The effect of Semantic Knowledge Expansion to Textual Entailment Recognition

Zornitsa Kozareva, Sonia Vázquez and Andrés Montoyo

Departamento de Lenguajes y Sistemas Informáticos
Universidad de Alicante
{zkozareva,svazquez,montoyo}@dlsi.ua.es

Abstract. This paper studies the effect of semantic knowledge expansion applied to the Textual Entailment Recognition task. In comparison to the already existing approaches we introduce a new set of similarity measures that captures hidden semantic relations among different syntactic categories in a sentence. The focus of our study is also centred on the synonym, antonym and verb entailment expansion of the initially generated pairs of words. The main objective for the realized expansion concerns the finding, the affirmation and the enlargement of the knowledge information. In addition, we applied Latent Semantic Analysis and the cosine measure to tune and improve the obtained relations. We conducted an exhaustive experimental study to evaluate the impact of the proposed new similarity relations for Textual Entailment Recognition.

1 Introduction

The web is the largest text repository, where millions of people share and consult information daily. Given a natural language query, present search engines identify and return relevant documents to the query. However, the relevant information may be present in different forms and a search about "tropical fruit" may return a document where "mango" appears. Although neither "tropical" nor "fruit" appear, the document is still relevant because "mango" is a type of tropical fruit. Other Natural Language Processing (NLP) applications have to handle language variabilities in order to avoid redundant information or to find the correct answer which may be represented in indirect way. Therefore, to improve their performance, a textual entailment (TE) module [4] is needed.

This directed researchers toward the development of diverse approaches of TE recognition such as logic forms [1], WordNet similarities [6],[7], [9], edit distance between parsing trees [8] among others [5].

At present, the already existing semantic similarity TE approaches, measure the word similarity among noun-noun, verb-verb, adjective-adjective and adverb-adverb pairs. In this work, we focus our study on word similarity relations among different syntactic categories. We measure the degree of contribution of such pairs to the recognition of textual entailment. In order to strengthen the similarity between the two texts, we expand the already obtained word pairs with their synonyms, antonym and verb entailment relations.

Additionally, we measure the semantic similarity between two texts using Latent Semantic Analysis (LSA) and the cosine measure. Instead of using the traditional word frequency approaches, we propose to measure similarity through the usage of relevant domains [14].

The paper is organised in the following way. Section 2 describes the motivation of our work and the utilized resources for our TE approach. Section 3 shows the experiments which we conducted to establish the robustness of the proposed method. Finally, we conclude in Section 4 and mention some work in progress.

2 Motivation and resource description

Recent textual entailment (TE) approaches [7], [3],[9] that rely on semantic information use only relations between words of the same syntactic category. However, we realise that word pairs from different syntactic categories also give relevant information. Thus, the main goal of our approach focuses on the study of the effect of semantic similarity between different syntactic categories such as verb-noun, adjective-noun, among others.

Additionally, we propose a new semantic similarity approach where the cosine and LSA are employed and examined. These measures identify the semantic distance and hidden relations between the text (T) and the hypothesis (H). The relatedness of the sentences is determined with the resource of relevant domains, rather than using the traditionall word frequency methods. The next subsections present the resources we utilized in our approach.

2.1 Inter-syntactic relations

Already existing works measure the semantic similarity between words of the same syntactic category. These systems do not take advantage of inter-syntactic relations[1]. In our study we find out that pairs of different syntactic categories are very indicative and can lead to a better textual entailment recognition. For example, in order to determine that *"He died of blood loss"* and *"He died bleeding"*, infer the same meaning, we need to use inter-syntactic relations. The previous approaches take into account only word pairs of the same syntactic category, so they cannot determine that *blood-N* and *bleeding-V* are semantically related. In this example *blood* and *bleeding* are the most relevant word pairs for the two texts and they infer that the entailment relation between the two sentences holds. Therefore, one of our main purposes in this investigation work is to apply the inter-syntactic relations which extract this kind of information.

To measure the semantic similarity between two sentences, first the parts of speech tags [13] are determined. From them, we took the four most significant word groups: verbs, nouns, adjectives and adverbs. The similarity between the different syntactic word pairs is determined with the WordNet::Similarity package [12].

[1] noun-verb, verb-noun, adjective-noun, noun-adjective, adverb-noun, noun-adverb, adjective-verb, verb-adjective, adverb-verb, verb-adverb

For each word pair [2], the *lin* and *path* similarity measures are applied. The reason of their usage is due to the different word senses and similarity scores that the WordNet::Similarity assigns. For example, the word pair "bank-money" with the measure of *lin* disambiguates the words with the senses bank#3–money#2 and establishes their similarity as 0.46. While the measure of *path* disambiguates the words as bank#8–money#2 with 0.14 similarity. In this example the first measure is more indicative.

2.2 Sentence expansion

To the previously extracted word pairs (noun–verb, noun–adjective, verb–adverb, etc), a synonym, antonym and verb entailment expansion is applied. For this expansion we use the WordNet[3] lexical resource.

The purpose of the word expansion is to provide to the original text (T) and hypothesis (H) sentences more relevant semantic information. The synonym expansion includes words that have the same meaning in the same context (arm–weapon). The antonym extracts words with opposite meaning (high–low). Verb entailment looks for verbs whose action can not be done unless the previous is accomplished (breathe–inhale, divorce–marry).

We come across some limitations associated to these expansions – the increase of computational cost and the degree of relevance for the new word pairs. The first obstacle is due to the large amount of possible combinations. The other is related to the appearance of a great number of synonyms that can transform the entailment relation from positive to negative and vice versa.

In order to reduce the noise of knowledge expansion, we used word sense disambiguation [12]. All words in T-H sentences are disambiguated and then expanded through WordNet. For example, for the pair bank-money, instead of including all synonyms related to all possible senses, we considered only the synonyms associated to senses bank#3–money#2 according to the measure of *lin*, and the senses bank#8–money#2 according to the measure of *path*.

2.3 Latent Semantic Analysis

LSA [10] has been applied in different NLP tasks. LSA consists in the construction and usage of a term-document matrix which describes the occurrences of terms in documents where each row corresponds to one term and each column corresponds to one document.

For our approach, we modify the space model of LSA. Instead of representing the columns as documents, we represent them as domains. These domains are extracted from the WordNet domain resource [11]. Thus, a new conceptual space with words and domains is obtained. This new space establishes the relevance among the words and the domains.

[2] a word pair consists of a word from the first sentences which is called the text and a word from the second sentence called the hypothesis

[3] http://wordnet.princeton.edu/

We use LSA technique to measure the similarity between two sentences. First, we obtain for each sentence the different constituents (noun, verb, adjective and adverb). Then, we apply the LSA over the words of the text T and the words of the hypothesis H. Thus, two different sets are obtained. These new sets contain a list of related words ordered by their similarity. The final step is to normalise the number of words that coincide between the T and the H.

Moreover, we use LSA in another approximation. Instead of using our conceptual space over terms and domains, we construct a new space, where the corpus is represented by the set of text sentences in the experimental data. Later, we use this new LSA space to determine the similarity between the T and H sentences. In the LSA experiments, we also study the effect of lemmatized and non lemmatized text.

2.4 Cosine measure

In our work, the cosine measure is used to establish the semantic relevance between T and H sentences. The most known usage of the cosine measure is taking the frequency of the words from the text and the hypothesis. In this work, we introduce a new interpretation of the cosine measure. Instead of word frequency, we consider Relevant Domains (RD).

The RD resource contains automatically extracted word-domain pairs, ordered by their association ratio. For each word in T/H, the set of RD is determined. Once this information is obtained, the T/H vectors are constructed and their similarity is measured with the formula 1.

$$\cos(T, H) = \frac{T \cdot H}{|T|\,|H|} = \frac{\sum_{i=1}^{n} T_i \cdot H_i}{\sqrt{\sum_{i=1}^{n} {T_i}^2} \cdot \sqrt{\sum_{i=1}^{n} {H_i}^2}} \tag{1}$$

The values of the cosine vary from 0 to 1, where a 1 indicates that T and H are very similar and 0 indicates that T and H have different meanings.

3 Experiments and Evaluation

This section concerns the experimental evaluation of the significance of the different knowledge representations which are described in the previous section.

All experiments are conducted with the Support Vector Machine (SVM) [2] algorithm. We selected this machine learning approach, because of its ability to manage high data scarcity problems and multidimensional attribute space.

3.1 Data Set

For our experiments, we use the development and test data sets provided by the Second Recognising Textual Entailment Challenge (RTE 2)[4]. The examples in

[4] http://www.pascal-network.org/Challenges/RTE2/

these data sets have been extracted from real Information Extraction, Information Retrieval, Question Answering and Text Summarization applications.

The development set consists of 800 text-hypothesis pairs, used as training examples. The other set of 800 text-hypothesis pairs is used for testing. The provided data sets are for the English language. The performances of the different knowledge representation sets are evaluated with the RTE2 evaluation script[5]. According to the script, systems are ranked and compared by their accuracy scores.

3.2 Experiment with knowledge expansion

The experiment knowledge expansion section presents two aspects – the contribution of the inter-syntactic word pairs and the effect of synonym, antonym and verb entailment relation expansions for the recognition of Textual Entailment.

We start our experiment with the measurement of the similarity for words of the same syntactic category. This approach is similar to the one presented in [9]. Next, we expand the initial noun, verb, adjective and adverbs pairs with their synonyms and verb entailment, as previously described in subsection 2.2. The obtained results are shown in Table 1.

In this table, we show the results for the development and the test data sets, so that a general overview of the behaviour of the knowledge features can be obtained. Without the expansion, the development set obtains 60.12% accuracy, while after the expansion, the performance increases with 0.53%. For the test set the performance improves with 1.38%.

sets	Acc.	IE	IR	QA	SUM
devWithoutExp	60.12	54.00	61.00	59.00	66.50
devWitExp	60.75	53.50	58.00	61.50	70.00
devAllAttr	59.62	**57.50**	60.00	57.50	63.50
devExpARNVent	61.38	55.50	60.50	62.00	67.50
devExpARNV_NpCd	59.62	50.50	59.00	59.00	70.00
testWithoutExp	54.25	50.00	55.50	47.50	64.00
testWitExp	**55.63**	52.00	56.50	57.00	57.00
testAllAttr	53.50	**52.50**	53.50	53.00	55.00
testExpARNVent	53.75	48.00	54.50	54.50	58.00
testExpARNV_NpCd	**55.37**	52.50	57.50	56.50	55.00

Table 1. *Results for the knowledge expansion experiments*

Considering the general scope of TE resolution, the performance of the already existing systems varies from 49% of accuracy as a minimum to 60% of accuracy as a maximum [5]. Therefore 1.38% of improvement can be considered as significant for a Textual Entailment system.

Once we demonstrated that the inclusion of synonym and verb entail expansion aided the TE recognition, we added the antonym and all inter-syntactic category information. In Table 1 this experiment is denoted with *AllAttr*. This information decreased the performance for the development and test data sets. The low performance is due to the antonym relations and to the accumulated noise introduced by the expansion of the inter-syntactic word groups. Additionally, not all sentence pairs express negative fact or event, therefore there is

[5] http://www.pascal-network.org/Challenges/RTE2/Evaluation/

no need to measure the antonym relation for each sentence. The synonym and antonym attributes contradict each other, therefore they sparse the example vector space of the SVM and hamper the classification of the employed machine learning algorithm.

An observation related to the *AllAttr* experiment concerns the performance of the Information Extraction (IE) task. Compared to the other sets, IE obtains 57.50% of accuracy. This shows that the inter-syntactic information is significant and important for the IE task, rather than to the other NLP tasks.

In order to confirm that the limitations of the *AllAttr* set are caused by the antonyms, we conduct an experiment where only the synonym expansion and verb entail information is included. For the development set, this combination obtains the highest accuracy of 61.38%. In addition, we add two more attributes: proper names and numbers. With them the performance of the development data decreases to 59.62%, however, the test data obtains 55.37%. This accuracy is the second highest score for the test data.

In this experimental subsection, we show that the expansion of synonym and verb entailment improves the score for the test data with around 1%. We also discover that the inter-syntactic relations are very informative for the IE task. In conclusion, we can affirm that semantic knowledge expansion has a positive effect over the performance of a TE system.

3.3 Experiment with LSA and the cosine measure

In respect to the previous experiments, in this section we study how entailments can be resolved using the LSA and the cosine measure. For all experiments, the results are shown in Table 2.

sets	Acc.	IE	IR	QA	SUM
devLSI_Lema	49.38	52.50	48.50	49.00	47.50
devLSI_NoLema	53.37	50.50	54.00	49.00	60.00
devCosine	54.25	50.50	48.00	57.00	61.50
devLSI_Lema_Cosine	53.63	52.50	50.00	50.00	62.00
devLSI_NoLema_Cosine	53.63	52.50	50.00	50.00	62.00
devBexpCosine	60.75	53.50	58.00	61.50	70.00
devBexpLSI_Lema_Cosine	63.38	55.50	63.50	62.50	72.00
devBexpLSI_NoLema_Cosine	61.50	57.00	60.00	61.00	68.00
testLSI_Lema	53.37	51.00	53.50	51.00	58.00
testLSI_NoLema	53.00	48.00	55.00	50.00	59.00
testCosine	54.00	46.50	56.50	56.00	57.00
testLSI_Lema_Cosine	52.38	47.00	54.50	52.50	55.50
testLSI_NoLema_Cosine	52.88	46.50	53.50	53.00	58.50
testBexpCosine	**55.63**	52.00	56.50	57.00	57.00
testBexpLSI_Lema_Cosine	52.88	51.50	55.00	51.50	53.50
testBexpLSI_NoLema_Cosine	**56.13**	53.50	57.00	58.00	56.00

Table 2. *Results for the LSI and cosine measures*

The experimental setup starts with the observation of the performance of the LSA with and without a lemmatizer. For the development data, the accuracy score increases with 4% in favour of the non lemmatized sentences, while for the test set the accuracy increase only with 0.37%. From the four different NLP tasks, the lemmatizer affects the performance of the information retrieval and summarisation.

The next experiments represent the different combinations of LSA, the cosine and the feature set with the synonym and verb entailment expansion. When only the LSA and cosine are combined the accuracy for the test set is decreased, because both measures depend only on the information of the relevant domains. However, combined with the expanded features, the final performance increases.

The best score for the whole Textual Entailment experiment are obtained after the combination of the LSA without a lemmatizer, the cosine, the synonym and verb entailment expansion. For the test data, this score is 56.13%. In comparison with the initial approach where simply the similarity of words from the same syntactic category are considered, the improvement is 2%. This shows that the incorporation of various semantic knowledge sources is beneficial and can help a semantic textual entailment module.

4 Conclusions and work in progress

The main contributions of this paper are related to the study of new semantic knowledge resources for the recognition of Textual Entailment.

First, we study the effect of word similarity across different syntactic categories. We discover that inter-syntactic information is very important for the text entailment recognition of the IE task.

On a second place, we take into account the word pair expansion with synonym, antonym and verb entailment relations. Such expansion lead to 1% of improvement compared to a system which does not use knowledge expansion. The performance of our system is lowered by the introduced noise of the newly incorporated irrelevant words. Although we used a word sense disambiguation method, by which words whose word senses do not correspond to the initial words are discarded, the experiments show that the computational time is highly increasing and the obtained knowledge is still noisy. At the moment, we are developing a method to discard and reduce these irrelevant word pairs, by the help of the LSA and the cosine measure.

Furthermore, we propose a novel approach to establish the semantic similarity of two sentences. For this approach we use the LSA and cosine measure, where the source of information is the relevant domain recourse, instead of the traditional word frequency methods. In addition, we have done different experiments, where the role of word lemmatization for the textual entailment recognition is demonstrated.

Finally, the combination of different semantic knowledge resources is explored. Among all experiments, the inclusion of synonym expansion, the verb entailment, the LSA and cosine measure yielded the highest score.

In conclusion, we can say that the effect of semantic knowledge expansion is significant for the textual entailment recognition. Following the development of our approach, the 2% improvement that is reached is significant, considering the global performance of the already existing systems.

In the future, in order to avoid the dispersion introduced by the expanded word pairs, we want to work with noun phrases. This will diminish the word

similarity combinations. With the same intention, LSA will be used to determine the most relevant synonym pairs.

Acknowledgements

This research has been funded by the Spanish Government under project CICyT number TIC2003-07158-C04-01 and PROFIT number FIT-340100-2004-14, and by the Valencia Government under project GV04B-276.

References

1. E. Akhmatova. Textual entailment resolution via atomic propositions. In *Proceedings of the PASCAL Challenges Workshop on Recognising Textual Entailment, 2005.*
2. R. Collobert and S. Bengio. Svmtorch: support vector machines for large-scale regression problems. *The Journal of Machine Learning Researc*, 2001.
3. C. Corley and R. Mihalcea. Measures of text semantic similarity. In *Proceedings of the ACL workshop on Empirical Modeling of Semantic Equivalence, 2005.*
4. I. Dagan and O. Glickman. Probabilistic textual entailment: Generic applied modeling of language variability. In *PASCAL Workshop on Learning Methods for Text Understanding and Mining,2004.*
5. I. Dagan, O. Glickman, and B. Magnini. The pascal recognising textual entailment challenge. In *Proceedings of the PASCAL Challenges Workshop on Recognising Textual Entailment,2005.*
6. J. Herrera, A. Peñas, and F. Verdejo. Textual entailment recognition based on dependency analysis and wordnet. In *Proceedings of the PASCAL Challenges Workshop on Recognising Textual Entailment,2005.*
7. V. Jijkoun and M. de Rijke. Recognizing textual entailment using lexical similarity. In *Proceedings of the PASCAL Challenges Workshop on Recognising Textual Entailment,2005.*
8. M. Kouylekov and B. Magnini. Recignizing textual entailment with tree edit distance algorithm. In *Proceedings of the PASCAL Challenges Workshop on Recognising Textual Entailment, 2005.*
9. Z. Kozareva and A. Montoyo. Mlent: The machine learning entailment system of the university of alicante. In *Proceedings of the PASCAL Challenges Workshop on Recognising Textual Entailment,2006.*
10. T. Landauer and S. Dumais. A solution to plato's problem: The latent semantic analysis theory of acquisition. In *Psychological Review*, pages 211–240, 1997.
11. B. Magnini and G. Cavaglia. Integrating Subject Field Codes into WordNet. In *Proceedings of LREC-2000, Second International Conference on Language Resources and Evaluation*, pages 1413–1418, 2000.
12. T. Pedersen, S. Patwardhan, and J. Michelizzi. Wordnet: : Similarity - measuring the relatedness of concepts. In *AAAI*, pages 1024–1025, 2004.
13. H. Schmid. Probabilistic part-of-speech tagging using decision trees. In *Proceedings International Conference on New Methods in Language Processing.*, pages 44–49, Manchester, UK, 1994.
14. S. Vázquez, A. Montoyo, and G. Rigau. Using relevant domains resource for word sense disambiguation. In *IC-AI*, pages 784–789, 2004.