Investigating Transfer Learning Approaches for Mining Opinions in the Electoral Domain

Anonymous Author(s) *

Abstract

The use of social media data to mine opinions during elections has emerged as an alternative to traditional election polls. However, relying on social media data in electoral scenarios comes with a number of challenges, such as tackling sentences with domain specific terms, texts full of hate speech, noisy, informal vocabulary, sarcasm and irony. Also, in Twitter, for instance, loss of context may occur due to the imposed limit of characters to the posts. Furthermore, prediction tasks that use machine learning require labeled datasets and it is not trivial to reliably annotate them during the short period of campaigns. Motivated by these issues, we investigate how to boost and speed-up the performance of opinion mining tasks during elections. We start by proposing a transfer learning approach that leverages curated datasets from other domains. To avoid negative transfer, *i.e.* introducing a knowledge from the other domains that could end up by disturbing the task, we propose to use similarity metrics (Jaccard, Cosine and Euclidean distance based on word embeddings) to point out whether or not the dataset should be used. Our preliminary results show that taking into account the (dis)similarity between different domains, it is possible to achieve results closer to the ones that would be achieved with classifiers trained with annotated datasets of the electoral domain.

1 Research Problem and Motivation

In democratic systems, *election polls* play an essential role. Once they measure voting intention [6], they are used by the candidates and their parties to adjust their campaigns and better communicate proposals [6]. In turn, their results can affect election outcomes [7], by influencing people who have not yet decided in which candidate to vote. However, predicting electorate preferences following the traditional poll methodology brings two main drawbacks [12]: (i) it demands much time to be conducted; and (ii) it demands high monetary costs. In order to overcome these drawbacks, a number of approaches in the literature have proposed to predict voting intention by applying machine learning and sentiment analysis techniques to data collected from social media [13], [3, 1], [14]. The negative/positive sentiment towards the candidates is inferred from the social media sample and, from that, it is possible to point out the one that seems to be the favorite among people. Existing approaches for predicting electoral trends/outcomes based on social media usually rely on Twitter as the source of opinions and present many pitfalls. In summary, the difficulty of collecting and labeling a large number of tweets during the short period of elections caused that many approaches choose to conduct a post-hoc analysis of electoral tweets, *i.e.* they only can analyze tweets *after* the occurrence of the real elections [9]. In this way, most of the approaches that try to predict election results do not consider information specific from the domain to assign polarities, relying only on generic lexical dictionaries [3], [19], [18] or using methods in which tweets are automatically labeled according to emoticons [10], [6].

In this research we are proposing to use existing sentiment analysis datasets from other domains as *starting point* to construct models for sentiment analysis to be applied in electoral scenarios. Ideally, this would avoid (or at least reduce) the need for manually label electoral datasets and would enable

* This research is being conducted by three Latin individuals.

the analysis/prediction of elections during their course. This task can be seen as an instance of domain transfer learning [15]. Taking advantage of existing datasets from other domains is not trivial, because electoral data collected from social media have several particularities, as for example: they contain specific electoral/political terms that change over time [4]; in addition, when collected from *social media*, this kind of data may contain characteristics that do not necessarily occur (or that occur but with less intensity [8], [21]) in other domains, such as *hate speech*, *data noise* (spam) due to political bots and fake users, high levels of *sarcasm/ irony*. Motivated by those particular issues and to avoid transferring negative knowledge, we propose to rely on similarity metrics to select the most promising existing datasets.

2 Preliminary Experiments and Conclusions

The case study adopted in this research was based on predicting sentiment of data about the 2018 Brazilian Presidential Elections. We selected five sentiment analysis datasets written in Brazilian Portuguese to serve as source data. They include different domains, namely TV shows, urban problems, restaurants, movies, and a dataset of general domain. It is worth noticing that dealing with the datasets written in Brazilian Portuguese introduces another challenge, in contrast to English language, as the number of existing tools for text preprocessing and the existing datasets labeled for sentiment analysis in non-english languages are very limited [2].

Before building the machine learning classifier to predict the sentiment of tweets related to the electoral sample (*target domain*), all datasets were balanced containing about 2000 instances per class. We adopted the Support Vector Machine [20] (SVM) algorithm with linear kernel to train the classifiers, which were built to each one of the datasets and were applied on the target dataset for comparison purposes. We get the vocabulary of each one of the datasets when vectorizing them with the TF-IDF method. After that, three similarity metrics were considered, (i) Jaccard distance (d_J): a lemmatization step is performed to shrink each word of the vocabulary to its root form. Next, Jaccard was calculated between each dataset and the target dataset (elections) taking into account their vocabulary; (ii) Cosine distance (d_{Cos}): the Cosine distance between the bag of word vectors of the datasets was calculated; (iii) Euclidean distance (d_E): we are proposing to use a pretrained word embedding which was trained using the Glove [16] algorithm with a huge corpus of portuguese texts. For each dataset, we get the embeddings values (based on the pretrained word embeddings) of each word of the vocabulary and compute the average of all these embeddings values. The distances between each dataset and the target dataset were calculated, with the the well-known metric Euclidean distance, taking as input the average of embedding values of each dataset. Several combinations of datasets were considered to train classifiers: (i) a classifier was trained by merging data from the two most similar datasets; (ii) a classifier was trained by merging data of the two most dissimilar datasets; (iii) a classifier was trained by merging data of the three most similar datasets; (iv) a classifier was trained by merging data of the four most similar datasets; and (v) a classifier was trained by merging data of all (five) datasets. The 10-fold cross-validation technique was adopted for evaluating each classifier when tested on its own domain. The comparison of the classifiers was conducted based on the results of the metric *F1-score*.

Our preliminary results showed that taking advantage of existing labeled datasets from other domains is a strategy that can help one to achieve better results when the similarity between domains is exploited. On the other hand, combining data of disparate domains can reduce the classifier's results when the similarity between them is low, reflecting in the negative transfer learning. In this context, our experiments show that the results achieved were very different according to the distance between the datasets merged and the target dataset. Analyzing the similarity between datasets before using them for training classifiers can be very helpful independently of the domain because it can prevent one for training a classifier (task that may be time-costly and computationally-costly) using unrelated data. Due to space limitations, the detailed results of our experiments are not presented.

As a future work, we intend to investigate other similarity methods that measure not only similarity between datasets vocabularies but also between context/semantics of the words in each domain, *i.e.*, considering cases of words that appear in one domain with positive/negative connotation and appear in another domain associated with opposite or neutral sentiment. In addition, we also intend to investigate the usage of transfer learning techniques related to language models such as ULMFit [11], ELMo [17] and BERT [5] to verify if they can be useful to improve the results of the task of predicting sentiment for Brazilian presidential elections.

References

[1] S. Asur and B. A. Huberman. Predicting the future with social media. In *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology-Volume 01*, pages 492–499. IEEE Computer Society, 2010.

[2] V. Basile et al. Sentiment analysis on italian tweets. In *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 100–107, 2013.

[3] P. Burnap, R. Gibson, L. Sloan, R. Southern, and M. Williams. 140 characters to victory?: Using twitter to predict the uk 2015 general election. *Electoral Studies*, 41:230–233, 2016.

[4] P. H. Calais Guerra, A. Veloso, W. Meira Jr, and V. Almeida. From bias to opinion: a transfer-learning approach to real-time sentiment analysis. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 150–158. ACM, 2011.

[5] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[6] N. Dwi Prasetyo and C. Hauff. Twitter-based election prediction in the developing world. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, pages 149–158. ACM, 2015.

[7] R. Forsythe, R. B. Myerson, T. A. Rietz, and R. J. Weber. An experiment on coordination in multi-candidate elections: The importance of polls and election histories. *Social Choice and Welfare*, 10(3):223–247, 1993.

[8] L. Gao, A. Kuppersmith, and R. Huang. Recognizing explicit and implicit hate speech using a weakly supervised two-path bootstrapping approach. *arXiv preprint arXiv:1710.07394*, 2017.

[9] D. Gayo-Avello. No, you cannot predict elections with twitter. *IEEE Internet Computing*, 16(6):91–94, 2012.

[10] B. Heredia, J. Prusa, and T. Khoshgoftaar. Exploring the effectiveness of twitter at polling the united states 2016 presidential election. In *2017 IEEE 3rd International Conference on Collaboration and Internet Computing (CIC)*, pages 283–290. IEEE, 2017.

[11] J. Howard and S. Ruder. Universal language model fine-tuning for text classification. *arXiv preprint arXiv:1801.06146*, 2018.

[12] M. Ibrahim, O. Abdillah, A. F. Wicaksono, and M. Adriani. Buzzer detection and sentiment analysis for predicting presidential election results in a twitter nation. In *Data Mining Workshop (ICDMW), 2015 IEEE International Conference on*, pages 1348–1353. IEEE, 2015.

[13] M. Maldonado and V. Sierra. Can social media predict voter intention in elections? The case of the 2012 dominican republic presidential election. In *AMCIS-2015*, 2015.

[14] E. Martínez-Cámara, M. T. Martín-Valdivia, L. A. Urena-López, and A. R. Montejo-Ráez. Sentiment analysis in twitter. *Natural Language Engineering*, 20(1):1–28, 2014.

[15] S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.

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

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

[18] A. Tsakalidis, S. Papadopoulos, A. I. Cristea, and Y. Kompatsiaris. Predicting elections for multiple countries using twitter and polls. *IEEE Intelligent Systems*, 30(2):10–17, 2015.

[19] S. Unankard, X. Li, M. Sharaf, J. Zhong, and X. Li. Predicting elections from social networks based on sub-event detection and sentiment analysis. In *International Conference on Web Information Systems Engineering*, pages 1–16. Springer, 2014.

[20] L. Wang. *Support vector machines: theory and applications*, volume 177. Springer Science & Business Media, 2005.

[21] S. C. Woolley. Automating power: Social bot interference in global politics. *First Monday*, 21(4), 2016.