Chapter 10

Application to an Accident Domain

10.1 Applying RHB$^+$

This section describes results of applying ψ-RHB$^+$ to the learning of IE rules in an accident domain [51].

10.1.1 Setting of Experiments

Articles relating to accidents were extracted from Mainichi Newspapers articles published in 1992. Forty two articles are related to accidents that resulted in some numbers of death and injury. The sentences were parsed with the commercial-quality parser of a machine translation system [20], and semantically analyzed to extract the case frames in the sentences. Twenty five articles were well parsed and semantically analyzed. Case frames were converted into literals and two kinds of positive examples were created from those literals. One was death_toll which represents the number of deaths in each article and the other was injury which represents the number of injuries in an article. In pre-possessing, literals unrelated to positive examples were removed. The result was that three literals for each article were selected as background knowledge. The data set consisted of 25 positive examples. The background knowledge contained 78 literals with 26 kinds of predicates and 124 sorts and constants.

10.1.2 Results

Table 10.1: Learning Results of "death toll"

Learner	Time (sec)	$\|Hypo\|$	$\|\hat{E}^+\| / \|Q(T)\|$
FOIL	587.4	3	25/25
Progol	59.1	3	25/26
RHB$^+$	172.7	4	25/25

Table 10.2: Learning Results of "injury"

Learner	Time (sec)	$\|Hypo\|$	$\|\hat{E}^+\| / \|Q(T)\|$
FOIL	1684.5	9	24/24
Progol	346.9	9	25/37
RHB$^+$	508.5	8	25/25

Table 10.1 and Table 10.2 show the learning results of FOIL, Progol and RHB$^+$. $|Hypo|$ shows the number of clauses in hypotheses $Hypo$. T is the union of the hypotheses and background knowledge. $|\hat{E}^+|/|Q(T)|$ shows the MCR, the ratio between covered positives and the empirical content of T, that is, the size of the set of all instances of the head provable from T. Handling sort information seriously degraded the learning speed of FOIL. Progol was relatively fast; however it's results were over-general in both experiments. RHB$^+$ recorded both good performance and appropriate generality of the outputs.

Because of space limits, only the learning results from data "death toll" are shown here as follows.

- **FOIL Result**

```
death_toll(A,B) :- dead(A,B,C,D).
death_toll(A,B) :- confirmed(A,C,B), anything(B).
death_toll(A,B) :- null(B), is_article_1(A).
```

• Progol Result

```
death_toll(a1,'').
death_toll(A,B) :- dead(A,B,C,D).
death_toll(A,B) :- confirmed(A,C,B).
```

•RHB$^+$ Result

```
death_toll( A:any,
            _B:number ) :-
     dead( _A,
           _B,
           any,
           any).
death_toll( a1, '').
death_toll( _A:a4, _B:four) :-
    confirmed( _A, accident, _B).
death_toll( a23, 'driver').
```

10.2 Applying ψ-RHB$^+$

This section describes results of applying ψ-RHB$^+$ to the learning of IE rules in an accident domain [55].

10.2.1 Setting of Experiments

For the purpose of estimating the performance of our system, experiments on the learning of IE rules were conducted. The IE tasks here involved the MUC-4 style IE, and the template elements to be filled had two items [1]. Articles related to accidents were extracted from a one-year newspaper corpus written in Japanese. Forty-two articles were related to accidents which resulted in some deaths and injuries. The template we used consisted of two slots: the number of deaths and injuries. One template was filled for each article. After parsing the sentences, tagged parse trees were converted into atomic formulae representing case frames.

[1] This is a relatively simple setting compared to state-of-the-art IE tasks.

Table 10.3: Comparison between RHB$^+$ and ψ-RHB$^+$

[deaths]	Time (sec)	$\|Hypo\|$	$\|\hat{E}^+\|$ / $\|Q(T)\|$
RHB$^+$	172.7	4	25/25
ψ-RHB$^+$	954.2	4	25/25

[injuries]	Time (sec)	$\|Hypo\|$	$\|\hat{E}^+\|$ / $\|Q(T)\|$
RHB$^+$	508.5	8	25/25
ψ-RHB$^+$	3218.0	10	25/25

As shown in Figure 8.1, case frames were given to our learner as background knowledge. *All of the 42 articles were able to be represented as case frames for the sake of the representation power of ψ-terms, while only 25 articles were able to be represented using τ-terms.* Each slot of a filled template was given as a positive example. For the precision and recall, the standard evaluation metrics for IE tasks, Evaluation method is four-fold cross validation on the 42 examples.

10.2.2 Results

Table 10.3 and Table 10.4 show the results of our experiments. Table 10.3 shows the experimental results of RHB$^+$ and ψ-RHB$^+$ using the same 25 examples as used in [51]. Test bed is a SparcStation 20 for this experiment. ψ-RHB$^+$ showed a high accuracy like RHB$^+$ but slowed down in exchange for its extended representation power in the hypothesis language.

Table 10.4 shows the experimental results on forty two examples. An AlphaStation 500/333MHz is used for this experiment. Overall, a very high precision, 90-97%, was achieved. 63-80% recall was achieved with all case frames including errors in the case selection and semantic tag selection. These selections had an error range of 2-7%. With only correct case frames, 67-88% recall was achieved.

It is important to note that the extraction of two different pieces of information showed good results. This indicates that our learner has high potential in IE tasks.

Table 10.4: Learning Results of Accidents

	deaths	injuries
Precision (all case frames)	96.7%	89.9%
Recall (all case frames)	80.0%	63.2%
Recall (correct case frames)	87.5%	66.7%
F-measure (all case frames)	87.6%	74.2%
F-measure (correct case frames)	91.9%	76.6%
Average time (sec.)	966.8	828.0

10.3 Discussion

The benefits of ψ-term capability, not just τ-term capability, depend on the writing style of the topic. In English, the expression "ABC Corp.'s printer" is commonly used and the logical term representation can be $printer(pos \Rightarrow$ "ABC Corp."$)$. However, if the expression "ABC Corp. released a printer and ..." were very common, the case frame could be *release*("ABC Corp.", *printer*). In this case, since the required representation is within the τ-term, extending the language from τ-term based to ψ-term based does not pay for the higher computing cost.

INDIE [6] learns a set of feature terms equivalent to ψ-terms. The learning is equivalent to the learning of a set of atomic formulae based on ψ-terms which cover all positive examples and no negative examples. Because its hypotheses are generated so as to exclude any negatives, it might be intolerant of noise.

The experimental results of applying ψ-RHB$^+$ and RHB$^+$ to the learning of information extraction rules from small amount of data showed very high precision and good recall. Recall will be improved by increasing the size of data because the two ILP systems could learn IE rules that cover rare cases which will be newly included in larger data.

There is also room for improving learning algorithms of the extended ILP systems. Currently, the extended ILP systems employ the greedy covering algorithm that does not guarantee optimal results. More advanced algorithms would achieve higher recall and precision of learned

IE rules.

10.4 Summary

This chapter described a use of semantic representations for generating information extraction rules by applying a hierarchically sorted ILP system. Experiments were conducted on the data generated from 100 news articles in the domain of accidents. ψ-RHB$^+$ showed very high precision of 90%-97%, recall of 63-80% with all semantic representations, and 67-88% recall with correct semantic representations. RHB$^+$ also showed both good performance and appropriate generality of the learned results.

Because of the modest robustness and performance of current natural language analysis techniques (for Japanese texts), errors were found in parsing, case selection, and semantic tag selection. The experimental results, however, show that learned rules achieve high precision and recall in IE tasks. Moreover, an important point is that all of the 42 articles related to the topic were able to be represented as case frames, which demonstrates the representation power of ψ-terms. This indicates that applying ILP to the learning from case frames will become more practical as natural language processing techniques progress in the near future.