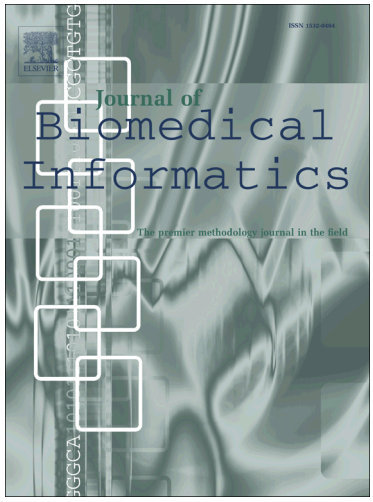

A Graph Kernel Based on Context Vectors for Extracting Drug-Drug Interactions

Wei Zheng, Hongfei Lin, Zhehuan Zhao, Bo Xu, Yijia Zhang, Zhihao Yang, Jian Wang

PII: S1532-0464(16)30005-3
DOI: http://dx.doi.org/10.1016/j.jbi.2016.03.014
Reference: YJBIN 2544

To appear in: *Journal of Biomedical Informatics*

Received Date: 2 December 2015
Revised Date: 14 March 2016
Accepted Date: 18 March 2016

Please cite this article as: Zheng, W., Lin, H., Zhao, Z., Xu, B., Zhang, Y., Yang, Z., Wang, J., A Graph Kernel Based on Context Vectors for Extracting Drug-Drug Interactions, *Journal of Biomedical Informatics* (2016), doi: http://dx.doi.org/10.1016/j.jbi.2016.03.014

A Graph Kernel Based on Context Vectors for Extracting Drug-Drug Interactions

Wei Zheng[1,2], Hongfei Lin[1], Zhehuan Zhao[1], Bo Xu[1], Yijia Zhang[1], Zhihao Yang[1], Jian Wang[1]

[1.] School of Computer Science and Technology, Dalian University of Technology, Dalian, China
[2.] College of software, Dalian JiaoTong University, Dalian, China
weizheng@mail.dlut.edu.cn
hflin@ dlut.edu.cn

Abstract—The clinical recognition of drug-drug interactions (DDIs) is a crucial issue for both patient safety and health care cost control. Thus there is an urgent need that DDIs be extracted automatically from biomedical literature by text-mining techniques. Although the top-ranking DDIs systems explore various features of texts, these features can't yet adequately express long and complicated sentences. In this paper, we present an effective graph kernel which makes full use of different types of contexts to identify DDIs from biomedical literature. In our approach, the relations among long-range words, in addition to close-range words, are obtained by the graph representation of a parsed sentence. Context vectors of a vertex, an iterative vectorial representation of all labeled nodes adjacent and nonadjacent to it, adequately capture the direct and indirect substructures' information. Furthermore, the graph kernel considering the distance between context vectors is used to detect DDIs. Experimental results on the DDIExtraction 2013 corpus show that our system achieves the best detection and classification performance (F-score) of DDIs (81.8 and 68.4, respectively). Especially for the Medline-2013 dataset, our system outperforms the top-ranking DDIs systems by F-scores of 10.7 and 12.2 in detection and classification, respectively.

Keywords—context vector; graph kernel; equivalent class; Drug-Drug Interactions

1. Introduction

Drug-drug interactions (DDIs) occur when one administered drug has an influence on the level or activity of another drug. The DDI problem is one of the most important causes of medical errors. Knowing all potential interactions is very important for physicians prescribing varying combinations of drugs for their patients.

Because it is becoming more common that multiple drugs are used simultaneously, DDIs are causing a great threat to public health [1]. The interactions among drugs are affected by many factors, such as the dose dependence of many DDIs, natural genetic and demographic variation. Therefore, the recognition of DDIs in the clinic is difficult even if a drug has been tested by stages. Today, most DDIs are discovered by accident in the clinic or during phases Ⅳ clinic trials that take place once a drug is already on the market [2]. Only a few drug combinations can be analyzed at a time by laboratory studies, which is not enough to recognize, understand and predict DDIs on a large scale. However, huge amounts of data relevant to DDIs are preserved in specialized databases and scientific documents. On the one hand, existing databases with structured data are generally incomplete because of their long update periods. On the other hand, the amount of biomedical literature with unstructured data, such as MEDLINE, is increasing exponentially. As a result, there is an urgent need that DDIs should be extracted automatically and effectively from biomedical literature. This contributes significantly not only to detecting known DDIs, but also to automating the database curation process, which is currently performed manually [3].

To promote the development of text-mining techniques and improve the performance for detecting DDIs from biomedical texts, Segura-Bedmar et al. [4, 5] developed valuable gold standard sets in the DDIExtraction challenges 2011 (DDI-2011) and 2013 (DDI-2013), which provided an opportunity to address the problem of DDI extraction for the community. While DDI-2011 corpus only focused on the identification of all possible pairs of interacting drugs, DDI-2013 corpus [6] also proposed, in addition to detection, a more fine-grained classification of each true DDI. In both challenges, the detection of DDIs was seen as a binary classification problem, where each candidate drug pair is