UPV-WSD : Combining different WSD Methods by means of Fuzzy Borda Voting

Davide Buscaldi and **Paolo Rosso**
DSIC, Dpto. Sistemas Informáticos y Computación
Universidad Politécnica de Valencia
Valencia, Spain
{dbuscaldi,prosso}@dsic.upv.es

Abstract

This paper describes the WSD system developed for our participation to the SemEval-1. It combines various methods by means of a fuzzy Borda voting. The fuzzy Borda vote-counting scheme is one of the best known methods in the field of collective decision making. In our system the different disambiguation methods are considered as experts that give a preference ranking for the senses a word can be assigned. Then the preferences are evaluated using the fuzzy Borda scheme in order to select the best sense. The methods we considered are the sense frequency probability calculated over SemCor, the Conceptual Density calculated over both hyperonyms and meronyms hyerarchies in WordNet, the extended Lesk by Banerjee and Pedersen, and finally a method based on WordNet domains.

1 Introduction

One of the lessons learned from our previous experience at Senseval-3[1] (Buscaldi et al., 2004; Vazquez et al., 2004) is that the integration of different systems usually works better than a standalone system. In our opinion this reflects the reality where humans do not apply always the same rule in order to disambiguate the same ambigue word; for instance, if we consider the sentences "*He hit a home run*" and "*The thermometer hit 100 degrees*", in the first case the sport domain helps in determining the right sense for *hit*, whereas in the latter the disambiguation is carried out mostly depending on the fact that the subject of the sentence is an object.

The combination of distinct methods represents itself a major problem. If the methods return different answers, how can we select the best one? In this sense the available choices are the following:

- *Rule-based* selection: a set of rules that can be both hand-made or automatically learned from examples;
- *Probability-based*: the output of the methods is normalized in the range $[0, 1]$ and is considered as a probability. Then the values are multiplied in order to obtain the sense with a maximum probability.
- *Vote-based*: the output of the methods is considered as a weighted vote. Then a voting scheme is used in order to obtain the most voted sense.

In our previous participation with the R2D2 project (Vazquez et al., 2004) the selection was rule-based, with hand-made rules that attempted to take into account the reliability of the various method. We subsequently attempted to learn automatically the rules, but the results of these experiments did not allow to determine clearly which method was to be used in each context.

Working with probabilities can be problematic due to the null probabilities that make necessary the adoption of smoothing techniques. Therefore, we opted for a voting scheme, in this case the fuzzy Borda (Nurmi, 2001; García Lapresta and Martínez

[1] http://www.senseval.org

Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007), pages 434–437,
Prague, June 2007. ©2007 Association for Computational Linguistics

Panero, 2002), one of the best known methods in the field of collective decision making. With this scheme the disambiguation methods are considered as experts providing a preference ranking over the sense of the word.

The methods we choose as experts are the sense probability calculated over SemCor, the Conceptual Density algorithm by (Rosso et al., 2003), the extended Lesk by (Banerjee and Pedersen, 2002), and an algorithm that takes into account the domains of the word to be disambiguated and the context words. In the following sections we describe in detail the fuzzy Borda scheme and each WSD expert.

2 The Fuzzy Borda voting scheme

The original Borda vote-counting scheme was introduced in 1770 by Jean Charles de Borda, and adopted by the French Academy of Sciences with the purpose of selecting its members. In the classical Borda count each expert gives a mark to each alternative, according to the number of alternatives worse than it. The fuzzy variant (Nurmi, 2001; García Lapresta and Martínez Panero, 2002) is a natural extension that allows the experts to show numerically how much some alternatives are preferred to the others, evaluating their preference intensities from 0 to 1.

Let $R^1, R^2, \ldots, R^m$ be the fuzzy preference relations of m experts over n alternatives $x_1, x_2, \ldots, x_n$. For each expert k we obtain a matrix of preference intensities:

$$\begin{pmatrix} r_{11}^k & r_{12}^k & \cdots & r_{1n}^k \\ r_{21}^k & r_{22}^k & \cdots & r_{2n}^k \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1}^k & r_{n2}^k & \cdots & r_{nn}^k \end{pmatrix}$$

where each $r_{ij}^k = \mu_{R^k}(x_i, x_j)$, with $\mu_{R^k} : X \times X \to [0,1]$ being the membership function of R^k. The number $r_{ij}^k \in [0,1]$ is considered as the degree of confidence with which the expert k prefers x_i to x_j. The final value assigned by the expert k to each alternative x_i is:

$$r_k(x_i) = \sum_{j=1, r_{ij}^k > 0.5}^{n} r_{ij}^k \qquad (1)$$

which coincides with the sum of the entries greater than 0.5 in the i-th row in the preference matrix. The threshold 0.5 ensure the relation $R^k$to be an ordinary preference relation (García Lapresta and Martínez Panero, 2002).

Therefore, the definitive fuzzy Borda count for an alternative x_i is obtained as the sum of the values assigned by each expert:

$$\mathbf{r}(x_i) = \sum_{k=1}^{m} r_k(x_i) \qquad (2)$$

In order to fill the preference matrix with the correct confidence values, the output weights $w_1, w_2, \ldots, w_n$ of each expert k are transformed to fuzzy confidence values by means of the following transformation:

$$r_{ij}^k = \frac{w_i}{w_i + w_j} \qquad (3)$$

An example of how fuzzy Borda is used to combine the votes in order to obtain the right sense of the target word is shown in Section 4.

3 WSD Experts

We considered five experts in order to carry out the disambiguation process. Sense probability and the extended lesk were available for every word, while the Conceptual Density was calculated only for nouns. Therefore, all the experts were available only for the nouns. For each expert different contexts were taken into account, depending on the specific characteristics of each expert.

3.1 Sense Probability

This expert is the simplest one: its votes are calculated using only the frequency count in SemCor of the WordNet senses of the word. The transformation of the frequency counts to the preference ranking is done according to Formula (3). Zero frequency are normalized to 1.

3.2 Conceptual Density

Conceptual Density (CD) was originally introduced by (Agirre and Rigau, 1996). It is computed on WordNet subhierarchies, determined by the *hypernymy* (or *is-a*) relationship. Our formulation (Rosso et al., 2003) of the Conceptual Density of a WordNet subhierarchy s is:

$$CD(m, f, n) = m^{\alpha} \left(\frac{m}{n} \right) \qquad (4)$$

Where m are the *relevant* synsets in the subhierarchy, n is the total number of synsets in the subhierarchy.The relevant synsets are both the synsets of the word to be disambiguated and those of the context words.

The WSD system based on this formula participated at the Senseval-3 competition as the CIAOSENSO system (Buscaldi et al., 2004), obtaining 75.3% in precision over nouns in the all-words task (baseline: 70.1%). These results were obtained with a context window of two nouns, the one preceding and the one following the word. In Senseval-3 the WSD system took also into account the frequency of senses depending on their rank. In SemEval-1 we do not, because of the presence of the Sense Probability expert.

The CD-based expert uses a context of two nouns for the disambiguation process too. The weights from Formula (4) are used for computing the fuzzy confidence values that are used to fill the preference matrix after they are transformed according to Formula (3).

A second CD-based expert exploits the *holonymy*, or *part-of* relationship instead of *hyperonymy*. This expert uses as context all the nouns in the sentence of the word to be disambiguated.

3.3 Extended Lesk

This expert is based on the algorithm by (Banerjee and Pedersen, 2002), a WordNet-enhanced version of the well-known dictionary-based algorithm proposed by (Lesk, 1986). The original Lesk was based on the comparison of the gloss of the word to be disambiguated with the context words and their glosses. This enhancement consists in taking into account also the glosses of concepts related to the word to be disambiguated by means of various WordNet relationships. Then similarity between a sense of the word and the context is calculated by means of *overlaps*. The word is assigned the sense obtaining the best overlap match with the glosses of the context words and their related synsets.

The weights used as input for Formula (3) are the similarity values between the senses of the world and the context words. The context for this expert consists of 4 WordNet words (disregarding their Part-Of-Speech) located in the same sentence of the word to be disambiguated, i.e., words with POS noun, verb, adjective or adverb that can be found in WordNet.

3.4 WordNet Domains

This expert uses WordNet Domains (Magnini and Cavaglià, 2000) in order to provide the system with domain-awareness. All WordNet words in the same sentence of the target word are used as context. The weight for each sense is obtained by counting the number of times the same domain of the sense appears in the context (all senses of context words are considered). We decided to not take into account the "factotum" domain.

4 Example

In this example we will consider only the sense probability and extended Lesk experts for simplicity.

Let us consider the following phrase: "*And he has kept mum on how his decision might affect a bid for United Airlines , which includes a big stake by British Airways PLC.*" with *affect* as target word. We can observe that in WordNet the verb *affect* has 5 senses. The sense count values are 43 for the first sense, 11 for the second, 4 for both the third and the fourth one, and 0 for the last one. We decided to normalize the cases with 0 occurrences to 1. After applying the transformation (3) to the sense counts, we obtain the following preference matrix for the sense probability expert:

$$\begin{pmatrix} 0.5 & 0.80 & 0.91 & 0.91 & 0.98 \\ 0.20 & 0.5 & 0.73 & 0.73 & 0.92 \\ 0.09 & 0.27 & 0.5 & 0.5 & 0.8 \\ 0.09 & 0.27 & 0.5 & 0.5 & 0.8 \\ 0.02 & 0.08 & 0.2 & 0.2 & 0.5 \end{pmatrix}$$

Therefore, the final fuzzy Borda counts by the sense probability expert are 3.60 for `affect(1)`, 2.38 for `affect(2)`, 0.8 for `affect(3)` and `affect(4)`, and 0 for `affect(5)`, obtained from the sum of the rows where the value is greater than 0.5.

The extended Lesk expert calculates the following similarity scores for thesenses of *affect*, with context words *decision*, *might*, *bid* and *include*: respectively 107, 70, 35, 63 and 71 for senses 1 to 5. After applying the transformation (3) to the weights, we obtain

the preference matrix for this expert:

$$\begin{pmatrix} 0.5 & 0.60 & 0.75 & 0.63 & 0.60 \\ 0.40 & 0.5 & 0.67 & 0.53 & 0.49 \\ 0.25 & 0.33 & 0.5 & 0.36 & 0.33 \\ 0.37 & 0.47 & 0.64 & 0.5 & 0.47 \\ 0.40 & 0.51 & 0.67 & 0.53 & 0.5 \end{pmatrix}$$

In this case the final fuzzy Borda counts are 2.58 for the first sense, 1.2 for sense 2, 0 for sense 3, 0.64 and 1.71 for senses 4 and 5 respectively.

Finally, the sum of Borda counts of every expert for each sense (see Table 4) are used to disambiguate the word.

sense no:	1	2	3	4	5
expert 1	3.60	2.38	0.80	0.80	0
expert 2	2.58	1.20	0	0.64	1.71
total:	**6.18**	3.58	0.80	1.44	1.71

Table 1: Borda Count for the verb *affect* in the example phrase.

5 Results

The system was not tested before SemEval. Our participation was limited to the All-Word and Coarse-Grained tasks (without the sense inventory provided by the organizers). The results are compared to the best system and the MFS (Most Frequent Sense) baseline. We calculated also the partial results over nouns in the all word task, obtaining that the MFS baseline in this case is about 0.633, whereas our system obtains 0.520.

task	upv-wsd	MFS	best system
coarse-grained	0.786	0.789	0.832
awt	0.420	0.471	0.537

Table 2: Recall obtained by our system (upv-wsd) in each task we participated in, compared with the most frequent sense baseline and the best system in the task.

6 Conclusions

The combination of different systems allowed us to attain higher recall than with our previous system used in Senseval-3. However, overall results were not as good as expected. Partial results over the nouns show that the CD expert did not perform as in the Senseval-3 and that the CD formula needs to include sense frequency ranking in order to achieve a good performance. As a further work we plan to add a weight reflecting the reliability of each expert.

Acknowledgements

We would like to thank the TIN2006-15265-C06-04 research project for partially supporting this work. We would also like to thank Prof. Eugene Levner of the Holon Institute of Technology for inspiring us to use the fuzzy Borda voting scheme.

References

Eneko Agirre and German Rigau. 1996. Word sense disambiguation using conceptual density. In *COLING*, pages 16–22.

Satanjeev Banerjee and Ted Pedersen. 2002. An adapted lesk algorithm for word sense disambiguation using wordnet. In *Proceedings of CICLing 2002*, pages 136–145, London, UK. Springer-Verlag.

Davide Buscaldi, Paolo Rosso, and Francesco Masulli. 2004. The upv-unige-CIAOSENSO WSD System. In *Proc. of Senseval-3 Workshop*, Barcelona (Spain), July. ACL.

José Luis García Lapresta and Miguel Martínez Panero. 2002. Borda Count Versus Approval Voting: A Fuzzy Approach. *Public Choice*, 112(1-2):167–184.

Michael Lesk. 1986. Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. In *Proc. of SIGDOC '86*, pages 24–26.

Bernardo Magnini and Gabriela Cavaglià. 2000. Integrating Subject Field Codes into WordNet. In *Proc. of the 2nd LREC Conference*, pages 1413–1418, Athens, Greece.

Hannu Nurmi. 2001. Resolving Group Choice Paradoxes Using Probabilistic and Fuzzy Concepts. *Group Decision and Negotiation*, 10(2):177–199.

Paolo Rosso, Francesco Masulli, Davide Buscaldi, Ferran Pla, and Antonio Molina. 2003. Automatic noun sense disambiguation. In *Proc. of CICLing 2003*, pages 273–276.

Sonia Vazquez, Rafael Romero, Armando Suarez, Andres Montoyo, Manuel García, M. Teresa Martin, M. Angel García, Alfonso Ureña, Davide Buscaldi, Paolo Rosso, Antonio Molina, Ferran Pla, and Encarna Segarra. 2004. The R2D2 Team at SENSEVAL-3. In *Proc. of Senseval-3 Workshop*.