

Unsupervised Word Sense Disambiguation with Multilingual Representations

Erwin Fernandez-Ordoñez, Rada Mihalcea, Samer Hassan

Department of Computer Science

University of North Texas

egf0011@unt.edu, rada@cs.unt.edu, samer@unt.edu

Abstract

In this paper we investigate the role of multilingual features in improving word sense disambiguation. In particular, we explore the use of semantic clues derived from context translation to enrich the intended sense and therefore reduce ambiguity. Our experiments demonstrate up to 26% increase in disambiguation accuracy by utilizing multilingual features as compared to the monolingual baseline.

Keywords: word sense disambiguation, multilingual natural language processing, lexical semantics

1. Introduction

Ambiguity has been always intertwined with human language and its evolution. Some argue that ambiguity of the human languages is a byproduct of its complexity, with words that are frequently used in language often being assigned to more than one reference in the real world, thus resulting in ambiguity. For instance, the word “bank” has several distinct interpretations, including that of “financial institution” and “edge of a river.” The following sentences illustrate the use of this word with the two meanings: *The van pulled up outside the bank and three masked men got out*, where the word “bank” has a different meaning as compared to its usage in the context *The boy leapt from the bank into the cold water*. While it is often easy for a human to identify the correct meaning of a word in a given context, the same task when performed by a computer is among the most difficult problems in natural language processing.

In this paper, we propose a method for word sense disambiguation (WSD), defined as the task of automatically assigning a meaning to an ambiguous word in a given context. Specifically, similar to one of the original WSD methods (Lesk, 1986), we formulate the task under an unsupervised setting, and assume that the only knowledge available is a dictionary with definitions for the various meanings of a given ambiguous word.

We investigate a new unsupervised WSD method that is able to take additional advantage of a multilingual representation of the word sense definitions and of the context where the ambiguous word occurs. We consequently try to identify the correct meaning of the word in several multilingual spaces. We show that by

using this multilingual representation, we are able to improve the performance of the WSD system by a significant margin, as compared to a traditional system that uses only monolingual features.

2. Related Work

Despite the large number of word sense disambiguation methods that have been proposed so far, targeting the resolution of word ambiguity in different languages, there are only a few methods that try to explore more than one language at a time.

The work that is perhaps most closely related to our is the WSD method with multilingual features proposed in (Banea and Mihalcea, 2011), where both training and test contexts are translated into French, German, and Spanish, and consequently a supervised WSD algorithm is applied on this multilingual feature space. In that work, the multilingual representation brings error rate reductions of up to 25%, as compared to the monolingual classifier. All the evaluations are however done in a supervised setting, which, unlike our method, assumes the availability of hand-annotated data.

Another closely related work is the bilingual bootstrapping method introduced in (Li and Li, 2002), where word translations are automatically disambiguated using information iteratively drawn from two languages. Unlike that approach, which iterates between two languages to select the correct translation for a given target word, in our method we simultaneously use the features extracted from several languages.

There have also been a number of attempts to exploit parallel corpora for word sense disambiguation (Resnik and Yarowsky, 1999; Diab and Resnik, 2002;

Ng et al., 2003), but in that line of work the parallel texts were mainly used as a way to induce word senses or to create sense-tagged corpora, rather than as a source of additional multilingual views for the disambiguation features. Another related technique is concerned with the selection of correct word senses in context using large corpora in a second language (Dagan and Itai, 1994), but as before, the additional language is used to help distinguishing between the word senses in the original language, and not as a source of additional information for the disambiguation context. Also related is the recent SEMEVAL task that has been proposed for cross-lingual lexical substitution, where the word sense disambiguation task was more flexibly formulated as the identification of cross-lingual lexical substitutes in context (Mihalcea et al., 2010). A number of different approaches have been proposed by the teams participating in the task, and although several of them involved the translation of contexts or substitutes from one language to another, none of them attempted to make simultaneous use of the information available in the two languages.

Finally, although the multilingual subjectivity classifier proposed in (Banea et al., 2010) is not directly applicable to the disambiguation task we address in this paper, their findings are similar to ours. In that paper, the authors showed how a natural language task can benefit from the use of features drawn from multiple languages, thus supporting the hypothesis that multilingual features can be effectively used to improve the accuracy of a monolingual classifier.

3. Motivation

To motivate our work and demonstrate the utility of using translation for WSD, we present several examples in Table 1. The sentences were collected from the web and they showcase three different senses of the word “capital.” Namely, capital as “main city,” as “available wealth or assets,” and as “letter represented in uppercase.” Along with each English sentence we also provide its Spanish translation. In the first example, we see that “capital” carries the meaning of city and the context provides some useful clues, among them “country,” geo-location attributes like “South West,” and some named entities like “Bangkok” and “Thailand.” The context also provides ambiguous clues such as “bank.” In the hypothetical scenario in which a WSD system would strongly favor the “bank” clue to support the “asset” sense of capital, the incorrect sense assignment would result. By translating the context to Spanish, we notice that the ambiguous notion of “bank” has been resolved to “orilla” (shore), hence resulting in more cohesive context clues to the intended sense. The notion repeats in the second example where “capital” indicates the sense of assets. This time, the ambiguous clue of bank has been cor-

rectly resolved to “bancos” and “bancarios” hence strengthening the financial aspect of the context and biasing the classification of the “capital” sense in the proper direction. Similarly, the last two examples exhibit an analogous pattern. While Example 3 carries the “capital” sense of assets, it emits ambiguous clues like “letter” which represents an alternative sense. The Spanish translation conveniently resolves this inconsistency by utilizing “carta” (paper). In Example 4, we see that not only did the translation help disambiguate polysemous contextual clues like “letter,” but it also disambiguated the target word (capital) by choosing “mayúsculas” (upper case) as its translation. This demonstrates the inherent power of alternative context representation through translation, as the ambiguity of the context is weakened with every additional language. The contextual clues permeate from every language and lend their disambiguating power to every monolingual fragment, thus allowing for a clearer relationship to transpire. This relationship manages to capture the quintessential meaning of the fragments in question.

<p>En: Bangkok is the capital of Thailand, in the South West part of the country, on the east <u>bank</u> of the Chao Phraya River, near the Gulf of Thailand.</p> <p>Es: Bangkok es la capital de Tailandia, en la parte sur oeste del país, en la <u>orilla</u> oriental del río Chao Phraya, cerca del Golfo de Tailandia.</p>
<p>En: Europe’s big <u>banks</u> will be forced to find €108bn (\$150bn) of fresh capital over the next six to nine months under a deal to strengthen the <u>banking</u> system agreed by European Union finance ministers.</p> <p>Es: Los grandes <u>bancos</u> europeos se verán obligados a encontrar €108bn (\$ 150 mil millones) de capital fresco en los próximos seis a nueve meses en virtud de un acuerdo para fortalecer el sistema <u>bancario</u> acordado por la Unión Europea a los ministros de finanzas.</p>
<p>En: Please find attached the reply of the Supervisory Board to the <u>letter</u> from the Capital Assets Management Agency of the Republic of Slovenia (AUKN) with regard to holiday allowance paid for 2011.</p> <p>Es: Se adjunta la respuesta del Consejo de Vigilancia a la <u>carta</u> de la Agencia de Gestión de Activos de Capital de la República de Eslovenia (AUKN) con respecto al subsidio de vacaciones pagadas para el año 2011.</p>
<p>En: The usage of <u>capital letters</u> can have different meanings in emails CAPS can sound rude, but when used properly, especially in web design they can be pretty effective to get attention and to showcase main information</p> <p>Es: El uso de <u>letras mayúsculas</u> puede tener significados diferentes - en los correos electrónicos CAPS pueden sonar grosero, pero cuando se utilizan correctamente, en especial en el diseño de paginas de web, pueden ser muy eficaces para llamar la atención y mostrar la información principal.</p>

Table 1: Examples for the ambiguous word “capital”

4. Multilingual Word Sense Disambiguation

We approach the WSD task using an unsupervised method based on the Lesk algorithm (Lesk, 1986). Given a sequence of words, the original Lesk algorithm attempts to identify the combination of word senses that maximizes the redundancy (overlap) across all corresponding definitions. The algorithm was later improved through a method for simulated annealing (Cowie et al., 1992), which solved the combinatorial explosion of word senses, while still finding an optimal solution. However, recent comparative evaluations of different variants of the Lesk algorithm have shown that the performance of the original algorithm is significantly exceeded by an algorithm variation that relies on the overlap between word senses and current context (Vasilescu et al., 2004). We are thus using this latter Lesk variant in our implementation, and select the meaning of an open-class word by finding the word sense that leads to the highest overlap between the corresponding dictionary definition and the current context.

One of the main drawbacks associated with the Lesk algorithm is the fact that often times no overlap is found between the word sense definitions and the given input context, which is primarily due to the small size of these definitions and the contexts, as well as to the diversity of language. Even if for a given example such as *The van pulled up outside the bank and three masked men got out*, it is clear that the intended meaning for “bank” was that of “financial institution,” it may be difficult for a computer to find any overlap with the corresponding definition of “a financial institution that accepts deposits and channels the money into lending activities.” We try to address this problem by expanding the representation into a multilingual space, and therefore seek to find an overlap between the context and the sense definitions under different linguistic realizations. In this way, as mentioned before, we are solving the ambiguity of several word representations by using their translation in other languages. At the same time, we are also identifying additional matches by using the representation of words in other languages.

In our experiments, we use four different languages: English (En), French (Fr), German (De), and Spanish (Es). For a given target word, we first identify its meaning definitions in the WordNet dictionary (Miller, 1995). Next, since we also need the corresponding definitions of these word meanings in the other languages under consideration, we first explored the idea of using EuroWordNet (Vossen, 1998). We ran however into several issues, the major one being the partial coverage of this resource for some of the word senses, which would have resulted in gaps in our word sense representations. We thus decided to use another solution, and gather definitions for the target

word senses in the other languages by using automatic machine translation. An alternative we may consider in future work is BabelNet (Navigli and Ponzetto, 2010), which combines WordNet and Wikipedia into a very large multilingual network.

For a target word, we use the Google Translate API and collect translations in the three languages (French, German, Spanish) for all its sense definitions. These translations, along with the original English definitions, form the *multilingual sense representations* for the target word. Given a context, we apply a similar process, and create a *multilingual context representation* by translating the text into the three languages, again by using the Google Translate API. Although the automatic translation is naturally error prone, based on previous work that compared automatic and manual translations and their role in language processing tasks (Banea et al., 2008), we do not expect the potential translation noise to play an important role in the overall quality of the disambiguation system.

Finally, the simplified Lesk algorithm is applied separately on each of the four language representations, and the sense that maximizes the overlap between its definition and the input context is selected. The final sense selection is then made using a voting among the senses chosen for the individual languages.

To measure the overlap, we use a simple metric that counts the number of common words between a definition and a context, after tokenizing the text and removing the function words. This metric is normalized with the length of the definition. We also experimented with stemming as a way to increase the number of word matches between the definition and the context, but we did not notice any improvements, and therefore our current implementation does not use stemming.

5. Experiments and Evaluations

In order to evaluate our approach, we use a subset of 30 ambiguous words from the SEMEVAL 2007 task (Pradhan et al., 2007). Although the task covered 100 ambiguous words, we decided to only use a subset in our experiments for two main reasons. First, this is the same dataset as used in the supervised multilingual WSD evaluations reported in (Banea and Mihaicea, 2011). Second, because we are using the Google Translate API to collect the translations, we have to account for the limitations imposed by this API, which only allows for a limited number of contexts to be translated daily.¹ Note however that since our approach is unsupervised, it can be potentially applied to any word that has a dictionary definition (e.g., more than 100,000 words in the English WordNet).

¹This limitation would not apply if one would use an offline translation tool.

For the entire set of 30 ambiguous words, there are a total of 19,200 contexts, with an average of 640 contexts per word. Each of these contexts is translated into French, German, and Spanish, thus resulting in a total of 76,800 contexts.

For these 30 ambiguous words, we also extract their WordNet definitions for all their senses, for a total of 173 definitions. All these definitions are then translated into the three languages, and we apply the simplified Lesk algorithm on each individual language. We evaluate the performance of the algorithm for each language, by determining the number of times that the correct word sense is selected for each of the 19,200 contexts. Finally, we create and evaluate a meta sense classifier, which for each sense chooses the sense that was selected by most of the individual language classifiers, using a random sense selection to break the ties. Table 2 shows the results obtained for the individual words, along with the overall results calculated as the micro-average over all 30 words.

6. Discussion

As seen in Table 2, the combination of four languages leads to significant improvements with respect to the accuracy obtained for the individual languages, resulting in an error rate reduction of up to 26%. This result demonstrates the usefulness of using a multilingual space for WSD, which is inline with previous findings concerned with the use of multilingual features for other language processing tasks, including supervised WSD (Banea and Mihalcea, 2011) and subjectivity analysis (Banea et al., 2010).

To place results in perspective, we calculate an unsupervised baseline, determined by randomly selecting a sense for each context, which results in an accuracy of 22.29%. Compared to this baseline, the results obtained for all the individual languages are significantly better.

As a reference, we also calculate the accuracy obtained with the most frequent sense heuristic, determined as 64.0%. It is important to note that this is a supervised algorithm, as it relies on sense annotated corpora, and thus it is not directly comparable with our fully unsupervised method.

By looking at the results, we observe that the individual systems for the English, German, and Spanish models perform well (55%-57%), however the French model seems to perform poorly in comparison. We investigated this outlier and identified the translation quality as a possible cause for this weaker performance. Given the fine granularity of some of the word senses, the distinction between senses becomes even more difficult in another language. This effect could be accentuated for languages with lower translation quality. If one takes the number of internet users as an indicator for the corpora size available on the web,

there are 153 million users utilizing Spanish as their language, versus 75 million for German and 60 million for French,² potentially corresponding to less data available for producing English-French translations.

7. Conclusions

In this paper, we introduced a method for unsupervised WSD that relies on multilingual representations. We extend the Lesk algorithm to other languages, and combine the knowledge drawn from multiple WSD systems covering several languages. Through experiments on a SEMEVAL dataset, we show that the use of a multilingual space can lead to up to 26% error rate reduction as compared to a monolingual WSD system.

Acknowledgments

This material is based in part upon work supported by the National Science Foundation CAREER award #0747340 and IIS award #1018613. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

8. References

- C. Banea and R. Mihalcea. 2011. Word sense disambiguation with multilingual features. In *International Conference on Semantic Computing*, Oxford, UK.
- C. Banea, R. Mihalcea, J. Wiebe, and S. Hassan. 2008. Multilingual subjectivity analysis using machine translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2008)*, Honolulu, Hawaii.
- C. Banea, R. Mihalcea, and J. Wiebe. 2010. Multilingual subjectivity: Are more languages better? In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 28–36, Beijing, China, August.
- J. Cowie, L. Guthrie, and J. Guthrie. 1992. Lexical disambiguation using simulated annealing. In *Proceedings of the 5th International Conference on Computational Linguistics (COLING 1992)*.
- I. Dagan and A. Itai. 1994. Word sense disambiguation using a second language monolingual corpus. *Computational Linguistics*, 20(4):563–596.
- M. Diab and P. Resnik. 2002. An unsupervised method for word sense tagging using parallel corpora. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL 2002)*, Philadelphia, PA, July.

²<http://www.internetworldstats.com/stats7.htm>

Languages	Number senses	Number contexts	Individual languages				Multilingual
			En	De	Es	Fr	
approve	2	53	88.68	94.34	60.38	50.94	94.34
ask	6	348	24.14	31.61	22.13	20.40	26.72
bill	9	404	45.54	51.24	61.39	60.15	61.63
buy	7	164	45.73	54.27	63.41	60.98	66.46
capital	5	278	55.4	60.07	71.94	82.73	81.29
care	3	69	75.36	69.57	76.81	71.01	78.26
effect	5	178	75.84	61.24	74.16	82.02	80.90
exchange	6	363	34.71	49.86	44.63	30.03	38.84
explain	2	85	62.35	60.00	29.41	49.41	64.71
feel	3	347	49.57	42.94	58.5	65.13	66.28
grant	3	19	63.16	57.89	47.37	73.68	63.16
hold	10	129	15.5	12.40	19.38	17.05	18.60
hour	4	187	54.01	48.66	54.55	55.61	64.71
job	10	188	65.43	38.83	51.60	70.74	69.68
part	7	481	64.86	54.68	56.76	81.91	76.51
people	6	754	62.33	71.22	79.71	89.52	88.59
point	14	469	17.27	27.29	14.71	4.90	18.34
position	7	268	26.49	18.28	28.36	7.84	20.52
power	4	251	27.89	44.22	43.43	20.72	41.43
president	3	879	57.45	59.50	76.22	3.87	65.53
promise	2	50	84.00	82.00	76.00	86.00	90.00
propose	3	34	73.53	73.53	85.29	88.24	85.29
rate	2	1009	62.14	57.48	42.02	28.34	59.46
remember	6	121	83.47	95.04	83.47	71.90	96.69
rush	4	28	78.57	78.57	82.14	57.14	89.29
say	5	2161	85.75	90.10	90.51	77.42	94.17
see	10	158	24.05	25.32	17.72	8.23	22.78
state	4	617	34.68	14.59	42.95	35.17	27.23
system	7	450	26.00	20.89	24.67	18.22	26.89
value	5	335	18.21	13.73	23.88	14.03	17.01
work	9	230	36.96	40.00	39.57	46.52	52.17
AVERAGE	5.5	640	54.31	54.52	57.74	47.86	61.66

Table 2: Disambiguation accuracy for the unsupervised WSD method. Results are shown for the individual languages, and for the combined method that includes all languages.

- M.E. Lesk. 1986. Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone. In *Proceedings of the SIGDOC Conference 1986*, Toronto, June.
- C. Li and H. Li. 2002. Word translation disambiguation using bilingual bootstrapping. In *Proceedings of 40th Annual Meeting of the Association for Computational Linguistics*, Philadelphia, Pennsylvania.
- R. Mihalcea, R. Sinha, and D. McCarthy. 2010. Semeval-2010 task 2: Cross-lingual lexical substitution. In *Proceedings of the ACL Workshop on Semantic Evaluations*, Uppsala, Sweden.
- G. Miller. 1995. Wordnet: A lexical database. *Communication of the ACM*, 38(11).
- R. Navigli and S. Ponzetto. 2010. Babelnet: Building a very large multilingual semantic network. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Uppsala, Sweden.
- H.T. Ng, B. Wang, and Y.S. Chan. 2003. Exploiting parallel texts for word sense disambiguation: An empirical study. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL 2003)*, Sapporo, Japan, July.
- S. Pradhan, E. Loper, D. Dligach, and M. Palmer. 2007. Semeval-2007 task-17: English lexical sample, srl and all words. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, Prague, Czech Republic, June.
- P. Resnik and D. Yarowsky. 1999. Distinguishing systems and distinguishing senses: new evaluation methods for word sense disambiguation. *Natural Language Engineering*, 5(2):113–134.
- F. Vasilescu, P. Langlais, and G. Lapalme. 2004. Evaluating variants of the Lesk approach for disambiguating words. In *Proceedings of the Conference of Language Resources and Evaluations (LREC 2004)*, Lisbon, Portugal.
- P. Vossen. 1998. *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*. Kluwer Academic Publishers, Dordrecht.