A Sub-topic Partition Method Based on Event Network

Wei Liu, Dong Wang, Wenjie Xu, Xujie Zhang, Zongtian Liu
School of Computer Engineering and Science, Shanghai University
Shanghai, China
{liuw, ming123, jiex, annie4sh, ztliu}@shu.edu.cn

***Abstract*—In TDT (Topic Detection and Tracking) of news stories, a topic usually contains lots of events. Because of small granularity of events, the relationships between events and topic are not closely enough to be distinguished. In this paper, topic and news stories are described by using event networks, and a network clustering algorithm EN-MST based on minimum spanning tree is proposed to discover event communities in the network. Each community is considered to be a sub-topic which could represent an aspect of the large topic as a coarse-grained concept. The experimental results show accuracy and reasonableness by using our method. In our further study, sub-topics obtained by the method proposed in this paper will be adopted to represent news stories in order to distinguish whether a news story belongs to a certain topic.**

Keywords-TDT; topic tracking; event network; community discovery; topic model; EN-MST

I. INTRODUCTION

Topic detection and tracking (TDT) is a research hotspot on information recognition, data mining and organization of news stories, so as to improve the efficiency of useful information acquisition on the Internet. A large topic usually contains many events, and there exist semantical relationships between event pairs, such as casual relations, accompany relations and follow relations. A network of events be formed based on the event relation information to represent the topic. By analyzing the network of events, some events which have more closely relationship are likely to describe one aspect of the topic, so, event clusters in the network are considered to be sub-topics. It is known that VSM (vector space model) is the main text representation method in TDT; however, the main shortcoming of this method is the lackness of semantic information. In this paper, VSM is replaced by event network to represent topics and news stories. A topic hierarchy structure is proposed including topics, sub-topics and events. Events with close relationships are put together by using network community discovery algorithm, to find sub-topics. The sub-topics will be a bridge connecting events and the topics to improve the accuracy of TDT in further study.

The remainder of the paper is organized as follows: Section II introduces the related work on TDT and network community discovery method. Section III discusses the definitions about event and event network and how to construct an event network. Section IV discusses a community discovery method based on MST to obtain sub-topics from an event network. Section V compares experiment results between the objective function in our method and other objective functions. Finally, the conclusions are given in Section VI.

II. RELATED WORK

TDT was proposed in 1996 [1] by the U.S Defense Advanced Research Projects Agency (DARPA). Then, researchers from DARPA, CMU, Dragon system company and Umass [2] began to define the main content of topic detection and tracking, and developed some initial technologies for the solution to these problems. TDT mainly focuses on three tasks[3]: topic tracking, topic detection and new event detection. A topic is considered to be only a collection of news stories. Although hierarchy was proposed for TDT in TDT-2003 [4], the subject of more fine-grained information extraction problems have not yet been considered. In [5], every news story was considered to be an event, and a news probability generation model was proposed. News event generated model including person, places, content and time was automatically built in a unified framework. In [6], a set of sudden outbreak lexical items were extracted in a concentrated time window, and further emergency was identified according to lexical items. In [7], W. Lam extended related words according to the statistical results of vocabulary in news stories, and events were identified by a method similar to Single-Pass cluster. A topic model based on event and event developing relations was proposed by Makkonen [8], but the detail for the method was not provided. Nallapati [9] gave a more specific concept of event identification and event relation extraction with a given topic, in addition, provided the corresponding evaluation methods and test data. Anicic introduced the concept of Event Processing (EP) and a stream reasoning method based on a language called EP-SPARQL [10], which provided syntax and formal semantics to detect compound events.

Another related research hotspot is network community discovery. Girvan-Newman algorithm was proposed by Girvan and Newman [11], whose method was to get communities by finding the edge with the highest score of betweenness and remove it from the network. Newman [12] proposed a weighted network community discovery method based on edge betweenness. He proposed that weighted graphs be mapped into multi-graphs, the betweenness of all the edges be calculated and the largest one be removed in turn until it reaches a reasonable structure. Noack [13] introduced two energy models whose minimum energy layouts represented the cluster structure, one based on repulsion between nodes and the other based on repulsion between edges. Grygorash proposed a graph cluster method based on MST (minimum spanning tree) [14], and removed the edges whose weights were above the threshold to get the community structures.

However, most of the TDT methods were based on traditional VSM which lacks semantic information and the above community discovery algorithms should be improved to be used in our research. In this paper, the content of text is described by using event network instead of the traditional VSM method, and the event network is divided into event clusters by a method based on MST to extract sub-topics from a large topic.

III. EVENT AND EVENT NETWORK

A. Definitions

In the field of the TDT, a topic is a collection of several related events, including a central event and some other events related to it. A story is a news report closely related to the topic, which contains two or more independent clauses for stating an event. A story is usually a statement of some aspects of a certain topic, while a topic contains all the content of the related stories. An event involves some participants, environment and some other elements. So we introduce the definitions of event, event class and event ontology in [15] which is the foundation of our work.

Definition 1 (Event): event is defined as a thing happens in a certain time and environment, which some actors take part in and show some action features. Event *e* can be defined as a 6-tuple formally:

$$e ::=_{def} < A, O, T, V, P, L >$$

where *A* means an action set happen in an event; it describes the process of the event happens. *O* means objects taking part in the event, *T* means the period of time that the event lasting. The time period includes absolute time and relative time. *V* means environment of the event, such as location of the event; *P* means assertions on the procedure of actions execution in an event; *L* means language expressions. In this paper, we use the event elements *A*, *T* and *V* to represent events. Different events have different elements. A word or a phrase which expresses an event happening can be called the denoter of the event. Each event has an event denoter in text. However, it is not adequate to distinguish events. Such as the *Sichuan earthquake* and *Japanese earthquake*. Although the event denoters *earthquake* are the same, they do not mean the same event due to the different places they occurred in. Therefore, action, time and location are important to represent events.

Definition 2 (Event Class): event class is a set of common characteristics of the event. It can be expressed by *EC*. $EC = (E, C_1, C_2, \cdots, C_6)$;

where E is the set of events, which is an extension of the event class. $C_i = \{c_{i1}, c_{i2}, \cdots, c_{im}, \cdots\}$ $(1 \le i \le 6, m \ge 0)$ is the intension of event class, and is the set of common characteristics in the ith element of E; c_{im} is one of the common characteristic in the ith element of each event.

In this paper, the definition of event network is proposed as follow.

Definition 3 (Event Network): an event network (*EN*) is a directed acyclic graph that consists of a set of nodes and edges. The nodes are events, and the edges are event relations.

$EN ::= (Events, Edges)$

$Events = \{e_1, e_2, \ldots, e_n\}$

$Edges = \{<e_i, e_j, r_{ij}>, <e_x, e_y, r_{xy}>, \ldots\}$ $(1 \le i, j, x, y \le n)$

$r = \{Correlation, Causal, Accompany, Follow\}$

where, *EN* denotes event network, which contains the set of event nodes *Events* and event relations *Edges*. In *Events*, e_i represents an event, r_{ij} represents the relation between e_i and e_j. We define four different relations between events in our event network model, including:

Correlation Relation：if two events appear in the same story and have common event elements, such as time, location or objects, they are correlated.

Causal Relation： if event e_1 causes event e_2, there exists causal relation between e_1 and e_2. Causal is the most important relation between two events. It not only reflects the interaction between events, but also reflects the time sequence of events. For example:

June 1, in the Afghan city of Kandahar, at least 40 people were killed and 60 wounded in an explosion at a mosque.

where *explosion* caused *killed* and *wounded*, so they have the relation of causal.

Accompany Relation：if two or several events almost happen at the same time, they have the relation of accompany. They are often series of actions caused by the same event. For example:

A large truck overturned in the corner, then the electrical tools are knocked out, and clips are thrown out to the ground.

where *overturned*, *knocked out* and *thrown* exist relations of accompany.

Follow Relation: two events have the relation of time sequence, such as *earthquake* and *rescue*, *wake up* and *teeth brushing*.

Definition 4 (Event Ontology). An event ontology is a formal, explicit specification of a shared event model that exists objectively, denoted as EO. The structure of event ontology can be defined as a 3-tuple：

$$EO : \{ECs, R, Rules\}$$

where *ECs* is the set of all events, R indicates all relations between events. *Rules* are expressed in logic languages, which can be used to describe the transformation and inference between events.

In this paper, by using event relations, we connect event instances that appear in text and construct an event network to represent the text. In contrast to word frequency method of VSM, the event network contains more semantic information: events and their relations.

B. Transmission rules of event relationship

The implicit relations will be extracted between events according to the transmission rules in the above event relations.

Causal Relation: The causal relation is transitive. If event B is caused by event A and event C is caused by event B, then, A causes C.

Accompany Relation: If the relationship between event A and event B are accompany, event B and event C are accompany, then there exists accompany relation between event A and event B.

Follow Relation: If event B follows event A and event C follows event B, then C follows A.

In CEC corpus [16] (An emergency corpus which contains 200 Chinese news stories annotated in Semantic Intelligence Lab of Shanghai University.), obvious relations between events have been annotated, and it is difficult to extract all the relationships manually. In order to extract event relations as many as possible to construct an event network, some implicit relations will be annotated by using these transmission rules above.

C. Quantified the relationship between events

While an event network with semantical relationships has been constructed, it is necessary to transform the event network into a weighted network in order to discovery communities of events. The method is introduced in Section IV. Each of the relationship will be mapped into a corresponding weight in the weighted graph, such as the causal relationship, events with causal relations usually describe the developing situation of the topic, which are more important to represent the theme of news stories. Therefore, the weight of causal relation should be larger than other types of relations.

In order to quantify the event relations with weights, The method in article [17] is introduced which was used to calculate the weights between event classes: Choose 200 stories as sample corpus, then add up each pair of event classes on the frequency and calculate the impact factor of them. For one text *d* in the text collection *N*, F_{ei} means the frequency the event class e_i appears in *d*, F_{ej} means the frequency the event class e_j appears in *d*. The formula for calculating impact factor between e_i and e_j is defined as follows:

$$w^d{}_{ij} = \begin{cases} \frac{F_{ej}}{F_{ei}}, & if\ \ F_{ei} \neq 0 \\ 0, & if\ \ F_{ei} = 0 \end{cases} \tag{1}$$

If $w^d{}_{ij}$ >1, it is normalized, $w^d{}_{ij}$ =1.

For the whole text collection *N*, the formula for calculating impact factor between e_i and e_j is defined as follow:

$$w_{ij} = \frac{\sum_{d \in M} w^d{}_{ij}}{|\ M\ |} \tag{2}$$

M is the text collection where each text contains event class e_i.

The steps of calculation are: ①Calculate the impact factors of event class pairs in the same text; ②Normalize and calculate the average impact factors of event class pairs in the text collection, which will avoid unreasonableness caused by the large impact factors in a single text.

IV. COMMUNITY DISCOVERY ALGORITHM BASED ON MINIMUM SPANNING TREE

A. Three-layers model structure of topic

In this paper, a three-layers model structure which contains *Event*, *Sub-topic* and *Topic* is proposed. In contrast to a large topic, a news story contains only a few sub-topics (usually less than 4 sub-topics). Therefore, a coming news story can be represented by some event communities and similarity degree between topic and new stories in the sub-topic level, thus to improve the accuracy in TDT.

Topic: a seminal event or activity, along with all directly related events and activities.

Sub-topic: one aspect of the whole topic. A large topic usually contains some sub-topics and each sub-topic focuses on a small aspect of the topic.

Event: a thing happens in a certain time and environment. In news stories, most of the events have implicit elements.

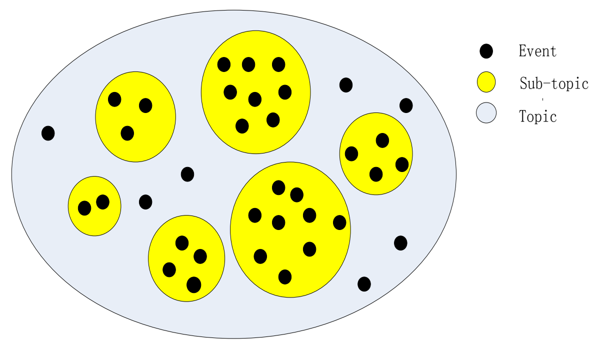

Figure 1. The three-layers model structure of topic.

As Figure 1 shows, a topic contains a lot of events. The granularity of events are too small to describe a topic. Therefore, the sub-topic level which is represented by event communities is proposed to connect events and topic.

B. Communities in network

Community is a description of the close relationships between events. A property of community structure, in which network nodes are joined together in tightly-knit groups and between which there are only looser connections. Currently, there is no recognized evaluation criteria for the community structure. In 2003, Newman proposed the concept of modularity (which is also represented by Q-function) [12]. This quantity is defined as the fraction of edges that fall within communities minus the expected value of the same quantity if edges are assigned at random. Partitioning result depends on the given community memberships and the degrees of vertices. Q-function is defined as follow:

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \tag{3}$$

As to undirected and unweighted graph, if there is connection between node *i* and node *j*, A_{ij} =1, otherwise,

A_{ij} =0. $m = \frac{1}{2}\sum_{ij} A_{ij}$ means the edge number in the network. c_i is the community which node *i* belongs to. If $c_i = c_j$, the value of δ function is 1, otherwise, the value of δ function is 0. $\frac{k_i k_j}{2m}$ means the probability of an edge existing between vertices *i* and *j*. Where k_i is the degree of *i*.

Although the Q-function is biased, most of the community partitioning algorithms choose Q-function as a measure means of the clustering quality. A weighted network can be mapped into a multi-graph, and A_{ij} can also represent the weight between *i* and *j*. That means Q-function is also suitable for weighted network. However, Q-function is difficult to obtain the optimal community structure in the sparse graph.

C. Communitiy discovery algorithm based on minimum spanning tree

Most of the community discovery algorithms are for undirected graphs which can not be utilized in event networks. In this paper, an improved algorithm based on minimum spanning tree (MST) is proposed for event clustering.

The traditional clustering based on MST is a splitting algorithm. In the traditional algorithm, the edges with larger weights are removed to form a forest, and every tree in the forest is a cluster. The time complexity of the algorithm is $O(mlogn)$ (*m* is the number of edges, *n* is the number of vertices). Any shape and high dimensional data clustering problem can be processed. However, the MST algorithm has a great complexity, and it is difficult to determine when the algorithm reaches the optimal community structure. Therefore, an algorithm *EN-MST* (*Event Network MST*) using the event relation information of event network for event clustering is proposed. By querying the event ontology, an event network may contain the four types of relations in Definition 3. Because causal relation is the most important relationship, event network will be simplified as follow: remove the relations which are not causal relation and if they have common neighbors, the edges between them can be removed. While utilizing MST algorithm, a simplified network will be faster. An example is shown in Figure 2：

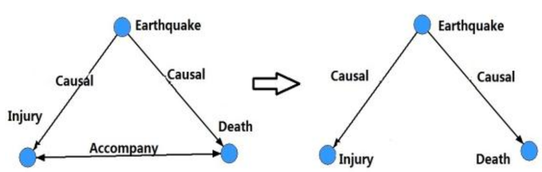

Figure 2. Simplify the event network

In the event network, the greater the weight, the relationship between the two events is more close, and the probability they are in the same community is greater. Therefore, the initial step of the algorithm is to construct a tree with the largest weights which is opposite to MST, that is, the edges with larger weights will be chosen rather than the smaller ones. The adjacent nodes in the tree have close relationship and they may be put in the same community. There exists several branches in a tree and nodes in the same branch are also probably in the same community. However, the granularity of community is too small just by using branch as the criteria, furthermore, the edge weight is not be considered. Perhaps, the community obtained is not accuracy.

In this paper, Q-function method is improved by considering the branches information in the MST, which is defined as follow to avoid the shortcomings of Q-function:

$$Q_{branch} = \frac{1}{2m}\sum_{ij}\left[A_{ij} - \frac{k_i k_j}{2m}\right]\delta(c_i, c_j)\mu(b_i, b_j) \qquad (4)$$

where b_i means the branch that node *i* belongs to. In MST, if there are not nodes whose degrees are more than 3 in the path from *i* to *j*, node *i* and *j* are in the same branch, $\mu(b_i, b_j)$ =1, otherwise, $\mu(b_i, b_j)$ =0.

In the community discovery algorithm EN-MST, there are two main processes. The first one is to generate a minimum spanning tree and the second one is to remove the edges in the tree to get event communities based on the Q_{branch} function.

The steps of EN-MST are described in the TABLE I:

TABLE I. STEPS OF EN-MST ALGORITHM

Algorithm: *EN-MST*

Step 1. *Remove the edges with accompany relation and follow relation;*

Step 2. *For the edges with causal relation, query the impact factors between two corresponding event classes from event ontology, and assign them to these edges, thus，get a weighted event network EN';*

Step 3. *Calculate all the paths between event pairs. A 2-dimensions matrix B is to save status whether the two events are in the same branch. If node i and j are in the same branch,* $B_{ij}=1$*, otherwise,* $B_{ij}=0$*;*

Step 4. *Generate a minimum tree ENTree from EN' by the method of Prim.*

Step 5. *Remove the edge with the lowest weight, query the matrix B to calculate the value of* Q_{branch} *and save the previous network condition.*

If (the value of Q_{branch} *is higher than before)*

Set Q_{branch} *as the largest one;*

Else

Recover the previous network condition and the next edge is set as the lowest weight ;

Step 6. *Repeat Step 5 until all the edges are checked. Each sub-graph is a community in the event network.*

Figure 3 shows a minimum spanning tree which is generated from an event network. Three event communities are partitioned by EN-MST is shown in Figure 4.

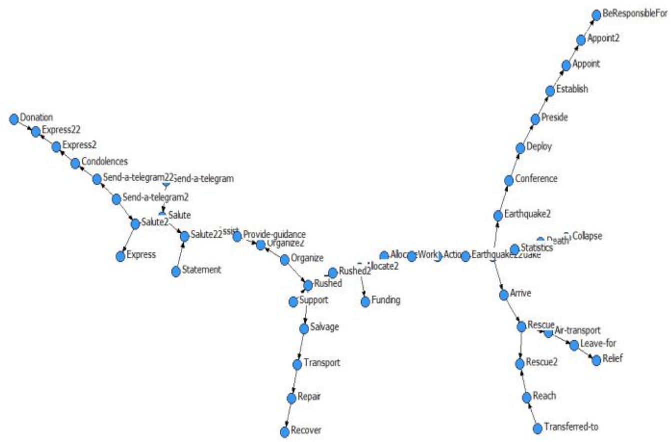

Figure 3. A minimum spanning tree by using prim algorithm

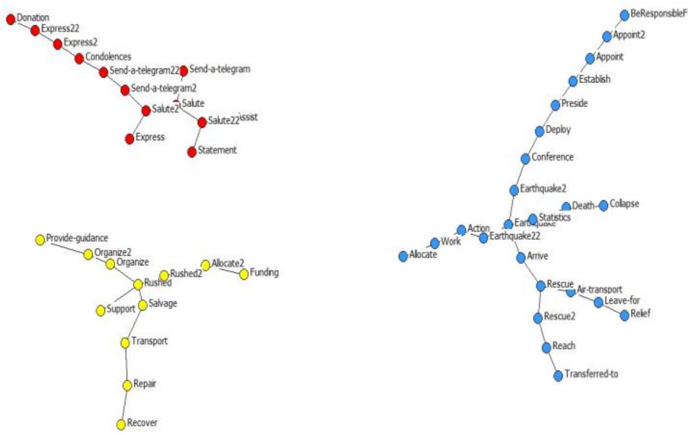

Figure 4. The communities of earthquake event network generated by the EN-MST algorithm

D. The efficiency of EN-MST

In event network, semantic information is considered and unimportant edges are removed, which will improve the efficiency of minimum spanning tree algorithm. Compared to Girvan-Newman algorithm whose complexity is $O(m^2n)$, EN-MST only removes the edges according to their weights rather than calculates the edge betweenness, and the number of the edges is the number of nodes minus 1. Therefore, EN-MST have higher efficiency. The time complexity of the algorithm in generating a minimum tree is $O(m\log n)$ (m is the number of edges, n is the number of vertices). In the process of initializing the matrix B, the time complexity is $O(n^2)$, and the time complexity in the removal of edges is $O(m)$.

V. EXPERIMENT AND RESULTS ANALYSIS

A. Experimental corpus

Impact factors between event classes are calculated by CEC corpus including different types of emergencies such as *earthquake*, *traffic accident*, *bromatoxism*, *fire disaster* and *terrorist attack*. Each of the stories from the corpus is considered to be a single topic. For each emergency, four stories are selected to be samples, and sub-topics are partitioned manually according to them. The rest of the stories in the corpus are selected as test corpus.

TABLE II. SUB-TOPICS IN SAMPLES AND THEIR CORRESPONDING NUMBER OF EVENTS

Sub-topics	The number of events
Emergency scene	58
The rescue	41
Remedial work	19
Cause of the incident	16
International concern	10
The donation	7

B. Experimental results and analysis

In the test corpus, an event network is constructed for each story, and EN-MST algorithm is used for sub-topic partition. In the experiment, the result of the sub-topic partition is represented by a relation between event nodes, that is, two events are either assigned to the same sub-topic or different sub-topics. In a data set with n events, there are $n(n-1)/2$ event pairs. *RI* is to evaluate the performance of the algorithm according to the correct event pairs which is defined as follow:

$$RI = \frac{\#CD}{n(n-1)/2} \tag{5}$$

where $\#CD$ is the number of the correct decision on the event pairs. $\#CD = A+C$. A is the number of the events pairs in which the two events belongs to the same sub-topic both according to the sample and the algorithm result. C is the number of the event pairs in which the two events belong to the different sub-topics both according to the sample and the algorithm result. It can be seen that $0<RI<1$，The value of *RI* is larger，the performance of the algorithm is better.

In the test set of the emergency corpus, 20 events are selected at random. The result of the sub-topic partition algorithm is compared to the sample corpus and the *RI* values are calculated by formula (5). In the step of generating a minimum tree, according to MST algorithm, the MST results are not unique, which may cause the variation of community discovery results. Therefore, we repeat this process 10 times, then take the average of *RI* values.

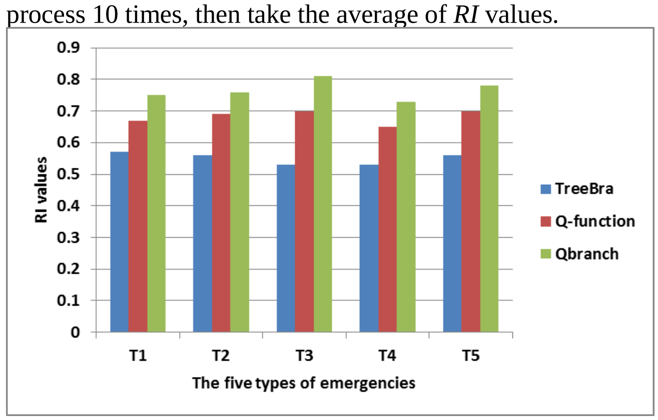

Figure 5. The result of sub-topic parition based on three methods.

TABLE III. THE AVERAGE NUMBER OF SUB-TOPICS USING THE THREE OBJECTIVE FUNCTION

Topics	The average number of sub-topics		
	TreeBra	Q-function	Q_{branch}
Earthquake	5.6	2.3	2.8
Traffic accident	5.4	3.1	3.9
Bromatoxism	4.8	2.5	3.4
Fire disaster	5.8	3.3	3.7
Terrorist attack	6.2	3.6	3.9

In Figure 5, the X-coordinate represents the five kinds of emergency corpus, the Y-coordinate represents *RI* values. The *RI* value by using the method of TreeBra is the lowest. Although it is reasonable to partition community by removing the edges connecting two branches in a MST, the size of each sub-topic is too small,. Sub-topics contain less event nodes compared to the sample sub-topics. Thus, the objective partition result only by choosing the branches in the tree is not corresponding to the sample. Considering the edge weights, the Q-function method has higher *RI* values than TreeBra, because in an event network, the weights of edges play an important part in the removal of node edges and distinguishing whether two events are in the same community. However, the sub-topic partition result by the Q-function is not very reasonable due to ignoring the information of nodes in the same branch. Compared to the above two methods, the algorithm using the objective function Q_{branch} reaches the highest RI value, in which the advantage of TreeBra and Q-function are taken.

Besides *RI* value, TABLE III shows the average number of sub-topics in each event network. A sub-topic may either be the core event which can represent a topic, or the collection of some related events. The granularity of sub-topic is a important factor to evaluating the partitioning result. In the method of TreeBra, the sub-topic number is the most, which means the granularity of the sub-topics is the smallest. In contrast to TreeBra, the size of the sub-topics obtained by the Q-function method is the largest, that means a sub-topic usually contains some events which do not belong to it. In comparison with TreeBra and Q-function, in the method of Q_{branch}, the number of events in each sub-topic is reasonable, and it is the most corresponding to the reality. Therefore, the EN-MST algorithm has the best performance in sub-topic partition.

VI. CONCLUSION AND FUTURE WORK

In this paper, topics and news stories are represented by event networks, which are divided into event clusters by EN-MST to extract sub-topics. A topic hierarchy structure is proposed, which includes topics, sub-topics and events. In comparison of the experiment results, EN-MST gets the highest *RI* values, and the granularity of the sub-topics are the most close to the sample. However, the MST is variation which will influence the results. In our further research, we will find a method to take the place of MST. Similarity calculation using sub-topic information will also be studied to improve the accuracy in TDT.

ACKNOWLEDGMENT

This paper is supported by the Internal Natural Science Foundation of China, No. 60975033 and No.61074135, and Postgraduate Innovation Fund of Shanghai University (SHUCX112161).

REFERENCES

[1] J. Allan, J. Carbonell, G. Doddington, J. Yamron, and Y. Yang, "Topic detection and tracking pilot study: Final report," in Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop, 1998, pp. 194-218.

[2] J. Allan, V. Lavrenko, D. Frey, and V. Khandelwal, "UMass at TDT 2000," TDT 2000 Workshop Notebook,2000, pp. 109-115.

[3] Y. Hong, Y. Zhang, T. Liu, and S. Li, "Topic Detection and Tracking Review," Journal of Chinese Information Processing, vol. 21, pp. 71-87, 2007.

[4] J. Fiscus, "Results of the 2003 Topic Detection and Tracking Evaluation," in Proc. LREC, 2003.

[5] Z. Li, B. Wang, M. Li, and W. Y. Ma, "A probabilistic model for retrospective news event detection," In Proceedings of SIGIR, 2005, pp. 106-113.

[6] G. P. C. Fung, J. X. Yu, P. S. Yu, and H. Lu, "Parameter free bursty events detection in text streams," In VLDB, 2005, pp. 181-192.

[7] W. Lam, H. Meng, K. Wong, and J. Yen, "Using contextual analysis for news event detection," International Journal of Intelligent Systems, vol. 16, pp. 525-546, 2001.

[8] J. Makkonen, "Investigations on event evolution in TDT," In Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, 2003, pp. 43-48.

[9] R. Nallapati, A. Feng, F. Peng, and J. Allan, "Event threading within news topics," In Proceedings of the Thirteenth ACM conference on Information and Knowledge Management, 2004, pp. 446-453

[10] D. Anicic, P. Fodor, S. Rudolph, and N. Stojanovic, "Ep-sparql: a unified language for event processing and stream reasoning," in Proceedings of The 20th International Conference on World Wide Web, New York, 2011, pp. 635-644.

[11] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," Proceedings of the National Academy of Sciences, vol. 99, p. 7821, 2002.

[12] M. E. J. Newman, "Analysis of weighted networks," Physical Review E, vol. 70, p. 056131, 2004.

[13] A. Noack, "Energy models for graph clustering," Journal of Graph Algorithms and Applications, vol. 11, pp. 453-480, 2007.

[14] O. Grygorash, Y. Zhou, and Z. Jorgensen, "Minimum spanning tree based clustering algorithms," IEEE International Conference on Tools with AI, 2006, pp. 73-81.

[15] W. Liu, W. Xu, J. Fu, Z. Liu, and Z. Zhong, "An extended description logic for event ontology," Advances in Grid and Pervasive Computing, pp. 471-481, 2010.

[16] J. Fu, W. Liu, Z. Liu, and S. Zhu, "A Study of Chinese Event Taggability," IEEE Computer Society 2010, 2010, pp. 400-404.

[17] Z. ZHONG, "Event Ontology and Its Application in Query Expansion," [PhD thesis] Shanghai University, 2010.