Shallow Semantics with Shallow Syntax

Nimar S. Arora

May 10, 2007

Abstract

Assigning semantic roles to the constituents of a natural language sentence is an important first step in translating natural language into a logical form for further processing. I present a statistical classifier which can perform this task using minimal syntactic cues. I use the syntactic and the semantic head of each constituent as the only features and present simple rules for extracting these heads. I also show how a hierarchical tree of semantic roles can be exploited to better estimate the probabilities in the model.

On the FrameNet 1.3 corpus, my system can label pre-segmented constituents with an accuracy of 79.6%.

1 Introduction

Translating natural language sentences into a logical form is known as semantic parsing. This is a critical first step for information extraction and question answering systems. Shallow semantics is the early step in semantic parsing where lexical units are labeled as predicates or arguments to predicates. In the context of the FrameNet [1] corpus, predicates are grouped into frames and the arguments into frame specific roles (also called frame elements). As an example, consider the following sentence from the *Duplication* frame:

> The teacher was demonstrating on the blackboard , [$_{\text{Creator}}$ the children] [$_{\text{Target}}$ **copying**] [$_{\text{Goal}}$ onto slates] . (Example 1)

The task that I took on was labeling the pre-segmented constituents with appropriate roles, given the frame and the target word.

The FrameNet 1.3 corpus contains roughly 140K sentences grouped into nearly 700 frames. I took every tenth sentence in the corpus as a test sentence and used the remaining for training. For the labeling task I ignore the null instantiated roles.

2 Baseline

As a simple baseline, I extracted the most frequent role occurring in each frame and labeled all the constituents with this role. This strategy performed at 43% element accuracy and 20% exact matches. Indicating that one-fifth of the sentences contain exactly one frame element.

Another simple baseline was to extract the most frequent role on the left and the right of the target word in each frame. This yielded 62% accuracy with 43% exact matches. However, since this corpus is restricted to the active voice (for size considerations), one can't generalize from any results using frame element order. Accordingly, I didn't use the element order in any subsequent experiments.

3 Syntactic Head

One of the features used in [4] for this task was the phrase type of each constituent. To extract the phrase type, they parse the sentence and use the highest node in the parse tree which contains the first word of the constituent and is contained within the constituent.

In order to simplify this feature, I only used the first word of the sentence. I call this feature the syntactic head since it capture information about the syntax of the constituent. For preposition phrases, for example, this is always a preposition and it captures information about the relation of the constituent with the target word. The preposition "onto" in the earlier example is a strong indicator that the role being played by the element is a *goal*. Thus this feature is much more powerful than just the phrase type. The performance of using this feature in the following model:

$$argmax_{role} P(syn.\ head|role) P(role|frame)$$

yields an element accuracy of 75.8% with 60% exact matches. This feature by itself is better than any combination of features used in [4].

4 Semantic Head

Looking at Example 1 we see that the syntactic head doesn't tell us too much about the element, *the children*. The syntact head here doesn't care if the element is *the drawing*, *the elephant* etc. All of which couldn't be the *creator* in the duplication frame. The *drawing* could be an *original* though. To bring out this difference, I added another feature which I'm calling the semantic head of the consituent. This word captures to a first order of approximation the entity being described by the consitutent. Consider another example from the *Duplication* frame.

> Duplication:: And [$_{\text{original}}$ the layout of the walled garden ; long narrow beds , with plants grouped according to their botanic family] , has been [$_{\text{target}}$ **copied**] [$_{\text{place}}$ worldwide] . (Example 4)

Here the long constituent is describing a layout and all the other words are of no further relevance as such.

To extract the semantic head I needed slightly more complicated rules than for the syntactic head. For starters, I needed the parts-of-speech of each word. Hence, I first built a simple HMM based part-of-speech tagger with the tags as the hidden states and the words as the observations. [1] I was able to tag all the test words with 94% accuracy.

Next, I scanned each constituent to find the first contiguous group of noun-like words (proper nouns, common nouns, pronouns, possessives and gerunds) and picked up the last of these to be the semantic head. If there was only one word I always picked up that one as the semantic head. Also the syntactic head was never allowed to be a noun like word or the only word in the constituent. This way I ensured that the two heads never overlapped. I labeled the frame elements using the following model:

$$argmax_{role} P(syn.\ head|role) P(sem.\ head|role) P(role|frame)$$

[1] The FrameNet corpus contains two different styles of part-of-speech tags, and hence I actually built two part-of-speech taggers

This had an accuracy of 76.9% with 62% exact matches.

5 Hierarchical Counting

For many semantic head words there was not enough data to estimate the probability of a role emitting that word. To handle this problem prior work typically has used an external data source to cluster the words. However, I decided to use the hierarchical relationship between the roles in different frames to estimate the emission probabilities.

For example the *Manufacturer* role in the *Manufacturing* frame is a subtype of the *Creator* role in the *Intentionally_Create* frame. Thus any semantic head word which appears as a *Manufacturer* is also counted as a *Creator*. One consequence of using the hierarchical information was that I could no longer condition the emission probabilities on the role alone. Instead, I had to condition on both the role and the frame. Using the following formula: (See Figure 1.)

$$argmax_{role} P(syn.\ head \mid frame,\ role)\ P(sem.\ head \mid frame,\ role)\ P(role \mid frame)$$

This had an accuracy of 79.6% with 65.4% exact matches.

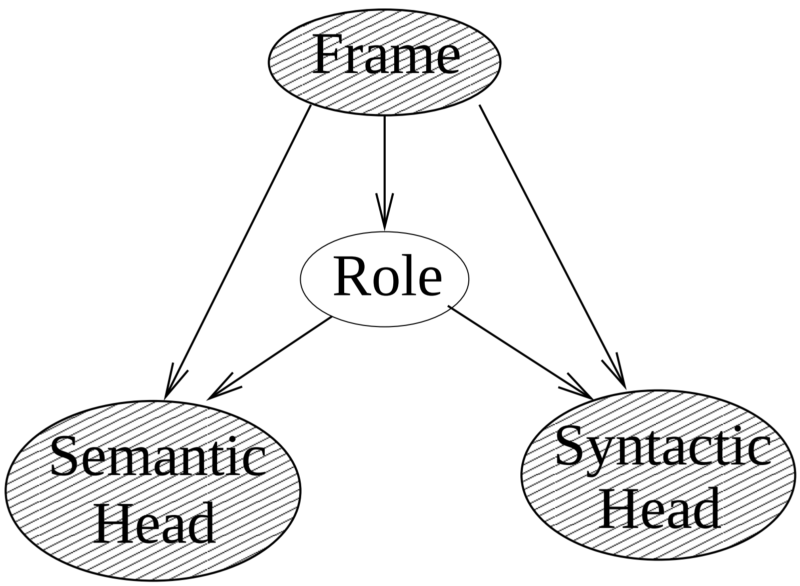

Figure 1: Graphical model for role classification

6 Related Work

The first semantic role classifier on FrameNet was described in [4]. In this work a number of different features like position of constituent w.r.t. target, voice, head word, parse tree path, governing category and phrase type were used. As I have mentioned that the FrameNet corpus doesn't contain examples of the passive voice, hence I have explicitly avoided using position or voice as a feature. The head word used by them is according to the rules in [3], this doesn't attempt to distinguish much between the syntactic and the semantic head. Also, The phrase type is usually captured in the syntactic head.

Finally, in my model I have entirely dropped all deeper syntactic features like parse tree path and governing category. But based on the very slight difference in performance (80.4%) it doesn't appear that these syntactic features help much. However, using external resources for clustering the words did give them a boost of upto 82%. The main difference in my work is the very simple classification model, unlike the complicated interpolation between the probability estimates from different features used by them. Also, I used the hierarchical relation between the frames to cluster words rather than using an external resource. Finally, in their work they condition on the target word, while I totally ignore the target word and focus entirely on the frame.

The other related work is a generative model in [2] which focuses primarily on the harder task of simultaneously segmenting and labeling. The key difference is that they use a single head word as returned by a parser rather than the two head words that I use.

7 Future Work

There were a number of obvious improvements that I couldn't make because of time constraints. First of all, my graphical model classifies each role independently of the others. In reality, most roles don't occur more than once in a frame, hence this independence assumption isn't quite accurate.

Secondly, I wasn't able to interpolate the emission probabilities with the parent frame, role's emission probability (hierarchical smoothing). It appears that doing this hurts performance because the common usage of ambiguous words dominate the rarer usages. I'm still investigating this phenomenon.

Finally, I need to work more on my rules for detecting the semantic head. Currently, I count gerunds as part of the head despite the fact that dropping them seems to marginally improve performance (79.7%). Intuitively, the constituent, *in the making of the movie*, is about *making* something rather than about a *movie*.

8 Conclusion

This work has shown that shallow semantic roles can be extracted fairly accurately with very simple syntactic methods. While there is certainly room for further improvement, it doesn't appear that deeper syntax would help. Most of the remaining errors are to do with ambiguous or rare words which could be better classified if more data (not necessarily frame annotated) was used.

References

[1] Baker, C. F., Fillmore, C. J., and Lowe, J. B. The Berkeley FrameNet Project. In *Proceedings of the Thirty-Sixth Annual Meeting of the Association for Computational Linguistics and Seventeenth International Conference on Computational Linguistics* (San Francisco, California, 1998), C. Boitet and P. Whitelock, Eds., Morgan Kaufmann Publishers, pp. 86–90.

[2] C. Thompson, R. L., and Manning., C. A generative model for semantic role labeling. In *Proceedings of ECML* (2003).

[3] Collins, M. J. *Head-Driven Statistical Models for Natural Language Parsing*. PhD thesis, University of Pennsylvania, Philadelphia, 1999.

[4] Gildea, D., and Jurafsky, D. Automatic Labeling of Semantic Roles. *Computational Linguistics 28*, 3 (2002), 245–288.