Real-time Local Topic Extraction using Density-based Adaptive Spatiotemporal Clustering for Enhancing Local Situation Awareness

Tatsuhiro Sakai, Keiichi Tamura, Shota Kotozaki, Tsubasa Hayashida and Hajime Kitakami
Graduate School of Information Sciences, Hiroshima City University, Hiroshima, Japan

Keywords: Spatiotemporal Analysis, Geotagged Tweets, Local Topic Extraction, Social Data Mining, Big Data Analysis, Spatiotemporal Clustering.

Abstract: In the era of big data, we are witnessing the rapid growth of a new type of information source. In particular, tweets are one of the most widely used microblogging services for situation awareness during emergencies. In our previous work, we focused on geotagged tweets posted on Twitter that included location information as well as a time and text message. We previously developed a real-time analysis system using the (ε,τ)-density-based adaptive spatiotemporal clustering algorithm to analyze local topics and events. The proposed spatiotemporal analysis system successfully detects emerging bursty areas in which geotagged tweets related to observed topics are posted actively; however the system is tailor-made and specialized for a particular observed topic, therefore, it cannot identify other topics. To address this issue, we propose a new real-time spatiotemporal analysis system for enhancing local situation awareness using a density-based adaptive spatiotemporal clustering algorithm. In the proposed system, local bursty keywords are extracted and their bursty areas are identified. We evaluated the proposed system using actual real world topics related to weather in Japan. Experimental results show that the proposed system can extract local topics and events.

1 INTRODUCTION

One of the most interesting emerging topics in big data analysis of social media is that social data posted on social media can be applied to situation awareness during real world topics and events (Yin et al., 2012). In particular, during natural disasters such as earthquakes, typhoons, floods, and heavy snow storms, people actively post messages that mention the situations they are facing through social media sites. This trend has been encouraged by the increasing popularity of a new type of data on social media: geo-annotated social data, which is also referred to as geo-referenced social data (Naaman, 2011). Moreover, it creates new technical challenges, such as how to show where and when events occur. These new techniques help users to better understand their local situation.

In our previous work we focused on geotagged tweets posted on Twitter that included location information as well as a time and text message. Geotagged tweets are referred to as spatiotemporal documents because we can analysis topics and events spatiotemporally using them. We proposed a spatiotemporal clustering algorithm called the (ε,τ)-density-based adaptive spatiotemporal clustering algorithm that allows us to extract spatiotemporal clusters in which geotagged tweets are actively posted. Moreover, we developed a real-time analysis system using the (ε,τ)-density-based adaptive spatiotemporal clustering algorithm to analyze local topics and events. The proposed spatiotemporal analysis system is successful in detecting emerging bursty areas in which geotagged tweets related to observed topics are posted actively.

The real-time analysis system proposed in our previous work allows us to enhance local situation awareness; however, the system requires the keywords related to the observed topics to be specified in advance. If the system is specialized for a particular observed topic, it cannot identify other topics, even though some local bursty keywords related to emerging local topics and events are posted around users. To address this issue, we propose a new real-time spatiotemporal analysis system for enhancing local situation awareness. This method is based on a density-based adaptive spatiotemporal clustering algorithm.

Our new real-time spatiotemporal analysis system is composed of two techniques: quartile-based outlier detection to identify bursty local keywords and density-based adaptive spatiotemporal clustering to identify bursty local areas. In our new system, lo-

Sakai, T., Tamura, K., Kotozaki, S., Hayashida, T. and Kitakami, H..
Real-time Local Topic Extraction using Density-based Adaptive Spatiotemporal Clustering for Enhancing Local Situation Awareness.
In *Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2015) - Volume 1: KDIR*, pages 203-210
ISBN: 978-989-758-158-8
Copyright © 2015 by SCITEPRESS – Science and Technology Publications, Lda. All rights reserved

cally frequent keywords in geotagged tweets that are within a particular distance from a user are first extracted. To determine whether local frequent keywords are bursty keywords or routine keywords, we utilize quartile-based outlier detection (Hyndman and Fan, 1996). Moreover, bursty local areas related to extracted local bursty keywords are identified using density-based adaptive spatiotemporal clustering.

The remainders of this paper is organized as follows. In Section 2, related work is reviewed. In Section 3, we propose a new real time spatiotemporal analysis system. In Section 4, we explain the method for detecting bursty local keywords. In Section 5, the (ε,τ)-density-based adaptive spatiotemporal clustering algorithm is described briefly. In Section 6, experimental results and case studies are reported. In Section 7, we conclude this paper.

2 RELATED WORK

In the era of big data, social media is expected to enhance the situation awareness of local topics. In particular, many researchers focus on natural disasters and have developed awareness systems for natural disasters such as earthquakes, typhoons, floods, and diseases (Yin et al., 2012). During natural disasters, users often post text messages through social media about things that they are witnessing(Hui et al., 2012), (Kreiner et al., 2013), (Mendoza et al., 2010).

Some of the most successful proposals concern crisis management systems for earthquakes, floods, and epidemics. Sakaki et al. (Sakaki et al., 2010) focused on a method for predicting earthquake epicenters by using geotagged tweets regarding earthquakes. Avvenuti et al. (Avvenuti et al., 2014) developed an earthquake alert and report system that can identify damage in earthquake-affected areas. Vieweg et al. (Vieweg et al., 2010) showed that information related to emergency situations is posted on Twitter during emergencies such as floods and fires. Moreover, Hwang et al. (Hwang et al., 2013) observed flu epidemics using a spatiotemporal analysis of social media geostreams.

Kim et al. (Kim et al., 2011) introduced mTrend, which constructs and visualizes spatiotemporal topic trends, referred to as "topic movements." mTrend is not a tailor-made system; however, it cannot analyze bursty areas of local topics and events. Thom et al. (Thom et al., 2012) presented a system that extracts anomalies from geolocated Twitter messages and visualizes them using term interactive clouds. This system does not address spatiotemporal analysis. Kumar et al. (Kumar et al., 2014) detected road hazards by aggregating hazard-related information posted by Twitter users. This system is was tailor-made and it could not extract any hazards other than road-related topics.

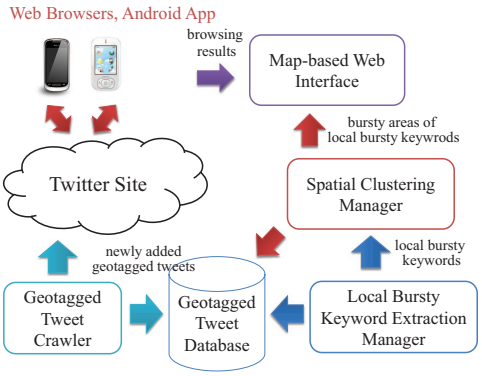

Figure 1: System Overview.

3 SYSTEM OVERVIEW

In this section, we present an overview proposed real-time spatiotemporal analysis system.

3.1 Sequence Geotagged Tweets

In this study, we focus on geotagged tweets posted on Twitter. Let gt_i denote the i-th geotagged tweet in a set of geotagged tweets $SGT = \{gt_1, gt_2, \cdots, gt_t\}$; then, gt_i consists of four items: $gt_i =< text_i, pt_i, pl_i, photo_i >$, where $text_i$ is a short text message, pt_i is the time when the geotagged tweet was posted, pl_i is the location where gt_i was posted or is located (i.e., latitude and longitude), and $photo_i$ is an attached photo.

3.2 Components

In our system, spatiotemporal clusters of local bursty keywords related to topics and events are extracted as bursty areas in real time. Moreover, to visualize bursty areas, our system provides a map-based user interface. There are four managers in our system (Figure 1).

- The *Geotagged Tweet Crawler* crawls geotagged tweets from Twitter feed via its Streaming APIs. Geotagged tweets are stored in a geotagged tweet database.
- The *Local Bursty Keyword Extraction Manager* generates a sequence of geotagged tweets that are located within r distance of a user. This sequence is stored in SGT. First, this manager extracts locally frequent keywords

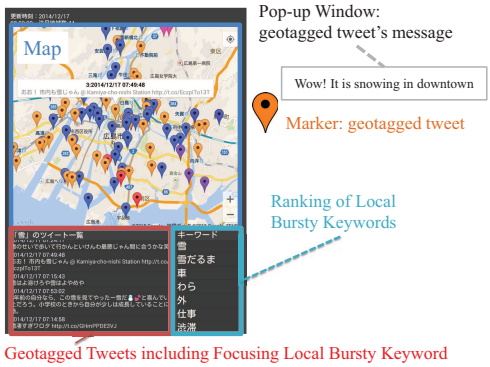

Figure 2: Interfaces.

$LFK^{(c)} = \{lfk_1^{(c)}, lfk_2^{(c)}, \cdots, lfk_l^{(c)}\}$ from $SGT^{(c)}$, where $SGT^{(c)} = \{gt_k | gt_k \in SGT$ and pt_k is within one hour of the current time$\}$. It then extract local bursty keywords $LBK^{(c)} = \{lbk_1^{(c)}, lbk_2^{(c)}, \cdots lbk_m^{(c)}\}$ from $LFK^{(c)}$ using quartile-based outlier detection. The details of this manager are further explained in Section 4.

- The *Spatial Clustering Manager* identifies bursty local areas as spatial clusters using (ε, τ)-density-based adaptive spatiotemporal clustering for each local bursty keyword. For each $lbk_i^{(c)} \in LBK^{(c)}$, (ε, τ)-density-based adaptive spatiotemporal clusters are extracted from $SKGT^{(key)} = \{gt_k | gt_k \in SGT$ *and* $key \in text_k\}$, where $key = lbk_i^{(c)}$. The details of this manage are further explained in Section 5.
- The *Map-based Web Interface* provides the user interface for browsing extracted spatial clusters.

3.3 User Interface

Figure 2 shows a screen shot of the *Map-based Web Interface* which is an Android application. There is a map, a list of geotagged tweets of a current focus local bursty keyword, and a ranking of local bursty keywords on the Android application interface. For each local bursty keyword, bursty local areas and each geotagged tweets in extracted spatiotemporal clusters within a particular distance from a user are shown on the Google map. A marker represents a geotagged tweet and a pop-up window appears when a user touches the marker.

4 LOCAL BURSTY KEYWORD EXTRACTION

We extract local bursty keywords from locally frequent keywords that appear frequently in geotagged tweets near a user. There are many routine keywords (e.g., greetings) in the set of locally frequent keywords. Therefore, we have to remove routine keywords from the set of local bursty keywords. For example, suppose that a set of frequent keywords is $\{$*"good morning," "heavy rainfall," "delay"*$\}$. The keyword *"good morning"* is a routine keyword that appears frequently everyday. Therefore, *"good morning"* is regarded as a routine keyword.

It is difficult to determine whether a keyword is a routine keyword or not because the frequencies of keywords change dynamically depending on time and location. To determine if locally frequent keywords are routine keywords, we utilize the quartile-based outlier detection. The local bursty keyword extraction comprises two steps: (1) keyword frequency counting and (2) routine keyword removal.

To count keyword frequency, for each keyword that appears in the geotagged tweets, the frequency of the keyword is counted per day and the total is stored in $FCD^{(key)}$. Let $FCD^{(key)} = (fcd_1^{(key)}, fcd_2^{(key)}, \cdots, fcd_t^{(key)})$ be a sequence of the daily frequencies of the keyword *key*. Moreover, for each keyword that appears in geotagged tweets, the frequency of keyword is counted par one hour. Let $FCH^{(key)} = (fch_1^{(key)}, fch_2^{(key)}, \cdots, fch_t^{(key)})$ be a sequence of the hourly frequencies of keyword *key*.

If $fch_c^{(key)}$ is larger than a user-given threshold $minf$ and $fch_c^{(key)} \geq \sum_{k=c-25}^{c-1} fch_k^{(key)}/24$, the keyword is a locally frequent keyword. Thus,

$$\begin{aligned} LFK^{(c)} = \{ & lfk_i^{(c)} | \\ & \exists gt_j^{(c)} \in SGT^{(c)} \text{ includes } lfk_i^{(c)}, \\ & fch_c^{(lfk_i^{(c)})} \geq minf \text{ and} \\ & fch_c^{(lfk_i^{(c)})} \geq \sum_{k=c-25}^{c-1} fch_k^{(lfk_i^{(c)})}/24\} \end{aligned} \quad (1)$$

In the second step, for each locally frequent keyword in $LFK^{(c)}$, locally frequent keyword $lfk \in LFK^{(c)}$ is removed if it is a routine keyword. Let the lower quartile, second quartile, and third quartile of $FCD^{(lfk)}$ be $Q_1^{(lfk)}$, $Q_2^{(lfk)}$, and $Q_3^{(lfk)}$, respectively. Consider as an example sequence $FCD^{(lfk_1)} = \{8, 12, 17, 10, 11, 13, 14\}$. The second quartile $Q_2^{(lfk_1)}$ is the median of $FCD^{(lfk_1)}$; therefore $Q_2^{(lfk_1)} = 12$. Moreover, $Q_1^{(lfk_1)} = 10$ and $Q_3^{(lfk_1)} = 14$.

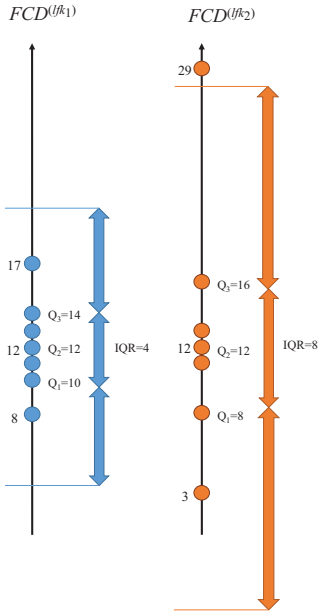

Figure 3: Detecting local bursty keywords.

The range $|Q_3^{(lfk)} - Q_1^{(lfk)}| = IQR^{(lfk)}$ is called the Interquartile Range (IQR). The distribution of $FCD^{(lfk)}$ is used to detect bursty keywords. If distribution of $FCD^{(lfk)}$ is small, this means that lfk appears constantly in geotagged tweets, but not constantly.

Definition 1 (Local Bursty Keyword)**.** If a locally frequent keyword lfk satisfies at least of the following conditions, lfk is a bursty local keyword.

(1) $\exists fcd_i^{(lfk)} \in FCD^{(lfk)}$, $fcd_i^{(lfk)} - Q_3^{(lfk)} > IQR^{(lfk)} \times 1.5$.

(2) $\exists fcd_i^{(lfk)} \in FCD^{(lfk)}$, $Q_1^{(lfk)} - fcd_i^{(lfk)} > IQR^{(lfk)} \times 1.5$

Figure 3 shows an example of local bursty keyword detection. Consider two example sequences $FCD^{(lfk_1)} = \{8,12,17,10,11,13,14\}$, and $FCD^{(lfk_2)} = \{3,12,29,8,11,13,16\}$. For the first sequence, as $IQR^{(lfk_1)} = 4$, there is no data that satisfies definition 1; therefore lfk_1 is a routine keyword. For the second sequence, $IQR^{(lfk_2)} = 8$ and $29-16 > IQR^{(lfk_2)} \times 1.5$; therefore lfk_2 is not a routine keyword.

5 ADAPTIVE SPATIOTEMPORAL CLUSTERING BASED ON (ε,τ)-DENSITY

Local bursty keywords help us to be aware of what is happing near users; however, it is difficult to determine in which places the keywords are occurring. The *Spatial Clustering Manager* identifies bursty local areas as spatial clusters using (ε,τ)-density-based adaptive spatiotemporal clustering.

Adaptive spatiotemporal clustering based on (ε,τ)-density is a natural extension of the density-based criteria proposed by Ester et al. (Ester et al., 1996),(Sander et al., 1998). In this section, (ε,τ)-density-based adaptive spatiotemporal clustering is explained briefly and we show how to use it to extract bursty areas of local bursty keywords.

In density-based spatial clustering, for each data point within a spatial cluster, the neighborhood of a user-defined radius must contain at least a minimum number of points; thus, the density in the neighborhood must exceed some predefined threshold. This density criterion allows us to recognize areas in which densities are higher than in other areas. However, it does not consider temporal changes correctly. It is important to analyze temporal changes to extract local topics and events to enhance situation awareness in real time. In contrast, the (ε,τ)-density-based adaptive spatiotemporal clusters cover spatiotemporal clusters that are both temporally and spatially separated from other spatiotemporal clusters.

5.1 Definitions

There are several density-based adaptive spatiotemporal criteria that can be used to define (ε,τ)-density-based adaptive spatiotemporal clusters.

Suppose that a keyword and the set of geotagged tweets including that keyword are *key* and $SKGT^{(key)} \in SGT$, respectively. Let $kgt_j^{(key)} = gt_{\phi^{(key)}(j)}$ be a keyword-related geotagged tweet that has keyword *key* included in its $text_{\phi^{(key)}(j)}$. A sequence of keyword-related geotagged tweets is $SKGT^{(key)} = (kgt_1^{(key)}, \cdots, kgt_m^{(key)}\}$, where $SKGT^{(key)} \in SGT$. Further, function $\phi^{(key)}(j)$ is an injective function:

$$\phi^{(key)}(j): \quad SKGT^{(key)} \to SGT;\ kgt_j^{(key)} \mapsto gt_{\phi^{(key)}(j)} \qquad (2)$$

Definition 2 ((ε,τ)-density-based Neighborhood)**.** The (ε,τ)-density-based neighborhood of a relevant geotagged tweet $kgt_p^{(key)}$, which is denoted by $KSTN_{(\varepsilon,\tau)}(kgt_p^{(key)})$, is defined as

$$KSTN_{(\varepsilon,\tau)}(kgt_p^{(key)}) = \{kgt_q^{(key)} \in SKGT^{(key)} \mid dist(kgt_p^{(key)}, kgt_q^{(key)}) \le \varepsilon \text{ and } iat(kgt_p^{(key)}, kgt_q^{(key)}) \le \tau\}, \qquad (3)$$

where the function $dist()$ returns the distance between $kgt_p^{(key)}$ and $kgt_q^{(key)}$, and the function $iat()$ returns the interarrival time between them.

The local spatiotemporal density of a keyword-related geotagged tweet $kgt_p^{(key)}$ is denoted as $lstd(kgt_p^{(key)})$. The spatiotemporal space is divided into several spatiotemporal grids in three-dimensional space. The number of spatiotemporal grids is $div_{lng} \times div_{lat} \times div_{time}$, where lng, lat and $time$ are longitude, latitude and posted time, respectively. For each spatiotemporal grid, the number of geotagged tweets posted in the past is calculated. The degree of local spatiotemporal density of a geotagged tweet is the normalized value of the number of geotagged tweets:

$$lstd(kgt_p^{(key)}) = \frac{stnum(geo\_gid(kgt_p^{(key)})) - stnum_{min}}{stnum_{max} - stnum_{min}}, \quad (4)$$

where $stnum(i)$ returns the number of geotagged tweets in the i-th grid. Function $geo\_gid(kgt_p^{(key)})$ returns the grid ID where $kgt_p^{(key)}$ is located. Furthermore, $stnum_{min}$ and $stnum_{max}$ are the minimum and maximum values, respectively.

Definition 3 (Adaptive Threshold)**.** The minimum number of keyword-related geotagged tweets is called the adaptive threshold ATH, defined as follows.

$$ATH(kgt_p^{(key)}, MinKGT) = (MinKGT - 1) \times lstd(kgt_p^{(key)}) + 1, \quad (5)$$

where the function $lstd()$ returns the degree of local spatiotemporal density $(0 \leq lstd(kgt_p^{(key)}) \leq 1.0)$.

A keyword-related geotagged tweet $kgt_p^{(key)}$ is called a core-keyword-related geotagged tweet if there is at least $ATH(kgt_p^{(key)}, MinKGT)$ in $KSTN_{(\varepsilon,\tau)}(kgt_p^{(key)})$ such that $(|KSTN_{(\varepsilon,\tau)}(kgt_p^{(key)})| \geq ATH(kgt_p^{(key)}, MinKGT))$. Otherwise, $kgt_p^{(key)}$ is called a border-keyword-related geotagged tweet.

Definition 4 ((ε,τ)-density-based Directly Adaptive Reachable)**.** Suppose that a keyword-related geotagged tweet $kgt_q^{(key)}$ is in the (ε,τ)-density-based neighborhood of $kgt_p^{(key)}$. If $|KSTN_{(\varepsilon,\tau)}(kgt_p^{(key)})| \geq ATH(kgt_p^{(key)}, MinKGT)$, $kgt_q^{(key)}$ is (ε,τ)-density-based directly adaptive reachable from $kgt_p^{(key)}$.

Definition 5 ((ε,τ)-density-based Adaptive Reachable)**.** Suppose that there is a sequence of keyword-related geotagged tweets $(kgt_p^{(key)}, kgt_{(p+1)}^{(key)}, \cdots, kgt_{(p+l)}^{(key)})$ and the $(p+i)$-th keyword-related geotagged tweet $kgt_{(p+i+1)}^{(key)}$ is (ε,τ)-density-based directly adaptive reachable from the $(p+i)$-th keyword-related geotagged tweet $kgt_{(p+i)}^{(key)}$. The keyword-related geotagged tweet $kgt_{(p+l)}^{(key)}$ is (ε,τ)-density-based adaptive reachable from $kgt_p^{(key)}$.

Definition 6 ((ε,τ)-density-based Adaptive Connected)**.** Suppose that the keyword-related geotagged tweets $kgt_p^{(key)}$ and $kgt_q^{(key)}$ are (ε,τ)-density-based adaptive reachable from an arbitrary keyword-related geotagged tweet $kgt_o^{(key)}$. If $|KSTN_{(\varepsilon,\tau)}(kgt_o^{(key)})| \geq ATH(kgt_o^{(key)}, MinKGT)$, $kgt_p^{(key)}$ is (ε,τ)-density-based adaptive connected to $kgt_q^{(key)}$.

5.2 Adaptive Spatiotemporal Clusters based on (ε,τ)-Density

An (ε,τ)-density-based adaptive spatiotemporal cluster is defined as follows:

Definition 7 ((ε,τ)-density-based adaptive spatiotemporal cluster)**.** An (ε,τ)-density-based adaptive spatiotemporal cluster for an keyword key ($ASTC^{(key)}$) in $SKGT^{(key)}$ satisfies the following restrictions:

- **i1j** $\forall kgt_p^{(key)}$, $kgt_q^{(key)} \in SKGT^{(key)}$, if and only if $kgt_p^{(key)} \in ASTC^{(key)}$, $kgt_q^{(key)}$ is (ε,τ)-density-based adaptive reachable from $kgt_p^{(key)}$, and $kgt_q^{(key)}$ is also in $ASTC^{(key)}$.
- **i2j** $\forall kgt_p^{(key)}$, $kgt_q^{(key)} \in ASTC^{(key)}$, $kgt_p^{(key)}$ is (ε,τ)-density-based adaptive connected to $kgt_q^{(key)}$.

5.3 Algorithm

Algorithm 1 describes the algorithm for (ε,τ)-density-based adaptive spatiotemporal clustering for extracting bursty areas of local bursty keyword key. In this algorithm, for each geotagged tweet gtp in $SKGT^{(key)}$, the function **IsClustered** checks whether gtp is already assigned to a spatiotemporal cluster. The (ε,τ)-density-based neighborhood of gtp is then obtained using the function **GetNeighborhood**. If gtp is a core keyword-related geotagged tweet, it is assigned to a new spatiotemporal cluster, and all the neighbors are queued to Q for further processing. The processing and assignment of keyword-related geotagged tweets to the current spatiotemporal cluster continues until the queue is empty. First, the next keyword-related geotagged tweet is dequeued from queue Q. If the dequeued keyword-related geotagged tweet is not already assigned to the current spatiotemporal cluster, it is assigned to the current

input : $SKGT^{(key)}$ - a set of geotagged tweet including keyword *key*, ε - neighborhood radius, τ - interarrival time, $MinKGT$ is threshold value
output: $STC^{(key)}$ - set of spatiotemporal clusters

$cid \leftarrow 1$;
$STC^{(key)} \leftarrow \phi$;
for $i \leftarrow 1$ **to** $|SKGT^{(key)}|$ **do**
  $gtp \leftarrow gt_i^{(key)} \in SKGT^{(key)}$;
  if `IsClustered`$(gtp) == false$ **then**
    $N \leftarrow$ `GetNeighborhood`(gtp,ε,τ);
    if $|N| \geq ATH(gtp, MinKGT)$ **then**
      $stc_{cid}^{(key)} \leftarrow$`MakeNewCluster`$(cid,gtp)$;
      $cid \leftarrow cid + 1$;
      `EnQueue`(Q,N);
      while Q *is not empty* **do**
        $gtq \leftarrow$ `DeQueue`(Q);
        $stc_{cid}^{(key)} \leftarrow stc_{cid}^{(key)} \cup gtq$;
        $N \leftarrow$ `GetNeighborhood`(gtq,ε,τ);
        if $|N| \geq ATH(gtp, MinKGT)$ **then**
          `EnNniqueQueue`(Q,N);
        end
      end
      $STC^{(key)} \leftarrow STC^{(key)} \cup stc_{cid}^{(key)}$;
    end
  end
end
return $STC^{(key)}$;

Algorithm 1: (ε,τ)-density-based adaptive spatiotemporal clustering algorithm.

spatiotemporal cluster. If the dequeued keyword-related geotagged tweet gtq is a core keyword-related geotagged tweet, the keyword-related geotagged tweets in the (ε,τ)-density-based neighborhood of gtq are then queued in queue Q using the function **EnNniqueQueue**, which places the input keyword-related geotagged tweets into queue Q if they are not already in it Q.

6 EXPERIMENTS

We implemented our proposed system and evaluated it.

6.1 Experiment Setup

We evaluated our proposed system by a case study. The parameters in the experiments were set as follows: ε is 5km, τ is 3600sec, and $MinKGT$ is 5. Moreover, the user location was set to (34.578618, 132.796105). In the experiments, we used geotagged tweets that were located within 70 km of the user. The user-given threshold $minf$ is 5.

In the (ε,τ)-density-based adaptive spatiotemporal cluster, local spatiotemporal densities are required. To calculate local spatiotemporal densities, we used 3,301,605 geotagged tweets from December 13 to December 23, 2013 and counted in each spatiotemporal grids. We considered the spatiotemporal space for local spatiotemporal densities, which is a rectangle consisted of the westernmost point (24.4494, 122.93361) and the northernmost point (45.5572, 148.752) of Japan. This rectangle was equally divided into several spatiotemporal grids of $div_{lng} = 1,000$, $div_{lat} = 1,000$ and $div_{time} = 24$.

6.2 Case Studies

In this experiment, we confirmed our new system could identify two local topics related to natural disasters. The first local topic is heavy snowfall on December 17, 2014. A explosive cyclogenesis called a “bomb” cyclone hit the region of Japan. This explosive cyclogenesis brought heavy snowfall in Hi-

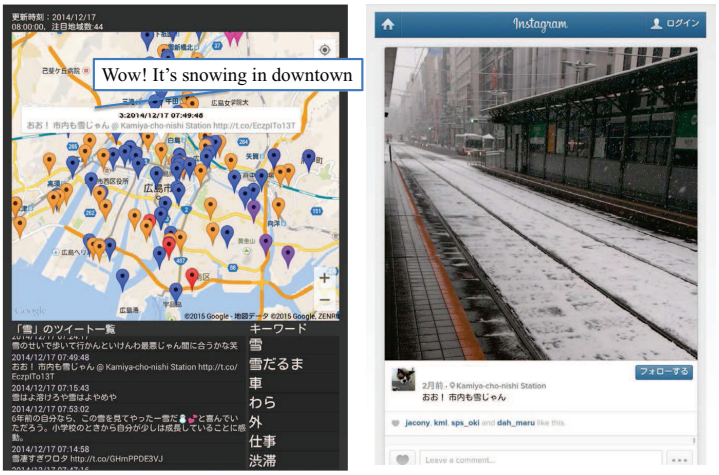

Figure 4: Case-1.

Figure 5: Case-2.

roshima, which is user's location. Figure 4 shows a screen shot of the Web-based interfaces on the Android application. The locally bursty keyword include "snow," "snow man" and "traffic jam." These keywords are related to the snowfall. In the Android application, attached photos can be opened. Figure 4 also shows a attached photo. If we look at this photo, we can be aware of the snowfall in downtown.

The second local topic is a power failure. Figure 5 shows a screen shot of the Web-based interfaces on the Android application. The power failure occurred in Hiroshima due to lightning strikes around 13:15 p.m. on February 12, 2015 (JST). Our system could detect this local topic in real time. The locally bursty keywords include "lightning" and "power failure."

7 CONCLUSIONS

In this paper, we proposed a new real-time spatiotemporal analysis system for enhancing local situation awareness using density-based adaptive spatiotemporal clustering. In the proposed system, local bursty keywords are extracted and their bursty areas are identified. In our new system, locally frequent keywords in geotagged tweets within a particular distance from a user are first extracted. To determine whether locally frequent keywords are bursty keywords or routine keywords, we utilize quartile-based outlier detection. Moreover, bursty local areas related to extracted local bursty keywords are identified using density-based adaptive spatiotemporal clustering. We evaluated the proposed system using actual real-world topics related to weather in Japan. Experimental results

showed that the proposed system could extract local topics and events. In our future work, we are going to evaluate using variety types of local topics and events. Moreover, some local bursty keywords are related each other; however, the proposed system shows these keywords individually. We are developing summarizing method for local bursty keywords.

ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI Grant Number 26330139 and Hiroshima City University Grant for Special Academic Research (General Studies).

REFERENCES

Avvenuti, M., Cresci, S., Marchetti, A., Meletti, C., and Tesconi, M. (2014). Ears (earthquake alert and report system): A real time decision support system for earthquake crisis management. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1749–1758.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Second International Conference on Knowledge Discovery and Data Mining*, pages 226–231.

Hui, C., Tyshchuk, Y., Wallace, W. A., Magdon-Ismail, M., and Goldberg, M. (2012). Information cascades in social media in response to a crisis: A preliminary model and a case study. In *Proceedings of the 21st International Conference Companion on WWW*, pages 653–656.

Hwang, M.-H., Wang, S., Cao, G., Padmanabhan, A., and Zhang, Z. (2013). Spatiotemporal transformation of social media geostreams: A case study of twitter for flu risk analysis. In *Proceedings of the 4th ACM SIGSPATIAL IWGS*, pages 12–21.

Hyndman, R. J. and Fan, Y. (1996). Sample Quantiles in Statistical Packages. *The American Statistician*, 50(4):361–365.

Kim, K.-S., Lee, R., and Zettsu, K. (2011). mtrend: discovery of topic movements on geo-microblogging messages. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in GIS*, pages 529–532.

Kreiner, K., Immonen, A., and Suominen, H. (2013). Crisis management knowledge from social media. In *Proceedings of the 18th ADCS*, pages 105–108.

Kumar, A., Jiang, M., and Fang, Y. (2014). Where not to go?: Detecting road hazards using twitter. In *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 1223–1226.

Mendoza, M., Poblete, B., and Castillo, C. (2010). Twitter under crisis: Can we trust what we rt? In *Proceedings of the First Workshop on SOMA*, pages 71–79.

Naaman, M. (2011). Geographic information from georeferenced social media data. *SIGSPATIAL Special*, 3(2):54–61.

Sakaki, T., Okazaki, M., and Matsuo, Y. (2010). Earthquake shakes twitter users: Real-time event detection by social sensors. In *Proceedings of the 19th International Conference on WWW*, pages 851–860.

Sander, J., Ester, M., Kriegel, H.-P., and Xu, X. (1998). Density-based clustering in spatial databases: The algorithm gdbscan and its applications. *Data Mining and Knowledge Discovery*, 2(2):169–194.

Thom, D., Bosch, H., Koch, S., Worner, M., and Ertl, T. (2012). Spatiotemporal anomaly detection through visual analysis of geolocated twitter messages. In *Pacific Visualization Symposium (PacificVis), 2012 IEEE*, pages 41–48.

Vieweg, S., Hughes, A. L., Starbird, K., and Palen, L. (2010). Microblogging during two natural hazards events: What twitter may contribute to situational awareness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1079–1088.

Yin, J., Lampert, A., Cameron, M., Robinson, B., and Power, R. (2012). Using social media to enhance emergency situation awareness. *IEEE Intelligent Systems*, 27(6):52–59.