Entity-level Classification of Adverse Drug Reactions: a Comparison of Neural Network Models

Ilseyar Alimova
Kazan Federal University,
Kazan, Russia
ISAlimova@kpfu.ru

Elena Tutubalina
Kazan Federal University,
Kazan, Russia
Samsung-PDMI Joint AI Center,
PDMI RAS, St. Petersburg, Russia
elvtutubalina@kpfu.ru

Abstract

This paper presents our experimental work on exploring the potential of neural network models developed for aspect-based sentiment analysis for entity-level adverse drug reaction (ADR) classification. Our goal is to explore how to represent local context around ADR mentions and learn an entity representation, interacting with its context. We conducted extensive experiments on various sources of text-based information, including social media, electronic health records, and abstracts of scientific articles from PubMed. The results show that Interactive Attention Neural Network (IAN) outperformed other models on four corpora in terms of macro F-measure. This work is an abridged version of our recent paper accepted to Programming and Computer Software journal in 2019.

1 Introduction

Detection of adverse drug reactions (ADRs) in the post-marketing period is becoming increasingly popular, as evidenced by the growth of ADR monitoring systems (Singh et al., 2017; Shareef et al., 2017; Hou et al., 2016). Information about adverse drug reactions can be found in the texts of social media, health-related forums, and electronic health records. This amount of information cannot be processed manually, therefore, methods based on natural language processing are actively developed (Harpaz et al., 2014; Sarker et al., 2015). There are two steps of ADR detection: named entity recognition and entity classification. Thus, at the first step, all information related to a state of health is extracted using named entity recognition systems. Then all obtained entities are classified in order to distinguish ADRs from indication and patient history. In this article, we focused on the task of entity-level ADR classification. This task is quite similar to aspect-based sentiment classification, which aims to determine the sentimental class of a specific aspect conveyed in user opinions. Inspired by recent successful methods in the field of sentiment analysis, we investigate state-of-the-art deep neural network models developed for aspect-based sentiment analysis for entity-level ADR classification task.

2 Corpora

CADEC (CSIRO Adverse Drug Event Corpus) consists of annotated user reviews written about Diclofenac or Lipitor on askapatient.com (Karimi et al., 2015). There are five types of annotations: 'Drug', 'Adverse effect', 'Disease', 'Symptom', and 'Finding'. We grouped diseases, symptoms, and findings as a single class called 'non-ADR'.

MADE corpus consists of de-identified electronic health record notes from 21 cancer patients (Liu et al., 2018). Each record annotated with medications and relations to their corresponding attributes, indications and adverse events. We grouped annotations corresponding to the diseases in class 'non-ADR', such as 'Indication' and 'SSLIF'.

TwiMed corpus consists of sentences extracted from PubMed and tweets. This corpus contains annotations of diseases, symptoms, and drugs, and their relations. If the relationship between disease and drug was labeled as 'Outcome-negative', we marked disease as ADR, otherwise, we annotate it as 'non-ADR' (Alvaro et al., 2017).

Twitter corpus include tweets about drugs. There are three annotations: 'ADR', 'Indication' and 'Other'. We consider 'Indication' and 'Other' as 'non-ADR' (Nikfarjam et al., 2015).

3 Models

We started with the basic LSTM-based model and then extend the model by attention mechanism and explicit memory. The discussed models are as follows:

LSTM: The basic neural network which models the semantic representation of a sentence without considering an entity.

TD-LSTM: The target-dependent model proposed in (Tang et al., 2015) which utilizes two LSTMs to model the left context with an entity and right context with an entity.

IAN: The interactive attention network proposed in (Ma et al., 2017) which consists of (i) LSTMs for handling an entity and its context and (ii) cross attention block where the final concatenation of attentions' outputs is fed to softmax for prediction.

MemNet: The deep memory network proposed in (Tang et al., 2016). It applies attention multiple times on the word embeddings (multi-hop property), and the last attentions output is fed to softmax for prediction.

RAM: recurrent attention on memory proposed in (Chen et al., 2017) is the more complex model than MemNet. RAM utilizes LSTMs for handling an entity and its context and multi-hop attention on memory.

4 Results

The models were evaluated on 5-fold cross-validation using standard classification quality metrics: precision (P), recall (R), and macro F-measure. The results are presented in Table 1. The results show that IAN outperformed other models on all corpora except Twitter. IAN obtained the most significant increase in results compared to other models on Twimed-Twitter and Twimed-Pubmed corpora with 81.9% and 87.4% of the macro F-measures, respectively. On the Twitter corpus, the RAM model achieved the best results with a macro F-measure of 83.4%.

TD-LSTM with the macro F-measures 75.8% and 73% on the Twitter and Twimed-Twitter corpora respectively outperformed the LSTM model with 61.3% and 70% of the macro F-measures. This result leads to the conclusion that the division of the input sentence into the right and left context of entity can improve the quality of classification for corpora consisting of tweets. Comparison of the results of the RAM and MemNet models shows that the presence of an LSTM layer before the memory layer turned out to be effective only on one Twitter corpus, where the RAM showed significantly high F-measure results (83.4%) compared to MemNet (76.3%). Due to IAN model outperformed RAM and MemNet models on four corpora we can conclude that the presence of additional memory does not give an advantage.

Corpora	**Twitter**	**Cadec**	**MADE**	**Twimed-Twitter**	**Twimed-PubMed**
LSTM	.613	.784	.771	.700	.839
TD-LSTM	.758	.772	.750	.730	.709
IAN	.794	**.815**	**.786**	**.819**	**.874**
RAM	**.834**	.734	.761	.780	.789
MemNet	.763	.758	.760	.795	.811

Table 1: Macro F-measure classification results of the compared methods for each datasets.

5 Conclusion

We explored the potential of state-of-the-art neural network models for entity-level ADR classification task. Based on our experimental results, we can note that the application of attention mechanism improves the performance of a model. IAN model performed the best results for entity-level ADR classification task in most of our experiments. RAM model obtained the best result on Twitter corpora.

References

Nestor Alvaro, Yusuke Miyao, and Nigel Collier. 2017. Twimed: Twitter and pubmed comparable corpus of drugs, diseases, symptoms, and their relations. *JMIR public health and surveillance*, 3(2).

Peng Chen, Zhongqian Sun, Lidong Bing, and Wei Yang. 2017. Recurrent attention network on memory for aspect sentiment analysis. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 452–461.

Rave Harpaz, Alison Callahan, Suzanne Tamang, Yen Low, David Odgers, Sam Finlayson, Kenneth Jung, Paea LePendu, and Nigam H Shah. 2014. Text mining for adverse drug events: the promise, challenges, and state of the art. *Drug safety*, 37(10):777–790.

Yongfang Hou, Xinling Li, Guizhi Wu, and Xiaofei Ye. 2016. National adr monitoring system in china. *Drug Safety*, 39(11):1043–1051, Nov.

Sarvnaz Karimi, Alejandro Metke-Jimenez, Madonna Kemp, and Chen Wang. 2015. Cadec: A corpus of adverse drug event annotations. *Journal of biomedical informatics*, 55:73–81.

Feifan Liu, Hong Yu, Abhyuday Jagannatha, and Weisong Liu. 2018. Nlp challenges for detecting medication and adverse drug events from electronic health records (made1.0).

Dehong Ma, Sujian Li, Xiaodong Zhang, and Houfeng Wang. 2017. Interactive attention networks for aspect-level sentiment classification. *arXiv preprint arXiv:1709.00893*.

Azadeh Nikfarjam, Abeed Sarker, Karen OConnor, Rachel Ginn, and Graciela Gonzalez. 2015. Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. *Journal of the American Medical Informatics Association*, 22(3):671–681.

Abeed Sarker, Rachel Ginn, Azadeh Nikfarjam, Karen OConnor, Karen Smith, Swetha Jayaraman, Tejaswi Upadhaya, and Graciela Gonzalez. 2015. Utilizing social media data for pharmacovigilance: a review. *Journal of biomedical informatics*, 54:202–212.

SM Shareef, CDM Naidu, Shrinivas R Raikar, Y Venkata Rao, and U Devika. 2017. Development, implementation, and analysis of adverse drug reaction monitoring system in a rural tertiary care teaching hospital in narketpally, telangana. *International Journal of Basic & Clinical Pharmacology*, 4(4):757–760.

Preeti Singh, Manju Agrawal, Rajesh Hishikar, Usha Joshi, Basant Maheshwari, and Ajay Halwai. 2017. Adverse drug reactions at adverse drug reaction monitoring center in raipur: Analysis of spontaneous reports during 1 year. *Indian journal of pharmacology*, 49(6):432.

Duyu Tang, Bing Qin, Xiaocheng Feng, and Ting Liu. 2015. Effective lstms for target-dependent sentiment classification. *arXiv preprint arXiv:1512.01100*.

Duyu Tang, Bing Qin, and Ting Liu. 2016. Aspect level sentiment classification with deep memory network. *arXiv preprint arXiv:1605.08900*.