DETECTING SIGNALS IN NOISY DATA - CAN ENSEMBLE CLASSIFIERS HELP IDENTIFY ADVERSE DRUG REACTION IN TWEETS?

MAJID RASTEGAR-MOJARAD

Department of Health Science Research, Mayo Clinic, USA
University of Wisconsin-Milwaukee, Milwaukee, WI, USA
Email: Mojarad.majid@mayo.edu

RAVIKUMAR KOMANDUR ELAYAVILLI

Department of Health Science Research, Mayo Clinic, USA
Email: komandurelayavilli.ravikumar@mayo.edu

YUE YU

Department of Health Science Research, Mayo Clinic, USA
Email: Yu.Yue@mayo.edu

HONGFANG LIU

Department of Health Science Research, Mayo Clinic, USA
Email: Liu.Hongfang@mayo.edu

In this paper, we describe our system for detecting adverse reactions from tweets, a task organized as part of Pacific Symposium of Biocomputing - Social Media Mining Shared Task. The shared task primarily involves a binary classification of the tweets, whether they contain description of adverse drug reaction or not. We propose an ensemble machine learning classifier to tackle the unbalanced distribution of the classes in the data provided for the task. A feature set containing unigrams, bigrams, and trigrams (a selected list, using mutual information), co-occurrence of drug and side effect, negation, sentiment score, etc. were used to train the classifiers in identifying the tweets with the description of adverse drug reactions. Our system obtained 0.4195 F-score and ranked first in the shared task.

1. Introduction

Social media provides a platform for patients to share their experience with illnesses and medications. Several studies showed that patients tend to share information regarding adverse drug events (ADEs) of medications in social media [1],[2]. These posts have become a valuable resource to identify potential adverse drug event after the launch of a drug in the market.

Task 1 of Pacific Symposium of Biocomputing (PSB) - Social Media Mining Shared Task (SMMST) primarily involves detection of tweets containing adverse drug reaction (ADR). For this purpose, the organizers provided tweets (with and without describing adverse drug reactions) 7754, 2349, and 4895 for training, development and test dataset respectively. Two main challenges for this task were 1) informal sentences (misspelling, incorrect grammar, incomplete sentences) 2) unbalanced distribution of the classes (11% of tweets in the training dataset contained ADR). To address the first issue, we used fuzzy matching instead of exact match search to detect drug names and side effects in tweets. To solve the second issue, we employed an ensemble classifier, comprising 9 different classifiers training each of them on selected equal numbers of positive and negative instances.

2. Related Works

In the current era of information technology, patients often post their experiences with conditions and medications to social media websites, which provides a valuable resource to study and analyze their perspectives [3]. One use of this resource is mining signals of ADRs. There have been several efforts to identify ADR in users' posts. Leman et al. [4] studied a medical forum to identify reported adverse drug events. They generated a collection of patients' posts and annotated the posts manually. Using natural language processing methods and machine learning techniques, they implemented a classifier to identify adverse drug reactions in the posts. Chee et al. [5] conducted a study on patients' posts on Health and Wellness Yahoo! Groups to identify drugs for further FDA scrutiny. They applied common natural language processing methods to predict adverse drug events. Freifeld et al. [6] evaluated the correlation between adverse drug events reported in Twitter (only 140 characters) and spontaneous reports received by a regulatory agency. Sharif et al. [7] proposed a sentiment classification framework to detect adverse drug reactions in medical blogs and forums. Karimi et al [8] provided a corpus of 1321 medical forum posts on patient reported adverse drug events. In the most recent efforts, Sarker and Gonzalez [9] developed a classifier for detecting adverse drug reactions in users posts using multiple corpus.

3. Method

For this task, we trained a binary classifier to categorize the tweets into two classes, positive (containing ADR) and negative (without ADR). As distribution of the classes in the training data is unbalanced, we proposed using an ensemble approach, which includes several classifiers trained with balanced datasets (equal distribution of positive and negative instances). Considering the number of tweets in the training dataset, our system is a set of 9 classifiers. All the classifiers are trained with all the positive instances in the training data and equal number of randomly selected negative instances. This approach is more biased towards positive instances, leading to higher recall and lower precision. As a supplementary measure, we built a list of post-processing rules in order to improve the precision of the system. We used Random Forest (RF), which is a combination of decision tress, as the learning model in all the classifiers. To aggregate the results of all the classifiers, we used common voting approach with different number of threshold.

Another challenge in this task was the informal expression of adverse drug reactions by the tweeters in free text. As tweets are informal, they often contain misspellings, grammatical errors, and incomplete sentences. The grammatical errors make it difficult to use syntactic approaches for finding relation between drugs and side effects. This also reflects in our choice of using Random Forest as the learning model, which is known to handle noisy data[10]. To further smoothen the misspelling of drug names and side effects, we implemented a fuzzy matching to identify any mention of drugs and side effects.

A feature set contains 6470 features is used in all the classifiers. Here is the list:

- **Words**: Unigrams, bi-grams, and tri-grams serve as features for the classifiers. Mutual information, as a feature selection approach, is implemented to select the most meaningful unigrams, bi-grams, and tri-grams.
- **Number of drugs and side effects:** The number of drugs and side effects in tweet is another feature. To recognize drugs and side effects in the tweets, a dictionary of drug names (from DrugBank[11]) and a dictionary of side effects (from public resources) are created. We implemented two methods a) exact b) fuzzy matching (using edit distance) to tag drugs and side effects.
- **Drug-Side effect pairs:** We generated a list of known drug-side effect pairs, and co-occurrence of the pairs in the tweet and used them as feature in the classifiers.
- **Side effects associated to drug class:** A list of side effects related to each drug class is generated and co-occurrence of a drug and a side effect associated to that drug's class is used as a feature.
- **Sentiment:** The sentiment score of the tweet is computed by the sentiment analysis module provided in LingPipe [12].
- **Negation:** We used an in-built negation detection tool [13] to detect whether the tweet is negated or not.
- **Question in tweet:** Simple rules are implemented to identify the questions in the tweet.
- **Change in tweet's tone [9]:** Certain features such as sub-ordinate clauses often indicate a change in the tone of the tweet. Identification of such features is helpful in ascertaining the adverse drug reaction mentioned in the tweet. In the following tweet one can observe the change of tone in the tweet:
  *"I took trazodone last night and it really helped- **but** it was difficult to wake up"*

Results:

Our system is implemented in Java and we used the Random Forest implementation in Weka [14] in our study. Table 1 shows the result of our system for different threshold.

Table 1: The performance of our system on the test data

Threshold	Precision	Recall	F-Score	Accuracy
3	0.3606	0.5013	0.4195	0.8931
4	0.3918	0.4084	0.4	0.9056
5	0.4309	0.3474	0.3847	0.9144
6	0.4590	0.2679	0.3383	0.9193

As the results show, increasing the threshold leads to higher precision and lower recall. Finding the best threshold is one of the challenges in our approach.

References

[1] R. Harpaz, W. DuMouchel, N. H. Shah, D. Madigan, P. Ryan, and C. Friedman, "Novel Data Mining Methodologies for Adverse Drug Event Discovery and Analysis," *Clin Pharmacol Ther*, vol. 91, no. 6, pp. 1010–1021, Jun. 2012.

[2] C. C. Yang, H. Yang, L. Jiang, and M. Zhang, "Social Media Mining for Drug Safety Signal Detection," in *Proceedings of the 2012 International Workshop on Smart Health and Wellbeing*, New York, NY, USA, 2012, pp. 33–40.

[3] M. Rastegar-Mojarad, Z. Ye, D. Wall, N. Murali, and S. Lin, "Collecting and Analyzing Patient Experiences of Health Care From Social Media," *JMIR Research Protocols*, vol. 4, no. 3, p. e78, 2015.

[4] R. Leaman, L. Wojtulewicz, R. Sullivan, A. Skariah, J. Yang, and G. Gonzalez, "Towards Internet-age Pharmacovigilance: Extracting Adverse Drug Reactions from User Posts to Health-related Social Networks," in *Proceedings of the 2010 Workshop on Biomedical Natural Language Processing*, Stroudsburg, PA, USA, 2010, pp. 117–125.

[5] B. W. Chee, R. Berlin, and B. Schatz, "Predicting Adverse Drug Events from Personal Health Messages," *AMIA Annu Symp Proc*, vol. 2011, pp. 217–226, 2011.

[6] C. C. Freifeld, J. S. Brownstein, C. M. Menone, W. Bao, R. Filice, T. Kass-Hout, and N. Dasgupta, "Digital drug safety surveillance: monitoring pharmaceutical products in twitter," *Drug Saf*, vol. 37, no. 5, pp. 343–350, May 2014.

[7] H. Sharif, F. Zaffar, A. Abbasi, and D. Zimbra, "Detecting Adverse Drug Reactions Using a Sentiment Classification Framework," Academy of Science and Engineering (ASE),USA, ©ASE 2014, Working Paper, Jun. 2014.

[8] S. Karimi, A. Metke-Jimenez, M. Kemps, and C. Wang, "Cadec: A corpus of adverse drug event annotations," *J Biomed Inform*, Mar. 2015.

[9] A. Sarker and G. Gonzalez, "Portable automatic text classification for adverse drug reaction detection via multi-corpus training," *Journal of Biomedical Informatics*, vol. 53, pp. 196–207, Feb. 2015.

[10] "Improving Random Forests." [Online]. Available: https://www.researchgate.net/publication/221112260_Improving_Random_Forests. [Accessed: 20-Aug-2015].

[11] C. Knox, V. Law, T. Jewison, P. Liu, S. Ly, A. Frolkis, A. Pon, K. Banco, C. Mak, V. Neveu, Y. Djoumbou, R. Eisner, A. C. Guo, and D. S. Wishart, "DrugBank 3.0: a comprehensive resource for 'omics' research on drugs," *Nucleic Acids Res.*, vol. 39, no. Database issue, pp. D1035–1041, Jan. 2011.

[12] Alias-i, *LingPipe 4.1.0*. 2008.

[13] M. Rastegar-Mojarad, "Extraction and Classification of Drug-Drug Interaction from Biomedical Text Using a Two-Stage Classifier," *Theses and Dissertations*, Dec. 2013.

[14] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *SIGKDD Explor. Newsl.*, vol. 11, no. 1, pp. 10–18, Nov. 2009.