Semantic Dependency Feature Selection

Lin Yao[1] Zhixin Hao[2] Tao Qian[2] Xuan Wang[3]

Shenzhen Graduate School, Harbin Institute of Technology,
Xili, Shenzhen 518055, China
[1]yaolin@hit.edu.cn
[2]{haozhix, canyboy}@cs.hitsz.edu.cn
[3]wangxuan@insun.hit.edu.cn

Abstract

This paper presents our solution for CoNLL 2008 shared task that jointly parses syntactic and semantic dependencies. The Maximum Entropy (ME) classifier has been selected to train the data used in this system. Also the Mutual Information (MI) model was utilized into feature selection of dependency labeling. Results show that the MI model allows the system to get better performance and less training hours.

Keywords: Maximum Entropy, Mutual Information, Semantic Role Labeling (SRL), Syntactic Parser

1. Introduction

In the past few years, both semantic role labeling (SRL) and Syntactic Parser have received considerable interest for their big contribution to many kinds of Natural Language Processing (NLP) applications, such as information extraction, question and answering, machine translation, paraphrasing, and etc . The Conference on Computational Natural Language Learning (CoNLL) featured an associated share task every year in the last four years, allowed the participants to train and test their SRL or Syntactic systems on the same date sets and share their experiences. In 2004 and 2005, the shared tasks of CoNLL were focus on SRL. In CoNLL-2006 and CoNLL-2007, the shared tasks were dedicated to the syntactic dependency parsing.

In 2008, The CoNLL consolidates past four years endeavor and unites the tasks into one, proposes a new challenge of the merging of both syntactic dependencies (extracted from the Penn Treebank) and semantic dependencies (extracted both from PropBank) under a unique unified representation.

We propose a solution that selects Maximum Entropy model (Berger et al., 1996) and Mutual Information to grow a joint syntactic and semantic structure. During the last years noticeable efforts have been devoted to semantic and syntactic areas, based on various machine learning techniques, several variants of the basic approaches using different features and different classifiers have been introduced. (Gildea and Palmer, 2002; Gildea and Hockenmaier, 2003; Surdeanu et. al., 2003; Fleischman and Hovy, 2003; Thompson et. al., 2003; Xue and Palmer 2004; Hacioglu, 2004; Carreras et al., 2006; Surdeanu et. al. 2007; Carreras, 2007 ;) The maximum entropy framework pales others by integrating facts without features independence hypothesis. (Liu et al. 2005) and (Li et al. 2008) have applied the maximum entropy classier in semantic task successfully. Different from prior systems, mutual information are choosing to decease the number of features and reduce the train time efficiently without big affection on result.

The rest of the paper is organized as follows. Section2 gives a detailed description of the corpora and classifier model used for our system. The following section (Section 3) takes a closer look at the system architecture and its realization. We present and analyze our results in section 4. Finally the conclusion was given.

2. Fundamental Knowledge

2.1. The Corpora

Large corpora that allow to automatically extracting information about language are beginning to serve researchers in NLP as critical tools. The Penn Treebank, a corpus annotated for part-of-speech (POS) and skeletal syntactic structure, was built at University of Pennsylvania from 1989. The Treebank is consisting of over 4.5 million American English words and with around 1 millions words from Wall Street Journal (WSJ) and 1 millions tokens from Brown Corpus. The type of syntactic information associated with verbs is a useful resource for automatic syntactic dependency parse and predicate argument tagging.

The Proposition Bank (PropBank) is a corpus adding a layer of semantic annotation onto the Penn Treebank. Different with FrameNet, PropBank gives each verb its own roles and gets more used since it's layered over the Treebank.

2.2. Maximum Entropy Model

In Shannon's Information Theory, entropy is defined as a measure of the uncertainty associated with a random variable. Maximum Entropy model (Berger et al., 1996) allows people to make an optimal choice among various states which all satisfy prior constraints by choosing the state with the maximum entropy, in other words, all unknown possibilities have equally opportunity. The maximum entropy framework makes different features be integrated into the same model without concerning the relationships among them and any unknown facts. These advantages make the ME model has been used in many NLP areas.

For the semantic role labeling, the main task will be picking up the right label from the semantic label sets under the particular condition. $P(y/X)$ denotes the conditional probability of getting the output y consistent with the context X. $f_i(X,y)$ describes the feature constrain with weighting parameter λ_i.

$$P(y/X) = \frac{1}{Z(X)} \exp\left[\sum_{i=1}^{k} \lambda_i f_i(X,y)\right] \quad (1)$$

The ME model will choose the value of $P(y/X)$ which makes $H(p)$ the gets the max value.

$$H(p) = -\sum_{X,y} P(y/X) \log p(y/X) \quad (2)$$

3. System Descriptions

3.1. System Architecture

Comparing with traditional Semantic Role Label, our system will produce a joint rich syntactic-semantic output to allow people getting semantic role annotation and syntactic structure at the same time. The main components of our system:
1) Syntactic parsing subtask.
2) Predicate tagging.
3) Features Selection using Mutual Information
4) Semantic dependency labeling.

The first stage is to create a labeled syntactic dependency parse y for input sentence x including words and their part of speeches (POS).Inspired by the parsing model presented in (McDonald, et al., 2005), we equates the problem of dependency parsing to finding maximum

spanning trees in directed graphs. The following stage, Predicate tagging, can also be transfer into a classification question since each predicate mark number can be seen as a particular type. The features used for both the syntactic parsing phase and predicate tagging are from (Li et. al. 2008).

3.2. Semantic Dependency labeling

This part of system is based on the output of syntactic parser and predicate tagging mentioned earlier and to find out the dependents of a given predicate in a sentence and label them with the one from the set of semantic dependency labels. This task can also be recognized as classification issue and work out with ME model. The chosen features are listed in the table 1.

3.3. Mutual Information

In dependency labeling, features have different knowledge in distinguishing class labels. How to reduce feature set without compromising classification accuracy? Mutual Information methods can be utilized as a pruning algorithm to choose an optimal subset from the earlier large candidate features. The mutual information represents the amount of uncertainty remaining about the system output Y that is resolved by observing the system input X. Mutual information between X and Y is given (A. Al-Ani et al., 2003) as follows.

$$I_{YX} = \sum_{x} \sum_{y} p(x, y) \log\left(\frac{p(y/x)}{p(y)}\right) \quad (3)$$

In our system, y is the semantic label from sets and x is the feature from table 1. The large value of mutual information between a feature and its output means the closer of their association. Usually A pruning value be set during the phrase of choosing feature, the features with the higher mutual information value will be kept.

No.	Features	Description
1	Dependent Relationship	Dependent relationship between dependent and predicate
2	Family Relation	Family relation between predicate and dependent in the syntactic dependent tree: Son, Father, Brother or Self.
3	Position of Dependent	Position of dependent according predicate: before, after or self.
4	POS pair	POS pair of predicate and dependent.
5	Dependency Path between Predicate and Dependent	Dependency label list of the Path from predicate to dependent.
6	Predicate Information	Lemma and POS of predicate,
7	Predicate Voice	Active or passive.
8	Dependent Information	Lemma and POS of dependent
9	Path POS	POS list of all words appearing on the path from predicate to dependent.
10	Dependency Path between Predicate and Parent	Syntactic dependency label of edge between predicate and its parent.
11	Predicate's Parent Info.	Lemma and POS of predicate's parent
12	POS List of Predicate's Siblings	
13	Syntactic Dependency Framework of Predicate's Parent	Syntactic dependency label list of the edges between predicate's parent and its siblings
14	Dependent's parent Info.	POS of dependent's parent.
15	POS List of Dependent 's Siblings	
16	Syntactic Dependency Framework of Dependent's Parent	Syntactic dependency label list of the edges between dependent's parent and its siblings.
17	Number of Dependent's Children	
18	Dependent Category	VerbNet class number of Dependent
19	Path length	Path length between predicate and dependent.
20	POS List of Predicate's Children	
21	POS of P-2	POS of the word 2 before Predict in the sentence
22	POS of P-1	POS of the word before predict in the sentence
23	POS of P+2	POS of the word 2 after predict in the sentence
24	POS of P+1	POS of the word after predict

Table 1: Features for semantic dependencies.

After calculation, feature numbers listed in Table 1 have been reordered according the MI value in descending order.

9, 6, 8, 5, 23, 11, 15, 4, 19, 18, 14, 21, 2, 16, 22, 24, 1, 10

4. Results and Experiments

4.1. The Data

The data used for this system is as the-CoNLL-2008 required, the Penn Tree-Bank for syntactic dependency, the PropBank for semantic dependency. The detail information about the data for dependency labeling is listed in Table 2.

	Train	Devel	Test1	Test2
corpus	WSJ	WSJ	WSJ	Brown
words	39832	1335	2400	426

Table 2: The data for dependency labeling

4.2. Performance

Utilizing the MI model, the features are chosen according their contribution to the results. First all features are grouped, then we collect the first 18 features and those with less contribution are removed, and the procedure are continued until the last group with 7 features. All the test results with different feature combinations are included in the Figure 1. F24 means 24 features were selected.
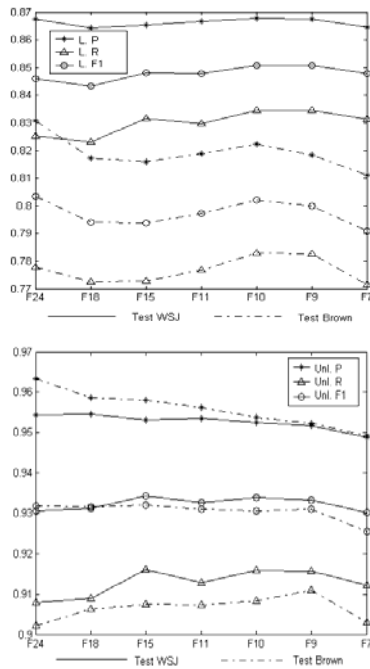

Figure 1: Testing Results and Features Combinations for Dependency labeling. L= Labeled; Unl= Unlabeled; P= precision; R= recall;

Figure 1 presents the best performance gotten with the 10 features combination. This result shows that more features do not mean the better performance. Usage of the MI model decreases the set from 24 features to 10 and the training time is cut sharply, therefore the performance is better.

From the test results, we can also see that our system gets much better performance on WSJ corpus than Brown corpus. The reason is the syntactic parser is constructed based on the WSJ corpus and we might get worse performance using other test corpora.

5. Conclusion

We present a semantic dependency system, which includes syntactic module, predicate tagging module and dependency labeling module. In the dependent role labeling subtask, we select its features with MI model; different feature combinations are evaluated, and get better performance and faster training speed.

6. References

[1] A. Berger, S. D. Pietra, and V. D. Pietra, "A Maximum Entropy Approach to Natural Language Processing," *Computational Linguistics*, pp. 39-71, 1996.

[2] Gildea D, Jurafsky D, "Automatic labeling of semantic roles". *Computational Linguistics*, pp. 245-288, 2002.

[3] A. Al-Ani, M. Deriche and J. Chebil, "A new mutual information based measure for feature selection", *Intelligent Data Analysis*, pp. 43-57, 2003.

[4] T. Liu, W. Che, S. Li, Y. Hu, and H. Liu, "Semantic role labeling system using maximum entropy classifier," *In Proceedings of CoNLL-2005*, pp. 189-192, 2005.

[5] N. Xue and M. Palmer, "Calibrating features for semantic role labeling," *In Proc. of the EMNLP-2004*, pp 88-94, 2004.

[6] D. Gildea and M. Palmer, "The Necessity of Parsing for Predicate Argument Recognition," *In Proceedings of the 40th Meeting of the Association for Computational Linguistics (ACL-02)*, pp 239-246, 2002.

[7] D. Gildea and J. Hockenmaier, "Identifying Semantic Roles Using Combinatory Categorical Grammar," *In Proc. of EMNL'03*, 2003.

[8] M. Surdeanu, S. Harabagiu, J. Williams, and P. Aarseth, "Using Predicate-Argument Structure for Information Extraction," *In Proc. of ACL03*, 2003.

[9] M. Fleischman and E. Hovy, "A Maximum Entropy Approach to FrameNet Tagging," *In Proc. of HLT/NAACL-03*, 2003.

[10] C. A. Thompson, R. Levy, and C. D. Manning, "A Generative Model for Semantic Role Labeling," *In Proc.of ECML-03*, pp. 397-408, 2003.

[11] K. Hacioglu and W. Ward, "Target word detection and semantic role chunking using support vector machines," *In Proc. Of HLT/NAACL-03*, 2003.

[12] X. Carreras, "Experiments with a Higher-Order Projective Dependency Parser," *In Proceedings of EMNLP-CoNLL*, pp. 957-961, 2007.

[13] X. Carreras, M. Surdeanu, and L. M`arquez, "Projective dependency parsing with perception," *In Proceedings of the 10th Conference on Computational Natural Language Learning (CoNLL-2006)*, 2006.

[14] M. Surdeanu, L. M`arquez, X. Carreras, and P. R. Comas, "Combination strategies for semantic role labeling," *Journal of Artificial Intelligence Research*, 2007.