

# Biomedical Event Extraction Using Convolutional Neural Networks and Dependency Parsing

Jari Björne and Tapio Salakoski

Department of Future Technologies, University of Turku

Turku Centre for Computer Science (TUCS)

Faculty of Science and Engineering, FI-20014, Turku, Finland

firstname.lastname@utu.fi

## Abstract

Event and relation extraction are central tasks in biomedical text mining. Where relation extraction concerns the detection of semantic connections between pairs of entities, event extraction expands this concept with the addition of trigger words, multiple arguments and nested events, in order to more accurately model the diversity of natural language.

In this work we develop a convolutional neural network that can be used for both event and relation extraction. We use a linear representation of the input text, where information is encoded with various vector space embeddings. Most notably, we encode the parse graph into this linear space using dependency path embeddings.

We integrate our neural network into the open source Turku Event Extraction System (TEES) framework. Using this system, our machine learning model can be easily applied to a large set of corpora from e.g. the BioNLP, DDI Extraction and BioCreative shared tasks. We evaluate our system on 12 different event, relation and NER corpora, showing good generalizability to many tasks and achieving improved performance on several corpora.

## 1 Introduction

Detection of semantic relations is a central task in biomedical text mining where information is retrieved from massive document sets, such as scientific literature or patient records. This information often consists of statements of interactions between named entities, such as signaling pathways between proteins in cells, or the combinatorial effects of drugs administered to a patient. *Relation*

and *event* extraction are the primary methods for retrieving such information.

Relations are usually described as typed, sometimes directed, pairwise links between defined named entities. Automated relation extraction aims to develop computational methods for their detection.

Event extraction is a proposed alternative for relation extraction. Events differ from relations in that they can connect together more than two entities, that they have an annotated trigger word (usually a verb) and that events can act as arguments of other events. In the GENIA corpus, a sentence stating “The binding of proteins A and B is regulated by protein C” would be annotated with two nested events *REGULATION(C, BIND(A, B))*. While events can capture the semantics of text more accurately, their added complexity makes their extraction a more complicated task.

Many methods have been developed for relation extraction, with various kernel methods such as the graph kernel being widely used (Mooney and Bunescu, 2006; Giuliano et al., 2006; Airola et al., 2008). For the more complex task of event extraction approaches such as pipeline systems (Björne, 2014; Miwa et al., 2010), semantic parsing (McClosky et al., 2011) and joint inference (Riedel and McCallum, 2011) have been explored.

In recent years, the advent of deep learning has resulted in advances in many fields, and relation and event extraction are no exception. Considerable performance increases have been gained with methods such as convolutional (Zeng et al., 2014) and recurrent neural networks (Miwa and Bansal, 2016). Some proposed systems have relied entirely on word embeddings (Quan et al., 2016), while others have developed various network architectures for utilizing parse graphs as an additional source of information (Collobert et al., 2011; Liu et al., 2015; Xu et al., 2015; Ma et al.,

2015; Peng et al., 2017a).

In this work we present a new convolutional neural network method for extraction of both events and relations. We integrate our network as a classification module into the Turku Event Extraction System (Björne, 2014)<sup>1</sup>, allowing it to be easily applied to corpora or texts stored in the TEES XML format. Our neural network model is characterized by a unified representation of input examples that can be applied to detection of both keywords as well as their relations.

## 2 Materials and Methods

### 2.1 Corpora

We develop and evaluate our method on a large number of event and relation corpora (See Table 1). These corpora originate from three BioNLP Shared Tasks (Kim et al., 2009, 2011; Nédellec et al., 2013), the two Drug–Drug Interaction (DDI) Extraction tasks (Segura-Bedmar et al., 2011, 2013) and the recent BioCreative VI Chemical–Protein relation extraction task (Krallinger et al., 2017). The BioNLP corpora cover various domains of molecular biology and provide the most complex event annotations. The DDI and BioCreative corpora use pairwise relation annotations, and one of the DDI corpora defines also a drug named entity recognition (NER) task.

All of these corpora are used in the TEES XML format and are installed or generated with the TEES system. The corpora are parsed with the TEES preprocessing pipeline, which utilizes the BLLIP parser (Charniak and Johnson, 2005) with the McClosky biomodel (McClosky, 2010), followed by conversion of the constituency parses into dependency parses with the Stanford Tools (de Marneffe et al., 2006). These tools generate the deep parse graph which is used as the source for our dependency path features.

### 2.2 TEES Overview

The TEES system is based around a graph representation of events and relations. Named entities and event triggers are nodes, and relations and event arguments are the edges that connect them. An event is modelled as a trigger node and its set of outgoing edges. For a detailed overview of TEES we refer to Björne (2014).

TEES works as a pipeline method that models relation and event extraction as four consecutive

classification tasks (See Figure 2). The first stage is *entity detection* where each word token in a sentence is classified as an entity or a negative, generating the nodes of the graph. This stage is used in NER tasks as well as for event trigger word detection. The second stage is *edge detection* where relations and event arguments are predicted for all valid pairs of named entity and trigger nodes. For relation extraction tasks where entities are given as known data this is the only stage used.

In the entity detection stage TEES predicts a maximum of one entity per word token. However, since events are *n*-ary relations, event nodes may overlap. The *unmerging* stage duplicates event nodes by classifying each candidate event as a real event or not. Finally, *modifier detection* can be used to detect event modality (such as speculation or negation) on corpora where this is annotated.

### 2.3 Neural Network Overview

In TEES the four classification stages are implemented as multiclass classification tasks using the SVM<sup>multiclass</sup> support vector machine (Tsochantzidis et al., 2005) and a rich set of features derived mostly from the dependency parse graph.

We develop our convolutional neural network method using the Keras (Chollet et al., 2015) package with the TensorFlow backend (Dean et al., 2015). We extend the TEES system by replacing the SVM-based classifier modules with our network, using various vector space embeddings as input features. Our neural network design follows a common approach in NLP where the input sequence is processed by parallel convolutional layers (Kim, 2014; Zeng et al., 2014; Quan et al., 2016).

We use the same basic network structure for all four TEES classification stages (See Figure 1). The input examples are modelled in the context of a sentence window, centered around the candidate entity, relation or event. The sentence is modelled as a linear sequence of word tokens. Each word token is mapped to relevant vector space embeddings. These embeddings are concatenated together, resulting in an *n*-dimensional vector for each word token.

This merged input sequence is processed by a set of 1D convolutions with window sizes 1, 3, 5 and 7. Global max pooling is applied for each convolutional layer and the resulting features are merged together into the convolution output vec-

<sup>1</sup><http://jbjorne.github.io/TEES/>

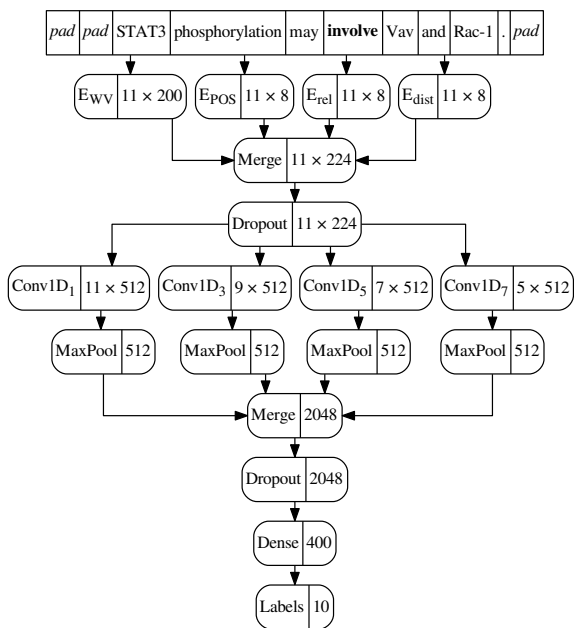


Figure 1: We use the same network architecture for all four classification tasks, with only the input embeddings and the number of predicted labels changing between tasks. This figure demonstrates *entity detection* where the word “involve” is being classified as belonging to one or more of 10 entity types. The input sentence is padded for the fixed example length (11 in this case) and the word tokens are mapped to corresponding embeddings, in this example Word Vectors, POS tags, relative positions and distances. The embedding vectors are merged together before the convolutional layers, and the results of the convolution operations are likewise merged before the dense layer, after which the final layer shows the predicted labels.

tor. This output vector is fed into a dense layer of 200–800 neurons, which is connected to the final classification layer where each label is represented by one neuron. The classification layer uses sigmoid activation, and the other layers use relu activation. Classification is performed as multilabel classification where each example may have 0–n positive labels. We use the *adam* optimizer with *binary crossentropy* and a learning rate of 0.001.

Dropout of 0.1–0.5 is applied at two stages in the system to increase generalization. Weights are learned for all input embeddings except for the word vectors, where we use the original weights as-is to ensure generalization to words outside the task’s training corpus.

## 2.4 Input Embeddings

All of the features used by our system are represented as embeddings, sets of vectors where each unique input item (such as a word string) maps to its own n-dimensional vector. The type and number of embeddings we use varies by classification task and is used to model the unique characteristics of each task (See Figure 2). The pre-made word vectors we use are 200-dimensional and the rest of the embeddings (learnt from the input corpus) are 8-dimensional.

**Word Vectors** are the most important of these embeddings. We use word2vec (Mikolov et al., 2013) vectors induced on a combination of the English Wikipedia and the millions of biomedical research articles from PubMed and PubMed Central by Pyysalo et al. (2013)<sup>2</sup>.

**POS** (Part-of-speech) tags generated by the BLLIP parser are used to define the syntactic categories of the words.

**Entity** features are used in cases where such information is already available, as in relation extraction where the pairs of entities are already known, or in event extraction where named entities are predefined.

**Distance** features follow the approach proposed by Zeng et al. (2014) where the relative distances to tokens of interest are mapped to their own vectors.

**Relative Position** features are used to mark whether tokens are located (B)efore, (A)fter or in the (M)iddle of the classified structure, or if they form a part of it as entities, event triggers or arguments. These features aim to model the context of the example in a manner somewhat similar to the shallow linguistic kernel of Giuliano et al. (2006).

**Path Embeddings** describe the shortest undirected path from a token of interest to another token in the sentence. Multiple sets of vectors (0–4), one for directed dependencies at each distance, are used for the dependencies of the path. For example, if paths of depth 4 are used, a shortest path of three directed dependencies connecting two tokens of interest could be modelled with four embedding vectors e.g.  $\leftarrow dobj, nsubj \rightarrow, nn \rightarrow, NULL$ . Our path embeddings are inspired by the concept of distance embeddings used by Zeng et al. (2014): Since it is possible to model linear distances between tokens in the input sentence, it

<sup>2</sup><http://evexdb.org/pmresources/vec-space-models/>

### 1. Entities

|       |         |                 |        |            |         |            |
|-------|---------|-----------------|--------|------------|---------|------------|
|       | Protein | Phosphorylation |        | Regulation | Protein | Protein    |
|       | STAT3   | phosphorylation | may    | involve    | Vav     | and Rac-1  |
|       | NN      | NN              | VB     | VB         | NN      | CC NN      |
|       |         | <nn             | <nsubj | <aux>      | <root>  | <conj_and> |
|       |         |                 |        | <dobj>     |         | <dobj>     |
| WV:   | STAT3   | phosphorylation | may    | involve    | Vav     | and Rac-1  |
| POS:  | NN      | NN              | VB     | VB         | NN      | CC NN      |
| dist: | -3      | -2              | -1     | 0          | 1       | 2 3 4      |
| rel:  | B       | B               | B      | E          | A       | A A A      |
| ent:  | Protein | -               | -      | -          | Protein | - Protein  |
| p1:   | nsubj>  | nsubj>          | aux>   | root       | dobj>   | - dobj>    |
| p2:   | nn>     | -               | -      | root       | -       | - -        |

### 2. Edges

|         |         |                 |        |            |         |            |
|---------|---------|-----------------|--------|------------|---------|------------|
|         | Protein | Phosphorylation |        | Regulation | Protein | Protein    |
|         | STAT3   | phosphorylation | may    | involve    | Vav     | and Rac-1  |
|         | NN      | NN              | VB     | VB         | NN      | CC NN      |
|         |         | <nn             | <nsubj | <aux>      | <root>  | <conj_and> |
|         |         |                 |        | <dobj>     |         | <dobj>     |
| WV:     | STAT3   | phosphorylation | may    | involve    | Vav     | and Rac-1  |
| POS:    | NN      | NN              | VB     | VB         | NN      | CC NN      |
| dist1:  | -3      | -2              | -1     | 0          | 1       | 2 3 4      |
| dist2:  | -1      | 0               | 1      | 2          | 3       | 4 5 6      |
| rel:    | B       | E2              | M      | E1         | A       | A A A      |
| ent:    | Protein | Phosphorylation | -      | Regulation | Protein | - Protein  |
| e1p1:   | nsubj>  | nsubj>          | aux>   | root       | dobj>   | - dobj>    |
| e1p2:   | nn>     | -               | -      | root       | -       | - -        |
| e2p1:   | nn>     | root            | <nsubj | <nsubj     | <nsubj  | <nsubj     |
| e2p2:   | -       | root            | aux>   | -          | dobj>   | - dobj>    |
| sp:     | o       | 1               | o      | 1          | 0       | 0 0 0      |
| sp:in:  | -       | nsubj>          | -      | [bgn]      | -       | - - -      |
| sp:out: | -       | [end]           | -      | nsubj>     | -       | - - -      |

### 3. Unmerging

|        |         |                 |        |            |         |            |
|--------|---------|-----------------|--------|------------|---------|------------|
|        | Protein | Phosphorylation |        | Regulation | Protein | Protein    |
|        | STAT3   | phosphorylation | may    | involve    | Vav     | and Rac-1  |
|        | NN      | NN              | VB     | VB         | NN      | CC NN      |
|        |         | <nn             | <nsubj | <aux>      | <root>  | <conj_and> |
|        |         |                 |        | <dobj>     |         | <dobj>     |
| WV:    | STAT3   | phosphorylation | may    | involve    | Vav     | and Rac-1  |
| POS:   | NN      | NN              | VB     | VB         | NN      | CC NN      |
| dist1: | -3      | -2              | -1     | 0          | 1       | 2 3 4      |
| dist2: | -1      | 0               | 1      | 2          | 3       | 4 5 6      |
| rel:   | B       | arg+            | M      | event      | arg+    | A arg- A   |
| ent:   | Protein | Phosphorylation | -      | Regulation | Protein | - Protein  |
| p1:    | nsubj>  | nsubj>          | aux>   | root       | dobj>   | - dobj>    |
| p2:    | nn>     | -               | -      | root       | -       | - -        |
| arg+:  | -       | Theme           | -      | -          | Cause   | - -        |
| arg-:  | -       | Theme           | -      | -          | -       | - Cause    |

### 4. Modifiers

|        |         |                 |        |            |         |            |
|--------|---------|-----------------|--------|------------|---------|------------|
|        | Protein | Phosphorylation |        | Regulation | Protein | Protein    |
|        | STAT3   | phosphorylation | may    | involve    | Vav     | and Rac-1  |
|        | NN      | NN              | VB     | VB         | NN      | CC NN      |
|        |         | <nn             | <nsubj | <aux>      | <root>  | <conj_and> |
|        |         |                 |        | <dobj>     |         | <dobj>     |
| WV:    | STAT3   | phosphorylation | may    | involve    | Vav     | and Rac-1  |
| POS:   | NN      | NN              | VB     | VB         | NN      | CC NN      |
| dist1: | -3      | -2              | -1     | 0          | 1       | 2 3 4      |
| rel:   | B       | B               | B      | E          | A       | A A A      |
| ent:   | Protein | Phosphorylation | -      | Regulation | Protein | - Protein  |
| p1:    | nsubj>  | nsubj>          | aux>   | root       | dobj>   | - dobj>    |
| p2:    | nn>     | -               | -      | root       | -       | - -        |

Figure 2: System stages. The TEES pipeline performs event extraction in four consecutive stages, generating first 1. the nodes (entities) and 2. edges (relations) of the event graph, which is then “pulled apart” in 3. unmerging, followed optionally by 4. modifier detection. The example being classified is shown with a dotted line in each image, and other examples in the same sentence with light gray dotted lines. We replace the four SVM classification stages in the TEES pipeline with convolutional neural networks. In place of the rich feature representations we use a sentence model where word token and dependency parse information is represented by embeddings. The Word Vector, POS and entities features are straightforwardly produced from the information of each token. The distance and relative position features model the position of the token in the sentence. The path features mark the dependencies connecting each token to a token of interest (candidate entity or relation endpoint). The shortest path features mark the set of dependencies forming the shortest path for a candidate relation. In the unmerging stage candidate event arguments are also used as features.

is also possible to model any other distance between these tokens, in our case paths in the dependency parse.

**Shortest Path Embeddings** follow the approach of the n-gram features used in methods such as the TEES SVM system. The shortest path consists of the tokens and dependencies connecting the two entities of a candidate relation. For each token on the path we define two embeddings, one for the incoming and one for the outgoing dependency. For example, if the shortest path would consist of three tokens and the two dependencies connecting them, the shortest path embedding vectors for the three tokens could be e.g. (*[begin]*,  $\leftarrow$ *nsubj*), ( $\leftarrow$ *nsubj*, *dobj* $\rightarrow$ ), (*dobj* $\rightarrow$ , *[end]*). Thus, our shortest path embeddings can be seen as a more detailed extrapolation of the “on dep path embeddings” of Fu et al. (2017).

**Event Argument Embeddings** are used only in the unmerging stage where predicted entities and edges are divided into separate events.

## 2.5 Parameter Optimization

When developing our system, we use the training set for learning, and the development set for parameter validation. We use the early stopping approach where the network is trained until the validation loss no longer decreases. We train for up to 500 epochs, stopping once validation loss has no longer decreased for 10 consecutive epochs.

Neural network models can be very sensitive to the initial random weights. Despite the relatively large training and validation sets, our model exhibits performance differences of several percentage points with different random seeds. In the current TensorFlow backend it is not possible to efficiently fix the random seed<sup>3</sup>, and in any case this would be unhelpful, as the impact of any given seed varies with the training data. Instead, we compensate for the inherent randomness of the network by training multiple models with randomized initializations and use as the final model the one which achieved the best performance on the validation set (measured using the micro-averaged F-score).

In addition to the random seed and optimal epoch, neural networks depend on a large number of hyperparameters. We use the process of training multiple randomized models also for param-

eter optimization. In addition to varying the random seed, we pick a random combination of hyperparameters from the ranges to be optimized, so that different models are randomized both in terms of initialization and the parameters. We test sizes of 200, 400 and 800 for the final dense layer, filter sizes of 128, 256 and 512 for the convolutional layers and dropout values of 0.1, 0.2 and 0.5. In addition, we experiment with path depths of 0–4 for the path embeddings.

Training a single model can still be prone to overfitting if the validation set is too small. To improve generalizability, we explore the use of model ensembles. Instead of using the best randomized model as the final one, we take n-best models, ranked with micro-averaged F-score on the validation set, and use their averaged predictions. These ensemble predictions are calculated for each label as the average of all the models’ predicted confidence scores.

With SVMs or random forests it is possible to “refit” a classifier after parameter selection, by retraining on the combined training and optimization sets, and this approach is also used by the TEES SVM classifiers. With the neural network, we cannot retrain with the validation set, as there would be no remaining data for detecting the optimal epoch. We approach also this issue using model ensembles. As the final source of randomization, we randomly redistribute the training and validation set documents before training each model. In this manner, the n-best models will together cover a larger part of the training data.

By training a large set of randomized models and using the n-best ones, we aim to address the effect of random initialization, parameter optimization and coverage of training data using the same process. However, with the size of the corpora used, training even a single model is relatively time consuming. In practice we are able to train only around 20 models for each of the four stages of the classification pipeline. Thorough parameter optimization comparable to the SVM system is thus not computationally feasible with the neural network, but good performance on varied corpora indicates that the current approach is at least adequate.

## 3 Results and Discussion

The results of applying our proposed system on the various corpora are shown in Table 2. We com-

<sup>3</sup><https://github.com/keras-team/keras/issues/2280>

| Corpus    | Domain                 | E  | I | S     |
|-----------|------------------------|----|---|-------|
| GE09      | Molecular Biology      | 10 | 6 | 11380 |
| GE11      | Molecular Biology      | 10 | 6 | 14958 |
| EPI11     | Epigenetics and PTM:s  | 16 | 6 | 11772 |
| ID11      | Infectious Diseases    | 11 | 7 | 5118  |
| REL11     | Entity Relations       | 1  | 2 | 11351 |
| DDI11     | Drug-Drug Interactions | -  | 1 | 5806  |
| DDI13 9.1 | Drug NER               | 4  | - | 9605  |
| DDI13 9.2 | Drug-Drug Interactions | -  | 4 | 10239 |
| GE13      | Molecular Biology      | 15 | 6 | 8369  |
| CG13      | Cancer Genetics        | 42 | 9 | 5938  |
| PC13      | Pathway Curation       | 24 | 9 | 5040  |
| CP17      | Chemical-Protein Int.  | -  | 5 | 6249  |

Table 1: The corpora used in this work are listed with their domain, number of event and entity types (E), number of event argument and relation types (I), and number of sentences (S).

pare our method with previous results from the shared tasks for which these corpora were introduced as well as with later research. In the next sections we analyse our results for the different corpus categories.

### 3.1 The BioNLP Event Extraction Tasks

The BioNLP Event Extraction tasks provide the most complex corpora with often large sets of event types and at times relatively small corpus sizes. The GENIA corpora from 2009 and 2011 have been the subject of most event extraction research. Our proposed method achieves F-scores of 57.84 and 58.10 on GE09 and GE11, respectively. Compared to the best reported results of 58.27 (Miwa et al., 2012) and 58.07 (Venugopal et al., 2014), our method shows similar performance on these corpora.

Our CNN reimplementation of TEES outperforms the original TEES SVM system on all the BioNLP corpora. In addition, we achieve to the best of our knowledge the highest reported performance on the GE11, EPI11, REL11, CG13 and PC13 BioNLP Shared Task corpora.

The annotations for the test sets of the BioNLP Shared Task corpora are not provided, instead the users upload their predictions to the task organizers’ servers for evaluation. While this method provides very good protection against overfitting and data leaks, unfortunately many of these evaluation servers are no longer working. Thus, we were able to evaluate our system on only a subset of all existing BioNLP Shared Task corpora.

### 3.2 The Drug-Drug Interactions Tasks

There have been two instances of the Drug-Drug Interactions Shared Task. The first one in 2011 concerned the detection of untyped relations for adverse drug effects. Unlike the other corpora, no official evaluator system or program exists for this corpus so we use our own F-score calculation. The lower performance compared to the original TEES system warrants further examination, but in any case the DDI11 corpus has been mostly superseded by the more detailed DDI13 corpora.

On the DDI13 corpora task 9.1, drug named entity recognition, our CNN system performs better than the original TEES entry, but neither of these TEES versions can detect more than single-token entities so they are not well suited for this task. Nevertheless, this result demonstrates the potential applicability of our method also to NER tasks.

Of all the DDI corpora the DDI13 task 9.2 corpus, typed relation extraction, has been the subject of much neural network based research in the past few years. A large variety of methods have been tested, and good results have been achieved by highly varying network models, some of which use no parsing or graph-like features, such as the multichannel convolutional neural network of Quan et al. (2016) which combines multiple sets of word vectors and achieves an F-score of 70.21. The highest result of 73.5 so far has been reported by Lim et al. (2018) who used a binary tree-LSTM model ensemble, with which our system achieves minutely higher, in practice comparable performance. Most recent DDI13 systems use corpus-specific rules for filtering negative candidate relations from the training data, which usually results in performance gains. As we aim to develop a generic method easily applicable to any corpus we did not implement these DDI filtering rules.

### 3.3 The CHEMPROT Task

Of all the evaluated corpora the CHEMPROT corpus used in the BioCreative VI Chemical-Protein relation extraction task is the most recent. Thus it provides an interesting point of comparison with current methods in relation extraction. All of our models outperform the task winning system combination of Peng et al. (2017b), with our mixed five model ensemble achieving a 5 pp increase over the shared task winning result. The CHEMPROT corpus is relatively large compared to its low number of five relation types, possibly making learning

| Corpus    | P     | R     | F                                | Team / System                                  | Method                   |
|-----------|-------|-------|----------------------------------|--|--------------------------|
| GE09      | 58.48 | 46.73 | 51.95 *                          | TEES (Björne et al., 2009)                     | SVM                      |
|           | -     | -     | 55.96                            | 6 x system ensemble (Kim et al., 2009)         | Metasystem               |
|           | 63.19 | 50.28 | 56.00                            | EventMine (Miwa et al., 2010)                  | SVM                      |
|           | -     | -     | 57.4                             | M3+enju (Riedel and McCallum, 2011)            | Joint inference          |
|           | 65.19 | 52.67 | <b>58.27</b>                     | EventMine+PR+FE (Miwa et al., 2012)            | SVM                      |
|           | 63.08 | 53.96 | 58.16                            | BioMLN (Venugopal et al., 2014)                | MLN and SVM              |
|           | 64.94 | 48.08 | 55.25                            | <i>Ours</i> (single)                           | CNN                      |
|           | 67.58 | 48.02 | 56.15                            | <i>Ours</i> (5 x ensemble)                     | CNN                      |
| 69.87     | 49.34 | 57.84 | <i>Ours</i> (mixed 5 x ensemble) | CNN  |                          |
| GE11      | 57.65 | 49.56 | 53.30 †                          | TEES (Björne and Salakoski, 2011)              | SVM                      |
|           | 64.75 | 49.41 | 56.04 *                          | FAUST (Riedel et al., 2011)                    | Joint inference / parser |
|           | 63.48 | 53.35 | 57.98                            | EventMine-CR (Miwa et al., 2012)               | SVM                      |
|           | 63.61 | 53.42 | 58.07                            | BioMLN (Venugopal et al., 2014)                | MLN and SVM              |
|           | 66.46 | 48.96 | 56.38                            | Stacked Generalization (Majumder et al., 2016) | SVM                      |
|           | 64.86 | 50.53 | 56.80                            | <i>Ours</i> (single)                           | CNN                      |
|           | 68.76 | 49.97 | 57.87                            | <i>Ours</i> (5 x ensemble)                     | CNN                      |
|           | 69.45 | 49.94 | <b>58.10</b>                     | <i>Ours</i> (mixed 5 x ensemble)               | CNN                      |
| EPI11     | 53.98 | 52.69 | 53.33 *                          | TEES (Björne and Salakoski, 2011)              | SVM                      |
|           | 54.42 | 54.28 | 54.35                            | EventMine multi-corpus (Miwa et al., 2013)     | SVM                      |
|           | 65.97 | 45.79 | 54.06                            | <i>Ours</i> (single)                           | CNN                      |
|           | 65.40 | 48.84 | 55.92                            | <i>Ours</i> (5 x ensemble)                     | CNN                      |
|           | 64.93 | 50.00 | <b>56.50</b>                     | <i>Ours</i> (mixed 5 x ensemble)               | CNN                      |
| ID11      | 48.62 | 37.85 | 42.57 †                          | TEES (Björne and Salakoski, 2011)              | SVM                      |
|           | 65.97 | 48.03 | 55.59 *                          | FAUST (Riedel et al., 2011)                    | Joint inference / parser |
|           | -     | -     | 55.6                             | M3+Stanford (Riedel and McCallum, 2011)        | Joint inference          |
|           | 61.33 | 58.96 | <b>60.12</b>                     | EventMine EasyAdapt (Miwa et al., 2013)        | SVM                      |
|           | 65.53 | 48.17 | 55.52                            | <i>Ours</i> (single)                           | CNN                      |
|           | 70.51 | 49.69 | 58.30                            | <i>Ours</i> (5 x ensemble)                     | CNN                      |
|           | 66.48 | 50.66 | 57.50                            | <i>Ours</i> (mixed 5 x ensemble)               | CNN                      |
| REL11     | 37.0  | 47.5  | 41.6 †                           | VIB - UGhent (Van Landeghem et al., 2011)      | SVM                      |
|           | 68.0  | 50.1  | 57.7 *                           | TEES (Björne and Salakoski, 2011)              | SVM                      |
|           | 70.87 | 59.56 | 64.72                            | <i>Ours</i> (single)                           | CNN                      |
|           | 76.30 | 48.09 | 58.99                            | <i>Ours</i> (5 x ensemble)                     | CNN                      |
|           | 73.65 | 61.17 | <b>66.83</b>                     | <i>Ours</i> (mixed 5 x ensemble)               | CNN                      |
| DDI11     | 58.04 | 68.87 | 62.99 †                          | TEES (Björne et al., 2011)                     | SVM                      |
|           | 60.54 | 71.92 | <b>65.74</b> *                   | WBI (Thomas et al., 2011)                      | kernels + CBR            |
|           | 69.83 | 55.49 | 61.84                            | <i>Ours</i> (single)                           | CNN                      |
|           | 69.78 | 57.21 | 62.88                            | <i>Ours</i> (5 x ensemble)                     | CNN                      |
|           | 77.57 | 49.93 | 60.75                            | <i>Ours</i> (mixed 5 x ensemble)               | CNN                      |
| DDI13 9.1 | 73.7  | 57.9  | 64.8 †                           | TEES (Björne et al., 2013)                     | SVM                      |
|           | 73.4  | 69.8  | <b>71.5</b> *                    | WBI-NER (Rocktäschel et al., 2013)             | CRF                      |
|           | 72    | 63    | 67                               | <i>Ours</i> (single)                           | CNN                      |
|           | 71    | 63    | 67                               | <i>Ours</i> (5 x ensemble)                     | CNN                      |
|           | 73    | 63    | 68                               | <i>Ours</i> (mixed 5 x ensemble)               | CNN                      |
| DDI13 9.2 | 73.2  | 49.9  | 59.4 †                           | TEES (Björne et al., 2013)                     | SVM                      |
|           | 64.6  | 65.6  | 65.1 *                           | FBK-irst (Chowdhury and Lavelli, 2013)         | kernels                  |
|           | 75.99 | 62.25 | 70.21                            | Multichannel CNN (Quan et al., 2016)           | CNN                      |
|           | 73.4  | 69.6  | 71.48                            | Joint AB-LSTM Model (Sahu and Anand, 2017)     | LSTM                     |
|           | 74.1  | 71.8  | 72.9                             | Hierarchical RNNs (Zhang et al., 2017)         | RNN                      |
|           | 77.8  | 69.6  | 73.5                             | One-Stage Model Ensemble (Lim et al., 2018)    | RNN                      |
|           | 75.80 | 70.38 | 72.99                            | PM-BLSTM (Zhou et al., 2018)                   | LSTM                     |
|           | 75.29 | 66.29 | 70.51                            | <i>Ours</i> (single)                           | CNN                      |
|           | 78.60 | 64.15 | 70.64                            | <i>Ours</i> (5 x ensemble)                     | CNN                      |
|           | 80.54 | 67.62 | <b>73.51</b>                     | <i>Ours</i> (mixed 5 x ensemble)               | CNN                      |
| GE13      | 56.32 | 46.17 | 50.74 †                          | TEES (Björne and Salakoski, 2013)              | SVM                      |
|           | 58.03 | 45.44 | 50.97 *                          | EVEX (Hakala et al., 2013)                     | TEES + rerank            |
|           | 59.24 | 48.95 | <b>53.61</b>                     | BioMLN (Venugopal et al., 2014)                | MLN and SVM              |
|           | 58.95 | 40.29 | 47.87                            | <i>Ours</i> (single)                           | CNN                      |
|           | 62.18 | 42.29 | 50.34                            | <i>Ours</i> (5 x ensemble)                     | CNN                      |
|           | 65.78 | 44.38 | 53.00                            | <i>Ours</i> (mixed 5 x ensemble)               | CNN                      |
| CG13      | 64.17 | 48.76 | 55.41 *                          | TEES (Björne and Salakoski, 2013)              | SVM                      |
|           | 55.82 | 48.83 | 52.09 †                          | EventMine (Miwa and Ananiadou, 2013)           | SVM                      |
|           | 60.45 | 51.34 | 55.52                            | <i>Ours</i> (single)                           | CNN                      |
|           | 63.92 | 51.00 | 56.74                            | <i>Ours</i> (5 x ensemble)                     | CNN                      |
|           | 66.55 | 50.77 | <b>57.60</b>                     | <i>Ours</i> (mixed 5 x ensemble)               | CNN                      |
| PC13      | 55.78 | 47.15 | 51.10 †                          | TEES (Björne and Salakoski, 2013)              | SVM                      |
|           | 53.48 | 52.23 | 52.84 *                          | EventMine (Miwa and Ananiadou, 2013)           | SVM                      |
|           | 58.31 | 47.08 | 52.10                            | <i>Ours</i> (single)                           | CNN                      |
|           | 58.66 | 48.49 | 53.09                            | <i>Ours</i> (5 x ensemble)                     | CNN                      |
|           | 62.16 | 50.34 | <b>55.62</b>                     | <i>Ours</i> (mixed 5 x ensemble)               | CNN                      |
| CP17      | 66.08 | 56.62 | 60.99 †                          | TEES (Mehryary et al., 2017)                   | SVM                      |
|           | 56.10 | 67.84 | 61.41 †                          | RSC (Corbett and Boyle, 2017)                  | LSTM                     |
|           | 72.66 | 57.35 | 64.10 *                          | NCBI (Peng et al., 2017b)                      | SVM, CNN, and RNN        |
|           | 71.40 | 61.86 | 66.28                            | <i>Ours</i> (single)                           | CNN                      |
|           | 74.38 | 60.44 | 66.69                            | <i>Ours</i> (5 x ensemble)                     | CNN                      |
|           | 75.13 | 65.07 | <b>69.74</b>                     | <i>Ours</i> (mixed 5 x ensemble)               | CNN                      |

Table 2: Results. Performance is shown in Precision, Recall and F-score, measured on the corpus test set for related work and our TEES CNN method (single best model, 5 model ensemble, or mixed 5 model ensemble with randomized train/validation set split). Shared task winning results are indicated with \* and shared task participant results with †. The highest F-score for each corpus is shown in bold. All of our results except for DDI11 are evaluated using the official evaluation program or server of each task.

| Corpus    | 0     | 1     | 2     | 3     | 4     |
|-----------|-------|-------|-------|-------|-------|
| GE11 (n)  | 69.85 | 71.45 | 72.34 | 72.87 | 72.64 |
| GE11 (e)  | 88.10 | 88.38 | 88.92 | 88.51 | 88.68 |
| DDI13 (e) | 59.46 | 59.23 | 58.22 | 56.50 | 54.37 |
| CP17 (e)  | 54.55 | 55.46 | 56.51 | 56.98 | 56.25 |

Table 3: The effect of path embeddings. The impact of using increasing depths of paths for embeddings is shown in terms of averaged F-score on the development set for entity (n) and edge (e) detection.

easier for our system.

### 3.4 Effect of Deep Parsing

Compared to neural networks which use either only word vectors, or model parse structure at the network level (e.g. graph-LSTMs), an interesting aspect of our method is that it can function both with and without parse information. By turning dependency path features on and off we can evaluate the impact of deep parsing on the system (See Table 3).

The path embeddings have the most impact on GE11 entity detection, where these paths link the entity candidate token to each other token. In GE11 event argument extraction the role of the path context embeddings is diminished. Surprisingly, on the DDI13 9.2 relation corpus path embeddings reduce performance, perhaps due to very long sentences and very indirect relations between the entity pairs. However, on another relation corpus, the CHEMPROT corpus, the path embeddings again increase performance, perhaps indicating that the CHEMPROT relation annotations follow more closely sentence syntax.

### 3.5 Computational Requirements

Our system improves on performance compared to the SVM-based TEES, but at the cost of increased computational requirements. The neural network effectively requires a specialized GPU for training and even then training times can be an issue.

For example, training the original TEES system on a four-core CPU for the GE09 task takes about 3 hours and classification of the test set with this model can be done in four minutes. For comparison, our GE09 neural network with 20 models for all four stages takes around nine hours to train on a Tesla P100 GPU. However, test set classification with a single model takes only about three minutes and using a five model ensemble about ten minutes.

Thus, while training the proposed method is much slower, classification can be performed relatively quickly. While the hardware and time requirements are much higher than with the SVM system, our proposed system can for some corpora achieve performance increases of even 10 pp. In most applications such gains are likely worth the increased computational requirements.

## 4 Conclusions

We have developed a convolutional neural network system that together with different vector space embeddings can be applied to diverse text classification tasks. We replace the TEES system’s event extraction pipeline components with this network and demonstrate considerable performance gains on a set of large event and relation corpora, achieving state-of-the-art performance on many of them and the best reported performance on the GE11, EPI11, REL11, CG13, PC13, DDI13 9.2 and CP17 corpora.

To the best of our knowledge our system represents the first application of neural networks to extraction of complex events from the BioNLP GENIA corpora. Our system uses a unified linear sentence representation where graph analyses such as dependency parses are fully included using our dependency path embeddings, and we demonstrate that these path embeddings can increase the performance of the convolutional model. Unlike systems where separate subnetworks are used to model graph structures, our network receives all of the information through the unified linear representation, allowing the whole model to learn from all the features.

The Turku Event Extraction System provides a unified approach for utilizing a large number of event and relation extraction corpora. As we integrate our proposed convolutional neural network method into the TEES system, it can be used as easily as the original TEES system, with the framework handling tasks such as preprocessing and format conversions. Our Keras-based neural network implementation can also be extended and modified, allowing continued experimentation on the wide set of corpora supported by TEES. We publish our method and our trained neural network models as part of the TEES open source project<sup>4</sup>.

<sup>4</sup><https://github.com/jbjorne/TEES/wiki/TEES-CNN-BioNLP18>



## Acknowledgments

We thank CSC – IT Center for Science Ltd for providing computational resources.

## References

- Antti Airola, Sampo Pyysalo, Jari Björne, Tapio Pahikkala, Filip Ginter, and Tapio Salakoski. 2008. All-paths graph kernel for protein-protein interaction extraction with evaluation of cross-corpus learning. *BMC bioinformatics*, 9(11):S2.
- Jari Björne. 2014. *Biomedical Event Extraction with Machine Learning*. Ph.D. thesis, University of Turku.
- Jari Björne, Antti Airola, Tapio Pahikkala, and Tapio Salakoski. 2011. Drug-drug interaction extraction from biomedical texts with svm and rls classifiers. *Proceedings of DDIEExtraction-2011 challenge task*, pages 35–42.
- Jari Björne, Juho Heimonen, Filip Ginter, Antti Airola, Tapio Pahikkala, and Tapio Salakoski. 2009. Extracting complex biological events with rich graph-based feature sets. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing: Shared Task*, pages 10–18. Association for Computational Linguistics.
- Jari Björne, Suwisa Kaewphan, and Tapio Salakoski. 2013. Uturku: drug named entity recognition and drug-drug interaction extraction using svm classification and domain knowledge. In *Second Joint Conference on Lexical and Computational Semantics (\*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, volume 2, pages 651–659.
- Jari Björne and Tapio Salakoski. 2011. Generalizing biomedical event extraction. In *Proceedings of the BioNLP Shared Task 2011 Workshop*, pages 183–191. Association for Computational Linguistics.
- Jari Björne and Tapio Salakoski. 2013. Tees 2.1: Automated annotation scheme learning in the bionlp 2013 shared task. In *Proceedings of the BioNLP Shared Task 2013 Workshop*, pages 16–25.
- Eugene Charniak and Mark Johnson. 2005. Coarse-to-fine n-best parsing and MaxEnt discriminative reranking. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL’05)*, pages 173–180. Association for Computational Linguistics.
- François Chollet et al. 2015. Keras. <https://github.com/keras-team/keras>.
- Md Faisal Mahbub Chowdhury and Alberto Lavelli. 2013. Fbk-irst: A multi-phase kernel based approach for drug-drug interaction detection and classification that exploits linguistic information. In *Second Joint Conference on Lexical and Computational Semantics (\*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, volume 2, pages 351–355.
- Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. 2011. Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, 12(Aug):2493–2537.
- P. Corbett and J. Boyle. 2017. Improving the learning of chemical-protein interactions from literature using transfer learning and word embeddings. In *Proceedings of the BioCreative VI Workshop*, pages 180–183.
- J Dean, R Monga, et al. 2015. Tensorflow: Large-scale machine learning on heterogeneous systems. *TensorFlow.org. Google Research*. Retrieved, 10.
- Lisheng Fu, Thien Huu Nguyen, Bonan Min, and Ralph Grishman. 2017. Domain adaptation for relation extraction with domain adversarial neural network. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, volume 2, pages 425–429.
- Claudio Giuliano, Alberto Lavelli, and Lorenza Romano. 2006. Exploiting shallow linguistic information for relation extraction from biomedical literature. In *11th Conference of the European Chapter of the Association for Computational Linguistics*.
- Kai Hakala, Sofie Van Landeghem, Tapio Salakoski, Yves Van de Peer, and Filip Ginter. 2013. **EVEX in ST’13: Application of a large-scale text mining resource to event extraction and network construction**. In *Proceedings of the BioNLP Shared Task 2013 Workshop*, pages 26–34, Sofia, Bulgaria. Association for Computational Linguistics.
- Jin-Dong Kim, Tomoko Ohta, Sampo Pyysalo, Yoshinobu Kano, and Jun’ichi Tsujii. 2009. **Overview of BioNLP’09 Shared Task on Event Extraction**. In *Proceedings of the BioNLP 2009 Workshop Companion Volume for Shared Task*, pages 1–9, Boulder, Colorado. Association for Computational Linguistics.
- Jin-Dong Kim, Sampo Pyysalo, Tomoko Ohta, Robert Bossy, Ngan Nguyen, and Jun’ichi Tsujii. 2011. **Overview of BioNLP Shared Task 2011**. In *Proceedings of BioNLP Shared Task 2011 Workshop*, pages 1–6, Portland, Oregon, USA. Association for Computational Linguistics.
- Yoon Kim. 2014. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.
- M. Krallinger, O. Rabal, S.A. Akhondi, M.P. Pérez, J. Santamaría, G.P. Rodríguez, G. Tsatsaronis, A. Intxaurrenondo, J.A. López, U. Nandal, E. Van Buel, A. Chandrasekhar, M. Rodenburg, A. Laegreid,

- M. Doornenbal, J. Oyarzabal, A. Lourenço, and A. Valencia. 2017. Overview of the biocreative vi chemical-protein interaction track. In *Proceedings of the BioCreative VI Workshop*, pages 141–146.
- Sangrak Lim, Kyubum Lee, and Jaewoo Kang. 2018. Drug drug interaction extraction from the literature using a recursive neural network. *PloS one*, 13(1):e0190926.
- Yang Liu, Furu Wei, Sujian Li, Heng Ji, Ming Zhou, and Houfeng Wang. 2015. A dependency-based neural network for relation classification. *arXiv preprint arXiv:1507.04646*.
- Mingbo Ma, Liang Huang, Bing Xiang, and Bowen Zhou. 2015. Dependency-based convolutional neural networks for sentence embedding. *arXiv preprint arXiv:1507.01839*.
- Amit Majumder, Asif Ekbal, and Sudip Kumar Naskar. 2016. Biomolecular event extraction using a stacked generalization based classifier. In *Proceedings of the 13th International Conference on Natural Language Processing*, pages 55–64.
- David McClosky. 2010. *Any domain parsing: automatic domain adaptation for natural language parsing*. Ph.D. thesis, Department of Computer Science, Brown University.
- David McClosky, Mihai Surdeanu, and Christopher D Manning. 2011. Event extraction as dependency parsing. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 1626–1635. Association for Computational Linguistics.
- Marie-Catherine de Marneffe, Bill MacCartney, and Christopher Manning. 2006. Generating typed dependency parses from phrase structure parses. In *Proceedings of LREC-06*, pages 449–454.
- Farrokh Mehryary, Jari Björne, Tapio Salakoski, and Filip Ginter. 2017. Combining support vector machines and LSTM networks for chemical-protein relation extraction. In *Proceedings of the BioCreative VI Workshop*, pages 175–179.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. [Distributed representations of words and phrases and their compositionality](#). In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems* 26, pages 3111–3119. Curran Associates, Inc.
- Makoto Miwa and Sophia Ananiadou. 2013. [NaCTeM EventMine for BioNLP 2013 CG and PC tasks](#). In *Proceedings of the BioNLP Shared Task 2013 Workshop*, pages 94–98, Sofia, Bulgaria. Association for Computational Linguistics.
- Makoto Miwa and Mohit Bansal. 2016. End-to-end relation extraction using lstms on sequences and tree structures. *arXiv preprint arXiv:1601.00770*.
- Makoto Miwa, Sampo Pyysalo, Tadayoshi Hara, and Jun’ichi Tsujii. 2010. [A comparative study of syntactic parsers for event extraction](#). In *Proceedings of the 2010 Workshop on Biomedical Natural Language Processing, BioNLP ’10*, pages 37–45, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Makoto Miwa, Sampo Pyysalo, Tomoko Ohta, and Sophia Ananiadou. 2013. Wide coverage biomedical event extraction using multiple partially overlapping corpora. *BMC bioinformatics*, 14(1):175.
- Makoto Miwa, Paul Thompson, and Sophia Ananiadou. 2012. Boosting automatic event extraction from the literature using domain adaptation and coreference resolution. *Bioinformatics*, 28(13):1759–1765.
- Raymond J Mooney and Razvan C Bunescu. 2006. Subsequence kernels for relation extraction. In *Advances in neural information processing systems*, pages 171–178.
- Claire Nédellec, Robert Bossy, Jin-Dong Kim, Jung-Jae Kim, Tomoko Ohta, Sampo Pyysalo, and Pierre Zweigenbaum. 2013. [Overview of BioNLP Shared Task 2013](#). In *Proceedings of the BioNLP Shared Task 2013 Workshop*, pages 1–7, Sofia, Bulgaria. Association for Computational Linguistics.
- Nanyun Peng, Hoifung Poon, Chris Quirk, Kristina Toutanova, and Wen-tau Yih. 2017a. Cross-sentence n-ary relation extraction with graph lstms. *arXiv preprint arXiv:1708.03743*.
- Yifan Peng, Anthony Rios, Ramakanth Kavuluru, and Zhiyong Lu. 2017b. Chemical-protein relation extraction with ensembles of svm, cnn, and rnn models. In *Proceedings of the BioCreative VI Workshop*, pages 147–150.
- S. Pyysalo, F. Ginter, H. Moen, T. Salakoski, and S. Ananiadou. 2013. [Distributional semantics resources for biomedical text processing](#). In *Proceedings of LBM 2013*, pages 39–44.
- Chanqin Quan, Lei Hua, Xiao Sun, and Wenjun Bai. 2016. Multichannel convolutional neural network for biological relation extraction. *BioMed research international*, 2016.
- Sebastian Riedel and Andrew McCallum. 2011. Fast and robust joint models for biomedical event extraction. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1–12. Association for Computational Linguistics.
- Sebastian Riedel, David McClosky, Mihai Surdeanu, Andrew McCallum, and Christopher D. Manning. 2011. [Model Combination for Event Extraction in BioNLP 2011](#). In *Proceedings of BioNLP Shared Task 2011 Workshop*, pages 51–55, Portland, Oregon, USA. Association for Computational Linguistics.

- Tim Rocktäschel, Torsten Huber, Michael Weidlich, and Ulf Leser. 2013. Wbi-ner: The impact of domain-specific features on the performance of identifying and classifying mentions of drugs. In *Second Joint Conference on Lexical and Computational Semantics (\*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, volume 2, pages 356–363.
- Sunil Kumar Sahu and Ashish Anand. 2017. Drug-drug interaction extraction from biomedical text using long short term memory network. *arXiv preprint arXiv:1701.08303*.
- I. Segura-Bedmar, P. Martínez, and D. Sánchez-Cisneros. 2011. The 1st DDIExtraction-2011 challenge task: Extraction of Drug-Drug Interactions from biomedical texts. In *Proceedings of the 1st Challenge Task on Drug-Drug Interaction Extraction 2011: 7 Sep 2011; Huelva, Spain*, pages 1–9.
- Isabel Segura-Bedmar, Paloma Martínez, and María Herrero Zazo. 2013. [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)*, pages 341–350, Atlanta, Georgia, USA. Association for Computational Linguistics.
- Philippe Thomas, Mariana Neves, Illés Solt, Domonkos Tikk, and Ulf Leser. 2011. [Relation Extraction for Drug-Drug Interactions using Ensemble Learning](#). In *Proc. of the 1st Challenge task on Drug-Drug Interaction Extraction (DDIExtraction 2011) at SEPLN 2011*, page 11–18, Huelva, Spain.
- Ioannis Tsochantaridis, Thorsten Joachims, Thomas Hofmann, and Yasemin Altun. 2005. Large margin methods for structured and interdependent output variables. *Journal of Machine Learning Research (JMLR)*, 6(Sep):1453–1484.
- Sofie Van Landeghem, Thomas Abeel, Bernard De Baets, and Yves Van de Peer. 2011. Detecting entity relations as a supporting task for biomolecular event extraction. In *Proceedings of the BioNLP Shared Task 2011 Workshop*, pages 147–148. Association for Computational Linguistics.
- Deepak Venugopal, Chen Chen, Vibhav Gogate, and Vincent Ng. 2014. Relieving the computational bottleneck: Joint inference for event extraction with high-dimensional features. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 831–843.
- Yan Xu, Lili Mou, Ge Li, Yunchuan Chen, Hao Peng, and Zhi Jin. 2015. Classifying relations via long short term memory networks along shortest dependency paths. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1785–1794.
- Daojian Zeng, Kang Liu, Siwei Lai, Guangyou Zhou, and Jun Zhao. 2014. Relation classification via convolutional deep neural network. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 2335–2344.
- Yijia Zhang, Wei Zheng, Hongfei Lin, Jian Wang, Zhihao Yang, and Michel Dumontier. 2017. Drug–drug interaction extraction via hierarchical rnns on sequence and shortest dependency paths. *Bioinformatics*.
- Deyu Zhou, Lei Miao, and Yulan He. 2018. Position-aware deep multi-task learning for drug–drug interaction extraction. *Artificial intelligence in medicine*.