UC3M: Classification of Semantic Relations between Nominals using Sequential Minimal Optimization

Isabel Segura Bedmar
Computer Science Department
University Carlos III of Madrid
isegura@inf.uc3m.es

Doaa Samy
Computer Science Department
University Carlos III of Madrid
dsamy@inf.uc3m.es

Jose L. Martinez
Computer Science Department
University Carlos III of Madrid
jlmartinez@inf.uc3m.es

Abstract

This paper presents a method for automatic classification of semantic relations between nominals using Sequential Minimal Optimization. We participated in the four categories of SEMEVAL task 4 (A: No Query, No Wordnet; B: WordNet, No Query; C: Query, No WordNet; D: WordNet and Query) and for all training datasets. Best scores were achieved in category B using a set of feature vectors including lexical file numbers of nominals obtained from WordNet and a new feature *WordNet Vector* designed for the task[1].

1 Introduction

The survey of the state-of-art reveals an increasing interest in automatically discovering the underlying semantics in natural language. In this interdisciplinary field, the growing interest is justified by the number of applications which can directly benefit from introducing semantic information. Question Answering, Information Retrieval and Text Summarization are examples of these applications (Turney and Littman, 2005; Girju et al., 2005).

In the present work and for the purpose of the SEMEVAL task 4, our scope is limited to the semantic relationships between nominals. By this definition, we understand it is the process of discovering the underlying relations between two concepts expressed by two nominals.

Within the framework of SEMEVAL, nominals can occur either on the phrase, clause or the sentence level. This fact constitutes the major challenge in this task since most of the previous research limited their approaches to certain types of nominals mainly the "compound nominals"(Girju et al. 2005).

The paper is divided as follows; section 2 is a brief introduction to SMO used as the classifier for the task. Section 3 is dedicated to the description of the set of features applied in our experiments. In section 4, we discuss the experiment's results compared to the baselines of the SEMEVAL task and the top scores. Finally, we summarize our approach, pointing out conclusions and future directions of our work.

2 Sequential Minimal Optimization

We decided to use Support Vector Machine (SVM), as one of the most successful Machine Learning techniques, achieving the best performances for many classification tasks. Algorithm performance and time efficiency are key issues in our task, considering that our final goal is to apply this classification in a Question Answering System.

Sequential Minimal Optimization (SMO) is a fast method to train SVM. SMO breaks the large quadratic programming (QP) optimization problem needed to be resolved in SVM into a series of smallest possible QP problems. These small QP problems are analytically solved, avoiding, in this way, a time-consuming numerical QP optimization as an inner loop. We used Weka (Witten and Frank, 2005) an implementation of the SMO (Platt, 1998).

[1] This work has been partially supported by the Regional Government of Madrid Ander the Research Network MAVIR (S-0505/TIC-0267)

Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007), pages 382–385,
Prague, June 2007. ©2007 Association for Computational Linguistics

3 Features

Prior to the classification of semantic relations, characteristics of each sentence are automatically extracted using GATE (Cunningham et al., 2002). GATE is an infrastructure for developing and deploying software components for Language Engineering. We used the following GATE components: English Tokenizer, Part-Of-Speech (POS) tagger and Morphological analyser.

The set of features used for the classification of semantic relations includes information from different levels: word tokens, POS tags, verb lemmas, semantic information from WordNet, etc. Semantic features are only applied in categories B and D.

On the lexical level, the set of word features include the two nominals, their heads in case one of the nominals in question or both are compound nominals (e.g. the relation between <e1>*tumor shrinkage*</e1> and <e2>*radiation therapy* </e2> is actually between the head of the first "*shrinkage*" and "*radiation_therapy*"). More features include: the two words before the first nominal, the two words after the second nominal, and the word list in-between (Wang et al., 2006).

On the POS level, we opted for using a set of POS features since word features are often too sparse. This set includes POS tags of the two words occurring before the first nominal and the two words occurring after the second nominal together with the tag list of the words in-between (Wang et al., 2006). POS tags of nominals are considered redundant information.

Information regarding verbs and prepositions, occurring in-between the two nominals, is highly considered. In case of the verb, the system takes into account the verb token and the information concerning the voice and lemma. In the same way, the system keeps track of the prepositions occurring between both nominals. In addition, a feature, called *numinter,* indicating the number of words between nominals is considered.

Other important feature is the path from the first nominal to the second nominal. This feature is built by the concatenation of the POS tags between both nominals.

The feature related to the query provided for each sentence is only considered in the categories C and D according to the SEMEVAL restrictions.

On the semantic level, we used features obtained from WordNet. In addition to the WordNet sense keys, provided for each nominal, we extracted its synset number and its lexical file number.

Based on the work of Rosario, Hearst and Fillmore (2002), we suppose that these lexical file numbers can help to determine if the nominals satisfy the restrictions for each relation. For example, in the relation Theme-Tool, the theme should be an object, an event, a state of being, an agent, or a substance. Else, it is possible to affirm that the relation is false.

For the Part-Whole relation and due to its relevance in this classification task, a feature indicating metonymy relation in WordNet was taken into account.

Furthermore, we designed a new feature, called *WordNet vector*. For constructing this vector, we selected the synsets of the third level of depth in WordNet and we detected if each is ancestor or not of the nominal. It is a binary vector, i.e. if the synset is ancestor of the nominal it is assigned the value 1, else it is assigned the value 0. In this way, we worked with two vectors, one for each nominal. Each vector has a dimension of 13 coordinates. Each coordinate represents one of the 13 nodes in the third level of depth in WordNet. Our initial hypothesis considers that this representation for the nominals could perform well on unseen data.

4 Experiment Results

Cross validation is a way to test the ability of the model to classify unseen examples. We trained the system using 10-fold cross-validation; the fold number recommended for small training datasets. For each relation and for each category (A, B, C, D) we selected the set of features that obtained the best results using the indicated cross validation.

We submitted 16 sets of results as we participated in the four categories (A, B, C, D). We also used all the possible sizes of the training dataset (1: 1 to 35, 2:1 to 70, 3:1 to 106, 4:1 to 140).

	A: No Query, No WordNet			B: No Query, WordNet			C: No Query, No WordNet			D: Query, Word-Net		
	Prec	Rec	F	Prec	Rec	F	Prec	Rec	F	Prec	Rec	F
Cause-Effect	50.0	51.2	50.6	66.7	73.2	69.8	42.9	36.6	39.5	59.0	56.1	57.5
Instrument-Agency	47.5	50.0	48.7	73.7	73.7	73.7	51.4	50.0	50.7	67.5	71.1	69.2
Product-Producer	65.3	51.6	57.7	83.7	66.1	73.9	67.4	50.0	57.4	74.5	61.3	67.3
Origin-Entity	50.0	27.8	35.7	63.0	47.2	54.0	54.5	33.3	41.4	63.3	52.8	57.6
Theme-Tool	50.0	27.6	35.6	50.0	48.3	49.1	47.4	31.0	37.5	40.9	31.0	35.3
Part-Whole	26.5	34,6	30.0	72.4	80.8	76.4	34.0	61.5	43.8	57.1	76.9	65.6
Content-Container	48.4	39.5	43.5	57.6	50.0	53.5	48.6	44.7	46.6	63.6	55.3	59.2
Avg for UC3M	48.2	40.3	43.1	66.7	62.8	64.3	49.4	43.9	45.3	60.9	57.8	58.8
Avg for all systems	59.2	58.7	58.0	65.3	64.4	63.6	59.9	59.0	58.4	64.9	60.4	60.6
Max Avg F			64.8			72.4			65.1			62.6

Table 1 Scores for A4, B4, C4 and D4

For some learning algorithms such as decision trees and rule learning, appropriate selection of features is crucial. For the SVM model, this is not so important due to its learning mechanism, where irrelevant features are usually balanced between positive and negative examples for a given binary classification problem. However, in the experiments we observed that certain features have strong influence on the results, and its inclusion or elimination from the vector, influenced remarkably the outcomes.

In this section, we will briefly discuss the experiments in the four categories highlighting the most relevant observations.

In category A, we expected to obtain better results, but the overall performance of the system has decreased in the seven relations. This shows that our system has over-fitted the training set. The contrast between the F score values in the cross-validation and the final test results demonstrates this fact. For all the relations in the category A4, we obtained an average of F=43.1% [average score of all participating teams: F=58.0% and top average score: F=64.8%].

In Product-Producer relation, only two features were used: the two heads of the nominals. In training, we obtained an average F= 60% using cross-validation, while in the final test data, we achieved an average score F=57.7%. For the relation Theme-Tool, other set of features was employed: nominals, their heads, verb, preposition and the list of word between both nominals. Based on the results of the 10-fold cross validation, we expected to obtain an average of the F=70%. Nevertheless, the score obtained is F =30%.

In category B, our system has achieved better scores. Our average score F is 64.3% and it is above the average of participating teams (F=63.6%) and the baseline.

Best results in this category were achieved in the relations: Instrument-Agency (F=73.7%), Product-Producer (F=73.9%), Part-Whole (F=76.4%). However, for the relation Theme-Tool the system obtained lower scores (F=49.1%).

It is obvious that introducing WordNet information has improved notably the results compared with the results obtained in the category A.

In categories C and D, only three groups have participated. In category C (as in category A), the system results have decreased obviously (F=45.3%) with respect to the expected scores in the 10-fold cross validation. Moreover, the score obtained is lower than the average score of all participants (F=58.4%) and the best score (F=65.1%). For example, in training the Instrument-Agent relation, the system achieved an average F=78% using 10-fold cross-validation, while for the final score it only obtained F=50.7%.

Results reveal that the main reason behind the low scores in A and C, is the absence of information from WordNet. Hence, the vector design needs further consideration in case no semantic information is provided.

In category D, both WordNet senses and query were used, we achieved an average score F=58.8%. The average score for all participants is F=60.6% and the best system achieved F=62.6%. However, the slight difference shows that our system worked relatively well in this category.

Both run time and accuracy depend critically on the values given to two parameters: the upper bound on the coefficient's values in the equation for the hyperplane (-C), and the degree of the polynomials in the non-linear mapping (-E) (Witten and Frank, 2005). Both are set to 1 by default. The best settings for a particular dataset can be found only by experimentation.

We made numerous experiments to find the best value for the parameter C (C=1, C=10, C=100, C=1000, C=10000), but the results were not remarkably affected. Probably, this is due to the small size of the training set.

5 Conclusions and Future Work

In our first approach to automatic classification of semantic relations between nominals and as expected from the training phase, our system achieved its best performance using WordNet information. In general, we obtained better scores in category 4 (size of training: 1 to 140), i.e., when all the training examples are used.

On the other hand, overfitting the training data (most probably due to the small size of training dataset) is the main reason behind the low scores obtained by our system.

These facts lead us to the conclusion that semantic features from WordNet, in general, play a key role in the classification task. However, the relevance of WordNet-related features varies. For example, lexical file numbers proved to be highly effective, while the use of the *WordNet Vector* did not improve significantly the results. Thus, we consider that a level 3 *WordNet Vector* is rather abstract to represent each nominal. Developing a *WordNet Vector* with a deeper level (> 3) could be more effective as the representation of nouns is more descriptive.

Query features, on the other hand, did not improve the performance of the system. This is due to the fact that the same query could represent both positive and negative examples of the relation. However, to improve results in categories A and C, more features need to introduced, especially context and syntactic information such as chunks or dependency relations.

To improve results across the whole dataset, wider use of semantic information is necessary. For example, the immediate hypernym for each synset obtained from WordNet could help in improving the system performance (Nastase et al., 2006). Besides, information regarding the entity features could help in the classification of some relations like Origin-Entity or Product-Producer. Other semantic resources such as VerbNet, FrameNet, PropBank, etc. could also be used.

Furthermore, we consider introducing a Word Sense Disambiguation module to obtain the corresponding synsets of the nominals. Also, information concerning the synsets of the list of the context words could be of great value for the classification task (Wang et al., 2006).

References

Hamish Cunningham, Diana Maynard and Kalina Bontcheva, Valentin Tablan, Cristian Ursu. 2002. *The GATE User Guide*. http://gate.ac.uk/

Roxana Girju, Dan Moldovan, Marta Tatu and Daniel Antohe. 2005. On the semantics of noun compunds. *Computer Speech and Language 19* pp. 479-496.

Vivi Nastase, Jelber Sayyad-Shirbad, Marina Sokolova and Stan Szpakowicz. 2006. Learning noun-modifier semantic relations with corpus-based and WordNet-based features. In *Proc. of the 21st National Conference on Artificial Intelligence (AAAI 2006)*. Boston, MA.

John C. Platt. 1998. Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines, *Microsoft Research Technical Report MSR-TR-98-14*.

Barbara Rosario, Marti A. Hearst, and Charles Fillmore. 2002. "The descent of hierarchy, and selection in relations semantics". In *Proceedings of the 40 th Annual Meeting of the Association for Computacional Linguistics (ACL'02),* Philadelphia, PA, pages 417-424.

Ian H. Witten, Eibe Frank. 2005. Data Mining: Practical machine learning tools and techniques. *Morgan Kaufmann*.

Peter D. Turney and Michael L. Littman. 2005. Corpus-based learning of analogies and semantic rela-tions. *Machine Learning, in press*.

Ting Wang, Yaoyong Li, Kalina Bontcheva, Hamish Cunningham and Ji Wang. 2006. Automatic Extraction of Hierarchical Relations from Text. In *Proceedings of the Third European Semantic Web Conference (ESWC 2006),* Lecture Notes in Computer Science 4011, Springer, 2006.