Make it go viral - Generating attractive headlines for distributing news articles on social media

ABSTRACT

A huge number of news articles are distributed on social media, and news consumers can access those articles at any time from hand-held devices. This means that news providers are under pressure to find ways of attracting the interest of news consumers. One key factor that has a great impact on the attractiveness of a news article is its headline as a way of guiding news consumers to news articles. This research explores the challenge of automatically generating attractive news headlines for social media, and to this end we focus on the problem of identifying key sentences that are useful for generating viral news headlines from a given news article. We show that this problem can be formulated as supervised sequence labeling that utilizes user activity on social media as supervised information, and we propose a neural network model for this purpose. Investigations with our corpus consisting of microblog posts and news articles demonstrate that lead sentences believed to be the most suitable for news summaries do not necessarily contribute to increased virality whereas our proposed method can accurately identify key sentences.

CCS Concepts

•Information systems → Social networks; •Computing methodologies → Information extraction; Neural networks;

Keywords

News headline generation; social media; recurrent neural network

1. INTRODUCTION

Social media are rapidly gaining in significance for news consumers. 62% of U.S. adults obtain news on social media, and 18% do so often, according to a survey by the Pew Research Center [1]. Reuters Institute Digital News Report [2] states that almost 50% of the population in the world use social media as a source of news each week and 12% say it is their main source. Social media also play an important role in communicating with news publishers and journalists [3].

News publishers and journalists also enjoy significant benefit from the growth of social media. It is becoming easier for journalists to find stimulating topics from social media and to create interesting articles by adding their own insights to the original topics [1]. Some journalists are highly motivated to engage with news consumers by leading social chat, responding to comments left on news articles posted to social media and using social media more broadly to develop relationships and guide people to news articles.

However, the recent growth of social media has also had a negative effect. Any type of news article published by any news provider has the potential to gain popularity on social media, and thus hundreds of thousands of news articles are now distributed on social media every day. Time limitations mean that consumers may select only a few of these news articles, and so news providers are under pressure to find ways of engaging consumer attention. This has also encouraged the commercial growth of new types of social platforms called viral media. These platforms employ a variety of strategies to attract consumers towards their own news content. Upworthy, a major viral medium in the U.S., presented several ways of making content go viral [10], including the use of web designs, impressive images, and interface optimization for target social platforms. Another viral medium in the U.S., Buzzfeed, introduces A/B testing to find promising news stories and to improve the user interface [4].

Of these approaches, headlines are the most significant factor influencing our impressions of news articles and they act as a gateway to the news articles. Upworthy recommends that we should write 25 headlines and test the best ones. The Japanese major news paper, Asahi Shinbun, designs news headlines for social media that conceal the most impotant information to entice readers to access the corresponding news articles [5]. However, the process of generating and identifying attractive news headlines relies heavily on "seat of the pants" work by highly skilled editors and

[1] http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/

[2] https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Digital-News-Report-2016.pdf

[3] http://journalistsresource.org/studies/society/news-media/audience-perception-journalists-use-social-media

[4] http://www.wired.co.uk/magazine/archive/2014/02/features/buzzfeed

[5] http://web-tan.forum.impressrd.jp/e/2015/07/14/20270

C+J '16 September 30 – October 1, 2016, Stanford, California, USA

DOI: 10.475/123_4

curators.

This paper tries to provide a novel solution to this problem by describing a way of automatically generating attractive news headlines that have the potential to go viral on social media. To this end, we focus on the problem of extracting key sentences from a given news article as material for generating attractive news headlines.

Our contributions can be summarized as follows: (1) We first reveal through manual investigations that lead sentences that are believed to be the most suitable for summarizing news articles might be insufficient as a resource for generating attractive news headlines. (2) On the basis of these investigations, we propose a novel method for identifying key sentences from a given news article. In contrast to standard extractive document summarization [7, 21, 27], we employ user activity on social media as supervised information, which we then use to formulate the problem as supervised sequence labeling. (3) We propose a recurrent neural network model for this supervised sequence labeling. This model is based on the bidirectional long-short term memory model [12, 13] which has achieved excellent performance as regards speech recognition.

2. RELATED WORK

Several studies have tried to understand, quantify and predict content virality. Park et al. [23] revealed that bad news spreads faster than other types of information such as apologies and sparks a great degree of negative sentiments in social networks. Gabielkov et al [9] undertook a large-scale study of social clicks to understand the effects of social media conversations mentioning an online article on actual clicks of the article. Resnick et al. [24] built a system that helps journalists identify rumors and their corrections. Suh et al. [26], Guerini [16] and Diakopoulos et al. [5] investigated the factors that contribute to the virality of Twitter messages from sociological standpoints. Jenders et al. [18] analyzed a wide range of tweet and user features with respect to their influence on virality, and built a model for predicting viral tweets. Guerini et al. [15] and Deza et al. [4] tried to understand and estimate the virality of image content. Kourogi et al. [20] attempted to reveal the source of virality from news headlines and the relationships between headlines and news articles. However, the above studies all focused solely on understanding content virality, and the generation of content with high virality has yet to be explored.

Headline generation for news articles has been well explored by the natural language processing community [30, 2]. Zajic et al. [32] proposed an HMM-based method for selecting headline words from the beginning of a given news article. Dorr et al. [6] and Wang et al. [29] focused on the use of lead paragraphs that are believed to be the most appropriate as a summary of news articles, and proposed methods for shortening a lead paragraph to generate a headline. However, as we describe in the following sections, information included in the lead paragraph might be insufficient to produce attractive news headlines for social media. Our research is unique in that it learns models for extracting key sentences from the whole news article taking user activity on social media into consideration.

Our approach to headline generation is similar to extractive document summarization, where news articles and corresponding headlines can be regarded as "documents" and "summaries", respectively. Extractive document summarization generally has two components: key sentence extraction, which is the main focus of this paper and sentence compression. State-of-the-art extractive document summarization methods formulate this task as a combinatorial optimization problem, for example, a knapsack problem [21], a maximum coverage problem [7, 27] or a budgeted median problem [28]. The rhetorical and discourse structures of documents can be incorporated into the summarization procedure by formulating the problem as integer linear programming [17, 31]. However, the aim of (extractive) document summarization is to retain the most important points contained in the original document, which might not be appropriate for generating viral news headlines.

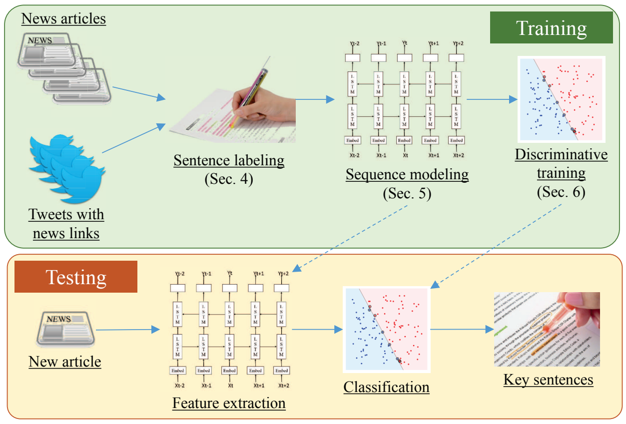

Figure 1: Framework of proposed method

3. FRAMEWORK

Figure 1 shows the framework of our proposed method. The method requires ground-truth data that contain news articles and sentence-wise labels representing the sentences that should be extracted as key sentences from a given news article. For this purpose, we make use of (1) tweets containing links to news articles published by a major news medium, (2) linked news articles and (3) the virality scores of tweets that can be computed from user activity on Twitter. The details for the dataset will be presented in Section 4. We used the obtained ground-truth data sets to build a neural network model for supervised sequence labeling, where the input is a news article and the target is its sentence-wise labels. The details for the labeling will be presented in Section 5. Although the proposed model itself can estimate key sentences from a given news article, we add a support vector machine (SVM) classifier as a post-processing technique to improve the performance. The details for the post-processing will be described in Section 6.

4. DATA ANALYSIS

To build the ground-truth data sets, we crawled Japanese Twitter posts through Twitter Firehose API [6] for a year (August 2014 - September 2015). From the tweets, we extracted those that (a) contained a link to an article in Mainichi Shin-

[6]We obtained almost half of all the Japanese tweets.

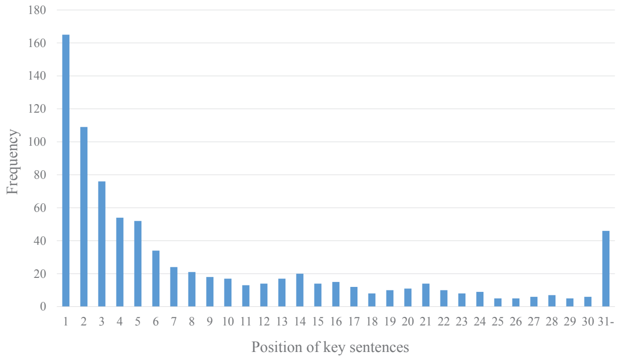

Figure 2: Statistics on positions of key sentences in news articles

bun [7], (b) were retweeted or favorited by someone, and (c) found other tweets with a link to the same article, resulting in a total of 800,000 tweets. Concurrently, we retrieved a total of 6,000 Mainichi Shinbun articles contained in those tweets. For every article, we selected the most viral tweet that contained a link to the article. We here adopt the sum of the retweet and favorite counts as a measure of virality. After removing the tweet-article pairs with low virality scores, we finally obtained about 300 pairs.

Next, for every sentence in every article we performed a manual annotation indicating whether or not the sentence is a key sentence for generating attractive news headlines. We gave positive labels to the sentences that included the same semantic meaning in part of the paired tweet, and negative labels for all the others.

Figure 2 shows the result of the annotations, where the horizontal axis is the position of key sentences and the vertical axis is its frequency. The first and second sentences were selected as key sentences for 55% (165 articles) and 36% (109 articles) of the articles, respectively. This follows a tradition of news styles whereby the lead paragraph, namely the opening paragraph of a news story, is optimized to articulate the most all-encompassing and interesting statements. We also found that more than two sentences were selected as key sentences for most (81%) of the articles. Meanwhile, the later parts of articles were often annotated as key sentences (34%), which implies that sentences other than those in the lead paragraphs should be included to improve virality.

5. NEURAL NETWORK MODEL

We built a recurrent neural network model to identify key sentences from a given news article, and trained it with the data set constructed in Section 4. Figure 3 shows our proposed model, which we built based on the bi-directional long-short term memory (BLSTM) model [14, 12, 13] that has recently been shown to provide state-of-the-art performance as an acoustic model for speech recognition. We take previous and future contexts into considerations by introducing bi-directional structures into a model to achieve key sentence extraction in a holistic manner. We also exploit the long-term dependencies introduced by LSTMs, since sentences from the later parts as well as from the lead paragraphs of a given article are frequently selected as key sentences, as shown in the previous section.

[7]http://mainich.jp/english

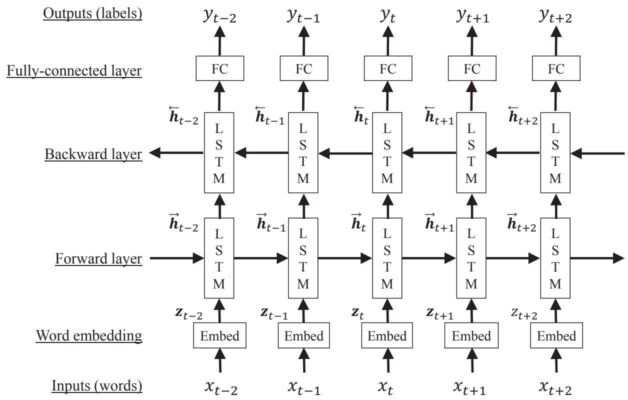

Figure 3: Proposed neural network model

Given an input sequence $\mathbf{X} = (x_1, x_2, \ldots, x_T)$ of words, our proposed model computes an output sequence $\mathbf{Y} = (y_1, y_2, \ldots, y_T)$ of labels as follows:

$$\begin{aligned}
\mathbf{z}_t &= \mathcal{E}(x_t), \\
\overrightarrow{\mathbf{h}}_t &= \mathcal{L}(\mathbf{W}_{z\overrightarrow{h}}\mathbf{z}_t + \mathbf{W}_{\overrightarrow{h}\overrightarrow{h}}\overrightarrow{\mathbf{h}}_{t-1} + \mathbf{b}_{\overrightarrow{h}}), \\
\overleftarrow{\mathbf{h}}_t &= \mathcal{L}(\mathbf{W}_{\overrightarrow{h}\overleftarrow{h}}\overrightarrow{\mathbf{h}}_t + \mathbf{W}_{\overleftarrow{h}\overleftarrow{h}}\overleftarrow{\mathbf{h}}_{t+1} + \mathbf{b}_{\overleftarrow{h}}), \\
\hat{y}_t &= \sigma(\mathbf{W}_{\overleftarrow{h}y}\overleftarrow{\mathbf{h}}_t + \mathbf{b}_y),
\end{aligned}$$

where the $\mathbf{W}$ terms denote weight matrices, the $\mathbf{b}$ terms denote bias vectors, $\mathcal{E}(\cdot)$ is a word embedding, $\mathcal{L}(\cdot)$ is a single LSTM cell, and $\sigma(\cdot)$ is an activation function such as sigmoid. Since our model requires word-wise labels for training, we annotate all the words in a sentence with the label of the sentence indicating whether or not the sentence should be selected as a key sentence. The function $\mathcal{E}(\cdot)$ for word embedding is trained by word2vec [22] with Wikipedia articles as a corpus, and the other parts of the proposed network are trained to minimize the cross-entropy error between the ground-truth labels $\mathbf{Y}$ and soft-max estimates $\hat{\mathbf{Y}}$ of outputs with ADAM [19].

6. KEY SENTENCE EXTRACTION

Key sentence extraction from a given new article can be achieved simply by decoding the article through the learned model and taking the most active output at every sentence. However, in this work, we introduce an additional classifier as a post-processing procedure to improve the performance. Specifically, we adopt support vector machines (SVMs) with RBF kernels for classification, and exploit $\overleftarrow{\mathbf{h}}_t$, a backward layer output, as the classifier input. With this strategy, our neural network model presented in Section 5 can be regarded as a feature extractor. Since estimates of the SVM classifier can be obtained for each word, we simply take the average of the estimates for all the words in a sentence to generate a final estimate that indicates whether or not the sentence should be identified as a key sentence.

7. EXPERIMENTS

We have undertaken experiments to evaluate the performance of our proposed method. The task here is to precisely

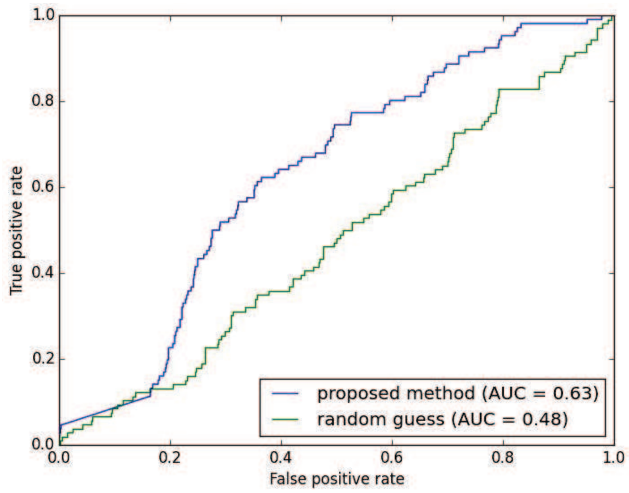

Figure 4: Experimental results

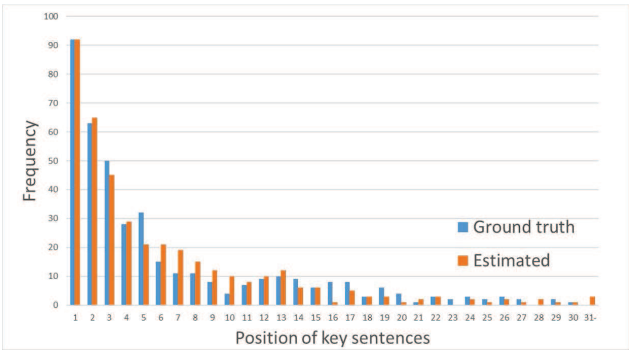

Figure 5: Positions of ground-truth and estimated key sentences

estimate key sentences for a given news article. The data set described in Section 4 was again used for this evaluation. We used the LIBSVM library [3] to implement SVM classifiers, and trained our neural network model and the SVM classifiers with four-fold cross validation. We introduced data augmentation to boost the performance with a limited number of training samples. More specifically, we selected long news articles with more than 10, 20 and 30 sentences and copied those articles after deleting any sentences after the 11th, 21st and 31st sentences if they included no ground-truth key sentences. We compared our proposed method with a simple baseline method that randomly extracts sentences of a given news article as a key sentence. We used a receiver operating characteristic (ROC) curve to evaluate the performance of the methods.

Figure 4 summarizes our experimental results. We observed that our proposed method worked reasonably well with the area under curve (AUC) = 0.63 and greatly outperformed the baseline (AUC= 0.48).

We also compared the positions of the ground-truth and estimated key sentences. Figure 5 shows the result, which indicates that the proposed method extracted key sentences not only from lead paragraphs but also from the later parts of news articles. We also observed that the proposed method often failed to extract the first and second sentences, which implies that several techniques should be introduced to bias the model towards selecting lead paragraphs.

8. CONCLUSION

This paper posed the problem of automatically generating attractive news headlines that have the potential to go viral on social media, and proposed a novel method for extracting key sentences from a given news article as material for generating attractive news headlines. We showed that the problem can be formulated as supervised sequence labeling that utilizes user activity on social media as supervised information, and proposed a method that integrated a recurrent neural network model for sequence modeling and SVM classifiers for key sentence extraction. Experiments with a corpus consisting of microblog messages and linked news articles showed the effectiveness of our proposed method. We have focused solely on the problem of identifying key sentences, however, we understand that several post-processing operations will be needed to generate news headlines, such as sentence compression [25, 8]. There is also the potential to improve the key sentence identification performance. Promising approaches for extending our model include deep bi-directional LSTM [13] and connectionist temporal classification [11].

9. REFERENCES

[1] J. Alejandro. Journalism in the age of social media. Reuters Institute Fellowship Paper, 2010.
[2] E. Alfonseca and D. Pighin. HEADY: News headline abstraction through event pattern clustering. In ACL2013.
[3] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. ACM Trans. IST, 2013.
[4] A. Deza and D. Parikh. Understanding image virality. In CVPR2015.
[5] N. Diakopoulos and A. Zubiaga. Newsworthiness and network gatekeeping on Twitter: The role of social deviance. In ICWSM2014.
[6] B. Dorr et al. Hedge trimmer: A parse-and-trim approach to headline generation. In HLT-NAACL2003 on Text Summarization Workshop.
[7] E. Filatova and V. Hatzivassiloglou. A formal model for information selection in multi-sentence extraction. In COLING2004.
[8] K. Filippova et al. Sentence compression by deletion with LSTM. In EMNLP2015.
[9] M. Gabielkov et al. Social clicks: What and who gets read on Twitter? In ACM SIGMETRICS 2016.
[10] M. Gidfer et al. Upworthy's how to make that one thing go viral, 2012. http://slidesha.re/QA1Tn9
[11] A. Graves et al. Connectionist temporal classification : Labelling unsegmented sequence data with recurrent neural networks. In ICML2006.
[12] A. Graves et al. Hybrid speech recognition with deep bidirectional LSTM. In ASRU2013.
[13] A. Graves et al. Speech recognition with deep recurrent neural networks. In ICASSP2013.
[14] A. Graves and J. Schmidhuber. Framewise phoneme classification with bidirectional LSTM networks. In IJCNN2005.

[15] M. Guerini et al. Exploring image virality in Google plus. In SocialCom2013.
[16] M. Guerini et al. Exploring text virality in social networks. In ICWSM2011.
[17] T. Hirao et al. Single-document summarization as a tree knapsack problem. In EMNLP2013.
[18] M. Jenders et al. Analyzing and predicting viral tweets. In WWW2013.
[19] D. Kingma and J. Ba. Adam: A method for stochastic optimization. In ICLR2015.
[20] S. Kourogi et al. Identifying attractive news headlines for social media. In CIKM2015.
[21] R. McDonald. A study of global inference algorithms in multi-document summarization. In ECIR2007.
[22] T. Mikolov et al. Distributed representations of words and phrases and their compositionality. In NIPS2013.
[23] J. Park et al. Managing bad news in social media: A case study on Domino's Pizza crisis. In ICWSM2012.
[24] P. Rasnick et al. RumorLens: A system for analyzing the impact of rumors and corrections in social media. In C+J2014.
[25] A. M. Rush et al. A neural attention model for abstractive sentence summarization. In EMNLP2015.
[26] B. Suh et al. Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network. In SocialCom2010.
[27] H. Takamura and M. Okumura. Text summarization model based on maximum coverage problem and its variant. In EACL2009.
[28] H. Takamura and M. Okumura. Text summarization model based on the budgeted median problem. In CIKM2009.
[29] R. Wang et al. Machine learning approach to augmenting news headline generation. In IJCNLP2005.
[30] S. Xu et al. Keyword extraction and headline generation using novel word features. In AAAI2010.
[31] Y. Yoshida Dependency-based discourse parser for single-document summarization. In EMNLP2014.
[32] D. Zajic et al. Automatic headline generation for newspaper stories. ACL Workshop on Text Summarization, 2002.