Emotional Influence Prediction of News Posts

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, on-line news agents post news articles on social media platforms with the aim to spread information as well as to attract more users and understand their reactions and opinions. Predicting the emotional influence of news on users is very important not only for news agents but also for users, who can filter out news articles based on the reactions they trigger. In this paper, we focus on the problem of emotional influence prediction of a news post on users before publication. For the prediction, we explore a range of textual and semantic features derived from the content of the posts. Our results show that *terms* is the most important feature and that features extracted from news posts' content allow to effectively predict the amount of emotional reactions triggered by a news post.

Introduction

Different types of news trigger different emotional reactions on users who may feel happy, sad, or even angry after reading a piece of news. Predicting the emotional reactions triggered from news articles and their influence (e.g., how many people will feel sad after reading a news article) are two very important problems. For example, a system able to automatically predict the influence of each emotional reaction can help journalists and advertisers understand what people think, and which kind of news trigger a large volume of emotional reactions. Consequently, this can be helpful in cases when journalists want to generate or prioritize news articles that trigger large volume of a specific reaction.

Predicting the influence of emotional reactions is not a trivial problem. The structure of the network or other factors such as the publication date may affect the amount of reactions that a news post triggers. However, content is one of the most important factors that influences the emotional reactions (Alam et al. 2016). Intuitively, terms are very important for predicting the emotional reaction influence, since words often convey emotions and feelings. However, the majority of terms fail to capture crucial information such as information that has to do with the discriminating power of each term. In addition, the constantly-changing vocabulary makes the problem for those approaches, that only rely on words, more challenging. Semantic features, such as *entities* and *concepts*, can be useful to address this problem, since they can determine the semantic similarity of two posts.

In this paper, we focus on the emotional reactions that online news articles trigger on users, and we attempt to predict the influence of different emotional reactions of news articles using features extracted from the content of the news posts. We address the prediction task for five different emotional reactions (*love, surprise, joy, sadness, anger*) as both 3-class and 5-class classification problem to capture different volume levels. The 3-class task aims to predict if a news post will receive *low, medium*, or *high* number of reactions while the 5-class assigns one of the following volume labels: *very low, low, medium, high, very high*. Each emotional reaction is addressed independently to others. Our dataset consists of news articles published on *the New York Times* Facebook group.

Related Work

Popularity prediction has attracted a lot of attention and several studies tried to predict the popularity of different web items prior and after their publication. Different features have been studied and those from early activity have shown to be the most informative (Cheng et al. 2014; Yang and Leskovec 2011). Pre-publication prediction is more challenging and has received little attention (Bandari, Asur, and Huberman 2012; Tsagkias, Weerkamp, and De Rijke 2009). Tsagkias et al. (2009) tackled the problem of news articles' popularity prediction as a binary classification task by using a set of surface (e.g., length of the article), cumulative (e.g., near duplicates articles), textual, semantic, and real-world (e.g., temperature) features. Bandari et al. (2012) tackled the task as both regression and classification, and used various features, such as the category of the article and named entities. The results of their study suggested that popularity prediction is feasible without any early activity signals. However, recently Arapakis et al. (2017) reproduced and expanded the study of Bandari et al. (2012), and showed that predicting the popularity of news articles prior to their publication is not a viable task.

A large part of our problem is also related to opinion and emotion analysis that mainly focus on predicting the sentiment polarity (positive, negative, or neutral) or the emotion (e.g., anger, sadness, happiness, etc.) expressed in a piece of text (Giachanou and Crestani 2016). A num-

ber of researchers have analyzed opinions and emotions on different social media platforms (Kiritchenko, Zhu, and Mohammad 2014; Giachanou, Harvey, and Crestani 2016; Giachanou et al. 2017). Clos et al. (2017) focused on predicting the probabilities of different emotional reactions that news posts would trigger. However, predicting the probabilities does not provide information on which posts trigger a large number of reactions. Similarly, Alam et al. (2016) explored different feature sets such as character, words, stylometric (i.e., lexical richness), and psycholinguistic (e.g., affect, cognition) features to predict the mood level (ranging from 0 to 1) of readers on news articles, and found that the n-grams and stylometric features are the most important. More recently, Goel et al. (2017) focused on predicting the intensity of emotions in tweets using an ensemble of three neural-network approaches.

Emotional Influence Prediction

In this paper, we focus on the problem of *emotional influence pre-publication prediction* of news posts published on a social network. The problem can be stated as: *Given a news article post, the task consists in predicting the amount of emotional reactions that the post will trigger*. 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 using features that can be extracted from the content of the news post before its publication. Given a news post we assign to it one of the labels *low, medium, high* for the 3-class and one of these labels *very low, low, medium, high, very high* for the 5-class task.

Term Frequencies

For the *terms* feature, we use the bag-of-words representation. Each term in the vector is weighed using the term frequency-inverse document frequency (TF-IDF) approach that considers how important is the term in a corpus. Contrary to other studies (Tsagkias, Weerkamp, and De Rijke 2009), that used only a small percentage of the vocabulary to represent textual features, we are using all the terms that appear in the collection (without stopwords). In the rest of the paper, we use *terms* to refer to the TF-IDF representation of the terms.

Similarities

For each news post we compute its content similarity with the documents that triggered a large number of each emotion. Let d_i be a news post that has to be classified into one of the k classes (e.g. low, medium, high). Also, let H be a hyper-document (i.e., aggregation of several documents) of the documents that attracted a large number of a specific emotional reaction e. We can calculate different similarity measures between d_i and H as described below.

Jaccard similarity. This measure computes the Jaccard similarity between a document d_i and the hyper-document H as: $Jaccard(d_i, H) = |W_d \cap W_H| / |W_d \cup W_H|$. In other words, the similarity is calculated using the set of common terms that appear in the document and the hyper-document.

Cosine similarity. Cosine similarity is a well known similarity measure and is estimated as:

$$cosine(d_i, H) = \frac{\sum_{w \in d_i} P(w|d_i) P(w|H)}{\sqrt{\sum_{w \in d_i} P(w|d_i)^2 \sum_{w \in H} P(w|H)^2}}$$

where $P(w|d_i)$ and $P(w|H)$ are the probabilities of a word w occurring in the post d_i and hyper-document H.

Symmetric Kullback-Leibler Divergence. This measure computes the similarity between the post d_i and the hyper-document H based on the distance function known as KL-divergence. We consider the symmetric version of the KL-divergence to compensate for terms that do not appear in any of the distributions. This measure is calculated as: $\delta(H, d_i) = 1/2[\, KLD(d_i|H) + KLD(H|d_i)]$, where

$$KLD(d_i|H) = \sum_{w \in d_i} P(w|d_i) \cdot \log \frac{P(w|d_i)}{P(w|H)}$$

Normalised Kullback-Leibler Divergence. The normalized version of the KL-divergence is calculated as follows:

$$KLD(d_i|H) = \sum_{w \in d_i} P(w|d_i) \cdot \log \frac{P(w|H)}{P(w|D)}$$

where $P(w|D)$ is estimated based on the background model of the entire collection.

Reactions Entropy

Terms do not capture the discriminative power of the words. Therefore, we also consider the *entropy* that can measure how well a term separates documents that attract a high number of emotional reactions from those with a lower number. This measure is inspired by the temporal entropy that was used to determine the time-stamp of a document (Kanhabua and Nørvåg 2008). First we divide the documents based on the number of the reactions they received given a specific emotion. Let $V = \{v_1, v_2, ..., v_i\}$ be a set of partitions where v_1 and v_i contain the documents with the highest and lowest number of reactions, respectively. Then the entropy of a word w_i is defined as follows:

$$VE(w_i) = 1 + \frac{1}{\log N_V} \sum_{v \in V} P(v|w_i) \cdot \log P(v|w_i)$$

where N_V is the number of partitions and $P(v|w_i)$ is the probability of the word w_i to occur in the partition v and is calculated as: $P(v_j|w_i) = tf(w_i, v_j) / \sum_{m=1}^{N_V} tf(w_i, v_k)$ where $tf(w_i, v_j)$ is the frequency of w_i in v_j. Following, we can calculate a score between a document d and each volume partition v as:

$$Score(d, v) = \sum_{w \in d} VE(w) \cdot P(w|d) \cdot \log \frac{P(w|v)}{P(w|D)}$$

Semantics

Semantics can capture the similarity of documents that do not share similar terms. The reason for introducing these features is that they can also be very useful for certain entities

and concepts which trigger specific reactions and specific volume of such reactions (e.g., music groups). The semantics features we explore are *entities*, *concepts*, and *sentiment*.

For *entities*, we extract names of *persons, companies, organizations, locations, facilities, crimes, drugs, sport events, movies,* and *health conditions* and we count how many times each entity category appears in each post. The second group of semantic features includes the *concepts* of the post such as *art and entertainment, technology and computing, food and drink*, etc. We use the primary concepts extracted from every post as additional features.

Finally, we use the EmoLex lexicon (Mohammad and Turney 2013) to calculate the *sentiment* (positive and negative) and *emotion* (anger, anticipation, disgust, fear, joy, sadness, surprise, trust) scores in each post. Even though there is no available lexicon for the reaction *love*, we believe that words that convey *joy* will be also useful for *love*. For generating the scores, we count the number of the predefined words provided by EmoLex in each post (for example, to compute the *sadness* score for a post, we count the words that, in the lexicon, are annotated to indicate *sadness*, and which appear in the post).

Publication Date

Regarding the publication date, we consider the following: *Day of month (1-31), month of publication (1-12), hour of the day (0-23), day of the week (1-7), week of the month.*

Experiments

We collected news posts from *The New York Times* group[1] in Facebook together with the amount of 5 different emotional reactions: *love, surprise, joy, sadness,* and *anger* for each post. We used Facebook API[2] to collect the posts, the reactions, and the comments. 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.

We divided the collection into 3 (and 5) balanced classes with regards to the amount of each emotional reaction. 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 the *very low, low, medium, high, very high*. We predicted the amount level of the following five different emotional reactions: *love*, *surprise*, *joy*, *sadness*, and *anger*. The emotional reactions, that are available on Facebook, were addressed individually.

For the classification we used Random Forest. To extract the entities and the concepts from each post we used Alchemy API[3]. We used the EmoLex lexicon (Mohammad and Turney 2013) to calculate the sentiment and emotion scores. Pre-processing of the posts involved stop-words removal and stemming with Porter stemmer. All the similarity measures were calculated between a post and a hyperdocument that included 10% of the posts appeared in the training set that collected the highest number of the reaction.

[1]https://www.facebook.com/nytimes/

[2]https://developers.facebook.com/

[3]https://console.bluemix.net/

We report F1 score for both 3-class and 5-class classifications and for each emotional reaction. We use the *publication date* feature as our baseline. The same baseline is used by Tsagkias et al. (2009). Significance is measured with the McNemar test which is appropriate for comparisons of nominal data.

Results and Discussion

Table 1 shows the results for the 3-class classification (i.e., *low, medium, high*) for each emotional reaction (*love, surprise, joy, sadness, anger*) using the different features. From the results, we observe that the most important feature is the *terms*. Indeed, *terms* manage to outperform all the rest of the features, for all the emotional reactions. *Terms* were also proved to be very important for news articles popularity prediction (Tsagkias, Weerkamp, and De Rijke 2009). However, we notice that in Tsagkias et al., the rest of the features performed slightly worse than *terms* whereas we found large negative differences (e.g. $\Delta = 24.97\%$ for *terms* over *date* for the classification regarding *love*). One explanation for this, is that in Tsagkias et al., the textual feature refers to the top 100 terms ranked by log-likelihood, whereas in our study we used the whole vocabulary (after removing stopwords) representing more than 20,000 unique terms.

In addition, we observe that the rest of the features manage to obtain similar performances and in most of the cases they significantly outperform the baseline. The feature *similarities* performs better than others. This result confirms the importance of content-based features for predicting the amount of emotional reactions. Also, we observe that *date* has little predictive power. That means that the publication date is not important for predicting the amount of emotional reactions that a news post will trigger. This result is consistent with previous studies that focused on popularity prediction of news articles (Arapakis, Cambazoglu, and Lalmas 2017; Tsagkias, Weerkamp, and De Rijke 2009).

Table 1 also shows the results when all the features are combined. Given the large difference in performance between *terms* and the rest of the features, we combine all the features for two settings: without (*All (- terms)*) and with (*All*

Table 1: Performance results (F1-scores) for the 3-class prepublication prediction. Scores with $*$ indicate statistically significant improvements with respect to the *date* approach. Scores in italics indicate the best performance per emotional reaction (i.e. per column).

	Love	Surprise	Joy	Sadness	Anger
Date	0.382	0.371	0.386	0.383	0.401
Terms	*0.491*$*$	*0.494*$*$	*0.578*$*$	*0.555*$*$	*0.597*$*$
Similarities	0.414$*$	0.424$*$	0.506$*$	0.475$*$	0.513$*$
Entropy	0.408$*$	0.418$*$	0.465$*$	0.450$*$	0.482$*$
Entities	0.388	0.402$*$	0.491$*$	0.423$*$	0.468$*$
Concepts	0.381	0.413$*$	0.448$*$	0.458$*$	0.489$*$
Sentiment	0.371$*$	0.390$*$	0.442$*$	0.435$*$	0.454$*$
All (-terms)	0.466$*$	0.472$*$	0.534$*$	0.527$*$	0.555$*$
All (+terms)	0.478$*$	0.486$*$	0.554$*$	0.543$*$	0.576$*$

(+ terms)) terms. Surprisingly, the model that is trained only on *terms* performs better compared to the combination of the features. This happens for both settings. However, as already mentioned *terms* represent more than 20,000 features, whereas the rest of the features refer to only 33 features.

Table 2 shows the results for the 5-class classification (*very low, low, medium, high, very high*) for each emotional reaction for the pre-publication prediction. This task is more difficult compared to the 3-class classification, and therefore, the performance results are lower. We notice that the results are consistent to the 3-class classification with *terms* to outperform the rest of the features. Similar to the 3-class classification *date* has the least predictive power. In addition, we observe that *entities* and *concepts* have inconsistent effects across the different emotional reactions. For example, *entities* contain some predictive power for the emotion *joy* (Δ = 33.22% over *date*), whereas their predictive power for *love* is little (Δ = 7.68% over *date*). Similar to the 3-class classification, the model that is trained only on the feature of *terms* performs better compared to the combination of the features.

Conclusions and Future Work

In this study, we presented a methodology for predicting emotional reactions of news posts using features extracted from the content of the news posts. For our study, we focused on the following five emotional reactions: *love, surprise, joy, sadness,* and *anger*. We explored content-based features that refer to *textual* and *semantic* features and we analyzed their effectiveness in predicting the amount of the emotional reactions. The results showed that *terms* is the most important feature. Other *textual* and *semantic* features also contain some predictive power but less than *terms*. Surprisingly, a model trained only on *terms* outperformed even the combination of all features.

As future work, 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 post-publication features that are usually extracted from users' early activity.

Table 2: Performance results (F1-scores) for the 5-class pre-publication prediction. Scores with $*$ indicate statistically significant improvements with respect to the *date* approach. Scores in italics indicate the best performance per emotional reaction (i.e. per column).

	Love	Surprise	Joy	Sadness	Anger
Date	0.238	0.228	0.241	0.234	0.245
Terms	*0.330*$*$	*0.336*$*$	*0.402*$*$	*0.371*$*$	*0.393*$*$
Similarities	0.262$*$	0.275$*$	0.328$*$	0.299$*$	0.319$*$
Entropy	0.247$*$	0.259$*$	0.288$*$	0.273$*$	0.289$*$
Entities	0.257$*$	0.262$*$	0.337$*$	0.262$*$	0.292$*$
Concepts	0.256$*$	0.269$*$	0.297$*$	0.283$*$	0.302$*$
Sentiment	0.240	0.253$*$	0.291$*$	0.265$*$	0.275$*$
All (-terms)	0.303$*$	0.306$*$	0.355$*$	0.333$*$	0.350$*$
All (+terms)	0.320$*$	0.326$*$	0.382$*$	0.354$*$	0.373$*$

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

Alam, F.; Celli, F.; Stepanov, E. A.; Ghosh, A.; and Riccardi, G. 2016. The social mood of news: Self-reported annotations to design automatic mood detection systems. In *PEOPLES '16*, 143–152.

Arapakis, I.; Cambazoglu, B. B.; and Lalmas, M. 2017. On the feasibility of predicting popular news at cold start. *Journal of the Association for Information Science and Technology* 68(5):1149–1164.

Bandari, R.; Asur, S.; and Huberman, B. A. 2012. The pulse of news in social media: Forecasting popularity. In *ICWSM '12*, 26–33.

Cheng, J.; Adamic, L.; Dow, P. A.; Kleinberg, J. M.; and Leskovec, J. 2014. Can cascades be predicted? In *WWW '14*, 925–936.

Clos, J.; Bandhakavi, A.; Wiratunga, N.; and Cabanac, G. 2017. Predicting emotional reaction in social networks. In *ECIR '17*, 527–533.

Giachanou, A., and Crestani, F. 2016. Like it or not: A survey of twitter sentiment analysis methods. *ACM Comput. Surv.* 49(2):28:1–28:41.

Giachanou, A.; Rangel, F.; Crestani, F.; and Rosso, P. 2017. Emerging sentiment language model for emotion detection. In *CLiC-it '17*.

Giachanou, A.; Harvey, M.; and Crestani, F. 2016. Topic-specific stylistic variations for opinion retrieval on twitter. In *ECIR '16*, 466–478.

Goel, P.; Kulshreshtha, D.; Jain, P.; and Shukla, K. K. 2017. Prayas at emoint 2017: An ensemble of deep neural architectures for emotion intensity prediction in tweets. In *WASSA' 17*, 58–65.

Kanhabua, N., and Nørvåg, K. 2008. Improving temporal language models for determining time of non-timestamped documents. In *ECDL '08*, 358–370.

Kiritchenko, S.; Zhu, X.; and Mohammad, S. M. 2014. Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research* 50(1):723–762.

Mohammad, S. M., and Turney, P. D. 2013. Crowdsourcing a word–emotion association lexicon. *Computational Intelligence* 29(3):436–465.

Tsagkias, M.; Weerkamp, W.; and De Rijke, M. 2009. Predicting the volume of comments on online news stories. In *CIKM '09*, 1765–1768.

Yang, J., and Leskovec, J. 2011. Patterns of temporal variation in online media. In *WSDM '11*, 177–186.