Semantic Role Labeling using Maximum Entropy Model

Joon-Ho Lim, Young-Sook Hwang, So-Young Park, Hae-Chang Rim
Department of Computer Science & Engineering Korea University
5-ka, Anam-dong, SEOUL, 136-701, KOREA
{jhlim, yshwang, ssoya, rim}@nlp.korea.ac.kr

Abstract

In this paper, we propose a semantic role labeling method using a maximum entropy model, which enables not only to exploit rich features but also to alleviate the data sparseness problem in a well-founded model. For applying the maximum entropy model to semantic role labeling, we take a incremental approach as follows: firstly, the semantic roles are assigned to the arguments in the immediate clause including a predicate, and then, the semantic roles are assigned to the arguments in the upper clauses by using previously assigned labels. The experimental result shows that the proposed method has about 64.76% (F1-measure) on the test set.

1 Introduction

The semantic role represents the relationship between a predicate and an argument. It provides a general semantic interpretation of the sentence, and it can play a key role in NLP. The shared task of CoNLL-2004 concerns the automatic semantic role labeling (Carreras, 2004). The challenge for this task is to come forward with machine learning approaches which based on only partial syntactic information such as words, POS tags, chunks, clauses, and named entities.

Some machine learning approaches for semantic role labeling have been previously developed (Gildea, 2002; Pradhan, 2003; Thompson, 2003). Gildea (2002) proposed a probabilistic discriminative model to assign a semantic roles to the constituent. However, it needs a complex interpolation for smoothing because of the data sparseness problem. Pradhan (2003) applied a support vector machine to semantic role labeling, but if it use a polynomial kernel function for the dependencies between features, it requires high computational complexity. Futhermore, becuase the SVM is a binary classifier, one-vs-rest or pairwise method is required for multi-class classification. Thompson (2003) proposed a probabilistic generative model which the constituents is generated by the semantic roles. In this model, because a constituent depends only on the role that generated it, and constituents are independent of each other, so this model can not utilize contextual information or a relational information between the constituent and the predicate.

In this paper, we propose a semantic role labeling method using a maximum entropy model. It is motivated by the thought of that for building a successful model, some knowledge of the task are reflected into the model based on the machine learning technique. In this method, we try to combine the structural linguistic knowledge linking syntax to semantics into the machine learning technique. It is realized in terms of two aspects: one is the model framework, the other is the design of feature sets. First of all, for the model framework, we utilize the syntactic knowledge of representing the semantic roles in a clause: the arguments of a predicate are located in the immediate clause or the upper clauses. Secondly, for the feature sets, we consider the relation between syntactic and semantic characteristics of a given context. For implementing the method with a machine learning algorithm, we take a maximum entropy model, which enables not only to exploit rich features but also to alleviate the data sparseness problem in a well-founded model.

The remaining of the paper is organized as follows: section 2 describes the proposed semantic role labeling method using a maximum entropy model. Section 3 presents feature sets for semantic role labeling. Section 4 shows some experimental results of the proposed method. Finally, section 5 concludes with some directions of future works.

2 Semantic Role Labeling using ME

In the maximum entropy framework (Berger, 1996), the conditional probability of predicting an outcome y given

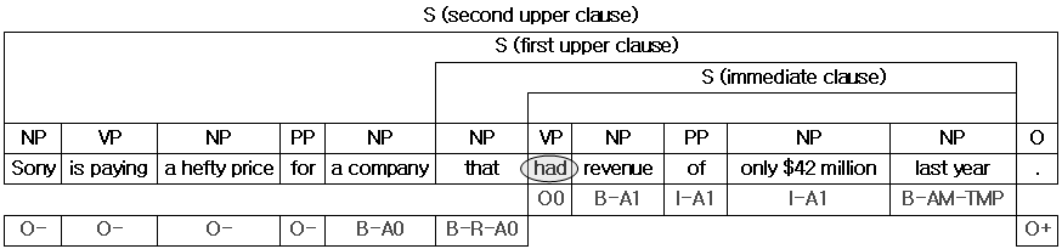

Figure 1: An example of the semantic role labels and an incremental approach.

a history *x* is defined as follows :

$$P(y|x) = \frac{1}{Z(x)} exp\left(\sum_{i=1}^{k} \lambda_i f_i(x,y)\right)$$

where $f_i(x,y)$ is the feature function, λ_i is the weighting parameter of $f_i(x,u)$, *k* is the number of features, and *Z(x)* is the normalization factor for $\sum_y p(y|x) = 1$.

Given a predicate and its partial parse tree represented by constituents such as chunks and clauses, the probabilistic model for semantic role labeling assigns the semantic role labels to the constituents as described in the equation (1).

$$\begin{aligned} R_{best} &= \text{argmax}_R P(R|c_{1n}, pred) \\ &= \text{argmax}_R \prod_{i=1}^{n} P(r_i|c_{1n}, pred, r_{1...i-1}) \end{aligned} \quad (1)$$

where R is a sequence of the semantic roles, c_{1n} is a sequence of constituents, $pred$ is the given predicate, r_i is the *i*-th semantic role, n is the number of constituents.

In order to apply the equation (1) to an incremental approach, we classify clauses into the *immediate clause* and the *upper clause*. The *immediate clause* is the clause which contains the target predicate, and the *upper clause* is the clause which includes the immediate clause. Generally, most of the arguments of the predicate are located in the immediate clause while some of them are located in the upper clauses, especially the first or second upper clauses. Since it is much easier and more reliable to identify the arguments in the immediate clause, the proposed method first assigns the semantic role labels to the constituents [1] in the immediate clause. Then, it assigns the semantic role labels to the constituents in the upper clauses by using previously assigned labels. This incremental approach is described in the equation (2) derived from the equation (1).

$$\begin{aligned} R_{best} &= \text{argmax}_R \prod_{i=1}^{n} P(r_i|c_{1n}, pred, r_{1...i-1}) \\ &\approx \text{argmax}_R \prod_{i=1}^{m} P(r_i|\Phi_1(c_{1n}, pred, r_{1...i-1})) \\ &\quad \times \prod_{i=m+1}^{n} P(r_i|\Phi_2(c_{1n}, pred, r_{1...i-1})) \end{aligned} \quad (2)$$

where m is the number of constituents covered by the immediate clause, Φ_1 is a feature set for immediate clause, and Φ_2 is a feature set for upper clauses.

A semantic role label(r_i) is represented by using a BIO notation such as *B-A**, *I-A**, etc. However, *O* is too frequently occurred than other semantic role labels, it can have a somewhat high probability than others. Therefore, to degrade its probability, we divide the single *O* into *O-*, *O+*, *O0* with respect to the position of a constituent which is relative to the predicate. Therefore, *B-A**, *I-A**, *O-*, *O+*, and *O0* are used as semantic roles as shown in Figure 1.

After processing the equation (2), we use some heuristic to attach the some semantic roles and to adjust the boundary of semantic arguments in the post-processing step. More specifically, we use some rules to attach the V, AM-MOD, and AM-NEG, and extend the boundary of core roles to include *to infinitive* of the VP chunk like "*expect/B-VP (A1 to/I-VP take/I-VP dive/B-NP)*".

3 Feature Sets for Semantic Role Labeling

For accurate semantic role labeling, we regard that the following information is important: the contextual information of the constituent, the syntactic information of the predicate, and the relation between the constituent and the predicate. Therefore, we use the features presented in Table 1 for semantic role labeling. For example, Figure

previous-label(*pl*)
predicate-POS(*predpos*), predicate-lex(*predlex*)
predicate-type(*predtype*)
tag(*ctag*), voice(*v*), position(*p*), path(*path*)
head-lex(*hl*), head-POS(*hp*), content-head(*chl*)
prev-tag(*ptag*), prev-head-lex(*phl*)
next-tag(*ntag*), next-head-lex(*nhl*)
path-immediate-clause(*path-im-cl*)
path-begin-end(*path-beg-end*)
level-of-clause(*l-cl*), is-clause-boundary(*cl-bn*)
immediate-clause-roles(*im-cl-roles*)

Table 1: Features for semantic role labeling.

[1] Here, we regard a chunk or a clause as a constituent.

	role-label	pl	predlex	predtype	ctag	v	p	hl	path	l-cl	cl-bn	path-target-clause
Immediate clause	O0	BEG	have	BEG	VP/p	act	0	had	VP/p			
	B-A1	O0	have	BEG	NP	act	+	revenue	VP/p+NP			
	I-A1	B-A1	have	BEG	PP	act	+	of	VP/p+NP+PP			
	I-A1	I-A1	have	BEG	NP	act	+	$	VP/p+NP+PP+NP			
	B-AM-TMP	I-A1	have	BEG	NP	act	+	year	VP/p+NP+PP+NP+NP			
Upper clauses	O-	BEG	have	BEG	NP	act	-	Sony		2	O	NP+VP+NP+PP+NP+NP+BEG
	O-	O-	have	BEG	VP	act	-	is		2	X	VP+NP+PP+NP+NP+BEG
	O-	O-	have	BEG	NP	act	-	price		2	X	NP+PP+NP+NP+BEG
	O-	O-	have	BEG	PP	act	-	for		2	X	PP+NP+NP+BEG
	B-A0	O-	have	BEG	NP	act	-	company		2	X	NP+NP+BEG
	B-R-A0	B-A0	have	BEG	NP	act	-	that		1	O	NP+BEG
	O+	BEG	have	BEG	O	act	+	.		2	X	BEG+O

Figure 2: Some instances extracted from example of Figure 1.

feature set Φ_1 for immediate clause	feature set Φ_2 for upper clause
pl, ctag, ctag+v+p, ctag+v+p+pl	pl, ctag, ctag+v+p
ptag+ctag, ctag+ntag	ptag+ctag, ctag+ntag
hp+p, hp+p+ntag	hp+p, hl+ctag,
predtype+ctag	predtype+ctag
predlex+hl, predlex+ctag+v+p, predlex+ctag+pl	predlex+hl, predlex+ctag+v+p, predlex+ctag+pl
predpos+p, predpos+hp+pl, predpos+ctag	
path, path+hp+v, path+nhl, path+predlex	path-im-cl, path-im-cl+ctag+v, path-beg-end
hl+p, hl+ctag, hl+ctag+predlex	ctag+l-cl, ptag+ctag+l-cl, ctag+ntag+l-cl
chl+pl, chl+pl+predlex	ctag+cl-bn, ptag+ctag+cl-bn, ctag+ntag+cl-bn
chl+phl, chl+phl+predlex	im-cl-roles

Table 2: Conjoined Feature Sets

2 shows how the features in Table 1 are used for labeling semantic roles to the proposition in Figure 1.

Because the maximum entropy model assumes the independence of features, we should conjoin the coherent features. As presented in Table 2, we use the conjoined feature sets to assign semantic roles to the constituents of the immediate clause and the upper clauses.

The predicate-type feature represents the predicate usage such as to-infinitive form (*TO*), the beginning of the immediate clause (*BEG*), and otherwise (*SEN*). The tag feature represents the tag of the current constituent. If it is a clause, it is subdivided into a relative pronoun, a infinitival relative clause, etc according to its represented form.

The path feature indicates the sequence of constituent tags between the current constituent and the predicate. The voice feature is determined to be an active or passive voice of the predicate, and the position feature is assigned by the constituent position with respect to predicate. These features implicitly represent the predicate-argument relation such as predicate-subject or predicate-object.

For the headword feature, we use the Collins' headword rules, and as a complementary feature to the head word feature, a content word feature[2] is used to represent the content of the PP, VP, or CONJP chunk.

The path-immediate-clause feature is the sequence of constituent tags between the current constituent and the immediate clause, and the path-begin-end feature is the sequence between current constituent and beginning/end of clause. The level-of-clause feature indicates whether the current constituent is located in the first upper clause or in the second upper clause, and the is-clause-boundary feature is the binary value which indicates the existence of the starting clause. The immediate-clause-roles features are the binary indicators to represent whether the core arguments exist in the immediate clause or not.

The path-immediate-clause, path-begin-end, level-of-clause, is-clause-boundary, and immediate-clause-roles features are used only in the second phase, and the others except the path feature and the content-head feature are used in common.

[2]For example, if the PP-chunk is *because of*, the headword feature is *of*, and the content word feature is *because*.

	Precision	Recall	$F_{\beta=1}$
Overall	68.42%	61.47%	64.76
A0	79.20%	75.73%	77.42
A1	67.41%	64.65%	66.00
A2	52.65%	45.94%	49.07
A3	52.53%	34.67%	41.77
A4	63.16%	48.00%	54.55
A5	0.00%	0.00%	0.00
AM-ADV	46.75%	35.18%	40.15
AM-CAU	57.69%	30.61%	40.00
AM-DIR	48.28%	28.00%	35.44
AM-DIS	60.11%	50.23%	54.73
AM-EXT	58.33%	50.00%	53.85
AM-LOC	35.56%	35.09%	35.32
AM-MNR	51.26%	23.92%	32.62
AM-MOD	89.77%	91.10%	90.43
AM-NEG	86.15%	88.19%	87.16
AM-PNC	48.98%	28.24%	35.82
AM-PRD	100.00%	33.33%	50.00
AM-TMP	59.51%	42.70%	49.73
R-A0	86.96%	75.47%	80.81
R-A1	57.89%	62.86%	60.27
R-A2	50.00%	33.33%	40.00
R-A3	0.00%	0.00%	0.00
R-AM-LOC	33.33%	25.00%	28.57
R-AM-MNR	0.00%	0.00%	0.00
R-AM-PNC	0.00%	0.00%	0.00
R-AM-TMP	42.86%	21.43%	28.57
V	97.99%	97.99%	97.99

Table 3: Experimental results on the test set.

4 Experiments

To test the proposed method, we have experimented on CoNLL-2004 datasets. For our experiments, we use the Zhang le's MaxEnt toolkit [3], and the L-BFGS parameter estimation algorithm with Gaussian Prior smoothing (Chen, 1999). The results on the test set are shown in Table 3, and Table 4 shows the overall results when the model is tested on the training set, the development set, and the test set.

From these experimental results, we can find that the proposed model has relatively high performance on the labels related to A0 and A1, while it has relatively low performance on the other labels. This may be caused by following two reasons. Firstly, the instances of A0 or A1 are provided enough for accurate semantic role labeling. Secondly, the thematic roles of A0 and A1 are more clear than other core semantic roles. For example, *agent* is labeled as mainly A0 while *benefactive* can be labeled as A2 or A3. Therefore, the maximum entropy model can get a good generalize performance in case of A0 or A1, but can't generalize well in other cases.

[3]http://www.nlplab.cn/zhangle/maxent_toolkit.html

	Precision	Recall	$F_{\beta=1}$
Overall(training)	96.40%	92.28%	94.29
Overall(dev)	69.78%	62.56%	65.97
Overall(test)	68.42%	61.47%	64.76

Table 4: The results when the model is tested on the training set, the development set, and the test set

5 Conclusion

In this paper, we propose a semantic role labeling method using a maximum entropy model. Because the maximum entropy model enables not only to exploit rich features but also to alleviate the data sparseness problem, we use it to model the probability of a semantic role label sequence. The proposed method has following characteristics: firstly, it assigns the semantic role labels to the constituents in the immediate clause, and then assigns role labels to the constituents in the upper clauses, and it utilizes the relation between syntactic and semantic characteristics of a given context.

For the future work, we will device a method of clustering for the path and predicate features, and include the clustering results as additional features.

References

Adam Berger, Stephen Della Pietra, and Vincent Della Pietra. 1996. *A maximum entropy approach to natural language processing* . Computational Linguistics, 22(1):39–71.

Xavier Carreras, and Lluis Marquez. 2004. *Introduction to the CoNLL-2004 Shared Task: Semantic Role Labeling* . Proceedings of CoNLL-2004.

S. Chen and R. Rosenfeld. 1999. *A Gaussian prior for smoothing maximum entropy models* . Technical Report CMUCS-99-108, Carnegie Mellon University.

Daniel Gildea, and Daniel Jurafsky. 2002. *Automatic Labeling of Semantic Roles* . Computational Linguistics, 28(3):245-288.

Sameer Pradhan, Kadri Hacioglu, Valerie Krugler, Wayne Ward, James H. Martin and Daniel Jurafsky. Oct 2003. *Support Vector Learning for Semantic Argument Classification* . Technical Report, TR-CSLR-2003-03.

Cynthia A. Thompson, Roger Levy and Christopher D. Manning. *A Generative Model for Semantic Role Labeling* . ECML 2003, 397-408.