

# Mining Twitter for Fine-Grained Political Opinion Polarity Classification, Ideology Detection and Sarcasm Detection

Sandeepa Kannangara  
 School of Computer Science and Engineering  
 University of New South Wales  
 Sydney NSW 2052, Australia  
 s.kannangara@unsw.edu.au

## ABSTRACT

In this paper, we propose three models for socio-political opinion polarity classification of microblog posts. Firstly, a novel probabilistic model, Joint-Entity-Sentiment-Topic (JEST) model, which captures opinions as a combination of the target entity, sentiment and topic, will be proposed. Secondly, a model for ideology detection called JEST-Ideology will be proposed to identify an individual's orientation towards topics/issues and target entities by extending the proposed opinion polarity classification framework. Finally, we propose a novel method to accurately detect sarcastic opinions by utilizing detected fine-grained opinion and ideology.

## KEYWORDS

Fine-Grained Opinion Mining; Ideology Detection; Sarcasm Detection

### ACM Reference Format:

Sandeepa Kannangara. 2018. Mining Twitter for Fine-Grained Political Opinion Polarity Classification, Ideology Detection and Sarcasm Detection. In *WSDM 2018: 11th ACM International Conference on Web Search & Data Mining*, February 5–9, 2018, Marina Del Rey, CA, USA. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3159652.3170461>

## 1 INTRODUCTION

Opinions in social media platforms provide worldwide access to what people think about daily life topics/issues. Thus exploiting such a source of information to understand public opinion can be very useful in many scenarios, for example by political parties interested in monitoring attitudes towards their policies.

The opinion mining research field aims to develop automated approaches to accurately analyse such opinion data. Although there is much previous work on opinion mining, the majority of early studies analysed text documents such as product and movie reviews [5]. Only a limited number of studies have attempted to analyse public opinion in a political context [4]. The nature of political discourse which often includes sarcasm and irony makes the analysis more challenging.

Most political opinion mining studies analysed election datasets with the aim of election result prediction. The approaches employed

range from lexicon based methods [8] to supervised machine learning methods using algorithms such as SVM [1] and Deep Learning approaches [11] in recent years. However, most work in this area focuses on overall sentiment classification, though Maynard and Funk [8] considered identification of the opinion target and Vijayaraghavan *et al.* [11] addressed classification of the topics along with the sentiment.

An *opinion* is defined as a tuple of four components [7]: sentiment orientation, sentiment target, opinion holder, and time. The sentiment target is defined as an entity or an entity with possibly an aspect of the entity that the sentiment has been expressed upon. As an example, if a tweet expresses a negative sentiment towards the Labor Party regarding Employment/Jobs, Labor Party is the target entity and Employment/Jobs is the aspect of the entity. However, the majority of early studies in opinion mining attempted to detect only the overall sentiment [6], regardless of the entities mentioned and their aspects. After Hu and Liu's [5] study, fine-grained opinion mining/aspect-based opinion mining became prominent. However, most of the aspect-based opinion mining studies [9] identified only aspect and sentiment by assuming a known target.

In this research, we aim to address the current challenges in socio-political opinion mining from Twitter focusing on the identification of target entity, aspect and sentiment for fine-grained opinion mining. Identifying an individual's political ideology/orientation from their reviews is also an important task for analysing political opinions. In general, opinions expressed toward the issues/topics of target entities are often indicative of political ideologies. For that, we propose a novel approach utilising the proposed opinion mining framework. Further, the presence of sarcasm is identified as a challenge for opinion polarity classification since sarcasm can potentially flip the polarity of the review. In this research, we will propose a novel method to detect sarcasm in tweets.

## 2 THE PROPOSED RESEARCH

Overall, the aim of this research is to develop a framework for political opinion mining focusing on fine-grained opinion mining, sarcasm detection and ideology detection. To evaluate each proposed model, we use tweets collected during the 2016 Australasian Federal Election.

### 2.1 Fine-Grained Political Opinion Polarity Classification

In this study, for fine-grained political opinion mining, we aim to extract the opinion target (e.g. candidates and political parties), topics discussed (e.g. Tax/Negative Gearing, Asylum Seekers/Refugees,

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

WSDM 2018, February 5–9, 2018, Marina Del Rey, CA, USA

© 2018 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-5581-0/18/02.

<https://doi.org/10.1145/3159652.3170461>

etc.) and sentiment (Positive, Negative or Neutral) towards the target from each tweet. For this, we propose a fine-grained opinion mining model, called JEST (Joint-Entity-Sentiment-Topic), to jointly model all three dimensions, based on LDA [3].

The strength of JEST is that all the component dimensions can help to improve each other during the joint modelling process. The rationale here is that it is possible to model each dimension as a latent variable in a graphical model. To distinguish the aspect and entity in the model, we utilize a knowledge base of entities which contains common names and hashtags used to represent entities in a predefined entity list. In this model, we assume that there is only one target entity and several or no topics in each tweet.

For evaluation, we use perplexity, which is the standard measure for evaluating topic models to compare the model with the selected baseline models. Furthermore, we will evaluate the document classification performance using distributions of the entity, target and sentiment generated by the model. However, since there is no prior work that models these three dimensions with similar assumptions, we need to select different baselines to evaluate each dimension. Furthermore, we can do qualitative analysis by manually inspecting the quality of the output.

## 2.2 Political Orientation Detection

For political orientation detection, we aim to identify a Twitter user's political alignment. However, our definition of political alignment is a user's set of political preferences over a range of policy positions of a party. For this, knowing the information such as user's opinion towards candidates or parties and topics that are discussed in his/her tweets is important. We model such evidence in our JEST model and here we propose an extension of JEST, the JEST-Ideology model, incorporating the ideology dimension.

The JEST model is capable of identifying an entity, sentiment and topics for each tweet in the corpus. The JEST-Ideology model has a top layer which can identify the favourable topics and entities for each user. However, there can be tweets that do not contain topics but generally express an opinion towards an entity. We consider such situations by modelling the general ideology of the user towards entities. The generative process runs through each user and then tweets of each user. For the evaluation of the model, we can use the same approaches proposed for the JEST model.

## 2.3 Sarcasm Detection in Political Tweets

Most previous approaches to sarcasm detection treat it as a purely linguistic matter, using information such as lexical cues and their corresponding sentiment as predictive features [2]. However, sarcasm has been extensively studied in the behavioural sciences and there are several theories explaining when, how and why sarcasm is expressed [10]. Those theories can be represented as contextual features by consider the user's current and previous posts.

In this research, we propose to use both linguistic and contextual features to detect sarcastic tweets. The accepted view is that sarcasm potentially flip the polarity of a review. Therefore, when we know the surface polarity of a tweet, to identify the situations where the actual polarity is different from surface polarity we can use the ideology of the author. If the ideology of the author is different from the surface polarity then we can infer that it is a sarcastic

tweet. Therefore, as contextual features for the supervised model, we can use the ideology of the Twitter user and surface polarity by utilizing the results of JEST and JEST-Ideology models.

The performance of the proposed approach can be measured using accuracy, recall, precision and F-measure. Further, we will compare our models with existing sarcasm detection techniques.

## 3 CONCLUSION AND FUTURE WORK

This paper gives a brief overview of ongoing research on political opinion mining. We have identified three main research directions to automatically analyse political opinions in Twitter: fine-grained opinion polarity classification, ideology detection and sarcasm detection. We propose novel topic models, JEST and JEST-Ideology, which are capable of fine-grained opinion mining and ideology detection respectively. Finally, with the aim of improving opinion polarity classification performance, we propose a novel sarcasm detection model which uses ideology and fine-grained opinion as features with other linguistic features to classify sarcastic opinions.

In the future, it is essential to build these proposed models and evaluate model performance. A number of experiments are planned to select best performing model for each problem. Although, the models are proposed for the political domain, they are also applicable for other domain. Therefore, we are planning to conduct experiments using other datasets from different domains to evaluate the performance of the proposed models.

## 4 ACKNOWLEDGMENTS

I would like to sincerely thank my supervisor Wayne Wobcke, and Michael Bain, Alfred Krzywicki and Bradford Heap for their continuous mentoring and constant support. The work is supported by the Data to Decisions Cooperative Research Centre.

## REFERENCES

- [1] Akshat Bakliwal, Jennifer Foster, Jennifer van der Puil, Ron O'Brien, Lamia Toussi, and Mark Hughes. 2013. Sentiment Analysis of Political Tweets: Towards an Accurate Classifier. In *Proceedings of the Workshop on Language in Social Media (LASM '13)*. 49–58.
- [2] David Bamman and Noah A. Smith. 2015. Contextualized Sarcasm Detection on Twitter. In *Proceedings of the Ninth International AACL Conference on Web and Social Media (ICWSM '15)*. 574–577.
- [3] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3 (2003), 993–1022.
- [4] Hsinchun Chen and David Zimbra. 2010. AI and Opinion Mining. *IEEE Intelligent Systems* 25, 3 (2010), 74–80.
- [5] Mingqing Hu and Bing Liu. 2004. Mining and Summarizing Customer Reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '04)*. 168–177.
- [6] Farhan Hassan Khan, Saba Bashir, and Usman Qamar. 2014. TOM: Twitter Opinion Mining Framework Using Hybrid Classification Scheme. *Decision Support Systems* 57 (2014), 245–257.
- [7] Bing Liu. 2012. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool.
- [8] Diana Maynard and Adam Funk. 2012. Automatic Detection of Political Opinions in Tweets. In *The Semantic Web: ESWC 2011 Workshops*, Raúl García-Castro, Dieter Fensel, and Grigoris Antoniou (Eds.). Springer-Verlag, Berlin, 88–99.
- [9] Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. 2011. Opinion Word Expansion and Target Extraction through Double Propagation. *Computational Linguistics* 37, 1 (2011), 9–27.
- [10] Ashwin Rajadesingan, Reza Zafarani, and Huan Liu. 2015. Sarcasm Detection on Twitter: A Behavioral Modeling Approach. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining (WSDM '15)*. 97–106.
- [11] Prashanth Vijayaraghavan, Soroush Vosoughi, and Deb Roy. 2016. Automatic Detection and Categorization of Election-Related Tweets. In *Proceedings of the Tenth International Conference on Web and Social Media (ICWSM '16)*. 703–706.