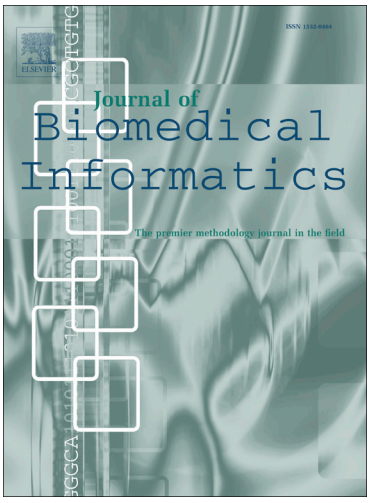

An effective neural model extracting document level chemical-induced disease relations from biomedical literature

Wei Zheng, Hongfei Lin, Zhiheng Li, Xiaoxia Liu, Zhengguang Li, Bo Xu, Yijia Zhang, Zhihao Yang, Jian Wang

PII: S1532-0464(18)30084-4
DOI: https://doi.org/10.1016/j.jbi.2018.05.001
Reference: YJBIN 2971

To appear in: *Journal of Biomedical Informatics*

Received Date: 9 October 2017
Revised Date: 14 March 2018
Accepted Date: 4 May 2018

Please cite this article as: Zheng, W., Lin, H., Li, Z., Liu, X., Li, Z., Xu, B., Zhang, Y., Yang, Z., Wang, J., An effective neural model extracting document level chemical-induced disease relations from biomedical literature, *Journal of Biomedical Informatics* (2018), doi: https://doi.org/10.1016/j.jbi.2018.05.001

An effective neural model extracting document level chemical-induced disease relations from biomedical literature

Wei Zheng[1,2], Hongfei Lin[1], Zhiheng Li[1], Xiaoxia Liu[1], Zhengguang Li[1,2], Bo Xu[1],Yijia Zhang[1], Zhihao Yang[1], Jian Wang[1]

[1] College 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

Since identifying relations between chemicals and diseases (CDR) are important for biomedical research and healthcare, the challenge proposed by BioCreative V requires automatically mining causal relationships between chemicals and diseases which may span sentence boundaries. Although most systems explore feature engineering and knowledge bases to recognize document level CDR relations, feature learning automatically is limited only in a sentence.

In this work, we proposed an effective model that automatically learns document level semantic representations to extract chemical-induced disease (CID) relations from articles by combining advantages of convolutional neural network and recurrent neural network. First, to purposefully collect contexts, candidate entities existing in multiple sentences of an article were masked to make the model have ability to discern candidate entities and general terms. Next, considering the contiguity and temporality among associated sentences as well as the topic of an article, a hierarchical network architecture was designed at the document level to capture semantic information of different types of text segments in an article. Finally, a softmax classifier performed the CID recognition.

Experimental results on the CDR corpus show that the proposed model achieves a good overall performance compared with other state-of-the-art methods. Although only using two types of embedding vectors, our approach can perform well for recognizing not only intra-sentential but also inter-sentential CID relations.

Keywords