ASSOCIATING EVENTS WITH PEOPLE ON SOCIAL NETWORKS USING A-PRIORI

Srijan Khare, Vyankatesh Agrawal, Gaurav Tiwari, Gourav Arora and
Bhaskar Biswas

Department of Computer Science & Engineering,
Indian Institute of Technology (BHU), Varanasi, India
`{srijan.khare.cse11, v.agrawal.cse11, gaurav.tiwari.cse11,`
`gourav.arora.cse11, bhaskar.cse} @iitbhu.ac.in`

ABSTRACT

*In social media, same news or events are associated with two or more people, sometimes with
different perspective. The representation of the news or events varies from person to person,
perspective to perspective or time to time. In this paper, we present a simple model to associate
events with different people (Personalities). To demonstrate our model, we have used real world
social networks data (i.e. from Twitter) and results show the accuracy of the model.*

KEYWORDS

Social Networks, Association rule, A-Priori.

1. INTRODUCTION

In the present times online social networks (social networks) plays an important role in modern
lives. People not only use this powerful media as a tool for entertainment, news, sharing emotions
and sentiments but also depend for news, political sentiments, and events. Most of the times,
news or an event is associated with many people. Many times people express their views on an
event based on their perspective. These expressions of views differ as they come from different
people with different perspectives. People in social media or in general make their opinion about
an event based on the views expressed by the others. This "word-of-mouth" is sometimes biased
towards a particular aspect of the news or event and this is due to the difference in perspective. To
get the holistic view of news or event, one may require getting many views/opinions associated
with that news/event. In general, words from the famous personalities about an event build up the
general opinion of that event. Another outlook is to judge or compare personalities or person
based on their views on a particular event. So, to understand the events, on whole, different views
are needed to be associated with views or personalities. In this paper, a model has been proposed
to address this challenge.

The proposed model uses Twitter as a source of views of personalities and events. Association of
different personalities/views with events has been done using A-Priori [17] algorithm. Twitter has
been used by many researchers. Event detection has been a major challenge that was addressed by
majority of researchers using twitter data. Using different approaches and methods, researchers
have proposed to detect event from Twitter and applied for different applications and usage. A
brief survey was conducted for the same ([2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13],

[14] and [16]). A-priori algorithm is a basic association rule generation method which has been used by many researchers for different applications. One such application is given in [15].

In [10], Hila Becker et al. proposed a method to identify real event from non event using twitter messages. From the twitter stream, authors make clusters of similar messages to identify events. In this proposed model, similar approach was followed to identify the tweets from different personalities without distinguishing the events. Luca Cagliero and Alessandro Fiori [1] presented TweM Tweet Miner framework which discover generalized association rules from Twitter. [1] gives a generalized association among the content of the tweets, whereas the proposed model tries to associate different personalities with events. The authors of this paper acknowledge the work presented in [1] and [10] along with all the other references.

2. PROPOSED MODEL

The proposed model is shown in Figure 1. Using the Twitter API, raw tweets were collected. Based on the different 'hash-tags' and the inputs from the different personalities, the raw tweets were sorted. Pre-processing was done to different inputs. This step mainly includes word extraction and removal of symbol and words less than four letters, using Regex words extraction in our model. The Association Rule Algorithm, which in this case is A-Priori algorithm, requires unique words. So, duplicate items were removed. The pre-processed tweets were now stored in database in the form of transactions from different personalities. These transactions were given input to the algorithm. Based on the association rule generated, the events were associated with different personalities.
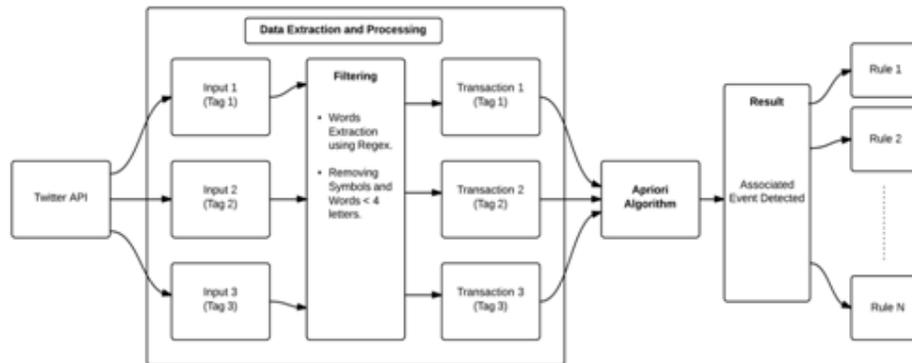

Figure 1. Proposed Model

3. EXPERIMENT AND RESULTS

To test the proposed model, tweets from famous personalities were extracted for a common time interval and limited to 100 tweets per day. Specifically, tweets about following were collected:

- Bill Gates
- Mark Zuckerberg
- Satya Nadella
- Jeff Bezos

These personalities represented 4 transactions. The tweets collected were broken into words using a simple regex pattern and words of length lesser than 4 were ignored. Moreover, special characters and words of non-English languages were discarded. The final input data comprised of

100 words within each. With Support 0.75 and Confidence: 0.5, a set of 666673 rules were identified. Similarly for Support 1.0 and Confidence: 0.5, a set of 223695 rules were found. The Ice Bucket Challenge which has been a recent trend was identified as a common link between all the transactions i.e. the personalities chosen.

Figure 2 shows a union of first 10 words of 4 transactions and their respective frequencies. The grey squares indicate the presence of a word in a particular transaction.
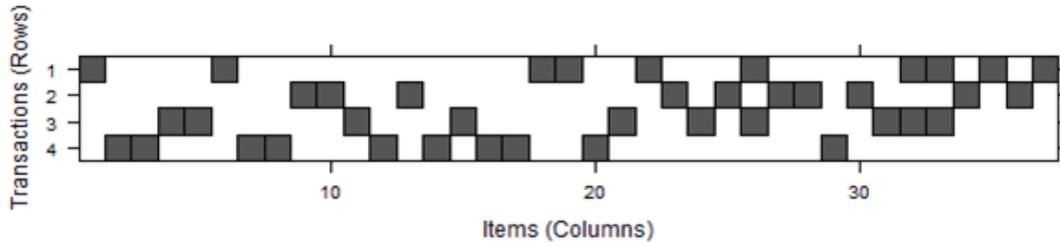

Figure 2. Items-Transaction Graph

Figure 3 shows some of the association rules found with support = 0.75 and confidence = 1.00.

Figure 3. Association Rules

4. CONCLUSIONS

A simple model to associate personalities with events has been proposed in this paper. Initial experiments depict the success of the model. This idea can also be applied to a situation where tweets from different News agencies form different transactions in a selected time interval. Applying Apriori algorithm on it could help identify a trending topic. The news agencies and others could themselves use this trend to focus more on the identified topic.

This proposed model is a simple model, where sentiments analysis, opinion mining and machine learning techniques can be used to improve the efficiency. Methods can be proposed for the dynamic nature of the twitter data.

REFERENCES

[1]   Luca Cagliero, Alessandro Fiori, " Discovering generalized association rules from twitter", Intelligent Data Analysis, Volume 17 Issue 4, July 2013, Pages 627-648.

[2]   Salvatore Orlando, Francesco Pizzolon and Gabriele Tolomei, "SEED: A Framework for Extracting Social Events from Press News", WWW 2013 Companion, May 13–17, 2013, Rio de Janeiro, Brazil.

[3]   Hila Becker, Dan Iter, Mor Naaman and Luis Gravano," Identifying Content for Planned Events Across Social Media Sites", WSDM'12, February 8–12, 2012, Seattle, Washington, USA.

[4]   Jeffrey Nichols, Jalal Mahmud and Clemens Drews," Summarizing Sporting Events Using Twitter", IUI'12, February 14-17, 2012, Lisbon, Portugal.

[5]   Arkaitz Zubiaga, Damiano Spina, Enrique Amigó and Julio Gonzalo," Towards Real-Time Summarization of Scheduled Events from Twitter Streams", HT'12, June 25–28, 2012, Milwaukee, Wisconsin, USA.

[6]   Rui Li, Kin Hou Lei, Ravi Khadiwala and Kevin Chen-Chuan Chang, "TEDAS : a Twitter-based Event Detection and Analysis System", 2012 IEEE 28th International Conference on Data Engineering.

[7]   Ana-Maria Popescu , Marco Pennacchiotti and Deepa Arun Paranjpe," Extracting Events and Event Descriptions from Twitter", WWW 2011, March 28–April 1, 2011, Hyderabad, India.

[8]   Adam Marcus, Michael S. Bernstein, Osama Badar,David R. Karger, Samuel Madden and Robert C. Miller," TwitInfo: Aggregating and Visualizing Microblogs for Event Exploration", CHI 2011,Session: Twitter Systems May 7–12, 2011,Vancouver, BC, Canada.

[9]   Akshaya Iyengar, Tim Finin and Anupam Joshi," Content-based prediction of temporal boundaries for events in Twitter", 2011 IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing.

[10] Hila Becker, Mor Naaman, and Luis Gravano," Beyond Trending Topics: Real-World Event Identification on Twitter", Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM '11), short paper.

[11] Yosuke Jimbo, Masaki Kohana and Shusuke Okamoto," A Twitter Service for School Event", 2011 Second International Conference on Networking and Computing.

[12] Ana-Maria Popescu and Marco Pennacchiotti, "Detecting Controversial Events from Twitter", CIKM'10, October 26–30, 2010, Toronto, Ontario, Canada.

[13] Makoto Okazaki and Yutaka Matsuo," Semantic Twitter: Analyzing Tweets for Real-Time Event Notification", BlogTalk 2008/2009, LNCS 6045, pp. 63–74, 2010.

[14] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo," Tweet Analysis for Real-Time Event Detection and Earthquake Reporting System Development", WWW '10 Proceedings of the 19th international conference on World wide web.

[15] Alexander Toshev, Francois Br´emond and Monique Thonnat," An APRIORI-based Method for Frequent Composite Event Discovery in Videos", Computer Vision Systems, 2006 ICVS '06. IEEE International Conference on 4-7 Jan. 2006.

[16] Luca Cagliero and Naeem A. Mahoto," Visualization of High-Level Associations from Twitter Data", www.irma-international.org/viewtitle/80217/.

[17] http://en.wikipedia.org/wiki/Apriori_algorithm