Generating Pattern-Based Entailment Graphs for Relation Extraction

Kathrin Eichler, Feiyu Xu, Hans Uszkoreit, Sebastian Krause
German Research Center for Artificial Intelligence (DFKI)
Berlin, Germany
`firstname.lastname@dfki.de`

Abstract

Relation extraction is the task of recognizing and extracting relations between entities or concepts in texts. A common approach is to exploit existing knowledge to learn linguistic patterns expressing the target relation and use these patterns for extracting new relation mentions. Deriving relation patterns automatically usually results in large numbers of candidates, which need to be filtered to derive a subset of patterns that reliably extract correct relation mentions. We address the pattern selection task by exploiting the knowledge represented by entailment graphs, which capture semantic relationships holding among the learned pattern candidates. This is motivated by the fact that a pattern may not express the target relation explicitly, but still be useful for extracting instances for which the relation holds, because its meaning *entails* the meaning of the target relation. We evaluate the usage of both automatically generated and gold-standard entailment graphs in a relation extraction scenario and present favorable experimental results, exhibiting the benefits of structuring and selecting patterns based on entailment graphs.

1 Introduction

The task of relation extraction (RE) is to recognize and extract relations among entities or concepts mentioned in texts. One common approach to RE is to learn and exploit extraction patterns (e.g., based on syntactic dependency trees), which express the targeted semantic relations. In order to circumvent the manual creation of patterns, numerous approaches have been investigated to derive patterns automatically. Automatic methods generally induce large numbers of unique candidate patterns, which only potentially express the target relation and need to be filtered in order to derive a subset of high-quality patterns for the relation extraction task. The task of filtering or selecting patterns can be tackled in various ways, e.g., based on frequency information, or by applying syntactic or semantic criteria.

For many RE applications, such as knowledge base population, patterns are not only relevant if they express the target relation explicitly, but also if they extract facts from which the target relation can be inferred. For example, all patterns below can be utilized for extracting pairs of people who are or were involved in a marriage relation:

> P1: PERSON_1 <marry> PERSON_2[1]
>
> P2: PERSON_1 <be widow of> PERSON_2
>
> P3: PERSON_1 <divorce from> PERSON_2

However, only pattern P1 expresses the target relation explicitly. Patterns P2 and P3 are semantically different from P1, but express a fact that *entails* the marriage relation[2]. As being aware of these semantic relationships holding among patterns can be of help in the pattern selection process, we propose to capture and exploit these relationships using pattern-based entailment graphs and show how technology from the area of recognizing textual entailment can be adapted to automatically generate these graphs. Finally, we apply the generated knowledge for relation extraction.

---

[1] PERSON_x refers to a slot filler for a person recognized in the input from which the pattern was extracted, <text> refers to a normalized form of the text part of the extracted pattern.

[2] We ignore tense aspects for the time being, which is also the approach taken in the RTE challenges (Dagan et al., 2005).

2   Related Work

The task of estimating the quality of automatically learned extraction patterns has been dealt with in various ways, for example based on integrity constraints (Agichtein, 2006), frequency heuristics (Krause et al., 2012) or lexical semantic criteria (Moro et al., 2013). Another line of research in RE groups similar patterns, e.g., by merging patterns based on syntactic criteria (Banko et al., 2007; Shinyama and Sekine, 2006; Thomas et al., 2011; Angeli et al., 2015), by clustering patterns that are semantically related (Kok and Domingos, 2008; Yates and Etzioni, 2009; Yao et al., 2011), or by identifying patterns associated to a given seed relation (Bauer et al., 2014). Such approaches help gain generalization; however, their ability to express semantic relationships is limited, as they cannot capture the asymmetric nature of these relationships. For example, clustering can help us identify pattern P4 below as being semantically related to patterns P1 to P3 in section 1.

P4: PERSON$_1$ <love> PERSON$_2$

However, it falls short of expressing that two entities linked by patterns P1 to P3 are mentions of the marriage relation, whereas this is not necessarily true of entities linked by pattern P4. Similarly, clustering can identify patterns P1 and P3 as semantically related. However, it cannot express that the relation expressed by pattern P3 entails the relation expressed by pattern P1, but not vice versa. These asymmetric relationships have been considered by Riedel et al. (2013), who learns latent feature vectors for patterns based on matrix factorization, and have also been studied extensively in the context of recognizing textual entailment (RTE). RTE is the task of determining, for two textual expressions T (text) and H (hypothesis), whether the meaning of H can be inferred from the meaning of T (Dagan and Glickman, 2004). In RE, RTE systems have been applied to validate a given relation instance (Wang and Neumann, 2008) and to extract instances entailing a given target relation (Romano et al., 2006; Bar-Haim et al., 2007; Roth et al., 2009).

As illustrated above, RE can clearly benefit from considering semantic relationships holding among extraction patterns. However, previous work in RE has either focussed on grouping related patterns without considering non-symmetric relations, or, on computing entailment decisions for individual T/H pairs. We propose to exploit entailment relationships holding among RE patterns by structuring the candidate set in an *entailment graph*. Entailment graphs are hierarchical structures representing entailment relations among textual expressions and have previously been generated for various types of expressions (Berant et al., 2010, 2012; Mehdad et al., 2013; Levy et al., 2014; Kotlerman et al., 2015). Entailment graphs can be constructed by determining entailment relationships between pairs of expressions or, as proposed by Kolesnyk et al. (2016), by generating entailed sentences from source sentences. Our work of building entailment graphs based on RE patterns is related to the work by Nakashole et al. (2012), who create a taxonomy of binary relation patterns. For their syntactic patterns, they compute partial orders of generalization and subsumption based on the set of mentions extracted by each pattern. In contrast to their work, we construct pattern-based entailment graphs using RTE technology. This is motivated by the fact that entailment is semantic and not mention-based, i.e., one pattern can entail another pattern even if they extract disjoint sets of mentions in a given text corpus.

3   Entailment Graph Generation

3.1   Pattern-Based Entailment Graphs

A pattern-based entailment graph refers to a directed graph, in which each node represents a unique RE pattern, and each edge ($\rightarrow$) denotes an entailment relationship. Bidirectional edges ($\leftrightarrow$) denote that the patterns represented by the two nodes are considered semantically equivalent. A sample subgraph for the *marriage* relation is given in Figure 1[3], which shows all entailment relations with respect to the pattern [PERSON$_1$ <marry> PERSON$_2$]. Automatic entailment graph generation is usually performed in two steps: First, entailment decisions for individual T/H pairs are computed (using an RTE engine); second, an optimization strategy is applied to derive a consistent, transitive graph (Berant et al., 2010).

3.2   RTE Engine

For recognizing entailment relations between individual T/H pairs of patterns, we make use

---

[3]For reasons of simplicity, the figure shows the text representation of the patterns, which are in fact represented as dependency structures. Since entailment is transitive, all edges are omitted that can be recovered in the transitive closure.
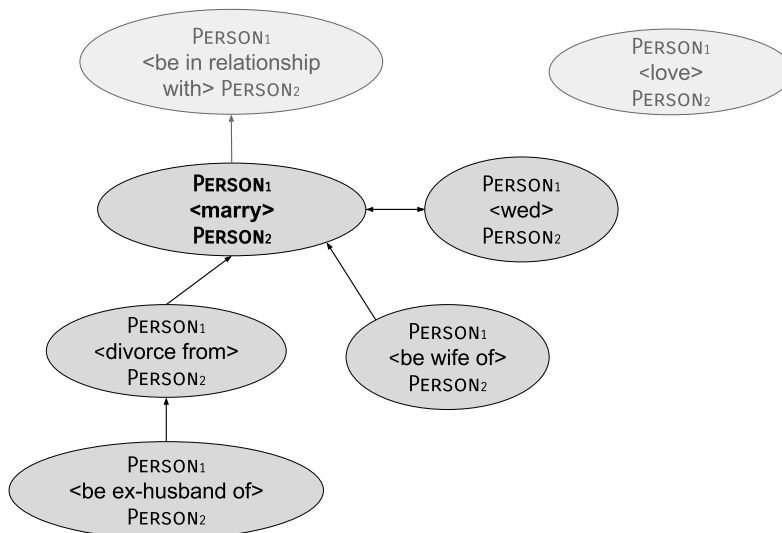
Figure 1: Subgraph showing entailment relations for the pattern [PERSON₁ <marry> PERSON₂]

of an RTE engine based on multi-level alignments. This RTE engine, referred to as *MultiAlign*, is available through the RTE platform EXCITEMENT (Magnini et al., 2014) and achieved state-of-the-art performance on several RTE corpora (Noh et al., 2015). We opted for this RTE system because it makes use of external knowledge resources and, unlike more recent systems based on neural networks (Rocktäschel et al., 2015; Bowman et al., 2015), is able to cope with the restricted amount of training data available for the task. *MultiAlign* uses shallow parsing for linguistic preprocessing and logistic regression for entailment classification. Features for the classifier are generated on the basis of multi-level alignments using four aligners: a lemma aligner (aligning identical lemmas found in T and H), an aligner based on the paraphrase tables provided by the METEOR MT evaluation package (Denkowski and Lavie, 2014), and two lexical aligners based on Wordnet (Fellbaum, 1998)[4] and VerbOcean (Chklovski and Pantel, 2004). As output, it produces a binary decision (entailment, non-entailment) along with a computed confidence score.

As the RTE engine was originally designed for sentences, rather than patterns, we converted each pattern into its textual representation. The variables expressing type and semantic role of the entities linked by the pattern were excluded in this rep-

resentation, as the resulting variable alignments would skew the RTE engine's entailment decision. For our experiments, we used the original *MultiAlign* implementation as well as an adapted version, in which we made some changes to the WordNet aligner. In particular, unlike in the original implementation, which only considered the first sense of each WordNet entry, we extended this to cover all senses. This allowed us to retrieve additional relevant alignments such as *wed ↔ marry*. In addition, rather than retrieving rules for all words in T, we only consider rules for full verbs, nouns, and adjectives. This way, we particularly filter out rules for auxiliary verbs, which tend to produce irrelevant alignments, especially when considering all senses. A sample set of decisions produced by our RTE engine among candidate patterns for the *marriage* relation is depicted in Figure 2.

3.3 Graph Optimization

Automatically derived entailment decisions may contradict each other. For example, as illustrated in Figure 2, our RTE engine correctly decides that [PERSON₁ <spouse of> PERSON₂] → [PERSON₁ <'s marriage to> PERSON₂] and that [PERSON₁ <'s marriage to> PERSON₂] → [PERSON₁ <marry> PERSON₂]. However, it misses the entailment relation between [PERSON₁ <spouse of> PERSON₂] and [PERSON₁ <marry> PERSON₂], because the relationship between *spouse* and *marry* is not covered by the semantic resources underlying the system. This leads

---

[4]Relations considered by the WordNet aligner: synonym, derivationally related, (instance) hypernym, member / part holonym, entailment, and substance meronym.
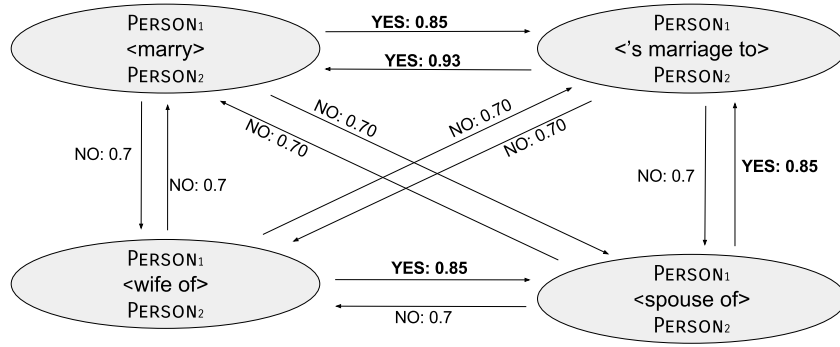
Figure 2: Sample set of RTE decisions (YES: entailment, NO: no entailment) with associated confidence

Figure 3: Sample outputs using greedy (left) and global (right) graph optimizer

to a set of decisions that is invalid given the transitivity of the entailment relation. For deriving a consistent graph, we applied two different strategies: First, a simple greedy strategy that assumes each computed positive entailment relation with a confidence exceeding a pre-defined threshold to be valid, and adds missing entailment edges to ensure transitive closure. Second, the global graph optimization algorithm by Berant et al. (2012), which searches for the best graph under a global transitivity constraint, approaching the optimization problem by Integer Linear Programming. The selection of the optimization strategy is crucial, as illustrated in Figure 3, which shows two sample outputs from each of the two strategies for the decisions in Figure 2.

4 Applying Pattern-Based Entailment Graphs for Relation Extraction

In order to exploit entailment graphs for relation extraction, we propose the following approach, which is depicted in Figure 4:

1. Create a set of candidate extraction patterns P (applying any method of choice).

2. Generate an entailment graph EG expressing entailment relations among the patterns in P.

3. Choose a base pattern[5], expressing the target relation explicitly and select all patterns entailing the base pattern according to EG.

4. Apply the selected patterns to extract relation mentions.

Given the sample graph in Figure 1 and the base pattern H: [PERSON$_1$ <marry> PERSON$_2$], our method would select all patterns entailing H, either directly or via transitivity, e.g., [PERSON$_1$ <be ex-husband of> PERSON$_2$], including patterns considered semantically equivalent, such as [PERSON$_1$ <wed> PERSON$_2$]. It would neither select [PERSON$_1$ <love> PERSON$_2$], as it has no entailment relation to H, nor [PERSON$_1$ <be in relationship with> PERSON$_2$], as it is entailed by, but not equivalent to H.

---

[5]Note that the selection of a base pattern can be done manually, but can also be automated. For example, for the relations at hand, the most frequent pattern candidate learned for a particular relation turned out to be an appropriate choice.
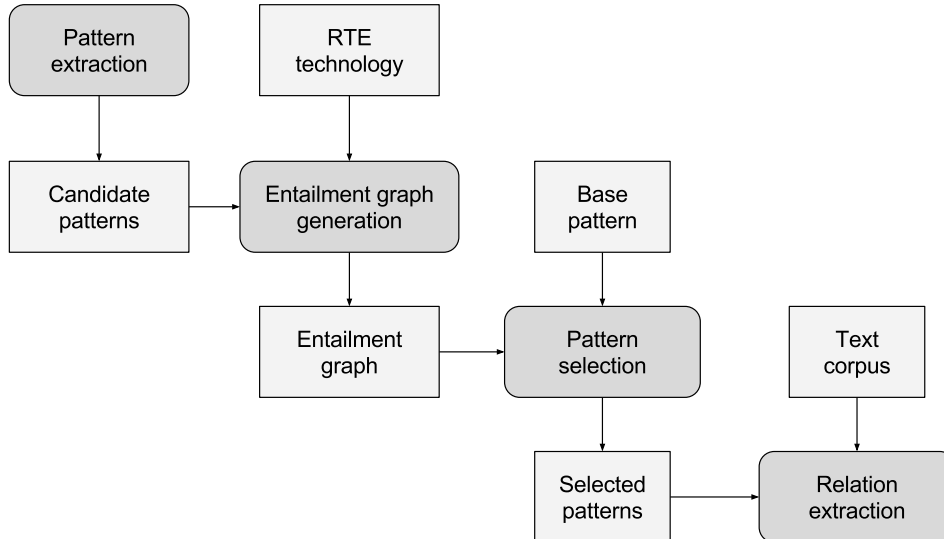
Figure 4: Procedure for relation extraction using pattern selection based on entailment graphs

5 Experiments

For evaluating our method on the relation extraction task, we conducted experiments on two freely available datasets: TEG-REP (Eichler et al., 2016) and FB15k-237 (Toutanova et al., 2015). On the TEG-REG dataset, we carried out a detailed evaluation of several pattern filtering strategies with respect to two semantic relations. On the FB15k-237 corpus, we evaluate the scalability of our method to other semantic relations.

5.1 TEG-REP

The TEG-REP corpus contains automatically derived relation extraction patterns as well as gold-standard entailment graphs created from these patterns for three relations typically considered in RE tasks: *marriage*, *acquisition*, and *award honor*. The patterns underlying this corpus are a subset of the patterns used by Moro et al. (2013) and were acquired automatically using the pattern discovery system by Krause et al. (2012). The system derives candidate patterns from dependency-parsed sentences extracted using distant supervision based on relation instances from the fact knowledge base Freebase (Bollacker et al., 2008). The TEG-REP corpus is the only available corpus of pattern-based entailment graphs and particularly suitable for our evaluation because it allows for a comparison of patterns selected based on both manually and automatically created entailment graphs. For our experiments on this corpus, we divided the full set of patterns in the corpus (around 500 per relation) into two equally-sized

portions, one for creating training data for the RTE engine, and one for evaluating pattern selection methods. For creating an evaluation dataset, we applied all patterns in the evaluation split to 14.5 million ClueWeb sentences (Lemur Project, 2009) linked to Freebase entities (Gabrilovich et al., 2013), and manually annotated around 3000 of the extracted mentions.[6] A mention was annotated as being correct if we found evidence for the target relation between the entities in the mention to hold. Evidence was drawn either from the source sentence itself, or, in cases were the source sentence did not express the relation explicitly, from external resources such as Freebase or Wikipedia.

Our experiments on the TEG-REP dataset are based on the relations *marriage* and *acquisition*[7]. For our experiments, we split the evaluation set into a development set for optimizing the graph building parameters and a test set for the final evaluation. In our experiments, we tested several strategies for selecting patterns and measured performance over the annotated relation mentions in the evaluation dataset. For evaluating the graph-based methods, we selected all patterns entailing the base patterns [PERSON$_1$ <marry> PERSON$_2$] (for *marriage*) and [ORGANIZATION$_1$ <acquire>

---

[6]The annotation was done by three annotators. About 10% of the mentions were annotated by two annotators in parallel, who achieved a very high interannotator agreement (Cohens Kappa > 0.9). The remaining mentions were annotated by a single person.

[7]We did not evaluate the *award honor* relation because the vast majority (> 98%) of mentions extracted using these patterns were correct, which would not have allowed for a meaningful evaluation.

ORGANIZATION$_2$] (for *acquisition*). In order to investigate the benefits of the graph structure, we compared the results to those achieved when computing entailment relations at a pair-wise level, i.e., using the base pattern of the relation as H and all other candidate patterns for the relation as T. We also applied the approach by Moro et al. (2013), who identify relation-relevant word senses based on lexical semantic subgraphs derived from BabelNet and filter out patterns not containing any relevant word sense. Based on a parameter k, they consider a word sense to be relevant, if it is at most k-step distant to the core word sense for the target relation.

5.2 FB15k-237

As training the RTE models requires appropriate training data, which may not be available, we ran additional experiments to investigate if the models trained on T/H pairs for one relation are general enough to be used for computing entailment relations among pattern candidates for other semantic relations. To this end, we used the FB15k-237 corpus (Toutanova et al., 2015), which contains knowledge-base relation triples and textual mentions of Freebase entity pairs. For our experiments on this corpus, we generated candidate patterns by extracting the first 1000 tuples matching a particular relation from the pattern files in the corpus, and then extracting all patterns linking any of the tuples in the textual triples used by Toutanova and Chen (2015). This way, our candidate pattern set contains both patterns expressing the target relation as well as patterns expressing other relations. For creating the entailment graph, we converted all patterns into a textual representation, removed patterns with no lexical item, and, from the remaining patterns, built an entailment graph applying the RTE engine described in 3.2 with the model trained on the *marriage* relation and the best parameter setting derived based on the TEG-REP corpus. For evaluating the result, we selected 10 relations, defined a base pattern for each of them, and checked, for each pattern in the graph, whether it entailed the base pattern according to the graph structure and whether the entailment decision was correct based on the semantics expressed by the pattern.

As in this setting, we evaluated the entailment relations expressed by the pattern graph rather than the usage of the patterns for relation extrac-

tion, the results are not directly comparable to the figures obtained on the TEG-REP corpus, but still allow for an assessment of the quality of the selected patterns.

6 Results and Discussion

6.1 TEG-REP

Table 1 shows results on the TEG-REP corpus and contains, for each of the following pattern selection methods, the computed precision, recall, and F1 scores:

- *All patterns* All patterns from the test split (baseline).

- *Lexical semantic filter (Moro et al., 2013)* Patterns selected using the lexical semantic filter.

- *Pair-wise entailment (MultiAlignAdapted)* Patterns selected based on pair-wise entailment decisions using the model of *MultiAlignAdapted*.

- *Entailment Graph (MultiAlignOriginal / MultiAlignAdapted)* Patterns selected based on entailment graphs generated with the original / adapted *MultiAlign* implementation.

- *Entailment Graph (TEG-REP gold standard)* Patterns selected based on gold-standard entailment graphs from the TEG-REP corpus.

For the lexical semantic filter method, we experimented with different levels of k and noted down the value achieving the best F1 score. The results in the table were produced setting k to 1 for the *marriage* relation and k to 5 for the *acquisition* relation. For the RTE-based methods, we experimented with the two different graph optimization strategies and, for each of them, with different confidence threshold values, and optimized these parameters based on the development split. The figures in the table show the results achieved on the test split using the parameter setting optimized on the development set: the greedy optimization strategy with thresholds of 0.71 (*MultiAlignOriginal*) and 0.77 (*MultiAlignAdapted*) for the *marriage* relation and thresholds of 0.74 (*MultiAlignOriginal*) and 0.75 (*MultiAlignAdapted*) for the *acquisition* relation. On our data, the greedy edge selection strategy produced better results than the global graph optimizer for both relations. This was because the global strategy, even with low confidence thresholds, was more restrictive and removed too many edges from the graph, thus yielding lower recall figures.

Configuration	Precision	Recall	F1
All patterns	0.15	1.00	0.27
Lexical semantic filter (Moro et al., 2013)	0.61	0.73	0.67
Pair-wise entailment (MultiAlignAdapted)	0.97	0.56	0.71
Entailment Graph (MultiAlignOriginal)	0.96	0.59	0.73
Entailment Graph (MultiAlignAdapted)	0.96	0.68	**0.80**
Entailment Graph (TEG-REP gold standard)	0.96	0.69	0.80

Table 1: Results for *marriage* relation (TEG-REP corpus)

Configuration	Precision	Recall	F1
All patterns	0.30	1.00	0.46
Lexical semantic filter (Moro et al., 2013)	0.30	0.97	0.46
Pair-wise entailment (MultiAlignAdapted)	0.82	0.49	0.62
Entailment Graph (MultiAlignOriginal)	0.81	0.53	0.64
Entailment Graph (MultiAlignAdapted)	0.59	0.93	**0.73**
Entailment Graph (TEG-REP gold standard)	0.82	0.49	0.62

Table 2: Results for *acquisition* relation (TEG-REP corpus)

For both relations, the best overall results were achieved using our proposed method based on entailment graphs generated automatically applying the adapted RTE engine. The results show that entailment-based pattern selection is in fact more powerful than the lexical semantic filter. It selects patterns yielding a much higher precision because it is able to successfully filter out non-entailing patterns, such as [PERSON$_1$ <be in relationship with> PERSON$_2$] for the *marriage* relation, which are wrongly selected using the lexical semantic filter. For the *marriage* relation, the results not only show that our RTE engine adaptations yielded a much higher recall (with almost no loss in precision) than the original implementation (thanks to an increased number of relevant alignments), but also that pattern selection can in fact benefit from the graph structure: Entailment graphs created using *MultiAlignAdapted* achieved much better performance than a selection based on pair-wise entailment computation using the same RTE model. This was due to a higher recall achieved because the graph structure allowed the algorithm to identify entailment relations that involved the combination of several inference steps and were missed when applying RTE in a pair-wise manner. An example is the relationship between *wife* and *marry*, as shown below:

$$wife \xrightarrow{\text{hyper-nym}} spouse \xrightarrow{\text{member holonym}} marriage \xrightarrow{\text{deri-vation}} marry$$

For the *acquisition* relation, we noticed that the lexical semantic filter performed quite poorly on our corpus. The relation requires a large k-value, i.e., $k >= 5$, since there are many ways in which an acquisition can be described. A company can for instance devour, take-over or purchase another company. Each increase of k allows many additional content words, thus increasing the danger of inappropriate ones. An example is [ORGANIZATION$_1$ <trademark of> ORGANIZATION$_2$]. Where it is plausible that in the training set, an acquired company may persist as a brand of its new owner, *trademark* does not express a take-over. Although the semantic filter by Moro et al. (2013) can provide useful hints and can be applied without manually annotating training data, it is not powerful enough to discriminate content words as to whether they provide strong evidence for an acquisition or not.

Also patterns selected based on the entailment graph gold-standard performed surprisingly low on the *acquisition* relation. Here, recall was affected negatively because some of the non-entailing patterns that were filtered out were in fact able to extract correct instances with good precision. In particular, patterns expressing a planned acquisition, such as [ORGANIZATION$_1$ <plan to purchase> ORGANIZATION$_2$], [ORGANIZATION$_1$ <be to acquire> ORGANIZATION$_2$], or [ORGANIZATION$_1$ <announce intention to acquire> ORGANIZATION$_2$] extracted many cor-

Relation	Base pattern	Precision	Recall	F1
award-award_honor-ceremony	win at	0.58 (0.31)	0.73 (1.00)	**0.65** (0.47)
base-locations-continents-countries_within	country in	0.67 (0.57)	0.84 (1.00)	**0.75** (0.59)
education-education-major_field_of_study	degree in	1.00 (0.47)	0.49 (1.00)	**0.65** (0.64)
film-...-film_regional_debut_venue	premiere at	0.73 (0.13)	0.89 (1.00)	**0.80** (0.23)
film-performance-film	star in	0.97 (0.57)	0.53 (1.00)	0.69 (**0.72**)
organization-place_founded	founded in	0.83 (0.16)	0.83 (1.00)	**0.83** (0.27)
people-marriage-location_of_ceremony	marry in	0.80 (0.07)	1.00 (1.00)	**0.89** (0.12)
people-marriage-spouse	marry	1.00 (0.15)	0.58 (1.00)	**0.73** (0.26)
people-person-place_of_birth	born in	1.00 (0.41)	0.84 (1.00)	**0.91** (0.58)
people-place_of_burial	buried at	1.00 (0.33)	0.71 (1.00)	**0.83** (0.50)

Table 3: Entailment graph based pattern selection (vs. baseline) for FB15k-237 relations

rect mentions, as the acquisition in fact happened at a later stage. Nevertheless, filtering out these cases is correct from a semantic point of view, even if many of the reported plans or attempts concerning acquisitions later become a reality.

Precision on the *acquisition* gold-standard was also lower than for the *marriage* relation, due to patterns annotated as entailing in the TEG-REP corpus, which extracted comparably many incorrect instances. One such pattern is [ORGANIZATION$_1$ <takeover of> ORGANIZATION$_2$], which yields low precision values because it often occurs in sentences expressing irrealis moods, such as *the proposed Microsoft takeover of Yahoo* or *is a Pfizer takeover of BMS realistic?*, and because of its generality often extracts non-company entities, e.g., *Republican takeover of Congress*. Detecting the embedding of correct patterns in irrealis contexts is a largely unsolved problem and calls for the development of general methods for recognizing nonfactual modalities along the lines of the NegEx algorithm for detecting negations in medical texts (Chapman et al., 2001) and its later extensions.

6.2 FB15k-237

Our experiments on the FB15k-237 corpus are presented in Table 3, showing the performance of our pattern selection method based on entailment graphs (with the adapted *MultiAlign* implementation) compared to a simple baseline (all patterns). The results show that, even using an RTE model trained on a completely different semantic relation, our method achieves decent performance on selecting meaningful patterns for a wide range of relations. The figures in the table were produced with the simple graph optimization strategy, but the global graph optimizer performed very similar on this dataset, achieving the same results for eight out of the ten relations. It performed

worse on the *award-award_honor-ceremony* relation (F1: 0.48), and better for the *base-locations-continents-countries_within* relation (F1: 0.91). Nevertheless, when dealing with larger numbers of patterns, the global graph optimizer should be the method of choice, as it is less prone to semantic drift.

7 Conclusions and Future Work

We presented an approach for structuring relation extraction patterns using entailment graphs and evaluated the usefulness of these graphs for pattern selection. For generating entailment graphs automatically, we employed and adapted an alignment-based entailment classifier, which makes use of external knowledge resources, and experimented with different graph optimization strategies. Our classifier was trained on a manageable amount of annotated patterns for a single semantic relation, resulting in a generic model that was shown to produce valid entailment decisions for a wide range of other semantic relations. Our experimental results suggest that meaningful pattern-based entailment graphs can be constructed automatically and that the derived knowledge is in fact valuable for selecting useful relation extraction patterns. In particular, entailment graph based filtering can help achieve higher precision than methods which do not take into account the asymmetric nature of semantic relations.

Acknowledgments

References

Eugene Agichtein. 2006. Confidence estimation methods for partially supervised information extraction. In *Proceedings of the Sixth SIAM International Conference on Data Mining*. Bethesda, MD, USA, pages 539–543.

Gabor Angeli, Melvin Jose Johnson Premkumar, and Christopher D. Manning. 2015. Leveraging linguistic structure for open domain information extraction. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*. Association for Computational Linguistics, Beijing, China, pages 344–354.

Michele Banko, Michael J. Cafarella, Stephen Soderland, Matt Broadhead, and Oren Etzioni. 2007. Open Information Extraction from the Web. In *Proceedings of the 20th International Joint Conference on Artifical Intelligence*. pages 2670–2676.

Roy Bar-Haim, Ido Dagan, Iddo Greental, and Eyal Shnarch. 2007. Semantic Inference at the Lexical-Syntactic Level. In *Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence*. AAAI Press, Vancouver, British Columbia, Canada, pages 871–876.

Sandro Bauer, Stephen Clark, Laura Rimell, and Thore Graepel. 2014. Learning a theory of marriage (and other relations) from a web corpus. In *Advances in Information Retrieval*, Springer International Publishing, pages 591–597.

Jonathan Berant, Ido Dagan, Meni Adler, and Jacob Goldberger. 2012. Efficient tree-based approximation for entailment graph learning. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*. Jeju Island, Korea, pages 117–125.

Jonathan Berant, Ido Dagan, and Jacob Goldberger. 2010. Global learning of focused entailment graphs. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. Uppsala, Sweden, pages 1220–1229.

Kurt D. Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the ACM SIGMOD international conference on Management of data*. pages 1247–1250.

Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning. 2015. A large annotated corpus for learning natural language inference pages 632–642.

Wendy W. Chapman, Will Bridewell, Paul Hanbury, Gregory F. Cooper, and Bruce G. Buchanan. 2001. A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of Biomedical Informatics* 34(5):301 – 310.

Timothy Chklovski and Patrick Pantel. 2004. VerbOcean: Mining the Web for Fine-Grained Semantic Verb Relations. In *Proceedings of EMNLP*. Barcelona, Spain, pages 33–40.

Ido Dagan and Oren Glickman. 2004. Probabilistic Textual Entailment: Generic Applied Modeling of Language Variability. In *Learning Methods for Text Understanding and Mining*.

Ido Dagan, Oren Glickman, and Bernardo Magnini. 2005. The PASCAL Recognising Textual Entailment Challenge. In *Proceedings of the PASCAL Challenges Workshop on Recognising Textual Entailment*.

Michael Denkowski and Alon Lavie. 2014. Meteor Universal: Language Specific Translation Evaluation for Any Target Language. In *Proceedings of the Ninth Workshop on Statistical Machine Translation*.

Kathrin Eichler, Feiyu Xu, Hans Uszkoreit, Leonhard Hennig, and Sebastian Krause. 2016. TEG-REP: A corpus of textual entailment graphs based on relation extraction patterns. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC)*.

Christiane Fellbaum. 1998. *WordNet: an electronic lexical database*. MIT Press, Cambridge, MA, USA.

Evgeniy Gabrilovich, Michael Ringgaard, and Amarnag Subramanya. 2013. FACC1: Freebase annotation of ClueWeb corpora, Version 1 (Release date 2013-06-26, Format version 1, Correction level 0).

Stanley Kok and Pedro Domingos. 2008. Extracting semantic networks from text via relational clustering. In *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases*. pages 624–639.

Vladyslav Kolesnyk, Tim Rocktäschel, and Sebastian Riedel. 2016. Generating natural language inference chains. *CoRR* abs/1606.01404.

Lili Kotlerman, Ido Dagan, Bernardo Magnini, and Luisa Bentivogli. 2015. Textual entailment graphs. *Natural Language Engineering* 21:699–724.

Sebastian Krause, Hong Li, Hans Uszkoreit, and Feiyu Xu. 2012. Large-scale learning of relation-extraction rules with distant supervision from the web. In *Proceedings of the 11th International Semantic Web Conference*. Springer.

The Lemur Project. 2009. Clueweb09. http://www.lemurproject.org.

Omer Levy, Ido Dagan, and Jacob Goldberger. 2014. Focused Entailment Graphs for Open IE Propositions. In *Proceedings of the Eighteenth Conference on Computational Natural Language Learning*. Ann Arbor, Michigan, pages 87–97.

Bernardo Magnini, Roberto Zanoli, Ido Dagan, Kathrin Eichler, Günter Neumann, Tae-Gil Noh, Sebastian Padó, Asher Stern, and Omer Levy. 2014. The Excitement Open Platform for Textual Inferences. In *Proceedings of the ACL 2014 System Demonstrations*. ACL.

Yashar Mehdad, Giuseppe Carenini, Raymond T. Ng, and Shafiq R. Joty. 2013. Towards topic labeling with phrase entailment and aggregation. In *Proceedings of the Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics*. Atlanta, Georgia, USA, pages 179–189.

Andrea Moro, Hong Li, Sebastian Krause, Feiyu Xu, Roberto Navigli, and Hans Uszkoreit. 2013. Semantic rule filtering for web-scale relation extraction. In *Proceedings of the 12th International Semantic Web Conference*. pages 347–362.

Ndapandula Nakashole, Gerhard Weikum, and Fabian M. Suchanek. 2012. PATTY: A taxonomy of relational patterns with semantic types. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. Jeju Island, Korea, pages 1135–1145.

Tae-Gil Noh, Sebastian Padó, Vered Shwartz, Ido Dagan, Vivi Nastase, Kathrin Eichler, Lili Kotlerman, and Meni Adler. 2015. Multi-level alignments as an extensible representation basis for textual entailment algorithms. In *Proceedings of *SEM 2015*. ACL.

Sebastian Riedel, Limin Yao, Andrew McCallum, and Benjamin M. Marlin. 2013. Relation Extraction with Matrix Factorization and Universal Schemas. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. pages 74–84.

Tim Rocktäschel, Edward Grefenstette, Karl Moritz Hermann, Tomás Kociský, and Phil Blunsom. 2015. Reasoning about entailment with neural attention. *CoRR* abs/1509.06664.

Lorenza Romano, Milen Kouylekov, Idan Szpektor, Ido Dagan, and Alberto Lavelli. 2006. Investigating a generic paraphrase-based approach for relation extraction. In *Proceedings of the 11st Conference of the European Chapter of the Association for Computational Linguistics*. Trento, Italy.

Dan Roth, Mark Sammons, and V. G. Vinod Vydiswaran. 2009. A framework for entailed relation recognition. In *Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing*. The Association for Computer Linguistics, Singapore, pages 57–60.

Yusuke Shinyama and Satoshi Sekine. 2006. Preemptive information extraction using unrestricted relation discovery. In *Proceedings of the Human Language Technology Conference of the NAACL*. New York City, USA, pages 304–311.

Philippe Thomas, Stefan Pietschmann, Illés Solt, Domonkos Tikk, and Ulf Leser. 2011. Not all links are equal: Exploiting Dependency Types for the Extraction of Protein-Protein Interactions from Text. In *Proceedings of the BioNLP Workshop*. pages 1–9.

Kristina Toutanova and Danqi Chen. 2015. Observed versus latent features for knowledge base and text inference. In *Proceedings of the Workshop on Continuous Vector Space Models and Their Compositionality*.

Kristina Toutanova, Danqi Chen, Patrick Pantel, Hoifung Poon, Pallavi Choudhury, and Michael Gamon. 2015. Representing text for joint embedding of text and knowledge bases.

Rui Wang and Günter Neumann. 2008. Relation validation via textual entailment. In *Proceedings of the 1st International and KI-08 Workshop on Ontology-based Information Extraction Systems*. volume 400, pages 26–37.

Limin Yao, Aria Haghighi, Sebastian Riedel, and Andrew McCallum. 2011. Structured relation discovery using generative models. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Stroudsburg, PA, USA, pages 1456–1466.

Alexander Yates and Oren Etzioni. 2009. Unsupervised methods for determining object and relation synonyms on the web. *Journal of Artificial Intelligence Research* 34:255–296.