Exploring the Effectiveness of Lexical Ontologies for Modeling Temporal Relations with Markov Logic

Eun Y. Ha, Alok Baikadi, Carlyle J. Licata, Bradford W. Mott, James C. Lester
Department of Computer Science
North Carolina State University
Raleigh, NC, USA
{eha,abaikad,cjlicata,bwmott,lester}@ncsu.edu

Abstract

Temporal analysis of events is a central problem in computational models of discourse. However, correctly recognizing temporal aspects of events poses serious challenges. This paper introduces a joint modeling framework and feature set for temporal analysis of events that utilizes Markov Logic. The feature set includes novel features derived from lexical ontologies. An evaluation suggests that introducing lexical relation features improves the overall accuracy of temporal relation models.

1 Introduction

Reasoning about the temporal aspects of events is a critical task in discourse understanding. Temporal analysis techniques contribute to a broad range of applications including question answering and document summarization, but temporal reasoning is complex. A recent series of shared task evaluation challenges proposed a framework with standardized sets of temporal analysis tasks, including identifying the temporal entities mentioned in text, such as events and time expressions, as well as identifying the temporal relations that hold between those temporal entities (Pustejovsky and Verhagen, 2009).

Our previous work (Ha et al., 2010) addressed modeling temporal relations between temporal entities and proposed a supervised machine-learning approach with *Markov Logic* (ML) (Richardson and Domingos, 2006). As novel features, we introduced two types of lexical relations derived from VerbOcean (Chklovski and Pantel, 2004) and WordNet (Fellbaum, 1998). A preliminary evaluation showed the effectiveness of our approach. In this paper, we extend our previous work and conduct a more rigorous evaluation, focusing on the impact of joint optimization of the features and the effectiveness of the lexical relation features for modeling temporal relations.

2 Related Work

Recently, data-driven approaches to modeling temporal relations for written text have been gaining momentum. Boguraev and Ando (2005) apply a semi-supervised learning technique to recognize events and to infer temporal relations between time expressions and their anchored events. Mani et al. (2006) model temporal relations between events as well as between events and time expressions using maximum entropy classifiers. The participants of TempEval-1 investigate a variety of techniques for temporal analysis of text (Verhagen et al., 2007).

While most data-driven techniques model temporal relations as local pairwise classifiers, this approach has the limitation that there is no systematic mechanism to ensure global consistencies among predicted temporal relations (e.g., if event *A* happens before event *B* and event *B* happens before event *C*, then *A* should happen before *C*). To avoid this drawback, a line of research has explored techniques for the global optimization of local classifier decisions. Chambers and Jurafsky (2008) add global constraints over local classifiers using Integer Linear Programming. Yoshikawa et al. (2009) jointly model related temporal classification tasks using ML. These approaches are shown to improve the accuracy of temporal relation models.

Our work is most closely related to Yoshikawa et al. (2009) in that ML is used for joint model-

ing of temporal relations. We extend their work in three primary respects. First, we introduce new lexical relation features. Second, our model addresses a new task introduced in TempEval-2. Third, we employ phrase-based syntactic features (Bethard and Martin 2007) rather than dependency-based syntactic features.

3 Data and Tasks

We use the TempEval-2 data for English for both training and testing of our temporal relation models. The data includes 162 news articles (totaling about 53,000 tokens) as the training set and another 11 news articles as the test set. The corpus is labeled with events, time expressions, and temporal relations. Each labeled event and time expression is further annotated with semantic and syntactic attributes. Six types of temporal relations are considered: *before*, *after*, *overlap*, *before-or-overlap*, *overlap-or-after*, and *vague*.

Consider the following example from the TempEval-2 data, marked up with a time expression t_1 and three events e_1, e_2, and e_3, where e_1 and e_2 are the main events of the first and the second sentences, respectively, and e_3 is syntactically dominated by e_2.

But a [minute and a half]^{t1} later, a pilot from a nearby flight [calls]^{e1} in. Ah, we just [saw]^{e2} an [explosion]^{e3} up ahead of us here about sixteen thousand feet or something like that.

In the first sentence, t_1 and e_1 are linked by a temporal relation *overlap*. Temporal relation *after* holds between the two consecutive main events: e_1 occurs *after* e_2. The main event e_2 of the second sentence *overlaps* with e_3, which is syntactically dominated by e_2.

In this paper, we focus on three subproblems of the temporal relation identification task as defined by TempEval-2: identifying temporal relations between (1) events and time expressions in the same sentence (*ET*); (2) two main events in consecutive sentences (*MM*); and (3) two events in the same sentence when one syntactically dominates another (*MS*), which is a new task introduced in TempEval-2.

4 Features

Surface features include the word tokens and stems of the words. In the TempEval-2 data, an event always consists of a single word token, but time expressions often consist of multiple tokens. We treat the entire string of words in a given time expression as a single feature.

Semantic features are the semantic attributes of individual events and time expressions described in Section 3. In this work, we use the gold-standard values for these features that were manually assigned by human annotators in the training and the test data.

Syntactic features include three features adopted from Bethard and Martin (2007): *gov-prep*, any prepositions governing the event or time expression (e.g., *'for'* in '*for ten years'*); *gov-verb*, the verb governing the event or time expression; *gov-verb-pos*, the part-of-speech (pos) tag of the governing verb. We also consider the pos tag of the word in the event and the time expression.

Lexical relations are the semantic relations between two events derived from VerbOcean (Chklovski and Pantel, 2004) and WordNet (Fellbaum, 1998). VerbOcean contains five types of relations (*similarity*, *strength*, *antonymy*, *enablement*, and *happens-before*) that commonly occur between pairs of verbs. To overcome data sparseness, we expanded the original VerbOcean database by calculating symmetric and transitive closures of key relations. With WordNet, a semantic distance between the associated tokens of each target event pair was computed.

5 Modeling Temporal Relations with Markov Logic

ML is a statistical relational learning framework that provides a template language for defining *Markov Logic Networks* (MLNs). A MLN is a set of weighted first-order clauses constituting a Markov network in which each ground formula represents a feature (Richardson and Domingos, 2006).

Our MLN consists of a set of formulae combining two types of predicates: *hidden* and *observed*. Hidden predicates are those that are not directly observable during test time. A hidden predicate is defined for each task: *relEventTimex* (temporal relation between an event and a time expression), *relMainEvents* (temporal relation between two main events), and *relMainSub* (temporal relation between a main and a dominated event). Observed predicates are those that can be fully observed during test time and represent each of the features described in Section 4.

The following is an example formula used in our MLN:

$$eventTimex(d, e, t) \wedge eventWord(d, e, w) \rightarrow relEventTimex(d, e, t, r) \quad (1)$$

The predicate *eventTimex(d, e, t)* represents the existence of a candidate pair of event *e* and time expression *t* in a document *d*. Given this candidate pair, formula (1) assigns weights to a temporal relation *r* whenever it observes a word token *w* in the given event from the training data. This formula is local because it considers only one hidden predicate (*relEventTimex*).

In addition to local formulae, we also define a set of global formulae to ensure consistency between local decisions:

$$relEventTimex(d, e_1, t, r_1) \wedge relEventTimex(d, e_2, t, r_2) \rightarrow relMainSub(d, e_1, e_2, r_3) \quad (2)$$

Formula (2) is global because it jointly concerns more than one hidden predicate (*relEventTimex* and *relMainSub*) at the same time. This formula ensures consistency between the predicted temporal relations r_1, r_2, and r_3 given a main event e_1, a syntactically dominated event e_2, and a time expression *t* shared by both of these events. Two additional global formulae (3) and (4) are similarly defined to ensure consistency as below.

$$relMainSub(d, e_1, e_2, r_3) \wedge relEventTimex(d, e_2, t, r_2) \rightarrow relEventTimex(d, e_1, t, r_1) \quad (3)$$

$$relMainSub(d, e_1, e_2, r_3) \wedge relEventTimex(d, e_1, t, r_1) \rightarrow relEventTimex(d, e_2, t, r_2) \quad (4)$$

6 Evaluation

To evaluate the proposed approach, we built and compared two models: one model (*NoLex*) used all of the features described in Section 4 except for the lexical relation features, and the other model (*Full*) included the full set of features. The features were generated using the Porter Stemmer and WordNet Lemmatizer in NLTK (Loper and Bird, 2002) and the Charniak Parser (Charniak, 2000). The semantic distance between two word tokens was computed using the path-similarity metric provided by NLTK. All of the models were constructed using Markov TheBeast (Riedel, 2008)

The feature set was optimized for each task on a held-out development data set consisting of approximately 10% of the entire training set (Table 1). Our previous work (Ha et al., 2010) observed that a local optimization approach that selects for each individual task (i.e., each hidden predicate in the given MLN) in isolation from the other tasks could harm the overall accuracy of a joint model because of resulting inconsistencies among individual tasks. In the new experiment described in this section, features were selected for each task to improve overall accuracy of the joint model combining all three tasks, similar to Yoshikawa et al. (2009).

Feature		Task		
		ET	MM	MS
Surface Features	*event-word*	√	√	√
	event-stem	√	√	√
	timex-word	√		
	timex-stem	√		
Semantic Attributes	*event-polarity*	√	√	√
	event-modal	√	√	√
	event-pos	√	√	√*
	event-tense	√	√	√
	event-aspect	√	√	√
	event-class	√	√	√
	timex-type	√		
	timex-value	√		
Syntactic Features	*pos*	√	√	√
	gov-prep	√	√	√
	gov-verb	√	√	√
	gov-verb-pos	√	√	√
Lexical Relations	*verb-rel*		√	√
	word-dist		√	

Table 1: Features used to model each task. *The feature is extracted only from the second event in the pair being compared.

Table 2 reports the resulting performance (*F1* scores) of the models. To isolate the potential effects of global constraints, we first compare the accuracies of the *Full* and the *NoLex* model, averaged from a ten-fold cross validation on the training data before global constraints are added. *Full* achieves relative 12% and 3% improvements over *NoLex* for temporal relation between events and time expressions (*ET*) and between two main events (*MM*), respectively. The improvement for *MM* was statistically significant ($p<0.05$) from a two-tailed paired *t*-test. Note that the *ET* task itself does not use lexical relation features but still achieves an improved result in *Full* over *NoLex*. This is an effect of joint modeling. There is a slight degradation (relative 2%) in the accuracy for temporal relations between main and syntactically dominated events (*MS*). Overall, *Full* achieves relative 5% improvement over *NoLex*. A similar trend of performance improvement in *Full* over *NoLex* was observed when the global formulae were added to each model. The second column (*Global Constraints*) of Table 2 compares the two models trained on the entire training set and tested on the test set after the global formulae were added. However, no statistical significance was found on these improvements. Compared to the state-

of-the-art results achieved by the TempEval-2 participants, *Full* achieves the same or better results on all three addressed tasks.

7 Conclusions

Temporal relations can be modeled with Markov Logic using a variety of features including lexical ontologies. Three tasks relating to the TempEval-2 data were addressed: predicting temporal relations between (1) events and time expressions in the same sentence, (2) two main events in consecutive sentences, and (3) two events in the same sentence when one syntactically dominates the other. An evaluation suggests that utilizing lexical relation features within a joint modeling framework using Markov Logic achieves state-of-the-art performance.

The results suggest a promising direction for future work. The proposed approach assumes events and time expressions are already marked in the data. To construct a fully automatic temporal relation identification system, the approach needs to be extended to include models that recognize events and time expressions in text as well as their semantic attributes. A data-driven approach similar to the one described in this paper may be feasible for this new modeling task. It will entail exploring a variety of features to further understand the complexity underlying the problem of temporal analysis of events.

Task	No Global Constraints		Global Constraints		State-of-the-art
	NoLex	Full	NoLex	Full	
Overall	0.60	0.63 (+5%)	0.59	0.61 (+3%)	NA
ET	0.52	0.58 (+12%)	0.62	0.65 (+5%)	0.63
MM	0.65	0.67 (+3%)*	0.52	0.56 (+8%)	0.55
MS	0.66	0.65 (- 2%)	0.66	0.66 (+0%)	0.66

Table 2. Performance comparison between models in *F1* score. *Statistical significance (p<0.05)

Acknowledgments

This research was supported by the National Science Foundation under Grant IIS-0757535.

References

S. Bethard and J. H. Martin. 2007. CU-TMP: Temporal relation classification using syntactic and semantic features. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 129-132, Prague, Czech Republic.

B. Boguraev and R. K. Ando. 2005. TimeML-compliant text analysis for temporal reasoning. In *Proceedings of the 19th International Joint Conference on Artificial intelligence*, pages 997-1003, Edinburgh, Scotland.

N. Chambers and D. Jurafsky. 2008. Jointly combining implicit constraints improves temporal ordering. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 698-706, Honolulu, HI.

E. Charniak. 2000. A maximum-entropy-inspired parser. In *Proceedings of the 1st North American Chapter of the Association for Computational Linguistics Conference*, pages 132-139, Seattle, WA.

T. Chklovski and P. Pantel. 2004. VerbOcean: Mining the web for fine-grained semantic verb relations. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 33-40, Barcelona, Spain.

E. Ha, A. Baikadi, C. Licata, and J. Lester. 2010. NCSU: Modeling temporal relations with Markov Logic and lexical ontology. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 341-344, Uppsala, Sweden.

E. Loper and S. Bird. 2002. NLTK: The natural language toolkit. In *Proceedings of ACL Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics*, pages 62-69, Philadelphia, PA.

J. Pustejovsky and M. Verhagen. 2009. SemEval-2010 task 13: Evaluating events, time expressions, and temporal relations. In *Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions*, pages 112-116, Boulder, CO.

S. Riedel. 2008. Improving the accuracy and efficiency of MAP inference for Markov Logic. In *Proceedings of the 24th Conference in Uncertainty in Artificial Intelligence*, pages 468-475, Helsinki, Finland.

M. Richardson and P. Domingos. 2006. Markov Logic networks. *Machine Learning*, 62(1):107-136.

M. Verhagen, R. Gaizauskas, F. Schilder, M. Hepple, G. Katz, and J. Pustejovsky. 2007. Semeval-2007 task 15: Tempeval temporal relation identification. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 75-80, Prague, Czech Republic.

K. Yoshikawa, S. Riedel, M. Asahara, and Y. Matsumoto. 2009. Jointly identifying temporal relations with Markov Logic. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 405-413, Suntec, Singapore.