

**Predicting Social Reaction and Its Labels\*  
for a Given News Article by Using Microblogs**

**Mikroblogları Kullanarak Verilen Bir Haber İçin  
Toplumsal Tepki ve Tepki Etiketlerinin Tahmini**

**TÜBİTAK Project No. 215E169**

**Fazlı Can (PI), Seyit Koçberber (I)**

**April 1, 2016**

**Abstract**

People's response to an event, that happens at a particular time and place, is defined as "social reaction." In today's society of multiple views it is difficult to estimate the *dimension* and *direction* of social reactions for events. For a news article such as "Fire Department saves cat from tree" no social reaction is expected. On the other hand, the protests of doctors after the "Enactment of Full-Time Employment Law" is an example of negative social reaction. Mass discontent of people in the social media on a news article such as "Bloody balance sheet of holiday accidents" is another example for negative reaction. However, the reaction associated with the holiday-accidents does not involve any protests in the streets. In addition to this dimension (high: a social event is happening or low: a social event not happening) aspect, its direction (negative vs. positive) is also important. An example of positive reaction is peaceful post match celebrations of football fans after a win.

In social media using microblog platforms individuals share their views in a data stream on the fly. In this study, our aim is early prediction of social reaction to an event mentioned in a given news article and the associated labels describing this reaction. For obtaining microblogs Twitter will be used.

Early prediction of social reaction for an event supports government institutions, individuals, and commercial organizations to prepare themselves in terms of precautions for future negative events and appropriate and advantageous responses for future positive events. There is a need for such event prediction algorithms that will make use of daily thoughts expressed in microblogs. In Twitter, Trend Topics show the social reactions after they do happen. There are also studies that aim to predict possible events and protests beforehand. In this study, different from earlier studies, the event prediction is based on a given arbitrary news article and is done in terms of time increments and its dimension/direction and labels.

In the first stage of the project tweets will be accumulated in the form of a data stream. Tweets are limited in size, noisy, unstructured, and involve spam and an unlimited jargon. For this reason, they have to be streamlined into a form which is suitable for estimating social events. In this stage we will also form a ground truth (gold standard) test collection that involves news articles marked in terms of social event dimension and direction and also in terms of their descriptive words or labels. The test collection will be used for the assessment of the developed algorithms. In the second stage of the project the processed tweets will be filtered to find the ones related to a given article.

In the third stage, which is the main thrust of the project, a two-part novel algorithm will be developed for early prediction of social reactions and their descriptive labels. The algorithm employs microblog meta features and tweet-based sentiment analysis and is language-independent. In the first part of the algorithm a time window-based ensemble of classifiers is used for predicting social reactions of people and in the second part cross-entropy concept and hidden Markov modeling is used for ranking and choosing topic headings selected by the topic modeling approach as social reaction labels.

In the fourth stage of the project an assessment of the social reaction prediction and label prediction algorithms will be performed by employing the test collection. In the fifth and final stage the prediction algorithms will be integrated into our large scale Bilkent News Portal. In various stages of the project several issues related to Turkish, such as effects of stemming, mapping microblog abbreviations into real words, and stopping of frequent terms will be considered.

The ground truth test collection will be in the form of a standard test collection like the ones in our previous TÜBİTAK projects. It will encourage other researchers to do new studies in the same or similar fields. The fact that the study investigates conditions related to Turkish gives additional importance to the project. The methods to be developed are adaptable to other microblogs, research and application areas. For example, in intelligence applications predictions based on intelligence reports can be used to develop precautions for the reported events. By this way the intelligence activities can be more effective.

\* Duration: April 1, 2016 – October 1, 2017

RA supported by the project: Çağrı Toraman (PhD Student)

Budget 142,497 TL (\$50,445 as of April 1, 2016: official project beginning date)