Using Syntactic Dependencies and WordNet Classes for Noun Event Recognition

Yoonjae Jeong and Sung-Hyon Myaeng

Korea Advanced Institute of Science and Technology
291 Daehak-ro (373-1 Guseong-dong), Yuseong-gu, Daejeon 305-701,
Republic of Korea

{hybris, myaeng}@kaist.ac.kr

Abstract. The goal of this research is to devise a method for recognizing TimeML noun events in a more effective way. TimeML is the most recent annotation scheme for processing the event and temporal expressions in natural language processing fields. In this paper, we argue and demonstrate that the dependencies and the deep-level WordNet classes are useful for recognizing events. We formulate the event recognition problem as a classification task using various features including lexical semantic and dependency-based features. The experimental results show that our proposed method outperforms significantly a state-of-the-art approach. Our analysis of the results demonstrates that the dependencies of direct object and the deep-level WordNet hypernyms play pivotal roles for recognizing noun events.

Keywords: Event Recognition, TimeML, TimeBank, WordNet, Natural Language Processing, Machine Learning

1 Introduction

Automatic event extraction from text is one of the important parts in text mining field. There are two types of definitions for events. In the area of topic detection and tracking (TDT), an event is defined as an instance of a document level topic describing something that has happened (Allan 2002). On the other hand, the information extraction (IE) field uses a more fine-grained definition of an event, which is often expressed by a word or phrase in a document. In TimeML, a recent annotation scheme, events are defined as situations that happen or occur and expressed by verbs, nominalizations, adjectives, predicative clauses or prepositional phrases (Pustejovsky, Castaño, et al. 2003). In this paper, we follow the view of IE, and focus on recognition of TimeML events.

Previous studies have proposed different approaches for automatic recognition of events, most notably adopting machine learning techniques based on lexical semantic classes and morpho-syntactic information around events (Bethard and Martin 2006; Boguraev and Ando 2007; Llorens, Saquete, and Navarro-Colorado 2010; March and Baldwin 2008; Saurí et al. 2005). In recognizing events, some of the past work used top level WordNet classes (Fellbaum 1998) to represent the meanings of events. It turns out, however, that such WordNet classes used as lexical semantic features are not sufficient. When WordNet hypernyms within the top four levels (Llorens, Saquete, and Navarro-Colorado 2010) or some selected classes (Bethard and Martin 2006) were used, they could not represent events well. For example, the WordNet class *event* is a representative level-4 class expressing events, but just 28.46% of *event* nouns, i.e., hyponyms of WordNet *event* class occurring in the TimeBank 1.2 corpus are annotated as TimeML events. TimeBank is a corpus containing news articles annotated based on the TimeML scheme (Pustejovsky, Hanks, et al. 2003).

Events can be recognized in different part-of-speech. In this paper, we focus on noun event recognition because the previous approaches showed low performances for recognizing noun events although nouns cover about 28% of all the events, according to our data analysis. For the problem of recognizing event nouns, we propose a method of using dependency-based features that exist between an event noun and its syntactically related words. In addition, we chose to use deeper level WordNet classes than those at the top-4 levels as in the previous work. We show that our proposed method outperforms the previous work by running experiments.

The rest of the paper is organized as follows. Section 2 introduces TimeML and TimeBank corpus as a representation and annotation scheme and as a test bed, respectively. It is followed by a discussion of related work for TimeML-based event recognition in Section 3. Section 4 presents our event recognition method using the deep-level WordNet classes and the dependency-based features. We then discuss our experiments and results in Section 5. Finally, the last section presents our conclusions.

2 TimeML and TimeBank Corpus

TimeML is a robust specification language for event and temporal expressions in natural language (Pustejovsky, Castaño, et al. 2003). It was first announced in 2002 in an extended workshop called TERQAS (Time and Event Recognition for Question Answering System)[1]. It addresses four basic problems:

1. Time stamping of events (identifying an event and anchoring it in time)
2. Ordering events with respect to one another (lexical versus discourse properties of ordering)
3. Reasoning with contextually underspecified temporal expressions (temporal functions such as "*last week*" and "*two weeks before*")
4. Reasoning about the persistence of events (how long does an event or the outcome of an event last)

Fig. 1. Four problems in event and temporal expression markup (Hobbs and Pustejovsky 2003)

There are four major data components in TimeML: EVENT, TIMEX3, SIGNAL, and LINK (Pustejovsky et al. 2007). TimeML considers event as a term for situations

[1] http://www.timeml.org/site/terqas/index.html

that happen or occur or elements describing states or circumstances in which something obtains or holds the truth (EVENT). Temporal expressions in TimeML are marked up with the TIMEX3 tags referring to dates, durations, sets of times, etc. The tag SIGNAL is used to annotate function words, which indicates how temporal objects (event and temporal expressions) are to be related to each other. The last component, LINK, describes the temporal (TLINK), subordinate (SLINK), and aspectual relationship (ALINK) between temporal objects.

Fig. 2 shows an example of TimeML annotation. For an event "*teaches*", its type is kept in class attribute, and its tense and aspect information is tagged in MAKEINSTANCE. The normalized value of temporal expressions "*3:00*" and "*November 22, 2004*" are stored in value attribute in TIMEX3 tag. The signal words "*at*" and "*on*" make links between events and temporal expressions through TLINK tags.

```
John
<EVENT eid="e1" class="OCCURRENCE"> teaches </EVENT>
<MAKEINSTANCE eiid="ei1" eventID="e1" tense="PRESENT"
    aspect="NONE" />
<SIGNAL sid="s1"> at </SIGNAL>
<TIMEX3 tid="t1" type="TIME" value="2004-11-22T15:00"
    temporalFunction="TRUE" anchorTimeID="t2"> 3:00
    </TIMEX3>
<SIGNAL sid="s2"> on </SIGNAL>
<TIMEX3 tid="t2" type="DATE value="2004-11-22">
    November 22, 2004 </TIMEX3>.

<TLINK eventInstanceID="ei1" relatedToTime="t1"
    relType="IS_INCLUDED" signalID="s1"/>
<TLINK timeID="t1" relatedToTime="t2"
    reltype="IS_INCLUDED" signalID="s2"/>
```

Fig. 2. An example of TimeML annotation (Pustejovsky et al. 2007)

Among several corpora[2] annotated with TimeML, TimeBank is most well-known as it started as a proof of concept of the TimeML specifications. TimeBank 1.2 is the most recent version of TimeBank, annotated with the TimeML 1.2.1 specification. It contains 183 news articles and more than 61,000 non-punctuation tokens, among which 7,935 are events.

We analyzed the corpus to investigate on the distribution of PoS (Part of Speech)[3] for the tokens annotated as events. As shown in Table 1, most events are expressed in verbs and nouns. Sum of the two PoS types covers about 93% of all the event tokens, which is split into about 65% and 28% for verb and nouns, respectively. The percentages for cardinal numbers and adjectives are relatively small. They usually express quantitative (e.g., "*47 %*") and qualitative (e.g., "*beautiful*") states. Adverbs and

---

[2] TimeML Corpora, http://timeml.org/site/timebank/timebank.html
[3] By Stanford PoS tagger, http://nlp.stanford.edu/software/tagger.shtml

prepositions indicate events when they appear in predicative phrases (e.g., “*he was here*” or “*he was on board*”).

Table 1. PoS distribution of event tokens

PoS tag	# Event	Coverage
VB (Verb)	5,171	65.17 %
NN (Noun)	2,183	27.51 %
CD (Cardinal Number)	279	3.52 %
JJ (Adjective)	223	2.81 %
RB (Adverb)	29	0.37 %
IN (Preposition)	46	0.58 %
Misc.	4	0.05 %
SUM	7,935	100.00 %

In finding verb events automatically from the TimeBank corpus, Llorens et al. (2010)’s work, a state-of-the-art approach, showed high effectiveness in terms of F1 (0.913). We note, however, its performance in recognizing noun events was just 0.584 in F1. This clearly indicates that noun even recognition, which is significant by itself, is a harder problem that needs to draw more attention and research.

3 Related Work

EVITA (Saur íet al. 2005) is the first event recognition tool for TimeML specification. It recognizes events by combining linguistic and statistical techniques. It uses manually encoded rules based on linguistic information as main features to recognize events. It also uses WorldNet classes to those rules for nominal event recognition, and checks whether the head word of noun phrase is included in the WordNet event classes. For sense disambiguation of nouns, it utilizes a Bayesian classifier trained on the SemCor corpus[4].

Boguraev and Ando (2007) analyzed the TimeBank corpus and presented a machine-learning based approach for automatic TimeML events annotation. They set out the task as a classification problem, and used a robust risk minimization (RRM) classifier (Zhang, Damerau, and Johnson 2002) to solve it. They used lexical and morphological attributes and syntactic chunk types in bi- and tri-gram windows as features.

Bethard and Martin (Bethard and Martin 2006) developed a system, STEP, for TimeML event recognition and type classification. They adopted syntactic and semantic features, and formulated the event recognition task as classification in the word-chunking paradigm. They used a rich set of features: textual, morphological, syntactic dependency and some selected WordNet classes. They implemented a Support Vector Machine (SVM) model based on those features.

Lastly, Llorens et al. (2010) presented an evaluation on event recognition and type classification. They added semantic roles to features, and built the Conditional Ran-

[4] http://www.gabormelli.com/RKB/SemCor_Corpus

dom Field (CRF) model to recognize events. They conducted experiments about the contribution of semantic roles and CRF and reported that the CRF model improved the performance but the effects of semantic role features were not significant. The approach achieved 82.4% in F1 in event recognition for the TimeBank 1.2 corpus. It is a state-of-the-art approach in TimeML event recognition and type classification.

4 Event Recognition

The main goal of our research is to devise an effective method for recognition of TimeML noun events. Our proposed method consists of three parts: preprocessing, feature extraction, and classification. The preprocessing part analyzes raw text to do tokenizing, PoS tagging, and syntactic parsing (dependency parsing). It is done by the Stanford CoreNLP package[5], which is a suite of natural language processing tools. Then, the feature extraction part converts the preprocessed data into the feature spaces. We explain the details of our feature extraction methods in Subsection 4.1. Finally, the classification part determines whether the given noun is an event or not using the MaxEnt classification algorithm.

4.1 Feature Sets

The feature sets to recognize events consist of three types: *Basic Features*, *Lexical Semantic Features*, *and Dependency-based Features*. The *Basic Features* are based on one of the TimeML annotation guidelines – prenominal noun is not annotated as events –, and the *Lexical Semantic Features* are the lemmas and all WordNet hypernyms of target nouns to be classified. Those hypernyms include the deep WordNet classes indicating the specific concept of nouns. The *Dependency-based Features* are adopted because syntactically related words tend to serve as important clues in determining whether or not a noun refers to an event.

Basic Features. *The Basic Features* include named entity (NE) tags and an indication of whether the target noun is prenominal or not. A personal name and a geographical location cannot be an event whereas prenominal nouns are not considered as events according to the TimeWML annotation guideline.

Lexical Semantic Features. The *Lexical Semantic Features* (*LS*) is the set of target nouns' lemmas and their all-depth WordNet semantic classes (i.e., hypernyms). Some nouns have high probabilities of indicating an event when they are included in a very specific WordNet classes. For example, a noun "*drop*" is always an event regardless of its context of a sentence. While the word sense-ambiguity problem arises in mapping a token to a synset in WordNet, we ignore the problem and simply use the WordNet hypernyms of all the senses.

---

[5] http://nlp.stanford.edu/software/corenlp.shtml

Dependency-based Features. We posit that nouns become events if they occur with a certain surrounding context, namely, syntactic dependencies. We use the words and their semantic classes related to the target noun through dependency relations. Four dependencies we consider are: direct object (OBJ), subject (SUBJ), modifier (MOD), and preposition (PREP).

- ***VB_OBJ type.*** A feature is formed with the governing verb, which has the OBJ relation with the target noun, and its hypernyms. In "... *delayed the game*...", for instance, the verb "*delay*" can describe the temporal state of its object noun, "*game*".
- ***VB_SUBJ type.*** It is the verb that has the SUBJ relation with the target noun and its hypernyms. For example, the verb "*occur*" indicates that the subject of the verb is an event because it actually occurs as in the definition of an event.
- ***MOD type.*** It refers to the dependent words and their hypernyms in MOD relation. This feature type is based on the intuition that some modifiers such as temporal expression reveal the noun it modifies has a temporal state and therefore is likely to be an event.
- ***PREP type.*** This is the preposition of a noun. Some prepositions such as "*before*" may indicate that the noun after them occurs at some specific time.

Sometimes, *Dependency-based Features* need to be combined with *Lexical Semantic Features* because a certain syntactic dependency may not be an absolute clue for an event by itself but only when it co-occurs with a certain lexical or semantic aspect of the target noun. As shown in Table 2, direct objects of "*report*" are not always events (about 32% are not events in the TimeBank corpus). However, then the direct object belongs to the WordNet *process* class, the target noun would be almost always an event. In this case, therefore, we need to use a combined feature.

Table 2. The *process* class as direct objects and its event ratio in TimeBank 1.2 corpus

Verb	Object (Noun)	# of Event (Ratio)
"*report*"	WordNet *process* class	14/14 (100.00%)
*	WordNet *process* class	153/325 (47.08%)
"*report*"	*	30/44 (68.18%)

[*] Indicates the any verbs or nouns

4.2 Classification

While the three different types of features make their own contributions in determining whether a noun is an event, their relative weights are all different. A strict classification algorithm categorizes the target nouns based on the weighted features.

We weight the features with Kullback-Leibler divergence (KL-divergence), which is a non-symmetric measure of the difference between two probability distributions (Kullback and Leibler 1951) and a popular weighting scheme in text mining. For a feature f, its weight is calculated using the formula in (1) where E and $\neg E$ are the distributions of event and non-event term. $P_E(f)$ and $P_{\neg E}(f)$ are the probabilities of f in E and $\neg E$, respectively.

$$W(f) = KL(E|\neg E) = P_E(f)\ln\frac{P_E(f)}{P_{\neg E}(f)} \tag{1}$$

Since we decided to use all the WordNet hypernyms as possible features, which cause the feature space too large to handle, we need to select more valuable ones from the candidate set. We use the weighing method using KL-divergence for this purpose and selected top 104,922 features because the cut-off value empirically showed the best performance in our preliminary experiment. We measured the performance when we applied top-*k* features, and it was maximized at k = 104,922.

For our classification algorithm, we considered four popular ones in machine learning: Naïve Bayes, Decision Tree (C4.5), MaxEnt, and SVM algorithms. Among them, the MaxEnt showed the best performance for our classification task. The packages we used are Weka (Witten, Frank, and Hall 2011) and Mallet machine learning tools (McCallum 2002).

5 Experiment

5.1 Comparison with Previous Work

We first evaluated the proposed method by comparing the previous work, whose result is shown in Table 3. We chose two baselines (Bethard & Martin 2006; Llorens et al. 2010) that were most recent ones using the TimeBank 1.2 corpus.

The proposed method shows an improvement of about 22% and 9% in terms of precision and recall than the state-of-the-art, respectively, the work of Llorens et al. Overall, the proposed method increased the F1 score by about 18% and 13% compared to the two baselines, respectively. The evaluation was done by 5-fold cross validation.

Our classifier used only 85,518 features within the top-8 WordNet classes among the 104,922 features mentioned in Section 4.2. In Section 5.3, we describe the cumulative level-8 features in detail.

Table 3. Comparison with the proposed method and previous works

Approach	Precision	Recall	F1
Bethard & Martin (2006)	0.729	0.432	0.543
Llorens et al. (2010)	0.727	0.483	0.584
Proposed Method	0.950	0.577	0.718

5.2 Contribution Analysis

We ran additional experiments to understand the roles of the individual feature types. In order to show relative importance of *Lexical Semantic Features* (*LS*), *De-*

pendency-based Features (*VB_OBJ*, *VB_SUBJ*, *MOD*, and *PREP types*), we measured performance changes caused by excluding one feature type at a time.

As shown in Table 4, *VB_OBJ* and *MOD* features are judged to be most important because the performance was decreased most significantly. The effects of the other features were not as great, but cannot be disregarded as they always contribute to the overall performance increase.

Table 4. Contributions of individual feature types

Feature Type	***Precision***	***Recall***	***F1***
ALL	0.950	0.577	0.718
- *LS*	0.958 (+0.8%)	0.561 (-1.6%)	0.708 (-1.0%)
- *VB_OBJ*	0.939 (-1.1%)	0.517 (-6.0%)	0.667 (-5.1%)
- *VB_SUBJ*	0.944 (-0.6%)	0.554 (-2.3%)	0.698 (-2.0%)
- *MOD*	0.941 (-0.9%)	0.524 (-5.3%)	0.673 (-4.5%)
- *PREP*	0.940 (-1.0%)	0.564 (-1.3%)	0.705 (-1.3%)

5.3 The Effect of Deep-level WordNet Classes

To investigate the effect of deep-level WordNet classes, we observed the performance changes incurred by increasing the cumulative WordNet depth within which features were generated. Depth fifteen, for example, means all the hypernyms of the matched word are considered as features. The results are presented in Fig. 3.

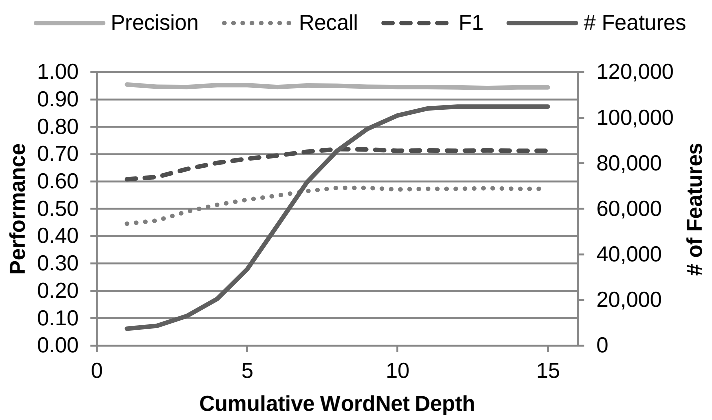

Fig. 3. Performance per cumulative WordNet depth

In this figure, the *y*-axis on the left represents the performance of event recognition in terms of precision, recall, or F1, and the *y*-axis on the right shows the numbers of features that vary when we apply the cumulative WordNet depth, which is represented by the *x*-axis.

Regardless of the depth of WordNet classes, the classifier reached the high precision over 0.9, but the recall varied quite widely. Recall increased with the rise of class depth, and it rose to the peak at top-8 level. The recall and F1-scores were 0.577 and 0.718, respectively.

The number of features increased continuously up to the level 13, but stayed the same beyond that. The number of features was 104,922, but the classifier used only 85,518 features at level 8 (where the performance was the best). From these results, we expect that there is a proper level of ontology to recognize events, which is shown to be level 8 in WordNet classes.

6 Conclusion

In this paper, we propose a TimeML noun event recognition method using syntactic dependency and WordNet classes and show their effect using the TimeBank collection. We chose to focus on noun events because they were recognized poorly in the previous research although they constitute about 28% of the events. The problem of recognizing such events was formulated as a classification task using lexical semantic (lemma and WordNet hypernyms) and dependency-based features.

Our experimental results show that the proposed method is better than the previous approach in recognizing TimeML noun events. The performance increase in terms of F1 measure is from 0.584 to 0.718, which we consider very significant. Through our analysis, we arrive at the conclusion that using dependency-based features and deep-level WordNet classes are important for recognizing events. We also showed that recall was increased significantly by using the hypernym features from lower depth of the WordNet hierarchy. A performance increase in recall for event detection, mainly due to the accurate handling of nouns and to effectiveness of the proposed classification method, would be translated into wider coverage of event-related triples in Semantic Web.

Although the proposed method showed encouraging results compared to the previous approaches, it still has some limitations. One issue is on the level of WordNet or an ontology for expanding the feature set because the current method requires too large feature space. Another one is word sense disambiguation that we ignored entirely in the current work. Although we obtained some performance increase with deeper levels, it’s not clear how much more gain we will get with sense disambiguation. We are currently working on these two issues.

Acknowledgment

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2011-0027292).

Reference

1. Allan, James, ed. 2002. *Topic Detection and Tracking: Event-based Information Organization*. Springer.
2. Bethard, Steven, and James H Martin. 2006. "Identification of Event Mentions and Their Semantic Class." In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, 146–154. Association for Computational Linguistics.
3. Boguraev, Branimir, and Rie Ando. 2007. "Effective Use of TimeBank for TimeML Analysis." In *Annotating, Extracting and Reasoning About Time and Events*, ed. Frank Schilder, Graham Katz, and James Pustejovsky, 4795:41–58. Springer Berlin / Heidelberg. doi:10.1007/978-3-540-75989-8_4.
4. Fellbaum, Christiane, ed. 1998. *WordNet: An Electronic Lexical Database*. The MIT Press.
5. Hobbs, Jerry, and James Pustejovsky. 2003. "Annotating and Reasoning About Time and Events." In *AAAI Technical Report SS-03-05*.
6. Kullback, Solomon, and Richard A. Leibler. 1951. "On Information and Sufficiency." *The Annals of Statistics* 22 (1): 79–86.
7. Llorens, Hector, Estela Saquete, and Borja Navarro-Colorado. 2010. "TimeML Events Recognition and Classification: Learning CRF Models with Semantic Roles." In *Proceedings of the 23rd International Conference on Computational Linguistics*, 725–733. Association for Computational Linguistics.
8. March, Olivia, and Timothy Baldwin. 2008. "Automatic Event Reference Identification." In *Proceedings of the Australasian Language Technology Workshop*, 6:79–87.
9. McCallum, Andrew Kachites. 2002. "MALLET: A Machine Learning for Language Toolkit." http://mallet.cs.umass.edu/.
10. Pustejovsky, James, José Castaño, Robert Ingria, Roser Saurí, Robert Gaizauskas, Andrea Setzer, and Graham Katz. 2003. "TimeML: Robust Specification of Event and Temporal Expressions in Text." In *Proceedings of the 5th International Workshop on Computational Semantics*.
11. Pustejovsky, James, Patrick Hanks, Roser Saurí, Andrew See, Robert Gaizauskas, Andrea Setzer, Dragomir Radev, et al. 2003. "The TIMEBANK Corpus." In *Proceedings of the Corpus Linguistics 2003 Conference*, 647–656.
12. Pustejovsky, James, Robert Knippen, Jessica Littman, and Roser Saurí. 2007. "Temporal and Event Information In Natural Language Text." In *Computing Meaning*, ed. Harry Bunt, Reinhard Muskens, Lisa Matthewson, Yael Sharvit, and Thomas Ede Zimmerman, 83:301–346. Springer Netherlands. doi:10.1007/978-1-4020-5958-2_13.
13. Saurí, Roser, Robert Knippen, Marc Verhagen, and James Pustejovsky. 2005. "Evita: a Robust Event Recognizer for QA Systems." In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, 700–707. Association for Computational Linguistics. doi:10.3115/1220575.1220663.
14. Witten, Ian H., Eibe Frank, and Mark A. Hall. 2011. *Data Mining: Practical Machine Learning Tools and Techniques*. 3rd ed. Morgan Kaufmann.
15. Zhang, Tong, Fred Damerau, and David Johnson. 2002. "Text Chunking Based on a Generalization of Winnow." *The Journal of Machine Learning Research* 2 (March): 615–637.