Exploring Social Context from Buzz Marketing Site – Community Mapping Based on Tree Edit Distance –

Shoichi Higuchi
Kyushu Institute of Technology
Kawazu 680-4, Iizuka 820-8502, Japan
syou_hig@dumbo.ai.kyutech.ac.jp

Takako Hashimoto
Chiba University of Commerce
Kohondai 1-3-1, Ichikawa, Chiba, 272-8512, Japan
takako@cuc.ac.jp

Tetsuji Kuboyama
Gakushuin University
Mejiro 1-5-1, Toshima, Tokyo, 171-8588, Japan
kuboyama@tk.cc.gakushuin.ac.jp

Kouichi Hirata
Kyushu Institute of Technology
Kawazu 680-4, Iizuka 820-8502, Japan
hirata@ai.kyutech.ac.jp

***Abstract*—In this paper, we design a new method to explore the social context as a community mapping from a buzz marketing site. In this method, after extracting significant topical terms from messages in buzz marketing sites, first we construct a snapshot co-occurrence network at each time stamp. Next, we organize topic hierarchical structures from each co-occurrence network by using the modularity. Then, we explore a community mapping as an LCA-preserving mapping between topic hierarchical structures and a topic mapping as a correspondence in a community mapping. Hence, we can extract a topic transition as topic mappings for the same topic. Finally, we give experimental results related to the East Japan Great Earthquake in the buzz marketing site.**

***Keywords*-buzz marketing site, tracking topic transitions, community mapping, tree edit distance, LCA-preserving distance**

I. Introduction

A buzz marketing site, in which individual users post their opinions and gradually build their consensus, is recognized as one of pervasive collaboration. After the topical problems, especially after a disaster, people's behavior is influenced by this pervasive collaboration. Therefore, by tracking topic transitions over time on a buzz marketing site, we can gain a rich insight into exploring its social context.

However, time series topic exploration is difficult, because messages in buzz marketing sites are written in a colloquial style and topics flexibly come and go. An effective technique to analyze topic exploration is required, but existing techniques cannot provide efficient solutions yet because of the complexity for computing topic exploration over time.

To address the problems, this paper proposes a novel approach using a community mapping technique based on approximate tree pattern matching. In this method, first, we extract significant topical terms from messages in buzz marketing sites. Next, we construct a snapshot co-occurrence network at each time stamp. Then, we organize topic hierarchical structures for each snapshot by using a network modularity measure [1].

In this paper, structural correspondences over time-series networks are tracked using a distance measure between two trees, in particular, the LCA (least common ancestor)-preserving distance (or degree-2 distance [3]) as a kind of the tree edit distance [2]. We call the resulting LCA-preserving mapping through the LCA-preserving distance a community mapping between topic hierarchical structures and the correspondence of nodes a topic mapping. Hence, we visualize topic transitions as the sequence of topic mappings over time.

This paper is organized as follows. Section II refers to existing researches. Section III illustrates our proposed method to explore topic transitions by using a community mapping extracted from buzz marketing messages. Section IV shows the experimental results using buzz marketing messages after the East Japan Great Earthquake. Finally, Section V concludes this paper.

II. Related Works

Regarding the topic exploration in social media, Sekiguchi *et al.* [4] treated recent blogger posts and analyzed word co-occurrence and the repeating rate of word. They visualized the relation between words and showed topics in social media through the visualization results. Asur *et al.* [5] investigated trending topics on Twitter. They proposed the simple model based on the number of tweets and found that the resonance of the content with the users of the social network plays a major role in causing trends. Wang and Araki [6] proposed a graphic reputation analysis system for Japanese products. All of the above researches require that specific topics/products be targeted in advance.

On the other hand, our proposed method does not focus on specific topics. It can flexibly show unspecified topic transition by taking into account topic structure changes over time.

Regarding research on detecting temporal relations, Radinsky *et al.* [7] proposed Temporal Semantic Analysis (TSA), a semantic relatedness model, that captures the words' temporal information. They targeted words in news archives (New York Times, etc.) and used the dynamic time warping technique to compute a semantic relation between pre-defined words. Wang *et al.* [8] proposed time series analysis which has been used to detect similar topic patterns. They focus on specific burst topic patterns in coordinated text streams and try to find similar topics. Zhou *et al.* [9] addressed the community discovery problem in a temporal heterogeneous social network of published documents over time. They showed temporal communities by threading the statically derived communities in consecutive time periods using a new graph partitioning algorithm. Qiu *et al.* [10] focused on the problem of discovering the temporal organizational structure from a dynamic social network using a hierarchical community model.

The above existing methods focused on specific topics/communities and analyzed their transition. In our method, on the other hand, unspecified topic exploration can be analyzed by mapping communities over time using a distance measure between graph structures. As far as we know, no topic transition methods exist that use a community mapping based on a tree edit distance.

III. Proposed Method

Our proposed method consists of the following 4 steps. Note that Step A to C are same as our previous work [11].

- Step A: Extracting significant topical terms.
- Step B: Constructing co-occurrence networks.
- Step C: Clustering by modularity and organizing topic hierarchical structures.
- Step D: Tracing structural correspondences over time based on LCA-preserving mapping.

A. Extracting significant topical terms

Step A acquires messages $D = \{d_i\}$ that contain the topical term such as the *great earthquake* from the target buzz marketing site. One message is defined as one document and the step retrieves it as the following tuples:

$$d_i = (MID_i, Posted_i, Title_i, Content_i).$$

Here, MID_i is an ID of each document, $Posted_i$ is a datetime that each document was posted, $Title_i$ is a title of each document and $Content_i$ is a text of each document.

The step then extracts terms that are nouns, verbs, adjectives, and adverbs from $Content_i$ of each d_i by morphological analysis, and the score of an individual term in d_i is calculated using RIDF (residual IDF) [12] measure. Finally, the words with high RIDF value are selected as a list of keywords $KW = \{kw_i\}$.

$$kw_i = (MID_i, Posted_i, \{w_{ij}\}).$$

Here, $\{w_{ij}\}$ is a list of extracted terms from document d_i with high RIDF value greater than a threshold τ.

B. Constructing co-occurrence networks

Step B makes relevance network graphs of words appearing in $\{kw_i\}$. Network graphs of related words are obtained using co-occurrence frequencies in a document, that is, for a subset X of D and terms w_{ij} and w_{ik}, if X contains w_{ij} and w_{ik}, then connect w_{ij} and w_{ik} by an edge.

In our method, the posted date is delimited by an appropriate period (e.g. monthly, weekly, or daily), and D is grouped by the period as $\{X_k\}$. Step B then constructs a co-occurrence network $G = \{g_k\}$ from $\{X_k\}$, so that the graphs are made in time-series. Fig. 1 shows an example graph structure constructed by Step B.

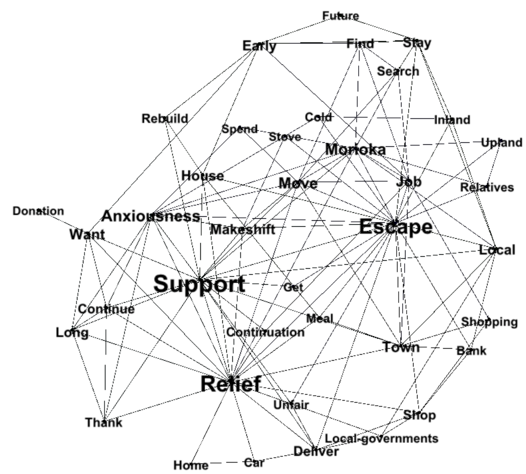


Figure 1. An example of co-occurrence network.

C. Clustering by modularity and organizing topic hierarchical structures

In Step C, we organize the topic hierarchical structures for each snapshot network by using a network modularity measure which expresses the quality of division of a network into modules or communities. In this paper, we define a community as a topic. By maximizing the modularity value, we can grasp the division of graph structure and detect topics in graphs. The modularity Q for a graph is defined as follows:

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

Here, A_{ij} is a weighted adjacent matrix of a graph, $k_i = \sum_j A_{ij}$ is the sum of the weights of the edges connected to a vertex i, c_i is the community to which vertex i is assigned, $\delta(u, v)$ is a function that $\delta(u, v) = 1$ if $u = v$

and 0 otherwise, and $m = \frac{1}{2}\sum_{i,j} A_{ij}$. It holds that $Q \in [0, 1]$, and, when Q is near to 1, we can decompose graphs with high quality.

Since the modularity optimization problem is computationally hard, we adopt the algorithm to find high modularity partitions from graphs in a short time; that unfolds a complete topic hierarchical structure [13]. Fig. 2 shows an example of a topic hierarchical structure. In Fig. 2, there are two level hierarchies. On level 1, there are 2 topics. On level 2, each level 1 topic is divided into 2 subtopics respectively.

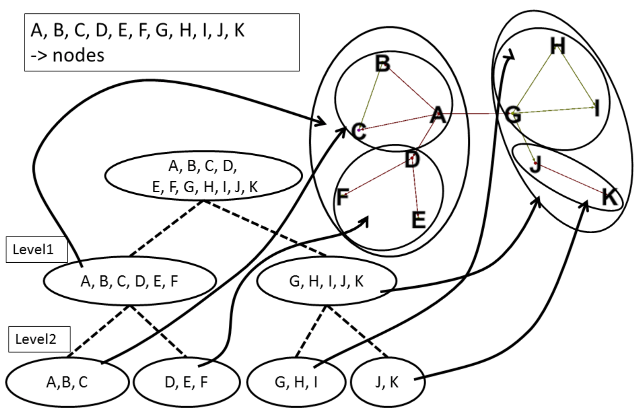


Figure 2. An example of a topic hierarchical structure obtained by clustering based on modularity.

D. Tracing structural correspondences over time based on LCA-preserving mapping

In STEP D newly introduced in this paper, we regard topic hierarchical structures as rooted (a tree has a root), labeled (every node is labeled by Σ) and unordered (a left-to-right order among siblings is not given) trees, and compare them by using a tree edit distance [2]. The tree edit distance is formulated as the minimum cost to transform from a tree to another tree by applying edit operations of a substitution, a deletion and an insertion to trees. However, it is known that computing the tree edit distance for unordered trees is intractable [14], [15].

For our purpose, we can assume that every correspondence between topic hierarchical structures preserves the least common ancestor (LCA, for short). The tree edit distance under this assumption is known as an LCA-preserving distance (or a degree-2 distance [3]). In operation, the LCA-preserving distance is the restricted tree edit distance that a deletion and an insertion are allowed either to leaves or nodes with one child.

For a tree T, we denote the set of nodes in T by $V(T)$. For $u, v \in V(T)$, we denote either v is an ancestor of u or $v = u$ by $u \leq v$. Then, we call the ancestor $w \in V(T)$ of u and v to which are nearest the least common ancestor (LCA) and denote it by $w = u \sqcup v$. Also the degree $d(v)$ of a node v is the number of children of v and the degree $d(T)$ of T is $\max\{d(v) \mid v \in V(T)\}$.

For trees T_1 and T_2, let $M \subseteq V(T_1) \times V(T_2)$. We say that M is a(n unordered Tai) mapping if, for every pair (v_1, w_1) and (v_2, w_2) in M satisfies that (1) $v_1 = v_2 \iff w_1 = w_2$ (one-to-one condition) and (2) $v_1 \leq v_2 \iff w_1 \leq w_2$ (ancestor condition).

Let M be a mapping between T_1 and T_2. Let I and J be the sets of nodes in T_1 and T_2 but not in M. Also let γ be a cost function from a pair of nodes (containing an empty node ε) to a positive number. Then, the cost $\gamma(M)$ of M is given as follows:

$$\gamma(M) = \sum_{(v,w)\in M} \gamma(v, w) + \sum_{v\in I} \gamma(v, \varepsilon) + \sum_{w\in J} \gamma(\varepsilon, w).$$

It is known that the minimum cost of all possible mappings coincides with the tree edit distance [2].

We say that a mapping M is an LCA-preserving mapping (or a degree-2 mapping [3]) if every pair (v_1, w_1) and (v_2, w_2) in M satisfies that $(v_1 \sqcup v_2, w_1 \sqcup w_2) \in M$. An LCA-preserving distance (or a degree-2 distance [3]) $\lambda(T_1, T_2)$ between T_1 and T_2 is defined as the minimum cost of all possible LCA-preserving mappings between T_1 and T_2.

According to the recurrence as Fig. 3, Zhang *et al.* [3] have designed an algorithm to compute $\lambda(T_1, T_2)$ running in $O(n^2 d \log d)$ time, where $n = \max\{|V(T_1)|, |V(T_2)|\}$ and $d = \min\{d(T_1), d(T_2)\}$.

$$
\begin{aligned}
&\lambda(\emptyset, \emptyset) = 0.\\
&\lambda(T_1[v], \emptyset) = \lambda(F_1[v], \emptyset) + \gamma(v, \varepsilon).\\
&\lambda(F_1[v], \emptyset) = \sum_{1\leq i\leq d(v)} \lambda(T_1[v_i], \emptyset).\\
&\lambda(\emptyset, T_2[w]) = \lambda(\emptyset, F_2[w]) + \gamma(\varepsilon, w).\\
&\lambda(\emptyset, F_2[w]) = \sum_{1\leq j\leq d(w)} \lambda(\emptyset, T_2[w_j]).\\
&\lambda(T_1[v], T_2[w])\\
&= \min \left\{ \begin{array}{l} \lambda(T_1[v], \emptyset)\\ \quad + \min\limits_{1\leq i\leq d(v)} \{\lambda(T_1[v_i], T_2[w]) - \lambda(T_2[v_i], \emptyset)\},\\ \lambda(\emptyset, T_2[w])\\ \quad + \min\limits_{1\leq j\leq d(w)} \{\lambda(T_1[v], T_2[w_j]) - \lambda(\emptyset, T_2[w_j])\},\\ \lambda(F_1[v], F_2[w]) + \gamma(v, w). \end{array} \right\}.\\
&\lambda(F_1[v], F_2[w])\\
&= \sum_{1\leq i\leq d(v)} \lambda(T_1[v_i], \emptyset) + \sum_{1\leq j\leq d(w)} \lambda(\emptyset, T_2[w_j])\\
&\quad - \sum_{(v,w)\in \mathtt{Ass}} \omega((v, w)).
\end{aligned}
$$

Figure 3. The recurrence for computing the LCA-preserving distance.

In Fig. 3, we denote the complete subtree of T rooted by v by $T[v]$, the forest obtained by deleting v from $T[v]$ by $F[v]$ and an empty tree or forest by $\emptyset$. Also $\mathtt{Ass}$ denotes the solution (i.e., the set of edges) of the assignment problem for a complete bipartite graph $(X \cup Y, E)$ [3], where X is the set of children of v in T_1, Y is the set of children of w in T_2 and the weight of an edge $e = (v', w') \in E$ is set to:

$$\lambda(T_1[v'], \emptyset) + \lambda(\emptyset, T_2[w']) - \lambda(T_1[v'], T_2[w'])\ [3].$$

To evaluate the trees as topic hierarchical structures of which node is a cluster as the set of labels, it is necessary to introduce the cost for $\gamma(v, \varepsilon)$, $\gamma(\varepsilon, w)$ and $\gamma(v, w)$ through the LCA-preserving mapping and distance.

First, the cost between labels is given as a cosine distance between words based on the frequency in a document-term matrix. We then set the costs of $\gamma(v, \varepsilon)$ and $\gamma(\varepsilon, w)$ to the number of labels in v and w, respectively, and the cost of $\gamma(v, w)$ as follows. Here, A is the set of labels in v, B is the set of labels in w and δ is a threshold.

1) If $|A| < |B|$, then add $|B| - |A|$ dummy labels to A. If $|A| > |B|$, then add $|A| - |B|$ dummy labels to B.
2) Construct a complete bipartite graph $(A \cup B, E)$.
3) For an edge $(a, b) \in E$, if either a or b is a dummy label or the cosine distance between a and b is greater than δ, then set $\omega(e)$ to 1; 0 otherwise.
4) Solve the assignment problem for $(A \cup B, E)$. Hence, the total weight of edges in the solution is $\gamma(v, w)$.

In the remainder of this paper, we call an LCA-preserving mapping between topic hierarchical structures a community mapping, and the correspondence in a community mapping a topic mapping. Also we call the sequence of topic mappings for the same topic over time a topic transition.

IV. Experimental Results

In this section, we give experimental results for tracing topic transitions related to the East Japan Great Earthquake over time in a buzz marketing site to explore the social context. Here, the target site of the experiments is the BBS in kakaku.com, which is the most popular buzz marketing site in Japan and target word is the "Great Earthquake." In Step A, we retrieve the messages from kakaku.com which contain the term "great earthquake" posted from 11th March to 5th April, 2011. As a result, we acquire $D = 436$ messages. Table I lists some examples. Also Table II lists keywords with high RIDF ($\tau = 2.0$).

Table I
Example of the result of Step A.

MID	*Posted*	*Title*	*Content*
73774	03/12 01:36	Earthquake and Donation	People who experienced the Hanshin Awaji earthquake are considering sending donations to afflicted people.....
77054	03/13 08:57	East Japan Great Earthquake	I'm sorry for people damaged by the Tohoku earthquake. What we can do is to send a donation. I would like to do some charity work.....
79662	03/13 23:46	East Japan Great Earthquake	Gasoline has been sold out even in Tokyo area gas stations. We have to send gasoline to the afflicted people.....
80646	03/13 12:08	East Japan Great Earthquake	As one of the afflicted people for the Hanshin, Awaji earthquake, I would like to encourage people affected by the East Japan Great Earthquake. Today, raising fund for damaged people has been started in Kobe.....
92249	03/18 13:35	East Japan Great Earthquake	The opening game of the Pacific League for the professional baseball was postponed on 12th April. But, the Central League is going to hold the opening game on 25th March as planned. I don't like it.....

Table II
Example of Keyword Extraction

MID	*Posted*	*Keywords*
73774	03/12 01:36	Hanshin, Awaji, earthquake, send, donation, afflict, people....
77054	03/13 08:57	sorry, people, damage, Tohoku, earthquake, send, donation, charity, work....
79662	03/13 23:46	Gasoline, sold-out, Tokyo, gas-station, send afflict, people...
80646	03/13 12:08	afflict, people, Hanshin, Awaji, earthquake, encourage people, today, raising-fund, damage, start, Kobe.
92249	03/18 13:35	opening-game, Pacific-League, professional-baseball, postpone, April, Central-League, hold, Mar, ..

In this experiment, documents are delimited daily. Then, Step B constructs snapshot co-occurrence networks over time from a set of keyword lists formed by Step A. Fig. 4 shows graph examples from 13th to 15th March.

For each snapshot co-occurrence network, Step C organizes topic hierarchical structures by clustering based on the modularity. In Step D, first we calculate cosine distances among keywords, and then compute the cost $\gamma(v, w)$ under the threshold $\delta = 0.7$.

Fig. 5 illustrates the community mapping between topic hierarchical structures of 13th and 14th March, and the costs of topic mappings in the community mapping.

Then, we compute the LCA-preserving distance between topic hierarchical structures by using the normalized cost $\gamma(v, w)$ by dividing $\max\{|v|, |w|\}$ and extract topic mappings from the LCA-preserving mapping (community mapping) of which distance is smaller than 1.3.

Table III shows the 36 extracted topic mappings. Here, we can find the following 6 meaningful topics T_i and 16 topic transitions for the topics.

- T_1: Sending monetary donation for afflicted people.
- T_2: Damages by the disaster (earthquake and Tsunami).
- T_3: The electricity problem.
- T_4: Prayers and visitations for afflicted people.
- T_5: Taking pictures of afflicted area.

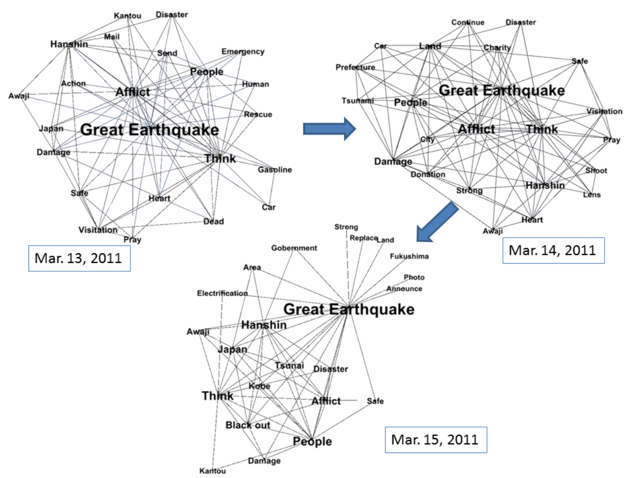


Figure 4. Co-occurrence networks from 13th to 15th March.

T_6: Problems on car manufacturing.

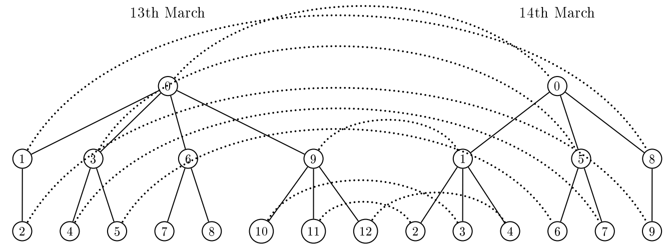
Fig. 6 visualizes the topic transitions obtained from the topic mappings in Table III. Topics T_1, T_2 and T_4 appear over a long duration on the time line. This means the topic about afflicted people was continuously discussed for these days. Topic T_3 about the electricity problem might have been triggered by a planned outage that happened 14th March, 2011. Topic T_6 about problems on car manufacturing loss from 25th March. This might mean that people started being concerned about other things besides afflicted people.

In our previous work [11], to trace topic transitions, we adopted the Matthews Correlation Coefficient (MCC), which is a similarity measure of two binary classifications [16], for topic hierarchical structures that were next to each other on the timeline. However, due to the computationally hard problem, we could not calculate it for all topic pairs among topic hierarchical structures. We had to manually select topic pairs as candidates of topic transitions in advance. Automatic detection for topic mappings over time has been achieved in this work.

V. Conclusion

In this paper, we proposed a novel method to explore the social context by using a community mapping as an LCA-preserving mapping between topic hierarchical structures. In this method, the correspondence between nodes in a community mapping is a topic mapping, and then the sequence of topic mappings over time represents the topic transition. Then, we gave the experimental results for the messages related to the East Japan Great Earthquake in a buzz marketing site.

It is a future work to conduct experiments for other social media data, and show the proposed method's effectiveness by comparing with existing techniques. We also plan to have scalability and performance tests as well, and finally propose the model for a topic life-cycle in buzz marketing sites.



node	label
313-00	Great Earthquake, visitation, Kantou, people, Japan, dead, worry, think, disaster, send, afflict, Hanshin, Awaji, safe, rescue, blackout, . . .
313-01	visitation, Kantou, people, Japan, dead, worry, disaster, send, . . .
313-03	blackout, rescue, think, afflict, safe, pray, . . .
. . .	. . .
314-00	Great Earthquake, blackout, Kobe, Hanshin, Awaji, Tsunami, car, afflict, . . .
314-01	Great Earthquake, blackout, Kobe, Hanshin, Awaji, shoot, pray, worry, . . .
314-05	afflict, land, Tsunami, city, prefecture, continue, rescue, donation, . . .
. . .	. . .

13th March	14th March	Cost	13th March	14th March	Cost
313-00	314-00	41	313-05	314-07	7
313-01	314-08	16	313-09	314-01	20
313-02	314-09	16	313-10	314-03	8
313-03	314-05	10	313-11	314-04	7
313-04	314-06	7	313-12	314-02	8

Figure 5. A community mapping between topic hierarchical structures of 13th and 14th March (upper), and the costs of topic mappings (lower).

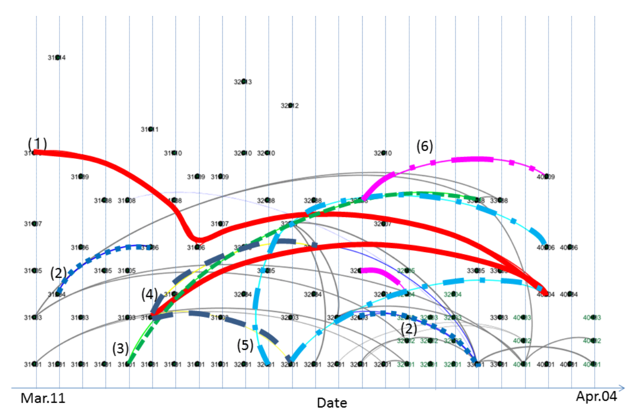


Figure 6. The extracted topic transitions.

Concerned with our experimental results, we extracted inappropriate topic mappings and could not extract some characteristic topic mappings. For example, a topic about the opening game of the professional baseball league in Japan was temporarily burst in the middle of March in

Table III
THE 36 EXTRACTED TOPIC MAPPINGS.

#	node		cost	topic
1	311-01	326-01	0.66	–
2	311-01	327-01	1.00	–
3	311-03	327-02	1.00	–
4	311-03	330-08	1.00	–
5	311-03	331-08	1.00	–
6	311-03	401-01	1.00	–
7	311-10	318-06	1.00	T_1
8	312-04	31-606	1.20	T_2
9	315-01	330-08	1.00	T_3
10	315-08	330-01	0.25	T_2
11	316-03	322-01	0.66	T_4
12	316-03	323-06	1.00	T_4
13	316-03	402-04	1.00	T_1
14	318-06	402-04	1.00	T_1
15	321-01	322-07	1.20	T_5
16	322-01	402-04	1.00	T_1
17	322-07	402-06	1.00	T_5
18	322-07	323-01	1.20	–
19	322-07	324-03	1.20	–
20	322-07	330-01	1.20	–
21	323-06	330-01	0.75	T_2
22	324-01	401-02	0.50	–
23	324-03	330-01	0.75	T_2
24	324-03	401-01	1.00	–
25	325-01	401-02	0.50	–
26	325-01	327-02	1.00	–
27	325-03	330-01	0.75	T_2
28	325-05	327-04	1.00	T_6
29	325-08	402-09	1.25	T_6
30	326-01	327-01	1.00	–
31	327-02	330-01	1.00	–
32	330-01	331-01	1.00	–
33	330-01	404-01	1.25	–
334	330-08	331-08	1.00	–
35	330-08	401-01	1.00	–
36	401-01	404-02	1.00	–

the target buzz marketing site. The topic was raised by one of the baseball club owners on 16th March, and it was concluded on 26th March after vehement opposition by people because of the electricity problem. Hence, it is a future work to improve our method not to extract inappropriate topic mappings and to extract characteristic topic mappings.

ACKNOWLEDGMENT

This work was supported by Grant-in-Aid for Scientific Researches 22240010, 23500185, 23680016, 24240021, 24300060 and 24500180 from the Ministry of Education, Culture, Sports, Science and Technology, Japan, and also partially supported by the Toyota Foundation and the Okawa Foundation for Information and Telecommunications.

REFERENCES

[1] Newman, M. E. J., *Modularity and community structure in networks*, National Academy of Science USA 103(23), 8577–8696, 2006.

[2] Tai, K.-C., *The tree-to-tree correction problem*, J. ACM **26**, 422–433, 1979.

[3] Zhang, K., Wang, J., Shasha, D., *On the editing distance between undirected acyclic graphs*, Int. J. Found. Comput. Sci. **7**, 43–58, 1995.

[4] Sekiguchi, Y., Kawashima, H., Uchiyama, T., *Discovery of related topics using series of blogsites' entries*, JSAI 2008, 2008 (in Japanese).

[5] Asur, S., Huberman, B. A., Szbaó, G., Wang, C., *Trends in social media: Persistence and decay,* ICWSM 2011, 2011.

[6] Wang, G., Araki, K., *A graphic reputation analysis system for mining Japanese weblog based on both unstructured and structured information*, AINA Workshops 2008, 1240–1245, 2008.

[7] Radinsky, K., Agichtein E., Gabrilovich, E., Markovitch, S., *A word at a time: Computing word relatedness using temporal semantic analysis*, WWW 2011, 337–346, 2011.

[8] Wang, X., Zhai C., Hu, X., Sproat, R., *Mining correlated bursty topic patterns from coordinated text streams*, KDD 2007, 784–793, 2007.

[9] Zhou, D., Councill, I., Zha, H., Giles, C. L., *Discovering temporal communities from social network documents*, ICDM 2007, 745–750, 2007.

[10] Qiu, J., Lin, Z., Tang, C., Qiao, S., *Discovering organizational structure in dynamic social network*, ICDM 2009, 932–937, 2009.

[11] Hashimoto, T., Kuboyama, T., Chakraborty, B., Shirota, Y., *Discovering Topic Transition about the East Japan Great Earthquake in Dynamic Social Media*, GHTC 2012, 259–264, 2012.

[12] Church, K. W., Gale, W. A., *Poisson mixtures*, Natural Language Engineering **1**, 163–190, 1995.

[13] Blondel, V. D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E., *Fast unfolding of communities in large networks*, J. Stat. Mech., P10008, 2008.

[14] Zhang, K., Jiang, T., *Some* MAX SNP-*hard results concerning unordered labeled trees*, Inform. Process. Lett. **49**, 249–254, 1994.

[15] Zhang, K., Statman, R., Shasha, D., *On the editing distance between unordered labeled trees*, Inform. Process. Letter. **42**, 133–139, 1992.

[16] Matthews, B.W., *Comparison of the predicted and observed secondary structure of T4 phage lysozyme*, Biochim. Biophys. Acta. **405**, 442–451,1975.