Detecting Unexpected Correlation between a Current Topic and Products from Buzz Marketing Sites

Takako Hashimoto[1], Tetsuji Kuboyama[2], and Yukari Shirota[3]

[1] Commerce and Economics, Chiba University of Commerce, Chiba, Japan
takako@cuc.ac.jp
[2] Computer Center, Gakushuin University, Tokyo, Japan
kuboyama@tk.cc.gakushuin.ac.jp
[3] Department of Management, Faculty of Economics, Gakushuin University, Tokyo, Japan
yukari.shirota@gakushuin.ac.jp

Abstract. This paper proposes a method to detect unexpected correlation from between a current topic and products word of mouth in buzz marketing sites, which will be part of a new approach to marketing analysis. For example, in 2009, the super-flu virus spawned significant effects on various product marketing domains around the globe. In buzz marketing sites, there had been a lot of word of mouth about the "flu." We could easily expect an "air purifier" to be correlated to the "flu" and air purifiers' shipments had grown according to the epidemic of flu. On the other hand, the relatedness between the "flu" and a "camera" could not be easily expected. However, in Japan, consumers' unforeseen behavior like the reluctance to buy digital cameras because of cancellations of a trip, a PE festival or other events caused by the epidemic of flu had appeared, and a strong correlation between the "flu" and "camera" had been found. Detecting these unforeseen consumers' behavior is significant for today's marketing analysis. In order to detect such unexpected relations, this paper applies the dynamic time warping techniques. Our proposed method computes time series correlations between a current topic and unspecified products from word of mouth of buzz marketing sites, and finds product candidates which have unexpected correlation with a current topic. To evaluate the effectiveness of the method, the experimental results for the current topic ("flu") and products ("air purifier", "camera", "car", etc.) are shown as well. By detecting unexpected relatedness from buzz marketing sites, unforeseen consumer behaviors can be further analyzed.

Keywords: Data mining, Marketing analysis, Web Intelligence, Dynamic time warping, Social media analysis.

1 Introduction

Data mining techniques for product marketing to analyze word-of-mouth in social media such as blogs and buzz marketing sites have recently become an active area of

S. Kikuchi et al. (Eds.): DNIS 2011, LNCS 7108, pp. 147–161, 2011.
© Springer-Verlag Berlin Heidelberg 2011

research[3-6, 11]. In analyzing product reviews or reputation in social media, almost all existing research focuses first on specific products, and extracts typical evaluation expressions such as “favorite,” “dislike,” “expensive,” and “useful.” They then calculate positive/negative degrees of extracted expressions. We have also researched data mining techniques on home electrical appliances such as air purifiers and front loading washing machines with automatic drying systems, and proposed a reputation analysis framework for buzz marketing sites[13]. It may be easy to analyze a specific product’s reputation, because the target product’s characteristics can be illustrated by the ontology for the product, which is constructed with relatively little effort.

On the other hand, for buzz marketing sites, it is very difficult to analyze unexpected consumer behavior for “unspecified products” which is caused by current topics such as an epidemic of flu and a great disaster. Because the target product is not explicit, it is not possible to prepare a specific ontology in advance. To detect unexpected consumer behavior, we also have proposed the graph-based consumer behavior analysis framework [15, 16]. Our previous proposed framework visualizes time series variation of unforeseen relations between a current topic and unspecified products from buzz marketing sites. In our previous experiments concerning the super-flu spawn in 2009, we could find an unexpected consumer behavior as follows; in threads about digital cameras, we discovered that many persons wrote that the flu made them cancel plans of children’s PE festivals and trips during Golden Week in Japan, since people had to be confined at home. The flu pandemic made consumers hold off buying digital camera, since people who had been planning to take photos at those events were reluctant to buy digital cameras due to the flu pandemic. The reluctance in buying digital cameras because of the flu was not something we’d expect. On the contrary, we could easily expect more air purifiers to be sold because of the flu pandemic. We can say that the relation between the flu and digital cameras can be recognized as an unforeseen and unexpected relationship.

The problem for finding unexpected relations between a current topic and unspecified products is how we can easily find the target unspecified products. To address the problem, this paper proposes a method to detect correlations between a current topic and unspecified products from buzz marketing sites. For detecting unexpected correlations, this paper applies the dynamic time warping techniques. Our method computes time series correlation based on the occurrence patterns of both a current topic and products, and selects product candidates which may cause consumers' unexpected behavior. By detecting unexpected correlation, unforeseen consumer behaviors can be further analyzed. Our proposed method will be part of a new approach to marketing analysis. This is the novel point in this paper.

The following section explains our previous method “Graph-based Consumer Behavior Analysis.” Section 3 refers to existing research. In section 4, our proposed method using the dynamic time warping technique is described. Section 5 shows experimental results. Finally, section 6 gives concluding remarks and describes the direction of future work.

2 Graph-Based Consumer Behavior Analysis

This section briefly explains our previous proposed method "Graph-based Consumer Behavior Analysis[15,16]." Our previous proposed method consists of the following six steps (Fig.1):

Step1. Crawling
Step2. Language processing
Step3. Graph transformation
Step4. Visualization
Step5. Graph edit distance calculation
Step6. Consumer behavior detection

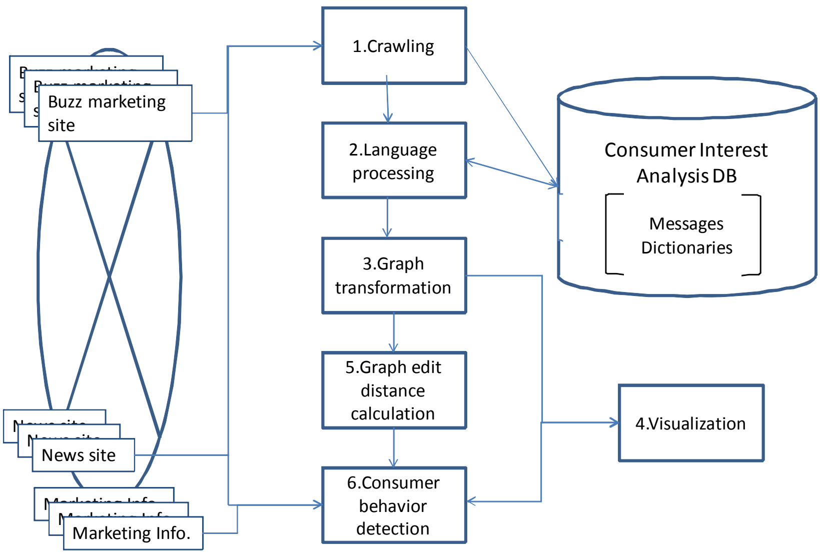

Fig. 1. Consumer behavior analysis framework

In Step1, word-of-mouth in buzz marketing sites is crawled according to the given query which is a term of a current topic like "flu." As the target buzz marketing site, we have selected the online bulletin board of kakaku.com[2], that is the most popular buzz marketing site in Japan.

In Step2, the crawling results of Step1 are analyzed by language processing technique. We've defined one word-of-mouth as one document. This step extracts keywords that are nouns, verbs, adjectives, and adverbs from one document using morphological analysis. Then the score of an individual keyword is calculated. As score calculation method, the step uses RIDF(residual IDF), LSA(latent semantic analysis), and tf-idf(Term Frequency- Inverse Document Frequency). According to our examination, at the moment, RIDF is appropriate for extracting keywords which indicate the document content.

In Step3, we construct directed graphs to show consumer behavior structure from the output of Step2, which is a matrix between message id and keyword id with high score. As a directed graph, we use the concept graph due to Hirokawa[1], which makes relevance hypernym relations of keywords appearing in a set of documents based on co-occurrence frequencies. In our framework, the posted date is delimited by appropriate period (e.g. monthly, weekly, or daily) and the graph structures are formed according to the period.

Step4 is a visualization module to show concept graph structures which is made by Step3. Fig.2 illustrates an example of the concept graph visualization related to the "flu" in 2009 from the kakaku.com BBS sites. There is a large island structure discussing digital camera and air purifier.

Our hypothesis is that major changes of the concept graph structures show consumers' behavior changes. To detect the consumers' behavior change, we employ graph topology-based distance for measuring changes in concept graph topology over time. In Step5, the graph edit distance[14] is calculated from a set of concept graph data (the outputs of Step3). By analyzing time series variation of the graph edit distance, people interest changes can be detected and unexpected consumer behavior can be analyzed. Fig.2 shows the results of the graph edit distance calculation as well. In Fig.2, graph structures about digital cameras are recognized in the concept graphs of January, May, July, August, September and October of 2009. With the plots in Fig.2, we recognize that the major structure changes happened in May and July, and a part of substructure emerged in July is preserved until October. Compared to real sales of digital cameras, sales increased in June (after May) and October (after September). We can guess, therefore, that the structure change in May and October illustrates consumer behavior change.

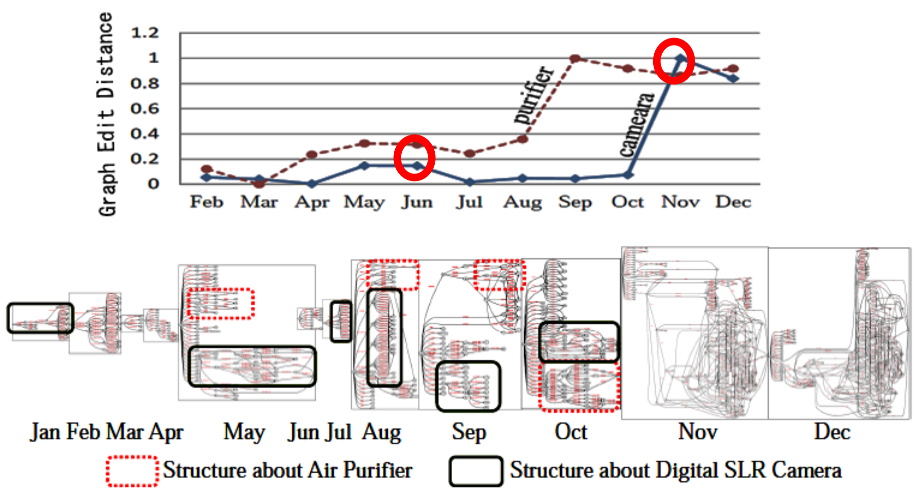

Fig. 2. Monthly concept graphs related to "flu" in 2009, and substructures related to "air purifier," "digital camera"

The problem for our previous method is how the pair of a current topic and unspecified products that would have unexpected relationship can be easily found. In our previous method, we've detected the unexpected correlation manually. However, to increase efficiency, the automatic detection method is needed. To address the problem, this paper proposes an unexpected correlation detection method using the dynamic time warping.

3 Related Work

3.1 Research on Reputation Analysis

Various researchers have analyzed product reviews and reputation from social media [3,4,5,6]. Nagano et. al[3] propose the word-of-mouth engine to present product reputation on the Web. In their system, users first specify the products by taking pictures using cell-phone cameras. The system then retrieves word-of-mouth information and extracts typical evaluation words like “favorite,” “dislike,” “expensive,” and “useful” about the specific product. It also calculates positive/negative degrees. Kobayashi et. al[4] define the main portions of an opinion as (object, attribute, opinion). Asano et. al[5] also define the basic element of reputation as (object, evaluation point, expression). To extract reputation from word-of-mouth information, both propose a technique for efficiently building an object name dictionary, an attribute expression dictionary (ontology), and an opinion word dictionary for the specific object domain. Spangler et. al[6] propose an automated way to monitor social media to analyze the specific corporate brand, reputation, consumer preferences and buying habits. They also offer a mechanism for developing the ontology, near-real-time gathering of word-of-mouth information and the calculation of positive/negative measures. This related work targets specific products, extracts evaluation expressions from word-of-mouth in social media and calculates sentiment orientations of extracted expressions to analyze product reviews and reputation. They require specific ontology. Our proposed method, however, does not target specific products, and a specific ontology is not needed. We focus on a current topic and visualize the unforeseen relations between a current topic and unspecified products from buzz marketing sites. Through the visualization, we can detect unexpected consumer behavior.

3.2 Research on Analyzing Correlation over Time

Various researchers have analyzed correlation over time. Zhu et. al[7] propose a mechanism for finding a song by humming part of the tune. They use the dynamic time warping (DTW) technique and improve both the retrieval precision and speed by introducing existing dimensionality reduction to DTW indexes. Our approach is also based on the dynamic time warping technique, but focuses on word correlation over time. Word correlation is not relatively complicated in comparison with the hum tune, so that we can concentrate on how to utilize the result of detecting word correlation.

Otanto et. al[8] propose the Dynamic Conditional Correlation model, which uses the idea of distance between dynamic conditional correlations, and the classical Wald test, to compare the coefficients of two groups of dynamic conditional correlations. They apply their method to a set of financial time series. Loy et. al[9] propose an approach to understanding activities from their partial observations monitored through multiple non-overlapping cameras separated by unknown time gaps. They use a new Cross Canonical Correlation Analysis (xCCA) to formulate to discover and quantify the time delayed correlations of regional activities observed within and across multiple camera views in a single common reference space. Unlike existing approaches, we focus on word correlation over time to detect the unexpected correlation between a current topic and unspecified products. Our target data is different from existing approaches' target data.

3.3 Research on Detecting Word Relation over Time

Regarding research on detecting word relation over time, Radinsky et. al[10] propose a semantic relatedness model, Temporal Semantic Analysis (TSA) which captures the words' temporal information. It targets words in news archives (New York Times, etc.) and utilizes the dynamic time warping technique to compute a semantic relatedness between pre-defined words. Our approach is also based on the dynamic time warping technique. But, our aim is to detect the unexpected correlation between a current topic and unspecified products. Wang et al[11] propose time series analysis which has been used to detect similar topic patterns. They focus on specific bursty topic patterns in coordinated text streams and try to find similar topics. Their aim is to detect similar topic patterns.

While our work also makes use of temporally evolving statistics, our target data is word of mouth in buzz marketing sites and the goal is different in that we seek unspecified products that consumers show unexpected behavior for the products. We do not pre-define the products which have unexpected correlation with a current topic. We propose a new marketing research framework. This is the novel point of our work.

4 Detecting Correlation between a Current Topic and Products Using Dynamic Time Warping

As we mentioned above, the problem for our previous proposed framework is that how we can easily find the unspecified products which have unexpected correlation with a current topic. To address the problem, this paper proposes the unexpected correlation detecting method using the dynamic time warping. Our proposed method targets buzz marketing sites and try to find unforeseen correlation between a current topic and unspecific products. This method will be inserted into the previous proposed framework as "Step3-2. Correlation calculation" after Step2 (Language processing) (Fig. 3).

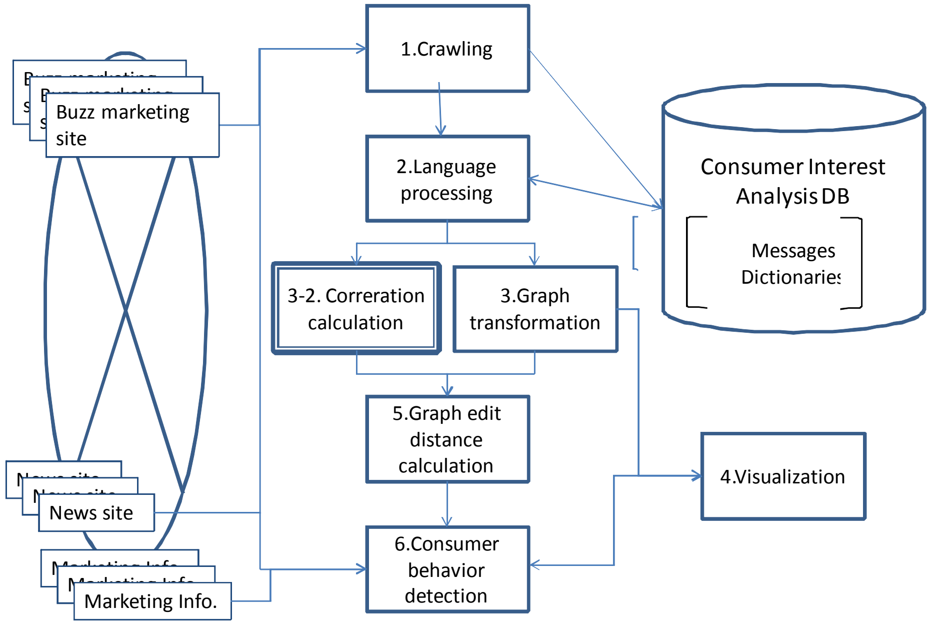

Fig. 3. New consumer behavior analysis framework including our proposed method

4.1 Dynamic Time Warping

This section explains the dynamic time warping (DTW) definition.

The dynamic time warping distance measures the similarity [17] between two time series that may differ in time scale, but similar in shape. For example, in speech recognition, this method is used to identify similar sounds between different speakers whose speech speed and pitch might be different. We use this technique to detect the correlation between a current topic and products. The influence by a current topic sometimes follows the development of the current topic. To address this time lag, DTW which can measure similarity between two sequences that may vary in time or speed is appropriate.

The standard definition of dynamic warping distance is as follow;

- **Definition 1.** Local cost matrix $C \in R^{|ts_1| \times |ts_2|}$ between two time series ts_1, ts_2 as

$$C_{i,j} \in \| ts_1[i] - ts_2[j] \|, i \in \langle 1 \ldots |ts_1| \rangle, j \in \langle 1 \ldots |ts_2| \rangle \tag{1}$$

where $\| ts_1[i] - ts_2[j] \|$ is a distance metric between two points of the time series.

Given this cost matrix, DTW constructs an alignment path that minimizes the cost over this cost matrix. This alignment p is called the "warping path," and defined as follows;

- **Definition 2.** Sequence of points pairs as

$$Pair_l = (pair_1, \ldots pair_k) \tag{2}$$

where $Pair_l = (i, j) \in \langle 1 \ldots |ts_1| \rangle \times \langle 1 \ldots |ts_2| \rangle$ is a pair of indexes in ts_1 and ts_2 respectively. Each consequent pair preserves the ordering of the points in ts_1 and ts_2, and enforces the first and last points of the warping path to be the first and last points of ts_1 and ts_2. For each warping path p we compute its cost as follows;

- **Definition 3.** Cost of warping path p as

$$c(p) = \sum_{i=1}^{k} c(pair_l) \tag{3}$$

The DTW is defined to be the minimum optimal warping path as follows;

- **Definition 4.** *DTW* between two time series ts_1, ts_2 as

$$DTW(ts_1, ts_2) = min\left\{ c(p) \mid p \in P^{|ts_1| \times |ts_2|} \right\} \tag{4}$$

where P are all possible warping paths. A dynamic programming algorithm (similar to the one in Fig. 6) is usually applied to compute the optimal warping path of the two sequences.

4.2 Step3-2: Correlation Calculation

Our proposed method will be inserted into the previous proposed framework as "Step3-2." Inputs for the step are results of Step2. The method will be done in parallel with Step3 and find product candidates as unspecified products that would have unexpected relationship with the current topic. In order to detect unexpected correlations, the method applies the above-mentioned dynamic time warping techniques for analyzing word correlation over time. The following is the process of our method.

1. At first, as for the results of Step2, the number of occurrences of a target current topic w such as "flu", "great earthquake", etc. in buzz marketing sites is counted according to the appropriate period (daily, weekly, monthly, etc.) delimited in advance.

2. We've decided products categories in kakaku.com as the products for calculating correlation with a current topic. It is not possible to calculate correlation for all existing products. Kakaku.com provides around 2000 product category list on their sites. Since the product categories provided by kakaku.com are well-organized and reliable, we've supposed that they are appropriate for calculating correlation. As for the results of Step2, we count the number of occurrences of these 2000 product categories, as well as the current topic. The number of occurrences is counted according to the appropriate period delimited in advance.

3. For both the current topic and each product category, we've calculated the occurrence pattern based on the following formula.

$$C_{tl} = \frac{n_{tl}}{\sum_{i=1}^{n} n_{ti}} \tag{5}$$

$$C_{pjl} = \frac{n_{pjl}}{\sum_{i=1}^{n} n_{pji}}, \quad \left\{ n_{pj} \in 2000 \ product \ categories \ in \ kakaku.com \right\} \tag{6}$$

Where n_{tl} is the number of occurrence for the lth period of the current topic and n_{pjl} is the number of occurrence for the lth period of the jth product.

4. Using dynamic time warping technique, we calculate correlations between the current topic occurrence pattern and each product occurrence pattern, then compute distances $D_{t\text{-}pj}$ for each correlation.

5. Products with high distance ($D_{t\text{-}pj} <= T$) will be extracted as product candidates which has a strong correlation with the current topic. Where T is a threshold for the distance.

These product candidates will be the result of this Step3-2 (Correlation calculation). As for the product candidates, in Step5, we seek substructures which include terms of product candidates from the concept graph structures. If there are substructures, they will be extracted, and the graph edit distance between substructures will be calculated. This graph edit distance calculation will be done based on our previous proposed method. Products with major graph structure changes will be recognized unspecified products which cause unexpected consumer behavior. Then, Step6 (Consumer behavior detection) refers a marketing data such as product shipments as an evidence of unexpected consumer behavior to confirm whether there are correlation between the product occurrence pattern and a marketing data.

5 Experimental Result

Based on our method, we conducted the experiments. This section shows the results of our experiment.

In our experiment, we set w="flu" and T=1.0. The delimited period is one week. At first, we retrieve post documents from January 2009 to December 2009 in buzz marketing sites of kakaku.com by the query "flu." As a result, 857 documents were retrieved. As for these 857 documents, we calculate the correlation between the "flu" and product categories of kakaku.com. To calculate the correlation based on the dynamic time warping, we used R[18] which is a free software environment for statistical computing and graphics. As an example, we select main products, "camera", "air purifier", "car", "printer", "mobile" and "television." Table 1 shows the distance calculation results between the "flu" and example products based on our calculation.

Fig. 4 shows the occurrence pattern of "flu", "air purifier" and "camera". And Fig. 5 shows the occurrence pattern of "flu", "car", "printer", and "television."

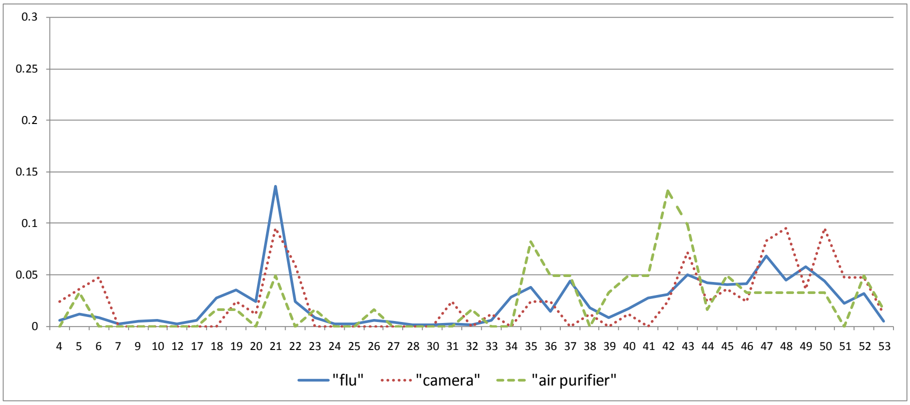

Fig. 4. Occurrence pattern of "flu", "air purifier", "camera", "car", "printer", and "television"

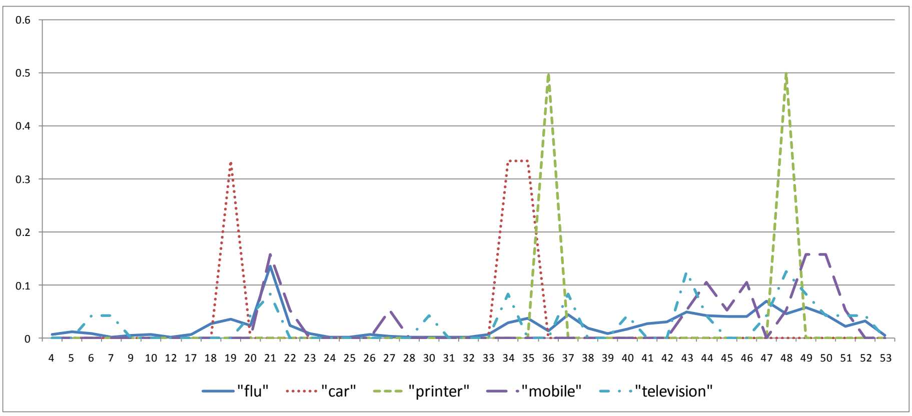

Fig. 5. Occurrence pattern of "flu", "air purifier", "camera", "car", "printer", and "television"

In Fig.4 and Fig.5, the horizontal axises show the number of week in 2009. The vertical axises show the occurrence ratio for each product derived by the formula (5) and (6). According to Fig.4, there seem correlations between the "flu" and the "camera"/ the "air purifier." On the other hand, in Fig.5, there seem less correlations between "flu" and other products. By computing the dynamic time warping paths, we confirm these correlations.

Fig. 6-11 show the warping path of each pair ("flu" and "camera", "flu" and "air purifier", "ful" and "car", "ful" and "printer", "ful" and "mobile", and "ful" and "television") in a time warping grid. Warping path distances of Fig.6 ("ful" and "camera"), and Fig. 7 ("flu" and "air purifier") are short. This means that both the relation between "ful" and "camera", and the relation between "flu" and "air purifier" are correlated. On the contrary, warping path distances of other products are longer. This means these products are not correlated with the "ful."

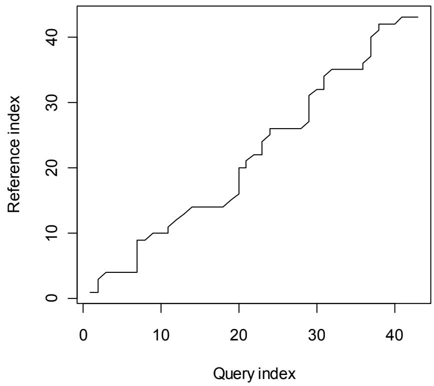

Fig. 6. Dynamic time warping path between the “ful” and the “camera”

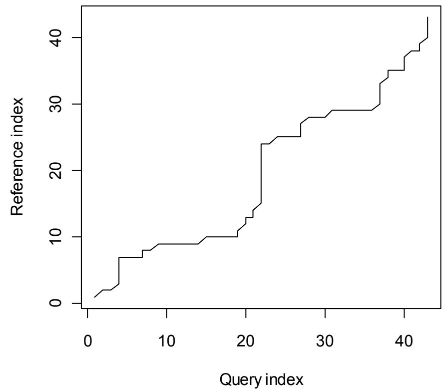

Fig. 7. Dynamic time warping path between the “ful” and the “air purifier”

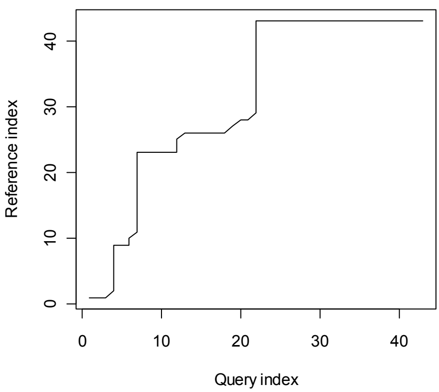

Fig. 8. Dynamic time warping path between the “ful” and the “car”

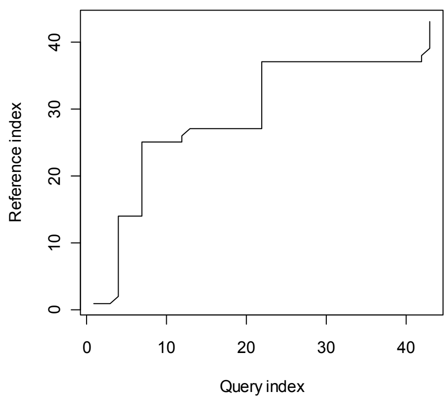

Fig. 9. Dynamic time warping path between the “ful” and the “printer”

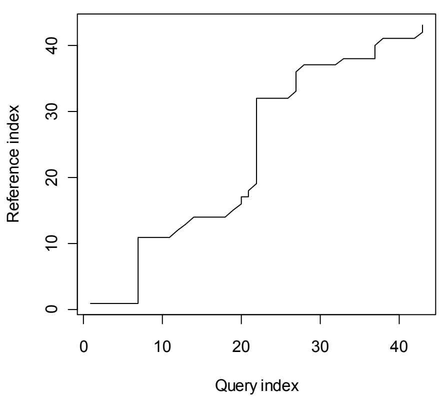

Fig. 10. Dynamic time warping path between the “ful” and the “mobile”

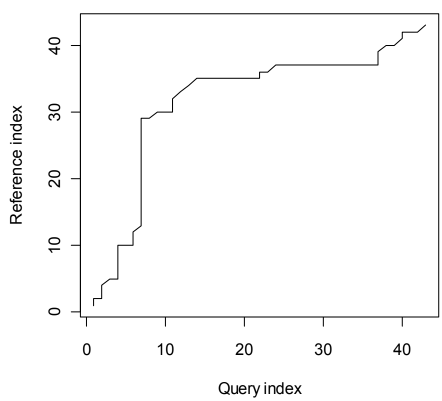

Fig. 11. Dynamic time warping path between the “ful” and the “television”

Table 1 shows the distance derived from the dynamic time warping for each pair. The distance of "camera" and "air purifier" are less than *T* (=1.0). On the other hand, the distance of "car", "printer", "mobile" and "television" are larger than *T*. Therefore, both "camera" and "air purifier" seem to correlate with the "flu." Based on the results, we can recognize "camera" and "air purifier" as product candidates which have correlations with the "flu."

Table 1. The distance based on dynamic time warping between "flu" and major products

	″camera″	″air purifier″	″car″	″printer″	″mobile″	″television″
″flu″	0.821	0.870	1.816	1.948	1.049	1.027

To confirm our result, we compare the occurrence pattern of product candidates with the real product shipments. Fig. 12 shows the volume of shipments for digital single-lens reflex camera in 2008 and 2009 and the occurrence pattern of camera in kakaku.com BBS sites. In 2008, an ordinary year, the lack sales in March, April, May, September and November do not exist. On the contrary, in March to May and September to November 2009, the volume of shipments in 2009 is negatively correlated with the occurrence pattern of camera. We can say that for a camera, consumers' unexpected behavior appears. On the other hand, Fig. 13 also illustrates the volume of shipments for air purifiers in 2009 and the occurrence pattern of air purifiers in kakaku.com BBS sites. We easily detect a correlation between the volume of shipments and the occurrence pattern. We recognize this kind of explicit relationships as expected consumer behavior.

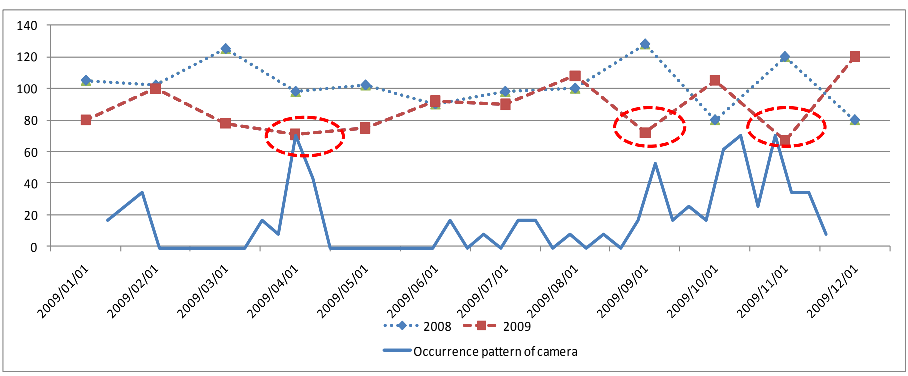

Fig. 12. The occurrence pattern of camera and the volume of shipments for digital single-lens reflex camera in 2008 & 2009. (Cited: The Camera Information Center: Camera information Center Report, http://www.camera-info.net/index.htm).

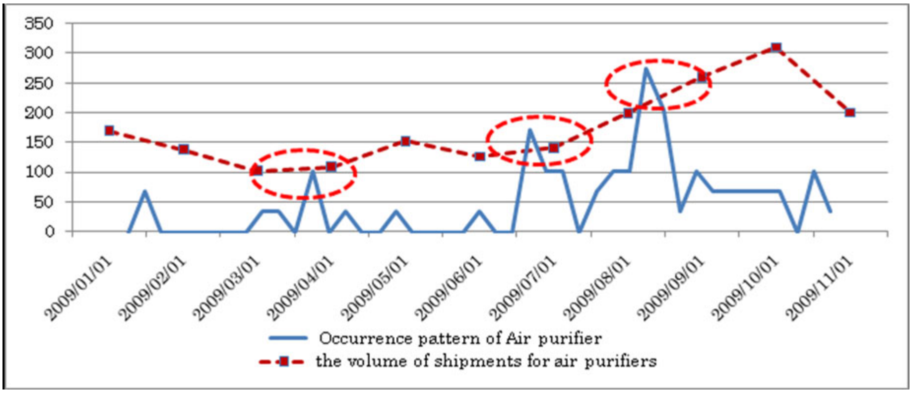

Fig. 13. The occurrence pattern of air purifier and the volume of shipments for air purifier in 2009. (Cited: : GfK Marketing Services Japan Ltd., http://www.gfkjpn.co.jp/).

Regarding the scalability, we have not applyed our proposed method for a large data yet. In fact, our method does not need to evaluate all time series data to find the optimal alignment. We plan to detect the specific bursty topic patterns , and then find correlations between the current topic and products. As for the algorithm for detection bursty structures in data streams, there are several researches[19, 20]. And there are also several proposals to improve the computational efficiency of the dynamic time warping[21, 22]. Especially, Salvador et. al[21] proposed FastDTW, an approximation of DTW that has a linear time and space complexity. We are going to evaluate these algorithms and introduce appropriate algorithms into our method.

6 Conclusion

In this paper, we proposed a method to detect correlations between a current topic and unspecified products from word of mouth in buzz marketing sites. For detecting unexpected correlations, this paper applies the dynamic time warping techniques to analyze over time. Our method computes time series correlation based on the occurrence patterns of both a current topic and products, and selects product candidates which may cause consumers' unexpected behavior. Our proposed method uses the dynamic time warping technique to compute time series correlation between a current topic and products. The method calculates the dynamic time warping paths for the 2000 product categories, which are classified in kakaku.com, and extracts product candidates which will have a unexpected correlation with a current topic. According to our method, we conducted the experiments and confirm the effectiveness. By our method, we could detect the unforeseen correlation between the “flu” and “camera”, while there are no relatedness between the “flu” and “car”, “printer”, and “television.” Of course, we confirmed expected correlations between “flu” and “air purifier.” We indicated the real marketing data about the "camera" as an evidence of unexpected consumer behavior to confirm unforeseen correlation.

As future work, we will acquire other data examples that can express unexpected consumer behavior from buzz marketing sites, and evaluate the effectiveness of our proposed method using the dynamic time warping. In addition, at the moment, from the product candidates which have correlations with the current topic, we manually decide products which have unexpected correlations. To automatically detect products with unexpected correlations, we investigate the measure for judging unexpected correlations.

Our proposed method is part of a marketing analysis framework which can detect unexpected correlation between a current topic and products. We can say that by detecting unexpected correlation, unforeseen consumer behaviors can be further analyzed and we can achieve the new marketing analysis.

References

1. Shimoji, Y., Wada, T., Hirokawa, S.: Dynamic Thesaurus Construction from English-Japanese Dictionary. In: The Second International Conference on Complex, Intelligent and Software Intensive Systems, pp. 918–923 (2008)
2. kakaku.com, `http://kakaku.com/`
3. Nagano, S., Inaba, M., Mizoguchi, Y., Iida, T., Kawamura, T.: Ontology-Based Topic Extraction Service from Weblogs. In: IEEE International Conference on Semantic Computing, pp. 468–475 (2008)
4. Kobayshi, N., Inui, K., Matusmoto, Y., Tateishi, K., Fukushima, S.: Collecting evaluative expressions by a text mining technique. IPSJ SIG NOTE 154(12), 77–84 (2003)
5. Asano, H., Hirano, T., Kobayasi, N., Matsuno, Y.: Subjective Information Indexing Technology Analyzing Word-of-mouth Content on the Web. NTT Technical Review 6(9), 1–7 (2008)
6. Spangler, W.S., Chen, Y., Proctor, L., Lelescu, A., Behal, A., He, B., Griffin, T.D., Liu, A., Wade, B., Davis, T.: COBRA - mining web for COrporate Brand and Reputation Analysis. Web Intelligence and Agent Systems (WIAS) 7(3), 243–254 (2009)
7. Zhu, Y., Shasha, D.: Warping Indexes with Envelope Transforms for Query by Humming. In: Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data, pp. 181–192 (2003)
8. Otranto, E.: Identifying financial time series with similar dynamic conditional correlation. Journal of Computational Statistics & Data Analysis archive 54(1), 1–15 (2010)
9. Loy, C., Xiang, T., Gong, S.: International Journal of Computer Vision 90(1), 106–129 (2010)
10. Radinsky, K., Agichtein, E., Gabrilovich, E., Markovitch, S.: A Word at a Time: Computing Word Relatedness using Tremporal Semantic Analysis. In: WWW 2011, pp. 337–346 (2011)
11. Wang, G., Araki, K.: A Graphic Reputation Analysis System for Mining Japanese Weblog Based on both Unstructured and Structured Information. In: AINA Workshops 2008, pp. 1240–1245 (2008)
12. Iino, Y., Hirokawa, S.: Time Series Analysis of R&D Team Using Patent Information. In: Velásquez, J.D., Ríos, S.A., Howlett, R.J., Jain, L.C. (eds.) KES 2009. LNCS, vol. 5712, pp. 464–471. Springer, Heidelberg (2009)

13. Hashimoto, T., Shirota, Y.: Semantics Extraction from Social Computing: A Framework of Reputation Analysis on Buzz Marketing Sites. In: Kikuchi, S., Sachdeva, S., Bhalla, S. (eds.) DNIS 2010. LNCS, vol. 5999, pp. 244–255. Springer, Heidelberg (2010)
14. Bunke, H.: On a relation between graph edit distance and maximum common subgraph. Pattern Recognition Letters 18(8), 689–694 (1997)
15. Hashimoto, T., Kuboyama, T., Shirota, Y.: Graph-based Consumer Behavior Analysis from Buzz Marketing Sites. In: Proc. of 21st European Japanese Conference on Information Modelling and Knowledge Bases, pp. 60–71 (2011)
16. Kuboyama, T., Hashimoto, T., Shirota, Y.: Consumer Behavior Analysis from Buzz Marketing Sites over Time Series Concept Graphs. In: Proc. of 15th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems, pp. 73–83 (2011)
17. Berndt, D., Clifford, J.: Using dynamic time warping to find patterns in time series. In: Proc. of Advances in Knowledge Discovery and Data Mining, pp. 229–248. AAAI/MIT (1994)
18. R, a language and environment for statistical computing and graphics, `http://www.r-project.org/`
19. Zhu, Y., Shasha, D.: Efficient Elastic Burst Detection in Data Streams. In: Proc. of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 336–345 (2003)
20. Kleinberg, K.: Bursty and hierarchical structure in streams. In: Proc. of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 91–101 (2002)
21. Salvadore, S., Chan, P.: FastDTW: Toward accurate dynamic time warping in linear time and space. In: Proc. of 3rd Workshop on Mining Temporal and Sequential Data, pp. 561–580 (2007)