JU_CSE: A CRF Based Approach to Annotation of Temporal Expression, Event and Temporal Relations

Anup Kumar Kolya[1], Amitava Kundu[1], Rajdeep Gupta[1]

[1]Dept. of Computer Science & Engineering
Jadavpur Univeristy
Kolkata-700 032, India
{anup.kolya,amitava.jucse,
rajdeepgupta20}@gmail.com

Asif Ekbal[2], Sivaji Bandyopadhyay[1]

[2]Dept. of Computer Science & Engineering
IIT Patna
Patna-800 013, India
asif@iitp.ac.in,
sivaji_ju_cse@yahoo.com

Abstract

In this paper, we present the JUCSE system, designed for the TempEval-3 shared task. The system extracts events and temporal information from natural text in English. We have participated in all the tasks of TempEval-3, namely Task A, Task B & Task C. We have primarily utilized the Conditional Random Field (CRF) based machine learning technique, for all the above tasks. Our system seems to perform quite competitively in Task A and Task B. In Task C, the system's performance is comparatively modest at the initial stages of system development. We have incorporated various features based on different lexical, syntactic and semantic information, using Stanford CoreNLP and Wordnet based tools.

1 Introduction

Temporal information extraction has been a popular and interesting research area of Natural Language Processing (NLP) for quite some time. Generally, a lot of events are described in a variety of newspaper texts, stories and other important documents where the different events described happen at different time instants. The temporal location and ordering of these events are either specified or implied. Automatic identification of time expressions and events and annotation of temporal relations constitute an important task in

text analysis. These are also important in a wide range of NLP applications that include temporal question answering, machine translation and document summarization.

A lot of research in the area of temporal information extraction has been conducted on multiple languages, including English and several European languages. The TimeML was first developed in 2002 in an extended workshop called TERQAS (Time and Event Recognition for Question Answering Systems) and, in 2003, it was further developed in the context of the TANGO workshop (TimeML Annotation Graphical Organizer). Since then most of the works in this research arena have been conducted in English. The variety of works include TimeML (Pustejovsky et al., 2003), the development of a temporally annotated corpus Time-Bank (Pustejovsky et al., 2003), the temporal evaluation challenges TempEval-1 (Verhagen et al., 2007), TempEval-2 (Pustejovsky and Verhagen, 2010). In the series of Message Understanding Conferences (MUCs) that started from 1987 and the Sheffield Temporal Annotation scheme (STAG) (Setzer &Gaizauskas, 2000) the aim was to identify events in news text and determine their relationship with points on a temporal line.

In the series of TempEval evaluation exercises, TempEval-1 was the first one where the focus was on identification of three types of temporal relation: relation between an event and a time expression in the same sentence, relation between an

event and the document creation time, and relation between two main events in consecutive sentences.

TempEval-2 was a follow up to TempEval-1 and consisted of six subtasks rather than three. It added (i) identification of time expressions and determination of values of the attributes TYPE and VAL (ii) identification of event expressions and determination of its attribute values. It included the previous three relation tasks from TempEval-1 and an additional task of annotating temporal relation between a pair of events where one subordinates the other.

We have participated in all three tasks of TempEval-3- Task A, Task B and Task C. A combination of CRF based machine learning and rule based techniques has been adopted for temporal expression extraction and determination of attribute values of the same (Task A). We have used a CRF based technique for event extraction (Task B), with the aid of lexical, semantic and syntactic features. For determination of event attribute values we have used simple rule based techniques. Automatic annotation of temporal relation between event-time in the same sentence, event-DCT relations, mainevent-mainevent relations in consecutive sentences and subevent-subevent relations in the same sentences has been introduced as a new task (Task-C) in the TempEval-3 exercise. We have adopted a CRF based technique for the same as well.

2 The JU_CSE System Approach

The JU_CSE system for the TempEval-3 shared task uses mainly a Conditional Random Field (CRF) machine learning approach to achieve Task A, Task B & Task C. The workflow of our system is illustrated in Figure 1.

2.1 Task A: Temporal Expression Identification and Normalization

Temporal Expression Identification:

We have used CRF++ 0.57[1], an open source implementation of the Conditional Random Field (CRF) machine learning classifier for our experiments. CRF++ templates have been used to capture the relation between the different features in a sequence to identify temporal expressions. Temporal expressions mostly appear as multi-word entities such as "the next three days". Therefore the use of CRF classifier that uses context information of a token seemed most appropriate.

Initially, all the sentences have been changed to a vertical token-by-token level sequential structure for temporal expressions representation by a B-I-O encoding, using a set of mostly lexical features. In this encoding of temporal expression, "B" indicates the 'beginning of sequence', "I" indicates a token inside a sequence and "O" indicates an outside word. We have carefully chosen the features list based on the several entities that denote month names, year, weekdays, various digit expressions (*day, time, AM, PM* etc.) In certain temporal expression patterns (*several months, last evening*) some words (*several, last*) act as modifiers to the following words that represent the time expression. Temporal expressions include time expression modifiers, relative days, periodic temporal set, year-eve day, month name with their short pattern forms, season of year, time of day, decade list and so on. We have used the POS information of each token as a feature. We have carefully accounted for a simple intuition revelation that most temporal expressions contain some tokens conveying the *"time"* information while others possibly conveying the "quantity" of time. For example, in the expression *"next three days"*, *"three"* quantifies *"days"*. Following are the different temporal expressions lists that have been utilized:

- A list of time expression modifiers: *this*, *mid*, *recent*, *earlier*, *beginning*, *late* etc.
- A list of relative days: *yesterday*, *tomorrow* etc.
- A list of periodic temporal set: *hourly*, *nightly* etc.
- A list of year eve day: *Christmas Day*, *Valentine Day* etc.
- A list of month names with their short pattern forms: *April*, *Apr.* etc.
- A list of season of year: *spring*, *winter* etc.
- A list of time of day: *morning*, *afternoon*, evening etc.
- A list of decades list: *twenties*, *thirties* etc.

---

[1] http://crfpp.googlecode.com/svn/trunk/doc/index.html

Raw Text:

For his part, Fidel Castro is the ultimate political survivor. People have predicted his demise so many times, and the US has tried to hasten it on several occasions. Time and again, he endures.

- Tokenize with Stanford CoreNLP
- Obtain POS tags of tokens
- Extract features from tokens
- Identify the features for event annotation and temporal annotation separately

Event & Time Features

CRF

Tag TIMEX3 tokens

For……… OTHERS
nearly ……….. TIMEX3
forty…. … TIMEX3
years…….. TIMEX3
.
.

Tag EVENT tokens

Determine Event Class

<MAKEINSTANCE eiid="ei1" eventID="e1" pos="VERB" tense="PRESENT" aspect="PERFECTIVE" polarity="POS" />

<MAKEINSTANCE eiid="ei2" eventID="e2" pos="NOUN" tense="PRESENT" aspect="PERFECTIVE" polarity="POS" />

<MAKEINSTANCE eiid="ei3" eventID="e3" pos="VERB" tense="PRESENT" aspect="PERFECTIVE" polarity="POS" />

Rule based approach to obtain tense, aspect, polarity, modality etc. for events

People……… OTHERS
have ……….. OTHERS
predicted …. … EVENT
his ………….. OTHERS
.
.

CoreNLP for "type" & "velue"

Annotated Text

For his part, Fidel Castro is the ultimate political survivor. People have <EVENT class="I_ACTION" eid="e1">predicted</EVENT> his <EVENT class="OCCURRENCE" eid="e2">demise</EVENT> so many times, and the US has <EVENT class="I_ACTION" eid="e3">tried</EVENT> to <EVENT class="OCCURRENCE" eid="e4">hasten</EVENT> it on several occasions.

Temporal Relations:

<TLINK lid="l1" relType="BEFORE" eventInstanceID="ei1" relatedTo-Time="t0" />

<TLINK lid="l2" relType="BEFORE" eventInstanceID="ei2" relatedToEvent-tInstance="ei1" />

CRF

Enlist entity pairs with features
<mainevent-mainevent>
<event-event>
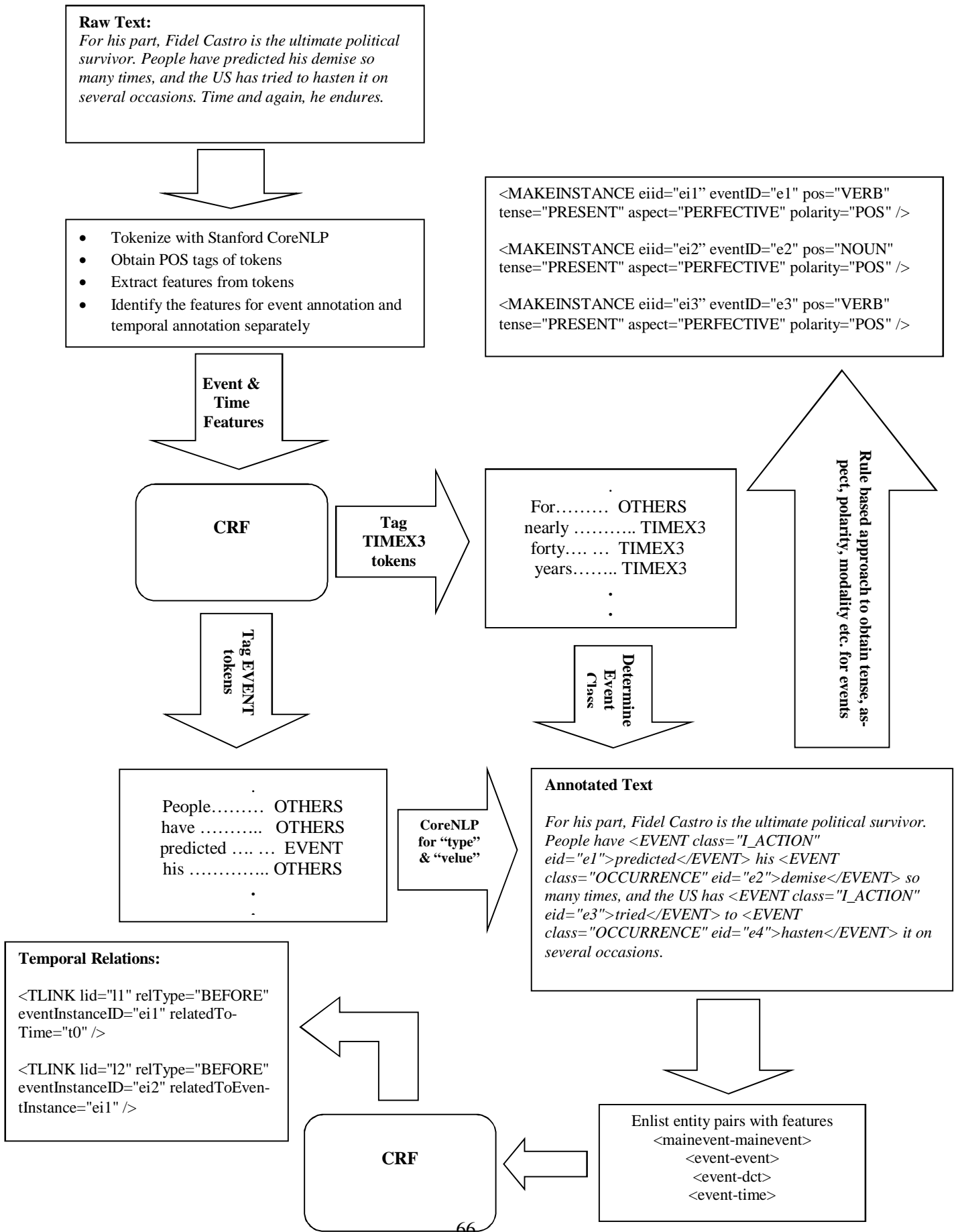<event-dct>
<event-time>

66

Figure 1.The JU_CSE System Architecture

Determination of Normalized *value* and *type* of Temporal Expressions:

Temporal expressions in documents are generally defined with the *type* and *value* attributes. All the temporal expressions can be differentiated into three types (i) explicit (ii) relative and (iii) implicit temporal expressions. For example, the expression *"October 1998"* refers to a specific month of the year which can be normalized without any additional information. On the other hand, the relative expression *"yesterday"* can't be normalized without the knowledge of a corresponding reference time. The reference time can either be a temporal expression or the Document Creation Time marked in the document. Consider the following piece of text: *"Yesterday was the 50th independence of India". The First Independence Day of India is 15th august 1947."* Here *"Yesterday"* can be normalized as *"15-08-1997"*. It may be noted that information such as *"First Independence Day of India"* can be directly accessed from the timestamp calendar, through the metadata of a document. The third type of temporal expressions includes implicit expressions such as names of festival days, birthdays and holidays or events. These expressions are mapped to available calendar timeline to find out their normalized values.

Temporal Expression	Type	Value
A couple of years	DURATION	P2Y
October	DATE	"1997-10"
Every day	SET	P1D
2 P.M.	TIME	2013-02-01T14:00
Now	DATE	PRESENT_REF"

Table 1: TimeML normalized type and value attributes for temporal expressions

We have implemented a combined technique using our handcrafted rules and annotations given by the Stanford CoreNLP tool to determine the 'type'-s and 'value'-s. Four types TIME, DATE, DURATION and SET of temporal expressions are defined in the TimeML framework. Next, we have evaluated the normalized value of temporal expressions using Document Creation Time (DCT) from the documents. In this way, values of different dates have been inferred e.g. *last year*, *Monday*, and *today*.

2.2 Task B: Extraction of Event Words and Determination of Event Attribute Values

Event Extraction

In our evaluation framework, we have used the Stanford CoreNLP tool extensively to tokenize, lemmatize, named-entity annotate and part-of-speech tag the text portions of the input files. For event extraction, the features have been considered at word level, where each word has its own set of features. The general features used to train our CRF model are:

Morphological Features: Event words are represented mostly as verbs and nouns. The major problem is detecting the events having non-verbal PoS labels. Linguistically, non-verbal wordforms are derived from verbal wordforms. Various inflectional and derivational morphological rules are involved in the process of evolving from verbal to non-verbal wordforms. We have used a set of handcrafted rules to identify the suffixes such as (*'-ción', '-tion' or '-ion')*, i.e., the morphological markers of word token, where Person, Location and Organization words are not considered. The POS and lemma, in a 5-window (-2, +2), has been used for event extraction.

Syntactic Feature: Different event words notions are contained in the sentences such as: verb-noun combinations structure, the complements of aspectual prepositional phrases (PPs) headed by prepositions and a particular type of complex prepositions. These notions are captured to be used as syntactic features for event extraction.

WordNet Feature: The RiTa Wordnet[2] package has been effectively used to extract different properties of words, such as Synonyms, Antonyms, Hypernyms, & Hyponyms, Holonyms, Meronyms, Coordinates, & Similars, Nominalizations, Verb-Groups, & Derived-terms. We have used these Wordnet properties in the training file for the CRF in the form of binary features for verbs and nouns indicating if the words like "act", "activity", "phenomenon" etc. occur in different relations of the Wordnet ontology.

---

[2] http://www.rednoise.org/rita/wordnet/documentation/

Features using Semantic Roles: We use Semantic Role Label (SRL) (Gildea et el, 2002; Pradhan et al, 2004; Gurevich et al, 2006) to identify different useful features for event extraction. For each predicate in a sentence acting as event word, semantic roles extract all constituents; determine their arguments (agent, patient, etc.) and adjuncts (locative, temporal, etc.). Some of the other 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 nominalizations of verbs such as *agreement* for *agree* or *construction* for *construct*. Event nominalizations often share the same semantic roles as verbs, and often replace them in written language. Noun words, morphologically derived from verbs, are commonly defined as deverbal nouns. Event and result nominalizations 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 denote both the event and result (e.g., selection). A smaller class is agent/patient nominalizations that are usually identified by suffixes such as -er, -or etc., while patient nominalizations end with -ee, -ed (e.g. employee).

Object information of Dependency Relations (DR): We have developed handcrafted rules to identify features for CRF training, based on the object information present in the dependency relations of parsed sentences. Stanford Parser (de Marneffe et al., 2006), a probabilistic lexicalized parser containing 45 different Part-of-Speech (PoS) tags of Penn Treebank is used to get the parsed sentences with dependency relations. The dependency relations are found out for the predicates "dobj" so that the direct object related components in the "dobj" predicate is considered as the feature for the event expression. Initially the input sentences are passed to the dependency parser[3]. From the parsed output verb noun combination direct object (dobj) dependency relations are extracted. These dobj relations basically inform us that direct object of a VP is the noun phrase which is the (accusative) object of the verb; the direct object of a clause is the direct object of the VP

which is the predicate of that clause. Within the dobj relation governing verb word and dependent noun words are acting as important features for event identification when dependent word is not playing any role in other dependency relation (nsubj, prep_of, nn ,etc.) of the sentence.

In this way, we have set list of word tokens and its features to train the recognition model. Then the model will give to each word one of the valid labels.

Determination of various Event Attribute Values:

Values of different event attributes have been computed as follows:

Class: Identification of the class of an event has been done using a simple, intuitive, rule based approach. Here too, the hypernym list of an event token from RitaWordnet has been deployed to determine the class of the respective event. In this case, OCCURRENCE has been considered the default class.

Tense, Aspect, POS: These three attributes are the obligatory attributes of MAKEINSTANCE tags. To determine the tense, aspect and polarity of an event, we have used the "parse" annotator in CoreNLP. We annotated each sentence with the Stanford dependency relations using the above annotator. Thereafter various specific relations were used to determine the tense, aspect and POS of an event token, with another rule based approach. For example, in the phrase "*has been abducted*", the token "*been*" appears as the dependent in an "aux" relation with the event token "*abducted*"; and hence the aspect "PERFECTIVE" is inferred. The value "NONE" has been used as the default value for both tense and aspect.

Polarity and Modality: Polarity of event tokens are determined using Stanford dependency relations too; here the "neg" relation. To determine the modality we search for modal words in "aux" relations with the event token.

2.3   Task C: Temporal Relation Annotation

We have used the gold-standard TimeBank features for events and times for training the CRF. In the present work, we mainly use the various combinations of the following features:

---

[3] http://nlp.stanford.edu:8080/parser/

(i)   Part of Speech (POS)
(ii)   Event Tense
(iii)   Event Aspect
(iv)   Event Polarity
(v)   Event Modality
(vi)   Event Class
(vii)   Type of temporal expression
(vii)   Event Stem
(viii) Document Creation Time (DCT).

The following subsections describe how various temporal relations are computed.

Event-DCT

We take the combined features of every event present in the text and the DCT for this purpose.

Derived Features: We have identified different types of context based syntactic features which are derived from text to distinguish the different types of temporal relations. In this task, following features help us to identify the event-DCT relations, specially "AFTER" temporal relations:

(i)Modal Context: Whether or not the event word has one of the modal context words like- *will*, *shall*, *can*, *may*, or any of their variants (*might*, *could*, *would*, etc.).In the sentence: *"The entire world **will** [EVENT see] images of the Pope in Cuba"*. Here *"will"* context word helps us to determine event-DCT relation 'AFTER'.

(ii)Preposition Context: Any prepositions preceding an event or time expression. We consider an example:*"Children and invalids would be permitted **to** [EVENT leave] Iraq"*. Here the preposition *to* helps us to determine event-DCT relation 'AFTER'. The same principle goes for time too: in the expressions *on Friday* and *for nearly forty years*, the prepositions *on* and *for* governs the time.

(iii)Context word before or after temporal expression: context words like *before*, *after*, *less than*, *greater than* etc. help us to determine event-time temporal relation identification. Consider an example: *"**After** ten years of [EVENT boom] ...."*

Event-Time

Derived Features: We extract all events from every sentence. For every temporal expression in a sentence, we pair an event in the sentence with the former so that the temporal relation can be determined.

Similar to annotation of event-DCT relations, here too, we have identified different types of context based temporal expression features which are derived from text to distinguish the different types of temporal relations. In this task, the following features help us to distinguish between event and time relations, specially "AFTER" and "BEFORE" temporal relations. The following features are derived from text.

(i)Type of temporal expression: Represents the temporal relationship holding between events, times, or between an event and a time of the event.

(ii)Temporal signal: Represents temporal prepositions "*on*" (*on this coming Sunday*) and slightly contribute to the overall score of classifiers

(iii)Temporal Expression in the target sentence: Takes the values *greater than*, *less than*, *equal* or *none*. These values contribute to the overall score of classifiers.

Mainevent-Mainevent and Subevent-Subevent

The task demands that the main event of every sentence be determined. As a heuristic decision, we have assumed that the first event that appears in a sentence is its main event. We pair up main events (if present) from consecutive sentences and use their combined features to determine their temporal relation. For the events belonging to a single sentence, we take into account the combined features of all possible pairs of sentential events.

Derived Features: We have identified different types of context based syntactic features which are derived from text to distinguish the different types of temporal relations.

(i)Relational context: If a relation holding between the previous event and the current event is "AFTER", the current one is in the past. This information helps us to identify the temporal relation between the current event and successive event.

(ii)Modal Context: Whether or not the event word has one of the context words like, *will*, *shall*, *can*, *may*, or any of their variants (*might*, *could*, *would*, etc.). The verb and auxiliaries governing the next event play as an important feature in event-event temporal relation identification.

(iii)Ordered based context: In event-event relation identification, when EVENT-1, EVENT-2, and EVENT-3 are linearly ordered, then we have assigned true/false as feature value from tense and aspect shifts in this ordered pair.

(iv) Co-reference based feature: We have used co-referential features as derived feature using our in-house system based on Standford CoreNLP tool, where two event words within or outside one sentence are referring to the same event, i.e. two event words co-refer in a discourse.

(v)Event-DCT relation based feature: We have included event-document creation times (DCT) temporal relation types as feature of event-event relation identification.

(ii) Preposition Context: Any prepositions before the event or time, we consider an example:*"Children and invalids would be permitted **to** [EVENT leave] Iraq"*. Here the preposition *to* helps us determine the event-DCT relation 'AFTER'.

(vi) Context word before or after temporal expression: Context words like *before*, *after*, *less than*, *greater than* help us determine event- event temporal relations .We consider an example:*"**After** ten years of [EVENT boom] …."*

(vii)Stanford parser based clause boundaries features: The two consecutive sentences are first parsed using Stanford dependency parser and then clause boundaries are identified. Then, considering the prepositional context and tense verb of the clause, temporal relations are identified where all temporal expressions are situated in the same clause.

3 Results and Evaluation

For the extraction of time expressions and events (tasks A and B), precision, recall and F1-score have been used as evaluation metrics, using the following formulae:

$$\text{precision (P)} = tp/(tp + fp)$$
$$\text{recall (R)} = tp/(tp + fn)$$
$$\text{F-measure} = 2 *(P * R) / (P + R).$$

Where, tp is the number of tokens that are part of an extent in keys and response, fp is the number of tokens that are part of an extent in the response but not in the key, and fn is the number of tokens that are part of an extent in the key but not in the response. Additionally attribute accuracies computed according to the following formulae have also been reported.

Attr. Accuracy = Attr. F1 / Entity Extraction F1
Attr. R = Attr. Accuracy * Entity R
Attr. P = Attr. Accuracy * Entity P

Performance in task C is judged with the aid of the Temporal Awareness score proposed by UzZaman and Allen (2011)

The JU_CSE system was evaluated on the TE-3 platinum data. Table 2 reports JU_CSE's performance in timex extraction Task A. Under the relaxed match scheme, the F1-score stands at 86.38% while the strict match scheme yields a F1-score of 75.41%. As far as TIMEX attributes are concerned, the F1-scores are 63.81% and 73.15% for *value* and *type* respectively.

Timex Extraction						Timex Attribute			
F1	P	R	Strict F1	Strict P	Strict R	Value F1	Type F1	Value Accuracy	Type Accuracy
86.38	93.28	80.43	75.49	81.51	70.29	63.81	73.15	73.87	84.68

Table 2:JU_CSE system's TE-3 Results on Timex Task A

Event Extraction			Event Attribute					
F1	P	R	Class F1	Tense F1	Aspect F1	Class Accuracy	Tense Accuracy	Aspect Accuracy
78.57	80.85	76.41	52.65	58.58	72.09	67.01	74.56	91.75

Table 3:JU_CSE system's TE-3 Results on Event Task B

Table 3 reports the system's performance in event extraction (Task B) on TE-3 platinum data. F1-score for event extraction is 78.57%. Attribute F1-scores are 52.65%, 58.58% and 72.09% for *class*, *tense* and *aspect* respectively.

In both entities extraction tasks recall is notably lower than precision. The F1-scores for event attributes are modest given that the attributes were computed using handcrafted rules. However, the handcrafted approach can be treated as a good baseline to start with. Normalization is proved to be a challenging task.

Task	F1	P	R
Task-ABC	24.61	19.17	34.36
Task-C	26.41	21.04	35.47
Task-C-relation-only	34.77	35.07	34.48

Table 4: JU_CSE system's TE-3 Temporal Awareness results on Task ABC, TaskC-only & TaskC-relation-only

Table 4 presents the Temporal Awareness F1-score for TaskABC, TaskC and TaskC-relation-only. For TaskC-only evaluation, the event and timex annotated data was provided and one had to identify the TLINKs and classify the temporal relations. In the TaskC-relation-only version the timex and event annotations including their attributes as well as TLINKs were provided save the relation classes. Only the relation classes had to be determined. The system yielded a temporal awareness F1-score of 24.6% for TaskABC, 26.41% for TaskC-only and 34.77% for TaskC-relation-only version.

4 Conclusions and Future Directions

In this paper, we have presented the JU_CSE system for the TempEval-3 shared task. Our system in TempEval-3 may be seen upon as an improvement over our earlier endeavor in TempEval-2. We have participated in all tasks of the TempEval-3 exercise. We have incorporated a CRF based approach in our system for all tasks. The JU_CSE system for temporal information extraction is currently undergoing a lot of extensive experimentation. The one reported in this article seemingly has a significant scope of improvement. Preliminarily, the results yielded are quite competitive and encouraging. Event extraction and Timex extraction F1-scores at 78.58% and 86.38% encourage us to further develop our CRF based scheme. We expect better results with additional features and like to continue our experimentations with other semantic features for the CRF classifier. Our rule-based approach for event attribute determination however yields modest F1-scores- 52.65% & 58.58% for class and tense. We intend to explore other machine learning techniques for event attribute classification. We also intend to use parse tree based approaches for temporal relation annotation.

Acknowledgments

References

A. Setzer, and R. Gaizauskas. 2000. Annotating Events and Temporal Information in Newswire Texts. In LREC 2000, pages 1287–1294, Athens.

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

James Pustejovsky, José Castano, Robert Ingria, Roser Sauri, Robert Gaizauskas, Andrea Setzer, Graham Katz, and Dragomir Radev. 2003. TimeML: Robust specification of event and temporal expressions in text. New directions in question answering, 3: 28-34.

Marc Verhagen, Robert Gaizauskas, Frank Schilder, Mark Hepple, Graham Katz, and James Pustejovsky. 2007. Semeval-2007 task 15: Tempeval temporal relation identification. In Proceedings of the 4th International Workshop on Semantic Evaluations, pages 75-80, ACL.

Marc Verhagen, Roser Sauri, Tommaso Caselli, and James Pustejovsky. 2010. Semeval-2010 task 13: Tempeval-2. In Proceedings of the 5th International Workshop on Semantic Evaluation, pages 57- 62. ACL.

Olga Gurevich, Richard Crouch, Tracy H. King, and V. de Paiva. 2006. Deverbal Nouns in Knowledge Representation. Proceedings of FLAIRS, pages 670–675, Melbourne Beach, FL.

Sameer Pradhan, Wayne Ward, Kadri Hacioglu, James H. Martin, and Daniel Jurafsky. 2004. Shallow Semantic Parsing using Support Vector Machine. Proceedings of HLT/NAACL-2004, Boston, MA.

UzZaman, N. and J.F. Allen (2011), "Temporal Evaluation." In Proceedings of The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (Short Paper), Portland, Oregon, USA.