Journal of the KNST

2023; Vol.6, No.3; pp. 293-297

ISSN: 2635-4926 https://doi.org/10.31818/JKNST.2023.09.6.3.293

Check for updates

Received: 2023/07/18
Revised: 2023/08/01
Accepted: 2023/08/31
Published: 2023/09/30

*Corresponding Author:

Hyeon-Ju Seol

School of Integrated National Security, Chungnam National University
Room 216, College of Natural Sciences Bldg. II,
99 Daehak-ro, Yuseong-gu, Daejeon 34134,
Republic of Korea

Tel: +82-42-821-8531

Fax: +82-42-821-8531

E-mail: hjseol@cnu.ac.kr

Abstract

설명 가능한 인공지능(XAI)은 기존 머신러닝의 성과와 설명력을 한 단계 더 향상시킬 수 있는 기술로 최근 주목받고 있다. 본 연구에서는 이 XAI 방법을 전훈분석이라는 국방정책에 정량적으로 새롭게 적용하고자 연구를 진행하였다. 특히 인공지능 자연어처리를 위한 LDA 토픽모델링을 적용한 새로운 전훈분석 방법을 제시한다. 이를 통해 도출된 정보를 예측하고 그 근거를 분석할 뿐만 아니라 인공지능 비전문가와 전투현장 지휘관의 신속한 판단에 도움이 되고자 한다.

eXplanable Artificial Intelligence(XAI) has recently gained attention as a technology that can enhance the performance and explanatory power of existing machine learning systems. In this study, we conducted research to quantitatively apply XAI methods to a defense policy known as Lessons Learned(LL) Analysis. Specifically, we propose a new LL analysis method that applies LDA topic modeling for AI NLP. The objective is not only to predict the information obtained from this method and analyze its basis, but also to assist non-expert AI users and combat field commanders in making quick judgments.

Keywords

전쟁교훈분석(Lessons Learned Analysis),
전투발전체계 분야(DOTMLPF-P),
인공지능(Artificial Intelligence),
LDA 토픽 모델링(LDA Topic Modeling),
설명 가능한 인공지능(XAI)

Acknowledgement

이 논문은 2023년도 한국해군과학기술학회 하계학술대회 발표 논문임.

설명 가능한 인공지능(XAI)을 이용한 확률 임베딩 벡터 예측에 관한 연구

Prediction of Probability Embedding Vector Using eXplainable Artificial Intelligence(XAI)

양성실[1], 설현주[2*]

[1]충남대학교 군사학과 박사과정
[2]충남대학교 국가안보융합학부 교수

Seong-Sil Yang[1], Hyeon-Ju Seol[2*]

[1]Ph.D. student, Department of Military Studies, Chungnam National University
[2]Professor, School of Integrated National Security, Chungnam National University

1. 서론

한국 국방부는 첨단과학기술 강군 육성을 위해 '국방 AI'를 미래 전장의 게임 체인저로서 적극적으로 검토하고 있다. 본 연구는 이런 배경하에서 '국방 AI'를 주목하여, 기존 국방정책 분야 중 전쟁교훈분석(이하 전훈분석)에 인공지능을 새롭게 적용하고자 연구를 진행하였다. 전훈분석은 각종 작전·전투, 훈련 등 군사활동에서 얻은 문제점에 관한 교훈을 전투발전체계 개선 소요에 적용함으로써 현재 운용되는 전력을 최적화하는 모든 활동이다. 이처럼 전훈을 도출하기 위해서는 과거 발생한 전쟁의 분석사례, 관련 교리·교범, 훈령·규정 외에도 작전계획 등 검토해야만 하는 데이터양이 매우 많고 결과적으로 분석 절차에 다량의 시간뿐만 아니라 관련된 예산이 다수 소요된다.

본 연구는 앞서 언급한 문제점에 초점을 맞추어 기존 정성적인 방법 외 인공지능을 기반으로 한 정량적인 연구를 적용하려 한다. 한편 최근 인공지능 모델을 통해 도출된 결과는 그 의미가 무엇인지 이해하거나 근거를 찾기가 쉽지 않으며, 도출과정의 논리적 설명이 매우 어렵다. 이 때문에 미국 국방고등연구계획국(DARPA: Defense Advanced Research Projects Agency)은 '설명 가능한 인공지능(XAI: eXplainable AI)' 연구를 진행하고 있으며, 현재 인공지능이 가진 한계점을 파악하고 XAI가 그것을 극복할 수 있으리라 판단하고 있다.

이처럼 XAI는 머신러닝의 성과와 설명력을 한 단계 더 향상시킬 수 있으며[1], 한국 국방부 역시 미래 유망 기술군으로 XAI를 선정한 바 있다.

따라서 본 연구는 인공지능 자연어처리를 위한 LDA 토픽 모델링을 적용한 새로운 전훈분석 방법을 제시한다. 이를 통해 도출된 정보를 예측하고 그 근거를 분석할 뿐 아니라 인공지능 비전문가인 국방부 및 각 군 본부의 정책결정자와 전투현장 지휘관의 신속한 판단에 도움이 되고자 한다.

2. 본론

2.1 전훈분석 및 전투발전체계

전훈분석은 전·평시 전쟁을 준비 및 시행하며 전쟁·전투·작전 등과 관련한 개선소요를 찾아내고자 노력해 얻는 결과이다. 최근 군사영역은 지상, 해양 및 공중, 우주, 사이버 공간 등으로 점진적으로 확대되고 있으며, 그 확장 속도는 더욱더 급변하고 있다. 이로 인해 군은 전쟁양상 변화에 효과적으로 적응하고 미래전에서 전승을 보장하고자 전력에 대한 전투발전체계 개선소요의 수집·분석, 개선소요 창출 등을 추진하게 되는데 그 핵심정책 중 한 가지 방안이 바로 전훈분석이다.

전투발전체계는 전훈분석 결과 개선되는 핵심이며 합동성 차원의 총체적인 산물로서, 현재 운용전력 및 미래 전력을 개선하고 최적화하는 목적을 가진다. 이는 교리(doctrine), 구조·편성(organization), 훈련(training), 무기·장비·물자(material), 리더십·교육(leadership & education), 인적자원(personnel), 시설(facilities), 정책(Policy) 등 분야로 구분할 수 있으며, 요약해서 DOTMLPF-P라 지칭한다[2].

2.2 LDA 토픽 모델링 기반 확률임베딩 벡터 추출

전훈분석의 데이터를 추출 및 예측하고자 연구에 필요한 데이터 수집, 데이터 전처리, 토픽 모델링, XAI분석의 단계로 진행하였으며, 그 과정은 Fig. 1과 같다. 이때 파이썬(Python)을 토픽 모델링 및 XAI 분석에 적용하였다.

기본적으로 전훈 데이터는 문자화된 이산형 데이터로서 과거 분석사례, 교범, 훈령·규정, 전쟁에 관한 작전계획 등을 기반으로 한다. 본 연구에서는 이를 어떻게 연속형 데이터로 바꿀 것인지에 관한 방법을 집중적으로 검토하였다. 결론적으로 LDA 토픽모델링에 이산형 전훈 데이터를 삽입하여 적용하였고, 토픽모델링을 통해 확률 임베딩 벡터값을 추출함으로써 연속형 데이터인 숫자 형태로 바꿀 수 있다.

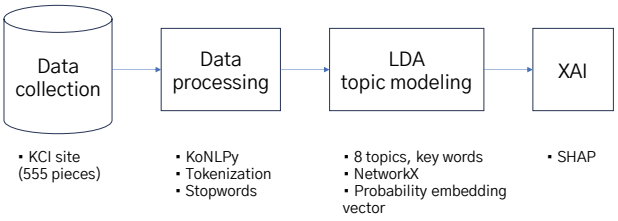

Fig. 1. Lessons learned analysis process

맨 처음 단계에서 전훈 데이터를 수집하기 위해 한국학술지인용색인(Korea Citation Index: KCI) 사이트에 게재된 한국어 논문 초록(abstract)을 수집하였다. 데이터 수집 시 '전쟁교훈', '전투교훈', '전투발전체계' 등 주요 키워드를 사용해 연구에 필요한 데이터 논문 초록 총 555편을 추출하였다.

이후 한국어 데이터를 사용하고자 KoNLPy(Korean NLP in Python)를 적용하여 데이터 전처리 단계를 진행한다. 데이터에 포함된 텍스트를 필요한 단위로 나누는 토큰화(tokenization) 과정을 통해 분석 시 요구되는 형태로 분류하였다. 또 토픽 모델링 연구에 필요하지 않은 한국어의 불용어, 예를 들어 의성어, 감탄사, 동사, 부사 등 총 1,342개를 분석 데이터상에서 제거하였다.

LDA 토픽 모델링 결과, 토픽별 주요 핵심단어 5개를 Table 1과 같이 추출하고, 최종적으로 토픽이 의미하는 주제를 전투발전체계 8가지 분야별로 추론할 수 있었다.

Table 1. Result of LDA topic modeling

Topic no.	Key words	Subject
1	시설, 정비, 보호, 군사시설, 공사	Facilities
2	체계, 전투, 발전, 장비, 미래	Material
3	갈등, 정책, 지역, 관계, 외교	Policy
4	전력, 정신, 교육, 자원, 군인	Personnel
5	부대, 사단, 전쟁, 지역, 군단	Organization
6	전쟁, 군사, 전략, 연구, 분석	Doctrine
7	드론, 항공, 로봇, 무인항공기, 전문	Training
8	리더십, 분석, 역량, 교육, 장교	Leadership, Education

토픽별 추출된 결과를 통해 주제를 추론해 보면, 예를 들어 토픽 1은 시설, 정비, 보호, 군사시설, 공사 등의 핵심단어를 가지므로 그 주제를 전투발전체계 중 시설(facilities)로 판단하였다. 다른 나머지 토픽에 대해서도 도출된 핵심단어를 고려해서 전투발전체계 중 적합한 주제를 결정할 수 있다.

LDA 토픽 모델링에서 추출한 토픽 8개와 핵심단어 도출한 결과 토픽별 주제가 어느 정도 전투발전체계와 일치하나, 완벽하게 일치하지 않는 사례도 있음을 알 수 있다. 그 이유는 LDA 토픽 모델링이 가지는 비지도학습(unsupervised learning)으로서 한계 때문이다. 즉, 비지도학습은 정답(label)이 없는 데이터를 알고리즘이 계산에 의해 처리하므로 정확한 예측이 쉽지 않고 때로는 모델을 직접 해석해야 하는 경우도 있다. 또 지도학습과 비교해 비지도학습은 출력값이 주어지지 않기 때문에 모델의 평가가 어렵고 생성된 토픽을 조정하는 데 문제점이 있다. 이와 같은 LDA 토픽 모델링의 문제점을 해결하기 위해 기존 연구에서 확장된 연구방안으로 지도학습(supervised learning)의 방법을 제시하기도 한다[3].

한편 LDA 토픽 모델링 결과 데이터의 정보를 잠재적(latent)으로 추론한 확률 임베딩 벡터값 추출 결과는 Table 2와 같다. Table 2에서 행은 555편의 논문 데이터이며, 열은 전투발전체계 분야(주제)별로 나열한 것이다. 이 값은 확률값이므로 각 행의 합은 1이다.

본 연구에서는 XAI를 적용하기 위해 LDA 토픽 모델링이라는 비지도학습방법을 실시한 결과 토픽과 논문 데이터가 일치하지 않는 경우를 확인하였다. 그 결과 논문 555편 중 428편, 약 77.1 %가 정확히 일치하는 반면, 일부 127편, 약 22.9 %가 일치하지 않음을 확인하였다. 이 결과를 출력값으로 두고 XAI 예측에 관한 추가 연구를 진행하였다.

수집된 데이터 간의 관계구조를 이해하고자 노드(node)와 링크(link)로 모형화하고 시각적으로 표현되는 네트워크 분석을 시행하였다. 이를 통해 데이터의 상호관계 및 연관성을 해석할 수 있었으며 그 결과는 Fig. 2와 같다. Fig. 2에서 교리는 가장 많은 링크를 가지므로 요소 간 영향력이 크고 연결 중심성(degree centrality)이 높았다. 즉 전훈 데이터에서 매우 중요하게 고려해야만 하는 속성이다.

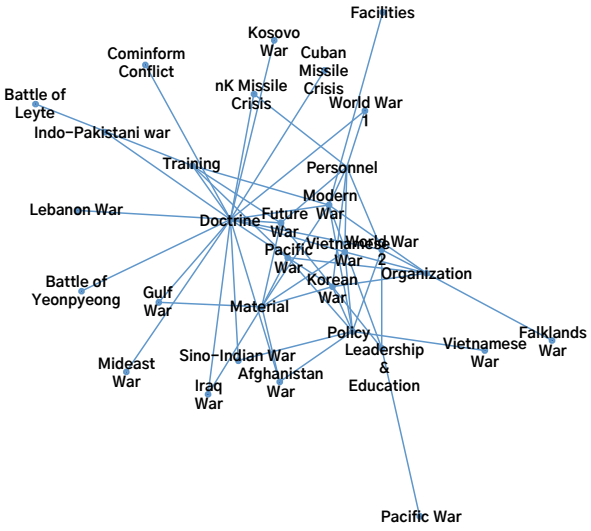

Fig. 2. Result of network analysis

Table 2. Result of deriving probability embedding errors

Data no.	Doctrine	Organization	Training	Material	Leadership & Education	Personnel	Facilities	Policy
1	0.7702595	0.07762446	0	0.02595359	0	0.06410624	0	0.05882305
2	0.13266823	0.47922707	0	0.04073082	0.33095792	0.01431501	0	0
3	0.38323134	0	0.11369541	0.07326552	0.06923365	0.35865301	0	0
4	0.21766208	0.28755257	0	0.09859805	0.02765369	0.2744329	0	0.09138203
5	0.68940282	0.04432707	0	0.19217272	0.06795599	0	0	0
6	0.24	0.21	0.01	0.03	0	0.47	0	0.04
…	…	…	…	…	…	…	…	…
554	0.06466052	0	0	0.61026549	0.3005639	0.02191217	0	0
555	0.07282072	0.01487061	0	0.09275485	0.3837547	0.02974654	0.0137443	0.39160138

2.3 SHAP을 이용한 XAI 기반 예측 결과

이제 SHAP(SHapley Additive exPlanations) 알고리즘을 사용해 전체 데이터의 상관관계가 어떠한지 해석하였다[4]. 전체 데이터 555개 중 127개의 학습용 데이터 세트를 세로로 누적하여 부분 의존성 플롯을 활용해 선형적으로 가시화할 수 있는데 그 결과는 Fig. 3과 같다. Fig. 3에서 x축이 37번째 데이터인 위치에 마우스를 올리면, SHAP 라이브러리는 자바스크립트 툴킷을 사용한 결과를 보여준다. 예컨대 37번째 데이터의 예측값은 5.843이며, 데이터에 긍정적인 요소와 부정적인 요소를 파악할 수 있다. 즉 전투발전체계 연구에 긍정적인 요소는 교리, 리더십·교육, 무기·장비·물자, 시설, 구조·편성이며, 부정적인 요소는 훈련임을 전체 데이터를 통해 확인할 수 있다.

다음으로 각 요소가 모델 내에 반영한 특정 데이터에 대한 영향력을 상세하게 분해하고 시각화한 결과는 Fig. 4와 같다. 각 요소 간 영향력에 관한 결과로 붉은색은 긍정적인 영향을 나타내며 녹색은 부정적인 의미이다. 제안된 모델이 제시한 예측값은 4.52인데, 요소별 영향력으로 긍정적인 영향은 교리, 구조·편성, 정책, 시설, 리더십·교육의 순서이고, 부정적인 요소에는 인적자원, 훈련이 있다. 즉 전투발전체계 연구에 긍정적인 영향을 준 요소 중에서는 교리가 가장 영향력이 크며, 부정적인 요소는 인적자원이라고 해석할 수 있다. 결과적으로 최근 전훈 관련 연구 중에서는 교리 분야 연구가 많고, 상대적으로 인적자원에 대한 연구가 적음을 추론할 수 있다.

각 데이터를 구성하는 요소 간의 독립성을 바탕으로 각 요소가 전체에 얼마나 공헌했는지 시각화한 결과는 Fig. 5와 같다. 해당 결과에 따르면 전투발전체계에서 예측에 가장 큰 영향을 주는 변수는 교리이다. 그 다음으로 고려해야 할 핵심분야는 리더십·교육, 무기·장비·물자, 인적자원, 시설, 구조·편성, 정책, 훈련 등의 순서임을 알 수 있다. 이를 통해서 교리, 리더십·교육, 무기·장비·물자 등을 연구한 데이터가 기존 다른 분야의 연구에 비해 다양하고 비중 있게 다루어진다는 것도 추론할 수 있다. 요소별 중요도를 통해서는 각 요소 간 의존성을 간과할 수밖에 없으므로 그 상관관계를 별도로 확인할 필요가 있다.

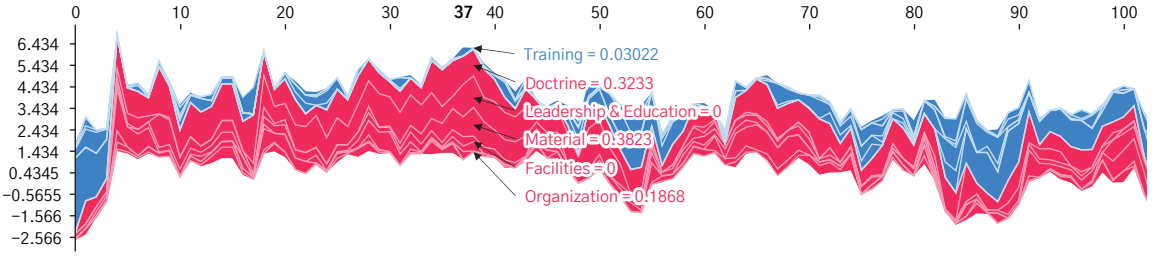

Fig. 3. Clustering with SHAP feature attribution

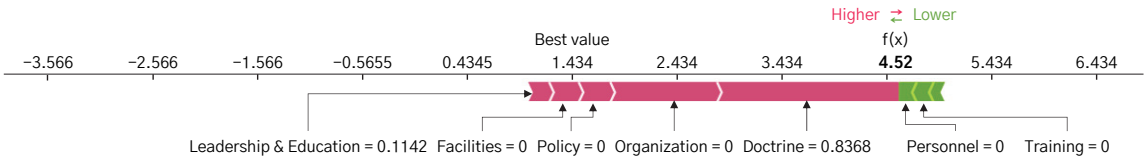

Fig. 4. SHAP feature attribution on specific data (No.78)

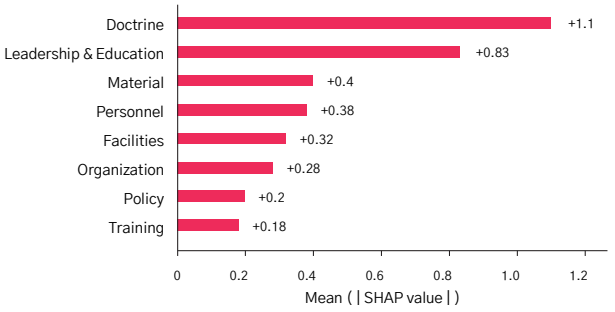

Fig. 5. Feature importance

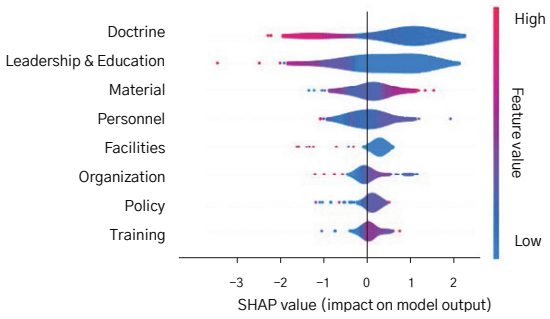

Fig. 6. Summary plot

모델 해석을 위하여 전체 요소가 전투발전체계 분석에 어떤 영향을 미치는지 시각화한 결과는 Fig. 6와 같다. 이 결과는 요소별 분포의 실제 데이터 및 전체 형상을 보여준다. Fig. 6에서 붉은색은 전투발전체계 연구분야 별로 해석할 때 큰 영향을 미친다는 의미이고, 파란색은 적은 영향을 미쳤다는 의미이다. 즉 교리, 리더십·교육은 분산의 정도가 타 분야와 비교해 상당히 크며 그 다음으로 무기·장비·물자, 인적자원, 시설, 구조·편성의 순이다. 즉 교리, 리더십·교육, 무기·장비·물자, 인적자원 순으로 전투발전체계 분야별 연구방향을 결정하는 데 큰 역할을 한다는 의미로 해석할 수 있다.

3. 결론

본 연구에서는 기존 전훈 연구에서 도출된 문제점을 해결하고자 새롭게 인공지능 기반 머신러닝, 자연어처리 기법을 적용하여 정량적 연구를 진행하였다.

특히 자연어로 이뤄진 전훈 데이터를 LDA 토픽 모델링을 이용해 전투발전체계 분야별 토픽 및 핵심단어로 구분하여 추출하고, 확률 임베딩 벡터값으로 전환할 수 있었다. 또 XAI를 적용함으로써 인공지능 관련 비전문가인 군의 작전·전투 지휘관과 국방부 및 각 군 본부의 정책결정자가 인공지능이 제시하는 결과를 신뢰할 수 있도록 근거를 도출하였다.

따라서 새롭게 적용한 전훈분석의 정량적인 방법을 통해서 전투발전체계 각 분야에 수렴한다는 연구결과 외에도 요소별 중요도 및 영향력, 상관관계 등을 시각적으로 확인하였다. 또한, 기존 연구에서 교리, 리더십·교육, 무기·장비·물자, 인적자원 등의 주제가 연구 결과도 많고 비중 있게 다뤄지고 있음을 데이터 분석을 통해 추론할 수 있었다.

향후 전훈에 대한 다양한 인공지능 방법이 적용되어야 하며, 특히 인공지능을 적용하는 데 보다 많은 가용 데이터 생성 및 추출을 위한 전반적인 국방분야 전훈의 빅데이터 확보가 필수적이라 판단한다. 더불어 인공지능 머신러닝을 적용하기 위해서 비지도학습 외에도 지도학습, 강화학습 등 다양한 기법을 적용하는 지속적인 연구가 필요하다.

참고문헌

[1] Gunning, David and Aha, David W., “DARPA’s Explainable Artificial Intelligence Program,” AI Magazine, Association for the Advancement of Artificial Intelligence, 2019. pp. 44–58.
[2] Yang, S-S. and Soel, H-J., “A Study on the Lessons Learned Analysis Using Artificial Intelligence Technique: Based on LDA Topic Modeling,” Review of Korean Military Studies, Vol. 12, No. 1, 2023, pp. 29–48.
[3] Blei, David M., “Probabilistic Topic Models,” Communications of the ACM, Vol. 55, No. 4, 2012, pp. 77–84.
[4] Lundberg, Scott M., Erion, Gabriel G. and Lee, Su-In, “Consistent Individualized Feature Attribution for Tree Ensembles,” arXiv preprint arXiv:1802.03888, 2018.