BINARY CLASSIFICATION OF TWITTER POSTS FOR ADVERSE DRUG REACTIONS

JITENDRA JONNAGADDALA [1, 2], TONI ROSE JUE [2], HONG-JIE DAI [3,*]

[1] *School of Public Health and Community Medicine*
University of New South Wales, Australia

[2] *Prince of Wales Clinical School*
University of New South Wales, Australia
Email: { z3339253, t.jue}@unsw.edu.au

[3] *Department of Computer Science and Information Engineering*
National Taitung University, Taiwan
Email: hjdai@nttu.edu.tw

Nowadays, social media is often being used by users to create public messages or posts that are related to their health. With the increasing number of social media usage, a trend has been observed of users creating posts related to adverse drug reactions. Mining social media data for these information can be used for pharmacological post-marketing surveillance and monitoring. We developed a binary classifier using linear support vector machines to automatically classify Twitter posts assertive of adverse drug reactions. Two runs were devised to evaluate the classifier. Our classifier achieved an F-score of 0.29 and 0.33 for the first and second run respectively.

1. Introduction

Social media users publicly share various types of health related information. Public messages or posts related to adverse drug reactions (ADR) are often noticed on social media platforms. With the growing number of social media users and usage, public health researchers now have an opportunity to mine social media data for surveillance and monitoring [1-4]. However, it is hard to perform text mining on social media data as messages are short and often include many acronyms, abbreviations, misspellings and special characters like hashtags in Twitter posts. Previous studies reported good classification performance using support vector machines (SVM) [5-7]. In this study, we describe our methods to automatically classify ADR-assertive Twitter posts. We developed a binary classifier using linear support vector machines (SVM). We aimed to increase the performance of SVM classifier by introducing topic model based features. The performance of our classifier increased when topic model distribution weights per Twitter post were added. We developed these methods as part of our participation in PSB 2016 Social Media Mining shared task.

* Corresponding author

2. Methods

We developed a binary classifier which categorizes a given Twitter post as positive or negative for ADR assertion. The shared task organizers provided a training set with over 5,000 Twitter posts. We used this training set to develop our classifier. The classifier was evaluated on a held-out test set of the same size using precision, recall and F-measure metrics. Initially, we pre-processed the Twitter posts to remove the Twitter specific characters like hashtags, usernames, and repetition of certain alphabets. This was followed by stemming using the Snowball stemmer[†] and tokenizing using the Stanford PTBTokenizer[‡]. Twitter posts are very short and in order to preserve the information expressed, we did not remove any stop words.

After the pre-processing, we extracted various features and classified the posts using the SVM classifier with a linear kernel. The feature types included syntactic, lexicon, polarity and topic modelling based features. We extracted unigram, bigram, trigram, POS tags and word-POS pairs syntactic features. The next set of features were lexicon-based features generated using pattern matching. We used an ADR lexicon which had ADRs and drug names listed from various resources such as SIDER, CHV and COSTART [8, 9]. We automatically searched for drug names or ADR mentions from the lexicon in the Twitter posts and marked their presence for these features. In other words, we had two binary features for a Twitter post: i) presence of drug names and ii) ADR mentions.

Previous studies reported that usage of polarity features improved ADR classification performance in Twitter posts [9]. Sarker et al. developed polarity cues categorized into four groups and extracted features in a window of four words [9]. However, in this study we reduced them into just two groups (good and bad), followed by representation of the count of polarity cues in the Twitter post.

Topic model features were derived from the Twitter posts using the Stanford Topic Modelling Toolbox[§]. We applied unsupervised topic modelling technique to extract five topics for each Twitter post [10]. The distribution weights of each topic per Twitter post were added as numeric features.

3. Results

Using the features described above our classifier achieved an F-score of 0.29 and 0.33 for the first and second run, respectively. The first run (traditional features) was done with syntactic, lexicon and

[†] http://snowball.tartarus.org/

[‡] http://nlp.stanford.edu/software/tokenizer.shtml

[§] http://nlp.stanford.edu/software/tmt/tmt-0.4/

polarity features. The second run, on the other hand, was done with syntactic, lexicon, polarity and topic modelling based features. The precision, recall and F-score of both runs are presented in Table 1.

Table 1. Official results of the ADR binary classifier on the test set

Run	ADR Precision	ADR Recall	ADR F-score
Traditional Features	0.35	0.24	0.29
Traditional features and Topic modelling	0.35	0.30	0.33

4. Discussion

With the addition of topic modelling based features, the performance of the classifier has increased. This improvement due to addition of topic distribution weights per instance is consistent with findings from a previous study in automatic text classification of electronic health records [11]. Overall, eight teams participated in this shared task with twenty different runs submitted. The best performing run achieved an F-score of 0.42. The shared task organizers did not release gold standard annotations for the test set making it not possible to analyze our classifier's performance in depth. The topic modelling features run on the training set using a ten-fold cross validation achieved an F-score of 0.42. On the training set, we noticed that the classifier resulted in a large number of false negatives. We believe this is due to a large class imbalance observed in the training set. SVM based classifier tend to be biased towards the majority class in an imbalanced dataset. We could overcome this issue by applying ensemble classifiers [12]. The training and test sets included several ill-formed words. Normalizing the ill-formed words would also improve the classifier's performance.

Acknowledgments

The authors would like to thank the organizers of PSB 2016 Social Media Mining Shared Task and anonymous reviewers for their valuable feedback and comments.

References

1. Wang, G., et al., *A method for systematic discovery of adverse drug events from clinical notes.* Journal of the American Medical Informatics Association, 2015: p. ocv102.
2. Yom-Tov, E. and E. Gabrilovich, *Postmarket drug surveillance without trial costs: discovery of adverse drug reactions through large-scale analysis of web search queries.* Journal of medical Internet research, 2013. **15**(6).
3. Sarker, A., et al., *Utilizing social media data for pharmacovigilance: A review.* Journal of biomedical informatics, 2015. **54**: p. 202-212.
4. Feldman, R., et al. *Utilizing Text Mining on Online Medical Forums to Predict Label Change due to Adverse Drug Reactions*. in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2015. ACM.
5. Karimi, S., et al., *Text and Data Mining Techniques in Adverse Drug Reaction Detection.* ACM Comput. Surv., 2015. **47**(4): p. 1-39.

6. Harpaz, R., et al., *Text mining for adverse drug events: the promise, challenges, and state of the art.* Drug Safety, 2014. **37**(10): p. 777-790.
7. Bian, J., U. Topaloglu, and F. Yu. *Towards large-scale Twitter mining for drug-related adverse events*. in *Proceedings of the 2012 international workshop on Smart health and wellbeing*. 2012. ACM.
8. Nikfarjam, A., et al., *Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features.* Journal of the American Medical Informatics Association, 2015: p. ocu041.
9. Sarker, A. and G. Gonzalez, *Portable automatic text classification for adverse drug reaction detection via multi-corpus training.* Journal of biomedical informatics, 2015. **53**: p. 196-207.
10. Blei, D.M., A.Y. Ng, and M.I. Jordan, *Latent dirichlet allocation.* the Journal of machine Learning research, 2003. **3**: p. 993-1022.
11. Jonnagaddala, J., et al., *A preliminary study on automatic identification of patient smoking status in unstructured electronic health records.* ACL-IJCNLP 2015, 2015: p. 147.
12. Galar, M., et al., A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 2012. 42(4): p. 463-4