Topic Sketch: Real-time Bursty Topic Detection from Twitter

Twitter has become one of the largest micro blogging platforms for users around the world to share anything happening around them with friends and beyond. A bursty topic in Twitter is one that triggers a surge of relevant tweets within a short period of time, which often reflects important events of mass interest. How to leverage Twitter for early detection of bursty topics has therefore become an important research problem with immense practical value. Despite the wealth of research work on topic modelling and analysis in Twitter, it remains a challenge to detect bursty topics in real-time. As existing methods can hardly scale to handle the task with the tweet stream in real-time, we propose in this paper Topic Sketch, a sketch-based topic model together with a set of techniques to achieve real-time detection. We evaluate our solution on a tweet stream with over 30 million tweets. Our experiment results show both efficiency and effectiveness of our approach. Especially it is also demonstrated that Topic Sketch on a single machine can potentially handle hundreds of millions tweets per day, which is on the same scale of the total number of daily tweets in Twitter, and present bursty events in finer-granularity.