Minimally Supervised Rule Learning for the Extraction of Biographic Information from Various Social Domains

Hong Li **Feiyu Xu** **Hans Uszkoreit**

German Research Center for Artificial Intelligence (DFKI), LT-Lab
Alt-Moabit 91c, 10559 Berlin, Germany
{lihong,feiyu,uszkoreit}@dfki.de
http://www.dfki.de/lt/

Abstract

This paper investigates the application of an existing seed-based minimally supervised learning algorithm to different social domains exhibiting different properties of the available data. A systematic analysis studies the respective data properties of the three domains including the distribution of the semantic arguments and their combinations. The experimental results confirm that data properties have a strong influence on the performance of the learning system. The main results are insights about: (i) the effects of data properties such as redundancy and frequency of argument mentions on coverage and precision (ii) the positive effects of negative examples if used effectively (iii) the different effects of negative examples depending on the domain data properties and (iv) the potential of reusing rules from one domain for improving the relation extraction performance in another domain.

1 Introduction

Domain adaptation is very important for information extraction (IE) systems. IE systems in the real world are often required to work for new domains and new tasks within a limited adaptation or tuning time. Thus, automatic learning of relation extraction rules for a new domain or a new task has been established as a relevant subarea in IE research and development (Muslea, 1999; Tsujii, 2000; Uszkoreit, 2011), in particular for minimally supervised or semi-supervised bootstrapping approaches (e.g., (Brin, 1998; Agichtein and Gravano, 2000; Yangarber, 2001; Sudo et al., 2003; Bunescu and Mooney, 2005; McDonald et al., 2005; Greenwood and Stevenson, 2006; Jones, 2005; Xu et al., 2007; Xu, 2007; Kozareva and Hovy, 2010a; Kozareva and Hovy, 2010b)). The advantage of the minimally supervised approaches for IE rule learning is that only initial seed knowledge is needed. Therefore the adaptation might be limited to substituting the seed examples. However, different domains/corpora exhibit rather different properties of their learning/extraction data with respect to the learning algorithm. Depending on the domain, the need for improving precision by utilizing negative examples may differ. An important research goal is the exploitation of more benign domains for improving extraction in less suitable domains.

Xu et al. (2007) and Xu (2007) present a minimally supervised learning system for relation extraction, initialized by a so-called semantic seed, i.e., examples of the target relations. We dub our system DARE for Domain Adaptive Relation Extraction. The system supports the domain adaptation with a compositional rule representation and a bottom-up rule discovery strategy. In this way, DARE can handle target relations of various complexities and arities. Relying on a few examples of a target relation as semantic seed dispenses with the costly acquisition of domain knowledge through experts or specialized resources.

In practice, this does not work equally well for any given domain. Xu (2007) and Uszkoreit et al. (2009) concede that DARE's performance strongly depends on the specific type of relation and domain. In our experiments, we apply DARE to the extraction of two different 4-ary relations from different domains (Nobel Prize awards and MUC-6 management succession events (Grishman and Sundheim, 1996)). In the data set of the first domain, the connectivity between relation instances and linguistic patterns (rules) approximates the small world property (Amaral et al., 2005). In MUC-6 data on the other hand, the redundancy of both mentions of instances and patterns as well as their connectivity are very low.

DARE achieves good performance with the first data set even with a singleton seed, but cannot deal nearly as well with the MUC-6 data.

A systematic comparative analyses was not possible since the two experiments differ in several dimensions: domain, relation, size of data sets, origin of data sets and the respective distribution of mentions in the data. In this paper, a much more systematic analysis is performed in order to understand the differences between domains represented by their respective data sets. We decide to use DARE because of its domain-adaptive design and because of its utilization of negative examples for improving precision (Uszkoreit et al., 2009). At the same time, this is the first study comparing the effects of the DARE utilization of negative examples relative to different domains. In order to secure the significance of the results, we restrict our experiments to one simple symmetric binary relation, i.e. the biographic relation "married to", a single text sort, i.e., Wikipedia articles, and three biographic domains exhibiting different data properties, i.e., entertainers, politicians and business people.

The three data sets are compared with respect to relation extraction performance with and without negative examples in relation to certain data properties. Furthermore, the potential for porting rules from one domain to another and the effects of merging domains are investigated. Our data analysis and experiments give us interesting insights into the relationship between the distribution of biographic information in various social domains and its influence on the learning and extraction task. Given the same target relation "married to", the entertainment domain contains most mentions and owns better data properties for learning than others. But, in the parallel, there are often multiple relations reporting about the same married couples in the entertainment domain, leading to the learning of spurious rules and finally bad precision.

The remainder of the paper is organized as follows: Section 2 explains the DARE system. In section 3, we represent our research idea and our experiments and evaluations. In section 4, we close off with summary and conclusion.

2 DARE

DARE is a minimally supervised machine learning system for relation extraction on free texts, consisting of two parts: 1) rule learning and 2) relation extraction (RE). Rule learning and RE feed each other in a bootstrapping framework. The bootstrapping starts from so-called "semantic seeds", which is a small set of instances of the target relation. The rules are extracted from sentences automatically annotated with semantic entity types and parsing results (e.g., dependency structures), which match with the seeds. RE applies acquired rules to texts in order to discover more relation instances, which in turn are employed as seed for further iterations. The core system architecture of DARE is depicted in Figure 1. The entire bootstrapping stops when no new rules or new instances can be detected. Relying entirely on semantic seeds as domain knowledge, DARE can accommodate new relation types and domains with minimal effort.

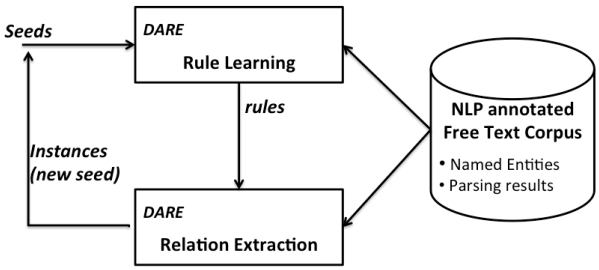


Figure 1: DARE core architecture

DARE can handle target relations of varying arity through a compositional and recursive rule representation and a bottom-up rule discovery strategy. A DARE rule for an *n*-ary relation can be composed of rules for its projections, namely, rules that extract a subset of the *n* arguments.

Let us consider an example target relation from (Xu, 2007). It contains prize award events at which a person or an organization wins a particular prize in a certain area and year. The relation can be presented as follows:

(1) *<recipient, prize, area, year>*

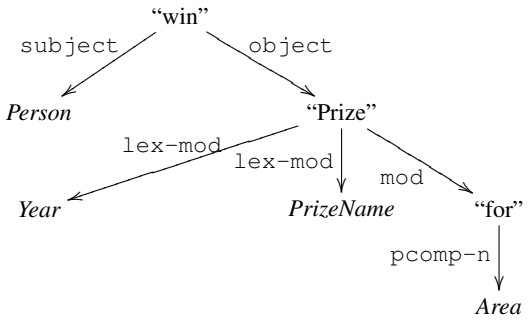
(2) is an example relation instance of (1), referring to an event mentioned in the sentence (3).

(2) *<Mohamed ElBaradei, Nobel, Peace, 2005>*

(3) *Mohamed ElBaradei won the 2005 Nobel Prize for Peace on Friday for his efforts to limit the spread of atomic weapons.*

(4) is a simplified dependency tree of the parsing result of (3).

(4)



From the tree in (4), DARE learns three rules in a bottom-up way. The first rule is dominated by the preposition "for", exacting the argument *Area*. The second rule is dominated by the noun "Prize", extracting the arguments *Year* and *PrizeName*, and calling the first rule for the argument *Area*. (5) and (6) show the first and second DARE rules.

(5) extracts the semantic argument *Area* from the prepositional phrase headed by the preposition "for", while (6) extracts the three arguments *Year*, *Prize* and *Area* from the complex noun phrase and calls the rule (5) for the semantic argument *Area*.

(5)
Rule name :: area_1
Rule body ::
[head [pos noun; lex-form "for"]
  daughters < [pcomp-n [head 1 Area]] >]
Output :: < 1 *Area* >

(6)
Rule name :: year_prize_area_1
Rule body ::
[head [pos noun; lex-form "prize"]
  daughters < [lex-mod [head 1 Year]],
  [lex-mod [head 2 Prize]],
  [mod [rule area_1 :: < 3 *Area* >]] >]
Output :: < 1 *Year*, 2 *Prize*, 3 *Area* >

(7) is the third rule that extracts all four arguments from the verb phrase dominated by the verb "win" and calls the second rule to handle the arguments embedded in the linguistic argument "object".

(7)
Rule name :: recipient_prize_area_year_1
Rule body ::
[head [pos verb; mode active; lex-form "win"]
  daughters < [subject [head 1 Person]],
  [object [rule year_prize_area_1 :: < 4 *Year*, 2 *Prize*, 3 *Area* >]] >]
Output :: < 1 *Recipient*, 2 *Prize*, 3 *Area*, 4 *Year* >

During the bootstrapping, the confidence values of the newly acquired rules and instances are calculated by DARE in the spirit of the "Duality principle" (Brin, 1998; Yangarber, 2001; Agichtein and Gravano, 2000), i.e., the confidence values of the rules are dependent on the truth value of their extracted instances and on the seed instances from which they stem. The confidence value of an extracted instance makes use of the confidence value of its ancestor seed instances. DARE employs two NLP modules: a named-entity recognizer SProUT (Drozdzynski et al., 2004) and a parser (De Marneffe et al., 2006). SProUT is adapted to new domains by adding rules for new NE types and extending the gazetteers.

3 Learning a General Relation from Single and Multiple Domains

The motivation of this work is to learn as many extraction rules as possible for extracting instances of the marriage relation between two persons, to fill, for instance, a biographic database about popular persons from different social domains. We employ DARE to learn the extraction rules from texts for three social categories: entertainment, politicians and business people.

3.1 Data Set and Data Properties

For each domain, we collect 300 Wikipedia documents, each document about one person. For the entertainment domain, we choose pages about actors or actresses of the Oscar academy awards and grammy winners. Pages about the US presidents and other political leaders are selected for the politician domain. American chief executives covered by the Wikipedia are candidates for the business people corpus. In Table 1, we show the distribution of persons, their occurrences and sentences referring to two persons. We immediately observe that the business texts mention much fewer persons or relationships between persons than the texts on politicians. Most mentions of persons and relationships can be found in the entertainment texts so that we can expect to find more extraction rules there than in the other domains.

3.2 Challenges without Gold Standard

Uszkoreit et al. (2009) discussed the challenge of seed selection and its influence on performance in a minimally supervised learning system, e.g., one randomly selected seed is sufficient to find most mentions in the Nobel Prize corpus, but many seeds cannot improve the performance for the MUC-6 corpus. Although we are aware of this problem, we still have to live with the situation

Domain	Entertainer	Politician	Business Person
Number of documents	300	300	300
Size (MB)	4.8	6.8	1.6
Number of person occurrences	61450	63015	9441
Number of person entities	9054	6537	1652
Sentences containing person-person-relations	9876	11111	1174

Table 1: Data Properties of the three Domain Corpora

that all three corpora selected here are unlabeled free texts and their data properties for learning are unknown to us. Furthermore, as pointed out by Agichtein and Gravano (2000), without annotated data, the calculation of recall is infeasible. Therefore, our evaluation can only provide the precision value and the number of the correctly extracted instances.

3.3 Experiments

In the first experiment, we begin by learning from each domain separately starting with positive examples from the domain. Then we merge the seeds and learn from the merged data of all three domains. The performance and the quality of the top ranked rules lead us to the second experiment, where we add negative seed in order to improve the ranking of the good rules. In the third experiment, we apply the good rules from the most fertile domain, i.e. entertainment, to the other two domains in order to find more relation instances in these texts.

3.3.1 Positive Seed

We decide to run 10 experiments, initialized each time with one positive example of a marriage instance for each respective domain, in order to obtain a more objective evaluation than only one experiment with a randomly selected seed. In order to operationalize this obvious and straightforward strategy, we first selected ten prominent married persons from the three sets of 300 persons featured in our Wikipedia articles. For finding the most prominent persons we simply took the length of their Wikipedia article as a crude indication. However, these heuristics are not essential for our experiments, since an increase of the seed set will normally substitute for any informed choice. For the runs with one example, the figures are the rounded averages over the ten runs with different seeds. For the merged corpus only one run was executed based on the three best seeds merged from the three domains.

Table 2 presents all figures for precision and number of correctly extracted instances for each domain and merged domains. The average precision of the business person domain is the highest, while the entertainment domain extracts the most correct instances but with the lowest precision. The politician domain has neither good precision nor good extraction gain.

Single domain	1 positive seed (each)	
	Precision	Extracted Correct Instances
Entertainer	5.9%	206
Politician	16.19%	159
Business Person	70.45%	31
Multiple domains	3 positive seed (merged)	
	Precision	Correct instances
merged corpus	8.91%	499

Table 2: Average values of 10 runs for each domain and 1 run for the merged corpus with best seeds

As expected, the distribution of the learned rules and their rankings behave differently in each domain. We got 907 rules from the entertainment domain, 669 from the politician domain, but only 7 from the business person domain. For illustration we only present the top-ranked rules from each domain cutting off after rank 15. The rules are extracted from the trees generated by the Stanford Dependency Parser for the candidate sentences of our corpora (De Marneffe et al., 2006). Here, we present the rules in a simplified form. The first elements in the rules are *head*, followed by their daughters. *A* and *B* are the two person arguments for the target relation. The good rules are highlighted as bold.

- **Top 15 rules in the entertainment domain:**
  1. <person>: dep(A), dep(B)
  2. (“meet”, VB): obj(A), subj(B)
  3. **(“divorce”, VB): subj(A, dep(B))**
  4. **(“wife”, N): mod(A), mod(B)**
  5. **(“marry”, VB): dep(A), nsubj(B), aux(“be”,VB)**
  6. (“star”, VB): dep(A), subj(B)
  7. **(“husband”,N): mod(A), mod(B)**

8. <position>: dep(A), dep(B)
9. (“attraction”, N): mod(A), mod(B)
10. <person>: mod(A), mod(A)
11. (“include”, VB): obj(A , dep(B))
12. **(“marry”, VB): obj(A), subj(B)**
13. (“star”, VB): obj(A , dep(B))
14. <person>: dep(A, dep(B))
15. **(“marriage”, N): dep(A), mod(B)**

- **Top 15 rules in the politician domain:**
  1. <person>: dep(A), dep(B)
  2. (“children”, N): dep(A, dep(B))
  3. **(“wife”, N): mod(A), mod(B)**
  4. **(“marry”, VB): obj(A), subj(B)**
  5. (“son”, N): mod(A), mod(B)
  6. <position>: mod(A), mod(B)
  7. (“include”, VB): obj(A , dep(B))
  8. <person>: mod(A), mod(B)
  9. <person>: dep(A), mod(B)
  10. (“defeat”, VB): obj(A), subj(B)
  11. (“successor”, N): mod(A), mod(B)
  12. (“lose”, VB): subj(A), dep(B)
  13. (“with”, IN): obj(A, dep(B))
  14. (“father”, NN): mod(A), mod(B)
  15. (“appoint”, VB): nsubj(A), dep(B), aux(“be”, VB)
- **Top rules in the business-person domain**
  1. (“children”, N): dep(A), dep(B)
  2. (“have”, VB): subj(A, dep(B))
  3. (“give”, VB): subj(A), obj(B)
  4. (“date”, VB): subj(A), obj(B)
  5. (A): dep((“wife”, NN), mod(B))
  6. (“student”, N): dep(A , dep(B))
  7. **(“marry”, VB): obj(A), subj(B)**
- **Top 15 rules in the merged corpus:**
  1. <person>: dep(A), dep(B)
  2. **(“wife”, N): mod(A), mod(B)**
  3. (“son”, N), mod(A): mod(B)
  4. **(“marry”, VB): obj(A), subj(B)**
  5. (“meet”, VB), obj(A): subj(B)
  6. (“include”, VB): obj(A), dep(B)
  7. <position>: mod(A), mod(B)
  8. (“children”, N): dep(A), dep(B)
  9. <person>: dep(A , mod(B))
  10. <person>: dep(A), mod(B)
  11. **(“marry”, VB): dep(A), nsubj(B), aux(“be”,VB)**
  12. (“father”, N): dep(A), dep(B)
  13. (“tell”, VB): obj(A), subj(B)
  14. **(“husband”,N): mod(A), mod(B)**
  15. <person>: mod(A), mod(B)

In all experiments, the good rules are not ranked highest. Although many good rules can be learned from the entertainment domain, several dangerous rules (such as the rule extracting instances of the “meet”-relation) are ranked higher because they are mentioned more frequently and often match with a seed person pair standing in marriage relation. In this domain, the married persons are often mentioned together in connection with other popular activities. This overlap of marriage with other relations causes many wrong rules. For example, the top ranked rule is learned from the following sentence (8) matching the seed (*Charles Laughton*, *Elsa Lanchester*).

(8) In total, he (Billy Wilder) directed fourteen different actors in Oscar-nominated performances: Barbara Stanwyck, . . . , Audrey Hepburn, Charles Laughton, Elsa Lanchester, Jack Lemmon, . . .

Many couples are mentioned in such coordination constructions. Therefore, this rule has a high connectivity and produces more than 2000 relation instances, boosting the rank of the rule to the top. Yet most instances extracted by this rule are incorrect. Several rules of similar type are the reason for the low precision in the entertainer and the politician domains. On the other hand, all three domains share the good rule:

(9) **(“marry”, VB): obj(A), subj(B)**

The extraction results from the merged corpus are comparable to the entertainment domain: low precision and high gain of instances. The increase of the data size supports higher recall.

Driven by our scientific curiosity, we increase the number of our positive seed to 10 with 10 runs too. Table 3 shows that the average precision for entertainer and politician domains do not improve significantly. All three domains yield a higher recall because more good rules could be learned from the larger seed.

Single domain domain	10 positive seed (each)	
	Precision	Correct instances
Entertainer	6.12%	264
Politician	17.32%	185
Business Person	78.95%	60
Multiple domains	30 positive seed (merged)	
	Precision	Correct instances
merged corpus	8.93%	513

Table 3: Experiments with 10 positive seeds for every corpus and 30 seeds for the merged corpus

But enlarged seeds could not help in finding more highly ranked good rules. On the contrary, some good rules disappear from the top positions. The reason is that different seeds produce different good rules but sometimes share the same bad rules, thus unfortunately boosting these bad rules

in rank. Bad rules are rules which extract wrong instances.

It is interesting to observe that the merged corpus in both experiments extracts more correct instances than the sum of the single domains together, in particular, in the one seed experiment, 499 (merged) vs. 396 (the sum of the single domains). In the case of the 10 seed experiment, the merged corpus extracted 513 correct instances while the single domains together 509. This indicates that both the enlargements of seeds and corpus size raise recall.

3.3.2 Negative Seed for Learning Negative Rules

Next we improve precision by accounting for other relations in which married couples are frequently mentioned:

1. Laurence Olivier saw Vivien Leigh in The Mask of Virtue.
2. Olivier and Leigh began an affair after acting as lovers in Fire Over England .
3. In the June 2006 Ladies' Home Journal, she said she (Nicole Kidman) still loved Cruise.
4. She (Nicole Kidman) became romantically involved with actor Tom Cruise on . . ..
5. He (Tom Cruise) and Kidman adopted two children.

Table 4 shows the average number of different relations reported about the extracted couples involved in the three domains. Thus, given a person pair as seed, DARE also learns rules which mention other relationships, especially in the entertainment domain.

Entertainer	Politician	Business Person
5.10	2.85	1.59

Table 4: Average number of various relations reported about the extracted couples

There are several approaches to negative samples for rule learning. Most of them ((Etzioni et al., 2005), (Lin et al., 2003), (Yangarber, 2003) and (Uszkoreit et al., 2009)) use the instances of other target relations as their negative examples or negative seed. Inspired by them, we employ negative seed examples to weed out dangerous rules. The dangerous rules are rules which extract incorrect instances in addition to the correct instances. We apply the negative seed to learn so-called negative rules and hope that the negative rules will cover the dangerous rules learned by the positive

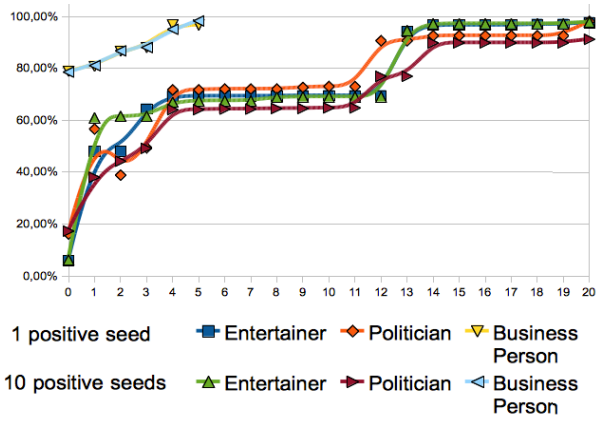


Figure 2: Average precision of experiments in 3 domains with 1 or 10 positive seeds and 1 to 20 negative seeds: *x axis* for negative seed, *y axis* for precision

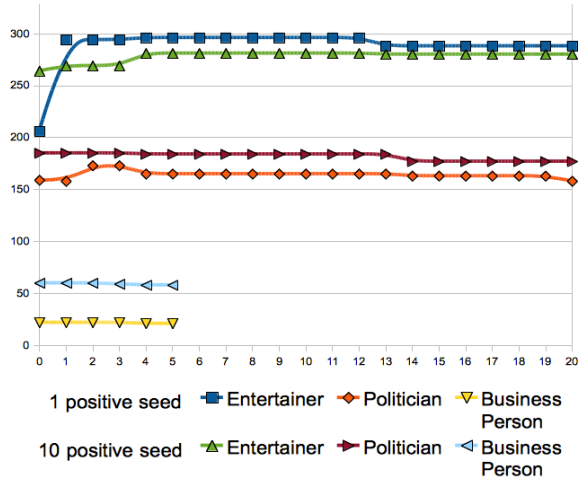


Figure 3: Correct instances of experiments in 3 domains with 1 or 10 positive seeds and 1 to 20 negative seeds: *x axis* for negative seed, *y axis* for number of extracted correct instances

seed. For the negative seed construction, we develop a new approach. Negative seed for our target relation contains person pairs who do not stand in a marriage relation, but who are extracted by the top 20 ranked rules produced from positive seed. The learning of the negative rules works just like the learning of the positive ones, but without any iterations. Once we have obtained rules from negative examples, we only use them for subtracting any identical rules from the rule set learned from positive seed.

Figure 2 shows the improvement of precision after the utilization of negative seed for 1 positive and 10 positive seed situations, while Figure 3 depicts the development of the extracted corrected instances. It appears that the number of the positive seeds does not make a significant difference of the performance development. For the business person domain, only a few negative seeds suffice for getting 100% precision. For both entertain-

ment and politician domains, the negative seeds considerably improve precision. There are several jumps in the curves. In the entertainment domain, the first negative seed removes the strongest bad rule. As a side-effect some good rules move upwards so that both precision and recall increase significantly and at the same time some other bad rules move downwards which are connected to subsequent negative seeds. Therefore, the second negative seed does not lead to big jump in the performance. Similar phenomena can be observed by analysing other flat portions of the curve.

In the following, we show only the top 10 rules learned from the entertainment domain with 1 positive seed and 20 negative seeds because of the limit of space.

(10) *top 10 rules learned from the entertainment domain:*

1. **(“wife”, N): mod(A), mod(B)**
2. **(“divorce”, VB): subj(A, dep(B))**
3. **(“marry”, VB): obj(A), subj(B)**
4. **(“husband”,N): mod(A), mod(B)**
5. **(“marry”, VB): dep(A), nsubj(B), aux(“be”,VB)**
6. **(“marriage”, N): dep(A), mod(B)**
7. (“appear”, VB): dep(A), subj(B)
8. <person>: dep(A), mod(B)
9. <position>: mod(A), mod(B)
10. (“friend”, N): mod(A), mod(B)

The entertainment domain has taken the biggest advantage of the negative seed strategy. The top 6 rules are all good rules. The other two domains contain only a subset of rules.

3.3.3 Exploitation of Beneficial Domains for Other Domains

The above experiments show us that the entertainment domain provides a much better resource for learning rules than the other two domains. As it will often happen that relevant application domains are not supported by beneficial data sets, we finally investigate the exploitation of data from a more popular domain for RE in a less beneficial domain. We apply rules learned from entertainment domain to the politician and business person domains. Table 5 shows that applying the top six rules in (10) learned from the entertainment domain discover many additional correct instances from the other two domains.

	Precision	new instances
Politician	98.48%	27
Business person	96.72%	17

Table 5: Additional instances extracted by the learned top six rules from the entertainment domain

4 Summary and Conclusion

In this paper we provide new evidence for the successful application of a minimally supervised IE approach based on semantic seed and bottom-up rule extraction from dependency structures to new domains with varying data properties. The experiments confirm and illustrate some hypotheses on the role of data properties on the learning process. A new approach to gathering and exploiting negative seed has been presented that considerably improves precision for individual and merged domains. Some positive effects of merging domains could be demonstrated.

An important observation is the successful exploitation of data from a related but different domain for a domain that does not possess suitable learning data. Thus we can cautiously conclude that the underlying minimally supervised bootstrapping approach to IE is not necessarily doomed to failure for domains that do not possess beneficial data sets for learning. Just as Xu (2007) already observed when they were able to use extraction rules learned from Nobel Prize news to detecting instances of other award events, we could now obtain first evidence for the effective reusability of rules learned from a combination of positive and negative examples.

Future research will have to confirm that the observed improvements of RE, especially the gain of precision obtained by the new method for using negative examples will actually scale up to much larger data sets and to more complex relations. We have already successfully applied the learned rule sets for the detection of marriage instances to collecting biographical information from other web data. However because of the inherent problems associated to measuring precision and especially recall in web-based IR/IE tasks, a rigid evaluation of these extractions will only be possible after extensive and expensive hand labelling efforts.

Acknowledgements

This research was conducted in the context of the German DFG Cluster of Excellence on Multimodal Computing and Interaction (M2CI), projects Theseus Alexandria and Alexandria for Media (funded by the German Federal Ministry of

Economy and Technology, contract 01MQ07016), and project TAKE (funded by the German Federal Ministry of Education and Research, contract 01IW08003).

References

Eugene Agichtein and Luis Gravano. 2000. Snowball: Extracting relations from large plain-text collections. In *Proceedings of the 5th ACM International Conference on Digital Libraries (DL'00)*, San Antonio, TX, June.

LAN Amaral, A. Scala, M. Barthélémy, and HE Stanley. 2005. Classes of small-world networks. *Proceedings of the National Academy of Sciences*, 102(30):10421–10426.

Sergey Brin. 1998. Extracting patterns and relations from the world wide web. In *WebDB Workshop at 6th International Conference on Extending Database Technology, EDBT'98*.

R. C. Bunescu and R.J Mooney. 2005. A shortest path dependency kernel for relation extraction. In *Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 724–731, Vancouver, B.C., October.

M.C. De Marneffe, B. MacCartney, and C.D. Manning. 2006. Generating typed dependency parses from phrase structure parses. In *LREC 2006*.

Witold Drozdzynski, Hans-Ulrich Krieger, Jakub Piskorski, Ulrich Schäfer, and Feiyu Xu. 2004. Shallow processing with unification and typed feature structures — foundations and applications. *Künstliche Intelligenz*, 1.

O. Etzioni, M. Cafarella, D. Downey, A.M. Popescu, T. Shaked, S. Soderland, D.S. Weld, and A. Yates. 2005. Unsupervised named-entity extraction from the web: An experimental study. *Artificial Intelligence*, 165(1).

Mark A. Greenwood and Mark Stevenson. 2006. Improving semi-supervised acquisition of relation extraction patterns. In *Proceedings of the Workshop on Information Extraction Beyond The Document*. Association for Computational Linguistics.

Ralph Grishman and Beth Sundheim. 1996. Message understanding conference - 6: A brief history. In *Proceedings of the 16th International Conference on Computational Linguistics*, Copenhagen, June.

R. Jones. 2005. *Learning to Extract Entities from Labeled and Unlabeled Text*. Ph.D. thesis, University of Utah.

Zornitsa Kozareva and Eduard Hovy. 2010a. Learning arguments and supertypes of semantic relations using recursive patterns. In *Proceedings of COLING 2010*, Uppsala, Sweden.

Zornitsa Kozareva and Eduard Hovy. 2010b. Not all seeds are equal: Measuring the quality of text mining seeds. In *Proceedings of HLT/NACL 2010*, Los Angeles, California.

W. Lin, R. Yangarber, and R. Grishman. 2003. Bootstrapped learning of semantic classes from positive and negative examples. In *Proceedings of ICML-2003 Workshop on The Continuum from Labeled to Unlabeled Data*, pages 103–111.

Ryan McDonald, Fernando Pereira, Seth Kulick, Scott Winters, Yang Jin, and Pete White. 2005. Simple algorithms for complex relation extraction with applications to biomedical IE. In *Proceedings of ACL 2005*. Association for Computational Linguistics.

Ion Muslea. 1999. Extraction patterns for information extraction tasks: A survey. In *AAAI Workshop on Machine Learning for Information Extraction*, Orlando, Florida, July.

K. Sudo, S. Sekine, and R. Grishman. 2003. An improved extraction pattern representation model for automatic IE pattern acquisition. *Proceedings of ACL 2003*, pages 224–231.

Junichi Tsujii. 2000. Generic nlp technologies: language, knowledge and information extraction. In *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics (ACL)*.

Hans Uszkoreit, Feiyu Xu, and Hong Li. 2009. Analysis and improvement of minimally supervised machine learning for relation extraction. In *14th International Conference on Applications of Natural Language to Information Systems*.

Hans Uszkoreit. 2011. Learning relation extraction grammars with minimal human intervention: Strategy, results, insights and plans. In Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing*, volume 6609 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg.

Feiyu Xu, Hans Uszkoreit, and Hong Li. 2007. A seed-driven bottom-up machine learning framework for extracting relations of various complexity. *Proceedings of ACL 2007*, pages 584–591.

Feiyu Xu. 2007. *Bootstrapping Relation Extraction from Semantic Seeds*. Phd-thesis, Saarland University.

Roman Yangarber. 2001. *Scenarion Customization for Information Extraction*. Dissertation, Department of Computer Science, Graduate School of Arts and Science, New York University, New York, USA.

R. Yangarber. 2003. Counter-training in discovery of semantic patterns. In *Proc. ACL-2003*. Association for Computational Linguistics.