An approach based on Logic Forms and WordNet relationships to Textual Entailment performance

Ó. Ferrández, R. M. Terol, R. Muñoz, P. Martínez-Barco and M. Palomar
Natural Language Processing and Information Systems Group
Department of Software and Computing Systems
University of Alicante
Alicante, Spain
{ofe,rafamt,rafael,patricio,mpalomar}@dlsi.ua.es

Abstract

This paper outlines the approach adopted by the PLSI research group at University of Alicante in the PASCAL-2006 second Recognising Textual Entailment challenge. Our system is composed of several components. On the one hand, the first component performs the derivation of the logic forms of the text/hypothesis pairs and, on the other hand, the second component provides us with a similarity score given by the semantic relations between the derived logic forms. In order to obtain this score we apply several measures of similitude and relatedness based on the structure and content of WordNet.

1 Introduction

This paper describes our participation in the second Recognising Textual Entailment (RTE) challenge, organized within the PASCAL network. Textual entailment is defined as a relation between two natural language expressions (Dagan and Glickman, 2004), a text (T) and an entailment hypothesis (H) that is entailed by T. For example,

T: His family has steadfastly denied the charges.
H: The charges were denied by his family.

is a true textual entailment.

The task of recognising this phenomenon is, without doubt, a complex task and great obstacle for many applications in the domain of natural language processing (Szpektor et al., 2004). For example, in a Question Answering (QA) system the same answer could be expressed in different syntactic and semantic ways and a RTE module could help QA system to identify the forecast answers that entail the expected answer. Similarly, in other natural language applications such us Information Retrieval, multi-document summarization and Information Extraction a RTE tool would be profitable for a better performance of each application.

We propose an approach based on knowledge as opposed to other authors who solve the problem of textual entailment by means of machine learning techniques. Our approach attempts to recognise textual entailment by determining if the text and the hypothesis are related using the derived logic forms from the text and the hypothesis, and by finding relations between their predicates using WordNet.

The architecture of our system is provided in Section 2, our results and performance analysis are presented in Section 3, and the conclusions of our participation in the challenge are drawn in Section 4.

2 System Architecture

Our approach is focused on the development of a textual entailment system based on knowledge techniques. Our system consists of two main components: the first produces the derivation of the logic forms and the other one computes the similarity measures between logic forms. The former embodies various advanced natural language processing techniques that derives from the text and the hypothesis the associated logic forms. The latter com-

ponent realizes a computation of similarity measures between the logic forms associated with the text and the hypothesis. This computation provides us with a score illustrating the similarity of the derived logic forms. Depending on the value of this score, we will decide if the two logic forms (text and hypothesis) are related or not. If the logic forms are related then the entailment between the text and the hypothesis is true.

An overview of our system is depicted in Figure 1. The following sections will describe in detail the main components of our system.

2.1 Derivation of the Logic Forms

The logic form of a sentence is derived through an analysis of dependency relationships between the words of the sentence. Our approach employs a set of rules that infer several aspects such as the assert, its type, its identifier and the relationships between the different asserts in the logic form.

This technique is clearly distinguished from other logic form derivation techniques such as Moldovan's (Moldovan and Rus, 2001) that constructs the logic form through the syntactic tree obtained as output of the syntactic parser. Our logic form, similar to Moldovan's logic form, is based on the logic form format defined in the eXtended WordNet (Harabagiu et al., 1999).

As an example, the logic form "*story:NN(x14) of:IN(x14, x13) variant:NN(x10) NNC(x11, x10, x12) fly:NN(x12) and:CC(x13, x11, x6) emergency:NN(x5) NNC(x6, x5, x7) rescue:NN(x8) NNC(x7, x8, x9) committee:NN(x9) who:NN(x1) save:VB(e1, x1, x2) thousand:NN(x2) in:IN(e1, x3) marseille:NN(x3)*" is automatically inferred from the analysis of dependency relationships between the words of the sentence "*The story of Variant Fly and the Emergency Rescue Committee who saved thousands in Marseille*". In this format of logic form each assert has at least one argument. The first argument is usually instantiated with the identifier of the assert and the rest of the arguments are identifiers of other asserts related to it. For instance, the assert "*story:NN(x14)*", has the type noun *(NN)* and the identifier *x14*; the assert "*NNC(x11, x10, x12)*", has the type complex nominal *(NNC)*, and its identifier is *x11*, and the other two arguments indicate the relationships to other asserts: *x10* and *x12*.

2.2 Computation of Similarity Measures between Logic Forms

In this section we are presenting the method followed by our system in order to obtain a similarity score between the logic forms. This method is focused on initially analysing the relation between the verbs of the two logic forms derived from the text and the hypothesis respectively. And secondly, if there is a relation between the verbs, then the method will analyse the similarity relations between all predicates which depending on the two verbs. If there is a NNC predicate that depends on the verb, the NNC predicate is explored until its NN associated predicates are obtained. All the weights provided by the analysis of the relations are summed and then normalized, thus obtaining the final normalized-relation score.

The aforementioned method is implemented as shown in the pseudo-code below.

```
simWeight = 0
Tvb = obtainVerbs(T)
Hvb = obtainVerbs(H)
for i = 0 ... size(Tvb) do
   for j = 0 ... size(Hvb) do
      if calcSim(Tvb(i),Hvb(j)) ≠ 0 then
         simWeight += calcSim(Tvb(i),Hvb(j))
         Telem = obtainElem(Tvb(i))
         Helem = obtainElem(Hvb(j))
         simWeight += calcSim(Telem,Helem)
      end if
   end for
end for
if simWeight > threshold then
   return TRUE
else
   return FALSE
end if
```

In order to obtain the similarity between the predicates of the logic forms (*calcSim(x,y)*), two approaches have been implemented: one based on WordNet relations and the other one based on Lin's measure (Lin, 1998). Both of them are based on WordNet, and they are described in detail below.

A Word Sense Disambiguation module was not employed in deriving the WordNet relations be-

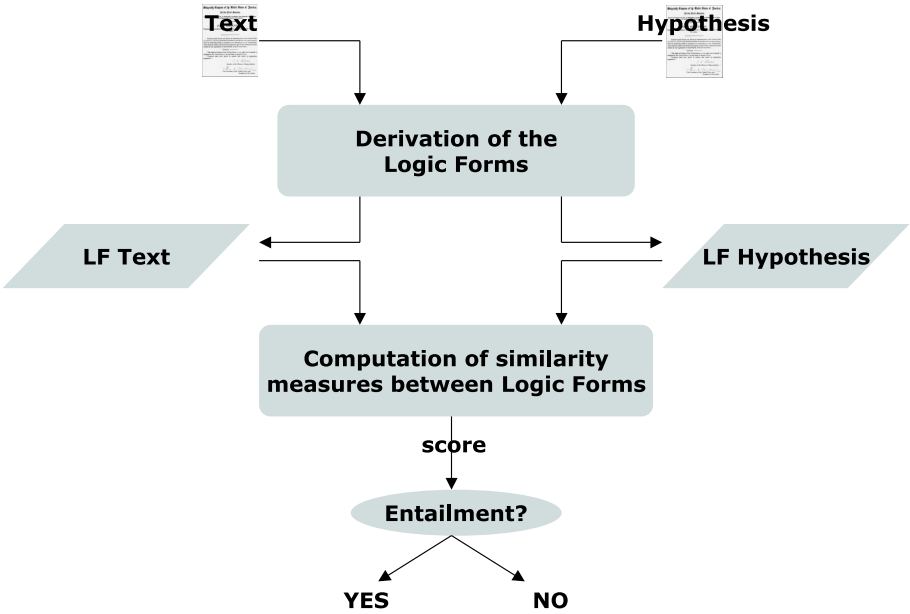

Figure 1: System architecture

tween any two predicates. Only the first 50% of the WordNet senses were taken into account. The threshold, which above one can consider that the text entails the hypothesis, has been obtained empirically using the provided development data. The Figure 2 in the section 3 presents this process in detail.

2.2.1 Approach Based on WordNet Relations

In the WordNet lexical database (Miller, 1990), a synset is a set of concepts that express the same meaning. A concept is defined as the use of one word in one determined context (sense). Thus, this task deals determining if two different concepts are related through the composition of different WordNet relations: hypernymy, hyponymy, entailment, similarity, meronymy and holonymy. The length of the path that relates the two different concepts must be lower or equal than 4 synsets. A weight has been assigned to each one of the WordNet relations: 0.8 for the hypernymy relationship, 0.7 for the hyponymy and entailment relationships, 0.9 for the similarity relationship, and 0.5 for the meronimy and holonymy relationships. Then, the weight of the path between two different concepts is calculated as the product of the weights associated to the relations connecting the intermediate synsets. This technique is different from the SpreadWeights algorithm (Moldovan and Novischi, 2002), even though derived from it.

2.2.2 Approach Based on Lin's Measure

In this case, the similarities were computed using Lin's similarity measure (Lin, 1998) as implemented in WordNet::Similarity[1] (Pedersen et al., 2004). WordNet::Similarity is an open source software package developed at the University of Minnesota. It allows the user to measure the semantic similarity or relatedness between a pair of concepts, as well as between a pair of words. WordNet::Similarity provides three measures of relatedness and six measures of similarity based on the WordNet lexical database. The similarity measures are based on analysing the WordNet *is-a* relations.

The similarity measures of WordNet::Similarity are divided into two groups: path-based and information content-based. For our experiments, we have chosen an information content-based similarity measure called Lin's similarity measure.

Lin's similarity measure augments the information content of the least common subsumer (LCS[2]) of the two concepts with the sum of the information

[1] http://www.d.umn.edu/ tpederse/similarity.html

[2] LCS is the most specific concept that two concepts share as an ancestor

content of the concepts themselves. The Lin's measure scales the information content of the LCS by this sum.

3 Result Analysis

Our participation in the RTE2 Challenge comprised two submissions. Both submissions were based on deriving the logic forms from the text and the hypothesis. However, our submission called *run1* computes the similarity measures between logic forms by means of Lin's similarity measure, whereas the *run2* uses our approach based on WordNet relations.

The officials results and the results achieved on the development data are shown in Table 1.

In order to adjust the threshold that determines if the text entails the hypothesis, we have carried out several experiments using the development data.

The Figure 2 shows an empirical increasing of the threshold in order to obtain the best performance one. The best threshold for both runs had a value of *0.24*.

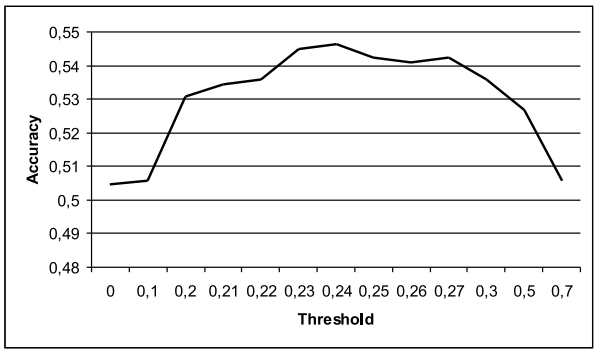

Figure 2: Adjusting the threshold on the development data

As we can deduct from Table 1, the run using Lin's similarity measure achieves better results than the approach based on WordNet relations, both when tested on development, as well as test data. This slight loss of accuracy is due to the fact that our WordNet relations approach (see Section 2.2.1) attempts to establish an objective semantic comparison between the logic forms rather than an entailment relation. Nevertheless, Lin's similarity measure, although not a pure entailment measure, seems to adapt good to the RTE task. In order to improve the results, a deeper study about more suitable existing WordNet relations should be performed.

Our system fails in many cases because it encounters good semantic matching between the logic forms of the text and the hypothesis, even if the two have got different meanings. In the case of the following example:

T: Jose Reyes scored the winner for Arsenal as they ended a three-game league losing streak with a victory over battling Charlton.

H: Jose Reyes scored the winner against Arsenal.

our system produces a true textual entailment due to a huge similarity score.

The reason for this is that the text's and the hypothesis' verbs and dependent predicates are the same or very similar semantically. However, in the hypothesis *against* causes a different meaning with respect to the text. Hitherto, our system is not able to recognise these cases. Therefore, a more detailed syntactic processing is needed in order to recognise the words that affect the meaning of the sentence.

4 Conclusions and Future Work

This paper presents a system dealing with the Textual Entailment phenomenon. Our system derives the logic forms for the text/hypothesis pair and computes the similarity between them. The similarity is computed using two different measures: Lin's similarity measure and WordNet relation-based similarity measure. Our system provides a score showing the semantic similarity between two logic forms. Although our system does not provide a specific entailment score, we found it challenging to evaluate it in a Textual Entailment competition.

We have achieved promising results for the RTE task (see Section 3), and the next step is to focalize our system for recognising only textual entailment.

As future work, we aim to perform a deeper study about the most suitable WordNet relations for recognising textual entailment. Perhaps only hypernymy, synonymy and entailment relations between the text and the hypothesis would be more suitable for the entailment phenomenon.

On the other hand, we are also interested in testing how other natural language processing tools can help in detecting textual entailment. For example,

run1	development data				
	overall	IE	IR	QA	SUM
Accuracy	0.5462	0.5421	0.5440	0.5722	0.5260
run2	development data				
Accuracy	0.5273	0.5510	0.5345	0.4677	0.5686
run1	test data				
Accuracy	0.5563	0.4950	0.5800	0.6100	0.5400
Average Precision	0.6089	0.5722	0.6159	0.6431	0.6215
run2	test data				
Accuracy	0.5475	0.4750	0.5850	0.6150	0.5150
Average Precision	0.5743	0.5853	0.6113	0.5768	0.5589

Table 1: *Final results for the development and test data*

using a Named Entity Recognizer could help in detecting entailment between two segment of text.

Acknowledgements

This research has been partially funded by the Spanish Government under project CICyT number TIC2003-07158-C04-01 and by the Valencia Government under project number GV04B-276.

References

Ido Dagan and Oren Glickman. 2004. Probabilistic textual entailment: Generic applied modeling of language variability. In *PASCAL Workshop on Learning Methods for Text Understanding and Mining*, pages 26–29, Grenoble, France, January.

S. Harabagiu, G.A. Miller, and D.I. Moldovan. 1999. Wordnet 2 - a morphologically and semantically enhanced resource. In *Proceedings of ACL-SIGLEX99: Standardizing Lexical Resources*, pages 1–8, Maryland, June.

Dekang Lin. 1998. An information-theoretic definition of similarity. In *ICML '98: Proceedings of the Fifteenth International Conference on Machine Learning*, pages 296–304, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

G.A. Miller. 1990. Wordnet: An on-line lexical database. In *International Journal of Lexicography 3, 4*, pages 235–312.

Dan Moldovan and Adrian Novischi. 2002. Lexical chains for question answering. In *Proceedings of the 19th international conference on Computational linguistics - Volume 1*, pages 1–7, Taipei, Taiwan.

Dan Moldovan and Vasile Rus. 2001. Logic form transformation of wordnet and its applicability to question-answering. In *Proceedings of 39th Annual Meeting of the Association for Computational Linguistics*, Toulouse, France, July.

Ted Pedersen, Siddharth Patwardhan, and Jason Michelizzi. 2004. Wordnet::similarity - measuring the relatedness of concepts. In *Proceedings of the Nineteenth National Conference on Artificial Intelligence (AAAI-04)*, San Jose, CA, July.

Idan Szpektor, Hristo Tanev, Ido Dagan, and Bonaventura Coppola. 2004. Scaling web-based acquisition of entailment relations. In *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP 2004)*, pages 41–48, Barcelona, Spain, July. Dekang Lin and Dekai Wu.