ALTN: Word Alignment Features for Cross-lingual Textual Entailment

Marco Turchi and Matteo Negri
Fondazione Bruno Kessler
Trento, Italy
{turchi,negri}@fbk.eu

Abstract

We present a supervised learning approach to cross-lingual textual entailment that explores statistical word alignment models to predict entailment relations between sentences written in different languages. Our approach is language independent, and was used to participate in the CLTE task (Task#8) organized within Semeval 2013 (Negri et al., 2013). The four runs submitted, one for each language combination covered by the test data (*i.e.* Spanish/English, German/English, French/English and Italian/English), achieved encouraging results. In terms of accuracy, performance ranges from 38.8% (for German/English) to 43.2% (for Italian/English). On the Italian/English and Spanish/English test sets our systems ranked second among five participants, close to the top results (respectively 43.4% and 45.4%).

1 Introduction

Cross-lingual textual entailment (CLTE) is an extension of the Textual Entailment task (Dagan and Glickman, 2004) that consists in deciding, given two texts T and H written in different languages (respectively called text and hypothesis), if H can be inferred from T (Mehdad et al., 2010). In the case of SemEval 2013, the task is formulated as a multi-class classification problem in which there are four possible relations between T and H: forward ($T \rightarrow H$), backward ($T \leftarrow H$), bidirectional ($T \leftrightarrow H$) and "no entailment".

Targeting the identification of semantic equivalence and information disparity between topically related sentences, CLTE recognition can be seen as a core task for a number of cross-lingual applications. Among others, multilingual content synchronization has been recently proposed as an ideal framework for the exploitation of CLTE components and the integration of semantics and machine translation (MT) technology (Mehdad et al., 2011; Mehdad et al., 2012b; Bronner et al., 2012; Monz et al., 2011).

In the last few years, several methods have been proposed for CLTE. These can be roughly divided in two main groups (Negri et al., 2012): *i)* those using a *pivoting* strategy by translating H into the language of T and then using monolingual TE components[1], and those directly using cross-lingual strategies. Among this second group, several sources of cross-lingual knowledge have been used, such as dictionaries (Kouylekov et al., 2012; Perini, 2012), phrase and paraphrase tables (Mehdad et al., 2012a), GIZA++ (Och and Ney, 2003) word alignment models (Wäschle and Fendrich, 2012), MT of subsegments (Esplà-Gomis et al., 2012), or semantic Wordnets (Castillo, 2011).

In this work we propose a CLTE detection method based on a new set of features using word alignment as a source of cross-lingual knowledge. This set, which is richer than the one by (Wäschle and Fendrich, 2012), is aimed not only at grasping information about the proportion of aligned words, but also about the distribution of the alignments in both

[1] In the first CLTE evaluation round at Semeval 2012, for instance, the system described in (Meng et al., 2012) used the open source EDITS system (Kouylekov and Negri, 2010; Negri et al., 2009) to calculate similarity scores between monolingual English pairs.

*Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 128–132, Atlanta, Georgia, June 14-15, 2013.

H and T. This set of features is later used by two support vector machine (SVM) classifiers for detecting CLTE separately in both directions ($T \rightarrow H$ and $T \leftarrow H$). We use the combined output of both classifiers for performing the CLTE detection.

The paper is organized as follows: Section 2 describes the features used and the classification method; Section 3 explains the experimental framework and the results obtained for the different language-pair sets; finally, the conclusions obtained from the results are summarised in Section 4.

2 ALTN System

In our approach we have implemented a system based on supervised learning. It takes an unlabeled sentence pair as input (T and H) and labels it automatically with one of the possible four valid entailment relations. The architecture is depicted in Figure 1.

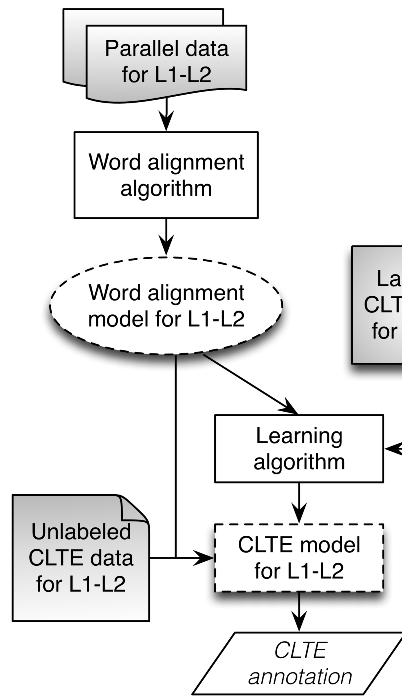

Figure 1: System architecture

A key component to our approach is the word alignment model. In a preprocessing step it is trained on a set of parallel texts for the target language pair. Next, different features based on the word alignment are extracted. Taking the features and the target language pair labels as input, a supervised learning algorithm is run to fit a model to the data. The last step is to use the model to automatically label unseen instances with entailment relations.

2.1 Features

What characterizes our submission is the use of word alignment features to capture entailment relations. We extract the following features from a word alignment model for a given sentence pair (all features are calculated for both T and H):

- proportion of aligned words in the sentence (baseline);
- number of unaligned sequences of words normalized by the length of the sentence;
- length of the longest sequence of aligned words normalized by the length of the sentence;
- length of the longest sequence of unaligned words normalized by the length of the sentence;
- average length of the aligned word sequences;
- average length of the unaligned word sequences;
- position of the first unaligned word normalized by the length of the sentence;
- position of the last unaligned word normalized by the lenght of the sentence;
- proportion of aligned n-grams in the sentence (n varying from 1 to 5).

These features are language independent as they are obtained from statistical models that take as input a parallel corpus. Provided that there exist parallel data for a given language pair, the only constraint in terms of resources, the adoption of these features makes our approach virtually portable across languages with limited effort.

2.2 CLTE Model

Our CLTE model is composed by two supervised binary classifiers that predict whether there is entailment between the T and H. One classifier checks for forward entailment ($T \rightarrow H$) and the other checks for backward entailment ($T \leftarrow H$). The output of both classifiers is combined to form the four valid entailment decisions:

- forward and backward classifier output `true`: "bidirectional" entailment;
- forward is `true` and backward is `false`: "forward" entailment;
- forward is `false` and backward is `true`: "backward" entailment;
- both forward and backward output `false`: "no entailment" relation.

Both binary classifiers were implemented using the SVM implementation of Weka (Hall et al., 2009).

3 Experiments

In our submission we experimented with three standard word alignment algorithms: the *hidden Markov model* (HMM) (Vogel et al., 1996) and *IBM models* 3 and 4 (Brown et al., 1993). They are implemented in the MGIZA++ package (Gao and Vogel, 2008). Building on a probabilistic lexical model to establish mappings between words in two languages, these models compute alignments between the word positions in two input sentences S_1 and S_2. The models are trained incrementally: HMM is the base for IBM model 3, which is the base for IBM model 4. To train our models, we used 5 iterations of HMM, and 3 iterations of IBM models 3 and 4.

Word alignments produced by these models are asymmetric ($S_1 \rightarrow S_2 \neq S_2 \rightarrow S_1$). To cope with this, different heuristics (Koehn et al., 2005) have been proposed to obtain symmetric alignments from two asymmetric sets ($S_1 \leftrightarrow S_2$). We experimented with three symmetrization heuristics, namely: *union*, *intersection*, and *grow-diag-final-and*, a more complex symmetrization method which combines intersection with some alignments from the union.

To train the word alignment models we used the Europarl parallel corpus (Koehn, 2005) concatenated with the News Commentary corpus[2] for three language pairs: English-German (2,079,049 sentences), English-Spanish (2,123,036 sentences), English-French (2,144,820 sentences). For English-Italian we only used the parallel data available in Europarl (1,909,115 sentences) since this language pair is not covered by the News Commentary corpus.

For our submitted run the SVM classifiers were trained using the whole training set. Such dataset consists of 1,000 pairs for each of the four language combinations, resulting from a concatenation of the training and test sets used for the first round of evaluation at SemEval 2012 (Negri et al., 2012; Negri et al., 2011). We have set a polynomial kernel with parameters empirically estimated on the training set: $C = 2.0$, and $d = 1$. After some preliminary experiments we have concluded that the HMM model in conjunction with the *intersection* symmetrization provides the best results.

Our results, calculated over the 500 test pairs provided for each language combination, are presented in Table 3. As can be seen from the table, our system consistently outperforms the best average run of all participants and is the second best system for Spanish/English and Italian/English. For the other two languages, French/English and German/English, it is the 3rd best system with a larger distance from top results. The motivations for such lower results, currently under investigation, might be related to lower performance in terms of word alignment, the core of our approach. The first step of our analysis will hence address, and in case try to cope with, significant differences in word alignment performance affecting results.

Overall, considering the small distance from top results, and the fact that our approach does not require deep linguistic processing to be reasonably effective for any language pair for which parallel corpora are available, our results are encouraging and motivate further research along such direction.

4 Conclusion

In this paper we presented the participation of the Fondazione Bruno Kessler in the Semeval 2013 Task#8 on Cross-lingual Textual Entailment for Content Synchronization. To identify entailment relations between texts in different languages, our system explores the use of word alignment features

[2] `http://www.statmt.org/wmt11/translation-task.html#download`

Features / Language pair	German/English	Spanish/English	French/English	Italian/English
Avg best runs	0.378	0.404	0.407	0.405
ALTN	0.388	0.428	0.420	0.432
Best system	0.452	0.434	0.458	0.454

Table 1: Accuracy results for the language pairs evaluated for the average of the best runs of the participating systems, our submission and the best systems.

within a supervised learning setting. In our approach, word alignment models obtained by statistical methods from parallel corpora leverage information about the number, the proportion, and the distribution of aligned terms in the input sentences. In terms of accuracy results over the SemEval 2013 CLTE test data, performance ranges from 38.8% (for German/English) to 43.2% (for Italian/English). On the Italian/English and Spanish/English test sets our systems ranked second among five participants, close to the top results (respectively 43.4% and 45.4%). Such results suggest that the use of word alignment models to capture sentence-level semantic relations in different language settings represents a promising research direction.

Acknowledgments

This work has been partially supported by the EC-funded project CoSyne (FP7-ICT-4-248531).

References

Amit Bronner, Matteo Negri, Yashar Mehdad, Angela Fahrni, and Christof Monz. 2012. CoSyne: Synchronizing Multilingual Wiki Content. In *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration*, WikiSym '12, pages 33:1–33:4, New York, NY, USA. ACM.

Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer. 1993. The Mathematics of Statistical Machine Translation: Parameter Estimation. *Computational Linguistics*, 19(2):263–311.

Julio J. Castillo. 2011. A WordNet-based Semantic Approach to Textual Entailment and Cross-lingual Textual Entailment. *International Journal of Machine Learning and Cybernetics*, 2(3):177–189.

Ido Dagan and Oren Glickman. 2004. Probabilistic Textual Entailment: Generic Applied Modeling of Language Variability. In *Proceedings of the PASCAL Workshop of Learning Methods for Text Understanding and Mining*, Grenoble, France.

Miquel Esplà-Gomis, Felipe Sánchez-Martínez, and Mikel L. Forcada. 2012. UAlacant: Using Online Machine Translation for Cross-Lingual Textual Entailment. In *Proceedings of the 6th International Workshop on Semantic Evaluation (SemEval 2012)*, pages 472–476, Montréal, Canada.

Qin Gao and Stephan Vogel. 2008. Parallel Implementations of Word Alignment Tool. In *Software Engineering, Testing, and Quality Assurance for Natural Language Processing*, pages 49–57, Columbus, Ohio, USA.

Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The WEKA Data Mining Software: an Update. *SIGKDD Explorations*, 11(1):10–18.

Philipp Koehn, Amittai Axelrod, Alexandra Birch Mayne, Chris Callison-Burch, Miles Osborne, and David Talbot. 2005. Edinburgh System Description for the 2005 IWSLT Speech Translation Evaluation. In *Proceedings of the International Workshop on Spoken Language Translation*, Pittsburgh, Pennsylvania, USA.

Philip Koehn. 2005. Europarl: a Parallel Corpus for Statistical Machine Translation. In *Proceedings of Machine Translation Summit X*, pages 79–86, Phuket, Thailand.

Milen Kouylekov and Matteo Negri. 2010. An Open-source Package for Recognizing Textual Entailment. In *Proceedings of the ACL 2010 System Demonstrations*.

Milen Kouylekov, Luca Dini, Alessio Bosca, and Marco Trevisan. 2012. CELI: an Experiment with Cross Language Textual Entailment. In *Proceedings of the 6th International Workshop on Semantic Evaluation (SemEval 2012)*, pages 696–700, Montréal, Canada.

Yashar Mehdad, Matteo Negri, and Marcello Federico. 2010. Towards Cross-Lingual Textual Entailment. In *Proceedings of the 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL HLT 2010)*.

Yashar Mehdad, Matteo Negri, and Marcello Federico. 2011. Using Bilingual Parallel Corpora for Cross-Lingual Textual Entailment. In *Proceedings of the 49th Annual Meeting of the Association for Compu-*

tational Linguistics: Human Language Technologies (ACL HLT 2011).

Yashar Mehdad, Matteo Negri, and José Guilherme C. de Souza. 2012a. FBK: cross-lingual textual entailment without translation. In *Proceedings of the 6th International Workshop on Semantic Evaluation (SemEval 2012)*, pages 701–705, Montréal, Canada.

Yashar Mehdad, Matteo Negri, and Marcello Federico. 2012b. Detecting Semantic Equivalence and Information Disparity in Cross-lingual Documents. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL 2012)*.

Fandong Meng, Hao Xiong, and Qun Liu. 2012. ICT: A Translation based Cross-lingual Textual Entailment. In *Proceedings of the 6th International Workshop on Semantic Evaluation (SemEval 2012)*.

Christoph Monz, Vivi Nastase, Matteo Negri, Angela Fahrni, Yashar Mehdad, and Michael Strube. 2011. CoSyne: a Framework for Multilingual Content Synchronization of Wikis. In *Proceedings of Wikisym 2011, the International Symposium on Wikis and Open Collaboration*, pages 217–218, Mountain View, California, USA.

Matteo Negri, Milen Ognianov Kouylekov, Bernardo Magnini, Yashar Mehdad, and Elena Cabrio. 2009. Towards Extensible Textual Entailment Engines: the EDITS Package. In *AI*IA 2009: XIth International Conference of the Italian Association for Artificial Intelligence*.

Matteo Negri, Luisa Bentivogli, Yashar Mehdad, Danilo Giampiccolo, and Alessandro Marchetti. 2011. Divide and Conquer: Crowdsourcing the Creation of Cross-Lingual Textual Entailment Corpora. *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP 2011)*.

Matteo Negri, Alessandro Marchetti, Yashar Mehdad, Luisa Bentivogli, and Danilo Giampiccolo. 2012. Semeval-2012 Task 8: Cross-Lingual Textual Entailment for Content Synchronization. In *Proceedings of the 6th International Workshop on Semantic Evaluation (SemEval 2012)*, pages 399–407, Montréal, Canada.

Matteo Negri, Alessandro Marchetti, Yashar Mehdad, Luisa Bentivogli, and Danilo Giampiccolo. 2013. Semeval-2013 Task 8: Cross-lingual Textual Entailment for Content Synchronization. In *Proceedings of the 7th International Workshop on Semantic Evaluation (SemEval 2013)*.

Franz J. Och and Hermann Ney. 2003. A Systematic Comparison of Various Statistical Alignment Models. *Computational Linguistics*, 29(1):19–51.

Alpár Perini. 2012. DirRelCond3: detecting textual entailment across languages with conditions on directional text relatedness scores. In *Proceedings of the 6th International Workshop on Semantic Evaluation (SemEval 2012)*, pages 710–714, Montréal, Canada.

Stephan Vogel, Hermann Ney, and Christoph Tillmann. 1996. HMM-based word alignment in statistical translation. In *Proceedings of the 16th International Conference on Computational Linguistics (ACL'96)*, pages 836–841, Copenhagen, Denmark.

Katharina Wäschle and Sascha Fendrich. 2012. HDU: Cross-lingual Textual Entailment with SMT Features. In *Proceedings of the 6th International Workshop on Semantic Evaluation (SemEval 2012)*, pages 467–471, Montréal, Canada.