

Making Sense of Social Media Streams through Text Summarization (PI: Dr. Li

Wenjie Maggie; 2014/15)

The social media world has experienced tremendous growth in recent years. On social media platforms, people all around the world are connected together. They actively interact with each other and collectively create a massive amount of online text streams. Such text streams, if explored well, contain a great wealth of information and offer significant opportunities for timely emergency response and appropriate actions. As an essential technique to deal with the problem of information explosion, text summarization aims to automatically extract the most important information from source documents to produce a condensed version. Although research on text summarization already has more than half a century of history, when confronted with online social media text, traditional summarization approaches are no longer adequate. Unlike carefully authored news articles or scientific papers, social media text is characterized by the short, noisy, large scale, and dynamic natures. The traditional summarization approaches that are dominated by content analysis inevitably suffer from its sparse and noisy text representation. Moreover, these approaches normally focus on static document collections and thus are insufficient to handle the continuously growing and dynamically changing text streams. New methods and new approaches are highly demanded to meet these new challenges. In this project, we propose a novel

summarization framework to address the problems raised by the less-informative and dynamic social media content. Our objective is to develop adaptive models and scalable algorithms based up this framework to explore three key intrinsic dimensions of social media text stream, i.e., the temporal, social and content dimensions. We integrate together the content analysis and time-series modeling techniques to automatically determine “what and when is worth to summarize” by tracking and modeling the dynamic changes in the stream over time. With emphasis on the roles of social influence and social interactions, which are critical to information dissemination on social media networks, we aim to alleviate the sparse and noisy text presentation problem in summary content selection and make further efforts to generate more informative summaries that not only depict topic evolvement but also explain how key people push forward the development of topic. The deliverables of the project include innovative solutions to dynamic and social content modeling and summarization and benchmark datasets to support research development and performance evaluation. They will contribute to promoting the research in social media summarization and add value to the practical applications like marketing management and emergency response.