

# Detecting Heavy Rain Disaster from Social and Physical Sensors

Tomoya Iwakura\*, Seiji Okajima\*, Nobuyuki Igata\*  
Kunihiko Takeda†, Yuzuru Yamakage†, Naoshi Morita†  
FUJITSU LABORATORIES LTD.\*

{tomoya.iwakura, okajima.seiji, igata}@jp.fujitsu.com  
FUJITSU LIMITED†

{takeda.kunihiko, yamakage.yuzuru, morita.naoshi}@jp.fujitsu.com

## Abstract

We present our system that assists to detect heavy rain disaster, which is being used in real world in Japan. Our system selects tweets about heavy rain disaster with a document classifier. Then, the locations mentioned in the selected tweets are estimated by a location estimator. Finally, combined the selected tweets with amount of rainfall given by physical sensors and a statistical analysis, our system provides users with visualized results for detecting heavy rain disaster.

## 1 Introduction

Every year, Japan suffers natural disasters due to torrential rains that cause river overflows, inundation, and landslides. Information on disasters has conventionally been obtained by using physical sensors such as water/rain gauges, weather radars, and meteorological satellites. However, install of physical sensors requires time and cost, and installing a sufficient number of them is not always possible.

Meanwhile, information about natural disasters began being circulated on social media by users nationwide. Among social media, Twitter<sup>1</sup> has received more attention. A characteristic of Twitter is its real-time nature. For example, Twitter has been used as sensors to detect earthquake (Sakaki et al., 2010), flu epidemic (Aramaki et al., 2011), the amount of pollen (Takahashi et al., 2011), and so on. Due to the nature, Twitter is called as a social sensor.

We present our system that identifies heavy rain disaster with social and physical sensors. Our system selects tweets about heavy rain disaster with natural language processing technologies. By combining selected tweets with data obtained from physical sensors and a statistical analysis, our system provides users with visualized results for detecting heavy rain disaster.

This paper is organized as follows. Section 2 describes an overview of our system. In section 3, NLP analyzers used in our system are described. We compare our system with previous systems in Section 4.

## 2 An Overview of Our System

This section describes our system that can identify heavy rain disaster in the real world from Twitter. Our basic idea is that by selecting tweets that mentioned in tweets and combining the selected tweets with additional information, we can know heavy rain disaster in the real world.

Figure 1 shows an overview of our system. First our system collects tweets with keywords related to heavy rain disaster such as flood, landslide, and so on. Then the searched tweets are selected with NLP analyzers. Our system consists of two main NLP analyzers described in Section 3; a heavy rain information filter and a location estimator. A filter is used to select tweets mentioning heavy rain information and a location estimator annotates the filtered tweets with locations. Then, the filtered tweets are displayed on a map. In addition, combined the

This work is licensed under a Creative Commons Attribution 4.0 International License.  
License details: <http://creativecommons.org/licenses/by/4.0/>.

<sup>1</sup><https://twitter.com/>

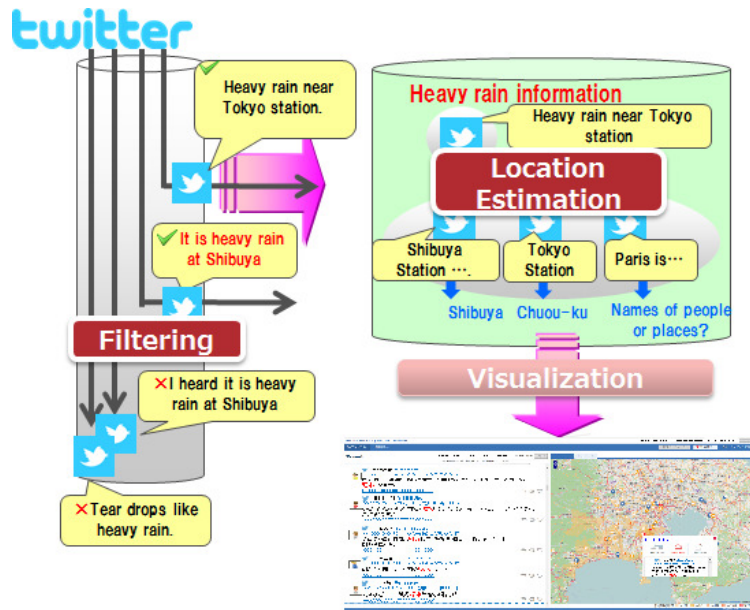


Figure 1: An overview of our system.

selected tweets with amount of rainfall obtained from physical sensors, our system provides users with visualized results for analyzing rain disaster.

Figure 2 shows a snapshot of the display that shows tweets with rainfall data obtained from physical sensors. The display can be used on a Web browser. The screen is divided into the following two displays.

- Time-line display: The left pane of the display shows the time-line. Selected Tweets annotated with location information are displayed in chronological order to facilitate an easy understanding of the latest postings.
- Map display: The map display shows the selected tweets with GPS information, tweets with estimated locations, results of disaster estimation, and the data from physical sensors. The tabs at the top are used to switch the display to a list of images (in thumbnails) that are tagged to the selected tweets. The images are helpful for understanding disaster situations. When a disaster is detected with an anomaly detection described in Section 3, the disaster alert is turned on. The regions estimated as disaster are displayed with a different color on the map like a heatmap.

### 3 Basic Technologies

This section describes two NLP, filtering and location estimation, and an anomaly detection.

#### 3.1 Filtering

A filter selects heavy rain information from tweets. Even if tweets include words that indicate heavy rain, all the such tweets are not useful for detecting heavy rain disaster. Therefore, we select tweets about reports of sighting of heavy rains; in other words, we exclude information of heavy rains included in news, TV programs, hearsay, and so on.

For example, “Heavy rain made me wet at Shibuya.” is a report of sighting of heavy rains. However, “I watched a news about heavy rain on TV.” is not. We distinguish such news because such tweets other than sighting of heavy rains do not contribute to know information about heavy rains and become noises.

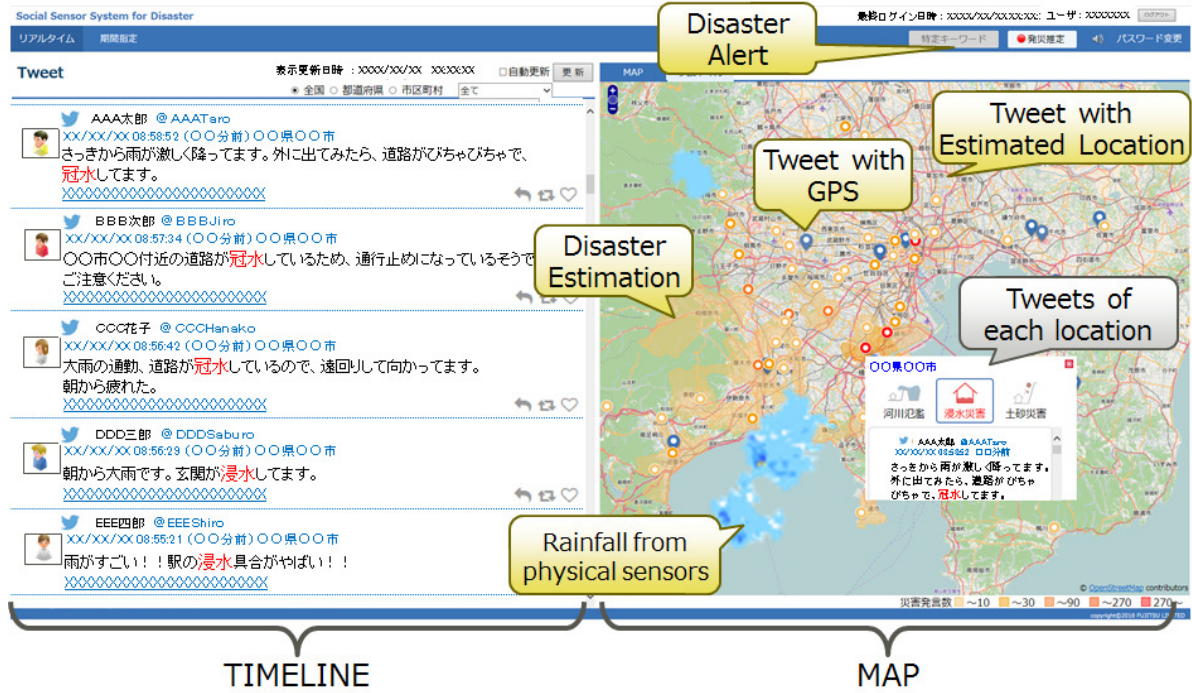


Figure 2: Our interface for displaying a map with selected tweets and rainfalls.

We used a document classifier based on an extended version of (Iwakura, 2017) as a filter. The filter accepts directed graph-based texts, represented as lattices of words with their part-of-speech (POS) tags as an input. A set of  $n$ -grams that consist of words and POS tags are learned as rules by the machine learning algorithm.

### 3.2 Location Estimation

Our location estimator annotates tweets with the latitude and longitude of each location that each tweet mentions. To estimate a location or locations mentioned in a tweet, we use a location estimator that uses dictionary and a machine-learning (Okajima and Iwakura, 2018).

First, our method recognizes candidate locations with a dictionary. Then, in order to filter out irrelevant location given by the dictionary, we use Japanese prefectures referred by a given text. For example, while there are Minato wards in Tokyo, Nagoya, and Osaka, the system can identify which city a mentioned Minato ward is in based on context. We can efficiently filter out irrelevant locations of tweets given by the dictionary with Japanese prefectures referred by tweets because there are almost no cities that have the same name in the same prefectures in Japan.

Prefectures referred by tweets is estimated by a classifier created from automatically generated training data. Considering words included in tweets, we estimated Japanese prefectures referred by tweets.

### 3.3 Anomaly Detection

We use a heavy rain detection-based on an anomaly detection. As described above, tweets about heavy rains from news, TV programs and hearsay become noises. Therefore, our method identifies heavy rains of each prefecture from tweets selected with the filter and the location estimator. As described in Figure 1, our filter selects tweets mentioned to heavy rains first. Then the location estimator identifies locations mentioned by the selected tweets. Finally, a heavy rain detector estimates whether each Japanese prefecture has heavy rain disaster or not. The heavy rain detector assumes a Poisson distribution and a probability given to a number of

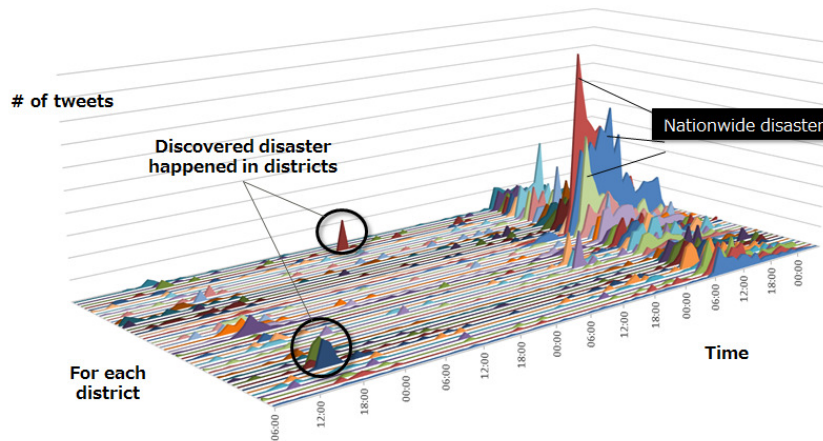


Figure 3: An image of heavy rain detection of each Japanese prefecture with an anomaly detection

tweets of each prefecture on the distribution is used for detecting heavy rains.

Figure 3 shows an image of the heavy rain detector. With estimated locations of tweets, we can identify heavy rain of each prefecture.

## 4 Related Work

One of the differences from the previous systems is the machine-learning method for filtering. Previous systems (Sakaki et al., 2010; Takahashi et al., 2011) use machine learning algorithms such as SVMs (Platt, 1999) and a boosting-based learner (Iwakura and Okamoto, 2008) that learn models for classifying texts represented by bag-of-words.

In contrast, our system uses a machine-learning-algorithm handles semi-structured texts. By handling semi-structured text, we can consider important substructures of semi-structured texts (Iwakura, 2017).

Another difference is the location estimation. Previous systems also used location information. However, they used locations extracted from user profiles such as GPS information of tweets and user profile (Sakaki et al., 2010), or prefecture level locations identified from user profiles (Takahashi et al., 2011). In contrast, when identifying heavy rain events, we have to identify detailed location information that the events happened. To identify detailed location information, we used a dictionary and context-information (Okajima and Iwakura, 2018).

DISANNA (Mizuno et al., 2016) also analyzes tweets in real time, discovers disaster-related information, and presents it in organized formats based on given queries.

Compared with DISANNA, our system can be used without specifying queries because our system focuses on heavy rain detection and tweets are selected by a filter for the heavy rain detection. In addition, our system also incorporates additional information such as rainfall amounts obtained from real sensors and alerting information.

## 5 Conclusion

This paper has presented our system that identifies heavy rains by analyzing social media with Natural Language Processing technologies combined with rainfall amounts obtained from physical sensors.

## References

Eiji Aramaki, Sachiko Maskawa, and Mizuki Morita. 2011. Twitter catches the flu: Detecting influenza epidemics using twitter. In *EMNLP*, pages 1568–1576.

- Tomoya Iwakura and Seishi Okamoto. 2008. A fast boosting-based learner for feature-rich tagging and chunking. In *Proc. of CoNLL'08*, pages 17–24.
- Tomoya Iwakura. 2017. Efficient training of adaptive regularization of weight vectors for semi-structured text. In *Proc. of PAKDD'17*, pages 261–272.
- Junta Mizuno, Masahiro Tanaka, Kiyonori Ohtake, Jong-Hoon Oh, Julien Kloetzer, Chikara Hashimoto, and Kentaro Torisawa. 2016. WISDOM x, DISAANA and D-SUMM: large-scale NLP systems for analyzing textual big data. In *Proc. of COLING'16 (demo)*, pages 263–267.
- Seiji Okajima and Tomoya Iwakura. 2018. Japanese place name disambiguation based on automatically generated training data (to appear). In *Proc. of CICLING'18*.
- John C. Platt. 1999. Fast training of support vector machines using sequential minimal optimization. In Bernhard Schölkopf, Christopher J.C. Burges, and Alexander J. Smola, editors, *Advances in Kernel Methods: Support Vector Learning*, pages 185–208. MIT Press.
- Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. 2010. Earthquake shakes twitter users: real-time event detection by social sensors. In *WWW*, pages 851–860.
- Tetsuro Takahashi, Shuya Abe, and Nobuyuki Igata. 2011. Can twitter be an alternative of real-world sensors? In *HCI (3)*, pages 240–249.