

RESEARCH ARTICLE

Open Access



BO-LSTM: classifying relations via long short-term memory networks along biomedical ontologies

Andre Lamurias^{1,2*} , Diana Sousa¹, Luka A. Clarke² and Francisco M. Couto¹

Abstract

Background: Recent studies have proposed deep learning techniques, namely recurrent neural networks, to improve biomedical text mining tasks. However, these techniques rarely take advantage of existing domain-specific resources, such as ontologies. In Life and Health Sciences there is a vast and valuable set of such resources publicly available, which are continuously being updated. Biomedical ontologies are nowadays a mainstream approach to formalize existing knowledge about entities, such as genes, chemicals, phenotypes, and disorders. These resources contain supplementary information that may not be yet encoded in training data, particularly in domains with limited labeled data.

Results: We propose a new model to detect and classify relations in text, BO-LSTM, that takes advantage of domain-specific ontologies, by representing each entity as the sequence of its ancestors in the ontology. We implemented BO-LSTM as a recurrent neural network with long short-term memory units and using open biomedical ontologies, specifically Chemical Entities of Biological Interest (ChEBI), Human Phenotype, and Gene Ontology. We assessed the performance of BO-LSTM with drug-drug interactions mentioned in a publicly available corpus from an international challenge, composed of 792 drug descriptions and 233 scientific abstracts. By using the domain-specific ontology in addition to word embeddings and WordNet, BO-LSTM improved the F1-score of both the detection and classification of drug-drug interactions, particularly in a document set with a limited number of annotations. We adapted an existing DDI extraction model with our ontology-based method, obtaining a higher F1 score than the original model. Furthermore, we developed and made available a corpus of 228 abstracts annotated with relations between genes and phenotypes, and demonstrated how BO-LSTM can be applied to other types of relations.

Conclusions: Our findings demonstrate that besides the high performance of current deep learning techniques, domain-specific ontologies can still be useful to mitigate the lack of labeled data.

Keywords: Text mining, Drug-drug interactions, Deep learning, Long short term memory, Relation extraction

Background

Current relation extraction methods employ machine learning algorithms, often using kernel functions in conjunction with Support Vector Machines [1, 2] or based on features extracted from the text [3]. In recent years, deep learning techniques have obtained promising results in various Natural Language Processing (NLP) tasks

[4], including relation extraction [5]. These techniques have the advantage of being easily adaptable to multiple domains, using models pre-trained on unlabeled documents [6]. The success of deep learning for text mining is in part due to the high quantity of raw data available and the development of word vector models such as word2vec [7] and GloVe [8]. These models can use unlabeled data to predict the most probable word according to the context words (or vice-versa), leading to meaningful vector representations of the words in a corpus, known as word embeddings.

*Correspondence: alamurias@lasige.dl.fc.ul.pt

¹LASIGE, Faculdade de Ciências, Universidade de Lisboa, 1749 016, Lisboa, Portugal

²University of Lisboa, Faculty of Sciences, BiolSI - Biosystems & Integrative Sciences Institute, Campo Grande, C8 bldg, 1749 016, Lisboa, Portugal



A high volume of biomedical information relevant to the detection of Adverse Drug Reactions (ADRs), such as Drug-Drug Interactions (DDI), is mainly available in articles and patents [9]. A recent review of studies about the causes of hospitalization in adult patients has found that ADRs were the most common cause, accounting for 7% of hospitalizations [10]. Another systematic review focused on the European population, identified that 3.5% of hospital admissions were due to ADRs, while 10.1% of the patients experienced ADRs during hospitalization [11].

The knowledge encoded in the ChEBI (Chemical Entities of Biological Interest) ontology is highly valuable for detection and classification of DDIs, since it provides not only the important characteristics of each individual compound but also, more importantly, the underlying semantics of the relations between compounds. For instance, dopamine (CHEBI:18243), a chemical compound with several important roles in the brain and body, can be characterized as being a catecholamine (CHEBI:33567), an aralkylamino compound (CHEBI:64365) and an organic aromatic compound (CHEBI:33659) (Fig. 1). When predicting if a certain drug interacts with dopamine, its ancestors will provide additional information that is not usually directly expressed in the text. While the reader can consult additional materials to better understand a biomedical document, current relation extraction models are trained solely on features extracted from the training corpus. Thus, ontologies confer an advantage to relation extraction models due to the semantics encoded in them regarding a particular domain. Since ontologies are described in a common machine-readable format, methods based on ontologies can be applied to different domains and incorporated with other sources of knowledge, bridging the semantic gap between relation extraction models, data sources, and results [12].

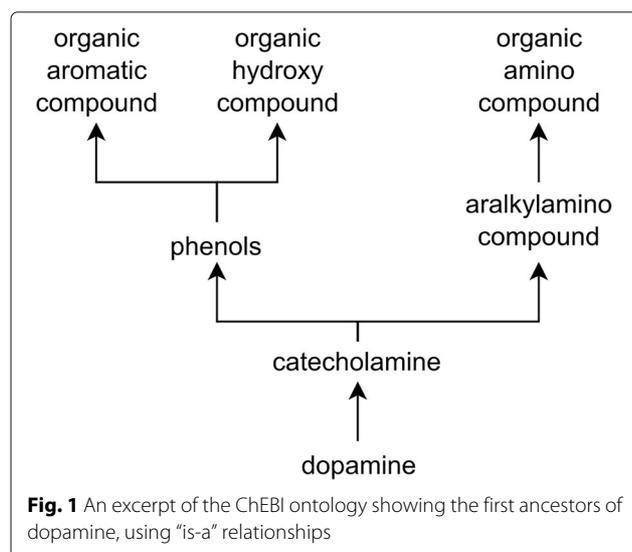


Fig. 1 An excerpt of the ChEBI ontology showing the first ancestors of dopamine, using “is-a” relationships

Deep learning for biomedical NLP

Current state-of-the-art text mining methods employ deep learning techniques, such as Recurrent Neural Networks (RNN), to train classification models based on word embeddings and other features. These methods use architectures composed of multiple layers, where each layer attempts to learn a different kind of representation of the input data. This way, different types of tasks can be trained using the same input data. Furthermore, there is no need to manually craft features for a specific task.

Long Short-Term Memory (LSTM) networks have been proposed as an alternative to regular RNN [13]. LSTMs are a type of RNN that can handle long dependencies, and thus are suitable for NLP tasks, which involve long sequences of words. When training the weights of an RNN, the contribution of the gradients may vanish while propagating for long sequences of words. LSTM units account for this vanishing gradient problem through a gated architecture, which makes it easier for the model to capture long-term dependencies. Recently, LSTMs have been applied to relation extraction tasks in various domains. Miwa and Bansal [14] presented a model that extracted entities and relations based on bidirectional tree-structured and sequential LSTM-RNNs. The authors evaluated this model on three datasets, including the SemEval 2010 Task 8 dataset, which defines 10 general semantic relations types between nominals [15].

Bidirectional LSTMs have been proposed for relation extraction, obtaining better results than one-directional LSTMs on the SemEval 2010 dataset [16]. In this case, at each time step, there are two LSTM layers, one that reads the sentence from left to right, and another that reads from right to left. The output of both layers is combined to produce a final score.

The model proposed by Xu et al. [17] combines Shortest Dependency Paths (SDP) between two entities in a sentence with linguistic information. SDPs are informative features for relations extraction since these contain the words of the sentence that refer directly to both entities. This model has a multichannel architecture, where each channel makes use of information from a different source along the SDP. The main channel, which contributes the most to the performance of the model, uses word embeddings trained on the English Wikipedia with word2vec. Additionally, the authors study the effect of adding channels consisting of the part-of-speech tags of each word, the grammatical relations between the words of the SDP, and the WordNet hypernyms of each word. Using all four channels, the F1-score of the SemEval 2010 Task 8 was 0.0135 higher than when using only the word embeddings channel. Although WordNet can be considered an ontology, its semantic properties were not integrated in this work, since only the word class is extracted, and the relations between classes are not considered.

Deep learning approaches to DDI classification have been proposed in recent years, using the SemEval 2013: Task 9 DDI extraction corpus to train and evaluate their performance. Zhao et al. [18] proposed a syntax convolutional neural network for DDI extraction, using word embeddings. Due to its success on other domains, LSTMs have also been used for DDI extraction [19–22]. Xu et al. [21] proposed a method that combines domain-specific biomedical resources to train embedding vectors for biomedical concepts. However, their approach uses only contextual information from patient records and journal abstracts and does not take into account the relations between concepts that an ontology provides. While these works are similar to ours, we present the first model that makes use of a domain-ontology to classify DDIs.

Ontologies for biomedical text mining

While machine learning classifiers trained on word embeddings can learn to detect relations between entities, these classifiers may miss the underlying semantics of the entities according to their respective domain. However, the semantics of a given domain are, in some cases, available in the form of an ontology. Ontologies aim at providing a structured representation of the semantics of the concepts in a domain and their relations [23]. In this paper, we consider a domain-specific ontology as a directed acyclic graph where each node is a concept (or entity) of the domain and the edges represent known relations between these concepts [24]. This is a common representation of existing biomedical ontologies, which are nowadays a mainstream approach to formalize knowledge about entities, such as genes, chemicals, phenotypes, and disorders.

Biomedical ontologies are usually publicly available and cover a large variety of topics related to Life and Health Sciences. In this paper, we use ChEBI, an ontology for chemical compounds with biological interest, where each node corresponds to a chemical compound [25]. The latest release of ChEBI contains nearly 54k compounds and 163k relationships. Note that, the success of exploring a given biomedical ontology for performing a specific task can be easily extended to other topics due to the common structure of biomedical ontologies. For example, the same measures of metadata quality have been successfully applied to resources annotated with different biomedical ontologies [26].

Other authors have previously combined ontological information with neural networks, to improve the learning capabilities of a model. Li et al. [27] mapped each word to a WordNet sense disambiguation to account for the different meanings that a word may have and the relations between word senses. Ma et al. [28] proposed the LSTM-OLSI model, which indexes documents based on the word-level contextual information from

the DBpedia ontology and document-level topic modeling. Some authors have explored graph embedding techniques, converting relations to a low dimensional space which represents the structure and properties of the graph [29]. For example, Kong et al. [30] combined heterogeneous sources of information, such as ontologies, to perform multi-label classification, while Dasigi et al. [31] presented an embedding model based on ontology concepts to represent word tokens.

However, few authors have explored biomedical ontologies for relation extraction. Textpresso is a project that aims at helping database curation by automatically extracting biomedical relations from research articles [32]. Their approach incorporates an internal ontology to identify which terms may participate in relations according to their semantics. Other approaches measure the similarity between the entities and use the value as a feature for a machine learning classifier [33]. One of the teams that participated in the BioCreative VI ChemProt task used ChEBI and Protein Ontology to extract additional features for a neural network model that extracted relation between chemicals and proteins [34]. To the best of our knowledge, our work is the first attempt at incorporating ancestry information from biomedical ontologies with deep learning to extract relations from text.

In this manuscript, we propose a new model, BO-LSTM that can explore domain information from ontologies to improve the task of biomedical relation extraction using deep learning techniques. We 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 LSTM networks. This model was evaluated on a publicly available corpus of 792 drug descriptions and 233 scientific abstracts annotated with DDIs relevant to the study of adverse drug effects. Using the domain-specific ontology in addition to word embeddings and WordNet, BO-LSTM improved the F1-score of the classification of DDIs by 0.0207. Our model was particularly efficient with document types that were less represented in the training data. Moreover, we improved the F1-score of an existing DDI extraction model by 0.022 by adding our proposed ontology information, and demonstrated its applicability to other domains by generating a corpus of gene-phenotype relations and training our model on that corpus. The code and results obtained with the model can be found on our GitHub repository (<https://github.com/lasigeBioTM/BOLSTM>), while a Docker image is also available (<https://hub.docker.com/r/andrelamurias/bolstm>), simplifying the process of training new classifiers and applying them to new data. We also made available the corpus produced for gene-phenotype relations, where each entity is mapped to an ontology concept. These results support our hypothesis that domain-specific information is useful

to complement data-intensive approaches such as deep learning.

Methods

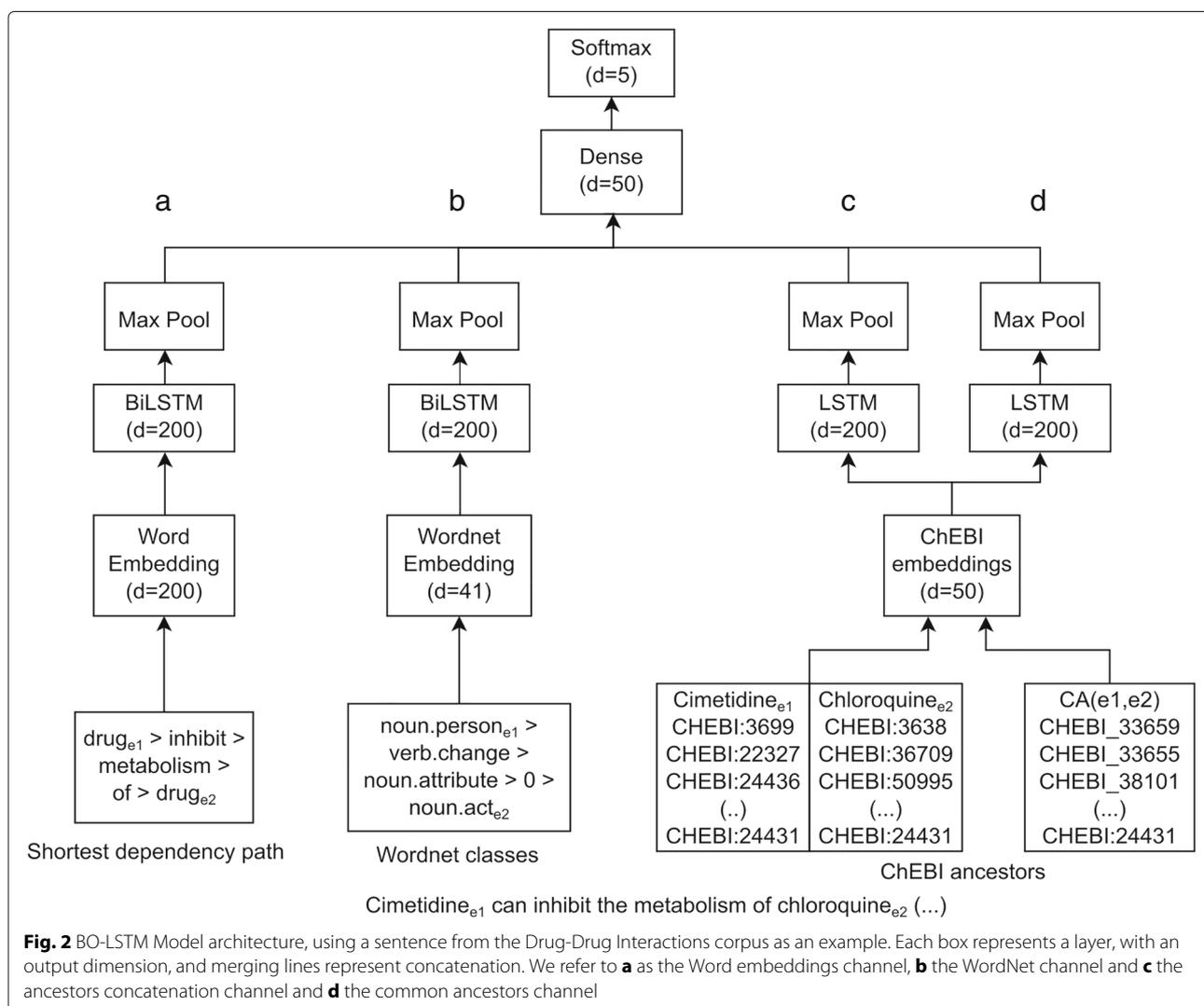
In this section, we describe the proposed BO-ISTM model in detail, as shown in Fig. 2, with a focus on the aspects that refer to the use of biomedical ontologies.

Data preparation

The objective of our work is to identify and classify relations between biomedical entities found in natural language text. We assume that the relevant entities are already recognized. Therefore, we process the input data in order to generate instances to be classified by the model. Considering the set of entities E mentioned in a sentence, we generate $\binom{E}{2}$ instances of that sentence. We refer to each instance as a candidate pair, identified by the two entities that constitute that pair, regardless of the

order. A relation extraction model will assign a class to each candidate pair. In some cases, it is enough to simply classify the candidate pairs as negative or positive, while in other cases different types of positive relations are considered.

An instance should contain the information necessary to classify a candidate pair. Therefore, after tokenizing each sentence, we obtain the Shortest Dependency Path (SDP) between the entities of the pair. For example, in the sentence “Laboratory Tests Response to Plenaxis_{e1} should be monitored by measuring serum total testosterone_{e1} concentrations just prior to administration on Day 29 and every 8 weeks thereafter”, the shortest path between the entities would be Plenaxis - Response - monitored - by - measuring - concentrations - testosterone. For both tokenization and dependency parsing, we use the spaCy software library (<https://spacy.io/>). The text of each entity that appears in the SDP, including the candidate entities,



is replaced by the generic string to reduce the effect of specific entity names on the model. For each element of the SDP, we obtain the WordNet hypernym class using the tool developed by Ciaramita and Altun [35].

To focus our attention on the effect of the ontology information, we use pre-trained word embedding vectors. Pyysalo et al. [36] released a set of vectors trained on PubMed abstracts (nearly 23 million) and PubMed Central full documents (nearly 700k), with the word2vec algorithm [7]. Since these vectors were trained on a large biomedical corpus, it is likely that its vocabulary will contain more words relevant to the biomedical domain than the vocabulary of a generic corpus.

We match each entity to an ontology concept so that we can then obtain its ancestors. Ontology concepts contain an ID, a preferred label, and, in most cases, synonyms. While pre-processing the data, we match each entity to the ontology using fuzzy matching. The adopted implementation uses the Levenshtein distance to assign a score to each match.

Our pipeline first attempts to match the entity string to a concept label. If the match has a score equal to or higher than 0.7 (determined empirically), we accept that match and assign the concept ID to that entity. Otherwise, we match to a list of synonyms of ontology concepts. If that match has a score higher than the original score, we assign the ID of the matched synonym to the entity, otherwise, we revert to the original match. It is preferable to match to a concept label since these are more specific and should reflect the most common nomenclature of the concepts. This way, every entity was matched to a ChEBI concept, either to its preferred label or to a synonym. Due to the automatic linking method used, we cannot assume that every match is correct, but fuzzy matching has been used for similar purposes [37], so we can assume that the best match is chosen. We matched 9020 unique entities to the preferred label and 877 to synonyms, and 1283 unique entities had an exact match to either a preferred label or synonym.

The DDI corpus used to evaluate our method has a high imbalance of positive and negative relations, which hinders the training of a classification model. Even though only entities mentioned in the same sentence are considered as candidate DDIs, there is still a ratio of 1:5.9 positive to negative instances. Other authors have suggested reducing the number of negative relations through simple rules [38, 39]. We excluded from training and automatically classify as negative the pairs that fit the following rules:

- entities have the same text (regardless of case): in nearly every case a drug does not interact with itself;
- the only text between the candidate pair is punctuation: consecutive entities, in the form of lists

and enumerations, are not interacting, as well as instances where the abbreviation of an entity is introduced;

- both entities have anti-positive governors: we follow the methodology proposed by [38], where the headwords of entities that do not interact are used to filter less informative instances.

With this filtering strategy, we used only 15,697 of the 27,792 pairs of the training corpus, obtaining a ratio of 1:3.5 positive to negative instances.

We developed a corpus of 228 abstracts annotated with human phenotype-gene relations, which we refer to as the HP corpus, to demonstrate how our model could be applied to other relation extraction tasks. This corpus was based on an existing corpus that were manually annotated with 2773 concepts of the Human Phenotype Ontology [40], corresponding to 2170 unique concepts. The developers of the Human Phenotype Ontology made available a file that links phenotypes and genes that are associated with the same diseases. Each gene of this file was automatically annotated on the HP corpus through exact string matching, resulting in 360 gene entity mentions. Then, we assumed that every gene-phenotype pair that co-occurred in the same sentence was a positive instance if this relation existed in the file. While the phenotype entities were manually mapped to the Human Phenotype Ontology, we had to employ an automatic method to obtain the most representative Gene Ontology [41, 42] concept of each gene, giving preference to concepts inferred from experiments. We applied the same pre-processing steps as for the DDI corpus, except for entity matching and negative instance filtering. This corpus is available at <https://github.com/lasigeBioTM/BOLSTM/tree/master/HP%20corpus>.

BO-LSTM model

The main contribution of this work is the integration of ontology information with a neural network classification model. A domain-specific ontology is a formal definition of the concepts related to a specific subject. We can define an ontology as a tuple $\langle C, R \rangle$, where C is the set of concepts and R the set of relations between the concepts, where each relation is a pair of concepts (c_1, c_2) with $c_1, c_2 \in E$. In our case, we consider only subsumption relations (is-a), which are transitive, i.e. if $(c_1, c_2) \in R$ and $(c_2, c_3) \in R$, then we can assume that (c_1, c_3) is a valid relation. Then, the ancestors of concept c are given by

$$Anc(c) = a : (c, a) \in T \quad (1)$$

where T is the transitive closure of R on the set E , i.e., the smallest relation set on E that contains R and is transitive. Using this definition, we can define the common ancestors of concepts c_1 and c_2 as

$$CA(c_1, c_2) = Anc(c_1) \cap Anc(c_2) \quad (2)$$

and the concatenation of the ancestors of concepts c_1 and c_2 as

$$Conc(c_1, c_2) = Anc(c_1) \oplus Anc(c_2) \quad (3)$$

We consider two types of representations of a candidate pair based on the ancestry of its elements: the first consisting of the concatenation of the sequence of ancestors of each entity; and second, consisting of the common ancestors between both entities. Each set of ancestors is sorted by its position in the ontology so that more general concepts are in the first positions and the final position is the concept itself. Common ancestors are also used in some semantic similarity measures [43–45], since they normally represent the common information between two concepts. Due to the fact that in some cases there can be almost no overlap between the ancestors of two concepts, the concatenation provides an alternative representation.

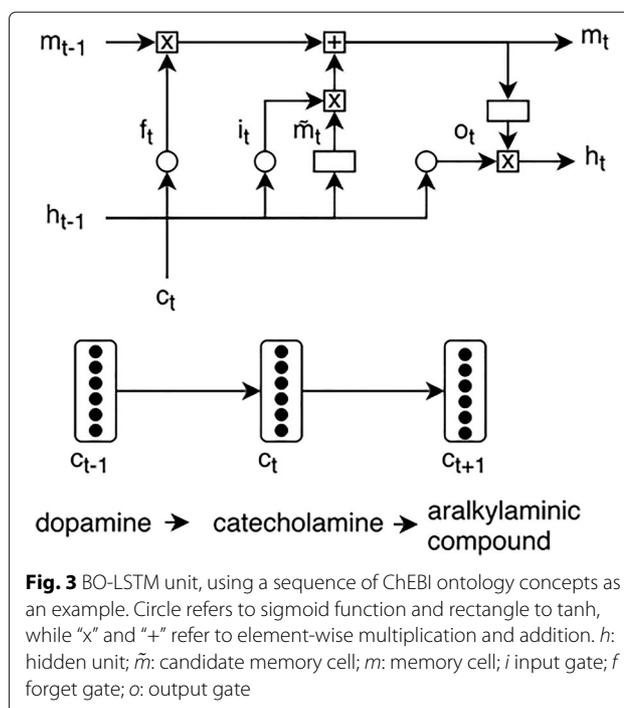
We first represent each ontology concept as a one-hot vector v_c , a vector of zeros except for the position corresponding to the ID of the concept. The ontology embedding layer transforms these sparse vectors into dense vectors, known as embeddings, through an embedding matrix $M \in \mathbb{R}^{D \times C}$, where D is the dimensionality of the embedding layer and C is the number of concepts of the ontology. Then, the output of the embedding layer is given by

$$f(c) = M \cdot v_c$$

In our experiments, we set the dimensionality of the ontology embedding layer as 50, and initialized its values randomly. Then, these values were tuned during training through back-propagation.

The sequence of vectors representing the ancestors of the terms is then fed into the LSTM layer. Figure 3 exemplifies how we adapted this architecture to our model, using a sequence of ontology concepts as input. After the LSTM layer, we use a max pool layer which is then fed into a dense layer with a sigmoid activation function. We experimented with bypassing this dense layer, obtaining inferior results. Finally, a softmax layer outputs the probability of each class.

Each configuration of our model was trained through mini-batch gradient descent with the Adam algorithm [46] and with cross-entropy as the loss function, with a learning rate of 0.001. We used the dropout strategy [47] to reduce overfitting on the trained embeddings and weights. We used a dropout rate of 0.5 on every layer except the penultimate and output layers. We tuned the hyperparameters common to all configurations using only the word embeddings channel on the validation set. Each model was trained until the validation loss stopped decreasing. The experiments were performed on an Intel Xeon



CPU (X3470 @ 2.93 GHz) with 16 GB of RAM and on a GeForce GTX 1080 Ti GPU with 11GB of RAM.

The ChEBI and WordNet embedding layers were trained along with the other layers of the network. The DDI corpus contains 1757 of the 109k concepts of the ChEBI ontology. Since this is a relatively small vocabulary, we believe that this approach is robust enough to tune the weights. For the size of the WordNet embedding layer, we used 50 as suggested by Xu et al. [17], while for the ChEBI embedding layer, we tested 50, 100 and 150, obtaining the best performance with 50.

Baseline models

As a baseline, we implemented a model based on the SDP-LSTM model of Xu et al. [17]. The SDP-LSTM model makes use of four types of information: word embeddings, part-of-speech tags, grammatical relations and WordNet hypernyms, which we refer to as channels. Each channel uses a specific type of input information to train an LSTM-based RNN layer, which is then connected to a max pooling layer, the output of the channel. The output of each channel is concatenated, and connected to a densely-connected hidden layer, with a sigmoid activation function, while a softmax layer outputs the probabilities of each class.

Xu et al. show that it is possible to obtain high performance on a relation extraction task using only the word representations channel. For this reason, we use a version of our model with only this channel as the baseline.

We employ the previously mentioned pre-trained word embeddings as input to the LSTM layer.

Additionally, we make use of WordNet as an external source of information. The authors of the SDP-LSTM model showed that WordNet contributed to an improvement of the F1-score on a relation extraction task. We use the tool developed by Ciaramita and Altun [35] to obtain the WordNet classes of each word according to 41 semantic categories, such as “noun.group” and “verb.change”. The embeddings of this channel were set to be 50-dimensional and tuned during the training of the model.

We adopted a second baseline model to make a stronger comparison with other DDI extraction models, based on the model presented by Zhang et al. [48]. Their model uses the sentence and SDP of each instance to train a hierarchical LSTM network. This model is constituted by two levels of LSTMs which learn feature representations of the sentence and SDP based on word, part-of-speech and distance to entity. An embedding attention mechanism is used to weight the importance of each word to the two entities that constitute each pair. We kept the architecture and hyperparameters of their model, and added another type of input, based on the common ancestors and concatenation of each entity’s ancestors. We applied the same attention mechanism, so that the most relevant ancestors have a larger weight on the LSTM. We ran the original Zhang et al. model to replicate the results, and then ran again with ontology information.

Results

We evaluated the performance of our BO-LSTM model on the SemEval 2013: Task 9 DDI extraction corpus [49]. This gold standard corpus consists of 792 texts from DrugBank [50], describing chemical compounds, and 233 abstracts from the Medline database [51]. DrugBank is a cheminformatics database containing detailed drug and drug target information, while Medline is a database of bibliographic information of scientific articles in Life and Health Sciences. Each document was annotated with pharmacological substances and sentence-level DDIs. We

refer to each combination of entities mentioned in the same sentence as a candidate pair, which could either be positive if the text describes a DDI, or negative otherwise. In other words, a negative candidate is a candidate pair that is not described as interacting in the text. Each positive DDI was assigned one of four possible classes: mechanism, effect, advice, and int, when none of the others were applicable.

In the context of the competition, the corpus was separated into training and testing sets, containing both DrugBank and Medline documents. We maintained the test set partition and evaluated on it, as it is the standard procedure on this gold standard. After shuffling we used 80% of the training set to train the model and 20% as a validation set. This way, the validation set contained both DrugBank and Medline documents, and overfitting to a specific document type is avoided. It has been shown that the DDIs of the Medline documents are more difficult to detect and classify, with the best systems having almost a 30 point F1-score difference to the DrugBank documents [52].

We implemented the BO-LSTM model in Keras, a Python-based deep learning library, using the TensorFlow backend. The overall architecture of the BO-LSTM model is presented in Fig. 2. More details about each layer can be found in the “Methods” section. We focused on the effect of using different sources of information to train the model. As such, we tuned the hyperparameters to obtain reasonable results, using as reference the values provided by other authors that have applied LSTMs to this gold standard [18, 19]. We first trained the model using only the word embeddings of the SDP of each candidate pair (Fig. 2a). Then we tested the effect of adding the WordNet classes as a separate embedding and LSTM layer (Fig. 2b) Finally, we tested two variations of the ChEBI channel: first using the concatenation of the sequence of ancestors of each entity (Fig. 2c), and second using the sequence of common ancestors of both entities (Fig. 2d).

Table 1 shows the DDI detection results obtained with each configuration using the evaluation tool provided by the SemEval 2013: Task 9 organizers on the gold standard, while Table 2 shows the DDI classification results,

Table 1 Evaluation scores obtained for the DDI detection task on the DDI corpus and on each type of document, comparing different configurations of the model

Configuration	DDI test			DrugBank			Medline		
	P	R	F	P	R	F	P	R	F
Word embeddings	0.7551	0.6865	0.7192	0.7620	0.7158	0.7382	0.6389	0.377	0.4742
+ WordNet	0.716	0.6936	0.7046	0.7267	0.7143	0.7204	0.5800	0.4754	0.5225
+ Common Ancestors	0.7661	0.6738	0.7170	0.7723	0.7003	0.7345	0.6667	0.3607	0.4681
+ Concat. Ancestors	0.7078	0.7489	0.7278	0.7166	0.7578	0.7366	0.6032	0.623	0.6129
+ WordNet + Ancestors	0.6572	0.8184	0.7290	0.6601	0.8385	0.7387	0.5574	0.5574	0.5574

Evaluation metrics used: Precision (P), Recall (R) and F1-score (F). Each row represents the addition of an information source to the initial configuration. Boldface indicates the configuration with highest score for each measure

Table 2 Evaluation scores obtained for the DDI classification task on the DDI corpus and on each type of document, comparing different configurations of the model

Configuration	DDI test			DrugBank			Medline		
	P	R	F	P	R	F	P	R	F
Word embeddings	0.5819	0.5291	0.5542	0.5868	0.5512	0.5685	0.5000	0.2951	0.3711
+ WordNet	0.5754	0.5574	0.5663	0.5845	0.5745	0.5795	0.4600	0.3770	0.4144
+ Common Anc.	0.5968	0.5248	0.5585	0.6045	0.5481	0.5749	0.5152	0.2787	0.3617
+ Concat. Anc.	0.5282	0.5589	0.5431	0.5286	0.5590	0.5434	0.4921	0.5082	0.5000
+ WordNet + Anc.	0.5182	0.6454	0.5749	0.5171	0.6568	0.5787	0.4590	0.4590	0.4590

Evaluation metrics used: Precision (P), Recall (R) and F1-score (F). Each row represents the addition of an information source to the initial configuration. Boldface indicates the configuration with highest score for each measure

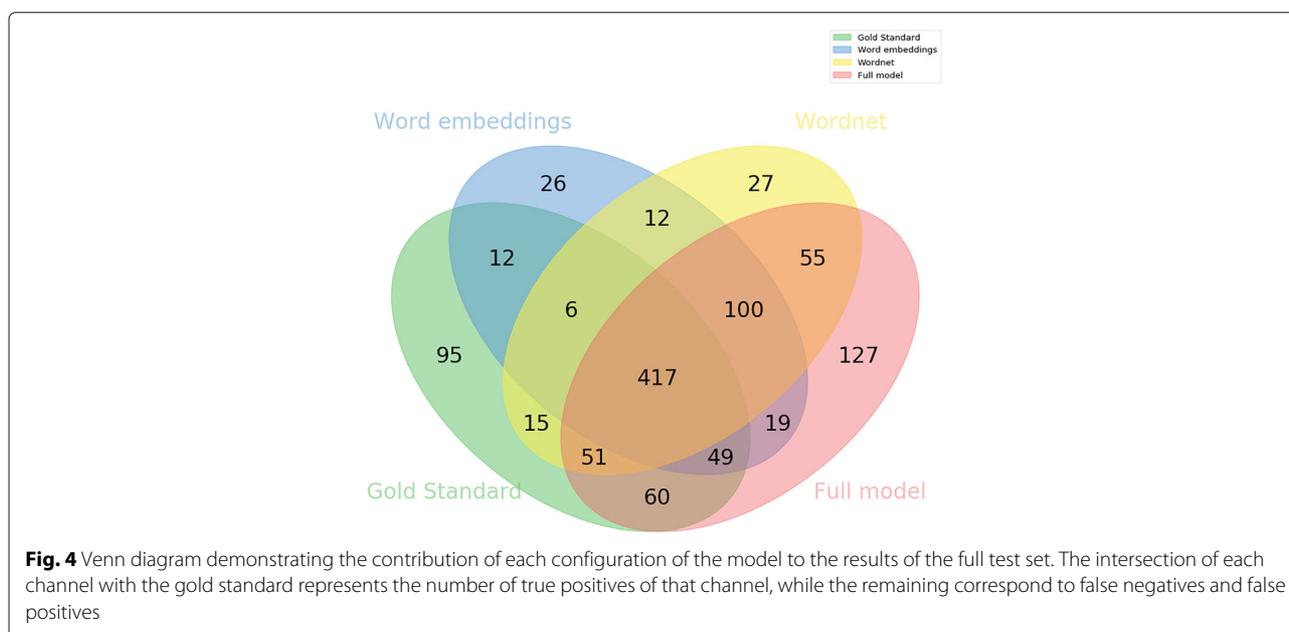
using the same evaluation tool and gold standard. The difference between these two tasks is that while detection ignores the type of interactions, the classification task requires identifying the positive pairs and also their correct interaction type. We compare the performance on the whole gold standard, and on each document type (DrugBank and Medline). The first row of each table shows the results obtained using an LSTM network trained solely on the word embeddings of the SDP of each candidate pair. Then, we studied the impact of adding each information channel on the performance of the model, and the effect of using all information channels, as shown in Fig. 2.

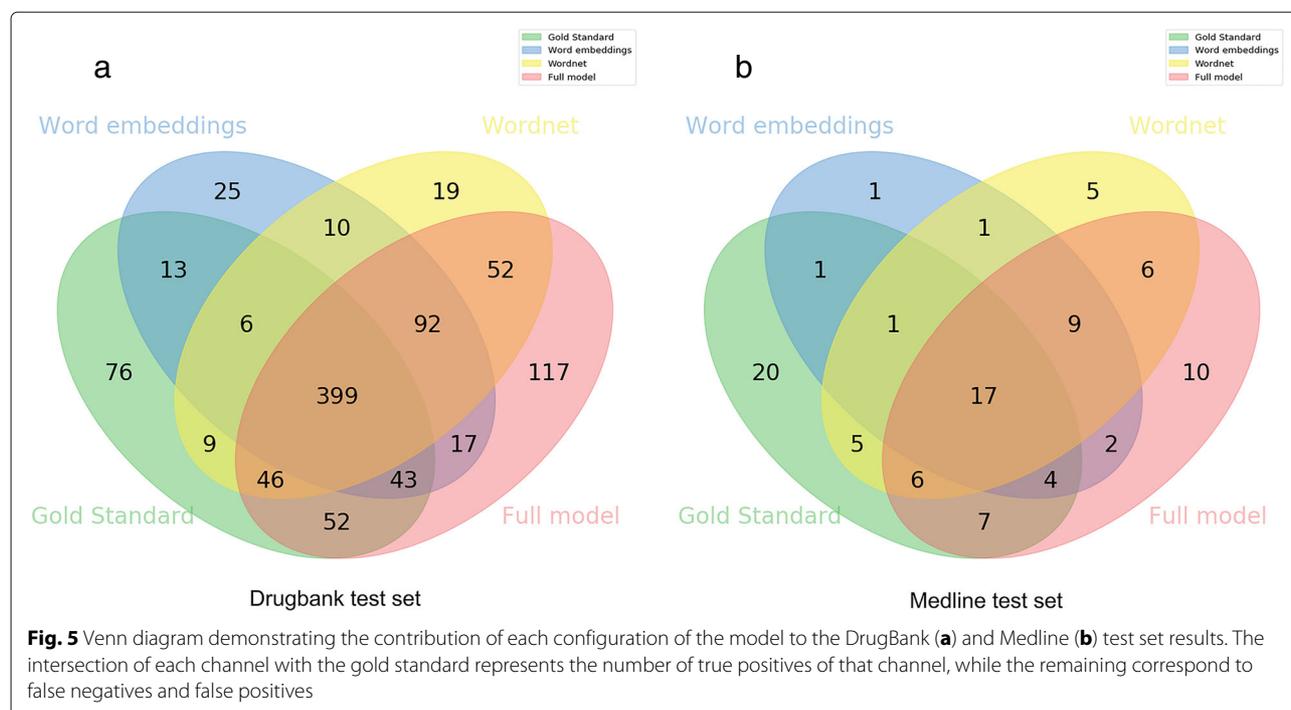
For the detection task, using the concatenation of ancestors results in an improvement of the F1-score in the Medline dataset, contributing to an overall improvement of the F1-score in the full test set. The most notable improvement was in the recall of the Medline dataset, where the concatenation of ancestors increased this score by 0.246.

The usage of ontology ancestors did not improve the F1-score of detection of DDIs in the DrugBank dataset. In every test set, it is possible to observe that the concatenation of ancestors results in a higher recall while considering only the common ancestors is more beneficial to precision. Combining both approaches with the WordNet channel results in a higher F1-score.

Regarding the classification task (Table 2), the F1-score was improved on each dataset by the usage of the ontology channel. Considering only the common ancestors led to an improvement of the F1-score in the DrugBank dataset and on the full corpus, while the concatenation improved the Medline F1-score, similarly to the detection results.

To better understand the contribution of each channel, we studied the relations detected by each configuration by one or more channels, and which of those were also present in the gold standard. Figures 4 and 5 show the





intersection of the results of each channel in the full, DrugBank, and Medline test sets. We compare only the results of the detection task, as it is simpler to analyze and show the differences in the results of different configurations. In Fig. 4, we can visualize false negatives as the number of relations unique to the gold standard and the false positives of each configuration as the number of relations that does not intersect with the gold standard. The difference between the values of this figure and the sum of their respective values in Fig. 5 is due to the system being executed once for each dataset. Overall 369 relations in the full test set were not detected by any configuration of our system, out of a total of 979 relations in the gold standard. We can observe that 60 relations were detected only when adding the ontology channels.

In the Medline test set, the ontology channel identified 7 relations that were not identified by any other configuration (Fig. 5b). One of these relations was the effect of quinpirole treatment on amphetamine sensitization. Quinpirole has 27 ancestors in the ChEBI ontology, while amphetamine has 17, and they share 10 of these ancestors, with the most informative being “organonitrogen compound”. While this information is not described in the original text, but only encoded in the ontology, it is relevant to understand if the two entities can participate in a relation. However, this comes at the cost of precision, since 10 incorrect DDIs were classified by this configuration.

To empirically compare our results with the state-of-the-art of the DDI extraction, we compiled the most

relevant works on this task in Table 3. The first line refers to the system that obtained the best results on the original SemEval task [38, 53]. Since then, other authors have presented approaches for this task, most recently using deep learning algorithms. In Table 3 we compare the machine learning architecture used by each system, and the results reported by the authors. Since some authors focused only on the DDI classification task, we could not obtain the DDI detection results for those systems, hence the missing values. We were only able to replicate the results of Zhang et al. [48]. Since this system followed an architecture similar to ours, we adapted the model with our ontology-based channel, as described in the “Methods” section. This modification to the model resulted in an improvement of 0.022

Table 3 Comparison of DDI extraction systems

System	Architecture	Best F1-score
FBK-irst [38]	SVM	0.651
SCNN [18]	CNN	0.686
Joint AB-LSTM [19]	LSTM	0.6939
Att-BLSTM [22]	LSTM	0.773
DLSTM [20]	LSTM	0.6839
BR-LSTM [21]	LSTM	0.7115
Zhang et al. 2018 [48]	LSTM	0.729
Zhang et al. 2018 + BO-LSTM	LSTM	0.751

The architectures mentioned are Support Vector Machines (SVM), Convolutional Neural Networks (CNN) and LSTMs

to the F1-score. Our version of this model is also available on our page along with the BO-LSTM model.

We used the HP corpus to demonstrate the generalizability of our method. This case-study served only as a proof-of-concept, it was not our intent to measure the performance of the model, given the limited number of annotations and the dependence on the quality of using exact string matching to identify the genes. For example, we may have missed correct relations in the corpus, because they were not in the reference file or the gene name was not correctly identified.

Therefore, we used 60% (137 documents) of the corpus to train the model and 40% (91 documents) to manually evaluate the relations predicted with that model. For example, in the following sentence:

Multiple angiofibromas , collagenomas , lipomas , confetti-like hypopigmented macules and multiple gingival papules are cutaneous manifestations of MEN1 and should be looked for in both family members of patients with MEN1 and individuals with hyperparathyroidism of other MEN1-associated tumors .

the model identified the relation between the phenotype “angiofibromas” and the gene “MEN1”. One recurrently identified relation by our model that was not present on the phenotype-gene associations file is between the phenotype ‘neurofibromatosis’ and the gene ‘NF2’:

Clinical and genetic data of 10 patients with neurofibromatosis 2 (NF-2) are presented .

Despite this relation not being described in the previous sentence, it is predicted given its presence in the phenotype-gene associations files. With a larger number of annotations in the training corpus, we expect this error to disappear.

Discussion

Comparing the results across the two types of documents, we can observe that our model was most beneficial to the Medline test set. This set contains only 1301 sentences from 142 documents for training, while the DrugBank set contains 5675 sentences from 572 documents. Naturally, the patterns of the DrugBank documents will be easier to learn than the ones of the Medline documents because more examples are shown to the model. Furthermore, the Medline set has 0.18 relations per sentence, while the DrugBank set has 0.67 relations per sentence. This means that DDIs are described much more sparsely than in the DrugBank set. This demonstrates that our model is able to obtain useful knowledge that is not described in the text.

One disadvantage of incorporating domain information in a machine learning approach is that it reduces its applicability to other domains. However, biomedical ontologies have become ubiquitous in biomedical research. One of the most successful cases of a biomedical ontology is the Gene Ontology, maintained by the Gene Ontology Consortium [54]. The Gene Ontology defines over 40,000 concepts used to describe the properties of genes. This project is constantly updated, with new concepts and relations being added every day. However, there are ontologies for more specific subjects, such as microRNAs [55], radiology terms [56] and rare diseases [57]. BioPortal is a repository of biomedical ontology, currently hosting 685 ontologies. Furthermore, while manually labeled corpora are created specifically to train and evaluate text mining applications, ontologies have diverse applications, i.e., they are not developed for this specific purpose.

We evaluate the proposed model on the DDI corpus because it is associated with a SemEval task, and for this reason, it has been the subject of many studies since its release. However, while applying our model to a single domain, we designed its architecture so it can fit any other domain-specific ontology. To demonstrate this, we developed a corpus of gene-phenotype relations annotated with Human Phenotype and Gene ontology concepts, and applied our model to it. Therefore, the methodology proposed can be easily followed to apply to any other biomedical ontology that describes the concepts of a particular domain. For example, the Disease Ontology [58], that describes relations between human diseases, could be used with the BO-LSTM model on a disease relation extraction task, as long as there is an annotated training corpus.

While we studied the potential of domain-specific ontologies based only on the ancestors of each entity, there are other ways to integrate semantic information from ontologies into neural networks. For example, one could consider only the ancestors with the highest information content, since those would be the most helpful to characterize an entity. The information content can be estimated either by the probability of a given term in the ontology or in an external dataset. Alternatively, a semantic similarity measure that accounts for non-transitive relations could be used to find similar concepts to the entities of the relation [59], or one that considers only the most relevant ancestors [60]. The quality of the ontology embeddings could also be improved by pre-training on a larger dataset, which would include a wider variety of concepts.

Conclusions

This work demonstrates how domain-specific ontologies can improve deep learning models for classification of biomedical relations. We developed a model, BO-LSTM

which combines biomedical ontologies with LSTM units to detect and classify relations in text. In this manuscript, we demonstrate that ontologies can improve the performance of deep learning techniques for biomedical relation extraction, in particular for situations with a limited number of annotations available, which was the case of the Medline dataset. Furthermore, we explored how it can be adapted to other relation extraction domains, for example, gene-phenotype relations. Considering that biomedical ontologies are openly available and regularly updated as the knowledge on the domain progresses, they should be considered important information sources for relation extraction.

Abbreviations

ADR: Adverse Drug Reactions; ChEBI: Chemical Entities of Biological Interest; DDI: Drug-drug interactions; LSTM: Long short-term memory; NLP: Natural Language Processing; RNN: Recurrent Neural Networks; SDP: Shortest Dependency Paths

Acknowledgements

We acknowledge the help of Nuno Dionisio in setting up the machine to run the experiments.

Funding

This work was supported by FCT through funding of the DeST: Deep Semantic Tagger project, ref. PTDC/CCI-BIO/28685/2017, LaSIGE Research Unit, ref. UID/CEC/00408/2013 and BioSI, ref. ID/MULTI/04046/2013. AL is recipient of a fellowship from BioSys PhD programme (ref PD/BD/106083/2015) from FCT (Portugal).

Availability of data and materials

The data and code used for this study are available at <https://github.com/lasigeBioTM/BOLSTM>.

Authors' contributions

All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 25 July 2018 Accepted: 12 December 2018

Published online: 07 January 2019

References

- Zelenko D, Zelenko D, Aone C, Aone C, Richardella A, Richardella A. Kernel Methods for Relation Extraction. *J Mach Learn Res.* 2003;3:1083–106. <https://doi.org/10.3115/1118693.1118703>.
- Reichartz F, Korte H, Paass G. Semantic relation extraction with kernels over typed dependency trees. Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '10; 2010, p. 773. <https://doi.org/10.1145/1835804.1835902>.
- Kambhatla N. Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations. Proceedings of the ACL 2004 on Interactive poster and demonstration sessions; 2004, p. 22. <https://doi.org/10.3115/1219044.1219066>.
- Collobert R, Weston J, Bottou L, Karlen M, Kavukcuoglu K, Kuksa P. Natural language processing (almost) from scratch. *J Mach Learn Res.* 2011;12(Aug):2493–537.
- Lamurias A, Couto FM. Text mining for bioinformatics using biomedical literature. In: Ranganathan S, Gribskov M, Nakai K, Schönbach C, editors. *Encyclopedia of Bioinformatics and Computational Biology*. Oxford: Academic Press; 2019. p. 602–611. <https://doi.org/10.1016/B978-0-12-809633-8.20409-3>. <http://www.sciencedirect.com/science/article/pii/B9780128096338204093>.
- Erhan D, Bengio Y, Courville A, Manzagol P-A, Vincent P, Bengio S. Why does unsupervised pre-training help deep learning? *J Mach Learn Res.* 2010;11:625–60.
- Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. In: *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2. NIPS'13. USA: Curran Associates Inc.* 2013. p. 3111–9. <http://dl.acm.org/citation.cfm?id=2999792.2999959>.
- Pennington J, Socher R, Manning CD. Glove: Global vectors for word representation. In: *Empirical Methods in Natural Language Processing (EMNLP)*; 2014. p. 1532–43. <http://www.aclweb.org/anthology/D14-1162>.
- Huang CC, Lu Z. Community challenges in biomedical text mining over 10 years: Success, failure and the future. *Brief Bioinform.* 2016;17(1):132–44. <https://doi.org/doi:10.1093/bib/bbv024>.
- Al Hamid A, Ghaleb M, Aljadhey H, Aslanpour Z. A systematic review of hospitalization resulting from medicine-related problems in adult patients. *Br J Clin Pharmacol.* 2014;78(2):202–17. <https://doi.org/10.1111/bcp.12293>.
- Bouvy JC, De Bruin ML, Koopmanschap MA. Epidemiology of adverse drug reactions in europe: a review of recent observational studies. *Drug Saf.* 2015;38(5):437–53.
- Dou D, Wang H, Liu H. Semantic data mining: A survey of ontology-based approaches. In: *Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015)*; 2015. p. 244–51. <https://doi.org/10.1109/ICOSC.2015.7050814>.
- Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput.* 1997;9(8):1735–80.
- Miwa M, Bansal M. End-to-end Relation Extraction using LSTMs on Sequences and Tree Structures. In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Vol. 1 (Long Papers)*; 2016. p. 10. <https://doi.org/doi:10.18653/v1/P16-1105.1601.0770>.
- Hendrickx I, Kim SN, Kozareva Z, Nakov P, Ó Séaghdha D, Padó S, Pennacchiotti M, Romano L, Szpakowicz S. Semeval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals. In: *Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions*; 2009. p. 94–9. Association for Computational Linguistics.
- Zhang S, Zheng D, Hu X, Yang M. Bidirectional long short-term memory networks for relation classification. In: *Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation*; 2015. p. 73–8.
- Xu Y, Mou L, Li G, Chen Y. Classifying Relations via Long Short Term Memory Networks along Shortest Dependency Paths. In: *Proceedings of Conference on Empirical Methods in Natural Language Processing*; 2015. p. 1785–94. <https://doi.org/doi:10.18653/v1/D15-1206.1508.03720>.
- Zhao Z, Yang Z, Luo L, Lin H, Wang J. Drug drug interaction extraction from biomedical literature using syntax convolutional neural network. *Bioinformatics.* 2016;32(November):486. <https://doi.org/doi:10.1093/bioinformatics/btw486>.
- Sahu SK, Anand A. Drug-Drug Interaction Extraction from Biomedical Text Using Long Short Term Memory Network. *CEUR Work Proc.* 2017;1828:53–9. <https://doi.org/10.1145/2910896.2910898.1701.08303>.
- Wang W, Yang X, Yang C, Guo X, Zhang X, Wu C. Dependency-based long short term memory network for drug-drug interaction extraction. *BMC Bioinforma.* 2017;18(Suppl 16) <https://doi.org/10.1186/s12859-017-1962-8>.
- Xu B, Shi X, Zhao Z, Zheng W. Leveraging biomedical resources in bi- lstm for drug-drug interaction extraction. *IEEE Access.* 2018;6:33432–9. <https://doi.org/10.1109/ACCESS.2018.2845840>.
- Zheng W, Lin H, Luo L, Zhao Z, Li Z, Zhang Y, Yang Z, Wang J. An attention-based effective neural model for drug-drug interactions extraction; 2017, pp. 1–11. <https://doi.org/10.1186/s12859-017-1855-x>.

23. Couto FM, Lamurias A. Semantic similarity definition. In: Ranganathan S, Gribskov M, Nakai K, Schönbach C, editors. *Encyclopedia of Bioinformatics and Computational Biology*. Oxford: Academic Press; 2019. p. 870–6. <https://doi.org/10.1016/B978-0-12-809633-8.20401-9>. <http://www.sciencedirect.com/science/article/pii/B9780128096338204019>.
24. Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, Goldberg LJ, Eilbeck K, Ireland A, Mungall CJ, et al. The obo foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotechnol*. 2007;25(11):1251.
25. Hastings J, De Matos P, Dekker A, Ennis M, Harsha B, Kale N, Muthukrishnan V, Owen G, Turner S, Williams M, Steinbeck C. The ChEBI reference database and ontology for biologically relevant chemistry: Enhancements for 2013. *Nucleic Acids Res*. 2013;41(D1):456–63. <https://doi.org/10.1093/nar/gks1146>.
26. Ferreira JD, Inácio B, Salek RM, Couto FM. Assessing public metabolomics metadata, towards improving quality. *J Integr Bioinforma*. 2017;14(4).
27. Li Q, Li T, Chang B. Learning word sense embeddings from word sense definitions. In: Lin C-Y, Xue N, Zhao D, Huang X, Feng Y, editors. *Natural Language Understanding and Intelligent Applications*. Cham: Springer; 2016. p. 224–35.
28. Ma N, B H-tZ, Xiao X. An Ontology-Based Latent Semantic Indexing Approach Using Long Short-Term Memory Networks. *Web and Big Data*. 2017;1(0366(2)):185–99. <https://doi.org/10.1007/978-3-319-63579-8>.
29. Goyal P, Ferrara E. Graph embedding techniques, applications, and performance: A survey; 2017. arXiv preprint arXiv:1705.02801.
30. Kong X, Cao B, Yu PS. Multi-label classification by mining label and instance correlations from heterogeneous information networks. In: *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '13*. New York: ACM; 2013. p. 614–22. <https://doi.org/10.1145/2487575.2487577>. <http://doi.acm.org/10.1145/2487575.2487577>.
31. Dasigi P, Ammar W, Dyer C, Hovy E. Ontology-aware token embeddings for prepositional phrase attachment. Stroudsburg: Association for Computational Linguistics; 2017. pp. 2089–2098. <https://doi.org/doi:10.18653/v1/P17-1191>. <http://www.aclweb.org/anthology/P17-1191>.
32. Müller H-MM, Kenny EE, Sternberg PV. Textpresso: an ontology-based information retrieval and extraction system for biological literature. *PLoS Biol*. 2004;2(11):309. <https://doi.org/10.1371/journal.pbio.0020309>.
33. Lamurias A, Ferreira JD, Couto FM. Identifying interactions between chemical entities in biomedical text. *J Integr Bioinforma*. 2014;11(3):1–16.
34. Tripodi I, Boguslav M, Haylu N, Hunter LE. Knowledge-base-enriched relation extraction. In: *Proceedings of the Sixth BioCreative Challenge Evaluation Workshop*. Bethesda, MD USA, vol. 1; 2017. p. 163–6.
35. Ciaramita M, Altun Y. Broad-coverage sense disambiguation and information extraction with a supersense sequence tagger. In: *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*; 2006. p. 594–602. Association for Computational Linguistics.
36. Pyysalo S, Ginter F, Moen H, Salakoski T, Ananiadou S. *Distributional Semantics Resources for Biomedical Text Processing*; 2013.
37. Bhasuran B, Murugesan G, Abdulkadhar S, Natarajan J. Stacked ensemble combined with fuzzy matching for biomedical named entity recognition of diseases. *J Biomed Inform*. 2016;64:1–9. <https://doi.org/10.1016/j.jbi.2016.09.009>.
38. Chowdhury MFM, Lavelli A. FBK-irst: A multi-phase kernel based approach for drug-drug interaction detection and classification that exploits linguistic information. Atlanta, Georgia, USA. 2013;351:53.
39. Kim S, Liu H, Yeganova L, Wilbur WJ. Extracting drug–drug interactions from literature using a rich feature-based linear kernel approach. *J Biomed Inform*. 2015;55:23–30.
40. Köhler S, Vasilevsky NA, Engelstad M, Foster E, McMurry J, Aymé S, Baynam G, Bello SM, Boerkoel CF, Boycott KM, Brudno M, Buske OJ, Chinnery PF, Cipriani V, Connell LE, Dawkins HJS, DeMare LE, Devereau AD, de Vries BBA, Firth HV, Freson K, Greene D, Hamosh A, Helbig I, Hum C, Jähn JA, James R, Krause R, F Laulederkind SJ, Lochmüller H, Lyon GJ, Ogishima S, Olry A, Ouwehand WH, Pontikos N, Rath A, Schaefer F, Scott RH, Segal M, Sergouniotis PI, Sever R, Smith CL, Straub V, Thompson R, Turner C, Turro E, Veltman MWM, Vulliamy T, Yu J, von Ziegenweid J, Zankl A, Züchner S, Zemojtel T, Jacobsen JOB, Groza T, Smedley D, Mungall CJ, Haendel M, Robinson PN. The human phenotype ontology in 2017. *Nucleic Acids Res*. 2017;45(D1):865–76. <https://doi.org/doi:10.1093/nar/gkw1039>.
41. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. *The Gene Ontology Consortium*. *Nat Genet*. 2000;25(1):25–9.
42. authors listed N. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res*. 2017;45(D1):331–8.
43. Resnik P. Using information content to evaluate semantic similarity in a taxonomy. In: *International Joint Conference on Artificial Intelligence*, vol. 14. San Francisco: Citeseer; 1995. p. 448–53.
44. Jiang JJ, Conrath DW. Semantic similarity based on corpus statistics and lexical taxonomy. *CoRR* cmp-lg/9709008. 1997;19–33.
45. Lin D. An information-theoretic definition of similarity. In: *Proceedings of the Fifteenth International Conference on Machine Learning, ICML '98*. San Francisco: Morgan Kaufmann Publishers Inc.; 1998. p. 296–304. <http://dl.acm.org/citation.cfm?id=645527.657297>.
46. Kingma DP, Ba J. Adam: A method for stochastic optimization. *CoRR*. 2014;abs/1412.6980 <https://doi.org/1412.6980>.
47. Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov R. Improving neural networks by preventing co-adaptation of feature detectors. *CoRR*. 2012;abs/1207.0580 <https://doi.org/1207.0580>.
48. Zhang Y, Zheng W, Lin H, Wang J, Yang Z, Dumontier M. Drug-drug interaction extraction via hierarchical rnns on sequence and shortest dependency paths. *Bioinformatics*. 2018;34(5):828–35. <https://doi.org/10.1093/bioinformatics/btx659>, http://arxiv.org/abs/oupp/backfile/content_public/journal/bioinformatics/34/5/10.1093_bioinformatics_btx659/2/btx659.pdf.
49. Herrero-Zazo M, Segura-Bedmar I, Martínez P, Declerck T. The DDI corpus: An annotated corpus with pharmacological substances and drug-drug interactions. *J Biomed Inform*. 2013;46(5):914–20. <https://doi.org/10.1016/j.jbi.2013.07.011>.
50. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, Assempour N, Iynkkaran I, Liu Y, Maciejewski A, Gale N, Wilson A, Chin L, Cummings R, Le D, Pon A, Knox C, Wilson M. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res*. 2018;46(D1):1074–82.
51. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetverin V, Church DM, DiCuccio M, Edgar R, Federhen S, et al. Database resources of the national center for biotechnology information. *Nucleic Acids Res*. 2006;35(suppl_1):5–12.
52. Segura-Bedmar I, Martínez P, Herrero-Zazo M. Lessons learnt from the DDIExtraction-2013 Shared Task. *J Biomed Inform*. 2014;51(May):152–64. <https://doi.org/10.1016/j.jbi.2014.05.007>.
53. Segura-Bedmar I, Martínez P, Zazo MH. Semeval-2013 task 9: Extraction of drug-drug interactions from biomedical texts (ddiextraction 2013). In: *Second Joint Conference on Lexical and Computational Semantics (*SEM)*, Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), vol. 2; 2013. p. 341–50.
54. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. *Nat Genet*. 2000;25(1):25.
55. Dritsou V, Topalis P, Mitraka E, Dialynas E, Louis C. mirnao: An ontology unfolding the domain of microRNAs. In: *IWBIO*; 2014. p. 989–1000.
56. Langlotz CP. RadLex: a new method for indexing online educational materials. *Radiological Society of North America*; 2006.
57. Rath A, Olry A, Dhombres F, Brandt MM, Urbero B, Ayme S. Representation of rare diseases in health information systems: the orphanet approach to serve a wide range of end users. *Hum Mutat*. 2012;33(5):803–8.
58. Kibbe WA, Arze C, Felix V, Mitraka E, Bolton E, Fu G, Mungall CJ, Binder JX, Malone J, Vasant D, et al. Disease ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. *Nucleic Acids Res*. 2014;43(D1):1071–8.
59. Ou M, Cui P, Wang F, Wang J, Zhu W. Non-transitive hashing with latent similarity components. In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM; 2015. p. 895–904.
60. Lamurias A, Ferreira J, Couto F. Improving chemical entity recognition through h-index based semantic similarity. *J Cheminformatics*. 2015;7(Suppl 1):13–120. <https://doi.org/10.1186/1758-2946-7-S1-513>.