

토픽모델링을 활용한 국방정책 연구 동향 분석

Trend Analysis in Defense Policy studies using Topic Modeling

이지선*¹⁾,

Jisun Lee*¹⁾,

[초 록]

국내외 안보환경의 변화와 주변국의 전략에 따라 국방정책을 수립해야 한다. 국방기술을 비롯하여 운영분야에서까지 정책의 고도화가 이루어지고 있다. 본 연구는 토픽모델링 기법을 이용하여 국방정책연구 동향을 분석하였다. LDA분석을 이용하여 방위산업, 조직/구성, 민군/병영을 포함한 14개의 토픽 주제어를 도출하였다. 도출된 토픽 주제에 유사도를 확인하기 위해 IDM 분석을 수행하였다. 본 연구를 통해 파악된 국방정책 연구 동향은 국방기술 분야의 이해를 높이고, 국방 분야의 제도적 정책 마련을 위한 기초자료로 활용될 수 있을 것으로 기대된다.

[ABSTRACT]

National defense policies should be established in accordance with changes in the national and international security environment and strategies of neighboring countries. In addition to defense technology policies are being advanced in the field of operation. This study analyzed the trend of defense policy research using the topic modeling. Using the LDA analysis, 14 topics including defense industry organization/composition, and civil/military/barracks were derived. The IDM analysis was performed to confirm the degree of similarity to the derived topic. The trends in defense policy research identified through this study are expected to be used as basic data for improving understanding of the defense technology field and preparing institutional policies in the defense field.

Key Words : 국방정책(Defense policy), 토픽모델링(topic modeling), LDA(Latent Dirichlet Allocation), IDM(Intertopic Distance Map)

1. 서론

대내외 국방정책 환경 변화로 중장기 국방안보 전략 및 정책 수립에 대한 전제적 대응에 필요성이 증대되고 있다. 또한, 국가는 국방정책을 실현하기 위해 외교부, 국방부, 병무청과 방위사업청 등의 다양한 부서를 두고 국방외교안보, 국방전력, 병무 행정, 국방획득 등의 정책을 시행하고 있다. 국방부 내에서도 기획조정, 국방정책, 인사복지, 전력자원, 국방개혁 등의 부서로 나누어 정책을 집행하고 있다.

하지만, 한국 사회의 역사적 특수성으로 인해 국익과 안보에 영향을 줄 수 있다는 명분에 따라 군사 및 국방정보가 폐쇄적으로 관리되고 외부로의 공개가 제한되면서 다양한 관점에서

의 정책을 시행하는 데 한계가 있다. 이러한 이유로 민간 국방국책기관 중심으로 편중, 독점된 국방정책연구는 전문연구역량과 국제 경쟁력을 제한하는 결과로 이어지게 되었다.

최근의 급격한 국방 분야의 환경변화로 정책에 대한 수요는 증가하고 있지만, 여러 제약으로 인해 원활한 연구가 이어지지 못하고 있다. 이에 현재까지의 정책연구를 통하여 국방정책을 분류하고 공급대비 정책의 수요가 높은 분야를 중장기적으로 지원할 수 있는 연구를 시행하기 위해서 연구 동향을 분석할 필요가 있다.^[1]

본 연구에서는 국방정책의 연구 동향을 분석하고자 한다. 국방에서 적용 가능한 기술과 관련한 연구 동향 분석은 다수 진행되었지만, 정책의 동향을 분석한 연구는 수행된 바 없다. 따라서 본 연구는 행정안전부 정책연구관리시스템(이하 PRISM)에서 주제가 국방인 지난 10년간(2009년~2018년)의 연구 내용을 수집하여 LDA(Latent Dirichlet Allocation)기반의 토픽모델링 분석을 수행함으로써 국방정책의 주요 연구 주제가 무엇인지 분류하여 국방정책연구 동향을 분석하고자 한다.

1) 고려대학교 산업경영공학과 박사과정(Department of Industrial Management Engineering, Korea University)

* Corresponding author, E-mail: jisun0319@gmail.com

Copyright © The Korean Institute of Defense Technology

Received: November 4, 2019

Revised:

Accepted: November 25, 2019

2. 국방정책연구사업 분류

국방부의 정책연구용역은 군 내에서 주로 이용되어 각 군의 소요를 제기하며 연구를 수행하는 ‘군사학술용역연구사업’과 그 외의 일반 국방정책을 연구하고, 국방정책연구훈령에 따라 관리되는 ‘국방정책연구용역사업’으로 나뉜다. 국방정책 연구 용역은 국방정책연구훈령 제2조(국방부 훈령 제 2369호)에서 표 1과 같은 4개의 사업으로 분류되어 운영되고 있다.^[2]

표 1. 국방정책연구용역 분류

Table 1. Classification of defense policy research

과제	내용
기본연구 과제	국방정책의 대안제시 및 전문성 향상을 위해 과제 제기, 수행, 평가 등 과제에 대한 제반 관리를 기본연구기관의 장에게 위임한 과제
정책연구 부여과제	국방정책 개발과 발전을 위하여 정책수립·집행부서가 제기하는 과제 중 국방부가 기본연구기관에 별도 예산 지원 없이 매년 연구를 부여하는 과제
정책연구 용역과제	국방부의 정책개발 및 수행을 위해 제17조 및 제24조의 소요제기에 따라 위원회 심의를 거쳐 선정한 정책연구과제 중 국방부가 용역연구기관과 연구 수행에 대한 대가를 정책연구비로 지급하는 내용의 계약을 체결하여 연구를 추진하는 과제
개별연구 용역과제	부서별 소요에 따라 자체적으로 발주하여 관리하는 연구용역 과제사업

기본연구과제와 정책연구부여과제의 경우, 국방부의 지시에 따라 기본 연구기관에서 과제를 바로 수행하게 되며, 연구 결과물은 외부로 공개되지 않는다. 하지만, 정책연구 및 개별연구 용역과제의 경우, 내부기관을 비롯하여 외부기관도 연구에 참여할 수 있다. 입찰을 통하여 연구기관을 선정하게 된다. 용역과제는 PRISM을 통해 연구 제목 및 수행기관을 필히 공개되며, 연구결과물은 공개를 원칙으로 하나, 담당 기관에 따라 외부로 공개하지 않을 수도 있다.

본 연구에서는 PRISM에 공개된 연구 제목을 통하여 용역과제의 연구 동향을 파악하고자 한다.

3. 연구 방법론

토픽모델링이란 기계학습 및 자연어 처리 분야에서 토픽이라는 문서 집합의 추상적인 주제를 발견하기 위한 통계적 모델 중 하나로서, 텍스트 본문의 숨겨진 의미 구조를 발견하기 위해 사용되는 텍스트 마이닝 기법 중 하나이다. 토픽모델링은 토픽과 함께 숨겨진 구조를 추출한다.

LDA는 주어진 문서에 대하여 각 문서에 어떤 주제들이 존재하는지를 서술하는 확률적 토픽 모델 기법의 일종이다. LDA

방법을 통해 주어진 문서에서 발견된 단어 수 분포를 분석함으로써 해당 문서가 주로 어떤 주제들을 함께 다루고 있을지를 예측할 수 있다. 이때, 토픽이란 LDA 결과 도출 시 동일한 주제를 가지고 있는 단어들의 집합을 의미한다.

그림 1은 토픽모델링과 LDA의 원리를 그림으로 표현한 것이다. 문서에서 Topic이 존재한다고 가정을 하여, 각 주제에 대한 단어들의 분포를 보는 것이다.^[3]

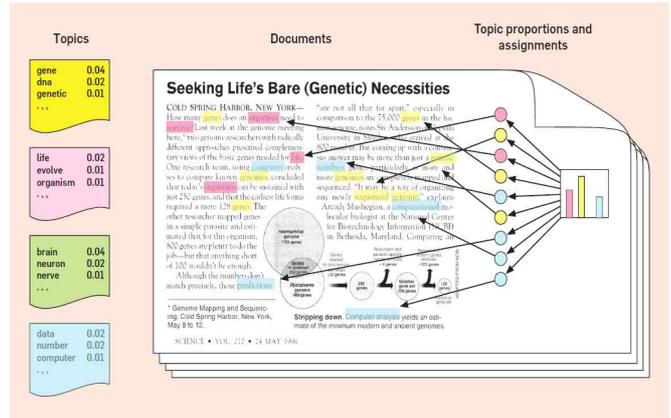


그림 1. 토픽모델링과 LDA

Fig1. Topic modeling and LDA

LDA의 분석결과로는 각 토픽의 주요 단어들의 집합으로만 추출되기 때문에, 이를 시각적으로 도식화하기 위해 IDM(Intertopic Distance Map) 방법을 활용하였다. IDM은 각 토픽의 크기와 토픽간의 유사의 정도를 파악하고, 토픽 모델을 시각화하는 방법의 하나이다.^[4]

토픽모델링은 기계학습을 통해 다량의 문서 혹은 문장에서 추상적인 주제를 확률적으로 도출하는 것 가능하다는 장점이 있다. 또한, 각 문서 혹은 문장에서의 주요 단어 목록을 추론하여 주제화할 수도 있다.

4. 연구결과

본 연구에서는 토픽모델링의 장점을 활용하여 국방정책 연구 동향의 결과물을 도출하였다. 주요 분석은 Python 3.7에서 Gensim package를 활용하였다.

PRISM에서 주제가 국방이며, 2009년에서 2018년도까지 수행된 연구과제는 총 787개였다. 787개의 연구 제목을 분석에 사용하였다. 분석결과 정확도를 높이기 위하여, 데이터는 숫자, 특수문자, 조사(은,는,이,가 등)와 같이 불필요하다고 판단되는 부분은 제거하였으며, 영어는 한글화하는 전처리 과정을 거쳤다.

앞서 설명한 토픽모델링을 활용하여 2가지 실험을 수행하였다. 첫째, LDA분석을 통하여 국방정책연구를 14가지의 토픽으로 분류하였다. 둘째, 14가지의 토픽결과물을 IDM분석을 통하여 유사도를 분석하였다.

4.1 LDA분석 결과

LDA기법을 적용하기 전에, 토픽 수를 정해야 한다. 적절한 토픽 수 추출을 위하여 Perplexity 값을 이용하였다. 2개부터 20개의까지 perplexity 값의 차가 14개일 때 가장 최소로 나타났다. 14개의 토픽을 나타내는 주요 구성단어를 각 10개씩 추출하였다.

LDA 기법을 이용하여 분석한 결과는 표 2와 같다.

표 2. LDA 분석 결과
Table 2. Results of LDA analysis

no.	단어	비율
Topic 1 조직/구조	교육, 무기체계, 육군, 성과, 공군, 구조, 표준화, 전시, 능력, 시험	0.102
Topic 2 미래전	미래, 환경, 전력, 지상군, 시설, 정보화, 부대, 해군, 획득, 모델	0.094
Topic 3 연구/기술개발	기술, 전투, 수준, 연구개발, 종합, 보호, 능력, 방산, 기준, 확보	0.082
Topic 4 작전	합동, 작전, 통제, 검증, 무기체계, 조직, SW, 위협, 능력, 협력	0.080
Topic 5 전투력 운용	훈련, 기준, 원가, 적정, 절타, 예비군, 검증, 시스템, 고도화, 획득	0.075
Topic 6 계약/획득	계약, 함정, 서울안보대화, 상호, 획득, 무기체계, 연구개발, 비용, 기준, 업체	0.074
Topic 7 전력획득	육군, 소요, 정보, 장병, 산정, 측정, 포럼, 실태, 혁신, 부품	0.074
Topic 8 시험평가	상호, 방위력, 시험, 모델, 기준, 육군, 방산, 공군, 적정, 예산	0.066
Topic 9 표준/기준	직무, 기준, 예산, 표준, 방위산업, 법령, 기술, 적정, 보수, 전력	0.064
Topic 10 정비	서비스, 병무행정, 국가, 만족도, 자격, 정비, 신설, 교육과정, 해군, 주파수	0.062
Topic 11 기술획득/품질유지	기술, 품질, 획득, 정보, 기준, 보증, 예비군, 전투, 군수, 핵심	0.061
Topic 12 비전력체계/안보교류	방위, 피복, 급식, 만족도, 협력, 안보, 포럼, 시험, 인력	0.057
Topic 13 민군/병영	자원, 민간, 동원, 데이터, 병영, 활성화, 전술, 전시, 검증, 링크	0.055
Topic 14 방위산업	작성, 계획, 방위산업, 기술, 정보, 무기체계, 육성, 타당성, 방산, 기준	0.055

그 결과, Topic 1은 교육, 표준화, 육공군, 구조 등을 포함한 조직/구조, Topic 2는 미래, 환경, 전력 등을 포함한 미래전, Topic 3은 기술, 전투, 연구개발, 확보 등의 키워드를 포함한 연구/기술개발, Topic 4는 합동, 작전, 통제, 협력등을 포함한 작전, Topic 5는 훈련, 예비군, 고도화 등을 포함한 전

투력 운용, Topic 6은 획득, 무기체계, 비용, 업체 등을 포함한 계약/획득, Topic 7은 육군, 장병, 측정, 실태 등을 포함한 전력획득, Topic 8은 상호, 방위력, 시험 등을 포함한 시험평가, Topic 9는 기준, 예산, 표준, 보수 등을 포함한 표준/기준, Topic 10은 서비스, 병무행정, 자격 등을 포함한 정비, Topic 11은 기술, 품질, 획득을 포함한 기술획득/품질유지, Topic 12는 방위, 피복, 급식 등을 포함한 비전력체계/안보교류, Topic 13은 자원, 민간, 동원, 병영을 포함한 민군/병영, Topic 14는 기술, 방위산업, 육성등을 포함한 방위산업으로 주제를 정의하였다.

조직/구조(Topic 1)과 관련된 연구가 10.2%였으며 미래전(Topic 2)와 관련된 연구가 9.4%로 높게 나타났다. 이에 비해 민군/병영(Topic 13), 방위산업(Topic 14)와 관련된 연구는 각 5.5%로 낮게 나타났다.

내부에서 관리되어야 하는 안보나 병영문제보다는 미래전을 준비하는 기술 및 무기체계와 관련한 연구가 외부에서 많이 진행된 것으로 판단되었다. 외부기관들은 내부 연구기관에 비해 기술의 동향을 파악하기에 수월하고, 정보의 습득이 빠르기 때문에 관련한 연구들이 외부에서 진행되었을 것으로 여겨졌다.

4.2 IDM분석 결과

LDA 분석 결과로 추출된 14개 주제를 활용하여 IDM 분석을 실시하였으며 결과는 그림 1과 같다.

그림 1에서 표기된 숫자는 LDA 결과의 Topic을 의미한다. 즉, 1은 Topic 1, 조직/구조를 말하며, 원의 크기는 비율에 따라 크기로 표현되었다.

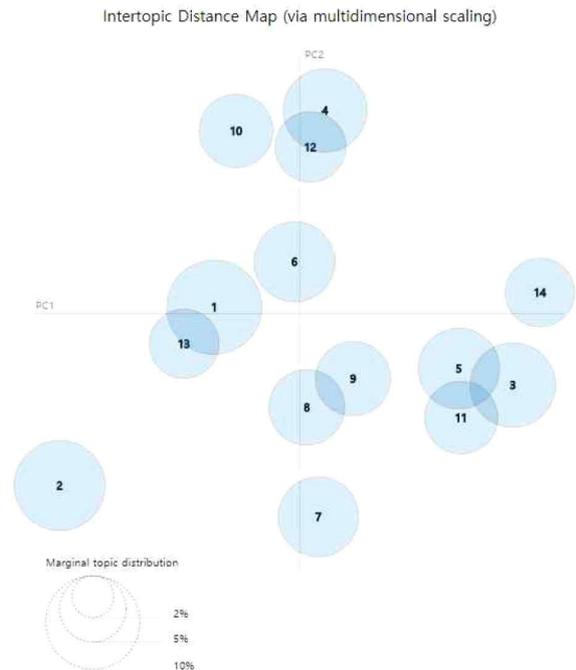


그림 2. IDM 분석 결과
Fig 2. Results of IDM analysis

References

- [1] Defense White Paper 2016, Ministry of National Defense, Republic of Korea
- [2] <https://www.law.go.kr/LSW/admRulLsInfoP.do?admRulSeq=2100000185081>
- [3] Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84.
- [4] Alghamdi, R., & Alfalqi, K. (2015). A survey of topic modeling in text mining. *Int. J. Adv. Comput. Sci. Appl. (IJACSA)*, 6(1).

조직/구조(Topic 1)과 민군/병영(Topic 13)은 병영, 표준화, 구조의 단어가 겹치는 것으로 나타났다. 군을 구성하는 전체의 조직에서 민군과 병영에 관련된 연구를 좀 더 심층적으로 하는 것으로 해석하였다.

작전(Topic 4), 비전력체계/안보교류(Topic 12)은 협력, 조직, 통제와 같은 단어가 겹치는 것으로 나타났다. 이는 국가 안보를 위한 전력체계와 비전력체계가 상황을 교류하고 있는 것으로 판단하였다.

연구/기술개발(Topic 3), 전투력 운용(Topic 5), 기술획득/품질유지(Topic 11)는 훈련, 전투, 예비군, 전술과 같은 단어가 겹치는 것으로 나타났다. 이는 전투력 강화를 위한 군의 훈련 시스템을 만들고 이를 유지하는 것으로 해석하였다.

시험평가(Topic 8), 표준/기준(Topic 9)는 예산, 기준, 적정과 같은 단어가 겹치는 것으로 나타났다. 시험평가와 표준정립 과정에서 발생하는 예산문제에 대한 기준 및 적정성을 평가하고 체계를 확립하는 것으로 해석하였다.

14개의 Topic으로 분류되었지만, 연구과제들이 서로 연관되어 정책연구가 수행된 것을 확인하였다. 특히, 각 주제들간의 관계를 보았을 때, 비율이 높게 나타난 Topic과 적게 나타난 Topic들이 서로 연관되어 있었다. 이는 관련한 연구들이 상호 관계를 가지고 있다는 것으로 파악하였으며, 정책연구에서의 동향을 결정하는 데 참고해야 할 것으로 판단되었다.

5. 결론

국방정책은 대내외적 환경에 따라 합리적이며 핵심적으로 변화되어야 할 필요가 있다. 이러한 정책변화를 위해서는 다양한 시각에서의 정책연구가 이루어져야 한다. 하지만, 국방 분야의 특수성으로 인해, 다른 영역에 비해 폐쇄성을 보이고 있다. 본 연구에서는 이러한 특수한 환경에서의 국방정책연구 동향을 확인하고, 국방기술의 이해를 높여, 제도적 정책마련을 위한 기초자료로서의 활용을 위한 연구를 수행하였다. 토픽모델링 기법을 적용하여 국방정책연구 제목을 중심으로 지난 10년 동안의 국방정책과 관련한 연구의 동향을 살펴보았다. 그 결과, 조직, 미래전, 연구, 작전, 획득 등 다양하게 연구가 되고 있음을 확인하였다.

하지만, 비공개로 된 기초과제나 부여과제에 대한 자료 수집에 불가하여, 본 결과가 지난 10년간의 연구 동향으로 보기에는 한계가 있다. PRISM에 공개된 자료들처럼, 연구 제목이라도 공개가 되었다면, 더욱 정확한 연구 동향을 파악할 수 있을 것으로 판단된다. 또한, 용역과제의 수집된 자료 내에서도 제목만을 이용하여 분석을 수행하다 보니 문헌 내 포함된 주제를 정확하게 추론하지 못한 한계가 있었다.

추후, 이런 부분을 보완하여 관련한 연구가 지속이 된다면, 중복되는 연구들을 최소화하고, 부족한 정책연구 주제를 도출할 수 있을 것으로 여겨진다. 또한, 이를 기반으로 하여 미래의 정책 연구주제를 선정할 수 있을 것으로 판단되며, 이는 나아가 국방정책연구 생태계를 조성하는데 이바지 할 수 있을 것으로 기대 된다.