NPR: a News Portal Recommendations dataset

Joel Pinho Lucas[1,*,†], João Felipe Guedes da Silva[1,2,†] and Letícia Freire Figueiredo[1,3,†]

[1]Grupo Globo, Av. das Américas, 1650, Rio de Janeiro, Brazil
[2]Universidade Federal do Rio de Janeiro, Av Athos da Silveira Ramos, 149 - Cidade Universitária - CT Bloco H sala 220, Rio de Janeiro, Brazil
[3]Universidade Federal Fluminense, Av Gal. Milton Tavares de Souza, s/nº - São Domingos, Niterói, RJ, Brazil

Abstract

Recommender systems have become key applications for news websites to filter relevant articles to users among an ever-growing catalog. However, building such applications brought challenges yet to be solved like filter bubbles and addressing diversity. In this way, publicly available datasets play a central role in solving these problems since they bring both academic and industrial researchers to a common ground for proposing new solutions. Yet, not only are news recommendation datasets scarce but also most of them lack the necessary content for research towards news diversity. In this paper, we introduce the News Portal Recommendations (NPR) dataset for news recommendation. NPR is an improvement of a previously published dataset, which lacked the information needed for normative diversity analysis. In this sense, we make use of the RADio framework in order to calculate diversity metrics on the dataset. Differently from other publicly available data, such as the MIND dataset, in this work, we are focusing on data tracked from frequent user interactions in hard news (i.e. users with more interactions with the portal). The NPR dataset is available in a Kaggle repository[1].

Keywords

Public dataset, News recommendations, Normative diversity

1. Introduction

News portals provide content to millions of users in current days, from topics like sports to politics. With such a wide range of themes and a massive number of possible articles to read, news recommender systems play a central role in filtering which items are more suited for a specific user at a given time [1]. However, several challenges still need to be overcome when building such systems, both in the societal and technical domains.

The definition of what is suited for a user is somewhat relative. Some may design news recommender systems to optimize for user engagement, which should lead to higher click rates or reading time on a platform [2, 3]. However, in this scenario, news recommender may not be

[1]https://kaggle.com/datasets/joelpl/news-portal-recommendations-npr-by-globo

NORMalize 2023: The First Workshop on the Normative Design and Evaluation of Recommender Systems, September 19, 2023, co-located with the ACM Conference on Recommender Systems 2023 (RecSys 2023), Singapore

*Corresponding author.
†These authors contributed equally.
joel.pinho@g.globo (J. P. Lucas); joao.guedes@g.globo (J. F. G. d. Silva); leticia.freire@g.globo (L. F. Figueiredo)
https://www.linkedin.com/in/joelplucas (J. P. Lucas); https://www.linkedin.com/in/joao-felipe-guedes/ (J. F. G. d. Silva); https://www.linkedin.com/in/le-freire/ (L. F. Figueiredo)
0000-0002-6789-1376 (J. P. Lucas); 0000-0001-5496-7936 (J. F. G. d. Silva); 0000-0001-7613-7423 (L. F. Figueiredo)

suited to keep the users informed on relevant aspects of society other than those they are more leaned to consume, which raises concerns about the democratic roles of these systems [4].

This algorithmic influence is amplified due to the fact that readers tend to engage more on contents that confirm their own worldview [5, 6]. This phenomenon prompts recommenders to limit the diversity of suggested items, potentially leading to user segregation and biased opinions [7, 8].

Other than these societal issues, technical challenges still need to be addressed on news recommender. As new items are added every minute with fresh information, old items are inactivated for recommendation, yielding a short item shelf life [9]. As a consequence, traditional user-item matrix used by algorithms are commonly very sparse, which sets additional challenges to model user's preferences [2, 3]. This scenario is aggravated with anonymous users who usually have few past interactions in the system [10].

Besides these sparsity challenges, news recommenders heavily rely on rich feature engineering to represent items and model users' consumption from previous behaviors [11, 2]. Although simpler forms of metadata can be used, such as news article categories, representing items from its textual content requires applying complex natural language processing techniques to the article's title or body content [12, 13].

Creating solutions to these issues with news recommender systems requires contributions from both industrial and academic players. In this scenario, proper datasets need to be publicly available for researchers to discuss and explore solutions on a common ground. Several benchmarks have been proposed so far and each of them has had its own share of contribution towards research development.

However, most datasets so far lack the proper information structure for research on news recommender to be properly conducted [14]. To the best of our knowledge, the most suited released so far is the MIND [11] dataset, which has enabled several works to be published towards technical challenges[15, 16, 17]. Nonetheless, even this last benchmark still has its own limitations when it comes to normative diversity research [1].

In order to fill the gaps in previously published datasets, this paper introduces the News Portal Recommendations dataset, an improved version of a past dataset that aims to provide the necessary information for research on normative diversity in news recommender systems. Therefore, it is structured as follows.

Section 2 revisits related works published in the past related to public datasets used for research development on news recommender systems. Then Section 3 describes how the proposed dataset was constructed and what are, as well as its main characteristics that contribute to bridging the gap of past datasets. Later, Section 4 brings normative diversity metrics from the proposed dataset and, finally, all results are discussed and concluded in Section 5.

2. Related Works

In the past years, a few datasets have been made public to foster research in news recommenders. A dataset from *Globo.com* [18] was built by sampling user interactions from G1[1], a Brazilian news portal. It contains $3M$ records distributed in $46k$ news articles and $314k$ users extracted

[1]https://g1.globo.com/

from October 1 to October 16, 2017. However, instead of having text information from the articles (like its title or body content), it contains the article's word embeddings generated by a neural model trained on classification tasks[18], which considerably limited the use of recent natural language processing tools and other types of diversity-oriented explorations.

The *MIcrosoft News Dataset* (MIND) [11] dataset was later published, providing news' textual information as metadata for 161k news items, MIND contains 24.1M logs for 1M randomly sampled users from Microsoft News[2] who had at least 5 clicks in the period between October 12 and November 22, 2019. In addition, the dataset is associated with a public competition[3] in which the goal was to predict the click scores of candidate news based on user interests.

A few other datasets were published before the aforementioned [14]. *Plista* [19] contains activity logs from 13 German news portals, recorded in June 2013 by $\approx 1M$ sampled records to $\approx 70k$ items. *Adressa* [20] included 27M click interactions from 3M users to 48k news articles, extracted in a ten weeks period from Adresseavisen[4]. However, each of these datasets has its own limitations like size or lack of metadata, as thoroughly described in the MIND original article [11].

Considering these aforementioned datasets, MIND became a reference benchmark due to its size and textual components. Nonetheless, despite its contributions, some of its limitations towards recommender diversity were brought to light.

Firstly, as it contains a considerable amount of soft news, it may compromise research in normative diversity metrics which are more tailored towards the so-called "hard" news [1]. Secondly, the dataset is split among training, validation, and test sets, in which the validation set only contains data from November 15, 2019. In this case, it is unlikely how the users and the recommender's behavior towards those users change over time [21]. Finally, nearly half of the anonymous user IDs have unique visits, which makes it unlikely to model the long-term effects of the recommender diversity on user consumption.

In order to improve the research possibilities on news recommendations towards normative diversity, this article proposes the News Portal Recommendations (NPR) dataset, a restructured version of a previous dataset aiming to provide a qualified dataset for research purposes. In particular, we list the following main contributions compared to the MIND dataset:

- Focus on hard news
- Ranked recommendation lists
- Distinction between logged and returning users from anonymous ones
- Longer periods of data

3. News Portal Recommendations Dataset

3.1. Dataset Construction

The NPR dataset was built by sampling users from G1[5], the largest Brazilian news portal maintained by the Globo media company. It contains 1162802 randomly sampled users who

[2] https://microsoftnews.msn.com
[3] https://msnews.github.io/competition.html
[4] https://reclab.idi.ntnu.no/dataset/
[5] https://g1.globo.com/

received recommendations in the period between January 3rd, 2023, and May 1st, 2023, where nearly 73% are non-logged users. All users were anonymized in order to protect data privacy. NPR was developed following the same structure as the MIND [11] dataset. Therefore, it is composed of the following files: behaviors, articles, and impressions.

The **behaviors** file contains 1402576 impressions logs regarding which sequence of items was recommended at a given time and which items users consumed before receiving such recommendations. Unlike the previous dataset, NPR also includes statistics on user behavior regarding the articles' page, such as the number of clicks, time spent on the page, and scroll percentage.

The **articles** file contains metadata for 148099 news items that were either consumed or recommended to users in the previous behaviors file. It contains news URLs, their title text, and a list of topics associated with each article assigned by a specialized editorial board. The complete schema for the article's file is displayed in the dataset's repository in Kaggle.

Finally, the **recommendations** contains three files on 92700 randomly sampled recommendations generated by means of the following algorithmic approaches: Collaborative Filtering, the most Recent publications, and the Top consumed articles. Each file refers to recommendations provided by one specific algorithm, but all files share the same impression IDs and, as a consequence, refer to the same users.

3.2. Dataset Analysis

Since textual information plays a central role in news datasets, Figure 1 displays the distribution from some of the article's features.

The top left and top right plots already show a language distinction between the NPR and MIND datasets. In terms of the number of words in the article's titles, NPR has an average of 14.9 words while its counterpart presents 11.52 [11]. However, the biggest difference is in the article's body length. While NPR has a single-modal and skewed distribution, with articles having 471.7 words in their body on average, MIND's body length is multimodal, with averages around 20 and 80 [11]. This indicates that NPR contributes to much richer textual information to be assessed with natural language processing techniques.

In addition, most news articles are associated with a topic assigned manually by the editorial board, which consists of multiple teams spread out in different geographic regions of the country. Such scenario potentially results in non-uniform categorization of news articles. Solutions to address this challenge are further discussed in the Future Work section. From a total of 94 topics (bottom left plot in Figure 1), most of the news is related to *sp*, *mg*, and *rj*, which are acronyms for Brazil's states. Since these are some of the most populated states in Brazil, it indicates the predominance of regional content. In fact, the topic's distribution is so unequal that it reaches a 74% Gini index [22] of distribution inequality. Other generic themes like "*mundo*" (world), "política" (politics), and "economia" (economy) also have a significant share of news articles.

Based on these topics, news articles can be associated with hard and soft news. As explored by Vrijenhoek [1], the MIND dataset has a higher share of soft news, which may not be the best scenario for research on normative diversity. In order to evaluate the differences in those news types, Table 1 shows comparisons between NPR and MIND datasets.

Considering all articles in the catalog, NPR presents 91.0% of hard news items while MIND

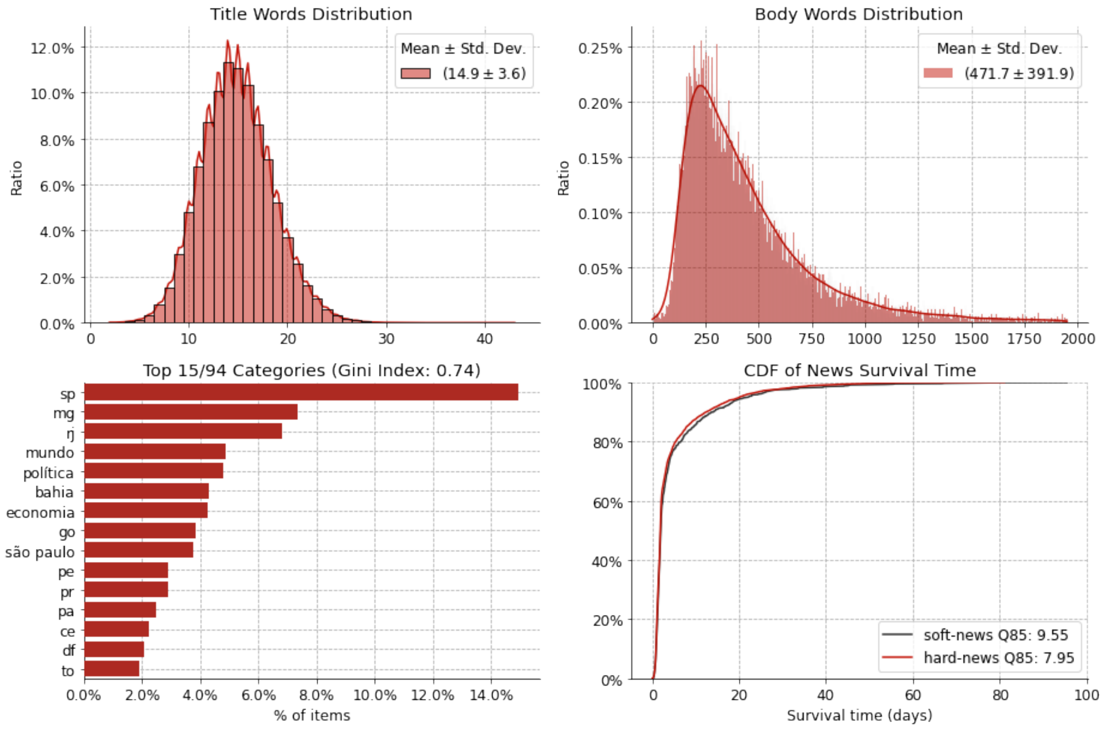

Figure 1: (top left) Title number of words. **(top right)** Body number of words. **(bottom left)** Top 15 out of 94 most common articles' categories. **(bottom right)** Cumulative Distribution Function (CDF) of hard/soft news survival time in days and their 85% quantile.

Dataset	News Type	All	History	Clicked	Candidate
MIND	Soft News	63.6%	62.2%	73.0%	69.8%
	Hard News	36.3%	34.8%	26.9%	30.2%
NPR	**Soft News**	**9.0%**	**17.8%**	**13.8%**	**15.4%**
	Hard News	**91.0%**	**82.2%**	**86.2%**	**84.6%**

Table 1
Share of news types between NPR and MIND datasets (MIND numbers extracted from [1]).

has 36.3%. This distinction expands to other aspects of the datasets such as the user's historical consumption (82.2% on NPR against 34.8% on MIND), clicked items (86.2% against 26.9%) and consumption candidate lists, i.e., recommendations (84.6% against 30.2%).

Finally, the news in the G1 ecosystem has a short survival time. It can be observed in Figure 1, where the consumption of the news is more concentrated on the first days after the publication. In this case, more recent news is more consumed.

Despite having a higher share of hard news, items on NPR seem to last longer than the ones presented in MIND. Figure 1 shows on the bottom right plot the cumulative distribution function (CDF) of news survival time (number of days between the article's publish date and last click). It can be seen that 85% of items are clicked up to 7.95 days for hard news and 9.55 days for soft news, which reinforces the characteristic of soft news to last longer.

4. Normative Diversity

4.1. Theoretical Background

As aforementioned, news recommender systems play a central democratic role in keeping users informed by unlocking the diversity of online information [4, 8]. However, the definition of diversity is plural, especially when contrasting the fields of computer science and normative literature. For instance, while technical metrics such as intra-list distance of recommended items [23, 24] or gini index [25, 26] may be a proxy to diversity in computer science, normative literature might lean towards concepts of democracy, freedom of expression and cultural inclusion [27, 4].

To bridge the gap between technical and normative literature, a framework by the name of RADio (*Rank-Aware Divergence Metrics*) [28] has been proposed to translate normative goals into a set of quantifiable metrics grounded in democratic theory. The framework works under five metrics which are summarized as follows (for a thorough description of the metrics, refer to [28, 27]):

- *Calibration*: assesses the degree to which the issued recommendations align with the user's preferences. The further from 0, the greater the deviation from the user's preferences.
- *Fragmentation*: quantifies the level of overlap among recommendations presented to distinct users. The closer to 0, the greater the overlap.
- *Activation*: gauges the extent to which the issued recommendations aim to motivate users into action. The closer to 0, the more neutral the content.
- *Representation*: indicates how different opinions or perspectives are expressed. The closer to 0, the more balanced the content whereas higher scores measure larger discrepancies.
- *Alternative Voices*: measures to which extent minority groups are represented in the content. The closer to 0, the fewer the presence of minority voices.

Based on the different values extracted from a news recommender for these five metrics, it can be assigned to four democratic models described by Helberger [4]: liberal, participatory, deliberative, and critical. A reference table overview for each model is documented in [28, 27].

4.2. Experimental RADio Metrics

All five RADio metrics aforementioned were applied to the MIND dataset for 6 different algorithms [28]. However, given that some of these metrics rely on applying natural language processing techniques to extract aspects such as entity recognition of minority voices or content neutrality, we focus the analysis on the *calibration* metric.

The NPR dataset contains recommendations for three different kinds of algorithms. The first is an Alternating Least Square ("ALS") strategy, which is a classical recommendation algorithm based on the factorization of the user-item matrix [29]. The other two algorithms are the "Top" algorithm, which recommends the most consumed news articles from the past 48 hours, and "Recents", which recommends the most recent news articles published by the editorial board. Both of these later algorithms are non-personalized, meaning that all users who access the news portal at a given time receive the same recommendation.

Considering these three algorithms, Figure 2 provides two plots generated after extracting the calibration metrics on different recommendation scenarios. The left plot shows how calibration is distributed among three different algorithms after recommending 5 items, whereas the right plot shows the average calibration considering multiple recommendation list sizes.

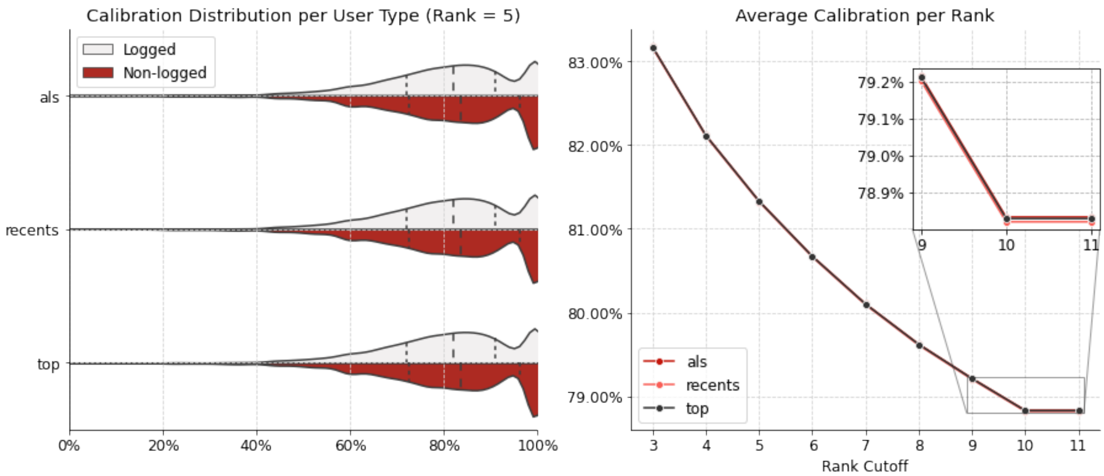

Figure 2: (left) Calibration distribution per user type for a 5-items recommendation. Dashed lines on the violin plots represent quantiles 25%, 50%, and 75%. **(right)** Average calibration considering multiple recommendation list sizes.

At first glance, no major differences can be observed between algorithms. A first hypothesis for such an overlap is that the "ALS" algorithm may converge to the "Top" or "Recents" strategy, especially when considering hard news with a short life span. Additionally, the dynamic behavior of hard news might make it more difficult to model user preferences, since users are likely to change their interests rapidly. For instance, users may consume distinct categories of hard news due to the fact that they are breaking news, yielding a more general type of consumption profile. This scenario can be addressed by more robust algorithms that are more suited for news recommendation, which are already in place in Globo portals. We discuss their use in the scope of this dataset later in the future work section.

However, the plot shows a noticeable difference when comparing recommendations between logged versus non-logged users. Recall that as calibration approaches 0, recommendations are more tailored towards the user's preferences. Since logged users tend to have more historical data, it is reasonable to see lower calibration values when compared to non-logged users, which can be seen by comparing the distributions' quantiles.

By expanding the analysis to ranks different than 5, the right plot in Figure 2 provides how the average calibration changes according to different recommendation list sizes. For longer lists, it is more probable to find items tailored to the users' preferences. Therefore, it is reasonable to observe a descending calibration variation as recommendation lists get larger. Since NPR contains up to 10 items in the recommendation lists, a lower limit of calibration can be observed around 78.9%.

Based on this lower limit, we can establish a calibration comparison between MIND and NPR using the results reported for the "top" algorithm in [28] (referred to as the "most popular"). By recommending a list of 10 items to the users, a calibration of 65.3% was observed after using news article topics. Therefore, we can observe that even a non-personalized algorithm may present significantly different calibration results depending on the dataset, which reinforces the need for several datasets to be employed as benchmarks when analyzing the diversity capabilities of a recommendation algorithm.

Besides the calibration analysis, Figure 3 also shows preliminary results on the representation (left plot) and fragmentation (right plot) metrics.

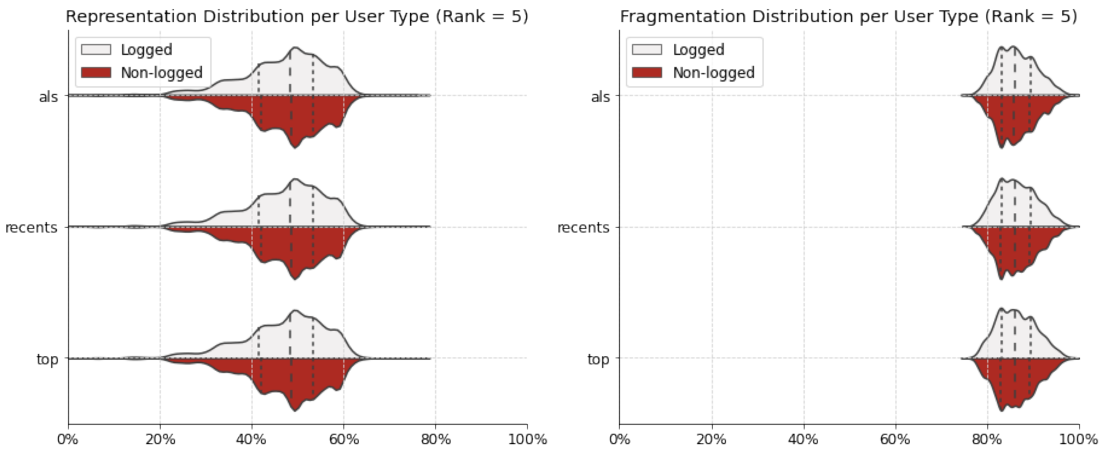

Figure 3: Representation **(left)** and Fragmentation **(right)** distributions per user type for a 5-items recommendation. Dashed lines on the violin plots represent quantiles 25%, 50%, and 75%.

It can also be seen from Figure 3 that no clear difference is observed among algorithms, which reinforces the aforementioned hypothesis that the "ALS" algorithm converges to the "Top" approach. Notice that, for all algorithms, the average representation is close to 50%, whereas average fragmentation approaches 85%. A hypothesis for such behavior is the lack of proper natural language processing tools towards Portuguese texts which better indicates how different opinions or perspectives are expressed.

5. Conclusion

Taking into account the need for proper datasets to be publicly available for both Academy and Industry researchers to discuss and explore solutions on common ground, this paper introduced the News Portal Recommendations (NPR) dataset. The dataset provides data on recommendation impressions, user behavior (consumption history), and also metadata about published articles.

Section 3 analyzed specific characteristics of the dataset and also compared it with the MIND dataset from Microsoft News [11]. Other related datasets are also described. Besides providing much richer textual information in comparison with MIND, NPR has a considerably greater proportion of hard news consumption than the MIND dataset. Subsequently, section 4 showed that the NPR dataset could be applied to the RADio framework [28], translating normative goals into quantifiable metrics.

The first version of the dataset is already publicly available, opening the horizon for continuous updates and improvements based on feedback from the community. Some improvements are already planned for Future Work.

6. Future Work

When looking at news categories, the NPR dataset also presents acronyms for Brazillian states, indicating the predominance of regional content. As stated in section 3.2, the need for manual tagging potentially results in non-uniform categories. In this sense, we are currently developing automatic extraction of semantic metadata from news articles, which will enrich the current categories already in place in the dataset. In this context, we also aim to explore content-representation techniques in order to remove any possible differences resulting from the Portuguese language.

Finally, in addition to the ALS algorithm, we will also incorporate recommendation impressions resulting from other, and more advanced, personalization algorithms. Although other algorithms are already being employed for providing recommendations to the final user, engineering efforts are needed to extract multiple algorithms' outputs to the same users due to Globo's AB platform. Since in Globo any information delivered to the final user is subjected to an AB test, the recommendation algorithm employed, as well as its resulting impressions, will vary depending on the AB testing alternative that has been employed for that specific user.

Acknowledgments

Thanks to Mateo Gutierrez Granada, Johannes Kruse, and Gabriel Benedict for implementing the code that made it possible to run the RADio metrics on Globo's dataset. We would also like to acknowledge Globo for providing this dataset for the academic community, especially to the Recommendation team for preparing the original dataset from the G1 Portal.

References

[1] S. Vrijenhoek, Do you mind? reflections on the mind dataset for research on diversity in news recommendations, 2023. doi:doi.org/10.48550/arXiv.2304.08253. arXiv:2304.08253.

[2] J. Liu, P. Dolan, E. R. Pedersen, Personalized news recommendation based on click behavior, in: Proceedings of the 15th International Conference on Intelligent User Interfaces, IUI '10, Association for Computing Machinery, New York, NY, USA, 2010, p. 31–40. URL: https://doi.org/10.1145/1719970.1719976. doi:10.1145/1719970.1719976.

[3] L. Li, L. Zheng, F. Yang, T. Li, Modeling and broadening temporal user interest in personalized news recommendation, Expert Systems with Applications 41 (2014) 3168–3177.

URL: https://www.sciencedirect.com/science/article/pii/S0957417413009329. doi:https://doi.org/10.1016/j.eswa.2013.11.020.

[4] N. Helberger, On the democratic role of news recommenders, Digital Journalism 7 (2019) 1012 – 993. URL: https://api.semanticscholar.org/CorpusID:197796153.

[5] T. Donkers, J. Ziegler, The dual echo chamber: Modeling social media polarization for interventional recommending, in: Proceedings of the 15th ACM Conference on Recommender Systems, RecSys '21, Association for Computing Machinery, New York, NY, USA, 2021, p. 12–22. URL: https://doi.org/10.1145/3460231.3474261. doi:10.1145/3460231.3474261.

[6] D. Frey, Recent research on selective exposure to information, volume 19 of *Advances in Experimental Social Psychology*, Academic Press, 1986, pp. 41–80. URL: https://www.sciencedirect.com/science/article/pii/S0065260108602129. doi:https://doi.org/10.1016/S0065-2601(08)60212-9.

[7] A. Tommasel, J. M. Rodriguez, D. Godoy, I want to break free! recommending friends from outside the echo chamber, in: Proceedings of the 15th ACM Conference on Recommender Systems, RecSys '21, Association for Computing Machinery, New York, NY, USA, 2021, p. 23–33. URL: https://doi.org/10.1145/3460231.3474270. doi:10.1145/3460231.3474270.

[8] T. T. Nguyen, P.-M. Hui, F. M. Harper, L. Terveen, J. A. Konstan, Exploring the filter bubble: The effect of using recommender systems on content diversity, in: Proceedings of the 23rd International Conference on World Wide Web, WWW '14, Association for Computing Machinery, New York, NY, USA, 2014, p. 677–686. URL: https://doi.org/10.1145/2566486.2568012. doi:10.1145/2566486.2568012.

[9] Ö. Özgöbek, J. A. Gulla, R. C. Erdur, A survey on challenges and methods in news recommendation, in: International Conference on Web Information Systems and Technologies, 2014. URL: https://api.semanticscholar.org/CorpusID:19984721.

[10] G. de Souza Pereira Moreira, CHAMELEON: A deep learning meta-architecture for news recommender systems [phd. thesis], CoRR abs/2001.04831 (2020). URL: https://arxiv.org/abs/2001.04831. arXiv:2001.04831.

[11] F. Wu, Y. Qiao, J.-H. Chen, C. Wu, T. Qi, J. Lian, D. Liu, X. Xie, J. Gao, W. Wu, M. Zhou, MIND: A large-scale dataset for news recommendation, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online, 2020, pp. 3597–3606. URL: https://aclanthology.org/2020.acl-main.331. doi:10.18653/v1/2020.acl-main.331.

[12] M. An, F. Wu, C. Wu, K. Zhang, Z. Liu, X. Xie, Neural news recommendation with long- and short-term user representations, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Florence, Italy, 2019, pp. 336–345. URL: https://aclanthology.org/P19-1033. doi:10.18653/v1/P19-1033.

[13] C. Wu, F. Wu, S. Ge, T. Qi, Y. Huang, X. Xie, Neural news recommendation with multi-head self-attention, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, Hong Kong, China, 2019, pp. 6389–6394. URL: https://aclanthology.org/D19-1671. doi:10.18653/v1/D19-1671.

[14] C. Wu, F. Wu, Y. Huang, X. Xie, Personalized news recommendation: Methods and challenges, 2022. doi:https://doi.org/10.48550/arXiv.2106.08934. arXiv:2106.08934.
[15] C. Wu, F. Wu, T. Qi, Y. Huang, Two birds with one stone: Unified model learning for both recall and ranking in news recommendation, 2022. arXiv:2104.07404.
[16] C. Wu, F. Wu, T. Qi, Y. Huang, Empowering news recommendation with pre-trained language models, in: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '21, Association for Computing Machinery, New York, NY, USA, 2021, p. 1652–1656. URL: https://doi.org/10.1145/3404835.3463069. doi:10.1145/3404835.3463069.
[17] T. Qi, F. Wu, C. Wu, Y. Huang, News recommendation with candidate-aware user modeling, in: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '22, Association for Computing Machinery, New York, NY, USA, 2022, p. 1917–1921. URL: https://doi.org/10.1145/3477495.3531778. doi:10.1145/3477495.3531778.
[18] G. de Souza Pereira Moreira, F. Ferreira, A. M. da Cunha, News session-based recommendations using deep neural networks, in: Proceedings of the 3rd Workshop on Deep Learning for Recommender Systems, ACM, 2018. URL: https://doi.org/10.1145%2F3270323.3270328. doi:10.1145/3270323.3270328.
[19] B. Kille, F. Hopfgartner, T. Brodt, T. Heintz, The plista dataset, in: Proceedings of the 2013 International News Recommender Systems Workshop and Challenge, NRS '13, Association for Computing Machinery, New York, NY, USA, 2013, p. 16–23. URL: https://doi.org/10.1145/2516641.2516643. doi:10.1145/2516641.2516643.
[20] J. A. Gulla, L. Zhang, P. Liu, O. Özgöbek, X. Su, The adressa dataset for news recommendation, in: Proceedings of the International Conference on Web Intelligence, WI '17, Association for Computing Machinery, New York, NY, USA, 2017, p. 1042–1048. URL: https://doi.org/10.1145/3106426.3109436. doi:10.1145/3106426.3109436.
[21] L. Michiels, J. Leysen, A. Smets, B. Goethals, What are filter bubbles really? a review of the conceptual and empirical work, in: Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization, UMAP '22 Adjunct, Association for Computing Machinery, New York, NY, USA, 2022, p. 274–279. URL: https://doi.org/10.1145/3511047.3538028. doi:10.1145/3511047.3538028.
[22] H. Dalton, The measurement of the inequality of incomes, The Economic Journal 30 (1920) 348–361. URL: http://www.jstor.org/stable/2223525.
[23] P. Castells, N. J. Hurley, S. Vargas, Novelty and diversity in recommender systems, in: Recommender Systems Handbook, 2015. URL: https://api.semanticscholar.org/CorpusID:45086523.
[24] S. Vargas, P. Castells, Rank and relevance in novelty and diversity metrics for recommender systems, in: Proceedings of the Fifth ACM Conference on Recommender Systems, RecSys '11, Association for Computing Machinery, New York, NY, USA, 2011, p. 109–116. URL: https://doi.org/10.1145/2043932.2043955. doi:10.1145/2043932.2043955.
[25] W. Sun, S. Khenissi, O. Nasraoui, P. Shafto, Debiasing the human-recommender system feedback loop in collaborative filtering, in: Companion Proceedings of The 2019 World Wide Web Conference, WWW '19, Association for Computing Machinery, New York, NY, USA, 2019, p. 645–651. URL: https://doi.org/10.1145/3308560.3317303. doi:10.1145/

3308560.3317303.

[26] S. Raza, S. R. Bashir, U. Naseem, Accuracy meets diversity in a news recommender system, in: Proceedings of the 29th International Conference on Computational Linguistics, International Committee on Computational Linguistics, Gyeongju, Republic of Korea, 2022, pp. 3778–3787. URL: https://aclanthology.org/2022.coling-1.332.

[27] S. Vrijenhoek, M. Kaya, N. Metoui, J. Möller, D. Odijk, N. Helberger, Recommenders with a mission: Assessing diversity in news recommendations, in: Proceedings of the 2021 Conference on Human Information Interaction and Retrieval, CHIIR '21, Association for Computing Machinery, New York, NY, USA, 2021, p. 173–183. URL: https://doi.org/10.1145/3406522.3446019. doi:10.1145/3406522.3446019.

[28] S. Vrijenhoek, G. Bénédict, M. Gutierrez Granada, D. Odijk, M. De Rijke, Radio – rank-aware divergence metrics to measure normative diversity in news recommendations, in: Proceedings of the 16th ACM Conference on Recommender Systems, RecSys '22, Association for Computing Machinery, New York, NY, USA, 2022, p. 208–219. URL: https://doi.org/10.1145/3523227.3546780. doi:10.1145/3523227.3546780.

[29] Y. Koren, R. Bell, C. Volinsky, Matrix factorization techniques for recommender systems, Computer 42 (2009) 30–37. URL: https://doi.org/10.1109/MC.2009.263. doi:10.1109/MC.2009.263.