Real-Time Knowledge Management and Data-mining with Twitter

Martin Moghadam, Kalle Grafström

May 29, 2010

Abstract

Real-time tracking of human behavior on Twitter for survey, marketing and analysis.
The research focus (i.e. statement of the problem(s)/research issue(s) addressed); * The research methods used (experimental research, case studies, questionnaires, etc.); * The results/findings of the research; and * The main conclusions and recommendations

1 Introduction

Knowledge management used to identify and represent insights and experiences gained from data collected from Twitter, can prove useful in many circumstances. Some of which will be discussed in this article, others are left for further development, see discussion section.
The validity and credibility of the users identity is not discussed in detail, we simply assume the user has a valid identity, see references for further details on "The Credibility of Digital Identity Information of the Social Web".
The project uses Twitter as a data source to comprise the knowledge, the Tweets all have dates and time allowing use to develop a systems that is updated minute by minute, gathering Information in real-time. This makes it possible to examine a period of time to extract knowledge from the Tweets in that period, for instance a debate performance[10].

1.1 Twitter

Twitter is a microblogging and social network service allowing users to send and read small text messages of up to 140 characters. The service has currently more than 100 million users worldwide. Users create accounts with profiles, the basic accounts are provided for free, premium account with a higher number of characters are purchasable.
The Tweets are associated to a specific user profile, the profile can have many followers that read the tweets. Each tweets has a time and date information,

and can include; profile tags, topic tags and urls , this information can be used to structure and gain a greater understanding of the tweets.

The tweets contain different kinds for information as shown in figure 1. Searching the entire web for this kind of information would yield poor results for to the wast amount of trash data that need to be sorted though.
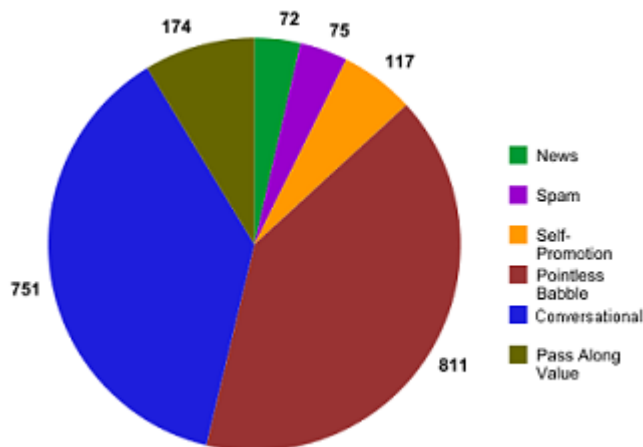

Figure 1: Content of Tweets: Source Wikipedia

1.2 Premise and Usage

Twitter can and has been used to gather knowledge for:

- Real-time recommendations for topical news. [5]

- Marketing and research.

- Disaster monitoring, earthquakes, volcanoes, hurricanes. [7]

- Political demonstrations. (The Iran demonstrations of 2009)

- Counter-terrorism.

- Many other possibilities.

A few people have used Twitter to make bomb treats, and encouraged assassination of politicians, these treats were not detected or monitored by an application, most of the threats were reported to the authorities by other users of the service and investigated but led to no results, because the threats were not serious just pointless babble.

2 Knowledge Management

2.1 Real-time Knowledge

3 Data Source

3.1 Twitter Streams

3.2 Real-time Data

3.3 Code Example

3.4 Test and Development

4 Discussion

4.1 Further Development

4.2 Conclusion

References

[1] Satyen Abrol and Latifur Khan. *TWinner: Understanding News Queries with Geo-content using Twitter.* ACM, 2010.

[2] Marc Cheong and Vincent Lee. *Integrating Web-based Intelligence Retrieval and Decision-making from the Twitter Trends Knowledge Base.* ACM, 2009.

[3] Balachander Krishnamurthy, Phillipa Gill, and Martin Arlitt. *A Few Chirps About Twitter.* ACM, 2008.

[4] Matthew Rowe. *The Credibility of Digital Identity Information on the Social Web: A User Study.* ACM, 2010.

[5] Owen Phelan, Kevin McCarthy, and Barry Smyth. *Using Twitter to Recommend Real-Time Topical News.* ACM, 2009.

[6] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. *What is Twitter, a Social Network or a News Media?* ACM, 2010.

[7] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. *Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors.* ACM, 2010.

[8] Vivek K. Singh, Mingyan Gao, and Ramesh Jain. *Situation Detection and Control using Spatio-temporal Analysis of Microblogs.* ACM, 2010.

[9] Vivek K. Singh and Ramesh Jain. *Structural Analysis of the Emerging Event-web.* ACM, 2010.

[10] Nicholas A. Diakopoulos and David A. Shamma. *Characterizing Debate Performance via Aggregated Twitter Sentiment.* ACM, 2010.

[11] Bernard J. Jansen, Gerry Campbell, and Matthew Gregg. *Real Time Search User Behavior.* ACM, 2010.

[12] Amanda Lenhart and Susannah Fox. *Twitter and status updating.* ACM, 2009.