Abstract:

Microblogging services have revolutionized the way people exchange information, and have emerged as an essential forum for people to air their views on topics of common interests. Therefore, monitoring and analyzing the rich and continuous flow of user-generated contents in microblog networks can yield unprecedentedly valuable information, which would not have been available from traditional media outlets. In particular, microblogs naturally unfolds events occurring in the real-world. By monitoring on "viral topics" in microblog networks, i.e., topics that receive a large volume of discussion as well as a large number of participants within a short period, we can make microblog a valuable source of information for individuals and organizations to stay informed of "what is happening and hot now". In this thesis, we aim to carry out a thorough study on viral topics in microblog networks. Specifically, the main aim of our study is to design and develop a viral topic monitoring system for microblog networks, which is able to predict, detect and summarize viral topics.

First, we investigate the prediction of viral microblogs by learning the influences among users in a microblog network. This component targets at predicting whether a piece of microblog will become viral, and which part of the network will participate in propagating this message. To facilitate the prediction ability, we firstly define three types of influences that will affect a user's decision on whether to perform a diffusion action, and propose a diffusion-targeted influence model to differentiate and quantify various types of influence. The problem of diffusion prediction is then modeled as factorizing a user's intention to transmit a microblog into these influences. In this way, a prediction model is achieved, which is able to predict the virality of incoming new microblogs.

Second, we explore the problem of microblog tracking. Due to the existence of celebrity effect, advertising needs and zombie accounts, a large portion of the viral messages predicted in the previous component are not topic-related, which cannot lead to sufficient follow-up discussions. Therefore, we further propose the second component, where the
previously predicted viral microblogs are utilized to monitor on the incoming microblog stream. In this component, a novel dictionary learning based method is proposed for tracking an individual microblog. This component aims to filter out non-topic microblogs, and detect the occurrence of viral topics in the early stage.

Finally, we examine how to conduct summarization for the predicted viral topics, which are in the form of a collection of related microblogs, with too much information to be presented. To express the contents concisely and make the topic readable to people, in this step, a multimedia topic summarization scheme is proposed. Given a collection of microblogs related to a topic, this scheme is able to automatically generate a summary for this topic, which contains both textual and visual information.

Through extensive experiments conducted on the large-scale real-world datasets, the experimental results have demonstrated that our study could yield significant gains in providing users with timely and concise information about the occurrence of incoming viral topics.