Borda-Based Voting Schemes for Semantic Role Labeling

Vladimir Robles[1,2], Antonio Molina[2], and Paolo Rosso[2]

[1] Universidad Politécnica Salesiana, Cuenca, Calle Vieja 12-30, Ecuador
vrobles@ups.edu.ec
[2] ELiRF, DSIC, Universidad Politécnica de Valencia,
Valencia 46022, Camino de Vera s/n, Spain
{amolina,prosso}@dsic.upv.es

Abstract. In this article, we have studied the possibility of applying Borda and Fuzzy Borda voting schemes to combine semantic role labeling systems. To better select the correct semantic role, among those provided by different experts, we have introduced two measures: the first one calculates the overlap between labeled sentences, whereas the second one adds different scoring levels depending on the verbs that have been parsed.

Key words: Semantic role labeling, Borda voting schemes.

1 Introduction

Previous studies shown that the semantic role labeling is a task that allows to improve the performance of many Natural Language Processing (NLP) applications. A semantic role is the underlying relationship between a syntactic constituent (consisting of a word or sequence of words) and the main verb of a sentence. The role is the function that assigns the predicate to its arguments. A clear example of what has been mentioned is shown in the following sentence: *"Hurricane-force winds demolished much of the town"*. If we review the sentence, it would have the following roles: *[Hurricane-force winds]*$_{\text{cause}}$ demolished *[much of the town]*$_{\text{theme}}$. The syntactic constituent *"Hurricane-force winds"* is the cause that leads to a certain event, while *"much of the town"* constituent represents the argument that undergoes a change of state. The main thematic roles are: agent (argument that produces the action), experiencer (argument that is subjected to a sensory, cognitive or emotional experience), container (argument that is good or bad in a situation), location (argument representing sites), action (argument expressing some dimension) and item (argument that undergoes a change of state).

The task of semantic role labeling has been studied from several machine learning approaches, including the use of probabilistic and statistical techniques, such as Maximum Entropy or Conditional Random Fields and methodologies based on artificial intelligence such as Support Vector Machines. These methodologies depend on choosing the relevant characteristics, representing information of various kinds: lexical, syntactic and probabilistic, among other types [1].

In this paper we review the possibility of applying Borda and Fuzzy Borda voting schemes [2], to determine the feasibility of combining various systems of semantic

Petr Sojka, Aleš Horák, Ivan Kopeček and Karel Pala (Eds.): TSD 2010, LNAI 6231, pp. 185–192, 2010.

role labeling. To accomplish this task we have worked with the data set published in the shared task of the conference CoNLL 2005 (Conference on Computational Natural Language Learning)[3]. We worked with the corpus tagged by the 5 best systems. We defined two measures of analysis: the level of role overlapping and the role scoring tables contained in each sentence.

The rest of the paper is organized as follows. In Section 2 we review the features of the used corpus. The Borda voting scheme and its variant Fuzzy, and the possibility of using it to combine two or more role labeling systems are described in Section 3. In Section 4 we review the steps we used to combine the results generated by the CoNLL 2005 systems. In Section 5 we show the results and analyze them. Conclusion and future work are described in Section 6.

2 Corpus CoNLL 2005

In CoNLL 2005, the corpus used is based on Section 02 - 21 (training), Section 24 (development) and Section 23 (test) of the Wall Street Journal (WSJ). More precisely, the corpus is based on PropBank 1.0, which is a part of the Penn Treebank with enriched structures (predicate and argument). The corpus has different type of arguments, (i.e., Semantic Roles), Numbered Arguments (A0-A5, AA), Adjuntcs (AM-), References (R-), and Verbs (V) [3].

In Table 1 we can see a list of the characteristics of the 5 best systems. The systems are ordered by F-Measure. The table lists the name of participation of each system, as well as precision, recall and F-Measure.

Table 1. The best five systems from the CoNLL 2005 competition

System	Short Name	Precision	Recall	F-Measure
punyakanok	S_1	82.28%	76.78%	79.44
pradhan	S_2	82.95%	74.75%	78.63
haghighi	S_3	79.54%	77.39%	78.45
marquez	S_4	79.55%	76.45%	77.97
surdeanu	S_5	80.32%	72.95%	76.46

3 Borda Voting Schemes

The Borda voting schemes is a technique that has been used in several NLP tasks: word sense disambiguation [5], geographical information retrieval [4], named entity recognition [6]. In this context, we consider that this methodology can improve the performance of semantic role labeling by combining different systems. For example, in this sentence of the tWSJ corpus: *"As a result, the link between the futures and stock markets ripped apart."*, the best CoNLL 2005 three labeling systems produce the following results (Table 2):

[3] `http://www.lsi.upc.edu/~srlconll/st05/st05.html`

Table 2. Comparison of labeling process performed by the systems S_1, S_2 and S_3

Constituent	S_1	S_2	S_3
As	(AM-CAU*	(AM-CAU*	(AM-DIS*
a	*	*	*
result	*)	*)	*)
,	*	*	*
the	(A1*	(A1*	(A1*
link	*	*	*
between	*	*	*
the	*	*	*
futures	*	*	*
and	*	*	*
stock	*	*	*
markets	*)	*)	*)
ripped	(V*)	(V*)	(V*)
apart	(AM-DIR*)	(AM-MNR*)	(AM-DIR*)
.	*	*	*

If we want to apply a Borda voting scheme, each system should provide a determined amount of candidate roles for each sentence argument. In the example described in Table 3, the role AM-CAU is assigned to the argument *"As a result"*[4]. This argument must have been assigned to two or more candidate roles by each system. This allows the creation of the necessary matrices to apply the Borda voting scheme.

We calculate the general voting results considering role AM-CAU as candidate1, AM-LOC as candidate2 and AM-DIS as candidate3. For example, to calculate M_{S_1}, we fill with 1 in row 1 and column 2 which indicates that the system prefers candidate1 than candidate2. Doing so for all candidates and by filling 0 in the rest of positions, we obtain the matrix. The final vote is the sum of the rows of systems matrices.

$$M_{S_1} = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad M_{S_2} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad M_{S_3} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix} \quad Final_{Vote} = \begin{bmatrix} 3 \\ 4 \\ 2 \end{bmatrix}$$

Table 3 shows the preference candidate order for each system and the general order of the Borda voting scheme.

Table 3. Order of preference after applying Borda voting scheme

S_1	S_2	S_3	Borda Preference
AM-CAU	AM-LOC	AM-DIS	**AM-LOC**
AM-LOC	AM-CAU	AM-LOC	**AM-CAU**
AM-DIS	AM-DIS	AM-CAU	**AM-DIS**

[4] To better illustrate our example, we add two candidate roles and we change the preference order for every role in the system S_2.

To apply a Fuzzy Borda voting scheme, we must add weights for each candidate role, as is shown in Table 4.

Table 4. Preference order for roles labeled by each system, using weights

S_1	S_2	S_3
AM-CAU: **8.2**	AM-LOC: **7.3**	AM-DIS: **9.2**
AM-LOC: **7.2**	AM-CAU: **5.2**	AM-LOC: **3.2**
AM-DIS: **6.7**	AM-DIS: **4.7**	AM-CAU: **2.7**

According to the Fuzzy Borda voting scheme, the element $r^i_{j,k}$ (row j, column k of the matrix M_{S_i} for the role labelling systems S_i) can be calculated using the following formula:

$$r^i_{j,k} = \frac{w^i_j}{w^i_j + w^i_k} \tag{1}$$

Using Formula 1, and the weights from Table 4, we calculate the preference matrix of Fuzzy Borda voting scheme:

$$M_{S_1} = \begin{bmatrix} 0.5 & 0.53 & 0.55 \\ 0.47 & 0.5 & 0.52 \\ 0.45 & 0.48 & 0.5 \end{bmatrix} \quad M_{S_2} = \begin{bmatrix} 0.5 & 0.42 & 0.53 \\ 0.58 & 0.5 & 0.6 \\ 0.47 & 0.39 & 0.5 \end{bmatrix} \quad M_{S_3} = \begin{bmatrix} 0.5 & 0.46 & 0.23 \\ 0.54 & 0.5 & 0.26 \\ 0.77 & 0.74 & 0.5 \end{bmatrix}$$

$$Final_{Vote} = \begin{bmatrix} 4.2 \\ 4.5 \\ 4.8 \end{bmatrix}$$

The resulting preference order, from the Fuzzy Borda scheme, is shown in Table 5.

Table 5. Preference order after applying Fuzzy Borda voting scheme

S_1	S_2	S_3	Preference Order
AM-CAU	AM-LOC	AM-DIS	**AM-DIS**
AM-LOC	AM-CAU	AM-LOC	**AM-LOC**
AM-DIS	AM-DIS	AM-CAU	**AM-CAU**

As we have seen, if we do not have the number of candidates or alternatives required by the Borda voting schemes, we can not apply them. We consider the following observations:

- In order to create the Borda matrix, each system must label roles as part of a single domain. All systems must assign the same candidate roles for each argument, ordered according to their preference or weights.
- The verb and its meaning are the parameters that help us to define candidate roles, which create the Borda matrix. Weights must be inferred from the level of precision, recall or F-Measure of a system.

4 Overlapping and Scored Verb Analysis

4.1 Overlapping

The level of overlap is a measure that allows us to analyze the level of matching between two or more role labeling systems. Therefore, high value of overlapping indicates that the criteria of these systems is closer. This allows us to select those two systems that have the greatest value of matching. The system that has the highest score of verb analysis is selected.

To illustrate how we calculate the overlapping we took the sentence "*'It screwed things up,' said one major specialist.*" from tWSJ corpus and the roles proposed by the systems S_1 and S_2 (Table 6).

Table 6. Sentence of corpus tWSJ labeled by the systems S_1 and S_2

	S_1		S_2	
Constituent	Verb screwed	Verb said	Verb screwed	Verb said
"	*	*	*	*
It	*	(A1*	(A0*)	*
screwed	(V*)	*	(V*)	(A1*
things	(A1*)	*	(A1*)	*
up	*	*)	*	*)
,	*	*	*	*
"	*	*	*	*
said	*	(V*)	*	(V*)
one	*	(A0*	*	(A0*
major	*	*	*	*
specialist	*	*)	*	*)
.	*	*	*	*
Verb Score	0.8225	0.8523	0.8	0.8

As shown in Table 6, for the analysis of the verb *"screwed"*, the system S_2 assigns A0 role to the constituent *"It"*, while the system S_1 does not. For the constituent *"things"*, both systems agree to assign A1 as role. In this case, there is an overlapping in a single argument. For the verb *"said"* there is a partial overlapping in A1 role, because for the system S_1 the argument is made up of *"It screwed things up"* constituents, whereas the system S_2 is made up of *"screwed things up"*. For the A0 role both systems assign the same constituents. In this case, there is a partial overlap (A1) and a full overlap (A0).

To calculate the overlaps that occur in arguments consisting of a single constituent, we assign a value of 1 and we add the other arguments that have a single constituent. For the verb *"screwed"*, the overlapping value 1.

In the case of partial overlapping, we consider how many overlapping constituents of an argument (CS) and how many constituents make up that argument (CF). To calculate this value we have derived the following formula:

$$Overlap = \sum_{1}^{N} \left(\frac{CS_{S1}}{CF_{S1}} \right) \cdot \left(\frac{CS_{S2}}{CF_{S2}} \right) . \tag{2}$$

Where:

- CS_{S1} and CS_{S2} represent the constituents that overlap in the labeled argument by the systems S_1 and S_2.
- CF_{S1} and CF_{S2} represent the constituents that make up the argument of the systems S_1 and S_2.
- N is the total number of roles in the sentence.

The level of overlapping for the verb *"said"* is calculated as follows:

$$Overlap = overlap_{A0-Role} + overlap_{A1-Role} .$$

$$Overlap = \left[\left(\frac{3}{3} \cdot \frac{3}{3} \right) + \left(\frac{3}{4} \cdot \frac{3}{3} \right) \right] = 1.75 .$$

4.2 Scored Verb Analysis

To calculate this value we use a scoring system for each labeled role in a sentence. The basis of this metric is the overall level of precision of each system for role labeling (recall and F-Measure could also be used). For example, the system S_1 labels A0 roles with a precision of 88.22%, recall of 87.88% and an F-Measure of 88.05.

Experiments were carried out using the precision values. By combining two or more role labeling systems, we are expanding the coverage level that the system has. To calculate the verb scores (Table 6), we obtain an average value of the precision that each system has to label the arguments of a specific verb. For example, in the system S_2, the score of the verb *"screwed"* is calculated as follows: [0.8(precision of labeling the role A0)+0.8 (precision of labeling the role A1)]/2=0.8.

For the verb *"screwed"*, the system S_2 has labeled two roles, while the system S_1 has labeled a single role. Our system selects the labeled roles of systems S_2 and S_1 for verbs *"screwed"* and *"said"*, respectively.

5 Experimental Results

In this section we show the main results that were obtained after applying our voting approach. We have used two schemes of overlapping and scoring, partial and complete overlapping. The first scheme does not discard the arguments that do not completely overlap. The second scheme discards those that do not have a complete overlapping (all its constituents, similar to a simple voting scheme). In Figure 1 we see the number of arguments correctly classified by each of the combinations that we have tested. The best result is achieved combining all the systems with partial overlapping. Figure 1 shows the roles that were misclassified and also those that the system was not able to classify. The

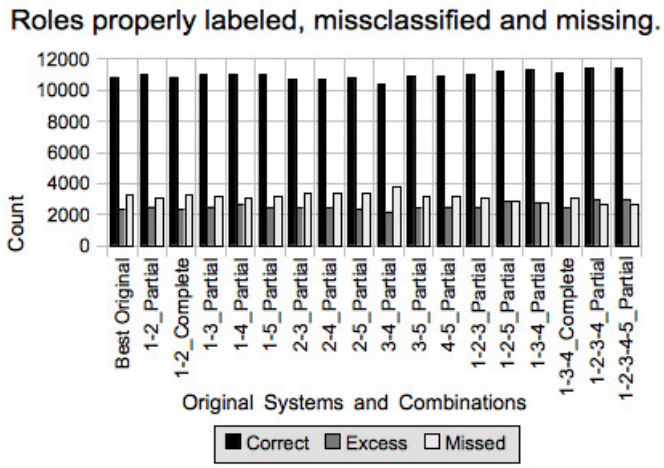

Fig. 1. Roles properly labeled, misclassified and missing

combination that produces the best results, considering the two values together (roles misclassified and non-tagged), is S_1-S_3-S_4 with partial overlapping.

In Figure 2 we observe precision, recall and F-Measure for all system combinations. The combination that gets the best F-Measure value is S_1-S_3-S_4 with partial overlapping. The precision is affected by the number of systems involved in the combination, because not all the systems have optimal values of this measure.

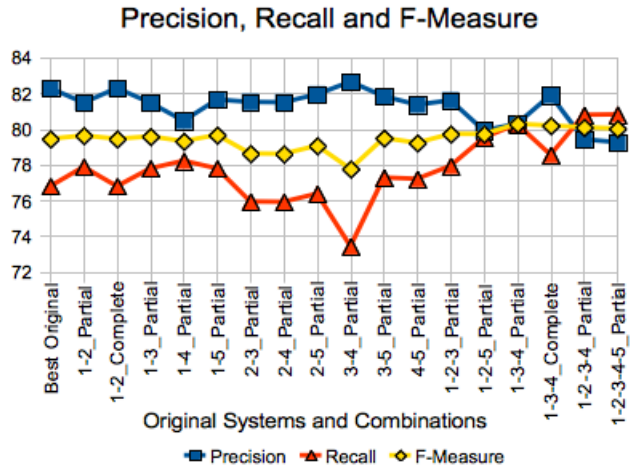

Fig. 2. Precision, recall and F-Measure of all system combinations

6 Conclusions

In this paper we have established an alternative measure of combinations between labeling systems, based on the Borda voting schemes. It has been shown that combining two or more systems together, better results can be achieved.

When we combine too many labeling systems, the precision become lower if these systems don not have similar values of precision. By contrast, the level of recall is enriched by the diversity of labeling schemes. One factor that improves the measurement of overlapping and especially the scored verb analysis, is to review the arguments that must have each verb. The implementation of this factor will help decrease the amount of roles that are misclassified or ignored.

As future work we propose to test scored verbs based on their level of matching with the arguments in PropBank and FrameNet, to apply Linear Integer Programming techniques which enrich the measurement process of overlapping and scored verb analysis, and include in the calculation of overlapping the values of precision, recall and F-Measure and verify their efficiency.

Acknowledgements

This work was partially funded by the TEXT-ENTERPRISE 2.0 TIN2009-13391-C04-03 and Telmosis PAID-06-08-3294 research projects.

References

1. Gildea, D., Jurafsky, D.: Automatic Labeling of Semantic Roles. Computational Linguistics, vol. 28, pp. 245–288 (2002)
2. García-Lapresta, J.L., Martínez-Panero, M.: Borda Count versus Approval Voting: A Fuzzy Approach. Public Choice, vol. 112, pp. 167–184 (2002)
3. Carreras, X. Màrquez, L.: Introduction to the CoNLL-2005 Shared Task: Semantic Role Labeling. In: Proceedings of the Ninth Conference on Computational Natural Language Learning, pp. 152–164 (2005)
4. Buscaldi, D., Perea, J.M., Rosso, P. Ureña, L.A., Ferrés, D. Rodríguez, H.: GeoTexMess: Result Fusion with Fuzzy Borda Ranking in Geographical Information Retrieval. In: CLEF, pp. 867–874. Revised Selected Papers (2008)
5. Buscaldi, D., Rosso, P.: UPV-WSD: Combining Different WSD Methods by Means of Fuzzy Borda Voting. In: Fourth International Workshop on Semantic Evaluations, pp. 434–437 (2007)
6. Benajiba, Y., Diab, M., Rosso, P.: Arabic Named Entity Recognition using Optimized Features Sets. EMNLP, Hawaii, USA (2008)