Artificial Intelligence in Medicine xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

journal homepage: www.elsevier.com/locate/aiim

Position-aware deep multi-task learning for drug–drug interaction extraction

Deyu Zhou[a,*], Lei Miao[a], Yulan He[b]

[a] School of Computer Science and Engineering, Key Laboratory of Computer Network and Information Integration, Ministry of Education, Southeast University, Nanjing 210096, China
[b] School of Engineering and Applied Science, Aston University, UK

ARTICLE INFO

Article history:
Received 19 November 2017
Received in revised form 26 February 2018
Accepted 11 March 2018

Keywords:
Classification
Multi-task learning
Drug–drug interaction extraction
Long short-term memory network

ABSTRACT

Objective: A drug–drug interaction (DDI) is a situation in which a drug affects the activity of another drug synergistically or antagonistically when being administered together. The information of DDIs is crucial for healthcare professionals to prevent adverse drug events. Although some known DDIs can be found in purposely-built databases such as DrugBank, most information is still buried in scientific publications. Therefore, automatically extracting DDIs from biomedical texts is sorely needed.
Methods and material: In this paper, we propose a novel position-aware deep multi-task learning approach for extracting DDIs from biomedical texts. In particular, sentences are represented as a sequence of word embeddings and position embeddings. An attention-based bidirectional long short-term memory (BiLSTM) network is used to encode each sentence. The relative position information of words with the target drugs in text is combined with the hidden states of BiLSTM to generate the position-aware attention weights. Moreover, the tasks of predicting whether or not two drugs interact with each other and further distinguishing the types of interactions are learned jointly in multi-task learning framework.
Results: The proposed approach has been evaluated on the DDIExtraction challenge 2013 corpus and the results show that with the position-aware attention only, our proposed approach outperforms the state-of-the-art method by 0.99% for binary DDI classification, and with both position-aware attention and multi-task learning, our approach achieves a micro *F*-score of 72.99% on interaction type identification, outperforming the state-of-the-art approach by 1.51%, which demonstrates the effectiveness of the proposed approach.

1. Introduction

A drug–drug interaction (DDI) is a situation in which a drug affects the activity of another drug synergistically or antagonistically when being administered together. Concomitant medications might alter drug transportation abruptly in individuals who have previously taken a particular dose of a drug. Such an abrupt alteration might change the known safety and efficacy of a drug. For example, *terfenadine* was a common antihistamine intended to block the effects of an allergic rhinitis. Unfortunately several people who took *terfenadine* concomitantly with *ketoconazole*, an antifungal, suffered cardiac problems which often lead to death [1]. Therefore, it is crucial to extract the information about DDIs. Although some known DDIs can be found in drug-related databases such as DrugBank,[1] most information is still buried in scientific articles. Automatic DDI extraction, aiming to automatically discover DDIs from text with high efficiency and accuracy, is becoming an increasingly well understood alternative to manual DDI discovery. Without automated DDI extraction tools, it is hard for doctors, pharmacists and researchers to keep up with the most recent discoveries described in biomedical literature.

To tackle the DDI extraction problem, several evaluation tasks, such as DDIExtraction 2011 [2] and DDIExtraction 2013 [3] shared tasks, have been proposed in recent years to provide common benchmarking datasets for the evaluation of DDI detection from biomedical text. An example of a sentence and its corresponding DDI annotations selected from DDIExtraction 2013 is presented in Table 1. The sentence contains three drug entities: "*neomycin*

---

* Corresponding author.
E-mail addresses: d.zhou@seu.edu.cn (D. Zhou), miaolei@seu.edu.cn (L. Miao), y.he@cantab.net (Y. He).

[1] http://www.drugbank.ca/.

https://doi.org/10.1016/j.artmed.2018.03.001
0933-3657/© 2018 Elsevier B.V. All rights reserved.

Please cite this article in press as: Zhou D, et al. Position-aware deep multi-task learning for drug–drug interaction extraction. Artif Intell Med (2018), https://doi.org/10.1016/j.artmed.2018.03.001

Table 1
An example of one sentence and its corresponding DDI annotations selected from DDIExtraction 2013.

Sentence	Oral **neomycin sulfate** may enhance the effect of **coumarin** in **anticoagulants** by decreasing vitamin K availability.
Annotation	Drug1: neomycin sulfate, Drug2: coumarin, Type: effect Drug1: neomycin sulfate, Drug2: anticoagulants, Type: effect Drug1: coumarin, Drug2: anticoagulants, Type: other

sulfate", "*coumarin*" and "*anticoagulants*". The DDI with the interaction type "*effect*" exists between *neomycin sulfate* and *coumarin*, and also, *neomycin sulfate* and *anticoagulants*. However, for the drug pair *coumarin* and *anticoagulants*, there is no interaction between them. As such, its interaction type is annotated as "*other*". There are two DDI extraction tasks defined, first, given a drug pair, determine whether there exists a DDI between them; second, determine the type of DDI. The former task is essentially a binary classification problem while the latter is a multi-class classification problem since there are more than two interaction types.

Early approaches to automatic extraction of DDIs are mostly based on hand-crafted rules due to the lack of annotated datasets [4]. With the introduction of DDIExtraction challenges in 2011 [2] and 2013 [3] and the availability of annotated DDI datasets, more and more machine learning based methods have been proposed [5]. These approaches typically rely on a set of carefully designed features to train supervised classifiers such as support vector machine (SVM). The results of DDI extraction largely depend on the quality of the features used, as evidenced by the submitted systems to the DDIExtraction 2013 challenge. To avoid the tedious process for feature design, in recent years, deep learning techniques have been proposed to automatically learn feature representations from abundant unannotated data [6]. Different neural network based methods, such as convolutional neural networks (CNNs) [7] and recurrent neural networks (RNNs) [8], have been proposed to automatically extract feature vectors from sentences for DDI extraction and get good performance.

In this paper, we propose a novel position-aware deep multi-task learning approach which is built upon bidirectional long short-term memory (LSTM) networks, called PM-BLSTM, for extracting DDIs from biomedical texts. Our contributions are summarized below: (1) We incorporate the position-aware attention mechanism by combining position embeddings with the hidden states of BiLSTM to generate the position-aware attention weights. Drug pairs occurring in different position give different indications of DDIs. Therefore, the position where a drug occurs in a sentence could capture syntactic information, which could be important for DDI extraction. (2) We propose a multi-task learning framework to tackle jointly the tasks of predicting whether or not two drugs interact with each other and further distinguishing the types of interactions. The binary (classify whether DDI exists within the given drugs or not) and multi-class (classify the interaction of a drug pair into one of the DDI types) DDI classifications are related tasks and jointly learning both tasks allows the capture of shared features effectively which could benefit both tasks; (3) The proposed approach has been evaluated on the DDIExtraction challenge 2013 corpus and the results show that with the position-aware attention only, our proposed approach outperforms the state-of-the-art method by 0.99% for binary DDI classification, and with both position-aware attention and multi-task learning, our approach achieves a micro *F*-score of 72.99% on interaction type identification, outperforming the state-of-the-art approach by 1.51%, which demonstrates the effectiveness of the proposed approach.

The rest of the paper is organized as follows. Section 2 surveys existing approaches for DDI extraction. Section 3 describes the proposed method, which consists of four components: the embedding layer, the BiLSTM layer, the position-aware attention layer and the multi-task output layer. Section 4 presents experimental setup and results. Finally, Section 5 concludes the paper and outlines future research directions.

2. Related work

Most existing approaches to DDI extraction are based on machine learning. In order to predict the relation between a given pair of drugs, classifiers are typically trained on lexical, syntactic and semantic features extracted from manually labeled corpora. Based on the way of feature construction, approaches can be roughly divided into two categories, feature-based and kernel-based methods [5]. Feature-based approaches focus on finding potentially discriminative features to represent data characteristics. Apart from the basic bag-of-words features, researchers have explored the use of various types of features including context features [9], a combination of lexical, semantic and domain features [10], and heterogeneous features consisting of lexical, syntactic, semantic and negation features derived from parse trees [11]. Kernel-based approaches employ different kernels to calculate the similarity between two instances by exploiting the structural representations of data instances such as syntactic parse trees or dependency graphs [12]. In the past DDIextraction challenges, the most commonly used kernels are all-paths graph kernel [13], shallow linguistic kernel [14] and path-enclose tree kernel [15]. It is also possible to combine multiple kernels in order to compensate the weakness of each individual kernel. For example, in the work of [16], three different kernels were combined to form a hybrid kernel which gives a better performance compared to those using a single kernel. However, these approaches typically rely on feature engineering to generate a list of discriminative features for training supervised classifiers. As observed in DDIExtraction 2013 challenge, different approaches adopted different feature engineering techniques and there is no standard way in generating features. Moreover, features often need to be redesigned when previously developed systems are adapted for the task of DDI extraction. For example, UTurku, a system originally developed for biomedical event extraction was adapted for DDI extraction by redesigning the features specifically for DDI extraction [17].

In recent years, deep learning techniques have been proposed to automatically learn feature representations from abundant unannotated data [6]. Features for DDI extractor can be learned automatically using deep neural networks without expensive manual feature engineering. Based on the structure of neural network, these methods can be roughly classified into two categories: CNN based models and RNN based models. Liu et al. [7] attempted to use CNN for DDIs extraction. They adopted a shallow CNN and combined word embeddings with position embeddings in the CNN model. Liu et al. [18] introduced the structure information of sentences with dependency-based CNN for DDI extraction. Quan et al. [19] incorporated semantic information of multiple word embeddings with multichannel CNNs. Zhao et al. [20] combined CNN with some traditional features. Apart from CNNs, RNNs have also been used for biomedical relation classification. Sahu and Anand [8] used LSTMs for DDI extraction and Yi et al. [21] extracted DDI at the corpus level via gated recurrent unit networks (GRUs). In both approaches, the attention mechanism was incorporated.

3. Method

In this section, we present the proposed position-aware multi-task deep learning method based on bidirectional LSTMs (PM-BLSTM), which is illustrated in Fig. 1, for DDI extraction. It can be observed from Fig. 1 that in embedding layer and BiLSTM layer, the representations for each word and its position in a sentence are generated. Then, in position-aware attention layer, these

Please cite this article in press as: Zhou D, et al. Position-aware deep multi-task learning for drug–drug interaction extraction. Artif Intell Med (2018), https://doi.org/10.1016/j.artmed.2018.03.001