Learning to Automatically Discover Meronyms

Austin Shoemaker and Varun Ganapathi

December 17, 2005

Abstract

We present a system for automatically discovering meronyms (noun pairs in a part-whole relationship) from text corpora. More precisely, our system begins by parsing and extracting dependency paths similar to (but not the same as) those used by (Snow et al., 2004). For each noun pair we calculate an empirical distribution over dependency relations, which are then used as features of a Support Vector Machine classifier. Noun pairs are labeled as meronyms if there exists a path traversing only meronym and hypernym links between the nouns. Since the method of labeling training examples treats sentences as bags of words, our training examples are extremely noisy. However, we are able to find a classifier that performs better than similar previous work.

1 Introduction

Many natural language processing applications depend on ontologies such as WordNet in order to obtain prior knowledge about the semantic relationships between words. Unfortunately, the domain of WordNet is limited in scope, and is time-consuming and expensive to maintain and extend. Furthermore, WordNet has no concept of probability. For a given word, WordNet stores a list of its relations to other words, but does not store the probability of the occurrence of that relationship in normal usage. Recently, substantial interest has been directed toward the idea of automatic detection of semantic relations between words.

Automatic extraction of semantic relations between nouns from text corpora is important to many Natural Language Processing tasks. Search would benefit from the ability to perform shallow semantic queries. For example one would like to be able to search for all terms that bear semantic relation to some other terms. One important semantic relation to extract is the *part-whole* relation, otherwise known as *meronymy*. One question that one would like to be able to ask is "Does X contain Y?" or "What does X contain?" (The latter of course being a harder question.)

The meronymy relation is surprisingly ambiguous as the authors discovered when trying to manually label sentences. Unlike other relations, meronymy relations often apply only to a specific instantiation of an entity rather than the general case. For instance, "The dog had brown fur and floppy ears." In this case, it is only the dog in this specific sentence that *has* brown fur and floppy ears. Should one label brown fur as a meronym of dog? Or merely label "fur" and "ear" as a meronym of dog, since most dogs have fur and ears. It's clear we're making decisions based on background knowledge rather than knowledge conveyed by the sentence.

Our task is to detect when pairs of words could possibly be meronyms under normal circumstances. In other words, even though in one particular setting a man may have a tail, we would hope that our system would not strongly imply a meronymic relationship between man and tail. In general, a car has windows(even though there are jeeps), so we would want the algorithm to identify window as a meronym of car.

1.1 Related Work

While part-of-whole relations have been studied in great depth in the liberal-arts, we were unable to find many papers investigating automatic methods for discovery of meronyms. Hearst (Hearst, 1998) investigated the use of high-precision lexico-syntactic patterns indicative of hypernymy relations. However, this approach was not successful when applied to meronymy, because the common patterns were also indicative of other semantic constraints.

Berland and Charniak identified five lexical patterns that tend to indicate part-whole relations by selecting the patterns that maximized a likelihood function $p(w|p)$ and a variation of same, where w is the outcome of the random variable generating wholes, and p is that for parts. Meronyms were selected by a "surprise" metric, which is high when $p(w|p) \gg p(w)$, or a significant- difference test. The proposed system was about 55% accurate on the top 50 proposed parts.

Girju et al. worked on improving the results of Hearst's approach. Hearst's method of using lexico-syntactic patterns filters the input corpus resulting in noun pairs that are more likely to be in meronymic relations. In that paper, they learned semantic constraints using WordNet class labels as features in order to remove false-positives from the result of applying lexico-syntactic patterns. They achieved 83% accuracy, but involved the use of much manual work; The corpus was first filtered by hand-picked lexico-syntactic patterns, word-sense disambiguated by hand, and finally class-labels were extracted from WordNet. As such, we do not see them as a fair comparison to the work we describe here.

Snow et al. (2005) propose an automatic method using dependency-paths for automatically identifying noun pairs in a hypernymy relationship. In spirit our approach to the problem is similar, although we modify the feature set and use a different classifier.

2 Methodology

Our approach is as follows:

1. Training:
   (a) Collect noun pairs from corpora
   (b) For each noun pair, collect sentences where the nouns occur together.
   (c) Parse the sentences and extract dependency paths relating pairs.
   (d) Label each noun pair with its features using WordNet
   (e) Train a classifier based on this data
2. Test:
   (a) For any pair of nouns, extract features and apply classifier.

2.1 Generation of Features

We mostly follow Snow et al. in our feature extraction. Based on the assumption that local syntactic structure can help predict certain semantic relationships, such as meronymy, we selected features that encapsulate the syntactic relationship between a pair of nouns as they occur in a given sentence. These patterns are called *dependency paths*. We apply a dependency parser, which produces a directed acyclic graph of syntactic relations between words. The dependency path, then, is the shortest path separating the noun pair in this graph. A single semantic relation is expressed as $relation(word1[pos], word2[pos])$, where $word\mathbf{n}[pos]$ represents the word in a specific syntactic class, and relation marks the manner in which one word governs the other.

We make one modification to the feature extraction algorithm as follows. Even with moderately large corpora, sparsity can still be a problem, making it more difficult to classify new examples. Our system compensates for this by introducing *anonymized* dependency path features, which describe the grammatical structure of the syntactic relationship while leaving the identity of the specific words unspecified. This improved results significantly.

2.2 Generation of Labels

We labeled each of the noun pairs extracted from our corpora automatically using WordNet. Of course, we lost examples if WordNet did not contain the pairs of nouns, which happened quite often. Given a noun X and Y, we label Y a meronym of X if in WordNet it is possible to reach Y from X using only hypernyms and meronymy relations. Consider the following example, where $x < y$ means that x is a meronym of y, and $x > y$ means that x is a hypernym of y:

```
management~{management}
> {directorate,board_of_directors}
< {board}
```

This was derived from "*Peter Cawdron, group strategy development director, and Bill Shardlow, group personnel director, will become part of the board 's management committee.*"

Beginning from the word in the sentence "management", the system looks up the sense (or senses) in WordNet coresponding to that word and then search for senses matching "board". In this case we see that management contains a board_of_directors which is a type of board. Hence, the management contains a board. This doesn't always work out so well:

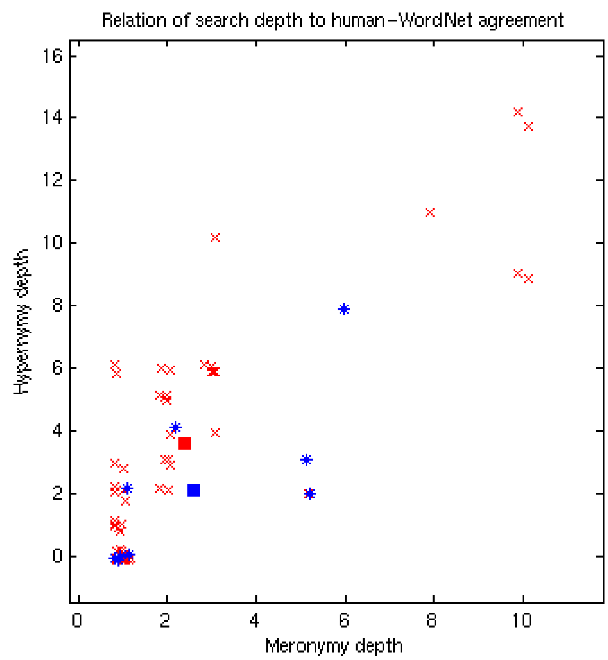
```
level ~ {floor,level,storey}
< {structure,construction}
> {foundation, base}
```

This noun pair was derived from the following sentence: "This has increased the risk of the government being forced to increase base rates to 16% from their current 15% level to defend the pound, economists and foreign exchange market analysts say."

As one can see, word sense ambiguity results in spurious labelings of sentences. Such spurious sentences create dependency paths between nouns that are not indicative of meronymy. Despite such examples, it's the case that while this particular sentence may not be informative of the relation bewteen level and base, other sentences may contain more valuable relations, and our classifier may be able to tease out the relationship. We provide this as an example of the noisyiness of the data on which we are operating.

In order to examine the value of the WordNet labels, we hand-labeled 3998 noun pairs as meronyms. We consider a pair of nouns to be meronyms if the nouns actually were semantically related in the sentence. Of these, a small number actually were in a meronymic relationship. Figure 2.2 displays the confusion matrix of WordNet's labels vs the human labels. It's clear that WordNet is adding significant noise to the training data.

Figure 1: WordNet Compared To Human Labeling



2.3 Training

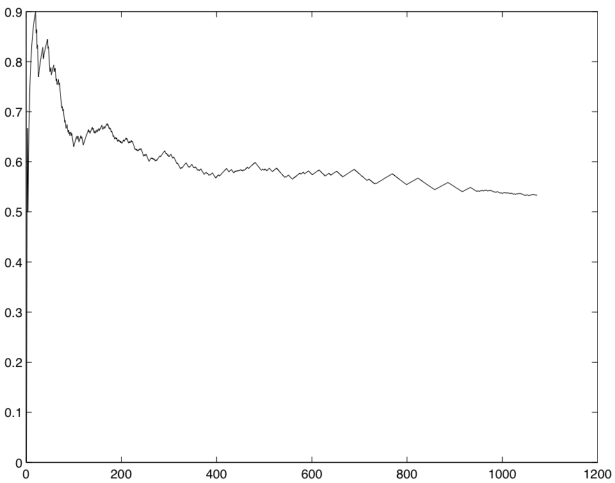
We used SVMLight to train several classifiers using different kernel and gamma values. Since the frequency of meronyms in data is extremely low, on the order of 1%, we created balanced training and test sets by subsampling the negative examples.

We began with a 500k-sentence corpus from the 1996 New York Times. After processing, this resulted in a total data set size with 160,200 pairs. Of these, about 2000 pairs were in a meronymy relationship. We created a balanced training set of 1500 positive and 1500 negative example, a balanced test set of 500 positive and 500 negative examples.

	Precision	Recall
Linear	62.67%	23.73%
Gaussian (RBF), $\gamma = 0.002$	61.26%	27.05%

Figure 2: Comparison of SVM kernels

Figure 3: Precision/Recall for RBF kernel, $\gamma = 0.002$



3 Results

In the table below you can see the precision/recall as reported by SVMLight for various parameter settings. We experimented with various kernels, including linear, polynomial, and found that the RBF kernel performed best. We selected the γ parameter for the RBF kernel to be .002 using `LOOCV`. Figure 3 displays the precision recall curve. As you can see, the classifier can achieve quite high accuracy rates when the recall is limited. This is reasonable considering that WordNet is an extremely noisy sample of the truth. Therefore, many of the positive examples are actually random, in a sense, because none of the dependency patterns are actually indicative of meronymy.

We plan to test the classifier against human labeled data to see how it fares against WordNet.

4 Future Work

This research has illustrated the strengths of using syntactic structure to predict semantic structure, while underlining the reality that complex relationships such as meronymy can often only be inferred with a sufficient understanding of the local semantic context.

While our classifier was trained using WordNet, we recognized that with our human-labeled data as ground truth, WordNet achieves only 15% precision and 5.4% recall. WordNet, aside from error resulting from annotation mistakes, by definition cannot represent semantic relations in terms of context, and thus limits the ability of our classifier to grasp human intuition. Moving forward, it is clear that the next step is to collect more human-annotated data and train on much larger text corpora.

The authors speculate that incorporating a probabilistic model for co-reference resolution would likely improve classification accuracy substantially, because it would enable syntactic relations to be traced across sentences. Part of the challenge lies in the fact that meronymy relations are often assumed to be background knowledge, and the words frequently occur together with no explicit mention of the relationship. Analyzing context across sentences could potentially capture these patterns of discourse.

Our analysis using WordNet shows that the ratio of the number of holonyms and meronyms stemming from a word sense was strongly indicative of its own meronymy; e.g. if an entity contains many things, then in general it is more likely to be a holonym. Conversely, items that are part of many entities are more likely to be a meronym in any given situation. This could perhaps be estimated by counting the number of occurrences of "of X" or "X's" without considering the other noun. The next step is to add such properties to the feature lists of the noun pairs themselves.

5 Conclusion

We have investigated the challenges of automatically extracting holonym-meronymy relationships from text corpora, and proposed an SVM classifier based on an extension of syntactic dependency paths. The best classifier had 61.26% precision at 27.05% recall. Our efforts have illustrated the challenge and promise of identifying meronymy relationships using global syntactic structure. The authors believe that significant improvements in accuracy can be made with a larger quantity of human-labeled data, and by leveraging the intrinsically context-dependent nature of meronymy in more sophisticated ways. This problem is emblematic of many of the challenges faced in NLP research today, and its solution will enable an order of magnitude improvement in information access and discovery technologies.

References

Berland and E. Charniak. 1999. Finding Parts in Very Large Corpora. *In Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, pages 57Ð64, College Park, MD.

R. Girju, A. Badulescu, and D. Moldovan. 2003. Learning Semantic Constraints for the Automatic Discovery of Part-Whole Relations. *In Proceedings of the Human Language Technology Conference / North American Chapter of the Association for Computational Linguistics Conference*, Edmonton, Canada.

Marti A. Hearst. 1998. Automated discovery of wordnet relations. In Christiane Fellbaum, editor, WordNet: An Electronic Lexical Database. MIT Press, Cambridge, MA.

T. Joachims, Making large-Scale SVM Learning Practical. *Advances in Kernel Methods - Support Vector Learning*, B. Schlkopf and C. Burges and A. Smola (ed.), MIT-Press, 1999.

Rion Snow, Daniel Jurafsky, and Andrew Y. Ng, "Learning syntactic patterns for automatic hypernym discovery". NIPS 2004.