Proceedings of the International MultiConference of Engineers and Computer Scientists 2014 Vol I,
IMECS 2014, March 12 - 14, 2014, Hong Kong

Measurement of Bloggers' Buzzword Prediction Ability Based on Analyzing Frequency of Early Mentions of Past Buzzwords

Seiya Tomonaga, Shinsuke Nakajima, Yoichi Inagaki, Reyn Nakamoto, Tsukasa Ogura, and Jianwei Zhang

Abstract—**The goal of our research is to discover factors which predict which words will become buzzwords–terms representing topics that have become popular–within the blogosphere. In this paper, we propose a method which evaluates bloggers' buzzword prediction ability by analyzing how early bloggers mentioned past buzzwords. We do so by measuring how early a buzzword is first mentioned until the buzzword's peak in popularity. We describe this method and also report the evaluation on buzzword classification.**

Index Terms—**blog mining, blogger's buzzword prediction ability, trend analysis**

I. Introduction

Discovering buzzword before they become popular is very difficult task. However, it is very beneficial to do so, especially from say a marketing point of view. Therefore, we are working on developing methods to discover buzzwords early based on analyzing consumer generated media such as blogs and social media. By analyzing these such sources, we believe it is possible to discover future buzzword candidates.

There has been several related works in regards to predicting which topics would become buzzwords. Nakajima et al.[1][2] implemented a system which predicted what topics would likely become buzzwords in the future and found moderate success in our experiments. In the paper, we proposed a method to discover bloggers who frequently mentioned buzzwords before they became popular and used these bloggers to find new buzzword candidates. Here, bloggers who were good predictors of buzzwords in a certain field or category were assumed to continue being a good predictor of buzzwords in the future.

In our own previous research[3][4], we looked into measuring a blogger's buzzwords prediction ability. However, we did not look into automatic detection of past buzzwords and additionally, we did not take into account the time from the first mention of a buzzword until the peak popular of a buzzword.

In this paper, we propose automatic detection method of past buzzwords and a method for estimation of the time from the first appearance to the peak of popularity. We utilize not only the buzzword itself, but also consider the related co-occurrence words, which captures the depth and richer information of the topic.

For example, in the case of the iPhone 5, we would value a blogger who mentions the specs and features of iPhone5 over a blogger who just mentions that they want a iPhone 5. The blogger who mentioned the specs would most likely be more knowledgeable and therefore a better predictor of future buzzwords.

We measure the growth period of buzzword, and then we rate a blogger's buzzword predictive ability by rating how early a blogger mentions a buzzword and it's related co-occurrence words. With this, we can provide a method to find bloggers with high buzzword predictive ability. In addition, in order to identify areas of bloggers' buzzword prediction ability, we categorize buzzword candidate.

In Section 2, we describe related work. In Section 3, we describe the measurement of bloggers' buzzword prediction ability based on analyzing frequency of early mentions of past buzzwords. In Section 4, we describe the evaluation about category classification of buzzword candidate. In the last Section, we describe our future work as well as our conclusions.

II. Related work

Research for discovering buzzwords by analyzing blog are as follows. Okumura et al, proposed a system for measure how popular a keyword has become by looking at the changes in frequency of a keyword [6]. Furukawa et al looked into determining the most important words by tracking how a topic propagated[7].

There also existing systems for detecting buzzwords or trends, such as Yahoo! Buzz Index [8], BlogPulse [9]. Yahoo! Buzz Index [8] calculates a topic's buzz score based on the percentage of Yahoo! users searching for that subject on a given day, and identifies "leaders" (subjects with the highest buzz scores) and "movers" (subjects with the highest percentage increase in buzz scores from one day to the next). BlogPulse [9] extracted key phrases and key people from blog entries by calculating the ratio of the frequency of occurrence of a phrase or a person name to its average frequency over the past two weeks.

These systems analyze the number of bloggers who write a word as well as the change in frequency in order to extract buzzwords. In contrast, our proposed method first discovers bloggers with good buzzword prediction ability and then extracts buzzwords candidates from their blogs.

Seiya Tomonaga, Shinsuke Nakajima, Tsukasa Ogura are with Faculty of Computer Science and Engineering, Kyoto Sangyo University, Motoyama, Kamigamo, Kita-ku, Kyoto-city, Kyoto-City, 603-8555, Japan e-mail: ({i1358068, nakajima, g1044228}@cse.kyoto-su.ac.jp).

Yoichi Inagaki, Reyn Nakamoto are with Kizasi Company, Inc. 6th floor Tomoe Nihonbashi Building, 20-14 Hakozaki-Cho, Nihonbashi, Chuo-ku, Tokyo, 103-0015 Japan e-mail: ({inagaki,reyn}@kizasi.jp).

Jianwei Zhang is with Faculty of Industrial Technology, Tsukuba University of Technology Amakubo 4-3-15, Tsukuba-City, Ibaraki, 305-8520 Japan e-mail: (zhangjw@a.tsukuba-tech.ac.jp).

III. Measurement of Bloggers' Buzzword Prediction Ability

We extract seed words (buzzwords of past) by analyzing past blog articles and then bloggers who mention these words early are determined to have high buzzword predictive ability. Our data set comes from kizasi.jp, a company which mines blog data. As of September 6, 2013, the data set consisted of 12,103,387 bloggers and 172,018,786 entries.

The steps of our method is as follows and will be explained in each subsection:

1) Classification of bloggers into knowledge groups（Section Ⅲ-A）
2) Extraction of seed words by analyzing blog set（Section Ⅲ-B）
3) Estimation of growth period of seed words（Section Ⅲ-C）
4) Calculation of bloggers' buzzword prediction score for individual entries（Section Ⅲ-D）
5) Calculation of bloggers ' buzzword prediction ability（Section Ⅲ-E）

A. Classification of bloggers into knowledge groups

Our purpose is to discover bloggers who pickup and discuss new topics before they popular. In order to do so, it is determine which areas each blogger is knowledgeable in. We assume that a blogger who possess a high level of knowledge about a certain field or category should be a good predictor of future buzzwords.

In the same line of thinking, a blogger who is knowledgeable about one category would not necessarily be a good predictor of buzzword in another category. For example, a blogger who frequency writes about internet would not necessarily be a good predictor of buzzwords in the topic of the economy.

Therefore, we classify each blogger into knowledge groups and then rank the bloggers within that group according to the knowledge level.

We use knowledge groups which were determined from our previous research. This system grouped bloggers into knowledge groups and then ranked their knowledge within that particular group.

B. Extraction of seed words by analyzing blog set

1) Extraction of seed word candidates: We first extract seed word candidates. We start with using a keyword ranking system provided kizasi.jp[11] which rankings popular keywords written in blogs. This list shows the top keywords written in blogs for that day.

We first take the top hundred keywords for each day over a two year period. Next, we exclude repeated words, common words, certain seasonal keywords, and lastly infrequently occurring words. Seasonal keywords are words that appear regularly depending on the season. For example, "New Year's" or "Summer Festival". Also, every four years, there would keywords such as "Olympic", "FIFA World Cup" and "WBC" etc. The remaining keywords become our seed word candidates.

2) Category classification of seed word candidates: Since it is necessary to classify buzzwords into categories, we must determine which categories each seed word candidate belongs to. In addition, in order to efficiently find bloggers who are good buzzword predictors, we focus on bloggers who are knowledgeable in the categories in which these seed words belong.

Thus, it is necessary to find the category which matches the each keyword the best. For categories, we use the knowledge groups which will be described in Section Ⅲ-A .

In order to classify a seed word, we first calculate the semantic proximity of a seed word candidate and knowledge groups. We do so by calculating the similarity of co-occurrence word set of seed word candidate and the co-occurrence word set of a knowledge group. Each co-occurrence word set of a seed word candidate consists of the top 400 co-occurring words in all blog posts in the data set. Each co-occurrence word set for a knowledge group is the top 400 co-occurring words in blog posts which were written by bloggers who belong to the knowledge group. We considered using three different measures for calculation similarity between these word sets: Jaccard similarity measure, Simpson diversity Index, and lastly, cosine similarity.

3) Recognition of seed words based on degree of influence: Since we are trying to extract bloggers who mentioned past buzzword before they become popular, it is necessary to focus on influential seed words.

To calculate a seed word's influence, we use the total number blog posts of a word during period T after the peak of the number of blog entries containing the word.

Here are the steps for calculating the degree of influence: First, we investigate the number of posts of each of the seed word candidates. Second, we calculate the moving average of the number of posts for each seed word candidate for the past two years. We then confirm the peak of the number of posts for that seed word candidate.

We assume that the peak is the point of highest recognition by the general population. If the number of posts are very low during period T after the peak, we assume the seed word candidate is quickly forgotten and is of low influence. If the number of posts does not decrease during period T following the peak, that keyword's influence is high. These words become seed words（fig1）.

In addition, we find knowledge groups of high relevance to each seed word candidate. Since we identify areas of seed word as learning data.

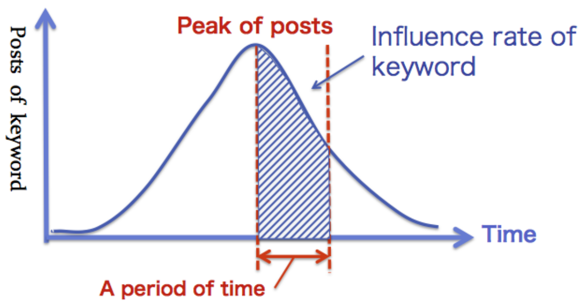

Fig. 1. Conceptual view of influence rate

ISBN: 978-988-19252-5-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

C. Estimation of growth period of seed words

We now describe the process for the estimation of the growth period of a seed word. The growth period is the period from when a seed word first starts to be posted in blogs until the peak of a seed word's popularity.

1) Similarity by word set of peak and previous peak: In order to estimate the growth period of the seed words, we first find when a seed word starts to be mentioned. In doing so, it is necessary to confirm that the topic contents at that time matches the content of when it is at its peak.

For example, if "iOS5" is the seed word, the peak time would contains keywords about the specifications or features the new iOS. If a blog post just asking when iOS 5 is to be released rather than talking about specifics, it would not match the peak time and thus, we cannot consider the blogger to be have good predictive ability.

In other words, to use only the seed word "iOS5" and not write any specific or useful information is not considered to be a early mention of a buzzword.

On the other hand, if a blogger posts that the "voice assistant feature named Siri will be included on the iOS 5", we can considered that the blogger has some prediction ability about the related category.

We estimate the time when the topic of a seed word begins including the nuances of the peak of the epidemic by calculating the similarity between the co-occurrence word set of the seed word before its peak to the co-occurrence word set of its peak. For example, here is "iOS5" (fig 2).

In addition, the co-occurrence word set of seed words are top 400 co-occurring words after removal of generic terms during the topics peak popularity. Words that appear more frequently as we approach the peak of popularity are of particular importance.

In order to discover specific keywords at the time of peak popularity, we use Spearman's rank correlation to determine if there is an increasing number of blog posts containing a co-occurrence word.

If there is a correlation of fairly strong positive (more than 0.6 coefficient ρ), this co-occurrence word has increasing number of posts as we approach the peak and it is considered to be a important keyword at the time of its peak popularity.

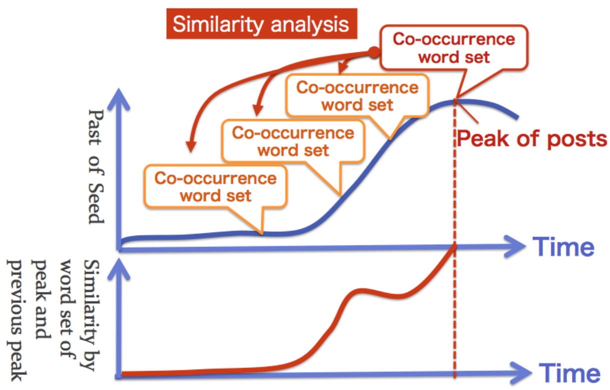

Fig. 2. Similarity by word set of peak and previous peak

2) Determination of the growth period of seed word: It is possible to examine the change in content from a specific period to the peak period by calculating similarity between topics of that specific period to the peak period.

If there is a rise in similarity in the vicinity of the peak, we consider this point to be when the seed word began.

However, there are no guarantee that the similarity is zero during other points. We much consider points where there is some, but low similarity to be background noise.

Therefore, we consider the following two methods for determining beginning of the growth period of seed word.

1) Sections of the first-order approximation line of the similarity of word set of peak to previous peak is the start of the growth period of the seed word(fig 3).
2) Set the appropriate threshold θ for the similarity of the popularity peak and previous peak and the point at which it exceeds the first threshold θ is the start of the growth period(fig 4).

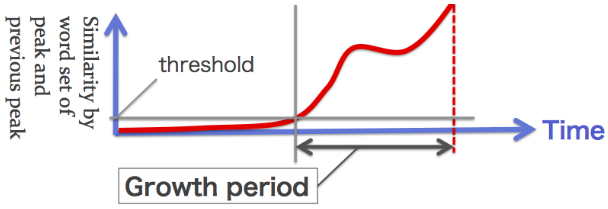

Fig. 3. Growth period of seed words for method 1

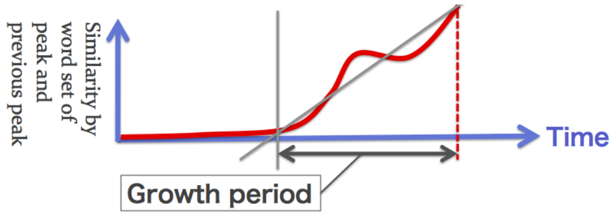

Fig. 4. Growth period of seed words for method 2)

Once we have determined the start point of the growth period of the seed word, we now determine the duration of the growth period from the start until the blog entry peak point.

D. Calculation of bloggers' buzzword prediction score for individual entries

In this section, we explain the method for calculating bloggers' buzzword prediction score for blogger entries.

The bloggers' buzzword prediction score is calculated for each blog article posted in the growth period of the seed word. Highest scores are given to entries near the beginning of the growth period. The lowest scores are given to entries at the end of the growth period, i.e. the peak.

The calculation for a blogger's buzzword prediction score $PredictionScoreEntry_i$ for entry $entry_i$ is shown in formula 1.

$$PredictionScoreEntry_i = \frac{(entry_{all} + 1) - order_i}{entry_{all}} \quad (1)$$

$entry_{all}$ is the number of entries that a blogger posted about the seed word during the growth period of seed word.

$order_i$ is the value representing how early a blogger early posted about the seed word during the growth period. It is

ISBN: 978-988-19252-5-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

normalized value from 0 to 1 based on the particular entry order in that blogger's posts during that period.

For example, say the number of entries $entry_{all}$ is 100. $order_i$ would be assigned to entries 1 to 100 as as $1, 0.99, \cdots, 0.02, 0.01$.

E. Calculation of bloggers' buzzword prediction ability

In this section, we explain calculation method of bloggers' buzzword prediction ability of each seed word.

Below, we show conditions which reflect a high prediction ability for blogger's buzzword prediction.

- Posting articles related to the target seed word early in the growth period.
- Posting many articles related to the target seed word in the growth period.

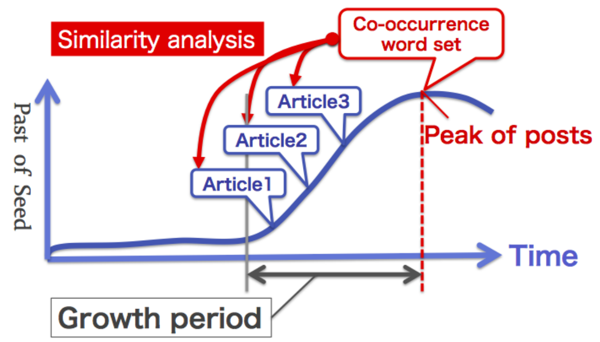

Fig. 5. Similarity analysis of blog posts for bloggers' buzzword prediction ability

The calculation of bloggers' buzzword prediction ability $PredictionScore_{(A,x)}$ for a seed words A is show in formula 2

$$PredictionScore_{(A,x)} = \sum_{k=1}^{N} Sim(D_A, entry_k) \times PredictionScoreEntry_k \tag{2}$$

N is the number of articles that blogger x posted in the growth period. D_A is the co-occurrence word set at the peak of seed word A. $Sim(D_A, entry_k)$ is the similarity between the topic of peak of seed word A and topic of article $entry_k$. $PredictionScoreEntry_k$ is the bloggers' buzzword prediction score for entry $entry_k$ as described in Section Ⅲ-D.

IV. Category Classification Evaluation Experiment of Buzzword Candidates

In this section, we discuss our evaluation experiment for category classification of seed word candidates as described in Section Ⅲ-B-2.

It is necessary that we categorize seed word candidates in order to determine which category each buzzword is related to. In addition, in order to efficiently discover bloggers who are good buzzword predictors, we focus on bloggers who are knowledgeable in topics related to the seed word candidates. From this reason, it is necessary that we determine which knowledge groups are semantically similar to each seed word candidate.

The categories used are taken from "Blog Ranking Based on Bloggers' Knowledge Level for Providing Credible Information"[10]. However, in that paper, many of the categories are too specific. Therefore, we use only the top 120 categories from that.

A. Category classification method of seed words

The purpose of our experiment is to confirm the method can accurately classify seed words into appropriate categories. In other words, the goal is to evaluate if we can assigned a strongly related group to each seed word.

The relevance measure, which represents how semantically similar a seed word and a knowledge group is, is calculated by the similarity between the co-occurrence word set of the seed word candidate and the knowledge group. After that, we take the knowledge group which has the high similarity. Similarities measures used in this evaluation experiment are jaccard coefficient (3), simpson coefficient (4), cosine similarity (5), and co-occurrence ranking points weighted cosine similarity(6).

$$jaccard(S, C) = \frac{|S \cap C|}{|S \cup C|} \tag{3}$$

$$simpson(S, C) = \frac{|S \cap C|}{min(|S|, |C|)} \tag{4}$$

$$cosSim(S, C) = \frac{|S \cap C|}{\sqrt{|S|^2}\sqrt{|C|^2}} \tag{5}$$

$$cosSim_{Rank}(S, C) = \frac{\sum_{i=1}^{N}(r_{Si} \cdot r_{Ci})}{\sqrt{\sum_{i=1}^{N} {r_{Si}}^2}\sqrt{\sum_{i=1}^{N} {r_{Ci}}^2}} \tag{6}$$

Here, S is the word set of the top 400 words co-occurring with the seed word candidate with in the set of crawled blog entries C is the word set of the top 400 words co-occurring with the knowledge group's representative keyword.

In addition, r_S is ranking score from 400 to 1 that we allocate to each keyword S in order of co-occurrence degree. r_C is ranking score from 400 to 1 that we assign to each keyword C in order of their co-occurrence degree. N is number of keywords to be compared–in this experiment, it is 400.

Using each of the four methods, we determine which seed words and knowledge groups have high similarities, and then determine which of the four methods gave the best results

As an additional baseline comparison, we used a method from "ratios of bloggers in a community who talked about the target seed word to all the bloggers in the community[1][2]". The method is based on the idea that if the majority of knowledge group P mentions seed words Q, that word is related to group P.

B. Method of experiment and evaluation

Seed words which were used this experiment are "AKB48", "Android" , "Facebook" , "Joshikai" , "K-POP" and "Smartphone". These words were popular keywords in the past. The goal of this experiment is to investigate that proposed method can automatically determine the correct knowledge group for each keyword. We first manually created a ranking of the top ten related categories for each of the six seed words. We call the rankings "ideal knowledge group ranking". The similarity method that produced categories most similar to our ideal knowledge group would be the best method.

Blog posts to be used in this experiment were separated into three month periods from July 2012 to July 2013 as such: "2012.07.01 - 2012.09.30 (period 1)", "2012.09.30 - 2012.12.31 (period 2)", "2012.12.31 - 2013.04.01 (period 3)" and "2013.04.01 - 2013.07.02 (period 4). Then, co-occurrence lists were created for each period for each knowledge group. The reason for having multiple periods is that we assume that over time a seed words related categories would change, and thus, we wished to confirm the changing of related categories. In addition, we use $nDCG$ and DCG as comparative evaluation measure. DCG is discounted cumulative gain, and it evaluates whether a correct ranking, including order, can be reproduced. In other words, it is a measure in which not only whether the ideal ranking is included, but also considers whether ranking order is correct as well. Additionally, $nDCG$ is normalized discounted cumulative gain, and its values are additionally normalized to 1.

We show the formula of DCG and $nDCG$ below.

$$DCG_p = rel_1 + \sum_{i=2}^{p} \frac{rel_i}{log_2\,(i)} \tag{7}$$

$$nDCG_p = \frac{DCG_p}{IDCG_p} \tag{8}$$

rel_i is a item's score of the i th item in the target ranking. p shows the number of items in the target ranking. In this experiment, p = 10. In addition, $IDCG_p$ shows the ideal value of DCG_p. Also, it is possible to add a higher weight to the top items in a ranking, and thus, the value of DCG is higher if a method can rank a top item of ranking higher. In our experiment, we used 10, 9, 8, ..., 1 for our ideal ranking and then calculated DCG. We regard this ideal ranking's DCG as $IDCG_p$.

C. Results of evaluation of knowledge group ranking based on relevance

Here are the results of our experiment. TableI〜VI shows the $nDCG$ of previous studies as well as each method. A is simpson's coefficient, B is jaccard's coefficient, C is cosine similarity, and D is weighted cosine similarity based on co-occurrence point.

Looking at tableI〜VI, ranking result is exactly the same result; however, each score of method A,B and C is different. Therefore, the value of $nDCG$ is also the same value. On the other hand, value of $nDCG$ of method D is different value of $nDCG$ of method A, B, and C. By contrast, results of ranking of table VI and IV are better than method A, B, and C. Lastly, the results of ranking other seed words are lower than method A, B, and C.

At first, we had expected that method D to produce the best results; however, it did not always do so. In the future, we will consider tuning method D and further investigating other seed words. We plan to looking to find which method

TABLE I
nDCG OF EACH METHOD TO IDEAL RANKING (AKB48)

	Proposed method				Previous method
Period	A	B	C	D	
Period1	**0.617**	**0.617**	**0.617**	0.549	0.091
Period2	0.605	0.605	0.605	**0.627**	0.144
Period3	**0.549**	**0.549**	**0.549**	0.300	0.104
Period4	**0.572**	**0.572**	**0.572**	0.547	0.102
Average	**0.586**	**0.586**	**0.586**	0.506	0.110

TABLE II
nDCG OF EACH METHOD TO IDEAL RANKING (ANDROID)

	Proposed method				Previous method
Period	A	B	C	D	
Period1	0.834	0.834	0.834	**0.838**	0.584
Period2	**0.880**	**0.880**	**0.880**	0.819	0.584
Period3	**0.806**	**0.806**	**0.806**	**0.806**	0.579
Period4	**0.813**	**0.813**	**0.813**	0.765	0.531
Average	**0.833**	**0.833**	**0.833**	0.807	0.569

TABLE III
nDCG OF EACH METHOD TO IDEAL RANKING (FACEBOOK)

	Proposed method				Previous method
Period	A	B	C	D	
Period1	**0.476**	**0.476**	**0.476**	0.403	0.327
Period2	0.384	0.384	0.384	**0.402**	0.327
Period3	0.384	0.384	0.384	**0.401**	0.327
Period4	**0.491**	**0.491**	**0.491**	0.455	0.327
Average	**0.433**	**0.433**	**0.433**	0.415	0.327

TABLE IV
nDCG OF EACH METHOD TO IDEAL RANKING (JOSHIKAI)

	Proposed method				Previous method
Period	A	B	C	D	
Period1	**0.391**	**0.391**	**0.391**	0.373	0.033
Period2	0.302	0.302	0.302	**0.397**	0.034
Period3	0.443	0.443	0.443	**0.469**	0.033
Period4	0.333	0.333	0.333	**0.385**	0.033
Average	0.367	0.367	0.367	**0.406**	0.033

TABLE V
nDCG OF EACH METHOD TO IDEAL RANKING (K-POP)

	Proposed method				Previous method
Period	A	B	C	D	
Period1	**0.829**	**0.829**	**0.829**	0.683	0.486
Period2	0.676	0.676	0.676	**0.708**	0.506
Period3	**0.658**	**0.658**	**0.658**	0.618	0.503
Period4	**0.689**	**0.689**	**0.689**	0.645	0.550
Average	**0.713**	**0.713**	**0.713**	0.663	0.511

TABLE VI
nDCG OF EACH METHOD TO IDEAL RANKING (SMARTPHONE)

	Proposed method				Previous method
Period	A	B	C	D	
Period1	0.775	0.775	0.775	**0.909**	0.050
Period2	0.678	0.678	0.678	**0.869**	0.071
Period3	0.553	0.553	0.553	**0.754**	0.071
Period4	0.676	0.676	0.676	**0.700**	0.071
Average	0.670	0.670	0.670	**0.808**	0.066

works best in differing conditions.

In any case, the accuracy of the proposed method is higher than accuracy of previous studies. Therefore, we consider the basis of the method to be effective. We experimented with four periods of each three month. However, as the value of $nDCG$ differed, the ranking which was calculated for each period had differing results. The implication is that blog content differ for each period.

Also, we were able to rank the top ten relevant knowledge group using each of the proposed methods, and we particular found success with the top one or two relevant rankings. However, our proposed method still has areas that need improving. If we consider limiting the the ranking to the top few, we believe the method is quite effective.

V. Conclusion

Our goal is develop to method for discovering bloggers with buzzword prediction ability, and then using them to detect buzzword candidates from those bloggers' contents". In this paper, we proposed a method for determining a bloggers' prediction ability by analyzing content and usage period for past buzzwords. Also, we experimented with buzzword candidate category classification and we were able to get positive results.

In the future, we will implement and evaluate the estimation of growth period of seed words " (Section Ⅲ-C), the calculation of bloggers' buzzword prediction score for individual entries " (Section Ⅲ-D), the calculation of bloggers' buzzword prediction ability " (Section Ⅲ-E). We are also working towards implementing a system of practical use.

Acknowledgment

This work was supported in part by the MEXT Grant-in-Aid for Scientific Research(C) (#23500140).

References

[1] Shinsuke Nakajima, Jianwei Zhang, Yoichi Inagaki and Reyn Nakamoto. Early Detection of Buzzwords Based on Large-scale Time-Series Analysis of Blog Entries, 23rd ACM Conference on Hypertext and Social Media (ACM Hypertext 2012), pp.275-284, June 2012.

[2] Shinsuke Nakajima, Jianwei Zhang, Yoichi Inagaki, Reyn Nakamoto, Early Detection of Gradual Buzzwords Based on Large-scale Time-series Analysis of Blog Entries, Information Processing Society of Japan:Database(TOD56), 2013.

[3] Shinsuke Nakajima, Adam Jatowt, Yoichi Inagaki, Reyn Nakamoto, Jianwei Zhang, Katsumi Tanaka: "Finding Good Predictors in Blogsphere Based on Temporal Analysis of Posting Patterns", DBSJ Journal, Vol.10, No.1, pp.13-18, June 2011.

[4] Seiya Tomonaga, Shinsuke Nakajima, Adam Jatowt, Yoichi Inagaki, Reyn Nakamoto, Jianwei Zhang, Katsumi Tanaka, Investigating predictability levels in Blogosphere based on Temporal Analysis of Blog Articles, Proceedings of SoC2012, pp.79-84, June 2012.

[5] Seiya Tomonaga, Shinsuke Nakajima, Jianwei Zhang, Yoichi Inagaki, Reyn Nakamoto, Method for Measuring Bloggers' Buzzword Prediction Ability Based on Analysis of Past Buzzword Mention Frequency, DEIM Forum 2013, C1-2, March 2013.

[6] Manabu Okumura, Blog Mining : Towards Trend and Sentiment Analysis on the Web(<Special Issue>Web Interaction in the Era of Social Network), The Japanese Society for Artificial Intelligence, Vol.21, No.4, pp.424-429, 2006.

[7] Tadanobu Furukawa, Yutaka Matuo, Ikki Ohmukai, Koki Uchiyama, Mitsuru Ishizuka . Keyword Discrimination using Topic Diffusion Process in Weblogs, journal of Japan Society for Fuzzy Theory and Intelligent Informatics, Vol.21, No.4, pp.557–566, 2009.

[8] Yahoo! Buzz Index. http://buzzlog.yahoo.com/overall/

[9] N. S. Glance, M. Hurst and T. Tomokiyo, "BlogPulse: Automated Trend Discovery for Weblogs," In WWW 2004 workshop.

[10] Shinsuke Nakajima, Jianwei Zhang, Yoichi Inagaki, Tomoaki Kusano, and Reyn Nakamoto, Blog Ranking Based on Bloggers' Knowledge Level for Providing Credible Information. Proc. of the 10th International Conference on Web Information Systems Engineering. WISE2009, pp.227-234, October 2009.

[11] kizasi.jp : http://kizasi.jp/

ISBN: 978-988-19252-5-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)