Cannot Predict Comment Volume of a News Article before (a few) Users Read It

Lihong He,[1] Chen Shen,[1] Arjun Mukherjee,[2] Slobodan Vucetic,[1] Eduard Dragut [1]
[1]Department of Computer and Information Sciences, Temple University, Philadelphia, PA
[2]Department of Computer Science, University of Houston, Houston, TX
{lihong.he, chen.shen, vucetic, edragut}@temple.edu,
arjun@cs.uh.edu

Abstract

Many news outlets allow users to contribute comments on topics about daily world events. News articles are the seeds that spring users' interest to contribute content, i.e., comments. An article may attract an apathetic user engagement (several tens of comments) or a spontaneous fervent user engagement (thousands of comments). In this paper, we study the problem of predicting the total number of user comments a news article will receive. Our main insight is that the early dynamics of user comments contribute the most to an accurate prediction, while news article specific factors have surprisingly little influence. This appears to be an interesting and understudied phenomenon: collective social behavior at a news outlet shapes user response and may even downplay the content of an article. We compile and analyze a large number of features, both old and novel from literature. The features span a broad spectrum of facets including news article and comment contents, temporal dynamics, sentiment/linguistic features, and user behaviors. We show that the early arrival rate of comments is the best indicator of the eventual number of comments. We conduct an in-depth analysis of this feature across several dimensions, such as news outlets and news article categories. We show that the relationship between the early rate and the final number of comments as well as the prediction accuracy vary considerably across news outlets and news article categories (e.g., politics, sports, or health).

Introduction

Commenting on news is a common form of participation in contemporary news consumption, and it is one of the most common forms of citizen engagement online (Emmer, Vowe, and Wolling 2011). A key indicator of user participation in daily news events is the volume of user comments reacting to a news article (Prochazka, Weber, and Schweiger 2018; Ziegele et al. 2018). Several works propose methods to predict it (Tsagkias, Weerkamp, and De Rijke 2009; Balali, Asadpour, and Faili 2017). They use a large number of features, which can be broadly categorized into article content, meta-article (e.g., outlet or category), temporal (e.g., date and time of publication), and semantic (e.g., named entities). They model the prediction problem as a classification problem. For example, they determine if an article will receive a "high" or "low" volume of comments. One of their key findings is that *(i) predictors based on article content features alone are the best performers and (ii) one may even achieve high accuracy with such predictors.*

The social science community, in particular the Communication community, argues that quality discourse emerges only when many users participate in commenting on a news article and when there is interactivity among users, i.e., users comment/reply to prior comments (Kiousis 2002). The general questions pursued in this space aim to understand the factors affecting participation and interactivity in the comment section of an article (Altheide and Schneider 2012; Weber 2014). Some studies (via face-to-face interviews) show that factors from previously posted user comments affect the involvement of new users and ultimately increase users' willingness to engage in online news discussions (Mishne, Glance et al. 2006; Ziegele and Quiring 2013). *They conclude that a large fraction of the comments an article receives– up to 50%– do not respond to the journalistic value of a news article, but rather to a previously posted user comment* (Singer 2009; Ruiz et al. 2011).

Our work in this paper is motivated by the apparent disagreement between the findings from different communities. We aim to understand the factors– ranging from article content to observed dynamics of user comments – on predicting the eventual comment volume an article receives. One may notice a problem here. On the one hand, one would like to predict the comment volume before an article's publication. On the other hand, one has access to user comments only after the article has been online for some time. Nonetheless, the number of eventual total comments an article will get remains relevant. Thus, we relax the problem by formulating it as follows:

Problem: Given a news article A and its first α user comments, predict N_A, the eventual number of comments A receives.

We can draw a parallel to the problem of predicting the distance traveled by a ball (news article), say in soccer. The distance depends on the ball itself and the person who kicks it (news outlet and author), but it also depends on factors, like launch angle and exit speed, unrelated to the ball. Those are only known shortly after the ball was kicked and traveled a short distance. Similarly, we expect the first α comments to give us the missing information necessary to predict the eventual number of comments the article receives.

One may notice that N_A is not well defined: theoretically, it may continue to grow endlessly with time. In practice, however, this does not happen for news articles. We monitored each article for 3 months. The articles accumulated 99.84% of their overall comment volumes within a week and 99.97% within a month. *Consequently, hereafter N_A is the number of comments accumulated in the first week by news article A*, which empirically is almost identical to the true total number of comments that A receives in practice.

The magnitude of α and its relation to N_A may trivialize the problem, say, "look at the first α = 1,000 user comments and predict if the article will receive N_A = 1,050." We study the dynamics among the very first few comments and aim to predict if N_A will reach 1,050. We empirically test α = 5, 10, 15, 20, and 50. The accuracy of prediction increases by about 9% from α = 5 to α = 10, and by less than 2% from α = 10 to α = 50. We set α = 10 in all our empirical studies. Thus, we aim to predict the eventual number of comments an article will receive based on the observations among the first 10 comments. It takes 25 minutes on average for the 10$^{\text{th}}$ comment to arrive since the posting of the first comment.

After evaluating and comparing the prediction performance of various models on 19K articles and over 9M comments from 6 news outlets, we show that signals gathered from the early dynamics of user comments largely influence the ability to predict the eventual number of comments on a news article, while the contribution from article features is small. This finding is consistent with the conclusion from the social science community. We study the user comment features and identify the feature of "the arrival rate of *early* comments" (`rate`), which is defined as the number of comments per minute, as a key missing link in accurately predicting the comment volume.

We study `rate` across six major U.S. and U.K. news outlets. We notice that the performance of the rate-based model varies across news outlets. It is highly accurate for news articles published by Wall Street Journal, but less accurate for Fox News and the Guardian. With regard to the analysis of `rate` by categories, we note that the characteristic of the rate model differs across categories. For example, "Politics" is particularly sensitive to the rate of early comments. This is common across all news outlets. "Health," on the other hand, is less sensitive to the rate of early comments.

We also consider the relationship between news outlets and categories. We study the characteristic of `rate` model for each outlet-category pair. We find that the rate model performs the best at Wall Street Journal in most of the categories, and the eventual user activity is more affected by the initial engagement in political areas across all outlets.

We believe that our findings are of interest to social scientists because they reveal the relationship between the early user commenting behavior and the total comment volume of a news article, across news outlets and news categories. *This appears to be a trait unique to news readership communities. Features based on early arrival pattern in other social communities, such as Twitter and Facebook, on related prediction tasks have limited predictive power* (Backstrom et al. 2013; Weng, Menczer, and Ahn 2014). Network topology features are more effective in those tasks; most of those features are not applicable to news readership communities as they lack an underlying network.

We make the following contributions in this paper:

- We postulate that one cannot predict the comment volume of an article unless one considers (early) user commenting activity.
- We identify the importance of (early arrival) `rate` in the task of predicting the comment volume of a news article.
- We perform extensive empirical studies by news outlets and news categories, and show additional novel insights.

Related Work

We review several lines of research about user generated content in news domain and social networks.

News Domain. Mining and analyzing the content produced by users in news media are popular research directions. Some of the explored problems include examining the relationships between news comment topicality, temporality, sentiment, and quality (Diakopoulos and Naaman 2011); analyzing the sentiment of comments and headlines of news article (Dos Rieis et al. 2015); news propagation (Tan, Friggeri, and Adamic 2016); personalized recommendation of news stories (Shmueli et al. 2012); topic clustering of news articles (Aker et al. 2016); and modeling and predicting comment volume (Tsagkias, Weerkamp, and De Rijke 2010; Balali, Asadpour, and Faili 2017; Rizos, Papadopoulos, and Kompatsiaris 2016; Tatar et al. 2011). The prediction of comment volume is treated as a (binary) classification problem (e.g., "High"/"Low" volume) (Tsagkias, Weerkamp, and De Rijke 2009) and regression classification problem (Balali, Asadpour, and Faili 2017) in previous studies. (Tatar et al. 2011) uses a simple linear regression model with early user activity during a short observation period after publication to predict the comment volume of articles. Besides, (Aragón et al. 2017) describes few current models of the growth of comment threads.

Social Networking Platforms. Understanding user behavior in social networking platforms, e.g. Twitter and Facebook, has attracted large interest. Some studies aim to understand user conversations and their evolution over time (Wang, Ye, and Huberman 2012), while others study the commenting and comment rating behavior (Siersdorfer et al. 2014). A number of works address problems related to the prediction of reply volume. They employ a variety of features, such as bag of words (Yano and Smith 2010), the arrival patterns of early comments (Backstrom et al. 2013), network specific, like "followship," and historical behavior in retweet (Artzi, Pantel, and Gamon 2012). The prediction problem itself is modeled in a variety of ways. Some model it as a binary classification problem (Artzi, Pantel, and Gamon 2012; Backstrom et al. 2013). Other formulations include regression (Tsur and Rappoport 2012), multi-label classification (Weng, Menczer, and Ahn 2014), cascades size prediction (Cheng et al. 2014; Kobayashi and Lambiotte 2016), self-exciting point process (Mishra, Rizoiu, and Xie 2016), or Hawkes process modeling (Zhao et al. 2015; Rizoiu et al. 2018). The popularity prediction

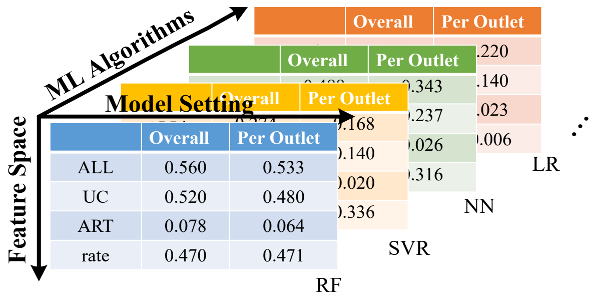


Figure 1: Methodology illustration.

Outlets	Aw.C	Mean Vol. (STD)	Mean Log Vol. (STD)
Washington Post	6,470	364.8 (942.2)	1.88 (0.76)
Daily Mail	6,046	264.1 (560.4)	1.99 (0.62)
Wall Street Journal	2,516	189.4 (346.5)	1.74 (0.7)
Fox News	1,739	1,896.5 (3790.2)	2.47 (0.94)
the Guardian	1,697	504.4 (716.8)	2.46 (0.45)
New York Times	965	481.4 (530.9)	2.38 (0.6)
Overall	19,433	465.8 (1400.6)	2.02 (0.74)

Table 1: Data summary. Aw.C = Articles with Comments.

and modeling on social media is also a fruitful research (Liao et al. 2019; Mishra 2019; Lin et al. 2019).

Our Work. While our work shares some commonalities with these lines of work, it also distinguishes from them in several important ways. The main difference with the work in the news domain is that we focus on the analysis of (early) user commenting activity and its importance on the prediction of comment volume of a news article. The study of using user comments in early stage to predict the final comment volume has been analyzed in (Tatar et al. 2011), which only evaluates a simple linear prediction model with limited factors from articles and comments. While our work in this paper explores more features related to article and early user commenting activity (both new and old). Moreover, we consider multiple machine learning techniques, both linear and nonlinear. According to our analysis, we show that (1) the prediction problem is difficult and, thus, nonlinear models are better suited to solve the prediction problem; and (2) the proposed new feature `rate` remains its dominant power across machine learning techniques. The key distinction with the work in social networks is that the social communities at news outlets are *not* networked. The works in social networks make heavy use of the network topology and the community around a user, e.g., followers and friends. These are not applicable in our setting. While arrival patterns of user posts are considered in previous works, they are not as consequential in their respective prediction tasks as `rate` is in ours. For instance, arrival patterns as defined in (Backstrom et al. 2013) contribute less than 4.4% to the overall performance, compared to 90% on average for `rate`. The family of features "growth rate" (Weng, Menczer, and Ahn 2014), which includes a feature similar to `rate`, has a much weaker predictive power than that of their other features. Since their `rates` are inconsequential, these works do not pursue any in depth studies of their `rates`. We present a study of `rate` along several dimensions, such as news outlet and news category.

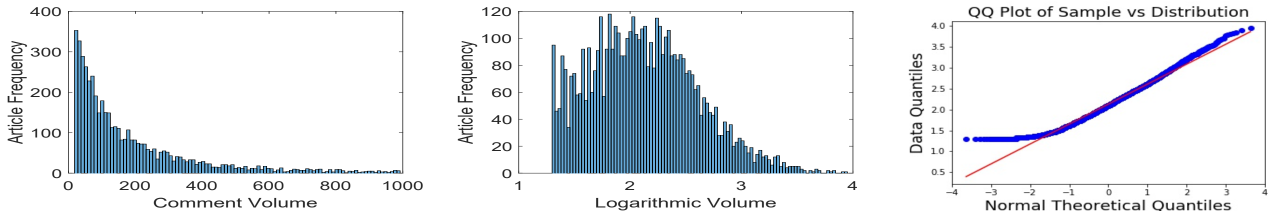
Methodology

Our goal is to understand the feature subset most important for predicting the comment volume of a news article. The prediction of the eventual comment volume is a regression problem. Figure 1 summarizes our methodology. We conduct our study along three dimensions: (i) Feature Space, (ii) Model Setting, and (iii) Machine Learning (ML) Algorithms. In (i), we analyze the entire feature space (denoted as ALL), the user comment only (UC) and news article only (ART) features, as well as `rate` (which is a single feature in UC) alone. In (ii), we consider two settings: *global* and *local*. The global dataset has the news articles from all the news outlets. The local dataset has the news articles grouped by news outlet. In (iii), we use 4 representative ML algorithms for regression: Random Forest (RF), Support Vector Regression (SVR), Neural Network (NN), and Linear Regression (LR). In the figure, a slice represents an instance from the cross product of (i), (ii), and (iii).

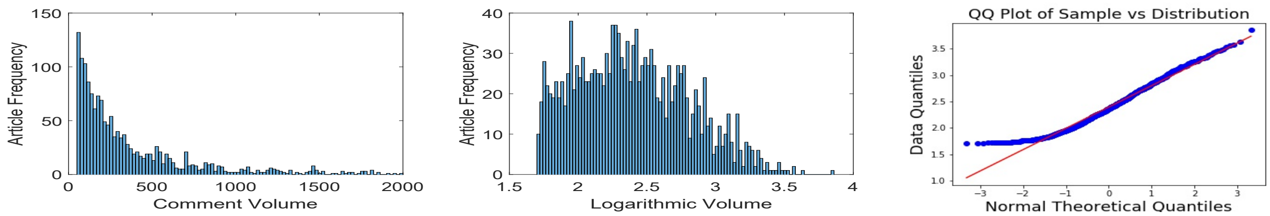
Data

We collected news articles with comments from Oct. 2015 to Feb. 2017 from the following six news outlets: Washington Post, Daily Mail, Wall Street Journal, Fox News, the Guardian, and New York Times. We crawled the topics of the collected articles from Google News and monitored their duration there as well. The dataset has over 19K articles with comments and 9M comments (including replies). We monitored each article for 3 months. We observed that on average each article accumulates 99.84% of its overall comment volume within a week. Recall that in this paper N_A is the number of comments accumulated in the first week by news article A. This is the number we try to predict.

Figure 2 illustrates the distribution of news articles at these outlets in our dataset (we give 2 outlets due to space constraints). We observe a heavy-tailed distribution in each outlet when we plot the distribution by number of comments (the first graph per outlet). If we plot the comment volume in the log scale, the distributions are (or close to be) bell-shaped. It seems that the volume distributions are nearly log-normal. To test this hypothesis, we provide the Q-Q plot of the logarithmic volume as the third column of graphs in Figure 2 for each outlet. We can see that most of the points stay on or very close to the straight line, except for some head and tail data points. It shows that the distribution of user



(a) Daily Mail



(b) the Guardian

Figure 2: The distribution of comment volume and logarithmic volume per news outlet. In the first column of graphs, the articles with more than 1,000/2,000 comments are discarded to make the graphs visible. For each outlet, article frequency on the y-axis of the first two graphs is the number of articles, the third graph provide the Q-Q plot of the logarithmic volume.

comment volume over news articles is well approximated by the log-normal distribution.

Table 1 describes the number of articles in each news outlet. We give the mean and standard deviation (STD) of comment volume and logarithmic volume, displayed in the last two columns of Table 1. Considering the log-normal distribution of comment volume, we will work on the prediction of comment volume in log scale.

Predicting Comment Volume

In this section, we show that factors drawn from (early) user commenting activity are the keys to accurately predict the comment volume a news article receives. We describe the feature set and the experimental setting, and report on the prediction performance in this section. We follow the methodology described above. Finally, we demonstrate that `rate` is the dominant feature.

Features

Table 2 summaries the set of features. There are five groups of features: topic, article, comment, news factors, and misc features. We introduce 11 new features.

Topic features. We observe that some topics, such as Ebola Outbreak or Paris (terrorist attack), trigger more discussion than others. Therefore, we include these finer grain topics as one of the predictive features. The fine grain topics are rarely provided by the news outlets. We extract them from Google News along with their parent categories, e.g., Health and World. We collect 768 distinct topics in total.

Article features. All features in this group are related to news articles. They can be categorized into metadata and text features. The metadata features include *month*, *day*, *hour*, *wom* (week of the month), and *dow* (day of the week) of the publication. Previous work argues that the time of publication may affect the comment volume an article receives (Tsagkias, Weerkamp, and De Rijke 2010). The rest of the features in this group are extracted from article title and content. The features *art_question* and *art_exclaim*, suggested in (Backstrom et al. 2013), show whether there are '?' and '!' in article title. The features *art_num_ne_loc*, *art_num_ne_per*, *art_num_ne_org*, and *art_num_ne_misc*, proposed in (Tsagkias, Weerkamp, and De Rijke 2009), provide the number of locations, persons, organizations, and miscellaneous named entities mentioned in the article content. We utilize Stanford NER to extract named entities.

The feature *art_senti_score* gives the sentiment score of article content. To calculate this sentiment score, we make use of an effective document representation and sentiment analysis model proposed in (Yang et al. 2016), which is a word-sentence-document level bi-directional GRU neural network with two levels of attention. We initialize the 100 dimensional word embeddings with pre-trained Glove word vectors. The model is trained on IMDB dataset (25K reviews with positive or negative rates) for 10 epochs. We use the output from the prediction layer of the deep model as sentiment score, which is in the range of [0, 1]. Articles with score close to 0 are predicted with overall negative sentiment, while articles with score close to 1 are predicted with overall positive sentiment.

Comment features. The comment features are extracted from the first α comments of an article. The feature *rate* measures the number of comments per unit of time, which is computed as

$$rate = \frac{i}{t_i - t_1}$$

Here, $t_i - t_1$ is the elapsed time (in minutes) between the first and *i*-th comment (as in (Weng, Menczer, and Ahn 2014)); it is 25 minutes on average for the 10^{th} comment. The feature *fc_mid* is the absolute difference between the time of

Feature	Description
Topic features	
topic*	Topic of article.
Article features	
month	Published month of article (1-12).
day	Published day of the month (1-31).
hour	Published hour of the day (0-23).
wom	Week of the month (1-5).
dow	Day of the week (1-7).
author	Author of article.
art_length	Article content length.
art_question	Whether there is a '?' in article title.
art_exclaim	Whether there is a '!' in article title.
art_num_ne_loc	Number of location-type named entities in article content.
art_num_ne_per	Number of person-type named entities in article content.
art_num_ne_org	Number of org.-type named entities in article content.
art_num_ne_misc	Number of miscellaneous-type named entity in article content.
art_senti_score*	Sentiment score of article content.
Comment features	
rate*	Arriving rate of the first α comments.
fc_mid*	Time of first comment - 12am (in min.)
uniq_com	Number of unique commenters.
num_reply	Number of replies.
num_thread	Number of threads.
num_question	Number of '?'.
num_exclaim	Number of '!'.
num_words	Number of words.
complexity	Complexity of the first α comments.
has_url	Whether there is a link.
num_ne_com	Number of named entities.
depth*	Depth of the comment tree.
width*	Width of the comment tree.
avg_senti_score*	Average of sentiment scores of the first α comments.
num_likes	Aggregated number of likes.
num_dislikes	Aggregated number of dislikes.
News Factors	
continuity*	Time difference between article's publication and its topic's appearance.
aggression*	Fraction of aggressive words.
position	NA.
MISC features	
pub_resp	Time difference (in minutes) of first comment to article's publication.
inter_art*	Defined as $\frac{\lvert NE_{art} \bigcap NE_{com_i} \rvert}{\lvert NE_{art} \rvert}$
inter_com*	Defined as $\frac{\lvert NE_{art} \bigcap NE_{com_i} \rvert}{\lvert NE_{com_i} \rvert}$

Table 2: Features utilized in prediction experiments. We organize them into 5 groups. The underlined ones belong to two groups, one of which is news factors. The features labeled with * are new features in the prediction task.

the first comment and midnight. The feature *uniq_com* gives the number of unique commenters in the first α comments, which is one of the indicators for the arrival pattern of the first α comments (Backstrom et al. 2013). The features *num_reply* and *num_thread* give the number of replies and discussion threads, respectively.

The features *num_question*, *num_exclaim*, *num_words*, *complexity*, *has_url*, *num_ne_com*, and *avg_senti_score* study the text of comments. The meaning of these features are provided in Table 2. *Complexity* measures the cumulative entropy of terms within the first α comments (Rowe and Alani 2014). It is given by:

$$C(c) = \frac{1}{|T(c)|} \sum_{t \in T(c)} tf(t,c)(\log|T(c)| - \log tf(t,c))$$

Here, $T(c)$ is the set of unique terms in comment c and $tf(t,c)$ is the frequency of each term $t \in T(c)$.

The feature *avg_senti_score* is the average of the sentiment scores of the first α comments. The sentiment score of a comment is given by the deep model in (Yang et al. 2016).

The features *depth* and *width* are extracted from the comment reply tree T_A of an article A. T_A is constructed as follows. An article A is the root of T_A. Comments that are not replies (responses) of any previous comments are the children of A (the article). The replies of a comment are its children nodes. The *depth* of the reply tree T_A is the number of levels of T_A. If L denotes the levels of the reply tree T_A, the *width* is given by

$$WIDTH = \max_{j \in \mathbf{L}} \sum_{i=1}^{m_j} s_{ji},$$

where m_j is the number of sibling groups in level j, and s_{ji} is the count of nodes in the i-th sibling group in level j. A feature named *depth* appears in (Cheng et al. 2014), but its definition and meaning are different from ours.

The features *num_likes* and *num_dislikes* count the aggregated number of likes and dislikes received by the first α comments. The consideration of number of likes is proposed in (Backstrom et al. 2013).

News factors. We implement a number of novel features based upon *news value theory*, which states that journalists and media users select news items depending on news factors such as continuity, negativity, and aggression (Weber 2014; Ziegele and Quiring 2013). These dimensions were confirmed after extensive face-to-face interviews with users who commented on news stories online (Ziegele, Breiner, and Quiring 2014). We create novel features to quantify many of the news factors. Some of them are encountered in other studies, e.g., climate change (Olteanu et al. 2015), but with different definitions. We quantify the factor *continuity* (if a news article continues issues that are already on the media agenda) (Weber 2014) as the time difference (in minutes) between article's publication time and its topic's appearance in Google News. The intuition is that a user's interest to comment on an article diminishes the farther its publication time is from the time when the news event first broke in. We additionally consider the factors *negativity* and *aggression* both in the article text and user comments. To quantify *negativity*, we calculate the sentiment score of a piece of text (article or comment) by applying an effective document representation and sentiment analysis model proposed in (Yang et al. 2016). A piece of text with sentiment score closer to 0 shows stronger negativity, while a text with score closer to 1 indicates stronger positivity. For the sentiment of article and comments, we propose features *art_senti_score* and *avg_senti_score*, which are already present in the group of article and comment features, respectively. We use the lexicon LIWC (Tausczik and Pennebaker 2010) to quantify *aggression*. Given a piece of text, *aggression* is defined as the count of entries in the category "Hostile" together with the ones under "anger" that appear in the text. Additional news factors, such as *time of publication*, *uncertainty*, *length*, and *facticity*, are considered in the previous feature groups. They are underlined in the table. Following their definitions (Ziegele, Breiner, and Quiring 2014), *uncertainty* is measured by the count of question marks in a piece of text and *facticity* is a binary feature, which is 1 if the piece of text contains an URL, and 0 otherwise. The factor *position* of comment in the discussion thread is not applicable in our case, since we only analyze the first α comments.

MISC features. The feature *pub_resp* describes how fast users respond to an article, which is similar to some of the features in (Backstrom et al. 2013; Cheng et al. 2014). Some works argue that the longer users delay their response to an article, the less overall user activity the article receives (Backstrom et al. 2013). The features *inter_art* and *inter_com* quantify the ratio of overlap between the sets of named entities in an article and its first α comments. Let NE_{art} and NE_{com_i} be the sets of named entities that appear in an article and its first α comments, respectively. We define *inter_art* and *inter_com* as

$$inter\_art = \frac{|NE_{art} \bigcap NE_{com_i}|}{|NE_{art}|}$$

$$inter\_com = \frac{|NE_{art} \bigcap NE_{com_i}|}{|NE_{com_i}|}.$$

Experimental Setup

The experimental study employs the cross-validation methodology. We split the dataset into five folds randomly. The training set consists of articles in four folds. The articles in the remaining fold are used for testing. For a given set of features, we build a model based on the training set, and apply it on a disjoint testing set. The process is repeated five times, each time selecting a different fold for testing. We report the average performance.

Evaluation Metrics We treat the task of predicting the comment volume of a news article as a regression problem. We evaluate each model in the experiments based on R^2 and the mean absolute error (MAE), which are defined as

$$R^2 = 1 - \frac{MSE}{Variance} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y_i})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y_i}|$$

We calculate the MAE instead of MAPE (mean absolute percentage error) since MAE is more robust to outliers (the long tail in the first graph per outlet in Figure 2). Target variable y_i in the calculation of R^2 and MAE is the logarithm of the number of comments because the distribution of comment volumes resembles lognormal distribution, as shown in Figure 2.

	RF	**SVR**	**NN**	**LR**
ALL	**0.560/0.282**	0.472/0.324	0.499/0.310	0.413/0.338
UC	0.520/0.294	0.479/0.303	0.502/0.301	0.400/0.342
ART	0.078/0.439	0.021/0.452	0.016/0.459	0.020/0.458
`rate`	0.470/0.316	0.465/0.311	0.459/0.323	0.370/0.354

Table 3: Comparison of R^2/MAE results on the overall dataset. ART is the baseline for each algorithm. The row ART gives the outcome of the baselines.

Hyperparameter Tuning and Setting We consider the first α = 10 user comments for each article when we compute the comment features. We reached α = 10 after we studied the variation in prediction accuracy for α = 5, 10, 15, 20, and 50, respectively. The accuracy of prediction increases by about 9% from α = 5 to α = 10, and by less than 2% from α = 10 to α = 50. Therefore, we set α = 10 in all our empirical studies. We explore multiple machine learning (ML) algorithms for performance comparisons on the proposed comment volume prediction task as listed in Figure 1. The results indicate that the feature `rate` does consistently well across the board, thereby indicating it to be a strong algorithm independent feature that inherently captures the prediction task.

We give a brief overview of the hyperparameter setting for the three nonlinear ML algorithms in our methodology: Random Forest (RF), Support Vector Regression (SVR), and Neural Network (NN). We tune the number of trees (`ntrees`) for RF. We use SVR with kernel 'rbf' and tune the hyperparameters `C` and ϵ. We choose Multi-layer Perceptron to implement the NN and tune the hidden layer sizes (`hsize`) and the initial learning rate (`lr`); the activation function for the hidden layer is set to be 'relu'. The choices for these hyperparameters are drawn from: $ntrees \in [50, 100, 200, 300]$, $C \in [0.1, 0.5, 1, 5, 10]$, $\epsilon \in [0.01, 0.05, 0.1, 0.5]$, $hsize \in [10, 20, 30, 50, 100, 200]$, and $lr \in [0.001, 0.005, 0.01, 0.05, 0.1]$.

Experimental Results

We report the performance for the four algorithms (i.e., Random Forest, Support Vector Regression, Neural Network, and Linear Regression) along the four sets of features (i.e., ALL, UC, ART, and `rate`), in the *global* setting in Table 3. We omit the outcome with the *local* setting because of the page limitation. But, the conclusion is very similar to the one drawn in the *global* setting.

User Factors Matter In Table 3, the combined use of all features (ALL) along with Random Forest achieves the best accuracy. We observe that the R^2 value for ART in each testing scenario is near zero, suggesting that the article features alone are not useful signals for predicting the comment volume an article will receive. The prediction models yield much better results when the commenting behavior from early users are taken into consideration (e.g., compare UC row to ART row). *This proves that attempting to predict the*

Model	ALL	UC	$ALL - UC$	*rate*	$ALL - \{rate\}$
WSP	0.533	0.480	0.147	0.471	0.152
DM	0.541	0.498	0.170	0.477	0.193
WSJ	0.737	0.726	0.327	0.651	0.359
FN	0.449	0.408	0.142	0.378	0.134
Gd	0.468	0.428	0.168	0.416	0.170
NYT	0.631	0.612	0.213	0.484	0.280
Overall	0.560	0.520	0.185	0.470	0.209

Table 4: R^2 results for feature ablation and selection for both global and local settings with machine learning algorithm Random Forest. Acronyms: WSP: Washington Post, DM: Daily Mail, WSJ: Wall Street Journal, FN: Fox News, Gd: the Guardian, NYT: New York Times.

eventual volume of user comments on the merits of a news article itself is a futile endeavour. The reason is that a large fraction of the users post comments are triggered by other users' comments instead of the content of the news article itself (Singer 2009; Ruiz et al. 2011). The article features cannot account for the user commenting dynamics. Thus, it is necessary to look into the early user commenting behavior after the publication of a news article to improve our ability to approximate the eventual comment volume of the article.

Since the difference among MAEs across the four feature spaces for a specific algorithm is not as clear as R^2, we use R^2 to discuss additional issues about the prediction performance in the subsequent sections. We use Random Forest in the remaining experiments.

Non-linearity Contrasting the performances of the linear algorithm (Linear Regression) and other nonlinear algorithms (Random Forest, Support Vector Regression, and Neural Network), the R^2 of Linear Regression is consistently worse no matter which feature space (except for ART) is considered. This suggests that Linear Regression alone cannot solve the task of predicting the eventual comment number in a news article. We need to look into more complex models to improve accuracy.

Dominant Feature Discovery

We perform feature ablation by removing one set of features at a time to understand the strengths of the feature families described in Table 2. We report the outcome for both the global and local settings. Table 4 summarizes the outcome of this study. We observe that the decrease in R^2 is no more than 0.055 when we include only the comment features (compare the columns ALL and UC). The performance drops dramatically if we remove the comment features (see column $ALL - UC$). This is another supporting evidence on our account that signals gathered from early user comments largely influence the ability to predict comment volume.

We also study the importance of the individual features by applying the stepwise forward feature selection method (Guyon and Elisseeff 2003). This study shows that `rate` (from the group of comment features) is the most useful predictive variable in the prediction of comment volume. To further understand its importance, we redo the experiments

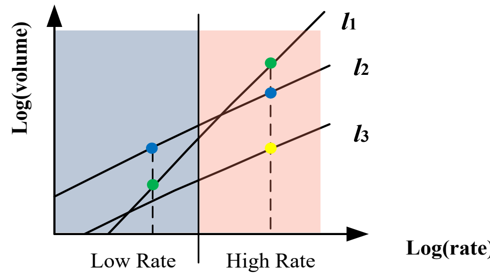


Figure 3: Comparison of regression lines.

with leaving out `rate`. The last column in Table 4 displays the outcome. Compared with the results in the column ALL, there is a dramatic drop in R^2, from 0.560 to 0.209, in the global setting. The decrease ranges between 0.298 - 0.381 across the outlets. This further illustrates the importance of `rate` in the prediction task at hand.

The outcome of feature ablation and selection is consistent with the results in Table 3. We also draw the same conclusion from the other algorithms: `rate` is the dominant single feature in the prediction task.

Rate Analysis

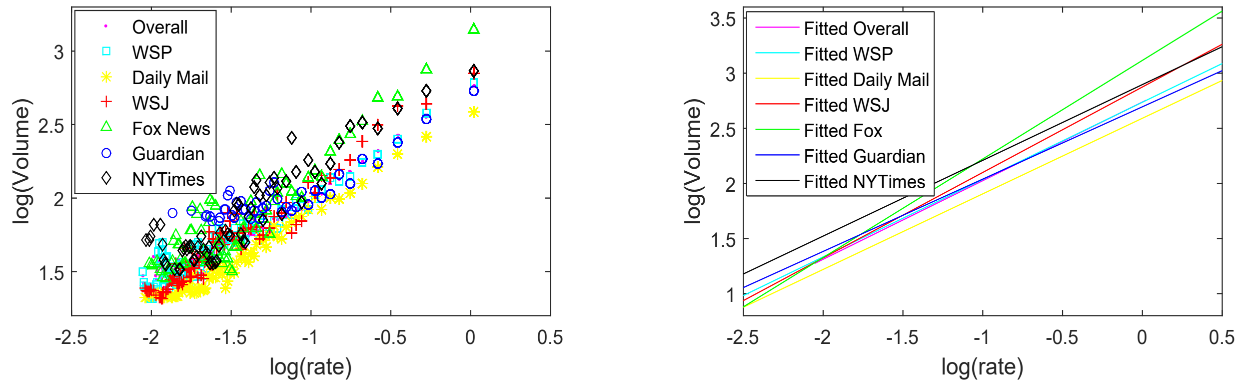
In this section, we focus on the prediction models trained only with `rate`, and investigate their characteristics across news outlets and news categories. We use Random Forests in the experiments reported in this section.

Study of Rate across Outlets

We build rate models for both the global and local settings with Random Forest, considering the observed `rate`'s among the first α = 10 user comments, and use them to predict the eventual comment volume of a news article. We also repeat the studies of rate models built on other values of α (the results are provided in Appendix), and the results are consistent with the 10 comment threshold.

Rate Modeling Figure 4a shows the shape of the prediction models. Points in each dataset are fitted by a linear regression line, as shown in Figure 4b. We use the cartoon example in Figure 3 to describe the chief points we seek to convey in this study about the rate models. There are three regression lines in Figure 3. *l1* is the fitted regression line of points in outlet 1, *l2* is for outlet 2, and *l3* for outlet 3. We distinguish two interesting cases: (1) the regression lines cross each other, as in *l1* and *l2*; and (2) the regression lines are parallel, as in *l2* and *l3*.

Consider the lines *l1* and *l2*. The rate area can be split into two parts: low rate and high rate, based on their intersection. If we carefully compared the points on the two lines, we gather that the user commenting behaviors in outlets 1 and 2 vary across areas. In the low rate area, users in both outlets show less interest at the beginning, reflected by the small values of rate, but the users in outlet 2 keep commenting more than those in outlet 1 as indicated by the larger eventual comment volume. However, in the high rate area, articles in



(a) Rate and Volume in the log scale

(b) Fitted Lines

Figure 4: The plot of the prediction model. If we plot the points based on the value of rate and logarithmic volume, points are too dense around the origin. Therefore, we draw the graph in the log scale for rate and volume.

Model	**Slope**	**Intercept**	**Slope Interval**	**Intercept Interval**	**MoPV**
WSP	0.758	2.740	[0.755, 0.762]	[2.766, 2.768]	2.163
DM	0.703	2.606	[0.701, 0.705]	[2.604, 2.608]	2.131
WSJ	0.841	2.885	[0.832, 0.849]	[2.875, 2.895]	2.111
FN	**0.963**	**3.201**	[0.953, 0.972]	[3.194, 3.209]	**2.577**
Gd	0.656	2.728	[0.649, 0.663]	[2.723, 2.732]	2.396
NYT	0.707	2.935	[0.695, 0.719]	[2.924, 2.945]	2.454
Overall	0.777	2.767	[0.776, 0.778]	[2.766, 2.768]	2.224

Table 5: Statistics of the prediction model trained by *rate*. The values in column Slope (Intercept) Interval are the lower and upper confidence limits for 95% confidence intervals of Slope (Intercept). Column MoPV = Mean of Predicted Volume in the log scale. We reuse the acronyms for news outlets in Table 4.

outlet 1 attract more commenting activity than those in outlet 2, even though the commenting activity early on is the same.

We can draw another useful observation by studying the lines *l1* and *l2*: the same rate fluctuation leads to different variation in comment volume. Since the slope of *l1* is larger than that of *l2*, *l1* will grow faster. Therefore, we conclude that the comment volume in outlet 1 is more sensitive to the rate in the early commenting stream than the comment volume in outlet 2.

Consider the parallel lines *l2* and *l3*. This scenario suggests that the same initial rate leads to different commenting volume in outlets 2 and 3. More precisely, the comment volume of a news article from outlet 2 is larger than that of a news article from outlet 3.

Comparison across Outlets We plot the shape of rate models in Figure 4 and provide the slopes and intercepts of regression lines in Table 5. We provide the lower and upper confidence limits for the 95% confidence intervals of slope and intercept of regression lines for each dataset. The bounds of slopes and intercepts show that their observed differences are due to actual differences between outlets rather than random chance. We also calculate the mean of the predicted logarithm of the comment volume for each dataset (column MoPV in Table 5). We present the results in the light of the discussion in the previous section. The plot shows that the users at Fox News are more active than those at the other news outlets— the logarithm of volume for Fox News is 2.577, which is close to the true value 2.47 in Table 1.

The regression line of Fox News (Figure 4b) behaves as line *l1* in Figure 3. Its high slope suggests that the total comment volume is very sensitive to the `rate` of the initial comments at this outlet. If the rate is small, the discussion dies out quickly. If the rate is large, the readers become very engaged. The comment volume of an article from the Guardian is the least sensitive to the `rate` of early user comments. Comparing the commenting activity at these outlets, the commenting activity at New York Times has the longest attention, because the commenting persists longer and this results in higher volume. We think that the observed behavior might be caused by factors such as the quality or temporal relevance decay of articles, which are difficult to extract from the article content. New York Times behaves as *l2* and Daily Mail as *l3* in Figure 3. The commenting attention at Daily Mail seems to be lower. Interestingly, Wikipedia calls Daily News “middle-market tabloid,” “that attempts to cater to readers who want some entertainment from their newspaper,” so it might indicate a more fleeting character of their articles compared to other outlets.

Article Examples for Rate Models We provide a few concrete examples to illustrate the rate-to-volume behavior across pair of news outlets in this section. We illustrate two scenarios: (i) pairs of news article for outlets whose regression lines cross each other (e.g., Fox News vs the Guardian) and (ii) pairs of news article for outlets whose regression lines are parallel (e.g., Daily Mail vs New York Times) in Table 6. For each of (i) and (ii) we showcase pairs of news articles from both the high and low rate zones.

Crossing regression lines. We have two pairs of news

		Fox News	the Guardian
Pair 1	Article	FN_1	GD_1
	Rate	0.769	0.769
	N_A	2,768	705
Pair 2	Article	FN_2	GD_2
	Rate	0.092	0.092
	N_A	30	117
		Daily Mail	**New York Times**
Pair 3	Article	DM_1	NYT_1
	Rate	0.667	0.667
	N_A	227	407
Pair 4	Article	DM_2	NYT_2
	Rate	0.07	0.07
	N_A	50	108

Table 6: Article examples for outlets with the regression lines crossing each other (Fox News vs the Guardian) and in parallel (Daily Mail vs New York Times).

articles (FN_1, GD_1)[1,2] and (FN_2, GD_2)[3,4] from Fox News and the Guardian, respectively. The rates of (FN_1, GD_1) are both high at 0.769, while the rates of (FN_2, GD_2) are low at 0.092. According to our analysis of the rate-to-volume behavior, since the rates of FN_1 and GD_1 are located in the high rate area and the slope of the regression line for Fox News is larger than that of the Guardian, we expect N_A, the eventual number of comments, of FN_1 to be larger than that of GD_1: FN_1 receives more comments than GD_1, 2,768 versus 705 (in a week). See Pair 1 in Table 6. The effect is reversed for FN_2 and GD_2: FN_2 receives fewer comments than GD_2, 30 versus 117. See Pair 2 in Table 6.

Parallel regression lines. For the parallel case (Daily Mail vs New York Times), we also provide two pairs of news articles: (DM_1, NYT_1)[5,6] and (DM_2, NYT_2)[7,8]. The rates of (DM_1, NYT_1) are both high at 0.667, while the rates of (DM_2, NYT_2) are low at 0.07. Since the regression lines of these two outlets are parallel, we do not expect a reversal as in the previous case. Hence, we expect both the N_A of NYT_1 to be larger than that of DM_1 (407 versus 227, Pair 3 in the table), and the N_A of NYT_2 to be larger than that of DM_2 (108 versus 50, Pair 4 in the table).

Study of Rate across Categories

We now study the performance and characteristics of rate models by news category. We show that rate model behaves quite differently across the major news categories. The present results are from the rate models built on the first 10 user comments, but we have consistent observations from other values of α.

Categorizing Articles To analyze `rate` in different news categories, we need to assign each article to its corresponding categories first. We observe that news outlets assign category labels to their articles. Our initial idea was to make use of these category labels. However, we soon noticed that labels are not consistent across news outlets. For example, the categories such as "U.S. Showbiz" in Daily Mail, "Local" and "National" in Washington Post, "Soccer" in the Guardian, and "Magazine" in New York Times are unique to these outlets. It is difficult to confidently align these categories over time in general, because news outlets periodically reorganize their news categories. We thus resort to the category labels in Google News, which are more stable over a longer period of time. We set the set of labels C = {"Politics", "US", "World", "Sports", "Entertainment", "Technology", "Business", "Science", "Health"} in our analysis. We add "Politics" because it is a category that appears uniformly in all news outlets. We encounter many articles which are not explicitly assigned to any of these categories. We use their topics to determine their categories. We first categorize the topics of an article and then propagate the category labels to the article. We describe the process below.

Categorizing Topics. We follow the method of categorizing topics proposed in (Zhao et al. 2011). For a topic t, its probability of belonging to category $q \in C$ is

$$p(q|t) = \frac{p(q,t)}{p(t)} = \frac{|D_{q,t}|}{|D_t|} \propto |D_{q,t}|,$$

where D_t is the union of articles whose topic is t, $D_{q,t}$ denotes the subset of articles in D_t that are labeled with category q. For a specific topic t, D_t is constant. We select the top three categories for t if it belongs to over three ones. For example, the topic "Donald Trump" is assigned to the categories "US," "World," and "Politics."

Assigning News Categories to News Articles. Our crawler extracts topic information for each news article. Once we determine the news categories C_t of a topic t, we place each article on topic t in each of the categories in C_t.

Comparison across Categories We first partition the overall dataset across the news categories C, and then train a rate model in each category. The analysis follows the steps of the analysis across news outlets. Table 7 gives the article count, performance (R^2), and the characteristics of the rate models per news category.

The first observation is that R^2 differs very little across the news categories, hovering about 0.5. This is in stark contrast to the observations made about R^2 in news outlets. Comparing the linear regression in each category, "Politics" has the

[1] www.foxnews.com/politics/2016/01/12/in-gop-response-haley-pans-obama-presidency-makes-case-for-new-direction.html

[2] theguardian.com/politics/2015/nov/26/labour-whip-email-vote-against-syria-airstrikes

[3] www.foxnews.com/opinion/2015/12/01/in-paris-obama-worships-at-altar-europes-real-religion-climate-change.html

[4] www.theguardian.com/science/2015/oct/28/us-approval-for-drug-that-turns-herpes-virus-against-cancer

[5] www.dailymail.co.uk/news/article-3296561/Syrian-anti-ISIS-activist-blogged-terrible-conditions-Raqqa-decapitated-Turkey-alongside-beheaded-corpse-friend.html

[6] www.nytimes.com/2015/11/12/us/politics/republicans-ted-cruz-marco-rubio.html

[7] www.dailymail.co.uk/sciencetech/article-3311075/Anomalies-thermal-scanning-Egypt-pyramids.html

[8] www.nytimes.com/2016/02/09/sports/basketball/knicks-fire-derek-fisher-as-coach.html

Category	# of Articles	R^2	Slope (Interval)	Intercept (Interval)
Politics	8,491	0.442	0.729 ([0.727, 0.730])	2.818 ([2.816, 2.819])
US	6,202	0.457	0.667 ([0.665, 0.668])	2.719 ([2.717, 2.720])
World	2,548	0.490	0.700 ([0.696, 0.704])	2.735 ([2.732, 2.739])
Sports	1,839	0.484	0.570 ([0.566, 0.575])	2.448 ([2.444, 2.453])
Entmt	1,828	0.535	0.645 ([0.640, 0.650])	2.549 ([2.545, 2.554])
Tech	469	0.528	0.633 ([0.620, 0.647])	2.567 ([2.553, 2.581])
Business	342	0.472	0.664 ([0.649, 0.679])	2.639 ([2.623, 2.654])
Science	235	0.468	0.652 ([0.626, 0.678])	2.626 ([2.599, 2.653])
Health	156	0.493	0.562 ([0.525, 0.600])	2.485 ([2.440, 2.530])

Table 7: Statistics of the datasets and the global models trained by *rate* in category domains. The values in column Slope (Intercept) shows the slope (intercept) of the *rate* model together with its lower and upper confidence limits for 95% confidence intervals. Entmt: Entertainment, Tech: Technology.

highest slope, indicating that the predicted comment volume for articles in this category is more sensitive to the early rate than for articles in other news categories. "Health" has the lowest comment volume and the prediction for its articles is less sensitive to the rate of early comments.

Interplay between Outlets and Categories

Given the different performance of the rate models across outlets and news categories, a natural follow up question is whether there is some mutual effect between outlets and categories over rate. For this study, we summarize the distribution of article sizes among outlets and categories in Table 8. Then, we repeat the prediction analysis for rate models. Table 9 displays the results per outlet and news categories.

Comparing Table 9 (local rate models) to Table 7 (global rate models), we observe that the local models at Washington Post achieve similar R^2 as the global one in "Politics" and "US." For "Sports," local models at Daily Mail perform similarly to the global one in Table 7. This is because either Washington Post or Daily Mail dominates the article size in these categories (see Table 8).

We notice that commenting activity in the "Politics" articles at Fox News is quite distinct from others: by far the largest slope and intercept. In other outlets, "Politics", "US," and "World" are comparable. This indicates that the total comment volume of an article from "Politics" is very sensitive to the `rate` of the initial comments at Fox News.

User commenting preference per category within each outlet reveals additional properties about news outlets (Table 8). In Table 9, we notice that the value of intercept in

	WSP	DM	WSJ	FN	Gd	NYT
Politics	3,374	2,077	1,212	940	408	480
US	2,257	1,548	966	283	655	493
World	635	878	362	188	351	134
Sports	403	1,060	112	62	154	102
Entmt	199	1,242	53	80	215	39
Tech	80	195	88	42	48	16
Business	80	105	89	8	41	19
Science	33	100	33	31	29	9
Health	33	72	18	25	4	4

Table 8: The distribution of article count by outlet and category. We reuse the acronyms for news outlets in Table 4. Entmt: Entertainment, Tech: Technology.

	WSP	DM	WSJ	FN	Gd	NYT
Politics	0.450 0.701 2.772	0.366 0.671 2.675	**0.657** 0.783 2.934	0.352 **0.891** **3.202**	0.396 0.657 2.775	0.400 0.647 2.953
US	0.438 0.652 2.692	0.393 0.652 2.612	**0.628** 0.772 2.921	0.324 0.690 2.932	0.432 0.744 2.764	0.372 0.628 2.925
World	0.490 0.714 2.765	0.361 0.707 2.744	**0.587** 0.706 2.722	0.236 0.663 2.772	0.458 0.634 2.714	0.380 0.640 2.781
Sports	0.389 0.566 2.478	0.489 0.588 2.426	**0.565** 0.694 2.722	-	0.337 0.433 2.481	-

Table 9: The results of the local models trained by *rate* among outlets and categories. The three values in each cell are R^2, then slope, and intercept. Some categories are removed because of insufficient articles.

"Politics" is the largest across all categories and outlets, except for Daily Mail where "World" has the largest intercept. Besides, the highest slope appears three times in "Politics", twice in "World", and once in "US". Considering that most of the "US" and "World" articles are related to politics, it is reasonable to conclude that the comment volume is *sensitive to rate* (suggested by high slope) and *higher* (reflected by large intercept) in political area at most outlets.

Conclusion

In this paper, we study the problem of predicting the total number of user comments a news article will receive. We compile and analyze a large set of features, which we group by topic, article, user comment, news factor, and miscellaneous. Our main insight is that the early dynamics of user comments contribute the most to an accurate prediction, while news article specific factors have surprisingly little influence. Furthermore, we show that the early arrival rate of comments is the best indicator of the eventual number of comments. We conduct an in-depth analysis of this feature across several dimensions, such as news outlets and news article categories. We show that the prediction of comment volume is very sensitive to the early user commenting ac-

	α=5		α=10		α=15		α=20		α=50	
	Slope	Intercept	Slope	Intercept	Slope	Intercept	Slope	Intercept	Slope	Intercept
Washington Post	0.788	2.692	0.758	2.740	0.728	2.741	0.700	2.735	0.574	2.725
Daily Mail	0.702	2.553	0.703	2.606	0.707	2.595	0.684	2.589	0.595	2.575
Wall Street Journal	0.869	2.812	0.841	2.885	0.809	2.886	0.779	2.876	0.665	2.856
Fox News	0.993	3.217	0.963	3.201	0.915	3.149	0.893	3.113	0.736	2.973
the Guardian	0.609	2.722	0.656	2.728	0.657	2.713	0.651	2.693	0.566	2.668
New York Times	0.699	2.933	0.707	2.935	0.701	2.910	0.684	2.895	0.603	2.815
Overall	0.805	2.744	0.777	2.767	0.739	2.749	0.713	2.738	0.594	2.698

Table 10: Slopes and Intercepts of regression lines for rate models based on different values of α.

tivity in some news outlets (e.g., Fox News) and categories (e.g., Politics).

We believe that our findings shed new light on the unique characteristics of readership community compared to other online communities, e.g., those at Twitter or Facebook. This is particularly emphasized by the strong role of early user posting activity on the eventual comment volume for a news article. This has important implications in social media and user behavioral response process understanding, which are key components of the social-news media ecosystem. We believe that this insight is also of value to news analytics, which may lead to better understanding of user participation motives and engagement in commenting news items online.

Acknowledgments

This work was supported in part by the following grants: U.S. NSF Award #1838145 #1838147, and #2026513.

Appendix

A. Rate Models Built on Different Values of α

To investigate the characteristics of rate models across news outlets, we build rate models with the observed `rate`'s among the first α (α = 5, 10, 15, 20, and 50) user comments, for both the global and local settings with Random Forest. The slopes and intercepts of regression lines based on different values of α are provided in Table 10.

According to the results in Table 10, the behavior observed for $\alpha = 10$ is observed for the rest of the values of α. We summarize the key observations below:

- The slope and intercept of the regression line at Fox News are the largest.
- The slope of the regression line at the Guardian is the smallest.
- Regression lines at Daily Mail and New York Times are almost parallel, the intercepts at New York Times are larger than those at Daily Mail.

Therefore, our findings in this paper for $\alpha = 10$ hold for any $\alpha \in [5, 10, 15, 20, 50]$.

References

Aker, A.; Kurtic, E.; Balamurali, A.; Paramita, M.; Barker, E.; Hepple, M.; and Gaizauskas, R. 2016. A graph-based approach to topic clustering for online comments to news. In *ECIR*.

Altheide, D. L.; and Schneider, C. J. 2012. *Qualitative media analysis*, volume 38. Sage.

Aragón, P.; Gómez, V.; García, D.; and Kaltenbrunner, A. 2017. Generative models of online discussion threads: state of the art and research challenges. *Journal of Internet Services and Applications* 8(1): 15.

Artzi, Y.; Pantel, P.; and Gamon, M. 2012. Predicting responses to microblog posts. In *NAACL*, 602–606. ACL.

Backstrom, L.; Kleinberg, J.; Lee, L.; and Danescu-Niculescu-Mizil, C. 2013. Characterizing and curating conversation threads: expansion, focus, volume, re-entry. In *WSDM*, 13–22.

Balali, A.; Asadpour, M.; and Faili, H. 2017. A supervised method to predict the popularity of news articles. *Computación y Sistemas* 21(4): 703–716.

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

Diakopoulos, N.; and Naaman, M. 2011. Topicality, time, and sentiment in online news comments. In *CHI*.

Dos Rieis, J. C. S.; de Souza, F. B.; de Melo, P. O. S. V.; Prates, R. O.; Kwak, H.; and An, J. 2015. Breaking the news: First impressions matter on online news. In *ICWSM*.

Emmer, M.; Vowe, G.; and Wolling, J. 2011. *Bürger online: die Entwicklung der politischen Online-Kommunikation in Deutschland*. UVK Verlagsgesellschaft.

Guyon, I.; and Elisseeff, A. 2003. An introduction to variable and feature selection. *JMLR* 3(Mar): 1157–1182.

Kiousis, S. 2002. Interactivity: a concept explication. *New media & society* 4(3): 355–383.

Kobayashi, R.; and Lambiotte, R. 2016. TiDeH: Time-Dependent Hawkes Process for Predicting Retweet Dynamics. In *ICWSM*.

Liao, D.; Xu, J.; Li, G.; Huang, W.; Liu, W.; and Li, J. 2019. Popularity Prediction on Online Articles with Deep Fusion of Temporal Process and Content Features. In *AAAI*.

Lin, Z.; Huang, F.; Li, Y.; Yang, Z.; and Liu, W. 2019. A layer-wise deep stacking model for social image popularity prediction. *WWW* 22(4): 1639–1655.

Mishne, G.; Glance, N.; et al. 2006. Leave a reply: An analysis of weblog comments. In *Weblogging at WWW*.

Mishra, S. 2019. Bridging Models for Popularity Prediction on Social Media. In *WSDM*, 810–811.

Mishra, S.; Rizoiu, M.-A.; and Xie, L. 2016. Feature driven and point process approaches for popularity prediction. In *CIKM*, 1069–1078.

Olteanu, A.; Castillo, C.; Diakopoulos, N.; and Aberer, K. 2015. Comparing events coverage in online news and social media: The case of climate change. In *ICWSM*, 288–297.

Prochazka, F.; Weber, P.; and Schweiger, W. 2018. Effects of Civility and Reasoning in User Comments on Perceived Journalistic Quality. *Journalism Studies* 19(1): 62–78.

Rizoiu, M.-A.; Mishra, S.; Kong, Q.; Carman, M.; and Xie, L. 2018. SIR-Hawkes: Linking Epidemic Models and Hawkes Processes to Model Diffusions in Finite Populations. In *WWW*, 419–428.

Rizos, G.; Papadopoulos, S.; and Kompatsiaris, Y. 2016. Predicting news popularity by mining online discussions. In *WWW*, 737–742.

Rowe, M.; and Alani, H. 2014. Mining and comparing engagement dynamics across multiple social media platforms. In *WebSci*, 229–238.

Ruiz, C.; Domingo, D.; Micó, J. L.; Díaz-Noci, J.; Meso, K.; and Masip, P. 2011. Public sphere 2.0? The democratic qualities of citizen debates in online newspapers. *The International Journal of Press/Politics* 16(4): 463–487.

Shmueli, E.; Kagian, A.; Koren, Y.; and Lempel, R. 2012. Care to comment?: recommendations for commenting on news stories. In *WWW*, 429–438.

Siersdorfer, S.; Chelaru, S.; Pedro, J. S.; Altingovde, I. S.; and Nejdl, W. 2014. Analyzing and mining comments and comment ratings on the social web. *TWEB* 8(3): 17.

Singer, J. B. 2009. Separate spaces: Discourse about the 2007 Scottish elections on a national newspaper web site. *The International Journal of Press/Politics* 14(4): 477–496.

Tan, C.; Friggeri, A.; and Adamic, L. 2016. Lost in Propagation? Unfolding News Cycles from the Source. In *ICWSM*.

Tatar, A.; Leguay, J.; Antoniadis, P.; Limbourg, A.; de Amorim, M. D.; and Fdida, S. 2011. Predicting the popularity of online articles based on user comments. In *WIMS*.

Tausczik, Y. R.; and Pennebaker, J. W. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *JLS* 29(1): 24–54.

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

Tsagkias, M.; Weerkamp, W.; and De Rijke, M. 2010. News comments: Exploring, modeling, and online prediction. In *ECIR*, 191–203.

Tsur, O.; and Rappoport, A. 2012. What's in a hashtag?: content based prediction of the spread of ideas in microblogging communities. In *WSDM*, 643–652.

Wang, C.; Ye, M.; and Huberman, B. A. 2012. From user comments to on-line conversations. In *SIGKDD*, 244–252.

Weber, P. 2014. Discussions in the comments section: Factors influencing participation and interactivity in online newspapers' reader comments. *New Media & Society* 16(6): 941–957.

Weng, L.; Menczer, F.; and Ahn, Y.-Y. 2014. Predicting Successful Memes Using Network and Community Structure. In *ICWSM*.

Yang, Z.; Yang, D.; Dyer, C.; He, X.; Smola, A.; and Hovy, E. 2016. Hierarchical attention networks for document classification. In *NAACL-HLT*, 1480–1489.

Yano, T.; and Smith, N. A. 2010. What's Worthy of Comment? Content and Comment Volume in Political Blogs. In *ICWSM*.

Zhao, Q.; Erdogdu, M. A.; He, H. Y.; Rajaraman, A.; and Leskovec, J. 2015. Seismic: A self-exciting point process model for predicting tweet popularity. In *KDD*, 1513–1522.

Zhao, W. X.; Jiang, J.; Weng, J.; He, J.; Lim, E.-P.; Yan, H.; and Li, X. 2011. Comparing twitter and traditional media using topic models. In *ECIR*. Springer.

Ziegele, M.; Breiner, T.; and Quiring, O. 2014. What creates interactivity in online news discussions? An exploratory analysis of discussion factors in user comments on news items. *Journal of Communication* 64(6): 1111–1138.

Ziegele, M.; and Quiring, O. 2013. Conceptualizing online discussion value: A multidimensional framework for analyzing user comments on mass-media websites. *Annals of the International Communication Association* 37(1): 125–153.

Ziegele, M.; Weber, M.; Quiring, O.; and Breiner, T. 2018. The dynamics of online news discussions: effects of news articles and reader comments on users' involvement, willingness to participate, and the civility of their contributions. *Information, Communication & Society* 21(10): 1419–1435.