Semantic Tagging Using WordNet Examples

Sherwood Haynes
Illinois Institute of Technology
Department of Computer Science
Chicago, Illinois, 60616 USA
skhii@mindspring.com

Abstract

This paper describes IIT1, IIT2, and IIT3, three versions of a semantic tagging system basing its sense discriminations on WordNet examples. The system uses WordNet relations aggressively, both in identifying examples of words with similar lexical constraints and matching those examples to the context.

1 Introduction

The ability of natural language understanding systems to determine the meaning of words in context has long been suggested as a necessary precursor to a deep understanding of the context (Ide and Véronis, 1998; Wilks, 1988). Competitions such as SENSEVAL (Kilgarriff and Palmer, 2000) and SENSEVAL-2 (SENSEVAL-2, 2001) model the determination of word meaning as a choice of one or more items from a fixed sense inventory, comparing a gold standard based on human judgment to the performance of computational word sense disambiguation systems.

Statistically based systems that train on tagged data have regularly performed best on these tasks (Kilgarriff and Rosenzweig, 2000). The difficulty with these supervised systems is their insatiable need for reliable annotated data, frequently called the "data acquisition bottleneck."

The systems described here avoid the data acquisition bottleneck by using only a sense repository, or more specifically the examples and relationships contained in the sense repository.

WordNet version 1.7 (Miller 1990; Fellbaum 1998; WordNet, 2001) was chosen as the sense repository for the English Lexical Sample task (where systems disambiguate a single word or collocation in context) and the English All Word task (where systems disambiguate all content words) of the SENSEVAL-2 competition. WordNet defines a word sense (or synset) as a collection of words that can express the sense, a definition of the sense (called a gloss), zero or more examples of the use of the word sense, and a set of tuples that define relations between synsets or synset words.

2 General Approach

This paper describes three systems that were entered in SENSEVAL-2 competition, IIT1, IIT2, and IIT3. IIT1 and IIT2 were entered in both the English All Word task and the English Lexical Sample task. IIT3 was entered in the English All Word task only. All three systems use the same unsupervised approach to determine the sense of a target word:

1. for each syntactically plausible sense, find the set of WordNet examples that appear in that synset or a related synset.
2. for each example, compare the example to the context, scoring the quality of the match.
3. choose the sense whose synset is responsible for the inclusion of the highest scoring example.

Hereafter, **target words** identify the words to be disambiguated (so identified by the SENSEVAL-2 task). The **context** identifies the text surrounding and including a target word.

2.1 Collecting Examples of a Sense

The systems first collect a set of example sentences and phrases from WordNet for each synset matching a target word (or its canonical or collocational form). The set includes examples from the synset itself as well as those of related synsets. Table 1 lists the relations available in WordNet 1.7. The application of direct relations includes only the examples of the related synset (or synsets of related words). The transitive closure of relations additionally

WordNet Relation	Relation Type	Relation Operands	Application of Relation
Antonym		Word	Direct
Hypernym	Parent	Synset	Transitive Closure
Hyponym	Child	Synset	Direct
Entailment		Synset	Transitive Closure
Similarity	Set	Word	Transitive Closure
Member	Child	Synset	Direct
Stuff	Child	Synset	Direct
Part	Child	Synset	Direct
Has Member	Parent	Synset	Transitive Closure
Has Stuff	Parent	Synset	Transitive Closure
Has Part	Parent	Synset	Transitive Closure
Holonym	Parent	Synset	Transitive Closure
Meronym	Child	Synset	Direct
PPL		Word	Transitive Closure
See Also		Word	Direct
Pertains		Word	Transitive Closure
Attribute		Synset	Transitive Closure
Verb Group	Set	Synset	Not Used

Table 1
Use of WordNet Relations

includes examples from repeated application of the relation. That is, for the hypernym relation, examples from all ancestor synsets are included.

Table 2 lists the examples identified for the synset for *faithful - steadfast in affection or allegiance*. WordNet 1.7 displays the synset as:

faithful (vs. unfaithful)
=> firm, loyal, truehearted, fast(postnominal)
=> true
Also See-> constant#3; true#1; trustworthy#1, trusty#1

This *faithful* synset contributes 3 examples, the *see also* relation contributes examples for *constant*, *true,* and *trustworthy*, the *similarity* relation contributes the examples from the *firm* synset and the *antonym* relation contributes the *unfaithful* example.

2.2 Comparing Examples to the Context

Each example is compared to the context. Consider the first example in Table 2, *a man constant in adherence to his ideals*. Since each example contains a word being defined, the systems consider that this word matches the target word, so *constant* is assumed to match *faithful*. Call this word the **example anchor.**

The remaining words of the example are compared to the words surrounding the target word. The comparison begins with the word to

Synset Words	Example
constant	a man constant in adherence to his ideals
	a constant lover
	constant as the northern star
faithful	years of faithful service
	faithful employees
	we do not doubt that England has a faithful patriot in the Lord Chancellor
firm, loyal, truehearted, fast	"the true-hearted soldier...of Tippecanoe" - Campaign song for William Henry Harrison;
	a firm ally
	loyal supporters
	fast friends
true	true believers bonded together against all who disagreed with them
	the story is true
	"it is undesirable to believe a proposition when there is no ground whatever for supposing it true" - B. Russell;
	the true meaning of the statement
trustworthy	a trustworthy report
	an experienced and trustworthy traveling companion
unfaithful	an unfaithful lover

Table 2
Examples Relate to Synset *faithful – steadfast in affection or allegiance*

the left of the example anchor followed by the word immediately to the right of the anchor, the second word to the left of the anchor, the second word to the right of the anchor, and so on. So the order of comparison of the example words is *man, in, a, adherence, to, his, ideals.*

Each example word is compared to the unmatched context words in a similar sequence. So, for example, the example word *man* would first be compared to the word immediately to the left of the context word followed by the word to its left, and so on, until a match is found.

Word matches also use the WordNet relations as described in Table 1. Under parent relations, two words match if they have a common ancestor. Other transitive closure relations generate a match if either word appears in the other's transitive closure. The words also match if there is a direct relation between the words.

2.3 Scoring the Match

Once the words of an example have been matched to the context, the result is scored. The score for all systems is computed as:

Characteristic	Description
Distance	Magnitude of the difference in the word position of the matching example and context words relative to the position of the example and context anchors
Direction Change	1 if the example words adjacent to a word match context words both occurring before or after its matching context word, 0 otherwise.
Lexical Proximity	0 for exact matches; 1 for matches based on non-parent relation matches; sum of the distances to the closest common ancestor for matches under parent relations
Maximum and Minimum Lexical Generalization	0,0 for exact matches; 1,0 for matches based on non-parent relation matches; maximum and minimum distance to the closest common ancestor for matches under parent relations
Alignment Skew	Ratio of the matching phrase length to the example length.
Match Failure	1 for example words with no matching context word, 0 otherwise

Table 3
Scoring Penalty Characteristics

$$score = \frac{1}{1 + \sum_i \sum_j s(w_i, c_j, d_i)}$$

The scoring function s generates a non-negative value for each example word w_i, penalty characteristic c_j (Table 3), distance d_i of w_i from the example anchor. In IIT1, d_i is not considered, so a penalty calculation is independent of the word position in the example. In IIT2, d_i reduces penalties for w_i further away from the example anchor.

If an example anchor alignment with the context word is the only open-class match for an example, the example receives a zero score.

Haynes (2001) describes these calculations in more detail.

A sense of a target word receives the maximum score of the examples related to that sense. The systems suggest the sense(s) with the highest score, with multiple senses in the response in the event of ties. (If a tie occurs because the same example was included for two senses, the other senses are eliminated, the common example is dropped from the example set of the remaining senses, and the sense scores are recomputed.) If no sense receives a score greater than zero, the first sense is chosen.

IIT1 and IIT2 match a context word independent of other sense assignment decisions. The IIT3 system (English All Word task only) uses the IIT1 scoring algorithm for target words, but limits the senses of preceding context words to the sense tags already assigned.

System	Course Grained Precision/Recall	Fine Grained Precision/Recall
IIT1 Lexical Sample	34.1% / 33.6%	24.3% / 23.9%
IIT2 Lexical Sample	34.6% / 34.1%	24.7% / 24.4%
Baseline Lesk	33.1% / 33.1%	22.6% / 22.6%
Best Non-Corpus	36.7% / 36.7%	29.3% / 29.3%

Table 4
SENSEVAL-2 English Lexical Sample Results

System	Course Grained Precision/Recall	Fine Grained Precision/Recall
IIT1 All Word	29.4% / 29.1% *	28.7% / 28.3% *
IIT2 All Word	33.5% / 33.2% *	32.8% / 32.5% *
IIT3 All Word	30.1% / 29.7% *	29.4% / 29.1% *
Best Non-Corpus	46.0% / 46.0%	45.1% / 45.1%

Table 5
SENSEVAL-2 English All Word Results

3 Results

Table 4 and Table 5 show the results for IIT1, IIT2 and IIT3 as well as that of the Lesk Baseline (English Lexical Sample task) and the best non-corpus based system, the CRL DIMAP system. The SENSEVAL-2 (2001) website presents the complete competition results as well as the CRL DIMAP and baseline system descriptions.

The IIT1 and IIT2 performed better than the comparable baseline system but not as well as the best system in its class. The IIT3 approach improves on the performance of IIT1 by using its prior annotations in tagging subsequent words.

Due to time constraints, the English All Word submissions only processed the first 12% of the corpus. The recall values marked * consider only those instances attempted.

4 Discussion

Many of the examples in WordNet were the result of lexicographers expanding synset information to clarify sense distinctions for the annotators of the Semcor corpus (Fellbaum, 1998). This makes a compelling argument for the use of these WordNet examples to assist in a computational disambiguating process.

The examples for rare word senses could be used to provide corpus-based statistical methods with additional evidence. Such an approach should help address the knowledge acquisition bottleneck.

The implementation and results presented here do not seem to justify this optimism. There are several reasons, though, why the method should not be dismissed without further investigation:

- The example sets were empty for a number of the candidate word senses. When this occurred, the system constructed a pseudo example by appending the WordNet gloss to the target word. This was sufficient for most collocation senses and some non-collocation senses such as *call* as in *calling a square dance* (where the gloss includes *square* and *dance*, one of which is highly likely to occur in any use of the sense). Others such as *day* as in *sidereal day* or *turn off* (gloss *cause to feel intense dislike or distaste*) competed at a disadvantage.
- The pattern matching and scoring methods were never tuned against any corpus data. This allowed the algorithm to have few competitors in the class of untrained systems, but scoring methods relied on intuition-founded heuristics. Such tuning should improve precision and recall.
- The approach was developed to be used in tandem with statistical approaches. Further research is required before its additive value can be fully assessed. IIT3 would have done better to be based on IIT2 and an approach maximizing the scores for a sentence should do even better.
- The best-matching example was chosen regardless of how bad a match was involved. The system also defaulted to the first sense encountered when all examples had a zero score. Using threshold score values may well provide substantial precision improvements (at the expense of recall).
- Semantic annotation of the WordNet examples should improve the results.

In addition, the following programming errors affected the precision and recall results:

- The generated answers for many adjective senses (those with similarity relations) were incorrectly formatted and were therefore always scored as incorrect. For example, in the IIT1 entry for the English Lexical Sample, 7.1% of all annotations were incorrectly formatted. Scoring only the answers that were correctly formatted raises the course-grained precision for IIT1 to 36.7% and fine-grained precision to 26.1%, competitive with the course-grained performance of the best non-corpus system.
- No annotations were generated for target words preceded by the word *to*. This results in recall ≠ precision as seen in Table 4 and Table 5.
- In a few rare cases, the system identified the incorrect example word as the example anchor. One such occurrence was the synset *art, fine art* and the example *a fine collection of art*. The system considered it an example of the *fine art* collocation and chose *fine* as the anchor.

5 Conclusion

The approach presented here does not appear to be sufficient for a stand-alone word sense disambiguation solution. Whether this method can be combined with other methods to improve their results requires further investigation.

References

Christiane Fellbaum, ed. (1998) *WordNet: An Electronic Lexical Database*. The MIT Press, Cambridge, Massachusetts

Sherwood Haynes (2001) *http://skhii.home.mindspring.com*

Nancy Ide and Jean Véronis (1998) *Introduction to the Special Issue on Word Sense Disambiguation: The State of the Art.* Computational Linguistics, 24/1, pp. 1 – 40.

Adam Kilgarriff and Martha Palmer (2000) *Introduction to the Special Issue on SENSEVAL.* Computers and the Humanities 34/1, pp. 1 – 13.

Adam Kilgarriff and Joseph Rosenzweig (2000) *Framework and Results for English SENSEVAL.* Computers and the Humanities 34/1, pp. 15 – 48.

George Miller, ed. (1990) *WordNet: An On-line Lexical Database.* International Journal of Lexicography, 3/4

SENSEVAL-2 (2001) *http://www.sle.sharp.co.uk/senseval2*

Yorick Wilks (1988) *Forward.* In "Lexical Ambiguity Resolution: Perspectives from Psycholinguistics, Neuropsychology, and Artificial Intelligence," Small, S., Cottrell, G., & Tanenhaus, M. ed., Morgan Kaufmann Publishers, Inc., San Mateo, California, pp. iii – ix.

WordNet (2001) version 1.7, available at *http://www.cogsci.princeton.edu/~wn/*