Original Article

A study on extracting disaster information from tweets

Toshihiko Shimauchi (Department of Regional Design and Development, Komatsu College, shimauchi@komatsu-c.ac.jp)
Naoki Taguchi (Mitsubishi Electric Engineering Company, taguchi0313123@gmail.com)
Hidetaka Nambo (Graduate school of Natural Science & Technology, Kanazawa University, nambo@blitz.ec.t.kanazawa-u.ac.jp)
Haruhiko Kimura (Department of Regional Design and Development, Komatsu College, hkimura@komatsu-c.ac.jp)

Abstract
In Japan, where natural disasters occurs frequently, obtaining and delivering accurate information promptly when a disaster occurs is essential to minimize damage. Information from traditional mass media contain a number of general information unrelated to disaster, so there are limitations in delivering necessary information to the resident in affected area. On the other hand, Twitter, one of the popular social media, is expected to play an important role during disaster because of its simplicity, promptness and wide propagation. However, because of its huge size of users, there are too many tweets which hinders timely extraction of relevant information. Disaster information is also useful for business travellers and tourists. They are less informed about the area and the challenge is to provide them with accurate information promptly. Our study proposes to establish a system to assist real time understanding of disaster by extracting relevant information efficiently from messages tweeted during two typhoons. First, binary classification is applied to classify and extract disaster tweets from tweets group. By using BNS method, the improvement in accuracy is confirmed. Then clustering is applied to the disaster tweets. The tweets are classified by 15 clusters generated. The result yields F measure of 0.59.

Keywords
disaster tweets, disaster prevention, morphological analysis, clustering, classification

1. Introduction

For its physical geography and location along the Pacific Ring of Fire, Japan is highly susceptible to natural disasters. [CAO, 2010]. Recently, there is a clear trend to indicate the increase in the occurrence of short-time heavy rain [JMA, 2017]. Based on these facts and to minimize the possible damage from disasters, the importance of obtaining and delivering accurate information promptly when a disaster occurs has been pointed out. However, traditional mass media such as newspapers and television transmit general information not related to natural disasters. So there are limitations in delivering necessary information to the residents in the affected areas [Toriumi et al., 2013]. On the other hand, Twitter, one of the popular social media, is expected to play an important role in disaster because of its simplicity, promptness and wide propagation. In the Great East Japan Earthquake of 2011, twitter played an important role in exchanging information among the users. A huge amount of information were transmitted through the Twitter, such as the tweets from users with on-site knowledge, which the traditional mass media could not report in timely fashion [Matsumoto et al., 2015]. However, because of its huge size of the users, there are too many tweets circulating on the internet which hinders timely extraction of useful information. In fact, on March 11, 2011 alone, when the Great East Japan Earthquake hit Japan, there were about 33 million tweets transmitted [Rokuse et al., 2015].

Natural disasters affect not only the residents but also business and non-business travellers visiting the areas. In 2016, the number of inbound tourists visiting Japan exceeded 20 million and developing a system to provide disaster information to these visitors is one of the policy priorities receiving more attention than ever before [Japan Tourism Agency, 2013]. However, there are little, if not any, discussions and considerations for the provision of such information to the Japanese travellers doing domestic travel [Nakatani, 2016]. In 2015, the number of domestic tourists are about 504 million guest nights with Japanese national accounting for 438.5 million [Japan Tourism Agency, 2016]. One study focused on the relation between tourists and disaster information and concluded that there is a lack of information on danger zones in the area where tourists are visiting temporarily [Akita Prefecture, 2015]. It is essential to provide the visitors who do not have sufficient on-site knowledge with accurate information in timely manner in order to minimize the damage and to prevent social unrest/disorder when a disaster occurs.

Our study focuses on typhoon, one of the most common natural disasters in Japan, in order to develop a system to assist real time understandings for current situation by efficiently extracting relevant information from tweets during typhoons. Specifically, the authors propose a method to classify tweets into two group: disaster tweets and non-disaster tweets. Furthermore, clustering is applied to the disaster tweets in order to develop a method to collect relevant information efficiently during disasters.

There are many existing studies on classifying huge amount of data in tweets: a study to extract events from twitter [Ritter et al., 2010]; a study to extract operation status of railroad systems [Tsuchiya et al., 2013]; studies to extract road traffic information [Hanifah et al., 2014; Sakaki et al., 2015]; and a study to extract tourist information [Obara et al., 2015]. As for the classification of disaster tweets, there is a study by Rokuse [Rokuse et al., 2015]. They classified the disaster tweets of the

Great Eastern Earth Quake. They used several key words expected to be used intensively during earthquakes for filtering the tweets. They used 5 categories, "information on tsunami", "information on evacuation", "information on utilities (gas, water, electricity)" "information on public transport" and "information on road status", as classes to classify the tweets. However, one can assume there are much wider variety of information in tweets during disasters.

Our study uses actual tweets during disasters to apply clustering method to investigate the contents of the tweets. And based on the analysis, the disaster tweets are classified in more detail.

2. Method

2.1 Outline

The authors propose to develop a system to classify unknown tweets during disasters into categories generated by clustering.

For this purpose, a classifier needs to be developed in advance to extract and classify disaster tweets from whole tweets data. In our study, two separate tweet groups tweeted during two typhoons are used. One tweet group (for typhoon 16) is used as learning data to develop a classifier. The other data group (for typhoon 13) is used as test data to assess the accuracy of the classifier. Figure 1 shows the outline of our study. The flow in the red box indicates an actual process when the proposed system becomes operational.

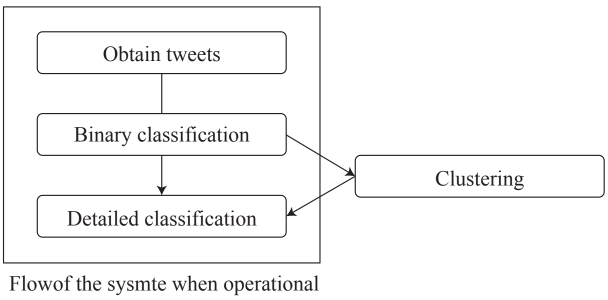

Figure 1: Outline of our study

2.2 Data

Table 1 shows the tweet data groups used in our study. Appropriate keywords are required to extract disaster tweets. To select the keywords, the authors employ dynamic time warping for time-series comparison between the occurrence of the words in the learning data and the precipitation of the cities hit by typhoon 16. As a result the comparison, the authors decided to use three keywords: Rain, Typhoon, and Heavy Rain. Table

Table 1: Tweet data obtained

	Period
Typhoon 13	2016/9/6 21:00–2016/9/7 21:59
Typhoon 16	2016/9/19 22:00–2016/9/20 11:59

Table 2: Tweet data used

	Number of tweets
Typhoon 13(test data)	5,474
Typhoon 16(learning data)	7,167

2 shows the tweet data used in our study.

2.3 Vectorization

In order to analyze the contents of the tweets expressed in natural language, the texts in the tweets are vectorized as follows:

- Pre-process the tweets for morphological analysis. Specifically, alphabets and numbers are unified into half-width and katakana into full-width.
- Extract noun, verb, adjective and adverb through morphological analysis using MeCab.
- Vectorize the tweets through bag of words method using the extracted words as features
- Weighting the words in each text by using TF-IDF.

3. Extracting disaster tweets: binary classification

3.1 Outline

Binary classification is applied to the tweet data vectorized by the feature words to classify the tweets into disaster tweets and non-disaster tweets. A classifier is developed by using the training data and assess it by using the test data. Support Vector Machine is used for a classifier. (Cortes and Vapnik 1995). As for training data, the authors balance the number of disaster and non-disaster tweets. Table 3 shows the data sets used. For the assessment, F-measure is used.

Table 3: Breakdown of the data sets

	Disaster tweets	Non disaster tweets
Training data	1,347	1,347
Test data	4,86	4,988

3.2 Developing a classifier

The training data is used to develop a classifier. The number of the feature words is 12466. 10-fold cross validation is applied for training the classifier. The result shows the existence of unnecessary words for classifying the tweets and words that function as noise. So an additional experiment is run by using the words selected by Bi-Normal Separation method. BNS is often used to select features in binary classification. In our study, the threshold is set to 0.33. The feature words with BNS value higher than the threshold are used. The number of the feature words is 1953. As a result of selecting the feature words by using BNS method, F-measures for both disaster tweets and non-disaster tweets are improved.

3.3 Assessing the classifier

The classifier is assessed by using all the test data. Specifically, the authors assess the classifier by using all the feature words extracted from the training data and by using only the feature words selected through BNS method. The number of the feature words when using all the words is 5280 and 1,014 when using only the selected words through BNS.

3.4 Result of the experiment

Table 4 shows the result of the experiment.

Table 4: F measures of the binary classification experiment

All words	BNS method
0.85	0.89

3.5 Discussion

The experiment shows the improvement in F-measure by selecting the feature words through BNS method. However, visual checks on the tweets classified incorrectly show that there are some tweets which are correctly classified when using all the words but incorrectly classified when using the selected words through BNS. The checks also show that by using BNS method, the tweets with lesser information are more likely to be classified incorrectly. As already mentioned, the threshold of BNS is set to 0.33 based on F measure of the experiment using the training data only. The incorrect classifications indicate a necessity to reconsider a proper level of threshold when data other than the training data (in our study, tweets of typhoon 13) is used as the test data. Also, one word tweets which tend to be classified incorrectly could be excluded from classification since they contain little information.

Without selecting the features words by BNS method, by using two data sets from different disasters (typhoon 13 and 16) for training data and test data, the feature words used frequently during disasters are selected. Hence, using disaster data obtained from similar situations can lead to a better selection of feature words and improvement in accuracy.

4. Contents analysis of disaster tweets by clustering

4.1 Outline

Clustering is applied to the training data to investigate the contents of the tweets during disasters. The experiment conducted in this chapter is shown in the red box in Figure 2

The tweets that may function as noise are excluded by using DBSCAN before conducting k-means++ clustering. DBSCAN is effective against noises. Also, the training data is weighed by TF-IDF. In this case, for the feature value to vectorize the tweets, the author do not adopt the feature words without processing as in the case of BNS. Rather dimension reduction is conducted by using t-SNE, which is suitable for multi-dimensional, non-linear analysis. (van der Maaten and Hinton, 2008). As a result, the dimensions are reduced from 12467 to 2.

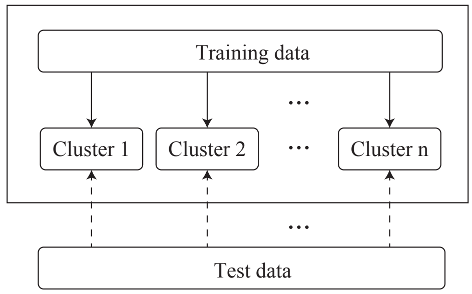

Figure 2: Application of the result of clustering to the test data

4.2 Experiment: Clustering

Following the method explained in 4.1, clustering is conducted by using the disaster tweet (1347 tweets) only. Tweets contain various expressions and many noises are expected to exist in them. To overcome these problems, the first stage clustering is applied by using DBSCAN, which is effective against noises. As a result, 68 clusters are generated. Then the second stage clustering is applied by using k-means++. 875 tweets are used. The remaining 472 tweets, which belong to 5 clusters with negative value after conducting silhouette analysis, are excluded. Silhouette coefficient and elbow method are employed to find appropriate numbers of cluster necessary for applying k-means++. The results yield k value of 15.

4.3 Result

Table 5 shows a part of the contents generated by DBSCAN followed by k-means ++. Figure 3 shows the silhouette plot.

Table 5: Contents of the clusters by k-means++ (partial)

Cluster	Silhouette coefficient	Contents
0	0.333	Tweets about flight
		Tweets about railroad operation
		Traffic information
		Tweets about going out
		Comments about precipitation
1	0.330	Comments about precipitation
		Comments about evacuation advice
		Comments about typhoon
2	0.457	Warning, caution (include comments on caution)
3	0.899	Comments about precipitation (short)
• • •	• • •	• • •
14	0.299	Precipitation in Kochi
		Fear of typhoon
		This is from (location). Weather report etc.
		Fear of flooding
		Thunder
		Fierce
		Comments about heavy rain

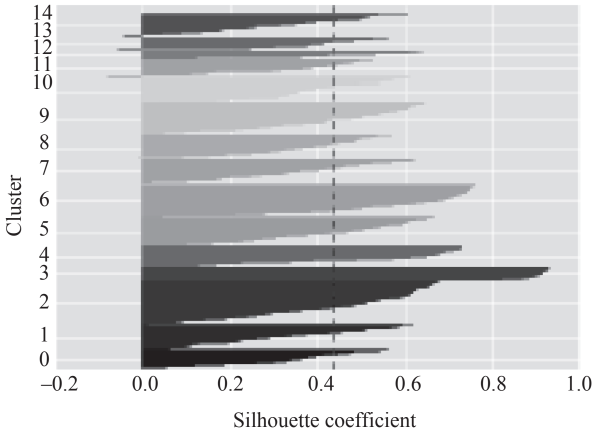

Figure 3: Silhouette plot of k-means++

4.4 Discussion

There are variation in silhouette coefficient among the clusters generated. The clusters with more tweet members tend to have low silhouette coefficient. However, cluster 14 is exceptional: It has small number of tweets (56) and the lowest silhouette coefficient. Visual checks on the content of the tweets in cluster 14 does not reveal any consistent messages. Hence, it is estimated that there are some relationship between silhouette coefficient and the consistency of the content of the tweets in the cluster.

In the following chapter, detailed classification is conducted by using 15 clusters generated in this experiment as classes. But a cluster such as cluster 14 can lower accuracy rate of classification.

Our study generates 15 clusters. This means the proposed system can encompass comprehensive tweets with various topics. However, there are several clusters which contain many information not necessary in a disaster. Also, in some cases, selectively providing specific topics is beneficial to a user who wants to obtain relevant information efficiently. Hence it is important for our study to take this user friendly perspective into account when presenting the information. In order to reflect this perspective, the clusters shown in Table 5 can be arranged in descending order of silhouette coefficient. A user can obtain necessary information quickly by focusing on clusters with larger coefficient.

5. Detailed classification

5.1 Outline

The test data are used to classify 15 clusters extracted in the experiment with the training data. The results of the clustering in the previous chapter are used as correct labels. The results of the classification are assessed by using the test data. F measure is used for the assessment.

5.2 Developing a classifier

A classifier for the detailed classification is developed by using the training data. One-versus-rest SVM, which is used in multi class classification, is applied to develop the classifier. The words in each tweets are used as the feature words to vectorize the tweets. TF-IDF is used to conduct weighting. The number of feature words in the training data is 6137. 10-fold cross validation method is applied for training. The result of the experiment yields F measure of 0.85.

5.3 Result

The test data are used to assess the classifier. All the feature words extracted in 5.2 are used for the assessment. Table 6 shows the data used. The number of the feature words in the test data is 6138. F measure are used for the assessment. Figure 4 shows the result of classification. F measure is 0.59.

Table 6: Training data and test data

Class	Training data	Test data
0	48	18
1	63	4
2	107	124
3	36	39
4	52	11
5	73	180
6	82	45
7	60	49
8	62	5
9	78	30
10	61	15
11	43	13
12	21	2
13	33	3
14	56	26

5.4 Discussion

The reasons for incorrect classifications of the tweets in class 5 and 7 are investigated.

There are many cases that the tweets in class 5 were incorrectly classified as class 7. The similarity between the two classes is they are both related to weather information about typhoon. The difference is that the class 5 is for national broadcast and class 7 is for affected residents with information about secondary disaster warning. Visual checks on the contents of the tweets in both classes show that even if a tweet is intended for national broadcast, when a specific disaster, such as “mud sliding” and “flooding in low land”, or the wording for warning are included, the classification may be conducted incorrectly.

For the incorrect classifications of tweets in class 14, there are many cases that tweets in this class were misclassified in class 2. The partial similarity between the two classes is they both contain the word “heavy rain”. The difference is that class 2 is about “warning and/or caution”, but the class 14 contains

0	16	0	0	0	0	0	0	1	0	1	0	0	0	0	0
1	0	2	0	0	0	0	1	0	0	0	0	0	0	0	1
2	0	0	21	0	0	0	0	3	0	0	0	0	0	0	0
3	2	0	1	14	0	0	2	1	6	9	3	0	0	1	0
4	0	0	0	0	4	0	0	1	1	2	2	0	0	0	1
5	4	2	12	0	3	88	0	31	4	2	2	10	0	0	24
6	1	1	0	0	0	0	30	0	1	1	1	0	1	0	0
7	0	3	2	0	0	5	0	34	0	0	1	1	1	0	2
8	0	0	0	0	0	0	0	0	4	0	0	0	0	0	1
9	1	0	0	1	0	0	2	0	1	23	0	0	0	0	2
10	0	0	0	1	0	3	0	1	0	0	10	0	0	0	0
11	0	0	0	0	0	1	0	0	2	0	0	8	0	0	2
12	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	1	0	0	0	2	1
14	1	1	14	0	3	1	0	0	2	0	1	0	0	0	3
	0	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*

Figure 4: Result of detailed classification

wide range of tweets. This partial matching leads to incorrect classifications. As mentioned before, class 14 shows the lowest silhouette coefficient. This also shows low consistency of tweets in class 14, which leads to a higher rate of incorrect classification.

6. Conclusion

The study tries to verify the potential of Twitter for disaster information source with its simplicity, promptness, and wide propagation. Specifically, a classification system is developed based on the tweets contents in order to extract useful information from a huge number of tweets. Based on the assumption that under similar disaster (in our study typhoon), words contained in disaster tweets have correlation, the author conduct several experiments. The data of typhoon 16 are analyzed through classification and clustering. For the data of typhoon 13, the authors apply the feature words to binary classify tweets data into disaster and non-disaster tweets, and conduct clustering and in-depth classification. F measures exceeds 0.8 when using BNS and about 0.6 when using 15 value classification.

As for clustering, DBSCAN is applied to exclude the data that can function as noise when clustering or classification, before applying k-means++ to generate 15 clusters. However, there are several different expressions to convey similar meaning and this lowers the accuracy of clustering.

In order to improve the accuracy of clustering of disaster tweets, following three challenges are needed to be overcome:

First, one word tweets have less information and hence prone to be classified incorrectly. In our study, these short tweets are included in disaster tweets. However, to put our proposed system into operation, it is advisable not to include short tweets in disaster tweets. Rather it should be treated as non-disaster tweets.

Second, the scope of study needs to be expanded. Specifically, in this study, the tweets from typhoon 13 and 16 are used. In order to improve the accuracy of the system, it is necessary to use larger data from various disasters to train the system with various tweet data.

Third, unifying various expressions with same meaning in advance is necessary to improve the accuracy of the system.

References

Akita Prefecture (2015). Guideline of disaster management concerning tourists. accessed on 2017/2/28 at http://www.bousai-akita.jp/uploads/user/system/File/guidelines/kankoukyaku.pdf.

CAO (2010). *White paper on disaster management 2010*. Cabinet Office of Japan, accessed on 2017/01/22 at http://www.bousai.go.jp/kaigirep/hakusho/h22/bousai2010/html/honbun/2b_1s_1_01.htm.

CAO (2016). *Flood responses manual for municipalities*. Cabinet Office of Japan, retrieved on 2017/1/27 from http://www.

bousai.go.jp/taisaku/chihogyoumukeizo-ku/pdf/suigai_tebiki 1.pdf.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, Vol. 20, No. 3, 273-297.

Hanifah, R., Supangkat, S.H., and Purwarianti, A. (2014). Twitter information extraction for smart city. *Proceedings of 2014 International Conference on ICT For Smart Society*, 295-299.

Ito, A., Kunitomo, M., Kamiyama J., Qiu, T., Araki, T., and Miyakawa T. (2015). Research on possibility of grasping situation in the areas affected by land slide disasters by utilizing tweets. *Proceedings of Japan Society of Erosion Control Engineering*, Vol. B, 132-133.

Japan Meteorological Agency. (2017). *Long term trend of the occurrence of short-time heavy rain from AMeDAS data*. accessed on 2017/1/22 at http://www.jma.go.jp/jma/kishou/info/heavyraintre-nd.html.

Japan Tourism Agency (2013). *A proposal for the establishment of natural disasters information service to the foreign tourists to Japan*, accessed on 2017/2/28 at http://www.mlit.go.jp/common/001000495.pdf.

Japan Tourism Agency (2016). *Accommodation statistics for 2015*, accessed on 2017/2/28 at http://www.mlit.go.jp/common/001136323.pdf.

Matsumoto, S., Kawaguchi, H., and Toriumi, F. (2015). Quantifying twitter users around the great East Japan earthquake based on their posting activities and its application for automatic user filtering. *Transactions of the Japanese Society for Artificial Intelligence*, Vol. 30, No. 1, 393-402.

Müller, M. (2007). "Dynamic time warping." *Information Retrieval for Music and Motion*. Springer, 69-84.

Nakatani, Y. (2016). Disaster information systems for tourists. *Systems, Control and Information*, Vol 60. No. 4, 160-165.

Obara, M., Morita K., Fuketa, M., and Aoe, J. (2015). Extraction of tourist information from contents of tweets and building an analysis system. *Proceedings of 29th Annual Conference of the Japanese Society for Artificial Intelligence*, 1-3.

Ritter, A. Mausam, Etzioni, O., and Clark, S. (2010). Open domain event extraction from twitter. *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1104-1112

Rokuse T., Uchida, O., Tomita, M., Kajita, Y., Yamamoto, Y., and Toriumi, F. (2015). A research on clarifying place names' ambiguity for facilitating information distribution during large scale disaster. *Proceedings of the Association for Natural Language Processing*, Vol. 21, 220-225.

Sakaki, T., Yanagihara, T., Nawa, K., and Matsuo, Y. (2015). Driving information extraction from twitter. *IEICE Transactions on Information and Systems*, Vol. 98, 1019-1032.

Toriumi, F., Sakaki, T., Shinoda, K., Kurihara, S., Kazama, K., and Noda, I. (2013). Classification of information in disaster situation from network structures. *Proceedings of 27th Annual Conference of the Japanese Society for Artificial Intelligence*, 1-4.

Tsuchiya, K., Toyoda, M., and Kitsuregawa, M. (2013). Detecting occurrences and continuation status of train troubles from microblogs. *Fundamental Problems of Artificial Intelligence*, Vol. 91, 53-58.

van der Maaten, L. and Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 2579-2605.

(Received June 22, 2017; accepted September 14, 2017)