Realtime Analysis of Tweet Streams with EmoTwitter

Yuki Kobayashi   Maxim Mozgovoy

The University of Aizu
Tsuruga, Ikki-machi,
Aizuwakamatsu, Fukushima, Japan
+81-242-37-2664
{m5201158, mozgovoy}@u-aizu.ac.jp

ABSTRACT

With the raise of social media, its abilities and influence on people continues to increase. It is being used for many purposes, such as communication with friends and business partners, advertisement, news delivery, information spread, and so on. To support the analysis of social media content, we developed EmoTwitter, a tool for emotion analysis of twitter messages. However, the current release of EmoTwitter supported only the analysis of a particular user's tweets, thus limiting its application area. In the present paper we outline the extension of EmoTwitter that allows realtime Twitter streams in any given geographical location.

Categories and Subject Descriptors

J.4.3 [**Computer Applications**]: Social and Behavioral Sciences – *sociology*.

General Terms

Algorithms, Design.

Keywords

Twitter, social media analysis, streaming.

1. INTRODUCTION

On the Internet, many of communication tools are used. Among them, Twitter is one of the most famous microblogs in the world, having more than 30 million active users. Twitter also became an important tool for marketing. According to research, many users follow some companies' accounts and take action that is useful for companies [1]. This trend is expected to continue in the future, so tools for twitter analysis can be of interest to sociologists and business analysts.

EmoTwitter is a tool for emotion analysis of Twitter messages, developed at the University of Aizu [2]. It is written in C# with the support of several third-party libraries, and it supports several types of text analysis. When a user inputs a Twitter account name into a textbox, EmoTwitter displays the retrieved tweets of the given account holder. The most recent version of EmoTwitter is powered by the modern Tweetinvi library [5], and supports OAuth authentication.

However, EmoTwitter has a Twitter-imposed limit of downloading only the most recent 3200 tweets of each user, and the total hourly limit of queries. At the same time, a sociologist or a business analyst might be interested not in a particular users' messages, but in the trends in the Twittersphere in general in a particular time in a particular geographical location. The purpose of the present work is to implement such functionality.

2. EMOTWITTER IMPROVEMENTS
2.1 Streaming API and Google Maps API

Twitter provides two kinds of APIs: REST API and Streaming API. The REST API is the most basic API that allows to retrieve and analyze individual messages. The Streaming API gives the developers access to global streams of tweet data. In other words, the Streaming API is able to provide realtime twitter datastream. Figure 1 shows the comparison between REST API process and Streaming API process. Both APIs use HTTP connection, but Streaming API requires keeping a persistent HTTP connection open [3]. The Streaming API is asynchronous: the downloaded tweets are returned to the caller as they arrive, and the download process continues to work.

The Streaming API lets the user to filter the stream by geographical location and tweet language. Unfortunately, the streams available to applications only contain small extract of the real global Twitter stream. The users have to consider this limitation unavoidable.

Google Maps API for .NET is a C# interface for interacting with the backend web service of Google Maps [4]. EmoTwitter uses this API to display a Google Map object by with a web browser, search arbitrary location and get its coordinates.

2.2 Realtime Tweet Analysis

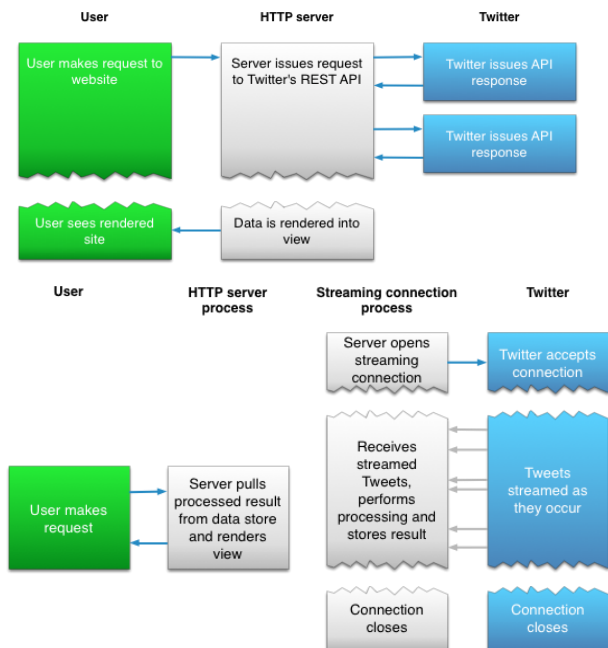EmoTwitter is able to retrieve tweets from an arbitrary geographical area. This functionality works as follows.

Figure 1. REST API process (upper) and Streaming API process (lower).
Source: https://dev.twitter.com/streaming/overview

First, the user inputs a location name to be analyzed. Then a Google Maps object is used to retrieve geographical coordinates to EmoTwitter. The actual search range is a square area with the corners at the following four coordinates: (latitude + 0.5 mile, longitude + 0.5 mile), (latitude + 0.5 mile, longitude − 0.5 mile), (latitude − 0.5 mile, longitude + 0.5 mile), and (latitude − 0.5 mile, longitude − 0.5 mile). Next, the user has to initiate stream analysis by clicking the "Start stream" and "Show" buttons.

EmoTwitter displays realtime tweets of the people in this location in a textbox. At the same time the most frequent words of the tweets are displayed in a word cloud (see Figure 2 and Figure 3).

3. CONCLUSION

EmoTwitter is able to retrieve realtime tweet streams of any arbitrary location in the world, providing a convenient way of accessing Twitter services for interested users. Streaming API helps to analyze ongoing trends in Twittersphere. In particular, the word cloud helps to grasp the topics under active discussion.

Our preliminary experiments show that in most cases words in a word cloud are not stable, and quickly change over time. However, sometimes it is possible to catch a somewhat longer trend where a number of keywords stay in active use for minutes and even hours. We are going to investigate this phenomenon in further experiments.
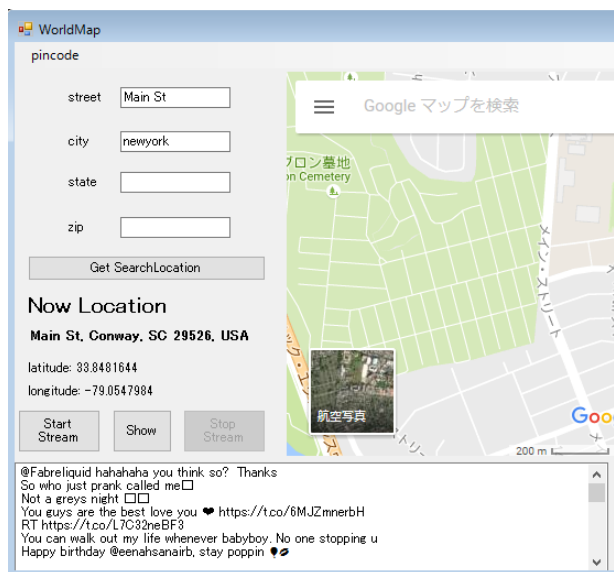

Figure 2. Choosing a location around New York Main St.

Figure 3. Example word cloud of a realtime stream.

4. REFERENCES

[1] Twitter ユーザー調査 2016: https://cdn.cms-twdigitalassets.com/content/dam/business-twitter/resources/customerinsightsloc/Customer_insights_2016_EN_06.28.2016_JA.pdf.

[2] M. Munezero, C. Montero, M. Mozgovoy, E. Sutinen. EmoTwitter — A Fine-Grained Visualization System for Identifying Enduring Sentiments in Tweets. *Lecture Notes in Computer Science*, 2015, vol. 9042, pp. 78-91.

[3] The Streaming APIs Overview. https://dev.twitter.com/streaming/overview.

[4] Google Maps API for .NET. https://github.com/ericnewton76/gmaps-api-net.

[5] Tweetinvi homepage: https://github.com/linvi/tweetinvi.