SINAI at RTE-6: Personalized Page Rank Vectors and Named Entities

Arturo Montejo-Ráez, José Manuel Perea-Ortega,
Fernando Martínez-Santiago, L. Alfonso Ureña-López

Department of Computer Science
University of Jaén
23071 Jaén, Spain

February 17, 2011

Abstract

This document describes the participation of the SINAI Research Group in the 6^{th} challenge on Recognition of Textual Entailment (RTE). Our approach is a supervised one, by applying a learning process on the development data. The model trained was a stacking of two SVM models, each one over different attributes. One of the SVMs was trained on the distance between Personalized Page Rank vectors, and the other on Named Entities matching counts. The results obtained are low, mainly due to the simplicity of the system. In any case, the use of stacking for merging different features results in a relevant increment in performance.

1 Introduction

The SINAI Research group has participated in the 6^{th} RTE challenge, organized as a workshop within the Text Analysis Conference in 2010 (TAC 2010). This document describes the system implemented for resolving the task of recognizing textual entailment. The approach followed consists of three main modules: two feature extraction modules and a machine learning algorithm. The features that model both main text and hypothesis are based on the distance between Personalized Page Rank vectors and different counts on the coincidence between named entities from both texts. The features served to train two Support Vector Machine models [4] that are used as input to a stacking process based in a Naïve Bayes learner. The results obtained are below the average scores in RTE-6 as reported by the organization, but the application of stacking proposes an interesting and worth to explore approach for mixing heterogeneous features in a supervised manner.

2 RTE-6 challenge

Recognizing textual entailment is a task that has attracted the attention of a large group of researchers in the area of Natural Language Processing during the last years. From 2008, the organizer of a related challenge has been the Information Technology Laboratory, at the National Institute of Standards and Technology, becoming RTE a track at the Text Analysis Conference[1]. This has brought the opportunity to take RTE challenge to more realistic and more application-oriented scenarios.

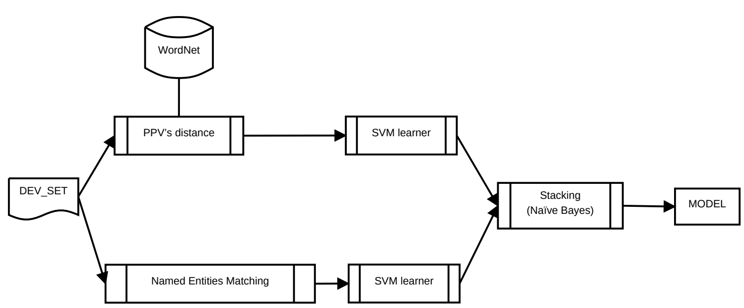
This year, two tasks have been proposed within the summarization scenario: the *Main task* and the *Novelty detection sub-task* task. We have participated in both. Also, ablation tests were required to participants in order to provide analysis of the different modules involved in the systems proposed, so a better understanding of the effect of each component to the final performance of a system.

3 System architecture

Our system follows a supervised approach, as we did in our participation in the Pascal RTE 2007 challenge [5], though this time the features are totally different. Two features per pair have been generated: distance between Personalized Page Rank vectors (PPVs) and counts on the matching of Named Entities between text and hypothesis.

Personalized Page Rank vectors consists in a ranked sequence of WordNet [2] synsets weighted according to a random walk algorithm. A similar approach has been used recently by [7] to compute text semantic similarity in RTE environments, and also as solution for word sense disambiguation [1]. We have used the UKB software from this last citation to generate the PPVs used in our system. Random walk algorithms are inspired originally by the Google PageRank algorithm [6].

Figure 1: Training processes



[1] http://www.nist.gov/tac

In figure 1, the learning process is unveiled. This approach is the same for the Main task and Novelty detection sub-task. This process can be summarized as follows:

1. First, for each *sentence* in the text, a vector of weighted nodes is computed. In our approach, when a text or a hypothesis contains more than one sentence, the weights of the nodes are averaged. Besides, only the 1,000 nodes with the highest score are kept (for dimensionality reduction). Therefore, the text and the hypothesis are processed so a vector of weighted synsets is generated for each of one by means of the PPV random walk algorithm.

2. Finally, the distance between both calculated PPVs is computed applying the cosine formula. That is the first feature.

3. In parallel, a simple named entities detection is performed. For named entities, just three values are computed:

    - Number of matched NE's, i.e. the number of named entities present in both the text and the hypothesis
    - Number of NE's present in the hypothesis but not in the text
    - Number of NE's present in the text but not in the hypothesis

4. The learning algorithm takes one feature, in the case of the PPVs distance, as input to the first SVM learner, and three integer values as input features to the second SVM learner. The results of both models are used in a Naïve Bayes stack model. That is, SVM outputs from both models are used as input for the probabilistic classifier. The Rapid Miner tool[2] was used to the machine learning phase.

As an example of a PPV vector, when processed, the following text: *"Overall, we're still having a hard time with it, mainly because we're not finding it in an early phase."* becomes the vector of weighted synsets: `[02190088-a:0.0016, 12613907-n:0.0004, 01680996-a:0.0002, 00745831-a:0.0002, ...]`

4 Experiments and results

Two runs were submitted for each task (Main and Novelty detection tasks): **SINAI1** and **SINAI2**. The main difference between both runs is in the WordNet version used. In the first run we used WordNet 1.7. For the second one we used WordNet 3.0. Table 1 illustrates the results of our submissions for the Main task based on the settings described above.

In the RTE-6 Main task, 18 teams submitted a total of 48 runs. Statistics over 48 runs ranked by micro-averaged F1 determine that the highest value was 0.4801, the lowest was 0.1160 and the median was 0.3372. Therefore, our best

[2] `http://rapid-i.com`

		Microaveraged			Macroaveraged		
Run	**WN**	**Precision**	**Recall**	**F1**	**Precision**	**Recall**	**F1**
SINAI1	1.7	23,4	24,76	24,06	25,66	26,40	24,67
SINAI2	3.0	23,27	30,69	26,47	25,07	32,07	26,74
SINAI1-abl1	1.7	4,51	44,34	8,19	4,43	44,62	8,03
SINAI1-abl2	1.7	15,14	17,99	16,44	14,31	18,41	15,74
SINAI2-abl1	3.0	4,51	44,34	8,19	4,43	44,62	8,03
SINAI2-abl2	3.0	3,34	44,02	6,22	3,37	44,61	6,22

Table 1: Main task results

result achieved with the *SINAI2* experiment (0.2647) is far from the highest F1 (ours is 81.38% lower) and from the median F1 (27.39% lower). Complete evaluation measures are detailed in Table 1, along with associated ablation tests.

The ablation study has consisted of applying independently the PPV learning algorithm and the Named Entity recognition/resolution in our RTE system. Both options have been applied to the two runs explained above (SINAI1 and SINAI2), generating four ablation tests results:

- **SINAI1_abl-1**: it only uses the PPV learning algorithm on WordNet 1.7
- **SINAI1_abl-2**: it only uses Named Entity recognition on WordNet 1.7
- **SINAI2_abl-1**: the same as SINAI1_abl-1 but using WordNet 3.0
- **SINAI2_abl-2**: the same as SINAI1_abl-2 but using WordNet 3.0

The improvement achieved when our system used the WordNet 3.0 is about 10% with regard to the use of WordNet 1.7. From these results, we may realize that the gain in recall is due to the larger size in vocabulary and relations in WordNet version 3.0 compared to 1.7 (see 2). Of course, there is an associated but small penalty in precision.

	1.7	**3.0**
# terms	136,972	147,306
# relations	150,896	252,392

Table 2: WordNet 1.7 and 3.0 comparison

Table 1 also shows the results of the ablation tests for the Main task described above. The best result is obtained with the SINAI1_abl-1 ablation test, which only uses the PPV learning algorithm on WordNet 1.7. The difference with the other ablation tests is important: 0.0825 points better than only using Named Entity recognition on WordNet 1.7 or 3.0 and 0.1022 points better than only using the PPV learning algorithm on WordNet 3.0.

Regarding the Novelty Detection task,the results on the main submissions along with ablation runs are summarized in Table4.The behavior observed is similar to that of the Main task: it is noticeable a strong synergy when both sets of features (PPVs distances and name entities matching) are considered under a stacking based learner. Although these results are far from other participants measurements, The effect of combining different features under stacking schemes seems promising.

		Micro-averaged			**Macro-averaged**		
Run	**WN**	**Precision**	**Recall**	**F1**	**Precision**	**Recall**	**F1**
SINAI1	1.7	20,93	22,41	21,64	21,32	23,55	20,83
SINAI2	3.0	20,72	23,79	22,15	22,08	25,89	22,24
SINAI1-abl1	1.7	4,84	35,41	8,51	4,70	34,69	8,17
SINAI1-abl2	1.7	3,98	53,67	7,41	3,94	56,45	7,30
SINAI2-abl1	3.0	3,40	49,52	6,36	3,23	49,00	6,02
SINAI2-abl2	3.0	3,98	53,67	7,41	3,94	56,45	7,30

Table 3: Novelty Detection task results

5 Conclusions and further work

A simple approach based on Personalized PageRank Vectors and Named Entities has been proposed. The combination of the cosine distance between the PPVs of text and hypothesis and the matching count between entities of both texts are used as features in a supervised learning model based on stacking. The results obtained are of low performance, but suggest us that both features have useful informational content. Furthermore, the proposed system profits from a significant combined effect due to the stacking approach.

This work is a preliminary step in the construction of a more advanced system whose core will be Semantic Roles [3]. Our idea is to represent the elements in a semantic role (agent, action, patient), as PPVs, and explore different distance measurements between these tuples.

6 Acknowledgements

This work has been partially supported by a grant from the Spanish Government, project TEXT-COOL 2.0 (TIN2009-13391-C04-02), the “Fondo Europeo de Desarrollo Regional” (FEDER), a grant from the Andalusian Government, project GeOasis (P08-TIC-41999), and a grant from the University of Jaén, project RFC/PP2008/UJA-08-16-14.

References

[1] Eneko Agirre and Aitor Soroa. Personalizing pagerank for word sense disambiguation. In *EACL '09: Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pages 33–41, Morristown, NJ, USA, 2009. Association for Computational Linguistics.

[2] Christiane Fellbaum, editor. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA, 1998.

[3] Daniel Gildea and Daniel Jurafsky. Automatic labeling of semantic roles. *Comput. Linguist.*, 28(3):245–288, 2002.

[4] T. Joachims. Text categorization with support vector machines: learning with many relevant features. In *European Conference on Machine Learning (ECML)*, 1998.

[5] Arturo Montejo-Ráez, Jose Manuel Perea, Fernando Martínez-Santiago, Miguel Ángel García-Cumbreras, Maite Martín-Valdivia, and Alfonso Ure na-López. Combining lexical-syntactic information with machine learning for recognizing textual entailment. In *RTE '07: Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, pages 78–82, Morristown, NJ, USA, 2007. Association for Computational Linguistics.

[6] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford University, 1999.

[7] Daniel Ramage, Anna N. Rafferty, and Christopher D. Manning. Random walks for text semantic similarity. In *TextGraphs-4: Proceedings of the 2009 Workshop on Graph-based Methods for Natural Language Processing*, pages 23–31, Morristown, NJ, USA, 2009. Association for Computational Linguistics.