

# Emotional Reactions Prediction of News Posts

Anastasia Giachanou<sup>1</sup>, Paolo Rosso<sup>2</sup>, Ida Mele<sup>3</sup>, Fabio Crestani<sup>1</sup>

<sup>1</sup>Faculty of Informatics, Università della Svizzera italiana, Lugano, Switzerland

<sup>2</sup>PRHLT Research Center, Universitat Politècnica de València, Spain

<sup>3</sup>ISTI-CNR, Pisa, Italy

anastasia.giachanou@usi.ch, proso@dsic.upv.es,  
ida.mele@isti.cnr.it, fabio.crestani@usi.ch

**Abstract.** Nowadays, on-line news agents post news articles on social media platforms with the aim to attract more users. Different types of news trigger different emotions on users who may feel surprised or sad after reading some piece of news. In this paper, we are interested in predicting the amount of emotional reactions triggered on users after reading a news post. To address the problem, we propose a model that is trained on features extracted from users' early commenting activity. Our results show that users' early activity features are very important and that combining those features with terms can effectively predict the amount of emotional reactions triggered on users by a news post.

## 1 Introduction

Social media platforms such as Facebook and Twitter allow news agents to post news articles online which are accessible to users to read, comment or express their opinion about them. Some of the news articles trigger a large amount of emotional reactions whereas others do not. Predicting the amount of emotional reactions is a very important problem for dealing with the problem of information overload. For example, a system that can predict the amount of emotional reactions that are triggered by news articles allows a user to filter the articles she would like to read based not only on the articles' content but also on the emotions they trigger.

Predicting the amount of triggered emotional reactions is not a trivial problem. Network properties such as the structure of the platform or other external factors such as user's location may affect the reactions of the users towards a specific news post. Intuitively, content is one of the most important factors that influences the emotional reactions [1] since there are certain terms that convey sentiment and emotion.

A related problem to *emotional reactions prediction* is the online content popularity prediction. Most of prior work was based on early-stage measurements, whereas little effort has been given on the pre-publication prediction [4, 3]. Bandari et al. [4] tackled the task as both regression and classification, and reached the conclusion that the prediction is feasible without any early activity signals. However, recently Arapakis et al. [3] extended the work of [4] and showed that predicting the popularity of news articles prior to their publication is not yet a viable task.

A closely related work is performed by Clos et al. [5] who proposed a unigram mixture model to create an emotional lexicon which was then used to predict the probabilities of different emotional reactions. More recently, Giachanou et al. [6] focused on predicting the amount of emotional reactions triggered on users. However, they only explored pre-publication features including content based similarities and frequencies of entities.

## 2 Emotional Reactions Prediction

The problem of *emotional reactions' amount prediction* of news posts published on a social network is defined as: *Given a news article post and data about early activity, the task consists in predicting the amount of emotional reactions that the post will trigger on users.* Note that our aim is to classify a news post with regards to the amount of the emotional reactions (e.g., love, surprise, joy, sadness, anger) it will trigger on users. We address the problem as a 3-class task. Given a news post we assign to it one of the labels *low, medium, high* that refer to the amount of each emotional reaction that the post will trigger.

### 2.1 Features

Intuitively, the content of the post is very important for predicting if a news article will trigger a high number of a certain emotional reaction. To this end, in our study we start with *terms*. Furthermore, we extract features from *users' early commenting activity* to investigate if there are temporal patterns in commenting activity.

**Frequencies.** The simplest textual feature is the *terms* that the news post contains. Although this is a simple feature, it is one of the most important features for news articles' popularity prediction [1, 9] as well as similar information retrieval tasks [2, 7]. We use the bag-of-words representation to model the terms. Each term in the vector is weighted using the term frequency-inverse document frequency (TF-IDF) approach that considers how important is the term in a corpus. In the rest of the paper, we use *terms* to refer to the TF-IDF representation of the terms.

**Commenting Activity.** Once a news post is published on a social network, the users are allowed to publish their comments about the specific post. These comments which are published below the news post are very important because they are an early indicator of users' interest and reaction regarding the news post. We use the activity of users in publishing comments regarding the news post to extract our early activity features regarding three time range scenarios: 10, 20, and 30 minutes after the publication of the news article. We use the following features:

1. *First comment*: time difference in seconds between publication date of the news post and the first comment, if the first comment is published within the specified time range.
2. *Number of comments*: number of comments published within the specified time range.
3. *Commenting ratio*: mean time of commenting for those published within the specified time range.

### 3 Experimental Setup

We used the same dataset as in Giachanou et al. [6] which contains news posts from *The New York Times* group in Facebook together with the amount of 5 different emotional reactions: *love*, *surprise*, *joy*, *sadness*, and *anger* for each post. The collection consists of 26,560 news posts that span from April 2016 to September 2017. We used a 10-fold cross validation to perform the experiments. We kept training and test sets separate.

We performed a 3-class classification task according to which a news post can get one of the following labels: *low*, *medium*, *high*. We predicted the amount level of the following emotional reactions: *love*, *surprise*, *joy*, *sadness*, and *anger*, which were addressed individually. For all the experiments, we used the Random Forest classifier. We report F1 score for each emotional reaction. We compare our results with *terms* that is based only on the terms of the posts and the *All (+terms)* that is based on the approach proposed in Giachanou et al. [6]. Significance is measured with the McNemar test.

### 4 Results and Discussion

From the results in Table 1 we observe that *terms* are better predictors compared to using only the *early activity*. This suggests that for the specific task *terms* contain more predictive power compared to *early activity*, that is considered the most important feature for popularity prediction [8]. When the *early activity* features are used alone, the best performance is obtained for *joy*. In addition, we observe that for the emotions *surprise* and *joy* the difference between *terms* and *early activity* is smaller compared to the rest of the emotions. Indeed, in case of *joy*,  $early_{t=30}$  obtains a slightly worse performance compared to *terms*. One possible explanation is that in case of news that trigger *joy* and *surprise*, users post more comments compared to the rest of emotions.

**Table 1.** Performance results (F1-scores) using early activity features. Scores with \* and † indicate statistically significant improvements with respect to the *terms* and *All (+terms)* approaches.

	Love	Surprise	Joy	Sadness	Anger
Terms	0.491	0.494	0.578	0.555	0.597
All (+terms) [6]	0.478	0.486	0.554	0.543	0.576
early <sub>t=10</sub>	0.416	0.476	0.549	0.435	0.509
early <sub>t=20</sub>	0.415	0.477	0.557	0.444	0.518
early <sub>t=30</sub>	0.415	0.480	0.574	0.448	0.535
Terms+early <sub>t=10</sub>	0.534* †	0.563* †	0.644* †	0.586* †	0.633* †
Terms+early <sub>t=20</sub>	0.536* †	0.573* †	0.652* †	0.592* †	0.642* †
Terms+early <sub>t=30</sub>	0.541* †	0.577* †	0.653* †	0.593* †	0.645* †

<https://www.facebook.com/nytimes/>

Table 1 shows that, in most of the cases, the performance improves when the time range is increased. The only exception is the reaction *love* for which the performance slightly decreases. For some emotions (e.g., *surprise*), the improvement is little, whereas for other emotions (e.g., *anger*) the improvement is larger. However, we expect that extracting features from even the first ten minutes is very useful for the prediction while keeping the advantage of quick access after the post is published. Finally, Table 1 shows that combining *terms* with *early commenting activity* is the most effective approach and leads to significant improvements over both *terms* and *All (+terms)* approaches.

## 5 Conclusions and Future Work

In this study, we presented a methodology for predicting the amount of emotional reactions that will be triggered towards a specific news post. Our results suggested that *early commenting activity* is very important for the emotional prediction task. However, *terms* contain more predictive power compared to using only *early activity* predictors. More importantly, we showed that models trained on both *terms* and *early commenting activity* can effectively address the problem.

As future work, we plan to address the task as an ordinal classification or a regression problem and we will try to predict the exact number of each emotional reaction.

**Acknowledgments.** This research was partially funded by the Swiss National Science Foundation (SNSF) under the project OpiTrack.

The work of the second author was partially funded by the the Spanish MINECO under the research project SomEMBED (TIN2015-71147-C2-1-P).

## References

1. Alam, F., Celli, F., Stepanov, E.A., Ghosh, A., Riccardi, G.: The social mood of news: Self-reported annotations to design automatic mood detection systems. In: PEOPLES '16. pp. 143–152 (2016)
2. Aliannejadi, M., Crestani, F.: Venue suggestion using social-centric scores. CoRR abs/1803.08354 (2018)
3. Arapakis, I., Cambazoglu, B.B., Lalmas, M.: On the feasibility of predicting popular news at cold start. *Journal of the Association for Information Science and Technology* 68(5), 1149–1164 (2017)
4. Bandari, R., Asur, S., Huberman, B.A.: The pulse of news in social media: Forecasting popularity. In: ICWSM '12. pp. 26–33 (2012)
5. Clos, J., Bandhakavi, A., Wiratunga, N., Cabanac, G.: Predicting emotional reaction in social networks. In: ECIR '17. pp. 527–533 (2017)
6. Giachanou, A., Rosso, P., Mele, I., Crestani, F.: Emotional influence prediction of news posts. In: ICWSM'18 (2018)
7. Paltoglou, G., Giachanou, A.: *Opinion Retrieval: Searching for Opinions in Social Media*, pp. 193–214. Springer International Publishing (2014)
8. Shulman, B., Sharma, A., Cosley, D.: Predictability of popularity: Gaps between prediction and understanding. In: ICWSM '16. pp. 348–357 (2016)
9. Tsagkias, M., Weerkamp, W., De Rijke, M.: Predicting the volume of comments on online news stories. In: CIKM '09. pp. 1765–1768 (2009)