

Prediction of Thematic Rank for Structured Semantic Role Labeling

Weiwei Sun and Zhifang Sui and Meng Wang

Institute of Computational Linguistics

Peking University

Key Laboratory of Computational Linguistics

Ministry of Education, China

weiwusun@gmail.com; {wm, szf}@pku.edu.cn

Abstract

In Semantic Role Labeling (SRL), it is reasonable to globally assign semantic roles due to strong dependencies among arguments. Some relations between arguments significantly characterize the structural information of argument structure. In this paper, we concentrate on thematic hierarchy that is a rank relation restricting syntactic realization of arguments. A log-linear model is proposed to accurately identify thematic rank between two arguments. To import structural information, we employ re-ranking technique to incorporate thematic rank relations into local semantic role classification results. Experimental results show that automatic prediction of thematic hierarchy can help semantic role classification.

1 Introduction

In Semantic Role Labeling (SRL), it is evident that the arguments in one sentence are highly correlated. For example, a predicate will have no more than one *Agent* in most cases. It is reasonable to label one argument while taking into account other arguments. More structural information of all arguments should be encoded in SRL approaches.

This paper explores structural information of predicate-argument structure from the perspective of rank relations between arguments. Thematic hierarchy theory argues that there exists a language independent rank of possible semantic roles, which establishes priority among arguments with respect to their syntactic realization (Levin and Hovav, 2005). This construct has been widely implicated in linguistic phenomena, such as in the subject selection rule of Fillmore's Case Grammar (1968): "If there is an A [=Agent], it becomes the subject; otherwise, if there is an I [=Instrument],

it becomes the subject; otherwise, the subject is the O [=Object, i.e., Patient/Theme]". This rule implicitly establishes precedence relations among semantic roles mentioned and can be simplified to:

$$Agent \succ Instrument \succ Patient/Theme$$

Emerging from a range of more basic semantic properties of the ranked semantic roles, thematic hierarchies can help to construct mapping from semantics to syntax. It is therefore an appealing option for argument structure analysis. For example, if the the rank of argument a_i is shown higher than a_j , then the assignment [$a_i=Patient$, $a_j=Agent$] is illegal, since the role *Agent* is the highest role.

We test the hypothesis that thematic rank between arguments can be accurately detected by using syntax clues. In this paper, the concept "thematic rank" between two arguments a_i and a_j means the relationship that a_i is prior to a_j or a_j is prior to a_i . Assigning different labels to different relations between a_i and a_j , we formulate prediction of thematic rank between two arguments as a multi-class classification task. A log-linear model is put forward for classification. Experiments on CoNLL-2005 data show that this approach can get an good performance, achieving 96.42% accuracy on gold parsing data and 95.14% accuracy on Charniak automatic parsing data.

Most existing SRL systems divide this task into two subtasks: Argument Identification (AI) and Semantic Role Classification (SRC). To add structural information to a local SRL approach, we incorporate thematic hierarchy relations into local classification results using re-ranking technique in the SRC stage. Two re-ranking approaches, 1) hard constraint re-ranking and 2) soft constraint re-ranking, are proposed to filter out unlike global semantic role assignment. Experiments on CoNLL-2005 data indicate that our method can yield significant improvement over a state-of-the-art SRC baseline, achieving 0.93% and 1.32%

absolute accuracy improvements on hand-crafted and automatic parsing data.

2 Prediction of Thematic Rank

2.1 Ranking Arguments in PropBank

There are two main problems in modeling thematic hierarchy for SRL on PropBank. On the one hand, there is no consistent meaning of the core roles (i.e. Arg0-5/ArgA). On the other hand, there is no consensus over hierarchies of the roles in the thematic hierarchy. For example, the *Patient* occupies the second highest hierarchy in some linguistic theories but the lowest in some other theories (Levin and Hovav, 2005).

In this paper, the proto-role theory (Dowty, 1991) is taken into account to rank PropBank arguments, partially resolving the two problems above. There are three key points in our solution. First, the rank of Arg0 is the highest. The *Agent* is almost without exception the highest role in proposed hierarchies. Though PropBank defines semantic roles on a verb by verb basis, for a particular verb, Arg0 is generally the argument exhibiting features of a prototypical *Agent* while Arg1 is a prototypical *Patient* or *Theme* (Palmer et al., 2005). As being the proto-Agent, the rank of Arg0 is higher than other numbered arguments. Second, the rank of the Arg1 is second highest or lowest. Both hierarchy of Arg1 are tested and discussed in section 4. Third, we do not rank other arguments.

Two sets of roles closely correspond to numbered arguments: 1) referenced arguments and 2) continuation arguments. To adapt the relation to help these two kinds of arguments, the equivalence relation is divided into several sub-categories. In summary, relations of two arguments a_i and a_j in this paper include: 1) $a_i \succ a_j$: a_i is higher than a_j , 2) $a_i \prec a_j$: a_i is lower than a_j , 3) $a_i ARa_j$: a_j is the referenced argument of a_i , 4) $a_i RAa_j$: a_i is the referenced argument of a_j , 5) $a_i ACA_j$: a_j is the continuation argument of a_i , 6) $a_i CAa_j$: a_i is the continuation argument of a_j , 7) $a_i = a_j$: a_i and a_j are labeled as the same role label, and 8) $a_i \sim a_j$: a_i and a_j are labeled as the Arg2-5, but not in the same type.

2.2 Prediction Method

Assigning different labels to possible rank between two arguments a_i and a_j , such as labeling $a_i \succ a_j$ as " \succ ", identification of thematic rank can be formulated as a classification problem. De-

<i>lemma</i> , <i>POS Tag</i> , <i>voice</i> , and <i>SCF</i> of predicate
<i>categories</i> , <i>position</i> of two arguments; <i>rewrite rules</i> expanding subroots of two arguments
<i>content</i> and <i>POS tags</i> of the boundary words and head words
<i>category path</i> from the predicate to candidate arguments
<i>single character category path</i> from the predicate to candidate arguments
<i>conjunction of categories</i> , <i>position</i> , <i>head words</i> , <i>POS of head words</i> <i>category</i> and <i>single character category path</i> from the first argument to the second argument

Table 1: Features for thematic rank identification.

note the set of relations \mathcal{R} . Formally, given a score function $S_{TH} : \mathcal{A} \times \mathcal{A} \times \mathcal{R} \mapsto \mathbb{R}$, the relation r is recognized in argmax flavor:

$$\hat{r} = r^*(a_i, a_j) = \arg \max_{r \in \mathcal{R}} S_{TH}(a_i, a_j, r)$$

A probability function is chosen as the score function and the log-linear model is used to estimate the probability:

$$S_{TH}(a_i, a_j, r) = \frac{\exp\{\psi(a_i, a_j, r) \cdot \mathbf{w}\}}{\sum_{r \in \mathcal{R}} \exp\{\psi(a_i, a_j, r) \cdot \mathbf{w}\}}$$

where ψ is the feature map and \mathbf{w} is the parameter vector to learn. Note that the model predicts the rank of a_i and a_j through calculating $S_{TH}(a_i, a_j, r)$ rather than $S_{TH}(a_j, a_i, r)$, where a_i precedes a_j . In other words, the position information is implicitly encoded in the model rather than explicitly as a feature.

The system extracts a number of features to represent various aspects of the syntactic structure of a pair of arguments. All features are listed in Table 1. The *Path* features are designed as a sequential collection of phrase tags by (Gildea and Jurafsky, 2002). We also use *Single Character Category Path*, in which each phrase tag is clustered to a category defined by its first character (Pradhan et al., 2005). To characterize the relation between two constituents, we combine features of the two individual arguments as new features (i.e. conjunction features). For example, if the category of the first argument is *NP* and the category of the second is *S*, then the *conjunction of category* feature is *NP-S*.

3 Re-ranking Models for SRC

Toutanova et al. (2008) empirically showed that global information is important for SRL and that

structured solutions outperform local semantic role classifiers. Punyakanok et al. (2008) raised an inference procedure with integer linear programming model, which also showed promising results.

Identifying relations among arguments can provide structural information for SRL. Take the sentence "[*Arg0* She] [*V* addressed] [*Arg1* her husband] [*ArgM-MNR* with her favorite nickname]." for example, if the thematic rank of *she* and *her husband* is predicted as that *she* is higher than *her husband*, then *her husband* should not be assigned the highest role.

To incorporate the relation information to local classification results, we employ re-ranking approach. Assuming that the local semantic classifier can produce a list of labeling results, our system then attempts to pick one from this list according to the predicted ranks. Two different policies are implemented: 1) hard constraint re-ranking, and 2) soft constraint re-ranking.

Hard Constraint Re-ranking The one picked up must be strictly in accordance with the ranks. If the rank prediction result shows the rank of argument a_i is higher than a_j , then role assignments such as [$a_i=Patient$ and $a_j=Agent$] will be eliminated. Formally, the score function of a global semantic role assignment is:

$$S(\mathbf{a}, \mathbf{s}) = \prod_i S_l(a_i, s_i) \prod_{i,j,i < j} I(r^*(a_i, a_j), r(s_i, s_j))$$

where the function S_l locally scores an argument; $r^* : \mathcal{A} \times \mathcal{A} \mapsto \mathcal{R}$ is to predict hierarchy of two arguments; $r : \mathcal{S} \times \mathcal{S} \mapsto \mathcal{R}$ is to point out the thematic hierarchy of two semantic roles. For example, $r(Agent, Patient) = \succ$. $I : \mathcal{R} \times \mathcal{R} \mapsto \{0, 1\}$ is identity function.

In some cases, there is no role assignment satisfies all predicted relations because of prediction mistakes. For example, if the hierarchy detection result of $\mathbf{a} = (a_1, a_2, a_3)$ is ($r^*(a_1, a_2) = \succ$, $r^*(a_2, a_3) = \succ$, $r^*(a_1, a_3) = \prec$), there will be no legal role assignment. In these cases, our system returns local SRL results.

Soft Constraint Re-ranking In this approach, the predicted confidence score of relations is added as factor items to the score function of the semantic role assignment. Formally, the score function in soft constraint re-ranking is:

$$S(\mathbf{a}, \mathbf{s}) = \prod_i S_l(a_i, s_i) \prod_{i,j,i < j} S_{TH}(a_i, a_j, r(s_i, s_j))$$

4 Experiments

4.1 Experimental Settings

We evaluated our system using the CoNLL-2005 shared task data. Hierarchy labels for experimental corpora are automatically set according to the definition of relation labels described in section 2.1. Charniak parser (Charniak, 2000) is used for POS tagging and full parsing. UIUC Semantic Role Labeler¹ is a state-of-the-art SRL system. Its argument classification module is used as a strong local semantic role classifier. This module is re-trained in our SRC experiments, using parameters described in (Koomen et al., 2005). Experiments of SRC in this paper are all based on good argument boundaries which can filter out the noise raised by argument identification stage.

4.2 Which Hierarchy Is Better?

	Detection	SRL (S)	SRL (G)
Baseline	–	94.77%	–
A	94.65%	95.44%	96.89%
A & P \uparrow	95.62%	95.07%	96.39%
A & P \downarrow	94.09%	95.13%	97.22%

Table 2: Accuracy on different hierarchies

Table 2 summarizes the performance of thematic rank prediction and SRC on different thematic hierarchies. All experiments are tested on development corpus. The first row shows the performance of the local semantic role classifier. The second to the fourth rows show the performance based on three ranking approach. A means that the rank of *Agent* is the highest; P \uparrow means that the rank of *Patient* is the second highest; P \downarrow means that the rank of the *Patient* is the lowest. Column *SRL(S)* shows SRC performance based on soft constraint re-ranking approach, and column *SRL(G)* shows SRC performance based on gold hierarchies. The data shows that the third thematic hierarchy fits SRL best, but is harder to learn. Compared with P \uparrow , P \downarrow is more suitable for SRL. In the following SRC experiments, we use the first hierarchy because it is most helpful when predicted relations are used.

4.3 Results And Improvement Analysis

Table 3 summarizes the precision, recall, and F-measure of this task. The second column is frequency of relations in the test data, which can be

¹<http://l2r.cs.uiuc.edu/~cogcomp/srl-demo.php>

seen as a simple baseline. Moreover, another natural baseline system can predict hierarchies according to the roles classified by local classifier. For example, if the a_i is labeled as Arg0 and a_j is labeled as Arg2, then the relation is predicted as \succ . The third column *BL* shows the F-measure of this baseline. It is clear that our approach significantly outperforms the two baselines.

Rel	Freq.	BL	P(%)	R(%)	F
\succ	57.40	94.79	97.13	98.33	97.73
\prec	9.70	51.23	98.52	97.24	97.88
\sim	23.05	13.41	94.49	93.59	94.04
=	0.33	19.57	93.75	71.43	81.08
AR	5.55	95.43	99.15	99.72	99.44
AC	3.85	78.40	87.77	82.04	84.81
CA	0.16	30.77	83.33	50.00	62.50
All	–	75.75		96.42	

Table 3: Thematic rank prediction performance

Table 4 summarizes overall accuracy of SRC. Baseline performance is the overall accuracy of the local classifier. We can see that our re-ranking methods can yield significant improvements over the baseline.

	Gold	Charniak
Baseline	95.14%	94.12%
Hard	95.71%	94.74%
Soft	96.07%	95.44%

Table 4: Overall SRC accuracy.

Hierarchy prediction and re-ranking can be viewed as modification for local classification results with structural information. Take the sentence "[Some 'circuit breakers' installed after the October 1987] crash failed [their first test]." for example, where phrases "*Some ... 1987*" and "*their ... test*" are two arguments. The table below shows the local classification result (column *Score(L)*) and the rank prediction result (column *Score(H)*). The baseline system falsely assigns roles as *Arg0+Arg1*, the rank relation of which is \succ . Taking into account rank prediction result that relation \sim gets a extremely high probability, our system returns *Arg1+Arg2* as SRL result.

Assignment	Score(L)	Score(H)
Arg0+Arg1	78.97% × 82.30%	\succ :0.02%
Arg1+Arg2	14.25% × 11.93%	\sim :99.98%

5 Conclusion and Future Work

Inspired by thematic hierarchy theory, this paper concentrates on thematic hierarchy relation which

characterize the structural information for SRL. The prediction of thematic rank is formulated as a classification problem and a log-linear model is proposed to solve this problem. To improve SRC, we employ re-ranking technique to incorporate thematic rank information into the local semantic role classifier. Experimental results show that our methods can construct high-performance thematic rank detector and that identification of arguments' relations can significantly improve SRC.

Acknowledgments

This work is supported by NSFC Project 60873156, 863 High Technology Project of China 2006AA01Z144 and the project of Toshiba (China) Co., Ltd. R&D Center.

References

- Eugene Charniak. 2000. A Maximum-Entropy-Inspired Parser. In *Proceedings of NAACL-00*.
- David R. Dowty. 1991. Thematic proto-roles and argument selection. *Language*, 67:547–619.
- Charles Fillmore. 1968. The case for case. In Emmon Bach and Richard Harms, editors, *Universals in Linguistic Theory*, pages 1–90. Holt, Rinehart and Winston, New York, New York.
- Daniel Gildea and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Computational Linguistics*, 28:245–288.
- Peter Koomen, Vasin Punyakanok, Dan Roth, and Wen-tau Yih. 2005. Generalized inference with multiple semantic role labeling systems. In *Proceedings of the CoNLL-2005*, pages 181–184, June.
- Beth Levin and Malka Rappaport Hovav. 2005. *Argument Realization*. Research Surveys in Linguistics. Cambridge University Press, New York.
- Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31.
- Sameer Pradhan, Kadri Hacioglu, Valerie Krugler, Wayne Ward, James H. Martin, and Daniel Jurafsky. 2005. Support vector learning for semantic argument classification. In *Machine Learning*.
- Vasin Punyakanok, Dan Roth, and Wen-tau Yih. 2008. The importance of syntactic parsing and inference in semantic role labeling. *Comput. Linguist.*
- Kristina Toutanova, Aria Haghighi, and Christopher D. Manning. 2008. A global joint model for semantic role labeling. *Comput. Linguist.*