Use case and example usage scenario description of tender prototype

In recent years, biomedical relation extraction has been the focus of various community challenges due to the high volume of information published in papers, patents and other biomedical documents that could be converted into structured data. One biomedical relation type that has high impact in healthcare is drug-drug interactions (DDIs). When developing novel drugs, it is necessary to consider the effects that it might have when co-administered with other drugs. Therefore, it is highly valuable to automate the extraction of DDIs from text, to accelerate the drug development. Applying automatic relation extraction methods to a large corpus enables the generation of an explorable knowledge graph about DDIs, which could then be integrated with other sources of information such as diseases, genes and phenotypes.

While machine learning classifiers trained on word embeddings can learn to detect relations between entities, these classifiers do not consider the semantics of their respective domain. Ontologies aim at providing a structured representation of the knowledge about a domain. A domain-specific ontology frequently consists of a directed acyclic graph where each node is a concept (or entity) of the domain and the edges represent known relations between these concepts. This graph can be used to measure the semantic similarity between chemical entities, which can then be used to classify chemical compounds (1) and improve named entity recognition (2). ChEBI is an ontology where each node corresponds to a chemical compound (3). Assuming that, two chemical compounds that interact share more semantics than compounds that do not interact, then semantic similarity measures should capture this correlation and be able to improve results. For example, clozapine is subclass of piperazines, and as such, it should share some characteristics with other piperazines such as ziprasidone and sildenafil. It is more likely than a drug that has an interaction with clozapine also interacts with other piperazines when compared to a random chemical compound.

This project aims at demonstrating how biomedical ontologies can improve deep learning models for classification of biomedical relations. We will compare the effect of using ChEBI, a domain specific ontology, and Wordnet, a generic English language ontology, as external sources of information to train a classification model based on recurrent neural networks using Long Short-Time Memory (LSTM) units. This model will be evaluated on the SemEval 2013 Task 9: (DDI extraction) gold standard (4). Our baseline consists of a simple LSTM-based network trained on the shortest-dependency paths between every pair of drug entities mentioned in a sentence. Our hypothesis is that adding domain specific information to the network should improve the performance when compared to general sources of information. To this end, we will develop a software component to train relation classification models based on biomedical texts and ontologies, and classify text using the trained models.

References

1. Ferreira JD, Couto FM. Semantic Similarity for Automatic Classification of Chemical Compounds. Mitchell JBO, editor. PLoS Comput Biol [Internet]. 2010;6(9):e1000937.
2. Lamurias A, Ferreira JD, Couto FM. Improving chemical entity recognition through h-index based semantic similarity. J Cheminform [Internet]. 2015;7(Suppl 1):S13.
3. Hastings J, De Matos P, Dekker A, Ennis M, Harsha B, Kale N, et al. The ChEBI reference database and ontology for biologically relevant chemistry: Enhancements for 2013. Nucleic Acids Res. 2013;41(D1):456–63.
4. Segura-Bedmar I, Martínez P, Herrero-Zazo M. Lessons learnt from the DDIExtraction-2013 Shared Task. J Biomed Inform [Internet]. 2014;51(May):152–64.