Extraction of Temporal Events in Natural Language Text

VANITHA GUDA  DR SURESH KUMAR SANAMPUDI

Research Scholar of JNTU, Department of CSE, Chaitanya Bharathi Institute of Technology, Gandipet, Hyderabad, TG 9959959585, vanithaguda@gmail.com
Department of IT, JNTUH College of Engg. Jagityal, Nachupally, Karimnagar 9440936885, sureshsanampudi@gmail.com

ABSTRACT

With the exponential growth of dynamic information on the World Wide Web, automatic recognition and extraction of temporal Event has become a penetrating research in Natural Language Processing (NLP). Temporal events form the basic component of knowledge for many of the applications such as question answering system, Temporal summarization etc. Temporal Events are defined as actions that happen or occur at a particular time and place. The resultant temporal Event extracted can be applied on the texts that belongs to various domains like news data, manuscripts, blogs, biological, legal etc. This paper explains a survey of various strategies that are used so far for the extraction of temporal events.

Key words: — Temporal Information, Temporal Events.

I. INTRODUCTION

Information system has to deal with enormous data where natural text analysis has become essential, for this the text represented in natural language [1] form divides the narratives into events, from events to corresponding simple assertions. Basically events are the assertions, which may falls within a time boundary. Events related to time are called as temporal events or temporal expressions. Recognizing temporal events can helps to investigate the information in better way [2], and also helps to attain the information dynamically.

For example, consider the following text:

(1) “APJ Abdul Kalam served as the 11th President of India, succeeding K.R. Narayanan. He won the 2002 presidential election with an electoral vote of 922,884. Expresses three events (served, succeeding, won), temporal event 2002.”

In the example depicted in (1) events are mostly temporal entities, but they also possess a rich non-temporal structure that is important for information access systems (information retrieval, question answering [3], summarization [4], etc.). Without information about the events what happened, where, and to whom, temporal information about an event may not be very useful. The sparsity of research on temporal information extraction is very interesting, since so many statements are temporally qualified. In particular, most of the sources such as newswire text [5] or Wikipedia [6] predominantly express the temporal form of events in the sentences.

Most of the Information extraction system focuses on extracting static facts, encoding the extracted facts into some binary relations. But real world context involves majority of the facts
are fluent. Fluents are dynamic relations whose truth is a function of time these are represented with time constraint. *E.g.* President *(Modi, India)* with temporal scope *(16/05/2014-1/05/2019).* When user wants to extract information of a particular period of times, it is observed to be difficult to extract the information from unstructured open domain sources.

In information system, data may be available in semi-structured or unstructured form, most of the present Information Extraction (IE) systems extracts events from semi-structured data or from closed domain. Temporal event identification and extraction for unstructured data could be beneficial for IE systems in various ways, these events could enhance the performance of personalized news systems, based on user preferences or identified topics or concepts. Further, they can be useful in risk analysis applications [7], monitoring systems [8], and decision making support tools [9].

The remaining part of the paper continues with an elaboration of temporal event extraction in **Section 2.** Subsequently, **Section 3** presents a survey of temporal events extraction strategies. **Section 4** presents the concluding remarks of the survey.

II. TEMPORAL EVENT EXTRACTION

The events which fall on a time line are called as temporal events, extracting temporal information will benefit most of the applications in natural language text. For example, in a news domain[10] related question answering systems, if the user post a query to know “**who was the Prime Minister of India during January 1983**”, then the system has to retrieve the documents that tell us about the Prime ministers data from 1980 to 1985. From the documents extracted a temporal question answering system infers the prime minister details related to January 1983.

Consider another example related to medical domain [11], where patient’s history record written by doctors is not accessible in chronological order. Extracting a temporal structure of the medical record will help practitioner to understand the patient’s medical history effortlessly. Extracting temporal events can be applied to a number of applications like text summarization, question answering systems and decision making process etc. Temporal events mentioned in the text can be classified into two forms namely precise explicit form and implicit form. For example, consider the following sentences,

(1) Two years ago Rohan worked in Microsoft after that he joined in TCS company last month.

(2) Raja Gopal Reddy, is the first Indian who won the Turing award in 1994.

(3) The Palar Blast was a landmine attack on 9 April 1993 in Karnataka, India. The attack, organized by the Forest Brigand Veerappan, killed 22 people, it is explosive attack in Karnataka during the 20th century.

In the example (2) “joined” is referred as temporal event since it is attached with a time interval of “*last month*”. In example (3) “won” is a temporal event as it is attached to a time period 1994. In example (4) events like attack, organized, killed, are found to have temporal nature as they are connected to a time duration of 9 april 1993 and 20th century.

Several challenging issues [12] were found in recognition and extraction of temporal events. Extraction of events and times in natural language text, identification of temporal events [13] from the generated events and mapping the relations amongst events and Times.
III. METHODS FOR TEMPORAL EVENT EXTRACTION

Researchers have worked to extract time of occurrence of an event from natural language text. Then time information processing has been envisioned to compute temporal relations among the events from event-time data extracted. The following are some of the systems that recognize event-times and establish the temporal relations among the events.

- **Lisp Form**: Kahn and gory [14] presents a time specialist program, which accepts input about time relations in a LISP form, It computes temporal relations among events and checks for inconsistencies in input to answer the questions about time relations among the events given in the input.

- **Symbolic Form**: Findler and chen [15][16] presents a model for processing durational and time point information which accepts imprecise and partial temporal specifications in symbolic form for a set of events. From this unfinished information the program checks the consistencies of the input and retrieve the relations such as causality, coexistence and relative duration for remaining pair of events.

- **Tense Aspect**: Bruce [16] presents theoretical model for the time information expressed by tense aspect. QACHRONOS [17] which accepts interactive input in a stylized form of English and answers the questions about the relative time of the events given as input.

- **Annotation Method**: Extracting explicitly obtainable temporal information was attempted through some standard schemes in the context of Message Understanding Conference (MUC)[18]. Other annotation methods like TIDE, STAG and TIMEML[19] mark-up language were introduced as standards for annotations of events, times and temporal relations in natural language text. These systems were emerged in order to annotate only temporal expressions and their normalized values in text, Morph syntactic (MS) knowledge were used as a part of these systems.

After analyzing the methods used for recognizing and extraction of event-times in natural language text. It is found that all the methods can be broadly classified into two categories namely closed domain and open domain. Closed domain means domain specific like medical, finance etc, and open domain means independent of the domain. In general three strategies were commonly used in both the methods they are categorized as Rule based, Data Driven and Hybrid methods.

In Rule based system hand coded rules that make use of low level techniques of NLP were used. The major struggle of this technique lies in writing up the rules part which is found to be critical to interpret all the aspects of natural language text. Rule based systems focuses only on syntactic features of the language where more amount of human interpretation is involved for analyzing the text.

In Data Driven methods syntactic and semantic features of the natural language text were analyzed, but this methods need statistical measures to recognize and classify the events and times.

Hybrid method is one where the issue of rule based and data driven methods are combined to improve the performance of the system, it considers the syntactic nature of rule based methods and semantic features of data driven methods. Linguistic knowledge and human interpretability is involved for all the three methods.
Table 2. Overview of the approaches discussed for temporal event extraction

<table>
<thead>
<tr>
<th>Strategy Based</th>
<th>System name</th>
<th>Tasks</th>
<th>Linguistic Knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rule Based</td>
<td>JU_CSE_TEMP[20]</td>
<td>Recog &amp; Classi</td>
<td>MorphoSyntactic (MS)</td>
</tr>
<tr>
<td></td>
<td>Heidel Time[21]</td>
<td>Recog &amp; Classi</td>
<td>MorphoSyntactic</td>
</tr>
<tr>
<td></td>
<td>TERSEO[22]</td>
<td>Timexes Recog &amp; Classi</td>
<td>MorphoSyntactic</td>
</tr>
<tr>
<td></td>
<td>TTK [23]</td>
<td>Timexes Recog &amp; Classi</td>
<td>MorphoSyntactic</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DataDriven</td>
<td>STEP[24]</td>
<td>Events Recog &amp; classi</td>
<td>MS,Semantic</td>
</tr>
<tr>
<td></td>
<td>TIPSEM [25]</td>
<td>Events Recog &amp; classi</td>
<td>MS,Semantic</td>
</tr>
<tr>
<td>Hybrid</td>
<td>KUL</td>
<td>Timexes Recog &amp; Classi</td>
<td>MS,semantic</td>
</tr>
<tr>
<td></td>
<td>TRIOS [26]</td>
<td>Timexes Recog &amp; Classi</td>
<td>MS, semantic</td>
</tr>
</tbody>
</table>

Table explains the overview of all the approaches discussed for temporal event extraction. Displaying the type of approaches as strategy (Strategies). The systems evolved that fall under this method were termed as system name (System name) and (Tasks) tasks for events in this recognition and classification of events as (Recogi & classi). Also, the type of required linguistic knowledge is termed as (Linguistic knowledge).

In this survey hybrid methods were found to be appropriate for temporal event extraction. This is due to the reason that hybrid methods were formed with the combination of the best features of both rule based and data driven methods. The hybrid approach of the planned work uses the subsequent steps for the extraction of event-times form natural language text. The steps include usage of efficient parsing strategies to parse the text, applying hand-coded rules to extract events and times from the logical forms acquired by the parser.

IV. CONCLUSION

This paper provided the investigations of different approaches used for extraction of temporal events in natural language text. The approaches were broadly classified into open and closed domain. These methods are further classified into rule based, data driven and hybrid methods based on the application of linguistic knowledge. Domain-independent system forms a generalized platform where it can extract significant amount of event-time information from any given document. Hybrid strategies are found to be suggestable for domain-independent methods.

REFERENCES

[4] Barzilay, R.; Elhadad, N.; and McKeown, K. Inferring strategies for sentence ordering in
[22] Event ordering using TERSEO system Application of natural language to information
systems (NLDB04): Application of natural language to information systems (NLDB04).

[23]  www.timeml.org/site/tarsqi/toolkit

