SentiDiff Combining Textual Information and Sentiment Diffusion Patterns for Twitter Sentiment Analysis

ABBURI.RAMESH ASSOC.PROFESSOR Narasaraopeta Institute of Technology

POTU BHARATH ASST.PROFESSOR Narasaraopeta Institute of Technology

ABSTRACT

In recent years, Twitter sentiment analysis has become a popular research subject. Most current Twitter sentiment analysis tools only accept textual input from Twitter tweets and fail to work well when confronted with brief and vague Twitter messages. Recent research indicates that emotion diffusion trends on Twitter are closely related to the sentiment polarities of Twitter tweets. As a result, in this article, we concentrate on how to combine textual knowledge from Twitter messages and sentiment diffusion trends to improve sentiment analysis output on Twitter details. To that end, we first investigate a process known as sentiment reversal and discover certain intriguing properties of sentiment reversals. Then, taking into account the interrelationships between textual knowledge in Twitter messages and sentiment diffusion trends, we suggest SentiDiff, an iterative algorithm for predicting sentiment polarities conveyed in Twitter messages. To the best of our understanding, this is the first study to use emotion diffusion trends to aid in Twitter sentiment analysis. Extensive tests on real-world datasets show that, as compared to state-of-the-art textual information-based sentiment analysis algorithms, our proposed algorithm improves PR-AUC on Twitter sentiment classification tasks by 5.09% and 8.38 percent.

Key words: Sentidiff, Sentiment analysis, sentiment diffusion, social networks, feature, graph analysis.

INTRODUCTION

Twitter, a common microblogging site around the globe, has shaped and transformed the way users receive knowledge from people or organizations of interest to them. Tweets are status update updates that users will send to their friends to let them know what they are thinking, doing, or what is going on around them.

Users may also communicate with other users by replying to or reposting their messages. Since its inception in 2006, Twitter has grown to become one of the world's leading online social networking sites [1]. Mining users' sentiment polarities reflected in Twitter messages has become a popular research subject due to its wide applications, given the ever-increasing volume of data accessible from Twitter [2]. Several methods have been developed to include democratic election tactics, for example, by examining Twitter users' opinion polarities on political parties and candidates [3]. (4th). Twitter opinion analysis is now used by businesses as a quick and efficient way to track people's thoughts towards their goods and brands [5]. The aim of sentiment analysis on Twitter data is to categories a Twitter message's sentiment polarity as positive, favorable, or negative. One approach for doing Twitter sentiment analysis is to use standard text sentiment analysis approaches [6].

Twitter tweets, on the other hand, are often brief and vague, in contrast to other text types such as press stories and book posts. Furthermore, because of the informal nature of Twitter tweets, there are more slangs, acronyms, misspelt sentences, and modal particles. [7, 8] Formalized paraphrase as a consequence, when used to estimate the emotion polarities of Twitter tweets, standard text sentiment analysis algorithms do significantly worse. Many novel sentiment analysis approaches for Twitter tweets have been created to address this problem. This method is broadly classified into two types: directly supervised methods and distantly supervised methods [9]. Completely controlled methods are designed to acquire sentiment classifiers from manually labelled data and sentiment lexicons [10] [11]. One major issue with completely supervised approaches is that manually building sentiment lexicons and labelling data is time-consuming and labor-intensive, and as a result, the sentiment lexicons and labelled data used by most methods are often insufficient to ensure decent results. Furthermore, fully controlled approaches typically focus on handcrafted features, and designing successful features remains a difficult challenge. The distantly supervised methods learn sentiment classifiers from noisy labels such as emoticons and hashtags in results. These approaches use emoticons such as “:)” and “:(” as loud identifiers for sentiment analysis, assuming that a message containing “:(” is more likely to convey a negative sentiment polarity and one containing “:)” is more likely to express a favorable sentiment polarity [7]. [12] Formal paraphrase While these distantly monitored approaches prevent labor-intensive manual annotation, their efficiency is inadequate due to mark noise [9]. [13] Formal paraphrase

Pre-processing approaches may help to mitigate the issue of noisy labels in sentiment analysis [14]. Recent research, however, has shown that there are no efficient pre1041-processing approaches for both datasets and algorithms [15]. When one pre-processing approach is efficient with one algorithm and one dataset, it may result in a decrease in sentiment analysis efficiency when extended to another dataset or algorithm. In general, all directly controlled and distantly supervised Twitter sentiment analysis solutions mainly rely on textual knowledge from Twitter messages and are unable to obtain adequate efficiency due to the particular characteristics of Twitter messages. Sentiment diffusion, which is primarily concerned with studying how emotions influence knowledge diffusion in social networks, has already piqued the interest of several academic groups [16]. [17] Formal paraphrase [18] Formalized paraphrase Users on Twitter will repost another Twitter user's message and share it (i.e., retweet) with their own followers by pressing the retweet button inside the tweet (or simply typing "RT" or "through" at the beginning of a tweet to signify that they are reposting anyone else's content).

When reposting a message, users will leave a note about it and paste it alongside the initial tweet (some tweets are reposted without any added comments, and these retweets are often ignored in sentiment diffusion studies as it is hard to know the sentiments expressed in these retweets). Tweets and retweets may therefore relay details regarding their authors' sentiment polarities on a given subject. As a result, we should look at how emotion polarities vary from a tweet to its retweets to explore sentiment diffusion on Twitter [19]. Recently, organically fusing information from different realms (but theoretically connected) has opened up new avenues for study in several machine learning and data mining tasks [20] [21].

Recent research on opinion diffusion indicate that individual’s users follow on Twitter affect their sentiment polarities [22], as do their roles inside knowledge dissemination processes [23]. Despite the fact that sentiment diffusion trends are closely related to sentiment polarities in Twitter messages, recent work on Twitter sentiment analysis focuses mostly on the textual details of Twitter messages and lacks sentiment diffusion information. Given the drawbacks of current Twitter sentiment analysis solutions that only recognize textual details, as well as the near relationships between sentiment diffusion trends and sentiment polarities of Twitter messages, we claim that combining textual knowledge from Twitter messages and sentiment diffusion information in a supervised learning environment is the best approach. However, integrating these two types of knowledge organically within the same learning system remains a problem.

In this paper, we propose SentiDiff, a novel algorithm for dealing with this issue. This paper's major contributions are outlined below. We investigate sentiment reversal, the process under which a message and its retweet have opposite sentiment polarities, on Twitter. We investigate the properties of sentiment reversals and suggest a model for sentiment reversal prediction. To predict the sentiment polarity of each Twitter message, we propose SentiDiff, an iterative algorithm that takes into account the interrelationships between textual knowledge in Twitter messages and sentiment diffusion trends. If the sentiment polarities expected by a textual information based sentiment classifier for a tweet and its retweet are compatible with the prediction outcome of sentiment reversal, the chance of tweets being identified correctly by a textual information based sentiment classifier increases. Otherwise, the likelihood would decline. Sentiment reversals may therefore be paired with textual knowledge from Twitter tweets. To test the accuracy of our proposed algorithm, we run a series of experiments. The experimental findings indicate that our proposed SentiDiff algorithm aids state-of-the-art textual information-based sentiment analysis algorithms in achieving PR-AUC improvements ranging from 5:09 to 8:38 percent.

LITERATURE REVIEW

Arampatzis, D. Effrosynidis, and S. Symeonidis[1]: Sentiment analysis in microblogging networks is a critical method for both science and market applications. Machine learning processes that analyse human emotion and understand human writings assist us in drawing valuable conclusions regarding human behaviour. Pre-processing is the first phase in text Sentiment Analysis, and utilising suitable techniques such as Linear SVC, Bernoulli Nave Bayes, Logistic Regression, and Convolutional Neural Networks may increase classification effectiveness. However, the detection precision of this paper is poor since it worked on lemmatization, deleting quantities, and replacing contractions techniques.

J. Zhao and X. Gui[2]: This paper explored the impact of text pre-processing methods on sentiment classification results in two categories of classification activities, and summed up the classification outputs of six pre-processing methods on five Twitter databases utilising two function models and four classifiers. However, since the author operated with static Twitter info, the training output is poor.

X. Zhang, D.-D. Han, R. Yang, and Z. Zhang[3]: The authors of this paper use analytical evidence crawled from Twitter to explain the topology and knowledge spreading mechanisms of Online Social Networks. Propose a calculation of three steps to state Twitter users' attempts to distribute their content, centred on Twitter's specific processes for information retransmission. It has been observed that a limited percentage of users with exceptional participation output will wield significant power, whereas the majority of other users serve as middleware during knowledge dissemination. However, deleting the missing data would result in the loss of user profile and user action records.

F. Frasincar and K. Schouten[4]: This paper described. In this survey, an overview of the state-of-the-art of aspect level sentiment analysis is given, and it is evident that the field has progressed beyond its early stages. While in some instances a comprehensive solution capable of jointly performing aspect detection and sentiment analysis is proposed, in others dedicated algorithms for either of those two tasks are given. The majority of methods discussed in this survey use machine learning to model language, which is not shocking considering that language is a non-random, very complicated phenomena for which a large amount of data is accessible. This article, on the other hand, introduces cutting-edge techniques for emotion analysis.

H. Ohsaki and S. Tsugawa[5]: They examined the relationship between a tweet's sentiment and its virality in terms of dissemination volume and speed by reviewing 4.1 million tweets on Twitter. They measured tweet virality using the amount of retweets and the N-retweet period. They discovered that when the diffusion volume was high, negative tweets spread more broadly than positive and neutral tweets, and that negative tweets spread faster than positive and neutral tweets.

However, the author investigated the relationship between the sentiment of each tweet and its virality. Calculating the connection function method is extremely challenging.

S. M. Mohammad and S. Kiritchenko[6] equate the efficiency of multiple term and character-based recurrent and convolutional neural networks with the performance on bag-of-words in this article. We also explore the transferability of the final secret state representations through various emotional classifications, and whether it is feasible to construct a unison model that predicts all of them using a shared representation. The poet, on the other hand, focused on bag of words techniques.

D. Hovy and B. Plank[7]: This paper examines two basic NLP tasks: discourse parsing and sentiment analysis. The development of three separate recurrent neural nets: two for the main subtasks of discourse parsing, structure prediction and connection prediction, and one for emotion prediction. However, since this job is performed by hand, it is time consuming and costly.

J. Martin, S. M. Mohammad, X. Zhu, S. Kiritchenko, and S. M. Mohammad[8]: This paper investigated the use of deep recurrent neural networks for sentence-level opinion speech extraction. DSEs (direct subjective expressions) are overt references of private states or speech activities expressing private states, while ESEs (expressive subjective expressions) are expressions that signify feeling, emotion, and so on without directly conveying them. Nonetheless, primarily For the device, this is both time and money intensive.

J. Bollen, H. Mao, and X.-J. Zeng[9]: In this article, we examine electoral tweets for more overtly articulated details such as mood (positive or negative), emotion (joy, sorrow, indignation, etc.), meaning or motive (to point out an error, to help, to mock, etc.), and tweet design (simple statement, sarcasm, hyperbole, etc.).

METHODOLOGY

The method suggests a new algorithm named SentiDiff for this challenge in the proposed scheme. This paper outlines the major contributions as follows. The method explores the diffusión of sentiments on

Twitter by examining the reversal of a feeling, the phenomenon of a message and its retwit. The method examines the characteristics of feeling changes and offers a prediction reversal model. We propose an iterative algorithm named SentiDiff to anticipate the sensitivity polarity of each Twitter message which takes account of the interrelationships among Twitter textual knowledge and sentiment propagation patterns. When a tweet and retweet is present, the likelihood of tweets being identified correctly by text-based sentiment classifier increases if their feelings polarity predictions are compatible with the prediction resulting from sentiment reversal. The likelihood would decrease otherwise. This can mix feeling reversals with Twitter textual content. The machine performs a series of experiments to test our proposed algorithm's accuracy. The experimental findings indicate that our proposed SentiDiff algorithm allows state-of-the-art sentiment analysis algorithms to boost PR-AUC between 5/9% and 8:38%.

Advantages

- The SentiDiff an algorithm is a general construct, which can quickly be applied in order to forecast the feelings polarities of messages from other social networks online.

- A sentiment reversal paradigm is suggested by the processes in the proposed scheme.

RESULTS & EVALUATION

(i) PERFORMANCE OF SENTIMENT REVERSAL PREDICTION

Here are the following: We can conclude that we have an efficient predictive model for reverse feelings with a PR-AUC classification of 81.63% after all of the function sets have been used. We also examine how all features (such as the cascading tree, the diffusion network and user history) can influence the prediction output only with regard to one feature at a time. We will find that the cascade trees will produce the highest output by themselves, which confirms that the flexibility of the cascade tree is the main aspect for reversing the feelings. If we just take account of diffusion network characteristics or past behavior, prediction accuracy of feeling reversals is not adequate. Our data collection contains additional times for new Twitter users and thus a number of users will not be obtained. We cannot remove your diffused network and historical behavior attributes from the training set by estimating reversals of perception between two new Twitter users, contributing to a low forecast of reversals of perceptions.

(ii) Effect of Fusing Textual and Sentiment Diffusion Information

For all sentiment analytics which only take Twitter textual information into consideration, we may observe that, following the combination of textual and sentiment diffusion information in a controlled learning algorithm, they can achieve dramatically improved results. After this mix SentiDiff delivers gains in Twitter sentiment classification between 5.09 and 8.38 percent for PR-AUC tasks that verify the efficiency of SentiDiff by combining textual and emotional knowledge for Twitter feeling research. The Fast Text model will produce the best results of six textual knowledge models if only texts for Twitter messages are taken into consideration. However, after the fusion of textual and sentimental content, the highest PR-AUC is achieved via the deep CNN-based model.

(iii) Effect of Amount of Training Data

We carry out PR-AUC experiments for a classifier, reverse prediction model of sentiment, and SentiDiff Algorithm based on textual knowledge. The deep CNN model [43] in this segment is used as a textual classifier of feelings dependent on details. We carry out PR-AUC computer experiments with classifier of feelings based on textual knowledge, model of a reverse predictor of the sentiment and algorithm of SentiDiff. In this segment, we use the deep CNN model to classify feelings based on textual details. Combining textual knowledge and information regarding emotion diffusion would have a detrimental effect on our sentiment research on Twitter. And if feeling reverse forecast outcomes are not accurate, feeling diffusion knowledge reduces the likelihood that Twitter messages are correctly identified by the textual information dependent classifier.

CONCLUSION

Mining sentiment polarities conveyed in Twitter tweets is a worthwhile but difficult job. Most current Twitter sentiment analysis tools still accept textual input from Twitter messages and are unable to achieve adequate results due to the peculiar characteristics of Twitter messages. Despite recent research indicating that sentiment diffusion trends are closely related to sentiment polarities of Twitter messages, current methods mostly concentrate on textual details of Twitter messages while ignoring sentiment diffusion information. We take a first step toward integrating textual and sentiment diffusion details to improve the efficiency of Twitter sentiment analysis, inspired by recent work on knowledge fusion from multiple domains. To that end, we first investigate a phenomena known as sentiment reversal on Twitter and discover some fascinating properties of sentiment reversals using repost cascade trees and repost diffusion networks. We then build a sentiment reversal prediction model and SentiDiff, a novel Twitter sentiment classification algorithm. SentiDiff takes into account the interrelationships between textual knowledge from Twitter messages and sentiment diffusion trends, and the textual information-based sentiment classifier and sentiment reversal prediction model are integrated in a supervised learning environment. Experiments on real-world datasets show that our proposed SentiDiff algorithm will assist state-of-the-art textual information-based sentiment analysis algorithms in achieving PR-AUC improvements ranging from 5:09 to 8:38 percent. In the future, we intend to investigate how sentiment diffusion trends vary across topics, as well as take into account the subject details of Twitter messages while fusing textual and sentiment diffusion data.

REFERENCES

[1] H. Li, L. Dombrowski, and E. Brady, “Working toward empowering a community: How immigrant-focused nonprofit organizations use twitter during political conflicts,” in Proceedings of the 2018 ACM Conference on Supporting Groupwork. ACM, 2018, pp. 335–346.

[2] D. Tang, B. Qin, F. Wei, L. Dong, T. Liu, and M. Zhou, “A joint segmentation and classification framework for sentence level sentiment classification,” IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 23, no. 11, pp. 1750–1761, 2015.

[3] H. Wang, D. Can, A. Kazemzadeh, F. Bar, and S. Narayanan, "A system for real-time twitter sentiment analysis of 2012 us presidential election cycle," in Proceedings of the ACL 2012 System Demonstrations. Association for Computational Linguistics, 2012, pp. 115–120.

[4] D. Paul, F. Li, M. K. Teja, X. Yu, and R. Frost, "Compass: Spatio temporal sentiment analysis of us election what twitter says!" in Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2017, pp. 1585–1594.

[5] F. Bravo-Marquez, E. Frank, and B. P fahringer, "Annotate-sample average (asa): A new distant supervision approach for twitter sentiment analysis," in 22nd European Conference on Artificial Intelligence (ECAI), vol. 285. IOS Press, 2016, pp. 498–506.

[6] B. Pang, L. Lee et al., "Opinion mining and sentiment analysis," Foundations and Trends R in Information Retrieval, vol. 2, no. 1–2, pp. 1–135, 2008.

[7] K.-L. Liu, W.-J. Li, and M. Guo, "Emoticon smoothed language models for twitter sentiment analysis," in Proceedings of the Twenty- Sixth AAAI Conference on Artificial Intelligence, 2012.

[8] J. Zhao, L. Dong, J.Wu, and K. Xu, "Moodlens: an emoticon-based sentiment analysis system for chinese tweets," in Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2012, pp. 1528–1531.

[9] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," CS224N Project Report, Stanford, vol. 1, no. 12, 2009.

[10] D.-T. Vo and Y. Zhang, "Target-dependent twitter sentiment classification with rich automatic features." in IJCAI, 2015, pp. 1347– 1353.

[11] E. Cambria, "Affective computing and sentiment analysis," IEEE Intelligent Systems, vol. 31, no. 2, pp. 102–107, 2016.

[12] D. Tang, F. Wei, N. Yang, M. Zhou, T. Liu, and B. Qin, "Learning sentiment-specific word embedding for twitter sentiment classification," in Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, 2014, pp. 1555–1565.

[13] K. Schouten and F. Frasincar, "Survey on aspect-level sentiment analysis," IEEE Transactions on Knowledge and Data Engineering, vol. 28, no. 3, pp. 813–830, 2016.

[14] J. Zhao and X. Gui, "Comparison research on text pre-processing methods on twitter sentiment analysis," IEEE Access, vol. 5, pp. 2870–2879, 2017.

[15] S. Symeonidis, D. Effrosynidis, and A. Arampatzis, "A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis," Expert Systems with Applications, 2018.

[16] M. Thelwall, K. Buckley, and G. Paltoglou, “Sentiment in twitter events,” Journal of the Association for Information Science and Technology, vol. 62, no. 2, pp. 406–418, 2011.

[17] N. Du, Y. Liang, M. Balcan, and L. Song, “Influence function learning in information diffusion networks,” in International Conference on Machine Learning, 2014, pp. 2016–2024.

[18] M. Tsytsarau, T. Palpanas, and M. Castellanos, “Dynamics of news events and social media reaction,” in Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2014, pp. 901–910.

[19] S. Stieglitz and L. Dang-Xuan, “Emotions and information diffusion in social media sentiment of microblogs and sharing behavior,” Journal of Management Information Systems, vol. 29, no. 4, pp. 217–248, 2013.

[20] Y. Fu, Y. Ge, Y. Zheng, Z. Yao, Y. Liu, H. Xiong, and J. Yuan, “Sparse real estate ranking with online user reviews and offline moving behaviors,” in 2014 IEEE International Conference on Data Mining (ICDM). IEEE, 2014, pp. 120–129.

[21] Y. Zheng, “Methodologies for cross-domain data fusion: An overview,” IEEE transactions on big data, vol. 1, no. 1, pp. 16–34, 2015.

[22] J. Tang and A. Fong, “Sentiment diffusion in large scale social networks,” in 2013 IEEE International Conference on Consumer Electronics (ICCE). IEEE, 2013, pp. 244–245.

[23] M. Miller, C. Sathi, D. Wiesenthal, J. Leskovec, and C. Potts, “Sentiment flow through hyperlink networks,” in ICWSM, 2011.

[24] X. Zhang, D.-D. Han, R. Yang, and Z. Zhang, “Users participation and social influence during information spreading on twitter,” PloS one, vol. 12, no. 9, p. e0183290, 2017.

[25] P. Nakov, A. Ritter, S. Rosenthal, F. Sebastiani, and V. Stoyanov, “Semeval-2016 task 4: Sentiment analysis in twitter,” in Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), 2016, pp. 1–18.

