A Sentiment Polarity Prediction Model Using Transfer Learning and Its Application to SNS Flaming Event Detection

Seiichi Ozawa*, Shun Yoshida*, Jun Kitazono*, Takahiro Sugawara†, Tatsuya Haga†
*Graduate School of Engineering, Kobe University
1-1, Rokkoudai-cho, Nada, Kobe, Hyogo
Email: ozawasei@kobe-u.ac.jp, kitazono@eedept.kobe-u.ac.jp
†Eltes Co., Ltd.
Shimbashi Square Building, 5-14-10, Shimbashi, Minato, Tokyo
Email: toiawaseeltes@eltes.co.jp

***Abstract*—In recent years, with the popularization of SNS, the incidents called *flaming*, in which a large number of negative comments are retweeted and spread to many followers on SNS, are increasing. Since a flaming event sometimes causes severe criticism by public people, it is becoming a great thread to companies and therefore it is important for companies to protect their reputation from such flaming events. In order to protect companies from serious damages in reputation, we propose a machine learning approach to the detection of flaming events by monitoring the sentiment polarity of SNS comments. From the nature of SNS comments such as the spread of a large number of retweets with the same content for a short time, the word distributions are often strongly biased and it leads to poor performance in sentiment polarity prediction. To alleviate this problem, we introduce transfer learning into the conventional Naive Bayes classifier. More concretely, in the Naive Bayes classifier, the occurrence probabilities of words on a target domain are recalculated using those on other domains, where a domain corresponds to a company to be protected. The experimental results demonstrate that the proposed transfer learning contribute to the improvement in the sentiment polarity prediction for SNS comments. In addition, we show that the proposed system can detect flaming events correctly by monitoring the number of negative comments.**

I. Introduction

Due to the recent popularization of Social Networking Service (SNS) such as Twitter and Facebook, the event so-called 'flaming' becomes serious problems especially for companies. The flaming is a phenomenon that negative comments explosively increases and spread for a short term as a trigger such as inappropriate speech and behavior. Once the flaming occurs, the company reputation could seriously be damaged unless a proper action such as apology and promise to explore incident causes is taken promptly. Therefore, in order to minimize the damage in reputation, it is important for companies to detect flaming events as early as possible. To monitor negative comments on SNS and to detect flaming events, reputation management (RM) companies have been established recently. In RM companies, human operators have to keep monitoring negative comments for 24 hours, 365 days because the SNS users submit comments anytime, day and night. Therefore, an automatic monitoring of negative comments has been solicited for such RM companies.

A straightforward way to detect flaming events is to carry out sentiment polarity prediction for each comment and to monitor the number of negative comments over time. Many researches on the sentiment polarity prediction have been studied so far [1], [2], [3], [4], [5] and it has been applied to the sentiment analysis for SNS comments [6], [7], [8]. However, most studies do not consider the following aspects of SNS.

- The distribution of frequently used words and their sentiment polarity can significantly be variated depending on domains (e.g., beverage company and car maker).
- A flaminig event on SNS is often caused by spreading a large number of retweets with the same content for a very short time.

Considering the above nature of SNS comments, it is very likely that the word distribution of a speccific domain could strongly be biased and this often causes the performance deterioration in the sentiment polarity prediction.

To alleviate the above problem, we propose a new sentiment polarity prediction model with a transfer learning function [9], [10], [11], [12]. For a target task, we first learn the occurrence probability of words in a Naive Bayes classifier. Then, the knowledge transfer is conducted by transferring the occurrence probabilities of selected words in other domains to a target domain. In the transfer learning, we select words with high entropy (i.e., commonly used words among different domains), and the occurrence probability of a target domain is recalculated by averaging over the probabilities of such words in other domains. Using the proposed sentiment polarity prediction, we build a flaming detection system where a time course of the number of predicted negative comments is monitored.

The rest of the paper is organized as follows. Section II first gives a brief explanation on preprocessing, feature selection, and Naïve Bayes classifier model, and then we propose an extension model using transfer learning. In Section III, we evaluate the accuracy of the proposed sentiment polarity prediction for the four domain datasets that consists of Japanese SNS comments on a specific company collected from various sources (e.g., Twitter, Facebook). Then, we study how the sentiment polarity prediction works in the flaming detection. Finally, we conclude this paper in Section IV.

II. Proposed Model

A. Preprocessing and Feature Selection

In Japanese, a morpheme is the smallest unit of meaningful words and it does not always consist of a single character. Therefore, to analyze a Japanese comment, a text document is first divided into morphemes using a morphological analysis, and then a part of speech or their inflectional form are given to each morpheme. In addition, to characterize the relation between morphemes, the dependency parsing is needed to distinguish the dependency structure that gives the ornamentation relations. Here, we adopt MeCab [13] and CaboCha [14] as the engines of morphological analysis and dependency parsing, respectively.

The morphological analysis provides the parts of speech of morphemes consisting of a comment. In general, not all the parts of speech are required to determine the sentiment polarity of a comment. Here, only noun, verb, adjective, adverb, adnominal, and conjunction are taken into consideration in the sentiment analysis; that is, other parts of speech are removed from features. In addition, we eliminate one byte characters and alphanumeric characters as noise features. Note that each morpheme is transformed into an end-form of conjugations and used as a feature.

The sentiment polarity of a morpheme may sometimes be changed by a negation word. For example, if a negation word “nai” or “n” in Japanese exists, the dependency parsing is first applied and the morpheme that gives ornamentation relation to the negation word would be transformed into its negation feature.

B. Naïve Bayes Classifier

Here, let us give a brief explanation of a multinomial Naïve Bayes model [15], [16] we adopt here. Suppose that we obtain a set of $|V|$ words $V = \{w_1, \cdots, w_{|V|}\}$ selected from training data whose number of classes is $|C|$. In the later experiment, we set $|C| = 3$ because the sentiment polarity is either positive, neutral, or negative.

In the multinomial Naïve Bayes model [15], it is assumed that the occurrence of all words are mutually independent and a comment $\boldsymbol{d}$ with $|d|$ words is generated by drawing a single word $|d|$ times from a vocabulary V. Further let $p_{i,c}$ be an occurrence probability of a word w_i $(1 \leq i \leq |V|)$ for a class c and let $\boldsymbol{\theta}_c = \{p_{1,c},\ p_{2,c},\ \cdots,\ p_{|V|,c}\}$ be a parameter vector for class c. Then, the likelihood $P(\boldsymbol{d}|\boldsymbol{\theta}_c)$ of a comment $\boldsymbol{d}$ is given by

$$P(\boldsymbol{d}|\boldsymbol{\theta}_c) = \frac{(\sum_i n_{i,d})!}{\prod_i n_{i,d}!} \prod_i p_{i,c}^{n_{i,d}} \tag{1}$$

where $n_{i,d}$ is the number of occurrences of a word w_i in a comment d. Further let p_c be a prior probability that a comment of class c is generated. Then, the output class c^* of the Naïve Bayes classifier is represented by the following equation:

$$\begin{aligned} c^* &= \operatorname*{argmax}_c \ p_c P(\boldsymbol{d}|\boldsymbol{\theta}_c) \\ &= \operatorname*{argmax}_c p_c \prod_i p_{i,c}^{n_{i,d}}. \end{aligned} \tag{2}$$

Here, we assume that the prior probabilities of parameters p_c and $p_{i,c}$ are subject to the Dirichlet distribution. Then, the probabilities p_c and $p_{i,c}$ are respectively estimated through the maximum posterior probability estimation as follows:

$$p_c = \frac{n_c + (\alpha - 1)}{\sum_c n_c + |C|(\alpha - 1)} \tag{3}$$

$$p_{i,c} = \frac{n_{i,c} + (\alpha - 1)}{\sum_i n_{i,c} + |V|(\alpha - 1)} \tag{4}$$

where n_c is the number of comments for class c and $n_{i,c}$ is the word frequency of w_i occurred in a training data set. In this paper, we adopt $\alpha = 2$ which is widely accepted as Laplace smoothing.

Using Eqs. (3) and (4), Eq. (2) is reduced to

$$c^* = \operatorname*{argmax}_c L(c) \tag{5}$$

$$\begin{aligned} L(c) &= \ln\left(\frac{n_c + 1}{\sum_c n_c + |C|}\right) \\ &\quad + \sum_i n_{i,d} \ln\left(\frac{n_{i,c} + 1}{\sum_i n_{i,c} + |V|}\right) \end{aligned} \tag{6}$$

The sentiment polarity prediction is carried out based on Eqs. (5) and (6).

C. Sentiment Polarity Prediction Using Transfer Learning

1) Basic Idea: Since the occurrence frequency of words and their sentiment polarity could significantly be different depending on domains, in general it is not easy to build a universal classifier that has high accuracy for arbitrary domains. Therefore, we adopt a *focused classifier* approach that is designed to estimate the sentiment polarity for specific domains.

Let us take an example of a car maker. Assume that a huge number of the following retweet comments are spread on twitter: *“That’s a good news for me! Discover a ’Recall’ in X carmaker! See what happens!”*. Apparently, a human supervisor would judge as a negative comment for this retweet and put a label as negative sentiment for the above comment. Learning the retweets with the class label (negative polarity), a trained classifier will predict negative for a word ‘Recall’ when it appears in the domain of a car maker. On the other hand, the sentiment polarity for positive words such as ‘good’ in the above retweet could also be modified as a negative word, and it may lead to a wrong estimation in the sentiment polarity. This could easily happen especially when the number of unique comments is limited and only a limited number of words appear in the comments in a specific domain. Thus, it will cause a large bias in the word distribution of training data (i.e., biased learning of a classifier). Actually, a flaming of negative comments on twitter usually includes lots of retweets with the same content that are caused by spreading an original tweet among followers. Therefore, the above conditions are often satisfied when training data are collected from the comments related to a specific company (domain) for a monitoring purpose. To alleviate the effects of learning training data with such a biased word distribution, we introduce *transfer learning* in which biased sentiment polarity of words are modified by using sentiment polarity information of other domains.

2) Proposed Transferred Learning: First, let us give a brief explanation on the entropy that is used for selecting commonly used words among different domains. Since a word like 'Recall' in the car domain appears only in the positive class (i.e., negative sentiment), the uniquness of a word w_i can be measured by the following entropy $H(i)$:

$$H(i) = -\frac{1}{\log |C|} \sum_c \frac{n_{i,c}}{\sum_c n_{i,c}} \log \frac{n_{i,c}}{\sum_c n_{i,c}} \tag{7}$$

where $0 \leq H(i) \leq 1$. As seen in Eq. (7), the entropy $H(i)$ has the maximum value of 1 if all words w_i have the same occurrence probabilities (i.e., w_i is a common word), and $H(i)$ is 0 if a word w_i appear in either class (positive, neutral, or negative).

Next, we explain the algorithm of the proposed transfer learning. Figure 1 shows the flow chart of the learning algorithm. First, we calculate the occurrence probability $p^t_{i,c}$ for a word w_i in the training set of comments related to a target domain t. If a word w_i appears in the comments included in the training set $\mathcal{D}^t$, $p^t_{i,c}$ is calculated from Eq. (4). If the entropy $H(i)$ of a word w_i is less than a threshold θ, w_i is considered as a unique word and the calculated probability $p^t_{i,c}$ is used to evaluate the sentiment polarity of a comment of a target domain t. Here, the threshold θ is determined through the cross-validation that will be mentioned in II-D.

If $H(i)$ is larger than or equal to θ, or a word w_i is not included in $\mathcal{D}^t$, the knowledge transfer is evoked so that the knowledge (i.e., the occurrence probability of words) of the target domain t can be modified or compensated with the knowledge of other domains to enhance the prediction accuracy. First, for every domains s where a word w_i appears more than once (i.e., $\sum_c n^s_{i,c} > 0$), a new domain probability $p^t_{i,c}$ for the domain t is obtained by averaging the probability $p^s_{i,c}$ as follow:

$$p^t_{i,c} \propto \left(\prod_{\{s | \sum_c n^s_{i,c} > 0\}} p^s_{i,c} \right)^{\frac{1}{\left|\sum_c n^s_{i,c} > 0\right|}} \tag{8}$$

where $\left|\sum_c n^s_{i,c} > 0\right|$ is the number of domains where a word w_i occurs. If a word w_i does not occur in any domains, it would not be used in the estimation.

The transfer learning is aiming for enhancing the prediction accuracy in sentiment polarity of comments in a targeted domain by modifying the occurrence probability with those in other domains. Therefore, it is expected that the proposed knowledge transfer effectively works especially when a sufficient number of training comments are given for other domains, while the number of comments in a target domain is scarce and/or a word frequency distribution has a strong bias.

D. Parameter Estimation

The sentiment polarity of words could be different depending on domains and their social backgrounds. Therefore, it might be useful to introduce the bias in the prediction of sentiment polarity. For this purpose, we change the decision function in Eq. (6) by considering a bias b as follow:

$$c^* = \begin{cases} 0 & \text{if } L(0) - L(1) - b > 0 \\ 1 & \text{otherwise.} \end{cases} \tag{9}$$

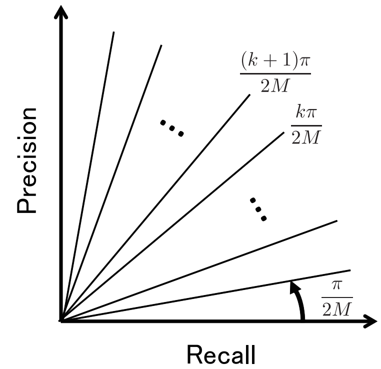


Fig. 2. Subregions to find optimal parameter pairs (θ^*_k, b^*_k) $(k = 1, \cdots, M)$ in a precision-recall plane.

where $c = 0$ is the class of positive or neutral sentiment and $c = 1$ is the class of negative sentiment.

In the proposed knowledge transfer, there are two parameters to be appropriately determined: the threshold θ in the knowledge transfer (see Fig. 1) and the bias b in Eq. (9). To choose proper parameters θ and b, we perform the K-fold cross-validation in the following way. First, consider a precision-recall plane in Fig. 2 and define M intervals with even angles: $\left[0, \frac{\pi}{2M}\right), \cdots, \left[\frac{(M-1)\pi}{2M}, \frac{\pi}{2}\right]$. For each interval $\left[\frac{k\pi}{2M}, \frac{(k+1)\pi}{2M}\right)$, we choose a pair of parameters (θ^*_k, b^*_k) so that the following Root Mean Square (RMS) of precision and recall is maximized:

$$RMS(\theta_k, b_k) = \sqrt{\text{Precision}^2 + \text{Recall}^2} \tag{10}$$

After obtaining M parameter pairs (θ^*_k, b^*_k) $(k = 1, \cdots, M)$, a parameter pair needs to be chosen depending on the user's preference. If a user wants to give priority to the precision rather than the recall, the user should select a parameter pair (θ^*_k, b^*_k) from the intervals on the precision side.

III. Performance Evaluation

In this section, we first evaluate the performance of the proposed sentiment polarity prediction system for various sources of Japanese SNS comments. Then, we demonstrate how the flaming detection system, in which negative comments are counted with the above prediction system, works well.

A. Experimental Setup

In the performance evaluation, we use the four domain datasets in Table I, each of which consists of Japanese SNS comments in a specific domain that are collected from the following sources: Twitter[1], Facebook[2], 2 Channel (bulletin board)[3], Yahoo! Answers[4], blogs, and news sites[5]. These domain datasets are collected and labelled by Eltes Co., Ltd. for their services called *reputation management*, which is a service to protect a company's brand from SNS flaming.

[1]https://twitter.com/
[2]https://www.facebook.com/
[3]http://www.2ch.net/
[4]http://chiebukuro.yahoo.co.jp/
[5]Yahoo! News and goo News

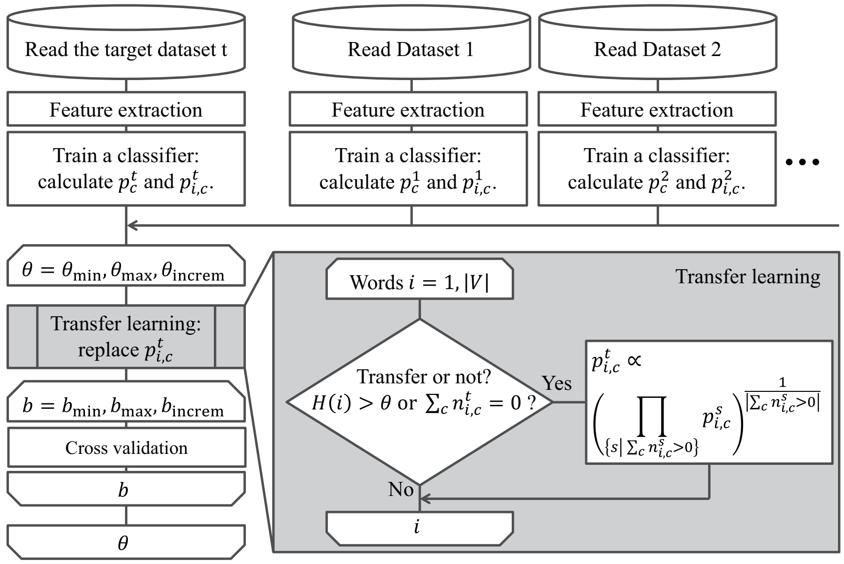


Fig. 1. The flowchart of the transfer learning.

TABLE I. DISTRIBUTIONS OF SENTIMENT POLARITY IN THE FOUR DOMAIN DATASETS AND THEIR DATA COLLECTION PERIODS.

	Positive	Neutral	Negative	Period of Data Collection
Dataset A	79,594	487,464	14,661	Jan.–Mar. 2013, Apr.–June 2014
Dataset B	109,542	366,542	10,380	Mar. 2013, Apr.–June 2014
Dataset C	193	23,652	9,837	Nov. 2013 to Dec. 2014
Dataset D	1,177	18,037	9,541	Mar.–June 2014

TABLE II. THE SOURCES OF SNS COMMENTS IN THE FOUR DOMAIN DATASETS.

	Twitter	Facebook	2 channel	Yahoo! Answers	Blogs	News sites
Dataset A	562,524	19,195	-	-	-	-
Dataset B	461,133	-	9,844	817	11,254	3416
Dataset C	7,674	-	26,008	-	-	-
Dataset D	24,793	-	2,677	160	1,035	90

Therefore, each domain dataset consists of comments collected by searching with some keywords related to a specific company (e.g., company name and its major products). Note that Company A and C in Table I are car makers, while Company B and Company D are a beverage company and a fast food chain, respectively.

Tables I and II show the number of comments in each class (i.e., positive, negative, neutral) and the number of comments collected from different sources, respectively. Since the purpose in this paper is to detect negative comments and to find a flaming incident as early as possible, actually we consider only the two classes: negative and non-negative classes.

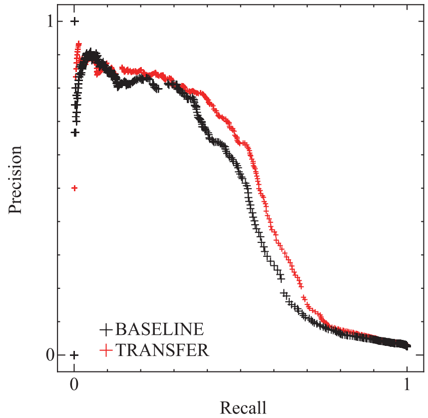
B. Evaluation of Sentiment Polarity Prediction

In this section, we study on how the transfer learning works effectively to enhance the prediction accuracy in the sentiment polarity of comments. In this experiment, we carry out the performance evaluation only for unique comments in the test datasets because the accuracy could strongly be affected by the prediction result for retweet comments that are spread over many users.

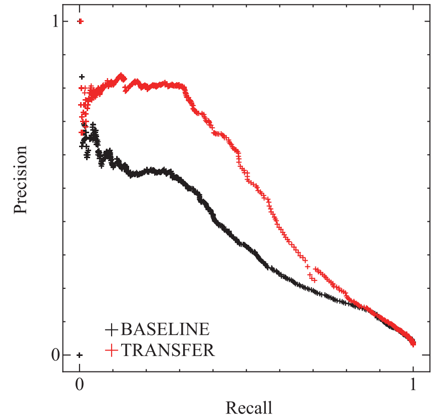
To see the effectiveness of transfer learning, we evaluate the performance of the Naïve Bayes classifier that has no knowledge transfer function as a baseline model. The baseline model learns only a domain dataset, while the proposed model learns not only a domain dataset but also other domain datasets by transferring the occurrence probabilities of words with high entropy (i.e., commonly used word over any domains). As discussed in II-C1, if a domain dataset has some bias in the word distribution (i.e., the frequency of specific words is distinctively high compared to other words), the proposed knowledge transfer could compensate for such a bias by transferring the word frequency information of other domains to a target domain. If such knowledge transfer works well, it is expected that the accuracy in sentiment polarity prediction should be improved. For the notational convenience, in the following, the baseline and the proposed model are denoted as BASELINE and TRANSFER, respectively.

To compare the performances of BASELINE and TRANSFER, we adopt a Precision-Recall (PR) Curve where horizontal and vertical axes correspond to the recall and the precision for negative comments. In a PR curve, if a curve exists upper right, it means that both precision and recall are high and the system has good performance in sentiment polarity prediction.

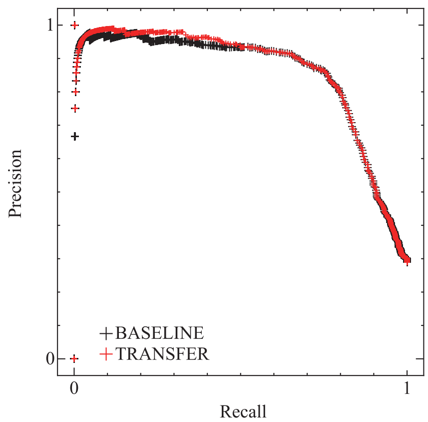
Figures 3(a)-(d) show the PR curves for the four datasets. The black and red points correspond to the average precision / recall in BASELINE and TRANSFER when the bias parameter b is fixed at a value that is changed from -300 to 300 with 0.1 intervals. Each point is obtained using the 10-fold cross-validation method. A dataset is first divided into 10 subsets,



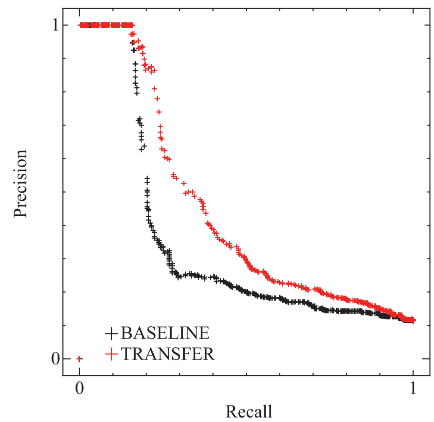
(a) Company A



(b) Company B



(c) Company C



(d) Company D

Fig. 3. PR curves of our method (red) and baseline (black).

and the 9 subsets are used for training and the remaining one subset is use for test. Using the 9 training subsets, M parameter sets $(\theta_k^*, b_k^*)(k = 1, \cdots, M)$ are obtained through the parameter estimation explained in II-D. Then, the precision and recall are calculated for the test subset. This iteration is repeated 10 times, and the average precision / recall in Figs. 3(a)-(d) are obtained. Note that there are two types of cross-validations: one for the performance evaluation and the other for the parameter estimation. This method is known as Nested Cross-Validation [17].

As seen in Figs. 3(a)-(d), for Datasets A, B, and D, the sentiment polarity prediction is clearly improved in both precision and recall by introducing the transfer learning. On the other hand, there is no clear effect of the knowledge transfer for Dataset C. This is because both precision and recall in BASELINE are relatively high for this dataset. Therefore, it is considered that there is little room for enhancement.

From the above experimental results, we can conclude that the proposed transfer learning contributes to the enhancement in the sentiment polarity prediction.

C. *Evaluation of Flaming Detection*

In this section, we verify whether the flaming of negative comments can successfully be detected by monitoring the transition of negative comments that are classified by the proposed Naïve Bayes classifier with transfer learning. To assume a practical environment, we train the proposed transfer learning model using the labelled comments that are collected one week before the end of the data collection (see Table I) and we test the performance for the last one week. The parameters θ and b are determined as the values of giving a break-even point (i.e., the point where the precision and the recall are equal), and these parameters are used for test.

Figures 4(a)-(d) illustrate the three time courses of the number of negative comments labelled by human (black solid line), the number of negative comments classified by the proposed system (red solid line), and the total number of comments on a target domain (black dotted line). Since we believe the correctness of human labelling, a black solid line is considered to be a true transition of negative comments. A black dotted line corresponds to a transition of comments on a target domain which includes non-negative comments as well as negative ones.

As seen in Fig. 4(a), a flaming event occurred around 6pm on June 23rd because the true number of negative comments are suddenly increasing. On the other hand, the red solid line is also increasing at the same time; therefore, we can say that the proposed system can detect the flaming event correctly. However, the proposed system wrongly detects a flaming event (i.e., false positive detection) around 10pm on June 28th. Then, we investigate the reason of the false potitive detection. As a result, we found that the following comment was misclassified as negative and it was spread to many twitter users as retweets: *"Company A suited Company Z that was owed to the product recall! X JPY compensation claim!."* The exact meaning of this comment is that a targeted Company A has sued another company that caused the product recall; thus, it was not a negative event for Company A. Since the proposed system takes a bug-of-word approach to sentiment polarity prediction, it is difficult to take into account the grammatical structure of comments, and the proposed system failed to predict correctly for which side of companies the words 'suit', 'recall', 'compensation' were used. However, this type of false positive detection is not a problem, rather it should be the case to make an alert to a human operator because the meanings of the above comment could be reversed in case Company A appears after Company X.

As seen in Fig. 4(b), two flaming events occurred around 8am and 10pm on June 24th, and the proposed system successfully detects these events. On the other hand, the dotted line has lot of peaks; thus, one can say that it is difficult to detect true flaming events only by monitoring the number of comments on a target domain. A flaming event occurred around 11pm on February 28th for Dataset C, and two flaming events occurred around 3pm on June 24th and 11am on June 25th for Dataset D are also successfully detected.

Currently, the detection of flaming events are usually conducted by human in a reputation management company, and they have to keep monitoring lots of comments for 24 hours, 365 days to protect a company's brand. In reality, however, a large part of such comments are not negative. Therefore, they are wasting time and money for monitoring. If the proposed system is deployed in real environments, the number of comments to be monitored would be significantly reduced, and it is also expected to reduce not only money but also human errors.

IV. Conclusions

In this paper, we develop a Naïve Bayes classifier model with a transfer learning function for accurate sentiment polarity prediction of SNS comments, and this model is applied to the detection of SNS flaming events. The knowledge transfer is conducted by transferring the occurrence probabilities of selected words in other domains to a target domain. In the transfer learning, for words with high entropy (i.e., commonly used words among different domains), the occurrence probability of a target domain is recalculated by averaging over the probabilities of other domains. If a domain dataset has some bias in the word distribution (i.e., the frequency of specific words is considerably higher than that of other words), it is expected that the proposed knowledge transfer would compensate for biased word distributions by transferring the word frequency information of other domains to a target domain.

To verify the effectiveness of the transfer learning, we conduct two experiments to evaluate the sentiment polarity prediction and flaming detection using the Japanese SNS comment datasets of four companies. The comments in each dataset were collected from various SNS sources such as Twitter and Facebook and they were labelled as either positive, neutral or negative by human operators. The experimental results demonstrate that the proposed transfer learning contributes to enhancing the sentiment polarity prediction especially when the classifier learning only with a target domain dataset has low performance. In addition, the proposed system can detect flaming events correctly for all companies without missing flaming events, although some false alarms are given.

There still remain several open problems. Considering the nature of SNS, it happens that a word distribution could still be strongly biased even after the transfer learning is conducted. This would cause poor sentiment polarity prediction and might potentially miss to detect important flaming events. Therefore, we consider that the flaming detection should not only rely on sentiment polarity prediction but also take other information into consideration such as the profiles of the users who give comments and the transitions in domains of comments [18]. These are left as our future work.

Acknowledgment

This research was partially supported by the Kayamori Foundation of Informational Science Advancement.

References

[1] P. D. Turney, "Thumbs Up or Thumbs Down?: Semantic Orientation Applied to Unsupervised Classification of Reviews," *Proc. of the 40th Annual Meeting of the Association for Computational Linguistics*, pp. 417–424 (2002)

[2] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs Up?: Sentiment Classification Using Machine Learning Techniques," Proc. of the ACL-02 Conf. on Empirical Methods in Natural Language Processing, Vol. 10, pp. 79–86 (2002)

[3] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis," *Proc. of the Conf. on Human Language Technology and Empirical Methods in Natural Language Processing*, pp. 347–354 (2005)

[4] A. Abbasi, H. Chen, and A. Salem, "Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums," *ACM Trans. on Information Systems*, Vol. 26, Issue 3, No.12 (2008)

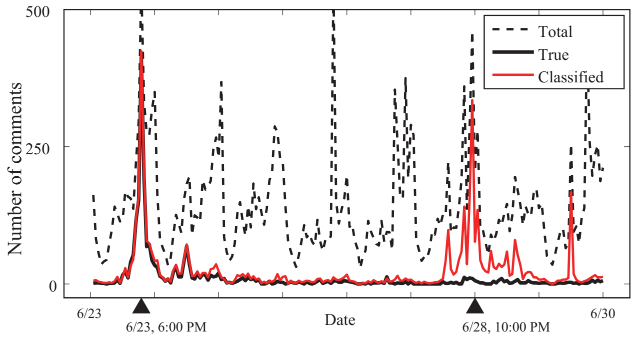
[5] H. Takamura, T. Inui, and M. Okumura, "Extracting Semantic Orientations of Words Using Spin Model," *Proc. of the 43rd Annual Meeting of the Association for Computational Linguistics*, pp. 133–140, 2005.

[6] A. Go, R. Bhayani, and L. Huang, "Twitter Sentiment Classification Using Distant Supervision," CS224N Project Report, Stanford, pp. 1–12 (2009)

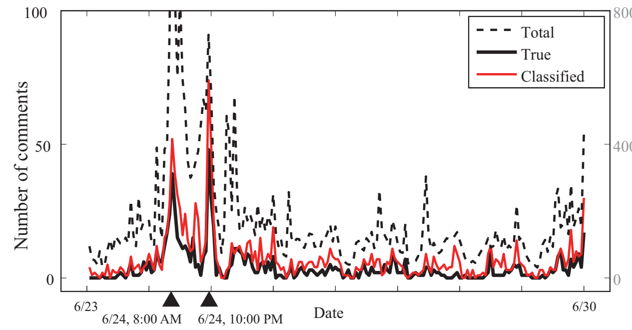
[7] A. Pak and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining," Proc. of the Int. Conf. on Language Resources and Evaluation, pp. 1320–1326 (2010)

[8] J. Spencer and G. Uchyigit, "Sentimentor: Sentiment Analysis of Twitter Data," *Proc. of the 1st Int. Workshop on Sentiment Discovery from Affective Data*, pp. 56–66 (2012)

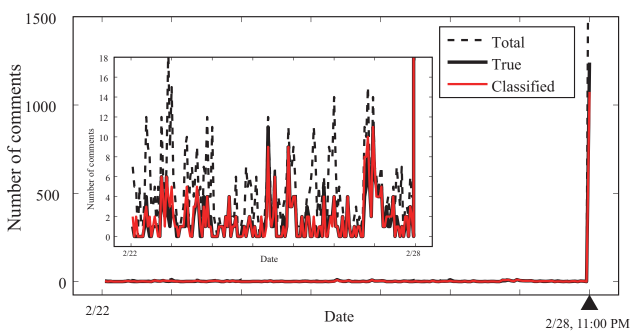
[9] J. Blitzer, M. Dredze, and F. Pereira, "Biographies, Bollywood, Boomboxes and Blenders: Domain Adaptation for Sentiment Classification," Proc. of the 45th Annual Meeting of the Association for Computational Linguistics, pp. 440–447 (2007)



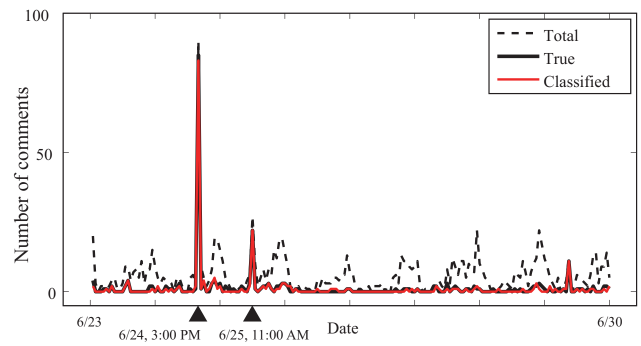
(a) Company A



(b) Company B



(c) Company C



(d) Company D

Fig. 4. Time course of the number of comments. Black and red lines indicate the true and the estimated numbers of negative comments, respectively. Dashed line indicates the total number of comments. The vertical scale in right-hand side of (b) represents the total number of comments. The inset in (c) is a zoom-in view.

[10] P. Melville, W. Gryc, and R. D. Lawrence, "Sentiment Analysis of Blogs by Combining Lexical Knowledge with Text Classification," *Proc. of the 15th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, pp. 1275–1284 (2009)

[11] S. J. Pan, X. Ni, J. Sun, Q. Yang, and Z. Chen, "Cross-Domain Sentiment Classification via Spectral Feature Alignment," *Proc. of the 19th Int. Conf. on World Wide Web*, pp. 751–760 (2010)

[12] S. Yoshida, J. Kitazono, S. Ozawa, T. Sugawara, T. Haga, "Improving the Accuracy of Sentiment Analysis of SNS Comments Using Transfer Learning and Its Application to Flaming Detection," (in Japanese) IEEJ Transactions on Electronics, Information and Systems, Vol.135 No.1 pp.1-8 (2016)

[13] T. Kudo, K. Yamamoto, and Y. Matsumoto, "Applying Conditional Random Fields to Japanese Morphological Analysis," *Proc. of the 2004 Conf. on Empirical Methods in Natural Language Processing*, pp. 230–237 (2004)

[14] T. Kudo and Y. Matsumoto, "Japanese Dependency Analysis using Cascaded Chunking," *Proc. of the 6th Conf. on Natural Language Learning 2002*, pp. 63–69 (2002)

[15] A. McCallum and K. Nigam, "A Comparison of Event Models for Naive Bayes Text Classification". *AAAI-98 Workshop on Learning for Text Categorization*, pp. 41–48 (1998)

[16] S. Kim, K. Han, H. Rim, and S. Myaeng, "Some Effective Techniques for Naive Bayes Text Classification," *IEEE Trans. on Knowledge and Data Engineering*, Vol. 18, Issue 11, pp. 1457–1466 (2006)

[17] S. Varma and R. Simon, "Bias in Error Estimation When Using Cross-validation for Model Selection," BMC Bioinformatics, Vol. 7, No. 1, p. 91 (2006)

[18] N. Awaya, J. Kitazono, T. Omori, and S. Ozawa, "Stochastic Collapsed Variational Bayesian Inference for Biterm Topic Model," Proc. of Int. Joint Conf. on Neural Networks 2016.