Public Sphere 2.0: Targeted Commenting in Online News Media

Abstract. With the increase in online news consumption, to maximize advertisement revenue, news media websites try to attract and retain their readers on their sites. One of the most effective tools for reader engagement is commenting, where news readers post their views as comments against the news articles. Traditionally, it has been assumed that the comments are mostly made against the full article. In this work, we show that present commenting landscape is far from this assumption. Because the readers lack the time to go over an entire article, most of the comments are relevant to only particular sections of an article. In this paper, we build a system which can automatically classify comments against relevant sections of an article. To implement that, we develop a deep neural network based mechanism to find comments relevant to any section and a paragraph wise commenting interface to showcase them. We believe that such a data driven commenting system can help news websites to further increase reader engagement.

1 Introduction

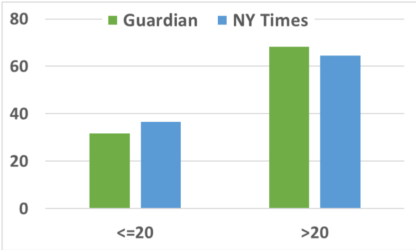
Recent years have witnessed a paradigm shift in the way people consume news. Online news media has become more popular than the traditional newsprint, especially to younger news readers [1]. To further engage them, in addition to presenting news, online news platforms also allow readers to comment and share their points of view on the matter reported in the story. Irrespective of concerns about the quality of the comments, especially their language and tone, comments are considered to be the most effective tool to increase reader engagements [1].

Several prior works in media and communication studies highlight the importance of discussions in the evolution of a democratic society. In a seminal work, Habermas established the notion of 'Public Sphere' where public opinion gets formed via *rational-critical debates* [2]. Ruiz *et al.* [3] argued that online news media provide a new manifestation of the public sphere – *Public Sphere 2.0*, where commenting acts as the facilitator of public debates.

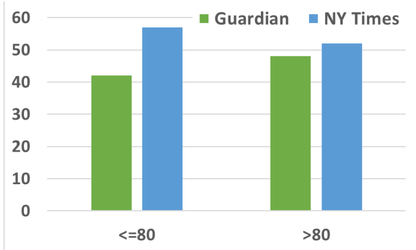
However, the myriad plethora of news websites and articles has resulted in a gradual decline of the attention span of an user to a particular news story. In an earlier work, Nielson [4] has noted that the readers predominately read online web pages in an F-shaped pattern i.e., two horizontal stripes in the top of the page followed by a vertical stripe along the page. This implies that the attention span of users wanes as they go through an article and most of their attention is focused on the initial paragraphs.

Hence, it is important to understand whether the commenting options in the online news media today can play the role of public sphere 2.0 and enable discussion on the news presented in an article. To investigate this issue, we have

[1] http://news.bbc.co.uk/2/hi/business/8542430.stm



(a) paragraphs vs comments



(b) sentences vs comments

Fig. 1: **Comment count varies with paragraph and sentence count.**

gathered articles and corresponding comments from two popular news websites, The Guardian (`theguardian.com`) and The New York Times (`nytimes.com`).

We observe that a large number of comments today are made targeting particular sections of an article, rather than the entire article itself. Yet, most news media websites allow their readers to comment only on the full article. In this paper, we propose to revamp the commenting UI by automatically placing the most relevant comments against each section of an article. For this, we develop a neural network based mechanism to map comments to particular paragraphs. Extensive evaluations show that our proposed methodology outperforms state-of-the-art baselines.

Finally, we build a system which can allow a reader to check for comments made against any section of an article and comment on the same. We believe this system can help news websites in increasing reader engagement further.

2 Motivation

In recent years, news media sites have seen huge increase in user engagement through commenting, liking, sharing etc. However, users do not spend similar time over the entire news article. The 'F-pattern' observed by J. Nielsen [4], shows that, for news articles, users mostly focus on initial paragraphs or few sentences of a paragraph to consume the summary of an article, possibly due to limited time to read the whole story.

To investigate how this influences the commenting behavior, we gathered news articles from two popular news websites - 'The Guardian' and 'The New York Times'. In total, we collected $1,352$ Guardian and $1,020$ NYTimes news articles encompassing various topics like Business, Technology, Politics, Sports and Editorials and all comments made against these articles.

Figure 1a and 1b show how the number of comments varies w.r.t. the number of paragraphs and sentences in an article (Y-axis is % distribution). Fig 1a points out that more than 60% comments are posted to the articles having more than 20 paragraphs. Fig 1b shows how comment distributions varies for 80 sentence threshold ($\sim$20 paragraphs) for two online news papers. Overall, we see that having more paragraphs in an article increases the number of comments posted against it. Thus, we can conclude that comment-paragraph relation is important.

Relevance Label	% in The Guardian	% in NY Times
1	31.05	40.11
2	19.09	10.23
3	17.77	17.50
4	19.08	12.29
5	13.01	19.87

Table 1: **Distribution of different labels for two datasets.**

From the articles collected, we randomly selected 50 articles from each media site for manual annotation, where two annotators were asked to give one of five possible relevance scores for a comment to a paragraph. The relevance scores are 1 (strongly irrelevant), 2(weakly irrelevant), 3(neutral), 4(weakly relevant) and 5(strongly relevant), where the relevance is judged by the presence and absence of common words or a common thought between the paragraph and the comment text. Both annotators provided a relevance score for each paragraph-comment pairs in all 100 articles. Inter-annotator agreement (Cohen κ) was 0.71. A particular relevance score to a comment-paragraph pair was granted when both the annotators agreed.

We observed that around 42.7% of the comments (in total) were relevant to the whole article as those were not mapped to a particular paragraph. We consider a comment to be related to the entire article if the comment has a relevance score ≥ 4 for at least 3 paragraphs or has a relevance score of ≤ 2 for all the paragraphs of the article.

However, approximately half of the comments (48.9% and 48.8%) of the Guardian and NYTimes articles are centered towards $2-3$ particular paragraphs as opposed to the entire article. Similar to [4], we also observe that the mean relevance of a comment decreases along the article's length. This exemplifies that more relevant comments are related to the beginning paragraphs of an article and such a trend holds true for both Guardian and NYT articles.

Thus it is an interesting problem to find out how comments are related to individual paragraphs rather than the whole article. To automatically find out this association, we created the gold standard annotated datasets of 1834 and 1114 comments for 'The Guardian' and 'New York Times' respectively. The detailed statistics of the different annotated labels are provided in Table 1. Using this data (after class balancing using the SMOTE [5] algorithm), we design an automated approach as explained next. .

3 Linking comments to paragraphs

In this paper, we propose an approach to correctly identify paragraph-comment pairs and encourage users to comment towards the paragraphs, instead of only commenting on the whole article. Our proposed framework is based on deep neural networks. We have used two different neural network models - Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) where inputs are paragraph and comment vectors. We have used the pre-trained 300 dimension

	The Guardian						New York Times					
	Precision			Recall			Precision			Recall		
Model	Macro	Micro	Weighted	Macro	Micro	Weighted	Macro	Micro	Weighted	Macro	Micro	Weighted
NB	46.2	42.6	61.6	42.6	42.5	42.6	33.9	35.2	60.9	40.9	35.2	35.2
DT	42.9	49.6	53.5	35.6	50.9	50.9	36.9	59.2	52.7	31.2	59.1	59.2
RF	37.9	44.7	45.1	24.4	44.6	44.7	17.5	57.6	37.8	20.2	57.3	57.6
K-NN	48.5	63.5	61.3	48.1	63.4	63.5	37.6	58.9	55.9	34.8	61.2	61.2
R-SVM	49.9	63.1	60.9	45.4	63.0	63.1	39.4	61.9	51.9	27.3	61.9	60.3
AdaBoost	38.3	49.3	48.2	35.7	49.3	49.2	29.3	56.1	48.1	28.5	55.1	54.6
LR	41.2	54.0	51.3	38.8	54.1	54.0	34.1	60.6	50.7	25.8	60.7	60.1
LSTM	64.1	74.4	74.5	63.6	74.5	73.3	56.6	76.8	76.1	57.8	76.9	76.8
GRU	**64.2**	**75.3**	**75.9**	**63.7**	**75.3**	**75.4**	**64.8**	**79.1**	**78.4**	**64.3**	**79.3**	**79.1**

Table 2: **Performance of different models on the two datasets.**

Google News Vectors for each word and in case a pre-trained embedding for a word is not found we take it to 0 (in 300 dimension space). In order to calculate the vector for the entire paragraph and comment, we take the average of all word vectors corresponding to each word in the paragraph and comment respectively. Deep neural network models - (i) LSTM and (ii) GRU were applied on top of the paragraph and comment vectors to get a 150 dimension vector for both paragraph and comment. Thereafter these two vectors were merged and on top of it a fully connected layer with 5 units (for five classes) and soft-max activation is applied to get the probability for each class. The proposed model is shown in Figure 2a. No explicit feature extraction, using POS Tagger or LIWC was required for these models.

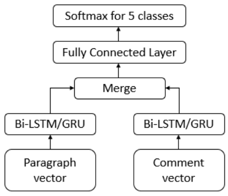
3.1 Baselines

Other than the deep neural network models, we have experimented with various traditional machine learning models - Naive Bayes (NB), Decision Tree (DT), Random Forest (RF), K-Nearest Neighbors (K-NN), RBF Support Vector Machine (R-SVM), Logistic Regression (LR) and Adaboost. We have extracted different features for these models, which can be grouped into three different categories.
POS Tag and Dependency Features: Stanford POS Tagger [6] and Stanford dependency parser [7] were used to get different Parts-Of-Speech based features. Total 45 features were extracted.
LIWC Features: Total 63 psycholinguistic features were extracted using the LIWC tool [8].
Others: Uni-gram, bi-gram, tri-gram features for paragraphs and comments.

After generating the feature matrix, dimensions were reduced using Latent Semantic Indexing (LSA) before feeding into the traditional ML-classifiers.

3.2 Evaluation

After feature extraction of the annotated datasets, various ML-classifiers were used to calculate 10-fold cross validation tests. For the deep learning model, we



(a) Architecture of the DNN model

(b) Snapshot of the system

Fig. 2: **Our proposed system.**

have trained for 5 epochs for each step in the 10-fold cross validation. Results are shown in terms of Macro, Micro and Weighted averaged precision and recall for 'The Guardian' and 'New York Times' datasets[2]. Table 2 shows that LSTM and GRU models outperform ML-classifier models in terms of all metrics and GRU model performs the best.

Figure 2b shows the snapshot of our model where top k (here k=3) relevant comments are highlighted when the cursor is placed around the second paragraph of a particular story.

To check the effectiveness of our system, we showed to 20 volunteers the same 10 Guardian news stories on the original website and through our system. At the end, the volunteers were asked to rate the interface better for commenting against the articles. 17 out of 20 volunteers gave higher rating to our system interface, and the main reason they cited is the ability to see old comments and post new comments against different portions of the articles.

4 Related Works

Earlier works on commenting can be categorized into four dimensions: (i) Comment to Article Mapping, (ii) Comment Ranking, (iii) Comment Recommendation and (iv) Comment Analysis.

(i) Comment to Article Mapping: Aker et al.[9] designed linear regression based graphical approach for topic clustering of online comments. Aker et al.[10] focused on linking comments to articles using similarity metrics, as defined by Das *et al.* [11].

[2] For ML-classifiers we have computed precision and recall for different combination of (i) POS Tag and Dependency, (ii) LIWC and (iii) Others features but due to space constraint only the best results were shown.

(ii) Comment Ranking: C.F. Hsu et al. [12] developed a regression model for identifying and ranking comments within a Social Web community based on the community's expressed preferences. Dalal et al. [13] built Hodge decomposition based rank aggregation technique to rank online comments on the social web.

(iii) Comment Recommendation: Bansal et al [14] proposed Collaborative Correspondence Topic Models (CCTM) to recommend comment-worthy articles (An article is defined to be comment-worthy for a particular user if that user is interested to leave a comment on it) like blogs, posts etc. to a particular user where user feature profile is generated by user contents analysis.

Shmueli et al. [15] built combined model of content-based approach with a collaborative-filtering approach (utilizing users' co-commenting patterns) for personalized recommendation of stories to users for discussing through comments. Agarwal et al. [16] focused on personalized user preference (i.e., whether the user is likely to like or dislike the comment) based ranking of the comments in an article.

(iv) Comment Analysis: X. Liu [17] ranked interest based news sections and articles by by using passage retrieval algorithm (user comments tagged news). Stroud et al. [18] analyzed population (e.g. features like - demographic makeup, attitudes, behaviors etc.) who comment or read online comment sections. Mullick et al. [19] classified online comments into opinion and fact. SM. Almgren and T. Olsson [20] provided comparisons among commenting, sharing, tweeting and measure participation of populations.

Our present work is complementary to these earlier works, as no other work has explored paragraph oriented commenting pattern and build a model showing relevant comments to a paragraph to ask user for more commenting.

5 Conclusion

To play the role of the Public Sphere, online news websites need to encourage readers to comment on their articles. In this paper, we argued for a revamp of the traditional commenting interface, and for enabling commenting on selective sections of an article. We developed a deep neural network approach to link comments to particular section. We showed that Gated Recurrent Unit (GRU) model provides best results in terms of macro and micro level precision and recall. Then, we built a basic user interface to increase user engagement in online comment sections. Our immediate future step is to develop an end-to-end system to show a user top K relevant comments (further divided into different sentiment expressed in the comments), while scrolling down the paragraphs. We believe such a data driven selective commenting system can bring more specific and targeted reader engagement for online publishing houses.

References

1. Park, D., Sachar, S., Diakopoulos, N., Elmqvist, N.: Supporting comment moderators in identifying high quality online news comments. In: CHI, ACM (2016)
2. Habermas, J.: Moral consciousness and communicative action. MIT press (1990)
3. Ruiz, C., Domingo, D., Micó, J.L., Díaz-Noci, J., Meso, K., Masip, P.: Public sphere 2.0? the democratic qualities of citizen debates in online newspapers. The International Journal of Press/Politics **16**(4) (2011)
4. Nielsen, J.: Usability 101: Introduction to usability (2003)
5. Chawla, N.V., et al.: Smote: synthetic minority over-sampling technique. Journal of artificial intelligence research (2002)
6. Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J.R., Bethard, S., McClosky, D.: The stanford corenlp natural language processing toolkit. In: ACL. (2014) 55–60
7. De Marneffe, M.C., Manning, C.D.: Stanford typed dependencies manual. Technical report, Technical report, Stanford University (2008)
8. Tausczik, Y.R., Pennebaker, J.W.: The psychological meaning of words: Liwc and computerized text analysis methods. Journal of language and social psychology (2010)
9. Aker, A., Kurtic, E., Balamurali, A., Paramita, M., Barker, E., Hepple, M., Gaizauskas, R.: A graph-based approach to topic clustering for online comments to news. In: ECIR. (2016)
10. Aker, A., Kurtic, E., Hepple, M., Gaizauskas, R., Di Fabbrizio, G.: Comment-to-article linking in the online news domain. In: SIGDIAL. (2015)
11. Das, M.K., Bansal, T., Bhattacharyya, C.: Going beyond corr-lda for detecting specific comments on news & blogs. In: WSDM. (2014)
12. Hsu, C.F., Khabiri, E., Caverlee, J.: Ranking comments on the social web. In: Computational Science and Engineering, 2009. CSE'09. International Conference on. Volume 4., IEEE (2009) 90–97
13. Dalal, O., Sengemedu, S.H., Sanyal, S.: Multi-objective ranking of comments on web. In: WWW, ACM (2012) 419–428
14. Bansal, T., Das, M., Bhattacharyya, C.: Content driven user profiling for comment-worthy recommendations of news and blog articles. In: Proceedings of the 9th ACM Conference on Recommender Systems, ACM (2015) 195–202
15. Shmueli, E., Kagian, A., Koren, Y., Lempel, R.: Care to comment?: recommendations for commenting on news stories. In: WWW, ACM (2012)
16. Agarwal, D., Chen, B.C., Pang, B.: Personalized recommendation of user comments via factor models. In: EMNLP, ACL (2011) 571–582
17. Liu, X.: Comment centric news analysis for ranking. Proceedings of the Association for Information Science and Technology **46**(1) (2009) 1–8
18. Stroud, N.J., Van Duyn, E., Peacock, C.: News commenters and news comment readers. Engaging News Project (2016)
19. Mullick, A., Ganguly, N., et al.: A generic opinion-fact classifier with application in understanding opinionatedness in various news section. In: World Wide Web. (2017)
20. Almgren, S.M., Olsson, T.: Commenting, sharing and tweeting news. Nordicom Review **37**(2) (2016) 67–81