COMP90049 Project 2: Identifying Tweets with Adverse Drug Reactions

1. Introduction

Adverse Drug Reactions (ADRs) are “appreciably harmful or unpleasant reactions” after taking a medication (Edward & Aronson, 2000). Traditionally, the detection of ADRs relies on spontaneous reporting and case-control studies (Ferner, 2016). With the increasing popularity of Twitter, a vast number of messages which contain drug usage are generated by users and could be utilised to identify ADRs.

This report aims to apply two supervised machine learning methods, the Naive Bayes and the Support Vector Machine (SVM), to detect ADRs in tweets with newly engineered features. The effectiveness of the machine learning methods will be assessed. Additionally, knowledge gained about the task, identifying tweets with ADRs, will be presented in the report.

2. Related Work

Previous work on using Twitter to detect ADRs includes applying Natural Language Processing (NLP) to preprocess the tweets and extract relevant attributes to build a machine learning-based classification system (Cieliebak, Egger & Uzdilli, 2016). Rajapaksha and Weerasinghe (2015) constructed a medical corpus to filter out unrelated tweets before the feature extraction since the dissemination of data set will affect the performance of the classifier. Ginn et al. (2014) used a binary classification approach to annotate the lexical features of tweets and suggested the potential of semantic analysis in locating ADRs. Based on the intuition that Twitter users post about ADRs with negative sentiments, Korkontzelos et al. (2016) analysed the effect of sentiment features and found that sentiment analysis feature improved the system performances.

3. Data set

The data set used for this report was altered from a subset of twitter corpus annotated for ADR mentions, which was curated by Sarker and Gonzalez (2015). The data set was partitioned into three sets, a training set, a development set and a test set. The training set and development set contained indications of the presence or absence of ADRs for each instance. 92 best terms, identified from the training set via a method of Mutual Information, will be used to represent the raw tweets.

Corpus	Features	
Training set	3166 instances	HasADRs: 373 instances
		NoADRs: 2793 instances
Development set	1076 instances	HasADRs: 114 instances
		NoADRs: 962 instances
Test set	1087 instances	

Table 1. The features of data set

4. Methodology

4.1 Learning and Classification Methods

Two supervised machine learning methods are used to predict labels of the data set via a binary classification: *hasADRs* or *noADRs*. Weka, a machine learning tool, is applied to conduct data pre-processing and classification for this task (Frank, Hall & Witten, 2016).

- ***Naïve Bayes***

The Naïve Bayes method is based on the conditional independence assumption with the decision rule, *maximum a posteriori* (MAP), to classify an instance (Erfani & Verspoor, 2017).

- ***Support Vector Machine (SVM)***

The SVM method constructs a maximum-margin hyperplane which aims to lower the generalisation error of the classification (Erfani & Verspoor, 2017).

4.2 Data Pre-processing

The original data representation includes 92 word-unigrams. Through data pre-processing function in Weka, nine attributes which are out of the task domain and lower the system precision are identified and deleted. They are ***am***, ***banana***, ***can***, ***i***, ***is***, ***it***, ***pic***, ***rt***, ***was***. Additionally, “***id***” attribute will also be deleted in the pre-processing step.

4.3 Identification of New Attributes

Two new attributes are identified, which are “*sentiment*” attribute and “*side-effect*” attribute.

“sentiment” attribute

Inspired by Korkontzelos et al. (2016), a sentiment analysis attribute, which indicates the polarity of tweets, is added as a new feature. Because tweets with ADRs potentially contain negative emotions.

A vanderSentiment package[1] is used to generate a compound sentiment score for each tweet. If the compound score is positive, then the tweet will be classified as positive with an attribute value of “1”. Otherwise, “0” will be assigned to tweets with negative compound scores.

“side-effect” attribute

The “side effect” bigram-word is observed in tweets with ADRs. It often comes along with drug names like effexor, lozenge, olanzapine with uncomfortable feelings. Although side effects do not equal to ADRs, the two terms are often mixed up in real life.

The “side-effect” attribute is applied to represent whether an instance contains this bigram-word or not. For tweets which contain “side effect”, the value of “side-effect” attribute will be 1. Otherwise, the attribute value will be 0.

5. Evaluation

In this section, Weka will be used to assess the effectiveness of Naïve Bayes and SVM. The development set will serve as a separate test set for the system evaluation.

5.1 Evaluation Metrics

- ***Accuracy*** calculates the proportion of tweets which have been correctly predicted the label, “has ADRs” or “no ADRs”.
- ***Precision*** calculates the proportion of positive predictions which are correct.
- ***Recall*** refers to the true positive rate.
- ***F-measure*** is a harmonic mean of precision and recall.

5.2 Results

5.2.1 Data pre-processing

As shown in Table 2, after deleting nine attributes and the tweet id, the performances of both Naïve Bayes and SVM classifiers improved. In addition, the overall performance of SVM is more effective than Naïve Bayes classifier.

Analysis

Dimensionality reduction in the pre-processing step includes *formal subject* like “it”; *verb BE* like “am”, “is” and “was”; *personal pronoun* like “i”; *abbreviations* like “rt” (retweet) and “pic” (picture); and *irrelevant features* like “banana”.

The results indicated that both Naïve Bayes and SVM are sensitive to appropriate feature selection for identifying tweets with ADRs. More fitted data representation improved the system accuracy from 82.15% to 84.38% for Naïve Bayes and from 89.59% to 89.68% for SVM.

> For instance, “`@needtobeskinny0 I was given 7 months worth of fluoxetine, at once. But my parents give it to me, I'm not trusted at all ;-;`” is correctly classified in the Naïve Bayes classifier after deleting the nine attributes and *id* attribute. Since “noADRs” tweets also contain a lot of expressions like *formal subject*, *personal pronoun*, and *verb BE*, these features could potentially lead to sparseness of training data and the decrease of system performance.

For high-dimensional data set like Twitter text, SVM performs better than Naïve Bayes on account of its theoretical maximum-margin hyperplane which avoids overfitting.

In addition, it is observed that the training data set is unbalanced and only 13.35% training data contains ADRs. As a result, the classifiers are prone to bias.

5.2.2 The “sentiment” attribute

The implementation of “sentiment” attribute improved the accuracy of Naïve Bayes system. However, the accuracy of SVM was slightly decreased.

Analysis

The vanderSentiment package combines lexical features with grammatical and syntactical conventions to assess the sentiment intensity of social media content (Hutto & Gilbert 2014).

In this report, the sentiment of tweets are divided into two polarities, the negative and positive. This binary classification is simple and effective for text which expressed intense emotion. It is observed that for Naïve Bayes classifier, the precision for “has ADRs” has improved 1.3%; but for “no ADRs”, the precision only improved 0.2%.

> For instance, “`I have got to stop taking my Vyvanse so late!! #nosleep #addproblems`” expresses a negative sentiment and it is correctly classified after taking sentiment attribute into account in the Naïve Bayes system.

[1] Imported from vanderSentiment Python Library (https://pypi.python.org/pypi/vaderSentimen)

		hasADRs			noADRs		
	Accuracy	Precision	Recall	F-measure	Precision	Recall	F-measure
92 attributes system							
Naïve Bayes	0.821561	0.293	0.482	0.364	0.934	0.862	0.896
SVM	0.8959	0.538	0.123	0.200	0.905	0.988	0.944
83 attributes system (after data pre-processing)							
Naïve Bayes	0.843866	0.325	0.439	0.373	0.931	0.892	0.911
SVM	0.89684	0.560	0.123	0.201	0.905	0.989	0.945
83 attributes system + sentiment attribute							
Naïve Bayes	0.847584	0.338	0.456	0.388	0.933	0.894	0.913
SVM	0.894981	0.517	0.132	0.210	0.905	0.985	0.944
83 attributes system + sentiment attribute + side-effect attribute							
Naïve Bayes	0.842937	0.331	0.474	0.390	0.934	0.887	0.910
SVM	0.895911	0.542	0.114	0.188	0.904	0.989	0.944

Table 2. Classification results

According to the error analysis, "Mirtazapine and olanzapine OD may cause drowsiness and small pupils. No response to naloxone #SAMCoventry" is wrongly classified as a tweet with ADRs in the Naïve Bayes classifier. From the lexical perspective, "drowsiness" and "no" indicate negative sentiments. But the later part of the tweet expressed positive usage feelings to naloxone. In this case, solely depending on lexical attributes, the semantic meaning of this instance is not correctly interpreted.

As for the SVM system, only recall and F-measure of "has ADRs" are improved, which indicated that sentiment analysis is effective for tweets with ADRs.

Moreover, since "neutral" sentiment is not considered in this project, some tweets which appreciably expressed positive or negative sentiments could be imprecisely represented by this dichotomous method.

5.2.3 The "sentiment" attribute + "side-effect" attribute

Comparing with the model with 83 attributes and a sentiment feature, the implementation of "side-effect" attribute slightly decreased the accuracy of Naïve Bayes system and increased the accuracy of SVM system.

For the error analysis, "The Fluoroquinolone Syndrome <-Have you experienced side effects from taking Quinolones antibiotics (e.g. Cipro) ?" is wrongly classified because it contains the bigram-word "side effect". However, the semantic meaning of this instance only expressed interrogation. This case indicates the importance of tweets' context.

6. Discussions and Future Improvements

Firstly, one of the main challenges of this task is that Tweets are written in an informal style and contain a lot of noise. As such, approximate matching should be conducted to optimise the generation of attribute values.

Secondly, feature selection is important for the effectiveness of classification. Due to the word limitation and the diversity of expression in tweets, the dimensionality of attributes should be carefully considered. Otherwise, the system will be prone to the "curse of dimensionality" and overfitting.

Thirdly, the classification results of sentiment feature indicate that intense emotions often associate with ADRs. Further improvement of the sentiment attribute could divide the sentiments into "strongly negative", "slightly negative", "neutral", "slightly positive" and "strongly positive". As such, the data representation will be more precise.

Moreover, due to the limitations of lexical attributes, the semantic analysis of the context of tweets is crucial for the ADRs identification. As for the imbalanced training data, resampling methods should be considered to improve the classification performances.

7. Conclusions

In summary, this report assessed the effectiveness of Naïve Bayes and SVM for the identification of ADRs in tweets with two newly engineered features, the "sentiment" and "side-effect". The SVM outperforms Naïve Bayes for the overall accuracy. The sentiment attribute had a positive effect on the classification of ADRs. Additionally, feature engineering is important for high-dimensional data set. Future improvements include semantic analysis of the context of tweets and optimisation of feature selection.

References

Cieliebak, M, Egger, D, & Uzdilli, F 2016, Twitter can Help to Find Adverse Drug Reactions. *ERCIM NEWS*, (104), 31-32.

Edwards, IR & Aronson, JK 2000, 'Adverse Drug Reactions: Adverse drug reactions: definitions, diagnosis, and management', *The Lancet*, vol. 356, pp. 1255-1259.

Erfani, S & Verspoor, K 2017, COMP90049 Knowledge Technologies: Classification [Lecture slides], The University of Melbourne, 9 Aug 2017.

Ferner, RE 2016, Adverse drug reactions. *Medicine*, 44(7), 416-421.

Frank, E, Hall, MA, Witten, IH 2016, The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.

Ginn, R, Pimpalkhute, P, Nikfarjam, A, Patki, A, O'Connor, K, Sarker, A, Smith, K & Gonzalez, G 2014, Mining Twitter for adverse drug reaction mentions: a corpus and classification benchmark. In *Proceedings of the fourth workshop on building and evaluating resources for health and biomedical text processing*.

Hutto, CJ & Gilbert, E 2014, Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth international AAAI conference on weblogs and social media*.

Hutto, CJ 2016, vaderSentiment 2.5, viewed 9th Octorber 2017, <https://pypi.python.org/pypi/vaderSentiment>.

Korkontzelos, I, Nikfarjam, A, Shardlow, M, Sarker, A, Ananiadou, S & Gonzalez, GH 2016, 'Analysis of the effect of sentiment analysis on extracting adverse drug reactions from tweets and forum posts', *Journal of Biomedical Informatics*, vol. 62, pp. 148-158.

Rajapaksha, P & Weerasinghe, R 2015, "Identifying Adverse Drug Reactions by analyzing Twitter messages," *2015 Fifteenth International Conference on Advances in ICT for Emerging Regions (ICTer)*, Colombo, 2015, pp. 37-42.

Sarker, A & Gonzalez, G 2015, *Portable automatic text classification for adverse drug reaction detection via multi-corpus training*. Journal of Biomedical Informatics, 53: 196-207.