Semantic Classes and Relevant Domains on WSD

Rubén Izquierdo[1], Sonia Vázquez[2], and Andrés Montoyo[2]*

[1] Computational Lexicology and Terminology Lab
Vrije University of Amsterdam. The Netherlands
ruben.izquierdobevia@vu.nl
[2] Department of Software and Computing Systems, University of Alicante, Spain
svazquez@dlsi.ua.es, montoyo@dlsi.ua.es

Abstract. Language ambiguities are a problem in various fields. For example, in Machine Translation the major cause of errors is ambiguity. Moreover, ambiguous words can be confusing for Information Extraction algorithms. Our purpose in this work is to provide a new approach to solve semantic ambiguities by dealing with the problem of the fine granularity of sense inventories. Our goal is to replace word senses with Semantic Classes that share properties, features and meanings. Also another semantic resources, Relevant Domains, is used to extract extract semantic information and enrich the process. The results obtained are evaluated in the Evaluation Exercises for the Semantic Analysis of Text (SensEval) framework.

1 Introduction

WSD is considered to be one of the most difficult problems in Artificial Intelligence and is also considered as an intermediate task within the Natural Language Processing (NLP) field [16]. Rather than an isolated problem, WSD often forms part of other NLP task. One of the most successful approaches in the last years has been Supervised Machine Learning, and in concrete *supervised learning from examples*. In this approach, Machine Learning models are induced from semantic annotated examples [8] usually using WordNet [3] as sense repository. Despite this wide use of WordNet, it has been criticized because the sense distinctions it provides are often too fine-grained for higher applications such as Machine Translation or Information Extraction, and represent too narrow distinctions that are not informative for NLP [9].

In this work we present a new supervised approach based on Semantic Classes in order to solve the problem of semantic ambiguity. Our goal is to provide a method with which obtain a hierarchy of semantic labels that represents a set of equivalent senses and use them into a supervised system. One of the main purposes of this work is to tackle the problems caused by semantic resources that are too fine-grained using another type of semantic indicators instead.

* This research work has been partially funded by the University of Alicante, Generalitat Valenciana, Spanish Government and the European Commission through the projects, "Tratamiento inteligente de la informacion para la ayuda a la toma de decisiones" (GRE12-44), ATTOS (TIN2012-38536-C03-03), LEGOLANG (TIN2012-31224), SAM (FP7-611312), FIRST (FP7-287607) and ACOMP/2013/067.

P. Sojka et al. (Eds.): TSD 2014, LNAI 8655, pp. 166–172, 2014.
This is preprint prepared by Proceedings editor for Springer International Publishing Switzerland.

Generally speaking, a semantic class is a concept that groups sub-concepts and word senses that share properties, features and meanings, such as vehicle with sub-concepts such as *car* or *helicopter* or building with *school* or *church* as sub-concepts. We will explore the performance of a WSD system when different types of semantic classes are used to build the semantic features. Firstly, we propose a basic system using traditional features, mainly contextual features. This system will be used as a reference to compare its results with other experiments. In order to improve this first approach, different semantic classes will be employed to generate the semantic features and the performance of this extended system will be analyzed.

2 State of the Art

Many researchers have focused their efforts on developing efficient WSD systems. However, WSD has been a challenging problem since [15] first introduced the question of ambiguity in the late 1950s, and it continues to be so. As last SensEval and SemEval[3] competitions have shown, there has been a very small improvement in the performance of the WSD systems.

Various approaches have been followed to tackle the problem of the excessively fine granularity of word senses. Some research has been focused on deriving different word–sense groupings to overcome the fine-grained description of word senses in WordNet [1,11,12]. However, these approaches only attempt to group senses of the same word, thus producing a coarse-grained repository of word senses, but only taking advantage of the reduction of the polysemy.

Like our approach, some research has been focused on using a predefined set of semantic classes for learning class–based classifiers for WSD [14,2]. Most of these approaches use the lexicographic files of WordNet (also known as SuperSenses (SS) as coarse-grained sense distinctions, but less attention has been paid to learning class-based classifiers from other available sense groupings.

The use of semantic classes rather than word senses in the development of a WSD system can also be found in unsupervised approaches, for instance in [7,4] which study the relations between meanings and domains.

3 Semantic Resources

In this section we explain the semantic resources used in our work. These can be divided into two groups: semantic classes and domains. A **semantic class** is a concept that subsumes other concepts that share common characteristics. The link between concepts and subconcepts is modelled by the relation “is-a”, for example a *car* “is-a” *vehicle*, (*vehicle* is a semantic class). A **domain** is a specific area of interest for humans. Each domain has a set of related words known as a terminology. The link between these words and their domain is modelled by the relation “belongs-to”. For example, the word *soccer* “belongs-to” the sports domain.

[3] `http://www.senseval.org`

There are different **Semantic Classes** repositories, each one groups the word senses in a different way, resulting in sets with a different granularity. We have used two types of Semantic Classes derived from WordNet: *Basic Level Concepts* (BLC) and *SuperSenses*. BLC are concepts that result from the compromise between representing as many concepts and features as possible. We have used the BLC sets created by [5]. SuperSenses are the name given to the WordNet Lexicographer Files within the framework of WSD. WordNet synsets are organized into 45 SuperSenses, based on syntactic categories (nouns, verbs, adjectives and adverbs) and logical groupings, such as person, phenomenon, feeling, location, etc.

On the other hand, we have used *WordNet Domains* (WND) and *Relevant Domains* (RD) as **domains** resources. WND [6] is a hierarchy of 165 domain labels which have been used to tag all WN synsets. This set of labels is organized into a taxonomy by following the Dewey Decimal Classification System[4].

WND is used as a basis to develop the resource called Relevant Domains (RD). The aim of this resource is to collect information from WordNet glosses in order to obtain new relations among words and domains. This resource has been previously evaluated in a WSD knowledge-based system [13]. The idea is to use the Mutual Information formula to calculate association ratios between words (or senses) and domains.

4 Semantic Class-based WSD System

In order to use the resources and features previously explained, we have followed a supervised Machine Learning approach to develop a set of class-based WSD taggers. Our systems use an implementation of a Support Vector Machine[5] algorithm to train the classifiers (one per class) in semantic annotated corpora in order to acquire positive and negative examples of each class and in the definition of a set of features to represent these examples. The system decides and selects from among the possible semantic classes defined for a word. In the sense approach, one classifier is generated for each word sense, and the classifiers choose between the possible senses of the word. The examples used to train a single classifier for a concrete word are all the examples of this word sense. In the class-based approach, one classifier is generated for each semantic class. Thus, when we wish to label a word, our system obtains the set of possible semantic classes for this word, and then launches each of the semantic classifiers related to these semantic categories. The most likely category is then selected for the word.

We believe that this semantic class-based approach has several advantages. First, semantic classes reduce the average polysemy degree of words (some word senses are grouped together within the same semantic class). Moreover, the problem of the acquisition bottleneck in supervised machine learning algorithms is attenuated, because the number of examples for each classifier is increased.

[4] It was selected because it provides a good coverage, is freely available and widely used to classify written data. `http://www.oclc.org/dewey`

[5] SVMLight

SemCor [10] was used for training while the corpora from the English all-words tasks of SensEval-2 (SE2)[6], SensEval-3 (SE3)[7] and SemEval–1(SEM1)[8] were used for testing. We also considered the SemEval-2007 coarse-grained task corpus for testing, but this dataset was discarded because this corpus is also annotated with clusters of word senses which are not comparable with Semantic Classes.

We have defined a set of features in order to represent the examples according to previous works in WSD and the nature of class-based WSD. Features widely used in literature have been selected. In particular, our systems use the following **basic features**. **Word-forms and lemmas** in a window of 10 words around the target word **PoS**: the concatenation of the preceding/following three/five PoS **bigrams and trigrams** formed by lemmas and word-forms and obtained in a window of 5 words. We use all tokens regardless of their PoS to build bi/trigrams. The target word is replaced with X in these features in order to increase their generalization for the semantic classifiers. We have also defined a set of **Semantic Features** with which explode different semantic resources in order to enrich the set of basic features: (1) the most frequent semantic class over SemCor for the target word, and the semantic classes of monosemic words around the target word.

Several types of semantic classes have been considered during the creation of these features, in particular, WordNet Domains, BLC–20[9] and Relevant Domains. We have selected these sets of semantic classes because they have been created by following different approaches and each one therefore has specific characteristics and a certain level of abstraction.

4.1 Experiments

We have defined several experiments focused on analyzing the influence of the semantic features in our semantic class-based system. In all the experiments, the classifiers have been built using BLC–20 semantic classes. In other words, one classifier is trained for each of the semantic classes within the BCL–20 set (in this set there are 649 semantic classes for nouns and 616 for verbs). Our classifiers therefore assign the proper semantic class to each ambiguos noun or verb in a text, as defined in the semantic-based WSD approach. Since the traditional evaluation in international WSD competitions (mainly SensEval and SemEval) is performed at a word sense level, we need to adapt the output of our system, and transform each semantic class assigned to a word into a word sense. We follow the same approach used in previous related work: our classifiers propose a Semantic Class, and the first sense for the word according to WordNet that belongs to that class is selected.

The evaluation was performed on SE2, SE3 and SEM1 All-Words tasks. We computed the standard measures for the evaluation of WSD systems (precision, recall

[6] `http://www.sle.sharp.co.uk/senseval2`
[7] `http://www.senseval.org/senseval3`
[8] `http://nlp.cs.swarthmore.edu/semeval`
[9] The value 20 refers to the threshold of minimum number of synsets that a possible BLC must subsume to be considered as a proper BLC. These BLC sets were built using all kind of relations.

and $F_{\beta=1}$ of each experiment on the three corpora. We have also included the baseline based on the most frequent sense ($F_{\beta=1}$ measure), and the result of the best system of the oficial participants in the task (also the $F_{\beta=1}$).

Table 1 contains the F1 value of our classifiers, when only one semantic class was used to build the semantic features (the type of this semantic class is represented in the header of the columns where RelDomW is Basic features and semantic features built with Relevant Domains based on words and RelDomS is Basic features and semantic features built with Relevant Domain based on word senses)

Table 1. F1 value over test corpora

Corpus	Basic	WND	BLC20	RelDomW	RelDomS	Baseline	Best
SV2	0.667	0.665	0.658	0.627	0.626	0.570	0.685
SV3	0.641	0.644	0.642	0.637	0.638	0.620	0.652
SEM1	0.511	0.511	0.543	0.511	0.521	0.514	0.591

It will first be observed that in three cases with different test corpus, the most frequent baselines are exceeded. As stated previously, these kinds of baselines are usually very hard to attain in WSD tasks. Moreover, despite the fact that our classifiers do not attain the result of the best participating systems in each task (SE2, SE3 and SEM1), in some cases our classifiers show a similar performance to the best system. In general, the results when semantic features are used are higher than if no semantic information is included, except in the case of SE2, in which the best result is obtained by the *Basic* experiment. This could be explained if we look at the domain of the evaluation set from SE2, which was built from three different sources, a mystery novel, a medical report and a paper on children education.

Furthermore, upon comparing *RelDomW* and *RelDomS* it will be noted that the performance is similar. This indicates that the statistic information provided by words, without considering senses, is sufficient to model the domains of a certain context.

Table 2 shows the performance of the system when pairs of semantic classes are combined to generate the semantic features. The aim of these experiments was to combine different sources of semantic information, in order to analyze whether they can provide complementary information.

Table 2. F1 value over test corpora

Corpus	Basic	BLC20+WND	BLC20+RelDomS	WND+SS	Baseline	Best
SV2	0.667	0.658	0.659	0.666	0.570	0.685
SV3	0.641	0.642	0.642	0.641	0.620	0.652
SEM1	0.511	0.541	0.543	0.519	0.514	0.591

On the one hand, we can see that the baselines are again exceeded. On the other hand, the results of these experiments when combining two semantic classes are worse than when only one class is used to build the semantic features. In general, it would appear that the information provided by semantic features can help up to a certain point, and that no combination of semantic classes improves these results. The information provided by different semantic classes therefore appears not to be complementary, and there is a kind of overlapping between the semantic features according to different semantic classes. This is normal if we consider that the main source from which all the semantic classes were derived was the same: WordNet.

5 Conclusions and Discussion

In this work we have presented a supervised approach to WSD using semantic classes rather than word senses. The goal of our proposal is to avoid the problem of the fine granularity of sense inventories. We propose to solve semantic ambiguities and to maintain semantic coherence by grouping together those concepts that share certain properties. This proposal has been evaluated in the SensEval framework and compared with the results of the best systems. According to the results obtained, we have demonstrated that semantic information extracted from resources based on WordNet leads to an improvement in the performance of our WSD system. However, there is a critical point at which the results are not improved and remain steady. One of the reasons for this behaviour is that the information added is redundant, despite the fact that it comes from different sources.

After the evaluation analysis and with regard to the results obtained, we can observe that the performance with the use of semantic attributes works better or worse depending on the evaluation corpora utilized. For instance, the use of WND as semantic attributes provides the best results in SensEval-3. However, BLC classes work better in SemEval. Again, we can observe the influence of domain specific corpora during the WSD process. Related to this question, although the resources should be adapted to a specific domain to improve the results, it is important to highlight that in several cases, the results obtained using semantic attributes built with an automatic process (BLC and Relevant Domains) work similarly or even better than manual attributes (WND). This fact supposes an advantage, since these kinds of attributes could be generated automatically and adapted to a new domain. On the other hand, the manual development of a resource for this purpose would be time consuming. Furthermore, our results are very close to the best systems in each competition, and in all cases exceed the baseline based on the most frequent sense.

References

1. Agirre, E., Calle, O.L.D.L.: Clustering Wordnet word senses. In: Proceedings of RANLP ’03. pp. 121–130. Borovets, Bulgaria (2003)
2. Ciaramita, M., Altun, Y.: 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). pp. 594–602. ACL, Sydney, Australia (2006)
3. Fellbaum, C. (ed.): WordNet. An Electronic Lexical Database. The MIT Press (1998)
4. Gliozzo, A.M., Strapparava, C., Dagan, I.: Unsupervised and supervised exploitation of semantic domains in lexical disambiguation. Computer Speech & Language 18(3), 275–299 (2004)

5. Izquierdo, R., Suárez, A., Rigau, G.: An empirical study on class-based word sense disambiguation. In: Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics. pp. 389–397. EACL '09, Association for Computational Linguistics, Stroudsburg, PA, USA (2009)
6. Magnini, B., Cavaglià, G.: Integrating subject field codes into Wordnet. In: Proceedings of LREC. Athens. Greece (2000)
7. Magnini, B., Pezzulo, G., Gliozzo, A.: The role of domain information in word sense disambiguation. Natural Language Engineering 8, 359–373 (2002)
8. Màrquez, L., Escudero, G., Martínez, D., Rigau, G.: Supervised corpus-based methods for WSD. In: E. Agirre and P. Edmonds (Eds.) Word Sense Disambiguation: Algorithms and applications. Text, Speech and Language Technology, vol. 33. Springer (2006)
9. Mccarthy, D.: Relating Wordnet senses for word sense disambiguation. In: In Proceedings of the ACL Workshop on Making Sense of Sense. pp. 17–24 (2006)
10. Miller, G., Leacock, C., Tengi, R., Bunker, R.: A Semantic Concordance. In: Proceedings of the ARPA Workshop on Human Language Technology pp. 303–308. (1993)
11. Navigli, R.: 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. pp. 105–112. Association for Computational Linguistics, Morristown, NJ, USA (2006)
12. Snow, R., et al.: Learning to merge word senses. In: Proceedings of Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL). pp. 1005–1014 (2007)
13. Vázquez, S., Montoyo, A., Rigau., G.: Método de desambiguación lingüistica basada en el recurso semántico dominios relevantes. In: Proceedings of the XIX Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN '03). pp. 141–149. Alcalá de Henares, Spain (2003)
14. Villarejo, L., Màrquez, L., Rigau, G.: 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. pp. 195–202. Granada, Spain (September 2005), ISSN 1136-5948
15. Weaver, W.: Translation. In: Locke, W.N., Booth, A.D. (eds.) Machine Translation of Languages, pp. 15–23. MIT Press (1957)
16. Wilks, Y., Stevenson, M.: The grammar of sense: Is word sense tagging much more than part-of-speech tagging? Tech. rep., CS-96-05, University of Sheffield. UK (1996)