Extraction of the Information from Tweet Data in Case of the Disaster using Visualization

T. Funayama[1], Y. Yamamoto[2], and O. Uchida[3]
[1]Graduate School of Science and Technology, Tokai University, Hiratsuka, Kanagawa, Japan
[2]School of Science, Tokai University, Hiratsuka, Kanagawa, Japan
[3]School of Information Science and Technology, Tokai University, Hiratsuka, Kanagawa, Japan

Abstract— *In Japan, when a disaster occurs, victims and local governments use a social network service such as Twitter as a means of information transmission. Therefore, in this study, we attempted to extract information from data contained in the posted tweets during a disaster. To this end, we implemented visualization using correspondence analysis. When a disaster occurs, many people tweet; consequently, the volume of tweet data becomes large and the computational cost increases. Owing to this problem, we decided upon the type of information to be extracted and the related keywords to be set for filtering the information. Thereby, only the tweet data including a particular keyword were analyzed. Moreover, we considered the trend of information from Twitter with regard to the time zone. In this study, we use tweet data to explain the aftermath of the April 2016 Kumamoto earthquake, whose epicenter was in the Kumamoto prefecture.*

Keywords: Twitter, Extraction of the information, Text mining, Data visualization, Correspondence analysis

1. Introduction

In recent years, many natural disasters such as earthquakes and floods have occurred in Japan. The Great East Japan Earthquake, which occurred in March 2011, is the most serious recent disaster. In this earthquake, extensive damage was caused by earthquake and tsunami. This earthquake occurred at 14:46(JST) on Friday, so many people were at a school, working place and so on. During that time, many people were forced to remain in evacuation shelters because transportation was disturbed or their residences had collapsed. The victims were able to obtain information from radio, television, and internet; however, they could not contact their families because they were unable to use the phone. At that time, social networking services (SNSs) such as Twitter were used for transmitting information. Consequently, the lives of many people were saved owing to the distribution of information via Twitter. Because many people were trying to contact their family members via phone calls, the phone lines were overloaded and could not be effectively used; however, internet could be used as usual. In natural disaster events such as earthquakes or floods following an earthquake SNS plays an active role in information transmission. Twitter has a large number of users; therefore, it receives various types of information. However, extracting useful information from Twitter is difficult because in the event of a disaster users tweet concurrently. Thereby, useful tweets regarding the state of the disaster become obscured. From the extensive tweet data, if the useful information can be extracted without being overlooked, disaster ’ s magnitude can be quickly assessed. For example, vital information pertaining to a location where a cliff is likely to collapse or a person who cannot evacuate on its own can be promptly grasped. By utilizing such information, loss of human life can be effectively averted. Therefore, the utilization of Twitter data during disasters is an active field of study. For example, during the disaster due to heavy snowfall in the San-in district, a Twitter user called for information-sharing using the hash tag “ #sanin_snow ”. There was also a case where a large-scale disaster community was formed. Research is also being conducted on classifying such Twitter data into weather information, damage report, calling, and encouraging victims to assess the disaster situation, reliability of information, and user characteristics [1]. In addition, an application[2] for sending accurate and useful information from Twitter was developed. In this application, a hashtag such as “ #平塚市災害 (Hiratsuka city disaster) ” or “ #救助 (rescue) ” as well as information on the current position and a code representing the position information, called the UTM point, were automatically inserted in the relevant tweets. “ #平塚市災害 (Hiratsuka city disaster) ” makes it easy to search for tweets containing information pertinent to the damage situation with regard to a specific city. Because the UTM point is a coordinate system, which the SDF also uses, it is useful in extracting the location where rescue efforts need to be focused. By utilizing Twitter in the event of a disaster, various types of information, relevant to the disaster, accumulate as tweet data. In this study, we attempted to extract useful information from these tweet data. We extracted information by visualizing tweet data using text mining. Furthermore, we divided data according to tweet date and time and examined the differences and the features of the extracted information with regard to each time zone. In this study, we use tweet data to explain the aftermath of the April 2016 Kumamoto earthquake, which had its epicenter in the Kumamoto prefecture.

Table 1: Collected tweet data

Keywords	Number of Tweets
#救助 (rescue)	48,222
デマ (false rumor)	150,314
救助 (rescue)	372,800
熊本 (Kumamoto)	2,310,704
地震 (earthquake)	9,235,067
東海大 (Tokai Univ.)	11,795
避難 (evacuation)	2,734,384
安否 (safety)	30,077
Total	14,893,363
After duplicate deletion	5,537,928

Table 2: Keywords related to evacuation life in the car

Keywords
車 (car)
車中泊 (stay in the car)
駐車場 (parking lot)
車中 (in the car)
車内 (in the car)
電車 (train)
車両 (vehicle)
車内泊 (stay in the car)
車中泊避難 (evacuation of stay in the car)
駐車場不足 (parking shortage)
自家用車 (private car)
自動車 (car)
車内避難 (evacuation into the car)
車避難 (evacuation into the car)
車生活 (living in the car)
車中避難 (evacuation into the car)
駐車 (parking)
満車 (The parking place is full.)
車泊 (stay in the car)
車内生活 (living in the car)
自動車内 (in the car)
避難車 (car of evacuees)
乗用車 (car)
車移動（To move the car）
車待機 (wait in the car)

2. Collection of Tweet data

In this study, eight keywords (Tab. 1), such as “地震 (earthquake),” “避難 (evacuation),” “安否 (safety),” “救助 (rescue),” “#救助 (rescue),” “熊本 (Kumamoto),” “東海大 (Tokai Univ.),” and “デマ (false rumor)” were set; tweet data were then collected using the Twitter API. The total amount of tweets collected was 14,893,363; when duplicate data were excluded, 4,868,308 tweets remained. The collected tweet data were tweeted during the period from April 12, 2016 at 1:58 to April 21, 2016 at 14:20.

3. Extraction of Disaster situation information

It was not possible to extract useful information by analyzing the collected tweet data because various types of information were included in the data; moreover, information other than the desired information became noise. For example, there were many automatic tweets of earthquake bulletins because “地震 (earthquake)” was the keyword; these tweets became noise. In addition, in the event of a disaster, many people tweet; therefore, the amount of tweet data becomes large. Consequently, the cost of computation increases; therefore, a high-spec computer is required. Owing to this problem, we had to decide upon the type information to be extracted and set keywords related to the desired type of information. Thereby, only tweet data including the keyword would be extracted for analysis. In the Kumamoto earthquake, there were many reports from mass media about the evacuation life in the car; therefore, we extracted information relevant to evacuation life in the car. In addition, to make the extraction of filtered information easier, re-tweet data were removed.

3.1 Extraction of information on evacuation life in the car

Because evacuation life in the car was related to evacuation, we used the collected tweet data by setting “evacuation” as the keyword. Then, we tried to extract information by morphological and correspondence analyses. Text Mining Studio (NTT DATA Mathematical Systems Inc.) was used for morphological analysis. Various words such as “車 (car),” “車内 (in the car),” “車中泊 (stay in the car),” and “駐車場 (parking lot)” could be thought of as words pertaining to the evacuation life in the car. Because many of these words contained the word “車 (car),” tweets including “車 (car)” were tabulated. Words (Tab. 2), which were considered to relate to evacuation life in the car from frequently appearing words, were referred to as keywords. The tweet data including the keyword were taken as the target data for analysis. By doing this, it was possible to reduce the size of data and to filter the extracted information to information relevant to evacuation life in the car. Looking at the time-series trend pertaining to the amount of this data (Fig. 1), the number of tweets was large during the foreshock and the main shock; moreover, the number of tweets was higher after the main shock in the daytime, than it was in the day after the foreshock. As shown in Table 3, information was extracted every 6 hours after the foreshock; information was extracted for each period and the change of information extracted in the particular time zone was considered. In this study, the co-occurrence matrices of nouns and adjectives were created in each period and visualized using correspondence analysis; relevant words were thus extracted. The nouns and adjectives used each of the top 50 words; however, when frequently appearing words were less than 50, all the words in that part of speech were used. Figure 2 shows the visualization by correspondence analysis in period no. 1. In this visualization, words which tended to be included in the same tweet data were displayed nearby, for example, “駐車場 (parking lot),” “痛い (painful),” “駐車場 (parking lot),” “広い (large),”

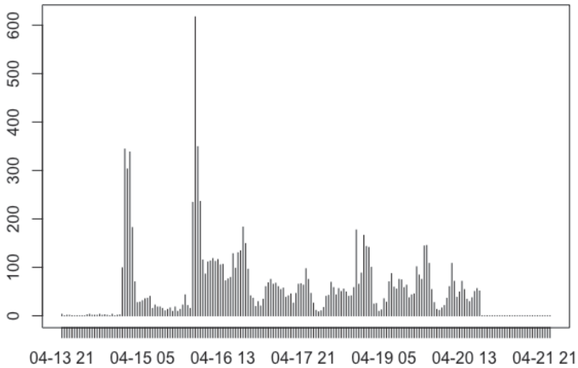

Fig. 1: Time-series change of number of tweets including “ 避難 (evacuation) ” and keywords

Table 3: Period

Period No.	Period	Remarks column
No. 1	From April 14, 2016 at 21:00(JST) to April 15, 2016 at 3:00(JST)	The foreshock occurred on April 14, 2016 at 21:26(JST).
No. 2	From April 15, 2016 at 3:00(JST) to April 15, 2016 at 9:00(JST)	
No. 3	From April 15, 2016 at 9:00(JST) to April 15, 2016 at 15:00(JST)	
No. 4	From April 15, 2016 at 15:00(JST) to April 15, 2016 at 21:00(JST)	
No. 5	From April 15, 2016 at 21:00(JST) to April 16, 2016 at 3:00(JST)	The main shock occurred on April 16, 2016 at 1:25(JST).
No. 6	From April 16, 2016 at 3:00(JST) to April 16, 2016 at 9:00(JST)	
No. 7	From April 16, 2016 at 9:00(JST) to April 16, 2016 at 15:00(JST)	
No. 8	From April 16, 2016 at 15:00(JST) to April 16, 2016 at 21:00(JST)	It will rain at night.

“ 心強い (encouraging), ” and “ 揺れ (shaking). ” From this, we could not understand the meaning of “ 駐車場 (parking lot) ” and “ 痛い (painful) ”; therefore, we decided to extract the tweet data including “ 駐車場 (parking lot) ” and “ 痛い (painful) ” and to obtain information from tweets. A part of tweet data including “ 駐車場 (parking lot) ” and “ 痛い (painful) ” is shown in Table 4. By checking Twitter, it was possible to extract information, from tweet data, about people who evacuated by laying blue sheets in the parking lot. Furthermore, because “ 駐車場 (parking lot) ” was also close to “ 広い (large), ” they tended to be included in the same tweet. These extracted tweet data (Tab. 4) contained enormous amount of tweets, whose content was “I evacuated to a large parking lot ”. This shows that there were many people evacuating to a large parking lot by car. In period no. 6 (Fig. 3), “ 外 (outside) ” and “ 寒い (cold) ” tended to be included in the same tweet. When the tweet data (Tab. 6) including these words were extracted, we were able to conclude that people did not evacuate because they did not have a car and that people changed their evacuation location, from the evacuation shelter to a car, because of the cold. Information regarding people who did not evacuate but were still at home was also extracted in this period. Furthermore, “ 揺れ (shaking) ” and “ 長い (long) ” tended to be included in the same tweet data. When these tweet data (Tab. 7) were extracted, information regarding the individuals who decided to evacuate was obtained. Because this was the period after the main shock, these would be individuals who did not evacuate when the foreshock occurred. In period no. 8 (Fig. 4), “ 早い (early) ” and “ そう (Sou) ” tended to be included in the same tweet data. Table 8 shows the tweets including “ 早い (early) ” and “ そう (Sou). ” These tweet data contain information regarding the state of the afflicted area conveyed by people not residing in the afflicted area. Additionally, people not residing in the afflicted area tweeted information from victims and requested relief for the victims. Furthermore, “ 風 (wind) ” and “ 強い (strong) ” tended to be included in the same tweet data. Table 9 shows the tweet data including “ 風 (wind) ” and “ 強い (strong). ” From these data, we concluded that there were people who changed their evacuation location from a car to an evacuation center and people who evacuated themselves. We were also able to extract information with regard to people who were evacuating in preparation for the night rain. By analyzing the period separately in this manner, it was possible to extract information particular to the time zone.

Table 4: Period no. 1 tweets including "駐車場 (parking lot)" and "痛い (painful)"

Tweet
@marira55 ほんとだよね、命懸け。 だんだん被害が明るみになってきていて心が痛い。益城町、対応大変そうだよね、避難所どこなのか知らないが駐車場にブルーシートって。
なんで学校の体育館とか開けないんだろう…なんでずっと市役所前や老人ホームの駐車場が避難所なんだろう…公共施設より個々の事業団のが判断早いのか… アスファルト座ってたらビニールシートの上では痛いよね、、
ブルーシートが敷かれた駐車場で毛布に包まって…胸が痛いね土日が雨降りそうな気配だから、できるだけ早くちゃんとした避難場所が確保される事を願うばかり https://t.co/WOpHtciBfd

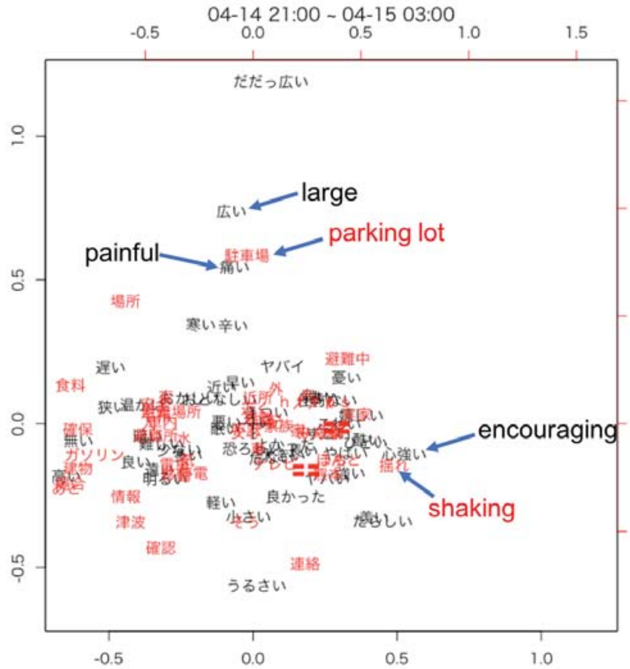

Fig. 2: Visualization by correspondence analysis of period no. 1

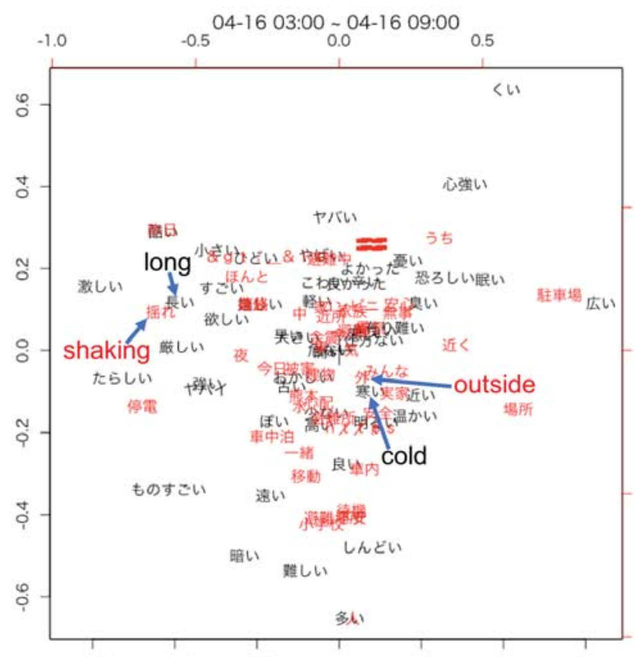

Fig. 3: Visualization by correspondence analysis of period no. 6

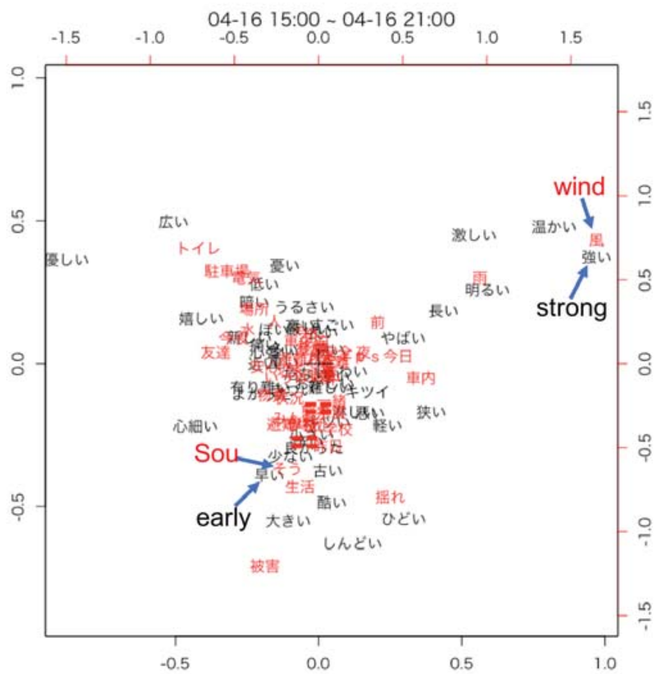

Fig. 4: Visualization by correspondence analysis of period no. 8

Table 5: Period no. 1 tweets including "駐車場 (parking lot)" and "広い (large)"

Tweet
外の広い駐車場に避難してるなう
広い駐車場に避難中。まだ揺れてやがる。
@marucat1119 ありがとー !!! o(T □ T)o いちお、念のため広い駐車場に避難しよる (T_T) そこで妹とも合流予定〓
@karagirl625 正直やっばいです。棚のもの全部落ちました…w 震源地からめちゃ近なんで広い駐車場に避難してます！
一応近くのお店の広い駐車場に家族みんなで車で避難してます〓 (〓゜〓〓 〓゜) 〓 マンション最上階だから何かあってからじゃ遅いからね
@umeshiho 今、帰宅途中なんだけど疲れたから広い駐車場にいる。避難してる人いっぱい。でも早くおウチ帰りたい
職場の駐車場が広くてたくさんの人が車で避難してきてます。私はそこにいます。大丈夫。

Table 6: Period no. 6 tweets including "外 (outside)" and "寒い (cold)"

Tweet
車もってないから避難しようにも勇気がない 身一つで外でるのも寒いのできついのが本音
避難場所の中学校は危険だと言われた。ショッピングモールの駐車場が一時避難場所だと言われたけど、外寒いし部屋で寝る
さすがに外での避難は寒いので、一旦アパートに戻り、車が動かせるのを確認。そのまま龍田中学校に避難場所を移動しました。

Table 7: Period no. 6 tweets including "揺れ（shaking）" and "長い (long)"

Tweet
寝てたら昨日よりも激しくて長い揺れ。停電して今も外は真っ暗なまま。近くのスシローの駐車場が広いから避難中。ほんとに怖い。
@misepsycho みせせんさん、ご心配本当にありがとうございました。前回よりも酷く長い揺れで、車に避難して夜を明かしました。お気持ちだけでも、すごく救われました…！
車中泊してやっと夜が明けた。 昨日の揺れは長く強く怖かった… もう家の中に居るのが怖過ぎて 今から家の状況見て避難所行こうと

4. Conclusion

In this study, tweets containing useful information were extracted by analyzing the words included in the same tweet by visualization using correspondence analysis. Because many people tweet concurrently during a disaster, enormous amount of relevant information accumulates on Twitter; therefore, it is difficult to extract the target information from the mixed information. However, by narrowing down on the data and setting relevant keywords for the same, it was easier to extract the desired information. Furthermore, it was possible to extract specific information related to the events before and after the disaster in that time zone by segmenting the time period to be analyzed.

Table 8: Period no. 8 tweets including "早い (early)" and "そう (Sou)"

Tweet
熊本の西山中学校に友達と小 2 の息子君が避難しています。車内生活を余儀なくされるそうです。 食べ物、飲み物を買うところがなく、大勢の方が飲まず食わずで困っているそうです。 どうか、少しでも早い物資の援助をお願いします。
意識的に自粛しているわけではないのですが…とてもそんな気になれない、という気持ちがあり浮上できません 私の親族は、最も被害が大きいとされる益城町のすぐ隣の地区にいます 避難所もいっぱいで、犬も猫もいる状況のため車で生活しているそうです どうか早く平穏を取り戻せますように
熊本の友達が困ってます 何もしてあげられない…。 指定避難場所からあふれ 免許センターの駐車場に避難してます 物資が届かないそうです 少しでも早く対応願います！
@a_km_2 そうなんよね（´・ω・｀）ショボーン 早く余震落ち着くといいよね !! うちの従兄弟のおばちゃんの実家が熊本で弟さん避難してるんだけど避難所が何処もいっぱいらしくて車で寝泊りしてるって言ってたって聞いたよ !!

Table 9: Period no. 8 tweets including "風 (wind)" and "強い (strong)"

Tweet
風も強くなってきたし 雨も降ってきたので 近くの駐車場じゃやっぱ怖いねってことで 小学校に避難して来ましたー 体育館の中はいっぱいで入れないから 車の中
@jiyongsarage88 風も強いのが恐怖すぎる泣 今、避難する準備した！車に詰めたよ！
また夜がくる、今日は風も強いし雨も降るって怖すぎる。ひとまず家族で道の駅に車で避難してるけど、ちゃんとした避難所がいいのかな。避難所の建物が壊れないかも心配。どうしよう。。
高専に避難。 車にいるけど風が強すぎて気持ち悪い。 体育館いこうかな
@caori_co 雨、意外と遅い降りですね。風やばくないですか？今日は私だけ避難所の駐車場にいます。私の住んでる地域は土砂崩れ危険地域でさすがに強かったので。ただ父と犬は残してきたので心配です。一人ぼっち。何事もないように祈ってます
とりあえず中学校のグラウンドに避難してる…体育館人いっぱいらしいから車内に！でも風が強くて車が余震で揺れてるのか風なのかわからない…

References

[1] T. Ishikawa, A. Kawasaki, and K. Meguro, "Investigation of information sharing by Twitter users during the 2011 heavy snow disaster in San-in region of western Japan", *Journal of social safety science / Institute of Social Safety Science*. no. 4, pp. 1–9, Jun. 2012.

[2] O. Uchida, M. Kosugi, G. Endo, T. Funayama, K. Utsu, S.Tajima, M. Tomita, Y. Kajita, and Y. Yamamoto, "A Real-Time Information Sharing System to Support Self-, Mutual-, and Public-Help in the Aftermath of a Disaster Utilizing Twitter", *IEICE Transactions on Fundamentals*. vol. E99-A, no. 8, pp.1551–1554, Aug. 2016..

[3] O. Uchida, T. Rokuse, M. Tomita, Y. Kajita, Y. Yamamoto, F. Toriumi, B. Semaan, S. Robertson, and M. Miller, "Classification and Mapping of Disaster Rel- evant Tweets for Providing Useful Information for Victims During Disasters", *IIEEJ Transactions on Image Electronics and Visual Computing*. vol. 3, no. 2, pp.224–232, Dec. 2015.

[4] O. Uchida, M. Kosugi, G. Endo, T. Funayama, K. Utsu, S. Tajima, M. Tomita, Y. Kajita, and Y. Yamamoto, "A Real-Time Disaster-Related Information Sharing System", *Proc. 5th International Conference on Social Media Technologies, Communication, and Informatics*. pp.22–25, Nov. 2015.