

GPLSI-IXA: Using Semantic Classes to Acquire Monosemous Training Examples from Domain Texts

Rubén Izquierdo & Armando Suárez
GPLSI Group
University of Alicante. Spain
{ruben, armando}@dlsi.ua.es

German Rigau
IXA NLP Group.
EHU. Donostia, Spain
german.rigau@ehu.es

Abstract

This paper summarizes our participation in task #17 of SemEval-2 (All-words WSD on a specific domain) using a supervised class-based Word Sense Disambiguation system. Basically, we use Support Vector Machines (SVM) as learning algorithm and a set of simple features to build three different models. Each model considers a different training corpus: SemCor (SC), examples from monosemous words extracted automatically from background data (BG), and both SC and BG (SCBG). Our system explodes the monosemous words appearing as members of a particular WordNet semantic class to automatically acquire class-based annotated examples from the domain text. We use the class-based examples gathered from the domain corpus to adapt our traditional system trained on SemCor. The evaluation reveal that the best results are achieved training with SemCor and the background examples from monosemous words, obtaining results above the most frequent baseline and the fifth best position in the competition rank.

1 Introduction

As empirically demonstrated by the last SensEval and SemEval exercises, assigning the appropriate meaning to words in context has resisted all attempts to be successfully addressed. In fact, supervised word-based WSD systems are very dependent of the corpora used for training and testing the system (Escudero et al., 2000). One possible reason could be the use of inappropriate level of abstraction.

Most supervised systems simply model each polysemous word as a classification problem

where each class corresponds to a particular synset of the word. But, WordNet (WN) has been widely criticized for being a sense repository that often provides too fine-grained sense distinctions for higher level applications like Machine Translation or Question & Answering. In fact, WSD at this level of granularity has resisted all attempts of inferring robust broad-coverage models. It seems that many word-sense distinctions are too subtle to be captured by automatic systems with the current small volumes of word-sense annotated examples.

Thus, some research has been focused on deriving different word-sense groupings to overcome the fine-grained distinctions of WN (Hearst and Schütze, 1993), (Peters et al., 1998), (Mihalcea and Moldovan, 2001), (Agirre and LopezDeLaCalle, 2003), (Navigli, 2006) and (Snow et al., 2007). That is, they provide methods for grouping senses of the same **word**, thus producing coarser word sense groupings for better disambiguation.

In contrast, some research have been focused on using predefined sets of sense-groupings for learning class-based classifiers for WSD (Segond et al., 1997), (Ciaramita and Johnson, 2003), (Villarejo et al., 2005), (Curran, 2005), (Kohomban and Lee, 2005) and (Ciaramita and Altun, 2006). That is, grouping senses of different words into the same explicit and comprehensive semantic class. Most of the later approaches used the original Lexicographical Files of WN (more recently called SuperSenses) as very coarse-grained sense distinctions.

We suspect that selecting the appropriate level of abstraction could be on between both levels. Thus, we use the semantic classes modeled by the **Basic Level Concepts**¹ (BLC) (Izquierdo et al., 2007). Our previous research using BLC empirically demonstrated that this automatically derived

¹<http://adimen.si.ehu.es/web/BLC>

set of meanings groups senses into an adequate level of abstraction in order to perform class-based Word Sense Disambiguation (WSD) (Izquierdo et al., 2009). Now, we also show that class-based WSD allows to successfully incorporate monosemous examples from the domain text. In fact, the robustness of our class-based WSD approach is shown by our system that just uses the SemCor examples (SC). It performs without any kind of domain adaptation as the Most Frequent Sense (MFS) baseline.

In section 2 semantic classes used and selection algorithm used to obtain them automatically from WordNet are described. In section 3 the technique employed to extract monosemous examples from background data is described. Section 4 explains the general approach of our system, and the experiments designed, and finally, in section 5, the results and some analysis are shown.

2 Semantic Classes

The set of semantic classes used in this work are the **Basic Level Concepts**² (BLC) (Izquierdo et al., 2007).

These concepts are small sets of meanings representing the whole nominal and verbal part of WN. BLC can be obtained by a very simple method that uses basic structural WordNet properties. In fact, the algorithm only considers the relative number of relations of each synset along the hypernymy chain. The process follows a bottom-up approach using the chain of hypernymy relations. For each synset in WN, the process selects as its BLC the first local maximum according to the relative number of relations. The local maximum is the synset in the hypernymy chain having more relations than its immediate hyponym and immediate hypernym. For synsets having multiple hypernyms, the path having the local maximum with higher number of relations is selected. Usually, this process finishes having a number of preliminary BLC. Figure 1 shows an example of selection of a BLC. The figure represents the hypernymy hierarchy of WordNet, with circles representing synsets, and links between them representing hypernym relations. The algorithm selects the D synset as BLC for J, due to D is the first maximum in the hypernymy chain, according to the number of relations (F has 2 hyponyms, D has 3, and A has 2, so D is the first maximum).

²<http://adimen.si.ehu.es/web/BLC>

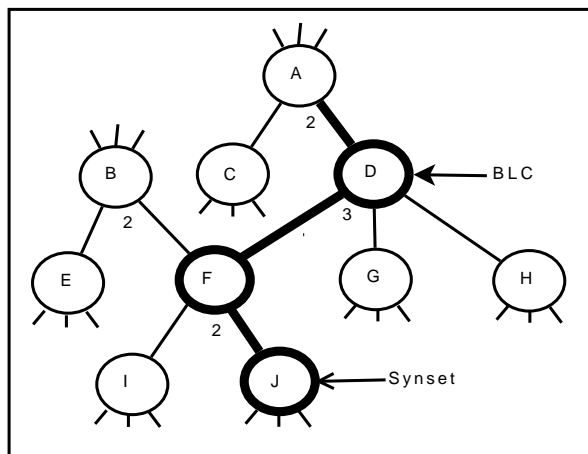


Figure 1: Example of BLC selection

Obviously, while ascending through this chain, more synsets are subsumed by each concept. The process finishes checking if the number of concepts subsumed by the preliminary list of BLC is higher than a certain threshold. For those BLC not representing enough concepts according to the threshold, the process selects the next local maximum following the hypernymy hierarchy. Thus, depending on the type of relations considered to be counted and the threshold established, different sets of BLC can be easily obtained for each WN version.

We have selected the set which considers WN version 3.0, the total number of relations per synset, and a minimum threshold of 20 concepts to filter out not representative BLC (BLC-20). This set has shown to reach good performance on previous SenseEval and SemEval exercises (Izquierdo et al., 2009). There are 649 different BLC for nouns on WordNet 3.0, and 616 for verbs. Table 2 shows the three most frequent BLC per POS, with the number of synsets subsumed by each concept, and its WordNet gloss.

3 Using Monosemous Examples from the Domain

We did not apply any kind of specific domain adaptation technique to our class-based supervised system. In order to adapt our supervised system to the environmental domain we only increased the training data with new examples of the domain. To acquire these examples, we used the environmental domain background documents provided by the organizers. Specifically, we used the 122 back-

PoS	Num.	BLC	Gloss
Nouns	4.792	person.n.01	a human being
	1.935	activity.n.01	any specific behavior
	1.846	act.n.02	something that people do or cause to happen
Verbs	1.541	change.v.01	cause to change; make different; cause a transformation
	1.085	change.v.02	undergo a change; become different in essence; losing one's or its original nature
	519	move.v.02	cause to move or shift into a new position or place, both in a concrete and in an abstract sense

Table 1: Most frequent BLC–20 semantic classes on WordNet 3.0

ground documents³. TreeTagger has been used to preprocess the documents, performing PoS tagging and lemmatization. Since the background documents are not semantically annotated, and our supervised system needs labeled data, we have selected only the monosemous words occurring in the documents. In this way, we have obtained automatically a large set of examples annotated with BLC. Table 3 presents the total number of training examples extracted from SemCor (SC) and from the background documents (BG). As expected, by this method a large number of monosemous examples can be obtained for nouns and verbs. Also as expected, verbs are much less productive than nouns. However, all these background examples correspond to a reduced set of 7,646 monosemous words.

	Nouns	Verbs	N+V
SC	87.978	48.267	136.245
BG	193.536	10.821	204.357
<i>Total</i>	<i>281.514</i>	<i>59.088</i>	<i>340.602</i>

Table 2: Number of training examples

Table 3 lists the ten most frequent monosemous nouns and verbs occurring in the background documents. Note that all these examples are monosemous according to BLC–20 semantic classes.

	Nouns		Verbs	
	Lemma	# ex.	Lemma	# ex.
1	biodiversity	7.476	monitor	788
2	habitat	7.206	achieve	784
3	specie	7.067	target	484
4	climate	3.539	select	345
5	european	2.818	enable	334
6	ecosystem	2.669	seem	287
7	river	2.420	pine	281
8	grassland	2.303	evaluate	246
9	datum	2.276	explore	200
10	directive	2.197	believe	172

Table 3: Most frequent monosemic words in BG

³We used the documents contained on the trial data and the background.

4 System Overview

Our system applies a supervised machine learning approach. We apply a feature extractor to represent the training examples of the examples acquired from SemCor and the background documents. Then, a machine learning engine uses the annotated examples to train a set of classifiers. Support Vector Machines (SVM) have been proven to be robust and very competitive in many NLP tasks, and in WSD in particular (Márquez et al., 2006). We used the SVM-Light implementation⁴ (Joachims, 1998).

We create a classifier for each semantic class. This approach has several advantages compared to word–based approach. The training data per classifier is increased (we can use examples of different target words for a single classifier, whenever all examples belong to the same semantic class), the polysemy is reduced (some different word senses can be collapsed into the same semantic class), and, finally, semantic classes provide higher levels of abstraction.

For each polysemous word occurring in the test corpus, we obtain its potential BLC–20 classes. Then, we only apply the classifiers corresponding to the BLC–20 classes of the polysemous word. Finally, our system simply selects the BLC–20 class with the greater prediction.

In order to obtain the correct WordNet 3.0 synset required by the task, we apply a simple heuristic that has shown to be robust and accurate (Kohomban and Lee, 2005). Our classifiers obtain first the semantic class, and then, the synset of the first WordNet sense that fits with the semantic class is assigned to the word.

We selected a simple feature set widely used in many WSD systems. In particular, we use a window of five tokens around the target word to extract word forms, lemmas; bigrams and trigrams of word forms and lemmas; trigrams of PoS tags,

⁴<http://svmlight.joachims.org>

and also the most frequent BLC–20 semantic class of the target word in the training corpus.

Our system is fully described in (Izquierdo et al., 2009). The novelty introduced here is the use of semantic classes to obtain monosemous examples from the domain corpus.

Following the same framework (BLC–20 semantic architecture and basic set of features) we designed three runs, each one using a different training corpus.

- SC: only training examples extracted from SemCor
- BG: only monosemous examples extracted from the background data
- SCBG: training examples extracted from SemCor and monosemous background data

The first run shows the behavior of a supervised system trained on a general corpus, and tested in a specific domain. The second one analyzes the contribution of the monosemous examples extracted from the background data. Finally, the third run studies the robustness of the approach when combining the training examples from SemCor and from the background.

5 Results and Discussion

A total of 29 runs has been submitted for the English All–words WSD on a Specific Domain. Table 5 shows the ranking results of our three runs with respect to the other participants. The figures for the most frequent sense (MFS) and random sense (Random) baselines are included.

In general, the results obtained are not very high. The best system only achieves a precision of 0.570, and the most frequent sense baseline reaches a precision of 0.505. This shows that the task is hard to solve, and the domain adaptation of WSD systems is not an easy task.

Interestingly, our worst result is obtained by the system using only the monosemous background examples (BG). This system ranks 23th with a Precision and Recall of 0.380. The system using only SemCor (SC) ranks 6th with Precision and Recall of 0.505. This is also the performance of the Most Frequent Sense baseline. As expected, the best result of our three runs is obtained when combining the examples from SemCor and the background (SCBG). This supervised system obtains

Rank	Precision	Recall
1	0.570	0.555
2	0.554	0.540
3	0.534	0.528
4	0.522	0.516
(SCBG) 5	0.513	0.513
<i>MFS</i>	<i>0.505</i>	<i>0.505</i>
(SC) 6	0.505	0.505
7	0.512	0.495
8	0.506	0.493
9	0.504	0.491
10	0.481	0.481
11	0.492	0.479
12	0.461	0.460
13	0.447	0.441
14	0.436	0.435
15	0.440	0.434
16	0.496	0.433
17	0.498	0.432
18	0.433	0.431
19	0.426	0.425
20	0.424	0.422
21	0.437	0.392
22	0.384	0.384
(BG) 23	0.380	0.380
24	0.381	0.356
25	0.351	0.350
26	0.370	0.345
27	0.328	0.322
28	0.321	0.315
29	0.312	0.303
<i>Random</i>	<i>0.230</i>	<i>0.230</i>

Table 4: Results of task#17

the 5th position with a Precision and Recall of 0.513 which is slightly above the baseline.

Possibly, the reason of low performance of the BG system is the high correlation between the features of the target word and its semantic class. In this case, these features correspond to the monosemous word while when testing corresponds to the target word. However, it also seems that class-based systems are robust enough to incorporate large sets of monosemous examples from the domain text. In fact, to our knowledge, this is the first time that a supervised WSD algorithm have been successfully adapted to an specific domain. Furthermore, our system trained only on SemCor also achieves a good performance, reaching the most frequent sense baseline, showing that class-based WSD approaches seem to be robust to domain variations.

Acknowledgments

This paper has been supported by the European Union under the project KYOTO (FP7 ICT-211423), the Valencian Region Government under PROMETEO project for excellence groups and the Spanish Government under the projects

KNOW2 (TIN2009-14715-C04-04) and TEXT-MESS-2 (TIN2009-13391-C04-04).

References

- E. Agirre and O. LopezDeLaCalle. 2003. Clustering wordnet word senses. In *Proceedings of RANLP'03*, Borovets, Bulgaria.
- M. Ciaramita and Y. Altun. 2006. Broad-coverage sense disambiguation and information extraction with a supersense sequence tagger. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'06)*, pages 594–602, Sydney, Australia. ACL.
- M. Ciaramita and M. Johnson. 2003. Supersense tagging of unknown nouns in wordnet. In *Proceedings of the Conference on Empirical methods in natural language processing (EMNLP'03)*, pages 168–175. ACL.
- J. Curran. 2005. Supersense tagging of unknown nouns using semantic similarity. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics (ACL'05)*, pages 26–33. ACL.
- G. Escudero, L. Màrquez, and G. Rigau. 2000. An Empirical Study of the Domain Dependence of Supervised Word Sense Disambiguation Systems. In *Proceedings of the joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora, EMNLP/VLC*, Hong Kong, China.
- M. Hearst and H. Schütze. 1993. Customizing a lexicon to better suit a computational task. In *Proceedings of the ACL SIGLEX Workshop on Lexical Acquisition*, Stuttgart, Germany.
- R. Izquierdo, A. Suarez, and G. Rigau. 2007. Exploring the automatic selection of basic level concepts. In Galia Angelova et al., editor, *International Conference Recent Advances in Natural Language Processing*, pages 298–302, Borovets, Bulgaria.
- Rubén Izquierdo, Armando Suárez, and German Rigau. 2009. An empirical study on class-based word sense disambiguation. In *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, pages 389–397, Athens, Greece, March. Association for Computational Linguistics.
- T. Joachims. 1998. Text categorization with support vector machines: learning with many relevant features. In Claire Nédellec and Céline Rouveirol, editors, *Proceedings of ECML-98, 10th European Conference on Machine Learning*, pages 137–142, Chemnitz, DE. Springer Verlag, Heidelberg, DE.
- Upali S. Kohomban and Wee Sun Lee. 2005. Learning semantic classes for word sense disambiguation. In *ACL '05: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 34–41, Morristown, NJ, USA. Association for Computational Linguistics.
- Ll. Màrquez, G. Escudero, D. Martínez, and G. Rigau. 2006. Supervised corpus-based methods for wsd. In E. Agirre and P. Edmonds (Eds.) *Word Sense Disambiguation: Algorithms and applications.*, volume 33 of *Text, Speech and Language Technology*. Springer.
- R. Mihalcea and D. Moldovan. 2001. Automatic generation of coarse grained wordnet. In *Proceeding of the NAACL workshop on WordNet and Other Lexical Resources: Applications, Extensions and Customizations*, Pittsburg, USA.
- R. Navigli. 2006. Meaningful clustering of senses helps boost word sense disambiguation performance. In *ACL-44: Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pages 105–112, Morristown, NJ, USA. Association for Computational Linguistics.
- W. Peters, I. Peters, and P. Vossen. 1998. Automatic sense clustering in eurowordnet. In *First International Conference on Language Resources and Evaluation (LREC'98)*, Granada, Spain.
- F. Segond, A. Schiller, G. Greffenstette, and J. Chanod. 1997. An experiment in semantic tagging using hidden markov model tagging. In *ACL Workshop on Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*, pages 78–81. ACL, New Brunswick, New Jersey.
- R. Snow, Prakash S., Jurafsky D., and Ng A. 2007. Learning to merge word senses. In *Proceedings of Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 1005–1014.
- L. Villarejo, L. Màrquez, and G. Rigau. 2005. Exploring the construction of semantic class classifiers for wsd. In *Proceedings of the 21th Annual Meeting of Sociedad Española para el Procesamiento del Lenguaje Natural SEPLN'05*, pages 195–202, Granada, Spain, September. ISSN 1136-5948.