Boosting Relation Extraction with Limited Closed-World Knowledge

Feiyu Xu **Hans Uszkoreit** **Sebastian Krause** **Hong Li**
Language Technology Lab
German Research Center for Artificial Intelligence (DFKI GmbH)
{feiyu,uszkoreit,sebastian.krause,lihong}@dfki.de

Abstract

This paper presents a new approach to improving relation extraction based on minimally supervised learning. By adding some limited closed-world knowledge for confidence estimation of learned rules to the usual seed data, the precision of relation extraction can be considerably improved. Starting from an existing baseline system we demonstrate that utilizing limited closed world knowledge can effectively eliminate ”dangerous” or plainly wrong rules during the bootstrapping process. The new method improves the reliability of the confidence estimation and the precision value of the extracted instances. Although recall suffers to a certain degree depending on the domain and the selected settings, the overall performance measured by F-score considerably improves. Finally we validate the adaptability of the best ranking method to a new domain and obtain promising results.

1 Introduction

Minimally supervised machine-learning approaches to learning rules or patterns for relation extraction (RE) in a bootstrapping framework are regarded as very effective methods for building information extraction (IE) systems and for adapting them to new domains (e. g., (Riloff, 1996), (Brin, 1998), (Agichtein and Gravano, 2000), (Yangarber, 2001), (Sudo et al., 2003), (Jones, 2005), (Greenwood and Stevenson, 2006), (Agichtein, 2006), (Xu et al., 2007), (Xu, 2007)). On the one hand, these approaches show very promising results by utilizing minimal domain knowledge as seeds. On the other hand, they are all confronted with the same problem, i.e., the acquisition of wrong rules because of missing knowledge for their validation during bootstrapping. Various approaches to confidence estimation of learned rules have been proposed as well as methods for identifying ”so-called” negative rules for increasing the precision value (e.g., (Brin, 1998), (Agichtein and Gravano, 2000), (Agichtein, 2006), (Yangarber, 2003), (Pantel and Pennacchiotti, 2006), (Etzioni et al., 2005), (Xu et al., 2007) and (Uszkoreit et al., 2009)).

In this paper, we present a new approach to estimating or ranking the confidence value of learned rules by utilizing limited closed-world knowledge. As many predecessors, our ranking method is built on the “Duality Principle” (e. g., (Brin, 1998), (Yangarber, 2001) and (Agichtein, 2006)). We extend the validation method by an evaluation of extracted instances against some limited closed-world knowledge, while also allowing cases in which knowledge for informed decisions is not available. In comparison to previous approaches to negative examples or negative rules such as (Yangarber, 2003), (Etzioni et al., 2005) and (Uszkoreit et al., 2009), we implicitly generate many negative examples by utilizing the positive examples in the closed-world portion of our knowledge. Rules extracting wrong instances are lowered in rank.

In (Xu et al., 2007) and (Xu, 2007), we develop a generic framework for learning rules for relations of varying complexity, called ***DARE*** (Domain Adaptive Relation Extraction). Furthermore, there is a systematic error analysis of the baseline system conducted in (Xu, 2007). We employ our system both as a baseline reference and as a platform for implementing and evaluating our new method.

Our first experiments conducted on the same data used in (Xu et al., 2007) demonstrate: 1) limited closed-world knowledge is very useful and effective for improving rule confidence estimation and precision of relation extraction; 2) integration of soft constraints boosts the confidence value of the good and relevant rules, but without strongly decreasing the recall value. In addition, we validate our method on a new corpus of newspaper texts about celebrities and obtain promising results.

The remainder of the paper is organized as follows: Section 2 explains the relevant related work. Sections 3 and 4 describe ***DARE*** and our extensions. Section 5 reports the experiments with two ranking strategies and their results. Section 6 gives a summary and discusses future work.

2 Related Work

In the existing minimally supervised rule learning systems for relation extraction based on bootstrapping, they already employ various approaches to confidence estimation of learned rules and different methods for identification of so-called negative rules. For estimation of confidence/relevance values of rules, most of the approaches follow the so-called "Duality Principle" as mentioned by Brin (1998) and Yangarber (2001), namely, the confidence value of learned rules is dependent on the confidence value of their origins, which can be documents or relation instances. For example, Riloff (1996), Yangarber (2001), Sudo et al. (2003) and Greenwood and Stevenson (2006) use domain relevance of documents in which patterns are discovered as well as the distribution frequency of these patterns in those relevant documents as an indication of good patterns. Their methods are aimed at detecting all patterns for a specific domain, but those patterns cannot be applied directly to a specific relation. In contrast, systems presented by Brin (1998), Agichtein and Gravano (2000), Agichtein (2006), Pantel and Pennacchiotti (2006) as well as our baseline system (Xu et al., 2007) are designed to learn rules for a specific relation. They start with some relation instances as their so-called "semantic seeds" and detect rules from texts matching with these instances. The new rules are applied to new texts for extracting new instances. These new instances in turn are utilized as new seeds. All these systems calculate their rule confidence based on the confidence values of the instances from which they stem. In addition to the confidence value of the seed instances, most of them also consider frequency information and include some heuristics for extra validation. For example, Agichtein (2006) intellectually defines certain constraints for evaluating the truth value of extracted instances. But it is not clear whether this strategy can be adapted to new domains and other relations. In (Xu et al., 2007) we make use of domain relevance values of terms occurring in rules. This method is not applicable to general relations.

Parallel to confidence estimation strategies, the learning of negative rules is useful for identifying wrong rules straightforwardly. Yangarber (2003) and Etzioni et al. (2005) utilize the so-called *Counter-Training* for detecting negative rules for a specific domain or a specific class by learning from multiple domains or classes at the same time. Examples of one certain domain or class are regarded as negative examples for the other ones. Bunescu and Mooney (2007) follow a classification-based approach to RE. They use positive and negative sentences of a target relation for a SVM classifier. Uszkoreit et al. (2009) exploit negative examples as seeds for learning further negative instances and negative rules. The disadvantage of the above four approaches is that the selected negative domains or classes or negative instances cover only a subset of the negative domains/classes/relations of the target domain/class/relation.

3 *DARE* Baseline System

Our baseline system ***DARE*** is a minimally supervised learning system for relation extraction, initialized by so-called "semantic seeds", i.e., examples of the target relations, labelled with their semantic roles. The system supports domain adaptation through a compositional rule representation and a bottom-up rule discovery strategy. In this way, ***DARE*** can handle target relations of varying arity. The following example is a relation instance of the target relation from (Xu, 2007) concerning Nobel Prize awards: *<Mohamed ElBaradei, Nobel, Peace, 2005>*. The target relation contains four arguments: WINNER, PRIZE_NAME, PRIZE_AREA and YEAR. This example refers to an event mentioned in the sentence in example (1).

(1) *Mohamed ElBaradei, won the 2005 Nobel Prize for Peace on Friday because of ....*

Figure 1 is a simplified dependency tree of example (1). ***DARE*** utilizes a bottom-up rule discovery strategy to extract rules from such dependency trees. All sentences are processed with named entity recognition and dependency parsing.

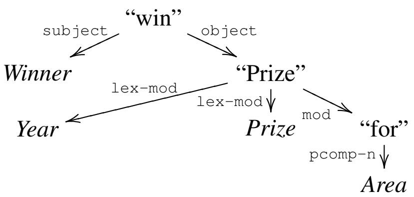

Figure 1: Dependency tree for example (1)

From the tree in Figure 1, ***DARE*** learns three rules. The first rule is dominated by the preposition "for", extracting the argument PRIZE_AREA (*Area*). The second rule is dominated by the noun "Prize", extracting the arguments YEAR (*Year*) and PRIZE_NAME (*Prize*), and calling the first rule for the argument PRIZE_AREA (*Area*). The rule "winner_prize_area_year_1" from Figure 2 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".

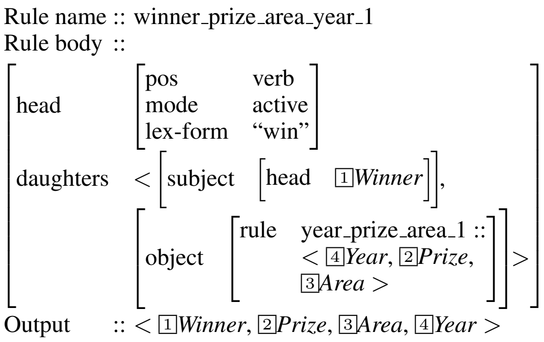

Figure 2: ***DARE*** extraction rule.

We conduct a systematic error analysis based on our experiments with the Nobel Prize award data (Xu, 2007). The learned rules are divided into four groups: *good*, *useless*, *dangerous* and *bad*. The good rules are rules that only extract correct instances, while bad ones exclusively produce wrong instances. Useless rules are those that do not detect any new instances. Dangerous rules are dangerous because they extract both correct and wrong instances. Most good rules are rules with high specificity, namely, extracting all or most arguments of the target relation. The 14.7% extraction errors are from bad rules and dangerous rules. Other errors are caused by wrong reported content, negative modality, parsing and named entity recognition errors.

4 Our Approach: Boosting Relation Extraction

4.1 Closed-World Knowledge: Modeling and Construction

The error analysis of ***DARE*** confirms that the identification of bad rules or dangerous rules is important for the precision of an extraction system. Using closed-world knowledge with large numbers of implicit negative instances opens a possibility to detect such rules directly. In our work, closed-world knowledge for a target relation is the total set of positive relation instances for entire relations or for some selected subsets of individuals. For most real world applications, closed-world knowledge can only be obtained for relatively small subsets of individuals participating in the relevant relations. We store the closed-world knowledge in a relational database, which we dub "closed-world knowledge database" (abbr. ***cwDB***). Thus, a ***cwDB*** for a target relation should fill the following condition:

> A ***cwDB*** must contain all correct relation instances ($insts$) for an instantiation value (***argValue***) of a selected relation argument ***cwArg*** in the target relation.

Given $\mathbf{R}$ (the total set of relation instances of a target relation), a ***cwDB*** is defined as follows:

$$\textbf{\textit{cwDB}}=\{inst \in \mathbf{R} : \textbf{\textit{cwArg}}(inst) = \textbf{\textit{argValue}}\}.$$

An example of a ***cwDB*** is the set of all prize winners of a specific prize area such as *Peace*, where PRIZE_AREA is the selected ***cwArg*** and ***argValue*** is *Peace*. Note that the merger of two ***cwDB***s, for example with PRIZE_AREAS *Peace* and *Literature*, is again a ***cwDB*** (with two ***argValue***s in this case).

4.2 Modified Learning Algorithm

In Algorithm 1, we present the modification of the ***DARE*** algorithm (Xu, 2007). The basic idea of ***DARE*** is that it takes some initial seeds as input and learns relation extraction rules from sentences in the textual corpus matching the seeds. Given the learned rules, it extracts new instances from the texts. The modified algorithm adds the **validate** step to evaluate the new instances against the closed-world knowledge ***cwDB***. Based on the evaluation result, both new instances and learned rules are ranked with a confidence value.

INPUT: *initial seeds*
$i \leftarrow 0$ (iteration of bootstrapping)
$seeds \leftarrow initial\ seeds$
$all\ instances \leftarrow \{\}$
while $(seeds \neq \{\})$
  $rules_i \leftarrow$ **getRules**$(seeds)$
  $instances_i \leftarrow$ **getInstances**$(rules_i)$
  $new\ instances_i \leftarrow instances_i - all\ instances$
  validate$(new\ instances_i,$ ***cwDB***$)$
  rank$(new\ instances_i)$
  rank$(rules_i)$
  $seeds \leftarrow new\ instances_i$
  $all\ instances \leftarrow all\ instances + new\ instances_i$
  $i \leftarrow i + 1$
OUTPUT: *all instances*

Algorithm 1: Extended ***DARE***

4.3 Validation against *cwDB*

Given a ***cwDB*** of a target relation and its ***argValue*** of its selected argument ***cwArg***, the validation of an extracted instance ($inst$) against the ***cwDB*** is defined as follows.

$$\begin{array}{rcll} inst\ correct & \Leftrightarrow & & inst \in \textbf{\textit{cwDB}} \\ inst\ wrong & \Leftrightarrow & & inst \notin \textbf{\textit{cwDB}} \wedge \\ & & & \textbf{\textit{cwArg}}(inst) = \textbf{\textit{argValue}} \\ inst\ unknown & \Leftrightarrow & (& inst \notin \textbf{\textit{cwDB}} \wedge \\ & & & \textbf{\textit{cwArg}}(inst) \neq \textbf{\textit{argValue}} \quad) \\ & \vee & (& inst \notin \textbf{\textit{cwDB}} \wedge \\ & & & \textbf{\textit{cwArg}}(inst) \text{ is } \textit{unspecified} \quad) \end{array} \tag{1}$$

4.4 Rule Confidence Ranking with *cwDB*

We develop two rule-ranking strategies for confidence estimation, in order to investigate the best way of integrating the closed-world knowledge: (a) **exclusive ranking**: This ranking strategy excludes every rule which extracts wrong instances after their validation against the closed-world knowledge; (b) **soft ranking**: This ranking strategy is built on top of the duality principle and takes specificity and the depth of learning into account.

Exclusive Ranking The exclusive ranking method is a very naive ranking method which estimates the confidence value of a learned rule (e.g., $rule$) depending on the truth value of its extracted instances (**getInstances**$(rule)$) against a ***cwDB***. Any rule with one *wrong* extraction is regarded as a bad rule in this method. This method works effectively in a special scenario where the total list of the instances of the target relation is available as the ***cwDB***.

$$\textbf{confidence}(rule) = \begin{cases} 1 & \text{if } \textbf{getInstances}(rule) \subseteq \textbf{\textit{cwDB}}, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Soft Ranking The soft ranking method works in the spirit of the "Duality Principle", the confidence value of rules is dependent on the truth value of their extracted instances and on the seed instances from which they stem. The confidence value of the extracted instances is estimated based on their validation against the ***cwDB*** or the confidence value of their ancestor seed instances from which their extraction rules stem. Furthermore, the *specificity* of the instances (percentage of the filled arguments) and the *learning depth* (iteration step of bootstrapping) are parameters too. The definition of instance scoring, namely, **score**$(inst)$, is given as follows:

$$\textbf{score}(inst) = \begin{cases} \gamma > 0 & \text{if } \textbf{validate}(inst, \textbf{\textit{cwDB}}) = \textit{correct}, \\ 0 & \text{if } \textbf{validate}(inst, \textbf{\textit{cwDB}}) = \textit{wrong}, \\ UN_{inst} & \text{if } \textbf{validate}(inst, \textbf{\textit{cwDB}}) = \textit{unknown}. \end{cases} \tag{3}$$

As defined above, if a new instance is confirmed as correct by the ***cwDB***, it will obtain a positive value. In our experiment, we set γ=10 in order to boost the precision. In the case of *unknown* about its truth value, the confidence value of a new instance ($inst$) is dependent on the confidence values of the seed instances (ancestor seeds) from which its mother rules (R_{inst}) stem. Below, the scoring of the *unknown* case, namely, UN_{inst}, is defined, where R_{inst} are rules that extract the new instance $inst$, while I_{rule} are instances from which a $rule$ in R_{inst} is learned and α is the specificity value of $inst$ while β is utilized to express the noisy potential of each further iteration during bootstrapping.

$$UN_{inst} = \frac{\sum_{rule \in R_{inst}} \left(\frac{\sum_{j \in I_{rule}} \textbf{score}(j)}{|I_{rule}|} \times \beta^{i_{rule}} \right)}{|R_{inst}|} \times \alpha$$

$$\begin{array}{l} \text{where} \\ R_{inst} = \textbf{getMotherRulesOf}(inst), \\ I_{rule} = \textbf{getMotherInstancesOf}(rule), \\ \alpha = \textit{specificity}, \\ \beta = 0.8, \\ i_{rule} = i\text{-th iteration where } rule \text{ occurs} \end{array} \tag{4}$$

Given the scoring of instance $inst$, the confidence estimation of a rule is the average score of all $inst$s extracted by this rule:

$$\textbf{confidence}(rule) = \frac{\sum_{inst \in \mathbb{I}} \textbf{score}(inst)}{|\mathbb{I}|} \quad \text{where } \mathbb{I} = \textbf{getInstances}(rule) \tag{5}$$

5 Experiments

5.1 Corpora and Closed-World Knowledge

We conduct our experiments with two different domains. We start with the Nobel Prize award domain reported in (Xu, 2007) and apply our method to the same corpus, a collection from various online newspapers. The target relation is the one with the four arguments as mentioned in Section 3. In this way, we can compare our results with those reported in (Xu, 2007). Furthermore, all Nobel Prize winners can be found from `http://nobelprize.org`, so it is easy to construct a ***cwDB*** for Nobel Prize winners. We take the PRIZE_AREA as our selected argument for closing sub-relations and construct various ***cwDB***s with the instantiation of this argument (e.g., all winners of Nobel *Peace* Prize). The second domain is about celebrities. Our text corpus is collected from tabloid newspaper texts, containing 6850 articles from the years 2001 and 2002. The target relation is the marriage relationship between two persons. We construct a ***cwDB*** of 289 persons in which we have listed all their (ex-)spouses as well as the time span of the marriage relation.

Table 1 summarizes the size of the corpus data of the two domains.

Domain	Space	#Doc.
Nobel Prize	18,4 MB	3328
Celebrity Marr.	16,6 MB	6850

Table 1: Corpus data.

5.2 Nobel Prize Domain

We apply the extended ***DARE*** system to the Nobel Prize corpus at first and conduct two rule ranking strategies with different sizes of the ***cwDB***. We conduct all our experiments with the seed *<Guenter Grass, Nobel, Literature, 1999>*. The ***DARE***-Baseline performance is shown in Table 2.

	Precision	Absolute Recall
Baseline	77.98%	89.01%

Table 2: ***DARE***-Baseline Performance

Exclusive Ranking

Given the complete list of Nobel Laureates, we can apply the exclusive ranking strategy to this domain. Our ***cwDB*** is the total list of Nobel Prize winners. The wrong instances will not be used as seed for the next iteration. Rules that extracted at least one wrong instance are marked as *bad*, the other rules as *good*. We utilize only the good rules for relation extraction.

Prec.	Rel. Recall	Rel. F-Measure
100.00%	82.88%	90.64%

Table 3: Performance of Exclusive Ranking in Nobel Prize award domain.

In comparison to the ***DARE*** baseline system, given the same seed setup, this experiment results in a precision boost from 77.98% to 100% (see Table 3). This is not surprising since the ***cwDB*** covers all relation instances for the target relation. Nevertheless, this experiment shows that the closed-world knowledge approach is effective to exclude bad rules. However, the recall decreases and is only 82.88% of the one of the baseline system. As we explain above, not all rules extracting wrong instances are bad rules because wrong extractions can also be caused by other error sources such as named entity recognition. Therefore, even good rules can be excluded because of other error sources. The exclusive ranking strategy is useful for application scenarios where people want to learn rules for achieving 100% precision performance and do not expect high recall. It is especially effective when a big ***cwDB*** is available.

Soft Ranking

This ranking strategy does not exclude any rules and assigns a score to each rule based on the definition in Section 4.4. Rules which extract correct instances, more specific relation instances and stem from high-scored seed instances obtain a better value than others. In our approach, the *specificity* is dependent on the number of the arguments in the extracted instances. For this domain, the most specific instances contain all four arguments. In the following, we conduct two experiments with two different sizes of the ***cwDB***: 1) with the total list of winners (*complete* ***cwDB***) and 2) with only winners in one PRIZE_AREA (*limited* ***cwDB***).

1) Complete closed-world database Figure 3 displays the correlation between the score of rules and their extraction precision performance. Each point stands for a set of rules with the same score and extraction precision. In this setup, the higher the score, the higher the precision. Given the scored rules, Figure 4 depicts precision, recall and F-Measure for different score thresholds. For a given threshold j we take all $rule$s with **score**$(rule) \geq j$ and use the instances they extract. The recall value here is the relative recall w. r. t. to the ***DARE*** baseline performance: i. e. the number of correct extracted instances divided by the number of correct instances extracted by the ***DARE*** baseline system. The F-Measure value is calculated by using the relative recall values, we therefore refer to it as the *relative* F-Measure. If the system takes all rules with score ≥ 7, the system achieves the best relative F-Measure.

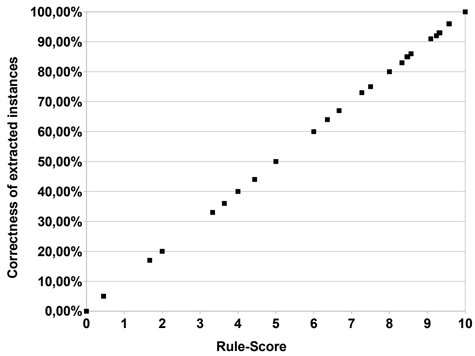

Figure 3: Rule scores vs. precisions with the complete closed-world database.

2) Limited closed-world database This experiment investigates the system performance in cases in which only a limited ***cwDB*** is available. This is the typical situation for most real world RE applications. Therefore, this experiment is much more

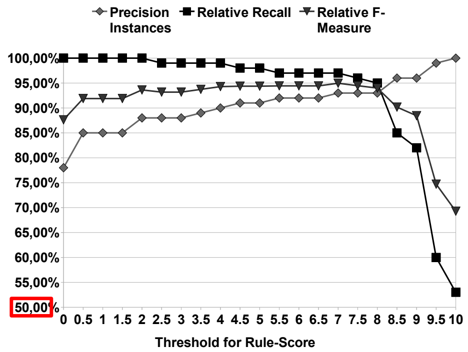

Figure 4: Performance with the complete closed-world database.

important than the previous one. We construct a smaller database containing only *Peace* Nobel Prize winners, which is about 1/8 of the previous complete ***cwDB***.

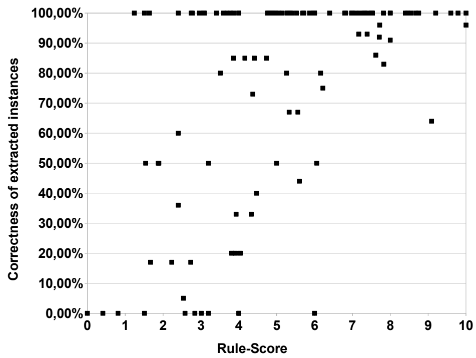

Figure 5: Rule score vs. precision with the limited closed-world database

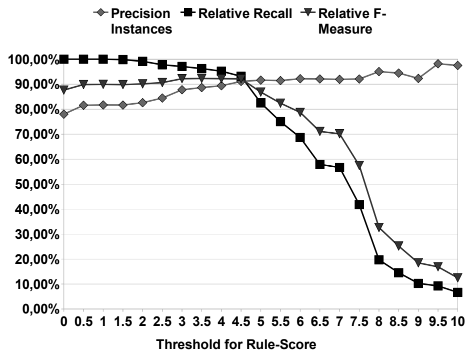

Figure 6: Performance with the limited closed-world database

Figure 5 shows the correlation between the score of the rules and their extraction precision. Although the development curve here is not as smooth as depicted in Figure 3, the higher scored rules have better precision values than most of the lower scored rules. However, we can observe that some very good rules are scored low, located in

Thresh.	Good	Dangerous	Bad
Baseline	58.94%	26.49%	14.57%
1	64.96%	29.20%	5.84%
2	66.67%	27.91%	5.43%
3	69.23%	26.50%	4.27%
4	**73.27%**	**23.76%**	**2.97%**
5	76.00%	22.67%	1.33%
6	77.59%	20.69%	1.72%
7	77.50%	22.50%	0.00%
8	87.50%	12.50%	0.00%
9	85.71%	14.29%	0.00%
10	90.00%	10.00%	0.00%

Table 4: Quality analysis of rules with the limited closed-world database

the left upper corner. The reason is that many of their extracted instances are *unknown*, even if their extracted instances are mostly correct.

As shown in Figure 6, even with the limited ***cwDB***, the precision values are comparable with the complete ***cwDB*** (see Figure 4). However, the recall value drops much earlier than with the complete ***cwDB***. With a threshold of score 4, the system achieves the best modified F-Measure 92,21% with an improvement of precision of about 11 percentage points compared to the ***DARE*** baseline system (89.39% vs. 77.98%). These results show that even with a limited ***cwDB*** this ranking system can help to improve the precision without loosing too much recall.

We take a closer look on the useful (actively extracting) rules and their extraction performance, using the same rule classification as (Xu, 2007). As shown in Table 4, more than one fourth of the extraction rules created by the baseline system are dangerous ones and almost 15% are plainly wrong. Applying the rule scoring with the limited ***cwDB*** increases the fraction of good rules to almost three quarters and nearly eliminates all bad rules at threshold 4. By choosing higher thresholds, surviving good rules raises to 90%. The total remaining set of rules then only consists of rules that at least partially extract correct instances.

5.3 Celebrity Domain

As presented above, the soft ranking method delivers very promising result. In order to validate this ranking method, we choose an additional domain and decide to learn marriage relations among celebrities, where the target relation consists of the following arguments: [NAME_OF_SPOUSE, NAME_OF_SPOUSE, YEAR].

The value of the marriage year is valid when the year is within the marriage time interval. The motivation of selecting this target relation is the large number of possible relations between two persons leading to dangerous or even bad rules. For example, the rule in Figure 7 is a very dangerous rule because "meeting" events of two married celebrities are often reported. A good confidence estimation method is very useful for boosting the good rules like the one in Figure 8. From our text corpus we extract 37.000 sentences that mention at least two persons. The ***cwDB*** consists of sample relation instances, in which one NAME_OF_SPOUSE is instantiated, i. e. we manually construct a database which contains all (ex-) spouses of 289 celebrities.

```
head([SPOUSE<ne_person>]),
mod({head(("meet", VB)),
    subj({head([SPOUSE<ne_person>])})})
```

Figure 7: A dangerous extraction rule example

```
head(("marry", VB)),
aux({head(("be", VB))}),
dep({head([SPOUSE<ne_person>]),
       dep({head([DATE<point>])})}),
nsubj({head([SPOUSE<ne_person>])})
```

Figure 8: Example of a positive rule

Since a gold standard of mentions for this corpus is not available, we manually validate 100 random samples from each threshold group. This evaluation gives us an opportunity to estimate the effect of a ***cwDB*** in this domain. Table 5 presents the performance of the rules with different thresholds. The precision value of the baseline system is very low. Threshold 3 slightly improves the precision of the ***DARE*** baseline without damaging recall too much. Step 4 excludes dangerous rules such as the one in Figure 7 which drastically boosts the precision. Unfortunately, the exclusion of such general rules leads to the loss of many correct relation instances too, therefore, the immense drop of recall from threshold 3 to 4 as well as from threshold 4 to 5. Positive extraction rules such as Figure 8 are quite highly scored. Because of the large number of rules and instances, we start the quality analysis of rules with score 3. As the table indicates, the use of the rule scoring in this domain clearly improves the quality of the created extraction rules. The error analysis shows that the major error resource for this domain is wrong coreference resolution or identity resolution. For ex-

Thresh.	# Instances	Prec.	Rel. Rec.	Rel. F-Meas.	# Rules	Good	Dangerous	Bad
Baseline	25183	9.00%	100.00%	16.51%	12258	———	———	———
1	19806	7.00%	61.17%	12.56%	562	———	———	———
2	14542	9.00%	57.75%	15.57%	159	———	———	———
3	11259	15.00%	74.51%	24.97%	121	19.83%	33.88%	46.28%
4	**788**	**65.00%**	**22.60%**	**33.54%**	**72**	**25.00%**	**27.78%**	**47.22%**
5	195	67.00%	5.76%	10.62%	29	37.93%	17.24%	44.83%
6	115	84.00%	4.26%	8.11%	11	45.45%	27.27%	27.27%
7	55	89.09%	2.16%	4.22%	6	50.00%	33.33%	16.67%
8	9	77.78%	0.31%	0.62%	4	75.00%	0.00%	25.00%
9	5	60.00%	0.13%	0.26%	3	66.67%	0.00%	33.33%
10	5	60.00%	0.13%	0.26%	3	66.67%	0.00%	33.33%

Table 5: Soft ranking for the celebrity marriage domain with a limited ***cwDB***.

ample, the inability to distinguish *Prince Charles* (former husband of British princess Diana) from *Charles Spencer* (her brother) is the reason that ***DARE*** crosses the border between the marriage and the sibling relation. In comparison to the Nobel Prize award event, the marriage relation between persons is often used as additional information to a person which is involved in a reported event. Therefore, anaphoric references occur more often in their mentionings, as the example relation in (3).

(3) *"My kids, I really don't like them to watch that much television," said Cruise, 40, who adopted Isabella and Connor while he was married to second wife Nicole Kidman.*

6 Summary

We propose a new way in which prior knowledge about domains can be efficiently used as additional criteria for confidence estimation of learned new rules or new instances in a minimally supervised machine learning framework. By introducing rule scoring on the basis of available domain knowledge (the cwDB), rules can be evaluated during the bootstrapping process with respect to their extraction precision. The results are rather promising. The rule score threshold is an easy way for users of an extraction system to adjust the precision-recall-trade-off to their own needs. The rule estimation method is also general enough to extend to integration of common sense knowledge. Although the relation instances in the closed-world knowledge database can also be used as seed in the beginning, the core idea of our research work is to develop a general confidence estimation strategy for discovered new information. As discussed in (Xu, 2007) and (Uszkoreit et al., 2009), the size of seed is not always relevant for the learning and extraction performance, in particular if the data corpus exhibits the small world property. Using all instances in the cwDB as seed, our experiments with the baseline system yield worse precision performance than the modified ***DARE*** algorithm with only one seed instance.

This approach is quite general and easily adaptable to many domains; the only prerequisite is the existence of a database with relation instances from the target domain with a fulfilled closed-world property on some relational argument. A database of this kind should be easily obtainable for many domains, e. g. by exploiting structured and semi-structured information sources in the Internet, such as *YAGO* (Suchanek et al. (2007)) and *DBpedia* (Bizer et al. (2009)). Furthermore, in some areas, such as Business Intelligence, there is nearly complete knowledge already present for past years, while the task is to extract information only from recent news articles. Constructing closed-worlds out of the present knowledge to improve the learning of new information is therefore a straightforward approach. Even the manual collection of suitable data might be a reasonable choice since appropriate closed worlds could be rather small if ***cwDB***is chosen properly.

Acknowledgments

The work presented here has been partially supported through the prject KomParse by the ProFIT program of the Federal State of Berlin which in turn is co-funded by the EFRE program of the European Union. It is additionally supported through a grant to the project TAKE, funded by the German Ministry for Education and Research (BMBF, FKZ: 01IW08003).

References

Agichtein, Eugene and Luis Gravano. 2000. Snowball: extracting relations from large plain-text collections. In *DL '00: Proceedings of the fifth ACM conference on Digital libraries*, pages 85–94, New York, NY, USA. ACM.

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

Bizer, Christian, Jens Lehmann, Georgi Kobilarov, Sören Auer, Christian Becker, Richard Cyganiak, and Sebastian Hellmann. 2009. DBpedia - a crystallization point for the web of data. *Journal of Web Semantics*, 7(3):154–165.

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

Bunescu, Razvan C. and Raymond J. Mooney. 2007. Learning to extract relations from the web using minimal supervision. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*.

Etzioni, Oren, Michael Cafarella, Doug Downey, Ana-Maria Popescu, Tal Shaked, Stephen Soderland, Daniel S. Weld, and Alexander Yates. 2005. Unsupervised named-entity extraction from the web: An experimental study. *Artificial Intelligence*, 165(1):91 – 134.

Greenwood, Mark A. and Mark Stevenson. 2006. Improving semi-supervised acquisition of relation extraction patterns. In *Proceedings of the Workshop on Information Extraction Beyond The Document*, pages 29–35, Sydney, Australia, July. Association for Computational Linguistics.

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

Pantel, Patrick and Marco Pennacchiotti. 2006. Espresso: Leveraging generic patterns for automatically harvesting semantic relations. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, Sydney, Australia, July. The Association for Computer Linguistics.

Riloff, Ellen. 1996. Automatically generating extraction patterns from untagged text. In *Proceedings of Thirteenth National Conference on Artificial Intelligence (AAAI-96)*, pages 1044–1049. The AAAI Press/MIT Press.

Suchanek, Fabian M., Gjergji Kasneci, and Gerhard Weikum. 2007. Yago: A Core of Semantic Knowledge. In *16th international World Wide Web conference (WWW 2007)*, New York, NY, USA. ACM Press.

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

Uszkoreit, Hans, 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*. Springer.

Xu, Feiyu, Hans Uszkoreit, and Hong Li. 2007. A seed-driven bottom-up machine learning framework for extracting relations of various complexity. In *Proceedings of ACL 2007, 45th Annual Meeting of the Association for Computational Linguistics*, Prague, Czech Republic, June.

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

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

Yangarber, Roman. 2003. Counter-training in discovery of semantic patterns. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 343–350, Sapporo Convention Center, Sapporo, Japan, July.