VOL 2 (2018) NO 4 - 2

e-ISSN : 2549-9904
ISSN : 2549-9610

INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

Estimation of Danger Signs in Regional Complaint Data

Yao Lin[#], Tsunenori Mine[#], Kohei Yamaguchi[#], Sachio Hirokawa[#]

[#] *Department of Advanced Information Technology, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka, 819-0395, Japan*
E-mail: yyays20130220@gmail.com, mine@ait.kyushu-u.ac.jp, yamaguchi.kohei.133@m.kyushu-u.ac.jp, hirokawa@cc.kyushu-u.ac.jp

Abstract— **Government 2.0 activities have become very attractive and popular. Using the platforms to support the activities, anyone can anytime report issues in a city on the Web and share the reports with other people. Since a variety of reports are posted, officials in the city management section have to give priorities to the reports. However, it is not easy task for the officials to judge the importance of the reports because importance judgments vary depending on the officials, and consequently the agreement rate becomes low. To remedy the low agreement rate problem of human judgment, it is necessary to create an intelligent agent which supports finding reports with high priorities. Hirokawa et al. employed the Support Vector Machine (SVM) with a word Feature Selection method (SVM+FS) to detect signs of danger from posted reports because the signs of danger is one of high priority issues to be dealt with. However they did not compare the SVM+FS method with other conventional machine learning methods and it is not clear if the SVM+FS method has better performance than the other methods. This paper explores methods for detecting the signs of danger through comprehensive experiments to develop an intelligent agent which supports officials in the city management sections. We explores conventional machine learning methods: SVM, Random Forest, Naïve Bayse using conventional word vectors, an LDA-based document vector, and word embedding by Word2Vec and compared the best method with SVM+FS. Experimental results illustrate the superiority of SVM+FS and invoke the importance of using multiple data sets when evaluating the methods of detecting signs of danger.**

Keywords— **Complement Report, Signs of Danger Detection, Government 2.0, Machine Learning.**

I. INTRODUCTION

Government 2.0 is the concept proposed by Tim O'Reilly, which means government as a platform, where citizens are encouraged to participate government activities. For example, FixMyStreet.com (https://www.fixmystreet.com) is a website which enables anyone to reporting street problems. Even in Japan, some platforms such as the Chiba citizen coordination report (https://chibarepo.secure.force.com) (ChibaRepo for short) have been established. Traditionally, it is necessary for citizens to call or visit officials in the city management section when reporting a complaint. The officials have to record and deal with it after checking its situation on site. As a result, it takes time for the officials before they start dealing with a complaint report after receiving it. While, using such the platforms, citizens are enabled to publish complaints that they hold about the region anytime and anywhere. They can report complaint with the location data and image data related to the complaint on the Web, check the correspondence situation of the the city management section side and argue with others about solving the problem. For the city management section side, it is convenient to check the situation from the data reported by citizens.

On the other hand, since a variety of reports are posted, officials in the city management section have to give priorities to the reports. However, it is not an easy task for the officials to judge the importance of the reports because importance judgments vary depending on the officials and consequently the agreement rate becomes low. This causes the delay of taking action by the local government. Therefore, it is indispensable to reduce the burden on the government side by developing an intelligent agent which supports automatic or semi-automatic judgment of the urgency to deal with complaint reports.

To realize the agent, we aim to find out a better method of detecting signs of danger in ChibaRepo. The danger is one of urgent issues in the city to be dealt with by the officials with the highest-priority. Actually, lots of researchers have studied about detection of emergency events such as disaster or criminal offense events from micro blogs or social networking services (e.g.[6],[4]).

Dealing with the signs of danger with high-priority can prevent citizens from facing accidents or emergency events forecasted by the signs. Hirokawa et al. employed Support Vector Machine (SVM)[8] with word Feature Selection (SVM+FS for short)[12],[1] to detect the signs of danger and achieved higher performance than the average performance of four human subjects[5].

However, since they only used SVM+FS in their paper, it can not be judged whether or not the problem of detecting signs of danger is a difficult one to be dealt with by machine learning methods in the first place, in other words, whether or not SVM+FS is the best method. In fact, effective machine learning methods change depending on the analysis target, so it is essential to compare with other methods. Therefore, this research aims to confirm if there are any better methods than SVM+FS. To this end, we conducted comprehensive experiments to compare SVM+FS with several machine learning methods such as SVM, Random Forest (RF)[3], and Nai"ve Bayse (NB)[10] using other feature selections than word feature selection such as part of speech, sentiment polarity words, Latent Dirichlet Allocation (LDA)[2] as a topic model, and Word2Vec (W2V)[11] as a word embedding method. We used two data sets in the experiments. Experimental results illustrate that SVM+FS was the best except only one case, which did not show any significant difference, and invoke the importance of using multiple data sets to evaluate methods for detecting signs of danger.

In what follows, Section 2 describes the literature of detecting emergency events and discusses the position of this research. Section 3 explains data set we used in this paper. Section 4 explains the methods we used in the experiments. Section 5 explains the way of experiments conducted, illustrates experimental results obtained by the methods, and discusses the experimental results. Finally we conclude the paper in Section 6.

II. RELATED WORK

Lots of researchers have studied about detection of emergency events such as disaster or criminal offense events from micro blogs or information network such as Twitter or social networking services such as Facebook.

Imran et al.[7] presented a system to automatically extract information nuggets from microblogging messages during disaster times. They divided the tweets into four types and extracted information from the tweets. Then they used the NB classifier implemented in Weka.

Sano et al.[13] suggested a method for classifying the category of complaint reports. They used Mutual Information with Term-Frequency Inverse-Document-Frequency (TFIDF) term weighting to weight the words and used RF as a classifier. They found that using data which have similar tendency as training data may increase the accuracy.

Hirokawa et al.[5] employed SVM+FS to estimate the danger signs of ChibaRepo. They asked four subjects to judge if there is a danger sign in a report and used the judgements given by the four subjects as gold standard. They used category information in ChibaRepo to tag words, and calculated a *svm-score* of each word to select words for feature selection, and used SVM^{light}[9] as a classifier. As a result, they achieved a higher result than human to satisfy decision by majority of agreements among human subjects. However, since Hirokawa et al. only used SVM+FS in their paper, it cannot be judged whether or not SVM+FS is the best method. There are many methods to make vectors. For example, Word2Vec[11], which is a distributed implementation, and LDA[2], which is a topic model. There are also many machine learning methods which sometimes achieve better results than SVM[8]. In fact, effective machine learning methods vary depending on the analysis target, so it is essential to compare with other methods. Therefore, in this research, we try to find out if there is a better method than SVM+FS by comparing with other machine learning methods using another data set in addition to the data set Hirokawa et al. [5] used.

III. DATA SET

A. *Input Datat*

ChibaRepo is a platform to enable citizens to report a variety of problems in the city using ICT, and enable the citizens and officials in the city management sections share and try to solve the problems. On the ChibaRepo Web site, citizens have issued 5,139 reports so far; 4,835 have already been dealt with, 50 is now being tackled, and 254 are in a waiting list. Among them, 1873 reports being in CSV format are open to the public (confirmed on June 29, 2018). Here each report data consists of 19 categories, which are summarized in Table 1.

Hirokawa et al.[5] used the data of ChibaRepo's 656 reports (CB656 for short), which were gathered by crawling the ChibaRepo Web site before the current open data was released. Then, CB656 does not include some latest reports in the ChibaRepo open data mentioned above.

They asked four subjects to read the report in the data and to put a mark on a report if signs of danger were included in the report. 36% reports in the whole were judged as reports including signs of danger (`danger report' for short), by at least one subject. Table 2 shows the numbers of danger reports in CB656 and their percentages in the whole judged by *N* or more subjects, where *N* varies from 1 to 4 in CB656.

TABLE I
DATA CATEGORIES OF CHIBAREPO OPEN DATA

Categori_Name	Comment
Address_c	Address
CBC_M_Sections_c	Section related to Address
CBC_M_WebUser_c	User ID
Category_c	Category ∈ {Road, Park, Garbage, Others }
Comment_c	Text message in a Report
CompleteDate_c	Date of completing the issue
CopeImage1Id_c	Photo image ID
CopeImage12d_c	Photo image ID
Image1Id_c	Photo image ID
Image2Id_c	Photo image ID
Image3Id_c	Photo image ID
LatitudeWGS84_c	latitude
LongitudeWGS84_c	longitude
ReportDateTime_c	Date when a report was issued
Status_c	State of Correspondence ∈ {Completed, In process, Unresponsive (Received, but not began)}
Subject_c	Report subject, Title
VideoURL_c	URL of video

TABLE II
PERCENTAGE OF DANGER REPORTS JUDGED BY N OR MORE SUBJECTS IN CB656[5]

N	count	percentage	comments
1	235	0.36	at least one subject judged as danger report
2	111	0.17	at least two subjects judged as danger report
3	66	0.10	three or four subjects judged as danger report
4	22	0.03	all the four subjects judged as danger report

In addition to CB656, in this paper, we use the ChibaRepo open data including 1873 reports (CB1873 for short). In CB1873, for each report, we asked five subjects, who were different from the subjects of CB656, to judge if it includes the signs of danger. The danger degree of each report, which is the number of subjects judged it as a danger report, ranges from 0 to 5 in CB1873. Table 3 shows the numbers of danger reports in CB1873 and their percentages in the whole judged by the five subjects. The agreement rate of detecting sings of dange in CB1873 is greater than that in CB656. However as the number of N increases, the agreement rate of at least *N* subjects tends to become decreased. This tendency of CB1873 is the same as that of CB656.

TABLE III
PERCENTAGE OF DANGER REPORTS JUDGED BY N SUBJECTS IN CB1873

N	count	percentage	comments
1	1456	0.78	at least one subject judged as danger report
2	1039	0.55	at least two subject judged as danger report
3	583	0.31	at least three subject judged as danger report
4	388	0.21	four or five subjects judged as danger report
5	179	0.10	all the five subjects judged as danger report

B. Data Tagging

To assign each word a tag, which is the name of a category where the word appeared, we used 5 categories: Comment_c, Subject_c, Status_c, Category_c, and CBC_M_Sections_c as Hirokawa et al. [5] did. Furthermore, in order to examine the effect of other tag data, we used the Japanese sentiment polarity dictionary (http://www.cl.ecei.tohoku.ac.jp/index. php?Open%20Resources%2FJapanese%20Sentiment%20Polarity%20Dictionary) to make tags, which contains about 8,500 nouns and 5,000 declinable words. The details of using tags are described later.

IV. EXPERIMENTAL METHODS

We applied several machine learning methods to the reports, built models to judge if a report is an danger report and compared the discrimination performance of the models with the model built by the SVM+FS method. For the comparison, we used two data sets: CB656 and CB1873.

A. SVM+FS

SVM+FS carries out word feature selction according to a *svm-score* of each word in the documents to be classified. The *svm-score* is calculated as follows:

1. Let D be a set of N documents, which are classified into M classes, where M=2 in this research.
2. When a document $d_i \in D$ ($1 \leq i \leq N$) includes n distinct words, SVM+FS produces n one-word documents from d_i, where each one-word document $d_{i,j}$ only includes one word $w_{i,j}$ ($1 \leq j \leq n$).
3. For each target class, SVM+FS assigns $d_{i,j}$ a positive flag if d_i belongs to the target class, otherwise assigns a negative flag.
4. SVM+FS converts each document $d_{i,j}$ to a word vector $vec(d_{i,j}) = \{v_1,\ldots,v_k,\ldots, v_m\}$, where m is the total number of distinct words in D, $v_k \in \{1, -1\}$ if $w_{i,j}$ corresponds to kth element in the vector, and $v_i = 0$ ($i \neq k$). If $d_{i,j}$ has a positive flag, v_k is set to 1; otherwise, v_k is set to -1.
5. SVM+FS builds a model using SVM from a set of word vectors produced in step 4, and obtains a score of each word calculated by the SVM model. We call this score *svm-score*. SVM+FS selects top K positive and top K negative words based on the *svm-score* of the words.
6. SVM+FS converts documents into input vectors only using the 2 K positive and negative words selected in step 5; if the input vector's element correspoinds to a word in the 2 K words, the value of the element is 1, otherwise 0.
7. SVM+FS builds a classification model using SVM from the input vectors produced in step 6

Please note that SVM+FS does not use parts of speech information of words in the data set and not remove stop words either. When applying SVM to calculate *svm-score*, we used default parameters of SVM.

B. Settings

1) Category Selection and Creating Tagged Words

We provided three cases for experiments.

- In **case 1**, we only used the Comment_c data to create input vector. Comparing the results in **case 2**, we can evaluate the effects of the categories.
- In **case 2**, we used the five category data. We denoted the five categories by "cOntent", "Title", "corresponDing state", "Genre", and "Region", and used one character in each category as a tag to distinguish the same word, say `danger,' appearing in different categories, such as "O:danger" in "cOntent" and "T:danger" in "Title."
- In **case 3**, we used sentiment polarity word tags from the Japanese Sentiment Polarity Dictionary in addition to the five category data used in case 2. Tags of "Negative" and "Positive" were added to the words used in the case 2. For example, if word "danger", which is a negative word by the Japanese Sentiment Polarity Dictionary, is appeared in "Title" category, we added **N**egative and **T**itle tag "NT" to "danger" and describe "NT:danger. "

We used parts of speech features, which automatically removes stop words.

2) Input Vector Creation

In all the three cases, we used one-hot bag-of-words (BoW for short), TFIDF, Word2Vec (W2V), and Latent Dirichlet A1llocation (LDA). When using W2V, we used other text data: the four year records of complaint calls from citizens about city parks in Kashiwa City (Kashiwa data, for short) to increase the text volume. We examined three data sets: 1) CB656, 2) CB1873, and 3) Kashiwa data consisting of 5665 reports. We combined and created the three data sets: 1)+2), 1)+3), 1)+2)+3).

3) Classifiers

In all the three cases, we applied the same three classifiers: RF, SVM and NB to input vectors. We used scikit-learn library to determine parameter values of each machine learning classifier. Since the numbers of positive and negative examples were imbalanced, we took **class_weight = "balanced"** option for RF and SVM and default values for other parameters. For NB, we took GaussianNB and used default parameters for the rest.

4) Experiment Procedure

Morphological Analysis: Since words in a Japanese sentence are not separated with each other, first, we performed morphological analysis using a Japanese morphological analyzer MeCab (http://taku910.github.io/mecab/) and extracted words and their parts of speech from sentences in the data set. Second, we normalized sentences by transforming several 2-byte characters such as alphanumeric characters, signs and spaces into 1-byte characters and 1-byte Japanese katakana characters into 2-byte characters, and deleting spaces and line breaks before and after the sentence. Then we extracted words with some specific parts of speech: adjectives, auxiliary verbs, verbs, and nouns, and reconstructed the sentences using the extracted words.

Input Vectors Creation: In **case 1**, we used the original form of a word in constructing each input vector by using BoW, TFIDF, W2V and LDA. In **case 2**, we added the category tag in front of the original form of a word and made input vectors as well as case 1. In **case 3**, we used the Japanese Sentiment Polarity Dictionary to judge a word's sentiment polarity: Positive or Negative, and added the tag of sentiment polarity in front of the word created in case 2, and made input vectors as well as case 1.

Input Vectors by BoW: We vectorized a report d_i as a vector $vec(d_i)=(s_1, \ldots, s_M)$, where M is the total number of words appeared in the reports and $s_j = 1$ if word w_j appears in d_i.

Input Vectors by TFIDF: We calculated the TFIDF value of word w_j in a report d_i as s_j and vectorize the report as a vector $vec(d_i)=(s_1, \ldots, s_M)$.

Input Vectors by W2V: In this research, we used the W2V library offered by gensim to make a model from the sentences consisting of separated words, using the default parameter of 200 dimensionality. From the model, the vector of a word w_i can be represented as

$vec(w_i)=(v_1, \ldots, v_{200})$

, where v_i is the *i*th feature value calculated by the W2V model. The vector of a report d_i can be represented as the average of vectors of words in d_i as $vec(d_i)= \sum_{i=1,n} (vec(w_i) / n)$. $vec(d_i)$ and n denote a report vector and the number of words appeared in the report, respectively.

Input Vectors by LDA: We used LDA to perform dimension reduction of an input vector; the number of dimensions before the dimension reduction was 2483 and 200 after performing the reduction. The vector of a report $vec(d_i)$ can be represented as $vec(d_i)=(s_1, \ldots, s_{200})$, where s_i denotes a feature value.

C. Pre-experiments by SVM+FS

We conducted pre-experiments using SVM+FS to reevaluate its performance. We used the 5-category data and used BoW to make vectors. The pre-experiments were performed using two data sets: CB656 and CB1873. First, we calculated the *svm-score* of words by SVM^{perf}. Then, we chose the top K positive words and top K negative words to reconstruct sentences. After that, we used LibSVM for classification. The procedures are described below.

1) Preprocessing

First, we performed Morphological analysis described in Section IV.B.4. Second, we tagged the words by one character of a category in which the word appeared. For example, the word "T:road" means the word "road" which appeared in the category "Title." Then, we reconstructed the sentences by the tagged words.

2) Feature Selection

We calculated *svm-score* by SVM^{perf} and chose top K positive words and top K negative words as feature selection. Here we chose one value as K from {10, 20, 30, ..., 100} to conduct the pre-experiments.

V. Experimental Results

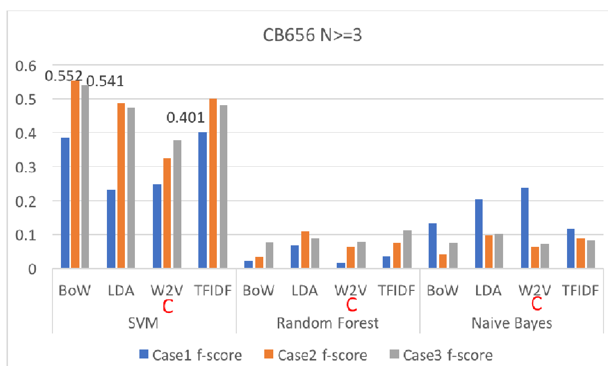

Fig. 1 Comparison of three cases in $N \geq 3$ of CB656

In order to compare the results of the SVM+FS method with other machine learning methods, we conducted experiments in the three cases by using the two data sets: CB656 and CB1873. In the experiments, we performed 10-fold cross validation for ten times and took the average. Comparison results are shown below. Since we focus on the decision by majority, we only show the experimental results on the cases of $N \geq 3$ and $N \geq 4$ for CB656 in Figures 1 and 2, and those of N ≥ 3, N ≥ 4 and N ≥ 5 for CB1873 in Figures 3, 4 and 5, respectively.

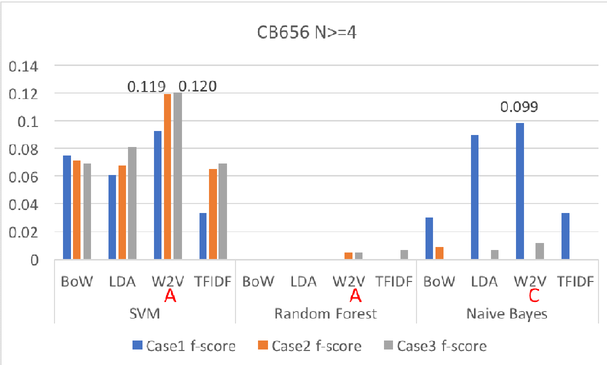

Fig. 2 Comparison of three cases in N ≥ 4 of CB656

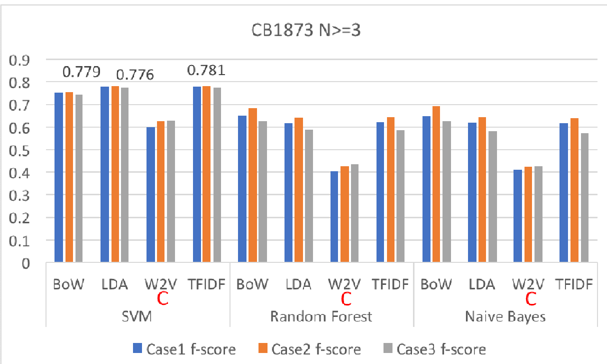

Fig. 3 Comparison of three cases in N ≥ 3 of CB1873

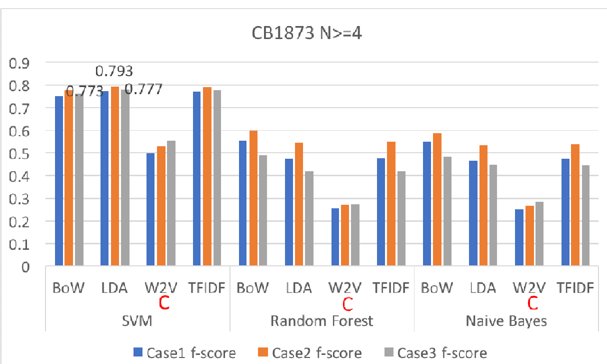

Fig. 4 Comparison of three cases in N ≥4 of CB1873

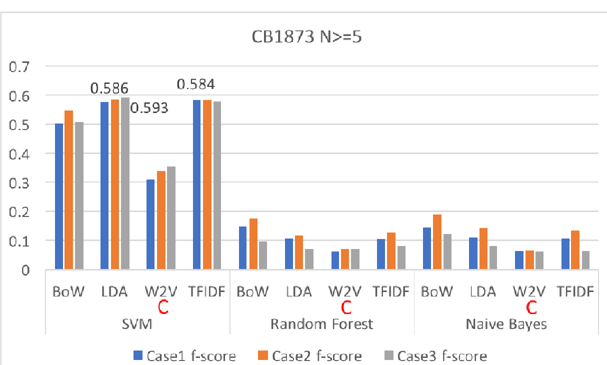

Fig. 5 Comparison of three cases in N ≥ 5 of CB1873

Results of Pre-experiments

We applied SVM+FS to the two data sets: CB656 and CB1873. We used K ={10, 20, ..., 100} to conduct experiments. The best results by K are shown in Table 4. The number in brackets denotes the value of K which took the best result.

TABLE IV
PRE-EXPERIMENT RESULTS BY SVM+FS

N	3	4	5
CB656	0.629(30)	0.355(10)	
CB1873	0.807(30)	0.810(10)	0.583(10)

Results in Case 1

In case 1, we only used the Comment_c data to perform experiment. As shown in Figures 1 and 2, F-scores of SVM+TFIDF and NB+W2C are the best in N ≥ 3, which are 0.401, and N ≥ 4, which are 0.099, respectively.

As shown in Figures 3, 4, and 5, F-score of SVM+LDA in CB1873 is the best in N ≥ 3, which are 0.779, while SVM+TFIDF is the best in N ≥ 4 and in N ≥ 5, which are 0.773 and 0.577, respectively.

However, in this case, the results of the SVM+FS method outperformed in both N ≥ 3, N ≥ 4 and N ≥ 5. For W2V, although we added another data: the four year records of complaint calls from citizens about city parks in Kashiwa city, to increase the data volume, the results of the three methods with W2V were worse than those of SVM with BoW.

Results in Case 2

In case2, we used the five categories: ChibaRepo's Comment_c, Subject_c, Status_c, Category_c, and CBC_M_Sections_c's, which means content, title, corresponding state, genre and section. The comparison results are shown below.

In CB656, F-scores of SVM+BoW and SVM+W2C are the best in N ≥ 3 and N ≥ 4, which are 0.552 and 0.119, respectively.

In CB1873, F-score of SVM+TFIDF is the best in N ≥ 3, while SVM+LDA is the best in N ≥ 4 and N ≥ 5, which are 0.781, 0.793 and 0.586, respectively.

Results in Case 3

In case 3, we used sentiment polarity word tags: "Negative" and "Positive" from the Japanese Sentiment Polarity Dictionary in addition to the five category data in case 2. In CB656, F-score of SVM+BoW is the best in N ≥ 3, which are 0.541, while SVM+W2C is the best in N ≥ 4, which are 0.120. In CB1873, F-score of SVM+LDA is the best in all of N ≥ 3, N ≥ 4 and N ≥ 5, which are 0.776, 0.777 and 0.593, respectively.

With the results in the three cases, we found that the average F-score values improved by using tagged data. We also found that different data has a different optimal method.

Comparison with SVM+FS and Human Subjects

Now we show comparison results between our proposed methods which took the best in each case, SVM+FS, and human subjects for N ≥ 3 and N ≥ 4 in CB656 in Figures 6 and 7, for N ≥ 3, N ≥ 4 and N ≥ 5 in CB1873 in Figures 8, 9, and 10, respectively.

In CB656, human subjects took the best results and SVM+FS was the second. On the other hand, in CB1873, SVM+FS took the best for N ≥ 3 and N ≥ 4, but SVM+LDA in case 3 took the best for N ≥ 5 although the differences were very small.

Interestingly, all the methods took the better results than human subjects in CB1873.

For CB1873 in N ≥ 5, because the differences of the results among all methods were not large, we performed F-test and t-test to examine if there are any significant differences in the results.

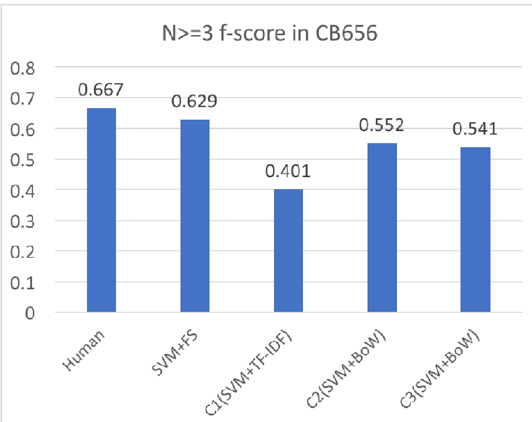

Fig. 6 Comparison in CB656 with SVM+FS in N ≥ 3

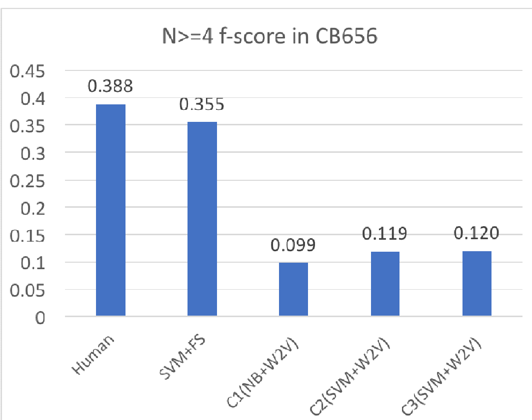

Fig. 7 Comparison in CB656 with SVM+FS in N ≥ 4

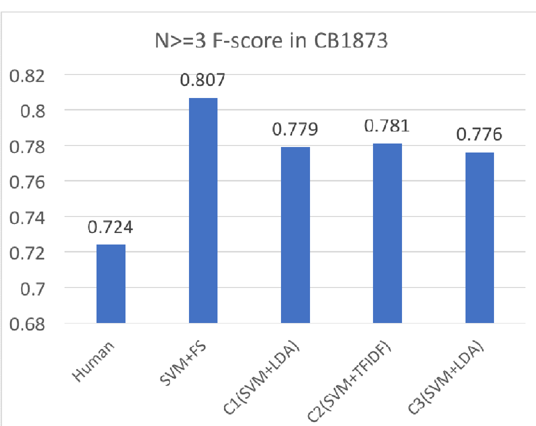

Fig. 8 Comparison in CB1873 with SVM+FS in N ≥ 3

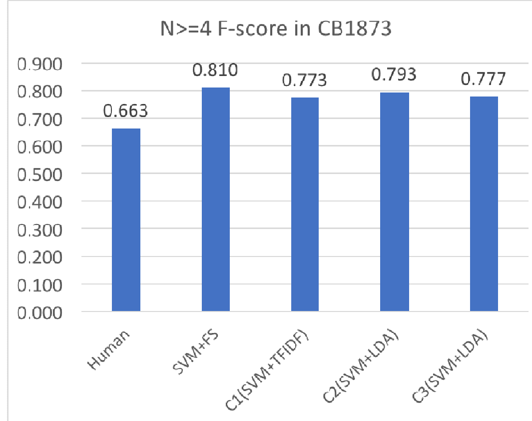

Fig. 9 Comparison in CB1873 with SVM+FS in N ≥ 4

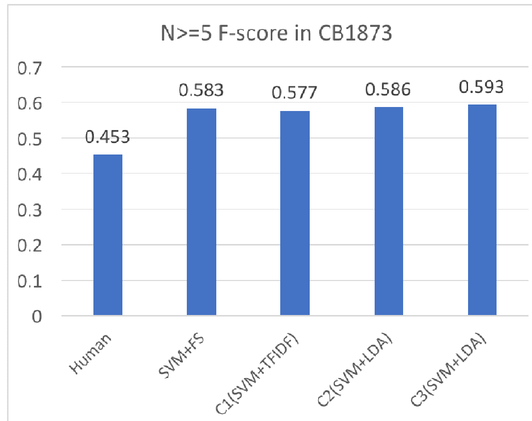

Fig. 10 Comparison in CB1873 with SVM+FS in N ≥ 5

The results of F-test and t-test are shown in Table 5. It can be seen that the difference between SVM+FS and SVM+LDA in case 3 obtained by using t-test is more than 0.05, which means there are almost no significant differences between the two methods.

TABLE V
RESULTS OF F-TEST AND T-TEST IN CB1873 (N≥5)

	FS&C1	FS&C2	FS&C3
F-test	0.124	0.011	0.009
t-test	0.901	0.654	0.120

VI. CONCLUSION

In this paper, we discussed comprehensive experiments to compare the SVM+FS method with several machine learning methods such as SVM, RF, and NB using other feature selections than word feature selection such as parts of speech, sentiment polarity words, LDA as a topic model, Word2Vec as a word embedding method.

We used two data sets: CB656 and CB1873, to conduct experiments considering the three cases to evaluate the effects of five categories and of the sentiment polarity words. In case 1, we only used the Comment_c data to create input vector by BoW, TFIDF, Word2Vec and LDA. Then we applied RF, SVM and NB to the vectors. We confirmed the effects of tagged words of five categories and sentiment polarity information in cases 2 and 3. We found that using tag data may increase the F-score.

Through the experiments, we have just used a few machine learning and vectorized methods. We will continue to perform experiments using other machine learning methods, vectorized methods and tags so as to find better results and develop an intelligent agent which can use the better methods for detecting the signs of danger. For example, we use bigger corpora to build Word2Vec models and provide a better method for using sentiment polarity information. Applying under and over sampling methods are also good alternatives. These are our future work.

ACKNOWLEDGMENT

This work was partially supported by JSPS KAKENHI Grant No. JP15H05708, JP16H02926, JP17H01843 and JP18K18656.

REFERENCES

[1] Adachi, Y., Onimura, N., Yamashita, T., Hirokawa, S.: "Standard measure and svm measure for feature selection and their performance effect for text classification." Proc. of the 18th International Conference on Information Integration and Web-based Applications and Services. pp. 262--266. 2016

[2] Blei, D.M., Ng, A.Y., Jordan, M.I.: "Latent dirichlet allocation." Journal of machine Learning research 3(Jan), pp. 993—1022, 2003

[3] Breiman, L.: "Random forests." Machine learning 45(1), pp.5—32, 2001

[4] Cresci, S., Cimino, A., Dell'Orletta, F., Tesconi, M.: "Crisis mapping during natural disasters via text analysis of social media messages." Proc. of International Conference on Web Information Systems Engineering. pp. 250--258. 2015

[5] Hirokawa, S., Suzuki, T., Mine, T.: "Machine learning is better than human to satisfy decision by majority." Proc. of Web Intelligence 2017. pp. 694—701, 2017

[6] Imran, M., Castillo, C., Diaz, F., Vieweg, S.: "Processing social media messages in mass emergency: A survey." ACM Computing Surveys (CSUR) 47(4), 67, 2015

[7] Imran, M., Elbassuoni, S.M., Castillo, C., Diaz, F., Meier, P.: "Extracting information nuggets from disaster-related messages in social media." Proc. of the 10th International Conference on Information Systems for Crisis Response and Management, 2013

[8] Joachims, T.: "Text categorization with support vector machines: Learning with many relevant features." Machine Learning: ECML-98. pp. 137—142, 1998

[9] Joachims, T.: "Learning to Classify Text Using Support Vector Machines." The Kluwer International Series in Engineering and Computer Science, Springer US, 2002

[10] McCallum, A., Nigam, K., et al.: "A comparison of event models for naive bayes text classification." Proc. of AAAI-98 workshop on learning for text categorization. vol. 752, pp. 41--48. 1998

[11] Mikolov, T.: "Statistical language models based on neural networks." Presentation at Google, Mountain View, 2nd April, 2012

[12] Sakai, T., Hirokawa, S.: "Feature words that classify problem sentence in scientific article." Proc. of the 14th International Conference on Information Integration and Web-based Applications & Services. pp. 360—367, 2012

[13] Sano, Y., Yamaguchi, K., Mine, T.: "Automatic classification of complaint reports about city park." Information Engineering Express 1(4), pp.119—130, 2015, [Online] Available: http://www.iaiai.org/journals/index.php/ IEE/article/view/35