Ontology-Aware Biomedical Relation Extraction

Ahmad Aghaebrahimian[0000−0003−0480−2505], Maria Anisimova[0000−0001−8145−7966], and Manuel Gil

Institute of Computational Life Sciences,
Department of Life Sciences and Facility Management,
Zurich University of Applied Sciences, 8820 Waedenswil, Switzerland
and Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland
{agha,anis,giln}@zhaw.ch

Abstract. Automatically extracting relationships from biomedical texts among multiple sorts of entities is an essential task in biomedical natural language processing with numerous applications, such as drug development or repurposing, precision medicine, and other biomedical tasks requiring knowledge discovery. Current Relation Extraction systems mostly use one set of features, either as text, or more recently, as graph structures. The state-of-the-art systems often use resource-intensive hence slow algorithms and largely work for a particular type of relationship. However, a simple yet agile system that learns from different sets of features has the advantage of adaptability over different relationship types without an extra burden required for system re-design.
We model RE as a classification task and propose a new multi-channel deep neural network designed to process textual and graph structures in separate input channels. We extend a Recurrent Neural Network with a Convolutional Neural Network to process three sets of features, namely, tokens, types, and graphs. We demonstrate that entity type and ontology graph structure provide better representations than simple token-based representations for Relation Extraction. We also experiment with various sources of knowledge, including data resources in the Unified Medical Language System to test our hypothesis. Extensive experiments on four well-studied biomedical benchmarks with different relationship types show that our system outperforms earlier ones. Thus, our system has state-of-the-art performance and allows processing millions of full-text scientific articles in a few days on one typical machine.

Keywords: Biomedical Relation Extraction · Graph Embedding · Deep Neural Network · Ontology · UMLS.

1 Introduction

The job of a biomedical Relation Extraction (RE) system is to identify semantic relationships among biomedical named entities such as genes, drugs, proteins, or chemical substances. There can be a large number of such relationships among different entities. Associations between genes and diseases, interactions among proteins and chemicals, or relationships among drugs and their side effects are a

few examples. RE plays an essential role in many biomedical applications such as clinical decision-making or information retrieval. Furthermore, RE is an integral component of Literature-Based Discovery (LBD) systems, commonly used to generate hypotheses for drug repurposing or drug discovery.

The advent of modern Machine Learning (ML) paradigms led to a significant boost in the performance of different RE systems, including Chemical-Protein Interactions (CPI) [19] or Chemical-Induced Diseases (CID) [14] to name a few. [27] use Support Vector Machines (SVMs) [3] for modeling Protein-Protein Interaction (PPI) and [14] use SVM and decision trees to model CID.

Deep Learning (DL) is the most recent and common class of ML techniques that attempted to address RE. Many studies on PPI extraction use variants of DL-based algorithms such as Recurrent Neural Network (RNN) [5]. [20, 9] employed DL to develop an end-to-end system for adverse drug event and drug-drug relationship detection. Using another DL-based algorithm named Convolutional Neural Network (CNN) [8], [12] proposed segment CNN for RE in clinical notes. [10] also made use of RNN to combine the feature vectors trained on MEDLINE with the semantic information obtained from external Knowledge Bases (KB) for relation and entity recognition.

Similar to our work, there are a few studies that attempted to integrate different neural architectures. The purpose is to benefit from the advantages and overcome the disadvantages of different shallow and deep algorithms. For instance, [28] combined RNN and CNN in a hybrid model or [15] combined RNN, CNN, and SVM as an ensemble system.

Contextualized language models help RE to obtain better results [11, 24]. However, they are considered highly resource-intensive algorithms. Dependence on massive machinery infrastructure usually raises concerns about scalability when considering large-scale RE. Aiming at developing a large-scale system, we avoid using any resource-intensive, hybrid, or ensemble system. Instead, we design a unified model that minimizes the load and complexity of the system via integrating ontology graph and typing information such that it can process millions of full-text articles in a reasonable time and on a sensible infrastructure.

We apply our method to four benchmarks with different biomedical relationship types and linguistic characteristics individually to ensure that our model handles agnostic datasets without requiring any particular tuning per dataset. These datasets include ChemProt [6], DDI [18], i2b2 [23], and AGAC [25]. Our method shows a substantial improvement (based on the F1 score) compared to the current SotA RE systems.

2 Methods

Instead of moving towards a more complex DL approach which is less effective [7], we use a simple architecture with several channels that allows us to integrate various sources of data into the training stream without over-complicating the problem.

Meantime to ensure optimum system throughput, and to benefit from graph-level and sentence-level information, we train an embedding space on a graph and integrate it into a sentence-level deep neural model. This way, we can enhance the system's performance while letting it process more than a thousand sentences a second. The required time would be higher by at least one order of magnitude if we would implement it in a graph neural network.

Three sets of features are integrated into our model, namely tokens, entity types, and graph structures extracted from ontologies in the form of graph embeddings. Assume the sentence $S = t_1, t_2, ..., t_n$ to consist of tokens t_i and to contain two named entities e_1 and e_2. We denote r as the relationship pointing to a pair of named entities e_1 and e_2.

In contrast to tokens which are merely occurrences of linguistic units (i.e., words, punctuation marks, symbols, etc.), named entities in life sciences are referred to well-recognized drugs, species, diseases, etc. They may consist of one or more consecutive tokens. Consider the following example:

> ... of the PDE inhibitors tested, dipyridamole was most effective, with IC50 values of 1.2 and 0.45 microM for inhibition of cAMP and cGMP hydrolysis, respectively.

The named entities are printed in red and blue. For the sake of brevity, we use entity to refer to a named entity from now on. In the ChemProt dataset, $CPR-9$ is the relationship between the two red entities. Note that there may be other relationships among the blue entities as well.

The task is then to find r such that

$$argmax_{r \in R} \quad p(r|S, e_i, e_j, T, G; \theta) \tag{1}$$

maximizes the probability of r where T is a set of associated entity types represented in t dimensional embedding space, and G is a graph consisting of all entities and relations available in the training data. Tokens in S, as well as the entities, are represented in d dimensional embedding space. G also is represented as g dimensional embeddings vectors. R is a set of relationships, and θ are the network parameters. We describe S, T, and G embeddings in more detail in sub-sections 2.1, 2.2, and 2.3 accordingly.

2.1 Token embedding

The most efficient way for representing tokens in almost all NLP tasks is via low-dimensional word vectors, also known as word embeddings. From a broad perspective, word embeddings can be of two types, namely static or dynamic. A static word embeddings algorithm (e.g., Word2Vec [13], Glove [16]) maps each token to a unique low-dimensional vector irrespective of the context where the token occurs. In contrast, a dynamic (i.e., contextual) word embeddings algorithm (e.g., ELMo [17], BERT [4]) maps each token to several different low-dimensional word vectors depending on their surrounding words. Due to the high

computational demand of the latter, we only use static embeddings to ensure a lean and scalable RE system. We use Word2Vec embeddings to represent S.

2.2 Type embedding

Typing information provides a mechanism for disambiguation when the system is not confident about the relationship between two entities. We integrate type embeddings into the system to examine their impact on the system performance.

In contrast to tokens, there are usually very few types available in a dataset. Consequently, a shallow embeddings technique known as the one-hot encoding (OHE) is sufficient for representing T.

2.3 Ontology graph embeddings

The idea in ontology graph embeddings is to map the graph of an ontology to low-dimensional vectors such that similar components in the graph are close to each other in the low-dimensional space. Therefore, in addition to isolated tokens represented via token embeddings, the network benefits from the information about the interaction of graph components and their neighbors. As the results show in Section 3, the embeddings of the ontology graph is a beneficial feature for RE. Graph structures provide three levels of features, namely node, link, and graph as a whole. We only estimate and use node-level embeddings to prove the concept and postpone the two other levels to further studies. To set up the input graph for embeddings generation, we construct a graph where entities (i.e., genes, diseases, drugs, etc.) are the vertices, and their relationships are the edges. Transforming this graph into a set of linear random walks (i.e., linearization) is the first step for embeddings generation. After setting the number and the length of random walks, we use a simple sampling agent to linearize the graph. The graph is a directed graph, hence backward moves are not possible. Therefore, at each vertex, the agent decides which outgoing edge to take using a uniform distribution.

Two hyper-parameters, namely the number and the length of random walks, control the agent's walking behavior. The model uses a portion of training data called the development data to tune these hyper-parameters. After transforming the graph into a set of random walks, we assume each walk as a sequence and use Word2Vec's Skip-gram algorithm to estimate the embeddings.

2.4 UMLS graph embeddings

The ontology graph provides a beneficial means of structured data for learning algorithms. However, for some datasets, the ontology graph is not available. A more robust way for generating ontology graph embeddings is to use external resources such as the Unified Medical Language System (UMLS) or Open Biomedical and Biological Ontology (OBO).

We consider the UMLS as an ontology of biomedical concepts. It consists of three main components, namely Metathesaurus, Semantic network, and Specialist lexicon. The Metathesaurus contains over four million biomedical concepts and their associated terms from over 200 source vocabularies. The Semantic network defines 133 broad types (e.g., disease, drug, disorder, etc.) and 54 relationships. It includes semantic types and semantic relationships such as "clinical drug A treats disease B or syndrome C". Finally, the Specialist lexicon provides lexical information for language processing.

Extracting the clusters of concepts from different vocabularies similar to the UMLS's Metathesaurus or extracting semantic typing information like the UMLS's Semantic network requires extensive querying among all available ontologies in the OBO Foundry. Given this constraint and for the sake of accessibility and reproducibility, in this study, we use UMLS and postpone OBO integration to further studies.

We extract the words and strings and their associations with their concepts from the UMLS 2021 package. Extracting the concepts, semantic types, and relationships, we construct a semantic graph. After the graph is constructed, a similar mechanism as described in the last subsection projects the concepts and relationships into an embedding space.

2.5 Architecture

Recent advances in DL have significantly enhanced RE. Here, we propose a new DL architecture to improve RE over biomedical data (see Figure 1 for the schema). This architecture complements an RNN with a CNN to extract two types of information that are deemed critical in RE.

On the one hand, Gated Recurrent Unit (GRU) [2] as an advanced variant of RNNs deals with strings with relatively long dependencies. GRUs in neural networks are often used in form of bidirectional units (i.e., BiGRU). Given a string, one GRU in a BiGRU unit extracts the textual features from right to left and the other from left to right and the resulting vectors are concatenated. CNN, on the other hand, is a great architecture for extracting keywords or key-phrases [8]. The combination of BiGRU and CNN assures that the model extracts the most informative sequential, local, and time-invariant features.

We hypothesize that combining GRU- and CNN-generated features provides RE with a more meaningful representation. Therefore, we propose a Bidirectional Gated Recurrent Unit-Convolutional Neural Network (BiGRU-CNN) multi-channel multi-input model for biomedical RE.

This architecture accepts a wide range of features. While tokens and their sequences are valuable features for RE, as we demonstrate via extensive experimentation (please refer to Section 6), entity types and ontology graph embeddings facilitate RE as well. Type information helps RE to disambiguate the detected relationships, while ontology embedding provides the model with implicit but beneficial information about entities and their connections in their ontology graph structure.

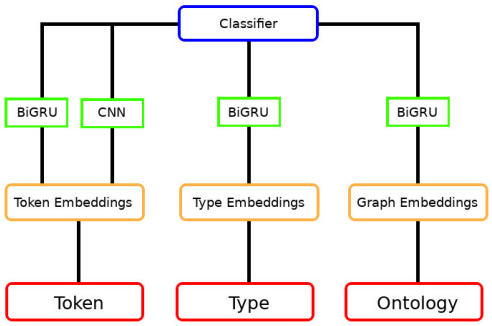

Fig. 1. Data-agnostic biomedical RE system architecture.

The first channel in Figure 1 is fed with the isolated token embeddings. While individual tokens provide strong signals for some relationships, the sequence of tokens known as n-grams allows better recognition of some other relationships. The combination of BiGRU and CNN ensures that both of these feature types are extracted. The model concatenates the resulting vectors of BiGRU and CNN to get the overall feature vector. The number of hidden layers for the BiGRU network, sequence length, CNN activation function, the dropout rate, and the optimizer are some of the hyperparameters for this channel.

More recent studies on RE use contextualized word embeddings. Computationally, such algorithms are highly demanding with hundreds of millions of parameters. Therefore, to estimate the S embeddings in the first channel, we use Word2Vec (Skip-gram) as a static word embeddings algorithm and train it on the PubMed abstracts released by BioASQ [22].

The second channel accepts the type embeddings, and the third channel receives the ontology graph embeddings. Sections 2.2, and 2.3 describe the procedure for estimating the embeddings representing T, and G required for these channels. The number and length of random walks for the ontology graph embeddings and the embeddings vector size are two other hyperparameters specific to these channels. Finally, the classifier on the top is a softmax function.

The hyperparameters in Table1 are reported to ensure reproducibility. All hyperparameters are optimized on the development set if available (the Chemprot dataset only), otherwise on randomly extracted 20% of the training set.

Table 1. System hyper-parameters

Hyper-parameter	Value	Hyper-parameter	Value
Emb. size d (tokens)	200	Optimizer	adam
Emb. g (Ontology)	128	hidden layers	64
Num. random walks	100	CNN filters	32
Length of walks	16	CNN kernel size	4
Drop-out	0.05	CNN activation	relu

3 Implementation and Results

A key motivation for our study is to enable to process millions of full-text articles while providing SotA accuracy. While many studies in RE focus on a particular dataset, we aim towards designing a dataset-agnostic system. To test the system, we selected four different benchmarks of relationship extraction tasks from various biomedical domains. They include the Active Gene Annotation Corpus (AGAC), the Informatics for Integrating Biology and the Bedside (i2b2), Drug-Drug Interaction (DDI), and CHEMical-PROTein interactions (ChemProt). This selection tries to reflect the thematic diversity, as well as the complexity of the task in terms of sequence length, number of classes, linguistic genre, and vocabulary. Training the models for different datasets takes from less than an hour to at most three hours on a standard machine with a Core-i7 CPU and 16 GB ram. Depending on the dataset and sequence length of the sentences, the models take a second to make inference over one thousand sentences with an average length of 70 to 120 tokens each. That makes relation extraction for the entire PubMed feasible in a few days and only using one typical machine. Tables 2, 3, 4, and 5 report the results of the system on AGAC, DDI, i2b2, and ChemProt datasets accordingly. The hyperparameters are tuned using the grid search strategy. The maximum length of all strings for each dataset is set as the length of the sequences for that dataset. If required, Micro F1, Macro F1, or both are reported to make comparison with earlier works possible.

Table 2. AGAC test results. The results of the current system are reported in the Micro F1 score with two significant figures. Samples without relationships are extracted as described in [21]

System	Without none relation			With none relation		
Relation/Score	P. (%)	R. (%)	F1 (%)	P. (%)	R. (%)	F1 (%)
No-Rel	-	-	-	95	93	94
COM	100	100	100	0	0	0
GOF	95	82	88	0.033	0.045	0.038
LOF	74	87	80	0.054	0.062	0.057
REG	100	25	40	0.031	0.042	0.035
Current system	**84**	**72**	**78**	**87**	**87**	**87**
[21]	-	-	-	86	86	86

The results in this section are reported based on the ontology graphs generated via the data-driven approach. Although the UMLS-based ontology graphs have a positive impact on the system performance, they yield inferior results compared to the data-driven approach. The distinction between the UMLS-based system and the data-driven approach is reported in the ablation study in Section 6. The reason for this inferiority comes from the fact that the coverage rate (i.e., the ratio of entities and relationships in a test set available in the relevant graph embeddings) of the data-driven approach is higher than the

Table 3. i2b2 test results. The results of the current system are reported with two significant figures due to the number of test samples. Similar to Table 10 in [26], a weighted F-Score is used to ensure a fair comparison. Since there are not enough training data in some classes in the i2b2 dataset, following [26], we did not use TrWP, TrIP, and TrNAP classes for training and development

System	[26]			Current system		
Relation/Score	P. (%)	R. (%)	F1 (%)	P. (%)	R. (%)	F1 (%)
TrCP	68	65	66	73	34	47
TrAP	79	82	81	86	94	90
TeRP	87	87	87	83	94	88
TeCP	63	63	63	64	46	54
PIP	73	67	70	100	100	100
Macro/Micro score	74/-	73/-	73/-	**81/89**	**74/89**	**76/89**

Table 4. DDI test results. The results of the current system are reported with three significant figures to account for the number of test instances. Similar to [1], the F1 score is Micro-averaged F1 score.

Relation/Score	P. (%)	R. (%)	F1 (%)
Advise	81.9	90.0	85.8
Effect	86.0	85.3	85.6
Int	94.4	35.4	51.5
Mechanism	89.9	91.7	90.8
Current system	**86.5**	**83.5**	**85.0**
[1]	85.36	82.83	84.08

Table 5. ChemProt results. The results of the current and SotA systems are reported in Macro/Micro F1 scores.

System	[19]	Current system		
Relation/Score	F1 (%)	P. (%)	R. (%)	F1 (%)
CPR:3	71.48	71.8	53.4	61.2
CPR:4	81.28	78.8	87.9	83.1
CPR:5	70.90	81.2	65.7	72.6
CPR:6	79.86	78.0	88.4	82.9
CPR:9	69.87	85.2	69.6	76.6
Macro/Micro score	-	**79/78.8**	**73/76.6**	**75.2/77.7**
Macro/Micro score	74.6/76.5	-	-	-

UMLS-based approach. Including other biomedical knowledge graphs leads to increasing the term coverage hence improving the performance. We postpone this integration to further studies.

3.1 Ablation

This section reports the impact of each layer and several design decisions on the system performance. We limit the parameters of this study to the BiGRU and CNN base models and the result of adding the type and ontology graph embeddings into the network. The ablation study is performed over all datasets to eradicate possible bias as much as possible.

Table 6. Ablation results; the impact of adding each network layer on the system performance. Statistically, significant changes are reported in bold. All scores are reported as Micro F1 score for the sake of consistency.

Config	Model-Dataset	AGAC(%)	DDI(%)	i2b2(%)	ChemProt(%)
1	Base CNN	71	76.1	80	70.9
2	Base GRU	72	77.2	81	72.8
3	1 + 2	73	78.6	82	74.1
4	3 + Type layer	75	81.4	85	75.4
5	4 + Ontology layer (UMLS)	77	83.2	88	77.4
6	4 + Ontology layer (data-driven)	**78**	**85**	**89**	**77.7**

The results in Table 6 show that the base BiGRU configuration consistently outperforms the CNN one, although the performance of the combined model is always higher than the sole BiGRU. It suggests that CNN captures some discriminative features which BiGRU encoders commonly lose. Our error analysis empirically shows that CNN does not work well for strictly directional relationships. For instance, CNN makes a lot of mistakes in recognizing CPR:5 and CPR:6 (Agonist and Antagonist relations) in the ChemProt dataset while it recognizes CPR:3 (Upregulator and activator) slightly better than BiGRU. The impact of type and ontology embeddings layers is also evident from the results.

4 Discussion

Biomedical relation extraction is a complex task. This complexity is partly due to the linguistic ambiguity and variability inherent in the biomedical entities. The difficulties involved in RE for different linguistic genres such as scientific papers (e.g., ChemProt) versus clinical texts (e.g., i2b2) add to this linguistic complexity. Another reason for such complexity is the wide range of ontologies in life sciences which lead to the definition of numerous relationships' types. Yet another source of complexity is added to RE because relationships are often directional connections between two entities. However, the text does not always preserve the order of the entities.

All studied datasets in this work are highly class-imbalanced that poses a significant issue in multi-class classification tasks. This includes an imbalance among classes as well as an imbalance between positive and negative instances of each class. Class imbalance usually works in favor of the majority class via disregarding the minority class at the training step. TeRP and TrAP in the i2b2 dataset are two evident examples of errors caused by class imbalance. TrCP and TeCP are the worst-performing classes in this dataset; TeCP is often misclassified with TeRP and TrCP is often misclassified with TrAP. In both cases, the class to which the true classes are wrongly assigned belongs to the majority classes.

Another reason for making errors in classification is that in both cases the misclassified classes are semantically similar to true classes; In the first case "Test Conducted to investigate Problem (TeCP)" and "Test Reveal Problem (TeRP)" and in the second case "Treatment Cause problems (TrCP)" and "Treatment Administered Problem (TrAP)" are considerably similar. Our experiments on various embeddings show that an embedding trained on biomedical data yields fewer misclassified instances of this type.

The worst-performing class in the DDI dataset is also the minority class *Int* which often is overshadowed by *Effect*. One reason for this is that *Int* is the super-class denoting any interaction which conveys the same semantics as *Effect* may do.

5 Conclusion

Relation Extraction is a fundamental task in biomedical text analytics. There is a wide range of domains within biomedical and health sciences. Therefore a universal model capable of extracting relationships across various biomedical subdomains is highly desirable since it reduces the time and effort required to design domain-specific architectures. Employing graph ontology and biomedical types represented as embeddings, we designed a deep neural network for relation extraction adaptable to various domains given the ontology and type information encoded as embeddings layers. The network takes this information directly from the datasets in a data-driven approach or indirectly from the UMLS as an external resource. Our system obtains state-of-the-art results on four datasets from different biomedical sub-domains, namely; Chemical Protein Interactions (CPI), Drug-Drug Interactions (DDI), Gene functions, and clinical problems and tests. Due to its uncomplicated yet quick encoders and classifier, it makes relation extraction feasible on a large volume of textual data and within a limited time.

Funding

This work was funded by the ZHAW Health@N initiative (grant 9710.3.01.5.0001.08 to M.G.).

References

1. Asada, M., Miwa, M., Sasaki, Y.: Using drug descriptions and molecular structures for drug–drug interaction extraction from literature. Bioinformatics **37**(12) (2020)
2. Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using rnn encoder-decoder for statistical machine translation (2014)
3. Cortes, C., Vapnik, V.: Support-vector networks. In: Machine Learning. pp. 273–297 (1995)
4. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. In: NAACL-HLT (2019)
5. Elman, J.L.: Finding structure in time. Cognitive science **14**(2), 179–211 (1990)
6. Hirschman, L., Yeh, A., Blaschke, C., Valencia, A.: Overview of biocreative: Critical assessment of information extraction for biology. BMC bioinformatics **6 Suppl 1** (2005)
7. Huynh, T., He, Y., Willis, A., Rueger, S.: Adverse drug reaction classification with deep neural networks. In: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. pp. 877–887. The COLING 2016 Organizing Committee, Osaka, Japan (Dec 2016)
8. Lecun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE **86**(11), 2278–2324 (1998). https://doi.org/10.1109/5.726791
9. Li, F., Liu, W., Yu, H.: Extraction of information related to adverse drug events from electronic health record notes: Design of an end-to-end model based on deep learning. JMIR Medical Informatics **6** (2018)
10. Li, Z., Lian, Y., Ma, X., Zhang, X., Li, C.: Bio-semantic relation extraction with attention-based external knowledge reinforcement. BMC Bioinformatics **21** (2020)
11. Lin, Y., Shen, S., Liu, Z., Luan, H., Sun, M.: Neural relation extraction with selective attention over instances. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 2124–2133. Association for Computational Linguistics, Berlin, Germany (Aug 2016)
12. Luo, Y., Cheng, Y., Uzuner, O., Szolovits, P., Starren, J.: Segment convolutional neural networks (Seg-CNNs) for classifying relations in clinical notes. Journal of the American Medical Informatics Association **25**(1), 93–98 (08 2017). https://doi.org/10.1093/jamia/ocx090
13. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space (2013)
14. Onye, S.C., Akkeleş, A., Dimililer, N.: relscan – a system for extracting chemical-induced disease relation from biomedical literature. Journal of Biomedical Informatics **87**, 79–87 (2018)
15. Peng, Y., Rios, A., Kavuluru, R., lu, Z.: Extracting chemical-protein relations with ensembles of svm and deep learning models. Database : the journal of biological databases and curation **2018** (07 2018)
16. Pennington, J., Socher, R., Manning, C.: GloVe: Global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 1532–1543. Association for Computational Linguistics, Doha, Qatar (Oct 2014)
17. Peters, M., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L.: Deep contextualized word representations. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Lin-

guistics: Human Language Technologies, Volume 1 (Long Papers). pp. 2227–2237. Association for Computational Linguistics, New Orleans, Louisiana (Jun 2018)
18. Segura-Bedmar, I., Martinez, P., Sanchez-Cisneros, D.: The 1st ddiextraction-2011 challenge task: Extraction of drug-drug interactions from biomedical texts. In: Challenge Task Drug-Drug Interact Extr. vol. 2011, pp. 1–9 (01 2011)
19. Sun, C., Yang, Z., Su, L., Wang, L., Zhang, Y., Lin, H., Wang, J.: Chemical-protein interaction extraction via gaussian probability distribution and external biomedical knowledge. Bioinformatics (2020)
20. Sänger, M., Leser, U.: Large-scale entity representation learning for biomedical relationship extraction. Bioinformatics **37**(2), 236–242 (07 2020)
21. Thillaisundaram, A., Togia, T.: Biomedical relation extraction with pre-trained language representations and minimal task-specific architecture. CoRR **abs/1909.12411** (2019)
22. Tsatsaronis, G., Balikas, G., Malakasiotis, P., Partalas, I., Zschunke, M., Alvers, M.R., Weissenborn, D., Krithara, A., Petridis, S., Polychronopoulos, D., Almirantis, Y., Pavlopoulos, J., Baskiotis, N., Gallinari, P., Artiéres, T., Ngomo, A.C.N., Heino, N., Gaussier, E., Barrio-Alvers, L., Schroeder, M., Androutsopoulos, I., Paliouras, G.: An overview of the bioasq large-scale biomedical semantic indexing and question answering competition. BMC Bioinformatics (2015)
23. Uzuner, O., South, B., Shen, S., DuVall, S.: 2010 i2b2/va challenge on concepts, assertions, and relations in clinical text. Journal of the American Medical Informatics Association : JAMIA **18**, 552–6 (06 2011)
24. Verga, P., Strubell, E., McCallum, A.: Simultaneously self-attending to all mentions for full-abstract biological relation extraction. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). pp. 872–884. Association for Computational Linguistics, New Orleans, Louisiana (Jun 2018)
25. Wang, Y., Zhou, K., Gachloo, M., Xia, J.: An overview of the active gene annotation corpus and the BioNLP OST 2019 AGAC track tasks. In: Proceedings of The 5th Workshop on BioNLP Open Shared Tasks. pp. 62–71. Association for Computational Linguistics, Hong Kong, China (Nov 2019)
26. Yadav, S., Ramesh, S., Saha, S., Ekbal, A.: Relation extraction from biomedical and clinical text: Unified multitask learning framework. IEEE/ACM transactions on computational biology and bioinformatics **PP** (2020)
27. Yan, C., Dobbs, D., Honavar, V.: Identification of surface residues involved in protein-protein interaction — a support vector machine approach. In: Abraham, A., Franke, K., Köppen, M. (eds.) Intelligent Systems Design and Applications. pp. 53–62. Springer Berlin Heidelberg, Berlin, Heidelberg (2003)
28. Zhang, Y., Lin, H., Yang, Z., Wang, J., Zhang, S., Sun, Y., Yang, L.: A hybrid model based on neural networks for biomedical relation extraction. Journal of Biomedical Informatics **81**, 83–92 (2018)