Using Periodic Patterns to Discover Events in Twitter: Extended Abstract

R. Uday kiran and Masaru Kitsuregawa

Institute of Industrial Science, The University of Tokyo, Komaba-ku, Tokyo, Japan.
uday_rage@tkl.iis.u-tokyo.ac.jp and kitsure@tkl.iis.u-tokyo.ac.jp.

Abstract. Finding events in Twitter feeds is an important task in disaster relief information system. Most of the previous study on event detection require user-specified keywords for filtering tweets and identifying events. When confronted with the problem of specifying keywords, researchers have tried to address this problem by extracting keywords from the articles of other sources, such as news sources and blogs. This approach of collecting keywords from other sources has two major limitations: (*i*) delay in identifying events and (*ii*) difficulty in developing domain-specific event detection application. This describes the automatic discovery of keywords from Twitter data using periodic pattern mining, which aims at finding all sets of keywords that have not only appeared frequently, but also occurring regularly in the data. Experimental results demonstrate that periodic patterns can provide useful information about the keywords, which in turn can be used for identifying events.

Keywords: Data mining, periodic pattern mining, rare item problem, irregular time series

1   Introduction

Event detection is an important task in Topic Detection and Tracking. Event detection can be considered as a clustering process, where tweets with similar contents are grouped in order to identify popular concepts. Several event detection algorithms have been discussed to find events in Twitter data [3–5, 1]. A major obstacle encountered by an event detection algorithm is the noisy nature of Twitter, which significantly affects the clustering performance. To filter noise, many event detection algorithms use user-specified keywords to filter tweets. When confronted with the problem of specifying keywords, researchers have tried to address this problem by extracting keywords from the articles of other sources, such as news sources and blogs [5]. This approach of collecting keywords from other sources has two major limitations: (*i*) delay in identifying events and (*ii*) difficulty in developing domain-specific event detection application.

Cordeiro [1] have used abrupt change in the frequency of hashtags as an interestingness criterion to find keywords in twitter data. We have observed that this criterion may not be appropriate as hashtags can have non-uniform frequencies. That is, some hashtags can appear more frequently in the data, while other hashtags can appear relatively less frequently in the data. As frequency dimension alone was not sufficient to

find keywords, we have investigated other dimensions to extract keywords. **During our investigation, we have observed that noise generally exhibits sporadic behavior, and therefore, periodicity detection or finding periodically occurring hashtags can provide useful information about the keywords.**

Based on our observation, we have employed partial periodic pattern model proposed in [2] to find periodic occurring hashtags in Twitter data. Using the generated periodic patterns as keywords, we have filtered the tweets and discovered events. We have been collecting twitter data for the past 6 years. As of 17 June 2016, our repository contains approximately 30 billion tweets of 20 million users. In this experiment, we considered a total of 2,813,450 (approximately 2.8 million) tweets, which were collected from 11 March 2011 to 30 March 2011. Most of these tweets are related to the Great East Japan Earthquake (GEJE) that took place on 11 March 2011. Table 1 shows some of the interesting periodic patterns discovered from Twitter data.

Table 1: Some interesting patterns discovered from Twitter data

S. No.	Pattern	Major causes for appearance of keywords
1	{Route,doctor}	(*i*) Many roads and highways were shutdown after earthquake; therefore, people were tweeting about closed routes and requesting information on alternative routes. (*ii*) People were requesting information for doctors who are working at night.
2	{rehabilitation, stress}	(*i*) People were requesting information on rehabilitation centers that were being set by various public and private institutions. (*ii*) People have reported their mental stress due to earthquakes and melt down of Fukushima nuclear plant.
3	{water,sleep}	(*i*) People were tweeting about the contamination of drinking water for collecting water bottles. (*ii*) People were tweeting about sleeplessness at nights.

2   Acknowledgement

References

1. CORDEIRO, M. Twitter event detection: Combining wavelet analysis and topic inference summarization. In *Doctoral Symposium on Informatics Engineering* (2012), pp. 11–16.
2. KIRAN, R. U., AND REDDY, P. K. An alternative interestingness measure for mining periodic-frequent patterns. In *DASFAA* (2011), pp. 183–192.
3. LI, R., LEI, K. H., KHADIWALA, R., AND CHANG, K. C.-C. Tedas: A twitter-based event detection and analysis system. In *ICDE* (2012), pp. 1273–1276.
4. OZDIKIS, O., SENKUL, P., AND OGUZTÜZÜN, H. Semantic expansion of tweet contents for enhanced event detection in twitter. In *ASONAM* (2012), pp. 20–24.
5. RITTER, A., MAUSAM, ETZIONI, O., AND CLARK, S. Open domain event extraction from twitter. In *KDD* (2012), pp. 1104–1112.