Survey on Event tracking and Event Evolution

S. S. Nandagaonkar  
Computer Department, VPCOE,Baramati  
susahma.nandagaonkar@gmail.com

D.B.Hanchate  
Computer Department, VPCOE,Baramati  
dineshbhanchate@gmail.com

S.N. Deshmukh  
ME-II-[Comp.Engg.] VPCOE,Baramati  
deshmukh.swarupa@gmail.com

Abstract:  
Due to the popularity of the Internet, most news stories have electronic versions published on newswires. Retrieving news of the same topic from multiple sources and keeping information updated becomes more convenient and easier. Techniques that are capable of extracting the underlying structure of the news events are desired. They are helpful to understand the evolution of events on the same topic. This paper surveys the different information retrieval techniques such as Topic detection and Tracking (TDT), event tracking and event evolution approaches. TDT is an active research area in recent years. TDT have different techniques on detecting news topics and tracking news stories for a news topic. TDT techniques have been attempting to detecting or clustering news stories into event, without defining or interpreting the association between these events. There also event evolution is a new concept developed recently, that was the first to conduct investigation on event. Event evolution is the transition development process of related events within the same topic. An event evolution graph is constructed to present events relationship.

Keyword: Topic detection and tracking (TDT), event evolution, evolution patterns, event evolution graph,

1. INTRODUCTION:

Due to the popularity of the internet, most news stories have electronic version published on newswires such as CNN, BBC, CBS, etc. inform diaries such as Google and Yahoo are available to search across multiple newswires to retrieve news stories of any ongoing incidents retrieving news of the same topic from multiple sources and keeping information updated become more convenient and easier. However, it also generates tremendous volume of news text stream. Managing, interpreting, and analyzing such a huge volume of information is a difficult task. Techniques that are capable of extracting the underlying structure of the news events are desired. The Topic Detection and Tracking Study is concerned with the detection and tracking of events. The input to this process is a stream of stories. This stream may or may not be pre-segmented into stories, and the events may or may not be known to the system (i.e., the system may or may not be trained to recognize specific events). This leads to the definition of three technical tasks to be addressed in the TDT study. These are namely the tracking of known events, the detection of unknown events, and the segmentation of a news source into stories. In the TDT research cluster news stories as hierarchical structure as shown in fig[7] below:
Consider the following example, the news topic of “26/11 Mumbai Attacks” as an illustration, it is list of 6 news event in this topic tracked by TDT technique without presenting how these events evolve as an event evolution graph. One can browse through the list and understand what happened according to the temporal order. The list of event under the topic “26/11 Mumbai Attacks”.

**Fig 3. Events under the topic “26/11 Mumbai Attacks”**

Now these events are represented in graphical format by using event evolution graph.

Events are required to construct event evolution graph. First track events by using event evolution pattern that technique as known as EPET (event pattern-supported event tracking).

**RELATED WORK:**

For event tracking mostly TDT technique is used. The objective of TDT is to organize news documents given a stream of constantly generated new stories coming from a wide range of resources such as text or audio in multiple languages. There have been several techniques on detecting news topics and tracking new stories for a news topic.

James Allan et. al [1] in “On line News event detection, Clustering And Tracking” considered each incoming news document as a query that was made on the previous clustered documents to determine if the incoming news document is similar to any of the clustered documents. A story was considered as a first story if no similar documents can be found. Here the problem of new event detection is related to the problem of on-line event clustering, in which news documents are grouped together if they discuss the same news event. The new event detection process finds the stories with which to seed a document clustering algorithm. This tech concentrated on the clustering of related document, it was failed to define the temporal relation between the events.

Yang et al. [2] employed the group average clustering and the single-pass cluster for topic detection. Yang in “Automated tracking” of events from chronologically ordered document streams is a new challenge for statistical text classification. For event detection and event tracking they used clustering of similar events, but did not defined event evolution.

Carthey [3], Allan [1], and Yang [2] used the natural language processing approach by combining lexical chains with keywords for topic tracking and extracted seven types of name entities. For finding similar keywords is not sufficient to decide it’s similarity between documents. They investigate a new approach consisting of the following components:

1. Classifying documents into broad topics each of which consists of multiple events;

2. Identifying Named Entities (NE’s), optimizing their weight relative to normal words for each topic, and computing a stop word list per topic;

3. Measuring the novelty of a new document conditions on the system-predicted topic for that document. This approach gives approximate result. To find event content similarity between two events and decaying factors required temporal proximity and document proximity.
Event evolution is a new concept developed recently. Makkonen[4] and Nallapati[5] conduct investigation on event evolution as a subtopic of TDT.

Makkonen [4] was the first to investigate event evolution. The news documents within a topic are temporally linearly ordered. A narrative begins when the first story of the topic is detected. A seminal event may lead to several other events. The events at the beginning may have more influence on the events coming immediately after than the events at the later time. As we go through the event in the temporal order, we may see the evolution of events within an incident. The events and the event evolution relationship can be represented as a graph structure. Makkonen[4] claimed to use the ontology’s, including general terms, locations, names, and temporal, to measure the similarity of events. But it did not define the concept of event evolution clearly and elaborate the structure of event evolution. It is lack of details in event evolution model and fails to present any experimental evaluation results.

Nallapati[5] (Event threading within news topics) was defined the concept of event threading, given a small number of documents and events in a news topic. Their definition of event threading is close to event evolution except that they consider event threading as a tree structure instead of a graph. The concept of event threading is indeed only part of the concept of event evolution, and their proposed technique cannot cope with the complexity in event evolution relationships. Their dependence modeling methods only consider the average document similarity between events. Such methods are not effective and sufficient to identify the event evolution relationships.

Wei and Y. Chang [6] proposed an event evolution pattern discovery technique that identifies event episodes together with their temporal relationships that occur frequently in a collection of events of the same type. Their study focuses on segmenting a sequence of news stories of specific event episode and their relationship. However, relationships among event episodes discussed only consist of temporal ones. But hear relationship among event episode only consists of temporal ones, which gives approximate results.

Wei and Chang [7] proposed an event evolution pattern discovery technique that identifies event episodes together with their temporal relationships that occur frequently in a collection of events of the same type. Their work differs from prior studies in that they focus on segmenting a sequence of news stories of a specific event into event episodes and generalizing event episodes across different events of similar topics. However, relationships among event episodes discussed only consist of temporal ones. In this work, we formally define the event evolution by three logical rules. Besides, introduce the temporal relationship, event similarity, temporal proximity and document distributional proximity to identify the event evolution relationships to construct the event evolution graphs.

Given such graphical representation of the underlying structure of events in a terror any incident, users can easily navigate the development of the incident and extract specific information for their needs.

Total survey:

I. Event Detection and tracking system (EPET)[6]:

Algorithm for Event Detection and tracking using event episodes/event patterns:

Algorithm:

Step 1: Parses the documents in each document in each document sequence to produce a list of nouns and noun phrases that exclude a set of pre-specified stop words by using Feature extraction technique.

Step 2: For each document sequence, the feature selection phase select the top k-feature with the highest feature selection metric score to represent the documents in the document sequence.

Step 3: Document representation phase, each document is then representing using the representative feature selected for the document sequence to which that document belongs.

Step 4: The documents clustering phase generates used HAC algorithm.

1. Initially, each data point forms a cluster by itself.
2. Repetitively merge the two closest clusters.
3. Output the hierarchical structure that is constructed.

II. Construction of event evolution Graph[7]:

After EPET process, represent event relationship in event evolution graph. The process to construct event evolution graph is as given below

Algorithm:

Step 1: First identify and generate events from news stories by using EPET or online sources e.g. CNN, BBC.

Step 2: Find event evolution relationship by using term vector, temporal proximity, and document proximity

Step 3: Event evolution relationship represent in Directed acyclic graph (DAG) by using pruning method.

III. Applications of event evolution graph:

1. We may integrate the event evolution graphs with automatic summarization and named entity recognition techniques to provide a well-equipped web news infomediary agency.
2. The named entity recognition techniques extract the names of persons or organizations and locations involved in an event. As a result, users can easily track the persons, organizations.

3. The interconnectivity of the graph structure can also support the construction of a convenient information-browsing platform for users to navigate through the development of that news affair.

**Comparision:**

We can detect and track different topics from document sequences by using TDT and EPET techniques. TDT techniques have been attempting to detecting or clustering news stories into event, without defining or interpreting the association between these events. Therefore event evolution is a new concept developed recently. EPET technique first finds event pattern and EP used in tracking process. Event evolution is the transition development process of related events within the same topic. Makkonen[4], Nallpati[5], Wei and Y.Chang[6,7] was defined concept of event evolution by using different methods. We compare these three approaches based on their used methods. Comparision is as given below:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporal order</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Temporal proximity</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Document Similarity</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Document Proximity</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Event Threading</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Event Joining</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

**Fig. 4. Comparison between event evolution approaches**

Wei and Y.Chang[6,7] used temporal order, temporal proximity, document proximity, event threading and event joining which gives best results. By using this approach we can discover event episodes and that are used in event tracking. By using Event threading and joining represent event evolution in directed acyclic graph. So that can be used for easy browsing.

**Conclusion:**

In this paper, we survey the major event tracking and event evolution approaches. The TDT generally focuses on how to detect topics and novel event as well as news stories into different topics or event. TDT paid less attention on interpreting the interrelationship between event. Interrelationship between event is out of scope of TDT research, and thus become novel problem i.e event evolution. Event evolution graphs shows the sophisticated event interrelationship in graphical structure for easy navigation and browsing.

**References:**


