

The Impact of metadata properties on News Event Detection on Twitter

Samer Aamar

Advisor: Dr. David Bhodof; Dr. Ossi Mokryn

Twitter is one of the leading micro-blogging systems today. Users on Twitter are discussing mostly irrelevant topics. However, they tend to report and discuss real world events instantly when they occur (e.g., earthquake, tsunami, death of a celeb, local events). Previous research works already proposed various methods for news event detection on Twitter. Mostly, they rely on the content of the tweets; while applying NLP methods, hashtags count, keywords and bursty words. Other methods proposed to focus mainly on non-content properties (metadata properties) such as geographic location, interactions between users, retweets and replies and user bio-attributes.

What are the unique metadata properties for news event diffusion in Twitter in comparison to other types of information? What types of events do we have in Twitter and what properties differentiate them one from the other? In this work we try to answer these questions by analyzing a dataset of 30 million tweets. We define a set of metadata properties and then check their ability to identify news event by using machine learning approaches. The main goals of this work are to reveal metadata properties that are informative for separating news events from non-news related discussions, and to give better understanding of the behavior of Twitter users when discussing news events.