ISSN 1462 2815

UNDERSTANDING THE INTERPLAY OF FAIRNESS, BIAS, AND CONTENT RECOMMENDATION ALGORITHMS: A FOUNDATIONAL STUDY

Mythili Boopathi [1], Harshita Sharma [2], A Sangeetha [3] and Tanushree Kaushik [4]

[1] School of Computer Science Engineering and Information Systems, Vellore Institute of Technology, Vellore, Tamil Nadu, India.
[2,4] School of Computer Science Engineering and Information Systems, Department of Computer Applications, Vellore Institute of Technology, Vellore, Tamil Nadu, India.
[3] School of Electronics Engineering, Vellore Institute of Technology Vellore, India.
Email: [1]nmythili@vit.ac.in, [2]harshita.sharma2023@vitstudent.ac.in, [3]asangeetha@vit.ac.in, [4]tanushree.kaushik2023@vitstudent.ac.in

DOI: 10.5281/zenodo.13884235

Abstract

In the ever-evolving digital landscape, algorithms wield remarkable influence over the information we encounter and the choices we make. This research aims to establish a foundational study as a valuable resource for novice researchers delving into the fundamental concepts of algorithms. Specifically, our study focuses on three significant aspects of algorithms: fairness, bias, and their role in content recommendation. We believe that these aspects are inherently interconnected, and our goal is to elucidate this relationship. By offering an introduction to these key components, we aim to equip researchers with the knowledge necessary to delve deeper into each of these domains and contribute to the discourse on ethical and effective algorithm deployment in the digital landscape.

Keywords: Algorithms, Bias, Fairness, Content Recommendation.

I. INTRODUCTION

Algorithms are becoming very potent and powerful tools. They help us to find information that we recommend things we might like. In short, algorithms are the invisible helpers that improve our daily experience. Compared to traditional media, online social media can connect more people in a cheaper and faster way[1]. Social media is a crucial way to understand human activity. Due to this increased activity, social media platforms provide a large opportunity for research about human behavior[2].

This paper was inspired by an urgent need: to provide an organized and basic introduction to the trinity of algorithmic fairness, bias, and content recommendation, with a focus on revealing the deep interplay that ties these parts. It is hard to overestimate the importance of algorithmic fairness. It underpins the ethical and fair deployment of algorithms, laying the groundwork for a just digital society. Algorithm biases have the ability to unknowingly perpetuate discrimination, promote preconceptions, and widen societal differences. In this context, content recommendation, the technique by which computers pick and provide digital information plays a critical role.

II. RELATED WORK

Bias can emerge at every stage of developing and using an algorithm: data collection, through coding, to application, and interpretation of output. The bias gets stronger in every cycle of the algorithm as stated by Iwas iński, Ł[3]. Training data bias is a widely discussed type of algorithmic bias. The training data is not representative of real-world situations which leads to unfairness in the results. Algorithms based on machine learning are not transparent, which is black box, so we do not have a chance to identify the discriminatory factors. Consumer bias is not directly caused by an algorithm. The ideas forbidden in the real world, are much more freely expressed in the digital environment and captured by algorithms as training data.

In the world of data analysis, two types of bias stand out: aggression and linking bias[4]. Aggression bias happens when people wrongly assume that everyone in a group is the same, even when there are differences among individuals. This can lead to mistaken generalizations from the data.Linking bias, on the other hand, occurs when data about how users connect and interact in a network doesn't really represent what users are doing. This bias comes from various factors like how the network data is collected and how users are observed. As a result, it can lead to inaccurate conclusions when studying how networks behave. These biases emphasize the importance of being careful when analyzing data and making decisions based on it.

Two notable forms of bias affect the relationship between algorithms and users. First, there's user interaction bias, which can originate from two sources: the user interface and the user's own tendencies. This type of bias is akin to presentation and ranking bias, as users often gravitate toward clicking on articles displayed at the top of their social media feeds, giving these articles more attention than less-visible ones.

Another form of bias, known as emergent bias, emerges as a result of real users' interactions. This bias comes into play when changes occur in the user population, cultural values, or social knowledge after an algorithm's design is already in place[4,5].

Fairness in the context of recommender systems calls for equitable treatment of all users and things. The algorithmic handling of material and content filtering are examples of social media fairness. Social media businesses also strive to make their algorithms impartial and fair.To address concerns of prejudice and discrimination, platforms are continually improving their policies and algorithms, ensuring that all users have an equitable experience regardless of their identity or background. When prejudice or unfairness are removed from systems, we may create recommender systems that are fair.

Recommendation systems rely on data from both explicit and implicit sources to understand user preferences. Explicit data includes user-provided information such as reviews and ratings, while implicit data is inferred from user behavior patterns. Data mining techniques are used to analyze these data sources and identify user tastes, enabling personalized recommendations.[6] The recommendation system serves as a filter, sifting through a wide array of items to provide tailored suggestions based on user preferences and behavior.

The various models of information propagation on social media include the subscription model, network model, and algorithmic model. In the subscription model, users receive content from creators they have subscribed to, akin to traditional broadcast media. The network model extends this by allowing users to see posts amplified by those they follow, potentially leading to information cascades. The algorithmic model, which is not solely dependent on the number of subscribers, optimizes the audience for each post based on topic and post quality, independent of past post-performance.[8] While no platform strictly adheres to one model,

ISSN 1462 2815

comprehending these fundamental models is essential for dissecting the dynamics of information propagation in the digital realm.

III. RESEARCH MODEL

Figure1. illustrates the algorithmic ecosystem, demonstrating the crucial connections between bias, fairness, and recommendation algorithms. It serves as a visual aid in comprehending how data bias and algorithmic bias interact with fairness measures and bias detection to promote a balanced and ethical content recommendation environment. This image encapsulates the core concepts underlying our research, opening the door to a more equitable digital landscape.

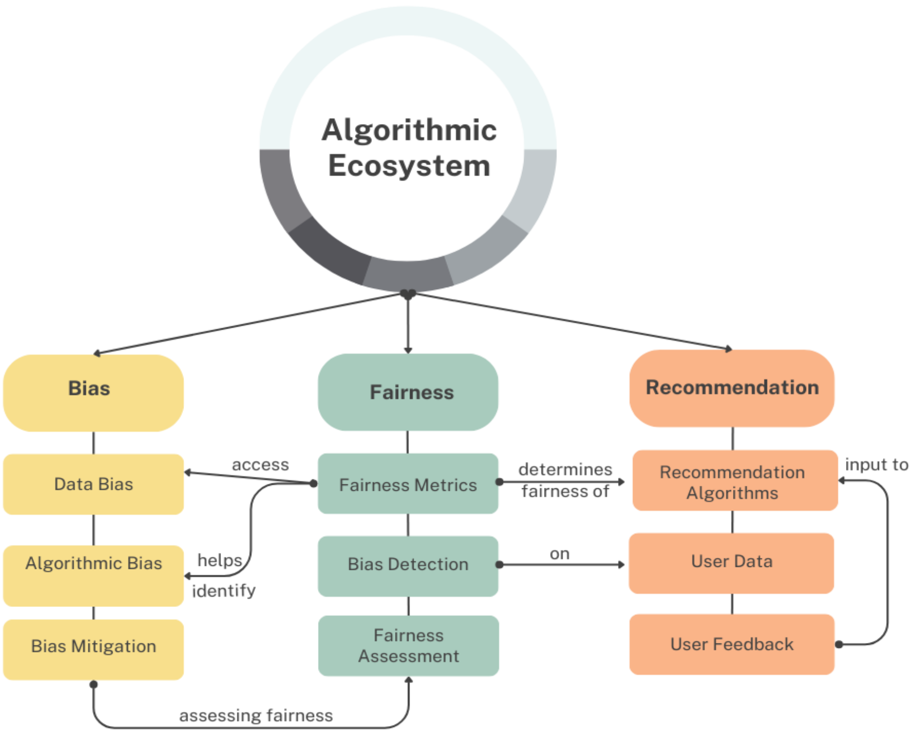

Figure 1: Architecture diagram of how Biasness, Fairness, and Recommendation are interrelated to each other

IV. METHODOLOGY

We have undertaken an in-depth qualitative examination of topics surrounding fairness, algorithms, and content recommendation algorithms. Our research primarily focuses on a comprehensive analysis of existing literature and research works. This qualitative study serves as a foundational approach to comprehending the intricate interplay between these crucial components, providing a basis for further investigations and discourse in the field.

A. Necessity of fairness in recommendation

The recommendation system is developed on a feedback loop with three characteristics: user, data, and model. These characteristics are used in three different levels. First, Collection: This refers to the process of obtaining user data, which includes interactions between users and items as well as unrelated data (such as user profiles, item characteristics, and contexts).Second, Learning: This is the process of creating recommendation models based on data that has been gathered. At its core, it makes predictions about a user's propensity to adopt a target item based on previous interactions. Third, Serving: In this stage, consumers receive the results of the proposal to fulfill their informational wants. Future actions and choices made by users will be influenced by this phase. These are the stages when the system is biased, which makes the recommendation system potentially unfair to either a subject or an object.

B. Fairness Data Processing

Pre-processing reduces biases in the training datasets. It is possible for biases to get magnified throughout a recommender system's lifetime and produce suggestions that are unjust to users. This is divided into 3 categories. First, Data Re-labeling which is used to eliminate biases in the input data and re-labeling involves altering the labels of the training dataset. Second, Data Re-sampling defines a less even distribution of data might reduce recommender system training effectiveness. With respect to the same percentage of users, this kind of resampling technique shrinks the dataset while preserving the dynamic aspect of user profiles. Lastly, Data Modification is the primary goal of data modification is to lessen bias by augmenting or changing biased data[7].

In-processing techniques adapt learning algorithms in order to remove discrimination during the model training process[4,7]. It is very useful when the bias in the data is not easily identifiable or when the dataset is too large to pre-process. If these changes in the machine learning model are allowed then in-processing can be used during the training of a model-either by incorporating changes into objective function.

Post-processing is performed after training by accessing a dataset which was not involved during the training of the model[7]. In the process of fairness, the recommendation results from target systems are not optimal, these results potentially don't take factors concerning fairness into consideration, these methods aim to rearrange the results of recommendation provided by target models, which are treated as black-box during the rearrangement after the training of recommender systems. This is categorized into several methods.

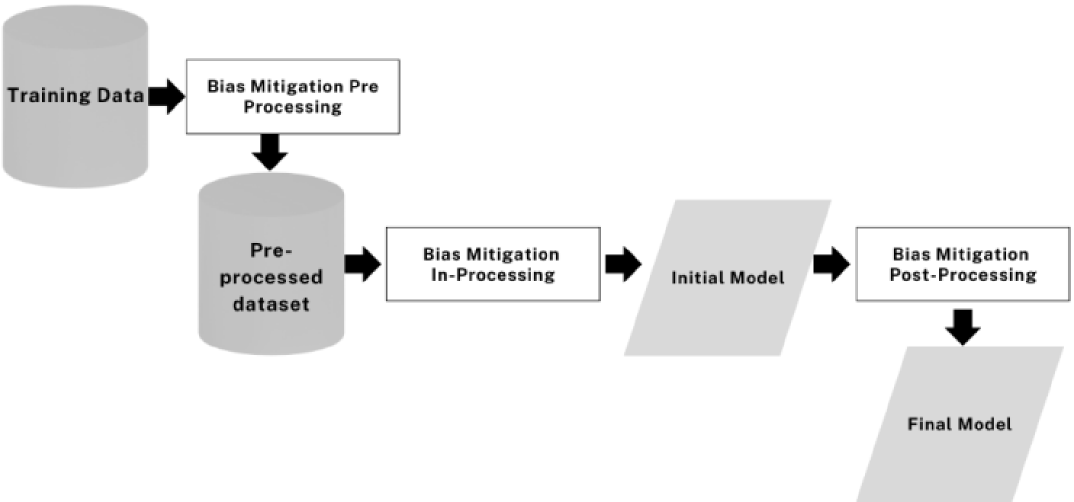

Figure 2: Fairness data processing diagram

C. Content Recommendation

Before embarking on a comprehensive exploration of recommendation systems, it is imperative to familiarize oneself with the foundational terminologies intrinsic to the entire process. Two core concepts prominently feature: 'Items,' which represent the entities proposed by the system, and 'Query,' denoting the information harnessed by the system to underpin its recommendations.[9]. These fundamental terms provide the necessary foundation for our ensuing investigation.

Following that, we come across the crucial concepts of 'Embedding' and 'Similarity Measures.' These principles are fundamental in the context of content recommendation. To be more specific, 'Embedding' refers to the strategic positioning of both items and user preferences within a shared embedding environment. Within this spatial framework, various techniques for determining similarity are employed, including 'Dot Product,' 'Cosine,' and 'Euclidean Distance.'[10]

In the first stage of a recommendation system, there is a crucial process called Candidate Generation. This process helps to narrow down the algorithm's search from a large, unwieldy collection to a much smaller set of options. The key feature of this phase is its ability to present a curated selection of suitable options in response to the user's query. Recommendation systems use different approaches to suggest items to users based on their interests. [10]

Here two dominant approaches emerge: Content-Based Filtering and Collaborative-Based Filtering. Content-based filtering matches content attributes to a user's previous interactions to suggest items that are conceptually related to their interests. On the other hand, Collaborative Filtering creates a user's profile by analyzing their resemblance to others. In Collaborative Filtering, the choice of a similarity measure is important to its effectiveness, as it assesses the strength of relationships between users or objects and establishes the foundation for recommendations. Often, recommendation systems combine both approaches into a Hybrid Filtering model, leveraging both approaches' capabilities to improve the quality and relevance of recommendations. [8]

A range of filtering models and techniques such as Text Mining, K-nearest neighbor (KNN), Clustering, Matrix Factorization, and Neural Networks. Among these models, Neural Networks are gaining popularity due to their advanced modeling capabilities that can enhance the performance of recommendation systems. However, it's important to note that different domains may require different filtering models for optimal results.[8] Therefore, it is essential to use tailored approaches while implementing these models. Our paper aligns with these research objectives and provides further insights on this topic.

V. RESULT

Our results illustrate the innate connection between algorithmic fairness, biases, and content recommendation. It draws attention to how crucial justice is to treating all users fairly and how prejudice must be overcome to stop inequalities and discrimination. Finding a way to recommend information while maintaining fairness and efficacy is a major challenge. This knowledge empowers new researchers to support the ethical and efficient implementation of algorithms, promoting diversity and confidence in the digital sphere.

VI. CONCLUSION

Fairness metrics, bias detection, and fairness assessment play a significant role in mitigating data and algorithmic bias, ultimately striving for a more equitable content recommendation platform. By providing essential insights into the operation of recommendation algorithms, user data, and user feedback, this research equips budding researchers with a fundamental knowledge to understand these critical aspects. This study acts as a cornerstone for promoting fairness and transparency in content recommendation, offering a strong foundation for future investigations and driving advancements toward a user-centric digital environment. In conclusion, this research provides valuable insights into the working of content recommendation systems and highlights the importance of fairness and transparency in the evolving digital landscape.

References

1) Morstatter, F., & Liu, H. (2017). Discovering, assessing, and mitigating data bias in social media. *Online Social Networks and Media*, *1*, 1-13.

2) Jackson, J. R. (2018). Algorithmic bias. Journal of Leadership, Accountability and Ethics, 15(4), 55-65.

3) Iwasiński, Ł. (2020). Social Implications of Algorithmic Bias.

4) Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM computing surveys (CSUR)*, *54*(6), 1-35.

5) Batya Friedman and Helen Nissenbaum. 1996. Bias in Computer Systems. ACM Trans. Inf. Syst. 14, 3 (July 1996), 330–347.

6) Ko H, Lee S, Park Y, Choi A. A Survey of Recommendation Systems: Recommendation Models, Techniques, and Application Fields. *Electronics*. 2022; 11(1):141. https://doi.org/10.3390/electronics11010141

7) Jin, D., Wang, L., Zhang, H., Zheng, Y., Ding, W., Xia, F., & Pan, S. (2023). A Survey on Fairness-aware Recommender Systems. arXiv preprint arXiv:2306.00403.

8) https://knightcolumbia.org/content/understanding-social-media-recommendation-algorithms

9) https://developers.google.com/machine-learning/recommendation