Knowledge Graph Augmented Political Perspective Detection in News Media

Shangbin Feng, Zilong Chen, Qingyao Li, Minnan Luo
Xi’an Jiaotong University, Xi’an, China
{wind,binteng,luoyangczl,ly890306}@stu.xjtu.edu.cn, minnluo@xjtu.edu.cn

Abstract

Identifying political perspective in news media has become an important task due to the rapid growth of political commentary and the increasingly polarized ideologies. Previous approaches only focus on leveraging the semantic information and leave out the rich social and political context that helps individuals understand political stances. In this paper, we propose a perspective detection method that incorporates external knowledge of real-world politics. Specifically, we construct a contemporary political knowledge graph with 1,071 entities and 10,703 triples. We then build a heterogeneous information network for each news document that jointly models article semantics and external knowledge in knowledge graphs. Finally, we apply gated relational graph convolutional networks and conduct political perspective detection as graph-level classification. Extensive experiments show that our method achieves the best performance and outperforms state-of-the-art methods by 5.49%. Numerous ablation studies further bear out the necessity of external knowledge and the effectiveness of our graph-based approach.

Introduction

The past decade has witnessed dramatic changes of political commentary in two major ways. Firstly, gone are the days that journalists and columnists in prestigious media outlets are the only source of news articles. The popularity of social media has provided viable venues for everyday individuals to express their political opinions freely, which leads to the dramatic increase in the volume of political commentary. Secondly, the increasing polarization of political ideologies has made it hard for news articles to remain impartial in nature. Detecting political perspectives in news media would help alleviate the issue of “echo chamber”, where only a single viewpoint is reiterated and further deepens the divide.

That being said, political perspective detection has become a pressing task which calls for further research efforts. Previous research proposals on news bias detection have focused on analyzing the textual content of news articles. Natural language encoders (Kiros et al. 2015, Yang et al. 2016) and pre-trained language models (Devlin et al. 2018) are adopted by (Li and Goldwasser 2021) to analyze news content to conduct stance detection. (Jiang et al. 2019) uses convolutional neural networks with Glove word embeddings (Pennington, Socher, and Manning 2014) for political perspective detection and achieves the best result in the hyperpartisan news detection task in SemEval 2019 (Kiesel et al. 2019). (Li and Goldwasser 2021) leverages the attention mechanism and entity mentions in the news document and achieves the state-of-the-art results.

However, an ample amount of external knowledge exist in these news articles and help individuals better identify political perspectives. For example, Figure 1 presents a news article that contains real-world entities such as political organizations, elected officials and geographical locations. External knowledge of these entities inform the reader that they are conservative republicans, which help to identify the liberal stance expressed in criticizing them. This reasoning demonstrates that external knowledge provides background information and serves as essential indicators of author perspectives. That being said, robust perspective detectors should leverage external knowledge about entities discussed in news articles to boost task performance.

Figure 1: Example of news piece and social and political entities that require external knowledge to understand.

Copyright © 2021, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.
learn representations for entities and relations in the knowledge graph with knowledge graph embedding techniques. After that, we construct a heterogeneous information network to represent news documents which include both textual content and mentioned entities in the knowledge graph. Finally, we apply gated relational convolutional networks for graph classification and conduct political perspective detection. Our main contributions are summarized as follows:

- We construct and publicize a knowledge graph of contemporary U.S. politics to serve as the external knowledge of political perspective detection. The knowledge graph could also serve as external knowledge for other tasks such as fake news detection.
- We propose to leverage the rich social and political context and model news documents as heterogeneous information networks for political perspective detection. Our approach is end-to-end, inductive and effectively framing the task of stance detection as a classification problem on graphs to incorporate external knowledge.
- We conduct extensive experiments to evaluate our method and competitive baselines. As a result, our method outperforms all state-of-the-art approaches. Further ablation studies bear out the necessity of external knowledge and the effectiveness of constructing a heterogeneous information network for political perspective detection.

**Related Work**

In this section, we briefly review related literature on political perspective detection and knowledge graph applications.

**Political Perspective Detection**

The problem of political perspective detection is generally studied in two different settings: social media and news media. For stance detection in social media, many works have studied the problem as a text classification task to identify stance in specific posts. They used techniques such as sentiment analysis (Jiang et al. 2011), Wang et al. 2017, n-gram language models (Mohammad et al. 2016) and different neural network architectures (Augenstein et al. 2016, Du et al. 2017, Xu et al. 2018). Other approaches explored the problem of identifying stance of social media users instead of individual tweets. Darwish et al. 2020 conducts user clustering to identify their stances. Magdy et al. 2016 tried to predict user perspectives on major events based on network dynamics and user interactions. Stefanov et al. 2020 adopted label propagation and proposed a semi-supervised approach.

For stance detection in news media, a supervised task is often formed to classify a news document into several perspective labels. Previous works have proposed to leverage semantic information in news articles for perspective detection, such as linguistic indicators (Field et al. 2018), bias features (Horne, Khedr, and Adali 2018) and deep language encoders (Li and Goldwasser 2021, Yang et al. 2016, Jiang et al. 2019). Later approaches have tried to enrich the textural content of news with graph structures of online communities that interact with different news outlets. Pan et al. 2016 study the problem of learning text representation with the help of social information. Li and Goldwasser 2019 supplement news articles with a network of Twitter users and their interaction with them. In this paper, we intend to focus on political perspective detection in news media and build on these works while also incorporating external knowledge of contemporary politics, since background knowledge is essential in how individuals identify political perspectives.

**Knowledge Graph Applications**

Knowledge graphs that represent diversified relations between entities have become popular in AI research. Many large-scale knowledge graphs have been constructed and a wide range of tasks and solutions are proposed, such as knowledge graph completion (Bordes et al. 2013, Sun et al. 2019) and modeling temporal knowledge graphs (Ma, Tresp, and Daxberger 2019, Liu et al. 2020b). Apart from being an effective data structure to represent real-world knowledge, knowledge graphs also serve as domain knowledge for a wide range of tasks, such as language representation learning (Liu et al. 2020a, He et al. 2019), question answering (Lin et al. 2019, Ding et al. 2019) and recommender systems (Wang et al. 2019, Xian et al. 2019). In this paper, we aim to leverage external knowledge in political perspective detection in the form of knowledge graphs.

**Knowledge Graph Construction**

We firstly build a knowledge graph of contemporary U.S. politics to serve as external knowledge for our political perspective detection model. We select important entities in the social and political context such as political ideologies, elected officials and political institutions. We then retrieve information from Wikipedia to determine relations between these entities and construct the knowledge graph with derived triples. A complete list of different nodes and relations are listed in Table 1. As a result, we construct a political knowledge graph with 1,701 entities and 10,703 triples. An overview of the knowledge graph is provided in Figure 2.
Table 1: List of entities and relations in our collected political knowledge graph.

| Entity Type         | Example                  | Relation Type | Example                                |
|---------------------|--------------------------|---------------|----------------------------------------|
| elected office      | the U.S. Senate          | affiliated_to | (Joe Biden, affiliated_to, Democratic Party) |
| time period         | 117th congress           | from          | (Ted Cruz, from, Texas)                |
| president           | Joe Biden                | appoint       | (Donald Trump, appoint, Amy Coney Barrett) |
| supreme court justice| Amy Coney Barrett        | overlap_with  | (Joe Biden, overlap_with, 117th congress) |
| senator             | Elizabeth Warren         | member_of     | (Dianne Feinstein, member_of, the U.S. Senate) |
| congressperson      | Nancy Pelosi             | strongly_favor| (Bernie Sanders, strongly_favor, liberal_values) |
| governor            | Ron DeSantis             | favor         | (Joe Cunningham, favor, liberal_values) |
| state               | Massachusetts            | neutral       | (Henry Cuellar, neutral, liberal_values) |
| political party     | Republican Party         | oppose        | (Lamar Alexander, oppose, liberal_values) |
| political ideology  | liberal values           | strongly_oppose| (Ted Cruz, strongly_oppose, liberal_values) |

Table 2: Three types of nodes in the HINs of news articles.

| Type            | #Nodes | Representation                                          |
|-----------------|--------|--------------------------------------------------------|
| document        | 1      | the news document as a whole                           |
| paragraph       | s      | s paragraphs in the document                           |
| entity          | n      | n entities in knowledge graph                          |

Embeddings (Li and Goldwasser 2021), LSTM (Yang et al. 2016) and BERT (Devlin et al. 2018), RoBERTa (Liu et al. 2019) is a pre-trained language model that could effectively encode text sequences, thus we adopt RoBERTa to encode news text at different levels. Specifically, let $D = \{d, p_1, \ldots, p_s\}$ be a news document of $s+1$ strings, where $d$ is the title of the article and $p_i$ is the $i$-th paragraph in the article. We encode the title and paragraphs as follows:

$$v_d = RoBERTa(d), \quad v^p_i = RoBERTa(p_i)$$

(2)

where $v_d$ is the representation vector of the title, $v^p_i$ is the representation of the $i$-th paragraph in the news article and $RoBERTa(\cdot)$ is the pre-trained language encoder.

**Graph Construction**

With the advent of graph neural networks, many works have used them for natural language processing tasks such as text classification (Yao, Mao, and Luo 2019), Zhang et al. 2020) and question answering (De Cao, Aziz, and Titov 2018). In this paper, we model news articles with heterogeneous information networks (HINs) to represent the diversified interactions between text and entities in the knowledge graph. Specifically, different types of nodes in the HINs are listed in Table 2. We transform the representation vector $v_d^d$, $v^p_i^d$ and $v^p_i^e$ with fully connected layers to obtain the initial features for these nodes on the HIN. Specifically, for document node and paragraph nodes:

1Political ideologies and individuals’ attitude towards them are adapted from aflcio.org/scorecard and heritageaction.com/scorecard
\[ x_0^{(0)} = \phi(W_S \cdot v^d + b_S), \quad x_i^{(0)} = \phi(W_S \cdot v^p + b_S) \]  
(3)

where \( W_S \) and \( b_S \) are learnable parameters and \( \phi \) is any non-linear function and we adopt Leaky-ReLU for \( \phi \) without further notice. Similarly, we use another fully connected layer to obtain initial features for entity nodes:

\[ x_j^{(0)} = \phi(W_E \cdot v^e + b_E) \]  
(4)

where \( W_E \) and \( b_E \) are learnable parameters.

After defining the nodes in the HIN, we connect them with three types of edges:

- **doc-para edge**: document node is connected with each paragraph node by doc-para edges. This ensures that every paragraph plays a role in analyzing the document.
- **para-para edge**: each paragraph node is connected with paragraphs that precede and follow it. These edges preserve the sequential flow of text in the news document.
- **para-ent edge**: each paragraph node is connected with its mentioned entities. We conduct coreference resolution \( [\text{Lee, He, and Zettlemoyer 2018}] \) to capture all entity mentions. These edges aim to incorporate external knowledge of the knowledge graph for perspective analysis.

We denote the three types of edges as \( O = \{ \text{doc–para, para–para, para–ent} \} \). We illustrate the HINs used to represent news articles in Figure 3.

**Learning and Optimization**

Our model uses gated relational graph convolutional networks (gated R-GCNs) to learn representations for the HIN and conduct political perspective detection on news articles. For the \( l \)-th layer of gated R-GCN, we firstly aggregate messages from neighbors as follows:

\[ u_i^{(l)} = \Theta_s \cdot x_i^{(l-1)} + \sum_{r \in O} \sum_{j \in N_r(i)} \frac{1}{|N_r(i)|} \Theta_r \cdot x_j^{(l-1)} \]  
(5)

where \( u_i^{(l)} \) is the hidden representation for the \( i \)-th node in the \( l \)-th layer, \( N_r(i) \) is node \( i \)'s neighborhood of relation \( r \), \( \Theta_s \) and \( \Theta_r \) are learnable parameters. We then calculate gate levels:

\[ a_i^{(l)} = \sigma(W_A \cdot [u_i^{(l)} \cdot x_i^{(l-1)}] + b_A) \]  
(6)

where \( \sigma(\cdot) \) is the sigmoid function, \( [\cdot, \cdot] \) denotes the concatenation operation, \( W_A \) and \( b_A \) are learnable parameters. We then apply the gate to \( u_i^{(l)} \) and \( x_i^{(l-1)} \):

\[ x_i^{(l)} = \tanh(u_i^{(l)}) \odot a_i^{(l)} + x_i^{(l-1)} \odot (1 - a_i^{(l)}) \]  
(7)

For the \( l \)-th gated R-GCN, we first aggregate messages from neighbors as follows:

\[ u_i^{(l)} = \Theta_s \cdot x_i^{(l-1)} + \sum_{r \in O} \sum_{j \in N_r(i)} \frac{1}{|N_r(i)|} \Theta_r \cdot x_j^{(l-1)} \]  
(8)

We use two methods to obtain representation for the entire graph \( v^g \):

- **document only**, where we use the representation of the document node to represent the entire graph:
  \[ v^g = x_0^{(L)} \]  
(9)

- **average paragraphs**, where we average representations of all paragraph nodes to represent the entire graph:
  \[ v^g = \frac{1}{s} \sum_{i=1}^{s} x_i^{(L)} \]  
(10)

We then transform \( v^g \) with a softmax layer to conduct political perspective detection:

\[ \hat{y} = \text{softmax}(W_O \cdot v^g + b_O) \]  
(11)

where \( \hat{y} \) is our model’s prediction, \( W_O \) and \( b_O \) are learnable parameters.

The loss function of our method is as follows:

\[ L = -\sum_{i=1}^{Y} \sum_{v \in \theta} y_i \log(\hat{y}_i) + \lambda \sum_{w \in \theta} w^2 \]  
(12)

where \( Y \) is the number of perspective labels, \( y_i \) is the annotation of news articles, \( \theta \) are all learnable parameters in our proposed model and \( \lambda \) is a hyperparameter.

**Experiments**

In this section, we compare our proposed solution with state-of-the-art methods on real-world political perspective detection datasets. We also study the effect of incorporating external knowledge, analyze the graph learning techniques in our method and present specific cases to better understand the mechanism and effectiveness of our proposed approach.

| Hyperparameter          | Value       |
|-------------------------|-------------|
| size of RoBERTa representation | 768         |
| size of TransE representation | 200         |
| size of gated R-GCN hidden state | 512         |
| optimizer               | Adam        |
| learning rate           | 10^{-3}     |
| batch size              | 16          |
| dropout                 | 0.5         |
| maximum epochs          | 50          |
| L2-regularization \( \lambda \) | 10^{-5}     |
| gated R-GCN layer count \( L \) | 2           |

Table 3: Implementation details and hyperparameter settings of our political perspective detection model.

Figure 3: Overview of our proposed heterogeneous information network structure to represent each news article.
Table 4: Statistics of the SemEval dataset.

|                          | # articles | avg. # word | avg. # paragraph | avg. # mentioned entity | avg. # entity mentions |
|--------------------------|------------|-------------|------------------|-------------------------|------------------------|
| # articles              | 645        |             |                  |                         |                        |
| # hyperpartisan         | 238        |             |                  |                         |                        |
| # not hyperpartisan     | 407        |             |                  |                         |                        |
| release year            | 2019       |             |                  |                         |                        |
| avg. # word             | 582.99     |             |                  |                         |                        |
| avg. # paragraph        | 16.16      |             |                  |                         |                        |
| avg. # mentioned entity | 20.75      |             |                  |                         |                        |
| avg. # entity mentions  | 42.70      |             |                  |                         |                        |

Table 5: Performance of our method and competitive baselines on the SemEval dataset. Entity and Knowledge indicate whether these models leverage entity mentions and external knowledge entailed in news articles.

| Model          | Entity | Knowledge | Accuracy |
|----------------|--------|-----------|----------|
| CNN_Glove      |        |           | 0.7963   |
| CNN_ELMo       |        |           | 0.8404   |
| HLSTM_Glove    |        |           | 0.8158   |
| HLSTM_ELMo     |        |           | 0.8328   |
| HLSTM_Emb     |        |           | 0.8171   |
| HLSTM_Output  |        |           | 0.8125   |
| BERT           | ✓      |           | 0.8341   |
| MEAN_ENT       | ✓      |           | 0.8451   |
| MEAN_REL       | ✓      |           | 0.8309   |
| MEAN_Emb      | ✓      |           | 0.8522   |
| OURS_DO        | ✓ ✓    |           | 0.8617   |
| OURS_AP        | ✓ ✓    |           | 0.9071   |

Dataset

While existing works in political perspective detection have provided many annotated datasets, a vast majority of them focus on detecting stances on social media rather than in news media. We make use of the news-based dataset SemEval, which is the training dataset from the SemEval 2019 Task 4: Hyperpartisan News Detection (Kiesel et al. 2019). The task aims to detect hyperpartisan perspectives in news articles which are manually annotated with binary labels. The statistics of the dataset is presented in Table 4. We follow the same 10-fold cross validation as in (Jiang et al. 2019) and (Li and Goldwasser 2021) so that our results are directly comparable with previous works.

Baselines

We compare our method with competitive baselines, which include both outstanding solutions in the SemEval 2019 contest and state-of-the-art proposals that come after the contest.

- **CNN** is the first place solution from the SemEval 2019 Task 4 contest (Kiesel et al. 2019). It uses Glove (Pennington, Socher, and Manning 2014) (CNN_Glove) and ELMo (Peters et al. 2018) (CNN_ELMo) word embeddings and convolutional layers for text representation learning and stance prediction. (Jiang et al. 2019)

- **HLSTM** stands for hierarchical LSTM (Yang et al. 2016). It encodes news articles with both word-level and sentence-level bidirectional LSTM and aggregate hidden states with self attention. (Li and Goldwasser 2019)

- **HLSTM_Emb** and **HLSTM_Output** are two proposals that leverage entity mentions. They use Wikipedia2Vec (Yamada et al. 2018) and BERT-inspired masked entity models to learn entity representations and concatenate them with word embeddings (HLSTM_Emb) or document representation (HLSTM_Output). (Li and Goldwasser 2021)

- **BERT** is a language model based on deep bidirectional transformers. BERT uses masked language model and next sentence prediction tasks to pre-train and could be fine-tuned on downstream NLP tasks. (Devlin et al. 2018)

- **MEAN** is a perspective detection method that learns representation for entities and relations and inject them into a hierarchical LSTM structure with multi-head entity-aware attention. Besides using entity representations (MEAN_ENT) and relation representations (MEAN_REL), an ensemble model (MEAN_Emb) that jointly leverages entity and relation mentions is also presented. (Li and Goldwasser 2021)

Implementation

We use pytorch (Paszke et al. 2019), pytorch lightning (Falcon 2019), torch geometric (Fey and Lenssen 2019) and the transformers library (Wolf et al. 2020) for an efficient implementation of our proposed political perspective detection model. We present our hyperparameter settings in Table 3 to facilitate reproduction. Our implementation is trained on a Titan X GPU with 12GB memory. We are committed to releasing our collected political knowledge graph and all implemented codes upon acceptance.

Experiment Results

The performance of our method and competitive baselines on the SemEval dataset are presented in Table 5. Ours_DO
and Ours\_AP represent our methods in the document only or average paragraphs settings respectively and we also report the standard deviation of detection accuracy across 10 folds. Table 5 demonstrates that:

- Our model, especially Ours\_AP, consistently outperforms all competitive baselines, including the state-of-the-art method MEAN (Li and Goldwasser 2021).
- Two methods that leverage entity mentions, MEAN and ours, outperform other baselines. Our method further incorporates external knowledge in political knowledge graphs and outperforms MEAN, thus the strategy of leveraging external knowledge is proved to be effective.
- For methods that use word embeddings, ELMo (Peters et al. 2018) often outperforms their Glove (Pennington, Socher, and Manning 2014) counterparts. This suggests that text analysis is essential in political perspective detection, where we adopt RoBERTa (Liu et al. 2019) as an up-to-date alternative to ELMo and Glove.

These observations substantiate the general effectiveness of our proposed model. We further examine how external knowledge and graph learning contribute to our model’s outstanding performance. We use OURS\_AP for these experiments without further notice.

**External Knowledge Study**

Our proposed model focuses on incorporating external knowledge of social and political context in the task of political perspective detection. We construct a knowledge graph about contemporary U.S. politics, learn representations for entities and relations and inject them into a heterogeneous information network. To examine whether real-world news articles actually contain entities in the political knowledge graph, we analyze news articles in the SemEval dataset and present our findings in Figure 4. Figure 4 illustrates that most news articles contain a considerable amount of entities in the political knowledge graph. In addition, news articles with the largest amount of entity mentions are generally correctly classified, suggesting that the inclusion of external knowledge boost model performance.

To further examine the effect of external knowledge, we randomly remove para-ent edges in HINs and present performance in Figure 5. It is demonstrated that our model reaches the best performance with full external knowledge and accuracy slightly decreases with reduced knowledge. Performance drops sharply when all external knowledge and entities are removed from the HIN, which suggests that as little as 10% of para-ent edges would help our model better understand the context of the news article.

As a result, incorporating external knowledge is essential in political perspective detection and our proposed approach of leveraging political knowledge graphs is effective in boosting task performance.

**Graph Learning Study**

Another contribution of our proposed model is to form the task of political perspective detection as classification on graphs. We construct heterogeneous information networks that contain both semantic contents of the news article and external knowledge of mentioned entities. To examine whether our proposed graph learning schema is essential in the model’s improved performance, we conduct a series of ablation study regarding the graph structure, node features and graph operators we adopted.

**Graph Structure** We propose to construct HINs to represent news articles, which are three types of nodes connected by three types of edges. In order to examine whether the heterogeneous graph structure is effective, we remove three types of edges and present our model’s performance in Figure 6. It is illustrated that removing any type of edge in the HIN would result in a considerable performance drop, which suggests that the HIN structure is essential in the model’s outstanding performance.
Table 6: Example of political perspective detection by our proposed method.

| Bias   | Gold | Ours | #Entity | #Paragraph | Title of the News Article |
|--------|------|------|---------|------------|---------------------------|
| lean right | True | True | 37      | 26         | THEY DON’T CALL IT ‘THE GREAT TWEET OF CHINA’ |
| lean right | True | True | 29      | 26         | Hanson: Progressive attacks on Trump are backfiring |
| lean left | True | True | 48      | 38         | When the Parades Are Over, Who Stands With Unions? |
| lean left | True | False | 13     | 24         | The G.O.P.’s Radical Supreme Court Talk |
| center | False | True | 5       | 6          | Why California is closer to becoming a sanctuary state |
| center | False | False | 26     | 33         | UK urged to follow Donald Trump’s vow to ‘deport illegal immigrants’ |

Table 7: Ablation study substituting RoBERTa and TransE initial features for nodes with trainable parameters.

| Ablation Settings                      | Accuracy       |
|----------------------------------------|----------------|
| all initial features                   | 0.9071 ± 0.0390|
| no RoBERTa features                    | 0.6341 ± 0.0371|
| no TransE features                     | 0.8776 ± 0.0405|
| no RoBERTa and TransE features         | 0.6326 ± 0.0372|

Node Features: Upon obtaining the graph structure of news articles, we adopt RoBERTa [Liu et al. 2019] and TransE [Bordes et al. 2013] to encode titles, paragraphs and entities in the knowledge graph. We then use these learnt representations as initial features for three types of nodes in the graph. To examine whether these initial features play an important role in our model’s performance, we substitute semantic and entity initial features with trainable parameters. The performance under these settings is presented in Table [2]. It is demonstrated that the RoBERTa features is a dominant factor in the model’s performance, which suggests that the task of political perspective detection still relies heavily on text analysis. TransE features are also beneficial in that they contribute to a 0.0305 increase in model accuracy.

Graph Operators: For the HINs we built to represent news documents, we adopt gated R-GCNs to propagate node messages and learn representations. Besides gated R-GCN, there are many graph operators on both homogeneous [Kipf and Welling 2016; Velickovic et al. 2017; Hamilton, Ying, and Leskovec 2017] and heterogeneous graphs [Schlichtkrull et al. 2018]. To validate the effectiveness of the graph heterogeneity and our adopted gated R-GCN, we substitute gated R-GCN layers with other graph operators and present their performances in Figure [7]. It is illustrated that R-GCN outperforms GCN, which indicates that heterogeneous graphs for news articles are better than its homogeneous counterparts. Apart from that, our adopted gated R-GCN performs best among all graph operators, validating our model’s design choices.

To sum up, the graph structure, node features and graph operators adopted in our model are effective and essential to the method’s outstanding performance.

Figure 7: Ablation study substituting gated R-GCN layers with other graph neural networks.

Case Study
To further understand the inclusion of external knowledge and how our method detects political perspective in news media, we select six representative news articles in the SemEval dataset and analyze their mentioned entities and model predictions. We present these cases in Table [6]. These examples demonstrate the correlation between external knowledge and stance detection performance, which suggests more external knowledge should be leveraged to identify subtle biases in news articles.

Conclusion and Future Work
Political perspective detection has become an important task due to the popularity of social media and the increasingly polarized political ideologies. We collected and publicized a knowledge graph of contemporary US politics to serve as external knowledge in political perspective detection. We proposed to incorporate external knowledge of social and political context, construct a heterogeneous information network to represent news articles and form the task of stance detection as classification on graphs. We conducted extensive experiments to demonstrate the superiority of our method in comparison to competitive baselines. Further ablation studies show that our proposals to leverage external knowledge, construct heterogeneous information networks to represent news articles and form the task as graph classification are essential in the model’s outstanding performance.

In the future, we will extend the ideas of using graphs to represent documents and leveraging domain knowledge to tasks such as fake news detection and Twitter bot detection.
References

[Augenstein et al. 2016] Augenstein, I.; Rocktäschel, T.; Vlachos, A.; and Bontcheva, K. 2016. Stance detection with bidirectional conditional encoding. arXiv preprint arXiv:1606.05464.

[Bordes et al. 2013] Bordes, A.; Usunier, N.; Garcia-Duran, A.; Weston, J.; and Yakhnenko, O. 2013. Translating embeddings for modeling multi-relational data. Advances in neural information processing systems 26.

[Darwish et al. 2020] Darwish, K.; Stefanov, P.; Aupetit, M.; and Nakov, P. 2020. Unsupervised user stance detection on twitter. In Proceedings of the International AAAI Conference on Web and Social Media, volume 14, 141–152.

[Du et al. 2017] Du, J.; Xu, R.; He, Y.; and Gui, L. 2017. Stance classification with target-specific neural attention networks. International Joint Conferences on Artificial Intelligence.

[Ma, Tresp, and Daxberger 2019] Ma, Y.; Tresp, V.; and Daxberger, E. A. 2019. Embedding models for episodic knowledge into pre-trained language models. arXiv preprint arXiv:1907.11692.

[Li and Goldwasser 2021] Li, C., and Goldwasser, D. 2021. Mean: Multi-head entity aware attention network for political perspective detection in news media. In Proceedings of the Fourth Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda, 66–75.

[Lin et al. 2019] Lin, B. Y.; Chen, X.; Chen, J.; and Ren, X. 2019. Kagnet: Knowledge-aware graph networks for commonsense reasoning. arXiv preprint arXiv:1909.02151.

[Liu et al. 2020b] Liu, Y.; Hua, W.; Xin, K.; and Zhou, X. 2020b. Context-aware temporal knowledge graph embedding. In International Conference on Web Information Systems Engineering, 583–598. Springer.

[Ma, Tresp, and Daxberger 2019] Ma, Y.; Tresp, V.; and Daxberger, E. A. 2019. Embedding models for episodic knowledge graphs. Journal of Web Semantics 59:100490.

[Magdy et al. 2016] Magdy, W.; Darwish, K.; Abokhodair, N.; Rahimi, A.; and Baldwin, T. 2016. #isisisnotislam or# deportallmuslims? predicting unspoken views. In Proceedings of the 8th ACM Conference on Web Science, 95–106.
[Mohammad et al. 2016] Mohammad, S.; Kiritchenko, S.; Sobhani, P.; Zhu, X.; and Cherry, C. 2016. A dataset for detecting stance in tweets. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), 3945–3952.

[Pan et al. 2016] Pan, S.; Wu, J.; Zhu, X.; Zhang, C.; and Wang, Y. 2016. Tri-party deep network representation. Network 11(9):12.

[Paszke et al. 2019] Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. 2019. Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems 32:8026–8037.

[Pennington, Socher, and Manning 2014] Pennington, J.; Socher, R.; and Manning, C. D. 2014. Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 1532–1543.

[Peters et al. 2018] Peters, M. E.; Neumann, M.; Iyyer, M.; Gardner, M.; Clark, C.; Lee, K.; and Zettlemoyer, L. 2018. Deep contextualized word representations. arXiv preprint arXiv:1802.05365.

[Schlichtkrull et al. 2018] Schlichtkrull, M.; Kipf, T. N.; Bloem, P.; Van Den Berg, R.; Titov, I.; and Welling, M. 2018. Modeling relational data with graph convolutional networks. In European semantic web conference, 593–607. Springer.

[Stefanov et al. 2020] Stefanov, P.; Darwish, K.; Atanasov, A.; and Nakov, P. 2020. Predicting the topical stance and political leaning of media using tweets. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 527–537.

[Sun et al. 2019] Sun, Z.; Deng, Z.-H.; Nie, J.-Y.; and Tang, J. 2019. Rotate: Knowledge graph embedding by relational rotation in complex space. arXiv preprint arXiv:1902.10197.

[Tang et al. 2020] Tang, Z.; Shen, Y.; Ma, X.; Xu, W.; Yu, J.; and Lu, W. 2020. Multi-hop reading comprehension across documents with path-based graph convolutional network. arXiv preprint arXiv:2006.06478.

[Veličković et al. 2017] Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; and Bengio, Y. 2017. Graph attention networks. arXiv preprint arXiv:1710.10903.

[Wang et al. 2017] Wang, B.; Liakata, M.; Zubiaga, A.; and Procter, R. 2017. Tdparse: Multi-target-specific sentiment recognition on twitter. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers, 483–493.

[Wang et al. 2019] Wang, X.; Wang, D.; Xu, C.; He, X.; Cao, Y.; and Chua, T.-S. 2019. Explainable reasoning over knowledge graphs for recommendation. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 33, 5329–5336.

[Wolf et al. 2020] Wolf, T.; Debut, L.; Sanh, V.; Chaumond, J.; Delangue, C.; Moi, A.; Cistac, P.; Rault, T.; Louf, R.; Funtowicz, M.; Davison, J.; Shleifer, S.; von Platen, P.; Ma, C.; Jernite, Y.; Plu, J.; Xu, C.; Scao, T. L.; Gugger, S.; Drame, M.; Lhoest, Q.; and Rush, A. M. 2020. Transformers: State-of-the-art natural language processing. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, 38–45. Online: Association for Computational Linguistics.

[Xian et al. 2019] Xian, Y.; Fu, Z.; Muthukrishnan, S.; De Melo, G.; and Zhang, Y. 2019. Reinforcement knowledge graph reasoning for explainable recommendation. In Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval, 285–294.

[Xu et al. 2018] Xu, C.; Paris, C.; Nepal, S.; and Sparks, R. 2018. Cross-target stance classification with self-attention networks. arXiv preprint arXiv:1805.06593.

[Yamada et al. 2018] Yamada, I.; Asai, A.; Sakuma, J.; Shindo, H.; Takeda, H.; Takefuji, Y.; and Matsumoto, Y. 2018. Wikipedia2vec: An efficient toolkit for learning and visualizing the embeddings of words and entities from wikipedia. arXiv preprint arXiv:1812.06280.

[Yang et al. 2016] Yang, Z.; Yang, D.; Dyer, C.; He, X.; Smola, A.; and Hovy, E. 2016. Hierarchical attention networks for document classification. 1480–1489.

[Yao, Mao, and Luo 2019] Yao, L.; Mao, C.; and Luo, Y. 2019. Graph convolutional networks for text classification. In Proceedings of the AAAI conference on artificial intelligence, volume 33, 7370–7377.

[Zhang et al. 2020] Zhang, Y.; Yu, X.; Cui, Z.; Wu, S.; Wen, Z.; and Wang, L. 2020. Every document owns its structure: Inductive text classification via graph neural networks. arXiv preprint arXiv:2004.13826.