Don’t burst blindly: for a better use of natural language processing to fight opinion bubbles in news recommendations
Evan Dufraisse, Célina Treuillier, Armelle Brun, Julien Tourille, Sylvain Castagnos, Adrian Popescu

To cite this version:
Evan Dufraisse, Célina Treuillier, Armelle Brun, Julien Tourille, Sylvain Castagnos, et al.. Don’t burst blindly: for a better use of natural language processing to fight opinion bubbles in news recommendations. First Workshop on Natural Language Processing for Political Science, Jun 2022, Marseille, France. pp.79-85. cea-04562587

HAL Id: cea-04562587
https://cea.hal.science/cea-04562587
Submitted on 29 Apr 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License
Don’t Burst Blindly: For a Better Use of Natural Language Processing to Fight Opinion Bubbles in News Recommendations

Evan Dufraisse*, Céline Treuillier*, Armelle Brun*, Julien Tourille†, Sylvain Castagnos*, Adrian Popescu†

† Université Paris-Saclay, CEA, List, F-91120, Palaiseau, France
* Université de Lorraine - CNRS - Loria, Vandoeuvre les Nancy Cedex, France
{evan.dufraisse, julien.tourille, adrian.popescu}@cea.fr
{celina.treuillier, armelle.brun, sylvain.castagnos}@loria.fr

Abstract

Online news consumption plays an important role in shaping the political opinions of citizens. The news is often served by recommendation algorithms, which adapt content to users’ preferences. Such algorithms can lead to political polarization as the societal effects of the recommended content and recommendation design are disregarded. We posit that biases appear, at least in part, due to a weak entanglement between natural language processing and recommender systems, both processes yet at work in the diffusion and personalization of online information. We assume that both diversity and acceptability of recommended content would benefit from such a synergy. We discuss the limitations of current approaches as well as promising leads of opinion-mining integration for the political news recommendation process.

Keywords: Political polarization, News processing, Recommender systems, Opinion bubbles

1. Introduction

The ubiquitous use of social media has revolutionized the way people consume news and get exposed to information. Social media give users access to a huge amount of news, and opportunities to engage with diverse opinions. However, the access to such a rich information landscape comes with important challenges. In particular, personalization tools, designed to help users access content they are interested in, filter and hide information, offering only news in line with a user’s opinion. This selective exposure limits the presentation of contrasting viewpoints, leading to the creation of opinion bubbles (Pariser, 2011; Bozdag and van den Hoven, 2015). In the political domain, such an exposure tends to polarize citizens’ opinions (Pariser, 2011; Zuiderveen Borgesius et al., 2016), often drifting them towards extreme viewpoints (Sunstein, 2009).

On the long run, polarization is detrimental to the political debate and ultimately to democratic societies. The High Level Expert Group on Media Diversity and Pluralism highlights that people need to confront opinions that differ from their own to develop themselves fully. In this context, opinion bubbles are a growing concern for researchers from different disciplines (e.g. political science, economics, computer science or media), with interests ranging from assessing the real impact of personalization (Zuiderveen Borgesius et al., 2016) to bursting these bubbles (Burbach et al., 2019). In this work, we focus on computer science research on opinion bubbles in two distinct but related domains. The News Recommender System (NRS) community has a long-term interest in designing recommendation algorithms that broaden users’ view about a given topic. To this end, researchers focus on forming a diversified set of news articles to make sure that users can access diversified opinions. These research efforts mainly differ in the way they measure diversity and how they use it (Kunaver and Pözl, 2017; Raza and Ding, 2021; Möller et al., 2018). However, as we will show, the news content analysis and the way diversity is managed are not adapted to the specificity of political NRS, which limits the impact of diversification.

For its part, the Natural Language Processing (NLP) community focuses on news content understanding through different tasks such as topic modeling, opinion mining or argument mining (Hemmatian and Sohrabi, 2019; Lawrence and Reed, 2020). The most effective strategies in those fields are supervised and heavily rely on annotated data. This makes the elaboration of solutions even more challenging considering the highly dynamic nature of news in terms of topics and opinions. In this paper, we discuss the importance of a fine-grained analysis of the opinions expressed in the news, combined with a NRS that handles these opinions, and not only topics, to create personalized sets of recommendations and efficiently burst bubbles. From an ethical point of view, the NRS should not favor any specific opinion but should guarantee that the recommendations are representative of the diversity of the opinions expressed about the topic (Helberger, 2019).

The rest of the paper is organized as follows. After presenting a review of the literature of NRS (Section 2), we focus on news content analysis in NLP (Section 3). In Section 4, we present our view of the characteristics that a system designed to burst bubbles should have, as well as the way it should be designed.

2. News Recommender Systems

Online platforms offer access to a vast amount of content, which needs to be ranked according to the pref-
erences of each reader. In this context, NRS are designed to help readers find relevant information among the large quantity of news available (Raza and Ding, 2021). The news recommendation task has three main characteristics that distinguish it from other recommendation tasks: news articles quickly become obsolete (1) and are characterized by a high turnover (2) as a large number of news articles is published every second (Lunardi et al., 2020). This induces the need for recommendation models that can be updated on the fly and that do not require many interactions between news and readers. Besides, news consumption trends to influence users’ opinion (3) (Helberger, 2019). This is highly critical when it comes to politics, as NRS have been shown to contribute to the creation of opinion bubbles, representing a threat to democracy (Pariser, 2011).

As a consequence, sets of recommendations have to be properly balanced to ensure that users can access news that convey diverse opinions. This refers to the accuracy-diversity dilemma (Zhou et al., 2010), which refers to the search of an optimal balance between a high level of accuracy, to keep users’ trust, and a sufficient diversification among recommendations. In the news domain, diversity is often viewed as mandatory since it represents a core principle for the development of a democratic society (Helberger, 2019). However, the literature does not offer a unique way to represent the news, to evaluate the diversity, nor to recommend diversified sets of news articles.

### 2.1. News representation

Before measuring their diversity, news articles are often pre-processed in order to build a representation of their content (e.g. body, title, preamble or keywords). This representation must be designed to precisely characterize news. Recent models in NRS use deep-learning approaches such as named entity recognition, entity-linking, or knowledge-graph to provide a complete representation of news (Wang et al., 2018; Joseph and Jiang, 2019; Zhang et al., 2021). These representations are optimized for high accuracy but do not meet the needs for a precise control of recommendation diversification. Moreover, NRS mainly promote diversity through the use of topic modeling, and rely on simple bag-of-words representation. For instance, they represent the discriminating power of each word in the news using TF-IDF (Gao et al., 2020) or build topic representation using LDA (Tintarev et al., 2018). Few studies focus on Sentiment Analysis (Wu et al., 2020) by identifying and representing positive and negative sentiments in the news. The authors hypothesize that topic and sentiment analysis can increase the diversity of recommended opinions, but no empirical evidence is provided. Models that focus on topic diversification and coarse sentiment diversification can barely capture opinions, even less their nuances. Their use to foster opinion diversity in NRS is thus limited. Apart from topics and opinions, particularities of textual content such as the style of the authors or the use of irony, are not processed by NRS either.

### 2.2. Diversity measures in NRS

NRS differ in their definition of diversity and its associated metrics. For the simplest cases, diversity is considered as the opposite of similarity. The similarity measure can be instantiated by cosine similarity or based on distance (e.g. Jaccard or Euclidean) (Möller et al., 2020; Lunardi et al., 2020). In rare cases, other metrics are used, such as entropy (Shannon index, Rao’s quadratic entropy) (Möller et al., 2018). Generally, diversity metrics are derived from other fields, but the literature rarely adapt them to the news domain, which may result in meaningless values. More importantly, as mentioned above, diversities computed from simple representations of news can hinder even more their accuracy.

### 2.3. Diversification processes

Diversity measures constitute the basis for diversification processes. Diversification is often implemented as a post-processing step which re-ranks a list of recommended contents by prioritizing contents with a high variety of topics (Lunardi et al., 2020; Ziegler et al., 2005). An important downside of this approach is that diversity cannot be improved if the initial list is homogeneous. To answer this limitation, researchers propose to incorporate diversity into the core of the recommendation algorithm. (Raza and Ding, 2020) presents a recommendation algorithm that weights both diversity and accuracy in a personalized way, and tries to answer the accuracy-diversity dilemma. As a result, diversification among recommendations is not simply based on the re-ranking of the news, but relies on the prior parameter optimization of the recommendation model.

To summarize, existing diversification processes are interesting, but their potential is limited by the use of simplistic news representations. Representations which are more adapted to the specificity of news – in particular political news – and diversification are needed. They can be obtained by applying advanced NLP techniques.

### 3. Natural Language Processing

Diversifying news content opinions requires a fine-grained understanding of articles’ stances. Several NLP sub-fields have developed approaches to meet this end. In the following, we distinguish proxy measures, that extract opinion cues, from finer-grained strategies.

#### 3.1. Proxy strategies to stance detection

A system able to extract fine-grained opinions from arbitrary textual contents is still an object of research. Facing this challenge, proxy measures that are easier to obtain can serve as coarse opinion indicators. The methods presented in this subsection are often gathered under the umbrella term of “media bias detection”. We refer the reader to (Nakov et al., 2021) for a complete survey of the subject.
3.1. Content-based strategies

Media biases can take several forms: (i) a subjective stylometry, (ii) a coverage restricted to a subset of topics (topic diversity), (iii) an unequal attention paid to certain aspects or facts in events (framing), or (iv) a constant leaning towards a political group.

Stylometry-based detection approaches revolve around the detection of subjective expressions using dictionaries. The latter are usually gathered using an unsupervised strategy (Riloff and Wiebe, 2003). Recently, (Patankar et al., 2019) used (Recasens et al., 2013) lexicon to make a bias-aware NRS. However, this strategy only detects a lack of stylometric neutrality but cannot help determine an article’s stance. Topic diversity bias is already discussed in the NRS literature, we redirect to Section 2 for further information.

Framing (Entman, 1993) implies a consistent focus on some aspects of an issue that leads to its partial comprehension and biased interpretation by the reader. An early NRS strategy implemented to counter this bias used keyword extraction and unsupervised clustering of articles (Park et al., 2009). Later, automatic approaches for framing identification have been developed around the detection of so-called frames, which are characteristic aspects of a particular issue. Considering a representative set of articles that address a same topic, the aim is to detect significant deviations in aspects distribution as an indicator of bias. However, the issue-specificity of aspects prevent their widespread computational use. (Boydston et al., 2014) solves this issue by developing a set of 15 generic frames that are pervasive in most subjects. This approach was later consolidated with a dataset (Card et al., 2015). More recently, (Kwak et al., 2021) offered a new approach around the use of “micro-frames”. They construct semantic axes based on the Glove embeddings (Pennington et al., 2014) of 1,621 antonyms selected from WordNet (Miller, 1995). Using the Glove embeddings of words within a text, they compute the cosine distance to the semantic axes, and derive bias indicators from the score distribution. Nonetheless, frame identification for stance detection is limited, as two articles could defend opposite views over the same set of aspects but with the same polarities.

If one can derive clues on a source’s political leaning using its topic diversity and framing, other cues based on semantically loaded expressions have proved their efficiency. Several methods use the U.S congressmen’s speeches as source data to extract politically differentiating expressions (Groseclose and Milyo, 2005; Gentzkow and Shapiro, 2010; Bayram et al., 2019). Recently, (D’Alonzo and Tegmark, 2021) led a comparative study over several media, solely using articles to extract such expressions.

3.1.2. Audience-based strategies

Stance detection through audience analysis is a complementary approach to text-based methods that stems from the homophily principle, which postulates that users principally interact with content they agree with. The overall political stance of a media can be derived by analyzing those interactions. Most research in this field revolves around the use of Twitter follow/retweet interactions to map media or entities in the political spectrum (Wong et al., 2016; Stefanov et al., 2020; Darwish et al., 2020). Other approaches use readily available media bias analyses from News Guard, All-Sides, or Media Bias/Fact Check to consolidate supervised datasets of articles (Baly et al., 2020).

The derived political stances could be used to diversify opinions through the diversification of news sources. Two underlying assumptions could preclude the success of such an approach, namely, that “News articles follow the political leaning of their source outlet” and that “Political leanings of news outlets do not change across topics”. (Ganguly et al., 2020) has recently shown that both of these assumptions are often violated on an article basis. To fully grasp the political stance of an article, one needs to rely on the content itself.

3.2. Opinion-Mining for stance detection

In computer science, “Opinion Mining” and “Sentiment Analysis” are often used interchangeably. Nonetheless, in everyday language, an opinion is rather defined as a “judgment formed about something”. Choices of subjective and sentimentally expressive words are indicative of such judgments, and form the basis of stance detection.

Among the three levels of sentiment-analysis usually distinguished (document, sentence and entity), only the finer-grained one is suitable for stance detection, as the document and sentence levels of analyses are too coarse to be informative of a writer’s stance. A document or a sentence could contain several entities upon which opinions are expressed.

3.2.1. Aspect-Level Opinion Mining

(Liu, 2010) defines opinions as quintuples $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$, where $e_i$ is the $i^{th}$ entity considered, $a_{ij}$ is the $j^{th}$ aspect of that entity, and $s_{ijkl}$ is the sentiment that the opinion holder $h_k$ expresses towards the aspect $a_{ij}$ at time $t_l$. Opinion mining at the aspect level is interested in extracting, sometimes partially, those quintuples. The two most studied extraction frames are “product aspect mining”, with the extraction of $(e_i, a_{ij}, s_{ijkl})$ triplets, and “stance detection” with the extraction of $(e_i, s_i)$ tuples. These frames can be split into two steps: Aspect-Extraction (AE), and Aspect-Based Sentiment Classification (ABSC). Aspect-based Opinion Mining has mostly been applied to product reviews datasets (Pontiki et al., 2011; Pontiki et al., 2015; Pontiki et al., 2016), but also to financial (Jangid et al., 2018; Gaillat et al., 2018), and more recently, news datasets (Steinberger et al., 2017; Hamborg and Donnay, 2021).

Aspect-Extraction (AE) Given an opinionated-text content as input, AE aims at extracting the targets...
and aspects \((a_{ij}, e_i)\) or \(e_i\) towards which a sentiment might be expressed. Approaches can be grouped into four types. Frequency-based approaches extract most frequent nouns and noun phrases after Part-of-Speech (PoS) tagging, notably missing less frequent aspects, and generating a large number of noisy targets. Other approaches such as \((\text{Qiu et al., 2011})\) extract domain-specific opinion words and targets using syntactic relationships. They alternatively expand both an opinion word lexicon and a set of candidate targets in a bootstrapping fashion. Supervised methods, by formalizing AE as a sequence-labeling task have also been implemented. Finally, non-lexicon-based unsupervised methods make use of topic-modeling approaches (e.g., LDA, PLSA). An example of such an approach is \((\text{Titov and McDonald, 2008})\), which models a document as a mixture of both local and global topics, local topics are derived at a window scale to capture target aspects. The above approaches assume explicitly mentioned aspects, however, aspects can also be implicit. We refer the reader to literature surveys \((\text{Hemmatian and Sohrabi, 2019}\) \cite{Hemmatian2019} \cite{Nazir2020}) for further information on the topic.

**Aspect-Based Sentiment Classification (ABSC)**

Once extracted aspects and entities, we need to determine the sentiment expressed towards them. In this context, machine learning methods are often distinguished from rule-based ones. Rule-based methods mostly rely on PoS tagging and the adjective-noun proximity heuristic to associate adjectives to aspects. These methods have been supplanted by machine learning methods, among which Deep-Learning recently took the lead. Most approaches are supervised but getting annotated data is expensive. This lack of annotated datasets is often balanced using hybrid approaches that integrate external knowledge using sentiment lexicons, ontologies, or discourse parser features. Datasets are scarce in the news domain, and a recently \((\text{Hamborg and Donnay,})\) supplied a high-quality dataset for the task. A challenge in the field, as in many, is the development of frugal approaches that require less or no annotations, especially regarding the difficulty of transferring ABSC capabilities between domains \((\text{Nazir et al., 2020})\).

If aspect-based opinion mining can help derive an article’s stance, it cannot explain its underlying reasons. Argument Mining \((\text{see the survey (Lawrence and Reed, 2020)})\) fills this gap by extracting argumentation structures in texts. Improvements in irony and sarcasm detection, or even negation handling could also be of use.

4. **Towards a NLP-RS Hand-to-Hand Approach for Political NRS**

We highlighted some limits in both NRS and NLP fields, that show existing approaches to be unsuitable to reduce the impact or to burst filter bubbles, by ensuring a diversity of themes and opinions in NRS. First, the diversity measures used in RS mostly evaluate the news content differences, only ensuring that the set of recommendations are not too similar. This is due to the simple representations of articles used, which do not allow accurate representation of opinions, and are thus inadequate to ensure fair recommendations by inclusively representing diverse opinions. Recent lines of thought promote the plurality of opinions in NRS using representation metrics \((\text{Vrijenhoek et al., 2021})\).

Nevertheless, the notion of opinion is still treated basically in the form of a positive, negative or neutral but intangible opinion on a statement. A step further, we support the idea of a temporal diversification to adapt to the shifts in opinion and the evolving needs of users during the recommendation process.

Second, due to the highly dynamic nature of news, opinion mining techniques must be generic and adaptable to new sources and opinions. State-of-the-art aspect-based opinion analysis systems implement supervised machine learning algorithms and need annotated data to be trained. This hinders the capabilities of such systems to adapt quickly to new topics and their associated aspects. Semi-supervised approaches need to be further investigated to remedy this issue and to cope with the dynamic nature of news.

These limitations show the need for a stronger cooperation between RS and NLP communities. News opinion mining tasks have to be driven by the need of a specific diversity/similarity temporal evaluation. Recommendation tasks have to rely on precise opinion representations to enable content diversification.

Future developments resulting from such a cooperation need to be extensively evaluated. However, NRS evaluation is a difficult and sensitive task. Especially when it comes to opinionated content diversification. Evaluation protocols should include both qualitative and quantitative metrics to build a complete view of the NRS performances. Quantitatively, topic and aspect-based opinion extraction models need to be evaluated through standard benchmarks available in the NLP community \((\text{Hamborg et al., 2021})\). This step guarantees that news representations contain all necessary information for diversification. If several RS benchmarks exist in the community, none has been developed to quantitatively evaluate diversification as defined in this paper. Qualitatively, effectiveness of NRS in bursting opinion bubbles must be evaluated. Longitudinal studies, enrolling several groups of people with different social and cultural backgrounds, could shed light on the performance and acceptability of such a system through time, but also on the evolution of the opinions of people. However, these questions are not discussed in either domain and requires an expansion of collaboration with political science researchers in the study.

5. **Acknowledgements**

This work has been funded by the BOOM ANR Project - ANR-20-CE23-0024.


Conference on Big Data (Big Data), pages 551–560. IEEE.


7. Language Resource References


Notes

1http://ec.europa.eu/digital-agenda/
sites/digital-agenda/files/HLG%20Final
2https://www.newsguardtech.com/
3https://www.allsides.com/
4https://mediabiasfactcheck.com/