Characterizing Twitter Users for Building Social Systems

by Parantapa Bhattacharya

Abstract

In this thesis, we focus on characterizing regular Twitter users in terms of their expertise, interests, and personality, and use that understanding to help them in better producing and consuming information and content on Twitter.

Earlier studies of the Twitter microblogging platform had suggested that Twitter is a social media platform consisting of a relatively small section of elite users, producing information on a few popular topics such as media, politics, and music. In this thesis, we show that the above characterization ignores a rich set of highly specialized topics, ranging from geology, neurology, to astrophysics and karate — each being discussed by their own topical groups. We present a semantic methodology to identify *topical groups* in Twitter on a large number of topics, each consisting of users who are experts on or are interested in a specific topic. We perform a detailed characterization of these topical groups based on their network structures and tweeting behaviors.

In Twitter, over 500 million tweets are posted daily, while on Facebook, over 4.75 billion pieces of content are shared daily. This sheer volume of unfiltered content makes it incredibly difficult for today's social media users, to filter out interesting and engaging content, from within the deluge of uninteresting ones. Using traditional recommendation system techniques to help social media users in this regard is difficult, as content recommendation systems for social media have markedly different challenges compared to traditional systems. The primary problem is with regard to the fact, that content in social media are produced at an amazingly fast pace and peoples interest in them dissipate rapidly. In this thesis we propose a novel system for performing topical recommendations in the Twitter network, that is scalable, trustworthy, and explainable.

Out of 500 million tweets posted daily on Twitter, 11% are deleted by the users posting them. This widespread deletion of tweets begs a number of questions: what kind of content posted by users makes them want to delete them later? Are users of certain predispositions more likely to post regrettable tweets, deleting them later? In this thesis we provide a detailed characterization of tweets posted and then later deleted by 200 thousand Twitter users, during a period of four weeks. Our characterization shows significant personality differences between users who delete their tweets and those who do not, and significant linguistic differences between tweets that are deleted later compared to those that are not. Additionally, we build a classifier which takes textual, contextual, as well as user features to predict whether tweet is likely to be deleted in the future.