

뉴스 기사의 정치적 성향 판단을 위한 지식 그래프 임베딩 기법의 효과 분석

류성은¹, 고윤용², 김상욱^{3*}

¹한양대학교 인공지능학과 석박사통합과정

²일리노이 대학교 어바나-삼페인 컴퓨터공학과 박사후연구원

³한양대학교 컴퓨터소프트웨어학과 교수

ryuseong@agape.hanyang.ac.kr, yyko@illinois.edu, wook@agape.hanyang.ac.kr

Knowledge Graph Embedding Methods for Political Stance Prediction: Performance Evaluation

Seongeun Ryu¹, Yunyong Ko², Sang-Wook Kim^{3*}

¹Dept. of Artificial Intelligence, Hanyang University

²Dept. of Computer Science, University of Illinois at Urbana-Champaign

³Dept. of Computer Science, Hanyang University

요 약

온라인 뉴스 플랫폼의 발전은 에코 챔버(echo chamber) 효과와 정치적 양극화를 심화시키며, 이를 완화하기 위한 선행 연구로 뉴스 기사의 정치적 성향을 판단하는 연구가 필요하다. 기존 연구는 외부 지식 그래프를 활용하여 뉴스 기사의 텍스트 정보를 더욱 풍부하게 표현한다. 그러나, 외부 지식을 임베딩하는 지식 그래프 임베딩(knowledge graph embedding, KGE) 방법은 다양하며, 각 KGE 방법이 정치적 성향 예측 정확도에 미치는 효과에 대해서 충분히 연구되지 않았다. 본 논문에서는 정치적 성향 예측에 외부 지식의 활용을 최대화하기 위한 다양한 KGE 방법들의 효과를 분석한다. 실험 결과, 외부 지식 그래프 내의 개체들 간 복잡한 관계를 간단하고 정확하게 표현 가능한 ModE 방법을 활용하는 것이 정치적 성향 예측에 가장 효과적이라는 것을 확인하였다.

1. 서론

최근 온라인 뉴스 플랫폼이 발전하면서, 해당 플랫폼의 사용자들은 양질의 다양한 정보를 쉽고 빠르게 접하고 있다. 일반적으로 사용자들은 자신의 의견과 비슷한 성향의 뉴스 기사를 선호하는 경향이 있다. 이러한 현상이 지속적으로 반복될 경우, 사용자는 특정한 성향을 갖는 의견에 간혀 다른 의견을 받아들이기 어려워 질 수 있으며, 이를 에코 챔버(echo chamber) 효과라 한다[1].

이는 사용자들의 건강한 뉴스 소비뿐만 아니라, 사용자들 간 건강한 의사소통 또한 저해할 수 있다. 특히 정치적 주제를 다루는 뉴스의 경우, 사용자들 간 정치적 양극화를 초래하여 전사회적 갈등을 심화시키는 근본적인 원인이 될 수 있다[2]. 설상가상으로 사용자의 과거 뉴스 소비 이력을 분석하여 새로운 뉴스를 추천하는 기계학습 기반의 뉴스 추천 시스템의 발전은 이러한 문제를 더욱 악화시키고 있다[3].

이러한 정치적 양극화 문제를 완화하기 위해서는 사용자들에게 다양한 성향의 뉴스 기사를 제공함으로써 사용자들이 자신의 생각과 다른 의견을 접하고 형성된 에코 챔버를 깰 수 있도록 유도하기 위한 노력이 필요하다[4]. 이러한 노력의 일환으로, 주어진 뉴스 기사의 정치 성향을 정확하게 판단하는 기술이 선행적으로 연구되어야 한다.

이러한 배경에서 뉴스 기사의 정치적 성향을 예측하기 위한 다양한 기술들이 활발하게 연구되고 있다[5-8]. 일반적으로 정치적 뉴스 기사는 전문적인 용어, 시의성이 있는 개념 등 텍스트 그 자체만으로는 이해하기 어려운 정보를 포함한다(예: 특정 정당의 이름, 정책의 이름 등). 따라서, 일반적인 언어모델은 정치적 뉴스 기사를 올바르게 모델링하기 어렵다는 한계를 갖는다[5]. 이러한 한계를 극복하기 위해, 정치와 관련된 외부 지식을 지식 그래프로 표현하여 그들 간의 관계를 학습한 후, 학습된 정보를 활용하여 뉴스 기사의 텍스트 정보를 보다 풍부하게 표현하는 지식 그래프 기반의 정치 성향 예측 기술들이 연구되고 있다[6-8].

* 교신저자

지식 그래프 기반의 정치 성향 예측 기법은 (1) 외부 지식 그래프 임베딩과 (2) 학습된 지식 기반의 정치 성향 예측, 두 단계로 구성된다. 이때, 단계 (1)에 대해 다양한 그래프 임베딩 기법들이 존재한다[9-11]. 그러나 각 지식 그래프 임베딩 기법이 최종 뉴스 기사의 정치 성향 예측 정확도에 미치는 효과에 대해서는 충분히 연구되지 않았다. 이러한 동기에서 본 논문은 뉴스 기사의 정치적 성향 예측에 미치는 지식 그래프 임베딩 기법들의 효과를 검증 및 분석하고자 한다.

2. 관련 연구

정치적 성향 예측 문제는 각 뉴스 기사의 텍스트 정보(예: 제목, 본문)를 기반으로 정치적 성향을 예측(예: 진보, 보수)하고, 예측한 성향에 따라 해당 뉴스를 분류하는 연구 분야이다[5-8].

이전의 정치 성향 분류 연구는 뉴스 기사에 등장하는 단어와 문장을 잘 이해하기 위해 사전 훈련된 언어 모델(pre-trained language model)(예: BERT, RoBERTa)을 활용한 방법들이 존재한다[5]. 그러나, 사전 훈련된 언어 모델은 정치적 성향을 파악하기 위해 정교하게 설계된 모델이 아니기 때문에, 정치적 성향을 파악하는 문제에서는 높은 정확도를 달성하기 어렵다[6].

정치적 뉴스는 정책, 정치적 인물과 같이 전문적인 배경 지식이 필요한 단어를 자주 포함한다. 따라서 텍스트 정보와 관련된 추가적인 배경 지식을 활용하여 뉴스 기사의 문맥을 파악하는 것은 정치적 성향을 예측하는데 도움이 된다[6-8].

최근 연구들은 정치적 지식 그래프를 구축하고, 구축된 지식 그래프로부터 지식 그래프 임베딩(knowledge graph embedding, 이하 KGE) 방법을 통해 외부 지식을 텍스트 정보에 효과적으로 주입하여 정치적 성향을 예측한다[6-8]. 특히 [8]에서는 진보와 보수에 대한 서로 다른 외부 지식 그래프를 활용한다. 이는 동일한 정치적 용어라도 서로 다른 의견을 공유하는 정당에 따라 해석과 의미가 다를 수 있음을 고려한다.

KGE는 지식 그래프 데이터를 저차원 벡터로 변환하는 방법으로, 그래프를 구성하는 개체(entity)와 관계(relation)를 벡터 공간에 표현하여 그래프 상에서의 연결 관계를 학습한다. 개체들 간의 관계를 표현하는 방법에 따라서 다양한 KGE가 존재하며, 대표적인 최근 방법들에는 RotatE, HAKE, ModE가

<표 1> SemEval, Allsides 데이터셋 통계

Dataset	# of articles	Class distribution
SemEval	645	407 / 238
Allsides	14.7k	6.6k / 4.6k / 3.5k

있다[10, 11]. RotatE는 각 개체들 간의 복잡한 관계를 회전 벡터(rotation vector)로 활용하는 방법으로, 소스(source) 개체의 회전 벡터와 타겟(target) 개체가 가까워지도록 학습한다[10]. HAKE는 개체들 간의 연결 정보를 극좌표계(polar coordinate system)에 표현하여, 극좌표계를 구성하는 여러 개의 동심원(concentric circle) 구조를 통해 각 개체들 간의 절대적인 위치(modulus) 정보와 계층적인 깊이(phase) 정보를 반영한다[11]. ModE는 HAKE 구조에서 하나의 동심원 좌표계만을 사용하여 개체들 간 절대적인 위치 정보를 활용한 KGE 방법이다[11].

지식 그래프로부터 개체들 간 관계를 학습하는 KGE 방법은 접근 방법에 따라 동일한 개체라도 서로 다른 관계를 학습하며, 지식 활용에 있어 서로 다른 효과를 보인다. 하지만, 기존 연구들은 정치 성향 예측 문제에서 다양한 KGE 방법이 정치 성향 예측 정확도에 미치는 효과에 대해 분석한 바가 없다. 따라서, 우리는 기존 연구를 확장하여 외부 지식의 활용 과정에서 어떠한 KGE 방법이 정치적 성향 예측에 가장 효과적인지에 대한 분석을 진행하였다.

3. 실험

3.1 실험 환경

본 논문에서는 공개 데이터셋인 SemEval과 Allsides 데이터셋을 사용하여 실험을 진행하였다[12, 13]. SemEval은 초당파(hyperpartisan)와 중립(center)의 성향을 나타내는 2개의 클래스로 구성되며, Allsides는 진보(left), 중립(center), 보수(right) 3개의 클래스로 구성된 정치적 뉴스 기사 데이터셋이다. SemEval과 Allsides 데이터셋에 대한 통계는 표 1에 나타내었다.

본 실험은 기존의 최신(state-of-the-art) 연구[8]에서 구축한 정치적 지식 그래프를 외부 지식으로 사용하였으며, 뉴스 기사에 등장하는 개체들 중 지식 그래프에 포함된 개체의 노드 임베딩을 학습하였다. 다양한 KGE 방법의 효과를 분석하기 위한 평가 지표로는 정확도(Accuracy)를 사용하였다. 이전 연구들과 마찬가지로, 10-fold 교차 검증 방법을 사용하여 모델 학습과 평가를 수행하였다.

<표 2> 기존 모델 및 KGE 방법에 따른 정확도

Method	Dataset	
	SemEval	Allsides
BERT[5]	0.8692	0.8246
RoBERTa[5]	0.8708	0.8535
KGAP[6]	0.8956	0.8602
KCD[7]	0.9087	0.8738
KHAN[8]-RotatE	0.9426	0.9151
KHAN[8]-HAKE	0.9395	0.9216
KHAN[8]-ModE	0.9521	0.9256

3.2 실험 1. 다양한 KGE 방법에 따른 정치적 성향 예측 효과 분석

표 2는 기존 모델의 예측 정확도 및 서로 다른 KGE 방법에 따른 정치적 성향 예측 정확도를 나타낸다. 실험을 위한 베이스라인 모델로는 기존의 정치적 성향 예측을 위한 최신 모델인 KHAN[8]을 사용하였으며, RotatE[10], HAKE[11], ModE[11] 세 가지 서로 다른 지식 인코딩 방법을 적용하였다. 실험 결과를 통해 ModE를 사용한 지식 인코딩 방법이 가장 우수함을 확인하였다. 이를 통해 단일 동심원 좌표계만을 활용하여 개체들 간의 절대적인 위치 정보를 임베딩하는 ModE 방법이 복잡한 개체들 간의 관계를 표현하는 간단하면서도 정확한 방법임을 확인할 수 있었다.

또한, 사전 훈련된 언어 모델들과 예측 정확도를 비교하였을 때, 외부 지식을 활용한 정치적 성향 예측 정확도가 크게 증가함을 확인하였다. 이를 통해 전문적인 배경 지식이 다수 포함되어 있는 정치적 뉴스 기사의 특징에 따라, 뉴스 기사와 관련된 외부 지식을 활용하는 것은 뉴스 기사를 이해하는데 보다 유용하다는 것을 확인할 수 있었다.

4. 결론 및 향후 연구

본 논문은 정치적 성향 예측을 위해 외부 지식 활용의 효과를 확인하고, 지식 활용 과정에서 어떠한 KGE 방법이 가장 효과적인지 분석하였다. 실험 결과를 통해 개체들 간의 복잡한 관계를 간단하게 표현 가능한 ModE 방법이 정치적 성향 예측을 위한 지식 인코딩에 가장 효과적임을 확인할 수 있었으며, 외부 지식의 활용 자체로도 예측 정확도를 크게 높일 수 있다는 결론을 도출하였다. 이러한 결과를 바탕으로, 시간적 특성을 고려하는 보다 고도화된 KGE 방법을 개발하는 것이 향후 연구 과제이다.

참고문헌

- [1] Kai Shu et al., "Fake news detection on social media: A data mining perspective," in ACM SIGKDD, 19, 1, 2017.
- [2] Wesley et al., "Quantifying echo chamber effects in information spreading over political communication networks," in EPJ Data Science, 8, 1, 2019.
- [3] Nabeel Gillani et al., "Me, my echo chamber, and I: introspection on social media polarization," in TheWebConf, 2018.
- [4] Q Vera Liao et al., "Can you hear me now? Mitigating the echo chamber effect by source position indicators," in CSCW, 2014.
- [5] Chang Li et al., "Using social and linguistic information to adapt pretrained representations for political perspective identification," in ACL-IJCNLP, 2021.
- [6] Shangbin Feng et al., "KGAP: Knowledge Graph Augmented Political Perspective Detection in News Media," arXiv preprint arXiv:2108.03861, 2021.
- [7] Wenqian Zhang et al., "KCD: Knowledge Walks and Textual Cues Enhanced Political Perspective Detection in News Media," in NAACL, 2022.
- [8] Ko, Yunyong et al., "KHAN: Knowledge-Aware Hierarchical Attention Networks for Political Stance Prediction," in TheWebConf, 2023.
- [9] Tim Dettmers et al., "Convolutional 2d knowledge graph embeddings," in AAAI, 2018.
- [10] Zhiqing Sun et al., "RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space," in ICLR, 2018.
- [11] Zhanqiu Zhang et al., "Learning hierarchy-aware knowledge graph embeddings for link prediction," in AAAI, 2020.
- [12] Johannes Kiesel et al., "Semeval-2019 task 4: Hyperpartisan news detection," in 13th International Workshop on Semantic Evaluation, 2019.
- [13] Chang Li et al. "Encoding social information with graph convolutional networks for political perspective detection in news media," in ACL, 2019.

사사

본 논문은 과학기술정보통신부의 재원으로 정보통신기획평가원(IITP)의 지원을 받아 수행된 연구임. (과제번호:RS-2022-00155586,2020-0-01373,2022-0-00352).