UNIVERSITY OF AMSTERDAM

Monitoring Social Media. Summarization, Classification and Recommendation

Z. Ren

A key characteristic of social media research is the ambition to monitor the content of social media, i.e., text from social media platforms, social relations among users, and changes in social media data over time. In this thesis, we present research on understanding social media along three dimensions: summarization, classification and recommendation.

Our first line of work concerns summarization of social media documents. Firstly, we address the task of time-aware tweets summarization, based on a user's history and collaborative influences from ``social circles.'' We propose a time-aware user behavior model to infer dynamic probabilistic distributions over interests and topics. Based on probabilistic distributions from our proposed model, we explicitly consider novelty, coverage, and diversity to arrive at an iterative optimization algorithm for selecting tweets. Secondly, we continue our research on summarization by addressing the task of contrastive theme summarization. We combine the nested Chinese restaurant process with contrastive theme modeling, which outputs a set of threaded topic paths as themes. We present the structured determinantal point process to extract a subset of diverse and salient themes. Based on probabilistic distributions of themes, we generate contrastive summaries subject to three key criteria: contrast, diversity and relevance. Lastly, we address the viewpoint summarization of multilingual streaming corpora. We propose a dynamic latent factor model to explicitly characterize a set of viewpoints through which entities, topics and sentiment labels during a time interval are derived jointly; we connect viewpoints in different languages by using an entity-based semantic similarity measure; and we employ an update viewpoint summarization strategy to generate a time-aware summary to reflect viewpoints.

Our second line of work is hierarchical multi-label classification of social text streams. Concept drift, complicated relations among classes, and the limited length of documents in social text streams make this a challenging problem. We extend each short document in social text streams to a more comprehensive representation via state-of-the-art entity linking and sentence ranking strategies. From documents extended in this manner, we infer dynamic probabilistic distributions over topics. For the final phase we propose a chunk-based structural optimization strategy to classify each document into multiple classes.

Our third line of work is explainable recommendation task via viewpoint modeling, which not only predicts a numerical rating for an item, but also generates explanations for users' preferences. We propose a latent variable model for predicting item ratings that uses user opinions and social relations to generate explanations. To this end we use viewpoints from both user reviews and trusted social relations. Our method includes two core ingredients: inferring viewpoints and predicting user ratings. We apply a Gibbs EM sampler to infer posterior distributions of our method.

In our experiments we have verified the effectiveness of our proposed methods for monitoring social media, showing improvements over various state-of-the-art baselines. This thesis provides insights and findings that can be used to

facilitate the understanding of social media content, for a range of tasks in social media retrieval.