Inferring semantic roles using sub-categorization frames and maximum entropy model

Akshar Bharati, Sriram Venkatapathy and Prashanth Reddy
Language Technologies Research Centre, IIIT - Hyderabad, India.
{sriram,prashanth}@research.iiit.ac.in

Abstract

In this paper, we propose an approach for inferring semantic role using sub-categorization frames and maximum entropy model. Our approach aims to use the sub-categorization information of the verb to label the mandatory arguments of the verb in various possible ways. The ambiguity between the assignment of roles to mandatory arguments is resolved using the maximum entropy model. The unlabelled mandatory arguments and the optional arguments are labelled directly using the maximum entropy model such that their labels are not one among the frame elements of the sub-categorization frame used. Maximum entropy model is preferred because of its novel approach of smoothing. Using this approach, we obtained an F-measure of 68.14% on the development set of the data provided for the CONLL-2005 shared task. We show that this approach performs well in comparison to an approach which uses only the maximum entropy model.

1 Introduction

Semantic role labelling is the task of assigning appropriate semantic roles to the arguments of a verb. The semantic role information is important for various applications in NLP such as Machine Translation, Question Answering, Information Extraction etc. In general, semantic role information is useful for sentence understanding. We submitted our system for closed challenge at CONLL-2005 shared task. This task encourages participants to use novel machine learning techniques suited to the task of semantic role labelling. Previous approaches on semantic role labelling can be classified into three categories (1) Explicit Probabilistic methods (Gildea and Jurafsky, 2002). (2) General machine learning algorithms (Pradhan et al., 2003) (Lim et al., 2004) and (3) Generative model (Thompson et al., 2003).

Our approach has two stages; first, identification whether the argument is mandatory or optional and second, the classification or labelling of the arguments. In the first stage, the arguments of a verb are put into three classes, (1) mandatory, (2) optional or (3) null. Null stands for the fact that the constituent of the verb in the sentence is not an semantic argument of the verb. It is used to rule out the false argument of the verb which were obtained using the parser. The maximum entropy based classifier is used to classify the arguments into one of the above three labels.

After obtaining information about the nature of the non-null arguments, we proceed in the second stage to classify the mandatory and optional arguments into their semantic roles. The propbank sub-categorization frames are used to assign roles to the mandatory arguments. For example, in the sentence ”John saw a tree”, the sub-categorization frame ”A0 v A1” would assign the roles A0 to John and A1 to tree respectively. After using all the sub-categorization frames of the verb irre-

Proceedings of the 9th Conference on Computational Natural Language Learning (CoNLL),
pages 165–168, Ann Arbor, June 2005.

spective of the verb sense, there could be ambiguity in the assignment of semantic roles to mandatory arguments. The unlabelled mandatory arguments and the optional arguments are assigned the most probable semantic role which is not one of the frame elements of the sub-categorization frame using the maximum entropy model. Now, among all the sequences of roles assigned to the non-null arguments, the sequence which has the maximum joint probability is chosen. We obtained an accuracy of 68.14% using our approach. We also show that our approach performs better in comparision to an approach with uses a simple maximum entropy model. In section 4, we will talk about our approach in greater detail.

This paper is organised as follows, (2) Features, (3) Maximum entropy model, (4) Description of our system, (5) Results, (6) Comparison with our other experiments, (7) Conclusion and (8) Future work.

2 Features

The following are the features used to train the maximum entropy classifier for both the argument identification and argument classification. We used only simple features for these experiments, we are planning to use richer features in the near future.

1. Verb/Predicate.
2. Voice of the verb.
3. Constituent head and Part of Speech tag.
4. Label of the constituent.
5. Relative position of the constituent with respect to the verb.
6. The path of the constituent to the verb phrase.
7. Preposition of the constituent, NULL if it doesn't exist.

3 Maximum entropy model

The maximum entropy approach became the preferred approach of probabilistic model builders for its flexibility and its novel approach to smoothing (Ratnaparakhi, 1999).

Many classification tasks are most naturally handled by representing the instance to be classified as a vector of features. We represent features as binary functions of two arguments, f(a,H), where 'a' is the observation or the class and 'H' is the history. For example, a feature $f_i(a, H)$ is true if 'a' is Ram and 'H' is 'AGENT of a verb'. In a log linear model, the probability function $P(a|H)$ with a set of features $f_1, f_2, ...f_j$ that connects 'a' to the history 'H', takes the following form.

$$P(a|H) = \frac{e^{\sum_i \lambda_i(a,H)*f_i(a,H)}}{Z(H)}$$

Here λ_i's are weights between negative and positive infinity that indicate the relative importance of a feature: the more relevant the feature to the value of the probability, the higher the absolute value of the associated lambda. Z(H), called the partition function, is the normalizing constant (for a fixed H).

4 Description of our system

Our approach labels the semantic roles in two stages, (1) argument identification and (2) argument classification. As input to our system, we use full syntactic information (Collins, 1999), Named-entities, Verb senses and Propbank frames. For our experiments, we use Zhang Le's Maxent Toolkit [1], and the L-BFGS parameter estimation algorithm with Gaussian prior smoothing (Chen and Rosenfield, 1999).

4.1 Argument identification

The first task in this stage is to find the candidate arguments and their boundaries using a parser. We use Collins parser to infer a list of candidate arguments for every predicate. The following are some of the sub-stages in this task.

- Convert the CFG tree given by Collins parser to a dependency tree.
- Eliminate auxilliary verbs etc.
- Mark the head of relative clause as an argument of the verb.

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

- If a verb is modified by another verb, the syntactic arguments of the superior verb are considered as shared arguments between both the verbs.

- If a prepositional phrase attached to a verb contains more than one noun phrase, attach the second noun phrase to the verb.

The second task is to filter out the constituents which are not really the arguments of the predicate. Given our approach towards argument classification, we also need information about whether an argument is mandatory or optional. Hence, in this stage the constituents are marked using three labels, (1) MANDATORY argument, (2) OPTIONAL argument and (3) NULL, using a maximum entropy classifier. For example, a sentence ”John was playing football in the evening”, ”John” is marked MANDATORY, ”football” is marked MANDATORY and ”in the evening” is marked OPTIONAL.

For training, the Collins parser is run on the training data and the syntactic arguments are identified. Among these arguments, the ones which do not exist in the propbank annotation of the training data are marked as null. Among the remaining arguments, the arguments are marked as mandatory or optional according to the propbank frame information. Mandatory roles are those appearing in the propbank frames of the verb and its sense, the rest are marked as optional. A propbank frame contains information as illustrated by the following example:

If Verb = play, sense = 01,
then the roles A0, A1 are MANDATORY.

4.2 Argument classification

Argument classification is done in two steps. In the first step, the propbank sub-categorization frames are used to assign the semantic roles to the mandatory arguments in the order specified by the sub-categorization frames. Sometimes, the number of mandatory arguments of a verb in the sentence may be less than the number of roles which can be assigned by the sub-categorization frame. For example, in the sentence

”MAN1 MAN2 V MAN3 OPT1”, roles could be assigned in the following two possible ways by the sub-categorization frame ”A0 v A1” of verb V1.

- A0[MAN1] MAN2 V1 A1[MAN3] OPT1

- MAN1 A0[MAN2] V A1[MAN3] OPT1

In the second step, the task is to label the unlabelled mandatory arguments and the arguments which are marked as optional. This is done by marking these arguments with the most probable semantic role which is not one of the frame elements of the sub-categorization frame ”A0 v A1”. In the above example, the unlabelled mandatory arguments and the optional arguments cannot be labelled as either A0 or A1. Hence, after this step, the following might be the role-labelling for the sentence ”MAN1 MAN2 V1 MAN3 OPT1”.

- A0[MAN1] AM-TMP[MAN2] V1 A1[MAN3] AM-LOC[OPT1]

- AM-MNC[MAN1] A0[MAN2] V1 A1[MAN3] AM-LOC[OPT1]

The best possible sequence of semantic roles ($\bar{R}$) is decided by the taking the product of probabilities of individual assignments. This also disambiguates the ambiguity in the assignment of mandatory roles. The individual probabilities are computed using the maximum entropy model. For a sequence $\vec{R}$, the product of the probabilities is defined as

$$P(\vec{R}) = \Pi_{R_i \in \vec{R}} P(R_i|Arg_i)$$

The best sequence of semantic roles $\bar{R}$ is defined as

$$\bar{R} = argmax\ P(\vec{R})$$

For training the maximum entropy model, the outcomes are all the possible semantic roles. The list of sub-categorization frames for a verb is obtained from the training data using information about mandatory roles from the propbank. The propbank sub-categorization frames are also appended to this list.

We present our results in the next section.

	Precision	Recall	$F_{\beta=1}$
Development	71.88%	64.76%	68.14
Test WSJ	73.76%	65.52%	69.40
Test Brown	65.25%	55.72%	60.11
Test WSJ+Brown	72.66%	64.21%	68.17

Test WSJ	Precision	Recall	$F_{\beta=1}$
Overall	73.76%	65.52%	69.40
A0	85.17%	73.34%	78.81
A1	74.08%	66.08%	69.86
A2	54.51%	48.47%	51.31
A3	52.54%	35.84%	42.61
A4	71.13%	67.65%	69.35
A5	25.00%	20.00%	22.22
AM-ADV	52.18%	47.23%	49.59
AM-CAU	60.42%	39.73%	47.93
AM-DIR	45.65%	24.71%	32.06
AM-DIS	75.24%	73.12%	74.17
AM-EXT	73.68%	43.75%	54.90
AM-LOC	50.80%	43.53%	46.88
AM-MNR	47.24%	49.71%	48.44
AM-MOD	93.67%	91.29%	92.46
AM-NEG	94.67%	92.61%	93.63
AM-PNC	42.02%	43.48%	42.74
AM-PRD	0.00%	0.00%	0.00
AM-REC	0.00%	0.00%	0.00
AM-TMP	74.13%	66.97%	70.37
R-A0	82.27%	80.80%	81.53
R-A1	73.28%	61.54%	66.90
R-A2	75.00%	37.50%	50.00
R-A3	0.00%	0.00%	0.00
R-A4	0.00%	0.00%	0.00
R-AM-ADV	0.00%	0.00%	0.00
R-AM-CAU	0.00%	0.00%	0.00
R-AM-EXT	0.00%	0.00%	0.00
R-AM-LOC	100.00%	57.14%	72.73
R-AM-MNR	25.00%	16.67%	20.00
R-AM-TMP	70.00%	53.85%	60.87
V	97.28%	97.28%	97.28

Table 1: Overall results (top) and detailed results on the WSJ test (bottom).

5 Results

The results of our approach are presented in table 1.

When we used an approach which uses a simple maximum entropy model, we obtained an F-measure of 67.03%. Hence, we show that the sub-categorization frames help in predicting the semantic roles of the mandatory arguments, thus improving the overall performance.

6 Conclusion

In this paper, we propose an approach for inferring semantic role using sub-categorization frames and maximum entropy model. Using this approach, we obtained an F-measure of 68.14% on the development set of the data provided for the CONLL-2005 shared task.

7 Future work

We have observed that the main limitation of our system was in argument identification. Currently, the recall of the arguments inferred from the output of the parser is 75.52% which makes it the upper bound of recall of our system. In near future, we would focus on increasing the upper bound of recall. In this direction, we would also use the partial syntactic information. The accuracy of the first stage of our approach would increase if we include the mandatory/optional information for training the parser (Yi and Palmer, 1999).

8 Acknowledgements

We would like to thank Prof. Rajeev Sangal, Dr. Sushama Bendre and Dr. Dipti Misra Sharma for guiding us in this project. We would like to thank Szu-ting for giving some valuable advice.

References

S. Chen and R. Rosenfield. 1999. A gaussian prior for smoothing maximum entropy models.

M. Collins. 1999. Head driven statistical models for natural language processing.

Daniel Gildea and Daniel Jurafsky. 2002. Automatic labeling of semantic roles.

Hwang Young Sook Lim, Joon-H and, So-Young Park, and Hae-Chang Rim. 2004. Semantic role labelling using maximum entropy model.

Sameer Pradhan, Kadri Hacioglu, Valerie Krugler, Wayne Ward, James. H. Martin, and Daniel Jurafsky. 2003. *Support Vector Learning for Semantic Argument Classification*.

Adwait Ratnaparakhi. 1999. Learning to parse natural language with maximum entropy models.

Cynthia A. Thompson, Roger Levy, and Christopher D. Manning. 2003. A generative model for semantic role labelling.

Szu-ting Yi and M. Palmer. 1999. The integration of syntactic parsing and semantic role labeling.