A Joint Syntactic and Semantic Dependency Parsing System based on Maximum Entropy Models

Buzhou Tang[1]   **Lu Li**[2]   **Xinxin Li**[1]   **Xuan Wang**[2]   **Xiaolong Wang**[2]
Shenzhen Graduate School
Harbin Institute of Technology
Shenzhen,518055, China
[1]{tangbuzhou,lixxin2}@gmail.com
[2]{lli,wangxuan,wangxl}@insun.hit.edu.cn

Abstract

A joint syntactic and semantic dependency parsing system submitted to the CoNLL-2009 shared task is presented in this paper. The system is composed of three components: a syntactic dependency parser, a predicate classifier and a semantic parser. The first-order MSTParser is used as our syntactic dependency pasrser. Projective and non-projective MSTParsers are compared with each other on seven languages. Predicate classification and semantic parsing are both recognized as classification problem, and the Maximum Entropy Models are used for them in our system. For semantic parsing and predicate classifying, we focus on finding optimized features on multiple languages. The average Macro F1 Score of our system is 73.97 for joint task in closed challenge.

1 Introduction

The task for CoNLL-2009 is an extension of the CoNLL-2008 shared task to multiple languages: English (Surdeanu et al., 2008), Catalan plus Spanish (Mariona Taulé et al., 2008), Chinese (Martha Palmer et al., 2009), Czech (Jan Hajič et al., 2006), German (Aljoscha Burchardt et al., 2006) and Japanese (Daisuke Kawahara et al., 2002). Compared to the CoNLL-2008 shared task, the predicates are given for us in semantic dependencies task. Therefore, we have only need to label the semantic roles of nouns and verbs, and the frames of predicates.

In this paper, a joint syntactic and semantic dependency parsing system submitted to the CoNLL-2009 shared task is presented. The system is composed of three components: a syntactic dependency parser, a predicate classifier and a semantic parser. The first-order MSTParser is used as our syntactic dependency parser. Projective and non-projective MSTParsers are compared with each other on seven languages. The predicate classifier labeling the frames of predicates and the semantic parser labeling the semantic roles of nouns and verbs for each predicate are both recognized as classification problem, and the Maximum Entropy Models (MEs) are used for them in our system. Among three components, we mainly focus on the predicate classifier and the semantic parser.

For semantic parsing and predicate classifying, features of different types are selected to our system. The effect of them on multiple languages will be described in the following sections in detail.

2 System Description

Generally Speaking, a syntactic and semantic dependency parsing system is usually divided into four separate subtasks: syntactic parsing, predicate identification, predicate classification, and semantic role labeling. In the CoNLL-2009 shared task, the predicate identification is not required, since the predicates are given for us. Therefore, the system we present is only composed of three components: a syntactic dependency parser, a predicate classifier and a semantic parser. The syntactic dependencies are processed with the MSTParser 0.4.3b. The predicates identification and semantic role label are processed with MEs-based classifier respectively. Unlike conventional systems, the predicates identifica-

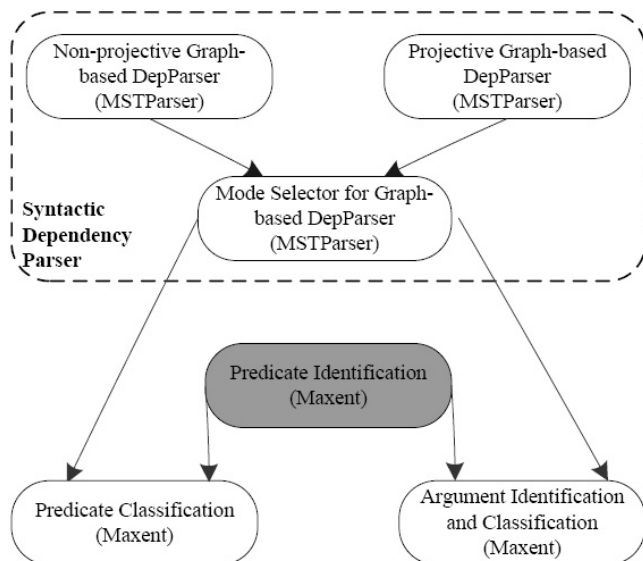tion and the semantic parser are independent with each other. Figure 1 is the architecture of our system.

Figure 1: System Architecture

No	Features	No	Features
1	w_0	20	*Lemma*
2	p_0	21	DEPREL
3	p_{-1}	22	CHD_POS
4	p_1	23	CHD_POS_U
5	$p_{-1}p_0$	24	CHD_REL
6	p_0p_1	25	CHD_REL_U
7	$p_{-2}p_0$	26	SIB_REL
8	p_0p_2	27	SIB_REL_U
9	$p_{-3}p_0$	28	SIB_POS
10	p_0p_3	29	SIB_POS_U
11	$p_{-1}p_0p_1$	30	VERB_V
12	w_0p_0	31	4+11
13	$w_0p_{-1}p_0$	32	Indegree
14	$w_0p_0p_1$	33	Outdegree
15	$w_0p_{-2}p_0$	34	Degree
16	$w_0p_0p_2$	35	ARG_IN
17	$w_0p_{-3}p_0$	36	ARG_OUT
18	$w_0p_0p_3$	37	ARG_Degree
19	$w_0p_{-1}p_0p_1$	38	Span

Table 1: Features for Predicate Classification.

In our system, we firstly select an appropriate mode (projective or non-projective) of Graph-based Parser (MSTParser) for each language, then construct the MEs-based predicates classification and the MEs-based semantic parser with syntactic dependency relationships and predicate classification respectively.

2.1 Syntactic Dependency Parsing

MSTParser (McDonald, 2008) is used as our syntactic dependency parser. It is a state-of-the-art dependency parser that searches for maximum spanning trees (MST) over directed graph. Both of projective and non-projective are supported by MSTParser. Our system employs the first-order framework with projective and non-projective modes on seven given languages.

2.2 Predicate Classification

In this phase, we label the sense of each predicate and the MEs are adopted for classification. Features of different types are extracted for each predicate, and an optimized combination of them is adopted in our final system. Table 1 lists all features. 1-20 are the features used in Li's system (Lu Li et al., 2008),

and 21-31 are a part of the optimized features presented in Che's system (Wanxiang Che et al., 2008)

In Table 1, "w" denotes the word and "p" denotes POS of the words. Features in the form of part1_part2 denote the part2 of the part1, while features in the form of part1+part2 denote the combination of the part1 and part2. "CHD" and "SIB" denote a sequence of the child and the sibling words respectively, "REL" denotes the type of relations, "U" denotes the result after reducing the adjacent duplicate tags to one, "V" denotes whether the part is a voice, "In" and "OUT" denote the in_degree and out_degree, which denotes how many dependency relations coming into this word and going away from this word,and "ARG" denotes the semantic roles of the predicate. The "Span" denotes the maximum length between the predicate and its arguments. The final optimized feature combination is :1-31 and 33-37.

2.3 Semantic Role Labeling

The semantic role labeling usually contains two subtasks: argument identification and argument classification. In our system, we perform them in a single

stage through one classifier, which specifies a particular role label to the argument candidates directly and assigns "NONE" label to the argument candidates with no role. MEs are also adopted for classification. For each word in a sentence, MEs gives each candidate label (including semantic role labels and none label) a probability for the predicate. The features except for the feature (lemma plus sense number of the predicate in (Lu Li et al., 2008)) and the features 32-38 in Table 1 are selected in our system.

3 Experiments and Results

We train the first-order MSTParser [1] with projective and non-projective modes in terms of default parameters respectively. Our maximum entropy classifiers are implemented with the Maximum Entropy Modeling Toolkit [2]. The default classifier parameters are used in our system except for iterations. All models are trained using all training data, and tested on the whole development data and test data, with 64-bit 3.00GHz Intel(R) Pentium(R) D CPU and 4.0G memory.

3.1 Syntactic Dependency Parsing

Table 2 is a performance comparison between projective parser and non-projective parser on the development data of seven languages. In Table 2, "LAS", "ULAS" and "LCS" denote as Labeled attachment score, Unlabeled attachment score and Label accuracy score respectively.

The experiments show that Catalan, Chinese and Spanish have projective property and others have non-projective property.

3.2 Predicate Classification

To get the optimized system, three group features are used for comparison.

- group 1: features 1-20 in Table 1.

- group 2: features 1-31 in Table 1.

- group 3: all features in Table 1.

The performance of predicate classification on the development data of the six languages, which contain this subtask, are given in Table 3. The results

[1] http://sourceforge.net/projects/mstparser.
[2] http://homepages.inf.ed.ac.uk/s0450736/maxent_toolkit.html.

	LAS(%)	ULAS(%)	LCS(%)
Catalan	84.18	88.18	91.76
	83.69	87.74	91.59
Chinese	72.58	77.06	82.07
	62.85	69.47	73.00
Czech	72.79	81.40	80.93
	73.18	81.86	81.30
English	86.89	90.29	91.50
	86.88	90.34	91.58
German	83.43	86.89	90.24
	84.00	87.40	90.61
Japanese	92.23	93.16	98.38
	92.23	93.14	98.45
Spanish	83.88	87.93	91.36
	83.46	87.46	91.37

Table 2: Performance of Syntactic Dependency Parsing with different modes. The above line is the performance of projective mode, while the below one is the performance of non-projective mode for each language.

	group 1	group 2	group 3
Catalan	75.51	80.90	82.23
Chinese	93.79	94.99	94.75
Czech	91.83	91.77	91.86
English	92.12	92.48	93.20
German	74.49	74.14	75.85
Spanish	74.01	76.22	76.53

Table 3: Performance of predicate classification (F1 scores) for different group features on the development data of the six languages.

show that Che's features and the degrees of the predicate and its arguments are useful for all languages, the former improves the labeled F1 measure by 0.3% to 5.4%, and the latter by 0.3% to 1.7%.

3.3 Semantic Role Labeling

In this phase, feature selection and performance lose caused by P-columns are studied. Firstly, we compare the following two group features:

- group 1: The features except for the lemma plus sense number of the predicate in (Lu Li et al., 2008).

	LF1	ULF1	PF1
Catalan	73.25	92.69	38.41
	72.71	91.93	35.22
	83.23	100.00	61.88
Chinese	69.60	82.15	28.35
	71.49	81.71	29.41
	85.44	95.21	58.20
Czech	80.62	92.49	70.04
	79.10	91.44	68.34
	85.42	96.93	77.78
English	73.91	87.26	33.16
	76.10	88.58	36.28
	79.35	91.74	43.32
German	64.85	88.05	27.21
	65.36	88.63	26.70
	72.78	94.54	41.50
Japanese	69.43	82.79	29.27
	69.87	83.31	29.69
	72.80	87.13	34.96
Spanish	73.49	93.15	39.64
	78.18	91.68	33.57
	81.96	99.98	59.20

Table 4: Performance of Semantic Role Labeling (F1 score) with different features.

- group 2: group1+the degrees of the predicate and its arguments presented in the last section.

Secondly, features extracted from golden-columns and P-columns are both used for testing.

The performance of them are given in Table 4, where "LF1", "ULF1" and "PF1" denote as Labeled F1 score, Unlabeled F1 score and Proposition F1 score respectively. The above line is the F1 scores of Semantic Role Labeling with different features. The uppermost line is the result of group1 features, the middle line is the result of group2 features extracted from P-columns, and the downmost one is the result of group2 features extracted from golden-columns for each language.

The results show that the features of degree also improves the labeled F1 measure by 3.4% to 15.8%, the different labeled F1 between golden-columns and P-columns is about 2.9%–13.9%.

	LAS	LF1	M_LF1
Catalan	84.18	72.71	81.46
	75.68	66.95	71.32
Chinese	72.58	71.49	72.20
	63.95	67.06	65.53
Czech	73.18	79.10	76.37
	72.60	79.08	75.85
Czech-ood	69.81	79.80	74.81
English	86.88	76.10	82.89
	86.61	77.17	81.92
English-ood	80.09	67.21	73.69
German	84.00	65.36	83.06
	79.85	61.98	70.93
German-ood	71.86	61.83	66.86
Japanese	92.23	69.87	83.77
	91.26	69.58	80.49
Spanish	83.88	71.18	80.74
	77.21	66.23	71.72

Table 5: Overall performance of our final joint system.

3.4 Overall Performance

In the final system, we select the optimized feature subset discussed in the former sections. The overall performance of the system on the development data , test data and Out-of-domain data are shown in Table 5 (all features are extracted from P-columns). The average Macro F1 Scores of our system are 73.97 on test data and 71.79 on Out-of-domain data.

In Table 5, "LAS", "LF1" and "M_LF1" denote as Labeled accuracy score for Syntactic Dependency Parsing, Labeled F1 score for Semantic Role Labeling, and Overall Macro Labeled F1 score respectively. The topmost line is the result on the development data, the middle one is the result on the test data for each language and the downmost one is the result on the Out-of-domain data if the data exist.

4 Conclusion and Discussion

We present a joint syntactic and semantic dependency parsing system for CoNLL2009 Shared Task, which composed of three components: a syntactic dependency parser, a predicate classifier and a semantic parser. All of them are built with some state-of-the-art methods. For the predicate classifier and the semantic parser, a new kind of features—

degrees, which reflect the activeness of the words in a sentence improves their performance. In order to improve the performance further, we will study new machine learning methods for semantic dependency parsing, especially the joint learning methods, which can avoid the information loss problem of our system.

Acknowledgments

References

Jan Hajič and Massimiliano Ciaramita and Richard Johansson and Daisuke Kawahara and Maria Antònia Martí and Lluís Màrquez and Adam Meyers and Joakim Nivre and Sebastian Padó and Jan Štěpánek and Pavel Straňák and Miahi Surdeanu and Nianwen Xue and Yi Zhang. 2009. The CoNLL-2009 Shared Task: Syntactic and Semantic Dependencies in Multiple Languages. *Proceedings of the 13th Conference on Computational Natural Language Learning (CoNLL-2009), June 4-5*. Boulder, Colorado, USA.

Mariona Taulé and Maria Antònia Martí and Marta Recasens. 2008. AnCora: Multilevel Annotated Corpora for Catalan and Spanish. *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC-2008)*. Marrakesh, Morroco.

Martha Palmer and Nianwen Xue. 2009. Adding semantic roles to the Chinese Treebank. *Natural Language Engineering*, 15(1),pages 143–172.

Jan Hajič and Jarmila Panevová and Eva Hajičová and Petr Sgall and Petr Pajas and Jan Štěpánek and Jiří Havelka and Marie Mikulová and Zdeněk Žabokrtský. 2006. Prague Dependency Treebank 2.0. *CD-ROM*, Cat. No. LDC2006T01, ISBN 1-58563-370-4. Linguistic Data Consortium, Philadelphia, Pennsylvania, USA. URL: http://ldc.upenn.edu.

Surdeanu, Mihai and Johansson, Richard and Meyers, Adam and Màrquez, Lluís and Nivre, Joakim. 2008. The CoNLL-2008 Shared Task on Joint Parsing of Syntactic and Semantic Dependencies. *Proceedings of the 12th Conference on Computational Natural Language Learning(CoNLL-2008)*.

Aljoscha Burchardt and Katrin Erk and Anette Frank and Andrea Kowalski and Sebastian Padó and Manfred Pinkal. 2006. The SALSA corpus: a German corpus resource for lexical semantics. *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC-2006)*, pages 2008–2013. Genoa, Italy.

Daisuke Kawahara and Sadao Kurohashi and Kôiti Hasida. 2002. Construction of a Japanese Relevance-tagged Corpus. *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC-2002)*, pages 2008–2013. Las Palmas, Canary Islands.

McDonald and Ryan. 2006. *Discriminative Learning and Spanning Tree Algorithms for Dependency Parsing*, Ph.D. thesis. University of Pennsylvania.

Lu Li, Shixi Fan, Xuan Wang, XiaolongWang. 2008. Discriminative Learning of Syntactic and Semantic Dependencies. *CoNLL 2008: Proceedings of the 12th Conference on Computational Natural Language Learning*, pages 218–222. Manchester.

Wanxiang Che, Zhenghua Li, Yuxuan Hu, Yongqiang Li, Bing Qin, Ting Liu, Sheng Li. 2008. A Cascaded Syntactic and Semantic Dependency Parsing System. *CoNLL 2008: Proceedings of the 12th Conference on Computational Natural Language Learning*, pages 238–242. Manchester.