Analyzing and Predicting News Popularity in an Instant Messaging Service

Mohammad Naseri
Shahid Beheshti University
Tehran, Iran
mohammadnaseri@acm.org

Hamed Zamani
University of Massachusetts Amherst
Amherst, MA 01003
zamani@cs.umass.edu

ABSTRACT

With widespread use of mobile devices, instant messaging (IM) services have recently attracted a great deal of attention by millions of users. This has motivated news agencies to share their contents via such platforms in addition to their websites and popular social media. As a result, thousands of users nowadays follow the news agencies through their verified channels in IM services. However, user interactions with such platforms is relatively unstudied. In this paper, we provide an initial study to analyze and predict news popularity in an instant messaging service. To this aim, we focus on *Telegram*, a popular IM service with 200 million monthly active users. We explore the differences between news popularity analysis in Telegram and typical social media, such as Twitter, and highlight its unique characteristics. We perform our analysis on the data we collected from four diverse news agencies. Following our analysis, we study the task of news popularity prediction in Telegram and show that the performance of the prediction models can be substantially improved by learning from the data of multiple news agencies using multi-task learning. To foster research in this area, we have made the collected data publicly available.

ACM Reference Format:
Mohammad Naseri and Hamed Zamani. 2019. Analyzing and Predicting News Popularity in an Instant Messaging Service. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '19), July 21–25, 2019, Paris, France.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3331184.3331301

1 INTRODUCTION

Recent years have witnessed a rapid growth in the use of mobile devices, enabling people to access various contexts via the Internet. Widespread use of smartphones and emergence of instant messaging (IM) services, such as Telegram, WeChat, and WhatsApp, have revolutionized online communication for millions of users, which has recently motivated researchers to study various aspects of these services, e.g. see [4].

In addition to one-to-one communication (i.e., private messaging), instant messaging services often provide one-to-many and many-to-many communications. This provides an appropriate basis for news agencies to share their contents. Many popular news agencies, such as CNN, Reuters World, and BBC, are currently using such platforms. In this paper, we focus on *Telegram*,[1] which has been one of the most growing and popular IM services in recent years with 200 million monthly active users.[2]

On the other hand, predicting news popularity (sometimes referred to as user engagement) is a vital task for news agencies; because they always prefer to attract more attention by their news posts [5, 6, 9]. Thus, understanding the reasons behind news popularity and predicting popular news beforehand are of significance for journalists and news agencies. Popularity has also played a key role in news ranking and recommendation, e.g., see [8].

Due to the importance of news popularity prediction, there is a line of research on this topic. Previous work predicted news popularity based on user interactions with the news websites [9] or the social media posts [1]. However, we argue that the conclusions made by previous work do not necessarily hold for IM services and their models cannot be directly applied to such platforms. Some evidence for differentiating news popularity in Telegram from typical social media services, e.g., Twitter, are listed below:

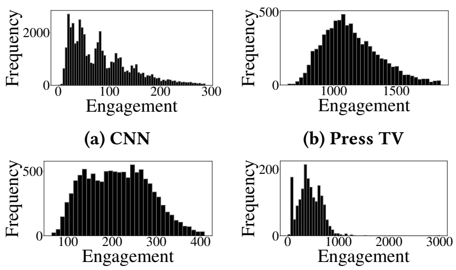
- As reported in Table 1, the *Obama-era* news was among the most popular tweets of CNN, while was less popular in the CNN's Telegram channel. The *Baby Charlie Grad* news was among the most popular CNN posts in Telegram, however, it did not draw attention in the CNN's Twitter account.
- The news popularity distributions for four news agencies in Telegram and Twitter for the same time period are depicted in

Table 1: Examples from CNN posts in Twitter and Telegram.

S.N.	Message	Popularity
Twitter	The EPA plans to propose the repeal of an Obama-era rule on power plants meant to curb greenhouse gas emissions http://cnn.it/2y9pa84.	top 5%
Telegram	Trump administration propose repealing Obama-era rule on greenhouse gas emissions.The Environmental Protection Agency plans to propose the repeal of a sweeping Obama-era rule on power plants meant to curb greenhouse gas emissions, according to a leaked proposal obtained by CNN.	top 25%
Twitter	Parents of baby Charlie Gard return to Britain's High Court to continue their fight to keep him on life support. http://cnn.it/2uhoE63.	bottom 10%
Telegram	Baby Charlie Gard's parents return to court to present new evidence. The parents of baby Charlie Gard have returned to Britain's High Court Thursday as they continue their fight to keep him on life support so they can take him to the United States for experimental treatment for a rare genetic disorder.	top 5%

[1] https://telegram.org
[2] https://telegram.org/blog/200-million

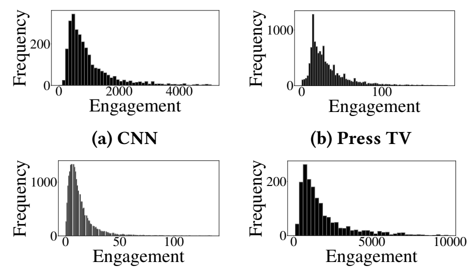
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
SIGIR '19, July 21–25, 2019, Paris, France
© 2019 Association for Computing Machinery.
ACM ISBN 978-1-4503-6172-9/19/07…$15.00
https://doi.org/10.1145/3331184.3331301



(a) CNN (b) Press TV

(c) Reuters World (d) BBC

Figure 1: Distribution of news popularity (i.e., total number of views) in Telegram for different news agencies.



(a) CNN (b) Press TV

(c) Reuters World (d) BBC

Figure 2: Distribution of news popularity (i.e., sum of favorite and retweets) in Twitter for different news agencies.

Figures 1 and 2. The plots demonstrate the significant differences between the ways users interact with these two platforms.

- In typical social media services, such as Twitter and Facebook, each user has a feed page to monitor all the activities in his/her network. However, in IM services, users can only read messages in individual groups or channels. This significantly affects the visibility of the content.
- The social network of users at each timestamp is well defined in typical social media platforms, which provides important signals for news popularity detection. However, in IM services, the relations are defined based on group and channel memberships and most importantly the one-to-one user communications, which are not known to popularity detection algorithms.

These reasons highlight the importance of studying the task of analysis and predicting news popularity in IM services. Therefore, in this paper, we study news popularity for a diverse set of news agencies: CNN (a US based agency), Press TV (a Middle East based agency), Reuters World (an international agency), and BBC (a UK based agency). These channels are among the most popular channels in Telegram. We further investigate the feasibility of predicting popular news in the mentioned Telegram channels. To measure the news popularity, we focus on the number of unique users that *viewed* a message[3] in Telegram. Note that by forwarding a message in the Telegram network, the views are accumulated for the original message. Therefore, higher views indicates higher engagements and thus more popular contents. Our major contributions include:

- Performing the first study on news popularity detection in IM services.
- Creating a collection of news posts in Telegram for multiple agencies. Our data is publicly available for research purposes.[4]
- Proposing a model to utilize information from multiple news agencies via multi-task learning. Our experiments show that news popularity prediction models can substantially benefit from transferring knowledge across agencies.

2 BACKGROUND

In this section, we briefly review previous work on news popularity prediction and further introduce Telegram and its features.

News Popularity. Predicting news popularity is important for journalists and editors so that they can efficiently allocate resources to support a better reading experience [5]. News popularity can be defined in various ways. For example, Tsagkias et al. [9] took the number of comments for a news post as a measurement for news popularity. In another study, Tsagkias et al. [10] explored news comment space and used comparison of its log-normal and binomial distribution models to predict comment volume as a measure of popularity. Lerman and Hogg [6] predicted the number of votes to each Digg post to measure popularity. This task is also related to user engagement prediction. For example, Zamani et al. [12] ranked tweets about movie ratings in terms of their engagements, i.e., total number of retweets and favorites. Previous work studied the popularity of posts in news websites or social media [7, 11]. However, in this work, we focus on instant messaging services. The unique characteristics of these services are enumerated in Section 1.

[3]Throughout this paper, the terms "message" and "news post" as well as "channel" and "news agency" are often used interchangeably.
[4]http://bit.ly/TelegramNewsData

Table 2: Statistics of the Telegram data.

Stats/Channel	CNN	Press TV	Reuters	BBC
# News Items (NI)	48458	7864	11306	1842
# Subscribers	9626	7209	3696	2519
Average View	89.83	1182.80	221.19	679.36
STD View	97.31	337.69	94.16	1885.33
# Hashtags / NI	0.001	0.0967	0	0.204
# Mentions / NI	0.470	0.316	0.403	0.406
# Media / NI	0.530	0.920	0.163	0.921
Distribution	*half-logistic*	*alpha*	*rice*	*burr*

Introducing Telegram. In this paper, we focus on Telegram, a popular instant messaging service for mobile devices and desktop computers. Users in Telegram can send text messages, media, files, and stickers. Telegram has three types of communication protocols: (1) one-to-one, (2) one-to-many, and (3) many-to-many.

Channels in Telegram, which are the focus of this work, are similar to public broadcast media and have been used for sharing textual content, advertisements, etc.

Each message in Telegram has an attribute "views", which demonstrates the number of unique users that has viewed the message. Note that forwarding a news post to other channels, groups, or individuals results in increasing the number of views in the original post. In other words, Telegram accumulates the number of views for the original message.

3 DATA COLLECTION AND ANALYSIS

We collected all the posts published by CNN, Press TV, Reuters World, and BBC through the API provided by Telegram. We selected these channels to study a diverse set of news agencies. In addition, they are among the most popular news channels in Telegram. They are all verified channels and publish their contents in English. In our study, we consider the news published in a 7 month period, from March 8, 2017 to October 8, 2017.

Table 2 reports the statistics of the data for different news channels. Surprisingly, the average views per news post are not correlated with the number of subscribers. These differences might be related to the type of news that is covered by each agency and the cultural differences between their audiences.

Hashtags and mentions are non-separable part of a message in IM services. As also explored by the studies on social media [1], hashtags are effective features for a wide variety of text mining tasks. In case of news popularity, hashtags can show the main purpose of a news item. In addition to hashtags, one can mention a Telegram user, group, or channel in a message. Table 2 reports the average number of hashtags and mentions for each channel.

In Telegram, media, e.g., video, image, and hyperlink, can be linked to a post. They can be useful in engaging the audiences. A post hyperlink is likely the link to the full news article. This information is also captured and can be potentially used in the future for studying different tasks, e.g., news summarization.

The views distributions for all channels are plotted in Figure 1. According to the plots, the views distribution significantly varies across news agencies. To estimate news popularity distribution, we perform a distribution hypothesis testing and examined 89 different distributions implemented in Scikit-learn.[5] We then chose the best fit for each, which is mentioned in the last row of Table 2. This suggests that news popularity identification varies across news agencies and one cannot generalize the findings on one agency to the others. Based on this observation, we claim that it is unlikely to train an effective news popularity detection model on multiple agencies by pooling all the data together, which is validated by our experiments. This has motivated us to propose a multi-task learning approach for this task.

4 NEWS POPULARITY PREDICTION

In this section, we focus on predicting the popularity of news posts for each news agency in Telegram. In more detail, we aim at predicting the top 5% and 25% popular posts for each channel. Therefore, this can be viewed as an imbalanced binary classification task (especially for the first setting).

In this paper, we are particularly interested in improving the performance of news popularity prediction by incorporating the data from multiple news agencies. Although our data analysis in Section 3 demonstrates significant differences across agencies, we hypothesize that there exists useful information that can be transferred between them in order to improve the news popularity prediction performance. From the different views distributions across news agencies (see Figure 1), it can be inferred that pooling the data from all agencies (i.e., putting all the data together) to train a prediction model is unlikely to be effective. The reason is that it destroys the training data distribution. Our experiments also verify this claim. Therefore, we propose to employ multi-task learning for identifying popular news items, where each news agency is a "task". The goal is to learn generalized individual model per news agency by transferring knowledge between them.

In the following subsections, we first introduce the multi-task learning approach used in our experiments, and further list the features extracted from each Telegram's news post. We then review our experimental setup and finish with reporting and discussing our results.

[5]https://docs.scipy.org/doc/scipy/reference/stats.html

Table 3: Channel- (top) and post-based (bottom) features.

Feature	Description
# subscribers	shows the popularity of channel
Avg # posts per hour	shows how active the channel is
Avg # posts per day	shows how active the channel is
channel age	channel's age in days
μ/σ/max/min of views	average, std. dev., max, and min of views
age	age of news item in seconds
Day of Week	what day of the week the post was published
Hour of day	exact hour (from 0 to 23) of publish time
Day/month/year	day, month, and year of publish time
Frequent n-grams	frequent n-grams that news item contains
Frequent hashtags	frequent hashtags that news item contains
# hashtags	how many hashtags that news item have
Frequent mentions	frequent mentions that news item contains
# mentions	how many hashtags that news item have
media and type	contains media or not? what type?
Has link	news item contains link or not

4.1 Multi-Task Learning

Multi-task learning (MTL) [2] is an inductive transfer learning technique whose goal is to improve the generalization performance by leveraging domain-specific signals from related tasks.

In this paper, we consider MTL CASO [3] as an effective multi-task learning model that uses a convex relaxed alternative structured optimization (CASO) that decomposes the model of each task into two components: task-specific and task-shared feature mappings. It has been shown that there is an equivalence relationship between CASO and clustered MTL (CMTL) which assumes that tasks have group structure in which all tasks in each cluster are related to each other. We use a linear model that optimizes a logistic loss function. Since the distribution of data in our problem is highly skewed, as suggested by Zamani et al. [12], we assign higher weights to the instances from the minority class. To do so, for each task i, an instance weighting matrix $\Lambda_i \in \mathbb{R}^{n_i \times 1}$ is defined, whose elements correspond to the weights of the training instances for the i^{th} task. The elements of Λ is computed as $\Lambda_{ij} = (1/n_i^{(j)})/(\sum_{k=1}^{n_i} 1/n_i^{(k)})$, where $n_i^{(j)}$ denotes the number of training instances in task i with the same label as the j^{th} instance. Therefore, the mentioned logistic loss function is modified by weighting each training instance using the above weighting scheme.

4.2 Features

Our features are listed in Table 3. The first part of the table includes **channel-based features** that only describe the channel's characteristics.[6] The second part of the table lists the **post-based features** that are those related to the particular news post.

4.3 Experimental Setup and Baselines

To have a fair and realistic evaluation, we split the data into the training and test sets in a chronological order, i.e., the first six months of the data is selected for training and the last month is selected for testing. As mentioned earlier in this section, we aim at predicting the top 5% and 25% popular news for each agency.

[6]These are useful for the models that are trained by the data from different channels.

Table 4: Performance of single-task and multi-task learning models for different channels.

Channel	Metric	Top 5%			Top 25%		
		STL	STL-All	MTL	STL	STL-All	MTL
CNN	Acc.	0.9079	0.9399	**0.9905**	0.3146	0.7285	**0.9399**
	BA	0.5082	0.5041	**0.5314**	0.5018	0.5719	**0.6801**
	Prec.	0.0625	0.0937	**0.3095**	0.2971	0.3384	**0.5517**
	Recall	0.0688	0.0176	**0.0724**	**0.9621**	0.4156	0.5573
	F1	0.0655	0.0297	**0.1173**	0.4540	0.3731	**0.5544**
Press TV	Acc	**0.9099**	0.2020	0.4538	0.5656	0.2516	**0.5791**
	BA	0.5417	0.5041	**0.6157**	**0.6039**	0.4983	0.5883
	Prec.	0.0200	0.0550	**0.0842**	0.2024	0.2516	**0.3775**
	Recall	0.1818	**0.9016**	0.3684	0.6628	**1.0000**	0.6117
	F1	0.0360	0.0347	**0.1370**	0.3101	0.4021	**0.4669**
Reuters	Acc.	0.5047	**0.9376**	0.1519	0.6619	0.2627	**0.8783**
	BA	**0.7580**	0.5041	0.5967	0.5117	0.5008	**0.8641**
	Prec.	0.2316	0.0800	**0.3030**	0.6715	0.2518	**0.8023**
	Recall	**0.2210**	0.0222	0.2072	0.9637	**0.9799**	0.8244
	F1	0.2262	0.0347	**0.2461**	0.7914	0.4007	**0.8132**
BBC	Acc.	0.8061	0.9118	**0.9622**	0.5814	0.2511	**0.8545**
	BA	0.4617	0.4813	**0.5811**	0.5166	0.4913	**0.8056**
	Prec.	0.0357	0.0000	**0.2727**	0.6729	0.2511	**0.7678**
	Recall	0.0555	0.0000	**0.2000**	0.7133	**1.0000**	0.6935
	F1	0.0434	0.0000	**0.2307**	0.6925	0.4014	**0.7288**

Therefore, we first found the thresholds that satisfy these percentages for training and test sets and assigned a binary label to each post ("popular" or "not popular"). We linearly normalize all features based on their maximum and minimum values in the training set.

In addition to the mentioned linear multi-task learning model (called MTL), we also consider two single-task learning models in our experiments. (1) An SVM classifier[7] trained in an in-domain fashion, i.e., for predicting news popularity for each news agency, the model is trained on the training data from the same agency. This method is called STL. (2) An SVM classifier trained on all the data from all the channels, called STL-All. Both of these models also use a sample weighting technique, exactly the same as the one used for MTL. The hyper-parameter c of the SVM classifiers was selected using 5-fold cross-validation over the training data. We also performed feature selection using randomized Lasso, however no substantial improvement has been observed. We use scikit-learn[8] for the single-task and MALSAR[9] for the multi-task models.

Evaluation Metrics. We use five standard metrics to evaluate our models for the news popularity prediction task: accuracy, balanced accuracy (i.e., the average of accuracy for popular and non-popular classes), precision, recall, and F1-measure. Precision, recall, and F1-measure are calculated for the "popular" class. Note that since we are tackling an imbalanced classification problem, balanced accuracy and F1-measure are the proper metrics for such scenarios [12].

4.4 Results and Discussion

Table 4 reports the results for predicting the top 5% and 25% popular news posts for the aforementioned learning models. The results show that predicting the top 5% news items is harder than predicting the top 25% most popular news posts. The reason is that the lower the percentage, the more imbalanced the data, and thus the more likely the model to bias towards the majority class.

The MTL method achieves the highest performance in most cases, specially in terms of balanced accuracy (BA) and F1-measure which are the proper metrics for evaluating imbalanced classification problems. STL-All often performs poorly, due to the fact that the views distribution varies across channels, and thus putting all the training data together destroys the training data distribution for all the channels. The relative improvements achieved by MTL are higher, when the threshold is set to 5% which is the hardest setting. The reason is that the training data from each channel only contains a few positive training instances and thus multi-task learning helps the model observe more positive instances from different channels.

Although both balanced accuracy (BA) and F1-measure are common metrics for imbalanced classification tasks, they do not agree in some cases. For example, by increasing the threshold from 5% to 25%, the balanced accuracy decreases for the Reuters channel, however, the F1-measure increases. This shows that in these cases the model cannot detect the popular news when the threshold is too low (i.e., 5%), however, it can detect the majority of non-popular news. This often happens in imbalanced settings.

5 CONCLUSIONS

This paper provided an initial study on news popularity analysis and prediction in Telegram, a popular instant messaging service with millions of users. We studied a wide range of news agencies: CNN, Press TV, Reuters World, and BBC. Our analysis investigated the differences between news popularity distribution across the agencies, and suggested that although most previous work on predicting news popularity in news websites and social media only focused on a single news agency, one cannot generalize the findings to all agencies. We further extracted several features in order to predict the top 5% and 25% popular news posts for each agency. The experiments showed that the performance of news popularity prediction can be improved by utilizing the data from multiple agencies via multi-task learning.

Acknowledgements. This work was supported in part by the Center for Intelligent Information Retrieval. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the sponsor.

REFERENCES

[1] R. Bandari, S. Asur, and B. A. Huberman. 2012. The Pulse of News in Social Media: Forecasting Popularity. In *ICWSM*. 26–33.
[2] R. Caruana. 1997. Multitask Learning. *Machine Learning* 28, 1 (1997), 41–75.
[3] J. Chen, L. Tang, J. Liu, and J. Ye. 2009. A Convex Formulation for Learning Shared Structures from Multiple Tasks. In *ICML*. 137–144.
[4] A. Dargahi Nobari, N. Reshadatmand, and M. Neshati. 2017. Analysis of Telegram, An Instant Messaging Service. In *CIKM*. 2035–2038.
[5] Y. Keneshloo, S. Wang, E.-H. Han, and N. Ramakrishnan. 2016. Predicting the popularity of news articles. In *SDM*. 441–449.
[6] K. Lerman and T. Hogg. 2010. Using a Model of Social Dynamics to Predict Popularity of News. In *WWW*.
[7] S. Phuvipadawat and T. Murata. 2010. Breaking News Detection and Tracking in Twitter. In *WI-IAT*, Vol. 3. 120–123.
[8] A. Tatar, P. Antoniadis, M. D. de Amorim, and S. Fdida. 2014. From popularity prediction to ranking online news. *SNAM* 4, 1 (2014), 174.
[9] M. Tsagkias, W. Weerkamp, and M. de Rijke. 2009. Predicting the Volume of Comments on Online News Stories. In *CIKM*. 1765–1768.
[10] M. Tsagkias, W. Weerkamp, and M. de Rijke. 2010. News Comments: Exploring, Modeling, and Online Prediction. In *ECIR*. 191–203.
[11] B. Wu and H. Shen. 2015. Analyzing and predicting news popularity on Twitter. In *IJIM*, Vol. 35. 702–711. Issue 6.
[12] H. Zamani, P. Moradi, and A. Shakery. 2015. Adaptive User Engagement Evaluation via Multi-task Learning. In *SIGIR*. 1011–1014.

[7] We tried different classifiers, including neural nets however, they underperform SVM.
[8] https://scikit-learn.org
[9] https://github.com/jiayuzhou/MALSAR