A System for Twitter User List Curation

Igor Brigadir
School of Computer Science and Informatics
University College Dublin, Ireland
igor.brigadir@ucdconnect.ie

Derek Greene, Pádraig Cunningham
School of Computer Science and Informatics
University College Dublin, Ireland
{derek.greene,padraig.cunningham}@ucd.ie

ABSTRACT

With increased adoption of social networking tools, it is becoming more difficult to extract useful information from the mass of data generated daily by users. Curation of content and sources is an important filter in separating the signal from noise. A good set of credible sources often requires painstaking manual curation, which often yields incomplete coverage of a topic. In this demo, we present a recommender system to aid this process, improving the quality and quantity of sources. The system is highly-adaptable to the goals of the curator, enabling some novel uses for curating and monitoring lists of users.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – Information filtering

Keywords

Content curation, Social media monitoring, Network analysis, Social network discovery

1. INTRODUCTION

Storyful[1] is a social media news agency established in 2010 with the aim of filtering newsworthy content from the vast quantities of noisy data on social networks such as Twitter and YouTube. To this end, Storyful invests considerable time into the manual curation of content on these networks. Twitter users can organise the users they follow into *lists*. Storyful maintains user lists as a means of monitoring breaking news. These lists can be constructed manually, but this process is time-consuming, and risks incomplete coverage of all aspects of a news story. Therefore, to support these curation tasks, we have developed and deployed a web-based system for exploring the Twitter network and recommending the important users that form the "community" around a news story (see Fig. 1). Currently the system is being used to monitor over 100 news stories, mining microblogging data for a diverse range of topics, from the United States 2012 presidential election to the political situation in Afghanistan. A video of the system in use is available online[2].

[1] `http://www.storyful.com`
[2] `http://www.youtube.com/watch?v=rMfN59bmEyc`

RecSys'12, September 9–13, 2012, Dublin, Ireland.
ACM 978-1-4503-1270-7/12/09.

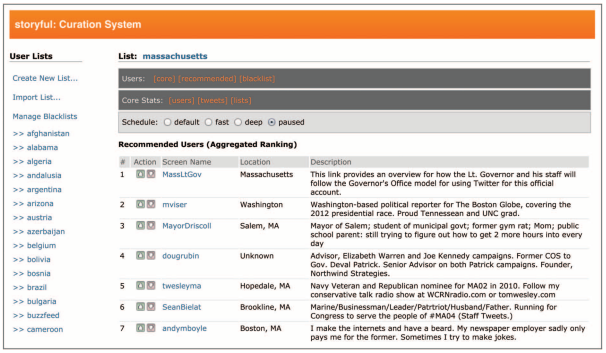

Figure 1: Screenshot of the curation system, showing candidate users recommended for addition to a user list covering the 2012 Republican Party nomination for the state of Massachusetts.

2. CURATION SYSTEM

The input to the system is an initial *seed list* of Twitter users that was manually labelled as being relevant to a particular news story, topic of interest or location. The bootstrap phase retrieves network structure around the egos in the seed list. Information retrieved consists of user profile information, friend and follower links, user list membership information, and tweets. The extent of the exploration process can be easily controlled by pre-set configuration settings – effectively controlling the trade-off between running time and accuracy.

After the initial bootstrap phase, the system maintains two distinct lists of users:

Core list: List of the actual Twitter user accounts used by journalists during content curation for the chosen news story. Initially this will contain the members of the seed list.

Candidate list: User accounts that are not in the core list, but may potentially be relevant for curation. Initially this will consist of the set of non-seed list users identified during the bootstrap phase.

Based on the candidate list, the recommendation engine will then produce a ranked list of potentially relevant users for promotion to the core list. Based on these recommendations, a human curator can select users to add to the core list, or filter incorrect recommendations. Once the core list has been modified, the system updates the network structure around the core list, to reflect (a) changes in membership of the core

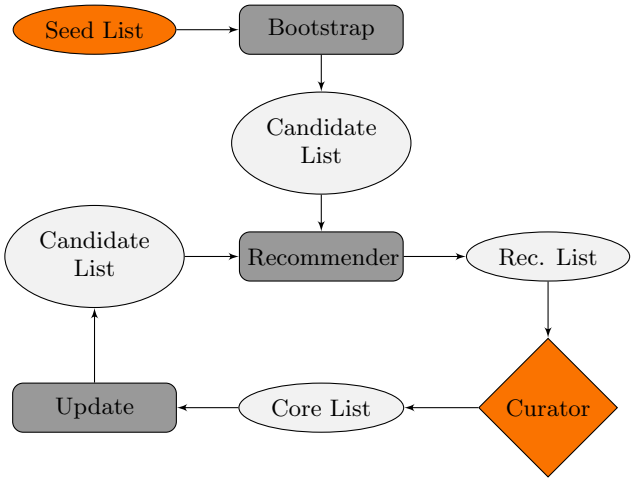

Figure 2: Overview of curation support system, illustrating workflow between phases.

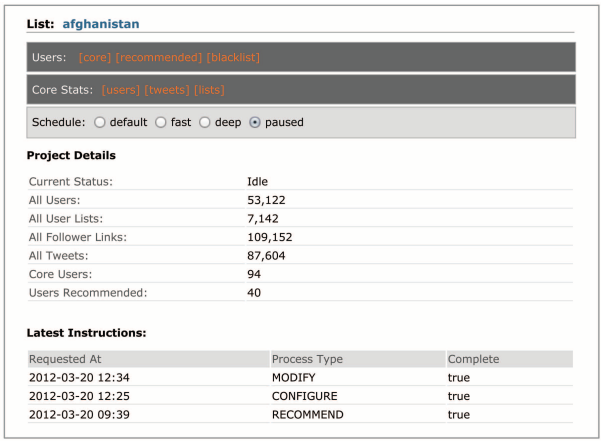

Figure 3: Screenshot presenting data statistics regarding a single user list monitoring the political situation in Afghanistan.

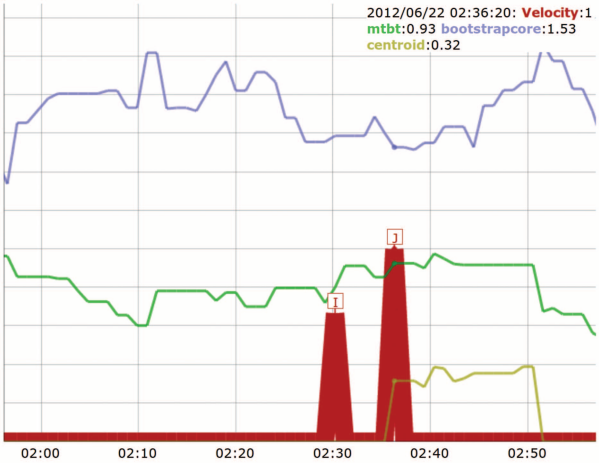

Figure 4: A *velocity chart* showing spikes of tweeting activity corresponding to breaking news events being discussed on the Afghanistan user list.

list, and (b) general changes in the larger Twitter network since the last update. Again the extent of the exploration during the update process can be controlled by user-defined settings. The system then iterates between recommendation and update phases (see Fig. 2).

Rather than using a single view of the network to produce recommendations, we employ a multi-view approach that produces recommendations based on different graph representations of the Twitter network surrounding a given user list, and combines them using an SVD-based aggregation approach [1]. Information from multiple views is also used to control the exploration of the Twitter network. This is an important consideration due to the limitations surrounding Twitter data access. These network views include friend and follower graphs, mention and retweet graphs and views based on how users are sorted into lists by other users – effectively crowd-sourcing the list curation in part.

The system runs on open source software and can be deployed on a single server or an Amazon EC2 instance for example. The specific hardware requirements will depend upon on the total number of lists being monitored. As of April 2012, Storyful are currently monitoring 124 stories, topics, and geopolitical regions using the deployed system.

The curation system is not limited to generating recommendations. When the system is used to monitor lists covering dozens or hundreds of news stories, it will often be important for journalists to focus their limited time on a subset of these lists, where breaking developments are occurring. To facilitate this kind of prioritisation, we monitor the *velocity* of the lists being monitored by the system. The velocity measure is a combination of several indicators including tweet similarity, the level of activity of the core users, and tweet frequency. The velocity measure can detect significant or unusual tweet activity, often indicating a breaking news story (see Fig. 4).

3. CONCLUSIONS

While the curation system presented here is primarily used as a support tool within Storyful for curating lists of sources for online journalism, we are currently investigating its use in other applications that involve social media exploration and insight. For instance, one current experiment uses the system to identify the presence of extremist groups on Twitter. Another application is the identification of spam accounts or bots that share common links in one or more Twitter network views.

Acknowledgments

This research was supported by SFI Grant 08/SRC/I1407 (Clique: Graph and Network Analysis Cluster). The authors thank Storyful for their participation in the development and testing of the system.

4. REFERENCES

[1] D. Greene, G. Sheridan, B. Smyth, and P. Cunningham. Aggregating Content and Network Information to Curate Twitter User Lists. Under Review, 2012.