Detecting New Concepts in Social Media using Co-burst Pattern Mining

Takako Hashimoto*, David Shepard†, Basabi Chakraborty‡ and Tetsuji Kuboyama§
*Commerce and Economics, Chiba University of Commerce, Chiba, Japan
Email: takako@cuc.ac.jp
†Center for Digital Humanities, University of California, Los Angeles, USA
Email: shepard.david@gmail.com
‡Software and Information Science, Iwate Prefectural University, Japan
Email: basabi@iwate-pu.ac.jp
§Computer Centre, Gakushuin University, Japan
Email: kuboyama@tk.cc.gakushuin.ac.jp

***Abstract*—This paper proposes a method for detecting new concepts in social media using co-burst pattern mining technique. The new concepts are defined as correlations between unexpected words. The target social media are viewers' comments attached to web videos and Twitter's tweets related to the East Japan Great Earthquake that happened on Mar. 11 in 2011. Our proposed method first crawls viewers' comments from web videos, and extracts words from them. Then it selects motive words candidates from words, and counts the occurrence numbers of tweets that include motive words candidates. To detect new concepts, it generates burst patterns based on occurrence numbers of motive words candidates over time and detects unexpected correlations between motive words candidates. By our method, after the earthquake, new unexpected correlations between motive words in social media are recognized as new concepts. For example, the method could extract motive words from web video comments, such as "escape, nuclear plant" and "Tokyo Electric Power Co., Inc.(TEPCO, that owns the nuclear plant), president." Then it could detect the new concept "escape (from) nuclear plant" and "TEPCO's president" on Twitter. In this paper, we provide the preliminary approximation results and discuss the effectiveness of our proposed method.**

Keywords*—*Social meda, burst pattern, unexpected words' correlation, video service, Twitter, East Japan Great Earthquake.

I. Introduction

Social media in which individual users post their opinions and gradually build new concepts together, is recognized as one of the important collaborations in today's information oriented society. After the East Japan Great Earthquake, we could detect discussions related to the nuclear plant, Tokyo Electric Power Company (TEPCO) and so on, that could not be recognized before the earthquake. We defined new concepts as correlations between unexpected words that could not be recognized before the earthquake, such as "nuclear plant, escape" and "TEPCO, president." Exploring newly-built concepts over time on social media is significant, so that we believe we can gain a rich insight into social context.

We already proposed the graph-based topic extraction method [1] using the modularity measure [2]. We also proposed an approach to extract hidden topics over time from social media messages using the latent semantic analysis (LSA) technique [3] , [4]. Our previous work targeted single social media such as a blog or a buzz marketing site. However, new concepts are sometimes created triggered by mutual relationships between different social media. This paper targets multiple social media and proposes the method to explore new concepts by analyzing multiple social media. The target social media are web video comments and tweets related to the East Japan Great Earthquake. Our proposed method first crawls viewers' comments attached to web videos, and extracts words from them. It selects motive words candidates from extracted word, and then counts the occurrence numbers of tweets that include the motive words candidates in Twitter. To detect new concepts, it generates burst patterns based on occurrence numbers of motive words candidates over time and analyzes burst correlations between them. By our method, new burst correlations between motive words triggered by social media are recognized as new concepts. For example, after the earthquake, we could extract "escape, nuclear plant" as motive words from web video comments, and detect the new concept "escape (from) nuclear plant" on Twitter.

The contributions of this paper are as follows:

- Propose the method for detecting new concepts from mutual correlation of multiple social media (cross-media analysis)
- Show concrete examples for new concepts that appeared after the East Japan Great Earthquake

This paper is organized as follows. Section II refers to existing researches. Section III introduces our target social media. Section IV illustrates our proposed method to explore new concepts by analyzing relationships between different social media. Section V shows the preliminary approximation result of our method that targets web video comments and tweets related to the East Japan Great Earthquake. Finally, Section VI concludes this paper.

II. Related Work

Most related works for detecting topics/concepts focus on single media, such as blogs, Twitter, and web videos respectively. Sekiguchi *et al.* [5] treated recent blogger posts and analyzed the word co-occurrence and the repeating rate of word. They visualized the relation between words and showed

 ISBN: 978-1-61208-324-7

topics in social media through the visualization results. Asur *et al.* [6] investigated trending topics on Twitter. They proposed a 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. Liu *et al.* [7] and Cao *et al.* [8] focus on web video analysis. Especially, Cao *et al.* [8] clusters video tags into groups to get small events and then link these events into topics based on textual and video similarity. On the other hand, our proposed method focuses on multiple social media and analyzes them. It can flexibly show concepts transition by taking into cross-media over time.

As for cross-media analysis, most existing works focus on co-clustering among multiple social media. Xue *et al.* [9] proposed the cross-media topic detection method that was based on co-clustering and detect new topics. Our proposed method focuses on characteristic words extracted from social media and then detect co-occurrence patterns among them that can be recognized as new concepts.

Regarding research on detecting temporal relations, Radinsky *et al.* [10] 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.* [11] 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.* [12] 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.* [13] 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 single media and analyzed their transition. In our method, on the other hand, new concepts exploration can be analyzed by investigating multiple social media over time based on co-burst pattern of characteristic words.

III. Target Social Media

The aim of our proposed method is cross-media concepts detection, so that it targets multiple social media. As the first targets, Nicovideo and Twitter have been selected in our work.

A. Nicovideo comments related to the East Japan Great Earthquake

Nicovideo is one of the most popular video sharing web sites in Japan [14]. In Nicovideo, users can upload, view and share videos, and also add comments while watching videos. Unlike other video sharing sites, comments are overlaid directly onto the video, synchronized to a specific playback time. Users can communicate each other through video comments and a sense of a shared watching experience could be created. After the East Japan Great Earthquake, Nicovideo provided live programs like the government press conferences, TEPCO press conferences and so on (Figure 1).

Fig. 1. An example of Nicovideo.

These live programs were not provided by major TV broadcasting companies and viewers could get actual information that they could not watch through mayor TV programs. Users' comments that were attached to the live program could be a trigger to produce new concepts related to the earthquake among users, and the concepts had an influence on users' behavior. Hence, the video sharing website may lead opinions in society. By analyzing comments on Nicovideo, we expect that the relationships between Nicovideo and other social media can be detected and the new concepts propagation can be illustrated.

B. Twitter's tweets related to the East Japan Great Earthquake

Tweets related to the East Japan Great Earthquake is also targeted in this paper. During the earthquake, people tweeted a lot of things about concerns for affected people and disaster situation, fear for future and so on (Table I).

TABLE I. Example of tweets related to the East Japan Great Earthquake.

Date	*Tweet (translated into English)*
2011/03/11	I can not contact my parents who live in Miyagi. #jishin, #miyage
2011/03/11	Be strong, we are with you #jishin
2011/03/11	The JR train service has returned to normal.. #jishin
2011/03/12	The government press conference has just started. #jishin#nhk
2011/03/12	My friend was almost to get robbed. Please take care.. #jishin

The social media monitoring company Hottolink [17] tracked users who used one of 43 hashtags (for example, #jishin, #nhk, and #prayforjapan) or one of 21 keywords related to the disaster. Later, they captured all tweets sent by all of these users between Mar. 9th and Apr. 2nd. This resulted in an archive of around 200 million tweets, sent by around 1 million users. Capturing programs searched tweets by hashtag, consequently, and many of these tweets contain useful information about users responses to the disaster. These tweets are one of big data and it is significant to analyze them to detect new concept generated after the quake.

IV. Proposed Method for Detecting New Concepts in Social Media using Co-burst Pattern Mining

Our proposed method focuses on detecting new concepts in social media. We define a new concept as new words' burst correlation. Suppose there are two words like "president" and "Tokyo Electric Power Co., Inc.(TEPCO, that owns the nuclear plant)." Before the East Japan Great Earthquake, we did not have the special meaning between "president" and "TEPCO," so that the correlations between "president" and "TEPCO" could not be recognized. But after the press conference by TEPCO, we suddenly began to have the new meaning of "president" and "TEPCO" as the person who was accountable for the nuclear accident. Actually, we could find new co-occurring patterns between "president" and "TEPCO" in Twitter. Our hypothesis is that new concepts are suddenly generated by communications in social media, and propagated quickly in social media. Hence, the objective of our method is to find new motive words candidates that can be basis of new concepts, and detect new correlations between them in social media.

There are two types of social media in our method. One generates motive words, and the other propagates motive words correlation (new concepts). As the social media for motive words generation (TriggerSM) and the social media for words correlation propagation (PropagateSM), in this paper, we use Nicovideo and Twitter respectively.

Our method consists of the following 3 steps.

STEP A: Find motive word candidates from TriggerSM.
STEP B: Count occurrence numbers of motive words candidates in PropagateSM.
STEP C: Analyze time series motive words' co-occurring patterns and detect new concepts on PropagateSM.

Figure 2 illustrates our proposed method's steps.

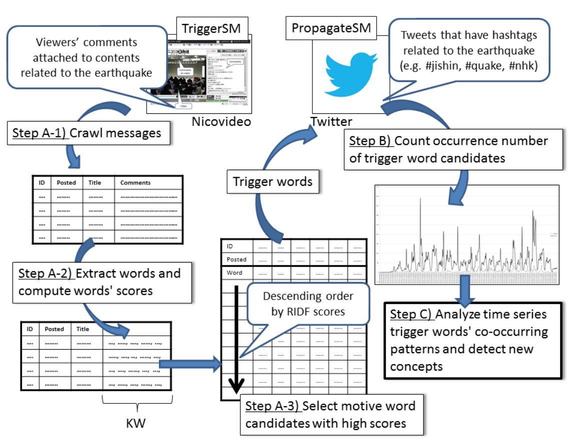

Fig. 2. Proposed method.

The following is the description of each step.

A. Find motive word candidates from TriggerSM.

This step consists of the following 3 sub-steps.

1) Crawl messages from TriggerSM.
2) Extract words candidates and compute their scores.
3) Select motive words candidates with high scores.

Each sub-step is explained in the following:

1) Crawl messages from TriggerSM: This sub-step crawls messages from TriggerSm. The target social media is viewers' comments attached to Nicovideo live contents (the press conferences related to the East Japan Great Earthquake). A set of comments attached to one video content were recognized as one document d_i as the following tuples:

$$d_i = (MID_i, Posted_i, Title_i, Content_i) \tag{1}$$

Here, MID_i is an ID of each document, $Posted_i$ is a broadcast date-time that the document (content), $Title_i$ is a title of each document (content) and $Content_i$ is a combined text of video comments. Table II shows some example of d_i.

TABLE II. Example of d_i.

MID	$Posted_i$	$Title_i$	$Content_i$
1	2011/03/11	Press Conference by Government	I can not believe, we should send something to affected people, It is really dangerous., ...
2	2011/03/14	Press Conference by TEPCO	The president should take responsibility, Where is the president? The vice president is also wired, ...
3	2011/03/15	Press Conference by TEPCO	The nuclear plant is really bad, melt down?, TEPCO is untrustworthy, ...

2) Extract words candidates and compute their scores: This sub-step extracts words that are nouns, verbs, adjectives, and adverbs from $Content_i$ of each d_i by morphological analysis. We use Mecab that is yet another Japanese Dependency Structure Analyzer [19] for word extraction. Then, the score of an individual word in d_i is calculated using RIDF [20] measure that is based on the poison distribution. We form a list of keywords $KW = \{kw_i\}$.

$$kw_i = (MID_i, Posted_i, \{w_{ij}, v_{ij}\}) \tag{2}$$

Here, $\{w_{ij}, v_{ij}\}$ is a list of a pair that consists of an extracted word w_{ij} from document d_i, and the corresponding RIDF value v_{ij} of w_{ij}.

3) Select motive word candidates according to their scores.: This sub-step sorts $\{w_{ij}, v_{ij}\}$ in descending order by v_{ij}. Then the step analyzes KW over time and finds newly appeared words that are high on the list $\{w_{ij}, v_{ij}\}$ of each kw_i. We focus on top n words of each kw_i and among those, we find characteristic words that did not seem appear before the earthquake as motive words candidates.

B. Count occurrence numbers of motive words candidates in PropagateSM

Then the method counts time series occurrence numbers of candidates words in PropagateSM. The occurrence number

 ISBN: 978-1-61208-324-7

is counted before and after the earthquake. If the occurrence number of the word is low before the quake, and becomes high after the quake, the word can be recognized to become burst after the quake.

C. Analyze time series motive words' co-occurring patterns and detect new concepts on PropagateSM.

The method checks the time series burst pattern for each motive word candidate from the occurrence number of each word. To detect the burst pattern, we adopt the method proposed by Zhu *et al.* [18]. Zhu *et al.* proposed the burst detection method using elastic windows over time. They propose the shifted wavelet tree as the data structure for efficient burst monitoring, so that their method can detect burst flexibly. The shifted wavelet tree uses the adjacent windows of the same level are half overlapping (Figure3). These additional windows provide valuable overlapping information for the time series. It will be better to analyze co-burst patterns between words than the conventional wavelet tree.

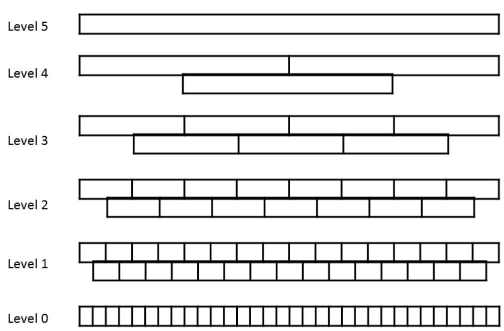

Fig. 3. Shifted Wavelet Tree proposed by Zhu *et al.*.

Any subsequence with length $w, w \leq 2^i$ is included in some subsequence(s) with length 2^i, and therefore is included in one of the windows at level $i+1$. We say that windows with size $w; 2^{i-1} \leq w \leq 2^i$, are monitored by level $i+1$ of the SWT.

The method computes the coefficient of correlation between burst patterns of motive words candidates. If the coefficient of correlation is larger than the threshold γ, the new concept is supposed to be generated.

V. Preliminary Approximation

We crawled around 94000 viewers' comments attached to 67 live videos (broadcasted from Mar. 13 to 24 in 2011) related to the East Japan Great Earthquake in Nicovideo (TriggerSM). Then words were extracted from crawled comments and the RIDF score for each word was computed. Table III shows some example of documents $\{d_i\}$ x words $\{w_{ij}\}$ matrix with the RIDF scores.

Then $\{w_{ij}\}$ in d_i were sorted in descending order by $\{v_{ij}\}$. We set $n = 10$, and top 10 words with high RIDF value in each d_i were extracted. Table IV shows some example of extracted top 10 words for each d_i.

For example, in the document of $MID = 1$, words such as "blackout", "TEPCO", "press", "escape", "stop" and

TABLE III. Example of $\{d_i\} - words\{w_{ij}\}$ matrix.

MID	$Posted_i$	escape	JSDF [15]	life	publish	president	...
1	2011/03/13 20:00	0	0	0	0	0.1	...
2	2011/03/15 8:30	0.027	0.026	0.013	0.01	0.11	...
3	2011/03/15 14:00	0	0	0	0	0.01	...
4	2011/03/15 21:00	0	0.01	0.01	0	0.03	...
5	2011/03/15 23:30	0.02	0.01	0.01	0	0.01	....
6	....	....	....	....	....	....	....

TABLE IV. Example of trigger words with high RIDF values in $\{d_i\}$.

MID	1	2	3	4	...	21	...
$Date_i$	$3/13$ 20:00	$3/15$ 8:30	$3/15$ 14:00	$3/15$ 21:00	... ...	$3/16$ 18:00	... ...
#1	blackout	ask	conference	blackout	...	vice-president	...
#2	TEPCO	Fukushima	NISA[16]	TEPCO	...	nuclear-plant	...
#3	press	president	TEPCO	nuclear-plant	...	president	...
#4	escape	escape	rain	NISA	...	TEPCO	....
#5	stop	nuclear-plant	field	electricity	...	measures	....
#6	president	planned-outrage	mass-media	conference	...	TEPCO	....
#7	measures	JSDF	measures	no-problem	...	problem	....
#8	fire-fighting	TEPCO	cover-up	power-saving	...	electricity	....
#9	Fukushima	field	officer	time	...	field	....
#10	nulcear-plant	fix	nuclear-plant	affected	...	remote	....

"president" were listed up. In the document of $MID = 6$, words such as "president", "planned-outrage", "mass-media", "TEPCO", "conference" and "TEPCO" were listed up. These words that characterize contents were recognized as trigger words candidates.

In this paper, we defined the follwoing 23 words as trigger words candidates.

"planned-outrage", "blackout", "field", "Edano", "JSDF", "employee", "fire-fighting", "government", "Shimizu", "power-saving", "measures", "stop", "power", "escape", "NISA", "radioactivity", "nuclear plant", "officer", "TEPCO", "Fukushima", "president", "vice-president", "director"

As for above trigger words candidates, we counted occurrence number of each candidate in Twitter data (from Mar. 9 to Apr. 2) provided by Hottolink [17]. Figure 4 and Figure 5 show some result of the occurrence number of each trigger word candidate.

In Figure 4, we can not find the explicit correlation between the occurrence pattern between president, vice-president and director in Twitter. On the other hand, in Figure 5, the occurrence pattern between nuclear plant and Fukushima in Twitter seems strongly co-related. Actually, "Fukushima nuclear plant" became the general word after the quake, so that these words must be co-related. However, correlations between other words were unexpected. To analyze correlations precisely, co-burst patterns were considered.

Then, we adopt the burst detection method proposed by Zhu *et al.* [18] to analyze co-burst patterns between trigger words candidates. Figure 6 shows results of burst patterns of trigger words candidates. The horizontal axis shows time (from Mar. 9 to Apr. 2), and black cells indicate burst periods for each word Figure 7 shows correlations between burst patterns of trigger words. We set the threshold value γ as 0.5 and the correlations larger than the threshold are shown by shaded region.

According to Figure 7, we could observe the following unexpected concepts

- escape $\rightarrow$ Fukushima, nuclear-plant
- TEPCO $\rightarrow$ president, vice-president, director, employee
- firefighting $\rightarrow$ power, JSDF, radioactivity, nulcear-plant
- measures $\rightarrow$ nuclear-plant, power, radioactivity, Fukushima

We could find the new concepts such as "escape (from) Fukushima", "TEPCO president/vice-president/... (for condemnation)" "(new relationships between) firefighting, power, JSDF, radioactivity, and nulcear-plant", and "(the importance of) measures for nuclear-plant, power, radioactivity and Fukushima."

VI. CONCLUSION

This paper proposed the method to detect new concepts in multiple social media after the topical problem like the East Japan Great Earthquake. As the preliminary approximation result, after the East Japan Great Earthquake, from web video service (Nicovideo) and Twitter, new concepts could be detected and shown as unexpected words co-occurrences. For example, after the earthquake, new concepts for the nuclear plant, TEPCO, and so on were recognized on social media.

As the future work, we plan to improve the method for automatically selecting trigger words candidates, and analyze the time series concepts detection using the video time line. Moreover, the method should be applied to other data and evaluated compared to the conventional method such as wavelet tree. As the future work, we are going to improve the technique for trigger words candidates selection and focus on precisely analyzing burst patterns over time using the time line of the video. In addition, we will improve the method by considering scalability.

REFERENCES

[1] T. Hashimoto, T. Kuboyama, B. Chakraborty, and Y. Shirota, Discovering Topic Transition about the East Japan Great Earthquake in Dynamic Social Media, GHTC 2012, Oct. 2012, pp. 259-264.

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

[3] T. K. Landauer and S. T. Dumais, A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge, Psychological Review, 104(2), 1997, pp. 211-240.

[4] T. Hashimoto, B. Chakraborty, T. Kuboyama, and Y. Shirota, Temporal Awareness of Needs after East Japan Great Earthquake using Latent Semantic Analysis, EJC2013 (the 23nd European-Japanese Conference on Information Modelling and Knowledge Bases), Jun. 2013, pp.214-226.

[5] Y. Sekiguchi, H. Kawashima, and T. Uchiyama, Discovery of related topics using series of blogsites' entries, JSAI 2008, 2I1-1, 2008 (in Japanese).

[6] S. Asur, B.A. Huberman, G. Szbaó, and C. Wang, Trends in social media: Persistence and decay, ICWSM 2011, Jul. 2011, pp. 434-437.

[7] L. Liu, L. Sun, Y. Rui, Y. Shi, and S. Yang, Web video topic discovery and tracking via bipartite graph reinforcement model, IWWWC 2011, Apr. 2011, pp. 1009-1018.

[8] J. Cao, C. W. Ngo, Y. D. Zhang, and J. T. Li, Tracking web video topics: discovery, visualization and monitoring, IEEE Transactions on Circuits and Systems for Video Technology 21(12), 2011, pp. 1835-1846, 2011.

[9] Z. Xue, Z. Jiang, G. Li, and Q. Huang, Cross-media topic detection associated with hot search queries, ICIMCS 2013, Aug. 2013, pp. 403-406.

[10] K. Radinsky, E. Agichtein, E. Gabrilovich, and S. Markovitch, A word at a time: Computing word relatedness using temporal semantic analysis, WWW 2011, Mar. 2011, pp. 337-346.

[11] X. Wang, C. Zhai, X. Hu, and R. Sproat, Mining correlated bursty topic patterns from coordinated text streams, KDD 2007, Aug. 2007, pp. 784-793.

[12] D. Zhou, I. Councill, H. Zha, and C. L. Giles, Discovering temporal communities from social network documents, ICDM 2007, Oct. 2007, pp. 745-750.

[13] J. Qiu, Z. Lin, C. Tang, and S. Qiao, Discovering organizational structure in dynamic social network, ICDM 2009, Dec. 2009, pp. 932-937.

[14] Nicovideo, [Online]. Available: http://www.nicovideo.jp/, [retrieved: Jan., 2014]

[15] JSDF, [Online]. Available: Japan Self-Defense Forces, http://www.mod.go.jp/asdf/English_page/, [retrieved: Jan., 2014]

[16] NISA, [Online]. Available: Nuclear and Industrial Safety Agency, http://www.nsr.go.jp/archive/nisa/english/, [retrieved: Jan., 2014]

[17] Hottolink,Inc., [Online]. Available: http://hottolink.co.jp, [retrieved: Jan., 2014]

[18] Y. Zhu and D. Shasha, Efficient elastic burst detection in data streams KDD 2003, Aug. 2003, pp. 336-345.

[19] T. Kudo, MeCab : Yet Another Part-of-Speech and Morphological Analyzer, [Online]. Available: http://mecab.sourceforge.net/, [retrieved: Jan., 2014]

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

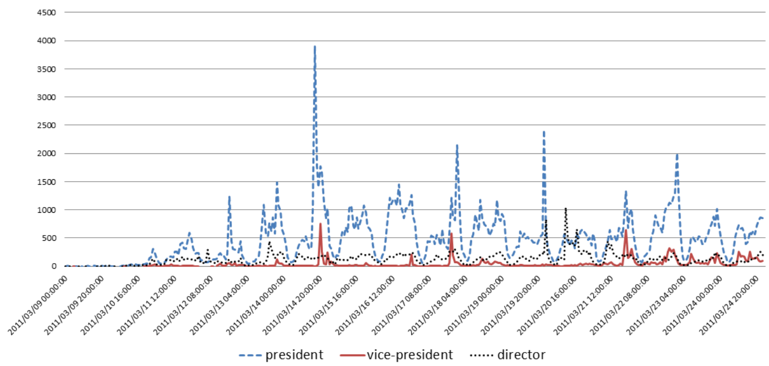

Fig. 4. Occurrence number of president, vice-president and director in Twitter

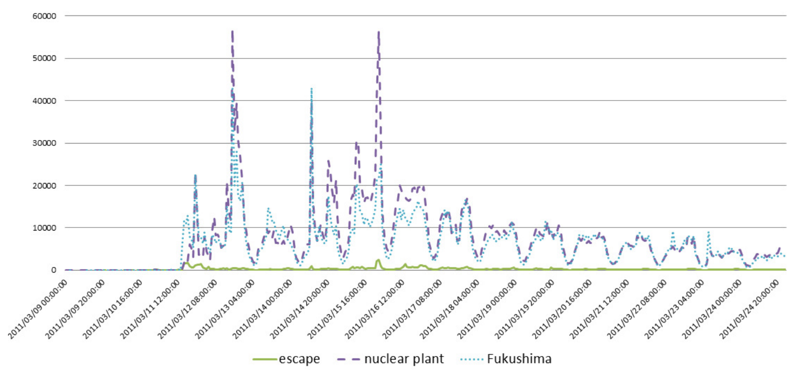

Fig. 5. Occurrence number of escape, nuclear plant and Fukushima in Twitter

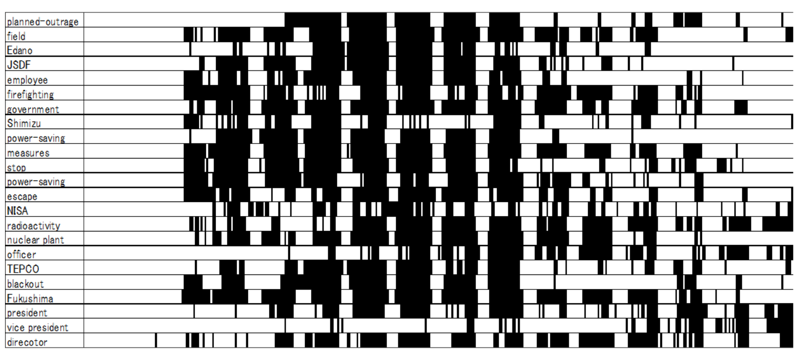

Fig. 6. Burst patterns of trigger words candidates

	planned-outrage	field	Edano	JSDF	employee	firefighting	government	Shimizu	power-saving	measures	stop	power	escape	director	NISA	radioactivity	nuclear-plant	officer	TEPCO	blackout	Fukushima	president	vice-president
planned-outrage	1.00																						
field	0.41	1.00																					
Edano	0.60	0.57	1.00																				
JSDF	0.49	0.69	0.67	1.00																			
employee	0.59	0.62	0.59	0.69	1.00																		
firefighting	0.31	0.64	0.37	0.49	0.50	1.00																	
government	0.52	0.75	0.54	0.67	0.63	0.61	1.00																
Shimizu	0.33	0.51	0.38	0.51	0.48	0.47	0.48	1.00															
power-saving	0.60	0.62	0.61	0.74	0.73	0.44	0.67	0.49	1.00														
measures	0.55	0.74	0.55	0.66	0.71	0.73	0.80	0.54	0.65	1.00													
stop	0.56	0.57	0.53	0.63	0.73	0.48	0.59	0.50	0.71	0.67	1.00												
power	0.43	0.62	0.42	0.60	0.67	0.68	0.64	0.54	0.65	0.71	0.60	1.00											
escape	0.43	0.65	0.47	0.60	0.65	0.59	0.63	0.42	0.52	0.67	0.55	0.62	1.00										
director	0.46	0.55	0.42	0.39	0.38	0.52	0.54	0.37	0.31	0.56	0.34	0.46	0.46	1.00									
NISA	0.29	0.43	0.38	0.34	0.29	0.33	0.43	0.20	0.36	0.36	0.29	0.28	0.37	0.30	1.00								
radioactivity	0.39	0.54	0.42	0.38	0.36	0.51	0.56	0.28	0.30	0.53	0.35	0.37	0.48	0.53	0.37	1.00							
nuclear-plant	0.50	0.75	0.61	0.69	0.58	0.67	0.74	0.45	0.62	0.78	0.53	0.66	0.67	0.50	0.48	0.57	1.00						
officer	0.25	0.41	0.23	0.17	0.12	0.33	0.36	0.13	0.11	0.29	0.10	0.20	0.30	0.47	0.23	0.55	0.37	1.00					
TEPCO	0.76	0.53	0.57	0.61	0.67	0.45	0.58	0.46	0.74	0.61	0.66	0.56	0.56	0.36	0.35	0.34	0.56	0.14	1.00				
blackout	0.76	0.51	0.57	0.61	0.76	0.41	0.55	0.47	0.70	0.62	0.71	0.59	0.54	0.36	0.25	0.34	0.45	0.14	0.74	1.00			
Fukushima	0.49	0.74	0.51	0.65	0.62	0.73	0.79	0.53	0.57	0.84	0.62	0.68	0.68	0.55	0.39	0.59	0.80	0.34	0.56	0.53	1.00		
president	0.49	0.42	0.39	0.36	0.40	0.38	0.46	0.35	0.50	0.42	0.36	0.33	0.42	0.55	0.26	0.55	0.39	0.47	0.50	0.41	0.44	1.00	
vice-president	0.16	−0.07	−0.05	−0.12	−0.09	−0.03	0.02	0.00	−0.10	−0.05	0.02	−0.09	0.01	0.17	0.05	0.29	−0.08	0.31	0.06	0.02	−0.01	0.61	1.00

Fig. 7. Correlations between trigger words candidates