A Hybrid Approach for Event Extraction

Anup Kumar Kolya, Asif Ekbal, and Sivaji Bandyopadhyay

***Abstract*—Event extraction is a popular and interesting research field in the area of Natural Language Processing (NLP). In this paper, we propose a hybrid approach for event extraction within the TimeML framework. Initially, we develop a machine learning based system based on Conditional Random Field (CRF). But most of the *deverbal* event nouns are not correctly identified by this machine learning approach. From this observation, we came up with a hybrid approach where we introduce several strategies in conjunction with machine learning. These strategies are based on semantic role-labeling, WordNet and handcrafted rules. Evaluation results on the TempEval-2010 datasets yield the precision, recall and F-measure values of approximately 93.00%, 96.00% and 94.47%, respectively. This is approximately 12% higher F-measure in comparison with the best performing system of SemEval-2010.**

***Index Terms*—About Event, TimeML, Conditional Random Field, TempEval-2010, WordNet.**

I. Introduction

TEMPORAL information extraction is, nowadays, a popular and interesting research area of Natural Language Processing (NLP). Generally, events are described in different newspaper texts, stories and other important documents where events happen in time and ordering of these events are specified. One of the important tasks of text analysis clearly requires identifying events described in a text and locating these in time. This is also important in a wide range of NLP applications that include temporal question answering, machine translation and document summarization.

In the literature, relation identification based on machine learning approaches can be found in [1, 2, 3] and some of the TempEval-2007 participants [4]. Most of these works tried to improve classification accuracies through feature engineering.

The performance of any machine learning based system is often limited by the amount of available training data. Mani *et al*. [2] introduced a temporal reasoning component that greatly expands the available training data. The training set was increased by a factor of 10 by computing the closures of the various temporal relations that exist in the training data. They reported significant improvement of the classification accuracies on event-event and event-time relations. However, this has two shortcomings, namely feature vector duplication caused by the data normalization process and the unrealistic evaluation scheme. The solutions to these issues are briefly described in [5]. In TempEval-2007 task, a common standard dataset was introduced that involves three temporal relations. The participants reported F-measure scores for event-event relations ranging from 42% to 55% and for event-time relations from 73% to 80%.

In TempEval-2007, the event-event relations were not considered discourse-wide like [2, 5]. Here, the event-event relations are restricted to events within the maximum of two consecutive sentences. Thus, these two frameworks produce highly dissimilar results for solving the problem of temporal relation classification.

One most common trend to apply machine learning algorithm for temporal information extraction is to formulate temporal relation as an event paired with a time or another event, and to transform these into a set of feature values. In most of the previous attempts, researchers have used some popular machine learning techniques like Naive-Bayes, Decision Tree (C5.0), Maximum Entropy (ME) and Support Vector Machine (SVM). Machine learning techniques alone cannot always yield good accuracies. In order to achieve reasonable accuracy, some researchers [6] used hybrid approach, where a rule-based component was added with machine learning. The system [6] was designed in such a way that they can take the advantage of rule-based as well as machine learning during final decision making. But, for a given instance, whether machine learning or rule-based component will be used, was not explained. They used either of the components in different situations in order to enjoy the advantage of the both the components.

In this work, we propose a hybrid approach for event extraction from the text under the TempEval-2010 framework. Initially, we develop a method for event extraction based on machine learning. We use Conditional Random Field (CRF) as the underlying machine learning algorithm. We observe that this machine learning based system often makes the errors in extracting the events denoted by *deverbial* entities. This observation prompts us to employ several strategies in conjunction with machine learning. These strategies are implemented based on semantic role labeling, WordNet and handcrafted rules. We experiment with the TempEval-2010 evaluation challenge setup [7]. Evaluation results yield the precision, recall and F-measure values of approximately 93.00%, 96.00% and 94.47%, respectively. This is approximately 12% higher F-measure in comparison to the best system [8] of TempEval-2010.

We use semantic role labels for event nominalizations. Events can be analyzed by these kinds of nominalizations. As our goal is on nominal Semantic Role Label (SRL), we concentrate on the event/target/results class. SRL for nominalization represents semantic roles to extract high level information that are more independent from the word tokens. On the other hand on verbal SRL [9, 10] there is relatively little work that specifically addresses nominal SRL. Nouns are generally treated like verbs. The task is split into two

Manuscript received November 2, 2010. Manuscript accepted for publication January 15, 2011.

A. Kolya, A. Ekbal, and S. Bandyopadhyay are with the Computer Science and Engineering Department , Jadavpur University, Kolkata; A. Ekbal is also with Computer science and engineering department,IIT Patna, India (e-mail: anup.kolya@gmail.com,asif.ekbal@gmail.com, sivaji_cse_ju@yahoo.com).

classification steps, argument recognition (telling arguments from non-arguments) and argument labelling (labelling recognized arguments with a role). Nominal SRL also typically draws on feature sets that are similar to those for verbs, i.e. comprising mainly syntactic and lexical-semantic information [11]. On the other hand, there is converging evidence that nominal SRL is somewhat more difficult than verbal SRL.

Hence, semantic roles may aid in learning a more general model. This learning model could improve the results of the approaches that are solely focused on lower-level information. Two frameworks for semantic roles have found wide use in the community, PropBank [12] and FrameNet [13]. Their corpora are used to train supervised models for semantic role labelling of new text [9][14]. The resulting analysis can benefit a number of applications, such as Information Extraction [15] or Question Answering [16].

The rest of the paper is structured as follows. Section 2 describes our Conditional Random Field (CRF) based event extraction approach. We describe our event extraction approaches with the use of semantic roles in Section 3, WordNet in Section 4 and hand-crafted rules in Section 5. Evaluation results under the experimental set up of TempEval-2010 evaluation challenge are reported in Section 6. Finally, Section 7 concludes the paper

II. CRF BASED APPROACH FOR EVENT EXTRACTION

Conditional Random Field (CRF) [17] is an undirected graphical model, which is a special case of which corresponds to conditionally trained probabilistic finite state automata. Being conditionally trained, these CRFs can easily incorporate a large number of arbitrary, non-independent features while still having efficient procedures for non-greedy finite-state inference and training. CRFs have shown success in various sequence modeling tasks including noun phrase segmentation [18] and table extraction [19]. The main advantage of CRF comes from that it can relax the assumption of conditional independence of the observed data often used in generative approaches, an assumption that might be too restrictive for a considerable number of object classes. Additionally, CRF avoids the label bias problem.

CRF is used to calculate the conditional probability of values on designated output nodes given values on other designated input nodes. The conditional probability of a state sequence $S =< s_1, s_2, ..., s_T >$ given an observation sequence $O =< o_1, o_2, ..., o_T)$ is calculated as:

$$P_\Lambda(s \mid o) = \frac{1}{Z_o} \exp(\sum_{t=1}^{T} \sum_{k=1}^{K} \lambda_k f_k(s_{t-1}, s_t, o, t))$$

where, $f_k(s_{t-1}, s_t, o, t)$ is a feature function whose weight λ_k is to be learned via training. The values of the feature functions may range between $-\propto .....+\propto$, but typically they are binary. To make all conditional probabilities sum up to 1, we must calculate the normalization factor,

$$Z_0 = \sum_s \exp(\sum_{t=1}^{T} \sum_{k=1}^{K} \lambda_k f_k(s_{t-1}, s_t, o, t)),$$

This, as in HMMs, can be obtained efficiently by dynamic programming.

Here, the CRF parameters are optimized using Limited-memory BFGS[16], a quasi-Newton method that is significantly more efficient, and results in only minor changes in accuracy due to changes in σ. CRFs generally can use real-valued functions but it is often required to incorporate the binary valued features. A feature function $f_k(s_{t-1}, s_t, o, t)$ has a value of 0 for most cases and is only set to 1, when s_{t-1}, s_t are certain states and the observation has certain properties. We have used the C^{++} based CRF++ package[1], a simple, customizable, and open source implementation of CRF for segmenting /labeling sequential data.

A. Features of CRF

We extract the gold-standard TimeBank features for events in order to train/test the CRF model. In the present work, we mainly use the various combinations of the following features:

(i). **Part of Speech (POS) of event terms**: It denotes the POS information of the event. The features values may be either of ADJECTIVE, NOUN, VERB, and PREP.

(ii). **Event Tense**: This feature is useful to capture the standard distinctions among the grammatical categories of verbal phrases. The tense attribute can have values, PRESENT, PAST, FUTURE, INFINITIVE, PRESPART, PASTPART, or NONE.

(iii). **Event Aspect**: It denotes the aspect of the events. The aspect attribute may take values**,** PROGRESSIVE, PERFECTIVE and PERFECTIVE PROGRESSIVE or NONE.

(iv). **Event Polarity**: The polarity of an event instance is a required attribute represented by the boolean attribute, polarity. If it is set to 'NEG', the event instance is negated. If it is set to 'POS' or not present in the annotation, the event instance is not negated.

(v). **Event Modality**: The modality attribute is only present if there is a modal word that modifies the instance.

(vi). **Event Class**: This is denoted by the 'EVENT' tag and used to annotate those elements in a text that mark the semantic events described by it. Typically, events are verbs but can be nominal also. It may belong to one of the following classes:

REPORTING: Describes the action of a person or an organization declaring something, narrating an event, informing about an event, etc.

PERCEPTION: Includes events involving the physical perception of another event. Such events are typically expressed by verbs like: *see*, *watch*, *glimpse*, *behold*, *view*, *hear*, *listen*, *overhear* etc.

ASPECTUAL: Focuses on different facets of event history.

[1]http://crfpp.sourceforge.net

I_ACTION: An intentional action. It introduces an event argument which must be in the text explicitly describing an action or situation from which we can infer something given its relation with the I_ ACTION.

I_STATE: Similar to the I_ACTION class. This class includes states that refer to alternative or possible words, which can be introduced by subordinated clauses, nominalizations, or untensed verb phrases (VPs).

STATE: Describes circumstances in which something obtains or holds true.

Occurrence: Includes all of the many other kinds of events that describe something that happens or occurs in the world.

Event Stem: It denotes the stem of the head event.

III. USE OF SEMANTIC ROLES FOR EVENT EXTRACTION

We use Semantic Role Label (SRL)[9] [20] to identify different features of the sentences of a document. These features help us to extract the events from the text. For each predicate in a sentence acting as event word, semantic roles extract all constituents, determining their arguments (agent, patient, etc.) and their adjuncts (locative, temporal, etc.). Some of the others features like predicate, voice and verb sub-categorization are shared by all the nodes in the tree. In the present work, we use predicate as an event. Semantic roles can be used to detect the events that are the nominalizations of verbs such as *agreement* for *agree* or *construction* for *construct*. Event nominalizations often afford the same semantic roles as verbs, and often replace them in written language [21]. Nominalisations (or, *deverbal nouns*) are commonly defined as nouns, morphologically derived from verbs, usually by suffixation [22]. They can be classified into at least three categories in the linguistic literature, event, result, and agent/patient nominalisations [23]. Event and result nominalisations constitute the bulk of *deverbal* nouns. The first class refers to an event/activity/process, with the nominal expressing this action (e.g., killing, destruction etc.). Nouns in the second class describe the result or goal of an action (e.g., agreement, consensus etc.). Many nominals have both an event and a result reading (e.g., selection). A smaller class is agent/patient nominalizations that are usually identified by suffixes such as *-er*, *-or* etc., while patient nominalisations end with *-ee*, *-ed* (e.g. employee). Let us consider the following example sentence to understand how semantic roles can be used for event extraction.

All sites were inspected to the satisfaction of the inspection team and with full cooperation of Iraqi authorities, Dacey said.

The output of SRL for this sentence is as follows:

[ARG1 All sites] were [TARGET inspected] to the satisfaction of the inspection team and with full cooperation of Iraqi authorities, [ARG0 Dacey] [TARGET said]

The sentence is traversed to find the argument-target relations. A sentence is scanned as many times as the number of target words in the sentence. In the first traversal, *inspected* is identified as the event. In the second pass, *said* is identified as an event. All the extracted target words are treated as the event words. We observed that many of these target words are identified as the event expressions by the CRF model. But, there exists many nominalised event expressions (i.e., *deverbal nouns*) that are not identified as events by the supervised CRF. These nominalised expressions are correctly identified as events by SRL. We observe performance improvement with the inclusion of this module.

IV. USE OF WORDNET FOR EVENT EXTRACTION

WordNet [23] features have been widely used to extract different lexical categories, such as *part-of-speech (POS)*, *stem*, *hypernym*, *meronym*, *distance* and *common-parents*, and demonstrated its worth in many tasks. Here, WordNet is mainly used to identify *non-deverbal event nouns*. We observed from the outputs of CRF and SRL that the event entities like '*war*', '*attempt*', '*tour*' etc. are not properly identified. These words have noun (NN) POS information, and the previous approaches, i.e. CRF and SRL can only identify those event words that have verb (VB) POS information. We know from the lexical information of WordNet that the words like '*war*' and '*tour*' are generally used as both *noun* and *verb* forms in the sentence. We design two following rules based on the WordNet:

Rule 1: The word tokens having Noun (NN) PoS categories are looked into the WordNet. If it appears in the WordNet with noun and verb senses, then that word token is also considered as an event. For example, *war* has both noun and verb senses in the WordNet, and thus considered as an event.

Rule 2: The *stems* of the noun word tokens are looked into WordNet. If one of the WordNet senses is verb then the token will be identified as verb. For example, the stem of *proposal*, i.e. *propose* has two different senses, noun and verb in the WordNet, and thus it is considered as an event.

We observe significant performance improvement on event extraction with the above mentioned two rules.

V. USE OF RULES FOR EVENT EXTRACTION

We used WordNet to extract the event expressions that appear in the WordNet with both noun and verb senses. Here, we mainly concentrate to identify the specific lexical classes like '*inspection*' and '*resignation'*. These can be identified by the suffixes such as ('*-ción*'), ('*-tion'*) or ('*-ion*'), i.e. the morphological markers of deverbal derivations.

Initially, we run the CRF based Stanford Named Entity (NE) tagger[2] on the TempEval-2 test dataset. The output of the system is tagged with *Person*, *Location*, *Organization* and *Other* classes. The words starting with the capital letters are

[2] http://nlp.stanford.edu/software/CRF-NER.shtml

also considered as NEs. Thereafter, we came up with the following rules for event extraction:

Cue-1: Nouns which are morphologically derived from verbs are commonly distinguished as nominalizations (or, deverbal nouns). The deverbal nouns are usually identified by the suffixes like '*-tion*', '*-ion*', '*-ing*' and '*-ed*' etc. The nouns that are not NEs, but end with these suffixes are considered as the event words.

Cue 2: The verb-noun combinations are searched in the sentences of the test set. The non-NE noun word tokens are considered as the events.

Cue 3: Nominals and non-deverbal event nouns can be identified by the complements of aspectual PPs headed by prepositions like *during*, *after* and *before*, and complex prepositions such as *at the end of* and *at the beginning of* etc. The next word token(s) appearing after these clue word(s)/phrase(s) are considered as events.

Cue 4: The non-NE nouns occurring after the expressions such as *frequency of, occurrence of* and *period of* are most probably the event nouns.

Cue 5: Event nouns can also appear as objects of aspectual and time-related verbs, such as *have begun a campaign* or *have carried out a campaign* etc. The non-NEs that appear after the expressions like "*have begun a*", "*have carried out a"* etc. are also most probably the events.

VI. EVALUATION RESULT

We use the TempEval-2010 datasets to report the evaluation results. We start with the development of a CRF based system. We develop a number of CRF models depending upon the various features included into it. We have a training data in the form W_i, T_i, where, W_i is the i^{th} pair along with its feature vector and T_i is its corresponding output label (i.e., *Event* or *Other*). Models are built based on the training data and the feature template. The procedure of training is summarized below:

1. Define the training corpus, C.
2. Extract the <*token, output*> relations from the training corpus.
3. Create a file of candidate features derived from the training corpus.
4. Define a feature template.
5. Compute the CRF weights λ_k for every f_K using the CRF toolkit with the training file and feature template as input.
6. Derive the best feature template depending upon the performance.
7. Select the best feature template obtained from Step 6.
8. Retrain the CRF model

We use various subsets of the template as shown in Figure 1 during our experiment. In the figure, w_i : Current <*token, output*> pair, $w_{(i-n)}$: Previous nth pair, $w_{(i+n)}$: Next nth pair, $t_{i-1.}$ previous pair.

The test data had 373 verbal and 125 non-deverbal event nouns. Overall evaluation results are reported in Table 1. The CRF based system shows the precision, recall and F-measure values of 75.3%, 78.1% and 76.87%, respectively. The performance increases by 1.39 percentage F-measure points with the use of semantic roles. Table shows very high performance improvement (i.e., 11.01%) with the use of WordNet. The rule-based component also shows the effectiveness with the improvement of 5.20 F-measure percentage points. Finally, the system achieves the precision, recall and F-measure values of 93.00%, 96.00% and 94.47%, respectively. This is actually an improvement of approximately 12% F-measure value over the best reported system [8].

$w_{(i-2)}$
$w_{(i-1)}$
w_i
w_{i+1}
$w_{(i+2)}$
Combination of w_{i-1} and w_i
Combination of w_i and w_{i+1}
Dynamic output tag (t_i) of the previous token
Feature vector of w_i of other features

Figure 1: Feature template used for the experiment

TABLE 1.
EVALUATION RESULTS OF EVENT EXTRACTION (PERCENTAGES)

Model	precision	Recall	F-measure
CRF	75.30	78.10	76.87
CRF+SRL	76.60	80.00	78.26
CRF+SRL+WordNet	88.56	90.00	89.27
CRF+SRL+WordNet+Rules	93.00	96.00	94.47

VII. CONCLUSION

In this paper, we have reported our work on event extraction under the TempEval -2010 evaluation exercise. Initially, we developed a CRF based supervised system for event extraction. This CRF based systems suffer mostly in identifying the deverbal nouns that denote the event expressions. Thereafter, we came up with several proposals in order to improve the system performance. We proposed a number of techniques based on SRL, WordNet and handcrafted rules. Evaluation results yield the precision, recall and F-measure values of 93.00%, 96.00% and 94.47%, respectively. This is an improvement of approximately 12 percentage F-measure points over the best performing system of TemEval-2010 evaluation challenge.

Future works include the identification of more precise rules for event identification and multiword events. Future works also include experimentations with other machine learning techniques like maximum entropy and support vector machine.

ACKNOWLEDGEMENTS

The work was partially supported by a grant from English to Indian language Machine Translation (EILMT) funded by Department of Information and Technology (DIT), Government of India.

REFERENCES

[1] Boguraev, B. and R. K. Ando. 2005. TimeMLCompliant Text Analysis for Temporal Reasoning. *Proceedings of Nineteenth International Joint Conference on Artificial Intelligence (IJCAI-05)*, pp. 997–1003, Edinburgh, Scotland.

[2] Mani, I., Wellner, B., Verhagen, M., Lee C.M., Pustejovsky, J. 2006. Machine Learning of Temporal Relation. *Proceedings of the 44th Annual meeting of the Association for Computational Linguistics*, Sydney, Australia.

[3] Chambers, N., S. Wang, and D. Jurafsky. 2007. Classifying Temporal Relations between Events. *Proceedings of the ACL 2007 Demo and Poster Sessions*, pp. 173–176, Prague, Czech Republic.

[4] Verhagen, M., Gaizauskas, R., Schilder, F., Hepple, M., Katz, G., Pustejovsky, and J.: SemEval-2007 Task 15: TempEval Temporal Relation Identification. *Proceedings of the 4th International Workshop on Semantic Evaluations (semEval-2007)*, pp. 75-80, Prague.

[5] Mani, I., B. Wellner, M. Verhagen, and J. Pustejovsky. 2007. Three Approaches to Learning TLINKs in TimeML. *Technical Report CS-07-268, Computer Science Department, Brandeis University*, Waltham, USA.

[6] Mao, T., Li. T., Huang, D.,Yang,Y. 2006. Hybrid Models for Chinese Named Entity Recognition. *Proceedings of the Fifth SIGHAN Workshop on Chinese Language Processing.*

[7] A. Kolya, A. Ekbal and S. Bandyopadhyay. 2010e. JU_CSE_TEMP: A First Step towards Evaluating Events, Time Expressions and Temporal Relations. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, ACL 2010, July 15-16, Sweden, PP. 345–350.

[8] Hector Llorens, Estela Saquete, Borja Navarro.2010. TIPSem (English and Spanish): Evaluating CRFs and Semantic Roles. *Proceedings of the 5th International Workshop on Semantic Evaluation, ACL 2010*, pages 284–291, Uppsala, Sweden, 15-16 July 2010.

[9] Gildea, D. and D. Jurafsky. 2002. Automatic Labeling of Semantic Roles. *Computational Linguistics*, 28(3):245–288.

[10] Fleischman, M. and E. Hovy. 2003. Maximum Entropy Models for FrameNet Classification. *Proceedings of EMNLP*, pages 49–56, Sapporo, Japan.

[11] Liu, C. and H. Ng. 2007. Learning Predictive Structures for Semantic Role Labeling of NomBank. *Proceedings of ACL*, pages 208–215, Prague.

[12] Palmer, M., D. Gildea, and P. Kingsbury. 2005. The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106.

[13] Fillmore, C., C. Johnson, and M. Petruck. 2003. Background to FrameNet. *International Journal of Lexicography*, 16:235–250.

[14] Carreras, X. and L. M`arquez, editors. 2005. Semantic Role Labeling. *Proceedings of the CoNLL-05 Shared Task.*

[15] Moschitti, A., P. Morarescu, and S. Harabagiu. 2003. Open-domain information extraction via automatic semantic labeling. *Proceedings of FLAIRS*, pages 397–401, St. Augustine, FL.

[16] Frank Schilder, Graham Katz, and James Pustejovsky. 2007. Annotating, Extracting and Reasoning About Time and Events (Dagstuhl 2005), *volume 4795 of LNCS. Springer*.

[17] Lafferty, J., McCallum, A., Pereira, F. 2001 Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. *Proceedings of 18th International Conference on Machine Learning*, pp.282-289.

[18] Sha, F., Pereira, F. 2003. Shallow Parsing with Conditional Random Fields. *Proceedings of HLT-NAACL.*

[19] Pinto, D., McCallum, A.,Wei, X., Croft, W.B. 2003. Table Extraction Using Conditional Random Fields. *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 235-242.

[20] Sameer S. Pradhan, Wayne Ward, Kadri Hacioglu, James H. Martin, Daniel Jurafsky, Shallow Semantic Parsing using Support Vector Machines . *Proceedings of the Human Language Technology Conference/North American chapter of the Association for Computational Linguistics annual meeting (HLT/NAACL-2004*), Boston, MA, May 2-7, 2004.

[21] Gurevich, O., R. Crouch, T. King, and V. de Paiva.2006. Deverbal Nouns in Knowledge Representation. *Proceedings of FLAIRS*, pages 670–675, Melbourne Beach, FL.

[22] Quirk, R., S. Greenbaum, G. Leech, and J. Svartvik. 1985. A Comprehensive Grammar of the English Language. *Longman*.

[23] Grimshaw, J. 1990. Argument Structure. MIT Press.

[24] George A. Miller. 1990. WordNet: An on-line lexical database. *International Journal of Lexicography*, 3(4): 235–312