Hotspots of News Articles: Joint Mining of News Text & Social Media to Discover Controversial Points in News

Ismini Lourentzou, Graham Dyer, Abhishek Sharma and ChengXiang Zhai
Department of Computer Science
University of Illinois at Urbana - Champaign
{lourent2,gdyer2,sharma51,czhai}@illinois.edu

***Abstract*—We propose and study a novel problem of mining news text and social media jointly to discover *controversial* points in news, which enables many applications such as highlighting controversial points in news articles for readers, revealing controversies in news and their trends over time, and quantifying the controversy of a news source. We design a controversy scoring function to discover the most controversial sentences in a news article by leveraging relevant comments in Twitter and comments on news web sites to assess the controversy of opinions about an issue mentioned in the news article. Multiple scoring strategies based on sentiment analysis and linguistic cues are proposed and studied. Experimental results show that the proposed algorithms can effectively discover controversial parts in news articles.**

I. Introduction

Managing, creating and sharing information has been reinvented in the Big Data era. Quickly locating interesting information in news articles can become time consuming. Although the headline of an article summarizes the content, there is an obvious need for automatically identifying sentences discussing controversial issues, which can be used to generate a digest of controversial issues in a given time period and help readers to quickly locate issues that triggered debates. In general, mining controversial issues from big text data is an important challenge behind many applications.

We propose a new idea for addressing this novel text mining problem, where we would mine news articles and the companion social media jointly to discover controversial points in news articles. We design and study two controversy scoring functions and use them to discover the most controversial sentences in a news article based on the relevant social media discussion of topics covered in each sentence in an *unsupervised* way. Such "companion" information of a news article is nowadays increasingly available on the Web.

By collecting news and social media comments in real time and nominating the top-K most controversial issues in a particular time period (e.g., during a day or a week), we can recognize new emerging controversial issues or further track the intensity of a particular controversial issue over time, and thus reduce the information clutter problem.

To the best of our knowledge, no previous work has studied how to rank sentences in a news article based on controversy. Furthermore, there is no research done on leveraging extreme words for controversy detection. However, there are several lines of related work. Vuong et al. [1] identify controversial articles in Wikipedia, Dori-Hacohen and Allan [2] detect controversial pages on the Web, while Pennacchiotti and Popescu [3], [4] focus on detecting controversies involving popular entities and events in Twitter data. Tsytsarau et al. [5] mine contradictions under multiple levels of time granularity from postings on a topic and Awadallah et al. [6] present a system, OpinioNetIt, that aims to automatically derive a map of the opinions-people network from news and other Web documents. More similar to our work, Choi et al. [7] discover controversial issues and their subtopics from news articles. The main difference of our method is that we "drill down" beyond the topic level in an unsupervised way. Although we do not retrieve opposing groups of people for each topic or each sentence in a news article, our method can be expanded to recognize stances in online debates, as well as create a timeline of controversy regarding an issue, topic or entity.

II. Problem Definition & Approach

We now formally define the proposed mining problem. Let $D = \{A_i, C_i\}_{i=1}^{N}$ be a corpus of N news articles with the associated comments, where A_i is a news article composed of a set of sentences $S_i = \{s_{i1}, s_{i2}, \ldots, s_{ij}\}$ and $C_i = \{c_{i1}, c_{i2}, \ldots, c_{im}\}$ is the collection of comments relevant to article A_i. The term "comments" in our case takes a broader meaning, as it is used to define posts (tweets) related to the article or comments made about the news article on a news website. Our task is to rank the news article sentences based on a scoring function $Score(s_{ij}|C_i)$ that can quantify controversy by leveraging a set of features extracted from the relevant comments in C_i that discuss the issue covered in sentence s_{ij}.

Our solution consists of three steps.

1. Retrieving relevant comments for each sentence: While comments are often linked to the full news article on news websites or by citing a news URL in a tweet, we would like to go further to associate comments with each sentence in a news article so that we can use them as a basis to assess the controversy of the corresponding sentence. That is, we need to identify for each sentence $S_{ij} \in A_i$ a subset of the most relevant comments $C_i' = \{c_{i1}, c_{i2}, \ldots, c_{ik}\} \subset C_i$.

This task is essentially a retrieval task. Thus we treat each sentence S_{ij} as a query and the comments C_i as the collection of documents, and leverage Okapi BM25 [8] to retrieve comments according to their relevance to the query sentence.

2. Extract controversy indicators based on the retrieved comments: We use sentiment, capitalization and extreme words. Controversial issues likely attract both positive and negative comments. To improve the scalability of our system, we chose SentiStrength, a fast sentiment analysis toolkit that can classify 16,000 social web texts per second and has human-level accuracy for short social web texts as described in [9], [10]. Text in social media often deviates from the standard rules of language. For example, punctuation and capitalization have been reinvented as ways to express emotions through text. The use of upper case terms (i.e. "all caps") in social media is commonly interpreted as yelling. Exaggeration can be found in the wide use of extreme words, such as "never, always" etc. We hypothesize that such linguistic cues can be potentially useful for identifying controversial issues and propose to leverage another set of features by counting in each comment the number of terms that have all upper case letters and the frequency of extreme words found in a manually compiled [1] "extreme lexicon", that enlists extreme adjectives and adverbs in English. Intuitively, we can expect the sentence to be controversial if a frequent use of extreme words and/or capitalization in both relevant positive and negative opinions is observed.

3. Design a scoring function to capture the existence of conflicting opinions in the comments: We experiment with two different statistical measures that we believe can incorporate dispute found in comments into a coherent ranking function for sentences; ratio and entropy.

Ratio: The first way to measure whether there are opposing opinions expressed in the comments is simply to compute a ratio for each feature. For example, in the case of sentiment, the expression is:

$$Sent(S_{ij}|C_i') = \left| \frac{\#C^-}{\#C^- + \#C^+} - 0.5 \right| \quad (1)$$

where $\#C^- \subseteq C_i'$ and $\#C^+ \subseteq C_i'$ are the numbers of negative and positive comments related to a sentence, respectively. Intuitively, the public opinion is turning towards a certain side as we move further from the centroid (0.5), that represents the balance of positive and negative comments. In order to combine sentiment and linguistic features in a common scale, we follow the same structure for defining the ratio of linguistic cues. The final ratio scoring function for controversy detection is the sum of the feature ratios with smaller values indicating more controversy.

Entropy: Entropy is a widely used measure of uncertainty of a random variable. Due to the fact that the output of the sentiment analysis is a scale of 9 classes (from -4 to 4, with zero (0) indicating neutral sentiment), it is easy to interpret sentiment as a discrete random variable X_{sent} with probability function $p(X_{sent} = x_i) = \frac{f(x_i) \in C_i'}{\sum_{i=1}^{z} f(x_i) \in C_i'}$ where z is the number of possible outcomes, in our case the 9 distinct aforementioned classes, $f(x_i) \in C_i'$ is the number of comments that have sentiment equal to x_i and $\sum_{i=1}^{z} f(x_i) \in C_i'$ is the total number of comments. Regarding the linguistic features, we can also use entropy to measure how uncertain the distribution of words satisfying a linguistic rule over comments is. In our case a high entropy represents the conflicting views in comments. Thus, high entropy would indicate high controversy; for example if extreme words are evenly spread over positive, negative, very positive, very negative etc. comments, then the amount of vertical exaggeration used in both opposing sides is the same. Additionally, in order to create a coherent scoring function for the sentiment and the linguistic features, we combine sentiment and the linguistic features as an unweighted sum of the three random variables, to keep the method unsupervised, however we keep in mind as future work to test the method in a supervised way and check whether some features could have dominant values.

III. Evaluation

For our experiments we extract a ranked list of controversial issues [2]. Using these controversial debate topics, we retrieve relevant news articles from CNN [3] and relevant comments from Disqus [4] and Twitter. We gather 192 news articles, 420,063 tweets and 403,947 comments. Annotators were asked to label each sentence as "controversial" or "non-controversial" based on the content of the news article, resulting in 6,218 labeled sentences: 2,730 sentences as controversial and 3,488 as non-controversial. We proceed by ranking the sentences of each news article using our controversial scoring functions and compare our results with the gold judgments. We report Mean Average Precision (MAP) as the main evaluation measure.

The main results are shown in Table I and figures 1-4, where "Linguistic" means the combination of the two linguistic features and "All" is combining all features together, both linguistic and sentiment. Moreover, we experiment with the level of retrieval, where the number of comments (tweets) retrieved varies from 50 to 3000 relevant posts. From these results, we can make a number of observations:

In most cases, entropy achieves higher performance compared to ratio when combined with tweets, however comments perform better with ratio as a scoring function. Using all features is the most effective approach in almost all cases except for extreme words that seem to be the best option

[1] from online English sources or lessons, such as http://www.espressoenglish.net/extreme-adjectives-in-english/

[2] http://www.debate.org/big-issues/

[3] http://www.cnn.com

[4] https://disqus.com/api/

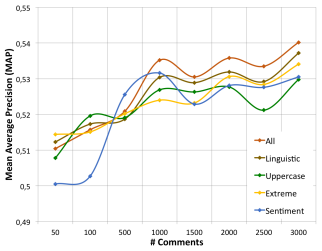

Figure 1. Comments - Entropy

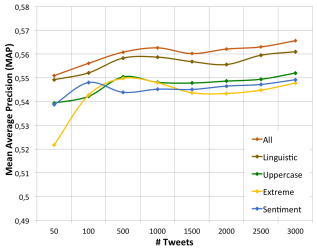

Figure 2. Tweets - Entropy

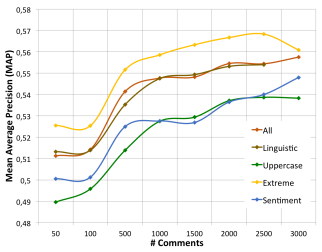

Figure 3. Comments - Ratio

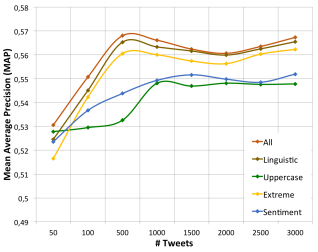

Figure 4. Tweets - Ratio

for comments combined with the ratio scoring function. We can also see in Table I that using twitter achieves higher performance for both scoring functions, despite the limited number of characters that tweets have; compared to comments in a news website, where such limitation does not exist. A main hypothesis in our work is that the number of social media posts used and the change in the performance of our methods are directly proportional. To test this hypothesis, we vary the amount of tweets (comments) available. Figures 1-4 showcase that increasing the number of social media posts used in our scoring functions indeed results in higher performance. Overall, evaluation results show that the proposed methods for controversy detection are effective and using all the features is in general beneficial.

MAP results	**Twitter**		**Comments**	
Method	**Entropy**	**Ratio**	**Entropy**	**Ratio**
All	0.5601	0.5585	0.5277	0.5411
Linguistic	0.5564	0.5558	0.5257	0.5403
Uppercase	0.5472	0.5410	0.5223	0.5212
Extreme	0.5426	0.5518	0.5237	0.5525
Sentiment	0.5454	0.5444	0.5211	0.5257

Table I
PERFORMANCE COMPARISON ON SCORING FUNCTIONS

IV. CONCLUSIONS AND FUTURE WORK

We proposed and studied a novel text mining problem that aims to mine news text data and the companion social media jointly to discover controversial points in news articles. Our work represents a step towards efficiently leveraging interconnected text data (in our case, news and the companion social media) for scalable retrieval of interesting information. The scoring functions proposed can be used as a highly versatile controversy ranking operator not only for sentences, but also for topics, entities, companies etc. Moreover, we have developed a web interface[5] for query-based retrieval of controversial news articles. The user can further click on a news article of his interest, where the most controversial sentences are highlighted with an additional option to explore relevant tweets. An important future work is to further evaluate our methods by incorporating the idea of weighting and explore supervised methods. Finally, leveraging additional features, especially social network features (e.g. number of followers) would help us improve the performance of these scoring functions.

ACKNOWLEDGMENTS

This work is supported in part by the National Science Foundation under grant number CNS-1027965 and by the National Institute of Health under grant number 1R56AI11450101A1.

REFERENCES

[1] B.-Q. Vuong, E.-P. Lim, A. Sun, M.-T. Le, H. W. Lauw, and K. Chang, "On ranking controversies in wikipedia: Models and evaluation," in *Proceedings of the 2008 ACM International Conference on Web Search and Data Mining*, 2008.

[2] S. Dori-Hacohen and J. Allan, "Detecting controversy on the web," in *Proceedings of the* 22^{nd} *ACM International Conference on Conference on Information & Knowledge Management*, 2013.

[3] M. Pennacchiotti and A.-M. Popescu, "Detecting controversies in twitter: A first study," in *Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics in a World of Social Media*, 2010.

[4] A.-M. Popescu and M. Pennacchiotti, "Detecting controversial events from twitter," in *Proceedings of the* 19^{th} *ACM International Conference on Information & Knowledge Management*, 2010.

[5] M. Tsytsarau, T. Palpanas, and K. Denecke, "Scalable discovery of contradictions on the web," in *Proceedings of the* 19^{th} *ACM International Conference on World Wide Web*, 2010.

[6] R. Awadallah, M. Ramanath, and G. Weikum, "Harmony and dissonance: Organizing the people's voices on political controversies," in *Proceedings of the* 5^{th} *ACM International Conference on Web Search and Data Mining*, 2012.

[7] Y. Choi, Y. Jung, and S.-H. Myaeng, "Identifying controversial issues and their sub-topics in news articles," in *Intelligence and Security Informatics*. Springer, 2010.

[8] S. E. Robertson, S. Walker, S. Jones, M. Hancock-Beaulieu, and M. Gatford", "Okapi at trec-3," in *TREC'94*, 1994.

[9] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas, "Sentiment in short strength detection informal text," *Journal of the Am. Society for Inf. Science & Tech.*, 2010.

[10] M. Thelwall, K. Buckley, and G. Paltoglou, "Sentiment strength detection for the social web," *Journal of the Am. Society for Inf. Science & Tech.*, 2012.

[5] http://controversy.2pitau.org/login