Proceedings of the First Workshop on Natural Language Processing for Political Sciences (PoliticalNLP), pages 79–85
Marseille, 24 June 2022

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[†]
[†] 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[1].

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 Požrl, 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 10; 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 preferences 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 tends 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., 2018; 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.

3.1.1. 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. (Boydstun 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 their cosine distances 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 congressman'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[2], AllSides[3], or Media Bias/Fact Check[4] 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_{ij}) 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., 2014; 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,).

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 (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 (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 (Hemmatian and Sohrabi, 2019 10; Nazir et al., 2020) 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 (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 (Nazir et al., 2020).

If aspect-based opinion mining can help derive an article's stance, it cannot explain its underlying reasons. Argument Mining (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 (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 on 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 (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.

6. Bibliographical References

Baly, R., Da San Martino, G., Glass, J., and Nakov, P. (2020). We can detect your bias: Predicting the political ideology of news articles. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4982–4991. Association for Computational Linguistics.

Bayram, U., Pestian, J., Santel, D., and Minai, A. A. (2019). What's in a word? detecting partisan affiliation from word use in congressional speeches. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.

Boydstun, A. E., Card, D., Gross, J. H., Resnik, P., and Smith, N. A. (2014). Tracking the development of media frames. *Work. Pap.*, pages 1–25.

Bozdag, E. and van den Hoven, J. (2015). Breaking the filter bubble: democracy and design. *Ethics and Information Technology*, 17(4):249–265.

Burbach, L., Halbach, P., Ziefle, M., and Calero Valdez, A. (2019). Bubble trouble: Strategies against filter bubbles in online social networks. In *Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. Healthcare Applications*, volume 11582, pages 441–456. Springer International Publishing.

D'Alonzo, S. and Tegmark, M. (2021). Machine-learning media bias. *arXiv:2109.00024 [cs]*.

Darwish, K., Stefanov, P., Aupetit, M., and Nakov, P. (2020). Unsupervised user stance detection on twitter. *Proceedings of the International AAAI Conference on Web and Social Media*, 14(1):141–152.

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4):51–58.

Ganguly, S., Kulshrestha, J., An, J., and Kwak, H. (2020). Empirical evaluation of three common assumptions in building political media bias datasets. *Proceedings of the International AAAI Conference on Web and Social Media*, 14(1):939–943.

Gao, Y., Zhao, H., Zhou, Q., Qiu, M., and Liu, M. (2020). An improved news recommendation algorithm based on text similarity. In *2020 3rd International Conference on Smart BlockChain (SmartBlock)*, pages 132–136. IEEE.

Gentzkow, M. and Shapiro, J. M. (2010). What drives media slant? evidence from u.s. daily newspapers. *Econometrica*.

Groseclose, T. and Milyo, J. (2005). A measure of media bias. *The Quarterly Journal of Economics*, 120(4):1191–1237.

Hamborg, F., Donnay, K., and Gipp, B. (2021). Towards target-dependent sentiment classification in news articles. In *Diversity, Divergence, Dialogue*, volume 12646, pages 156–166. Springer International Publishing.

Helberger, N. (2019). On the democratic role of news recommenders. *Digital Journalism*, 7(8):993–1012.

Hemmatian, F. and Sohrabi, M. K. (2019-10). A survey on classification techniques for opinion mining and sentiment analysis. *Artificial Intelligence Review*, 52(3):1495–1545.

Joseph, K. and Jiang, H. (2019). Content based news recommendation via shortest entity distance over knowledge graphs. In *Companion Proceedings of The 2019 World Wide Web Conference*, pages 690–699. ACM.

Kunaver, M. and Požrl, T. (2017). Diversity in recommender systems – a survey. *Knowledge-Based Systems*, 123:154–162.

Kwak, H., An, J., Jing, E., and Ahn, Y.-Y. (2021). FrameAxis: characterizing microframe bias and intensity with word embedding. *PeerJ Computer Science*, 7.

Lawrence, J. and Reed, C. (2020). Argument mining: A survey. *Computational Linguistics*, 45(4):765–818.

Liu, B. (2010). Sentiment analysis and subjectivity. In *Handbook of Natural Language Processing*, pages 627–666.

Lunardi, G. M., Machado, G. M., Maran, V., and de Oliveira, J. P. M. (2020). A metric for filter bubble measurement in recommender algorithms considering the news domain. *Applied Soft Computing*, 97.

Möller, J., Trilling, D., Helberger, N., and van Es, B. (2018). Do not blame it on the algorithm: an empirical assessment of multiple recommender systems and their impact on content diversity. *Information, Communication & Society*, 21(7):959–977.

Nakov, P., Sencar, H. T., An, J., and Kwak, H. (2021). A survey on predicting the factuality and the bias of news media. *arXiv:2103.12506 [cs]*.

Nazir, A., Rao, Y., Wu, L., and Sun, L. (2020). Issues and challenges of aspect-based sentiment analysis: A comprehensive survey. *IEEE Transactions on Affective Computing*.

Pariser, E. (2011). *The filter bubble: what the Internet is hiding from you*. Penguin Press.

Park, S., Kang, S., Chung, S., and Song, J. (2009). NewsCube: delivering multiple aspects of news to mitigate media bias. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 443–452. ACM.

Patankar, A., Bose, J., and Khanna, H. (2019). A bias aware news recommendation system. In *2019 IEEE 13th International Conference on Semantic Computing (ICSC)*, pages 232–238. IEEE.

Qiu, G., Liu, B., Bu, J., and Chen, C. (2011). Opinion word expansion and target extraction through double propagation. *Computational Linguistics*, 37(1):9–27.

Raza, S. and Ding, C. (2020). A regularized model to trade-off between accuracy and diversity in a news recommender system. In *2020 IEEE International*

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

Raza, S. and Ding, C. (2021). News recommender system: a review of recent progress, challenges, and opportunities. *Artificial Intelligence Review*, 55(1):749–800.

Recasens, M., Danescu-Niculescu-Mizil, C., and Jurafsky, D. (2013). Linguistic models for analyzing and detecting biased language. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1650–1659. Association for Computational Linguistics.

Riloff, E. and Wiebe, J. (2003). Learning extraction patterns for subjective expressions. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, pages 105–112.

Stefanov, P., Darwish, K., Atanasov, A., and Nakov, P. (2020). Predicting the topical stance and political leaning of media using tweets. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 527–537. Association for Computational Linguistics.

Sunstein, C. R. (2009). *Going to extremes: how like minds unite and divide*. Oxford University Press.

Tintarev, N., Sullivan, E., Guldin, D., Qiu, S., and Odjik, D. (2018). Same, same, but different: Algorithmic diversification of viewpoints in news. In *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization*, pages 7–13. ACM.

Titov, I. and McDonald, R. (2008). Modeling online reviews with multi-grain topic models. In *Proceeding of the 17th international conference on World Wide Web - WWW '08*. ACM Press.

Vrijenhoek, S., Kaya Independent Researcher, M., Metoui Delft, N. T., Möller, J., Odijk, D., and Helberger, N. (2021). Recommenders with a Mission: Assessing Diver-sity in News Recommendations. 11.

Wang, H., Zhang, F., Xie, X., and Guo, M. (2018). DKN: Deep knowledge-aware network for news recommendation. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18*, pages 1835–1844. ACM Press.

Wong, F. M. F., Tan, C. W., Sen, S., and Chiang, M. (2016). Quantifying political leaning from tweets, retweets, and retweeters. *IEEE Trans. Knowl. Data Eng.*, 28(8):2158–2172.

Wu, C., Wu, F., Qi, T., and Huang, Y. (2020). SentiRec: Sentiment diversity-aware neural news recommendation. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pages 44–53. Association for Computational Linguistics.

Zhang, X., Yang, Q., and Xu, D. (2021). Combining explicit entity graph with implicit text information for news recommendation. In *Companion Proceedings of the Web Conference 2021*, pages 412–416. ACM.

Zhou, T., Kuscsik, Z., Liu, J.-G., Medo, M., Wakeling, J. R., and Zhang, Y.-C. (2010). Solving the apparent diversity-accuracy dilemma of recommender systems. *Proceedings of the National Academy of Sciences*, 107(10):4511–4515.

Ziegler, C.-N., McNee, S. M., Konstan, J. A., and Lausen, G. (2005). Improving recommendation lists through topic diversification. In *Proceedings of the 14th international conference on World Wide Web - WWW '05*, page 22. ACM Press.

Zuiderveen Borgesius, F. J., Trilling, D., Möller, J., Bodó, B., de Vreese, C. H., and Helberger, N. (2016). Should we worry about filter bubbles? *Internet Policy Review*, 5(1).

7. Language Resource References

Card, D., Boydstun, A. E., Gross, J. H., Resnik, P., and Smith, N. A. (2015). The media frames corpus: Annotations of frames across issues. volume 2, pages 438–444.

Gaillat, T., Stearns, B., Sridhar, G., McDermott, R., Zarrouk, M., and Davis, B. (2018). Implicit and explicit aspect extraction in financial microblogs. In *Proceedings of the First Workshop on Economics and Natural Language Processing*, pages 55–61. Association for Computational Linguistics.

Hamborg, F. and Donnay, K.). NewsMTSC: A dataset for (multi-)target-dependent sentiment classification in political news articles. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 1663–1675. Association for Computational Linguistics.

Jangid, H., Singhal, S., Shah, R. R., and Zimmermann, R. (2018). Aspect-based financial sentiment analysis using deep learning. In *Companion of the The Web Conference 2018 on The Web Conference 2018 - WWW '18*, pages 1961–1966. ACM Press.

Miller, G. A. (1995). WordNet: a lexical database for english. volume 38, pages 39–41.

Pennington, J., Socher, R., and Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543. Association for Computational Linguistics.

Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., and Manandhar, S. (2014). SemEval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35. Association for Computational Linguistics.

Pontiki, M., Galanis, D., Papageorgiou, H., Manandhar, S., and Androutsopoulos, I. (2015). SemEval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 486–495. Association for Computational Linguistics.

Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., AL-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., De Clercq, O., Hoste, V., Apidianaki, M., Tannier, X., Loukachevitch, N., Kotelnikov, E., Bel, N., Jiménez-Zafra, S. M., and Eryiğit, G. (2016). SemEval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 19–30. Association for Computational Linguistics.

Steinberger, R., Hegele, S., Tanev, H., and Della Rocca, L. (2017). Large-scale news entity sentiment analysis. In *RANLP 2017 - Recent Advances in Natural Language Processing Meet Deep Learning*, pages 707–715. Incoma Ltd. Shoumen, Bulgaria.

Notes

[1] http://ec.europa.eu/digital-agenda/sites/digital-agenda/files/HLG%20Final

[2] https://www.newsguardtech.com/

[3] https://www.allsides.com/

[4] https://mediabiasfactcheck.com/