

# Opinion-aware Knowledge Graph for Political Ideology Detection

Wei Chen<sup>1</sup>, Xiao Zhang<sup>1</sup>, Tengjiao Wang<sup>1</sup>, Bishan Yang<sup>2</sup>, Yi Li<sup>1</sup>

<sup>1</sup>Key Lab of High Confidence Software Technologies (MOE), School of EECS, Peking University

<sup>2</sup>Computer Science Department, Carnegie Mellon University

{pekingchenwei, xiao.zhang, tjwang, liyix}@pku.edu.cn, bishan@cs.cmu.edu

## Abstract

Identifying individual’s political ideology from their speeches and written texts is important for analyzing political opinions and user behavior on social media. Traditional opinion mining methods rely on bag-of-words representations to classify texts into different ideology categories. Such methods are too coarse for understanding political ideologies. The key to identify different ideologies is to recognize different opinions expressed toward a specific topic. To model this insight, we classify ideologies based on the distribution of opinions expressed towards real-world entities or topics. Specifically, we propose a novel approach to political ideology detection that makes predictions based on an opinion-aware knowledge graph. We show how to construct such graph by integrating the opinions and targeted entities extracted from text into an existing structured knowledge base, and show how to perform ideology inference by information propagation on the graph. Experimental results demonstrate that our method achieves high accuracy in detecting ideologies compared to baselines including LR, SVM and RNN.

## 1 Introduction

Recognizing different political ideologies from user-generated texts is essential in analyzing public opinion and user behavior on social media. For example, Facebook analyzes how users interact with socially shared news and opinions by looking at content shared by different ideology groups [Bakshy *et al.*, 2015]. Conover studies how twitter facilitates communication between online communities with different ideological positions [Conover *et al.*, 2011]. Accurate detection of political ideologies from texts is critical for performing such studies at scale.

Detecting ideology in text is difficult, as the task requires picking up complex evidence that explains the ideological positions. Individuals with different ideologies generally differ in their stands on political issues. For example, in a political dialogue about economy, liberals tend to encourage the involvement of government in regulating the economy, while

conservatives tend to limit government controls and encourage free market. In general, we find that sentiment, beliefs and judgments (in general, opinions) expressed toward the discussed subjects are often indicative for political ideologies.

Most existing opinion mining methods usually employ bag-of-words assumptions and heavily rely on word-level statistics. They are limited in capturing the nuances of ideology as they do not distinguish between words that express opinions and words that describe entities or topics. Consider for example the two sentences, “We support government involvement in healthcare.” versus “We oppose government involvement in healthcare.”. The ideology differences are implied by only the opinion words “support” and “oppose”. Therefore similar word distributions do not indicate similar ideology labels. On an analysis of a real-world text corpus (the debate of the 2012 United States presidential election), we find that more than half of the top 50 frequent words (non-stop-words) used by Obama (Democrat, liberal) and Romney (Republican, conservative) are the same. We also examine the mentioned entities and find that they are highly related in the knowledge graph. Table 1 shows two comment snippets about abortion and highlights the entities in bold. The relations between these entities (based on DBPedia) are shown in Figure 1. Although the lexical forms of these entities do not exactly match, they are highly related to the topic “abortion”.

Obama	As we mark the 39th anniversary of <b>Roe v. Wade</b> , we must remember that this Supreme Court decision not only protects a <b>woman’s health</b> and reproductive <b>freedom</b> . . . I remain committed to protecting a <b>woman’s right</b> to choose and this fundamental <b>constitutional right</b> . . . (Statement by the President on Roe v. Wade Anniversary)
Romney	I am <b>pro-life</b> . I believe that <b>abortion</b> is the wrong choice. . . (“Why I vetoed contraception bill” written by Mitt Romney) I’d like to see <b>Roe v. Wade</b> overturned and allow the states and the elected representatives of the people. . . (Republican presidential primary debate on Sept. 5, 2007)

Table 1: Speech snippets on “abortion” by Obama and Romney.

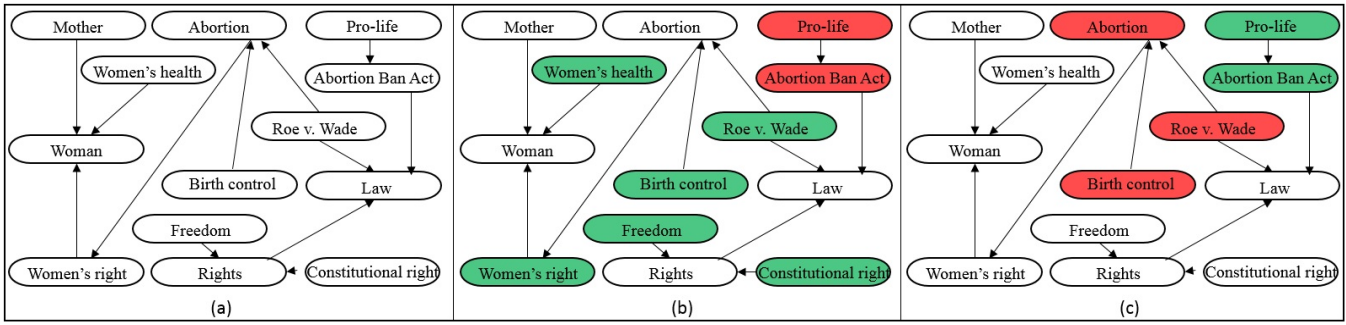


Figure 1: (a) shows a knowledge graph extracted from DBpedia based on the entities mentioned in Table 1; (b) and (c) show the same knowledge graph infused with opinions from Obama and Romney, respectively.

Sentiment information is critical for interpreting political ideologies. However, existing work on sentiment analysis largely predicts sentiment without its target (i.e., what the sentiment is about). This provides incomplete information for inferring ideologies. For example, figure 1 (b) and (c) illustrate the opinions of Obama and Romney based on their speeches (green indicates positive and red indicates negative). The evidence of their ideologies comes from their sentiment regarding different subjects related to “abortion”. Therefore in order to infer ideologies, we need to recognize not only sentiment but also the target entities or topics which the sentiment is expressed toward.

In this paper, we propose a novel political ideology detection method that makes predictions based on fine-grained information about opinions and their targeted entities. To relate text to world knowledge, we exploit entities and relations stored in large-scale knowledge graphs, facilitating knowledge-based inference for political ideologies. To the best of our knowledge, this is the first computational approach to enriching knowledge graphs with opinions for understanding political text data. We demonstrate that our approach is more effective than existing methods in political ideology detection on a variety of datasets, from formal political speeches to social media texts. Our contributions are as follows:

- We propose an opinion-aware knowledge graph, a knowledge graph that is infused with opinion information. Such a graph effectively integrates factual and subjective information and supports joint inference of texts and knowledge bases for ideology detection.
- We propose a graph-based inference framework for political ideology detection that can effectively propagate and aggregate ideology-dependent opinions for accurate ideology detection.
- We evaluate our approach on three datasets: US Congressional floor debate transcripts, Ideological Books Corpus (IBC) and tweets collections from Twitter. Experimental results demonstrate that our model is more effective than traditional text classifiers LR, SVM and an advanced neural network RNN.

## 2 Related Work

There has been a growing interest on automatic detection of political ideological beliefs from large text corpora. Early work [Thomas *et al.*, 2006] used standard text classifiers such as SVMs to classify support/against sentiment of legislators. To capture latent topical structure of text, researchers have explored topic models for ideology analysis. Ahmed and X-ing [2010] proposed a multiview LDA to model ideological perspective at the topic level. Gerrish and Blei [2011] applied ideal point topic models to infer political leanings of legislators. Nguyen *et al.* [2013] developed a hierarchical topic model to predict perspectives/ideologies in congressional debates. Sim *et al.*[2013] applied a Bayesian HMM to measure the proportions of ideologies that political candidates use in their campaigns. Mukherjee and Liu2013 presented a topic model based approach to discovering user interactions in ideological discussions. Qiu *et al.* [2015] proposed a regression-based latent factor model which jointly models user arguments, interactions, and attributes for stance prediction in online debate forums. Bamman and Smith [2015] inferred the political import of propositions from open information extraction. Recent works in sentiment analysis have used deep learning to discover compositional effects. Iyyer *et al.* [2014] applied a recursive neural network (RNN) to identify the political position at the sentence level and demonstrated RNNs efficacy on predicting ideological bias than traditional text classifiers trained with syntactic and semantic features. Different from these existing methods, our approach is a graph-based classification method that can learn from both texts and available knowledge bases. It can be applied to ideology detection at both sentence and document levels.

Detecting political ideology is closely related to research on analysis of subjectivity, sentiment and other private states [Ding *et al.*, 2008; Pang and Lee, 2008; Wiebe *et al.*, 2004]. Lexicon and sentiment-based features have been shown to be useful for recognizing political leanings [Mullen and Malouf, 2006; Somasundaran and Wiebe, 2010]. Our work also explores the use of opinion information but integrate it into a knowledge graph for ideology detection.

Knowledge bases such as DBpedia [Bizer *et al.*, 2009] and YAGO [Hoffart *et al.*, 2013] have been shown to be effective for language understanding tasks. Schuhmacher and Ponzetto [2014] proposed a graph-based semantic model for represent-

ing document content using information from DBpedia and shown the effectiveness of such model in entity ranking and document similarity modeling tasks. Wang and Domeniconi [2008] improved traditional bag-of-words representation for document classification by embedding background knowledge from Wikipedia into a semantic kernel. Our approach is also a knowledge graph base approach, which makes use of background knowledge from knowledge bases to help political ideology detection.

### 3 Problem Definition

In the social science literature, ideology has been defined as “a configuration of ideas and attitudes in which the elements are bound together by some form of constraint of functional interdependence” [Converse, 1964]. “With just the liberal or conservative label, we can quite accurately predict a politician’s stance on policy issues as disparate as taxes, health care, or abortion.” [Treier and Hillygus, 2009]

In this paper, the goal of the political ideology detection is to classify the texts into two categories (liberal and conservative) according the ideological positions in them (see examples in Table 2). The texts without ideology are not under consideration here. The features that explain the ideological positions are complex and difficult to extract. As an initial step towards this goal, we demonstrate that different political ideologies can be effectively distinguished by sentiment expressed towards certain topics. The ideology categories of the texts are determined neither by the opinion polarities (positive or negative), nor by topic contents. Instead, the key is to match the opinion positions on different topics. If the opinion polarities of the topics in one text (unknown ideology) are identical with the opinion polarities of the same topics in the training text (already known ideology), the ideology category of the unknown text will be labeled the same ideology category as the known text.

### 4 The Proposed Framework

In this paper, we propose an opinion-aware knowledge graph based framework for political ideology detection. It classifies political ideological leanings expressed in text based on a knowledge graph infused with ideology-dependent opinions. Figure 2 illustrates the proposed framework. The training set contains a set of text instances labeled with political ideology leanings (i.e., conservative vs. liberal). We also assume a background knowledge base which can be represented as a knowledge graph that consists of entities and their semantic relations. At the high level, we first construct an opinion-aware knowledge graph from the training set and the background knowledge base, and then perform ideology inference based on the constructed graph. In the following we describe the three main components of our approach:

- **Opinion Estimation:** given a knowledge graph, we estimated the opinions expressed toward the entities in the knowledge graph in terms of opinion (sentiment) distributions over two ideology categories. In figure 2, colorful nodes represent entities with ideology-dependent opinions. The left and right halves of the colorful nodes

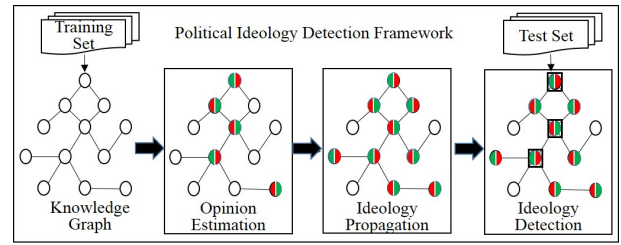


Figure 2: The framework of opinion-aware knowledge graph for political ideology detection

represent the opinions of entities over liberal and conservative respectively. We use red to indicate negative sentiment and green to indicate positive sentiment.

- **Ideology Propagation:** the opinion distributions are propagated across entities in the knowledge graph based on the semantic relation between entities. After this step, an opinion-aware knowledge graph is established.
- **Ideology Detection:** the entities in the test set are matched to the entities in the opinion-aware knowledge graph. In figure 2, the matched entities are labeled with black blocks. The ideology leanings of the test set are determined by aggregating ideology-dependent opinions toward entities on the opinion-aware knowledge graph.

## 5 Opinion-aware Knowledge Graph Construction

### 5.1 Definition

Modern knowledge bases usually use the W3C Resource Description Framework (RDF) [Cyganiak *et al.*, 2014], a generic graph-based data model for storing entities and their relationships. The standard RDF represents facts in the form of (*subject*, *predicate*, *object*) (SPO) triples, where *subject* and *object* are entities and *predicate* is the type of a relation. The existence of a particular SPO triple indicates an existing fact, i.e., that the respective entities are in a relationship of the respective type. In this paper, we represent an opinion-aware knowledge graph using an extended version of the SPO triples. The definition of opinion-aware knowledge graph is as follows:

**Definition:** an opinion-aware knowledge graph is a graph-based relational knowledge representation model. It can be described by following RDF triple:

$$([\textit{subject}, \langle \textit{libOp}, \textit{conOp} \rangle], \textit{predicate}, [\textit{object}, \langle \textit{libOp}, \textit{conOp} \rangle]),$$

where  $\langle \textit{libOp}, \textit{conOp} \rangle$  expresses a distribution of ideology-dependent opinions toward an entity over two ideology categories. *libOp* and *conOp* refer to opinions representing liberal and conservative views respectively. They can take values from  $\{> 0, = 0, < 0\}$ , which indicates positive, neutral, negative correspondingly.

Let  $G_b$  be the background knowledge graph and  $C_b$  be the set of entities and  $R$  be the set of semantic relations. Let  $C$

be the set of entities mentioned in the training sets. In the first phase of our approach, we create a directed graph  $G = (V, E)$  containing i) all entities in  $C$ ; ii) their relations matched in  $R$ . That is  $V = C$  and  $E \subseteq V \times R \times V$ . The initial opinion-aware knowledge graph  $G_1$  is created from  $G$  by extended the representation of entities in  $V$  to  $[\text{subject/object}, \langle \text{libOp}, \text{conOp} \rangle]$  structure. Initially, both  $\text{libOp}$  and  $\text{conOp}$  are set to be 0.

## 5.2 Opinion Estimation

After the construction of the initial knowledge graph  $G_1$ , we perform the first step of opinion-aware knowledge graph construction: opinion estimation. The aim of opinion estimation is to compute the opinions toward the entities in  $G_1$  over two ideology categories using the training set. We first apply a holistic lexicon-based approach [Ding *et al.*, 2008] to infer opinion scores toward entities identified in text. We then aggregate the resulting opinion scores to determine the final orientation of the opinion on each entity. Let  $v$  be an entity in  $G_1$ , which appear in a set of sentences  $L = \{L_1, \dots, L_m\}$  that express liberal views and a set of sentences  $S = \{S_1, \dots, S_n\}$  that express conservative views in the training set. Let  $op = \{op_1, \dots, op_k\}$  be a set of opinion words or phrases from an opinion lexicon. The opinion vector  $\langle \text{libOp}, \text{conOp} \rangle$  on each entity  $v$  is determined by the following opinion aggregation functions:

$$\begin{cases} \text{libOp} = \sum_{L_i \in L} \sum_{op_j \in L_i} \frac{op_j \cdot SO}{d(op_j, v)} \\ \text{conOp} = \sum_{S_i \in S} \sum_{op_j \in S_i} \frac{op_j \cdot SO}{d(op_j, v)} \end{cases}$$

where  $op_j$  is an opinion word in sentence  $L_i$  or  $S_i$ ,  $d(op_j, v)$  is the distance between entity  $v$  and opinion word  $op_j$  in  $L_i$  or  $S_i$ ,  $op_j \cdot SO = \{1, 0, -1\}$  means that the opinion orientation of  $op_j$  is  $\{\text{positive}, \text{neural}, \text{negative}\}$ .

## 5.3 Ideology Propagation

After the first step, the resulting opinion-aware knowledge graph  $G_1$  can already be used for ideology detection. However,  $G_1$  does not take advantage of the rich entities and relations that exist in the background knowledge base. Therefore, we introduce the second step of opinion-aware knowledge graph construction: ideology propagation, in order to perform more knowledge-intensive inference. We can extend the base graph  $G_1$  by the following criterion:

**Criterion:** entity  $v'_i$  is appended into the entity set  $V'$  and edge  $(v, r_j, v'_i)$  or  $(v'_i, r_j, v)$  is appended into edge set  $E'$ , if and only if i) there is a RDF triple  $(v, r_j, v'_i)$  or  $(v'_i, r_j, v)$  in the background knowledge graph  $G_b$ ; ii)  $v'_i \in C_b$  and  $v'_i \notin V$ . The extended opinion-aware knowledge graph  $G_{op} = (V_{op}, E_{op})$ , where  $V_{op} = V \cup V'$  and  $E_{op} = E \cup E'$ .

The opinions toward all entities in  $V$  are obtained during the opinion estimation step. Only the new entities in  $V'$  need to be estimated. We estimate the opinion distributions  $\langle \text{libOp}, \text{conOp} \rangle$  of all entities in  $V'$  based on their semantic relations in the knowledge graph.

If for every  $v'_i \in V'$ , there is a set  $V^s = \{v_1^s, \dots, v_p^s\} \subseteq V$ , satisfying  $(v_j^s, r_j, v'_i)$  or  $(v'_i, r_j, v_j^s) \in E'$ ,  $j = 1, \dots, p$ , then the opinion vector  $\langle \text{libOp}, \text{conOp} \rangle$  on entity  $v'_i$  is determined by the following opinion aggregation functions:

$$\begin{cases} \text{libOp} = - \sum_{j=1}^p \log[P(r_j)] \cdot \text{libOp}_j \\ \text{conOp} = - \sum_{j=1}^p \log[P(r_j)] \cdot \text{conOp}_j \end{cases}$$

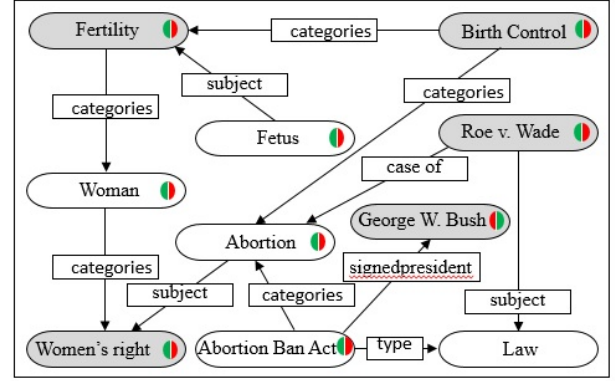


Figure 3: A subgraph of opinion-aware knowledge graph

Table 2: Samples from IBC dataset

Ideology	Text content
Liberal	Criminalizing abortion would not, for instance, address the needs of women.
Liberal	Birth control gives women the power to pace their own fertility, and access to good jobs.
Conservative	Abortion harms women physically and mentally and causes fetuses pain.

where  $\langle \text{libOp}, \text{conOp} \rangle$  is the opinion vector of entity  $v_j^s$ ,  $P(r_j)$  is the probability of the relation  $r_j$  in the background knowledge graph  $G_b$ . This measure makes the assumption that specificity is valuable for relevance, for example the “type” relation vs. “signedpresident” relation. The weight  $-\log[P(r_j)]$  is meant to capture the degree of relevance of semantic relation for the entities it connects. The key idea underlying our weighting is to reward, for a given source node, those edges and target nodes that are most specific to it. This idea is also applied in [Schuhmacher and Ponzetto, 2014].

Unlike lexical resources like WordNet, our backend ontology contains many different, fine-grained semantic relations. Figure 3 illustrates an example subgraph of  $G_{op}$ . Entities in  $V$  and their opinion distributions are shown in white oval frames. Gray oval frames represent extended entities in  $V'$ . Examples of corresponding training data are presented in Table 2. Although entities (e.g. Birth Control, Fertility, Roe v. Wade et al.) aren’t included in training data, their ideology-dependent opinion can also be determined through semantic relations.

## 5.4 Political Ideology Detection

So far, we construct an opinion-aware knowledge graph  $G_{op}$  with an ideology-dependent opinion vector on every entity.  $G_{op}$  can be applied to classify the ideology categories of unseen sentence or document  $T$  in the test set. Let  $C_T$  be a set of entities in  $T$ . First, the intersection  $V_T$  needs to be computed as  $V_T = C_T \cap V_{op}$  by matching the entities in  $C_T$  and  $G_{op}$ .

Then, the opinion orientation  $TOP_i$  of every entity  $v_{T_i} \in V_T$  is estimated based on the test set:

$$TOP_i = \text{sgn}\left(\sum_{op_j \in T} \frac{op_j \cdot SO}{d(op_j, v_{T_i})}\right),$$

where  $\text{sgn}()$  is sign function,  $op_j$  is an opinion word in sentence  $T$ .  $d(op_j, v_{T_i})$  is the distance between entity  $v_{T_i}$  and opinion word  $op_j$  in  $T$ . The ideology weight of  $T$  is defined as follows:

$$\text{ideology}W = \left(\sum_{v_{T_i} \in V_T} \text{libOp}_{v_{T_i}} \cdot TOP_i - \sum_{v_{T_i} \in V_T} \text{conOp}_{v_{T_i}} \cdot TOP_i\right) / |V_T|,$$

where  $\langle \text{libOp}, \text{conOp} \rangle$  is the opinion of entity  $v_{T_i}$  in  $G_{Op}$ , and  $|V_T|$  expresses the number of the matched entities in  $V_T$ . If  $\text{ideology}W > 0$ , then the ideology leaning of  $T$  is labeled as liberal. In contrast, if  $\text{ideology}W < 0$ , then the ideology leaning of  $T$  is labeled as conservative. The key idea is same opinion means same ideology.

## 6 Experimental Setup

### 6.1 Datasets

We use three real-world corpora to evaluate the opinion-aware knowledge graph based approach for political ideology detection.

1. The U.S. Congressional floor debate (Convote) dataset [Thomas *et al.*, 2006]: it consists of debate transcripts of the US Congress in 2005. All speakers are labeled with their political parties (*Democrat*, *Republican*, or *Independent*). We create the ideology labels for each speech (*Liberal* or *Conservative*) according to the speaker’s party labels (from *Democrat* to *Liberal*, from *Republican* to *Conservative*).
2. The IBC dataset [Iyyer *et al.*, 2014]: it includes sentences selected from the Ideological Books Corpus, a collection of books and magazine articles written between 2008 and 2012 by authors with well-known political leanings. Iyyer *et al.* [2014] uses the Crowdfunder crowdsourcing platform (crowdfunder.com.) to obtain human annotations of the IBC dataset for political bias at both the sentence and phrase level.
3. The Twitter dataset: Bakshy *et al.* [2015] lists the media with their ideology leanings. We further find their corresponding twitter accounts. In all, there are 127 liberal and 86 conservative twitter accounts. We collect tweets (provided by SasaPetrovic<sup>1</sup>) posted by these twitter accounts from Jul., 2011 to Sep., 2011. We create the ideology labels for tweets according to the twitter accounts.

We use DBPedia<sup>2</sup> [Bizer *et al.*, 2009] as a source of background knowledge graph.

### 6.2 Compared Methods

We validate the effectiveness of the proposed opinion-aware knowledge graph (OKG) for political ideology detection by comparing with the following methods:

<sup>1</sup>[http://demeter.inf.ed.ac.uk/cross/docs/fsd\\_corpus.tar.gz](http://demeter.inf.ed.ac.uk/cross/docs/fsd_corpus.tar.gz)

<sup>2</sup><http://wiki.dbpedia.org>

Table 3: The confusion matrix on Convote and IBC datasets

method	Convote	IBC
LR	101 107	96 74
	37 171	65 137
LR-KG	163 45	103 67
	54 154	40 162
SVM	109 99	108 62
	20 188	58 144
SVM-KG	163 45	116 54
	51 157	54 148
OKG	151 57	114 56
	20 177	13 189

1. LR: the logistic regression algorithm using bag-of-words features.
2. LR-KG: knowledge graph is incorporated as extended features. The bag-of-words features are extended by the strongly related entities by ranking them on the basis of their relevance in knowledge graph [Schuhmacher and Ponzetto, 2014]. The opinion polarities of the words are also extended as classification features.
3. SVM: the SVM with the RBF kernel using bag-of-words features.
4. SVM-KG: knowledge graph extended bag-of-words features and opinion features are used just as same as the features used by LR-KG.
5. RNN1, RNN1-(W2V) and RNN2-(W2V): ideology classification models proposed in [Iyyer *et al.*, 2014]. RNN1 initializes all parameters randomly. RNN1-(W2V) and RNN2-(W2V) are initialized using word2vec embeddings. RNN1 and RNN1-(W2V) are trained on only sentence-level labels. RNN2-(W2V) includes annotated phrase labels in its training.

We tuned the parameters of SVM and LR with cross-validation by using machine learning library scikit-learn<sup>3</sup>.

### 6.3 Experimental Results

We perform 10-fold cross-validation on Convote and IBC datasets for each method except for the RNN-serial model<sup>4</sup>. Table 3 and Table 4 show the results.

We report the confusion matrix for each method in Table 3.

The confusion matrix is defined as  $\begin{bmatrix} tc & fc \\ fl & tl \end{bmatrix}$ , which contains the number of test instances for true liberal ( $tl$ ), true conservative ( $tc$ ), false liberal ( $fl$ ) and false conservative ( $fc$ ). We also compute the accuracy, which is defined as:

$$\text{accuracy} = \frac{tl + tc}{tl + tc + fl + fc}.$$

<sup>3</sup><http://scikit-learn.org/stable/index.html>

<sup>4</sup>For the RNN-serial models, we use the accuracy results reported in [Iyyer *et al.*, 2014] on the same dataset. They did not provide the confusion matrix in the paper.



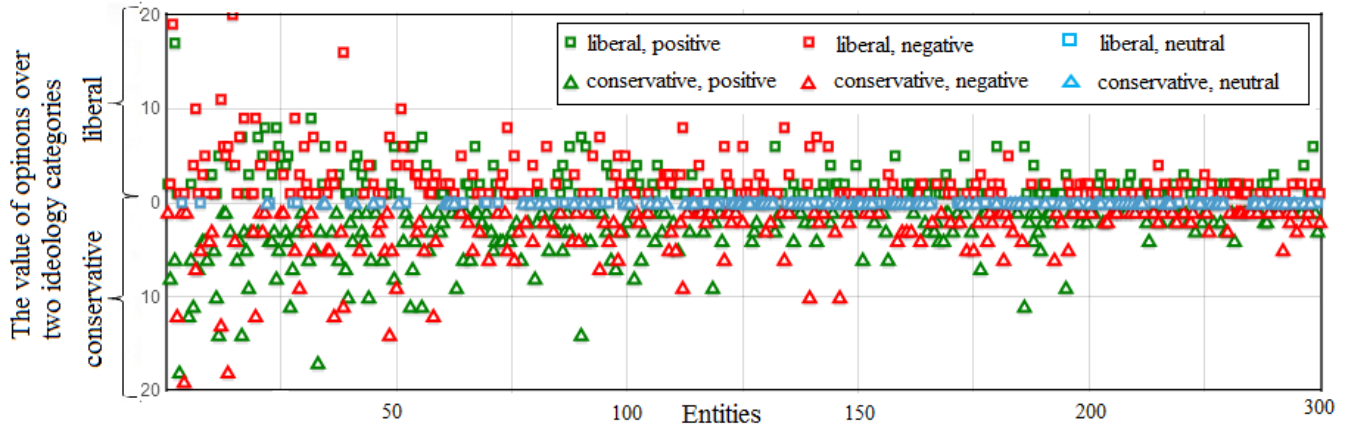


Figure 4: The samples of ideology dependent opinion distributions of entities obtained by OKG on IBC dataset

Table 4: The accuracy results on the Convote, IBC and Twitter dataset

	LR	LR-KG	SVM	SVM-KG	RNN1	RNN-(w2v)	OKG
Convote	68.3%	76.2%	71.4%	76.9%	69.4%	70.2%	<b>81.5%</b>
IBC	67.5%	71.2%	67.8%	71.0%	66.2%	69.3%	<b>81.4%</b>
Twitter	62.4%	68.2%	64.4%	72.7%	-	-	<b>81.8%</b>

We merge all tweets by the same Twitter account together. The ideology leanings of the Twitter accounts can be determined by classifying the ideology categories of the merged data. Table 4 shows the accuracy of all methods on the Convote, IBC and Twitter dataset. The increased accuracy and recall suggest that the OKG outperforms the baseline models on all three datasets.

Table 5 shows the examples of the ideology detection results, which are classified correctly by OKG and incorrectly by LR and SVM. The major reason for these results is that OKG integrates factual and subjective information and supports joint inference of texts and knowledge bases for ideology detection. It is difficult for LR and SVM to deal with the lack of key words and complex texts.

In figure 4 we choose entities mentioned by both liberals and conservatives and arrange them based on their frequencies from high to low. The squares above the horizontal axis indicate the estimated opinions toward the selected entities expressed by liberals. The triangles below the horizontal axis indicate the opinions toward the selected entities expressed by conservatives. The distance from each point to the horizontal axis is the absolute value of the opinion score libOp or conOp for the corresponding entity. The opinion polarities are illustrated with green (positive), red (negative) and blue (neutral) colors. In total, opinion distributions of the top 300 entities are illustrated. 68% of the entities associate with different opinion polarities over the two ideology categories. These results confirm that recognizing the difference of opinions expressed toward the same entity is essential for ideology detection.

Table 5: Samples classified correctly by OKG and incorrectly by LR and SVM.

Test Texts
Pro-abortion forces typically exaggerate the number of abortions performed prior to legalization and paint a dreadful historical picture where millions of desperate women met their demise at the hands of quack abortionists until legalization made the procedure safe. ( <b>Conservative</b> )
Bush created the Office of Faith-based and Community Initiatives, headed by the Faith-Based Czar, and even gave the nation a Birth Control Czar, formally the Deputy Assistant Secretary of Population Affairs in the Department of Health and Human Services. ( <b>Conservative</b> )
Even if South Dakota’s ban is voted down for want of a rape or incest exception, the woman-protective argument against abortion will spread, making ever more commonplace the 19th-century forms of reasoning about women that underwrote the law. ( <b>Liberal</b> )

## 7 Conclusion

Recognizing different political ideologies from user-generated texts is essential in analyzing public opinion and user behavior on social media. In this paper, we propose a novel political ideology detection method. It leverages opinions extracted from texts with existing knowledge bases, allowing better understanding of the political topics mentioned in text as well as the opinions expressed toward these topics.

## Acknowledgments

This research is supported by the Natural Science Foundation of China (61572043) and National Key Research and Development Program (2016YFB1000704).

## References

- [Ahmed and Xing, 2010] Amr Ahmed and Eric P. Xing. Staying informed: supervised and semi-supervised multi-view topical analysis of ideological perspective. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1140-1150, 2010.
- [Bakshy *et al.*, 2015] Eytan Bakshy, Solomon Messing, and Lada Adamic. Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348 (6239) : 1130-1132, 2015.
- [Bamman and Smith, 2015] David Bamman and Noah A. Smith. Open Extraction of Fine-Grained Political Statements. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 76-85, 2015.
- [Bizer *et al.*, 2009] Christian Bizer, Jens Lehmann, Georgi Kobilarov, Sören Auer, Christian Becker, Richard Cyganiak and Sebastian Hellmann. DBpedia- A crystallization point for the web of data. *Journal of Web Semantics*, 7 (3): 154-165, 2009.
- [Conover *et al.*, 2011] Michael D. Conover, Bruno Gonçalves, Jacob Ratkiewicz, Alessandro Flammini and Filippo Menczer. Predicting the political alignment of twitter users. In *Proceedings of 3rd IEEE Conference on Social Computing*, pages 192-199, 2011.
- [Converse, 1964] Philip E. Converse. The Nature of Belief Systems in Mass Publics. In *Ideology and Discontent*, The Free Press of Glencoe, New York 1964.
- [Cyganiak *et al.*, 2014] Richard Cyganiak, David Wood and Markus Lanthaler. RDF 1.1 Concepts and Abstract Syntax. [Online]. Available: <http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>
- [Ding *et al.*, 2008] Xiaowen Ding, Bing Liu and Philip S. Yu. A holistic lexicon-based approach to opinion mining. In *Proceedings of the Conference on Web Search and Web Data Mining*, pages 231-240, 2008.
- [Gerrish and Blei, 2011] Sean Gerrish and David M. Blei. Predicting legislative roll calls from text. In *Proceedings of the International Conference on Machine Learning*, pages 489-496, 2011.
- [Hoffart *et al.*, 2013] Johannes Hoffart, Fabian M. Suchanek, Klaus Berberich and Gerhard Weikum. YAGO2: a spatially and temporally enhanced knowledge base from Wikipedia. *Artificial Intelligence*, 194: 28-61, 2013.
- [Iyyer *et al.*, 2014] Mohit Iyyer, Peter Enns, Jordan Boyd-Graber and Philip Resnik. Political ideology detection using recursive neural networks. In *Proceedings Annual Meeting of Association for Computational Linguistics Conference*, pages 1113-1122, 2014.
- [Mukherjee and Liu, 2013] Arjun Mukherjee, Bing Liu. Discovering user interactions in ideological discussions. In *Proceedings Annual Meeting of Association for Computational Linguistics Conference*, pages 671-681, 2013.
- [Mullen and Malouf, 2006] Tony Mullen, Robert Malouf. A preliminary investigation into sentiment analysis of informal political discourse. In *Proceedings of the AAAI symposium on computational approaches to analyzing weblogs*, pages 159-162, 2006.
- [Nguyen *et al.*, 2013] Viet-An Nguyen, Jordan Boyd-Graber and Philip Resnik. Lexical and Hierarchical Topic Regression. In *Proceedings of the 2013 Neural Information Processing Systems (NIPS) Conference*, pages 1106-1114, 2013.
- [Pang and Lee, 2008] Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2 (1-2) : 1-135, 2008.
- [Qiu *et al.*, 2015] Minghui Qiu, Yanchuan Sim, Noah A. Smith and Jing Jiang. Modeling User Arguments, Interactions, and Attributes for Stance Prediction in Online Debate Forums. In *Proceedings of SIAM International Conference on Data Mining*, pages 855-863, 2015.
- [Schuhmacher and Ponzetto, 2014] Michael Schuhmacher, Simone Paolo Ponzetto. Knowledge-based graph document modeling. In *Proceedings of the 7th ACM international conference on Web search and data mining*, pages 543-552, 2014.
- [Sim *et al.*, 2013] Yanchuan Sim, Brice D. L. Acree, Justin H. Gross and Noah A. Smith. Measuring ideological proportions in political speeches. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 91-101, 2013.
- [Somasundaran and Wiebe, 2010] Swapna Somasundaran and Janyce Wiebe. Recognizing Stances in Ideological On-Line Debates. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 116-124, 2010.
- [Thomas *et al.*, 2006] Matt Thomas, Bo Pang and Lillian Lee. Get out the vote: Determining support or opposition from Congressional floor-debate transcripts. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 327-335, 2006.
- [Treier and Hillygus, 2009] Shawn Treier and D. Sunshine Hillygus. The Nature of Political Ideology in the Contemporary Electorate. *Public Opinion Quarterly*, 73 (4): 679-703, 2009.
- [Wang and Domeniconi, 2008] Pu Wang and Carlotta Domeniconi. Building Semantic Kernels for Text Classification using Wikipedia. In *Proceedings of the 14th ACM International Conference on Knowledge Discovery and Data Mining*, pages 713-721, 2008.
- [Wiebe *et al.*, 2004] Janyce Wiebe, Theresa Wilson, Rebecca Bruce, Matthew Bell and Melanie Martin. Learning subjective language. *Computational Linguistics*, 30 (3): 277-308, 2004.