Mitigating Media Bias: A Computational Approach

Souneil Park, Seungwoo Kang, Sangjeong Lee, Sangyoung Chung, Junehwa Song

Korea Advanced Institute of Science and Technology (KAIST)

373-1 Guseong-Dong, Yuseong-Gu, Daejeon, Korea

{spark, swkang, peterlee, sychung, junesong}@nclab.kaist.ac.kr

ABSTRACT

The bias in the news media is an inherent flaw of the news production process, spanning news gathering, writing, and editing stages. Producer's subjective valuation, wittingly or unwittingly, takes place during the daily production process. The resulting bias often causes a sharp increase in political polarization and in the cost of conflict on social issues such as Iraq war [3]. It is very difficult, if not impossible, for readers to have penetrating views on realities against such bias. We propose *NewsCube*, a novel Internet news service framework aiming at mitigating the effect of media bias. NewsCube is designed to automatically create and promptly provide readers with multiple classified viewpoints on a news event of interest. It helps readers to easily discover rich facts and compare diverse biased views on the event. In this paper, we discuss the design of the NewsCube framework and introduce novel approaches which are under development.

Categories and Subject Descriptors

J.7 [**Computers in Other Systems**]: Publishing; K.4 [**Computers and Society**]: Social Issues; H.4 [**Information Systems Applications**]: Communication Applications – *Information browsers*; I.7 [**Document and Text Processing**]: Document Capture – *document analysis*; I.5 [**Pattern Recognition**]: Clustering

General Terms

Design, Algorithm

1. INTRODUCTION

Bias in the news media is an inherent flaw of the news production process, spanning news gathering, writing, and editing stages. At every single stage, news is probably never free from the producer's subjective valuation and external forces including owners and advertisers. Such valuation and forces often lead to totally different *framing*[1] of realities. Producers select different aspects of reality; they cover these aspects in different tones; they present them in different styles. The resulting bias causes a sharp increase in political polarization, misunderstanding of critical issues such as the Iraq war [3], and a significant impact upon elections [17]. It is very difficult for readers to have a balanced view on realities. They mostly receive news from restricted channels, often times through a single channel. Advanced news production and delivery models are required to break the vicious cycle of bias.

In this paper, we propose NewsCube, a novel Internet news

Figure 1. Framing an event with different photos[2]
(While others selected a handshake photo, the three major conservative news outlets (left side) of Korea covered the Korea-US summit with a photo showing the fist of US president.)

service framework that mitigates the effects of media bias. NewsCube is designed to automatically create and promptly provide readers with multiple classified viewpoints on a news event of interest. It helps readers to easily discover rich facts and compare diverse biased views on the event. As such, it effectively helps readers understand the event from a plural of viewpoints and formulate their own, more balanced viewpoints free from specific biased views.

Dealing with the media bias problem from the consumers' perspective has clear advantages over doing it from the producers' perspective. News producers have made considerable efforts to reduce bias. They established and observed journalism ethics and standards[3]. More recently, producers of radio and TV news have adopted an adversarial format [4] such as point-counterpoint roundtable discussions. The format attempts to give readers a balanced perspective by incorporating multiple viewpoints and opinions on a news event. However, such efforts have not succeeded in fully resolving the problem; *media bias is still widespread*. They do not address the underlying structural factors[5] of media production. Moreover, they cannot ensure that the inherent bias of producers does not affect the framing of articles or roundtable discussions. Addressing the structural factors of media bias is a difficult, large-scale problem. It requires significant changes in the complex relationships between government, business and media over the long term. NewsCube

---

[1] To frame is to "select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation and/or treatment recommendation" [4].

[2] Photo from http://www.hani.co.kr/arti/society/media/157337.html

[3] http://en.wikipedia.org/wiki/Journalism_ethics

[4] http://en.wikipedia.org/wiki/Media_bias#Efforts_to_correct_bias

[5] Structural factors may include conflicts of interest due to corporate influence in the form of ownership and advertising and the reliance on government and corporate press releases and press conferences [5].

offers a practical means of mitigating media bias that can be applied immediately.

The NewsCube is designed to effectively achieve its goal by classifying aspects, presenting them fairly, and recommending different aspects revealing contrasting viewpoints. The core of NewsCube is *aspect-based clustering*, a method to provide readers with a classified view of a set of articles with different aspects. Depending on the news producer's viewpoints, they frame issues by covering different aspects (i.e. features of framing such as main agent, action, and attributed cause) at different depths [7]. Thus, to provide readers with multiple viewpoints of a news event, most importantly, diverse aspects of the event have to be delivered. In this paper, we discuss the design challenges of the NewsCube framework and present the aspect-based clustering method. To the best of our knowledge, our work is the first to develop a news service framework attempting to mitigate media bias with computational techniques.

2. Motivation

We believe that advanced computational techniques can open new opportunities to overcome the chronic and serious problems of media bias. Many studies have shown the problems of the inherent bias of news companies [5]. A survey by the American Society of Newspaper Editors [16] revealed that 78 percent of the public believed that there was bias in news reporting. The effects of media bias on political or socio-economic issues have also been reported. Bernhardt et al. gave an example of the conflicting views on the Iraq war among people who have different news sources [3]. Della Vigna et al. showed how the media bias affected voting in elections through empirical studies [17].

A number of researches are conducted in mass communication on analyzing biased media frames [4][15]. However, due to the lack of technical support, they can hardly be applied to an online service. Research efforts attempting to identify and measure media frames usually rely on manual analysis. On the other hand, many research efforts on news article processing cannot be directly adopted since they focused on helping users efficiently browse a vast quantity of news articles on the Internet. Research on news article processing such as news clustering [14], topic detection and tracking [1], news summarization [9], and news categorization [6] mostly saw the relation between news documents and readers. They did not look deeply into how the problems occurring in the news production process are propagated to the readers. On the contrary, we aim to take a broader view and examine the three-party relationship, i.e., producers, news articles, and readers.

Bias handling has not only been an issue in news media but also in shared authoring communities such as Wikipedia[6] and dooyoo[7]. Their approach to the bias problem is to collect evaluations on produced contents from the public and revise biased contents. Such an approach may not be directly applied to achieving the goal of NewsCube due to the immediacy of news. Raising timely and prompt participation for evaluations and reviews in a short period of time will be a challenging issue. On the other hand, we believe our approach is complementary since it can help users compare different point of views on produced contents and facilitate the evaluation process.
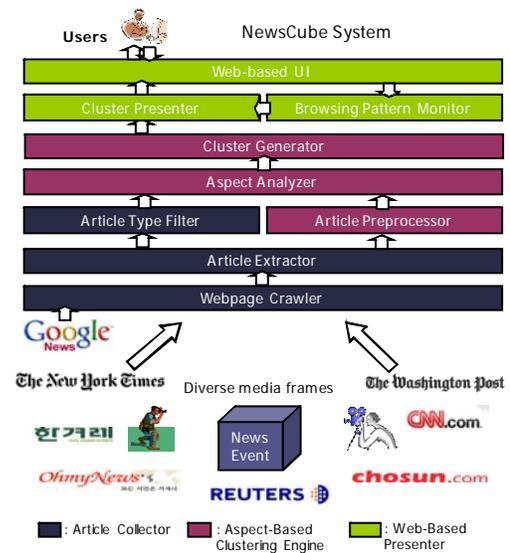

Figure 2. NewsCube Framework

3. NEWSCUBE DESIGN CHALLENGES

The unique goal of the NewsCube framework (Figure 2) raises a number of challenges and requires new approaches throughout the entire process of the service: article collection, processing, presentation, and personalization. In order to expose different viewpoints, diverse articles covering different aspects of a news event should be collected. Using the collected article set, producing a quality classification of articles based on aspect is important. Furthermore, to deliver the results more successfully, the system may create effective presentations and adapt to individual usage patterns. In this section, we discuss the challenges faced at each stage of the service process and considerations in implementing the components of the framework.

Wide Coverage Collection: It is necessary to collect articles from increasingly diverse sources to cover aspects of news events broadly. Heterogeneity of news providers is the first issue faced in article collection. The news articles are published from more than five hundred news providers, for example in Korea. This number greatly increases when non-traditional sources such as blogs are included. The Webpage Crawler of the framework, which is in charge of article collection, should be implemented to collect articles from various sources while considering their credibility, ideological characteristics, etc. Another consideration in source selection is the different webpage formats of the articles. The Article Extractor should extract pure news texts flexibly from different Web pages of wide range of sources.

Heterogeneity also lies in the types of news articles, e.g. straight news, interviews, and editorials. Bias is expressed differently depending on the types of articles. For example, bias is implicitly represented by filtering or emphasizing certain facts in straight news whereas it is more explicit in editorials. The Article Type Filter is necessary in order to separately collect and handle each type of articles.

Aspect-aware Article Processing: Identifying and classifying articles based on their aspects form the core of article processing for bias mitigation. Many types of bias in framing, including commission, omission, story selection and spin, commonly emphasize certain aspects of a news story while deemphasizing other aspects [2]. However, existing news processing techniques can hardly be applied since they are designed to classify articles

---

into a coarse-grained level, e.g. events or genre. News categorization [6] sort news articles into a pre-defined topic category, e.g. politics, business, and sports. A pre-defined category cannot be developed with aspects since they cannot be anticipated before occurrences of news events. News clustering [14] and News-event detection [1] techniques can be considered as an alternative since they organize news articles in an unsupervised manner. However, such techniques aim at providing grouped news articles related to an event, and thus they cannot effectively deliver multiple aspects under the heading of an event.

We have developed aspect-based clustering method to automatically organize articles based on aspect. The three components of the Aspect-Based Clustering Engine in Figure 2 performs clustering by extracting aspect of articles, and clustering the articles based on the similarity of the extracted aspects. The detail of the method is described in section 4.

Balanced Presentation: Presentation of processed results should be created carefully since bias may be implicit in the presentation itself. The goal of Cluster Presenter of the framework is to create an intuitive layout that enables users to easily compare contrasting viewpoints. Most importantly, articles revealing contrasting views should be located close to each other and such articles should be treated fairly in terms of space allocated. To support understanding of a selected news event, articles that provide crucial information should also be easy to find.

Creating such presentations is complex since it is impractical to allocate space equally to dozens of articles and locate them in eye-catching places in a webpage. Some articles which are redundant or less informative may be filtered out and articles may be ranked for locations if it is necessary. After articles are selected and ranked for display, their spaces and locations should be further adjusted carefully.

Counter-view Recommendation: To effectively achieve the goal of NewsCube, it is important to encourage users to view more articles covering different aspects. The counter-view recommendation approach aims at delivering articles that reveal aspects not being covered by the articles selected by the user. We are considering two methods, namely explicit recommendation and implicit recommendation. The explicit recommendation method displays a message suggesting articles covering different aspects for a user to read. Although this method is consistent with our goal, it may be inconvenient or annoying for the user. The implicit recommendation method dynamically adjusts the presentation of articles on the web page to increase the likelihood of users' selecting certain articles to read. Users may view more articles if unread aspects are prominently displayed or highlighted. For the recommendation, Browsing Pattern Monitor should be implemented to track users' browsing behavior and infer articles that a user reads.

4. ASPECT-BASED CLUSTERING

Classifying articles based on aspect is a technically involved problem, mainly due to the lack of semantic knowledge with which to abstract and classify the mainly covered aspects of articles. In this section, we present our model for aspects of news events and describe the proposed method for article clustering based on aspects.

4.1 Modeling Aspects of News Events

Among various facts of a news event, reporters select some facts for their articles. Aspects refer to the selected facts in the articles.

---

1. Title: Police, Soldiers Evict Protesters from U.S. Base Site
"Thousands of South Korean police and soldiers on Thursday evicted the remaining occupants and anti-U.S. civic activists from an area designated for expanded U.S. military facilities south of Seoul. ..."

2. Title: More Violence Looms in Planned Rally at U.S. Base Site
"Another violent showdown looms as groups protesting against the relocation of U.S. military bases to Pyeongtaek, Gyeonggi Province are gearing up for a massive 10,000-strong protest this weekend…"

Figure 3. Example articles

Main agent: Who does the act? (need not to be a human)
Action: What is the agent's action about?
Agency: How is the agent's action done?
Scene: What is the agent's background of the action?
Implications: What are the consequences?

Figure 4. Elements of story-line

Figure 3[8] shows two news articles reporting the same event but choosing different aspects in their story-lines.

Aspects are used to organize the story-line of news articles [4][10]. In detail, aspects can be understood as a model identifying the elements of news story-lines. Usually, a news story-line is constructed based on the five elements presented in Figure 4 [10]. Selecting different elements which will be importantly covered can lead to totally different story-lines. We conceptualize aspects as a set of answers to the selected elements, mostly consisting of a main agent and corresponding descriptions. In the example in Figure 3, the first article selects the police as a main agent and describes their stern actions evicting the protesters. However, the second article selects the protesters as a main agent and describes their plan for a violent protest.

Extracting such semantic information, i.e., aspects, is very challenging when it has to be performed computationally. To understand the aspects of articles correctly as humans do, machines should recognize the main agent and understand the meaning of the related descriptions. They will require much more complex semantic knowledge than just the meaning of words, e.g., relationship between participants, related events in the past, etc. Despite much effort to enable computational semantic processing, it seems very hard to completely encode such semantic knowledge into machines. This is because news topics generally span over various broad domains and are newly generated every day.

To be more practical, we adopt a keyword-based approach to approximate aspects. We note that the aspects of news articles are usually reflected in keywords in the articles. For example, in the above two articles, the keywords such as "police, soldiers" and "violent shutdown, protesting groups" directly reveal the main aspect of each article respectively. Furthermore, how importantly they are handled often reveals on which aspect is focused. Keywords related to the main aspect may be located in titles or may appear frequently.

Therefore, we define an aspect as a set of pairs of a keyword and a weight value (see Def. 1). The weight values are calculated according to the importance of keywords. Although the definition cannot fully capture the semantic meaning of texts, it is meaningful because we originally aim to group articles by

---

[8] Article 1 from http://english.hani.co.kr/arti/english_edition/12
1596.html
Article 2 from http://english.chosun.com/w21data/html/news/200605/20
0605090016.html

capturing the differences among aspects, not by perfectly understanding individual aspects.

Definition 1. Computational Aspect
For a given article, *a*,
Aspect(*a*) = {(*k_i*, *w_i*) | *k_i* ∈ W(*a*) and *w_i* ∈ **R**},
where W(*a*) denotes a set of words existing in the article

4.2 News Structure-Based Aspect Extraction

Selecting keywords and computing their weight values are more challenging since the difference of aspects should be captured within articles reporting a common news event. Existing methods such as TF-IDF are usually used to extract important terms from documents of diverse topics and distinguish the topics. Applied to the articles reporting a common event, they will detect the common topic of the event since the articles tend to share many important terms. We need a more sophisticated method that can handle the subtle difference presented in aspects. Keywords should be selected more carefully; their weight also should be analyzed in a finer-grained manner to distinguish the keywords holding the salient aspects of an article.

One of the famous news writing rules, called 'the inverted pyramid style of news writing', gives us a conclusive hint for aspect extraction; where the semantically meaningful keywords are located and how much they are importantly addressed. The style has been considered as an implicit standard for more than a century. Journalists are highly trained to follow this style regardless of nations or companies. The rule guides journalists to arrange information and facts in the descending order of importance within an article. Without reading to the end, readers will still understand the most important parts of the story. Most news articles are written in this style except informal types of articles such as columns, editorials or interviews.

The inverted pyramid structure is composed of head, subhead, lede and main text. The head and the sub-head are decisive factors in deciding whether or not to read an article. Therefore, they usually contains core terms which reflect the aspect being focused in the article. The lede is usually the first sentence of the article, or the first two sentences in some cases. To quickly inform readers, the information ordering principle especially is applied to the lede and guides reporters to reveal the most interesting facts. In addition, the lede covers the selected aspect clearly and concisely to avoid the unreadability of long sentences. Thus, we believe the lede will contain crucial terms which reflect the mainly covered aspect. The main text is organized with collected information which supports the lede in the order of diminishing importance.

The structure of news articles is exploited in both keyword selection and weight calculation. For keyword selection, we basically extract keywords from the core parts of an article, i.e., head, sub-head and lede. The weights of the extracted keywords are calculated in a more sophisticated manner using the inverted pyramid structure. Not merely counting the occurrences of keywords, the locations of occurrences and the amount of texts used to describe the keywords are taken into account. A keyword is considered important if the followings hold. First, the keyword is covered in the front part of the main text. Second, the keyword appears repeatedly. Third, the keyword appears in a long sentence or a long paragraph. The variables and equations for the weight calculation are summarized in Table 1 and Equations (1)-(3), respectively. For each occurrence, the weight is given based on the length of a paragraph *Lp(i, j)*, and a sentence *Ls(i, k)* relative to the length of the article *La(i)*. In order to differentiate weights

Table 1. Notations

Symbol	Description
A(i)	i-th article in a set of news articles
La(i)	Length of A(i)
Lp(i, j)	Length of j-th paragraph in A(i)
Ls(i, k)	Length of k-th sentence in A(i)
W(i)	Set of keywords of A(i)
w_n	n-th keyword in W(i)
v_n	Weight value of w_n
Va(i)	Set of keyword and its weight value pairs
d_{Pj}	Diminishing factor of j-th paragraph in A(i)
d_{Sk}	Diminishing factor of k-th sentence in A(i)

$$v = \sum_{j,k} \left(\frac{Lp(i,j)}{La(i)} d_{Pj} + \frac{Ls(i,k)}{La(i)} d_{Sk} \right) \qquad (1)$$

$$d_{Pj} = 1 - \sum_{m=1}^{j-1} \frac{Lp(i,m)}{La(i)} \quad , where \quad d_{P1} = 1 \qquad (2)$$

$$d_{Sk} = 1 - \sum_{m=1}^{k-1} \frac{Ls(i,m)}{La(i)} \quad , where \quad d_{S1} = 1 \qquad (3)$$

according to the locations in the structure, a diminishing factor d_{Pj} is multiplied to *Lp(i, j)* and d_{Sk} to *Ls(i, k)*. The values of the diminishing factors decrease as the keywords occur at the latter parts.

4.3 Major-minor Clustering Method

It is another key challenge to distinguish multiple aspects for an event and deliver them effectively. As described, our computational aspect model based on keywords and their weight values facilitates the extraction of aspects in a news article. However, extracted keywords and their weight values themselves do not distinguish different aspects in an article nor explicitly tell us what a main aspect of the article is. Thus, beyond the computational aspect extraction, it is imperative to develop an efficient method for the careful classification of articles in terms of their main aspect. We develop a clustering method which can group articles in an unsupervised manner since it is impossible to predict and predetermine a category with aspects before occurrences of news events.

Our clustering method, named *major-minor clustering*. Based on an extensive article set analysis [11], the proposed major-minor clustering method is designed to perform clustering in two main stages. For the first step, the method separates an article set into a major group and a non-major group. By carefully analyzing the distribution of keywords' weights over articles, articles which include popularly appearing keywords over the set are separated into the major group, otherwise into the non-major group. Second, the method partitions the non-major group into minor groups by analyzing the similarity of weights given to their keywords of the article set.

5. CONCLUDING REMARKS

We have been conducting research to address the challenges described above and have built a prototype system (See Figure 5). The prototype system automatically collects news articles reporting on a common event from various sources, identifies the main aspect of articles and performs clustering. Users can view various aspects by skimming through the groups. As in Figure 5, to encourage users to view more groups, groups which are read by a user are traced and grayed out to give emphasis on the unread groups. The system is providing a public service in order to demonstrate its practical value [18]. The details of the system and

clustering method and their evaluation can be found in [11] and [12].

Figure 5. Screenshot of the prototype system (Korean)

6. REFERENCES

[1] J. Allan, et al. Topic Detection and Tracking Pilot Study: Final Report. In *Proceedings of DARPA Broadcast News Transcription and Understanding Workshop*, 1998.

[2] Brant Baker. How to Identify, Expose and Correct Liberal Media Bias. Media Research Center, 1994.

[3] Dan Bernhardt, et al. Political Polarization and the Electoral Effects of Media Bias. *CESifo Working Paper Series No. 1798*, Available at SSRN: http://ssrn.com/abstract=892475, 2006.

[4] Robert M. Entman. Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication* 43 (4): 51-8, 1993.

[5] E. S. Herman, et al. Manufacturing Consent: the Political Economy of the Mass Media. New York: Pantheon Books, 1988.

[6] W.-L. Hsu, et al. Classification Algorithms for NETNEWS Articles. In *Proceedings of CIKM*, 1999.

[7] Jae-Kyung Lee. Basic News Reporting & Writing. Namoo wa Soop (in Korean), 1998.

[8] M. Mark, Miller. Frame Mapping and Analysis of News Coverage of Contentious Issues. *Social Science Computer Review* 15 (4): 367-78. 1997.

[9] D. R. Radev, et al. NewsInEssence: Summarizing Online News Topics. *Communications of the ACM Vol. 48, No. 10*, 2005.

[10] June Woong Rhee. Framing the Health Care Reform Campaign of 1993-94: News Frame, Interpretation, and Public Opinion Change. *Political Communication, vol. 15*: 274-76. 1998.

[11] Souneil Park, et al. Aspect-based Clustering of News to Mitigate Media Bias, *KAIST Technical Report CS-TR-2008-282*, Jan. 2008.

[12] Souneil Park, Seungwoo Kang and Junehwa Song, PolyNews: Delivering Multiple Aspects of News to Mitigate Media Bias, *KAIST Technical Report CS-TR-2006-263*, Nov. 2006.

[13] Michael Steinbach, et al. A Comparison of Document Clustering Techniques. *University of Minnesota, Technical Report #00-034*, 2000.

[14] H. Toda and R. Kataoka. A Clustering Method for News Articles Retrieval System. In *Proceedings of WWW Conference*, 2005.

[15] Tankard, James W. The Empirical Approach to the Study of Media Framing. In *Framing Public Life: Perspectives on Media and our Understanding of the Social World*. Stephen D Reese, Oscar H Gandy, and August E Grant. Mahwah, NJ: Lawrence Erlbaum Associates.

[16] C. Urban. Examining Our Credibility: Perspectives of the Public and the Press. *American Society of Newspaper Editors*, 1999.

[17] Stefano Della Vigna and Ethan Kaplan. The Fox News effect: Media bias and voting. *University of California, Berkeley Mimeograph*, 2005.

[18] NewsCube. http://newscube.kaist.ac.kr.