Proceedings of the 44th Hawaii International Conference on System Sciences - 2011

Sentiment Bias Detection in Support of News Credibility Judgment

Jianwei Zhang
Kyoto Sangyo University
zjw@cc.kyoto-su.ac.jp

Yukiko Kawai
Kyoto Sangyo University
kawai@cc.kyoto-su.ac.jp

Shinsuke Nakajima
Kyoto Sangyo University
nakajima@cse.kyoto-su.ac.jp

Yoshifumi Matsumoto
Kyoto Sangyo University
g738444@cc.kyoto-su.ac.jp

Katsumi Tanaka
Kyoto University
tanaka@dl.kuis.kyoto-u.ac.jp

Abstract

Recently, an increasing number of online news websites have come to provide news browsing and retrieval services. For certain topics, certain news websites may hold sentiment bias, and therefore select and edit information according to their own standpoints before delivering news articles. Lacking conscious awareness of websites' sentiment bias may result in blind obedience to the reported information. We focus on the sentiment aspect of news articles and develop a system which can detect and visualize sentiment tendencies of different websites. Given a topic, the system extracts relevant subtopics and presents sentiment difference between different subtopics. Once a subtopic is specified, sentiment difference between news websites is also provided. The background knowledge of sentiment difference between subtopics and between websites can assist users in judging the news credibility. In particular, the system analyzes four-dimension sentiment, which is more similar to human emotion than conventional positive-negative sentiment. Experimental evaluations show the accuracy of sentiment extraction and subtopic extraction is good, and our observation results show sentiment bias can be detected by the system.

1. Introduction

Reading newspapers is a major means of acquiring information in our daily lives. Recently, with the widespread of online news websites, users can browse and retrieve news more easily. In general, news websites' reports are expected to have a fair and impartial coverage of facts. However, in some cases, websites may have sentiment tendencies dissimilar from others according to their own standpoints or interests. In this paper, we call a website's particular sentiment tendency "sentiment bias". Generally speaking, a website's sentiment bias is not manifested, but may be embodied in picking up different subtopics for a topic, or reporting a same event in different tones. For example, for a topic "Obama", a website which supports him may only report his positive subtopics, such as "Nobel laureate", while a website which holds opposite impression on him may inimically write articles about his negative subtopics, such as "land deal with fundraiser". Even for a same subtopic related to "Obama", such as "BP oil spill", USA media and UK media may report this accident with different attitudes. Biased reports convey one-sided information, and may consequently lead to wrong opinions or wrong judgments. Producing the background knowledge of news websites' tendencies is necessary for assisting users in determining whether news articles are fair and credible.

We focus on the sentiment aspect of news articles, and develop a system which can detect and visualize websites' sentiment tendencies with respect to a given topic. More concretely, given a topic, the system can extract relevant subtopics and present sentiment difference between different subtopics. In general, a wide topic relates to multiple subtopics and news articles on different subtopics may have different sentiment. For example, for a topic "Yen appreciation", Japanese newspapers write some articles on its relevant subtopic "export trade" with negative sentiment, while they also provide news articles with positive sentiment on another relevant subtopic "travel industry". Extracting subtopics for a topic and grouping news articles into corresponding subtopics can help analyze news sentiment more meticulously. For a certain subtopic, the system can also provide sentiment difference between different websites. If users only browse news articles in a sentiment biased website, they may believe its reports and follow its viewpoint indiscriminately. By the system, users can obtain an overall perspective of the various sentiment with respect to different subtopics or held by different websites. When a website's sentiment tendency about a subtopic is obviously different from others, the sentiment bias is visually presented to users. Next, users can read news articles with the sentiment bias in mind, or browse articles from other websites which have different sentiment tendencies. In this way, users can acquire well-balanced information and consequently judge news credibility fairly. An idea for evaluating news credibility is that "information, which news websites with different sentiment tendencies reach agreement on, may be credible". For example, if both news websites usually with bright sentiment on the topic "economy" and news websites usually with dark sentiment on "economy" come to provide news articles referring to economic comeback, the information has high credibility.

We detect websites' sentiment bias based on sentiment analysis. In particular, unlike conventional positive-negative analysis of sentiment, we define more detailed sentiment. News about different topics have various kinds of sentiment, not restricted to positive-negative sentiment. For example, reports about "sports games" are often written with "happy" or "sad" sentiment, while articles about "Iraq war" mainly reflect the sentiment of "acceptance" or "disgust". Moreover, there exist news topics containing both positive and negative sentiment. For example, a website has an "expectation" (positive sentiment) about "economic comeback", but thinks the present situation is "sad" (negative sentiment). A psychologist Plutchik proposed an emotion model [1], which consists of four pairs of basic elements of human emotion: "Joy ⇔ Sadness", "Acceptance ⇔ Disgust", "Anticipation ⇔ Surprise", and "Fear ⇔ Anger". We adopt this model to our system due to its relative polarity and diversity of emotion.

Our system enables the following:

- Four-dimension sentiment of news articles beyond positive-negative analysis is extracted.
- Given a topic, sentiment tendencies are summarized for different subtopics.
- When a subtopic is specified, sentiment tendencies are summarized for different websites.
- Sentiment difference is visualized in an intuitive interface.
- Users' judgment on news credibility is supported by sentiment bias detection.

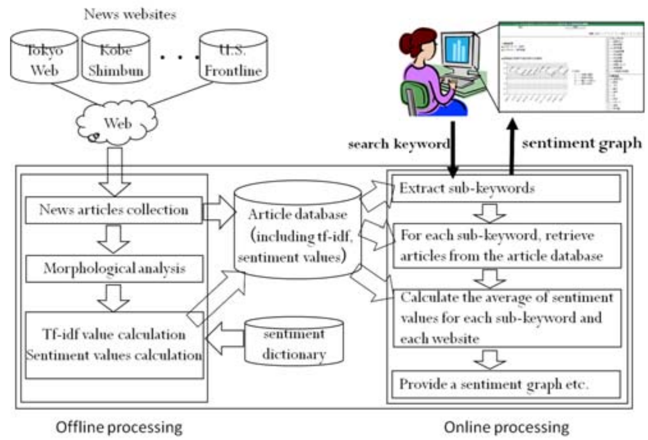

Figure 1: System overview.

The rest of this paper is structured as follows. Section 2 presents an overview of the system. Section 3 describes the offline processing of the system, and Section 4 describes the online processing. Section 5 evaluates the accuracy of sentiment extraction and subtopic extraction, and reports the observation results on sentiment difference. Section 6 reviews related research. Section 7 concludes the paper and describes future work.

2. System overview

Figure 1 shows the system overview . This system consists of two parts: offline processing (left-hand side) and online processing (right-hand side).

Collection and analysis of news articles are preprocessed offline, before online retrieval from a user. First, news articles from several pre-specified news websites are crawled. Then, a morphological analysis of the collected articles is conducted to extract the words with specific parts of speech, and their $tf \cdot idf$ values are calculated. We construct a sentiment dictionary in which the entries include sentiment values of words. A sentiment vector of four dimensions is attached to each news article by looking up the sentiment values of words appearing in the article from the sentiment dictionary and averaging these values. The collected news articles, the $tf \cdot idf$ values of the extracted words, and the sentiment vectors of news articles are stored in a database.

When a user enters a search keyword representing a news topic, the system first extracts sub-keywords representing its subtopics. The system then retrieves relevant news articles from the article database for each sub-keyword. The articles are grouped by news websites and ranked in each website based on the $tf \cdot idf$ values. Next, for each sub-keyword and each news website, the average of sentiment values of news articles is calculated. This average represents the sentiment tendency of the website with respect to the subtopic. Next, sentiment graphs are generated and presented when the user specifies a website, or a subtopic, or both of them. By browsing the visual results provided by the system, the user can grasp the sentiment tendencies of different websites with respect to different subtopics, and further determine the credibility of news articles by themselves.

3. Offline processing

In preprocessing, news articles are first collected and analyzed as follows:

1. n news articles (P_1, ..., P_i, ..., P_n) are crawled from several news websites listed in Table 1. The news websites are those associated with eight newspapers published in Japan and three newspapers' Japanese versions in other countries. HTML tags are eliminated from the crawled news articles.
2. The articles are morphologically analyzed to extract proper nouns, general nouns, adjectives, and verbs.
3. The weight $tf \cdot idf(w, P_i)$ of each extracted word w in news article P_i is calculated:

$$tf \cdot idf(w, P_i) = \frac{N(w, P_i)}{N(P_i)} \cdot log \frac{N}{N(w)} \quad (1)$$

   where $N(w, P_i)$ is the number of times that word w appears in article P_i, $N(P_i)$ is the number of words extracted from P_i, N is the number of all collected news articles, and $N(w)$ is the number of articles in which word w appears.
4. A sentiment dictionary, in which each entry indicates the correspondence of a target word and its sentiment value, is constructed.
5. A sentiment vector of four dimensions is generated for each article by averaging the sentiment values of words which appear in the article.

The following sections describe the details of the sentiment dictionary construction (step 4) and sentiment vector generation for news articles (step 5).

3.1. Sentiment dictionary construction

We consider sentiment values of four dimensions for news articles: "Joy ⇔ Sadness", "Acceptance ⇔ Disgust", "Anticipation ⇔ Surprise", and "Fear ⇔ Anger". A sentiment dictionary is constructed to extract the sentiment values of these four dimensions for

Table 1: News website list.

ID	News website	URL
1	Tokyo Web	www.tokyo-np.co.jp/
2	Kobe Shimbun	www.kobe-np.co.jp/
3	Hokkaido Shimbun	www.hokkaido-np.co.jp/
4	Kahoku Online	www.kahoku.co.jp/
5	Chunichi Web	www.chunichi.co.jp/
6	Chugoku Shimbun	www.chugoku-np.co.jp/
7	Okinawa Times	www.okinawatimes.co.jp
8	Journal Nagasaki	www.nagasaki-np.co.jp
9	People's Daily Online	j1.people.com.cn/
10	Chosun Online	www.chosunonline.com/
11	U.S. Frontline	www.usfl.com/

words by analyzing the Nikkei Newspaper Full Text Database from 1990 to 2001, which consists of two million articles in total. The basic idea is to compare the co-occurrence of each target word with two groups of original sentiment words for each dimension. The original sentiment words for the four dimensions are listed in Table 2. Each dimension e ($e \in \{a, b, c, d\}$) has two opposite sets e_1 and e_2 of original sentiment words. For example, for dimension a: "Joy ⇔ Sadness", $a_1 = \{pleasure,\ be\ pleased,\ ...,\ bless\}$ and $a_2 = \{sad,\ feel\ sorry,\ ...,\ sorrow\}$. Each entry of the sentiment dictionary (Table 3) consists of a target word w and its sentiment values (including a scale value $S_e(w)$ and a weight $M_e(w)$) of four dimensions.

The scale value $S_e(w)$ of one dimension is calculated using the following procedure. First, considering the Y (year) edition of the Nikkei newspaper, let the number of articles which include any word in the set e of original sentiment words in Table 2 be $df(Y, e)$, and let the number of articles which include both target word w and any word in e be $df(Y, e\&w)$ [1]. The joint probability $P(Y, e\&w)$ of e and w is calculated as follows:

$$P(Y, e\&w) = \frac{df(Y, e\&w)}{df(Y, e)} \quad (2)$$

Next, considering the two opposite sets e_1 and e_2 of original sentiment words, the interior division ratio $R_{e_1 \Leftrightarrow e_2}(Y, w)$ of $P(Y, e_1\&w)$ and $P(Y, e_2\&w)$ is :

$$R_{e_1 \Leftrightarrow e_2}(Y, w) = \frac{P(Y, e_1\&w)}{P(Y, e_1\&w) + P(Y, e_2\&w)} \quad (3)$$

where $R_{e_1 \Leftrightarrow e_2}(Y, w) = 0$ if the denominator is 0.

1 We compared our methods which counted co-occurrence on a document level with those on a paragraph or sentence level in our preliminary experiments. The results showed which the processing time of the methods on a paragraph or sentence level increased dramatically but the improvement of precision was not remarkable. Therefore, the document-level co-occurrence was chosen in our current implementation.

Table 2: Original sentiment words for four dimensions (translated from Japanese).

Dimensions (e)	Original sentiment words ($e_1 \Leftrightarrow e_2$)
a: Joy	pleasure, be pleased, glad, happy, enjoy, blessing, bless
$\Leftrightarrow$ Sadness	$\Leftrightarrow$ sad, feel sorry, sadness, sorrow
b: Acceptance	agreement, agree, consent, acknowledgment, acknowledge, acceptance, accept
$\Leftrightarrow$ Disgust	$\Leftrightarrow$ disgust, dislike, hate, be unpleasant, antipathy, have an antipathy, evasion, evade
c: Anticipation	expectation, expect, anticipation, anticipate, forecast
$\Leftrightarrow$ Surprise	$\Leftrightarrow$ surprise, be surprised, astonishment, astonish, admiration, admire
d: Fear	fear, be scary, misgivings, have misgivings, be frightened
$\Leftrightarrow$ Anger	$\Leftrightarrow$ anger, get angry, resentment, resent, rage, enrage

Table 3: Examples of sentiment dictionary entries (translated from Japanese).

	Joy	$\Leftrightarrow$ Sadness	Acceptance	$\Leftrightarrow$ Disgust	Anticipation	$\Leftrightarrow$ Surprise	Fear	$\Leftrightarrow$ Anger
Entry word w	$S_a(w)$	$M_a(w)$	$S_b(w)$	$M_b(w)$	$S_c(w)$	$M_c(w)$	$S_d(w)$	$M_d(w)$
childcare	0.604	1.273	0.336	1.199	0.285	1.346	0.404	1.105
get angry	0.274	1.300	0.170	1.179	0.107	1.304	0.021	1.622
ghost	0.395	0.869	0.416	0.617	0.338	0.849	0.793	0.803
new year's present	0.897	0.877	0.516	0.456	0.393	0.877	0.564	0.348
smell	0.485	1.309	0.098	1.205	0.133	1.304	0.469	1.113
strong	0.575	1.270	0.190	1.221	0.397	1.489	0.422	1.159

Finally, the scale value $S_e(w)$ is calculated as the average value of all editions,

$$S_e(w) = \frac{\sum_{Y=1990}^{2001} R_{e_1 \Leftrightarrow e_2}(Y, w)}{\sum_{Y=1990}^{2001} T_{e_1 \Leftrightarrow e_2}(Y, w)} \quad (4)$$

where $T_{e_1 \Leftrightarrow e_2}(Y, w)$ is 0 if both $df(Y, e_1 \& w)$ and $df(Y, e_2 \& w)$ are 0, and $T_{e_1 \Leftrightarrow e_2}(Y, w)$ is 1 otherwise. The introduction of the denominator tends to assign a relatively large $S_e(w)$ to those words, e.g., "Olympics", which appear only during specific years (rather than every year) but are strongly related to specific sentiment words. The scale value $S_e(w)$ of a word w is between 0 and 1. This value is close to 1 if w appears in many articles together with the original positive words in e_1, and is close to 0 if w and the original negative words in e_2 often appear in the same articles.

For different words, the numbers of editions in which they appear and the total number of occurrences may vary greatly. Therefore, we introduce the weight $M_e(w)$ of w, which is calculated as follows:

$$\begin{aligned} M_e(w) \quad = \quad & \log_{12} \sum_{Y=1990}^{2001} T_{e_1 \Leftrightarrow e_2}(Y, w) \times \quad (5) \\ & \log_{144} \sum_{Y=1990}^{2001} (df(Y, e_1 \& w) + df(Y, e_2 \& w)) \end{aligned}$$

$M_e(w)$ is proportional to the number of editions and the number of occurrence, which means words which appear multiple times and in several editions are assigned large weights. Specifically, the words, $M_e(w)$ of which are 0, are not appended to the sentiment dictionary. Since we use a large corpus, the number of such words is actually small and the coverage of words in the sentiment dictionary is high.

3.2. Sentiment vector generation for news articles

The sentiment vector $O(P)$ of a news article P has the form $(O_a(P), O_b(P), O_c(P), O_d(P))$. Consider P as a set of words extracted from it by the morphological analysis. A sentiment value $O_e(P)$ of article P on dimension e is calculated by averaging and inclining the sentiment values of words which appear in P. The calculation equation is as follows:

$$O_e(P) = \frac{\sum_{w \in P} S_e(w) \times |2S_e(w) - 1| \times M_e(w)}{\sum_{w \in P} |2S_e(w) - 1| \times M_e(w)} \quad (6)$$

where the scale value $S_e(w)$ and weight $M_e(w)$ can be looked up in the sentiment dictionary constructed as described above. Many general words may be independent of the sentiment of the text, and the scale values $S_e(w)$ of these words are approximately 0.5. The

$|2S_e(w) - 1|$ term of these words approach 0, so that the effect of the emotionless words is removed.

A sentiment value $O_e(P)$ of a news article calculated by Equation (6) is a value between 0 and 1. For the symmetry of the sentiment graphs which will be generated in the online processing, we normalize the sentiment value to a value $\in$ (-0.5, +0.5) by subtracting 0.5. A sentiment value close to +0.5 means the sentiment of the news article tends to be positive ("Joy", "Acceptance", "Anticipation", or "Fear"), while a value close to -0.5 means the sentiment tends to be negative ("Sadness", "Disgust", "Surprise", or "Anger").

4. Online processing

When a user enters a search keyword representing the topic which he is concerned about, the system performs the following procedure and returns sentiment graphs and relevant news articles.

1. The news articles which include the search keyword are retrieved from the article database.
2. Sub-keywords are extracted from the retrieved articles for representing the subtopics of the specified topic. Concretely, the words with the highest $tf \cdot idf$ values are used for the sub-keywords. For example, for a topic "soccer", the sub-keywords, such as "world cup", "champions league", "FIFA", and "Messi" etc., are extracted.
3. The news articles including both the search keyword and each extracted sub-keyword are retrieved and grouped by each news website.
4. Sentiment vectors for each news website are generated by averaging the sentiment vectors of news articles in that website, which are generated as described in Section 3.2. Each vector represents the sentiment tendency of each news website with respect to each sub-keyword.
5. When the user specifies a website and does not select any sub-keyword, the system generates a sentiment graph of the specified website, which visually presents sentiment difference between multiple relevant subtopics. The GD graphics library and pChart chart library are used for creating visual sentiment graphs . For example, a user inputs a search keyword "Hanshin", specifies a website "Kobe Shimbun", and selects no sub-keywords. The results are shown in Figure 2. The upper frame displays the search keyword and the search period of news articles. The lower-right frame displays the list of news websites and the list of extracted sub-keywords related to the search keyword. The lower-left frame displays the generated sentiment graph, which in this case shows that different subtopics of "Hanshin" have different sentiment on dimension "Joy $\Leftrightarrow$ Sadness".
6. When the user specifies a sub-keyword and does not select any website, the system generates a sentiment graph, which presents sentiment difference between different websites for the specified subtopic (Figure 3).
7. When the user selects both a website and a sub-keyword, the sentiment values of each news article related to the subtopic in the website are presented in the sentiment graph. The user can also browse the contents of news articles by clicking on the presented article list (Figure 4).

5. Experiments

We first evaluated the accuracy of sentiment extraction and subtopic extraction. Then we observed sentiment difference between different subtopics and between different websites.

5.1. Sentiment extraction accuracy

To evaluate the accuracy of sentiment extraction, we compared the sentiment values calculated by the system and the conversion values from the results of questionnaires filled out by 100 individuals.

We specified a website "Tokyo Web" and selected five search keywords "Beijing", "teacher", "Hashimoto governor", "Kyoto", and "Fukuda premier". For each keyword, the system retrieved related news articles from the website, and calculated the sentiment values of four dimensions by averaging the sentiment values of the 10 news articles with the highest $tf \cdot idf$ values of the search keyword.

For each keyword, the 100 individuals were asked to read the 10 news articles and evaluate the sentiment tendency of each article. For example, considering dimension "Joy $\Leftrightarrow$ Sadness", an individual should evaluate an article's sentiment from five levels: "joy", "close to joy", "neither joy nor sadness", "close to sadness", and "sadness". For a news article, n_1, n_2, n_3, n_4, and n_5 ($\sum_{i=1}^{5} n_i = 100$) were the numbers of the individuals who gave the five levels. We converted the evaluation of the individuals to a numerical value using the following scoring system: "joy = 1", "close to joy = 0.75", "neither joy nor sadness = 0.5", "close to sadness = 0.25", and "sadness = 0". The sentiment value of a news article on a dimension evaluated by the 100 individuals was $(n_1 {*} 1 + n_2 {*} 0.75 + n_3 {*} 0.5 + n_4 {*} 0.25 + n_5 {*} 0)/100$. Finally, the sentiment values of the 10 news articles were averaged as the conversion value for the keyword.

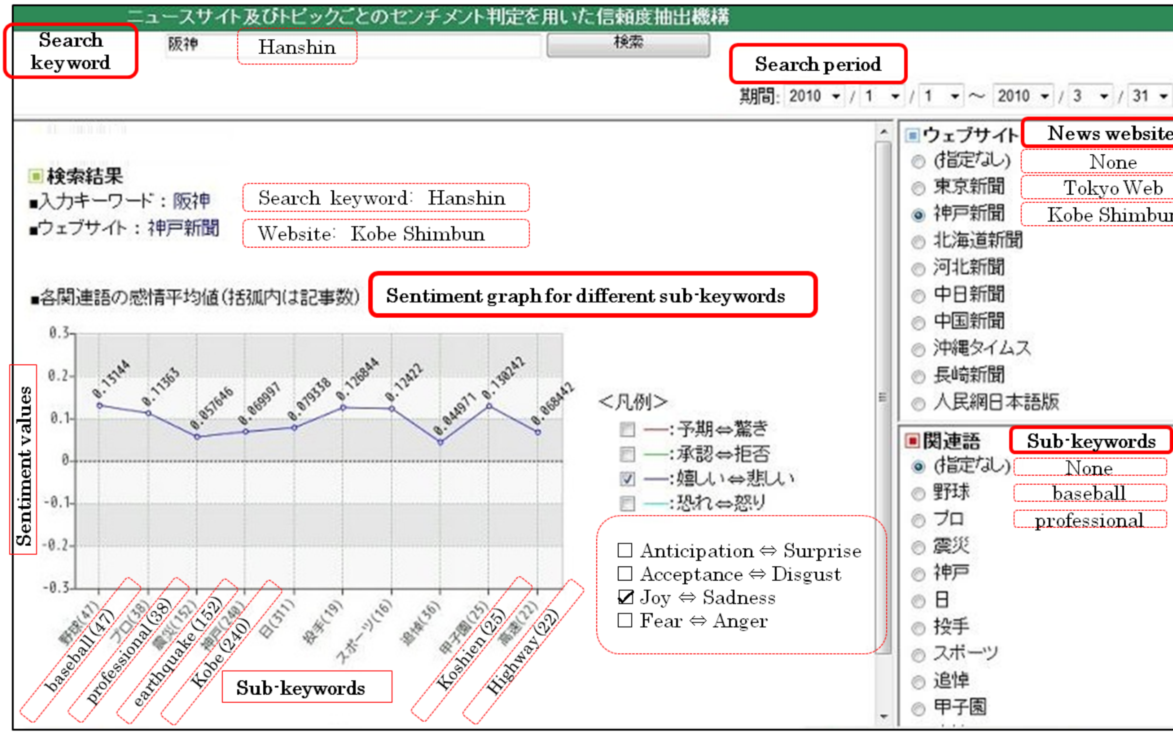

Figure 2: System snapshot 1: when a website is specified and no sub-keyword is selected.

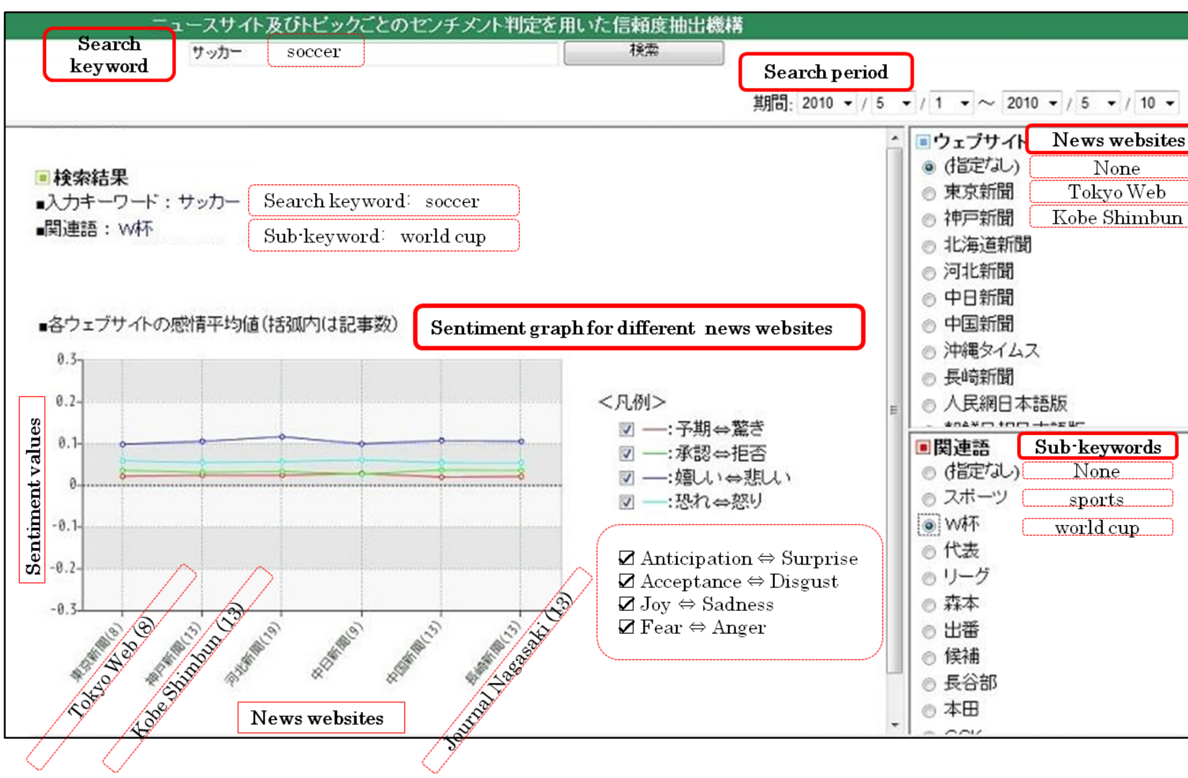

Figure 3: System snapshot 2: when a sub-keyword is specified and no website is selected.

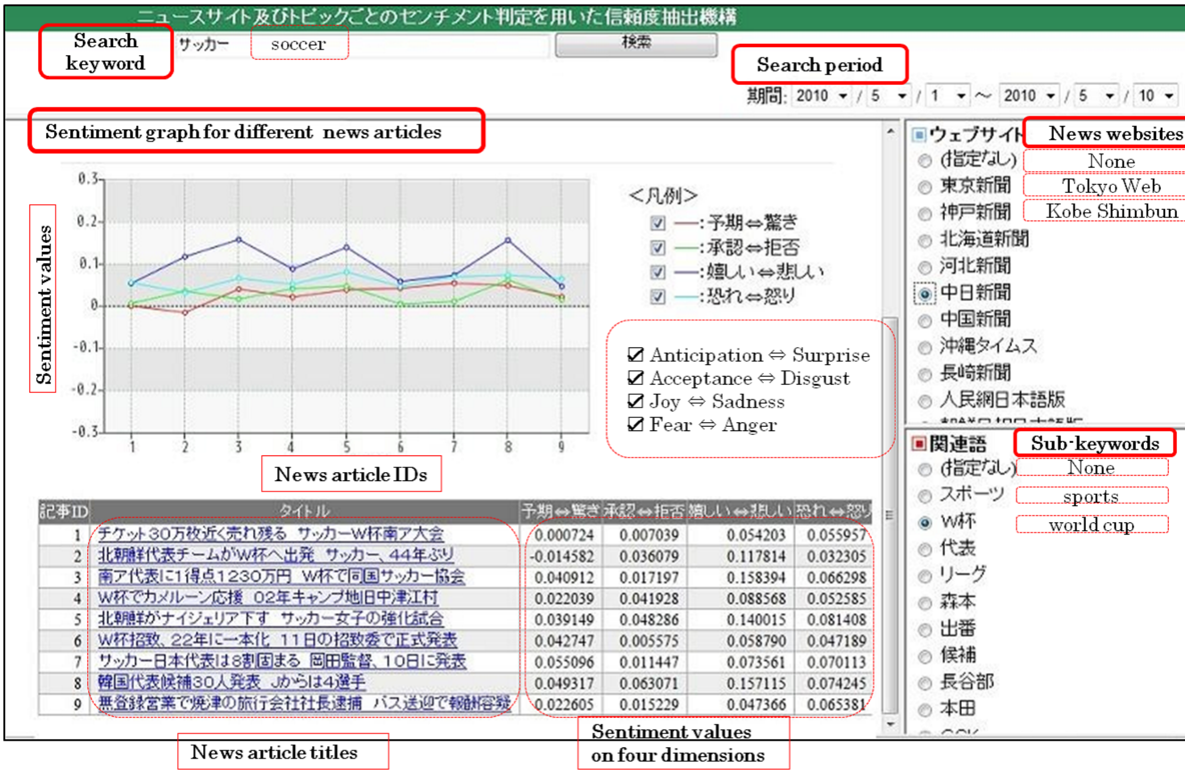

Figure 4: System snapshot 3: when both a website and a sub-keyword are specified.

Table 4: Evaluation of the error of sentiment values between sentiment values calculated by the system and sentiment values decided by individuals.

Query keyword		Average sentiment values of news articles related to the search keyword			
		Joy ⇔ Sadness	Acceptance ⇔ Disgust	Anticipation ⇔ Surprise	Fear ⇔ Anger
Beijing	user	0.5203	0.5815	0.5368	0.5165
	system	0.5110	0.5255	0.3766	0.5566
teacher	user	0.2533	0.3230	0.3528	0.7560
	system	0.4639	0.5135	0.4430	0.4952
Hashimoto governor	user	0.5080	0.5590	0.5115	0.5075
	system	0.4236	0.5244	0.5238	0.4571
Kyoto	user	0.4733	0.5903	0.5135	0.5120
	system	0.5299	0.5440	0.3983	0.5587
Fukuda premier	user	0.4418	0.4825	0.4453	0.5800
	system	0.4208	0.4692	0.4957	0.4519
Average error		0.07638	0.06814	0.08566	0.10522

The error of sentiment values between the system and the user evaluation is shown in Table 4. For most of search keywords and most of sentiment dimensions, the sentiment values calculated by the system and those evaluated by the individuals are similar. For example, for the keyword "Beijing", the sentiment value on dimension "Joy ⇔ Sadness" which our system calculated is 0.5110, a value close to the user evaluation's 0.5203. The average errors of all the search keywords are small, between 0.06814 and 0.10522, which indicates that the system can extract sentiment values similar to those decided by the individuals. The dimension "Fear ⇔ Anger" has the largest error, because fear-anger is a relatively weak contrast comparable to the other three pairs. Our research group is also designing new sentiment dimensions [2], which is geared to the needs of news analysis based on factor and cluster analysis.

5.2. Subtopic extraction accuracy

The same 100 individuals also evaluated that given a keyword (topic), whether the extracted top

$k_1 : Haneda$	$k_2 : Futenma$	$k_3 : Cheonan$
$k_4 : Yanba\ dam$	$k_5 : Hanshin$	$k_6 : Ozawa$
$k_7 : Nakamura$	$k_8 : Ishikawa$	$k_9 : Matsui$
$k_{10} : soccer$	$k_{11} : greenhouse$	$k_{12} : highway$
$k_{13} : influenza$	$k_{14} : aftosa$	$k_{15} : budget$

Figure 5: 15 topics used for the observation on sentiment difference (translated from Japanese).

10 sub-keywords were appropriate as representing its subtopics. These individuals freely inputted their search keywords five times and gave their overall impression on the extracted sub-keywords' appropriateness from four levels: "appropriate", "somewhat appropriate", "somewhat inappropriate", "inappropriate". The numbers of the individuals giving four evaluation levels are 7, 63, 28 and 2 respectively. 70% individuals thought the extracted sub-keywords were "appropriate" or "somewhat appropriate", which indicates that the extracted sub-keywords are effective for representing the subtopics.

5.3. Sentiment difference between different subtopics

We picked up 15 search keywords (15 topics) listed in Figure 5 for the observation on sentiment difference. Since the news articles were crawled from Japanese news websites or Japanese versions of other countries' news websites, the topics related to Japanese society were selected. k_1 - k_4 are the economic or political incidents, k_5 is a polysemous keyword, k_6 - k_9 are the names of politicians or sports players, and k_{10} - k_{15} are relatively common topics.

For each of the 15 topics, we specified a website and observed whether there existed sentiment difference between different subtopics of the topic. It was found that most of the 15 topics had more or less sentiment difference between their different subtopics. Especially, sentiment difference between subtopics was relatively obvious for the topics which were polysemous or had relatively independent subtopics.

For example, the polysemous search keyword "k_5 : Hanshin" is not only a baseball team name, but also a placename which a massive earthquake struck. As Figure 6 shows, for the subtopics related to the baseball, such as "baseball", "pitcher", and "Koshien", the website "Kobe Shimbun" has relatively positive sentiment on dimension "Joy ⇔ Sadness", while for the subtopics related to the earthquake, such as "earthquake" and "mourning", the website has relatively negative sentiment. Figure 7 shows sentiment difference between the subtopics of the search keyword "k_6 : Ozawa", which has relatively independent subtopics. "Ozawa", who

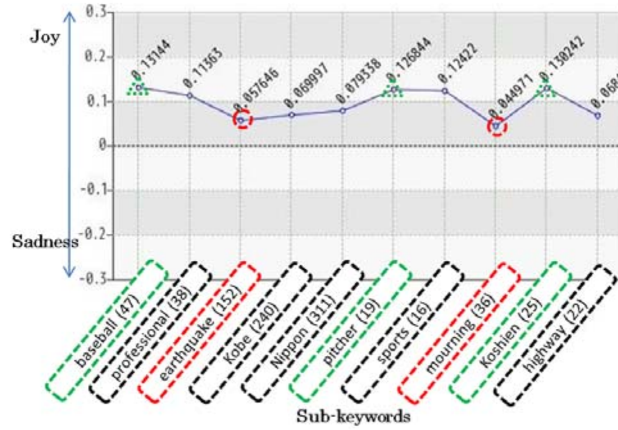

Figure 6: Sentiment graph of the specified website "Kobe Shimbun" on dimension "Joy ⇔ Sadness" with respect to the search keyword "Hanshin" and different sub-keywords.

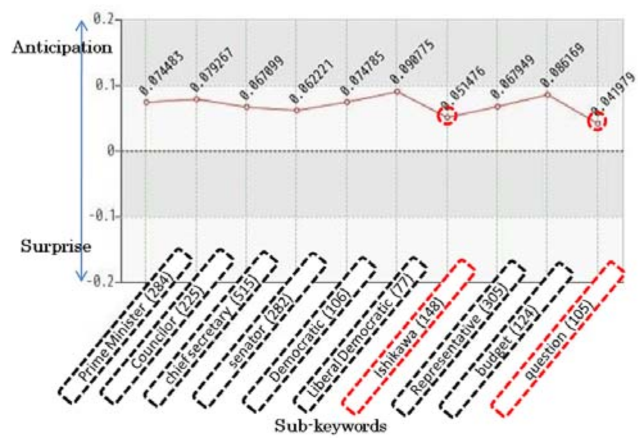

Figure 7: Sentiment graph of the specified website "Tokyo Web" on dimension "Anticipation ⇔ Surprise" with respect to the search keyword "Ozawa" and different sub-keywords.

was the secretary-general of the Democratic Party of Japan, made some achievements as people expected, but he also surprised the media because his former secretary "Ishikawa" was arrested and the prosecutors "question"ed him. Our system faithfully reflects that news' sentiment for the subtopics "Ishikawa" and "question" are relative "Surprise", while those for the other subtopics are relative "Anticipation".

The sentiment graph also presents the numbers of news articles related to each subtopic in the parentheses. By observing the numbers of news articles about positive subtopics and negative subtopics which a website is reporting, the user can also be somewhat aware of the website's tendency.

5.4. Sentiment difference between different websites

For each of the 15 search keywords listed in Figure 5, we selected the top one from the extracted sub-keywords and observed that whether there existed sentiment difference between different websites with respect to the search keyword and the sub-keyword. For the search keywords k_1 - k_5, our expectation was that sentiment difference between websites might exist, because certain news websites had their own interests about the five topics. For k_6 - k_9, it was difficult to predict whether sentiment difference between websites existed or not. For k_{10} - k_{15}, sentiment difference between websites might be inapparent, because they were not the topics which were easy to evoke conflicts between different websites.

Our observation results show that there is no obvious sentiment difference between websites for the search keywords k_4 - k_{15}. Because for most of news topics most of news websites report the facts objectively, news articles which they write have similar sentiment tendency. The sentiment graph in Figure 3 is a representative example, which shows that for the search keyword "k_{10} : soccer" and its top sub-keyword "world cup", all websites' sentiment is parallel on the whole. However, sentiment bias is exactly detected for certain websites and certain topics. Our system detects sentiment bias for the three search keywords "k_1 : Haneda", "k_2 : Futenma", and "k_3 : Cheonan". Their corresponding top sub-keywords were "hub", "move", and "sinking".

"Haneda hub" is an issue that Japan's transport minister planed to upgrade Tokyo's Haneda airport into a 24-hour international hub. Figure 8 shows that the news website "Tokyo Web" has more positive sentiment on dimension "Acceptance ⇔ Disgust" than the other websites. This is because the local newspaper "Tokyo Web" tends to support this plan which can bring local people benefit. Through the sentiment graph, users can understand the local website's tendency, and notice there are also other websites with different attitudes. Therefore, users can choose to browse news articles from different perspectives. "Futenma move" is an issue of relocating the U.S. Marine Corps Air Station Futenma in Okinawa, Japan. Okinawans increasingly felt angry because Japanese government did not give full consideration to their pain of accepting some of Futenma's functions. Figure 9 shows the website "Okinawa Times" has more obvious anger than the other websites. Similarly, for "Cheonan sinking", the sentiment of the Korean newspaper "Chosun Online" is angrier than the other websites.

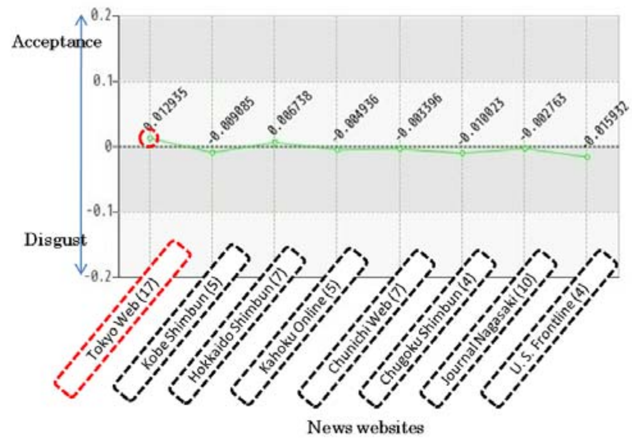

Figure 8: Sentiment difference between different websites on dimension "Acceptance ⇔ Disgust" with respect to the search keyword "Haneda" and the sub-keyword "hub".

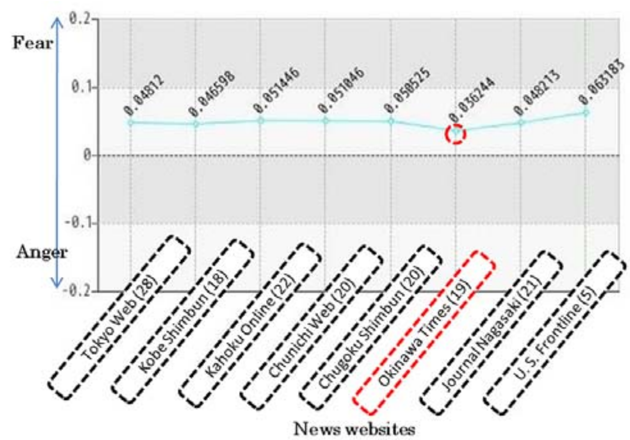

Figure 9: Sentiment difference between different websites on dimension "Fear ⇔ Anger" with respect to the search keyword "Futenma" and the sub-keyword "move".

Also, the numbers of news articles of each website related to the specified subtopic are presented in the parentheses of the sentiment graph, so that users can understand how concerned each website is about the subtopic. A website's intention may be revealed by presenting the sentiment bias and the numbers of news articles related to certain subtopics. Extracting this kind of background knowledge is helpful to users when they want to judge the news credibility.

6. Related research

Sentiment analysis [3, 4, 5] is increasingly important in the areas of NLP and text mining, which extracts sentiment from text such as movie reviews, book reviews, and production evaluations. Turney [6] proposed a method for classifying reviews into two categories: recommended and not recommended based on mutual information. Pang et al [7] extracted only the subjective portions of movie reviews and classified them as "thumbs up" or "thumbs down" by applying text-categorization techniques. Esuli et al [8] presented a method for determining the orientation of subjective terms based on quantitative analysis of the glosses of such terms. Liu et al [9] presented an autoregressive sentiment-aware model to utilize the sentiment information for predicting product sales performance. Lu et al [10] proposed to automatically integrate lots of opinions scattering in various sources using semi-supervised topic models. Su et al [11] detected sentiment association between product feature words and opinion words on Chinese reviews. Some generative models [12, 13] were proposed for sentiment retrieval considering both topic relevance and sentiment relevance.

However, these methods only consider positive-negative sentiment. Unlike these methods, the developed system captures more detailed sentiment aspects of four dimensions. Except for the model of emotion proposed by Plutchik [1] which is used by our current research, there also exist other models [2, 14, 15]. Extension based on these models is one of our future work.

Previous work mainly focuses on using sentiment analysis for opinion extraction or opinion integration. The sentiment of information senders such as news websites is also very useful for assisting users in judging the news credibility. Our effort is to detect and visualize sentiment difference between different subtopics and between different websites, and make use of them to evaluate the information credibility.

7. Conclusions and future work

In this paper, we described a system for detecting and visualizing sentiment bias of Web news. The system can dynamically summarize the sentiment for different subtopics and for different websites. As the experimental evaluations and the observation showed, the accuracy of sentiment extraction and subtopic extraction was good, and the news website's sentiment bias could be detected when its tendency was different from the others. The background knowledge can be used to assist users in determining news credibility.

News sentiment with respect to a topic may also vary with time. We will take into account the aspect of time in our future work. Furthermore, we plan to construct a model which can automatically calculate credibility scores for news articles based on sentiment difference between subtopics and between news websites.

Acknowledgments

This work was supported in part by the National Institute of Information and Communications Technology, Japan, and by the MEXT Grant-in-Aid for Young Scientists (B) (#21700120, Representative: Yukiko Kawai).

References

[1] R. Plutchik, "*The Emotions*," Univ Pr of Amer, 1991.

[2] T. Kumamoto, "Design of Impression Scales for Assessing Impressions of News Articles," In *SNSMW 2010*, pp. 285–295, 2010.

[3] C. Strapparava and R. Mihalcea, "Task 14: Affective Text," In *SemEval 2007*, 2007.

[4] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," *Foundations and Trends in Information Retrieval*, Vol. 2, Nos. 1-2, pp. 1–135, 2007.

[5] A. Wright, "Our Sentiments, Exactly," *Communications of the ACM*, Vol. 52, No. 4, pp. 14–15, 2009.

[6] P. D. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," In *ACL 2002*, pp. 417–424, 2002.

[7] B. Pang and L. Lee, "A Sentiment Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts," In *ACL 2004*, pp. 271–278, 2004.

[8] A. Esuli and F. Sebastiani, "Determining the Semantic Orientation of Terms through Gloss Classification," In *CIKM 2005*, pp. 617–624, 2005.

[9] Y. Liu, X. Huang, A. An, and X. Yu, "ARSA: A Sentiment-aware Model for Predicting Sales Performance Using Blogs," In *SIGIR 2007*, pp. 607–614, 2007.

[10] Y. Lu and C. Zhai, "Opinion Integration through Semi-supervised Topic Modeling," In *WWW 2008*, pp. 121–130, 2008.

[11] Q. Su, X. Xu, H. Guo, Z. Guo, X. Wu, X. Zhang, B. Swen, and Z. Su, "Hidden Sentiment Association in Chinese Web Opinion Mining," In *WWW 2008*, pp. 959–968, 2008.

[12] K. Eguchi and V. Lavrenko, "Sentiment Retrieval Using Generative Models," In *EMNLP 2006*, 2006.

[13] M. Zhang and X. Ye, "A Generation Model to Unify Topic Relevance and Lexicon-based Sentiment for Opinion Retrieval," In *SIGIR 2008*, pp. 411–418, 2008.

[14] J. A. Russell, "A Circumplex Model of Affect," *Journal of Personality and Social Psychology*, Vol. 39, No. 6, pp. 1161–1178, 1980.

[15] G. Pitel and G. Grefenstette, "Semi-automatic Building Method for a Multidimensional Affect Dictionary for a New Language," In *LREC 2008*, 2008.