International Journal of Pervasive Computing and Communications

Proposal of a system for visualizing temporal changes in impressions from tweets

Tadahiko Kumamoto Hitomi Wada Tomoya Suzuki

Proposal of a system for visualizing temporal changes in impressions from tweets

Tadahiko Kumamoto
Department of Information and Network Science, Chiba Institute of Technology, Narashino, Japan
Hitomi Wada
GI Technos Inc., Tokyo, Japan, and
Tomoya Suzuki
Inter Wave Ltd., Tokyo, Japan

Received 28 February 2015
Revised 28 February 2015
Accepted 5 April 2015

Abstract

Purpose – The purpose of this paper is to propose a Web application system for visualizing Twitter users based on temporal changes in the impressions received from the tweets posted by the users on Twitter.

Design/methodology/approach – The system collects a specified user's tweets posted during a specified period using Twitter API, rates each tweet based on three distinct impressions using an impression mining system, and then generates pie and line charts to visualize results of the previous processing using Google Chart API.

Findings – Because there are more news articles featuring somber topics than those featuring cheerful topics, the impression mining system, which uses impression lexicons created from a newspaper database, is considered to be more effective for analyzing negative tweets.

Research limitations/implications – The system uses Twitter API to collect tweets from Twitter. This suggests that the system cannot collect tweets of the users who maintain private timelines. According to our questionnaire, about 30 per cent of Twitter users' timelines are private. This is one of the limitations to using the system.

Originality/value – The system enables people to grasp the personality of Twitter users by visualizing the impressions received from tweets the users normally post on Twitter. The target impressions are limited to those represented by three bipolar scales of impressions: "Happy/Sad", "Glad/Angry" and "Peaceful/Strained". The system also enables people to grasp the context in which keywords are used by visualizing the impressions from tweets in which the keywords were found.

Keywords Twitter, Visualization, Sentiment analysis, Impression mining

Paper type Research paper

1. Introduction

The proliferation of mobile devices such as smartphones and tablet computers enables people to access the Internet anytime and anywhere. Consequently, a number of social networking services, such as Twitter and Facebook, are actively used. Twitter in

A portion of this work was supported by JSPS KAKENHI grant numbers 24500134 and 26330347, and donations from Mr Masaharu Fukuda.

International Journal of Pervasive Computing and Communications
Vol. 11 No. 2, 2015
pp. 193-211

1742-7371
DOI 10.1108/IJPCC-02-2015-0011

particular has a special feature that allows people to easily connect with celebrities, on-screen talent and people who can provide valuable information, as well as with friends and acquaintances. This is considered to be one of the reasons why Twitter is supported by several generations, regardless of age and gender.

On Twitter, many users post daily tweets on topics ranging from political and economic events to personal experiences. In their tweets, users frequently offer constructive suggestions, while a portion of the users sometimes enact mental abuse. Some users may post somber or lively tweets, while other users may tweet to express anger or joy. The impressions the majority of the people receive from reading tweets that an individual user normally posts can be inferred by carefully reading a large number of this user's posts, as tweets can reflect the personality of the Twitter user.

Therefore, this paper proposes a Web application system for visualizing Twitter users based on temporal changes in the impressions received from the tweets they post. When users input a Twitter user's account name and a period for analysis into our proposed system, the system will work as follows:

- The system collects the specified user's tweets posted during the specified period using the Twitter API (refer to https://dev.twitter.com/docs).
- The system rates each tweet based on three distinct impressions, using the impression mining system that Kumamoto *et al.* (2011) proposed and made available to the public. The target impressions are limited to those represented by three bipolar scales of impressions (Kumamoto, 2010): "Happy/Sad"; "Glad/Angry"; and "Peaceful/Strained".

The strength of each impression is computed as an "impression value", that is a real number between one and seven that denotes a position on the corresponding scale. For example, on the scale of "Happy/Sad", a score of one indicates "Happy", a middle score of four denotes "Neither happy nor sad" and a score of seven equals "Sad". If the impression value of a tweet is 2.5, then the average person will experience an intermediate impression between "Comparatively happy (2 points)" and "A little happy (3 points)" from reading the tweet.

The system generates pie and line charts to visualize the results of the previous processing using the Google Chart API (refer to https://developers.google.com/chart/). As a result, three pie charts are generated, each displaying an overall distribution of impression values from the tweets in the corresponding impression scale. A line chart with three polygonal lines is also generated, in which each line indicates the temporal changes in the impression values based on the tweet's posting date on the corresponding impression scale.

We also applied the aforementioned process to a retrieval result obtained through a search-by-keywords and added a function to our proposed system to visualize temporal changes in the impressions from the tweets that used the keywords. To summarize, our proposed system uses the Twitter API to collect the tweets in which the keywords were used, quantifies impressions from each tweet using the impression mining system and generates pie and line charts from the tweet's impression values using the Google Chart API.

The remainder of this paper is organized as follows. In Section 2, we present related work. In Section 3, we propose a system for visualizing impressions values computed from a specified user's tweets and their temporal changes, as well as for visualizing impression values computed from the tweets obtained from a search-by-keywords and their temporal changes. In Section 4, we reveal the results of the evaluation experiment we conducted in which 1,000 Twitter users used a nine-point scale to rate how well the pie charts generated by our proposed system represented impressions from the corresponding tweets. Finally, in Section 5, we conclude the paper and discuss future work.

2. Related work

There are several systems that have been developed to effectively use Twitter information, including a system that detects trends in the Twitter stream (Mathioudakis and Koudas, 2010), a system that recommends news articles based on Twitter-based user modeling (Abel *et al.*, 2013) and a system that detects earthquakes by monitoring tweets (Sakaki *et al.*, 2010).

There are also ongoing studies on the methods used to recommend users as candidates to follow, with the goal of facilitating connections between Twitter users. Weng *et al.* proposed a PageRank-like algorithm named TwitterRank to identify and recommend influential Twitter users (Weng *et al.*, 2010). Sadilek *et al.* (2012) proposed a system for suggesting users to follow by inferring real-world friendships. Pennacchiotti and Gurumurthy (2011) proposed a method that suggests users who have similar latent interests based on information extracted from their tweets. In Japan, a method for recommending users to follow was proposed based on the number of users' tweets registered as "Favorites" for a given topic (Watabe and Miyamori, 2012). On the official Twitter Web site, several users are suggested as "Who to follow" based on whom other users follow and additional criteria. In addition, users can easily follow popular users in a topic field by clicking on "Popular accounts" on the official Web site and selecting a topic that interests them. In contrast to these studies, our proposed system enables people to grasp Twitter users' personalities by examining the impressions from tweets the candidate users normally post on Twitter.

Research is ongoing on affective computing and sentiment analysis, such as studies on extracting subjective information, known as sentiments, emotions or impressions, from text data, such as reviews, news articles and Web pages. These results have been applied to various task domains, such as sentiment analysis (Pang and Lee, 2005), information visualization (Lin *et al.*, 2008) and annotation of impression tags (Kiyoki *et al.*, 1994). However, these studies only classify text data into emotion classes or attach impression tags to text data, without quantifying the impressions from the data.

Various visualization systems that use Twitter as an information source have been proposed. Many of these systems visualize social networks by representing following–follower relationships and users' behaviors such as posting times on Twitter, as well as DeepTwitter (Rotta *et al.*, 2013). Conversely, a system known as EmotionWatch (Kempter *et al.*, 2014) uses sentiment analysis techniques and visualizes Twitter users' affective reactions against a specified open event. Simply stated, EmotionWatch collects a set of tweets posted regarding an open event as input, computes a relative ratio for each of the 20 affective reaction types from the tweets, such as "Anger", "Happiness", "Relief", "Sadness", and "Regret", and visualizes their ratios for all of the affective reaction types using the "Emotion Wheel", a type of radar chart. In contrast to this system, our proposed system visualizes Twitter users' personality by examining the impressions received from the users' timelines.

3. Design and implementation of our visualization system

3.1 System flow

The flow of our proposed system is shown in Figure 1. First, system users are asked to specify the account name of a Twitter user or one or more keywords and a period for analysis. When a Twitter user is specified, the system uses the Twitter API to collect the tweets posted by the identified user during the specified period. When one or more keywords are specified, the system uses the Twitter API to collect the tweets in which the keywords were used during the specified period. Next, the system computes three impression values for each tweet using the impression mining system that Kumamoto *et al.* (2011) proposed and made available to the public, then stores the tweet and its impression values into a database along with the tweet's posting date. This impression mining system segments an input tweet into words and obtains values representing the effect of each word based on the impression lexicons that were automatically constructed from a given text database. The system then computes and outputs three impression values from the tweet based on these values. Finally, our proposed system uses the Google Chart API to generate three pie charts and a line chart with three polygonal lines from the impression values from all of the tweets.

As shown in Figure 1, please note that tweets from the users who have been registered in our proposed system are automatically collected at a consistent frequency and are stored in the database together with their impression values and posting dates.

3.2 User interface design

Our proposed system has been implemented as a Web application system and is designed to be used through a Web browser. Figure 2 illustrates a snapshot of the screen displayed when we specified *@MorinoKumazo* as a target user and a one-month period ending February 9, 2014 as the target period for analysis. Figure 3 illustrates a snapshot of the screen displayed

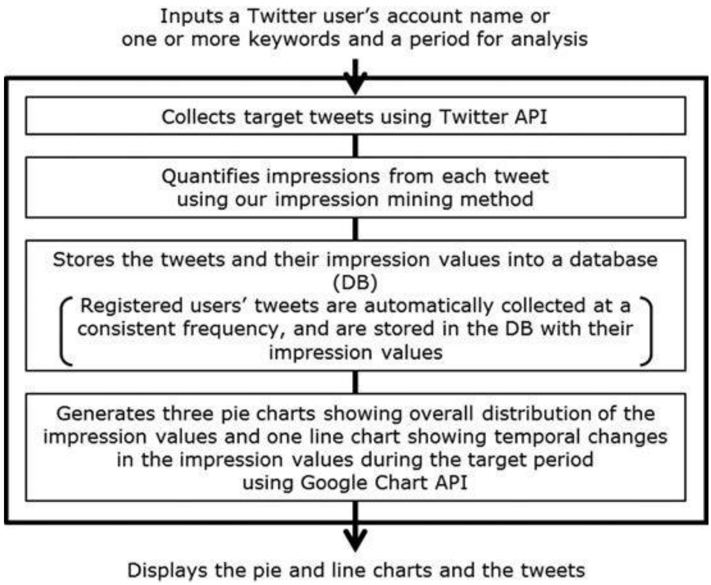

Figure 1.
Flow of our proposed visualization system

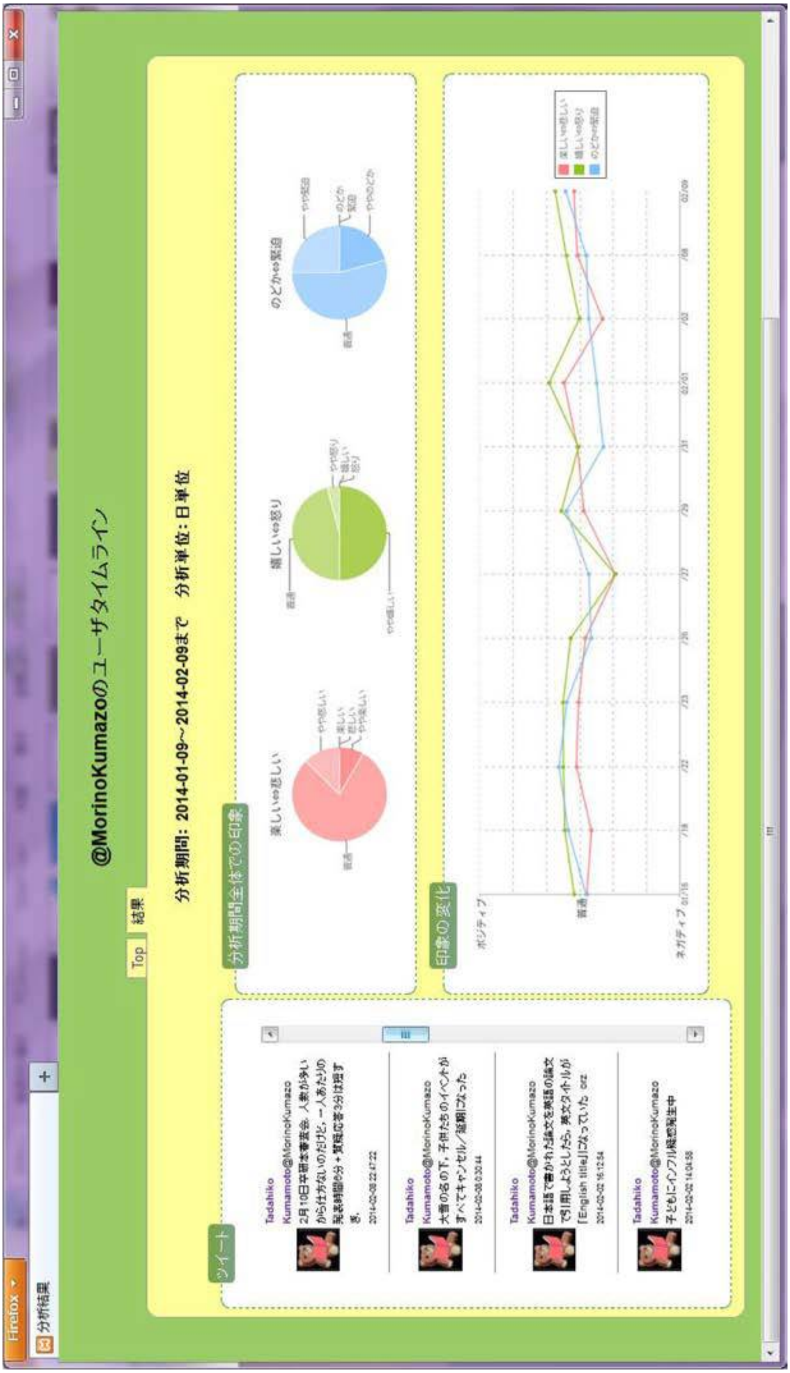

Figure 2.
A snapshot of the screen displayed when we specified *@MorinoKumazo* as a target user and a one-month period ending February 9, 2014 as the target period for analysis

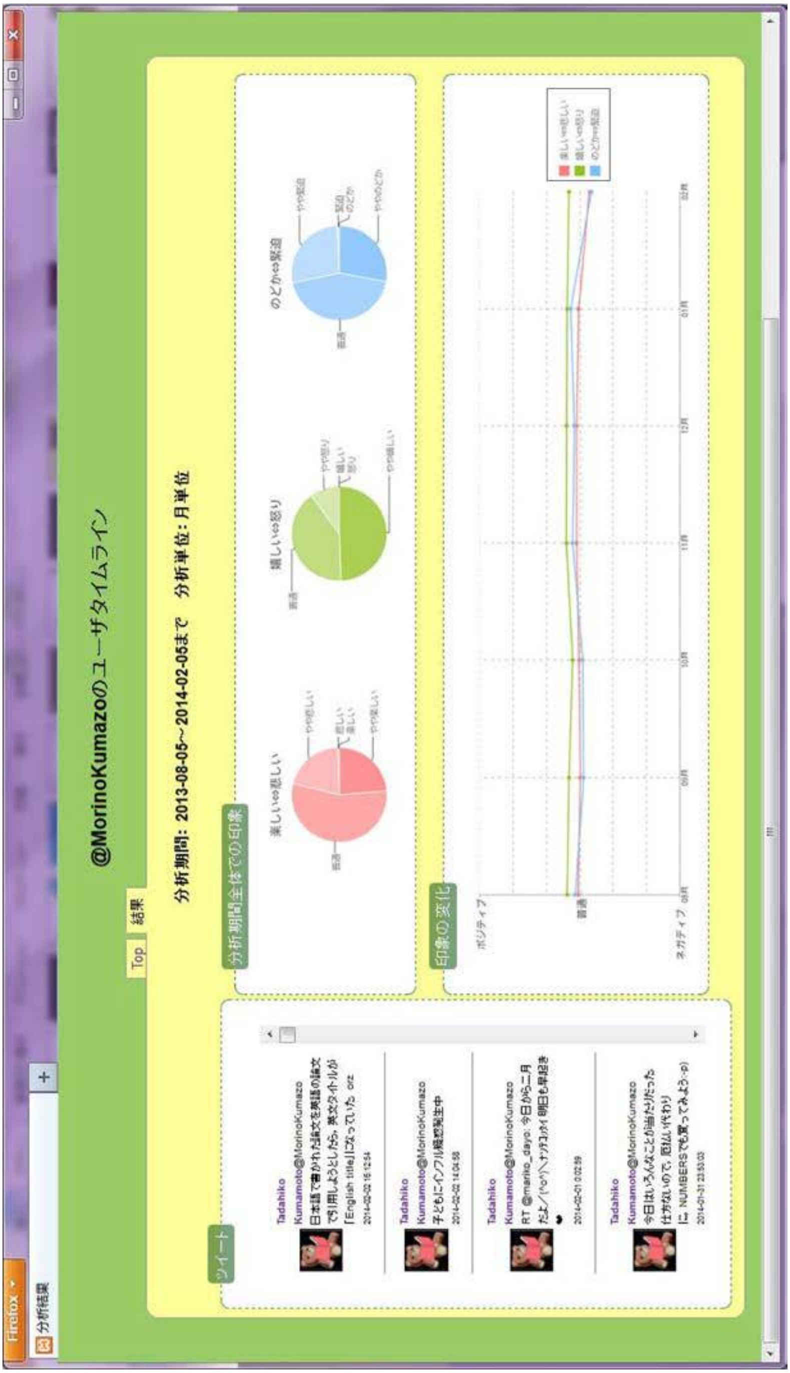

Figure 3.
A snapshot of the screen displayed when we specified *@MorinoKumazo* as a target user and a six-month period ending February 5, 2014 as the target period for analysis

when we specified *@MorinoKumazo* as a target user and a six-month period ending February 5, 2014 as the target period for analysis. Figure 4 illustrates a snapshot of the screen displayed when we specified the Japanese version of "Chiba Institute of Technology" as a keyword and a one-month period ending February 6, 2014 as the target period for analysis. In both Figures 2 and 3, the tweets that the specified user posted during the specified period are shown in reverse chronological order from the upper region of the screen, while in Figure 4, the tweets in which the keywords were used during the specified period are shown in reverse chronological order from the upper region of the screen. On each screen, three pie charts are displayed at the upper right section, each showing an overall distribution of impression values computed from the target tweets in each impression scale. Pie charts for impression values in the "Happy/Sad" scale, the "Glad/Angry" scale and the "Peaceful/Strained" scale are shown on the left, middle and right, respectively. A line chart is displayed at the lower right section of the screen. In the line chart, three polygonal lines are drawn, each line indicating the temporal changes in the impression values of each impression scale. The red line, the green line and the blue line, respectively, correspond with the scales "Happy/Sad", "Glad/Angry" and "Peaceful/Strained". The horizontal axis represents the date in which one or more tweets were collected. Dates for which no tweets were collected are simply skipped. The time span options for the target period for analysis are either "one month" or "six months". System users were asked to choose the length of the analysis period as well as specify an end date. The vertical axis corresponds with impression values. The points in the upper half of the line chart suggest that impressions from the tweets posted on the corresponding date are on average positive, while points in the lower half suggest that impressions from the tweets are on average negative.

3.3 Impressions to be visualized

Kumamoto (2010) designed six bipolar scales suitable for representing readers' impressions from news articles:

(1) Happy/Sad;

(2) Glad/Angry;

(3) Interesting/Uninteresting;

(4) Optimistic/Pessimistic;

(5) Peaceful/Strained; and

(6) Surprising/Common.

First, he conducted nine experiments. In each experiment, 100 subjects read ten news articles and estimated their impressions on a scale of 1 to 5 for each of 42 impression words. These 42 impression words were manually selected from a Japanese thesaurus (Ohno and Hamanishi, 1986) to represent words that expressed impressions from news articles. Next, a factor analysis was applied to the data obtained from the experiments, and the 42 words were divided into four groups:

(1) negative words;

(2) positive words;

(3) two words that were "uninteresting" and "common"; and

(4) two words that were "surprising" and "unexpected".

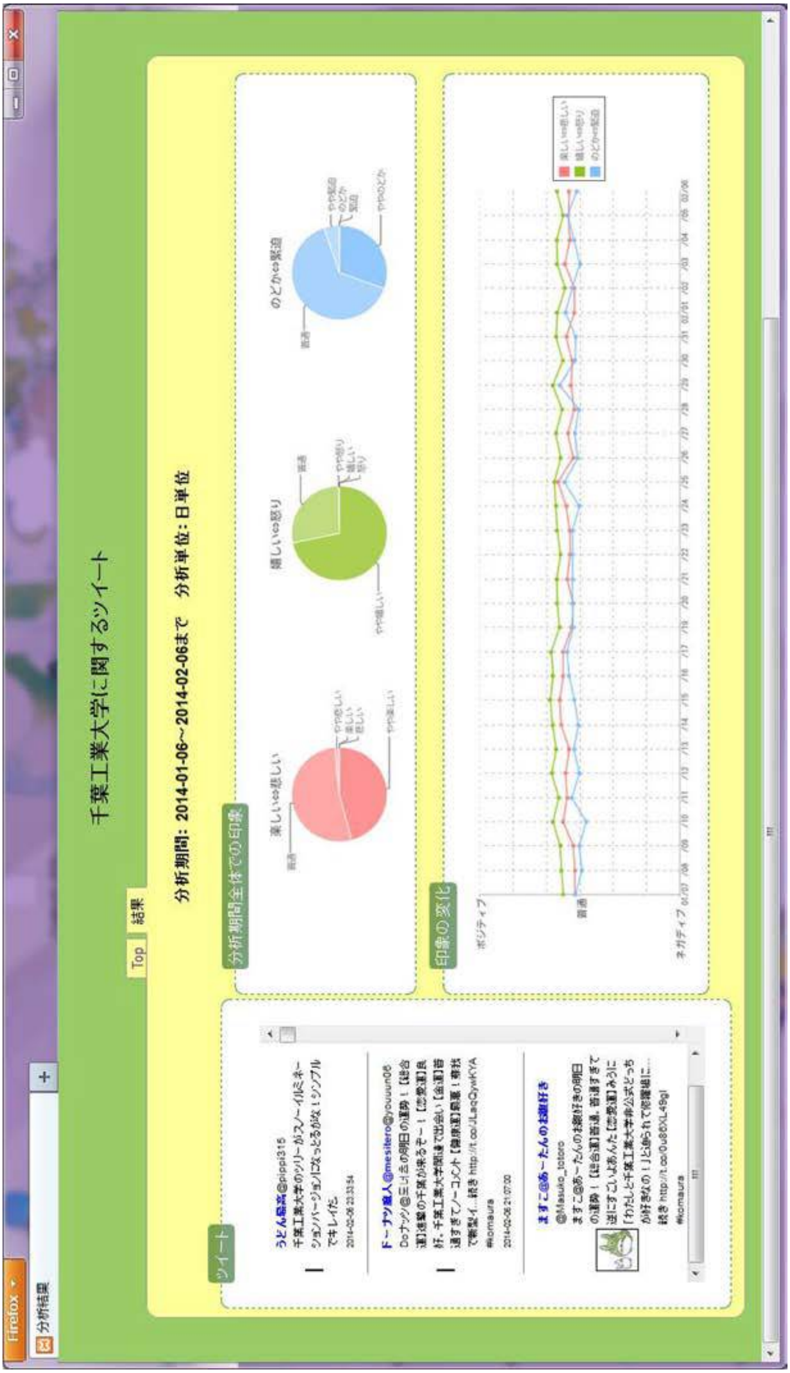

Figure 4.
A snapshot of the screen displayed when we specified the Japanese version of "Chiba Institute of Technology" as a keyword and a one-month period ending February 6, 2014 as the target period for analysis

Meanwhile, after a cluster analysis of the same data, the 42 words were divided into ten groups. The results of the two analyses were used to create the six bipolar scales previously mentioned. Kumamoto demonstrated that impressions on the "Surprising/Common" scale differed greatly among individuals in terms of their perspective. He also demonstrated that processing according to the individuals' background knowledge, interests and characters was required to categorize the impressions represented by the scales of "Interesting/Uninteresting" and "Optimistic/Pessimistic". As a result, he decided not to use these three scales at the present time and adopted the remaining three scales:

(1) Happy/Sad;

(2) Glad/Angry; and

(3) Peaceful/Strained.

Therefore, we also use only these three scales in this study.

3.4 Collecting target tweets

System users are asked to input the account name of a Twitter user or one or more keywords and a period for analysis. When an account name is input, the system collects tweets posted by the indicated user during the specified period using the Twitter API prepared for obtaining an arbitrary user's timeline. When one or more keywords are input, the system uses the Twitter API to collect the tweets that were obtained as a result of a search-by-keywords during the specified period.

3.5 Quantifying impressions from tweets

An impression lexicon plays an important role in quantifying impressions from tweets. Our impression lexicons were automatically constructed from our given text database through the impression mining system.

Two contrasting sets, each consisting of multiple reference words, are used to construct an impression lexicon for each scale. Therefore, we designated the reference word set that expresses an impression found on the left of the scale as S_L and designated the reference word set that expresses an impression found on the right of the scale as S_R. Articles including one or more reference words in S_L or S_R are extracted from a given text database, and the number of reference words belonging to each set is counted in each article. For this database, we used the 2002 to 2006 editions of the Yomiuri newspaper text database (Refer to www.nichigai.co.jp/sales/corpus.html). Next, we classify the articles in which the number of reference words belonging to S_L is larger than the number of reference words belonging to S_R as A_L and designate the number of articles in A_L as N_L. The articles in each of which the number of reference words belonging to S_L is smaller than the number of reference words belonging to S_R becomes A_R and number of articles in A_R becomes N_R. Next, all words are extracted from A_L and A_R, and the document frequency of each word is measured. Please note that we excluded particles, adnominal words and demonstratives in which the part of speech known as "adnominal word" exists only in Japanese, not in English. For example, "that", "so called" and "of no particular distinction" are considered adnominal words in Japanese. Next, the document frequency in A_L of a word w is designated as $N_L(w)$, and the document frequency in A_R of a word w is designated as $N_R(w)$. The revised conditional probabilities of a word designated as w are defined as follows:

$$P_L(w) = N_L(w)/N_L$$

$$P_R(w) = N_R(w)/N_R$$

In these equations, only articles that satisfy the assumptions described above were used to calculate $P_L(w)$ and $P_R(w)$.

Finally, the impression value $v(w)$ of a word designated as w is calculated using $P_L(w)$ and $P_R(w)$ as follows:

$$v(w) = P_L(w)\, W_L/(P_L(w)\, W_L + P_R(w)\, W_R)$$

Where:

$W_L = log_{10} N_L$; and
$W_R = log_{10} N_R$.

In other words, in an impression lexicon, a weighted interior division ratio $v(w)$ of $P_L(w)$ and $P_R(w)$ is calculated using these formulas and stored as an impression value of w in the scale "S_L/S_R". It is important to note that W_L and W_R denote weights. Therefore, the larger N_L and N_R are, the heavier the W_L and W_R will be.

The numbers of entries in the impression lexicons constructed above are shown in Table I, along with the obtained values of W_L and W_R. Further, the two contrasting sets of reference words used to create the impression lexicons for each scale are listed in Table II. We determined these words after some trial and error, as based on two criteria:

(1) the word is a verb or adjective that expressed either of two contrasting impressions represented by the scale; and
(2) as much as possible, the word did not suggest other types of impressions.

For each scale, the impression value of a tweet is calculated using the impression mining system. First, the tweet is segmented into words using "Juman" (Kurohashi *et al.*, 1994),

Table I. Specifications of the impression lexicons that we created from the Yomiuri newspaper text database

Scales	No. of entries	W_L	W_R
Happy/sad	387,428	4.90	3.80
Glad/angry	350,388	4.76	3.82
Peaceful/strained	324,590	3.91	4.67

Table II. Reference words prepared for each scale, which were translated into English by the authors

Scales	Reference words
Happy	*tanoshii* (happy), *tanoshimu* (enjoy), *tanosimida* (look forward to), *tanoshigeda* (joyous)
Sad	*kanashii* (sad), *kanashimu* (experience sadness), *kanashimida* (feel sad), *kanashigeda* (look sad)
Glad	*ureshii* (glad), *yorokobashii* (blessed), *yorokobu* (feel delight)
Angry	*ikaru/okoru* (get angry), *ikidooru* (become irate), *gekido-suru* (become enraged)
Peaceful	*nodokada* (peaceful), *nagoyakada* (friendly), *sobokuda* (simple), *anshinda* (feel easy)
Strained	*kinpaku-suru* (strained), *bukimida* (scared), *fuanda* (be anxious), *osoreru* (fear)

one of the most powerful Japanese morphological analysis systems. Because there are no boundary markers between words in Japanese, word segmentation is needed to identify individual words. An impression value for each word is obtained by consulting the impression lexicon constructed for the scale. Finally, an average x_{ave} of the impression values obtained for all of the words except for particles, adnominal words and demonstratives is calculated and presented as the impression value of the tweet. However, some errors are likely to occur between impression values computed in an unsupervised manner and those provided by human readers. The impression mining system corrects impression values computed by the system itself using a regression equation designated for each scale. Consequently, errors in these values are reduced. Note that the regression equations were designed based on the results of experiments with a total of 900 subjects which were conducted to identify the errors that actually occurred. The regression equations are shown in Table III, where x_{ave}, or a value between 0.0 and 1.0, is converted into a value between 1.0 and 7.0 using the following formula, and is represented as x in the regression equations:

$$x = 6\,(1 - x_{ave}) + 1$$

The impression mining system described here is applied to tweets, with three impression values computed from each tweet.

3.6 Generating pie and line charts

Three pie charts are generated using the Google Chart API, each displaying an overall distribution of impression values from each impression scale. First, all of the tweets collected in 3.4 are classified into five classes based on their impression values from each impression scale. For example, in the "Happy/Sad" scale, the tweets whose impression values are less than 2.5 are classified into the "Happy" class; the tweets whose impression values are 2.5 or more and less than 3.5 are classified into the "A Little Happy" class; the tweets with impression values of 3.5 or more and less than 4.5 are classified into the "Normal" class; the tweets with impression values of 4.5 or more and less than 5.5 are classified into the "A Little Sad" class; and the tweets with impression values of 5.5 or more are classified into the "Sad" class. For each impression scale, our proposed system counts the number of tweets classified into each class and generates a pie chart that illustrates the percentage of tweets among the classes.

A line chart with three polygonal lines is generated using the Google Chart API, and the temporal change in the impression values of each impression scale is represented by a polygonal line. The time span for the target period of analysis is either "one month" or "six months", as determined by the system users. When a "one month" period is selected, the impression value average computed from the tweets is calculated daily. When a "six

Scales	Regression equations (x: converted values)
Happy/sad	$-1.6355586x^3 + 18.971570x^2 - 70.68575x + 88.5147$
Glad/angry	$2.384741939x^5 - 46.87159982x^4 + 363.6602058x^3 - 1391.589442x^2 + 2627.06261x - 1955.3058$
Peaceful/strained	$-1.7138394x^3 + 21.942197x^2 - 90.79203x + 124.8218$

Table III. Regression equations designed for correcting impression values computed from tweets month" period is selected, the impression value average computed from the tweets is calculated monthly. The system users are also asked to specify an end date for the target period.

4. Evaluation

We conducted an evaluation experiment in which 1,000 Twitter users who generally viewed tweets once a week or more participated. The subjects were asked to compare the impressions that they received from reading our specified user's tweets with the three pie charts that were generated based on impression values computed from the tweets. They rated the tweets on a nine-point scale from "Very good" to "Very bad" based on each pie chart's overall distribution of the impression values from the corresponding impression scale.

4.1 Preliminary questionnaire

First, we conducted a preliminary questionnaire to identify the Twitter users who generally viewed tweets on Twitter once a week or more. The target respondents for this questionnaire were 50,000 Internet users over the age of 16.

Items for the preliminary questionnaire are listed in Table IV. In Q1, we asked potential subjects whether they had a Twitter account. They were given three options:

(1) Yes;

(2) No; and

(3) I am unfamiliar with Twitter.

Only the respondents who chose "Yes" progressed to *Q2* and *Q3*, while the remaining respondents were dismissed from the questionnaire. In *Q2* and *Q3*, we asked respondents about their viewing and posting frequency on Twitter, respectively. The options respondents were given for these two questions were the same:

- four or more times a day;
- two or three times a day;
- once a day;
- four or five times a week;
- two or three times a week;
- once a week;
- every once in a while; and
- almost zero.

In this preliminary questionnaire, a total of 50,000 Internet users responded to *Q1*. The results for this question, as shown in Table V, indicated that 15,880 (about 31.8 per cent)

Table IV. Items for preliminary questionnaire

Item no.	Questions
Q1	Do you have an account on Twitter?
Q2	How often do you usually view tweets on Twitter?
Q3	How often do you usually post tweets?

of the 50,000 respondents have a Twitter account. These 15,880 respondents answered *Q2* (question on viewing frequency) and *Q3* (question on posting frequency). The results for these questions were compiled and shown in Table VI, from which we obtained the following findings. Of the 15,880 respondents, 11,583 (about 72.9 per cent) reported that they actually viewed tweets on Twitter, while 4,297 of the respondents rarely viewed tweets, and 5,870 (about 37.0 per cent) of the respondents viewed tweets daily. Meanwhile, 7,573 (about 47.7 per cent) of the 15,880 respondents actually posted one or more tweets on Twitter, while 8,307 of the respondents rarely posted tweets, and 2,540 (about 16.0 per cent) of the respondents posted tweets daily. Overall, the posting frequency was lower than the viewing frequency. Because the impression mining system quantifies the impressions which people feel from reading tweets, 8,481 (about 53.4 per cent) who generally viewed tweets on Twitter once a week or more were extracted as candidate subjects for the following evaluation experiment.

4.2 Evaluation experiment

We conducted a Web-based evaluation experiment. We invited 8,481 people from the preliminary questionnaire, who answered that they viewed tweets on Twitter once a week or more to join the evaluation experiment, with considerations given to their ages and genders. Of those invited, 1,000 subjects participated in the experiment. The ages and genders of these subjects who participated are shown in Table VII.

First, the subjects were asked to input into our proposed system the account name *@ariyoshihiroiki*, which is the account of the famous Japanese comedian Hiroiki Ariyoshi. Next, the subjects viewed a screen in which they compared the impressions that they felt from reading Mr Ariyoshi's tweets with the three pie charts displayed on the screen and rated the extent to which each pie chart displayed an overall distribution of the impression values from each impression scale. The nine options prepared for this rating were:

(1) Very Good (9 points);
(2) Almost Good (8 points);
(3) Pretty Good (7 points);
(4) Somewhat Good (6 points);
(5) Neither Good Nor Bad (5 points);
(6) Somewhat Bad (4 points);
(7) Pretty Bad (3 points);
(8) Almost Bad (2 points); and
(9) Very Bad (1 point).

Options	No. of respondents	(%)
Yes	15,880	31.8
No	32,158	64.3
I am unfamiliar with Twitter	1,962	3.9
Total	50,000	100

Table V. Possession of Twitter accounts (*Q1*)

Table VI.
Number of respondents in each combination of viewing and posting frequency (*Q2*, *Q3*)

	Posting frequency								
	No. of times a day			No. of times a week					
Viewing frequency	Four or more	Two or three	Once	Four or five	Two or three	Once	Every once in a while	Almost zero	Total
Four or more times a day	1,001	358	207	207	244	84	443	454	2,998
Two or three times a day	56	310	145	118	184	65	330	321	1,529
Once a day	19	64	267	88	112	80	300	413	1,343
Four or five times a week	5	8	20	130	90	48	184	195	680
Two or three times a week	2	6	10	17	278	87	382	423	1,205
Once a week	1	4	4	5	20	180	222	290	726
Every once in a while	10	4	14	9	20	25	987	2,033	3,102
Almost zero	9	4	12	1	9	8	76	4,178	4,297
Total	1,103	758	679	575	957	577	2,924	8,307	15,880

We considered a score between 6 points and 9 points as a positive score, a score of 5 points as a medium score and a score between 1 point and 4 points as a negative score. We then calculated the percentages of the positive, medium and negative scores. The results are shown in Table VIII.

Next, the subjects were asked to input into our proposed system the account name *@yoichiomar*, which is the account of the famous Japanese war photographer Yoichi Watanabe. They were asked to rate the resulting three pie charts displayed in the same manner as they did for the previous evaluation of Mr Ariyoshi's tweets. We calculated the percentages of positive, medium and negative scores for this rating. The results are shown in Table IX.

Finally, the subjects were asked to input the account name of either themselves or one of their following users into our proposed system. They were asked to rate the resulting three pie charts displayed in the same manner as they did for the evaluations of Mr Ariyoshi's and Mr Watanabe's tweets. We calculated the percentages of positive, medium and negative scores for this rating. The results are shown in Table X.

Please note that in Tables VIII-X, the option "Others" indicates that the respondents could not rate the pie charts for some reason.

4.3 Results and consideration

From Table VIII, we can observe that 38.7 to 40.2 per cent of the 1,000 subjects gave a positive score, while just 15.6 to 15.8 per cent of them gave a negative score. From Table IX, we can observe that 42.1 to 43.0 per cent of the 1,000 subjects gave a positive

Age	Men	Women	Total
20-29	108	108	216
30-39	108	108	216
40-49	109	109	218
50-59	108	100	208
60	108	34	142
Total	541	459	1,000

Table VII. Age and gender of subjects who participated in our evaluation experiment

Scores	Happy/sad		Glad/angry		Peaceful/strained	
9 pts	31	40.2%	36	38.7%	32	39.1%
8 pts	66		53		54	
7 pts	124		119		131	
6 pts	181		179		174	
5 pts	391	39.1%	402	40.2%	403	40.3%
4 pts	77	15.8%	72	15.6%	74	15.6%
3 pts	37		28		32	
2 pts	20		27		20	
1 pt	24		29		30	
Others	49	4.9%	55	5.5%	50	5.0%
Total	1,000	100%	1,000	100%	1,000	100%

Source: Mr Hiroiki Ariyoshi's tweets

Table VIII. Results of a comparative evaluation between pie charts and tweets

Scores	Happy/sad		Glad/angry		Peaceful/strained	
9 pts	45	42.2%	37	42.1%	45	43.0%
8 pts	64		72		66	
7 pts	143		131		146	
6 pts	170		181		173	
5 pts	399	39.9%	387	38.7%	376	37.6%
4 pts	62	12.4%	77	13.9%	80	14.0%
3 pts	26		23		22	
2 pts	13		16		15	
1 pt	23		23		23	
Others	55	5.5%	53	5.3%	54	5.4%
Total	1,000	100%	1,000	100%	1,000	100%

Source: Mr Yoichi Watanabe's tweets

Table IX. Results of a comparative evaluation between pie charts and tweets

Scores	Happy/sad		Glad/angry		Peaceful/strained	
9 pts	29	36.2%	28	36.8%	27	35.1%
8 pts	42		46		46	
7 pts	114		115		113	
6 pts	177		179		165	
5 pts	378	37.8%	380	38.0%	406	40.6%
4 pts	95	18.3%	87	17.5%	77	16.6%
3 pts	40		38		39	
2 pts	16		20		21	
1 pt	32		30		29	
Others	77	7.7%	77	7.7%	77	7.7%
Total	1,000	100%	1,000	100%	1,000	100%

Source: Tweets of either subjects or one of their following users

Table X. Results of a comparative evaluation between pie charts and tweets

score, while just 12.4 to 14.0 per cent of them gave a negative score. From Table X, we can see that 35.1 to 36.8 per cent of the 1,000 subjects gave a positive score, while just 16.6 to 18.3 per cent of them gave a negative score. These results suggest that pie charts are effective for representing an overall distribution of impression values from tweets based on the three impression scales.

Moreover, in comparing the results of Tables VIII-X, the pie charts displaying impressions from Mr Watanabe's tweets were evaluated higher than other pie charts, and the pie charts displaying impressions from tweets of the subjects themselves or one of their following users were evaluated lower than the other pie charts. Mr Watanabe frequently posts negative tweets on war, poverty and similar topics, while Mr Ariyoshi often posts positive tweets on cheerful topics. Because there are more news articles featuring somber topics than those featuring cheerful topics, the impression mining system, which uses an impression lexicon created from a newspaper database, is considered to be more effective at analyzing negative tweets. In addition, we considered that tweets from the subjects themselves or one of their following users were evaluated more rigorously because the subjects either understood or were able to accurately guess the background of the target tweets, at least compared to the tweets posted by Mr Ariyoshi and Mr Watanabe.

4.4 Limitations on the use of our proposed system

We administered another questionnaire to grasp the limitations of using our proposed system. Target respondents to this questionnaire were 10,000 Internet users ranging in age from 20 to > 60 years.

In the first question, we asked the users whether they had a Twitter account. To this question, 3,321 of the users answered "Yes". Further, we asked these 3,321 users whether their timeline was open to the public. The results obtained are shown in Table XI.

Our proposed system uses the Twitter API to collect tweets from Twitter. This suggests that the system cannot collect tweets of the users who maintain private timelines. According to Table XI, 31.3 per cent of the 3,321 Twitter users' timelines were private; thus, the ratio is not low. This condition is considered to be one of the limitations to using our proposed system.

5. Conclusion

In this paper, we proposed a Web application system for visualizing Twitter users based on temporal changes in impressions from the tweets the users posted on Twitter. When system users input a Twitter user's account name and a period for analysis into the system, the system examines and visualizes the impressions from tweets the designated user posted during the specified period. The system enables people to grasp the personality of Twitter users by visualizing the impressions received from tweets the users normally post on Twitter. The target impressions are limited to those represented by three bipolar scales of impressions:

(1) Happy/Sad;

(2) Glad/Angry; and

(3) Peaceful/Strained.

We also applied this concept to a retrieval result obtained through a search-by-keywords, and added a function to our proposed system to visualize temporal changes in the impressions from the tweets that included the keywords. This system enables people to grasp the context in which keywords are used by visualizing the impressions from tweets in which the keywords were found.

Our future work is as follows. Because the impression mining system we used in our proposed system was designed for quantifying impressions from news articles

Options	No. of respondents	
My timeline is now open, but I had locked and closed my timeline to the public	284	8.6%
My timeline has been open from the beginning	1,885	56.8%
My timeline is now closed to the public	1,041	31.3%
Others	111	3.3%
Total	3,321	100%

Table XI. Is your timeline open to the public?

(Kumamoto *et al.*, 2011), the method's effectiveness with tweets has not been sufficiently verified. There are many grammatically incorrect sentences, short sentences consisting of one or two words and Twitter-dependent expressions such as emoticons and Internet slang words that are observed in tweets. Therefore, we recognize that the current lexicon-based approach to impression mining is not suitable for such tweets, and thus, we are currently designing and developing an impression mining method for social media. Impression scales should also be redesigned according to the impressions to be extracted from tweets.

References

Abel, F., Gao, Q., Houben, G.J. and Tao, K. (2013), "Twitter-based user modeling for news recommendations", *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*, *Beijing*, pp. 2962-2966.

Kempter, R., Sintsova, V., Musat, C. and Pu, P. (2014), "EmotionWatch: visualizing fine-grained emotions in event-related tweets", *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media*, *Michigan, MI*, pp. 236-245.

Kiyoki, Y., Kitagawa, T. and Hayama, T. (1994), "A metadatabase system for semantic image search by a mathematical model of meaning", *ACM SIGMOD Record*, Vol. 23 No. 4, pp. 34-41.

Kumamoto, T. (2010), "Design of impression scales for assessing impressions of news articles", *Proceedings of the 15th International Conference on Database Systems for Advanced Applications*, *Tsukuba*, pp. 285-295.

Kumamoto, T., Kawai, Y. and Tanaka, K. (2011), "Improving a method for quantifying readers' impressions of news articles with a regression equation", *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, *Portland, OR*, pp. 87-95.

Kurohashi, S., Nakamura, T., Matsumoto, Y. and Nagao, M. (1994), "Improvements of Japanese morphological analyzer JUMAN", *Proceedings of the International Workshop on Sharable Natural Language Resources*, *Nara*, pp. 22-28.

Lin, H.Y.K., Yang, C. and Chen, H.H. (2008), "Emotion classification of online news articles from the reader's perspective", *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, *Sydney*, pp. 220-226.

Mathioudakis, M. and Koudas, N. (2010), "TwitterMonitor: trend detection over the twitter stream", *Proceedings of the International Conference on Management of Data*, *Indianapolis, IN*, pp. 1155-1158.

Ohno, S. and Hamanishi, M. (Eds) (1986), *Ruigo-Kokugo-Jiten*, Kadokawa Shoten Publishing, Tokyo.

Pang, B. and Lee, L. (2005), "Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales", *Proceedings of the Annual Meeting on Association for Computational Linguistics*, *Morristown*, pp. 115-124.

Pennacchiotti, M. and Gurumurthy, S. (2011), "Investigating topic models for social media user recommendation", *Proceedings of the 20th International Conference Companion on World Wide Web*, *Hyderabad*, pp. 101-102.

Rotta, G., Lemos, V., Lammel, F., Manssour, I., Silveira, M. and Pase, A. (2013), "Visualization techniques for the analysis of twitter users' behavior", *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media*, *Boston, MA*, pp. 748-749.

Sadilek, A., Kautz, H. and Bigham, J.P. (2012), "Finding your friends and following them to where you are", *Proceedings of the Fifth International Conference on Web Search and Data Mining, Seattle, WA*, pp. 723-732.

Sakaki, T., Okazaki, M. and Matsuo, Y. (2010), "Earthquake shakes twitter users: real-time event detection by social sensors", *Proceedings of the 19th International Conference on World Wide Web, Raleigh, NC*, pp. 851-860.

Watabe, S. and Miyamori, H. (2012), "Twitter user recommender: the user recommendation system using the favorite function of twitter", *Proceedings of the Forum on Data Engineering and Information Management, Kobe*.

Weng, J., Lim, E.P., Jiang, J. and He, Q. (2010), "TwitterRank, finding topic-sensitive influential twitterers", *Proceedings of the Third International Conference on Web Search and Data Mining, New York, NY*, pp. 261-270.

About the authors

Tadahiko Kumamoto was born in Nagasaki, Japan. He received his BE, ME, and PhD in Engineering from the University of Tsukuba in 1988, 1990 and 1996, respectively. Since 1990, he has been affiliated to the National Institute of Information and Communications Technology. In 2007, he went to work at the Chiba Institute of Technology (CIT) as an Associate Professor. Since 2010, he has been a professor at the institute. His interests of research include impression mining and its application. Tadahiko Kumamoto is the corresponding author and can be contacted at: kumamoto@net.it-chiba.ac.jp

Hitomi Wada received her BSc (IT) degree from CIT in 2014. Now she works for an IT-related company.

Tomoya Suzuki received his BSc (IT) degree from CIT in 2014. Now he works for Inter Wave Ltd., an IT-related company.