

Incorporating Editorial Feedback in the Evaluation of News Recommender Systems

BILAL MAHMOOD, MediaFutures, University of Bergen, Norway

MEHDI ELAHI, MediaFutures, University of Bergen, Norway

SAMIA TOUILEB, MediaFutures, University of Bergen, Norway

LUBOS STESKAL, TV 2, Norway

CHRISTOPH TRATTNER, MediaFutures, University of Bergen, Norway

Research in the recommender systems field typically applies a rather traditional evaluation methodology when assessing the quality of recommendations. This methodology heavily relies on incorporating different forms of user feedback (e.g., clicks) representing the specific needs and interests of the users. While this methodology may offer various benefits, it may fail to comprehensively project the complexities of certain application domains, such as the news domain. This domain is distinct from other domains primarily due to the strong influence of editorial control in the news delivery process. Incorporation of this role can profoundly impact how the relevance of news articles is measured when recommended to the users. Despite its critical importance, there appears to be a research gap in investigating the dynamics between the roles of editorial control and personalization in the community of recommender systems. In this paper, we address this gap by conducting experiments where the relevance of recommendations is assessed from an editorial perspective. We received a real-world dataset from TV 2, one of the largest editor-managed commercial media houses in Norway, which includes editors' feedback on how news articles are being related. In our experiment, we considered a scenario where algorithm-generated recommendations, using the K -Nearest Neighbor (KNN) model, employing various text embedding models to encode different sections of the news articles (e.g., title, lead title, body text, and full text), are compared against the editorial feedback. The results are promising, demonstrating the effectiveness of the recommendation in fulfilling the editorial prospects.

ACM Reference Format:

Bilal Mahmood, Mehdi Elahi, Samia Touileb, Lubos Steskal, and Christoph Trattner. 2024. Incorporating Editorial Feedback in the Evaluation of News Recommender Systems. In *Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation, and Personalization (UMAP Adjunct '24)*, July 1–4, 2024, Cagliari, Italy. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3631700.3664866>

1 INTRODUCTION & BACKGROUND

These days, finding the right news article to read has become a challenge for the users of news platforms due to the significant volume and velocity of news articles published online. News Recommender Systems (NRSs) are digital tools equipped with AI algorithms that can deal with this challenge by analyzing the content of the news articles clicked and read by the users and offering suggestions of relevant articles [10, 19, 26]. There are a variety of approaches to news recommender systems. Despite their differences in operation, they are typically tuned to primarily optimize for an “individual” objective, *i.e.*, the preferences of the users (e.g., clicks) [2, 13]. Hence, these approaches typically analyze the historical behavioral data (e.g., user clicks) and generate recommendations tailored to the preferences of the users [4, 6].

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

© 2024 Copyright held by the owner/author(s).

Manuscript submitted to ACM

While such a viewpoint might be suitable for the majority of the application domains, it might be too narrow-sighted for the news recommendation domain. This domain is distinct from others due to its unique characteristics, complexities, and multi-stakeholder nature [8, 15]. Users, as important stakeholders in this domain, have specific needs and preferences that must be satisfied through personalized recommendations. However, news organizations, another key stakeholder in this domain, have their own goals and values that need to be fulfilled [15]. While part of these goals involves business-oriented objectives, the responsibility towards society and the commitment to editorial values are also crucial part of them [7, 9]. An example can be the British Broadcasting Corporation (BBC), which has formulated a set of principles regarding the provision of news to represent the various cultures of their audience [3]. As a matter of fact, traditional editorially driven media are often considered one of the pillars of democracy, which leads to an increased sense of responsibility towards informing the reader about important societal issues [5]. Newsrooms are thus often hesitant to implement recommender systems, as they might be afraid that this would decrease the plurality of opinions and views one would be exposed to [24].

Prior studies in recommender systems have predominantly followed a traditional methodology that depends on user preferences to generate and evaluate recommendations for relevance [22, 23]. This approach builds user profiles on news platforms from historical data, such as clicks, with recommendations produced via automated algorithms and evaluated against the user preferences, serving as the ground truth. While this approach has its advantages, it may not fully address the complexities of real-world news recommender systems where editorship is crucial. Arguably, integrating editorial feedback into the evaluation could better align with news organizations' objectives, offering a more balanced perspective than solely relying on user preferences. This integration helps to ensure that recommendations adhere to the editorial norms and values critical in editor-managed news platforms [7, 9, 15, 18].

In addressing this research gap, we formulate the following research questions:

- **RQ1:** Do the news articles, which editors select as “related”, represent similarity based on their text embeddings?
- **RQ2:** How are the news recommendations generated based on the latest text embedding models evaluated with respect to editorial feedback as ground truth?

In this paper, we address these research questions by employing a methodology that evaluates recommendations from the perspectives of news editors. We received a real-world dataset from TV 2, one of the largest *editor-managed* commercial media houses in Norway, which includes editors' feedback on how news articles are related. We conducted an experiment considering a scenario where a content-based algorithm, *K*-Nearest Neighbor (KNN), generates recommendations using some of the state-of-the-art embedding models, *i.e.*, OpenAI, NB-SBERT, and NorBERT3.

These models are employed to separately encode different sections of the news articles (*e.g.*, title, lead title, and body text) and create their embedding vectors. In addition to that, all the article sections are concatenated to form the full text and embedded accordingly using these models. The outputs (*i.e.*, embedding vectors) were utilized by a classical content-based recommender algorithm for generating recommendations. These recommendations were then evaluated based on the editors' feedback (serving as ground truth), taking into account Recall@5, Precision@5, and MAP@5 (Mean Average Precision), which are among the common evaluation metrics in this research field. The results of the experiment were promising, demonstrating the effectiveness of the recommendations in fulfilling the editorial prospects. The main contributions of this paper are as follows:

- Compared to previous works [22, 23, 27], we employ a novel methodology for evaluating news recommender systems that utilizes the feedback provided by editors at one of the largest news platforms in Norway, *i.e.*, TV 2;

- We consider various state-of-the-art text embeddings to represent the content of news articles (in the Norwegian language) for the purpose of similar item recommendations. Furthermore, we cluster the news articles and visualized them using the t-SNE technique [25];
- We provide comparisons of these embedding models, using different news sections and evaluation metrics, *i.e.*, Precision@5, Recall@5, and MAP@5. We show that different embeddings can result in varying qualities while they can still maintain a good level of relevance from an editorial perspective.

2 METHODOLOGY

2.1 Dataset

We received a real-world dataset from TV 2, one of the largest editor-managed commercial media houses in Norway, which includes editors' feedback on how news articles are related. For each article published, the editor selects related articles that the reader should consider reading further. These articles are selected based on the judgment of the editors who think those articles are relevant to the given article. Not all the articles published have related articles, and thus are excluded from this study.

The dataset contains 21,701 news articles published from January 1st, 2018, to January 3rd, 2023. The data for each article includes the title, lead title (with an extended description), and body text. On average, each news article in the dataset had 2.17 related articles (median = 2). These articles were selected by editors for their relevance to the original article. We found that more than 80% of the related articles were published within 30 days of the original article's publication date, and more than 90% within the past year. We initially preprocessed the dataset and discovered that the data for a number of article IDs selected by editors were missing. Filtering out these articles resulted in a total of 14,282 articles. Subsequently, we employed various text embedding models to construct a Vector Space Model (VSM) representation of news articles [12, 14], where each article is modeled as a vector in a multi-dimensional space derived from the embedding models. The embedding vectors have been computed for the title, lead title, body text, and full text of the articles, separately. The full text of the articles has been constructed by concatenating the title, the lead title, and the body text.

Compared to the previous works, such as [22], we experimented with a set of the state-of-the-art pre-trained Large Language Models (LLMs) to extract embeddings for the news articles to compute similar item recommendations employing *Cosine* similarity [14, 17]. We started with NorBERT3 [20] (more precisely, the largest version dubbed NorBERT3_{large}¹), since it has been considered state-of-the-art in natural language understanding (NLU) tasks according to the ScandEval benchmark [16]. NorBERT3_{large} has 353M parameters and produces a 1024-dimensional vector representation of the text. Additionally, we explored the use of OpenAI's embedding model (also known as Text-Embedding-3-Small [1]), as a comparative multilingual model to generate vector representations. OpenAI's embedding model (at the time of writing this paper) provides a 1536-dimensional representation for a given text. Finally, we used a sentence transformers model, NB-SBERT_{base}², which maps text to 768-dimensional representation, and has shown promising results in a recent A/B testing [11]. Sentence transformers models enable the computation of dense vector representations for longer sequences of texts as, *e.g.*, sentences and paragraphs. The main advantage of such models is their ability to capture the semantics of the text sequences, enabling a more contextual representation of the text.

¹<https://huggingface.co/lgt/norbert3-large>

²<https://huggingface.co/NbAiLab/nb-sbert-base>

2.2 Recommendation Scenario

To generate the recommendation list of related news articles, we applied a classical content-based approach using the K -Nearest Neighbor algorithm, where $K = 5$. We employed an Information Retrieval (IR) methodology, assuming that the target user has clicked on and read an article $i \in I$, where I denotes the entire catalog of news articles. Then, given the news article i , we identified the 5 most similar articles in the catalog by computing the cosine similarity between the given article and all other articles in the catalog. This computation utilizes the embedding vectors of the article i and j , *i.e.*, E_i and E_j according to the following formula:

$$s_{ij} = \frac{E_i^T E_j}{\|E_i\| \|E_j\|} \quad (1)$$

where s_{ij} represents the cosine similarity between E_i (the embedding vector of the article clicked and read by the target user) and E_j (the embedding vector for each of the other articles in the catalog). We ranked the articles based on the similarity score, s_{ij} , to generate the recommendation list related to the article i . Hence, assuming the target user has clicked on and read that article, we provided recommendations of the most similar articles for her—a common personalization technique adopted by news platforms [11]. These recommendations were then evaluated against the editor-selected related articles as the ground truth. In particular, the top 5 articles were selected for recommendation and evaluated using the considered evaluation metrics, *i.e.*, Precision@5, Recall@5, and MAP@5 [21]. The rationale for choosing a recommendation size of 5 was that more than 98% of the articles, for which editors selected as related articles, had at most 5 related articles. It is worth noting that our initial exploratory analysis revealed that more than 90% of the articles selected by editors as related were published within 1 year of the publication date of the news article.

3 RESULTS

3.1 Exploratory Data Analysis

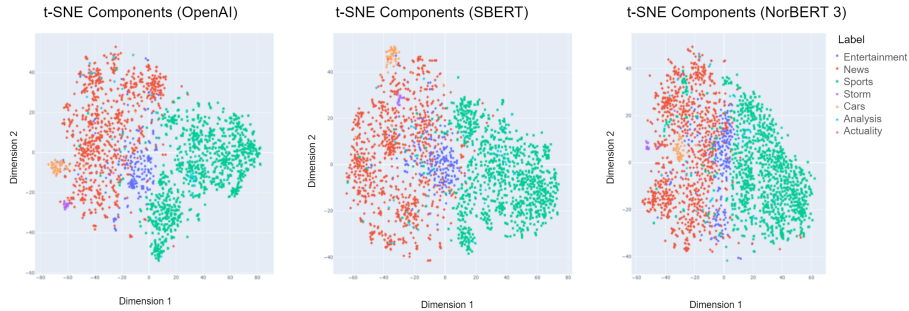


Fig. 1. Visualizing the high-dimensional vectors of news articles by applying a t-SNE dimensionality reduction technique.

Initially, we randomly sampled 1000 articles and conducted exploratory data analyses using a dimensionality reduction technique known as T-distributed Stochastic Neighbor Embedding (t-SNE) [25] to visualize the high-dimensional vectors from the embedding models.

The results are displayed in Figure 1, where each point represents the full text of the news article projected into 2D space. The colors correspond to the news articles’ topics—entertainment, news, sports, storm (weather), cars (automotive),

analysis, and actuality—with the distances between points indicating levels of (dis-)similarity. According to the figure, distinct clusters formed based on article topics, demonstrating that despite differences among the embedding models used to encode the news articles, all are effective in differentiating articles by topic. For instance, entertainment articles tend to cluster in the central areas across all models, while articles about cars appear more on the outer areas.

3.2 RQ1: Assessing the Similarities Among Editor-selected Articles

In addressing RQ1, we computed the similarity scores based on the embedding vectors of the news articles. For the sake of comparison, we also computed the similarities across the complete catalog of news articles. We wanted to investigate how closely the news articles, selected by editors as related, are similar concerning the embeddings computed for their title, lead title, body text, and full text. Due to space limitations, we only present the results based on the full text of the article.

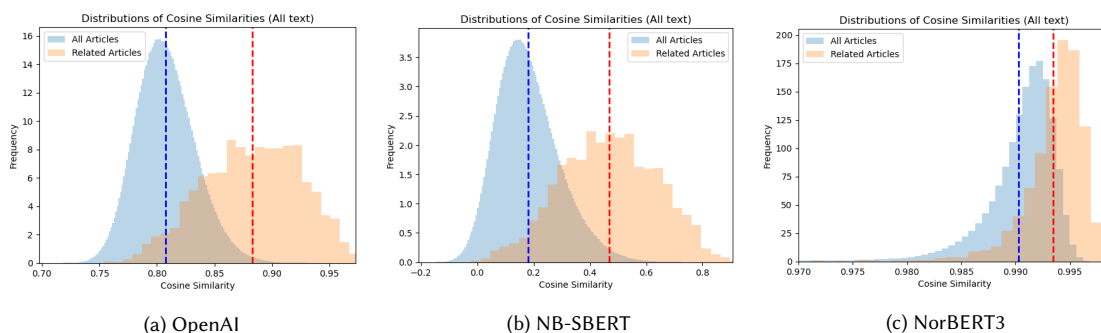


Fig. 2. Comparing the distribution of similarities among news articles computed based on different embedding models.

The results are depicted in Figure 2. It is evident that the similarities among editor-selected related articles (plotted in light orange color) are generally greater than those within the complete catalog of articles (plotted in light blue color). This can be apparent for each of the embedding models when comparing the corresponding vertical dashed lines, where the red lines indicate the average similarity score of editor-selected articles, and the blue lines indicate the average similarity score for the complete catalog of articles. A notable observation is that, for OpenAI and NB-SBERT, distribution curves resemble a tendency towards normal distributions, while for NorBERT3, the distribution curve is considerably skewed. A similar pattern can be observed for editor-selected articles.

To better present these differences, we computed statistics for these distributions, detailed in Table 1. For OpenAI, the average similarity for editor-selected articles is 0.883, compared to 0.808 for all articles, indicating a 9.2% higher similarity among related articles. The most significant difference is with NB-SBERT, where related articles show a 159.1% greater average similarity, with values of 0.469 versus 0.181 for the complete catalog of articles. For NorBERT3, the average similarities are closely matched at 0.993 for related articles and 0.990 for all articles, a marginal difference of 0.3%. These results suggest that editor-selected articles are, on average, more similar than the complete catalog of articles, as evidenced by the three embedding models.

Differences are more evident when analyzing similarity value ranges ($Max_{similarity} - Min_{similarity}$) across embedding models. NB-SBERT shows the largest range, exceeding 1, while OpenAI and NorBERT3 have narrower ranges of just over 0.2 and 0.1, respectively. All models reach a maximum similarity of 1, but minimum values vary significantly: they

Table 1. Statistics about distributions of similarities among related articles and the entire catalog of articles.

| Model | Articles | Min | Max | Range | Average |
|-----------------|----------|--------|-------|-------|---------|
| OpenAI | Related | 0.753 | 1.000 | 0.247 | 0.883 |
| | All | 0.697 | 0.972 | 0.275 | 0.808 |
| NB-SBERT | Related | -0.022 | 1.000 | 1.022 | 0.469 |
| | All | -0.207 | 0.908 | 1.115 | 0.181 |
| NorBERT3 | Related | 0.898 | 1.000 | 0.102 | 0.993 |
| | All | 0.874 | 0.999 | 0.125 | 0.990 |

are positive for OpenAI and NorBERT3, but unexpectedly negative for NB-SBERT, highlighting notable differences in how models process news content.

3.3 RQ2: Evaluating Recommendation based on Editors Feedback

In addressing RQ2, we conducted a set of experiments, the results of which are reported in Table 2. When comparing different embedding models, the best results were overall achieved by the OpenAI model, taking into account all of our metrics. Expectedly, using this model to create embeddings from the full text of news articles leads to substantially better results compared to embeddings created solely from the title and lead title. However, the outcomes of full text embeddings are similar to those obtained from body text embeddings alone. This suggests that although integrating more content from the news articles into the recommendation process enhances overall performance, the body text of the news articles can achieve a similar effectiveness compared to their full text.

Table 2. Results of evaluating the quality of content-based recommendation based on different embedding models considering the editorial feedback as ground truth (best values are highlighted in boldface).

| News Section | Embedding Model | Recall@5 | Precision@5 | MAP@5 |
|---|-----------------|--------------|--------------|--------------|
| Title | OpenAI | 0.057 | 0.018 | 0.052 |
| | NB-SBERT | 0.044 | 0.014 | 0.041 |
| | NorBERT3 | 0.019 | 0.006 | 0.019 |
| Lead title | OpenAI | 0.129 | 0.041 | 0.120 |
| | NB-SBERT | 0.098 | 0.031 | 0.092 |
| | NorBERT3 | 0.050 | 0.016 | 0.050 |
| Body text | OpenAI | 0.260 | 0.083 | 0.239 |
| | NB-SBERT | 0.148 | 0.047 | 0.140 |
| | NorBERT3 | 0.134 | 0.043 | 0.134 |
| Full text | OpenAI | 0.262 | 0.083 | 0.239 |
| | NB-SBERT | 0.166 | 0.052 | 0.153 |
| | NorBERT3 | 0.148 | 0.047 | 0.144 |
| Random recommendation (baseline) | | <0.001 | <0.001 | <0.001 |

In terms of Recall@5, the OpenAI model achieves a value of 0.262 when the embedding of the full text is used for recommendation. This essentially means that, on average, more than 26% of the editor-selected articles are included in the recommendations. This can be considered a promising result, especially taking into account the simplicity of the recommendation approach. For the other embedding models, there was a considerable reduction in the Recall@5, with values of 0.166 and 0.148 for encoding the full text based on NB-SBERT and NorBERT3 models, respectively. As expected, as more textual information in the news article is embedded, the better the Recall@5 metric gets.

In terms of Precision@5, the OpenAI embedding again performs the best among all the experimented embeddings, achieving a score of 0.083 when full text in the news articles is considered. Interestingly, the same result is observed for body text of the news articles. For the NB-SBERT and NorBERT3 models, the results are 0.052 and 0.047, respectively when the full text is considered. These values can be explained highlighting the fact that a considerable proportion of articles have only one or two editor-selected (related) articles. This insight can significantly limit the ideal values of Precision@5 to low levels, implying the effectiveness of the approach.

Considering the ranking quality, measured in terms of MAP@5, a similar pattern is observed, with the OpenAI embedding of both the full text and body text performing the best, each achieving the MAP@5 value of 0.239. For the NB-SBERT and NorBERT3 models, again, encoding the full text showed the best results, with values of 0.153 and 0.144, respectively. For the comparison, we also considered a random recommendation baseline, which consistently obtained scores of less than 0.001 for all metrics.

In a follow-up experiment, we extended the results of the overall best-performing model—OpenAI’s full-text embeddings—by applying different time filters. Specifically, we analyzed the performance of recommendations based on this model when the potential related articles were selected within 1, 7, 14, and 30 day(s) prior to the publication date of the article clicked on and read by a target user. We observed the best outcomes with a 14-day filter, noting improvements of 20.2% in Recall@5, 19.2% in Precision@5, and 18.8% in MAP@5 compared to the original 1-year time filter, previously reported. The observed results were 0.315 for Recall@5, 0.099 for Precision@5, and 0.284 for MAP@5. While these results are promising and highlight the importance of time sensitivity analysis, further investigation will be necessary.

4 DISCUSSION AND CONCLUSION

In this paper, we employed a novel methodology to evaluate the output of a news recommendation approach by utilizing the editorial feedback as a ground truth instead of the user feedback. We received a real-world dataset from TV 2, one of the largest editor-managed commercial media houses in Norway. We considered a set of state-of-the-art text embedding models to represent the news articles. The results of the experiment are promising, demonstrating the effectiveness of our approach in fulfilling editorial prospects.

We would like to acknowledge the potential limitations of our work. First, we considered a classical content-based algorithm for generating recommendations based on the similarity of news articles. While this approach might be simple, its effectiveness in the news domain has been evidenced in news websites in prior studies, including a recent A/B testing in a real-world setup [11]. Additionally, our focus in this paper has been on exploring the editorial perspective in the evaluation of recommendations, rather than on the algorithmic aspect. However, considering a broader range of recommender algorithms with different characteristics could enrich the analyses with a more extensive comparison.

It is important to note that our experimental design may deviate from traditional methodologies adopted in the recommender systems community. Our experiments follow an Information Retrieval (IR) centered methodology, considering relevance feedback from an editorial perspective, as a complementary signal rather than a substitution for user feedback. Regardless, in alignment with our devised methodology, we highlight the importance and necessity of personalizing recommendations according to the specific needs and interests of the end users. In future work, we plan to extend our current analysis by conducting a user study to evaluate a hybrid recommendation approach that can incorporate both user and editorial feedback and potentially enhance users’ experience and trust in the news domain.

ACKNOWLEDGMENTS

This work was in part supported by the Research Council of Norway (RCN) with funding to MediaFutures: Research Centre for Responsible Media Technology and Innovation, through the Centre for Research-based Innovation scheme, project number 309339.

REFERENCES

- [1] [n. d.]. OpenAI embedding model. <https://platform.openai.com/docs/guides/embeddings/what-are-embeddings>. Accessed: 2023-8-15.
- [2] Himan Abdollahpouri, Gediminas Adomavicius, Robin Burke, Ido Guy, Dietmar Jannach, Toshihiro Kamishima, Jan Krasnobebski, and Luiz Pizzato. 2020. Multistakeholder recommendation: Survey and research directions. *User Modeling and User-Adapted Interaction* 30 (2020), 127–158.
- [3] BBC. 2019. Mission, values and public purposes - About the BBC. <https://www.bbc.com/aboutthebbc/governance/mission>.
- [4] Amin Beheshti, Samira Ghodrattnama, Mehdi Elahi, and Helia Farhood. 2022. *Social Data Analytics*. CRC Press.
- [5] Elda BROGI, Daniella BORGES, Roberta CARLINI, Iva NENADIC, Konrad BLEYER-SIMON, Jan Erik KERMER, Urbano REVIGLIO DELLA VENARIA, Matteo TREVISAN, and Sofia VERZA. 2023. The European Media Freedom Act: media freedom, freedom of expression and pluralism.
- [6] Mehdi Elahi, Matthias Braunhofer, Tural Gurbanov, and Francesco Ricci. 2018. User Preference Elicitation, Rating Sparsity and Cold Start.
- [7] Mehdi Elahi, Dietmar Jannach, Lars Skjærven, Erik Knudsen, Helle Sjøvaag, Kristian Tolonen, Øyvind Holmstad, Igor Pipkin, Eivind Throndsen, Agnes Stenbom, et al. 2022. Towards responsible media recommendation. *AI and Ethics* (2022), 1–12.
- [8] Alireza Gharahighehi, Celine Vens, and Konstantinos Pliakos. 2021. Fair multi-stakeholder news recommender system with hypergraph ranking. *Information Processing & Management* 58, 5 (2021), 102663.
- [9] Natali Helberger. 2021. On the democratic role of news recommenders. In *Algorithms, Automation, and News*. Routledge, 14–33.
- [10] Mozghan Karimi, Dietmar Jannach, and Michael Jugovac. 2018. News recommender systems—Survey and roads ahead. *Information Processing & Management* 54, 6 (2018), 1203–1227.
- [11] Peter Kolbeinsen Klingenberg. 2023. Using content- and behavioural data for recommendations in the Norwegian news market.
- [12] Balaji Krishnamurthy, Nikaash Puri, and Raghavender Goel. 2016. Learning vector-space representations of items for recommendations using word embedding models. *Procedia Computer Science* 80 (2016), 2205–2210.
- [13] Jiahui Liu, Peter Dolan, and Elin Ronby Pedersen. 2010. Personalized news recommendation based on click behavior. In *Proceedings of the 15th International Con Intelligent User Interfaces*. 31–40.
- [14] Pasquale Lops, Marco De Gemmis, and Giovanni Semeraro. 2011. Content-based recommender systems: State of the art and trends. *Recommender systems handbook* (2011), 73–105.
- [15] Feng Lu, Anca Dumitrache, and David Graus. 2020. Beyond optimizing for clicks: Incorporating editorial values in news recommendation. In *Proceedings of the 28th ACM conference on user modeling, adaptation and personalization*. 145–153.
- [16] Dan Saatstrup Nielsen. 2023. Scandeval: A benchmark for Scandinavian natural language processing. *arXiv preprint arXiv:2304.00906* (2023).
- [17] Michael J. Pazzani and Daniel Billsus. 2007. *The Adaptive Web*. Springer-Verlag, Berlin, Heidelberg, Chapter Content-based Recommendation Systems, 325–341. <http://dl.acm.org/citation.cfm?id=1768197.1768209>
- [18] Christian Peukert, Ananya Sen, and Jörg Claussen. 2023. The editor and the algorithm: Recommendation technology in online news. *Management science* (2023).
- [19] Shaina Raza and Chen Ding. 2022. News recommender system: a review of recent progress, challenges, and opportunities. *Artificial Intelligence Review* (2022), 1–52.
- [20] David Samuel, Andrey Kutuzov, Samia Touileb, Erik Vellidal, Lilja Øvrelid, Egil Rønningstad, Elina Sigdel, and Anna Palatkina. 2023. NorBench – A Benchmark for Norwegian Language Models. In *Proceedings of the 24th Nordic Conference on Computational Linguistics (NoDaLiDa)*, Tanel Alumäe and Mark Fishel (Eds.). University of Tartu Library, Tórshavn, Faroe Islands, 618–633. <https://aclanthology.org/2023.nodalida-1.61>
- [21] Markus Schedl, Hamed Zamani, Ching-Wei Chen, Yashar Deldjoo, and Mehdi Elahi. 2018. Current challenges and visions in music recommender systems research. *International Journal of Multimedia Information Retrieval* 7 (2018), 95–116.
- [22] Alain D Starke, Sebastian Øverhaug Larsen, and Christoph Trattner. 2021. Predicting feature-based similarity in the news domain using human judgments. In *Proceedings of the 9th International Workshop on News Recommendation and Analytics (INRA 2021) co-located with 15th ACM Conference on Recommender Systems (RecSys 2021)*.
- [23] Christoph Trattner and Dietmar Jannach. 2020. Learning to recommend similar items from human judgments. *User Modeling and User-Adapted Interaction* 30 (2020), 1–49.
- [24] Hilde Van den Bulck and Hallvard Moe. 2018. Public service media, universality and personalisation through algorithms: mapping strategies and exploring dilemmas. *Media, culture & society* 40, 6 (2018), 875–892.
- [25] Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9, 11 (2008).
- [26] Chuhan Wu, Fangzhao Wu, Yongfeng Huang, and Xing Xie. 2023. Personalized news recommendation: Methods and challenges. *ACM Transactions on Information Systems* 41, 1 (2023), 1–50.

- [27] Yuan Yao and F Maxwell Harper. 2018. Judging similarity: a user-centric study of related item recommendations. In *Proceedings of the 12th ACM Conference on Recommender Systems*. 288–296.