KCD: Knowledge Walks and Textual Cues Enhanced Political Perspective Detection in News Media

Wenqian Zhang[♠*] **Shangbin Feng**[♠*] **Zilong Chen**[♢*]
Zhenyu Lei[♢] **Jundong Li**[♡] **Minnan Luo**[♠†]
School of Computer Science and Technology, Xi'an Jiaotong University[♠]
School of Science, Xi'an Jiaotong University[♢], University of Virginia[♡]
{2194510944,wind_binteng,luoyangczl,Fischer}@stu.xjtu.edu.cn
jundong@virginia.edu minnluo@xjtu.edu.cn

Abstract

Political perspective detection has become an increasingly important task that can help combat echo chambers and political polarization. Previous approaches generally focus on leveraging textual content to identify stances, while they fail to reason with background knowledge or leverage the rich semantic and syntactic textual labels in news articles. In light of these limitations, we propose KCD, a political perspective detection approach to enable multi-hop knowledge reasoning and incorporate textual cues as paragraph-level labels. Specifically, we firstly generate random walks on external knowledge graphs and infuse them with news text representations. We then construct a heterogeneous information network to jointly model news content as well as semantic, syntactic and entity cues in news articles. Finally, we adopt relational graph neural networks for graph-level representation learning and conduct political perspective detection. Extensive experiments demonstrate that our approach outperforms state-of-the-art methods on two benchmark datasets. We further examine the effect of knowledge walks and textual cues and how they contribute to our approach's data efficiency.

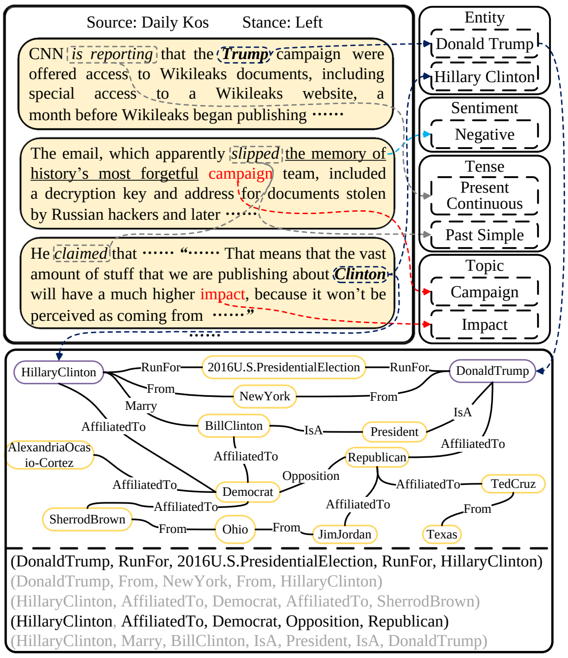

Figure 1: Multi-hop knowledge reasoning and implicit textual indicators that facilitate perspective detection.

1 Introduction

Political perspective detection aims to identify ideological stances of textual data such as social media posts and news articles. Previous approaches generally leverage the textual content of news articles with various text modeling techniques to identify political stances. Those works (Jiang et al., 2019; Li and Goldwasser, 2019, 2021; Feng et al., 2021a) leveraged diversified text models, such as recurrent neural networks (Yang et al., 2016), word embedding techniques (Pennington et al., 2014; Peters et al., 2018), convolutional neural networks (Jiang et al., 2019), and pre-trained language models (Devlin et al., 2019; Liu et al., 2019), to encode news paragraphs and classify them into different perspective labels. Later approaches incorporate information sources beyond text to facilitate argument mining and boost task performance. News discussion on social networks (Li and Goldwasser, 2019), social and linguistic information about news articles (Li and Goldwasser, 2021), media sources and information (Baly et al., 2020) as well as external knowledge from knowledge graphs (Feng et al., 2021a) are introduced in the task of political perspective detection and achieve better performance.

Although these methods attempted to leverage more than news content, they fail to present a framework capable of reasoning with background knowledge and leveraging implicit semantic and syntactic indicators such as sentiment and tense of news articles. For example, Figure 1 presents a typical news

[*]These authors contributed equally to this work.
[†]Corresponding author.

Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 4129 - 4140
July 10-15, 2022 ©2022 Association for Computational Linguistics

article from Daily Kos[1]. This article discusses remarks from the Trump campaign team about Wikileaks and its effect on Hillary Clinton's bid for president. Individuals often rely on the multi-hop reasoning that Clinton and Trump are from opposite political parties and run against each other to inform their perspective analysis process. Besides, the negative sentiment expressed in satiric tones and the quotation of Trump campaign staff also give away the author's denial and left-leaning perspective. That being said, knowledge reasoning and implicit textual indicators are essential in the news bias detection process.

In light of these limitations, we propose a political perspective detection framework **KCD** (**K**nowledge Walks and Textual **C**ues Enhanced Political Perspective **D**etection). Specifically, KCD generates multi-hop knowledge walks, aggregates them based on semantic relevance and incorporates them in textual representations with multi-head attention. KCD then constructs a heterogeneous information network to jointly model knowledge-enriched news content and diversified textual cues as paragraph-level labels. Finally, KCD learns graph representations with relational graph neural networks and conduct perspective detection with different aggregation strategies. Our main contributions are summarized as follows:

- We propose knowledge walks, a strategy to incorporate multi-hop knowledge reasoning in textual representations for knowledge-aware political perspective detection.
- We propose to construct a heterogeneous information network to represent news articles, which jointly models knowledge-enriched news content and implicit textual cues in news articles.
- Extensive experiments demonstrate that our approach consistently outperforms state-of-the-art methods on two widely adopted benchmarks. Further analysis bears out the necessity of knowledge walks and textual cues in our approach.

2 Related Work

2.1 Political Perspective Detection

Political perspective detection aims to identify the ideological stances of news articles, which is widely studied to help strengthen the online information landscape (Li and Goldwasser, 2019) and mitigate ideological echo chambers (Li and Goldwasser, 2021; Feng et al., 2021a). Early approaches leverage text analysis techniques for bias detection, such as sentiment analysis (Jiang et al., 2011; Wang et al., 2017), bias feature extraction (Horne et al., 2018), word embeddings (Jiang et al., 2019; Li and Goldwasser, 2019), and different neural network architectures (Augenstein et al., 2016; Du et al., 2017; Xu et al., 2018; Yang et al., 2016; Jiang et al., 2019; Feng et al., 2021b; Li and Goldwasser, 2021; Feng et al., 2021a). In addition to textual content of news articles, social media users also become the focus of perspective detection research (Bel-Enguix et al., 2021). User interactions (Magdy et al., 2016), user clustering (Darwish et al., 2020), and label propagation (Stefanov et al., 2020) are leveraged to identify the ideological preferences on social media. Fusing both news text and social network analysis directions, Li and Goldwasser (2019) propose to enrich news text with the content and structure of social media discussions about these news articles. Recent state-of-the-art approaches chart a new path by incorporating social and political external knowledge into stance detection. Baly et al. (2020) propose adversarial media adaptation and leverage source background knowledge for political perspective detection. Li and Goldwasser (2021) combine language encoders with pre-training tasks of social and linguistic information. Feng et al. (2021a) propose to construct and leverage political knowledge graphs as domain-specific external knowledge. In this paper, we build on these works to examine and explore the effect of multi-hop knowledge reasoning and diversified textual cues in the task of political perspective detection.

2.2 Knowledge Graph in NLP

Knowledge graphs (KGs) are effective representations of real-world entities, relations, and knowledge. Generic (Fellbaum, 2010; Pellissier Tanon et al., 2020; Bollacker et al., 2008; Speer et al., 2017) and domain-specific KGs (Feng et al., 2021a; Chang et al., 2020) are widely adopted in NLP tasks as external knowledge sources. These approaches could mainly be categorized into feature extraction, language model and graph-based methods. For feature extraction approaches, KG embedding technique TransE (Bordes et al., 2013) is leveraged to learn features for knowledge injecton (Ostendorff et al., 2019; Hu et al., 2021). For language model approaches, the adapter architecture is leveraged

[1] https://www.dailykos.com/

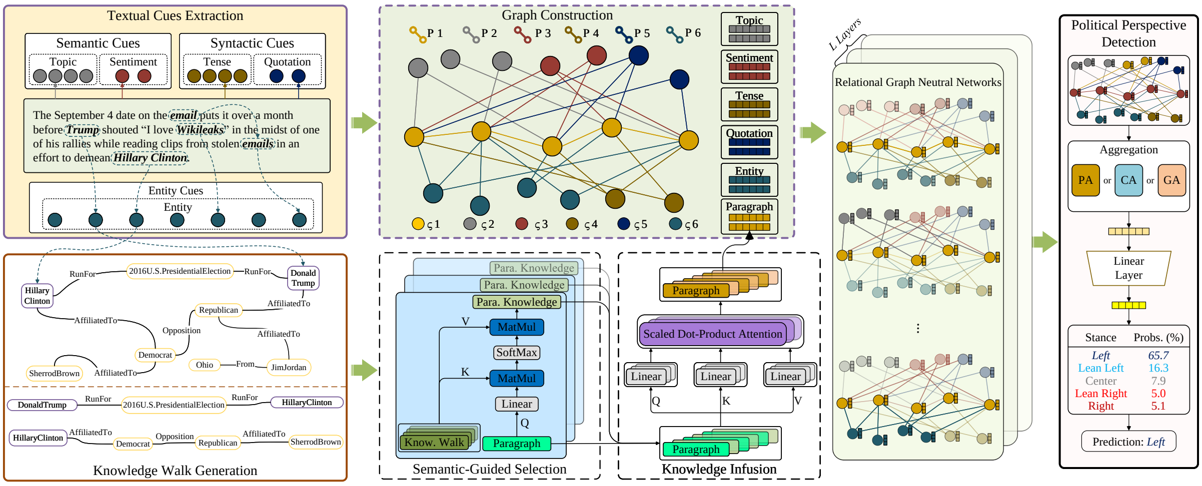

Figure 2: Overview of our proposed framework KCD.

to fine-tune on KG-related tasks (Majewska et al., 2021; Meng et al., 2021; Wei et al., 2021). In addition, Wang et al. (2021) propose a unified model to combine knowledge embedding with language representation pre-training. For graph-based approaches, KG entities and relations are injected into graphs and heterogeneous information networks (Hu et al., 2021; Feng et al., 2021a; Lu et al., 2021). Graph neural networks are then adopted to learn knowledge-aware text representations. In this paper, we propose knowledge walk, a novel strategy to infuse multi-hop knowledge reasoning into language representations and apply them in political perspective detection.

3 Methodology

Figure 2 presents an overview of our proposed political perspective detection framework **KCD** (**K**nowledge Walks and Textual **C**ues Enhanced Political Perspective **D**etection). We firstly generate knowledge walks on the external knowledge graph. These knowledge walks are then selected based on semantic relevance and injected into textual representations with multi-head attention. We then construct a heterogeneous information network to jointly model knowledge-enriched news content and diversified textual cues as paragraph-level labels and supernodes. Finally, we adopt relational graph neural networks and different aggregation strategies to learn graph-level representation and conduct political perspective detection.

3.1 Knowledge Walks and Infusion

We firstly propose the novel strategy of knowledge walks and combine them with textual representations to enable multi-hop knowledge reasoning. We partition an n-paragraph news document into different paragraphs and denote them as $\mathcal{S} = \{s_1, ..., s_n\}$. We encode each paragraph by averaging the the embeddings of words from pre-trained RoBERTa (Liu et al., 2019):

$$v_i^s = RoBERTa(s_i), \quad 1 \leq i \leq n \qquad (1)$$

We use a political knowledge graph[2] as external knowledge for perspective detection. Let the i-th triple in the knowledge graph be (e_{ih}, r_i, e_{it}), where e_{ih} and e_{it} denote the head and tail entity and r_i represents the relation of the i-th triple.

3.1.1 Knowledge Walk Generation

We firstly use TagMe (Ferragina and Scaiella, 2012) to identify mentioned KG entities in each paragraph s_i. For each mentioned entity, we use it as the starting point $e_{(0)}$ in a K-hop knowledge walk:

$$kw_i = \{e_{(0)}, r_{0,1}, e_{(1)}, ..., r_{K-1,K}, e_{(K)}\} \qquad (2)$$

where $e_{(i-1)}$ and $r_{i-1,i}$ denote the i-th triple's head entity and relation. Specifically, a knowledge walk is generated by adopting biased random walk of length K starting from $e_{(0)}$. The conditional probability of arriving at $e_{(i)}$ from $e_{(i-1)}$ through $r_{i-1,i}$ is formulated as

$$P(e_{(i)}|e_{(i-1)}, r_{i-1,i}) = \frac{exp(p(r_{i-1,i}))}{\sum_{j=1}^{|N_r(i-1)|} exp(p(r_j))} \qquad (3)$$

[2]https://github.com/BunsenFeng/news_stance_detection

where $N_r(i-1)$ denotes the neighboring relations of $e_{(i-1)}$, $p(r)$ is the importance score of KG relation r, which could be tuned by domain experts for human-in-the-loop knowledge walk generation. In this way, we generate multiple knowledge walks for each paragraph based on its mentioned entities, which models the multi-hop reasoning process with external knowledge.

3.1.2 Semantic-Guided Selection

After obtaining multiple knowledge walks for a single news paragraph, we propose a selection and aggregation process guided by textual content to differentiate essential knowledge walks from the irrelevant ones. We firstly transform each knowledge walk kw_i into a sentence t_i by concatenating the textual description of entities and relations. We then encode the knowledge walk sentence t_i with pre-trained RoBERTa (Liu et al., 2019):

$$v_i^k = RoBERTa(t_i) \tag{4}$$

Suppose a total of m knowledge walks $\{kw_{i,j}\}_{j=1}^m$ are generated for paragraph s_i, we then aggregate their knowledge walk sentence embeddings $\{v_{i,j}^k\}_{j=1}^m$ as follows:

$$v_i^p = \sum_{j=1}^m \frac{exp(\alpha \cdot v_{i,j}^k)}{\sum_{q=1}^m exp(\alpha \cdot v_{i,q}^k)} v_{i,j}^k \tag{5}$$

where α denotes the learnable attention vector guided by paragraph semantics:

$$\alpha = \phi(W_a v_i^s + b_a) \tag{6}$$

where W_a and b_a are learnable parameters of the attention module and we use Leaky-ReLU for ϕ. v_i^s is the sentence embedding from equation 1. In this way, we aggregate m knowledge walks based on semantic relevance to the paragraph to filter and retain important knowledge reasoning paths.

3.1.3 Knowledge Infusion

After representing multi-hop knowledge reasoning for paragraph s_i with v_i^p, we conduct document-wise multi-head self-attention to infuse knowledge walks into textual representations v_i^s. We concatenate knowledge walk and text representations:

$$T = concat([v_1^s, v_1^p, ..., v_n^s, v_n^p]) \tag{7}$$

where T is the input for multi-head self-attention:

$$\tilde{T} = MultiHead(Q, K, V) \tag{8}$$

where $Q = K = V = T$ and the output $\tilde{T} = concat([\tilde{v_1^s}, \tilde{v_1^p}, ..., \tilde{v_n^s}, \tilde{v_n^p}])$. In this way, we obtain language representations of news paragraphs $\{\tilde{v_i^s}\}_{i=1}^n$, which jointly models textual content and related multi-hop knowledge reasoning paths.

3.2 Textual Cues and Graph Construction

We construct a heterogeneous information network (HIN) as in Figure 2 "Graph Construction" to jointly represent knowledge-enriched news content and diversified textual cues in news articles. Specifically, we use paragraph nodes to represent the news content and connect them with different paragraph-level labels with heterogeneous edges. Firstly, for paragraph nodes:

$\mathcal{V}1$ and $\mathcal{R}1$: Paragraph Nodes We use one node in $\mathcal{V}1$ to represent each paragraph in the news article to partition the entire document and allow fine-grained analysis. We adopt the knowledge-enriched representations $\{\tilde{v_i^s}\}_{i=1}^n$ in Section 3.1 as initial node features for $\mathcal{V}1$. We then use relation $\mathcal{R}1$ to connect adjacent paragraphs to preserve the original flow of the news article.

3.2.1 Semantic Cues

We further analyze the topic and sentiment of news paragraphs, extract paragraph-level labels and inject them into our news HIN structure.

$\mathcal{V}2$ and $\mathcal{R}2$: Topic Cues The topics and frequent topic switching in news articles often give away the stance and argument of authors. We train LDA to extract the topics in each political perspective detection corpus and use one node to represent each topic. We then encode the topic text with pre-trained RoBERTa as node attributes. We then use $\mathcal{R}2$ to connect each paragraph node in $\mathcal{V}1$ with its affiliated topic node in $\mathcal{V}2$ with the help of BertTopic (Grootendorst, 2020).

$\mathcal{V}3$ and $\mathcal{R}3$: Sentiment Cues The sentiment of news articles signal the authors' approval or denial, which helps identify their stances towards individuals and issues. We use two nodes to represent positive and negative sentiment and we make their node attributes learnable. We then conduct sentiment analysis (Wolf et al., 2020) to identify paragraph sentiment and use $\mathcal{R}3$ to connect $\mathcal{V}1$ with their corresponding sentiment nodes in $\mathcal{V}3$.

3.2.2 Syntactic Cues

Apart from semantic cues, syntactic information in news articles also contribute to the perspective analysis process (Dutta et al., 2022). In light of this, we analyze the tense of news paragraphs and whether it contains direct quotation and use them as paragraph-level labels in our constructed HIN.

$\mathcal{V}4$ and $\mathcal{R}4$: Tense Cues The tense of news paragraphs helps separate facts from opinions. For example, simple past tense often indicates factual statements while simple future tense suggests opinions and projections that might not be factual. We use 17 nodes in $\mathcal{V}4$ to represent 17 possible tenses in our constructed news HIN. We use NLTK (Bird et al., 2009) to extract paragraph tenses and use $\mathcal{R}4$ to connect paragraph nodes in $\mathcal{V}1$ with $\mathcal{V}4$.

$\mathcal{V}5$ and $\mathcal{R}5$: Quotation Cues It is common for authors to directly quote others' words in news articles, which helps to identify the basis of the author's argument. We use two nodes to differentiate between whether a news paragraph quotes someone or not. Specifically, we identify quotation marks in news paragraphs and use $\mathcal{R}6$ to connect $\mathcal{V}1$ with $\mathcal{V}6$ based on whether direct quotation is detected.

3.2.3 Entity Cues

$\mathcal{V}6$ and $\mathcal{R}6$: Entity Cues We follow previous works (Feng et al., 2021a; Hu et al., 2021) to use one node to represent each entity in the external knowledge graph. We adopt TransE (Bordes et al., 2013) to learn knowledge graph embeddings and use them as initial node features for $\mathcal{V}6$. We then adopt entity linking tool TagMe (Ferragina and Scaiella, 2012) to align news paragraphs with their mentioned entities and use $\mathcal{R}6$ to connect $\mathcal{V}1$ with $\mathcal{V}6$ correspondingly.

In this way, we obtain a heterogeneous information network for news articles that jointly models knowledge-enriched news content and diversified textual cues in news articles. Our approach could be similarly extended to other textual cues and paragraph-level labels that would be helpful in political perspective detection and related tasks.

3.3 Learning and Optimization

Upon obtaining the news HINs, we adopt relational graph neural networks for representation learning and conduct political perspective detection as graph-level classification. Specifically, we follow Feng et al. (2021a) and use gated R-GCN to ensure a fair comparison and highlight the effectiveness of knowledge walks and textual cues. After L layers of gated R-GCN, we denote the learned node representations as $\overline{v}$ and obtain graph-level representation v_g with three different aggregation strategies: Paragraph Average (PA), Cue Average (CA) and Global Average (GA):

$$v_g = \begin{cases} \frac{1}{|\mathcal{V}1|}\sum_{v\in\mathcal{V}1}\overline{v} & \text{if Paragraph Average;} \\ \frac{1}{|\mathcal{V}-\mathcal{V}1|}\sum_{v\notin\mathcal{V}1}\overline{v} & \text{if Cue Average;} \\ \frac{1}{|\mathcal{V}|}\sum_{v\in\mathcal{V}}\overline{v} & \text{if Global Average.} \end{cases} \tag{9}$$

where $\mathcal{V} = \bigcup_{i=1}^{6}\mathcal{V}i$ represents the set of all nodes in our HIN. We then transform the graph-level representation v_g with a softmax layer and classify news articles into perspective labels:

$$\hat{y} = softmax(W_o \cdot v_g + b_o) \tag{10}$$

where W_o and b_o are learnable parameters and $\hat{y}$ is our model's prediction. We optimize the end-to-end process with cross entropy loss and L_2 regularization.

Hyperparameter	Value
GNN input size	768
GNN hidden size	512
GNN layer L	2
# epoch	150
batch size	16
dropout	0.6
# knowledge walk	30,114
$p(r)$ in Equ. (3)	constant c
# head in Equ. (8)	SE: 8, AS: 32
λ in Equ. (11)	1e-4
learning rate	1e-3
lr_scheduler_patience	20
lr_scheduler_step	0.1
# early stop epoch	40
Optimizer	Adam

Table 1: Hyperparameter settings of KCD. SE and AS denote the datasets SemEval and Allsides.

4 Experiments

4.1 Dataset

We make use of two real-world political perspective detection datasets SemEval (Kiesel et al., 2019) and Allsides (Li and Goldwasser, 2019), which are widely adopted in various previous works (Li and Goldwasser, 2019, 2021; Feng et al., 2021a). We follow the same evaluation settings as in previous works so that our results are directly comparable. Section B in the appendix provides more dataset details to facilitate reproduction.

4.2 Baselines

We compare KCD with the following competitive baselines and state-of-the-art methods:

- **CNN** (Jiang et al., 2019) is the first-place solution in the SemEval 2019 Task 4 contest (Kiesel et al., 2019). It combines convolutional neural networks with Glove (Jiang et al., 2019) and ELMo (Peters et al., 2018) for political perspective detection on the SemEval dataset.

- **HLSTM** (Yang et al., 2016) is short for hierarchical long short-term memory networks. Li and Goldwasser (2019) uses HLSTMs and different word embeddings for news bias detection.

- **HLSTM_Embed** and **HLSTM_Output** (Li and Goldwasser, 2021) leverage entity information with masked entity models in addition to news content for political perspective detection.

- **Word2Vec** (Mikolov et al., 2013), **GloVe** (Pennington et al., 2014), **ELMo** (Peters et al., 2018), pre-trained **BERT** (Devlin et al., 2019) and **RoBERTa** (Liu et al., 2019) are leveraged by Feng et al. (2021a) as textual features and political perspective detection is further conducted with two fully connected layers.

- **MAN** (Li and Goldwasser, 2021) incorporates social and linguistic information with pre-training tasks and conducts fine-tuning on the task of political perspective detection.

- **KGAP** (Feng et al., 2021a), short for **K**nowledge **G**raph **A**ugmented **P**olitical perspective detection, leverages knowledge graphs and graph neural networks for a knowledge-aware approach. We compare our gated R-GCN based approach with KGAP's gated R-GCN setting.

4.3 Implementation

We implement our KCD framework with pytorch (Paszke et al., 2019), pytorch lightning (Falcon and The PyTorch Lightning team, 2019), pytorch geometric (Fey and Lenssen, 2019), and the transformers library (Wolf et al., 2020). We present our hyperparameter settings in Table 1 to facilitate reproduction. We adhere to these settings throughout all experiments in the paper unless stated otherwise. Our implementation is trained on a Titan X GPU with 12GB memory. We make our code and data publicly available[3].

[3]https://github.com/Wenqian-Zhang/KCD

Method	Setting	SemEval Acc	SemEval MaF	AllSides Acc	AllSides MaF
CNN	GloVe	79.63	N/A	N/A	N/A
	ELMo	84.04	N/A	N/A	N/A
HLSTM	GloVe	81.58	N/A	N/A	N/A
	ELMo	83.28	N/A	N/A	N/A
	Embed	81.71	N/A	76.45	74.95
	Output	81.25	N/A	76.66	75.39
Text Model	Word2Vec	70.27	39.37	48.58	34.33
	GloVe	80.71	63.64	71.01	69.81
	ELMo	86.78	80.46	81.97	81.15
	BERT	86.92	80.71	82.46	81.77
	RoBERTa	87.08	81.34	85.35	84.85
MAN	GloVe	81.58	79.29	78.29	76.96
	ELMo	84.66	83.09	81.41	80.44
	Ensemble	86.21	84.33	85.00	84.25
KGAP	GRGCN	89.56	84.94	86.02	85.52
KCD	GA	88.52	84.13	86.02	85.53
	CA	89.77	85.26	81.28	80.39
	PA	**90.87**	**87.87**	**87.38**	**87.14**
KCD (PA)	- w/o TC	88.22	83.53	86.08	85.58
	- w/o KW	87.29	81.77	85.51	85.00

Table 2: Political perspective detection performance on two benchmark datasets. Acc and MaF denote accuracy and macro-averaged F1-score. N/A indicates that the result is not reported in previous works. TC and KW indicate textual cues and knowledge walks respectively.

4.4 Experiment Results

We present model performance on two benchmark datasets in Table 2, which demonstrates that

- KCD, especially with the PA aggregation strategy, consistently outperforms state-of-the-art methods on both benchmark datasets.

- KGAP and KCD, which incorporate knowledge graphs, outperform other baselines. This indicates that external knowledge is essential in providing background information and political context to analyze ideological perspectives. KCD without textual cues performs better than baseline methods except KGAP and performs close to KGAP. These suggests that KGAP's method of infusing knowledge as HIN nodes and our method of infusing knowledge as knowledge walks are both effective.

- PA outperforms CA and GA on both datasets, which suggests that the aggregation strategy is important since subsidiary nodes like textual cues may result in noise. As a result, we should focus on paragraph nodes in our heterogeneous information networks.

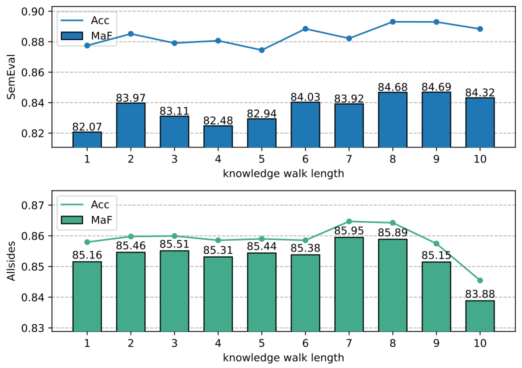

Figure 3: Our approach's performance when the maximum length of knowledge walk generation is specified from 1 to 10 knowledge graph triples.

- Removing textual cues and knowledge walks in KCD result in substantial performance drop, which demonstrates the effectiveness of textual cues and knowledge walks.

In the following, we examine the effect of knowledge walks and textual cues in our approach. We also explore how our approach performs with limited data compared to baseline methods.

4.5 Knowledge Walks Study

We propose knowledge walks, an approach to conduct multi-hop reasoning on knowledge graphs and inject them into textual representations. We study the effect of knowledge walk length and knowledge infusion strategies on our model's performance.

4.5.1 Knowledge Walks Length

Our proposed knowledge walks could be of any length, where shorter walks provide more condensed knowledge and longer walks provide more diverse knowledge. To examine the effect of knowledge walk length, we generate 5,088[4] knowledge walks of 1 to 10 triples and present model performance in Figure 3. It is illustrated that longer knowledge walks (8 or 9 for SemEval, 7 or 8 for Allsides) perform better than shorter ones, indicating the necessity of multi-hop knowledge reasoning in the task of political perspective detection.

4.5.2 Knowledge Infusion Strategy

We propose a two-step approach to infuse multi-hop knowledge reasoning into textual representations of news articles:

[4] so that there is a knowledge walk beginning with every possible (entity, relation) in the knowledge graph.

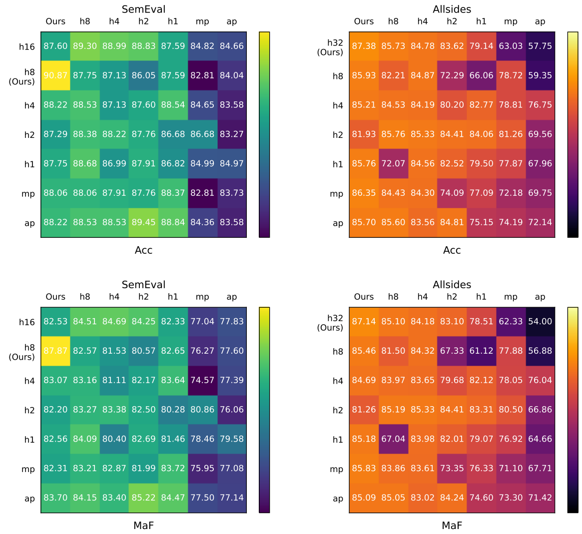

Figure 4: Model performance with different knowledge infusion strategies at two aggregation steps. The horizontal and vertical axis represent the first and second aggregation. h_k denotes multi-head attention with k heads, mp and ap stand for max and average pooling.

- First Aggregation: We firstly aggregate different generated knowledge walks based on semantic relevance in Equ. (5) and Equ. (6).

- Second Aggregation: We then use multi-head attention to aggregate all paragraphs and knowledge representations with Equ. (7) and Equ. (8).

To examine the effect of our knowledge infusion strategy, we substitute these two aggregation steps with different multi-head attention settings as well as max and average pooling. Results in Figure 4 demonstrate significant performance difference on the horizontal axis. This suggests that our semantic relevance-based knowledge walks aggregation strategy in Equ. (5) and Equ. (6) successfully filters out irrelevant knowledge reasoning and contributes to model performance. Besides, according to the vertical axis, our adopted multi-head attention in Equ. (7) and Equ. (8) is generally effective and does not rely on specific attention head settings.

4.6 Textual Cues Study

We propose to leverage semantic, syntactic and entity textual cues as paragraph-level labels to leverage implicit indicators in news articles for political perspective detection. To examine the effectiveness of these textual cues, we randomly remove them with probability p and present model performance in Figure 5. It is illustrated that:

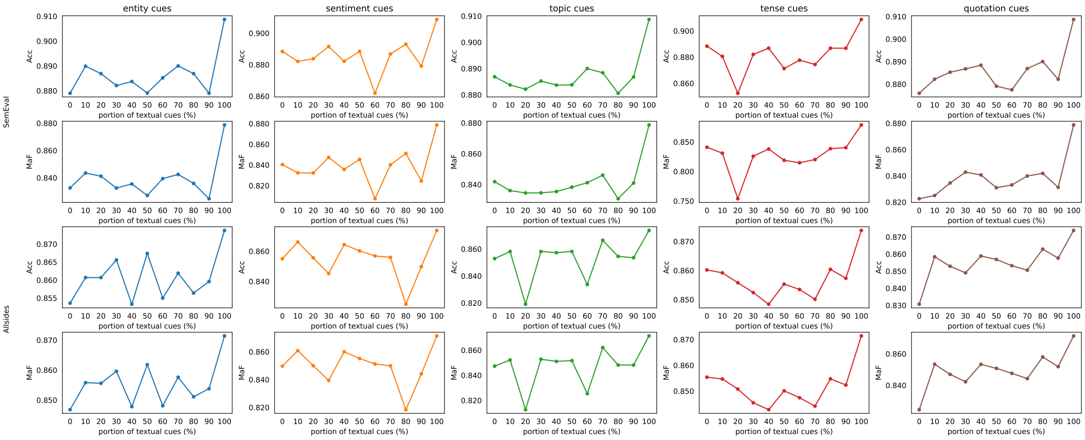

Figure 5: Model performance when five different types of textual cues are gradually removed.

- A performance boost is observed between 0% and 100% for all five textual cues, suggesting the necessity of modeling implicit textual indicators. Besides, adding only part of textual cues sometimes leads to a decrease in performance, which implies that incomplete cues may be counterproductive and introduce noise.
- Among five different cues, entity and quotation cues contribute more to model performance than others. This suggests some implicit textual cues are more important than others in analyzing the ideological perspectives of news articles.
- The effect of textual cues is larger on the dataset SemEval, which is significantly smaller than Allsides. This suggests that we alleviate the data-hungry problem by introducing diversified textual cues as paragraph-level labels and contribute to model performance.

4.7 Data Efficiency Study

As Li and Goldwasser (2021) point out, supervised data annotations could be difficult and expensive to obtain for the task of political perspective detection in news media. Our proposed knowledge walks and textual cues serve as additional information and might help mitigate this issue. To examine whether we have achieved this end, we train KCD, KGAP (Feng et al., 2021a) as well as various text models with reduced training sets of SemEval and Allsides. Results in Figure 6 demonstrate that

- KCD has better data efficiency and achieves steady performance with smaller training sets. This observation is especially salient on Allsides where the news articles are longer (Li and Goldwasser, 2021), thus more knowledge walks and textual cues could be extracted and incorporated to alleviate data dependence.

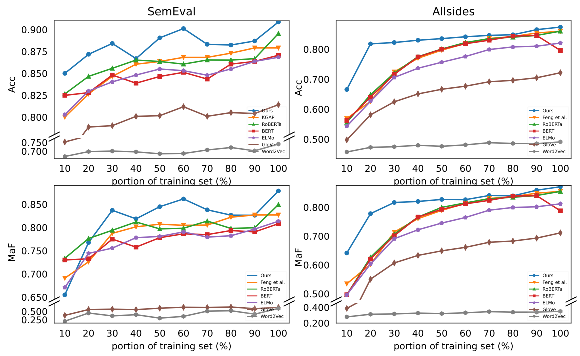

Figure 6: Model performance when KCD and various competitive baselines are trained with 10% to 100% of the training set on SemEval and Allsides.

- Both KCD and KGAP leverage external knowledge and are more robust to reduced datasets. Our approach further leverages textual cues and has better data deficiency. This suggests that a solution to limited data could be incorporating information in addition to news content.
- With only 10% training set, KCD outperforms all baselines by at least 5.68% and 9.71% in accuracy on two datasets. This suggests that our approach is simple, effective, and not data-hungry under limited data settings.

5 Conclusion

In this paper,we propose KCD, a political perspective detection approach that reasons with multi-hop external knowledge and leverages diversified implicit textual indicators. We firstly generate multi-hop knowledge walks, dynamically aggregate them based on semantic relevance and infuse into news text representations. We then construct a heterogeneous information network to jointly model knowledge-enriched news content and diversified textual cues as paragraph-level labels. Finally, we learn graph representations with relational graph neural networks under different aggregation settings and conduct political perspective detection as graph-level classification. Extensive experiments demonstrate that our approach consistently outperforms state-of-the-art baselines on two benchmark datasets. Further experiments also bear out the necessity of knowledge walks and textual cues in modeling political perspectives in news media.

6 Acknowledgment

We would like to thank the reviewers and chairs for their feedback and suggestions. We would also like to thank all LUD Lab members for our collaborative research environment. This study was supported by the National Key Research and Development Program of China (No. 2020AAA0108800), National Nature Science Foundation of China (No. 61872287, No. 62192781, No. 61937001, No. 62050194, No. 62137002), Innovative Research Group of the National Natural Science Foundation of China (61721002), Innovation Research Team of Ministry of Education (IRT_17R86), Project of China Knowledge Center for Engineering Science and Technology, and Project of Chinese academy of engineering "The Online and Offline Mixed Educational Service System for 'The Belt and Road' Training in MOOC China".

References

Isabelle Augenstein, Tim Rocktäschel, Andreas Vlachos, and Kalina Bontcheva. 2016. Stance detection with bidirectional conditional encoding. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 876–885, Austin, Texas. Association for Computational Linguistics.

Ramy Baly, Giovanni Da San Martino, James Glass, and Preslav Nakov. 2020. We can detect your bias: Predicting the political ideology of news articles. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4982–4991, Online. Association for Computational Linguistics.

Gemma Bel-Enguix, Helena Gómez-Adorno, Alejandro Pimentel, Sergio-Luis Ojeda-Trueba, and Brian Aguilar-Vizuet. 2021. Negation detection on mexican spanish tweets: The t-mexneg corpus. *Applied Sciences*, 11(9).

Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural language processing with Python: analyzing text with the natural language toolkit*. " O'Reilly Media, Inc.".

Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: A collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, SIGMOD '08, page 1247–1250, New York, NY, USA. Association for Computing Machinery.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc.

David Chang, Ivana Balažević, Carl Allen, Daniel Chawla, Cynthia Brandt, and Andrew Taylor. 2020. Benchmark and best practices for biomedical knowledge graph embeddings. In *Proceedings of the 19th SIGBioMed Workshop on Biomedical Language Processing*, pages 167–176, Online. Association for Computational Linguistics.

Kareem Darwish, Peter Stefanov, Michaël Aupetit, and Preslav Nakov. 2020. Unsupervised user stance detection on twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 141–152.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Jiachen Du, Ruifeng Xu, Yulan He, and Lin Gui. 2017. Stance classification with target-specific neural attention. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pages 3988–3994.

Subhabrata Dutta, Samiya Caur, Soumen Chakrabarti, and Tanmoy Chakraborty. 2022. Semi-supervised stance detection of tweets via distant network supervision. *arXiv preprint arXiv:2201.00614*.

William Falcon and The PyTorch Lightning team. 2019. PyTorch Lightning.

Christiane Fellbaum. 2010. Wordnet. In *Theory and applications of ontology: computer applications*, pages 231–243. Springer.

Shangbin Feng, Zilong Chen, Wenqian Zhang, Qingyao Li, Qinghua Zheng, Xiaojun Chang, and Minnan Luo. 2021a. Knowledge graph augmented political perspective detection in news media. *arXiv preprint arXiv:2108.03861*.

Shangbin Feng, Zhaoxuan Tan, Zilong Chen, Peisheng Yu, Qinghua Zheng, Xiaojun Chang, and Minnan Luo. 2021b. Legislator representation learning with social context and expert knowledge. *arXiv preprint arXiv:2108.03881*.

Paolo Ferragina and Ugo Scaiella. 2012. Fast and accurate annotation of short texts with wikipedia pages. *IEEE Software*, 29(1):70–75.

Matthias Fey and Jan Eric Lenssen. 2019. Fast graph representation learning with pytorch geometric. *arXiv preprint arXiv:1903.02428*.

Maarten Grootendorst. 2020. Bertopic: Leveraging bert and c-tf-idf to create easily interpretable topics.

Xu Han, Shulin Cao, Xin Lv, Yankai Lin, Zhiyuan Liu, Maosong Sun, and Juanzi Li. 2018. OpenKE: An open toolkit for knowledge embedding. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 139–144, Brussels, Belgium. Association for Computational Linguistics.

Benjamin D Horne, Sara Khedr, and Sibel Adali. 2018. Sampling the news producers: A large news and feature data set for the study of the complex media landscape. In *Twelfth International AAAI Conference on Web and Social Media*.

Linmei Hu, Tianchi Yang, Luhao Zhang, Wanjun Zhong, Duyu Tang, Chuan Shi, Nan Duan, and Ming Zhou. 2021. Compare to the knowledge: Graph neural fake news detection with external knowledge. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 754–763, Online. Association for Computational Linguistics.

Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu, and Tiejun Zhao. 2011. Target-dependent Twitter sentiment classification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 151–160, Portland, Oregon, USA. Association for Computational Linguistics.

Ye Jiang, Johann Petrak, Xingyi Song, Kalina Bontcheva, and Diana Maynard. 2019. Team bertha von suttner at SemEval-2019 task 4: Hyperpartisan news detection using ELMo sentence representation convolutional network. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 840–844, Minneapolis, Minnesota, USA. Association for Computational Linguistics.

Johannes Kiesel, Maria Mestre, Rishabh Shukla, Emmanuel Vincent, Payam Adineh, David Corney, Benno Stein, and Martin Potthast. 2019. SemEval-2019 task 4: Hyperpartisan news detection. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 829–839, Minneapolis, Minnesota, USA. Association for Computational Linguistics.

Chang Li and Dan Goldwasser. 2019. Encoding social information with graph convolutional networks forPolitical perspective detection in news media. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2594–2604, Florence, Italy. Association for Computational Linguistics.

Chang Li and Dan Goldwasser. 2021. Using social and linguistic information to adapt pretrained representations for political perspective identification. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 4569–4579, Online. Association for Computational Linguistics.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Yinquan Lu, Haonan Lu, Guirong Fu, and Qun Liu. 2021. Kelm: Knowledge enhanced pre-trained language representations with message passing on hierarchical relational graphs. *arXiv preprint arXiv:2109.04223*.

Walid Magdy, Kareem Darwish, Norah Abokhodair, Afshin Rahimi, and Timothy Baldwin. 2016. #isisisnotislam or #deportallmuslims? predicting unspoken views. In *Proceedings of the 8th ACM Conference on Web Science*, WebSci '16, page 95–106, New York, NY, USA. Association for Computing Machinery.

Olga Majewska, Ivan Vulić, Goran Glavaš, Edoardo Maria Ponti, and Anna Korhonen. 2021. Verb knowledge injection for multilingual event processing. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 6952–6969, Online. Association for Computational Linguistics.

Zaiqiao Meng, Fangyu Liu, Thomas Clark, Ehsan Shareghi, and Nigel Collier. 2021. Mixture-of-partitions: Infusing large biomedical knowledge graphs into BERT. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 4672–4681, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Tomás Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. In *1st International Conference* *on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings*.

Malte Ostendorff, Peter Bourgonje, Maria Berger, Julian Moreno-Schneider, Georg Rehm, and Bela Gipp. 2019. Enriching bert with knowledge graph embeddings for document classification. *arXiv preprint arXiv:1909.08402*.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Thomas Pellissier Tanon, Gerhard Weikum, and Fabian Suchanek. 2020. Yago 4: A reason-able knowledge base. In *The Semantic Web*, pages 583–596, Cham. Springer International Publishing.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar. Association for Computational Linguistics.

Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2227–2237, New Orleans, Louisiana. Association for Computational Linguistics.

Robyn Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, AAAI'17, page 4444–4451. AAAI Press.

Peter Stefanov, Kareem Darwish, Atanas Atanasov, and Preslav Nakov. 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*, pages 527–537, Online. Association for Computational Linguistics.

Bo Wang, Maria Liakata, Arkaitz Zubiaga, and Rob Procter. 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*, pages 483–493, Valencia, Spain. Association for Computational Linguistics.

Xiaozhi Wang, Tianyu Gao, Zhaocheng Zhu, Zhengyan Zhang, Zhiyuan Liu, Juanzi Li, and Jian Tang. 2021. Kepler: A unified model for knowledge embedding and pre-trained language representation. *Trans. Assoc. Comput. Linguistics*, 9:176–194.

Xiaokai Wei, Shen Wang, Dejiao Zhang, Parminder Bhatia, and Andrew Arnold. 2021. Knowledge enhanced pretrained language models: A compreshensive survey. *arXiv preprint arXiv:2110.08455*.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.

Chang Xu, Cécile Paris, Surya Nepal, and Ross Sparks. 2018. Cross-target stance classification with self-attention networks. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 778–783, Melbourne, Australia. Association for Computational Linguistics.

Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. Hierarchical attention networks for document classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1480–1489, San Diego, California. Association for Computational Linguistics.

A Limitations

Our proposed model has two minor limitations:

- We propose to model news articles with heterogeneous information networks. This graph-based approach might not fit well with shorter news articles with only a few paragraphs. This issue might be addressed by using sentence nodes instead of paragraph nodes for shorter articles.
- For very large knowledge graphs with many different types of relations, it might be hard for domain experts to help set $p(r)$ for every knowledge graph relation. This issue might be addressed by only setting a larger $p(r)$ for several important rs according to domain expert.

Dataset	# Articles	# Class	Class Distribution
SemEval	645	2	407 / 238
Allsides	10,385	3	4,164 / 3,931 / 2,290

Table 3: Details of two datasets SemEval and Allsides.

Datasets	Setting	Acc	MaF
SemEval	GA	88.62 ± 0.32	83.86 ± 0.27
	CA	89.30 ± 0.24	84.85 ± 0.30
	PA	89.90 ± 0.68	86.11 ± 1.18
AllSides	GA	84.88 ± 2.90	84.31 ± 2.96
	CA	79.85 ± 2.58	79.41 ± 2.56
	PA	87.17 ± 0.24	86.72 ± 0.35

Table 4: Average performance and standard deviation of three different aggregation strategies for five runs.

B Dataset Details

We used the same datasets as in previous works (Li and Goldwasser, 2019, 2021; Feng et al., 2021a), namely SemEval (Kiesel et al., 2019) and Allsides (Li and Goldwasser, 2019). We follow the same 10-fold setting for SemEval and 3-fold setting for Allsides (Li and Goldwasser, 2021). We use the exact same folds so that the results are directly comparable. A minor difference would be that we have to discard a few news articles on Allsides since their urls have expired and we could not retrieve their original news article. We report the statistical information of SemEval and Allsides in Table 3.

C Computation Details

C.1 Computational Resources

Our proposed approach has a total of 7.8M learnable parameters. It takes approximately 0.7 and 1.6 GPU hours to train our approach on two datasets respectively. We train our model on one Titan X GPU with 12GB memory.

C.2 Experiment Runs

We run our approach with three different aggregation strategies **five times** and report the average accuracy and macro F1-score with standard deviation in Table 4. For experiments in Section 4.5, 4.6 and 4.7, we do not have enough computational resources to run five times, thus we report the performance of a single run.

D Scientific Artifact Usage

We provide additional details about used scientific artifacts and specifically how we used them.

- NLTK (Bird et al., 2009): We use NLTK to extract the tense of news articles. Specifically, we first use NLTK POS-tagger to process new paragraphs and attach speech tag to each word. Then we align verb tags with NLTK tagset to identify the tense of paragraphs.

- BertTopic (Grootendorst, 2020): We use BertTopic to mine the topics of news corpus. Specifically, we use BertTopic topic model to learn dataset-specific topic models. For SemEval we obtained 197 topics and for Allsides we obtained 1225 topics. Next, we predict topics for each news paragraph. Each topic consists of ten topic words with scores and we select the top five to serve as the news paragraph's topic.

- Huggingface Transformers (Wolf et al., 2020): We use the pipeline module for sentiment analysis. Specifically, we use the sentiment analysis API in the text classification pipeline to generate a sentiment label and score for news paragraphs. We then use the sentiment label as the sentiment cues for news paragraphs.

- TagMe (Ferragina and Scaiella, 2012): We use TagMe to align news articles with entities in the knowledge graph. Specifically, we use TagMe to annotate named entities in news paragraphs and save the entities with a score higher than 0.1 for further alignment. We then calculate the similarity score between TagMe annotated entities and political knowledge graph entities. We recognize the entities with a score higher than 0.9 as entity cues in our constructed HIN.

- Political knowledge graph (Feng et al., 2021a): We use the political knowledge graph collected in Feng et al. (2021a) for external knowledge in political perspective detection.

- OpenKE (Han et al., 2018): We use OpenKE to train TransE (Bordes et al., 2013) knowledge graph embeddings for the political knowledge graph. Specifically, we set the TransE hidden size to 768 and train the model with other default hyperparameters in OpenKE.