# NCSU: Modeling Temporal Relations with Markov Logic and Lexical Ontology

Eun Young HaAlok BaikadiCarlyle LicataDepartment of Computer Science

James C. Lester

North Carolina State University Raleigh, NC, USA

{eha,abaikad,cjlicata,lester}@ncsu.edu

#### Abstract

As a participant in TempEval-2, we address the temporal relations task consisting of four related subtasks. We take a supervised machine-learning technique using Markov Logic in combination with rich lexical relations beyond basic and syntactic features. One of our two submitted systems achieved the highest score for the Task F (66% precision), untied, and the second highest score (63% precision) for the Task C, which tied with three other systems.

## **1** Introduction

Time plays a key role in narrative. However, correctly recognizing temporal order among events is a challenging task. As a follow-up to the first TempEval competition, TempEval-2 addresses this challenge. Among the three proposed tasks of TempEval-2, we address the temporal relations task consisting of four subtasks: predicting temporal relations that hold between events and time expressions in the same sentence (Task C), events and the document creation time (Task D), main events in adjacent sentences (Task E), and main events and syntactically dominated events, such as those in subordinated clauses (Task F). We are primarily concerned with Task C, E, and F, because D is not relevant to our application domain.<sup>1</sup> However, rather than eliminating Task D altogether, we build a very simple model for this task by using only those features that are shared with other task models (i.e., the document creation time data are not used because none of the other task models need them as features). It was expected that this approach would support more interesting comparisons with other systems that take a more sophisticated approach to the task. Further, we experiment with a joint modeling technique to examine if the communication with other task models brings a boost to a performance of the simple model.

Taking a supervised machine-learning approach with Markov Logic (ML) (Richardson and Domingos, 2006), we constructed two systems, NCSU-INDI and NCSU-JOINT. NCSU-INDI consists of four independently trained classifiers, one for each task, whereas NCSU-JOINT models all four tasks jointly. The choice of ML as learning technique for temporal relations is motivated both theoretically and practically. Theoretically, it is a statistical relational learning framework that does not make the *i.i.d.* assumption for the data. This is a desirable characteristic for complex problems such as temporal relation classification, as well as many other natural language problems, in which the features representing a given problem are often correlated with one another. Practically, ML allows us to build both individual and joint models in a uniform framework; individual models can be easily combined together into a joint model with a set of global formulae governing over them.

In previous work (Yoshikawa et al., 2009), ML was successfully applied to temporal relation classification task. Our approach is different from this work in two primary respects. First, we introduce new lexical relation features derived from English lexical ontologies. Second, our model addresses a new task introduced in TempEval-2, which is to identify temporal relations between main and syntactically dominated events in the same sentence. We also employ phrase-based syntactic features (Bethard and

<sup>&</sup>lt;sup>1</sup> Our application domain concerns analysis of narrative stories written by middle school students, with the analysis being conducted a single story at a time.

Martin 2007) rather than dependency-based syntactic features.

# 2 Features

We consider three types of features: basic, syntactic, and lexical relation features. Basic features represent the information directly available from the original data provided by the task organizer; syntactic features are extracted from syntactic parses generated by Charniak parser (Charniak, 2000); and lexical semantic relations that are derived from two external lexical databases, VERBOCEAN (Chklovski and Pantel, 2004) and WordNet (Fellbaum, 1998).

## 2.1 Basic Features

Basic features include the word tokens, stems of the words, and the manually annotated attributes of events and time expressions. In the TempEval-2 data, an event always consists of a single word token, but time expressions often consist of multiple tokens. We treat each word in time expressions as a different feature. For example, two word features, 'this' and 'afternoon', are extracted from a given time expression 'this afternoon'. Stemming is done with the Porter Stemmer in NLTK (Loper and Bird, 2002). The value attributes of time expressions are treated as symbolic features, rather than being decomposed into actual integer values representing dates and times.

## 2.2 Syntactic Features

Our syntactic features draw upon the features previously shown to be effective for temporal relation classification (Bethard and Martin, 2007), including the following:

- pos: the part-of-speech (pos) tags of the event and the time expression word tokens, assigned by Charniak parser.
- 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, similar to *gov*-*prep*.
- gov-verb-pos: the *pos* tag of the governing verb.

We also investigate both full and partial syntactic paths between a pair of event and time expressions, but including these features does not improve the classification results on our development data set.

#### 2.3 Lexical Relation Features

VERBOCEAN is a graph of semantic relations between verbs. There are 22,306 relations between 3,477 verbs that have been mined using Google searches for lexico-syntactic patterns. VERBOCEAN contains five different types of relations (Table 1). Verbs are stored in the lemmatized forms and senses are not disambiguated. A connection between two verbs indicates that the relation holds between some senses of the verbs.

VERBOCEAN'S database is presented as a list of verb pair relations, along with a confidence score. Both the transitive and symmetric closure over the relations were taken before storage in a SQLite database for queries. The transitive closure was calculated using the Warshall algorithm (Agrawal and Jagadish, 1990). The confidence score for the new arc was calculated as the average of the two constituents. The symmetric closure was calculated using a simple pass. The confidence score is the same as the reflected edge for symmetric relations. A set of VER-BOCEAN features were calculated for each target event pair within each of the temporal relations tasks. Each verb was lemmatized using the WordNet lemmatizer in NLTK before being compared against the database. Rather than focusing only on HAPPENS-BEFORE relation as in Mani et al. (2006), we consider all five verb relations in two different versions, unweighted and weighted. The unweighted version is a binary feature indicating the existence of an arc between the two target verbs in VERBOCEAN. In the weighted version, the existence of an arc is weighted by the associated confidence score.

In addition to VerbOcean, WordNet was used for its conceptual relations. WordNet is a large lexical database, which contains information on verbs, nouns, adjectives and adverbs, grouped into hierarchically organized cognitive synonym

Relation	Example
Similarity ‡†	produce :: create
STRENGTH †	wound :: kill
ANTONYMY ‡	open :: close
ENABLEMENT	fight :: win
HAPPENS-BEFORE $\dagger$	buy :: own

Table 1: Semantic relations between verbs in VERBOCEAN ( $\ddagger$  and  $\ddagger$  denotes symmetric and transitive closure, respectively, holds for the given relation)<sup>2</sup>

<sup>&</sup>lt;sup>2</sup> Examples are taken from

http://demo.patrickpantel.com/Content/Verbocean/.

sets (*synsets*). WordNet was accessed through the WordNetCorpusReader module of NLTK. For each target event pair within each of the temporal relations tasks, a semantic distance between the associated tokens was computed using the path-similarity metric present within the API. The *synset* chosen was simply the first synset returned by the reader. Similar to the VER-BOCEAN features, we consider both unweighted and weighted versions of the feature.

## 3 The Systems

ML is a probabilistic extension of first-order logic that allows formulae to be violated. It assigns a weight to each formula, reflecting the strength of the constraint represented by the formula. A *Markov logic network (MLN)* is a set of weighted first-order clauses, which, together with constants, defines a Markov network. We constructed two systems, NCSU-INDI and NCSU-JOINT using an off-the-shelf tool for ML (Riedel, 2008).

## 3.1 NCSU-INDI

NCSU-INDI consists of four independently trained MLNs, one for each task. Each MLN is defined by a set of local formulae that are conjunctions of predicates representing the features. An example local formula used for Task C is

 $eventTimex(e, t) \land eventWord(e, w) \\ \rightarrow relEventTimex(e, t, r)$ (1)

If a pair of event e and time expression t exists and the event consists of a word token w, formula (1) assigns a temporal relation t to the given pair of e and t with some weights.

For each task, the features described in Section 2 were examined on a held-out development data set (about 10% of the training data) for their effectiveness in predicting temporal relations and removed if they do not improve the results. Table 2 lists the features actually used for the tasks. Interestingly, none of the time expression features were effective on the development data.

## 3.2 NCSU-JOINT

As well as the local formulae from the four local MLNs, a set of global formulae are added to NCSU-JOINT as hard constraints to ensure the consistency between the classification decisions of local MLNs. For example, formula (2) ensures that if an event e1 happens before the document creation time (dct) and another event e2 happens

Feature		Task			
		С	D	Е	F
Event	event-word			$\sqrt{1}{e^2}$	$\sqrt{1}_{e1,e2}$
	event-stem			$\sqrt{1}_{e1.e2}$	$\sqrt{1}_{e1,e2}$
Event	event-polarity			$\sqrt{1}_{e1.e2}$	$\sqrt{1}{e_{1,e_{2}}}$
Attribute	event-modal			$\sqrt{1.e^{2}}$	$\sqrt{1}{e_{1,e2}}$
	event-pos			$\sqrt{1}_{e1.e2}$	$\sqrt{1}{e^2}$
	event-tense			$\sqrt{1.e^{2}}$	$\sqrt{1}{e_{1,e2}}$
	event-aspect			$\sqrt{1}$ e1.e2	$\sqrt{1}{e_{1,e_2}}$
	event-class			$\sqrt{1}_{e1.e2}$	$\sqrt{1}{e_{1,e2}}$
Timex	timex-word				
	timex-stem				
Timex	timex-type				
Attribute	timex-value				
Syntactic	pos		$\sqrt{e}$	$\sqrt{1.e^{2}}$	
Parse	gov-prep	$\sqrt{e,t}$	$\sqrt{e}$	$\sqrt{e_{1.e2}}$	$\sqrt{1}{e_{1,e_2}}$
	gov-verb	$\sqrt{v_{e,t}}$	$\sqrt{e}$	$\sqrt{e_{1.e2}}$	$\sqrt{1}$ e1,e2
	gov-verb-pos	$\sqrt{e,t}$		$\sqrt{1.e^{2}}$	$\sqrt{1}{e_{1,e2}}$
Verb-	verb-rel				
Ocean	verb-rel-w				
WordNet	word-dist				
	word-dist-w				

Table 2: Features used for each task (subscripts e and t mean event and time expression, respectively. Subscripts e1 and e2 mean the first and the second main events for the Task E and the main and the syntactically dominated events for the Task F, respectively)

after *dct*, then *e1* happens before *e2* and vice versa.

 $relDctEvent(e1,t,BEFORE) \land relDctEvent(e2,t, AFTER) \\ \rightarrow relEvents(e1, e2, BEFORE)$ (2)

A set of global constraints is defined between Tasks C and F, D and F, as well as D and E, respectively.

# 4 Results and Discussion

The predicted outputs from our systems exhibit mixed results. NCSU-INDI achieves the highest precision score on the test data for Task F by a relatively large margin (6%) from the secondplace system, as well as the second highest precision score on Task C, tied with three other systems. Given the encouraging result for Task F, we would preliminarily conclude that the VE-BOCEAN relations are effective predictors of temporal relations between main and syntactically dominated events. However, the same system does not achieve the same level of accuracy

System	Precision / Recall (%)					
	Task C	Task D	Task E	Task F		
NCSU-INDI	63/63	68/68	48/48	66/66		
NCSU-JOINT	62/62	21/21	51/51	25/25		

Table 3: Accuracy of the systems on each task

for Task E, even though it is closely related to Task F. The major difference between the models of Task E and F is that the Task E model uses weighted VERBOCEAN relations along with a WordNet feature, while the Task F model uses unweighted VERBOCEAN relations without the WordNet feature. We suspect these two features might negatively impact the classification decisions on the test data, even though they preliminarily appeared to be effective predictors on the development data.

NCSU-JOINT also yields mixed results. The performance on both Task D and F dramatically drops with the joint modeling approach, while there is a modest improvement on Task E. Manual examination of the results on the test data revealed that the majority of the relations in Task D and F were classified as *OVERLAP*, which may be due to overly strict global constraints; rather than violating global constraints, the system resorted to rather neutral predictions.

## 5 Conclusions

Temporal event order recognition is a challenging task. Using basic, syntactic, and lexical relation features, we built two systems with ML: NCSU-INDI models each subtask independently, and NCSU-JOINT models all four tasks jointly. NCSU-INDI was most effective in predicting temporal relations between main events and syntactically dominated events (66% precision), as well as temporal relations between time expressions and events (63% precision). Future directions include conducting a more rigorous examination of the predictive power of the features, as well as the impact of global formulae for the joint model.

#### Acknowledgments

This research was supported by the National Science Foundation under Grant IIS-0757535. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

# References

- R. Agrawal, S. Dar, and H. V. Jagadish. 1990. Direct transitive closure algorithms: design and performance evaluation. *ACM Transactions on Database Systems*, 15(3): 427-458.
- 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.
- E. Charniak. 2000. A maximum-entropy-inspired parser. In *Proceedings of the 1<sup>st</sup> North American chapter of the Association for Computational Linguistics conference*, pages 132-139, Seattle, WA.
- Y. Cheng, M. Asahara, and Y. Matsumoto. 2007. NAIST.Japan: Temporal relation identification using dependency parsed tree. In *Proceedings of the* 4<sup>th</sup> International Workshop on Semantic Evaluations, pages 245-248, Prague, Czech Republic.
- T. Chklovski and P. Pantel. 2004.VerbOcean: Mining the Web for Fine-Grained Semantic Verb Relations. In *Proceedings of Conference on Empirical Methods in Natural Language Processing*, pages 33-40, Barcelona, Spain.
- C. Fellbaum. 1998. *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.
- 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.
- I. Mani, M. Verhagen, B. Wellner, C. M. Lee, and J. Pustejovsky. 2006. Machine learning of temporal relations. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pages 753-760, Sydney, Australia.
- M. Richardson and P. Domingos. 2006. Markov Logic Networks. *Machine Learning*, 62(1): 107-136.
- 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.
- 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.