Imminent Event Extraction from Twitter in Closed Domain

Ravi Shankar
Arizona State University
Computer Sci & Engineering
ravs@asu.edu

Anil Kuncham
Arizona State University
Computer Sci & Engineering
akuncham@asu.edu

Abstract

Twitter has become a major platform for spreading word/publicizing any forthcoming event. Technical events such as Conferences, hack days, developer summit, e.t.c use Twitter to make developers aware of the event and market their events. Previous work on extracting events from Twitter has been very open and does not address to events which are very specific to certain users. In this paper we are going to devise **ConfTracker** - a forthcoming closed-domain technical event extraction system for Twitter. In addition, we present an approach for discovering trending events based on geolocation, rate the events using Sentiment Analysis and also suggest events based on user's tweeting pattern and his timeline analysis. We will then evaluate the extracted events against events from popular event registration sites such as Eventbrite, Lanyrd e.t.c. as well as against a corpus of test tweets.

Introduction

We propose a method for discovering imminent events of technical and professional genre from Social Media feeds such as Twitter. Twitter's popularity and its exponential growth has made it first choice of event organisers to market their event. We will devise an approach to mine events from Twitter's timeline data. Its very easy to get data from Twitter but the real challenge lies in interpreting the data. There has been continuous work on extracting events from Twitter [1, 2, 3]. All of these previous work mainly target news data, social events or open-domain events. We propose a very specific type of event extraction which targets only technical or professional events.

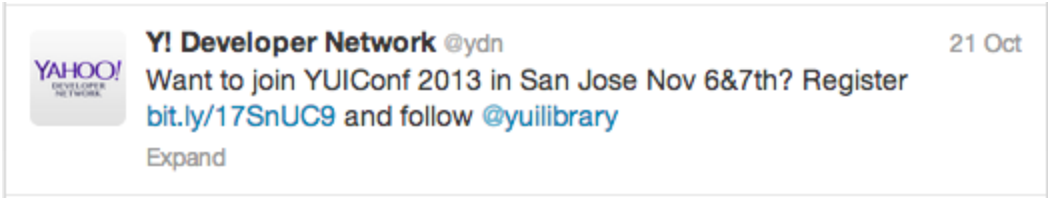
Social Networking sites such as Facebook and Twitter are good source of information. While they contain wealth of useful and important data, they are disorganized and unstructured. Data from Twitter can be at times very specific and useful but mostly it is redundant and noisy. Previous work of *Ritter et al.* [2] demonstrates that extracting events from Twitter can be achieved with 90% of accuracy.

Extracting Event Mentions

In order to extract explicit event mentions from Twitter's data we can use traditional approach of annotating the text, in our case tweets. We will use this annotated textual data to train our model

*This is a project proposal for course CSE598: Social Media Mining. As of now everything mentioned is a proposal except the references and ConfTracker (we proactively booked the domain name for it, conftracker.com).

to extract events.
Event tweets can consist of many different part of speech as illustrated in following example:



Verbs: join, Register
Nouns: YUIConf, San Jose
Temporal Data: Nov 6&7th

Extracting Events from tweets with hidden data

Sometimes organizers tweet about the event with a link to their site in which case there is no explicit information about event present in the tweet itself.

We propose a **novel approach** to identify tweets with hidden event information and to extract event from the them.
Following tweet is an example of tweet with hidden event information:

Samsung Developers Conference: 4 Things You Need to Know

October 15, 2013 | **0 Comments** | Tomorrow Works

SAMSUNG DEVELOPERS CONFERENCE

We have already talked about the first annual **Samsung Developers Conference** in San Francisco. This conference will feature in-depth insights to the latest tools, SDKs, and emerging future platforms to help developers create the next hottest application/service.

Here is what you need to know about the upcoming Samsung Developers Conference.

1. Time and Place

Samsung's first annual conference starts from **October 28-29** at the **Westin St. Francis Hotel, Union Square** in **San Francisco, California.** There is only 2 weeks left!

*This is a project proposal for course CSE598: Social Media Mining. As of now everything mentioned is a proposal except the references and ConfTracker (we proactively booked the domain name for it, conftracker.com).

Rating, Trending and Suggesting: Events

In addition to extraction of events, we propose to develop a system to rate events based on the tweets pertaining to the event. We would typically use sentiment analysis to analyze the tweets and rate events based on that.
We also propose an approach to find trending events based on geo-location of user, as well as generic trending events worldwide
Additionally, we propose an approach to suggest the events based on user's previous tweets and analysis of his/her timeline.

References

[1] E. Benson, A. Haghighi, and R. Barzilay. Event discovery in social media feeds. In ACL, 2011.

[2] A. Ritter, Mausam, O. Etzioni, S. Clark. Open Domain Event Extraction from Twitter. In KDD'12, August 12–16, 2012, Beijing, China.

[3] Bayar Tsolmon, A-Rong Kwon, and Kyung-Soon Lee. Extracting Social Events Based on Timeline and Sentiment Analysis in Twitter Corpus. In proceedings of the 17th International Conference on Applications of Natural Language to Information Systems, June 26-28, 2012.

*This is a project proposal for course CSE598: Social Media Mining. As of now everything mentioned is a proposal except the references and ConfTracker (we proactively booked the domain name for it, conftracker.com).