Detecting Agreement and Disagreement in Political Debates

Mahboubeh Ahmadalinezhad(✉) and Masoud Makrehchi

University of Ontario Institute of Technology, Oshawa, ON L1H 7K4, Canada
{mahboubeh.ahmadalinezhad, masoud.makrehchi}@uoit.ca

Abstract. In this paper, the task of agreement/disagreement detection in political debates is studied. The main goal of this study is to detect agreement/disagreement between two individuals on a topic based on their conversations. This is a challenging task due to the lack of annotated corpora in this field. A self-labeling method is introduced for data collection and generating the training data. A new approach based on text classification is proposed for this task. The experimental results on Canadian Parliamentary debates and the United State 1960 Presidential Campaign datasets have proven the efficiency of the developed methodology and outperforms the baseline methodologies. In addition, the validity of the proposed self-labeling method is evaluated, and its efficiency is confirmed.

Keywords: Sentiment analysis · Agreement/disagreement detection · Text classification · Political analytics

1 Introduction

Recently, analyzing social interactions and mining public opinions have attracted a great deal of attention due to its practical applications in providing better services to the users. A number of studies have focused on mining for dispute in online interactions [3,9]. One of the main argumentative dataset to analyze both agreement/disagreement is the political debates, which contain official and unofficial documents. Detecting agreement/disagreement in political debates in the US congress has been studied [2,13]. In these methods, determining the single stance of a debate participant with respect to a specific topic was investigated.

In this paper, the goal is to introduce a new approach to generate training data for political analytics and introduce a methodology to detect agreement/disagreement in political debates. The cost of gathering such a dataset is not very high, but handling this kind of information is difficult and requires further development.

In Sect. 2, related works are presented. The problem statement and the notion of agreement/disagreement are defined in Sect. 3. A new methodology is presented in Sect. 4. In Sect. 5, further details about collecting the data and labeling are discussed. Implementation results are demonstrated in Sect. 6. Finally, conclusion and future works are mentioned in Sect. 7.

R. Thomson et al. (Eds.): SBP-BRiMS 2018, LNCS 10899, pp. 54–60, 2018.
https://doi.org/10.1007/978-3-319-93372-6_6

2 Related Works

Two major categories of methods are used to classify political statements as supporting/opposing for a debated topic.

The first category are those approaches that utilize the common information of the text structure, in which the focus is on sentiment analysis and contextual information. Somasundaran and Wiebe [12] proposed an approach, which classifies a stance as approve or disapprove about a debated topic.

In [1], determining disagreement in online political forums between a pair of quoted text and a given response is studied. Anand et al. [2] improved the results of unigram and classification for various topics using contextual information and opinion dependencies. To classify controversial discussion topics on the political domain, an LM-based method is proposed [4]. In [6], U.S Congressional floor debate transcripts are used as a dataset and sentiment classification is applied to determine agreement/disagreement.

The second category is based on corpus-specific features. In [13], a new variant of Latent Semantic Analysis (LSA) is proposed to detect the support and opposition to legislation in congressional debates using information such as speech transcriptions, records on voting, and the relation between the speakers. Moreover, some approaches are proposed to identify agreement/disagreement in consecutive speech transcription segments. Different speakers talk either positively or negatively against the discussed topic by using lexical, structural, and prosodic features [7].

As opposed to the previous studies, in which the opinion of one speaker about a specific topic is investigated and his agreement/disagreement is detected, the objective of this study is to detect the type of the interaction between two speakers regardless of the topic.

3 Problem Statement

In Canadian Parliamentary, parties are categorized into two groups: governing party and opposition party. In this study, there are two main assumptions: first, a representatives of the governing party and a representative from the opposition party disagree on a topic, and second two representatives of the same party agree on the topic. In addition, it is assumed that each conversation between two individuals is a document. In this scenario, first, the document collections are processed and the text are extracted to compared and classified. Second, an approach is proposed to classify each pair of conversations based on supervised learning, which considers the features capturing the relevant dimensions.

4 The Proposed Method

The major problem in an agreement/disagreement task is to represent the conversation between two individuals. The core idea in the proposed approach is conversation modeling using Bag-of-Words representation of interpolation discussions. Three different operators are used to build a document representation and apply a text classifier such as Support Vector Machine to train the prediction model.

The objective is to learn the features for the automatic detection of agreement/disagreement that would provide useful information about the conversations between people without knowledge of their topic. We focus on oral speech, which has less information in comparison to written text such as punctuation, non-lexical features, and time between posts, etc. Since we work on Hansard - the printed version of what members of Parliament expressed in the House of Commons- we also lose utterance information of the oral conversations. In addition to the main proposed method, three other methods based on similarity and sentiment analysis are also implemented and compared to the text classification methods.

1. **Text classification:** Each document is considered separately and two vectors for each document are computed. Each document is converted to a fixed-size representation to be used as an input to the classifier. Three different operators are applied to interpolate a document for conversation modeling:
   Concatenation operator: The conversation between two members is represented by concatenation of two vectors. The length of the document is $2n$.
   OR operator: This operator is used to represent the conversation by using this operator, the length of the document is n.
   AND operator: The conversation is represented by applying AND operator. The length of the document after using AND operator is n.
2. **Lexicon based analysis (Sentiment):** In order to implement the sentiment analysis of a conversation, the Linguistics Inquiry Word Count tool (LIWC-2001) [11] is utilized.
3. **Cosine similarity:** This measure is used as a metric to compute similarity between two documents [14].
4. **Cosine similarity and Lexicon based analysis:** Both of them are combined and considered as features for a document.

5 Dataset

The proposed method is evaluated on three different datasets.

Parliament of Canada: The first one is the debates of Parliament of Canada are collected from January to May 2016. The data includes 55 debates and more than 5000 documents. Conservative Party, Liberal Party, and New Democratic Party are the three major political parties in Canada.

To analyze effectiveness of these assumptions, two other datasets are considered which have been annotated by independent annotators using the CrowdFlower crowd sourcing.

1960 Presidential Campaign Dataset: The transcription of discourses and official declarations issued by Nixon and Kennedy during 1960 presidential campaign are collected. This data includes 881 documents.

Table 1. Results of determining agreement/disagreement by running various methods on Parliament of Canada dataset

Methods	Accuracy	F-score
Sentiment	0.48	0.37
Cosine similarity	0.52	0.41
Sentiment and similarity	0.53	0.44
Text classification with concatenation	**0.81**	**0.80**
Text classification with OR	0.68	0.67
Text classification with AND	0.63	0.63

Extended 1960 Elections: This data is extended version of the second dataset includes 1,400 pairs.

6 Results

A Support Vector Machine (SVM) [10] is used to train a model. The results are based on 10-fold cross-validation and the average prediction accuracy and F-scores are reported for all experiments.

6.1 Classification and Self-labeling for Parliament of Canada Data

The classification accuracy is shown in Table 1, where the results are demonstrated for all methods. The accuracy of the proposed method, text classification, is significantly improved in comparison to the others. An interesting point about the text classification method is the reduction in the difference between accuracy and F-score. For other methods, the results of accuracy is 10% higher than the results of F-score. This observation shows a strong capability of text classification in detecting both agreement and disagreement in a conversation.

6.2 Evaluation the Sensitivity of the Classification to the Amount of the Training Data

In order to evaluate the sensitivity of our models, the proposed method is evaluated on varied percentage of the training data. This is done to investigate the effect of a change in the percentage of training and its impact on the results. Since the best result was obtained with text classification using concatenation operator, its sensitivity to the training percentage against the test sample is investigated. The results are reported in Fig. 1 which is proving the fact that the proposed approach can detect the agreement/disagreement without much dependency on the training percentage.

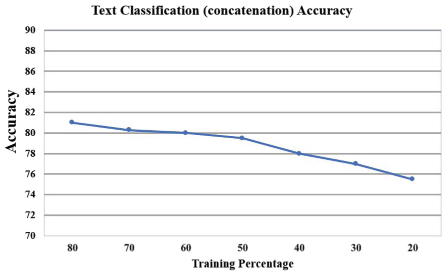

Fig. 1. Results of text classification (concatenation) at different training percentage on Parliament of Canada dataset

Table 2. Results of text classification (concatenation) on annotated dataset

Methods	1960 Elections	Extended 1960 Elections
Menini and Tonelli [8]	0.83	0.80
Self-labeling (Sect. 6.3)	0.79	0.74
Text classification (Sect. 6.4)	0.87	0.93

6.3 Evaluation the Proposed Self-labeling Method

Furthermore, the proposed self-labeling is evaluated. Therefore the US 1960 Presidential Campaign dataset is used [8] which is an annotated dataset and the transcription of discourses during the campaign. In this scenario, the proposed method (Text classification with concatenation) is run on this dataset. The model is constructed based on the two main assumptions and self-labeling, however, test phase is evaluated based on the goal labels which achieved 79% accuracy. The goal of this experiment is to test the proposed self-labeling method and the strength of the training model. The results are compared to [8] which uses negation/overlap, entailment, sentiment, cosine, word embeddings as features. According to Table 2 which is confirmed that the proposed self-labeling method works properly.

By applying transfer learning, we achieved 60% accuracy. The difference in the accuracy from the literature may be the result of the two political domains and the fact that language used has changed between 1960 and 2016.

6.4 Evaluation the Efficiency of Text Classification Method

In addition, the efficiency of the text classification approach by using interpolation compares to the proposed method of [8]. According to the Table 2 the results of text classification approach approve the efficiency of text classification method by using concatenation operator. We can conclude that our approach is a reliable solution to the task of detection both agreement and disagreement.

6.5 Domain Adaptation and Transfer Learning

In this experiment, transfer learning methods are applied and the training model is updated with the data which achieved high probability of predictions. By using transfer learning [5], the effort for annotating reviews for each document can be reduced, and the model which is based on training documents is used to learn classification models of other datasets. In this case, transfer learning can save a significant amount of labeling effort. The Parliament of Canada debates are considered as training data and the US 1960 Elections as test data. In each iteration, samples with high probability are added to improve the model. By applying transfer learning, we achieved 60% accuracy. The difference in the accuracy from the literature may be the result of the two political domains and the fact that language used has changed between 1960 and 2016.

7 Conclusion

In this paper, detection of agreement/disagreement in Canada's Parliament debates and the US 1960 election datasets were investigated. The detection was done by inputing the conversations and determining the agreement/disagreement of two individuals without respect to the topic. The input data were the oral debates between the parties of the Parliament without written information and utterance, which makes it a challenging task to detect the agreement/disagreement. As the data is not annotated, the data labeling was done based on two main assumptions: a representative of the governing party and a representative of the opposition party disagree on a topic, and two representatives of the same party agree on the topic. A new method was introduced for data collection and a novel algorithm based on classical text classification was proposed to detect agreement/disagreement. Different classification methods and different types of interpolation were examined and text classification with concatenation operator was found to be the best one by 81% accuracy.

Moreover, validity of two main hypotheses of this study was investigated. The US 1960 Presidential Campaign dataset was used to evaluate the assumptions, is labeled manually. We achieved 79% accuracy which proves efficiency of the proposed self-labeling and two assumptions.

In addition, text classification method is applied to the US 1960 Presidential Campaign dataset and observed significant improvement in comparison to proposed features and classifier of [8]. Overall, 87% accuracy is attained for the US 1960 elections and 93% for extended the US 1960 elections. Furthermore, semi-supervised learning and applying domain adaptation achieve acceptable results in comparison to the previous work. In the future, we also want to use other methods such as skip-gram to represent a document and apply advanced text classification algorithms.

References

1. Abbott, R., Walker, M., Anand, P., Fox Tree, J.E., Bowmani, R., King, J.: How can you say such things?!?: recognizing disagreement in informal political argument. In: Proceedings of the Workshop on Languages in Social Media, pp. 2–11. Association for Computational Linguistics (2011)
2. Anand, P., Walker, M., Abbott, R., Tree, J.E.F., Bowmani, R., Minor, M.: Cats rule and dogs drool!: classifying stance in online debate. In: Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, pp. 1–9. Association for Computational Linguistics (2011)
3. Andreas, J., Rosenthal, S., McKeown, K.: Annotating agreement and disagreement in threaded discussion. In: LREC, pp. 818–822. Citeseer (2012)
4. Awadallah, R., Ramanath, M., Weikum, G.: Language-model-based pro/con classification of political text. In: Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 747–748. ACM (2010)
5. Blitzer, J., Dredze, M., Pereira, F., et al.: Biographies, bollywood, boom-boxes and blenders: domain adaptation for sentiment classification. In: ACL, vol. 7, pp. 440–447 (2007)
6. Burfoot, C.: Using multiple sources of agreement information for sentiment classification of political transcripts. In: Australasian Language Technology Association Workshop, vol. 6, pp. 11–18 (2008)
7. Galley, M., McKeown, K., Hirschberg, J., Shriberg, E.: Identifying agreement and disagreement in conversational speech: use of Bayesian networks to model pragmatic dependencies. In: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, p. 669. Association for Computational Linguistics (2004)
8. Menini, S., Tonelli, S.: Agreement and disagreement: Comparison of points of view in the political domain. In: Proceedings of the 26th International Conference on Computational Linguistics, pp. 2461–2470 (2016)
9. Misra, A., Walker, M.A.: Topic independent identification of agreement and disagreement in social media dialogue. In: Conference of the Special Interest Group on Discourse and Dialogue, p. 920 (2013)
10. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al.: Scikit-learn: machine learning in python. J. Mach. Learn. Res. **12**(Oct), 2825–2830 (2011)
11. Pennebaker, J.W., Francis, M.E., Booth, R.J.: Linguistic inquiry and word count: LIWC 2001, vol. 71. Lawrence Erlbaum Associates, Mahway (2001)
12. Somasundaran, S., Wiebe, J.: 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, pp. 116–124. Association for Computational Linguistics (2010)
13. Thomas, M., Pang, B., Lee, L.: Get out the vote: determining support or opposition from congressional floor-debate transcripts. In: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, pp. 327–335. Association for Computational Linguistics (2006)
14. Turney, P.D., Pantel, P.: From frequency to meaning: vector space models of semantics. J. Artif. Intell. Res. **37**, 141–188 (2010)