

Abstract.

The wide adoption of social media platforms by a large number of users across the globe has provided a rich source of unstructured information for understanding users' behaviors, interests and opinions. An active area in this space is the detection of important real-world events from user-generated social content. The works in this area identify instances of events that impact a large number of users. However, a more nuanced form of an event, known as life event, is also of high importance, which in contrast to real-world events, does not impact a large number of users and is limited to at most a few people.

For this reason, life events, such as marriage, travel, and career change, among others, are more difficult to detect for several reasons: 1) they are specific to a given user and do not have a wider reaching reflection; 2) they are often not reported directly and need to be inferred from the content posted by individual users; and 3) many users do not report their life events on social platforms, making the problem highly class-imbalanced.

The aim of this thesis is to provide a method for identifying self-report life events using social media text data. Given that much of these text data is about the future events, the ability of this method to identify future events for the user is the main aim of this research. In future life event detection problem, there is no access to the content of the tweet in which the life event is mentioned and detection is done solely based on the past tweets published by the user, ie. user history. By discovering common signs in the users' history that will occur life event in future, it is possible to solve the problem of future life event detection.

The proposed approach includes two main parts: i) first, by modeling life event semantically based on word embedding technique and processing past tweets of users, their life events are extracted in order to build histories. ii) second, two approaches are introduced namely content generation and discovering relations between life events that determine the future personal life events of the users from their histories. Content generation based on recurrent neural network tries to predict future personal life events by processing short interval of tweets of users, where the second approach discovers common signs in long interval of tweets using machine learning techniques.

We have shown in our experiments that future life event detection problem is a non-trivial task and our work is able to provide reasonable performance on a gold standard dataset despite the highly class-imbalanced nature of the personal life event data.