Arturo.

ACL 2007

PRAGUE

ACL 2007

Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing

June 28-29, 2007
Prague, Czech Republic

The Association for Computational Linguistics

Production and Manufacturing by
Omnipress Inc.
2600 Anderson Street
Madison, WI 53704, USA

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

Organizers

Chairs:

General Session:

Satoshi Sekine, New York University
Kentaro Inui, Nara Institute of Science and Technology

PASCAL RTE-3 Challenge:

Ido Dagan, Bar Ilan University
Bill Dolan, Microsoft Research
Danilo Giampiccolo, CELCT
Bernardo Magnini, ITC-irst

Program Committee:

Caroline Brun, Xerox Research Centre Europe, France
Johan Bos, University of Rome "La Sapienza"
Robert Dale, Macquarie University
Mark Dras, Macquarie University
Anette Frank, University of Heidelberg
Ralph Grishman, New York University
Sanda Harabagiu, University of Texas at Dallas
Graeme Hirst, University of Toronto
Yves Lepage, Universite de Caen
Dekang Lin, Google
Katja Markert, University of Leeds
Chris Manning, Stanford University
Rada Mihalcea, University of North Texas
Dan Moldovan, University of Texas at Dallas
Patrick Pantel, ISI
Kiyonori Ohtake, ATR
Ellen Riloff, University of Utah
Dan Roth, University of Illinois at Urbana-Champaign
Satoshi Sato, Nagoya University
Hinrich Schuetze, University of Stuttgart
Donia Scott, Open University
Kentaro Torisawa, JAIST
Lucy Vanderwende, Microsoft Research
Kazuhide Yamamoto, Nagaoka University of Technology
Fabio Zanzotto, University of Rome "Tor Vergata"

Invited Speaker:

Oren Etzioni, University of Washington

Website:

http://nlp.cs.nyu.edu/WTEP

Table of Contents

Combining Lexical-Syntactic Information with Machine Learning for Recognizing Textual Entailment

Arturo Montejo-Ráez, Jose Manuel Perea, Fernando Martínez-Santiago, Miguel Ángel García-Cumbreras, Maite Martín-Valdivia, Alfonso Ureña-López
Dpto. de Informática, Universidad de Jaén
Campus de las Lagunillas s/n, 23071 - Jaén
{amontejo, jmperea, dofer, magc, maite, laurena}@ujaen.es

Abstract

This document contains the description of the experiments carried out by SINAI group. We have developed an approach based on several lexical and syntactic measures integrated by means of different machine learning models. More precisely, we have evaluated three features based on lexical similarity and 11 features based on syntactic tree comparison. In spite of the relatively straightforward approach we have obtained more than 60% for accuracy. Since this is our first participation we think we have reached a good result.

1 Approach description

We fill face the textual entailment recognition using Machine Learning methods, i.e. identifying features that characterize the relation between hypothesis and associated text and generating a model using existing entailment judgements that will allow us to provide a new entailment judgement agains unseen pairs text-hypothesis. This approach can be split into the two processes shown in Figures 1 and 2.

In a more formal way, given a text t and an hypothesis h we want to define a function e which takes these two elements as arguments and returns and answer to the entailment question:

$$e(t,h) = \begin{cases} YES & \text{if } h \text{ is entailed by } t \\ NO & \text{otherwise} \end{cases} \quad (1)$$

Now the question is to find that ideal function

Figure 1: Training processes

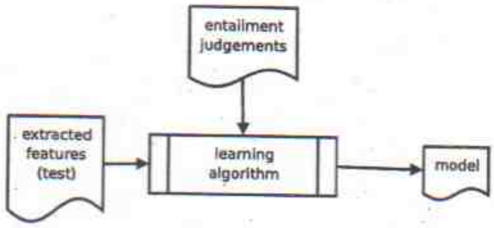
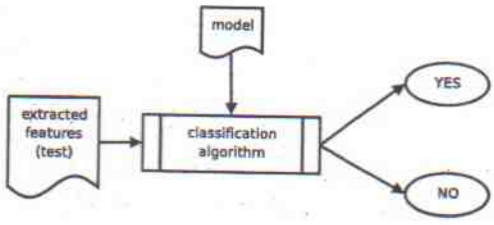


Figure 2: Classification processes



$e(t,h)$. We will approximate this function using a binary classifier:

$$\hat{e}(t,h) = bc(f,m) \quad (2)$$

where

bc is a binary classifier

f is a set of *features*

m is the learned model for the classifier

Therefore, it only remains to select a binary classifier and a feature extraction method. We have performed two experiments with different choices for both decisions. These two experiments are detailed below.

Proceedings of the Workshop on Textual Entailment and Paraphrasing, pages 78–82,
Prague, June 2007.

three words. If a hypothesis trigram matches in text, then the similarity weight value increases in one. The accumulated weight is normalized dividing it by the number of trigrams of the hypothesis.

1.2 Syntactic tree comparison

Some features have been extracted from pairs hypothesis-text related to the syntactic information that some parser can produce. The rationale behind it consists in measuring the similarity between the syntactic trees of both hypothesis and associated text. To do that, terms appearing in both trees are identified (we call this *alignment*) and then, graph distances (number of nodes) between those terms in both trees are compared, producing certain values as result.

In our experiments, we have applied the COLLINS (Collins, 1999) parser to generate the syntactic tree of both pieces of text. In Figure 3 the output of the syntactic parsing for a sample pair is shown. This data is the result of the syntactical analysis performed by the mentioned parser. A graph based view of the tree corresponding to the hypothesis is drawn in Figure 4. This graph will help us to understand how certain similarity measures are obtained.

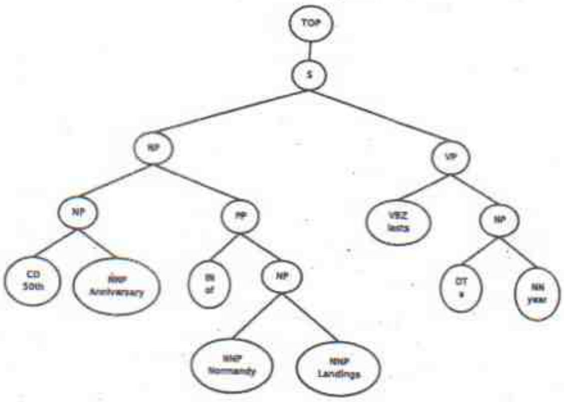
Figure 3: Syntactic trees of sample hypothesis and its associated text

```
<t>
(TOP (S (LST (LS 0302) (. .)) (NP (JJ Next) (NN year))
(VP (VBZ is) (NP (NP (DT the) (JJ 50th) (NN anniversary))
(PP (IN of) (NP (NP (DT the) (NNP Normandy) (NN invasion)
(, ,)) (NP (NP (DT an)(NN event)) (SBAR (IN that) (S (VP
(MD would) (RB n't) (VP (VB have) (VP (VBN been) (ADJP
(JJ possible)) (PP (IN without) (NP (NP (DT the) (NNP
Liberty) (NN ships.)) (SBAR (S (NP (DT The) (NNS
volunteers)) (VP (VBP hope) (S (VP (TO to) (VP (VB raise)
(NP (JJ enough) (NN money)) (S (VP (TO to) (VP (VB sail)
(NP (DT the) (NNP O'Brien)) (PP (TO to) (NP (NNP France)))
(PP (IN for)(NP (DT the) (JJ big) (NNP D-Day) (NN celebration)
(. .))))))))))))))))))))))))))))
</t>

<h>
(TOP (S (NP (NP (CD 50th) (NNP Anniversary)) (PP (IN of)
(NP (NNP Normandy) (NNP Landings)))) (VP (VBZ lasts) (NP
(DT a) (NN year) (. .)))))
</h>
```

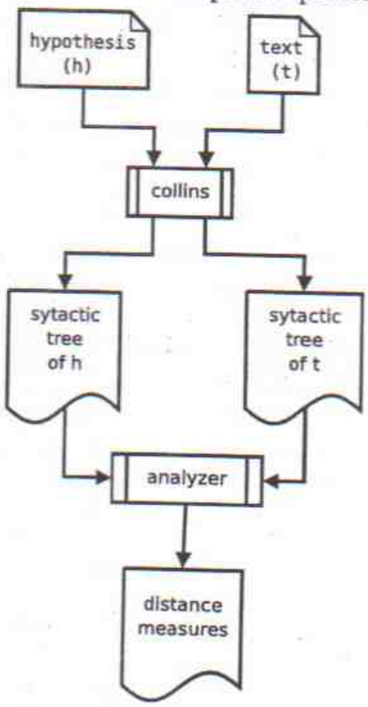
From the sample above, the terms *normandy*, *year* and *anniversary* appear in both pieces of text. We say that these terms are "aligned". Therefore, for the three possible pairs of aligned terms we can compute the distance, in nodes, to go from one term to the other at each tree. Then, the difference of these

Figure 4: Syntact tree of sample hypothesis



distances is computed and some statistics are generated. We can summarize the process of computing this differences in the algorithm detailed in Figure 6.

Figure 5: Tree comparison process



For instance, in the tree represented in Figure 4 we can see that we have to perform 5 steps to go from node *Anniversary* to node *Normandy*. Since there are no more possible occurrences of these two terms, then the minimal distance between them is 5. This value is also measured on the tree corre-

- **Exp2** uses five features: four lexical similarities ($SIM_{matching}$ + $CSS_{matching}$ + Trigrams + $BIN_{matching}$) and Syntactic tree comparison.
- **Exp3** uses only three lexical similarities ($SIM_{matching}$ + $CSS_{matching}$ + Trigrams).
- **Exp4** uses the four lexical similarities ($SIM_{matching}$ + $CSS_{matching}$ + Trigrams + $BIN_{matching}$)
- **Exp5** uses only three lexical similarities ($SIM_{matching}$ + $CSS_{matching}$ + Trigrams).
- **Exp6** uses four features: three lexical similarities ($SIM_{matching}$ + $CSS_{matching}$ + Trigrams) and Syntactic tree comparison.

As we expected, the best result we have obtained is by means of the integration of the whole of the features available. More surprising is the good result obtained by using lexical features only, even better than experiments based on syntactical features only. On the other hand, we expected that the integration of both sort of features improve significatively the performance of the system, but the improvement respect of lexical features is poor (less than 2%). . Similar topics share similar vocabulary, but not similar syntax at all. Thus, we think we should to investigate semantic features better than the syntactical ones.

3 Conclusions and future work

In spite of the simplicity of the approach, we have obtained remarkable results: each set of features has reported to provide relevant information concerning to the entailment judgement determination. On the other hand, these two approaches can be merged into one single system by using different features all together and feeding with them several binary classifiers that could compose a voting system. We will do that combining TiMBL, SVM and BBR.We expect to improve the performance of the entailment recognizer by this integration.

Finally, we want to implement a hierarchical architecture based on constraint satisfaction networks. The constraints will be given by the set of available features and the maintenance of the integration across the semantic interpretation process.

4 Acknowledgements

This work has been partially financed by the TIMOM project (TIN2006-15265-C06-03) granted by the Spanish Government Ministry of Science and Technology and the RFC/PP2006/Id_514 granted by the University of Jaén.

References

Alexander Budanitsky and Graeme Hirst. 2001. Semantic distance in wordnet: An experimental, application-oriented evaluation of five measures.

Michael Collins. 1999. *Head-Driven Statistical Models for Natural Language Parsing*. Ph.D. thesis, University of Pennsylvania.

Walter Daelemans, Jakub Zavrel, Ko van der Sloot, and Antal van den Bosch. 1998. Timbl: Tilburg memory based learner, version 1.0, reference guide.

Oscar Ferrandez, Daniel Micolo, Rafael Mu noz, and Manuel Palomar. 2007. Técnicas léxico-sintácticas para reconocimiento de inmplicación textual. . *Tecnologías de la Informacón Multilingüe y Multimodal. In press.*

Jay J. Jiang and David W. Conrath. 1997. Semantic similarity based on corpus statistics and lexical taxonomy. In *Proceedings of International Conference on Research in Computational Linguistics*, Taiwan.

Dekang Lin. 1998. An information-theoretic definition of similarity. In *Proceedings of the 15th International Conference on Machine Learning*.

Philip Resnik. 1995. Using information content to evaluate semantic similarity. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, Montreal.