# Early Commenting Features for Emotional Reactions Prediction

Anastasia Giachanou[1], Paolo Rosso[2], Ida Mele[3], Fabio Crestani[1]
[1]Faculty of Informatics, Università della Svizzera italiana, Lugano, Switzerland
[2]PRHLT Research Center, Universitat Politècnica de València, Spain
[3]ISTI-CNR, Pisa, Italy
anastasia.giachanou@usi.ch, prosso@dsic.upv.es,
ida.mele@isti.cnr.it, fabio.crestani@usi.ch

**Abstract.** Nowadays, one of the main sources for people to access and read news are social media platforms. Different types of news trigger different emotional reactions to users who may feel happy or sad after reading a news article. In this paper, we focus on the problem of predicting emotional reactions that are triggered on users after they read a news post. In particular, we try to predict the number of emotional reactions that users express regarding a news post that is published on social media. In this paper, we propose features extracted from users' comments published about a news post shortly after its publication to predict users' the triggered emotional reactions. We explore two different sets of features extracted from users' comments. The first group represents the activity of users in publishing comments whereas the second refers to the comments' content. In addition, we combine the features extracted from the comments with textual features extracted from the news post. Our results show that features extracted from users' comments are very important for the emotional reactions prediction of news posts and that combining textual and commenting features can effectively address the problem of emotional reactions prediction.

## 1 Introduction

In recent years, social media platforms have become an integral part of news industry. News agents post news articles on social media platforms such as Facebook[1] and Twitter[2]. These news articles are accessible to users who can comment or express their opinion about them. Some of the news articles posted on social networks trigger a large number of emotional reactions whereas others do not. Predicting the number of emotional reactions that will be triggered on users is very useful for information spreading and fake news detection. For example, fake news are written to attract users' attention and to trigger emotions to a large number of people [24]. Therefore, the number of emotional reactions can be used as an additional information for fake news or clickbait detection.

Emotional reactions prediction is a challenging problem. The structure of the network or other external factors such as users' location are some of the factors that can

[1] https://www.facebook.com/
[2] https://twitter.com/

affect the number of the triggered emotional reactions. Intuitively, the content of the news post is one of the most important factors that influences the emotional reactions that will be triggered [1]. However, content is not sufficient alone since there are other factors that may influence the triggered reactions. Information extracted from users' *early comments* (i.e., comments published within the first ten minutes after the publication of the news post) can be very useful for an effective emotional reaction prediction.

The problem of *emotional reactions prediction* is related to online content popularity prediction. Most prior work on news articles' popularity prediction is based on early-stage measurements, whereas little effort has been made on the pre-publication prediction scenario [5, 4]. Although the problem of predicting the number of emotional reactions has apparent similarities with predicting the popularity of a piece of news, the two problems are not the same. A piece of news that triggers emotional reactions has certainly higher probabilities of receiving attention compared to news articles that do not trigger any emotional reaction. However, predicting the triggered emotional reactions depends on many factors such as, for example, the affective words that the news post contains, the structure of the network and the early commenting activity. Therefore, for an effective prediction it is very important to combine features extracted from the news post content and the comments that are posted after the news post is published.

In this paper, we focus on the problem of predicting the ordinal level in regards to the number of the emotional reactions triggered on users after reading a news post per emotion (e.g., low, medium, high number of anger reactions). We propose two different sets of features extracted from users' early comments to perform the prediction on five standard emotional reactions (*love, surprise, joy, sadness, anger*). The two proposed sets of features capture two different aspects of information: the commenting activity (e.g., when the first comment is published) and the content of the comments (e.g., relevance to the post). In addition, we combine the features extracted from *early comments* with the *terms* of the news post and we show that this combination can effectively address the problem of emotional reactions prediction.

## 2 Related Work

One aspect that is relevant to the emotional reaction prediction is popularity prediction. Prior work tried to predict the popularity of different web items such as images, videos or tweets prior and after their publication. A wide range of features have been explored and the most informative have shown to be those extracted from early activity [8]. To this end, a large number of researchers tackled the online content prediction after publication by modeling the early users' behavior [18] or by using temporal patterns of online content [29].

Tsagkias et al. [27] explored different features such as the length of the article and the number of authors to address the problem of news articles' popularity prediction. Tsagkias et al. addressed the problem as a binary classification where the news articles were classified as having low or high popularity. Bandari et al. [5] tackled the prediction task as both regression and classification, and used various features including category of the article and named entities. Bandari et al. reached the conclusion that predicting the popularity of web items is feasible without any early activity signals. However, recently Arapakis et al. [4] extended the study of Bandari et al. and they showed that predicting the popularity of news articles prior to their publication is not a viable task.

The problem of the emotional reactions prediction is also related to opinion and emotion analysis that have been applied on different social media platforms, including blogs [10], forums [28] and microblogs [17, 14, 12]. Prior work on emotion and sentiment analysis include classification and lexicon-based approaches [11]. The classification based approaches [3, 22, 17] leverage classifiers that are trained on several features such as n-grams, stylistic features (e.g., number of exclamation and question marks), negation, or part-of-speech tags. Lexicon-based approaches use list of words known as opinion or sentiment lexicons which convey a specific sentiment or emotion to label the text [26].

Regarding emotional reactions, Clos et al. [9] proposed a unigram mixture model to create an emotional lexicon that was used to predict the probabilities of five different emotional reactions. In addition, Alam et al. [1] focused on mood level prediction of readers on news articles (ranging from 0 to 1) using features such as character, words and affect scores. Alam et al. showed that n-grams and stylometric features are the most important. More recently, Goel et al. [15] focused on predicting the intensity of emotions in tweets using an ensemble of three neural-network approaches. However, our problem is not the same as emotional intensity, since an article may trigger an emotion that is intense to only a small number of people. Consider the case of a strike in the means of transportation in a small city. In such a case, some people may feel very angry (e.g., "I got stuck in traffic for an hour and a half! #busStrike") but such intense emotion might be triggered only in a small number of people.

The study that is the most similar to ours is the one presented by Giachanou et al. [13] who also focused on predicting the ordinal level regarding the number of emotional reactions triggered by news posts. However, in their study they only explored pre-publication features including similarities and entities extracted from the article's content. Different from Giachanou et al., we focus on features that are extracted from the users' comments to understand how effective they are in predicting the emotional reactions of the news post. We study the effectiveness of two groups of features extracted from users' comments regarding the post. In addition, we propose combining simple textual and early commenting features for effectively predicting the triggered emotional reactions.

## 3 Problem Definition

The problem of *emotional reactions prediction* of news posts published on a social network is defined as: *Given a news article post and data about users' early comments published regarding the post, the task consists in predicting the qualitative ordinal level of emotional reactions that the post will trigger*. Note that the main aim is to classify a news post with regards to the volume of the emotional reactions it will trigger per emotion. We focus on the following five different emotions: *love, surprise, joy, sadness, anger*. We address the problem as both 3-class and 5-class ordinal classification task to capture the different levels of the reactions. Hence, given a news post we assign to it one of these labels: *low, medium, high* for the 3-class task and one of these labels *very* *low, low, medium, high, very high* for the 5-class task. The labels refer to the number of reactions that the post will collect per emotion.

## 4 Features

Intuitively, content 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* extracted from the news post. Terms can be very important to understand why a specific article triggered massive emotions. Furthermore, we extract features from users' comments published shortly after the publication of the post to investigate if there is any pattern in commenting that can be useful for predicting the emotional reactions' popularity.

### 4.1 Term Frequencies

The first feature we use is the *terms* of the news post. Although *terms* is a simple feature, it is one of the most important features for news articles' popularity prediction [1, 27] as well as similar information retrieval tasks [2, 20]. For *terms* feature we use the bag-of-words representation of a news post. In particular, we use the classic term frequency-inverse document frequency (TF-IDF) approach [23] that considers how important is a term in a corpus to represent the content of the post. On the contrary to other studies [27] that used only a small percentage of the vocabulary to represent textual features, we are using all the terms that appear in the collection after stopwords removal. In the rest of the paper, we use *terms* to refer to the TF-IDF representation of the post's content.

### 4.2 Early Commenting Features

As already mentioned, once a news post is published on a social network, the users can publish their comments regarding the post. These comments usually appear below the post. We explore two groups of features extracted from users' comments. The first group represents the commenting activity and includes features such as how fast the users publish a comment. For the second group we extract features from the content of the comments such as their relevance to the news post.

Here we should note that activity of emotional reactions can also be very useful (e.g., number of sadness reactions in the first ten minutes). However, we do not have access to these data. Therefore, we use features from the early comments of users to capture early patterns in the users' comments. To extract the commenting features we use three different time range settings: 10, 20, and 30 minutes after the publication date of the news post to explore how useful the different time ranges are and if there is any improvement in performance when a wider time range is used. Finally, we do not differentiate between comments and replies to comments.

**Early Commenting Activity.** The early commenting activity features aim to capture the patterns in the activity of publishing comments below the news post. We explore the following features:

1. *First comment.* Time difference in seconds between publication date of the 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.
4. *Unique authors.* Number of unique authors for the comments published within the specified time range. This feature can partially capture the discussion activity in the comments since a certain author will post more than one comments when there is a discussion.

**Early Comments' Content Features.** In this section we propose features that are extracted from the comments' content. These features can reveal if there is any pattern in the content of the comments that are posted about a news post and the emotional reactions it triggers. We propose three features: the length of the comments, the relevance to the post and the sentiment expressed in the comments.

1. *Length of comments.* This feature is calculated based on the average length of the comments published. The length of a comment is represented by the number of words it contains. This feature is useful because users might tend to post shorter or longer comments regarding the news posts that trigger specific emotional reactions. In addition, longer comments might express stronger emotional reactions that may relate to the reactions triggered regarding the news post.
2. *Relevance to the post.* This feature represents the average relevance of the comments published within the specified time range to the post. This feature is important since there may be comments not related to the post. To calculate the relevance, we use the word2vec model that is an embedding model proposed by Mikolov et al. [19] and which learns word vectors via a neural network with a single hidden layer. First, we calculate the average vector for all words in the comment and the post and then we use cosine similarity between the vectors to calculate the similarity score. We use the pre-trained word embeddings that are publicly available and which are generated from news articles[3] to generate the word vectors.
3. *Sentiment in comments.* We also measure the sentiment expressed in the comments published within the specified time range. In particular, we calculate the positive, neutral and negative sentiment ratio of the comments. We use an opinion lexicon [16] to calculate the sentiment expressed in a comment. More formally, let $N_t(z, s)$ be the number of comments that express a sentiment $s$ towards the news post $z$ posted during a particular time period $t$ and $N_t(z)$ the number of total comments posted regarding $z$ at $t$. Then, we can define the ratio of comments that share a common sentiment $s$ as:
$$r_t(z, s) = \frac{N_t(z, s)}{N_t(z)}$$
We calculate the ratio for all the three sentiment polarities: positive, neutral and negative.

[3] https://code.google.com/p/word2vec/

# 5 Experimental Setup

In this section, we describe the experimental setup of the study. First, we describe the dataset and next we present the experimental settings we applied for our study.

## 5.1 Dataset

For this study, we collected news posts from *The New York Times* group[4] in Facebook together with the number of 5 different emotional reactions: *love, surprise, joy, sadness,* and *anger* for each post. We used Facebook API[5] to collect the posts, the reactions, and the comments[6]. The number of reactions are used to determine how many reactions each post has triggered. Other types of posts, such as tweets, do not contain information about the emotional reactions, and therefore, they need to be manually annotated, a process that is very costly in time and resources.

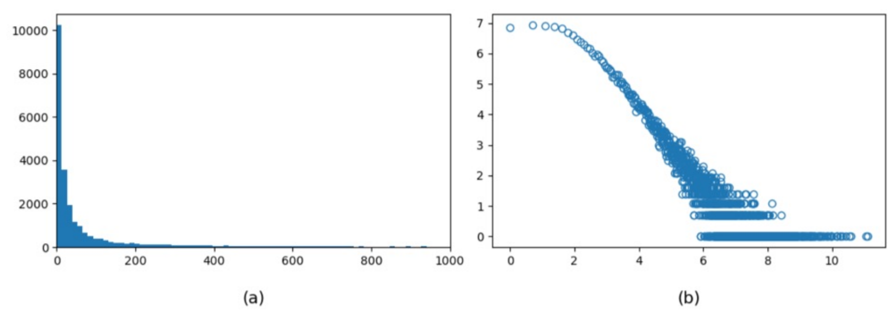



Fig. 1: (a). Frequency of posts versus number of the emotional reaction *love* (binned). (b). Number of *love* reactions per post versus number of posts with that number of *love* reactions (log-scale).

Our collection consists of 26,560 news posts that span from April 2016 to September 2017. We use a 10-fold cross validation to perform the experiments. We keep training and test sets always separate. As an example, Figure 1 shows the distribution of the posts with regards to the emotional reaction *love*. More specifically, Figure 1(a) shows the number of posts versus the number of the *love* reactions triggered. For clarity, we show only the first part of the distribution and cut the long tail after 1,000 *love* reactions. Figure 1(b) shows the number of *love* reactions per post versus the number of posts that triggered that number of *love* reactions. The other emotional reactions follow similar distributions. From the figures, we can observe that the number of reactions per post follows a long-tail distribution. In other words, few posts collect a high number of reactions, while the majority of posts get very few.

[4] `https://www.facebook.com/nytimes/`
[5] `https://developers.facebook.com/`
[6] Facebook allows users to select an emotional reaction with regards to a post.

### 5.2 Experimental Settings

In this study, we performed two tasks: a 3-class and 5-class emotional reaction ordinal classification task. For those tasks, we divided the collection into 3 (and 5) balanced classes with regards to the number of each emotional reaction. A balanced classification formulation has also been chosen by several prior studies on popularity prediction [25, 8]. For the 3-class task a news post can get one of the following labels: *low, medium, high*, while for the 5-class one of: *very low, low, medium, high, very high*. We predicted the number of the following five different emotional reactions: *love*, *surprise*, *joy*, *sadness*, and *anger*. The emotional reactions were addressed individually.

Table 1 shows the boundaries of the different classes. From the table, we observe that the range of the *high* and *very high* classes of the 3-class and 5-class task respectively is wide. For example, the class *very high* of the 5-class task contains posts that received from 122 to 67K *love* reactions. This is due to the long-tail distribution of the data and the balanced classes setting.

Table 1: Boundaries of the different classes.

| 3-class | | | | | |
|---|---|---|---|---|---|
| | Love | Surprise | Joy | Sadness | Anger |
| Low | 0-9 | 0-8 | 0-3 | 0-2 | 0-2 |
| Medium | 10-47 | 9-39 | 4-21 | 3-31 | 3-35 |
| High | 48-67K | 40-23K | 22-27K | 32-50K | 36-67K |
| 5-class | | | | | |
| | Love | Surprise | Joy | Sadness | Anger |
| Very Low | 0-5 | 0-4 | 0-1 | 0-0 | 0-0 |
| Low | 6-13 | 5-11 | 2-4 | 1-4 | 1-3 |
| Medium | 14-33 | 12-28 | 5-13 | 5-18 | 4-17 |
| High | 34-121 | 29-89 | 15-63 | 19-110 | 18-134 |
| Very high | 122-67K | 90-23K | 64-27K | 111-50K | 135-67K |

For the ordinal classification of the emotional reactions, we used Random Forest [7], a decision tree meta classifier[7]. For all the experiments, we used the open source machine learning toolkit scikit-learn[8]. To generate the word vectors we used publicly available pre-trained word embeddings[9]. To calculate the sentiment expressed in a comment, we used the opinion lexicon described in [16]. Pre-processing of the posts involved stop-words removal and stemming with Porter stemmer [21].

*Mean Absolute Error (MAE)* is reported for both 3-class and 5-class tasks and for each emotional reaction. We used the runs trained on *terms* and *activity+content*$_{t=10}$ as

[7] We use Random Forest because it obtained the best results on the run trained on terms among the various classifiers that we tried including SVM and Logistic Regression

[8] `http://scikit-learn.org/`

[9] https://code.google.com/p/word2vec/

our baselines. Significance is measured with the non-parametric Wilcoxon signed-rank test that is appropriate for the ordinal classification.

## 6 Results and Discussion

Tables 2 and 3 show the results using the early commenting features on predicting the number of emotional reactions triggered on users regarding a news post for the 3-class and 5-class ordinal classification respectively. The tables show the MAE scores (the lower the value, the better the approach performs) for three different groups of features: the commenting activity features (*activity*), the comments' content features (*content*) and their combination (*activity+content*). The approach based on post's terms is used as a baseline.

Table 2: Performance results (MAE) for the 3-class ordinal classification using early commenting features. Scores with $*$ indicate statistically significant improvements with respect to the *terms* approach.

| | Love | Surprise | Joy | Sadness | Anger |
|---|---|---|---|---|---|
| Post's terms | 0.629 | 0.649 | 0.542 | 0.565 | 0.503 |
| activity$_{t=10}$ | 0.743 | 0.631 | 0.517$*$ | 0.730 | 0.596 |
| activity$_{t=20}$ | 0.732 | 0.616 | 0.504$*$ | 0.699 | 0.560 |
| activity$_{t=30}$ | 0.724 | 0.602 | 0.493$*$ | 0.690 | 0.544 |
| content$_{t=10}$ | 0.697 | 0.655 | 0.556 | 0.633 | 0.507 |
| content$_{t=20}$ | 0.686 | 0.660 | 0.583 | 0.618 | 0.507 |
| content$_{t=30}$ | 0.683 | 0.664 | 0.590 | 0.609 | 0.505 |
| activity+content$_{t=10}$ | 0.612$*$ | 0.568$*$ | 0.448$*$ | 0.586 | 0.442$*$ |
| activity+content$_{t=20}$ | 0.581$*$ | 0.539$*$ | 0.426$*$ | 0.551$*$ | 0.408$*$ |
| activity+content$_{t=30}$ | 0.555$*$ | 0.534$*$ | 0.413$*$ | 0.539$*$ | 0.388$*$ |

From the results we observe that post's *terms* are better predictors compared to using only the *early commenting activity* or the *comments' content* in the case of *love*, *sadness* and *anger*. However, in case of *surprise* and *joy* the *early commenting activity* runs perform better compared to *terms* and in fact in some cases the difference is statistically better (e.g., 5-class classification of *surprise* and *joy*). Also, we observe, that in general the runs that use the *comments' content* features obtain a lower performance compared to *terms*. One exception is the case of *surprise* and *joy* on the 5-class task where there are runs that perform statistically better to *terms* (e.g., $content_{t=10}$ run).

Regarding the performance between the runs that are based only on the *activity* and those based only on the *comments' content*, we observe that the emotional reactions perform in a different way. More specifically, *activity* leads to a better performance compared to *comments' content* in case of *surprise* and *joy*, whereas regarding *love*, *sadness* and *anger*, the *comments' content* features are better predictors compared to

Table 3: Performance results (MAE) for the 5-class ordinal classification using early commenting features. Scores with $*$ indicate statistically significant improvements with respect to the *terms* approach.

| | Love | Surprise | Joy | Sadness | Anger |
|---|---|---|---|---|---|
| Post's terms | 1.232 | 1.269 | 1.101 | 1.059 | 0.982 |
| activity$_{t=10}$ | 1.396 | 1.195$*$ | 1.009$*$ | 1.334 | 1.122 |
| activity$_{t=20}$ | 1.377 | 1.161$*$ | 0.989$*$ | 1.300 | 1.070 |
| activity$_{t=30}$ | 1.362 | 1.142$*$ | 0.956$*$ | 1.275 | 1.044 |
| content$_{t=10}$ | 1.334 | 1.249 $*$ | 1.078$*$ | 1.175 | 0.989 |
| content$_{t=20}$ | 1.311 | 1.250$*$ | 1.114 | 1.151 | 0.972$*$ |
| content$_{t=30}$ | 1.298 | 1.256$*$ | 1.125 | 1.124 | 0.960$*$ |
| activity+content$_{t=10}$ | 1.177$*$ | 1.093$*$ | 0.895$*$ | 1.103 | 0.857$*$ |
| activity+content$_{t=20}$ | 1.112$*$ | 1.039$*$ | 0.846$*$ | 1.042$*$ | 0.794$*$ |
| activity+content$_{t=30}$ | 1.074$*$ | 1.021$*$ | 0.822$*$ | 1.014$*$ | 0.766$*$ |

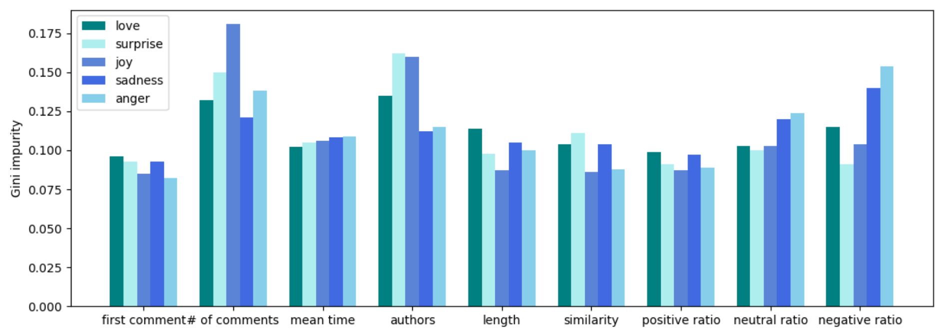



Fig. 2: Gini impurity score for the *activity+content*$_{t=10}$ run for the 3-class ordinal classification per each emotional reaction.

*activity*. This result shows that users' follow different patterns in commenting regarding the different emotional reactions and they probably tend to write more useful comments regarding *love*, *sadness* and *anger*.

More importantly, the majority of runs that use all the early commenting features (i.e. *activity+content*) perform statistically better compared to the ones trained on the *terms* of the post. The only exception is the case of *sadness* in the *activity+content*$_{t=10}$ run. This suggests that in case of *sadness* the terms from the post are stronger predictors compared to commenting activity. However, the results also prove that for most of the reactions the features that are extracted from the users' commenting activity shortly after the post is published can effectively predict the number of emotional reactions.

To understand the contribution of each feature on the prediction, we calculated the Gini impurity scores as described in [6]. Figure 2 shows the Gini impurity score for

Table 4: Performance results (MAE) for the 3-class ordinal classification on combining terms with early commenting features. Scores with $*$ and † indicate statistically significant improvements with respect to *terms* and *activity+content*$_{t=10}$ respectively.

| | Love | Surprise | Joy | Sadness | Anger |
|---|---|---|---|---|---|
| Post's terms | 0.629 | 0.649 | 0.542 | 0.565 | 0.503 |
| activity+content$_{t=10}$ | 0.612 | 0.568 | 0.448 | 0.586 | 0.442 |
| terms+activity+content$_{t=10}$ | 0.540$*$† | 0.510$*$† | 0.405$*$† | 0.499$*$† | 0.403$*$† |

Table 5: Performance results (MAE) for the 5-class ordinal classification on combining terms with early commenting features. Scores with $*$ and † indicate statistically significant improvements with respect to *terms* and *activity+content*$_{t=10}$ respectively.

| | Love | Surprise | Joy | Sadness | Anger |
|---|---|---|---|---|---|
| Post's terms | 1.232 | 1.269 | 1.101 | 1.059 | 0.982 |
| activity+content$_{t=10}$ | 1.177 | 1.093 | 0.895 | 1.103 | 0.857 |
| terms+activity+content$_{t=10}$ | 1.078$*$† | 1.012$*$† | 0.830$*$† | 0.949$*$† | 0.789$*$† |

each feature in the *activity+content*$_{t=10}$ run for the 3-class classification per each emotional reaction. From the figure we observe that the *number of comments* that have been published in the first ten minutes are good predictors for all the five emotional reactions. Indeed for the reaction *joy*, the *number of comments* is the best predictor. Similarly, the *number of unique authors* feature is important for the reactions *joy* and *surprise*.

An interesting observation is that in case of *sadness* and *anger*, the *negative ratio* has the highest Gini impurity score. This result suggests that users tend to express their feelings in comments to the posts that trigger *sadness* or *anger*.

Tables 4 and 5 show the performance of runs trained on combining the *terms* extracted from the news post with the *early commenting* features (*activity+content*$_{t=10}$) for the 3-class and 5-class tasks respectively. We use features from the first ten minutes (i.e. $t = 10$) because we believe that they are very important for the prediction while keeping the advantage of quick access after the post is published.

From the results, we observe that the performance after combining the *terms* with the *early commenting* features leads to significant improvements over both *terms* and *activity+content*$_{t=10}$ runs. Also, we notice that this improvement is not consistent across the emotional reactions. For example, the least improvements over *terms* are observed for the reaction *sadness* (e.g., regarding the 3-class classification, the improvement of *terms+activity+content*$_{t=10}$ over the *terms* is 12.41%) whereas the largest improvements are observed for *joy* (28.93%).

One possible explanation for this inconsistency could be that in case of news that trigger a large number of *anger* and *sadness*, the textual features are very important predictors regardless if they are extracted from the news post or the comments' content. To investigate if there are any different patterns in commenting across the different reactions, we display the boxplot of the number of comments published in the first ten

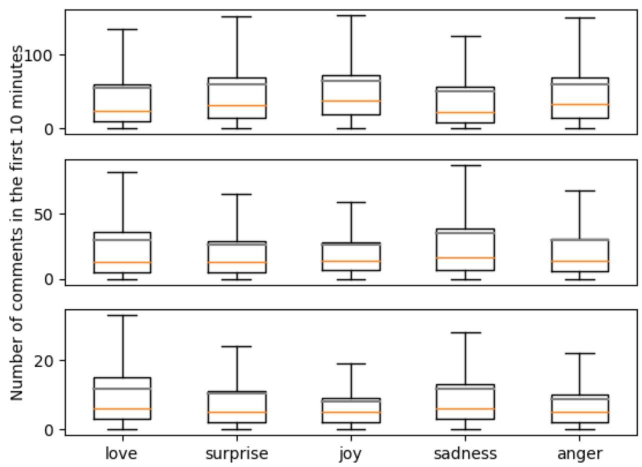
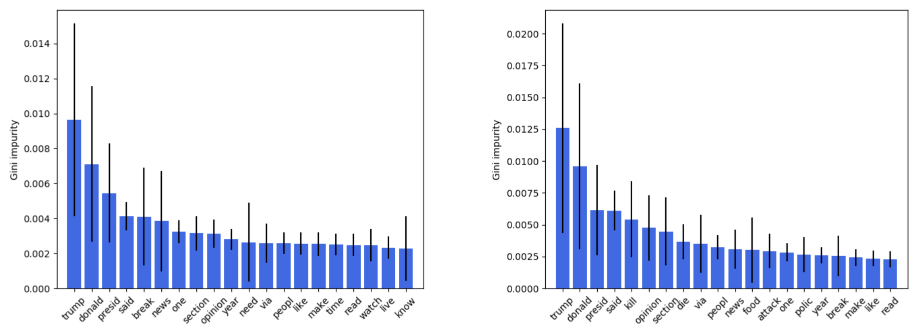



Fig. 3: Boxplot showing the number of comments published in the first ten minutes for the five emotional reactions and the classes *low, medium, high*. The yellow and black line refer to median and mean number of comments respectively.

minutes for each class and for each emotional reaction in Figure 3. The figure suggests that there is a difference in the distributions of *sadness* compared to *joy* and *surprise*. Therefore, we also calculate the statistical differences in the number of comments published in the first ten minutes for the posts that triggered a high number of *sadness* compared to *surprise* and *joy*. The results showed that there is a statistical difference between *sadness* and *surprise* (2-sample t-test, p-value $< 0.001$) as well as *sadness* and *joy* (2-sample t-test, p-value $< 0.001$). This suggests that users may have different commenting patterns on news posts that trigger *sadness* compared to those that trigger *surprise* or *joy*.

**Analysis on Terms.** We also carried out further analysis to explore which terms were the most informative for the prediction. As an example, we present the top 20 terms that are the most informative for the 3-class classification of the emotional reactions *surprise* and *sadness*. Figure 4 shows the most informative terms sorted by their Gini impurity score [6] for the reactions of (a) *surprise*, and (b) *sadness*. We observe that in both cases the most informative terms are *donald*, *trump* and *president*. We believe that this happens because of the time range of our collection that contains a lot of articles referring to US Elections 2016. In addition, we observe that there are also some terms that convey sentiment, such as the terms *kill* and *attack* that are informative for the emotional reaction *sadness*.

What is important to mention is that there are some words that are informative for both emotions (e.g., breaking, Donald, Trump, president). This observation suggests that there are terms that in general trigger either a large or a small number of emotional reactions regardless of the emotion. In addition to those terms, there are also terms (e.g., excited, attack, etc.) that are important only for a specific emotion (e.g. sadness).




(a) (b)

Fig. 4: Top 20 most important *terms* for the 3-class ordinal classification for (a) *surprise* and (b) sadness.

## 7 Conclusions and Future Work

In this study, we presented a methodology for predicting the ordinal level regarding the number of emotional reactions triggered on users by news posts. For our study, we focused on the following five emotional reactions: *love, surprise, joy, sadness,* and *anger*. We studied the prediction task by using features extracted from the *comments* of users. In addition, we studied the effectiveness of combining *early commenting* features with news posts' *terms* on predicting the emotional reactions.

Our results suggested that features extracted from *comments* are very important for the emotional prediction task. More importantly, we showed that the *commenting* features contain more predictive power compared to *terms* for all the reactions except for *sadness*. In addition, we showed that the different features extracted from *comments* are not equally important for the different emotional reactions because there are different commenting patterns across reactions. For example, we found that the negative ratio is the most important feature for *sadness* and *anger*. Finally, our results suggested that the most effective prediction models are those trained on both *terms* and *comments*.

In the future, we plan to address the prediction task as a regression problem and we will try to predict the exact number of each emotional reaction. In addition, we would like to explore the effect of time on the prediction task since news articles are extremely sensitive to time and temporal information can be very useful.

**Acknowledgments.** This paper 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).

This paper is partially supported by the BIGDATAGRAPES project (grant agreement N. 780751) that received funding from the European Union's Horizon 2020 research and innovation programme under the Information and Communication Technologies programme.

## 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. Alm, C.O., Roth, D., Sproat, R.: Emotions from text: machine learning for text-based emotion prediction. In: HLT '05. pp. 579–586 (2005)
4. 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)
5. Bandari, R., Asur, S., Huberman, B.A.: The pulse of news in social media: Forecasting popularity. In: ICWSM '12. pp. 26–33 (2012)
6. Breiman, L., Friedman, J., Stone, C., Olshen, R.: Classification and regression trees. The Wadsworth and Brooks-Cole statistics-probability series (1984)
7. Breiman, L.: Random forests. Mach. Learn. 45(1), 5–32 (Oct 2001)
8. Cheng, J., Adamic, L., Dow, P.A., Kleinberg, J.M., Leskovec, J.: Can cascades be predicted? In: WWW '14. pp. 925–936 (2014)
9. Clos, J., Bandhakavi, A., Wiratunga, N., Cabanac, G.: Predicting emotional reaction in social networks. In: ECIR '17. pp. 527–533 (2017)
10. Gerani, S., Carman, M., Crestani, F.: Aggregation methods for proximity-based opinion retrieval. ACM Transactions on Information Systems (TOIS) 30(4), 1–36 (2012)
11. Giachanou, A., Crestani, F.: Like it or not: A survey of twitter sentiment analysis methods. ACM Comput. Surv. 49(2), 28:1–28:41 (Jun 2016)
12. Giachanou, A., Harvey, M., Crestani, F.: Topic-specific stylistic variations for opinion retrieval on twitter. In: ECIR '16. pp. 466–478 (2016)
13. Giachanou, A., Rosso, P., Mele, I., Crestani, F.: Emotional influence prediction of news posts. In: ICWSM'18 (2018)
14. Go, A., Bhayani, R., Huang, L.: Twitter sentiment classification using distant supervision. Tech. rep., Standford (2009)
15. Goel, P., Kulshreshtha, D., Jain, P., Shukla, K.K.: Prayas at emoint 2017: An ensemble of deep neural architectures for emotion intensity prediction in tweets. In: WASSA' 17. pp. 58–65 (2017)
16. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: KDD '04. pp. 168–177 (2004)
17. Kiritchenko, S., Zhu, X., Mohammad, S.M.: Sentiment analysis of short informal texts. Journal of Artificial Intelligence Research 50(1), 723–762 (2014)

18. Lee, J.G., Moon, S., Salamatian, K.: Modeling and predicting the popularity of online contents with cox proportional hazard regression model. Neurocomputing 76(1), 134–145 (2012)
19. Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J.: Distributed representations of words and phrases and their compositionality. In: NIPS'13. pp. 3111–3119. USA (2013)
20. Paltoglou, G., Giachanou, A.: Opinion Retrieval: Searching for Opinions in Social Media, pp. 193–214. Springer International Publishing (2014)
21. Porter, M.F.: An algorithm for suffix stripping. Program 14(3), 130–137 (1980)
22. Roberts, K., Roach, M.A., Johnson, J., Guthrie, J., Harabagiu, S.M.: EmpaTweet: Annotating and detecting emotions on twitter. In: LREC '12. pp. 3806–3813 (2012)
23. Salton, G.: Automatic text processing: The transformation,analysis, and retrieval of information by computer. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA (1989)
24. Shu, K., Sliva, A., Wang, S., Tang, J., Liu, H.: Fake news detection on social media: A data mining perspective. SIGKDD Explor. Newsl. 19(1), 22–36 (Sep 2017)
25. Shulman, B., Sharma, A., Cosley, D.: Predictability of popularity: Gaps between prediction and understanding. In: ICWSM '16. pp. 348–357 (2016)
26. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M.: Lexicon-based methods for sentiment analysis. Computational linguistics 37(2), 267–307 (2011)
27. Tsagkias, M., Weerkamp, W., De Rijke, M.: Predicting the volume of comments on online news stories. In: CIKM '09. pp. 1765–1768 (2009)
28. Wen, M., Yang, D., Rose, C.: Sentiment analysis in mooc discussion forums: What does it tell us? In: EDM '14 (2014)
29. Yang, J., Leskovec, J.: Patterns of temporal variation in online media. In: WSDM '11. pp. 177–186 (2011)