

# OWNS: Cross-lingual Word Sense Disambiguation Using Weighted Overlap Counts and Wordnet Based Similarity Measures

**Lipta Mahapatra**   **Meera Mohan**   **Mitesh M. Khapra**   **Pushpak Bhattacharyya**  
Dharmasinh Desai University   Indian Institute of Technology Bombay  
Nadiad, India   Powai, Mumbai 400076, India  
lipta.mahapatra89@gmail.com   miteshk@cse.iitb.ac.in  
mu.mohan@gmail.com   pb@cse.iitb.ac.in

## Abstract

We report here our work on English French Cross-lingual Word Sense Disambiguation where the task is to find the best French translation for a target English word depending on the context in which it is used. Our approach relies on identifying the nearest neighbors of the test sentence from the training data using a pairwise similarity measure. The proposed measure finds the affinity between two sentences by calculating a weighted sum of the word overlap and the semantic overlap between them. The semantic overlap is calculated using standard Wordnet Similarity measures. Once the nearest neighbors have been identified, the best translation is found by taking a majority vote over the French translations of the nearest neighbors.

## 1 Introduction

Cross Language Word Sense Disambiguation (CL-WSD) is the problem of finding the correct target language translation of a word given the context in which it appears in the source language. In many cases a full disambiguation may not be necessary as it is common for different meanings of a word to have the same translation. This is especially true in cases where the sense distinction is very fine and two or more senses of a word are closely related. For example, the two senses of the word *letter*, namely, “*formal document*” and “*written/printed message*” have the same French translation “*lettre*”. The problem is thus reduced to distinguishing between the coarser senses of a word and ignoring the finer sense distinctions which is known to be a common cause of errors in conventional WSD. CL-WSD can thus be seen as a slightly relaxed version of the conventional

WSD problem. However, CL-WSD has its own set of challenges as described below.

The translations learnt from a parallel corpus may contain a lot of errors. Such errors are hard to avoid due to the inherent noise associated with statistical alignment models. This problem can be overcome if good bilingual dictionaries are available between the source and target language. EuroWordNet<sup>1</sup> can be used to construct such a bilingual dictionary between English and French but it is not freely available. Instead, in this work, we use a noisy statistical dictionary learnt from the Europarl parallel corpus (Koehn, 2005) which is freely downloadable.

Another challenge arises in the form of matching the lexical choice of a native speaker. For example, the word *coach* (as in, *vehicle*) may get translated differently as *autocar*, *autobus* or *bus* even when it appears in very similar contexts. Such decisions depend on the native speaker’s intuition and are very difficult for a machine to replicate due to their inconsistent usage in a parallel training corpus.

The above challenges are indeed hard to overcome, especially in an unsupervised setting, as evidenced by the lower accuracies reported by all systems participating in the SEMEVAL Shared Task on Cross-lingual Word Sense Disambiguation (Lefever and Hoste, 2010). Our system ranked second in the English French task (in the *out-of-five* evaluation). Even though its average performance was lower than the baseline by 3% it performed better than the baseline for 12 out of the 20 target nouns.

Our approach identifies the *top-five* translations of a word by taking a majority vote over the translations appearing in the nearest neighbors of the test sentence as found in the training data. We use a pairwise similarity measure which finds the affinity between two sentences by calculating a

<sup>1</sup><http://www.illc.uva.nl/EuroWordNet>

weighted sum of the word overlap and the semantic overlap between them. The semantic overlap is calculated using standard Wordnet Similarity measures.

The remainder of this paper is organized as follows. In section 2 we describe related work on WSD. In section 3 we describe our approach. In Section 4 we present the results followed by conclusion in section 5.

## 2 Related Work

Knowledge based approaches to WSD such as Lesk’s algorithm (Lesk, 1986), Walker’s algorithm (Walker and Amsler, 1986), Conceptual Density (Agirre and Rigau, 1996) and Random Walk Algorithm (Mihalcea, 2005) are fundamentally overlap based algorithms which suffer from data sparsity. While these approaches do well in cases where there is a surface match (*i.e.*, *exact word match*) between two occurrences of the target word (say, training and test sentence) they fail in cases where there is a semantic match between two occurrences of the target word even though there is no surface match between them. The main reason for this failure is that these approaches do not take into account semantic generalizations (*e.g.*, *train is a vehicle*).

On the other hand, WSD approaches which use Wordnet based semantic similarity measures (Patwardhan et al., 2003) account for such semantic generalizations and can be used in conjunction with overlap based approaches. We therefore propose a scoring function which combines the strength of overlap based approaches – frequently co-occurring words indeed provide strong clues – with semantic generalizations using Wordnet based similarity measures. The disambiguation is then done using  $k$ -NN (Ng and Lee, 1996) where the  $k$  nearest neighbors of the test sentence are identified using this scoring function. Once the nearest neighbors have been identified, the best translation is found by taking a majority vote over the translations of these nearest neighbors.

## 3 Our approach

In this section we explain our approach for Cross Language Word Sense Disambiguation. The main emphasis is on disambiguation *i.e.* finding English sentences from the training data which are closely related to the test sentence.

### 3.1 Motivating Examples

To explain our approach we start with two motivating examples. First, consider the following occurrences of the word *coach*:

- $S_1$ :...*carriage of passengers by **coach** and **bus**...*
- $S_2$ :...*occasional services by **coach** and **bus** and the transit operations...*
- $S_3$ :...*the Gloucester **coach** saw the game...*

In the first two cases, the word *coach* appears in the sense of a *vehicle* and in both the cases the word *bus* appears in the context. Hence, the surface similarity (*i.e.*, word-overlap count) of  $S_1$  and  $S_2$  would be higher than that of  $S_1$  and  $S_3$  and  $S_2$  and  $S_3$ . This highlights the strength of overlap based approaches – frequently co-occurring words can provide strong clues for identifying similar usage patterns of a word.

Next, consider the following two occurrences of the word *coach*:

- $S_1$ :...*I **boarded** the **last coach** of the **train**...*
- $S_2$ :...*I **alighted** from the **first coach** of the **bus**...*

Here, the surface similarity (*i.e.*, word-overlap count) of  $S_1$  and  $S_2$  is zero even though in both the cases the word *coach* appears in the sense of *vehicle*. This problem can be overcome by using a suitable Wordnet based similarity measure which can uncover the hidden semantic similarity between these two sentences by identifying that {bus, train} and {boarded, alighted} are closely related words.

### 3.2 Scoring function

Based on the above motivating examples, we propose a scoring function for calculating the similarity between two sentences containing the target word. Let  $S_1$  be the test sentence containing  $m$  words and let  $S_2$  be a training sentence containing  $n$  words. Further, let  $w_{1i}$  be the  $i$ -th word of  $S_1$  and let  $w_{2j}$  be the  $j$ -th word of  $S_2$ . The similarity between  $S_1$  and  $S_2$  is then given by,

$$\begin{aligned} Sim(S_1, S_2) = & \lambda * Overlap(S_1, S_2) \\ & + (1 - \lambda) * Semantic\_Sim(S_1, S_2) \end{aligned} \quad (1)$$

where,

$$Overlap(S_1, S_2) = \frac{1}{m+n} \sum_{i=1}^m \sum_{j=1}^n freq(w_{1i}) * \mathbf{1}_{\{w_{1i}=w_{2j}\}}$$

and,

$$Semantic\_Sim(S_1, S_2) = \frac{1}{m} \sum_{i=1}^m Best\_Sim(w_{1i}, S_2)$$

where,

$$Best\_Sim(w_{1i}, S_2) = \max_{w_{2j} \in S_2} lch(w_{1i}, w_{2j})$$

We used the *lch* measure (Leacock and Chodorow, 1998) for calculating semantic similarity of two words. The semantic similarity between  $S_1$  and  $S_2$  is then calculated by simply summing over the maximum semantic similarity of each constituent word of  $S_1$  over all words of  $S_2$ . Also note that the overlap count is weighted according to the frequency of the overlapping words. This frequency is calculated from all the sentences in the training data containing the target word. The rationale behind using a frequency-weighted sum is that more frequently appearing co-occurring words are better indicators of the sense of the target word (of course, stop words and function words are not considered). For example, the word *bus* appeared very frequently with *coach* in the training data and was a strong indicator of the *vehicle* sense of *coach*. The values of  $Overlap(S_1, S_2)$  and  $Semantic\_Sim(S_1, S_2)$  are appropriately normalized before summing them in Equation (1). To prevent the semantic similarity measure from introducing noise by over-generalizing we chose a very high value of  $\lambda$ . This effectively ensured that the  $Semantic\_Sim(S_1, S_2)$  term in Equation (1) became active only when the  $Overlap(S_1, S_2)$  measure suffered data sparsity. In other words, we placed a higher bet on  $Overlap(S_1, S_2)$  than on  $Semantic\_Sim(S_1, S_2)$  as we found the former to be more reliable.

### 3.3 Finding translations of the target word

We used GIZA++<sup>2</sup> (Och and Ney, 2003), a freely available implementation of the IBM alignment models (Brown et al., 1993) to get word level alignments for the sentences in the English-French

<sup>2</sup><http://sourceforge.net/projects/giza/>

portion of the Europarl corpus. Under this alignment, each word in the source sentence is aligned to zero or more words in the corresponding target sentence. Once the nearest neighbors for a test sentence are identified using the similarity score described earlier, we use the word alignment models to find the French translation of the target word in the top- $k$  nearest training sentences. These translations are then ranked according to the number of times they appear in these top- $k$  nearest neighbors. The top-5 most frequent translations are then returned as the output.

## 4 Results

We report results on the English-French Cross-Lingual Word Sense Disambiguation task. The test data contained 50 instances for 20 polysemous nouns, namely, *coach*, *education*, *execution*, *figure*, *job*, *letter*, *match*, *mission*, *mood*, *paper*, *post*, *pot*, *range*, *rest*, *ring*, *scene*, *side*, *soil*, *strain* and *test*. We first extracted the sentences containing these words from the English-French portion of the Europarl corpus. These sentences served as the training data to be compared with each test sentence for identifying the nearest neighbors. The appropriate translations for the target word in the test sentence were then identified using the approach outlined in section 3.2 and 3.3. For the *best evaluation* we submitted two runs: one containing only the top-1 translation and another containing top-2 translations. For the *oof evaluation* we submitted one run containing the top-5 translations. The system was evaluated using Precision and Recall measures as described in the task paper (Lefever and Hoste, 2010). In the *oof evaluation* our system gave the second best performance among all the participants. However, the average precision was 3% lower than the baseline calculated by simply identifying the five most frequent translations of a word according to GIZA++ word alignments. A detailed analysis showed that in the *oof evaluation* we did better than the baseline for 12 out of the 20 nouns and in the *best evaluation* we did better than the baseline for 5 out of the 20 nouns. Table 1 summarizes the performance of our system in the *best evaluation* and Table 2 gives the detailed performance of our system in the *oof evaluation*. In both the evaluations our system provided a translation for every word in the test data and hence the precision was same as recall in all cases. We refer to our system as OWNS (**O**verlap

and WordNet Similarity).

System	Precision	Recall
OWNS	16.05	16.05
Baseline	20.71	20.71

Table 1: Performance of our system in *best evaluation*

Word	OWNS (Precision)	Baseline (Precision)
coach	<b>45.11</b>	39.04
education	<b>82.15</b>	80.4
execution	<b>59.22</b>	39.63
figure	30.56	35.67
job	<b>43.93</b>	40.98
letter	<b>46.01</b>	42.34
match	<b>31.01</b>	15.73
mission	55.33	97.19
mood	35.22	64.81
paper	<b>48.93</b>	40.95
post	36.65	41.76
pot	26.8	65.23
range	16.28	17.02
rest	<b>39.89</b>	38.72
ring	<b>39.74</b>	33.74
scene	33.89	38.7
side	<b>37.85</b>	36.58
soil	<b>67.79</b>	59.9
strain	21.13	30.02
test	<b>64.65</b>	61.31
Average	43.11	45.99

Table 2: Performance of our system in *oof evaluation*

## 5 Conclusion

We described our system for English French Cross-Lingual Word Sense Disambiguation which calculates the affinity between two sentences by combining the weighted word overlap counts with semantic similarity measures. This similarity score is used to find the nearest neighbors of the test sentence from the training data. Once the nearest neighbors have been identified, the best translation is found by taking a majority vote over the translations of these nearest neighbors. Our system gave the second best performance in the *oof evaluation* among all the systems that participated in the English French Cross-Lingual Word Sense Disambiguation task. Even though the average performance of our system was less than the

baseline by around 3%, it outperformed the baseline system for 12 out of the 20 nouns.

## References

- Eneko Agirre and German Rigau. 1996. Word sense disambiguation using conceptual density. In *In Proceedings of the 16th International Conference on Computational Linguistics (COLING)*.
- Peter E Brown, Vincent J. Della Pietra, Stephen A. Della Pietra, and Robert L. Mercer. 1993. The mathematics of statistical machine translation: parameter estimation. *Computational Linguistics*, 19:263–311.
- P. Koehn. 2005. Europarl: A parallel corpus for statistical machine translation. In *In Proceedings of the MT Summit*.
- C. Leacock and M. Chodorow, 1998. *Combining local context and WordNet similarity for word sense identification*, pages 305–332. In C. Fellbaum (Ed.), MIT Press.
- Els Lefever and Veronique Hoste. 2010. Semeval-2010 task 3: Cross-lingual word sense disambiguation. In *Proceedings of the 5th International Workshop on Semantic Evaluations (SemEval-2010)*, Association for Computational Linguistics.
- Michael Lesk. 1986. Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. In *In Proceedings of the 5th annual international conference on Systems documentation*.
- Rada Mihalcea. 2005. Large vocabulary unsupervised word sense disambiguation with graph-based algorithms for sequence data labeling. In *In Proceedings of the Joint Human Language Technology and Empirical Methods in Natural Language Processing Conference (HLT/EMNLP)*, pages 411–418.
- Hwee Tou Ng and Hian Beng Lee. 1996. Integrating multiple knowledge sources to disambiguate word senses: An exemplar-based approach. In *In Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 40–47.
- Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1):19–51.
- Siddharth Patwardhan, Satanjeev Banerjee, and Ted Pedersen. 2003. Using measures of semantic relatedness for word sense disambiguation. In *In proceedings of the Fourth International Conference on Intelligent Text Processing and Computation Linguistics (CICLing)*.
- D. Walker and R. Amsler. 1986. The use of machine readable dictionaries in sublanguage analysis. In *In Analyzing Language in Restricted Domains*, Grishman and Kittredge (eds), LEA Press, pages 69–83.