Adaptive Topic Mining from Social Media Streams

Gopi Chand Nutakki & Olfa Nasraoui
University of Louisville

Abstract

October 2, 2017

Topic Models are statistical models that can be used for discovering the abstract "topics" that may occur in social media streams. However, these techniques face dramatic challenges when coping with very sparse and yet topically diverse micro-blog posts such as Tweets. In such data streams, not only are the topics very diverse, but also, the vocabulary is huge. This makes the sampling space for generative models vast. In this work, we propose a topic modeling framework built on top of an existing stream clustering framework, called Stream Dashboard, which can extract, isolate, and track evolving topics over any given time period of time. The Stream Dashboard module initially clusters the data stream points into homogeneous groups which are then subjected to online topic modeling, thus making our approach a compartmentalized topic modeling approach. The proposed evolving topic mining framework can handle the noise, size, dynamic nature, and diversity of the data stream by partitioning the data in both the content and temporal spaces, hence making the challenging topic modeling task easier and faster and making the cluster models resulting from Stream-Dashboard compatible with topic modeling, thus more appropriate for text data.