계출산율 1명 미만의 수치가 지속해서 이어지고 있다.

본 연구는 토픽 모델링을 활용하여 국내 물류 로봇 연구 동향에 대해 분석하여 정책 입안자 및 관련 연구자에게 유용한 정보를 제공하고자 한다.

2. 데이터 수집

데이터의 수집은 한국학술지인용색인에서 수행하였다. 검색어로 '물류' AND '로봇'으로 하였으며, 기간은 별도로 설정하지 않았다.

그 결과 2006년부터 ~ 2024년까지 논문 131편이 검색되었고, 2024년은 제외 처리 후 121편의 논문을 대상으로 서지 정

보를 추출하였다. 또한, 물류 로봇과 관련 없는 논문 3편과 저자 주제가 없는 논문 1편을 추가로 제거하여 총 117편을 분석 대상으로 도출하였으며, 연도별 게재 현황 [표 1]과 같다.

[표 1] 연도별 게재 현황

연도	2006	2007	2008	2009	2010	2011	2012	2013	2014
논문	3	1	0	0	0	2	0	4	1
연도	2015	2016	2017	2018	2019	2020	2021	2022	2023
논문	1	3	9	4	10	11	23	20	25

3. 데이터 전처리

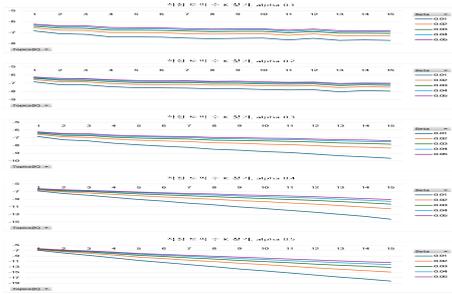
키워드 분석에서는 서로 같은 단어라도 대/소문자, 단/복수 등의 차이로 서로 다른 키워드로 인식하기 때문에 데이터 전처리는 매우 중요한 과정이며 분석의 정확도 및 정밀도를 높이기 위해 정제하려는 키워드 분야의 지식을 일정 수준 이상 필요로 한다. 또한, 분석의 효율을 높이기 위한 과정이며 데이터 특성에 따라 전처리 과정을 생략해도 된다.

한국학술지인용색인에서 DB를 상세내용 포함하여 엑셀 형식인 xls 확장자로 다운로드가 가능하며 해당 데이터를 엑셀 환경에서 오타 교정, 불필요한 띄어쓰기 제거, 특수문자 제거, 불용어 제거, 대/소문자 통일, 약어 통일, 국문, 한문, 영문을 국문으로 통일, 동의어, 유의어 통일, 복합어 분리 과정으로 키워드를 정제하였다.

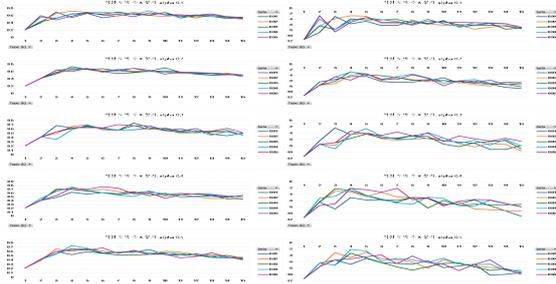
4. 토픽 모델링

토픽 모델링은 여러 가지 알고리즘과 분석 기법이 있는데, 본 연구에서는 LDA를 활용하였다[3]. 사용자가 지정해줘야 하는 하이퍼 파라미터 토픽 수 K와 α , β 를 튜닝하기 위해 복잡도(Perplexity)와 일관성(Coherence) c_v , U-mass를 활용하였다. 복잡도는 낮을수록 좋고, 일관성 c_v 는 높을수록 좋고, U-mass는 0에 가까울수록 좋다.

본 연구에서는 $K = 1 \sim 15$, $\alpha = 0.1 \sim 0.5$, $\beta = 0.01 \sim 0.05$ 로 조합하여 각 지표 값을 확인하였다.



[그림 1] 복잡도(Perplexity) 활용 적정 토픽 수 찾기



[그림 2] 일관성(Coherence) c_v , U-mass 활용 적정 토픽 수 찾기

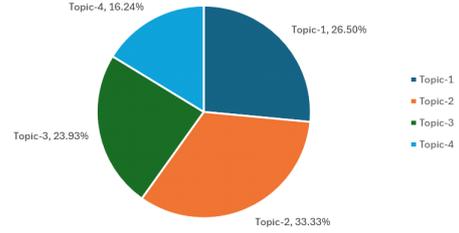
복잡도는 토픽 수가 증가할수록 낮은 값을 가진다. 감소세가 둔화하는 $K = 4$ 를 기준으로 c_v 값을 확인하였다[그림 1]. c_v 값 또한, $K = 4$ 일 때 대부분의 하이퍼 파라미터에서 상승세가 둔화함을 확인하였고, U-mass 값은 $K = 4$ 일 때 대부분의 하이퍼 파라미터에서 하락세로 전환함을 확인하였다[그림 2]. 즉, $K = 4$, $\alpha = 0.4$, $\beta = 0.04$ 를 기준값으로 하였다.

[표 2] 토픽 모델링 결과

구분	1st	2nd	3rd	4th	5th
Topic-1	4차산업혁명	기술	스마트	융합	산업
Topic-2	자동화	시스템	AMR	시뮬레이션	배송
Topic-3	인식	KIVA	제어	정확도	위치
Topic-4	다중	창고	경로	알고리즘	스마트팩토리

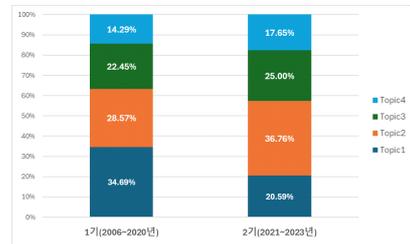
도출된 토픽 모델링 결과 [표 2]로 토픽별 분류된 논문 중 비중이 높은 논문을 바탕으로 주제 범주화를 하였으며, Topic-1 4차 산업혁명 기술과 물류의 융합연구, Topic-2 AMR 경로 최적화 및 도입 연구, Topic-3 이동형 로봇의 위치 추정 및 인식 연구, Topic-4 충돌 및 회피 알고리즘 연구로 나타났다.

토픽별 분류된 논문 비율은 [그림 3]에서 Topic-2(33.3%), Topic-1(26.5%), Topic-3(23.9%), Topic-4(16.2%)순으로 나타났다.



[그림 3] 토픽별 분류된 논문 비율

코로나 전과 후의 토픽 추이를 살펴보기 위해 1기를 2006~2020년, 2기를 2021~2023년으로 나누어 확인했다[그림 4]. 그 결과 Topic-1은 대폭 감소, Topic-2는 대폭 증가, Topic-3, 4는 소폭 증가하였다.



[그림 4] 시기별 토픽 비율의 변화

5. 결론

한국학술지인용색인 DB에 ‘물류’ AND ‘로봇’으로 검색된 전체 기간 연구를 대상으로 LDA 토픽 모델링을 통해 분석한 결과는 다음과 같다. 첫째, 주제는 4개로 분류되었다. 둘째, 4차 산업혁명 기술 융합연구는 감소 추세를 보이며 AMR 관련 연구는 증가 추세를 보였다. 셋째, 위치 추정 및 충돌회피 연구는 지속하여 연구되고 있다. 향후 국내외 더 많은 데이터에 대한 심도 있는 연구가 이루어져, 정책 입안자 및 관련 연구자에게 유용한 정보가 되기를 기대한다.

참고문헌

[1] 민연주 외 4인, “2019 물류기술정책지원사업 성과보고서”, 한국교통연구원, 2019년.
 [2] 박소연, 노승일, “물류 5.0 시대의 모듈형 스마트 물류 모델리티 디자인 콘셉트 제안”, 기초조형학연구, 24(3), pp. 49-64, 2023년.
 [3] Blei, D. M., Ng, A. Y., & Jordan, M. I, “Latent dirichlet allocation”, Journal of Machine Learning Research, 3(Jan), pp. 993-1022, 2003.