Improvement of n-ary Relation Extraction by Adding Lexical Semantics to Distant-Supervision Rule Learning

Hong Li[1], Sebastian Krause[1], Feiyu Xu[1], Andrea Moro[2], Hans Uszkoreit[1] and Roberto Navigli[2]

[1]*Language Technology Lab, DFKI, Alt-Moabit 91c, 10559 Berlin, Germany*
[2]*Dipartimento di Informatica, Sapienza Università di Roma, Viale Regina Elena 295, 00161 Roma, Italy*
{lihong, skrause, feiyu, uszkoreit}@dfki.de, {moro, navigli}@di.uniroma1.it

Abstract: A new method is proposed and evaluated that improves distantly supervised learning of pattern rules for n-ary relation extraction. The new method employs knowledge from a large lexical semantic repository to guide the discovery of patterns in parsed relation mentions. It extends the induced rules to semantically relevant material outside the minimal subtree containing the shortest paths connecting the relation entities and also discards rules without any explicit semantic content. It significantly raises both recall and precision with roughly 20% f-measure boost in comparison to the baseline system which does not consider the lexical semantic information.

1 INTRODUCTION

The task of relation extraction is to recognise and extract relations between entities or concepts in free texts. Parse trees have become a popular source for discovering extraction patterns, which encode the grammatical relations among the phrases that jointly express the instance of an n-ary relation. In rule-based relation extraction methods, the patterns are directly applied to extract relation mentions from parsed sentences of free texts (e.g., (Yangarber et al., 2000; Krause et al., 2012; Alfonseca et al., 2012)). Other methods treat relation extraction as a classification or sequence-labeling problem, but even for those techniques parse-tree patterns have proven useful as key features for the classifiers (e.g. (Zelenko et al., 2003; Bunescu and Mooney, 2005; Mintz et al., 2009)). Our work presented here belongs to the rule-based relation extraction methods.

In comparison to the statistical classifier methods, the rules should be able to incorporate a higher degree of structural complexity, and therefore provide more contextual information for correct extraction. However, the most widely used pattern discovery methods extract minimal subtrees containing the arguments of the relation or the shortest paths connecting them in dependency parses (e.g., (Zelenko et al., 2003; Bunescu and Mooney, 2005; Xu et al., 2007; Mintz et al., 2009; Krause et al., 2012)). In our own automatic pattern learning experiments, we have observed the following problems when using the smallest subtree approaches or even the radical shortest path method (Bunescu and

Mooney, 2005):

- except for the entities themselves, the minimal subtrees are often semantically empty and therefore not able to express explicit semantic relations between the entities. E. g., a pattern *and(Person, Person)* is a typical example of automatically learned patterns with minimal span for relations between two persons.

- the minimal subtrees indicate a relation different from the target relation. For example, a pattern *meet(Person, Person)* is not suited for extracting marriage relations between persons.

- a shortest path can be semantically incomplete. A pattern *celebrate(Person, with(Person), wedding)* indeed indicates a marriage. However, the shortest path method learns only *celebrate(person, with(person))* which extracts many events of celebration that are not weddings.

The major reason of the above problems is that the minimal subtree or the shortest path solution do not provide the sufficient semantic conditions for a correct extraction. In various statistical approaches (e.g., (Mintz et al., 2009; Chowdhury and Lavelli, 2012)), additional features such as words around entities, words between entities or trigger words are employed to compensate the problem.

In this paper, we propose an extension of the pattern discovery algorithm by integrating relation-relevant lexical semantic information. The algorithm is embedded in a distant supervision framework. Our

pattern rule discovery and learning system runs without any manual annotation. In previous distant supervision approaches, a sentence is a candidate of a relation mention if the entities of a relation instance occur in this sentence. In our new approach, a sentence is a candidate of a relation mention if it contains at least the two main entities of a relation instance and at least one other relation-relevant semantic term. Thus, in addition to a large base of factual knowledge such as freebase (Bollacker et al., 2008) utilised by several distant-supervision methods, we also utilize a large repository of lexical semantic knowledge, i.e., BabelNet (Navigli and Ponzetto, 2012). The input of our system for our experiment contains 1) around 17K seed facts taken from Freebase for three biographic relations; 2) a large volume of free texts crawled from the Web, totaling around 500K web documents that each contain the (main) entities of a Freebase seed fact; 3) for each target relation a large number of content words, actually word senses, that are semantically relevant to this relation. These were learned by connecting two data sets: (i) the content words of all sentences in the crawled texts that contain instance candidates and (ii) the lexical knowledge repository BabelNet that has been acquired by unsupervised learning from WordNet and Wikipedia (Navigli and Ponzetto, 2012). A core ingredient for the learning of BabelNet and for its application to our mention candidates is word-sense disambiguation (Navigli, 2009). The sentences with candidate mentions are preprocessed with named entity recognition, dependency parsing and BabelNet-based word sense disambiguation and entity linking (Moro et al., 2014). Then in each parsed sentence the entities of the Freebase facts and any occurrence of a semantically relevant word sense are automatically marked by annotation. Now the pattern extraction can extract from an annotated parse all minimal trees containing all argument entities and one or more semantically related terms. If the minimal tree spanning just the argument entities already contains a semantically relevant term, this minimal tree also qualifies as a pattern.

The experimental results show that the new patterns significantly improve both recall and precision for the selected biographic relations *marriage*, *parent-child* and *sibling* with some 20% f-measure boost in average. We choose the biographic relations because the linguistic constructions expressing relations between two persons often contain coordinations and appositions connecting entities without a directly attached overt marker of a semantic relations. These constructions posed a problem to previous methods that only consider minimal subtrees or shortest paths between the instance entities.

2 RELATED WORK

Our approach learns semantically enriched dependency pattern rules for relation extraction by distant supervision from large scale text volumes. We learn the dependency paths between the semantic arguments and the relation-relevant lexical semantic terms automatically without any human intervention.

In the last section, we pointed out that minimal subtrees or shortest paths connecting entities often do not provide sufficient semantic context for extracting target relations. But integrating domain or relation relevant terms into relation extraction rule is not a new idea. Many early information extraction systems use event trigger words for locating the relevant sentences or instances (e.g., (Grishman and Sundheim, 1996; Appelt and Israel, 1999; Grishman et al., 2005)). Many systems (e.g., (Ravichandran and Hovy, 2002; Agichtein, 2006)) learn automatically lexical syntactic patterns that include words between and around the entities of the instance. As mentioned before, words in the textual context of the entities of a relation mention are commonly employed as features in addition to the dependency patterns in statistical approaches, e.g., also in the recent distant supervision approaches to relation extraction (Mintz et al., 2009; Jean-Louis et al., 2013; Min et al., 2013). However, in these approaches, the words are selected on the basis of their textual distance to the entities, not because of their semantic domain relevance.

Because of space limitations, we just refer to two very closely related approaches. (Grishman et al., 2005) present a supervised pattern discovery approach to event extraction. Utilizing a training corpus annotated with both event arguments and event anchors, paths are learned between the event trigger and the individual arguments. The drawback is the need for manual labelling of training data with event triggers. The second related approach is (Xu et al., 2009), where dependency patterns are learned for the detection of binary relations. The patterns must contain at least three nodes: the two semantic arguments and a key word which indicates the semantic relation. Just as much as other early IE work, this approach is more suitable for learning from smaller manually annotated corpora. There is no indication of how the relation relevant keywords could be acquired without manual intervention.

Our work is not directly related to the Open Information Extraction approaches (e.g., (Banko and Etzioni, 2008; Wu and Weld, 2010; Etzioni et al., 2011; Moro and Navigli, 2013)), although these also learn patterns from the web and apply them to extract entities and facts from free texts (e.g., (Fader et al., 2011;

Mausam et al., 2012; Xu et al., 2013)). However, their objective is not to target specific relations, e.g., for feeding a knowledge base. Thus, their requirement of semantic disambiguation is limited.
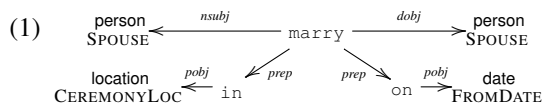
3 PATTERN DISCOVERY VIA SHORTEST DEPENDENCY PATH SPANNING N SEMANTIC ARGUMENTS

In this section, we describe our baseline relation extraction system, which learns relation extraction pattern rules from dependency tree structures in a distant-supervision manner. The rule discovery method makes use of the shortest paths between the entities of a relation instance mentioned in a sentence. In the following, we refer to this baseline system by SPL, for Shortest Path Learner.

3.1 SPL

As defined by most distant supervision systems (e.g., (Mintz et al., 2009; Alfonseca et al., 2012)), SPL regards a sentence as a candidate of a relation mention if it contains the (main) entities of a relation instance of the fact knowledge base. SPL utilizes facts from Freebase for annotating the relation mentions in the candidate sentences and learns automatically dependency pattern rules from the sentence parses. In comparison to most other relation extraction systems, SPL can deal with n-ary relations, not only binary relations. Furthermore, just as in the Snowball system (Agichtein, 2006), SPL rules assign the semantic role labels to the relation arguments.

The following example rule of SPL for the relation *marriage* contains four arguments, two married persons plus the wedding location and the starting date of the marriage. The notation person|SPOUSE represents a placeholder for an entity mention of type person, which is assigned the role label SPOUSE at extraction time.

(1)

3.2 SPL Rule Discovery Algorithm for N-ary Relations

SPL's rule learning identifies patterns in automatically annotated sentences and then induces extraction rules from these patterns. An annotated sentence contains named entity markup, its dependency parse tree and the marked semantic arguments of a relation instance if a Freebase fact matches this sentence.

Pattern extraction in SPL aims to find linguistic patterns that trigger the relations, locate the relation arguments and assign the corresponding semantic roles to these arguments. The pattern-extraction algorithm of SPL is outlined in Algorithm 1. Given a sentence with a mention of a target-relation instance, SPL learns one or more RE rules from it. The RE rules are effectively all sub-graphs of the sentence's dependency parse that satisfy the criteria listed in (3 a-c) of Algorithm 1.

Input:	▶ an instance $r(a_1,\ldots,a_n)$ of the n-ary target relation r ▶ a sentence s with mentions of $a_1,\ldots,a_n$
(1)	augment s with morphologic and syntactic information • create dependency-parse $d_s = (V,E)$ for s • attach lemmatization information to nodes V of d_s
(2)	process s & d_s with entity recognition • detect mentions of the arguments $a_1,\ldots,a_n$ and replace the corresponding nodes in V with place-holders for entity type and role label
(3)	find all sub-graphs C of d_s s.t. $\forall c \in C$, $c = (V_c,E_c)$: (a) V_c contains two or more of the argument mentions $a_1,\ldots,a_n$ (b) c is the minimal subtree of d_s containing shortest paths connecting the nodes defined by (a) (c) V_c contains a content word (i. e., nouns, verbs, adjectives, adverbs)
Output:	▶ a set of graphs C, which can be used for pattern-based relation extraction

Algorithm 1: SPL pattern-learning algorithm.

As an example, consider the 5-ary target relation *marriage* with the argument signature

(2) ⟨ person|SPOUSE, person|SPOUSE, location|CEREMONYLOC, date|FROMDATE, date|TODATE ⟩,

as well as the seed fact from Freebase:

(3) ⟨ Brad Pitt|SPOUSE, Jennifer Aniston|SPOUSE, –|CEREMONYLOC, –|FROMDATE, –|TODATE ⟩.

Given the example sentence (4) from the Web, SPL produces the analysis depicted in Figure 1. Here, the entity mentions (in blue) have already been assigned their semantic roles by exploiting the role mapping from the seed fact (3).

(4) In addition, a friend says, Brad Pitt's marriage to Jennifer Aniston wasn't the golden love story it appeared to be.

Processing this linguistic analysis of the input sentence, Algorithm 1 yields the following learned rule, namely, the shortest path connecting the two person names:
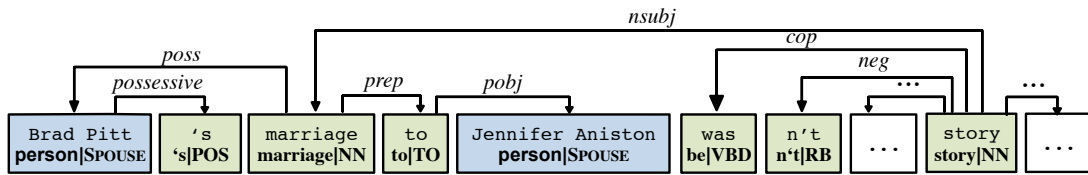
(5)

Figure 1: Dependency parse of the sentence in (4). Certain parts of the parse are left out for brevity. Blue nodes represent detected entity mentions, green nodes correspond directly to tokens of the input sentence.

3.3 The Problem of Relation Clues Outside of Minimal Subtrees

While the pattern-learning algorithm, described in the last section, works reasonably well for many sentences with target-relation mentions, the algorithm fails to extract the gist of the mention if important relation-relevant terms are not contained within the minimal component of the dependency parse which links the semantic arguments. In such cases, the algorithm extracts semantically underspecified rules not suitable for accurate RE. As an example, see the following sentence and its linguistic analysis in Figure 2:

(6) Brad Pitt celebrated a wonderful wedding with Jennifer Aniston.

Algorithm 1 from Section 3.2 identifies the sub-graph highlighted in purple as semantically relevant, but misses the path to the verb's object wedding (highlighted in red in Figure 2), thus returning a misleading pattern which only captures that a person celebrated with another person.

4 AUTOMATIC ACQUISITION OF RELATION-RELEVANT LEXICAL SEMANTICS

For determining relation-relevant terms, we use Babel-Net (Navigli and Ponzetto, 2012) as our initial lexical semantic knowledge base. BabelNet[1] is a large-scale multilingual semantic network that was automatically built through the algorithmic integration of Wikipedia, OmegaWiki, Wikidata, Wiktionary, WordNet (Fellbaum, 1998) and Open Multilingual WordNet (Bond and Kyonghee, 2012). Its core components are the Babel synsets which are sets of multilingual synonyms automatically extracted from the considered resources. Each Babel synset is related to other Babel synsets with semantic relations obtained from WordNet and Wikipedia, such as hypernymy, meronymy and semantic relatedness. BabelNet 2.5 contains roughly 9M
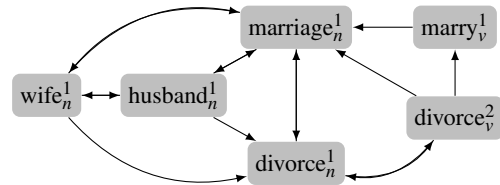
_____
[1] http://babelnet.org

Figure 3: An excerpt of the semantic graph associated with the relation *marriage*, see (Moro et al., 2013). Node labels refer to BabelNet synsets, for example "wife_n^1" represents the first sense of the noun "wife". Edges correspond to edges between these synsets in BabelNet.

synsets, 15M lexicalizations in 50 languages and 250M relation instances.

(Moro et al., 2013) presented a new method for creating so-called *relation-specific semantic graphs* by using this generic lexical semantic resource together with automatically learned relation extraction patterns and their sentence mentions for a semantic relation type. They applied word sense disambiguation to the content words of the automatically learned relation extraction patterns by using the sentence with mentions as semantic contexts. Then, the most frequent word senses were considered as key concepts for the target relation. Finally, relation-specific subgraphs were extracted from BabelNet starting from the key concepts and using simple neighborhood expansion. This knowledge-based approach works without any supervision, and can be applied to any semantic relation type for which lexicalized patterns exist. Moreover, it is a parametrized approach, i.e., there is a free parameter that allows application-specific fine tuning for better recall or precision. A small example graph for the relation *marriage* is depicted in Figure 3. For our experiments, we have created one subgraph for each of the three target relations.

5 A NEW PARADIGM

This section presents an extended version of the pattern-extraction approach from Section 3.2, which is able to identify relation-relevant parts of dependency graphs outside of the minimal subtree. Since Algorithm 1 can deal with *n* arguments, the extension is straightforward.
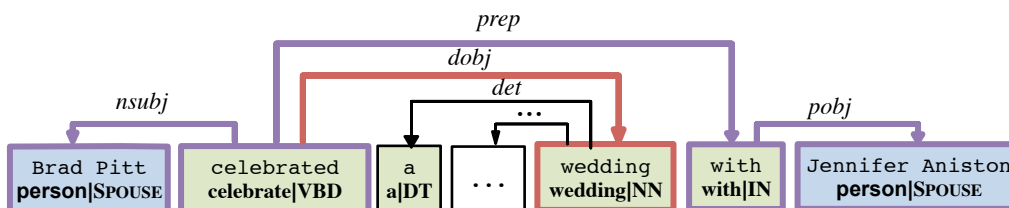
Figure 2: Dependency parse of the sentence in (6). Again, certain parts of the parse have been left out for brevity.
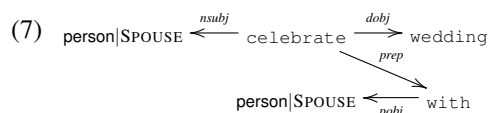
5.1 Idea & New Pattern-Learning Algorithm

Input:	▶ a relation instance $r(a_1, \ldots, a_n)$ and a sentence s (see Alg. 1)
	▶ the semantic graph $SG_r = (V_{SG_r}, E_{SG_r})$ of the target relation r
(1)	augment s with morphologic and syntactic information (see Alg. 1)
(2)	process s and its dependency parse d_s with entity recognition (see Alg. 1)
(3)	find all sub-graphs C of d_s s.t. $\forall c \in C,\ c = (V_c, E_c)$:
	(a) V_c contains two or more of the argument mentions $a_1, \ldots, a_n$
	(b) V_c contains one or more relation-specific semantic terms: $V_c \cap V_{SG_r} \neq \varnothing$
	(c) c is the minimal subtree of d_s containing the shortest paths connecting the nodes defined by (a) & (b)
Output:	▶ a set of graphs C, which can be used for pattern-based relation extraction

Algorithm 2: Proposed pattern learning, enhancing Algorithm 1 with lexical-semantic information.

We propose to enhance the pattern-extraction method of Algorithm 1 by injecting lexical-semantic information from the relation-specific semantic graphs we have acquired based on the method described in Section 4. The enhanced version is shown in Algorithm 2. The major improvement of the new algorithm over the original one is given by (3b), which allows the dependency subtree detection to make a lexico-semantically informed choice. For the example of the relation *marriage* this means that $V_{SG_{marriage}}$ contains terms like `bride`, `divorce`, `fiance`, `hubbie` and `wedding`, among others. The pattern-extraction process exploits this information during the identification of shortest paths linking the mentioned relation arguments in the parse trees, i. e. it extends subgraph until one or more of such terms are included. For the example in Figure 2, Algorithm 2 identifies the semantic term `wedding` and extracts the following relevant

pattern, which indeed catches the main content of the relation mention:

(7)

6 EXPERIMENTS & EVALUATION

In the following, we evaluate the impact of the proposed extension on the relation extraction performance for three semantic relations.

6.1 Setup

The experiments in this section were carried out using the gold-standard corpus *Celebrity* (Li et al., 2014). This corpus consists of 142 newspaper articles, annotated with gold mentions of three kinship relations: *marriage*, *parent-child*, *siblings*. The argument signature of *marriage* is given in (2), the ones of *parent-child* and *siblings* are similar, i. e., both relate sets of person mentions.

We compare the performance of patterns learned using Algorithms 1 & 2, as well as a third pattern set, which represents an alternative way to incorporate lexical semantics into pattern learning:

- **SPL:** patterns from Algorithm 1
- **SPL+SG-Filter:** patterns from Algorithm 1 after a subsequent pattern-filtering step. Only patterns containing the semantic terms in the lexical semantic subgraphs are kept.
- **SG-SPL:** patterns from Algorithm 2

In order to generate training examples for the pattern-learning step, we followed a distant-supervision approach, which included collecting instances (*seeds*) of the three target relations from Freebase (totally, around 17K seed facts) and finding web documents (*docs*) mentioning the seeds' arguments (around 500K documents). The first part of Table 1 lists details about the training data for the individual relations. In this table,

Table 1: Statistics about training data and relation extraction rules. "Matched patterns" refers to the amount of patterns which matched at least one sentence in the evaluation corpus.

	training data			learned patterns			matched patterns		
	# seeds	# docs	# synsets	SPL	SPL+ SG-Filter	SG-SPL	SPL	SPL+ SG-Filter	SG-SPL
marriage	5,993	211,186	54	88,456	33,822	79,178	498	112	166
parent-child	3,379	148,598	126	45,093	29,592	76,765	357	159	272
siblings	7,630	130,448	56	26,250	13,004	38,412	204	70	132

synsets refers to the nodes of the respective relation-specific semantic graph (i. e., "# synsets" $= |V_{SG_r}|$, for V_{SG_r} from Algorithm 2). We employ Maltparser for parsing the sentences (Nivre et al., 2007).

Table 1 also lists statistics about the number of relation extraction rules per pattern set and relation. The new approach generates a similar number of patterns as the original algorithm does, but compared to SPL+SG-Filter the amount of rules is more than doubled. Since all the rules in one set differ lexically and/or syntactically, an ideal evaluation of the rules would require an enormous annotated corpus in order to validate a larger fraction of the patterns. As such corpora are too expensive, we had to stay with the already mentioned Celebrity corpus and thus had to accept the low number of actually evaluated rules, as shown in the right half of Table 1.

6.2 Experimental Results

Table 2 lists statistics about the relation extraction performance of the three pattern sets on the Celebrity corpus. The new method has improved both precision and recall significantly for each target relation. The average precision improvement in comparison to the baseline system is 20.4%, while the improvement of the recall is 16.68% and the f-measure with 21.66%.

While applying lexical semantics to rule filtering does help improve precision (SPL+SG-Filter vs. SPL), it inevitably leads to a recall drop due to the sharply reduced number of rules. The new algorithm SG-SPL is naturally able to achieve the same precision improvement because it restricts the possible pattern set during pattern learning by utilizing the same lexical-semantic information as SPL+SG-Filter. However, SG-SPL is in addition capable of lifting recall to a higher level because it enables the learning of patterns from relation mentions that do not contain a content word on the shortest path between the arguments but nevertheless exhibit one or more semantically relevant words in the intrasentential context of the arguments. We discuss examples in the next section.

6.3 Result Analysis

In this section, we analyze differences in the pattern sets that bring about the increased recall of SG-SPL compared to the other approaches. We also give examples of cases where mistakes in the learning process led to the extraction of erroneous patterns.

Quite a number of target-relation mentions link the persons participating in the relation only by a conjunction, shifting relation triggers to the context of the argument mentions. Our new approach is in many cases able to identify a trigger word as being semantically relevant and thus incorporates it in the extracted pattern. Examples include the *marriage* patterns (8a.) and (9a.), maching the Celebrity-corpus' sentences (8b.) and (9b.), respectively:

(8)

a. $\text{wedding} \xrightarrow{nn} \text{person}|\text{SPOUSE} \xrightarrow{conj} \text{person}|\text{SPOUSE}$

b. The good feelings were on display the evening of <u>Scott and Laci</u>'s <u>wedding</u>.

(9)

a. $\text{marry} \xrightarrow{nsubj} \text{person}|\text{SPOUSE} \xrightarrow{conj} \text{person}|\text{SPOUSE}$

b. Two years after <u>Aniston and Pitt married</u>, ...

A similar example pattern from the same relation is (10), which again contains a semantic key term outside of the shortest path between the relation arguments. Corresponding rules were learned for the other two relations as well, i. e., for *parent-child* patterns like (11) and for *siblings* ones like (12):

(10) $\text{ex-husband} \xleftarrow{nn} \text{person}|\text{SPOUSE} \xrightarrow{poss} \text{person}|\text{SPOUSE}$

(11) $\text{person}|\text{PARENT} \xleftarrow{poss} \text{person}|\text{CHILD} \xrightarrow{nn} \text{daughter}|\text{son}|\text{child}$

(12) $\text{person}|\text{SIBLING} \xleftarrow{poss} \text{person}|\text{SIBLING} \xrightarrow{nn} \text{brother}|\text{sister}$

Sometimes patterns exclusively learned by SG-SPL fail for erroneous syntax in application sentences, despite being correct. For example, the following rule (13a.) mistakenly matches the sentence in (13b.) because of an incorrect dependency analysis:

Table 2: relation extraction performance on Celebrity corpus.

		SPL	SPL+SG-Filter	SG-SPL
precision	*marriage*	16.49%	40.00%	38.70%
	parent-child	17.89%	36.80%	33.30%
	siblings	4.89%	13.40%	27.70%
	macro-avg.	13.09%	**30.07%**	**33.23%**
recall	*marriage*	50.96%	43.80%	48.40%
	parent-child	40.76%	35.50%	49.50%
	siblings	18.36%	17.80%	62.20%
	macro-avg.	36.69%	**32.37%**	**53.37%**
F1 score	*marriage*	24.91%	41.81%	43.00%
	parent-child	24.86%	36.13%	39.81%
	siblings	7.72%	15.28%	38.33%
	macro-avg.	19.30%	**31.17%**	**40.96%**

(13)

a. person|SPOUSE $\xleftarrow{conj}$ person|SPOUSE $\xrightarrow{rcmod}$ marry

b. ... between <u>Amber and Scott</u>, who had told her he was not <u>married</u>.

Another issue resulting in false positive extractions can be attributed to the fact that the semantic graph for a relation may contain terms of slightly varying significance for the relation. For example, the following patterns (14) and (15) were learned for the relation *marriage*. The semantic terms in them may in some cases indeed indicate an embedded mention of this relation, but will usually not be of great utility to distinguish actual relation mentions from negative ones. These examples suggest that further work has to be invested into the creation of (stricter versions of) the relation-specific semantic graphs.

(14) person|SPOUSE $\xrightarrow{conj}$ person|SPOUSE $\xrightarrow{nn}$ partner| girlfriend

(15) relationship $\xrightarrow{prep}$ with $\xrightarrow{pobj}$ person SPOUSE $\xrightarrow{conj}$ person SPOUSE

7 CONCLUSION & FUTURE WORK

By our experiment we could demonstrate that apparent shortcomings of the structural rule-based approach could be overcome by adding lexical semantics to the rule discovery process. Although it may seem at first glance that the resulting extended rule induction mirrors the function of trigger word approaches, actually the effects of the additional terms is tamed through the structural constraints of the parse tree. Remember the example in (6): `Brad Pitt celebrated a wonderful wedding with Jennifer Aniston`. The rule induced from (6) would not extract a marriage between two cardinals from the following sentence in (16), while a statistical trigger word approach might well do this.

(16) `Cardinal Anderson celebrated after a wonderful wedding ceremony the holy communion with his German colleague Marx.`

For future work we suggest a more extensive evaluation of the impact of the new rules licensed by semantically relevant terms. It may well be that this set could be reduced again by structural distance or other structural constraints further improving precision without hurting recall. Another opportunity for improvement could be a more sophisticated treatment of different lexical semantic relations in the compilation of semantically related terms from BabelNet.

ACKNOWLEDGEMENT

REFERENCES

Agichtein, E. (2006). Confidence estimation methods for partially supervised information extraction. In *Proc. of the Sixth SIAM International Conference on Data Mining*.

Alfonseca, E., Filippova, K., Delort, J.-Y., and Garrido, G. (2012). Pattern learning for relation extraction with a hierarchical topic model. In *Proc. of ACL (2)*, pages 54–59.

Appelt, D. E. and Israel, D. J. (1999). Introduction to information extraction technology. A tutorial prepared for IJCAI-99.

Banko, M. and Etzioni, O. (2008). The Tradeoffs Between Open and Traditional Relation Extraction. In *Proc. of ACL/HLT*, pages 28–36.

Bollacker, K. D., Evans, C., Paritosh, P., Sturge, T., and Taylor, J. (2008). Freebase: a collaboratively created graph database for structuring human knowledge. In *Proc. of SIGMOD*, pages 1247–1250.

Bond, F. and Kyonghee, P. (2012). A survey of wordnets and their licenses. In *Proceedings of the 6th International Global WordNet Conference*, pages 64–71.

Bunescu, R. C. and Mooney, R. J. (2005). A Shortest Path Dependency Kernel for Relation Extraction. In *Proc. of HLT*, pages 724–731.

Chowdhury, M. F. M. and Lavelli, A. (2012). Combining tree structures, flat features and patterns for biomedical relation extraction. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, EACL '12, pages 420–429, Stroudsburg, PA, USA. Association for Computational Linguistics.

Etzioni, O., Fader, A., Christensen, J., Soderland, S., and Mausam (2011). Open Information Extraction: The Second Generation. In *Proc. of IJCAI*, page 310.

Fader, A., Soderland, S., and Etzioni, O. (2011). Identifying Relations for Open Information Extraction. In *Proc. of EMNLP*, page 15351545.

Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. MIT Press.

Grishman, R. and Sundheim, B. (1996). Message understanding conference - 6: A brief history. In *Proc. of the 16th International Conference on Computational Linguistics*, Copenhagen.

Grishman, R., Westbrook, D., and Meyers, A. (2005). Nyu's english ace 2005 system description. Technical report, Proteus Project, Department of Computer Science, New York University.

Jean-Louis, L., Besanon, R., Ferret, O., and Durand, A. (2013). Using Distant Supervision for Extracting Relations on a Large Scale. In Fred, A., Dietz, J., Liu, K., and Filipe, J., editors, *Knowledge Discovery, Knowledge Engineering and Knowledge Management*, volume 348 of *Communications in Computer and Information Science*, page 141155. Springer Berlin Heidelberg.

Krause, S., Li, H., Uszkoreit, H., and Xu, F. (2012). Large-scale learning of relation-extraction rules with distant supervision from the web. In *Proc. of 11th ISWC, Part I*, pages 263–278.

Li, H., Krause, S., Xu, F., Uszkoreit, H., Hummel, R., and Mironova, V. (2014). Annotating relation mentions in tabloid press. In *Proceedings of the 9th edition of the Language Resources and Evaluation Conference*.

Mausam, Schmitz, M., Soderland, S., Bart, R., and Etzioni, O. (2012). Open Language Learning for Information Extraction. In *Proc. of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 523–534, Jeju Island, Korea. Association for Computational Linguistics.

Min, B., Grishman, R., Wan, L., Wang, C., and Gondek, D. (2013). Distant supervision for relation extraction with an incomplete knowledge base. In *Proceedings of NAACL-HLT*, pages 777–782.

Mintz, M., Bills, S., Snow, R., and Jurafsky, D. (2009). Distant supervision for relation extraction without labeled data. In *Proc. of ACL/AFNLP*, page 10031011.

Moro, A., Li, H., Krause, S., Xu, F., Navigli, R., and Uszkoreit, H. (2013). Semantic rule filtering for web-scale relation extraction. In *International Semantic Web Conference (1)*, pages 347–362.

Moro, A. and Navigli, R. (2013). Integrating syntactic and semantic analysis into the open information extraction paradigm. In *Proc. of IJCAI*, pages 2148–2154.

Moro, A., Raganato, A., and Navigli, R. (2014). Entity linking meets word sense disambiguation: A unified approach. *Transactions of the Association for Computational Linguistics*, 2:231–244.

Navigli, R. (2009). Word Sense Disambiguation: A survey. *ACM Comput. Surv.*, 41(2):1–69.

Navigli, R. and Ponzetto, S. P. (2012). BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250.

Nivre, J., Hall, J., Nilsson, J., Chanev, A., Eryigit, G., Kübler, S., Marinov, S., and Marsi, E. (2007). Maltparser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering*, 13(2):95–135.

Ravichandran, D. and Hovy, E. H. (2002). Learning surface text patterns for a Question Answering System. In *Proc. of ACL*, pages 41–47.

Wu, F. and Weld, D. S. (2010). Open information extraction using wikipedia. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 118–127. Association for Computational Linguistics.

Xu, F., Uszkoreit, H., and Li, H. (2007). A seed-driven bottom-up machine learning framework for extracting relations of various complexity. In *Proc. of ACL*.

Xu, H., Hu, C., and Shen, G. (2009). Discovery of dependency tree patterns for relation extraction. In *PACLIC*, pages 851–858.

Xu, Y., Kim, M.-Y., Quinn, K., Goebel, R., and Barbosa, D. (2013). Open Information Extraction with Tree Kernels. In *Proc. of NAACL-HLT*, pages 868–877, Atlanta, Georgia. Association for Computational Linguistics.

Yangarber, R., Grishman, R., and Tapanainen, P. (2000). Automatic acquisition of domain knowledge for information extraction. In *Proc. of COLING*, pages 940–946.

Zelenko, D., Aone, C., and Richardella, A. (2003). Kernel methods for relation extraction. *The Journal of Machine Learning Research*, 3:1083–1106.