Enhancing Temporal Relation Classification by Features Extracted from a Syntactic Parser

Natsuda Laokulrat[1] Makoto Miwa[2] Yoshimasa Tsuruoka[1] Takashi Chikayama[1]
[1]School of Engineering, The University of Tokyo, Japan
[2]National Centre of Text Mining, the University of Manchester, UK
{natsuda, miwa, tsuruoka, chikayama}@logos.t.u-tokyo.ac.jp

1 Introduction

Temporal relation classification [1] refers to the task of identifying temporal relationship between a pair of events. It is one of the keys to deep language understanding and could help advance other NLP applications such as textual entailment, document summarization, and question answering.

Previous research on temporal relations used hand-coded rules. With the emerging of the annotated corpora, the Timebank corpus, machine learning approaches with different sets of features have been proposed.

This paper introduces new types of features extracted from a syntactic parser for classifying temporal relations between events in newswire documents. We propose to use paths between event words in syntactic trees and path lengths as features for temporal relation classification.

We first describe temporal relations and the corpora in Section 2. Section 3 describes briefly the related work. The features used for classification are presented in Section 4. Then, the evaluation and results are followed in Section 5. Lastly, we give the discussion and conclusion in Section 6 and 7.

2 Temporal relation classification

Temporal relation classification [1] is the task to classify relations between temporal entities (events or temporal expression). Possible pairs of entities are: (i) event and temporal expressions in the same sentence, (ii) event and document creation time, (iii) main events of consecutive sentences and (iv) pairs of events in the same sentence. In this paper, we only focus on (iii) and (iv).

2.1 Timebank corpus

The Timebank corpus [2] is annotated following TimeML specification [3] to indicate events, times, and temporal relations. It also provides five attributes, namely, *class, tense, aspect, modality,* and *polarity*, associated with each event. An example of the annotated events are shown below.

It wasn't until twenty years after the first astronauts were <EVENT eid="e6" class="OCCURRENCE"> chosen </EVENT> that NASA finally < EVENT eid="e7" class= "OCCURRENCE"> included </EVENT> six women, and they were all scientists, not pilots.

<MAKEINSTANCE eventID="e6" eiid="ei265" tense="PAST" aspect="NONE" polarity="POS" />

<MAKEINSTANCE eventID="e7" eiid="ei266" tense="PAST" aspect="NONE" polarity="POS" />

There is no modal word in the sentence so the attribute *modality* does not appear.

2.2 Temporal relations

The related work mentioned in Section 3 used subsets of TimeML temporal relations. However, we used a complete set of the relations, as described in [3], which has 14 types of temporal relations including *before, after, includes, is included, during, during inv, simultaneous, iafter (immediately before), ibefore (immediately after), identity, begins, ends, begun by, ended by.* Given the example mentioned above, the temporal relation is annotated as below.

<TLINK lid="l6" relType="AFTER" eventInstanceID="ei266" relatedToEventInstance="ei265" />

From the annotated relation above, the event **included (e7)** happens *after* the event **chosen (e6)**.

3 Related work

Chambers et. al [4] proposed a two-stage machine learning architecture using Support Vector Machine

(SVM) that automatically extracts features from raw text. In the first stage, temporal attributes of events, such as tense, aspect, class, are learned, and temporal relations are learned in the second stage. However, the system does not extract events automatically so, the events annotated in the corpus are used.

Llorens et a. [5] built a system to extract temporal information from raw texts focusing on semantic information. The system learns Conditional Random Field (CRF) models from training data to recognize events and uses semantic roles as one of the features for supervised learning to classify event temporal relations.

UzZaman et al. [6] applied around one hundred of hand-coded rules extraction to extract events and features. Then they used Markov Logic Networks for temporal relation classification.

4 Classifier

We use a machine learning approach for temporal relation classification. In the same way as described in [4], we do not extract events from raw text. We read the annotated event words from the corpus and use them directly, since our goal is to evaluate the features extracted from a syntactic parser.

4.1 Basic features

The features includes the event words (including two words before/after the event), part of speech tags and lemmas of the event words (including two words before/after the event), Wordnet synsets, and the appearance of auxiliaries and modals (before the event word). In addition, the five attributes associated with events *class, tense, aspect, modality,* and *polarity*, are used as features as well. In the case that the events are in different sentences, the order (precedence) of the event pair is also used as a feature.

We directly read the attributes as tagged in the corpus, different from the related work [4] that they automatically determine the attributes by SVM.

4.2 Features extracted from a syntactic parser

In this paper, in the case that the events are in the same sentence, we extract two new types of features from a syntactic parser, namely, paths between the event words in the syntactic parse tree, and up/down lengths of paths. We use 3-grams of paths as features instead of full paths since they are too sparse.

For a better understanding, an example is shown in Figure 1 below. In this case, the path between the event words, *estimates* and *worth*, is *VBZ-UP, VX-UP, VP-UP, VP-UP, VP, PP-DOWN, PX-DOWN, IN-DOWN.*

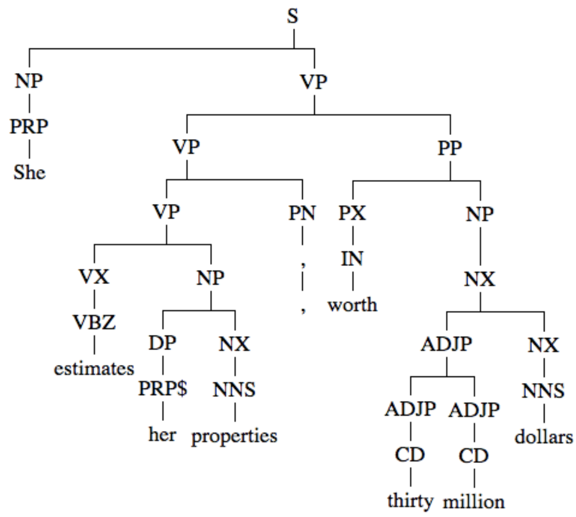

Figure 1: Syntactic parse tree

Relation	Number	Relation	Number
before	2117	iafter	43
after	827	ibefore	60
includes	309	identity	746
is included	418	begins	45
during	80	ends	17
during inv	0	begun by	42
simultaneous	523	ended by	62

Table 1: Number of occurrences of each relation type

So, the 3-grams of the path are {*VBZ-UP-VX-UP-VP-UP, VX-UP-VP-UP-VP-UP, VP-UP-VP-UP-VP, VP-UP-VP-PP-DOWN, VP-PP-DOWN-PX-DOWN, PP-DOWN-PX-DOWN-IN-DOWN*}. The up/down path lengths are 4 *(VBZ-UP, VX-UP, VP-UP, VP-UP)* and 3 *(PP-DOWN, PX-DOWN, IN-DOWN)* respectively.

5 Evaluation and results

We use the Enju parser [7] for parsing syntactic trees. For the classifier, we use the LIBLINEAR [8] and configure it to work as a L2-regularized logistic regression classifier.

5.1 Data

We used the TimeBank corpus and the AQUAINT corpus provided by TempEval-3 task at the Semeval-2012 competition. There are 256 newswire articles containing 5,289 event pairs in total.

The number of occurrences of each relation type is shown in the Table 1. We can see that the relation type *before* is the majority of the event pairs, while the relation type *during inv* does not occur at all.

Features	Accuracy	P	R	F
basic	57.61	57.18	28.89	38.39
basic + new	59.82	59.17	30.39	40.16

Table 2: F-score based evaluation results

Features	P	R	F
basic	28.42	31.31	29.79
basic + new	28.48	31.24	29.80

Table 3: Graph-based evaluation results

5.2 F-score based evaluation

Table 2 shows the classification performance based on 10-fold cross validation. We got about 2.41% improvement in classification accuracy and 1.78% improvement in F-score.

Table 4 and 5 show the confusion matrices of the classification results without and with using syntactic tree features respectively.

5.3 Graph-based evaluation

We also evaluated the system using graph-based evaluation metric. It uses temporal closure to reward relations instead of direct comparison, since it is possible to express a same temporal relation in different ways. The detail can be found in [9].

Table 3 shows the results by graph-based evaluation. We can see that it has no difference in precision, recall, and F-score.

6 Discussion

As we can see from Table 1, the distribution of the training data is very biased. Some relation types appear in the corpus less than 50 times while some of them appear more than 500 times. This is probably one of the major reasons that we could not obtain a high classification performance.

According to the results shown in Table 2 and 3, the features extracted from syntactic trees are not effective. Possibly the reason is that the syntactic tree features that we used do not imply a temporal relation of events in the sentence. For instance, the two following sentences give exactly the same path of the event words in the syntactic parse trees.

John **saw** *Mary before the* **meeting.**

John **saw** *Mary after the* **meeting.**

By using semantic structures, such as predicate-argument structure, or different methods of extracting features from syntactic trees may improve the performance.

7 Conclusion

We have presented a machine learning approach for classifying temporal relations. We proposed two new features and evaluated the results using F-score and graph-based evaluation. The new features extracted from the syntactic parser does not make a significant improvement in the performance of the classification. We will continue investigating and trying more new features from the Enju parser and other resources.

References

[1] N. UzZaman, H. Llorens, J. F. Allen, L. Derczynski, M. Verhagen, and J. Pustejovsky, “Tempeval-3: Evaluating events, time expressions, and temporal relations.” 2012.

[2] J. Pustejovsky, P. Hanks, R. Sauri, A. See, R. Gaizauskas, A. Setzer, D. Radev, B. Sundheim, D. Day, L. Ferro, and M. Lazo, “The TIMEBANK corpus,” in *Proceedings of Corpus Linguistics 2003*, Lancaster, Mar. 2003, pp. 647–656.

[3] R. S. and Jessica Littman, R. Gaizauskas, A. Setzer, and J. Pustejovsky, “Timeml annotation guidelines, version 1.2.1,” 2006.

[4] N. Chambers, S. Wang, and D. Jurafsky, “Classifying temporal relations between events,” in *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, ser. ACL ’07, 2007, pp. 173–176.

[5] H. Llorens, E. Saquete, and B. Navarro, “Tipsem (english and spanish): Evaluating crfs and semantic roles in tempeval-2,” in *Proceedings of the 5th International Workshop on Semantic Evaluation*, July 2010, pp. 284–291.

[6] N. UzZaman and J. F. Allen, “Trips and trios system for tempeval-2: Extracting temporal information from text,” in *Proceedings of the 5th International Workshop on Semantic Evaluation*, ser. SemEval ’10, 2010, pp. 276–283.

[7] Y. Miyao and J. Tsujii, “Feature forest models for probabilistic hpsg parsing,” *Comput. Linguist.*, vol. 34, no. 1, pp. 35–80, Mar. 2008.

[8] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, “Liblinear: A library for large linear classification,” *J. Mach. Learn. Res.*, vol. 9, pp. 1871–1874, Jun. 2008.

[9] N. UzZaman and J. F. Allen, “Temporal evaluation,” in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers - Volume 2*, ser. HLT ’11, 2011, pp. 351–356.

		Predicted												
	Relation type	before	simul.	after	inc.	begins	iden.	is inc.	ibefore	during	begun by	ended by	ends	iafter
	before	1774	72	141	31	0	65	28	2	0	2	2	0	0
	simul.	139	188	72	11	0	95	11	1	2	0	4	0	0
	after	328	84	335	7	0	49	22	0	0	0	1	0	1
	inc.	140	27	30	58	0	49	5	0	0	0	0	0	0
	begins	22	2	14	0	3	1	3	0	0	0	0	0	0
	iden.	120	44	43	11	0	502	25	0	0	0	1	0	0
Gold	is inc.	140	29	56	9	0	46	135	0	0	0	1	0	2
	ibefore	23	4	5	1	0	2	0	25	0	0	0	0	0
	during	39	8	12	2	0	7	5	1	6	0	0	0	0
	begun by	16	1	9	2	0	10	2	0	0	2	0	0	0
	ended by	26	5	14	0	0	5	1	0	0	0	11	0	0
	ends	6	1	7	0	0	0	3	0	0	0	0	0	0
	iafter	7	6	11	2	0	5	3	1	0	0	0	0	8

Table 4: Confusion matrix (no syntactic tree features, simul.: simultaneous, inc.: includes, iden.: identity, is inc.: is included)

		Predicted												
	Relation type	before	simul.	after	inc.	begins	iden.	is inc.	ibefore	during	begun by	ended by	ends	iafter
	before	1783	70	138	27	0	59	34	2	0	1	3	0	0
	simul.	135	194	72	11	0	97	12	0	0	0	2	0	0
	after	283	66	387	8	0	51	30	0	0	0	1	0	1
	incl.	134	29	29	55	0	54	8	0	0	0	0	0	0
	begins	23	1	11	0	3	5	2	0	0	0	0	0	0
	iden.	102	40	42	14	0	530	17	0	0	0	1	0	0
Gold	is inc.	122	26	52	6	1	51	157	0	1	0	0	0	2
	ibefore	20	4	5	2	0	6	0	23	0	0	0	0	0
	during	38	8	11	3	0	6	6	1	7	0	0	0	0
	begun by	14	2	14	1	0	6	3	0	0	2	0	0	0
	ended by	21	7	13	0	0	4	1	0	0	0	16	0	0
	ends	7	1	5	0	0	2	2	0	0	0	0	0	0
	iafter	5	8	13	1	0	4	3	2	0	0	0	0	7

Table 5: Confusion matrix (with syntactic tree features, simul.: simultaneous, inc.: includes, iden.: identity, is inc.: is included)