UTD-SRL: A Pipeline Architecture for Extracting Frame Semantic Structures

Cosmin Adrian Bejan and Chris Hathaway
Human Language Technology Research Institute
The University of Texas at Dallas
Richardson, TX 75083-0688, USA
{ady,chris}@hlt.utdallas.edu

Abstract
This paper describes our system for the task of extracting frame semantic structures in SemEval–2007. The system architecture uses two types of learning models in each part of the task: Support Vector Machines (SVM) and Maximum Entropy (ME). Designed as a pipeline of classifiers, the semantic parsing system obtained competitive precision scores on the test data.

1 Introduction
The SemEval–2007 task for extracting frame semantic structures relies on the human annotated data available in the FrameNet (FN) database. The Berkeley FrameNet project (Baker et al., 1998) is an ongoing effort of building a semantic lexicon for English based on the theory of frame semantics. In frame semantics, the meaning of words or word expressions, also called target words (TW), comprises aspects of conceptual structures, or frames, that describe specific situations. The semantic roles, or frame elements (FE), associated with a target word are locally defined in the frame evoked by the target word. Currently, the FN lexicon includes more than 135,000 sentences extracted from the British National Corpus containing more than 6,100 target words that evoke more than 825 semantic frames.

For this task, we extended our previous work at Senseval-3 (Bejan et al., 2004) by (1) experimenting with additional features, (2) adding new classification sub-tasks to accomplish all the requirements, and (3) integrating these sub-tasks into a pipeline architecture.

2 System Description
Given a sentence, the frame semantic structure extraction task consists of recognizing the word expressions that evoke semantic frames, assigning the correct frame to them and, for each target word, detecting and labeling the corresponding frame elements properly. The task also requires the determination of syntactic realizations associated to a frame element, such as grammatical function (GF) and phrase type (PT). The following illustrates a sentence example annotated with frame elements together with their corresponding grammatical functions and phrase types for the target word “tie”:

Frame = Make_Cognitive_Connection
evoked by

To extract semantic structures similar to those illustrated in the example we divide the SemEval–2007 task into four sub-tasks: (1) target word frame disambiguation (TWFD); (2) FE boundary detection (FEBD); (3) GF label classification (GFLC) and (4) FE label classification (FELC). The sub-tasks TWFD and GFLC are natural extensions of the approach described in (Bejan et al., 2004) for the task of semantic role labeling at Senseval-03. We design machine learning classifiers specific for each of the four sub-tasks and arrange them in a pipeline architecture such that a classifier can use information predicted by its previous classifiers. The system architecture is illustrated in Figure 1. In the data processing step, we parse each sentence into a syntactic tree using the Collins parser and extract named entities using an in...
In the remaining part of this section we describe in detail each classification sub-task and the features that have the most salient effect on improving the corresponding classifiers.

### 2.1 Frame Disambiguation

In FrameNet, some target words can evoke multiple semantic frames. In order to extract the semantic structure of an ambiguous target word, the first step is to assign the correct frame to the target word in a given context. This task is similar with the word sense disambiguation task.

We select from the FN lexicon 556 target words that evoke at least two semantic frames and have at least five sentences annotated for each frame, and assemble a multi-class classifier for each ambiguous target word. As described in Figure 3, for this task we extract features used in word sense disambiguation (Florian et al., 2002), lexical features of the target word, and NAMED ENTITY FLAGS associated with the root node in a syntactic parse tree. For the rest of the ambiguous target words that have less than five sentences annotated we randomly choose a frame as being the correct frame in a given context.

### 2.2 Frame Element Identification

The idea of splitting the automatic semantic role labeling task into FE boundary detection and FE label classification was first proposed in (Gildea and Jurafsky, 2002) and then adopted by other works in this task. The problem of detecting the FE boundaries is cast as the problem of deciding whether or not a constituent is a valid candidate for a FE.
We consider a binary classifier over the entire FN data and extract features for each constituent from a syntactic parse tree. Because this experimental setup allows training the binary classifier on a large set of examples, the best feature combination consists of a restrained number of features. Most of these features are from the set proposed by (Gildea and Jurafsky, 2002). Another feature that improved the prediction of FE boundaries in every feature selection experiment is the FRAME feature. Since the frame disambiguation is executed before the FE boundary detection in the pipeline architecture, we can use the FRAME feature at this step. This feature helps the binary classifier distinguish between frame element structures from different semantic frames.

### 2.3 Grammatical Function Classification

Once we identify the candidate boundaries for frame elements, the next step is to assign the grammatical functions to these boundaries. In FrameNet, the grammatical functions represent the manner in which the frame elements satisfy grammatical constraints with respect to the target word.

For this task we train a multi-class classifier over the entire lexicon to predict seven categories of GFs that exist in FN. In addition, we assign the NULL category for those FEs that double as target words.

The features are extracted only for the constituents that are identified as FEs in the previous FE boundary identification sub-task. The best feature set in this phase includes the features proposed by (Gildea and Jurafsky, 2002) and the FRAME feature.

### 2.4 Frame Element Classification

The task of FE classification is to assign FE labels to every constituent identified as FE. In order to predict the frame elements, which are locally defined for each semantic frame, we built 489 multi-class classifiers, where each classifier corresponds to a frame in FrameNet. This partitioning of the FN lexicon has the advantage of increasing the overall classification performance and efficiently learning the frame elements labels. On the other hand, this approach suffers from the lack of annotated data in some frames and hence it requires using a large set of features.

The advantage of designing the classifiers in a pipeline architecture is best illustrated in this subtask. Some of the most effective features for FE classification are extracted using information from previous sub-tasks: FRAME feature is made available by the TWFD sub-task, CONSTITUENTS NUMBER and CONSTITUENTS LIST are made available by the FEBD sub-task, and GF and GF LIST are made available by the GFLC sub-task.
3 Experimental Results

We report experimental results on all four classification sub-tasks. In our experiments we trained two types of classification models for each sub-task: SVM and ME. In order to optimize the performance measure of each sub-task and to find the best configuration of classification models we used 20% of the sub-tasks training data as validation data. Table 1 lists the best configuration of classification models as well as the best sub-task results when running the experiments on the validation data. For frame disambiguation, we obtained 76.71% accuracy compared to a baseline of 60.72% accuracy that always predicts the most annotated frame for each of the 556 target words. The results for GFLC and FELC sub-tasks listed in Table 1 were achieved by using gold FE boundaries.

Table 1: Task results on the validation set.

| Task                        | Best Model | Accuracy |
|-----------------------------|------------|----------|
| Frame Disambiguation        | SVM        | 76.71    |
| GF Label Classification     | ME         | 96.00    |
| FE Label Classification     | ME         | 86.93    |
| Frame Detection             | SVM        | 73.65    |

Table 2: System results on the test set.

| Options | Semantic Dependency Evaluation | Frame Detection Evaluation |
|---------|--------------------------------|---------------------------|
|         | Precision | Recall | F1-measure | Precision | Recall | F1-measure |
| ELY     | 51.10     | 27.74  | 35.85      | 69.16     | 42.73  | 52.71      |
| ELY     | 55.56     | 30.19  | 35.94      | 77.02     | 48.05  | 55.92      |
| PDI     | 57.24     | 29.05  | 35.11      | 71.69     | 44.83  | 54.74      |
| DLY     | 54.75     | 29.43  | 35.26      | 80.25     | 49.79  | 51.35      |
| ELY     | 57.85     | 27.09  | 35.94      | 69.16     | 42.73  | 52.71      |
| PLY     | 56.59     | 30.14  | 35.95      | 77.02     | 48.05  | 55.92      |
| PLY     | 51.36     | 26.90  | 35.29      | 71.69     | 44.83  | 54.74      |
| DLY     | 56.15     | 29.45  | 35.57      | 80.25     | 49.79  | 51.35      |

Although the system achieved good precision scores on the test data, the recall values caused the system to obtain unsatisfactory F1-measure values. We expect that the recall will increase by considering various heuristics for a better mapping of the frame elements to constituents in parse trees.

4 Conclusions

We described a system that participated in SemEval–2007 for the task of extracting frame semantic structures. We showed that a pipeline architecture of the SVM and ME classifiers as well as an adequate selection of the classification models can improve the performance measures of each sub-task.

References

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In Proceedings of the COLING-ACL, Montreal, Canada.

Cosmin Adrian Bejan, Alessandro Moschitti, Paul Morărescu, Gabriel Nicolae, and Sarda Harabagiu. 2004. Semantic Parsing Based on FrameNet. In Senseval-3: Workshop on the Evaluation of Systems for the Semantic Analysis of Text.

Radu Florian, Silviu Cucerzan, Charles Schafer, and David Yarowsky. 2002. Combining classifiers for word sense disambiguation. Natural Language Engineering.

Daniel Gildea and Daniel Jurafsky. 2002. Automatic Labeling of Semantic Roles. Computational Linguistic.

Sameer Pradhan, Kadri Hacioglu, Valeri Krugler, Wayne Ward, James H. Martin, and Daniel Jurafsky. 2005. Support vector learning for semantic argument classification. Journal of Machine Learning Research.

Mihai Surdeanu, Sarda M. Harabagiu, John Williams, and Paul Aarseth. 2003. Using predicate-argument structures for information extraction. In Proceedings of ACL.

Nianwen Xue and Marta Palmer. 2004. Calibrating features for semantic role labeling. In Proceedings of EMNLP.