

An Overview on Event Evolution Technique

Rutuja Ahirrao
Dept. of Information Technology
PCST, Indore (M.P.)

Sachin Patel
Head Dept. of Information Technology
PCST, Indore (M.P.)

ABSTRACT

Event evolution is the most challenging task now days. There are series of events which are incident in a sequence and event evolution is used to arrange and show these events in the exact way as they occur. Lack of appropriate integration between events information causes severe problem as they misguide users.

For example due to some reason if any flight is delayed and after a span of time it is rescheduled and not updated by various resources, and now if user searches the information of that flight via internet he will not able to find the correct status. So, the major concepts based on event evolution are: Collection of events, Validation of events instantly and Publishing of events. If events collected from various sources are not validated properly the incorrect information will be reached to the user and the verified events must be published instantly so that user can access most recent document. This paper proposes a model for event evolution and generates a graph for events as they occur.

Keywords

TDT, Event Evolution Graph, Event Tracking.

1. INTRODUCTION

Now a day's almost all news channels published their news in electronic versions also. Users of online news are increasing rapidly due to lack of time for reading hard copy of newspapers. These news also is available on various search engines and they will update the recent events as soon as they occur. Updating news straight away is the major issue and lots of research is continuously performed in this field. 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. They are helpful for user to understand the evolution of events on the same topic. Event is something that happens at some specific time. Although user are able to capture the major events. There are several techniques present for event evaluation, this paper focused on TDT (Topic Detection and tracking).

TDT are used to discover topic wise data and the information is collected from various sources and may be from different language The TDT technique have been attempting to detecting or clustering news

stories into these events, without defining or interpreting the association between these events. To present the development process of incident, we must model this kind of associations between events, which we define as event evolutions. This paper focused on all major components of event evaluation and shows the future direction in this field.

Section II focused on work done in the field of event evolution, Section III based of various challenges in the field

of Event evolution and proposed model for efficient event evolution, Section IV is a conclusion part.

2. BACKGROUND

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. The most important task in our proposed system is to construct the event evolution graph for identifying the event evolution relationships from the events. TDT is used to detect topics and tracking all necessary information related to it.

2.1 TDT:

In TDT, collected documents which contain information related to events are stored in hierarchical form. The main advantage of data store in hierarchical form is it is very easy to search data and time complexity of hierarchical data is less than linear data. Data in TDT is stored in the form of tree. There must be one root node which shows/indicate specific topic for news.

For example, movies, bollywood, jokes, etc. The subparts of root are various events which belong to that category.

Figure.1 shows the example for TDT , the root for TDT is news, news are divided into various topics, Example Jokes,bollywood,movies and recipe then the subtopics are also divided into another level known as events. The leaf node of TDT tree is episode which contains main data.TDT is used to show events only not the sequence of events or relationship between them.

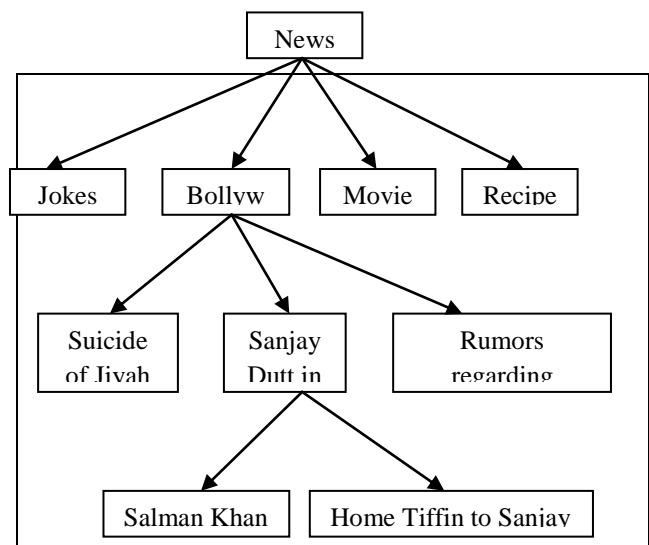


Fig. 1. Topic Detection & Tracking

2.2 Event Evolution Graph:

Event Evolution Graph represents the relationship between events. Events are some process which are some process which happens at some specific time. TDT technique uses to detecting or clustering new stories into these events, but not focused on the association between the events. The relationships between events are important in event evolution because it shows the sequence of events. Event evolution graph is used to show relationship between events in well defined structure figure 2 shows event evolution graph.

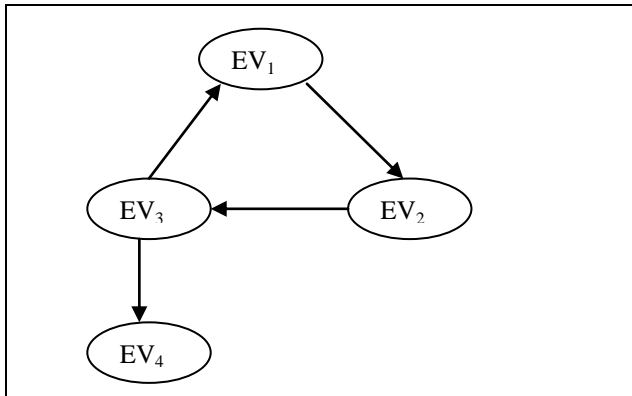


Fig. 2. Event Evolution Graph

2.2.1. Application of Event evolution

Graphs:

1. We may integrate the event evolution graphs with automatic summarization and named entity recognition techniques to provide a well-equipped web news information 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.

Event episode Identification:

Event Episode Identification is a technique to identify events and episodes related to them. There are several steps present for this task:

- a) Feature Extraction.
- b) Feature Selection.
- c) Document Representation
- d) Document Clustering.

1. Feature Extraction:

In feature extraction technique system parses the documents in each document sequence to produce a list of nouns and noun phrases that exclude a set of pre-specified stop words.

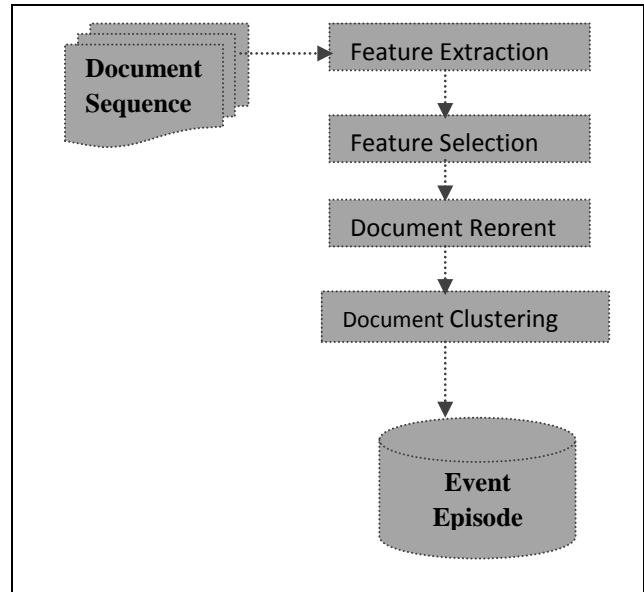


Fig. 3. Feature Extraction

2. Feature Selection:

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.

3. Document Representation:

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

4. Document Clustering:

The documents clustering phase generates episode as cluster of documents. A clustering algorithm is based on a measure of the similarity between the documents to be clustered and similarity is based on the representing feature vector.

2.3 Event Relationship Construction:

The Input for event relationship construction will be the Output of the first phase i.e. Event Episode Identification. This phase mainly include three sub-phases:

- A. Modeling Event evolution Relationship
- B. Pruning Event Evolution Graph
- C. Evolution Measures.

It should be noted that there are some online sources of well – generated news events. These sources after pre-processing, can serve as input to our proposed event evolution technique.

A. Modeling Event Evolution Relationships:

In modeling event evolution relationships, we propose to utilize

- 1 Vectors space
- 2 Event term vector
- 3 Temporal Proximity,
- 4 Document distributions proximity.

1. Vector space:

The vector space model use to measure the relatedness of events.

2. Event term vector:

In this case, we treat an event as a collection of stories/event episodes and view the information content of that event as the average of the term vectors of its stories. Thus, the content similarity between two events is the cosine similarity of their event term vectors.

3. Temporal proximity:

If two events are temporally close, an event evolution between them is more likely to exist. We use the temporal proximity between events to measure their relative temporal distance between two events.

4. Document Distributional Proximity:

The document distributional proximity is similar to temporal proximity except that it substitutes the time with the distribution of documents.

B. Pruning Event Evolution Graphs:

Given a complete graph with directed edges for each pair of events, a pruning method removes those directed edges corresponding to invalid or weak event evolution relationships and generates an event evolution graph.

There are three pruning methods, namely,

1. Static Thresholding.
2. Static Pruning.
3. Dynamic pruning.

The pruning of directed edges depends on the value of event evolution scores and the degree of event threading and event joining.

C. Evaluation Measures:

We adopt the measurement of precision and recall for our evaluation.

- 1) **Precision (P):** It is the ratio of the number of true and valid event evolution relationships retrieved by the automatic system to the total number of event evolution Relationships retrieved by the automatic system.
- 2) **Recall(R):** It is the ratio of the number of true and valid event evolution relationships retrieved by the automatic system to the total number of true and valid event evolution relationships interprets manually.

3. CHALLENGES AND FUTURE DIRECTIONS

Event Evolution can be performed using TDT & Event evolution graph but the major part of event evolution are events. If information is not collected properly for each event then we can't find out the efficient output for Event Evolution. The major challenges of Event Evolution are: Collection of events, Validation of events and Publishing of an event.

(a) Collection of Events:

The central concept of event evolution is events and the result efficiency based on event collection. For any incident the whole information (event) must be collected because each and every event has same priority.

(b) Validation of Events:

All collected events must be validated properly because any incorrect information may change the scenario of event.

(c) Publishing of event:

When events are supposed to be published they must be published properly and in a proper sequence

Example shows the event evolution process for Bodhgaya bomb blast. This example covers all the steps (phases) which are involved in Event Evolution.

3.1.Example: Serial Blast in Bodh Gaya

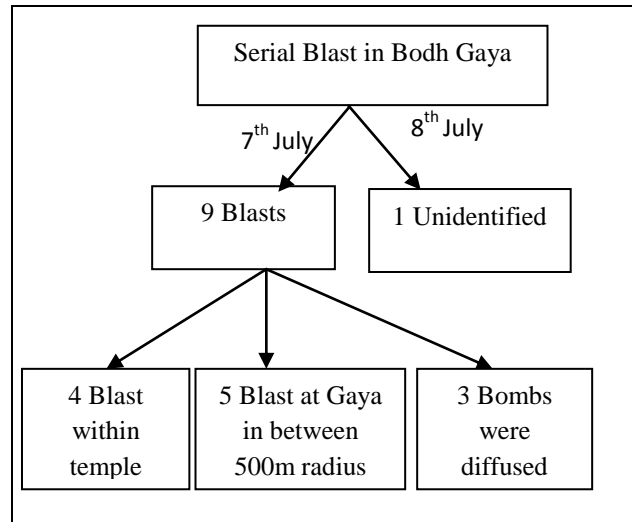


Fig. 4. Serial Blast in Bodh Gaya

Serial Blast in Bodh Gaya

Mahabodhi Temple Blast around the Mahabodhi temple complex and UNESCO world heritage site in Bodh Gaya: (7th July 2013)

On 7th July 2013, ten bombs exploded between 5:30 and 6.00 IST, at that time in the temple Sutra Chantry and meditation was running, because this time was the beginning time of their daily routine. Mahabodhi Mahavihara is an ancient and most famous temple for Buddhist. In this serrate place there is a holy Bodhi tree where Gautama Buddha is believed to have attained enlightenment.

There are 10 bomb blasts that occurred and 3 bombs got diffused by bomb disposal squad. The main incidents of these blasts have occurred in following sequence:

EV1: On 7th July 2013, the news regarding bomb blast in Mahabodhi Temple complex spread out.
EV2: News regarding 9 bomb explosions has founded.
EV3: There were 4 blasts occurred within the Mahabodhi temple complex.
EV4: There were 5 bomb blasts occurred in Gaya, in a 500-meter radius of Mahabodhi temple complex.
EV5: 3 bombs were diffused by bomb disposal squad.
EV6: On 8th July 2013, Indian home minister Mr. Sushil Kumar Shinde said that there had actually been ten blasts.

Sequence of each event i.e. EV₁ to EV₆ can be divided multiple events or sub events.

EV₃ can be divided into following events:

// **Event 3:** News regarding 9 bomb explosion has founded.

EV3₁: The first bomb exploded at 5:30 IST.

EV3₂: 2 minutes later, next bomb exploded.

EV3₃: Third bomb exploded on southern side

EV3₄: Fourth bomb exploded in the northern side of the complex.

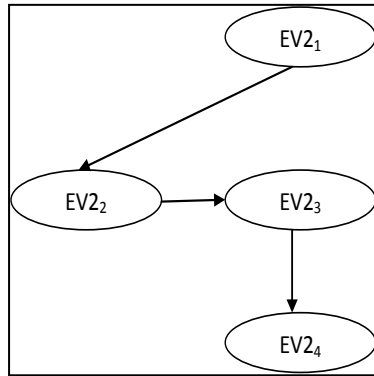


Fig. 5.

//**Event 4:** 5 blasts occurred within the Mahabodhi temple complex in a 500-meter radius.

EV4₁: A low intensity bomb exploded

EV4₂: Three bombs exploded at tergar (Tibetan) monastery.

EV4₃: One bomb exploded on a bus parked at Sujata Bypass.

//**Event 5:** three bombs were diffused by bomb disposal squads.

EV5₁: Three bombs were diffused by bomb disposal squads.

//**Event 6:** actual 10 blasts.

On 8th July 2013, Indian home minister Sushilkumar Shinde said that there had actually been 10 blasts.

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. The most important task in our proposed system is to construct the event evolution graph for identifying the event evolution relationships from the events.

In our proposed system first we identify event episodes for given a set of document sequence D pertaining to the same event type. Certain event episodes within D occur frequently and possess specific temporal relationships. After identification of an event episode we construct an event evolution graph. An event evolution graph is a directed acyclic graph (DAG) consisting of events as the nodes and event evolution relationships as the directed edges between nodes. Given a set of n distinct news stories (here we story is referred as an event episode) $S = \{s_1, s_2, \dots, s_n\}$ on a given news topic, we have a set of m events $E = \{e_1, e_2, \dots, e_n\}$ and their event timestamps $T = \{t_1, t_2, \dots, t_n\} = \tau(e_i)$. Each story is

assigned to one of the m events. A directed edge from vertex e_i to e_j is created in the event evolution graph if there is an event evolution relationship from e_i to e_j . Event e_i is the parent of event e_j and event e_j is the child of event e_i . L is the set of event evolution relationships, $L = \{(e_i, e_j) \text{ where } e_i, e_j \in E\}$. Therefore, the event evolution graph G is a directed acyclic graph, $G = \{E, L\}$.

4. CONCLUSION

The events together with their relationships are used to build a well-defined structure, *event evolution graph* which presents the blueprint of a news topic. Such event evolution graphs show the sophisticated event interrelationships in a graphical structure for easy navigation and browsing. Accordingly, users are able to capture the major events and understand the flow of the stories within the incident. The work can be extended to design an EP discovery technique that for mining EP from document sequence. EPET technique employs an EP-based filtering mechanism which is used for event tracking. So discovered episodes we present in evolutionary graph. EPET technique provides better tracking than traditional ET- technique.

5. REFERENCES

- [1] J. Allan, R. Papka, and V. Lavrenko, "On-line new event detection and tracking," in Proc. 21st Annu. Int. ACM SIGIR Conf. Res. Development Inf. Retrieval, Melbourne, Australia, 1998, pp. 37–45.
- [2] Y. Yang, T. Pierce, and J. Carbonell, "A study on retrospective and online event detection," in Proc. 21st Annu. Int. ACM SIGIR Conf. Res. Development Inf. Retrieval, Melbourne, Australia, 1998, pp. 28–36.
- [3] J. Carthy, "Lexical chains for topic detection," Dept. Comput. Sci., Univ. College Dublin—National Univ. Ireland, Dublin, Ireland, 2002. Tech. Rep.
- [4] J. Makkonen, "Investigations on event evolution in TDT," in Proc. Conf. North Amer. Chapter Assoc. Linguistics Human Language Technol., HLT-NAACL Student Research Workshop, Edmonton, AB, Canada, 2003, pp. 43–48.
- [5] R. Nallapati, A. Feng, F. Peng, and J. Allan, "Event threading within news topics," in Proc. 13th ACM Int. Conf. Inf. Knowl. Management, Washington, DC, 2004, pp. 446–453.
- [6] Wei and Y. Chang, "Discovering event evolution Graph from corpora," IEEE Trans. Syst., Man, Cybern. A, Syst., Humans, vol. 39, no. 4, July. 2009.
- [7] Wei and Y. Chang, "Discovering event evolution patterns from document sequences," IEEE Trans. Syst., Man, Cybern. A, Syst., Humans, vol. 37, no. 2, pp. 273–283, Mar. 2007.
- [8] R. Barzilay and L. Lee, "Catching the drift: Probabilistic content models, with applications to generation and summarization", in Proceedings of HLT-NAACL, 2004, vol. 2004.
- [9] C. C. Yang, X. Shi, and C. P. Wei, "Discovering event evolution graphs from news corpora", Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, vol. 39, no. 4, pp. 850–863, 2009

- [10] Q. Mei and C. X. Zhai, "Discovering evolutionary theme patterns from text: an exploration of temporal text mining", in Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining, 2005, pp. 198-207.
- [11] K. F. McCoy and M. Strube, "Taking time to structure discourse: Pronoun generation beyond accessibility", in Proc. of the 21th Annual Conference of the Cognitive Science Society, 1999, pp. 19-21.
- [12] E. Marsh and D. Perzanowski, "MUC-7 evaluation of IE technology: Overview of results", in Proceedings of MUC, 1998, vol. 7.
- [13] D. Ahn, "The stages of event extraction", in Proc. COLING/ACL 2006 Workshop on Annotating and Reasoning about Time and Events, 2006, pp. 1-8.
- [14] E. Filatova and V. Hatzivassiloglou, "Domain independent detection, extraction, and labeling of atomic events", in Proceedings of RANLP, 2003, pp. 145-152.
- [15] J.M. Wu, W. Li, Q. Lu, and B. Li, "CTEMP: A Chinese temporal parser for extracting and normalizing temporal information", in Proceedings of International Joint Conference on Natural Language Processing, 2005, pp. 694-706.
- [16] Z. Chen and H. Ji, "Graph-based event co reference resolution", in Proceedings of the 2009 Workshop on Graph-based Methods for Natural Language Processing, 2009, pp. 54-57.
- [17] H. Ji and R. Grishman, "Refining Event Extraction Through Cross-document Inference", Proc. ACL 2008, 2008.
- [18] Shi and J. Malik, "Normalized cuts and image segmentation", Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 22, no. 8, pp. 888-905, 2000.
- [19] "TDT 2004: Annotation Manual". <http://www.ldc.upenn.edu/Projects/TDT5/Annotation/TDT2004V1.2.pdf>