

Predicting Treatment Relations with Semantic Patterns over Biomedical Knowledge Graphs

Gokhan Bakal² and Ramakanth Kavuluru^{1,2*}

¹ Division of Biomedical Informatics, Department of Biostatistics

² Department of Computer Science

University of Kentucky, Lexington, KY

{mgokhanbakal, ramakanth.kavuluru}@uky.edu

Abstract. Identifying new potential treatment options (say, medications and procedures) for known medical conditions that cause human disease burden is a central task of biomedical research. Since all candidate drugs cannot be tested with animal and clinical trials, *in vitro* approaches are first attempted to identify promising candidates. Even before this step, due to recent advances, *in silico* or computational approaches are also being employed to identify viable treatment options. Generally, natural language processing (NLP) and machine learning are used to predict specific relations between any given pair of entities using the distant supervision approach. In this paper, we report preliminary results on predicting treatment relations between biomedical entities purely based on semantic patterns over biomedical knowledge graphs. As such, we refrain from explicitly using NLP, although the knowledge graphs themselves may be built from NLP extractions. Our intuition is fairly straightforward – entities that participate in a treatment relation may be connected using similar path patterns in biomedical knowledge graphs extracted from scientific literature. Using a dataset of treatment relation instances derived from the well known Unified Medical Language System (UMLS), we verify our intuition by employing graph path patterns from a well known knowledge graph as features in machine learned models. We achieve a high recall (92%) but precision, however, decreases from 95% to an acceptable 71% as we go from uniform class distribution to a ten fold increase in negative instances. We also demonstrate models trained with patterns of length ≤ 3 result in statistically significant gains in F-score over those trained with patterns of length ≤ 2 . Our results show the potential of exploiting knowledge graphs for relation extraction and we believe this is the first effort to employ graph patterns as features for identifying biomedical relations.

1 Introduction

Biomedical processes are inherently composed of interactions between various types of entities involved. Typically, these interactions are captured, for computational convenience, as binary relations connecting a subject entity to an

* corresponding author

object entity through a predicate (or relation type). For example, the relation (“Tamoxifen”, *treats*, “Breast Cancer”) indicates that the subject entity Tamoxifen is related to the object entity breast cancer via the relation type or predicate *treats*. Besides *treats*, relations with other types of associative predicates such as *causes*, *prevents*, and *inhibits* are also interesting for biomedical research. Often different relations are put together to derive new relations, also termed knowledge discovery. Given that we have established that relations are central to biomedical research, a natural question that arises is how we obtain these relations. Relations that are already discovered and considered common knowledge in the clinical and biomedical communities are typically manually recorded and distributed in public knowledge bases like the Unified Medical Language System (UMLS) Metathesaurus [9]. However, relations that are not well known and accepted by the scientific community but are being discovered by particular individuals are often reported in research articles that are subject to peer review. Given the exponential growth [3] of scientific literature, it is unrealistic to manually review all articles published on a given topic. Hence, natural language processing (NLP) techniques have been increasingly used to *extract* biomedical relations from free text documents. For instance, the treatment relation example discussed earlier in this paragraph may be extracted from the sentence – “We conclude that Tamoxifen therapy is more effective for early stage breast cancer patients.” However, NLP extractions are essentially based on evidence present in particular sentences and are prone to two types of errors. First, the NLP techniques themselves might not be foolproof and second the evidence found in a particular sentence might be circumstantial and not something that is universally accepted. However, extraction of the same relation from multiple sentences might be indicative of the strength of the relation if it is being reported by multiple research teams.

In this paper, we take a completely different approach to predict relations between arbitrary pairs of biomedical entities input to our predicate specific models. We refrain from the NLP approaches that simply look at individual sentences to extract a potential relation. Instead, we build a large graph of relations (given each relation translates to a labeled edge) extracted using NLP approaches from scientific literature and use semantic path patterns over this graph to build models for specific predicates. That is, instead of looking at what a particular sentence conveys, we model the prediction problem at a global level and build models that output probability estimates of whether a pair of entities participate in a particular relation. Our models are trained with graph pattern features over a well known knowledge graph extracted from scientific literature.

In our approach, a different model needs to be trained for each predicate. The ability to identify potentially viable drugs, procedures, and other therapeutic agents for treating different conditions that cause disease burden among humans is at the heart of biomedical research. So in this paper, we focus on the *treats* predicate and build models that achieve a recall of over 92% with a precision of 71% even with ten times as many negative examples in the test set. Our

method generalizes to other predicates and can also complement other lexical and syntactic pattern based distant supervision [4] approaches for relation extraction.

In the rest of the paper, we first discuss the primary motivation for our efforts and some related work in Section 2. We provide the details of the knowledge graph used in our experiments and specify the graph pattern features used in our models in Section 3. Next, we present the details of our experiments including a discussion of our results in Section 4 and conclude with some remarks on limitations and future work in Section 5.

2 Related Work

As we discussed in Section 1, NLP approaches can be used to extract relations from particular sentences using the linguistic structure of a sentence (syntactic/dependency parse) especially involving the spans of named entities that occur in it [2]. However, in our current approach, we take a global approach to predicting treatment relations between any two entities without looking at particular sentences that contain them.

The primary motivation for our effort is the distant supervision [4] paradigm (also called weak supervision) where any sentence containing a pair of entities known to participate in a relation (based on an external knowledge base) is assumed to manifest that relation in natural language. Hence an external knowledge base of relations is used to search for corresponding entity pairs in all sentences from a corpus. Many lexical and syntactic features surrounding both entity spans in those sentences form features for a multiclass classifier trained on sentences that contain pairs from the external knowledge base. At test time, each pair of entities that co-occurs in at least one sentence, but is not already known to be related according to the external knowledge base, is a candidate input pair to the classifier that predicts the best predicate for that pair using features extracted from all sentences that contain that pair. This is different from conventional approaches that look at every sentence that contains an entity pair and determine whether a particular relation is expressed in it. The distant supervision approach rather looks at all sentences containing the pair and makes a decision based on evidence gleaned from all of them.

Distant supervision offers a great alternative in cases where the number of predicates is large and when hand labeling thousands of sentences with relations is impractical. However, it suffers from two important issues.

1. Not all sentences that contain a pair of entities express the particular relation recorded in the knowledge base. That is, just because a sentence has a pair of entities, it does not automatically mean that the sentence is discussing the particular relation between them that we have in the knowledge base.
2. Negative examples for training, which are typically derived from entity pairs absent in the knowledge base, may not necessarily be true negative examples given the knowledge base could simply be incomplete.

Although later approaches [11, 13–15] addressed these issues to some extent, they still persist given they are inherent to the distant supervision paradigm.

Because our effort relies on graph patterns over knowledge graphs extracted from text, we obviate the issue of whether a particular sentence containing a training entity pair expresses our relation. Although we rely on NLP based extractions for building the knowledge graph over which the patterns are obtained, the graph pattern abstraction disassociates direct dependence on relations in particular sentences. Our method also largely avoids deriving negative examples based on potentially missing relations from the knowledge base by leveraging NLP extractions from over 20 million biomedical research article abstracts. Thus our approach, although inspired by distant supervision, overcomes its main issues.

3 Semantic Patterns over Knowledge Graphs

In this section, we describe our main approach to relation extraction using semantic graph patterns. Our basic intuition is simple: different entity pairs participating in a particular relation type (that is, linked via a specific predicate) are potentially connected in “similar” ways to each other where the connections are paths between them in knowledge graphs extracted from scientific literature. This is analagous to the NLP variant where a particular type of relation manifests with certain lexico-syntactic patterns surrounding the entity pair mentions in free text, which is the central idea exploited in distant supervision. Coming to our approach, we need two important components:

1. a broad scoped and large knowledge graph over which paths connecting candidate entity pairs can be obtained and
2. an approach to identify similar paths connecting entities so we can abstract or “lift” specific paths to high level semantic graph patterns to be subsequently used as features in a supervised classifier.

3.1 The SemMedDB Knowledge Graph

For this effort, we build a large knowledge graph of relations obtained from SemMedDB [1, 7], a large database of over 70 million binary relationships extracted from over 20 million biomedical citations (titles & abstracts). SemMedDB is a public resource made available by the US National Library of Medicine (NLM), which uses a rule based NLP program SemRep [8, 12] to extract “semantic predications” from biomedical text. SemMedDB is produced by running the SemRep program on all biomedical citations made available thorough the PubMed search system. The relations recorded in this database are called semantic predications given SemRep normalizes textual mentions of entities to unique UMLS Metathesaurus concepts (that is, performs named entity recognition) and the predicates are also based upon those available in the UMLS semantic network [5]. Each of the UMLS concepts also has at least one semantic type [6], which is essentially a classification system to categorize different biomedical entities. As such, the 70 million relations in SemMedDB represent a semantic summary of over 20 million biomedical citations. Our knowledge graph

is essentially a directed graph with labeled edges formed from the relations in SemMedDB. The scope of this graph is very broad in a thematic sense given its edges are not limited to a particular biomedical subject. It is also large in that it has around 13 million unique edges³ connecting over 3 million nodes.

3.2 Specific Paths to Semantic Graph Patterns

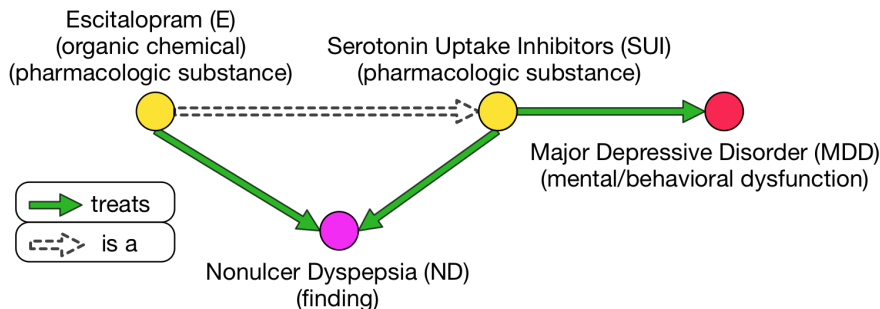


Fig. 1. A small subgraph of the SemMedDB knowledge graph

To abstract specific paths between entities over the SemMedDB graph to semantic patterns, we use an intuitive heuristic – simply replace the concepts along the path with their corresponding semantic type sets (given a concept can have more than one type) and retain the directions of the edges and edge labels as they are. For example, consider a segment of the SemMedDB graph showing a couple of paths between the drug **Escitalopram** (E) and the condition **major depressive disorder** (MDD) in Figure 1. We only employ simple paths (that is, without cycles) and ignore directionality when computing paths (but retaining it after paths are identified). Thus we have the following two paths between E and MDD: (E, is_a, SUI, treats, MDD) and (E, treats, ND, treats⁻¹, SUI, treats, MDD), where SUI stands for **Serotonin Uptake Inhibitors** and ND denotes **Nonulcer Dyspepsia**. For notational convenience we encode the reverse direction with a superscript of -1 on the predicate. To obtain the patterns, we replace the specific entities with their semantic type sets. Thus the corresponding two patterns are:

$$(\{oc, ps\}, is_a, \{ps\}, treats, \{md\})$$

$$(\{oc, ps\}, treats, \{f\}, treats^{-1}, \{ps\}, treats, \{md\}),$$

where oc , ps , f , and md are abbreviations of the semantic types organic chemical, pharmacologic substance, finding, and mental/behavioral dysfunction respectively. By replacing specific entities with their semantic types we aim to

³ Although SemMedDB has 70 million relations, there are many duplicates given a relation can be extracted from multiple sentences due to the semantic mapping to UMLS concepts and semantic network predicates.

capture high level patterns that connect candidate entity pairs. Although we just showed two paths, there are usually many others with a variety of edge types (over 50 different predicates) connecting related entities. We reiterate that our main hypothesis is that these patterns will act as highly discriminative features in identifying entity pairs that participate in a particular type of relationship.

In this paper, we are exclusively interested in predicting treatment relationships and hence we chose this particular example from Figure 1. The two example patterns we show here have a nice high level meaning. In the first pattern, we see that a pharmacologic substance (SUI) is a hypernym for another (E) and is known to treat a dysfunction (MDD). In the second pattern, two pharmacologic substances (SUI and E) both treat a common second condition (ND) while one of them (SUI) treats the target condition (MDD). However, in general, the patterns themselves do not need to have interesting or meaningful interpretations, but when considered together they should be reasonably predictive of the particular predicate that is of interest to us. In this specific example, it turns out that the treatment relationship also holds for our candidate pair (E, MDD). Essentially, we expect to leverage machine learned models to automatically weight different patterns based on their predictive power rather than human experts having to manually identify interesting patterns, a highly impractical task with the explosion of biomedical knowledge.

Before we move on to our experiments, we should mention that although we refer to the SemMedDB graph as a knowledge graph, the precision of NLM’s SemRep tool used to build SemMedDB is known to be around 75% [1]. However, the advantage of our approach is that our prediction is not directly dependent on the correctness of each and every relation in the knowledge graph, rather on the general patterns found within it. Hence any knowledge graph with reasonable quality will suffice although high quality graphs should yield better results.

4 Prediction with Graph Pattern Features

We describe experiments conducted and results obtained using semantic graph patterns connecting candidate pairs as discussed in Section 3. We derive our dataset from the UMLS Metathesaurus’s MRREL table [9, Chapter 2] that has over 11 million *manually curated* relations that are sourced from different biomedical terminologies. Among these we also have several TREATS relations which are used for our experiments. We needed an external human vetted resource like the relations in UMLS given our knowledge graph is derived from a computationally curated relation database. We curated a set of around **7000 unique treatment relations** (entity pairs connected through the *treats* predicate) connecting UMLS concepts from the MRREL table. We divided this positive example dataset into 80% (5600) training set and 20% (1400) test set splits.

4.1 Selection of Negative Examples

Selecting negative examples for distant supervision is a challenge as discussed in Section 2 given the incomplete nature of manually curated knowledge bases

– that is, just because a pair is not participating in a treatment relation in such a knowledge base, does not automatically mean it is a negative example. A second problem is that, in general, any two biomedical entities are not going to have a treatment relation between them. That is, it is not even worth exploring arbitrary pairs of entities to find potential treatment associations and researchers are not usually interested in such pairs. So we carefully choose negative training examples using the following two steps.

1. Every predicate in the UMLS semantic network, including *treats*, has a set of domain/range semantic type constraints. That is, NLM based on expert consultation prescribes which types of entities can take the role of the subject and object in treatment relations. All such possible and allowable subject-object semantic entity type combinations for each predicate are available in three tables with the SRSTR prefix [9, Chapter 5] in the UMLS. We first randomly select a pair of entities (from over 3 million unique UMLS concepts) that satisfies these domain/range constraints.
2. For each pair selected in step 1, we check if there is a *treats* relation between its entities either in the UMLS MRREL table or in the SemMedDB relation database. If it does not hold in our knowledge bases, we include it as a negative example in our dataset.

This two step process selects fairly hard-to-classify negative examples since they satisfy the domain/range constraints but don't participate in a treatment relationship. Checking for membership in both the UMLS and SemMedDB resources minimizes concerns surrounding incomplete knowledge bases. Since we want to predict treatment relations based on graph patterns, if the knowledge graph already has a *treats* edge between our candidate pairs, the prediction could become trivial and the whole process would be self-deceiving. So we deleted any existing *treats* edges between entities in all training/test positive pairs from the knowledge graph (note that negative example selection already ensures this) to guarantee a fair analysis of the predictive ability of graph patterns.

4.2 Experiments and Results

For the experiments, we selected the same number of positive and negative examples (80% of 7000 = 5600) for training. It is straightforward to see that the ground truth or the true treatment relation space is very sparse and most candidate pairs (even if they satisfy domain/range constraints) will belong to the negative class. However, if this sparsity is incorporated as is in the training data, the classifiers might not have enough to learn about predictive patterns for the positive class, which is of most interest to us. Thus we chose equal number of pairs in the training dataset. However, for testing the resultant models, we gradually increased the number of negative examples in the test set to as many as ten times that of the positive class size. We also conducted experiments with patterns of lengths one, two, and three to see the potential of patterns of varying lengths. From our literature review, there are no efficient implementations for

computing *all* simple paths of an arbitrary length between two given nodes in large graphs, although many well known algorithms (e.g., modified breadth first search) exist for identifying shortest paths. In general, finding all simple paths becomes extremely expensive with lengths greater than three simply because the number of such paths could increase drastically in dense graphs. Our implementation for lengths ≤ 3 is based on straightforward heuristics that maintain precomputed lists of neighbors for each node in the knowledge graph. Specifically, to determine length two paths between nodes $e1$ and $e2$, we simply look at nodes in $\mathcal{N}(e1) \cap \mathcal{N}(e2)$ where $\mathcal{N}(e)$ denotes neighbors of node e . To identify length three paths, we look for edge membership for pairs in $\mathcal{N}(e1) \times \mathcal{N}(e2)$ in our knowledge graph.

Table 1. Test set scores with 43,246 patterns of length ≤ 2

Imbalance	Precision	Recall	F-score
$ N = P $	0.970	0.857	0.910
$ N = 2 \cdot P $	0.900	0.857	0.878
$ N = 5 \cdot P $	0.816	0.857	0.836
$ N = 10 \cdot P $	0.716	0.857	0.780

Table 2. Test set scores with patterns of length ≤ 3 with feature minimum frequency thresholds of 1000 (97,864 patterns) and 500 (384,417 patterns)

Imbalance	Min. Frequency: 1000			Min. Frequency: 500		
	Precision	Recall	F-score	Precision	Recall	F-score
$ N = P $	0.951	0.919	0.935	0.955	0.922	0.939
$ N = 2 \cdot P $	0.900	0.919	0.909	0.911	0.922	0.916
$ N = 5 \cdot P $	0.809	0.919	0.861	0.825	0.922	0.871
$ N = 10 \cdot P $	0.695	0.919	0.792	0.708	0.922	0.801

We trained our models with the semantic pattern features using the well known logistic regression classifier available through the Python Scikit-Learn [10] machine learning library. When we conducted experiments with paths of only length one, we found a total of 25 unique patterns used as features. Regardless of the number of negative examples, the precision was over 99% with a recall of 23%. It is clear that length one pattern based models yield very low recall. Next we show the results on the test sets with patterns of lengths ≤ 2 in Table 1 and lengths ≤ 3 in Table 2. In both tables we keep the positive example set P constant and add many examples to increase the size of the negative example set N . From Table 1, we notice there are a total of 43,246 unique patterns for lengths one or two and the precision drops to 71.6% while the recall stays at 85.7% with the ten

fold increase in $|N|$. The number of patterns increased drastically when we also included those with length three. Hence we needed to apply a minimum feature frequency threshold which parameterizes the minimum number of training pairs that should be connected with a particular pattern for it to be included in the final feature space. This is akin to how a minimum frequency is imposed for n-gram features for text classification. We experimented with thresholds of 1000, 500, and 250 that result in feature sets with sizes 97,864, 384,417, and 1,101,106 patterns, respectively. We only show the results for the first two thresholds in Table 2, which indicates a recall improvement of nearly 7% compared with simply using length ≤ 2 patterns from Table 1. This recall gain was obtained at the expense of only 1-2% loss in precision. Between the two thresholds, we notice that the smaller 500 threshold gives a precision gain of 1.3% and a recall gain of 0.3% compared with the larger threshold from the last row of the table.

We did not show the results of experiments we did with a threshold of 250, which led to over a million features, because the performance slightly dipped compared with the larger thresholds shown in Table 2. In both tables, we see that the recall stays constant even as precision goes down when more negative examples are added to the test set. That is, the number of false negatives stayed constant as the imbalance in test set increased. This is not surprising given the model was not changed during these experiments and only new negative examples were added to increase class imbalance in the test set.

We conducted an additional set of experiments where we obtained repeated measurements for the last row cases in both tables (min. frequency 500 case for Table 2) with train-test splits from hundred distinct shuffles of our full dataset. Based on these experiments the 95% confidence intervals for precision are 71.11 ± 0.2 (length ≤ 2) and 68.30 ± 0.5 (length ≤ 3) and for recall are 86.14 ± 0.2 (length ≤ 2) and 92.80 ± 0.1 (length ≤ 3). These results establish the generalizability of our results and show that adding patterns of length 3 results in more than 6% improvement in recall with a less than 3% loss in precision, which is a reasonable compromise given the recall increase is double that of the loss in precision. The confidence intervals for the F-score are 77.90 ± 0.1 (length ≤ 2) and 78.66 ± 0.3 (length ≤ 3), which clearly indicate that the improvement in F-score is also statistically significant given the intervals do not overlap.

Although we cannot predict or quantify the true distribution of treatment relations, we think the imbalance will be more extreme than the 1:10 ratio we tried in our experiments. While additional experiments are justified, even with what we were able to show, we conclude that semantic graph patterns are definitely useful in predicting treatment relations and have the potential to complement NLP based approaches and also aid in relation extraction for other predicates.

5 Concluding Remarks

In this paper, we employed semantic graph patterns connecting pairs of candidate entities as the sole set of features to predict treatment relations between them. We exploited a well known biomedical relation database to build a knowl-

edge graph with over 13 million edges. We then used the knowledge graph to derive features and also select good negative training instances for experiments. Evidenced by the results presented in Section 4, we have successfully verified our hypothesis that semantic patterns over knowledge graphs can be powerful predictors of treatment relations. Our central idea is straightforward, intuitive, and naturally generalizes to other predicates besides *treats*, and also to other domains of interest where knowledge graphs of reasonable quality are available. Next, we identify some limitations of our effort and discuss future research directions to address them.

- When a node has multiple semantic types (like *Escitalopram* in Figure 1), instead of considering them as a compound set type in the semantic abstraction process (see Section 3.2), we can relax them into multiple semantic patterns each with the constituent primitive types. This can drastically prune the feature space and might result in semantically higher level patterns than we are able to obtain now.
- From Tables 1 and 2, it is clear that with increasing class imbalance, the precision significantly goes down – with a ten fold increase in negative test examples, precision dropped by 25%. This presents both an opportunity and a challenge: some of the false positives (FPs) leading to low precision might actually be indicating potential new treatments while others might be simply wrong. The challenge is to separate these two types of FPs; solution might involve semi-automatic approaches using classifier scores and biomedical domain experts. Features that are weighted heavily by the classifier might also give us insights into ranking FPs in descending order of their potential for representing an actual new treatment relation.
- A thorough manual analysis of FPs and false negatives (FNs) can also help us identify broad classes of such errors that can be targeted using specific tailored heuristics including some form of pre-processing of the knowledge graph and post-processing of the results. A more thorough analysis of top features and errors for different classes of treatment relations depending on each allowed semantic type combination can also help us further fine-tune these heuristics based on the semantics types of input pair.
- We strongly believe our work complements NLP approaches in that the graph patterns can be used as additional features along with the lexico-syntactic features typically employed for weakly supervised relation extraction. Besides combining our methods with NLP approaches, employing more sophisticated machine learning techniques especially ensemble approaches [16] might help in improving the overall results.

Acknowledgments

Thanks to anonymous reviewers for their helpful comments that helped improve the paper. The project described in this paper was supported by the National Center for Advancing Translational Sciences (UL1TR000117). The content is

solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

1. H. Kilicoglu, D. Shin, M. Fiszman, G. Roseblat, and T. C. Rindflesch. Semmeddb: a pubmed-scale repository of biomedical semantic predications. *Bioinformatics*, 28(23):3158–3160, 2012.
2. S. Kim, H. Liu, L. Yeganova, and W. J. Wilbur. Extracting drug–drug interactions from literature using a rich feature-based linear kernel approach. *Journal of biomedical informatics*, 55:23–30, 2015.
3. Z. Lu. PubMed and beyond: a survey of web tools for searching biomedical literature. *Database: the journal of biological databases and curation*, 2011, 2011.
4. M. Mintz, S. Bills, R. Snow, and D. Jurafsky. Distant supervision for relation extraction without labeled data. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 1003–1011. Association for Computational Linguistics, 2009.
5. National Library of Medicine. Current Hierarchy of UMLS Predicates. http://www.nlm.nih.gov/research/umls/META3_current_relations.html.
6. National Library of Medicine. Current Hierarchy of UMLS Semantic Types. http://www.nlm.nih.gov/research/umls/META3_current_semantic_types.html.
7. National Library of Medicine. Semantic MEDLINE Database. <http://skr3.nlm.nih.gov/SemMedDB/>.
8. National Library of Medicine. SemRep - NLM’s Semantic Predication Extraction Program. <http://semrep.nlm.nih.gov>.
9. National Library of Medicine. Unified Medical Language System Reference Manual. <http://www.ncbi.nlm.nih.gov/books/NBK9676/>.
10. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
11. S. Riedel, L. Yao, and A. McCallum. Modeling relations and their mentions without labeled text. In *Machine Learning and Knowledge Discovery in Databases*, pages 148–163. Springer, 2010.
12. T. C. Rindflesch and M. Fiszman. The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text. *Journal of Biomedical Informatics*, 36(6):462–477, 2003.
13. A. Ritter, L. Zettlemoyer, O. Etzioni, et al. Modeling missing data in distant supervision for information extraction. *Transactions of the Association for Computational Linguistics*, 1:367–378, 2013.
14. M. Surdeanu, J. Tibshirani, R. Nallapati, and C. D. Manning. Multi-instance multi-label learning for relation extraction. In *Proceedings of the 2012 Conference on Empirical Methods in Natural Language Processing*, pages 455–465. Association for Computational Linguistics, 2012.
15. W. Xu, R. Hoffmann, L. Zhao, and R. Grishman. Filling knowledge base gaps for distant supervision of relation extraction. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 665–670. Association for Computational Linguistics, 2013.
16. Z.-H. Zhou. *Ensemble methods: foundations and algorithms*. CRC Press, 2012.