Using Neural Networks for Relation Extraction from Biomedical Literature

Diana Sousa⋆, Andre Lamurias, and Francisco M. Couto

LASIGE, Faculdade de Ciências, Universidade de Lisboa, Portugal

Abstract. Using different sources of information to support automated extracting of relations between biomedical concepts contributes to the development of our understanding of biological systems. The primary comprehensive source of these relations is biomedical literature. Several relation extraction approaches have been proposed to identify relations between concepts in biomedical literature, namely, using neural networks algorithms. The use of multichannel architectures composed of multiple data representations, as in deep neural networks, is leading to state-of-the-art results. The right combination of data representations can eventually lead us to even higher evaluation scores in relation extraction tasks. Thus, biomedical ontologies play a fundamental role by providing semantic and ancestry information about an entity. The incorporation of biomedical ontologies has already been proved to enhance previous state-of-the-art results.

Keywords: Relation Extraction, Biomedical Literature, Neural Networks, Deep Learning, Ontologies, External Sources of Knowledge

1 Introduction

Biomedical literature is the main medium that researchers use to share their findings, mainly in the form of articles, patents, and other types of written reports [1]. A researcher working on a specific topic needs to be up-to-date with all developments regarding the work done on the same topic. However, the volume of textual information available widely surpasses the ability of analysis by a researcher, even if restricting it to a domain-specific topic. Not only that, but the textual information available is usually in an unstructured or highly heterogeneous format. Thus, retrieving relevant information requires not only a considerable amount of manual effort but is also a time-consuming task.

Scientific articles are the primary source of knowledge for biomedical entities and their relations. These entities include human phenotypes, genes, proteins, chemicals, diseases, and other biomedical entities inserted in specific domains. A comprehensive source for articles on this topic is the PubMed platform [2], combining over 29 million citations while providing access to their metadata. Processing this volume of information is only feasible by using text mining solutions.

⋆ dfsousa@lasige.di.fc.ul.pt

Automatic methods for information extraction (IE) aim at obtaining useful information from large data-sets [3]. Text mining uses IE methods to process text documents. Text mining systems usually include named-entity recognition (NER), named-entity linking (NEL), and relation extraction (RE) tasks. NER consists of recognizing entities mentioned in the text by identifying the offset of its first and last character. NEL consists of mapping the recognized entities to entries in a given knowledge base. RE consists of identifying relations between the entities mentioned in a given document.

RE can be performed by different methods, namely, by order of complexity, co-occurrence, pattern-based (manually and automatically created), rule-based (manually and automatically created), and machine learning (feature-based, kernel-based, and recurrent neural networks (RNN)).

In recent years, deep learning techniques, such as RNN, have proved to achieve outstanding results at various natural language processing (NLP) tasks, among them RE. The success of deep learning for biomedical NLP is in part due to the development of word vector models like Word2Vec [4], and, more recently, ELMo [5], BERT [6], GPT [7], Transformer-XL [8], and GPT-2 [9]. These models learn word vector representations that capture the syntactic and semantic word relationships and are known as word embeddings. Long Short-Term Memory (LSTM) RNN constitute a variant of artificial neural networks presented as an alternative to regular RNN [10]. LSTM networks deal with more complex sentences, making them more fitting for biomedical literature. To improve their results in a given domain, it is possible to integrate external sources of knowledge such as domain-specific ontologies.

The knowledge encoded in the various domain-specific ontologies, such as the Gene Ontology (GO) [11], the Chemical Entities of Biological Interest (ChEBI) ontology [12], and the Human Phenotype Ontology (HPO) [13] is deeply valuable for detection and classification of relations between different biomedical entities. Besides that, these ontologies make available important characteristics about each entity; they also provide us with the underlying semantics of the relations between those entities, such as is-a relations. For example, *neoplasm of the endocrine system* (HP:0100568), a phenotypic abnormality that describes a tumor (abnormal growth of tissue) of the endocrine system **is-a** *abnormality of the endocrine system* (HP:0000818), and **is-a** *neoplasm by anatomical site* (HP:0011793), which in turn **is-a** *neoplasm* (HP:0002664) (Fig. 1).

The information provided by the ancestors is not expressed directly in the text and can support or disprove an identified relation. Ontologies are formally organized in machinereadable formats, facilitating their integration in relation extraction models.

Using different sources of information, as additional data, to support automating searching for relations between biomedical concepts contributes to the development of pharmacogenomics, clinical trial screening, and adverse drug reaction identification [14]. Identifying new relations can help validate the results of recent research, and even propose new experimental hypotheses.
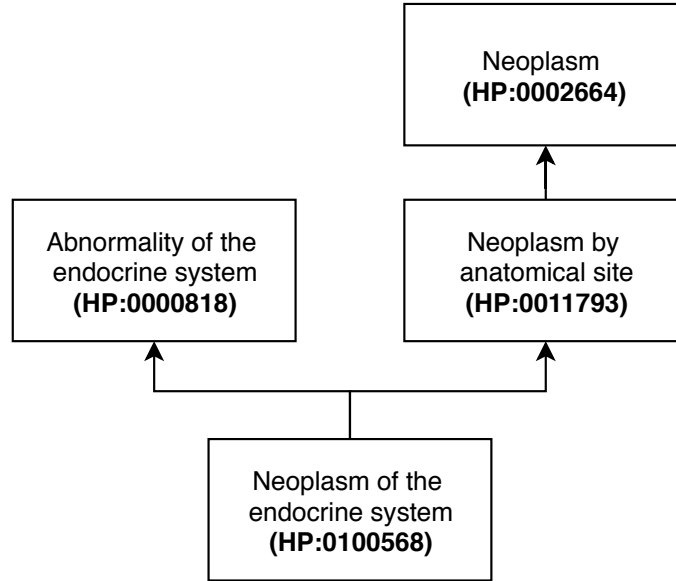
Fig. 1. An excerpt of the HPO ontology showing the first ancestors of *neoplasm of the endocrine system*, using **is-a** relationships

2   Related Work

This section presents the basic concepts and resources that support Relation Extraction (RE) deep learning techniques, namely, Natural Language Processing (NLP), Text Mining Primary Tasks, and initial approaches for Relation Extraction.

2.1   Natural Language Processing

Natural language processing (NLP) is an area in computer science that aims to derive meaning from unstructured or highly heterogeneous text written by humans. NLP covers several techniques that constitute preprocessing steps for the tasks described in Subheading 2.2. These NLP techniques have different goals and are often combined to obtain higher performance.

 – **Tokenization**: has the purpose of breaking the text into tokens to be processed individually or as a sequence. These tokens are usually words but can also be phrases, numbers and other types of elements. The most straightforward form of tokenization is breaking the input text by the use of whitespace or punctuation. However, with scientific biomedical literature, that is usually descriptive and formal, we have to account for complex entities like human phenotype terms (composed of multiple words), genes (represented by symbols), and other types of structured entities. These entities tend to be

morphologically complex and need specialized tokenization pipelines. Some researchers use a compression algorithm [15], byte pair encoding (BPE), to account for biomedical vocabulary variability. BPE represents open vocabularies through a fixed-size vocabulary of variablelength character sequences, making it suitable for neural networks models.

– **Stemming and Lemmatization**: aims at reducing the variability of natural language by normalizing a token to its base form (stem) [16]. It can also take into account the context of the token, along with vocabulary and morphological analysis to determine the canonical form of the word (lemma). The stem can correspond only to a fragment of a word, but the lemma is always a real word. For instance, the stem of the word *having* is *hav* and the lemma is *have*.

– **Part-of-Speech Tagging**: consists of assigning each word of a sentence to the category where it belongs taking into account their context (e.g., verb or preposition). Each word can belong to more than one category. This feature is useful to gain information on the role of a word in a given sentence.

– **Parse Tree**: represents the syntactic structure of a sentence. There are two different types of parse trees: constituency-based parse trees and dependency-based parse trees. The main difference between the two is that the first distinguishes between the terminal and non-terminal nodes and the second does not (all nodes are terminal). In constituency-based parse trees, each node of the tree is either a *root* node, a *branch* node, or a *leaf* node. For each given sentence there is only one *root* node. The *branch* node connects to two or more *child* nodes, and the *leaf* node is terminal. These leaves correspond to the lexical tokens [17]. Dependency-based parse trees are usually simpler because they only identify the primary syntactic structure, leading to fewer nodes. Parse trees generate structures that are used as inputs for other algorithms and can be constructed based on supervised learning techniques.

2.2   Text Mining Primary Tasks

Text mining has become a widespread approach to identify and extract information from unstructured or highly heterogeneous text [18]. Text mining is used to extract facts and relationships in a structured form that can be used to annotate specialized databases and to transfer knowledge between domains [19]. We may consider text mining as a sub-field of data mining. Thus, data mining algorithms can be applied if we transform text to a proper data representation, namely, numeric vectors. Even if in recent years text mining tools have evolved considerably in number and quality, there are still many challenges in applying text mining to scientific biomedical literature. The main challenges are the complexity and heterogeneity of the written resources, which make the retrieval of relevant information, that is, relations between entities, a nontrivial task.

Text Mining tools can target different tasks together or separately. Some of the primary tasks are Named Entity Recognition (NER), Named-Entity Linking (NEL) and Relation Extraction (RE).

- **Named Entity Recognition (NER)**: seeks to recognize and classify entities mentioned in the text by identifying the offset of its first and last character. The workflow of this task starts by splitting the text in tokens and then labeling them into categories (part-of-speech (POS) tagging).
- **Named-Entity Linking (NEL)**: maps the recognized entities to entries in a given knowledge base. For instance, a gene can be written in multiple ways and mentioned by different names or acronyms in a text. NEL links all these different nomenclatures to one unique identifier. There are several organizations dedicated to providing identifiers, among them the National Center for Biotechnology Information (NCBI) for genes, and the Human Phenotype Ontology (HPO) for phenotypic abnormalities encountered in human diseases.
- **Relation Extraction (RE)**: identifies relations between entities (recognized manually or by NER) in a text. Tools mainly consider relations by the co-occurrence of the entities in the same sentence, but some progress is being made to extend this task to the full document (taking into account a global context) [20].

The workflow of a typical RE system is presented in Fig. 2.

2.3   Initial Approaches for Relation Extraction

Through the years, several approaches have been proposed to extract relations from biomedical literature [22]. Most of these approaches work on a sentence level to perform RE, due to the inherent complexity of biomedical literature.

- **Co-occurrence**: assumes that if two entities are mentioned in the same sentence (co-occur), it is likely that they are related. Usually, the application of this approach results in a higher recall (most of the entities co-occurring in a sentence participate in a relation), and lower precision. Some methods use frequency-based scoring schemes to eliminate relations identified by chance [23]. Nowadays, most applications use co-occurrence as a baseline against more complex approaches [24].
- **Pattern-based**: uses manually defined patterns and automatically generated patterns to extract relations. **Manually defined patterns** require domain expertise knowledge about the type of biomedical entities, their interactions, and the text subject at hand. Initial systems made use of regular expressions to match word patterns that reflected a relation between two entities [25], making use of a dictionary of words that express relation, such as *trigger* and *stimulate*. Later systems introduce part-of-speech when applied to complex sentences, such as the ones that we typically find in biomedical literature [26]. Opposite to the co-occurrence approaches, manually defined patterns frequently achieve high precision but tend to have poor recall. This approach does not generalize well, and therefore is difficult to apply to new unseen data. **Automatically generated patterns** encompass two main approaches, bootstrapping with *seeds* [27] and leveraging of the corpora [28].
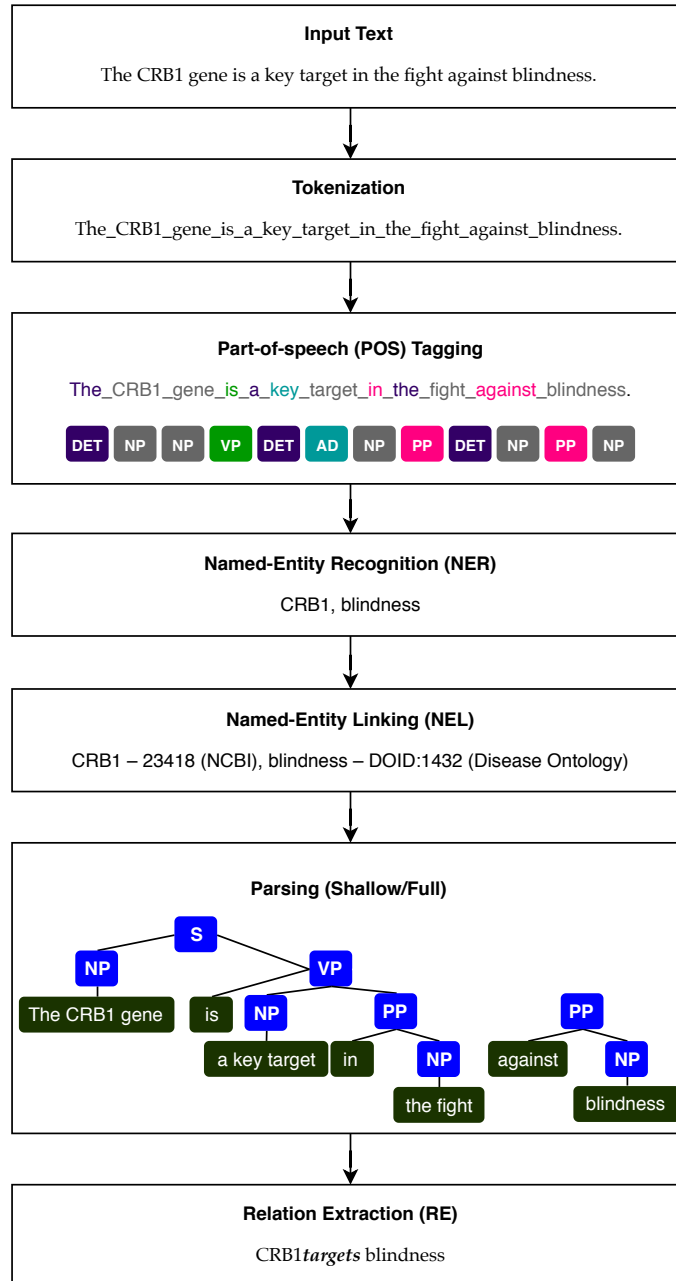
Fig. 2. Workflow of a simplified RE system. (Text obtained from [21])

The bootstrapping method uses a small set of relations known as *seeds* (e.g., gene-disease pairs). The first step is to identify the *seeds* in the data-set and map the relation pattern they describe. The second step is to try to apply the mapped patterns to the data-set to identify new pairs of relations that follow the same construction. Finally, the original set of relations is expanded by adding these new pairs. When after repeating all previous steps no more pairs are found, the process ends. Some systems apply distant supervision techniques to keep track of the validity of the added patterns. Distant supervision uses existing knowledge base entries as gold standards to confirm or discard a relation. This method is susceptible to noisy patterns, as the original set of relations grows. On the other hand, the leveraging of the corpora method makes immediately use of the entire data-set to generate the patterns. This method requires a higher number of annotated relations and produces highly specific patterns that are unable to match new unseen data. Automatically generated patterns can achieve a higher recall than manually defined patterns, but overall the noisy patterns continue to damage the precision. Nevertheless, there are a few efforts to reduce the number of noisy patterns [29].

– **Rule-based**: also uses manually defined and automatically generated rules from the training data to extract relations. Depending on the systems, the differences between pattern-based and rule-based approaches can be minor. Ruled-based approaches not only use patterns but also additional restraints to cover issues that are difficult to express by patterns, such as checking for the negation of the relations [30]. Some ruled-based systems distance themselves from pattern-based approaches by replacing regular expressions with heuristic algorithms and sets of procedures [31]. Like pattern-based approaches, ruled-based approaches tend to have poor recall, even though rules tend to be more flexible. The trade-off recall/precision can be improved using automatic methods for rule creation [32].

– **Machine Learning (ML)-based**: usually makes use of large annotated biomedical corpora (supervised learning) to perform RE. These corpora are preprocessed using NLP tools and then used to train classification models. Beyond Neural Networks, described in detail in Subheading 3, it is possible to categorize ML methods into two main approaches, Feature-based and Kernel-based. **Feature-based approaches** represent each instance (e.g., sentence) as a vector in an n-dimensional space. Support Vector Machines (SVM) classifiers tend to be used to solve problems of binary classification, and are considered *black-boxs* because there is no interference of the user in the classification process. These classifiers can use different features that are meant to represent the data characteristics (e.g., shortest path, bag-of-words (BOW), and POS tagging) [33]. **Kernel-based approaches** main idea is to quantify the similarity between the different instances in a data-set by computing the similarities of their representations [34]. Kernel-based approaches add the structural representation of instances (e.g., by using parse trees). These methods can use one kernel or a combination of kernels (e.g., graph, sub-tree (ST), and shallow linguistic (SL)).

3   Neural Networks for Relation Extraction

Artificial neural networks have multiple different architectures implementations and variants. Often data representations are used as added sources of information to perform text mining tasks, and ontologies may even be used as external sources of information to enrich the model.

3.1   Architectures

Artificial Neural Networks are a parallel combination of small processing units (nodes) which can acquire knowledge from the environment through a learning process and store the knowledge in the connections between the nodes [35] (represented by direct graphs [36]) (Fig. 3). The process is inspired by the biological brain function, having each node correspond to a *neuron* and the connections between the nodes representing the *synapses*.
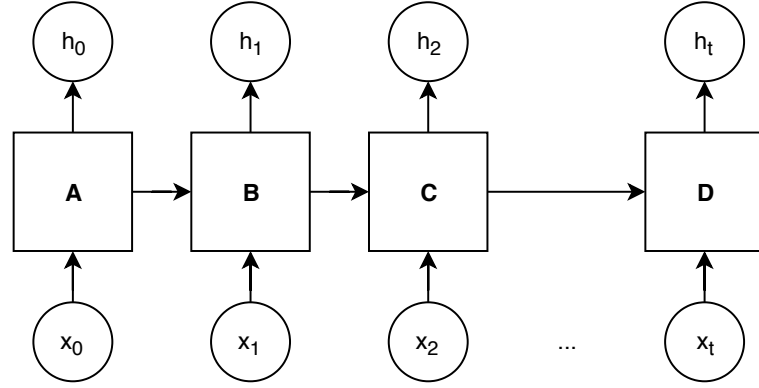

Fig. 3. Architecture representation of an artificial neural networks model, where x_{0-t} represents the inputs and h_{0-t} the respective outputs, for each module from A to D

   Recurrent Neural Networks (RNN) is a type of artificial neural network where the connections between the nodes are able to follow a temporal sequence. This means that RNN can use their internal state, or *memory*, to process each input sequence (Fig. 4). Deep learning techniques, such as RNN, aim to train classification models based on word embeddings, part-of-speech (POS) tagging, and other features. RNN classifiers have multilayer architectures, where each layer learns a different representation of the input data. This characteristic makes RNN classifiers flexible for application to multiple text mining tasks, without requiring task-specific feature engineering.

   Long Short-Term Memory (LSTM) networks are an alternative to regular RNN [10]. LSTMs are a type of RNN that handles long dependencies (e.g., sentences), making this classifier more suitable for the biomedical domain, where
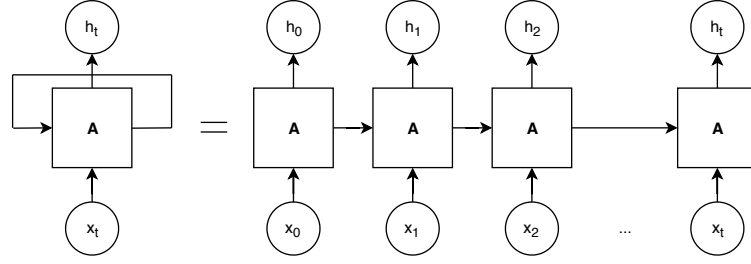
Fig. 4. Architecture representation of a recurrent neural networks model, where x_{0-t} represents the inputs and h_{0-t} the respective outputs, for the repeating module A

sentences are usually long and descriptive (Fig. 5). In recent years, the use of LSTMs to perform Relation Extraction (RE) tasks has become widespread in various domains, such as semantic relations between nominals [37]. **Bidirectional LSTMs** use two LSTM layers, at each step, one that reads the sentence from right to left, and other that reads from left to right. The combined output of both layers produces a final score for each step. Bidirectional LSTMs has yield better results than traditional LSTMs when applied to the same data-sets [38].
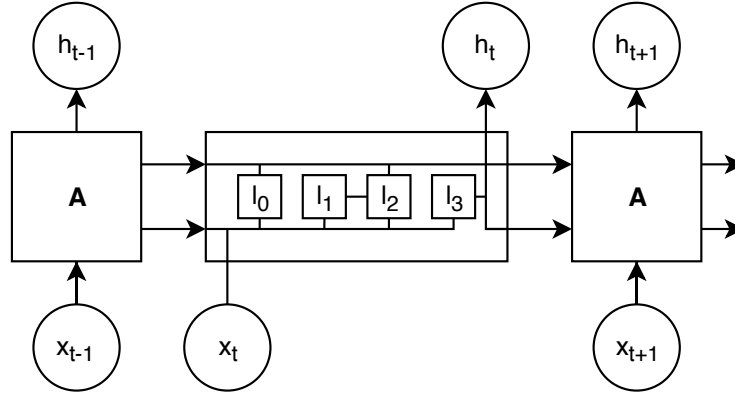

Fig. 5. Architecture representation of a long-short-term memory networks model, where x_{0-t} represents the inputs and h_{0-t} the respective outputs, for the repeating module A, where each repeating module has four interacting layers (l_{0-3})

3.2   Data Representations

The combination of multiple and different language and entityrelated data representations is vital for the success of neural network models dedicated to RE tasks. Some of these features were already described in Subheading 2.1, such as POS tagging and parse trees.

Shortest Dependency Path (SDP) is a feature that identifies the words between two entities mentioned in the text, concentrating the most relevant information while decreasing noise [39].

Word Embeddings are fixed-sized numerical vectors that aim to capture the syntactic and semantic word relationships. These word vectors models use multiple different pretraining sources, for instance, Word2Vec [4] uses English Wikipedia, and BERT [6] uses both English Wikipedia and BooksCorpus. Early models, such as Word2Vec, learned one representation per word, but this proved to be problematic due to polysemous and homonymous words. Recently, most systems started to apply one embedding per word sense. One of the reasons why BERT outperforms previous methods is because it uses contextual models, meaning that it generates a unique representation for each word in a sentence. For instance, in the sentences fragments, *they got **engaged***, and *students were very **engaged** in*, the word *engaged* for non-contextual models would have the same meaning. BERT also outperforms other word vector models that take into account the sentence context, such as ELMo [5] and ULMFit [40], due to being an unsupervised and deeply bidirectional pre-trained language representation.

WordNet Hypernyms are a feature that helps to hierarchize entities, structuring words similar to direct acyclic graphs [41]. For example, *vegetable* is a hypernym of *tubers*, which in turn constitutes a hyponym of *vegetable*. This feature is comparable to an ontology in the sense that a hierarchy relation is identified, but is missing the identification of the relations between the different terms.

Using different features as information sources feeding individual channels leads to multichannel architecture models. Multichannel approaches were already proven to be effective in RE tasks [39].

Regarding biomedical RE, LSTMs were successful in identifying drug-drug interactions [42], gene-mutation relations [43], drug-mutation relations [44], among others. Some methods use domain-specific biomedical resources to train features for biomedical tasks. BioBERT [45] is a domain specific language representation model pre-trained on large-scale biomedical corpora, based on BERT [6] architecture. BioBERT, using minimal task-specific architecture modifications, significantly outperforms previous biomedical state-of-the-art models in the text mining primary tasks of Named-Entity Recognition, Named-Entity Linking, and RE. The BR-LSTM [46] model uses a multichannel approach with pre-trained medical concept embeddings. Using the Unified Medical Language System (UMLS) concepts, BR-LSTM applies a medical concept embedding method developed by Vine et al. [47]. BO-LSTM [48] uses the relations provided by domain-specific ontologies to aid the identification and classification of relations between biomedical entities in biomedical literature.

3.3   Ontologies

An ontology is a structured way of providing a common vocabulary in which shared knowledge is represented [49]. Word embeddings can learn how to detect relations between entities but manifest difficulties in grasping the semantics of each entity and their specific domain. Domain-specific ontologies provide and

formalize this knowledge. Biomedical ontologies are usually structured as a directed acyclic graph, where each node corresponds to an entity and the edges correspond to known relations between those entities. Thus, a structured representation of the semantics between entities and their relations, an ontology, allows us to use it as an added feature to a machine learning classifier. Some of the biomedical entities structured in publicly available ontologies are genes properties/attributes (gene ontology (GO)), phenotypes (human phenotype ontology (HPO)), diseases (disease ontology (DO)), and chemicals (chemical entities of biological interest (ChEBI)). All of these entities participate in relations with different and same domain type entities. Hence, the information about each entity on a semantic level adds a new layer of knowledge to increase the performance of RE classifiers.

Non-biomedical models using ontologies as an added source of information to neural networks is becoming widespread for several tasks. Li et al. [50] propose using word sense definitions, provided by the WordNet ontology, to learn one embedding per word sense for word sense disambiguation tasks. Ma et al. [51] focus their work on semantic relations between ontologies and documents, using the DBpedia ontology. Some researchers explored graph embedding techniques [52] that convert relations to a low dimensional space which represents the structure and properties of the graph. Other researchers have combined different sources of information, including ontological information, to do multi-label classification [53] and used ontology concepts to represent word tokens [54].

However, few authors have used biomedical ontologies to perform RE. Textpresso [55] is a text-mining system that works as a search engine of individual sentences, acquired from the full text of scientific articles, and articles themselves. It integrates biomedical ontological information (e.g., of genes, phenotypes, and proteins) allowing for article and sentence search query by term. The integration of the ontological information allows for semantic queries. This system helps database curation by automatically extracting biomedical relations. The IICE [56] system uses kernel-based support vector machines along with an ensemble classifier to identify and classify drugdrug interactions, linking each chemical compound to the ChEBI ontology. Tripodi et al. [57] system focus on druggene/protein interaction discovery to aid database curation, making use of ChEBI and GO ontologies. BO-LSTM [48] is the only model until now that incorporates ancestry information from biomedical ontologies with deep learning to extract relations from the text, specifically drugdrug interactions and genephenotype relations.

4   Evaluation Measures

The evaluation of machine learning systems is done by applying the trained models to a gold standard test-set, manually curated or annotated by domain experts and unseen by the system. For a Relation Extraction (RE) task, the gold standard test-set should correspond to the list of pairs of entities (e.g.,

phenotypegene or genedisease pairs) that co-occur in the same sentences and their relation (*Known* or *Unknown*).

To any given information extraction system it is necessary to define what constitutes a positive and negative result. In RE tasks the types of results possible are shown in Table 1.

Table 1. Types of results obtained with an information extraction system for a RE task

Annotator (Gold Standard)	System	Classification
Relation	Relation	True Positive (TP)
	No Relation	False Negative (FN)
No Relation	Relation	False Positive (FP)
	No Relation	True Negative (TN)

The primary goal of a given information retrieval system is to maximize the number of TP (True Positives) and TN (True Negatives). To compare results obtained with different data-sets or different tools we have three distinct evaluation metrics: recall, precision and F-measure. Precision represents how often the results are correct, recall is the number of correct results identified and F-measure is a combination of both metrics that expresses overall performance, being the harmonic mean of precision and recall:

$$Recall = \frac{TP}{TP + FN} \qquad Precision = \frac{TP}{TP + FP} \qquad F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (1)$$

The performance of the most recent systems dedicated to biomedical RE, described in Subheading 3.2, is shown in Table 2. These systems are not comparable, since each system is focused on the relations between different biomedical entities, and even addresses more than binary relations, such as the Graph LSTM (GOLD) system.

Table 2. Biomedical RE systems current performance

System	Precision	Recall	F-Measure
DLSTM [42]	0.7253	0.7149	0.7200
Graph LSTM (GOLD) [43]	0.4330	0.3050	0.3580
BioBERT [45]	0.8582	0.8640	0.8604
BR-LSTM [46]	0.7152	0.7079	0.7115
BO-LSTM [48]	0.6572	0.8184	0.7290

For RE tasks a human acceptable performance is usually around 85/90% in F-measure [58]. To facilitate the creation of gold standards we should strive for semi-automation, that is, employ automatic methods for corpora annotation (creating silver standard corpora), and then correct those annotations using domain-specific curators.

5   Conclusions

The way we communicate scientific knowledge is through scientific literature. At the current rate of document growth, the only way to process this amount of information is by using computational methods. The information obtained through these methods can lead to a better understanding of biological systems. When creating biomedical text mining systems, it is essential to take into account the specific characteristics of biomedical literature. Biological information follows different nomenclatures and levels of sentence complexity.

Automatic biomedical Relation Extraction (RE) still has a long way to go to achieve human-level performance scores. Over recent years, some innovative systems have successively achieved better results by making use of multiple knowledge sources and data representations. These systems not only rely on the training data but make use of different language and entity-related features, to create models that identify relations in highly heterogeneous text. This chapter described the evolution of dedicated RE systems, up until neural networks. Neural networks, especially deep neural networks, which make use of multichannel architectures are composed of multiple features. Each system integrates different features distinctively. An optimal combination of features and the ideal features to perform biomedical RE tasks are still far from human level performance.

The knowledge encoded in biomedical ontologies plays a vital part in the development of deep neural networks systems, providing semantic and ancestry information for entities, such as genes, proteins, phenotypes, and diseases. One could also add semantic similarity measures as an additional information layer. Also, to accompany the growth of systems targeting different biomedical relations, there is a growing need for more domain-specific corpora, that can only be accomplished by automated corpus creation (silver standard corpora) [59].

Integrating different knowledge sources instead of relying solely on the training data for creating classification models will allow us not only to find relevant information for a particular problem quicker, but also to validate the results of recent research, and propose new experimental hypotheses.

References

1. Hearst MA (1999) Untangling text data mining. Paper presented at the 37th Annual Meeting of the Association for Computational Linguistics, College Park, Maryland, 2026 June 1999. https://doi.org/10.3115/1034678.1034679
2. PubMed (1996) United States National Library of Medicine. https://www.ncbi.nlm.nih.gov/pubmed/. Accessed 05 Apr 2019

3. Lamurias A, Couto F (2019) Text mining for bioinformatics using biomedical literature. In: Ranganathan S, Nakai K, Schnbach C, Gribskov M (eds) Encyclopedia of Bioinformatics and Computational Biology, vol 1. Elsevier, Oxford, pp 602-611. `https://doi.org/10.1016/B978-0-12-809633-8.20409-3`

4. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013) Distributed Representations of Words and Phrases and their Compositionality. Paper presented at the 26th International Conference on Neural Information Processing Systems, Lake Tahoe, Nevada, 05-10 December 2013

5. Peters M, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, Zettlemoyer L (2018) Deep Contextualized Word Representations. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, New Orleans, Louisiana, 01-06 June 2018

6. Devlin J, Chang M, Lee K, Toutanova K (2018) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, Minnesota, 06-07 June 2019

7. Radford A, Narasimhan K, Salimans T, Sutskever I (2018) Improving Language Understanding by Generative Pre-Training. OpenAI. `https://openai.com/blog/language-unsupervised`. Accessed 02 May 2019

8. Dai Z, Yang Z, Yang Y, Carbonell JG, Le QV, Salakhutdinov RR (2019) Transformer-XL: Attentive Language Models Beyond a Fixed-Length Context. CoRR, abs/1901.02860

9. Radford A, Jeffrey W, Child R, Luan D, Amodei D, Sutskever I (2019) Language Models are Unsupervised Multitask Learners. OpenAI. `https://openai.com/blog/better-language-models/`. Accessed 02 May 2019

10. Hochreiter S, Schmidhuber J (1997) Long Short-Term Memory. Neural Computation 9:1735-1780. `https://doi.org/10.1162/neco.1997.9.8.1735`

11. 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 (2000) Gene Ontology: tool for the unification of biology. Nature Genetics 25:25-29. `https://doi.org/10.1038/75556`

12. Hastings J, Owen G, Dekker A, Ennis M, Kale N, Muthukrishnan V, Turner S, Swainston N, Mendes P, Steinbeck C (2016) ChEBI in 2016: Improved services and an expanding collection of metabolites. Nucleic Acids Res D1(44):D1214-D1219. `https://doi.org/10.1093/nar/gkv1031`

13. Robinson PN, Mundlos S (2010) The human phenotype ontology. Clinical genetics 77(6):525-34. `https://doi.org/10.1111/j.1399-0004.2010.01436.x`

14. Luo Y, Uzuner , Szolovits P (2017) Bridging semantics and syntax with graph algorithms-state-of-the-art of extracting biomedical relations. Briefings in Bioinformatics 18(1):160178. `https://doi.org/10.1093/bib/bbw001`

15. Sennrich R, Haddow B, Birch A (2016) Neural Machine Translation of Rare Words with Subword Units. CoRR, abs/1508.07909

16. Manning CD, Raghavan P, Schtze H (2008) Introduction to Information Retrieval. Cambridge University Press, New York

17. Aho AV, Sethi R, Ullman JD (1986) Compilers: Principles, Techniques, and Tools. Addison Wesley, Boston

18. Westergaard D, Strfeldt HH, Tnsberg C, Jensen LJ, Brunak S (2018) A comprehensive and quantitative comparison of text-mining in 15 million full-text articles versus their corresponding abstracts. PLOS Computational Biology 14:116. `https://doi.org/10.1371/journal.pcbi.1005962`
19. Fleuren WWM and Alkema W (2015) Application of text mining in the biomedical domain. Methods 74:97106. `https://doi.org/10.1016/j.ymeth.2015.01.015`
20. Singhal A, Simmons M, Lu Z (2016) Text mining genotype-phenotype relationships from biomedical literature for database curation and precision medicine. PLoS Computational Biology 12(11):e1005017. `https://doi.org/10.1371/journal.pcbi.1005017`
21. Alves CH, Wijnholds J (2018) AAV Gene Augmentation Therapy for CRB1-Associated Retinitis Pigmentosa. In: Boon C., Wijnholds J. (eds) Retinal Gene Therapy. Methods in Molecular Biology, vol 1715. Humana Press, New York, NY
22. Lamurias A, Clarke LA, Couto FM (2017) Extracting microRNA-gene relations from biomedical literature using distant supervision. PLoS One 12(3):e0171929. `https://doi.org/10.1371/journal.pone.0171929`
23. Zweigenbaum P, Demner-Fushman D, Yu H, Cohen KB (2007) Frontiers of biomedical text mining: current progress. Briefings in Bioinformatics 8(5):358-75. `https://doi.org/10.1093/bib/bbm045`
24. Bunescu R, Mooney R, Ramani A, Marcotte, E (2006) Integrating co-occurrence statistics with information extraction for robust retrieval of protein interactions from MEDLINE. In: Proceedings of the HLT-NAACL BioNLP Workshop on Linking Natural Language and Biology, New York, NY, 8 June 2006
25. Zhou D, He Y, Kwoh CK (2008) From Biomedical Literature to Knowledge: Mining Protein-Protein Interactions. In: Smolinski TG, Milanova MG, Hassanien AE (eds) Computational Intelligence in Biomedicine and Bioinformatics. Studies in Computational Intelligence, vol 151. Springer, Berlin, Heidelberg
26. Hao Y, Zhu X, Huang M, Li M (2005) Discovering patterns to extract protein-protein interactions from the literature: Part II. Bioinformatics 21(15):32943300. `https://doi.org/10.1093/bioinformatics/bti493`
27. Wang HC, Chen YH, Kao HY, Tsai SJ (2011) Inference of transcriptional regulatory network by bootstrapping patterns. Bioinformatics 27:1422-8.73. `https://doi.org/10.1093/bioinformatics/btr155`
28. Liu H, Komandur R, Verspoor K (2011) From Graphs to Events : A Subgraph Matching Approach for Information Extraction from Biomedical Text. In: Proceedings of BioNLP Shared Task 2011 Workshop, Portland, Oregon, 24 June 2011
29. Nguyen QL, Tikk D, Leser U (2010) Simple tricks for improving pattern-based information extraction from the biomedical literature. Journal of Biomedical Semantics 1(1):9. `https://doi.org/10.1186/2041-1480-1-9`
30. Koike A, Niwa Y, Takagi T (2005) Automatic extraction of gene/protein biological functions from biomedical text. Bioinformatics 21:1227-36. `https://doi.org/10.1093/bioinformatics/bti084`
31. Rinaldi F, Schneider G, Kaljurand K, Hess M, Andronis C, Konstandi O, Persidis A (2007) Mining of relations between proteins over biomedical scientific literature using a deep-linguistic approach. Artificial Intelligence in Medicine 39:127-36. `https://doi.org/10.1016/j.artmed.2006.08.005`
32. Xu Y, Hong K, Tsujii J, Chang E I-C (2014) Feature engineering combined with machine learning and rule-based methods for structured information extraction from narrative clinical discharge summaries. Journal of the American Medical Informatics Association 19(5):824832. `https://doi.org/10.1136/amiajnl-2011-000776`

33. Kim MY (2008) Detection of gene interactions based on syntactic relations. Journal of Biomedicine & Biotechnology 2008:371710. https://doi.org/10.1155/2008/371710
34. Giuliano C, Lavelli A, Romano L (2006) Exploiting Shallow Linguistic Information for Relation Extraction from Biomedical Literature. In: Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics, Trento, Italy, 03-07 April 2006
35. HayKin S (1998) Neural Networks: A Comprehensive Foundation. Prentice Hall PTR, New Jersey
36. Guresen E, Kayakutlu G (2011) Definition of artificial neural networks with comparison to other networks. Procedia Computer Science 3:426-433. https://doi.org/10.1016/j.procs.2010.12.071
37. Miwa M, Bansal M (2016) 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, Berlin, Germany, 07-12 August 2016
38. Zhang S, Zheng D, Hu X, Yang M (2015) Bidirectional long short-term memory networks for relation classification. In: Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation, 30 October - 01 November 2015
39. Xu Y, Mou L, Li G, Chen Y (2015) Classifying Relations via Long Short Term Memory Networks along Shortest Dependency Paths. In: Proceedings of Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal, 17-21 September 2015
40. Howard J, Ruder S (2018) Universal Language Model Fine-tuning for Text Classification. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, 15-20 July 2018
41. Christiane Fellbaum (ed) (1998) WordNet: An Electronic Lexical Database. The MIT Press, Cambridge
42. Wang W, Yang X, Yang C, Guo X, Zhang X, Wu C (2017) Dependency-based long short term memory network for drug-drug interaction extraction. BMC Bioinformatics 18(16):578. https://doi.org/10.1186/s12859-017-1962-8
43. Song L, Zhang Y, Wang Z, Gildea D (2018) N-ary Relation Extraction using Graph-State LSTM. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October - 04 November 2018
44. Peng N, Poon H, Quirk C, Toutanova K, Yih W (2017) Cross-Sentence N-ary Relation Extraction with Graph LSTMs. Transactions of the Association for Computational Linguistics 5:101-115. https://doi.org/10.1162/tacl_a_00049
45. Lee J, Yoon W, Kim S, Kim D, Kim S, So C-H, Kang J (2020) BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics 36(4):1234-1240.https://doi.org/10.1093/bioinformatics/btz682
46. Xu B, Shi X, Zhao Z, Zheng W (2018) Leveraging Biomedical Resources in Bi-LSTM for Drug-Drug Interaction Extraction. IEEE Access 6:33432-33439. https://doi.org/10.1109/ACCESS.2018.2845840
47. Vine LD, Zuccon G, Koopman B, Sitbon L, Bruza P (2014) Medical Semantic Similarity with a Neural Language Model. In: Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management CIKM, Shanghai, China, 03-07 November 2014
48. Lamurias A, Sousa D, Clarke LA, Couto FM (2019) BO-LSTM: classifying relations via long short-term memory networks along biomedical ontologies. BMC Bioinformatics 20:10. https://doi.org/10.1186/s12859-018-2584-5

49. Gruber TR (1993) A translation approach to portable ontology specifications. Knowledge Acquisition 5(2):199220. `https://doi.org/10.1006/knac.1993.1008`

50. Li Q, Li T, Chang B (2016) Learning word sense embeddings from word sense definitions. In: Lin C-Y, Xue N, Zhao D, Huang X, Feng Y (eds) Natural Language Understanding and Intelligent Applications, vol 10102. Springer, Cham, pp 22435

51. Ma N, B H-tZ, Xiao X (2017) An Ontology-Based Latent Semantic Indexing Approach Using Long Short-Term Memory Networks. Web and Big Data 10366(2):18599.`https://doi.org/10.1007/978-3-319-63579-8`

52. Goyal P, Ferrara E (2018) Graph embedding techniques, applications, and performance: A survey. Knowledge-Based Systems 151:78-94. `https://doi.org/10.1016/j.knosys.2018.03.022`

53. Kong X, Cao B, Yu PS (2013) 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, Chicago, Illinois, 11-14 August 2013

54. Dasigi P, Ammar W, Dyer C, Hovy E (2017) Ontology-aware token embeddings for prepositional phrase attachment. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Vancouver, Canada, 30 July - 04 August 2017

55. Mller HM, Kenny EE, Sternberg PW (2004) Textpresso: An Ontology-Based Information Retrieval and Extraction System for Biological Literature. PLOS Biology 2(11):e309. `https://doi.org/10.1371/journal.pbio.0020309`

56. Lamurias A, Ferreira JD, Couto FM (2014) Identifying interactions between chemical entities in biomedical text. Journal of Integrative Bioinformatics 11(3):116. `https://doi.org/10.1515/jib-2014-247`

57. Tripodi I, Boguslav M, Haylu N, Hunter LE (2017) Knowledge-base-enriched relation extraction. In: Proceedings of the Sixth BioCreative Challenge Evaluation Workshop, Bethesda, Maryland, 18-20 October 2017

58. Aroyo L, Welty CA (2015) Truth Is a Lie: Crowd Truth and the Seven Myths of Human Annotation. AI Magazine 36:15-24. `https://doi.org/10.1609/aimag.v36i1.2564`

59. Sousa D, Lamurias A, Couto FM (2019) A Silver Standard Corpus of Human Phenotype-Gene Relations. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, Minnesota, 06-07 June 2019