Drug-Drug Interaction Extraction from Drug Labels

Chinmay Swami[†] Gursimrat Singh[†] Kumkum Kaushal[†*]

[†]Department of Computer Science and Engineering
State University of New York at Buffalo
{chinmayp,gsingh37,kumkumka}@buffalo.edu

Fall 2018

Abstract

Drug Drug interaction information extraction is important yet a complex task. Interaction between two administered drugs could lead to minor side effects or a major health threatening condition. Such types of adverse effects is found to be 8th leading cause of death is us as per a paper published in Internet Scientific Journals Goldstein et al. (2005). We plan on leveraging computational linguistics techniques paired with deep learning techniques to identify whether 2 drugs when administered to the same patient in the same time frame would cause any unexpected reactions.Currently we are looking at various deep learning techniques such as LSTM's, RNN's etc.

1 TODO

A script can convert the current xml into a vertical format (conll-like format), and another script for vice-versa for the next week.

1. find evaluation data (I emailed to the organizer);
2. convert training (and later eval) data into the vertical form (one word a line and an empty line for sentence boundaries if any), and learn the model using any sequence labeling algorithm
   (a) for now, split training data 80/10/10 or 90/10 for training/dev/eval or training/eval;
   (b) is ddi a directed or undirected graph? (if "undirected", can you find the "direction"?
3. use the eval script to get the results (for now train-split-data).
   (a) converted data (vertical form) as results should be restored to the original xml file format for evaluation
4. find out WHY/HOW previous work changed data (original vs. their own way modified data)
5. list previous work and their results

[*]Supervised by J. Park.

(a) previous results with table;

6. (todo later) add features for drug from `http://www.yapcwsoft.com/dd/padeldescriptor/`; Could you extract drug properties from this software, and use them as features? thanks, + features from (Asada et al., 2018) (see how to extract their features too);

for ddi training, your approach will be dependency parsing-like, which I will explain the next week; add the following paper in bib:

```
@inproceedings{dozat-manning:2017:ICLR,
address = {Toulon, France},
archivePrefix = {arXiv},
arxivId = {1611.01734},
author = {Dozat, Timothy and Manning, Christopher D.},
booktitle = {Proceedings of the 5th International Conference on Learning Representations (ICLR 2017)
eprint = {1611.01734},
month = {nov},
title = {{Deep Biaffine Attention for Neural Dependency Parsing}},
url = {http://arxiv.org/abs/1611.01734},
year = {2017}
}
```

2 Previous Work

2.1 Using GNN's and CNN's

This paper published on the same task used a combination of convolutional neural network and Graph convolutional network. The GCN's were used to predict interaction between two drugs using their molecular structures (Asada et al., 2018).

2.2 Using RNN's

This paper used Recursive neural networks(RNN's) to improve the performance of DDI extractions.This paper compared the results of RNN's with other conventional models used and concluded that the RNN's performed better than the other conventional methods (Lim et al., 2018).

2.3 Using LSTM's

This paper promulgates a technique which uses 3 models namely B-LSTM,AB-LSTM and joint AB-LSTM which are all based on long short-term memory. The paper discussed the performance of the 3 models and concluded that joint AB-LSTM's performed the best. The performance of the other two discussed models was also commensurate with the performance with existing state of the art models (Sahu and Anand, 2018).

3 Dataset details

Since this is a current task in TAC 2018 conference, the dataset has been provided by them. The training set consists of data about 22 drugs in XML format.Also test data set if about 50 drugs have been provided. Please find the link below for the datasets.

https://bionlp.nlm.nih.gov/tac2018druginteractions/

References

Asada, M., Miwa, M., and Sasaki, Y. (2018). Enhancing drug-drug interaction extraction from texts by molecular structure information. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 680–685. Association for Computational Linguistics.

Goldstein, J. N., Jaradeh, I. E., Jhawar, P., and Stair, T. O. (2005). ED drug-drug interactions: frequency & type, potential & actual, triage & discharge. *Internet Journal of Emergency & Intensive Care Medicine*.

Lim, S., Lee, K., and Kang, J. (2018). Drug drug interaction extraction from the literature using a recursive neural network. *PLoS ONE*.

Sahu, S. K. and Anand, A. (2018). Drug-drug interaction extraction from biomedical texts using long short-term memory network. *Journal of Biomedical Informatics*.