

# Macroanalysis of Microblogs: An Empirical Study of Communication Strategies on Twitter During Disasters and Elections

Yuuto Fukushima Fumito Masui Michal Ptaszynski Yoko Nakajima  
Keisuke Watanabe Ryotaro Kawaishi Taisei Nitta Ryoya Sato

Department of Computer Science, Kitami Institute of Technology  
165 Koen-cho, Kitami, 090-8507, Japan  
{fukushima,yoko,ksk67,kantannano,nitta,r.satou}@ialab.cs.kitami-it.ac.jp  
{f-masui,ptaszynski}@cs.kitami-it.ac.jp

## Abstract

In this paper we present our study on language behavior of Twitter users during two kinds of influential events, namely, natural disasters and political elections. In our study we analyze credibility of information included by users in tweets in the above-mentioned situations, by classifying the information in two kinds: Primary Information (first-hand reports) and Secondary Information (second-hand reports, retweets<sup>1</sup>, etc.). We also perform sentiment analysis of the data to check user attitude toward the occurring events.

## Introduction

The main purpose of Social Networking Services (SNS), such as facebook<sup>2</sup> or mixi<sup>3</sup>, is communication with friends and acquaintances. Compared to such SNS, twitter<sup>4</sup> is a kind of service often used for slightly different purpose, namely the dissemination and broadcasting of information. For example, by using the unique features such as “retweet” or “hashtags” it is possible to broadcast information to a countless number of users more easily than on other SNS. Nowadays, twitter has become an important source of information in everyday life. It affects people’s decision making and has become an object of numerous research (Kamishima 2007; Iwaki et al. 2009; Tanaka and Tajima 2010; Kazama et al. 2010; Fujisaka et al. 2010; Jiang et al. 2011; Umejima et al. 2011). For example, Kuwano et al. extracted from twitter tourism-related information (Kuwano et al. 2012). Aramaki et al. used twitter to predict the spread of influenza (Aramaki et al. 2011). Umejima et al. performed an analysis of factual information on twitter to filter out and prevent spreading false rumors (Umejima et al. 2011). It has been also noticed that the speed of information spreading on twitter can be useful in times of emergencies and natural disasters, such as the Great East Japan Earthquake, which occurred on 11th of

March 2011 (Tateiri 2011). Finally, the importance of information spreading on twitter was noticed by the government of Japan, and for the first time in history caused allowing the use of Internet as a place for election campaign during the recent Upper House elections on 21st July 2013 (Miyabe et al. 2011). All the above indicates that SNS is rapidly becoming an important part of the society infrastructure.

In general, appropriate selection of information is important especially when it comes to the gathering of information in times of emergency and making decisions based on such information. A lot of information appears on twitter, some of it contains private opinions about a variety of topics. This also includes appearing of various hoax tweets and false rumors unrelated to the general topic and mixed into the main thread. Therefore, a method for extracting only valid and useful tweets from a jumble of information on twitter becomes essential. It is important to ensure the accuracy and the uniformity of the extracted information.

One of the means to determine the accuracy of information is using the concept of primary and secondary information. Primary information refers to the kind of information that a person directly saw, heard or personally did. Secondary information refers to indirect information such as re-posting or re-telling what was described by someone else (third party), such as describing friend’s opinions about books, or what someone saw on TV.

Moreover, it has been pointed out by Kobayashi et al. that when making decisions or when evaluating something (books, movies, products), people are always subject to psychological effects caused by external information (Kobayashi et al. 2008). Kahneman and Tversky call this the “cognitive bias” which hinder the perception of pure facts (Kahneman and Tversky, 1972). In situations of decision making on the basis of ambiguous information the existence of the cognitive bias factor causes the “initial value” (person’s background, or what they experienced previously) to affect the final judgment through the “anchor effect” (taking our background for granted). This causes the person to collect or remember only the information that is convenient for them, or to reinforce the prejudicial information, which is also called the “confirmation bias”. The existence of the cognitive bias and related effects influencing the decision-making process of a person becomes a problem in situations of emergency or events of great importance, when obtaining

Copyright © 2014, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

<sup>1</sup>ReTweet, or RT is a function on twitter that allows users to highlight that a tweet contains a citation of a tweet posted by another user.

<sup>2</sup><https://www.facebook.com/>

<sup>3</sup><http://mixi.jp/>

<sup>4</sup><https://twitter.com/>

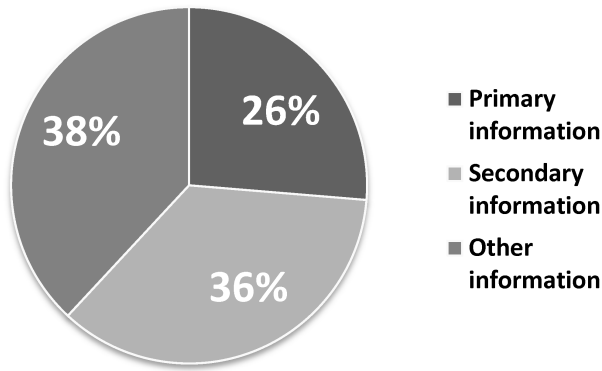


Figure 1: Ratio of primary, secondary and other information in tweets from the time of the Great East Japan Earthquake.

the accurate and unbiased information is crucial for making appropriate judgments.

In this study, in order to obtain accurate information of high uniformity to perform information triage (Marshall and Shipman 1997; Macskassy and Provost 2001; Macskassy et al. 2001), we firstly perform a preliminary study using a sample of tweet logs from the time of the Great East Japan Earthquake. The basic tweet classification rules defined in the preliminary study are further used to classify other tweets by dividing them into representing either primary or secondary information. We also use these rules to analyze tweets from the time of general elections. In the latter we found out that a third kind of information needs to be recognized. We call it the “sesquary” information, and place it between primary and secondary information. We analyze the tweet logs related to elections in accordance with the new rules, and investigate the effectiveness of the classification rules.

## Analysis of Tweets from the Great Earthquake

In this section we perform the study of tweets from the time of the Great East Japan Earthquake and describe the results of classification of those tweets.

We used the data provided by Twitter Japan in the Big Data Project<sup>5</sup>. The tweets represent a time period of one week from March 11 to March 17, from the time of the Great East Japan Earthquake.

After omitting 151 tweets from before the earthquake, we randomly extracted 6,000 tweets and analyzed them manually by six project members (expert annotators, each annotator analyzed 1,000 tweets). The manual analysis revealed that many of the tweets were actually retweets or contained other second-hand information. This lead us to divide the information into primary information and secondary information.

Primary information refers to the kind of information that a person directly saw, heard or personally did. Secondary information refers to indirect information such as re-posting or re-telling what was described by someone else, such as

Table 1: Examples of each type of information found in the Earthquake tweets.

Category	Example / Romanization / Translation
Primary	冷蔵庫あいちやって中全部落ちてきたー
Information	<i>Reizoko aichatte naka zenbu ochitekitayo-</i> My fridge opened and everything falled on the ground! しかしおなかすいたにやー <i>Shikashi onaka suita nya-</i> Oh, but I'm soo hungree 停電きたあ <i>Teiden kitaa</i> Here comes power cut
Secondary	RT @***: 東急戦、世田谷線以外は前線再開。
Information	RT @***:Tokyuu-sen, Setagaya-sen igai wa zen sen saikai. RT @***:Tokyuu-war [line]. All lines except Setagaya back on track. らしい。心配。@RT***:東京来て一番でかい <i>Rashii. Shinpai. RT@***:Tokyo kite ichiban dekai</i> Apparently. I'm worried. RT@***:The biggest since I came to Tokyo RT @***:RT してください!! 全国避難所一覧 <i>RT @***:RT shite kudasai!! Zenkoku hinanjo ichiran</i> RT @***:Please RT!! National list of shelters
Other	@*** (´ロ´) 笑 <i>@*** (´ロ´) warai</i> @*** (´ロ´) laugh くうう・・・ <i>Kuuuu...</i> Kuuuu... (sound of rumbling tummy) 花見だ？ <i>Hanami da to?</i> You wanna go view cherry blossoms [In a situation like this]!?

describing friend’s opinions about books, or what someone saw on TV. It is indirect information described on the Internet by the third party.

The primary information was represented by 1,539 tweets (26%), secondary information was represented by 2,083 tweets (36 %), other kinds of information not coming under the definition of either primary or secondary information was represented by 2,227 tweets (38%) (Fig. 1). An example of each type of information is represented in Table 1.

Primary information is represented by tweets in which users directly describe their own state, such as *samui* (“I’m cold”) or *tsurai* (“It’s so hard...”). There were many tweets of this kind. In such tweets the speaker refers to themselves using first person expressions, such as *watashi* (“I”), or *boku* (“I”, or “me”). Apart from this, a frequent expression appearing in primary information tweets was *nau* (“now”). Also an expression indicating that the user is describing their own present state was the use of rhetorical figure called *taigen-dome* (ending a sentence with a noun or noun phrase often used in Japanese poems, for example “shining stars” instead of “stars are shining”).

Comparing to primary information there was a larger amount of secondary information. However, much of this information was directed outside of the affected areas. Apart from that, there were numerous retweets containing information about the lifeline between the outside areas and the areas affected by the disaster, or in-tweet citations such as *...to no koto* (“it is said that/they say that...”), or *...rashii* (“apparently...”).

<sup>5</sup><https://sites.google.com/site/prj311/>

Table 2: Definition of primary information with examples form election tweets.

Classification criteria	Example / Romanization / Translation
Tweets containing facts which one could directly confirm such as things one personally saw, heard, or did.	青森県内の衆院選の立候補予定者動画を撮影しました / <i>Aomori-ken nai no shūinsen no rikkōho yoteisha dōga wo satsuei shimashita</i> / I took a video of an expected candidate for the House of Representatives elections in Aomori Prefecture
Tweets containing predicative expressions, such as <i>-da</i> , or <i>-dearu</i> .	
Tweets containing unofficial retweets and one's personal opinions about them.  Unofficial retweets with one's own other contents.	支持組織の自治労の不始末を税金突っ込んでリカバリした だけで自慢できる実績ではないよね。RT @  / <i>Shiji soshiki no jichirō no fushimatsu wo zeikin tsukkonde rikabari shita dake de jiman dekiru jisseki dewa nai yo ne. RT @</i> / Just by recovering by putting people's tax money to cover up the misconducts of self-governing body workers of one's supporting organizations is not yet an achievement of which one should be boastful. RT @
Original tweets posted as with an annotation <i>kakusan-kibo</i> ( <i>kakusan-kibō</i> ("spread the news")) is an annotation used only in Japanese tweets which inform other users that the tweet was written to be widely retweeted; similar in English to #RT).	【拡散希望】福岡10区の全候補者の政策を動画でチェック できます。 / <i>[Kakusan-kibō] Fukuoka 10-ku no zen kōho-sha no seisaku wo dōga de chekku dekimasu.</i> / <i>[Kakusan-kibo]</i> You can check the videos showing the policy of all candidates from Fukuoka District 10.
Tweets containing annotation #RT (similar to <i>kakusan-kibo</i> ) except those which have a possibility of rumor (containing phrases like <i>...rashii</i> , or <i>...mitai</i> )	選挙情報サイト「エレクトベディア」サイトの周知に ご協力お願い致します。#RT / <i>Senkyo jōho saito "Erekutopedia" saito no shūchi ni go kyōryoku onegai itashimasu # RT</i> / Please help in making the election information site "Electopedia" widely known. # RT
Tweets containing phrase <i>...nau</i> ("now") (A phrase indicating that a person is doing something at the moment of writing).	明日からの選挙に向けて学習なう。(と言いつつネットなう / <i>Ashita kara no senkyo ni mukete gakushū nau. (To iitsutsu netto nau</i> / Learning for the upcoming elections now. (Actually just surfing on the Web now

## Definition of Rules for Detection of Primary and Secondary Information

Based on the classification results of the tweet logs of the Great East Japan Earthquake, we defined rules (a set of criteria) for the classification of this kind of contents.

In tweets like these, posted during the times requiring urgent decision making, what is important is the accuracy of the information. If one wants to focus on the accuracy information it is useful to classify the contents into primary and secondary information.

Primary information refers to the kind of information that a person directly saw, heard or personally did. Secondary information refers to indirect information such as re-posting or re-telling what was described by someone else (third party), such as describing friend's opinions about books, or what someone saw on TV.

The detailed criteria for classification of primary information are represented in Table 2. The detailed criteria for classification of secondary information are represented in Table 3.

## Analysis of Election Tweets

We analyzed two types of elections taking place in different periods of time. Firstly, we analyzed tweets about Lower House elections (elections to the House of Representatives). Basing on the analysis of those tweets we redefined the classification criteria. Next, we performed additional analysis of tweets about Upper House elections (elections to the House of Councilors).

### Lower House Election Tweets

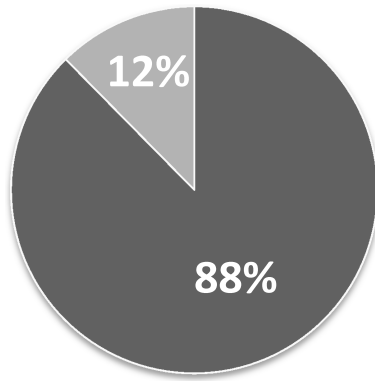
In order to classify the tweets about the general elections, we used a service site called Hashtag Cloud<sup>6</sup>. This site provides all tweets saved from a week back grouped by hashtags. Therefore, we started the search by looking for a hashtag "# general elections" and downloading all tweets containing this hashtag. We began the downloading process on 3rd December 2012 (the day of public announcement of the elections) and started collecting the tweets day by day (the process still continues).

In the analysis we used only the tweets which appeared between 3 to 4 December 2012. There were 1,503 tweets.

<sup>6</sup><http://hashtagcloud.net/>

Table 3: Definition of secondary information with examples form election tweets.

Classification criteria
Tweets citing or referring to news or news sites (by using URL address or phrases such as “News about...”).
Official RT: Tweets which represent a an official twitter citation form containing Somebody else’s tweet. This allows one to easily forward the particular tweet to one’s followers.
Tweets containing phrases indicating second-hand information, such as “I heard/saw that...”, or “Apparently...”.
Unofficial RT not containing one’s personal opinions nor other contents.



■ Primary information  
■ Secondary information

Figure 2: Ratio of primary and secondary information in Lower House election tweets

The tweets obtained by using the “# general elections” hash-tag were longer in number of characters than the length of usual tweets<sup>7</sup> (approximately 30 characters longer). They also contained more decisive and conclusive expressions. Moreover, compared to 23% of the average ratio of replies<sup>8</sup>, the number of replies to election-related tweets was very low, only 5%.

Among all 1,503 tweets, the primary information was represented by 88% (1,317) of the tweets, and the secondary information by merely 12% (186) of the tweets (see Fig.2.)

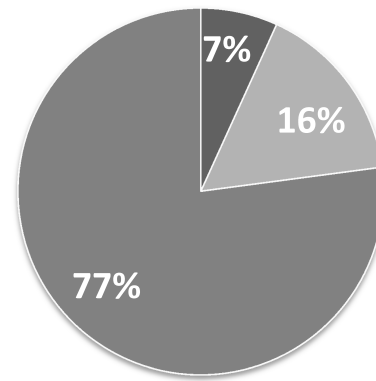
In addition, out of 82 unofficial retweets, 57 have been classified as primary information, out of 64 tweets containing annotation “# RT”, 52 was classified as primary information. In unofficial RT there were many comment supporting the tweets introducing political parties, and personal opinions about tweets regarding amendments to the constitution. In “# RT” there were also people who put a link to the site supporting the election.

In secondary information, there were official RT which copied the poster of the elections, and tweets containing phrases like “They say that there is a an election campaign speech at...”. Moreover, out of total number of 1,503 tweets, 997 contained some kind of URL information, and information about candidates which could be influential on the

<sup>7</sup><http://teapipin.blog10.fc2.com/blog-entry-294.html>

<sup>8</sup><http://b.hatena.ne.jp/entry/www.tommyjp.com/2010/10/7123rt6.html>

tom-



■ Positive ■ Negative ■ Neutral

Figure 3: Classification of Lower House election tweets into positive, negative and neutral.

choice of the a candidate. For example, there were 197 tweets containing links to a list of election candidates divided by each district. This kind of information is often included in primary information tweets and often contains positive or negative opinions about political parties.

Next, from the overall 1,317 primary information tweets we randomly extracted a thousand and classified them into three categories: positive, negative, neutral. Objective opinions were considered as neutral. In the result of classification, out of 1,000 tweets, there was 68 (7%) of positive tweets, 771 (77%) of neutral tweets and 161 (16%) of negative tweets (see Fig. 3).

In positive tweets there were many direct positive expressions about particular political parties, such as “I support the party of...!”. In negative tweets there were many expressions such as “I will never vote for the party of...”. There was also a great deal of neutral tweets which contained lists of candidates for each city district.

### Primary, Secondary and “Sesquiary” Information

Based on the analysis results of tweet logs from both the Great East Japan Earthquake and the general election tweet logs we redefined our classification criteria.

It was not possible to apply to the general election tweets the classification criteria from the Great East Japan Earthquake alone, due to the fact that in tweets about the general elections different information is often considered as important. In the disaster tweets the factual information is the most important. In the election tweets, users often write

Table 4: Redefinition of classification rules

Tweet type	Primary	Sesquiary	Secondary	Example
<ul style="list-style-type: none"> <li>- Factual Information</li> <li>- Description of an action</li> <li>- Decisive expressions</li> <li>- Interview contents</li> <li>- Policy</li> </ul>	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>			<p>These are historically first elections after raising the ban for Internet elections            I went to give my vote            Taro Yamamoto will be elected with no doubt            Making bad guys out of politicians leads to nothing / Interview with Sugawara Taku            Policy 1. of Tamiya Kaichi: Voice of the people is more important than the pressure of large companies            I'm going to the elections!!            Congratulations Taro Yamamoto!! I'm so happy for you!            Taro Yamamoto, talk in a more concise way!            Let's vote everyone without abstention!            Wanna know how Dietmen're really thinking? Check here!</p>
<ul style="list-style-type: none"> <li>- Expression of an intention</li> <li>- Emotional expressions</li> <li>- Opinions</li> <li>- A call to action</li> <li>- Introduction of an URL link</li> <li>- Official RT</li> <li>- Things seen on TV (including facts)</li> <li>- Expressions indicating a rumor</li> <li>- Written reproduction of original information</li> <li>- Citations</li> </ul>		<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/>  <input type="radio"/> <input type="radio"/>  <input type="radio"/>	<p>Here is again Ikegami the stabilizer with a quick report of votes counted!</p> <p>This arse interfering in the city council elections, they say he is a supporter of the opposition</p> <p>Intention to resign in case of defeat in elections in Tokyo - Asahi Shimbun Digital</p> <p>"Its possible to achieve something big only by helping each other". by Nakatomi Noka-mako</p>

about their political preferences, thus it was also important to take into consideration information from the borderline of pure fact and rumor, such as opinions, or emotional attitudes. This kind of information in the Earthquake tweets is mostly considered as noise. However, in election tweets, private opinions and emotional comments could become useful as a reference. Therefore it is important to distinguish this kind of information from the rest and annotate it separately.

To do this we have defined a third kind of information not being primary or secondary information, though keeping a structure of its own, namely, “sesquinary”<sup>9</sup> information (see Table4).

Many tweets represent mixed information. If in order to solve this conflict we applied the following heuristic rules.

1. If a tweet contains different kinds of information, priority is given to the kind of lower density;
2. In cases of only sesquinary and secondary information appearing in a tweet priority is given to secondary information.

## Additional Analysis

Having redefined the original rules we performed additional analysis of the Great Earthquake tweets and the Lower House Election tweets. Moreover we performed additional analysis on completely new data containing tweets from the time of Upper House Elections.

<sup>9</sup>“sesqui-” = 1.5

## Great Earthquake Tweets

Except the tweets we used as the data for classification, we randomly selected another 600 tweets. This data, unlike the data that was collected by limiting the topic to specific keywords, such as “#general elections”, etc., was gathered from all tweets that appeared after the earthquake. Therefore, the average string length of a tweet and the number of replies was close to the general average.

When categorized according to the original definitions, the data contained 30% of primary information (183 tweets), 25% of secondary information (148 tweets), and 45% of other (269 tweets). See Figure 4 for details. The tweets that stood out consisted of incomprehensible entries, such as “@—— njgo” or contained single words, such as *Yahoo-bokin* (“Yahoo donations”).

When categorized according to the redefined rules, the data contained 23% of primary information (141 tweets), 35% of sesquinary information (210 tweets), 25% of secondary information (148 tweets), and 17% of other (101 tweets). See Figure 5 for details. The tweets that stood out contained entries resembling talking to oneself, such as “*Souieba sotsugyoushiki dounaru no darou.*” (“BTW, What will happen with the graduation ceremony?”), or greetings sent to other users, such as “@\_\_\_\_\_ *Iiterasshaaai!*” (“Bon voyage!/See you later!”).

## Lower House Election Tweets

Next we categorized the previously used election tweets according to the redefined rules. The data contained 30% of primary information (449 tweets), 57% of sesquiary in-

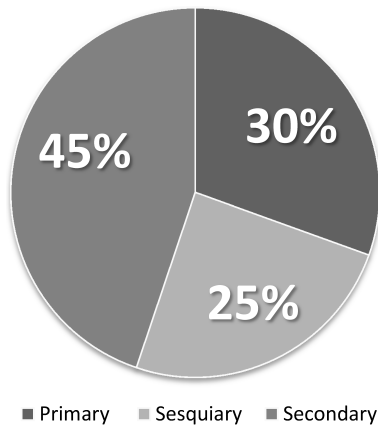


Figure 4: A breakdown of classification of the additional Earthquake tweets according to the original criteria.

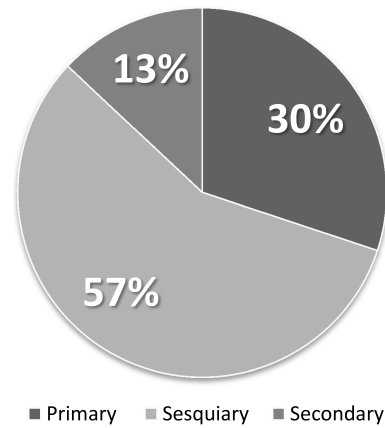


Figure 6: A breakdown of re-classification of the Lower House election tweets according to the redefined criteria.

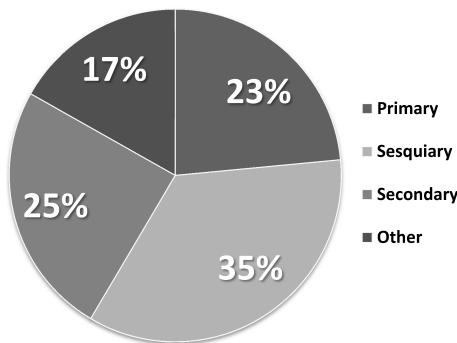


Figure 5: A breakdown of classification of the additional Earthquake tweets according to the redefined criteria.

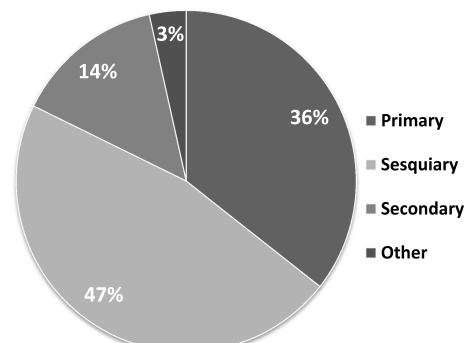


Figure 7: A breakdown of classification of Upper House election tweets (redefined criteria).

formation (849 tweets), and 13% of secondary information (194 tweets). See Figure 6 for details.

There was still a small number of tweets (11 cases, 0.7%) which did not fit into any of the information categories. There were no incomprehensible tweets or greetings like in the Earthquake tweets. However, there were tweets which used hashtags and unofficial retweets, such as “*#Sousenkyo RT @show-you-all : ishin/jimin no ushiro ni wa, Hashimoto Tōru ya...*” (“*#General\_elections RT @show-you-all : Behind the Restoration and the Liberal Democratic Party, Oh, its Tōru Hashimoto...*”)

### Upper House Election Tweets

Using Hashtag Cloud we have collected an additional set of tweets for classification. The set consisted of tweets from the 23rd regular election of members of the Japanese House of Councilors (Upper House election). We collected 22,176 tweets from the time period of 4th of July 2013 (official announcement of the election) to 21st of July 2013.

From the obtained data we extracted 93 Official retweets as secondary information and from the remaining 22,083 we randomly extracted 2,000 tweets for classification. Average length of election tweets this time was 45 characters, which is approximately 15 characters longer than in usual tweets.

This means that election tweets contain more information in general. Primary information was represented by 711 tweets (36%), sesquinary information by 933 tweets (47%) and secondary information by 286 tweets (14%) (see Fig.7).

There were also 70 tweets (3%) which did not fit in any of the three categories. These tweets consisted of only hastags, place names or greetings. Unofficial retweets were in most cases classified as sesquinary information (135 tweets), and sometimes as primary information (29 tweets). As an interesting remark, 90% of all data classified as primary information also contained some amount of secondary information. Moreover, none of the primary information tweets contained either positive or negative contents (all primary information tweets were neutral).

### Discussion

In general, the Great East Japan Earthquake tweets contained much of highly accurate and reliable information (primary information), but they also contained much noise especially with regards to the areas affected by the earthquake.

Within all unofficial retweets from general election for the Lower House (82 tweets), there were 57 tweets (69.5%) which contained primary information. Before the classifica-

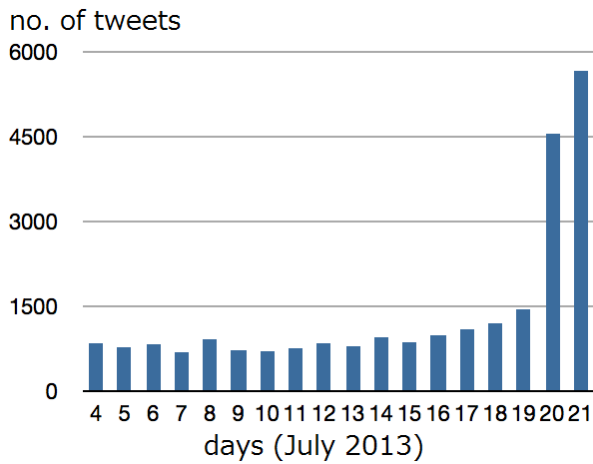


Figure 8: Tweets which appeared in each day till the day of election.

tion we assumed that all or almost all of unofficial retweets would contain primary information. The fact that more than 30% of unofficial retweets contained secondary information was an unexpected result. This means that there were many tweets for which the authors did not want to send an official retweet but still wanted the cited tweet to spread.

Also, in tweets containing URL information, there were many which contained information helpful for voters. This finding could be useful in specifying helpful information in the future.

When focused on the opinionated information within the primary information, 77 % of tweets contained neutral expressions. Neutral tweets contained objective opinions written from neutral perspective, which could become useful for other voters to make their choice in the election. This means that automatically extracting neutral tweets from primary information could help in extraction of useful information in the future. Tweets that contained either positive or negative information usually impose user's individual biased ideas and cannot be considered as useful in gathering information for elections. However, since ratio of positive and negative information for each political party could become a useful information (for example in predicting election results), a deeper analysis and more strict discrimination standards are required.

On 3rd of December (election official announcement day) there were 623 tweets, whereas on 16th December (voting day) there were as many as 13,093 tweets. Number of tweets per day increases as it approaches the voting day, reaches its peak on the two days of 16th and 17th of December and slowly decreases thereafter. From the fact that there was a large number of tweets during the voting days we infer that many users focused on the results of the election. Therefore, there is a possibility that the tweets that appeared around the official announcement day and the voting day are likely to contain different kinds of information.

In the election tweets that were classified by redefined rules based on the analysis results of the two first exper-

iments we were able to extract primary information more precisely. However, there were still tweets which remained unclassified, such as the ones containing greetings. In the future it is necessary to consider how to handle these cases as well

The classification results show that there was as much as 47 % of the sesquary information and the average length of one tweet was longer than usual. This means that people were more interested in expressing their own opinions than retweeting other people's opinions, which indicates an interest in politics in general. However, the third lowest voting rate in the history of postwar Japan<sup>10</sup> does not confirm this interest in actual behavior of voters.

Fig.8 shows that the closer it is to the election day, the number of tweets is increasing, which suggests that people are in general interested in the election as such. Also, the age of twitter users corresponds to the increase in the younger age groups taking part in the elections<sup>11</sup>, which means that SNS, such as twitter could positively influence social and political awareness. Because of this visible influence of SNS on social and political life, on April 19, 2013 Japan allowed for using Internet as a venue for political campaigns. However, this does not necessarily result in people's going to the elections and the number of election tweets cannot be considered as related in any way to the voting rate.

All primary information tweets were neutral. This indicates that by applying the idea of sesquary information we were able to reduce the noise in primary information. Although there were tweets containing both primary and secondary information (for example, "Historically first elections after raising the ban for Internet elections. Let's go voting everyone!"), we did not go into details of those this time. The presence of such cases means that sometimes tweets classified as primary information could contain some amount of positive or negative opinions. This could influence user's cognitive bias and therefore all tweets classified as primary information on the document level should be also in the future re-classified on sentence level.

## Conclusions

In this paper we presented our macroanalysis study of user linguistic behavior in tweets of two kinds: disaster tweets and election tweets. In particular we focused on the time of the Great East Japan Earthquake and the Lower and the Upper House elections.

As a basic idea, we assumed that people use SNS in their decision-making or present status determination. Since it is important in such situation to automatically extract valid primary information, we firstly analyzed the earthquake tweets and defined rules for our classification criteria. Next, we classified election tweets based on these criteria and found out that even in primary information tweets there could be other information which influences reader's cognitive bias, and thus it is necessary to further divide factual information

<sup>10</sup><http://sankei.jp.msn.com/politics/news/130722/elc13072202420023-n1.htm>

<sup>11</sup><http://web-tan.forum.impressrd.jp/e/2012/05/11/12694>

from opinionated information. This kind of information, neither being primary nor secondary information, though still preserving the structure of its own, we called “sesquiary information”. It becomes especially important in the time of elections when people look for opinions about the candidates. However, it is also useful in the time of a disaster. Right after the disaster occurs it is the most important to quickly obtain only the primary information. However, after the first phase of emergency passes and people begin to look for appropriate information to support their further planning and decision making, the sesquiary information greatly gains on importance as well.

Therefore, we redefined the classification rules based on the analysis of the Great East Japan Earthquake tweets and the tweets for the general election to the House of Representatives. We used the new classification rules to reanalyze the both types of data. Moreover, we gathered new data from the time of later election to the House of Councilors. We collected all tweets from 4th to 21 of July appearing under the hashtag “# elections” and analyzed accordingly a randomly selected sample of 2,000 tweets. We were able to confirm the effectiveness of the re-defined classification rules. In a further investigation into the opinionated contents of primary information tweets, we found out that all those tweets were neutral. However, since some of them contained sesquiary information, which could cause cognitive bias on readers, further post-processing of such cases is necessary.

## References

- Eiji Aramaki, Sachiko Maskawa, Mizuki Morita. 2011. Twitter Catches the Flu: Detecting Influenza Epidemics using Factuality Analysis [in Japanese]. *IPSJ SIG Notes*, 2011-SLP-86(1), 1-8, 2011-05-09.
- Tatsuya Fujisaka, Li Yong, Kazutoshi Sumiya. 2010. User Movement Pattern Analysis System Using a Real Space of Microbloggers for Local Event Detection and Property Verification, In *Proceedings of 72nd Annual Meeting of IPSJ*, 2010 (1), “1-845”-“1-846”, 2010-03-08.
- Yusuke Iwaki, Adam Jatowt, Katsumi Tanaka. 2009. Support for Discovery of Useful Articles on Microblogs [in Japanese]. In *Proceedings of the 1st Data Engineering and Information Management Forum (DEIM 2009)*, March 2009.
- Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu, Tiejun Zhao. 2011. Target-dependent Twitter Sentiment Classification. *ACL 2011*.
- D. Kahneman and A. Tversky. 1972. Subjective probability: A judgment of representativeness. *Cognitive Psychology*, pp. 430-454.
- Toshihiro Kamishima. 2007. Problems for Collaborative Filtering: Privacy, Shilling Attack, and Variability of Users’ Ratings [in Japanese]. *IPSJ Magazine*, 48(9), 2007, pp. 966-971.
- Kazuhiro Kazama, Miyuki Imada, Keiichiro Kashiwagi. 2010. Analysis of Information Propagation Network on Twitter [in Japanese]. In *Proceedings of The 24th Annual Meeting of the Japanese Society for Artificial Intelligence*.
- Takuya Kobayashi, Hiroaki Ohshima, Satoshi Oyama and Katsumi Tanaka. 2008. Credibility Improvement of Review Information based on Correction for Biases which resulted from Reviewer’s Profiles and their Regionality. In *Proceedings of the 19 Data Engineering Workshop (DEWS2008)*, B8-4.
- Takamitsu Kuwano, Tamotsu Mitamura, Isao Watanabe, Yasuhiro Suzuki, Takafumi Oobori. 2012. The Study of Tourism Informatics Using Twitter. *Tourism Information Society Journal*, Vol.8, No.1, pp.27-38(2012).
- Sofus A. Macskassy and Foster Provost. 2001. Intelligent Information Triage. *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 318-326.
- Sofus A. Macskassy, Haym Hirsh, Foster Provost, Ramesh Sankaranarayanan, Vasant Dhar. 2001. Information Triage using Prospective Criteria. In *User Modeling 2001 Workshop: Machine Learning, Information Retrieval and User Modeling*.
- Catherine C. Marshall, Frank M. Shipman, III. 1997. Spatial hypertext and the practice of information triage. *Proceedings of the eighth ACM conference on Hypertext (HYPERTEXT’97)*, pp. 124-133.
- Mai Miyabe, Eiji Aramaki, Asako Miura. 2011. Use Trend Analysis of Twitter after the Great East Japan Earthquake [in Japanese]. *IPSJ SIG Notes*, 2011-EIP-53(17), 1-7, 2011-09-08.
- Atsushi Tanaka, Takashi Tajima. 2010. A proposal of classification method for on tweets on twitter [in Japanese]. In *Proceedings of the 2nd Data Engineering and Information Management Forum (DEIM 2010)*, A5-4, 2010-3.
- Katsuyoshi Tateiri. 2011. *A study: What social media or said when the Great East Japan Earthquake happened?* Publ. by Discover 21 Inc.
- Ayana Umejima, Mai Miyabe, Eiji Aramaki, Akiyo Nadamoto. 2011. Tendency of Rumor and Correction Retweet on the Twitter During Disasters [in Japanese]. *IPSJ SIG Notes*, 2011-DBS-152(4), 1-6, 2011-07-26.