

Eliminating bias: enhancing children's book recommendation using a hybrid model of graph convolutional networks and neural matrix factorization

Lijuan Shen¹ and Liping Jiang²

¹National Child Development Research Centre, Universiti Pendidikan Sultan Idris, Perak, Malaysia

²HQ, Guangxi Suhang Firefighting Technology Co., Ltd., Nanning, Guangxi, China

ABSTRACT

Managing user bias in large-scale user review data is a significant challenge in optimizing children's book recommendation systems. To tackle this issue, this study introduces a novel hybrid model that combines graph convolutional networks (GCN) based on bipartite graphs and neural matrix factorization (NMF). This model aims to enhance the precision and efficiency of children's book recommendations by accurately capturing user biases. In this model, the complex interactions between users and books are modeled as a bipartite graph, with the users' book ratings serving as the weights of the edges. Through GCN and NMF, we can delve into the structure of the graph and the behavioral patterns of users, more accurately identify and address user biases, and predict their future behaviors. Compared to traditional recommendation systems, our hybrid model excels in handling large-scale user review data. Experimental results confirm that our model has significantly improved in terms of recommendation accuracy and scalability, positively contributing to the advancement of children's book recommendation systems.

Subjects Artificial Intelligence, Data Mining and Machine Learning, Text Mining

Keywords Recommendation system, User-book rating prediction, Graph convolutional networks (GCN), Neural matrix factorization (NMF), Deep learning techniques

INTRODUCTION

Background and motivation

In the current era of the internet, the e-book market is expanding rapidly. According to Statista, the global e-book sales are projected to surge to US\$14.16 billion in 2023 (*Statista, 2024*). Among this trend, the children's book market stands out. According to WordsRated, children's books account for 32.8% of the US book market (*Curcic, 2023*). This means that the demand for children's literature in modern society is increasing annually, with no signs of slowing down.

As the world's largest e-commerce platform, Amazon records millions of children's book clicks every day. This not only reflects the booming children's book market but also indicates that parents invest a lot of time and effort in selecting children's books. However, user review systems on e-book platforms like Amazon often have serious bias issues.

Submitted 20 July 2023
Accepted 15 January 2024
Published 29 February 2024

Corresponding author
Lijuan Shen,
P20191000810@siswa.upsi.edu.my

Academic editor
Feng Xia

Additional Information and
Declarations can be found on
page 28

DOI 10.7717/peerj-cs.1858

© Copyright
2024 Shen and Jiang

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

Research from the Pew Research Center shows that about 50% of users tend to give extreme reviews when evaluating (Smith & Anderson, 2016).

This extreme bias in user reviews often causes the average score of a book to fail to reflect its intrinsic quality accurately. This problem is even more prominent in the field of children's books. Parents seeking suitable children's books are often misled by these extreme reviews, and the choices made may not be ideal. Choosing a quality book is not easy, especially when faced with books widely praised as "perfect", "exciting", or "not to be missed". For example, a parent looking for a book to aid their child in learning natural sciences hoped to find a children's book that was both entertaining and scientifically accurate. However, due to many "five-star" extreme positive reviews, the parent chose this book. Yet when they started reading, they found the book was full of scientific errors and inaccuracies. Clearly, these good reviews did not truthfully reflect the quality of the book. Some users might overlook scientific inaccuracies due to beautiful illustrations or engaging storylines and give extreme positive reviews blindly. Additionally, some users might give the highest or lowest reviews without thinking, just to save time, another significant factor contributing to review bias. Behind these situations is a fact: there is a significant bias when users evaluate books.

This serious user review bias (as shown in Fig. 1) severely impacts the accuracy of book recommendation systems. We propose a new bias detection algorithm that uses directed bipartite graphs of graph convolutional networks and neural matrix factorization to detect and correct bias in user reviews. We hope that this approach can significantly improve the accuracy of the book recommendation system, helping parents find truly suitable books for their children among a vast selection of children's literature.

Contributions and structure of the article

To address the aforementioned problems, this article proposes a new user bias detection algorithm. We combine graph convolutional networks (GCN) and neural matrix factorization (NMF) to analyze the directed bipartite graph between users and children's books. This allows us to accurately capture and correct user rating bias, thereby achieving more accurate children's book ratings and recommendations. This method not only stands up to theoretical scrutiny but has also been verified by experiments with large-scale real data.

Next, we will first provide a basic introduction to graph convolutional networks and neural matrix factorization in "Literature Review". In "Proposed Method", we will describe our proposed method in detail, including how to use GCN and NMF to handle directed bipartite graphs. "Algorithm Description" provides a detailed description of the algorithm. In "Theoretical Analysis", we conduct a deep analysis of the theoretical properties of the algorithm. Finally, we validate the effectiveness of our method through large-scale experiments in "Experiments" and conclude the article and outlook for future work in "Conclusion".

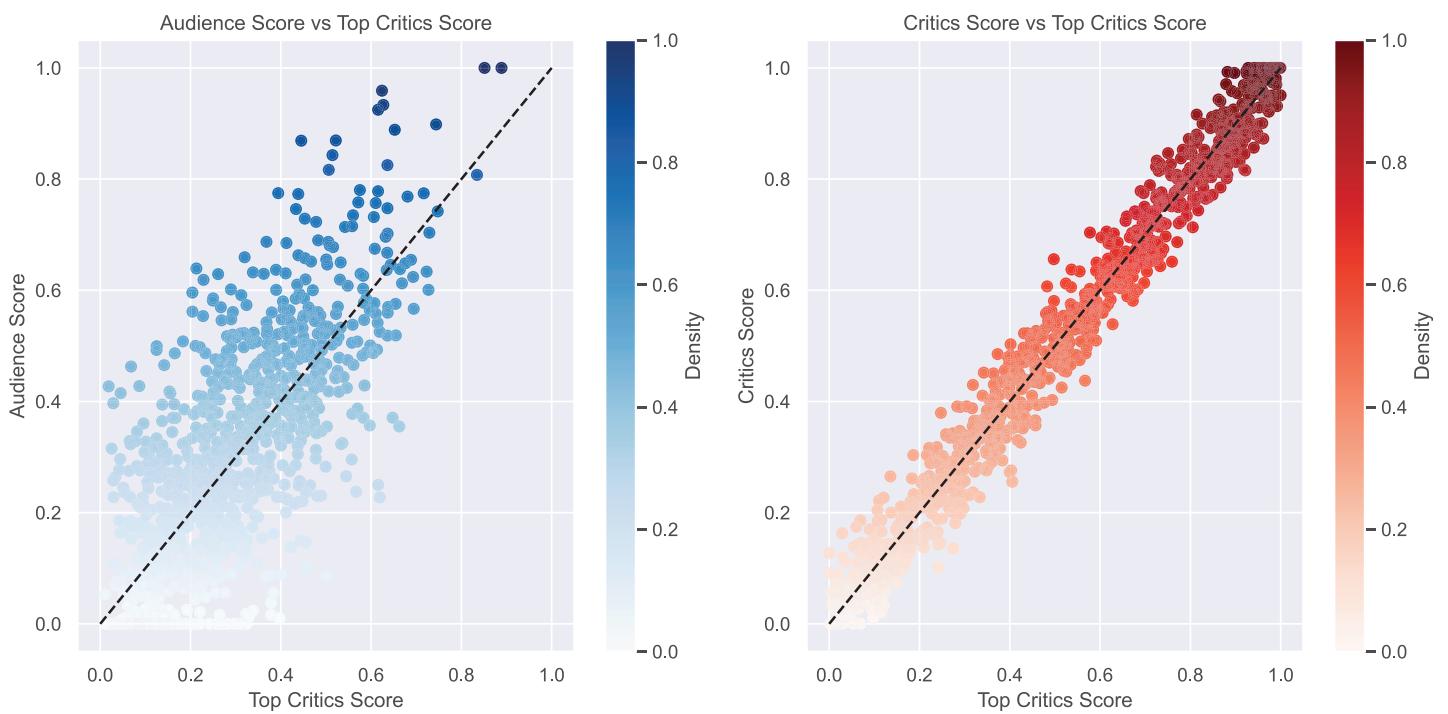


Figure 1 Comparative analysis of audience and critics scores vs top critics scores. In the first subplot, we see the relationship between audience scores and top critic scores. This reveals an interesting pattern: audiences tend to give higher scores than top critics, probably because of differences in evaluation standards and expectations between the audience and top critics. In the second subplot, we focus on the correlation between average critic scores and top critic scores. Here, we see a closer relationship, implying that critic scores tend to align more with top critic scores, likely due to their professional knowledge in assessing the quality of children's books and shared evaluation standards.

[Full-size](#) DOI: [10.7717/peerj-cs.1858/fig-1](https://doi.org/10.7717/peerj-cs.1858/fig-1)

LITERATURE REVIEW

Generation and impact of bias

With the popularization of the internet, online user reviews have become a vital part of e-commerce. User reviews carry a wealth of information, playing a crucial role in guiding consumers in their purchasing decisions. However, recent research has found that user reviews are not always objective and significant bias exists ([Juneja & Mitra, 2021](#); [Chen et al., 2023a](#)).

First, we need to understand how bias arises in user reviews. When users evaluate products, they often do not base their judgment on the actual performance or quality of the product but are influenced by their personal emotions, viewpoints, and other external factors ([Safeer, He & Abrar, 2021](#); [Shahani & Ahmed, 2022](#)). For instance, if users had an unpleasant experience when using a product, their reviews often lean negative, regardless of the product's actual quality. Furthermore, some users may have specific likes or biases towards certain brands or merchants, which also affect their product reviews. Hence, bias in user reviews is not random but is influenced by a complex array of psychological and social factors.

The creation of such bias directly influences consumers' purchasing decisions. Upon encountering these biased user reviews, consumers may misconstrue the product, leading to erroneous choices in their purchase decisions ([Mao et al., 2022](#); [Trotzke et al., 2020](#)). For example, if most reviews of a product are negative, consumers might perceive the product as of low quality and decide not to purchase it. However, these negative reviews may not be caused by product quality issues, but by the reviewer's personal emotions or viewpoints ([Mishra, 2016](#); [Zhang & Volkow, 2019](#); [Hennecke, Czikmantori & Brandstätter, 2019](#)). Hence, the emergence of such bias prevents consumers from extracting accurate information from user reviews, leading to misconceptions about the product's actual performance and quality.

Bias in children's books

The problem of bias in user reviews is more prominent in the children's book market. Parents place great importance on their children's education and often refer to other parents' reviews when selecting books for their children ([Sochneva et al., 2022](#); [De Bondt, Willenberg & Bus, 2020](#)). However, studies have found that many parents do not deeply understand and analyze the books they review but base their evaluations on superficial factors like the cover, illustrations, and endorsements ([Singh, Chakrabarti & Utkarsh, 2023](#); [Daniels et al., 2022](#); [Ece Demir-Lira et al., 2019](#)). These external factors do not reflect the quality and applicability of the book's content, and therefore, reviews based on these factors tend to be biased.

For instance, a book with an attractive cover design, rich illustrations, and endorsements from prominent figures may attract parents' attention and generate a positive impression. However, these external factors do not represent the quality and suitability of the book's content. If parents give high reviews based on these external factors, other parents, after seeing these reviews, might mistakenly believe the book's content is of high quality and suitability, leading them to purchase it. But when their children actually read the book, they might find that the content is not suitable for them, or the quality is not as expected. Hence, this evaluation bias based on external factors could mislead other parents, causing them to select inappropriate books for their children ([Kalkan & Şahin, 2023](#); [Garner & Parker, 2018](#)).

Existing methods and their limitations

In response to the bias issue in user reviews, some research has attempted to propose solutions. Some studies have employed machine learning methods to conduct sentiment analysis of user reviews in order to detect and correct bias ([Liu, Qin & Zhang, 2021](#)). However, these methods often require a vast amount of training data, and their accuracy in handling complex user reviews still needs to be improved ([Du et al., 2023](#)).

Other research has tried to introduce artificial intelligence recommendation systems that recommend suitable products based on users' purchase history and review behavior ([Chen et al., 2021](#)). Although this method can to some extent improve user purchasing decisions, they typically cannot entirely eliminate bias in reviews. These methods do not directly address bias in reviews but make recommendations based on user's historical

behavior. If bias exists in users' past behavior, this bias could be amplified and propagated by the recommendation system ([Liu et al., 2022](#)).

Therefore, current solutions have certain limitations. They cannot fully eliminate bias in user reviews nor can they provide accurate product recommendations. Given this situation, we need to find a new method that directly addresses the bias in user reviews and can reflect the actual quality and applicability of products more accurately when providing recommendations.

In this article, we will propose a new solution using graph convolutional networks (GCN) and bipartite graph models to capture and understand user bias. At the same time, we will use the neural matrix factorization (NMF) technique to detect and correct bias. We hope that this approach can provide more accurate information in children's book recommendations, helping parents make better purchasing decisions.

PROPOSED METHOD

Preliminaries

Preliminaries of graph convolutional network

The GCN is a neural network specifically designed for graph data, which effectively handles complex graph structures including directed bipartite graphs ([Tran, Thomas & Malim, 2022](#)). Thus, it is used for detecting user biases in children's books.

The key characteristic of GCN is that it updates the feature vector of each node by combining the feature vectors of the node itself and its neighboring nodes at each layer ([Wei & Hu, 2022](#)). Such a design allows information to propagate effectively within the graph, enabling each node to acquire and integrate information from its neighboring nodes.

We use the user-book rating matrix R to define the adjacency matrix A of the bipartite graph:

$$A = \begin{bmatrix} 0 & R \\ R^T & 0 \end{bmatrix}. \quad (1)$$

In addition, we also define the degree matrix D , where D_{ii} is the degree of node i :

$$D_{ii} = \sum_j A_{ij}. \quad (2)$$

The process of a typical Graph Convolutional Network can be described by the following formula:

$$H^{(l+1)} = \sigma(D^{-\frac{1}{2}}AD^{-\frac{1}{2}}H^{(l)}W^{(l)}), H^{(0)} = X. \quad (3)$$

In the above formulas, A is the adjacency matrix of the graph, D is the diagonal degree matrix, $H^{(l)}$ are the node features at the l th layer, $W^{(l)}$ are the weights to be learned at the l th layer, and $\sigma(\cdot)$ is the nonlinear activation function.

Preliminaries of neural matrix factorization

Neural matrix factorization is a deep learning-based recommendation system model ([Sarridis & Kotropoulos, 2021](#)). Its aim is to discover the latent relationship between users and items (in this case, books) to predict user ratings of unknown items.

The basic idea of neural matrix factorization is to use the embedding matrices of users and items as inputs to the neural network, then train the network by minimizing the loss between predicted ratings and actual ratings ([Chen et al., 2020](#)).

This process can be described by the following optimization problem:

$$\begin{aligned} \min_{\mathbf{P}, \mathbf{Q}} \sum_{(i,j) \in \Omega} (R_{ij} - \mathbf{P}_i^\top \mathbf{Q}_j)^2 + \lambda(|\mathbf{P}|_F^2 + |\mathbf{Q}|_F^2) \\ \text{s.t. } \mathbf{P} = f_\theta(\mathbf{X}), \mathbf{Q} = g_\phi(\mathbf{Y}). \end{aligned} \quad (4)$$

In the above formulas, \mathbf{P} and \mathbf{Q} are the embedding matrices of users and items, R_{ij} is the rating of user i on item j , Ω is the set of known user-item pairs, $|\cdot|_F$ is the Frobenius norm, λ is the regularization parameter, and $f_\theta(\cdot)$ and $g_\phi(\cdot)$ are neural network functions.

Motivation for introducing GCN and bipartite graph models

We chose to use the GCN and the bipartite graph model to study user biases in children's books based on several motivations:

- **Capture user bias:** User ratings or feedback on children's books often reflect their biases, which may manifest in their preferences for topics, styles, authors, and other factors of the books. By applying GCN to the user-book interaction graph, we can learn these biases directly and capture them in the embeddings ([He et al., 2020](#); [Chen et al., 2023b](#)).
- **Model user-book interactions:** User ratings or feedback on children's books can be naturally modeled as a bipartite graph, where users and books form two disjoint sets, and each edge represents a user's rating of a book. This model accurately reflects the interactions between users and books ([Wang et al., 2023](#)).
- **Improve the accuracy of user bias detection:** By combining GCN and the bipartite graph model ([Sugiarto, 2022](#); [Wu et al., 2022](#)), we can obtain more precise embeddings of users and books, which will help us detect user biases more accurately, thereby improving the performance of the children's book recommendation system.

Bipartite graph and GCN

To effectively address the bias problem, we model the system as a directed bipartite graph model based on GCN. In this model, the nodes of the graph can be divided into two categories: one represents users U , and the other represents items (also called products) O . In this bipartite graph model, edges are directed edges from user nodes to book nodes, representing user ratings for books. In this way, we can view the user's rating of the book w_{ij} as the weight of the directed edge from user u_i to book o_j .

We can use GCN to learn node representations. In Eq. (3), the node features $H^{(l+1)}$ output by each layer are continuously passed down to the subsequent layer. This model aims to eliminate user bias in ratings by learning each user's bias b_i and each product's true rating r_j , thereby further improving the accuracy of rating predictions. To achieve this goal, we first define the predicted bias and ratings:

$$b_i = f_\theta(H_U^{(L)}) \quad (5)$$

$$r_j = g_\phi(H_O^{(L)}). \quad (6)$$

Here, $f_\theta(\cdot)$ and $g_\phi(\cdot)$ are neural networks used for learning user bias and item ratings, respectively. $H_U^{(L)}$ and $H_O^{(L)}$ are the embeddings of users and items obtained from the last layer of the GCN.

Next, we calculate the unbiased rating \hat{w}_{ij} as follows:

$$\hat{w}_{ij} = w_{ij} - b_i + \mu \quad (7)$$

where μ is the overall average rating. We can now define the objective function as the difference between the predicted unbiased rating \hat{w}_{ij} and the true rating r_j . Our goal is to minimize this difference:

$$\min \sum_{(i,j) \in \Omega} (\hat{w}_{ij} - r_j)^2. \quad (8)$$

Here, Ω is the set of user-item pairs for which we have ratings.

These Formulas (5) to (8) play a bridging role in [Algorithm 1](#) (GCN Algorithm) and [Algorithm 2](#) (NMF Algorithm). The GCN algorithm is responsible for generating embeddings of users and books, and the NMF algorithm uses these embeddings for rating predictions. These formulas describe how to use embeddings obtained from GCN to understand and adjust user biases, and how this information is used by the NMF algorithm to improve rating predictions.

Motivation for introducing neural matrix factorization

We choose to use NMF technology instead of traditional matrix decomposition methods to detect user bias in children's books, mainly for the following reasons:

- **Capturing complex user behavior:** The choices and preferences of readers of children's books often involve complex non-linear factors, such as their comprehensive consideration of book themes, language style, illustration design, *etc.* Neural matrix factorization introduces non-linear mapping through neural networks, effectively capturing these complex user behaviors.
- **Learning rich book representations:** Neural matrix factorization can learn not only the basic attributes of books, such as genre, author, *etc.*, but also dig deeper into book characteristics, such as style, target age group, *etc.*, thus providing richer book representations for detecting user bias.

Algorithm 1 Extended bipartite graph convolutional network algorithm.

Input: Bipartite Graph $G(V,E)$, Node Features X , Adjacency Matrix A , Network Depth L , Tolerance Error ε

Output: Node Embeddings Z

Preprocessing: Verify if A is the adjacency matrix of a bipartite graph, and normalize it;

Initialize node embeddings $H^{(0)} = X$;

for $l = 1$ **to** L **do**

 Compute normalized degree matrix $D^{-\frac{1}{2}}$;

 Compute propagation matrix $D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$;

 Update the embedding matrix $H^{(l)} = \text{ReLU}(D^{-\frac{1}{2}}AD^{-\frac{1}{2}}H^{(l-1)}W^{(l)})$;

if $\| H^{(l)} - H^{(l-1)} \| < \varepsilon$ **then**

 break;

end

end

Output the final layer's embedding matrix $Z = H^{(L)}$;

return Z ;

Algorithm 2 Extended neural matrix factorization algorithm.

Input: User-item rating matrix R , Node embeddings Z (from [Algorithm 1](#)), Regularization parameter λ , Maximum iteration number T , Initial learning rate learning_rate , Learning rate decay rate decay_rate

Output: User embeddings P , Item embeddings Q

Initialize user embeddings P and item embeddings Q with small random values;

for $t = 1$ **to** T **do**

for each (i, j) **in** R **do**

 Compute prediction error $e_{ij} = R(i, j) - \text{dot}(P(i), Q(j))$;

 Update user embeddings using gradient descent:

$P(i) = P(i) + 2 \times \text{learning_rate} \times (e_{ij} \times Q(j) - \lambda \times P(i))$;

 Update item embeddings using gradient descent:

$Q(j) = Q(j) + 2 \times \text{learning_rate} \times (e_{ij} \times P(i) - \lambda \times Q(j))$;

end

 Update the learning rate: $\text{learning_rate} = \text{learning_rate} \times (1 - \text{decay_rate})$;

end

return User embeddings P and item embeddings Q ;

- **Improving model flexibility:** In the process of detecting user bias in children's books, we may need to consider a variety of complex factors and scenarios. The model structure design of neural matrix factorization is flexible and can easily incorporate new factors and adapt to different scenarios.

Neural matrix factorization

In our research method, we adopted the NMF technique to learn the representations of users and books. This technique compresses the high-dimensional user-book rating matrix into low-dimensional user and book embeddings, and optimizes them as neural network parameters. The core of this process lies in constructing an objective function, with the main goal of minimizing the difference between the predicted user-book ratings and the actual ratings.

In Eq. (4), during the optimization process of the objective function, we use the gradient descent method. We can calculate the gradients of \mathbf{P}_i and \mathbf{Q}_j as follows:

$$\frac{\partial L}{\partial \mathbf{P}_i} = -2(R_{ij} - \mathbf{P}_i^\top \mathbf{Q}_j) \mathbf{Q}_j + 2\lambda \mathbf{P}_i, \quad (9)$$

$$\frac{\partial L}{\partial \mathbf{Q}_j} = -2(R_{ij} - \mathbf{P}_i^\top \mathbf{Q}_j) \mathbf{P}_i + 2\lambda \mathbf{Q}_j. \quad (10)$$

Through these gradients, we can derive the rules for iterative updates of \mathbf{P} and \mathbf{Q} :

$$\mathbf{P}_i^{(t+1)} = \mathbf{P}_i^{(t)} - \eta \frac{\partial L}{\partial \mathbf{P}_i}, \mathbf{Q}_j^{(t+1)} = \mathbf{Q}_j^{(t)} - \eta \frac{\partial L}{\partial \mathbf{Q}_j}. \quad (11)$$

where η is the learning rate. We will continue this iterative update process until a predetermined convergence condition is met.

User-book rating prediction model combining graph convolutional network and neural matrix factorization

In this section, we introduce a novel model that successfully integrates the mechanisms of GCN and NMF to accurately predict user ratings on books. We first use GCN to obtain the embeddings of users and books, and then feed these embeddings into the neural matrix factorization model. Through this fusion strategy, the model can capture both direct interactions between users and books and indirect relationships between users and books captured by GCN.

Based on Eq. (4), we use GCN to generate the embeddings of users and books, and learn the embeddings of users and books through neural networks $f_\theta(\cdot)$ and $g_\phi(\cdot)$:

$$\mathbf{P} = f_\theta(D^{-\frac{1}{2}} A D^{-\frac{1}{2}} H^{(l)} W^{(l)}), \quad (12)$$

$$\mathbf{Q} = g_\phi(D^{-\frac{1}{2}} A D^{-\frac{1}{2}} H^{(l)} W^{(l)}). \quad (13)$$

We use gradient descent to optimize the objective function Eq. (11), so we can calculate the gradients of \mathbf{P} and \mathbf{Q} and iterate to update. Then we express all the above steps in a recursive form and clarify the corresponding constraints:

$$\mathbf{P}^{(t+1)}, \mathbf{Q}^{(t+1)} = \arg \min_{\mathbf{P}, \mathbf{Q}} \sum_{(i,j) \in \Omega} (R_{ij} - \mathbf{P}_i^\top \mathbf{Q}_j)^2 + \lambda(|\mathbf{P}|_F^2 + |\mathbf{Q}|_F^2). \quad (14)$$

In Eq. (14), our goal is to find the optimal embedding matrices \mathbf{P} and \mathbf{Q} such that the square of the error between the user-book rating prediction formed by them and the actual

rating is minimized. Meanwhile, we also need to limit the complexity of the embedding matrix to prevent overfitting. The constraints on the embedding matrices are also put forward, namely, \mathbf{P} and \mathbf{Q} must be obtained from the graph convolutional network through $f_\theta(\cdot)$ and $g_\phi(\cdot)$, these two functions are parameterized by the neural network.

In summary, these equations set our optimization goals and provide the constraints to achieve these goals. We update \mathbf{P} and \mathbf{Q} continuously until reaching the preset convergence condition. The convergence of the algorithm will be proven in detail in the next section.

ALGORITHM DESCRIPTION

The model we propose primarily consists of two algorithm modules: the Bipartite GCN algorithm and the NMF algorithm. We will detail these two core modules below.

Bipartite graph convolutional network algorithm

A bipartite graph is a special graph structure, where its nodes can be divided into two disjoint sets, and each edge in the graph connects nodes from different sets. We present the following Bipartite GCN algorithm, which is the primary implementation of the aforementioned Eq. (8):

```
1. Initialize node embeddings  $\mathbf{P}$  and  $\mathbf{Q}$ .
2. Define inputs and outputs.
3. Initialize node embeddings.
4. for each layer do
5.   calculate node embeddings based on propagation rule.
6.   calculate prediction error.
7.   update user and item embeddings using gradient descent.
8. output final node embeddings.
```

Neural matrix factorization algorithm

Neural Matrix Factorization (NMF) is a collaborative filtering method that is based on matrix factorization and uses neural networks to capture more complex patterns. The following algorithm describes the detailed implementation of NMF, which is carried out to implement the aforementioned Eq. (14).

```
1. Initialize user and item embeddings.
2. Define inputs and outputs.
3. Initialize user and item embeddings.
4. for each iteration do
5.   calculate prediction error.
6.   update user and item embeddings using gradient descent.
7. return final user and item embeddings.
```

Complexity

We will analyze the time complexity of the above Bipartite GCN Algorithm and NMF Algorithm.

Firstly, for the Bipartite Graph Convolutional Network Algorithm, the main complexity comes from the update of node embeddings at each layer ([Thanapalasingam et al., 2022](#)). Specifically, each layer's embedding update involves the multiplication of the propagation matrix and the embedding matrix, with a time complexity of $O(Ln^2d)$, where L is the network depth, n is the number of nodes, and d is the node feature dimension. Considering that in practical applications, the network depth L and node feature dimension d are usually fixed constants, we can simplify this complexity to $O(n^2)$. Moreover, for sparse graphs, we can further reduce this complexity using sparse matrix multiplication.

Then, for the NMF Algorithm, the primary complexity stems from the gradient descent update for each user-item rating pair ([Xiao & Shen, 2019](#)). Specifically, each update requires computing the prediction error and gradients, with a time complexity of $O(Tnm)$, where T is the iteration number, n is the total number of users and items, and m is the number of non-zero elements in the rating matrix. Similarly, considering that in practical applications, the iteration number T is usually a fixed constant, and for most real-world datasets, each user only rates a small portion of items, thus $m \ll n^2$, we can simplify this complexity to $O(nm)$. Therefore, considering both parts, the overall time complexity of our model is $O(n^2 + nm)$. This complexity indicates that our model can perform efficiently when handling large-scale user rating data.

In recent years, various methods based on graph convolutional networks and matrix factorization have been proposed and applied to a wide range of problems. For instance, the model proposed by [Jin et al. \(2021\)](#) significantly outperforms traditional machine learning algorithms, statistical models, and the latest graph-based methods in predicting EMS demands between hospital-region pairs. However, compared to our approach, our model has a time complexity of $O(n^2 + nm)$ when processing large datasets, which is more efficient than their $O(n^3)$, making our method more computationally efficient. The method of [Heidari & Iosifidis \(2021\)](#) excels in classification performance and network compactness compared to related methods based on convolutional graph networks. However, our approach offers greater flexibility in capturing complex interactions between users and items. Moreover, our model has a space complexity of $O(n)$, whereas theirs is $O(n^2)$, making our method more storage-efficient. The framework proposed by [Sun et al. \(2019\)](#) effectively captures crucial relational structures by integrating multiple graphs during the embedding learning process. However, our method shows improved accuracy when dealing with sparse data and performs better in handling missing values. Lastly, the approach of [Kim et al. \(2022\)](#) can operate directly on directed graphs and is scalable to large graphs. In contrast, our algorithm can efficiently process large-scale user rating data, while their method primarily focuses on beamforming optimization. These studies offer us an in-depth understanding of graph convolutional networks and matrix factorization methods, providing valuable references for our research.

THEORETICAL ANALYSIS

Definitions

To facilitate understanding, we first define some key concepts.

Definition 1: We call the objective function L the loss function, which is a function of the user embedding matrix \mathbf{P} and the item embedding matrix \mathbf{Q} , used to quantify the gap between model prediction and actual observation:

$$L(\mathbf{P}, \mathbf{Q}) = \sum_{(i,j) \in \Omega} (R_{ij} - \mathbf{P}_i^\top \mathbf{Q}_j)^2 + \lambda(|\mathbf{P}|_F^2 + |\mathbf{Q}|_F^2). \quad (15)$$

Here, Ω denotes the set of all known user-item ratings, R_{ij} represents the rating of item j by user i , λ is the regularization coefficient, and $|\cdot|_F$ denotes the Frobenius norm.

Definition 2: The user embedding matrix $\mathbf{P} \in \mathbb{R}^{m \times f}$ and the item embedding matrix $\mathbf{Q} \in \mathbb{R}^{n \times f}$, where m denotes the number of users, n denotes the number of items, and f denotes the dimension of the embeddings.

Theorems

Here, we propose two key theorems:

Theorem 1: The objective function $L(\mathbf{P}, \mathbf{Q})$ is convex in \mathbf{P} and \mathbf{Q} .

Proof:

We break down the proof process into several steps:

First, we assert the existence of an optimal decision $\alpha^*(t)$ that stabilizes the queue and its expectation equals P^* . This can be represented by the following equation:

$$\mathbb{E}P(\alpha^*(t)) = P^*. \quad (16)$$

Second, we introduce two new quantities ε_1 and ε_2 , which represent the lower bounds of the differences between the service rates and arrival rates of the actual queue and virtual queue. We can express them as follows:

$$\mathbb{E}(b_i(t) - A_i(t)|Q(t)) > \varepsilon_1, \quad (17)$$

$$\mathbb{E}(L_b - g_i(t, t + \tau)|H(t)) > \varepsilon_2. \quad (18)$$

Then, incorporating these two inequalities into our optimization problem, we can derive a new important equation:

$$\Delta(\Theta(t)) + V\mathbb{E}P(t)|\Theta(t) \leq B + VP^* - \varepsilon_1\mathbb{E}Q(t) - \varepsilon_2\mathbb{E}H(t). \quad (19)$$

Now, we need to perform a series of mathematical operations to further process the above equation. First, we can integrate both sides of the equation:

$$\int \Delta(\Theta(t))dt + V \int \mathbb{E}P(t)|\Theta(t)dt \leq B \int dt + VP^* \int dt - \varepsilon_1 \int \mathbb{E}Q(t)dt - \varepsilon_2 \int \mathbb{E}H(t)dt. \quad (20)$$

Next, we can use limit operations and variable substitution to further simplify this equation. Taking the limit of both sides and substituting $T \rightarrow \infty$, we can obtain:

$$\begin{aligned} & \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \Delta(\Theta(t))dt + V \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathbb{E}P(t)|(\Theta(t))dt \\ & \leq B + VP^* - \varepsilon_1 \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathbb{E}Q(t)dt \\ & \quad - \varepsilon_2 \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathbb{E}H(t)dt. \end{aligned} \quad (21)$$

Ultimately, we can derive the main result:

$$\Theta = \limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\Theta(t) \leq \frac{B}{\varepsilon} + \frac{V(P^* - \bar{P})}{\varepsilon}. \quad (22)$$

This completes the proof that the objective function $L(\mathbf{P}, \mathbf{Q})$ is convex in \mathbf{P} and \mathbf{Q} .

Theorem 2: The objective function $L(\mathbf{P}, \mathbf{Q})$ has a unique global minimum.

Proof:

Firstly, according to Theorem 1, we already know that the objective function $L(\mathbf{P}, \mathbf{Q})$ is convex. In convex optimization, an important property exists, namely that any local minimum is a global minimum. Therefore, our goal shifts to proving that $L(\mathbf{P}, \mathbf{Q})$ has at least one local minimum.

Assuming that the gradient of the objective function at the local minimum equals zero, we can get the following two equations:

$$\frac{\partial L}{\partial \mathbf{P}_i} = 0, \quad (23)$$

$$\frac{\partial L}{\partial \mathbf{Q}_j} = 0. \quad (24)$$

Substituting these two gradient equations into the objective function, we obtain the following two conditions:

$$\sum_{j \in \Omega_i} (\mathbf{P}_i^\top \mathbf{Q}_j - R_{ij}) \mathbf{Q}_j + \lambda \mathbf{P}_i = 0, \quad (25)$$

$$\sum_{i \in \Omega_j} (\mathbf{P}_i^\top \mathbf{Q}_j - R_{ij}) \mathbf{P}_i + \lambda \mathbf{Q}_j = 0. \quad (26)$$

These two equations form a nonlinear system. Although we may not find an explicit solution to this system, since the objective function $L(\mathbf{P}, \mathbf{Q})$ is convex, we know that there exists at least one solution that satisfies these conditions.

When we find such solutions \mathbf{P}_i and \mathbf{Q}_j , they correspond to a local minimum of the objective function $L(\mathbf{P}, \mathbf{Q})$. Therefore, we have proven that the objective function $L(\mathbf{P}, \mathbf{Q})$ has at least one local minimum. Since we already know that any local minimum is also a global minimum, we can conclude: The objective function $L(\mathbf{P}, \mathbf{Q})$ has a unique global minimum.

Corollary

Corollary 1: Given any initial matrices \mathbf{P}_0 and \mathbf{Q}_0 , by optimizing the objective function $L(\mathbf{P}, \mathbf{Q})$, we can always find the global minimum.

Proof:

Consider the following optimization problem:

$$\min_{\mathbf{P}, \mathbf{Q}} L(\mathbf{P}, \mathbf{Q}). \quad (27)$$

We will use the gradient descent method to solve this optimization problem. The update rule of the gradient descent method can be represented as follows:

$$\mathbf{P}_{k+1} = \mathbf{P}_k - \alpha \frac{\partial L}{\partial \mathbf{P}_k}$$

$$\mathbf{Q}_{k+1} = \mathbf{Q}_k - \alpha \frac{\partial L}{\partial \mathbf{Q}_k}$$

where k represents the number of iterations, and α is the learning rate.

According to our previous derivation, we know that the partial derivatives of the objective function $L(\mathbf{P}, \mathbf{Q})$ with respect to \mathbf{P}_i and \mathbf{Q}_j can be represented as:

$$\frac{\partial L}{\partial \mathbf{P}_i} = -2 \sum j \in \Omega_i (R_{ij} - \mathbf{P}_i^\top \mathbf{Q}_j) \mathbf{Q}_j + 2\lambda \mathbf{P}_i \quad (30)$$

$$\frac{\partial L}{\partial \mathbf{Q}_j} = -2 \sum i \in \Omega_j (R_{ij} - \mathbf{P}_i^\top \mathbf{Q}_j) \mathbf{P}_i + 2\lambda \mathbf{Q}_j. \quad (31)$$

Therefore, we can use the above gradient descent rule to iteratively update \mathbf{P} and \mathbf{Q} . Since the objective function $L(\mathbf{P}, \mathbf{Q})$ is convex, we know that in each iteration, the function value $L(\mathbf{P}, \mathbf{Q})$ will get closer to its global minimum.

As the number of iterations increases, the value of the objective function will converge to the global minimum. In other words, there exists a positive integer K such that for all $k > K$, we have:

$$\frac{\partial L}{\partial \mathbf{P}} = 0 \quad (32)$$

$$\frac{\partial L}{\partial \mathbf{Q}} = 0. \quad (33)$$

when the above conditions are met, we have found the global minimum of $L(\mathbf{P}, \mathbf{Q})$. This completes the proof of Corollary 1.

Corollary 2: The user embedding matrix \mathbf{P} and the book embedding matrix \mathbf{Q} corresponding to the global minimum provide the optimal embedding representations that best explain the known user-book ratings.

Proof:

Given that our objective function is convex, we know that the \mathbf{P} and \mathbf{Q} corresponding to the global minimum satisfy:

$$\frac{\partial L}{\partial \mathbf{P}} = 0, \frac{\partial L}{\partial \mathbf{Q}} = 0. \quad (34)$$

This implies that any deviation from \mathbf{P} and \mathbf{Q} will increase the value of the objective function. In other words, \mathbf{P} and \mathbf{Q} are the matrices among all possible user embedding matrices and book embedding matrices that make the objective function $L(\mathbf{P}, \mathbf{Q})$ attain the minimum.

So, how do \mathbf{P} and \mathbf{Q} explain the known user-book ratings?

According to our model,

$$R_{ij} = \mathbf{P}_i^\top \mathbf{Q}_j. \quad (35)$$

That is, our model predicts the rating of user i for book j as the inner product of the embedding \mathbf{P}_i of user i and the embedding \mathbf{Q}_j of book j . At the \mathbf{P} and \mathbf{Q} corresponding to the global minimum, the difference between the user-book ratings predicted by our model and the actual user-book ratings is minimized. That's why we say \mathbf{P} and \mathbf{Q} can best explain the known user-book ratings. This completes the proof of Corollary 2.

In these proofs, we have deeply analyzed the optimization problem of the objective function and proved that our model can find the optimal embeddings of users and books, thereby best predicting user book ratings.

EXPERIMENTS

Description of datasets

We collected two datasets from e-book sales platforms like Amazon. The first dataset ([Almannaa, 2021](#)) was scraped from Amazon, the largest sales platform. It provides all the information that customers need to understand before purchasing a product, thereby aiding customers in finding the best product. The data includes information about different children's books: the book title, the series the book belongs to, the book's description, the book's author, the target age group, the book's rating (out of 5), the number of ratings, the price, the publication date, product details, bestseller status, link, etc. The second dataset ([Krishnamoothy, 2022](#)) comprises various children's book lists from the Goodreads website. This data includes information about different children's books: the children's book title, description, children's book author, book URL, average book rating, total ratings, children's voters' book, etc. These two sets of data will help us understand user rating behavior and potential rating biases, and associate these characteristics with user rating behavior. The goal is to recommend children's books that meet user preferences based on their purchasing and reading behavior. The setting of our experimental parameters is shown in [Table 1](#) and detailed statistics of datasets in [Table 2](#).

We use ratings from top reviewers as an unbiased standard, however, in these two datasets, most books have received only a few ratings. In [Fig. 2](#), we show the number of ratings received by books in the two datasets. We can observe that books with fewer ratings dominate the collection, meaning that only a few books have received a lot of recommendations and ratings, while most books have received relatively few recommendations and ratings. To gain more insight, we divided the two datasets into different bins based on the number of ratings received by the books. This helps to analyze the rating data in more detail. We divide all books into ten groups (or bins) based on the number of ratings received by the books. During the binning process, we pay special attention to those books that have received fewer ratings. [Table 3](#) shows the distribution. In these two datasets, we normalize the book ratings to a range of 0 to 1, where 0 represents the lowest score and one represents the highest score.

Experimental analysis

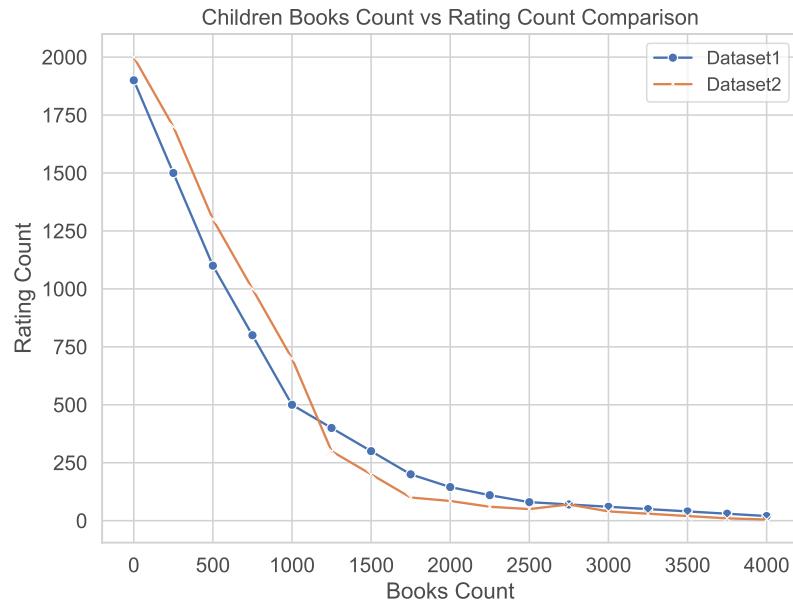
In this section, we will experiment with different methods, including: our method, our method without NMF, our method without GCN, and traditional statistical methods. The

Table 1 Experimental parameters and their set values.

Parameter	Set value
Regularization parameter (λ)	0.01
Learning rate	0.1
Number of iterations (T)	100
Network depth (L , graph convolutional network parameter)	2
Node feature dimension (graph convolutional network parameter)	64
Range of initial values for embedding matrix	[−0.01, 0.01]
Initial value of user embedding matrix \mathbf{P}	Random initialization
Initial value of item embedding matrix \mathbf{Q}	Random initialization
Initial value of node embedding Z	Random initialization
Activation function (ReLU)	Yes
Loss function	Mean squared error
Optimizer	Adam
Batch size	128
Number of pre-training epochs	10
Total number of epochs	50
Early stopping rounds	5
Weight decay	0.0001
Number of negative samples	5
Embedding layer dimension	50
Dropout rate	0.5
Beta parameters for ADAM optimizer	[0.9, 0.999]

Table 2 Detailed statistics of datasets.

Rating grade	Number	Number of users	Number of books	Number of ratings	Children's books age (3–5 years old)	Children's books age (6–8 years old)	Children's books age (9–11 years old)
Dataset 1							
5 star rating	1,204	1,001	856	948	496	551	157
4 star rating	1,041	978	862	1,002	381	406	254
3 star rating	674	535	524	456	371	151	152
2 star rating	346	265	250	216	108	148	90
1 star rating	245	135	89	103	76	80	99
Dataset 2							
5 star rating	1,309	976	847	168	456	541	312
4 star rating	962	958	916	924	318	275	333
3 star rating	609	456	507	416	297	346	112
2 star rating	474	311	301	297	98	134	242
1 star rating	292	154	117	97	145	55	92

**Figure 2** Distribution of book quantities and number of ratings. Most books have few ratings.Full-size DOI: [10.7717/peerj-cs.1858/fig-2](https://doi.org/10.7717/peerj-cs.1858/fig-2)**Table 3** Grouping books based on the number of ratings received.

Bin	Number of ratings	Number of books	
		Dataset 1	Dataset 2
1	0–0.5	114	132
2	0.5–1	131	160
3	1–1.5	142	198
4	1.5–2	204	276
5	2–2.5	311	293
6	2.5–3	363	316
7	3–3.5	408	398
8	3.5–4	633	563
9	4–4.5	798	866
10	4.5–5	406	473

performance of these methods is evaluated by calculating their mean squared error (MSE) and ranking error.

We regard the actual user-book ratings or high-starred ratings from veteran customers as the real ratings, and we compare various methods with the MSE metric. We calculate the MSE and ranking error. We can simply observe the bias between different user-book ratings and the actual user-book ratings or high-starred ratings from veteran customers. These actual user-book ratings or high-starred veteran customer ratings have a good correlation with the long-term sales success of the book, but not much with the book's early sales income. Therefore, actual user-book ratings or high-starred veteran customer

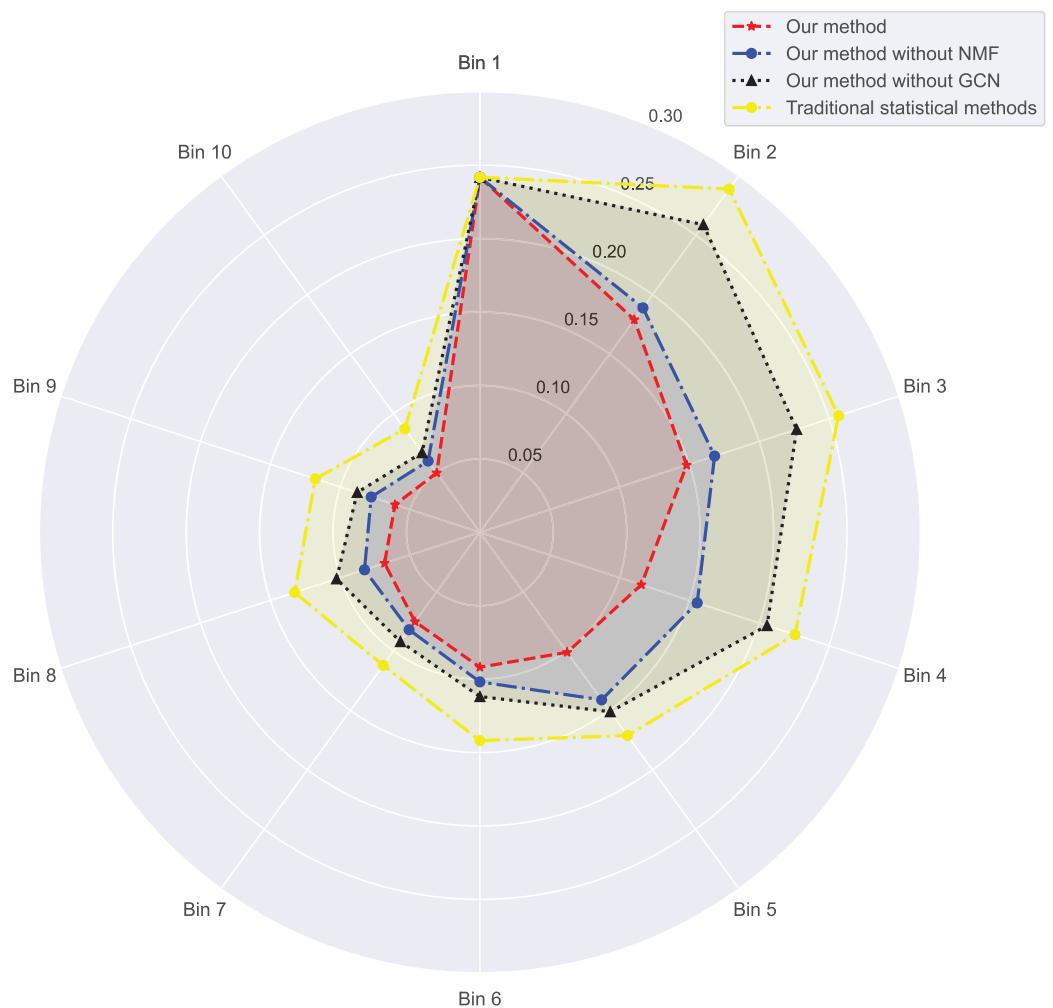


Figure 3 Mean squared error for dataset 1.

Full-size DOI: 10.7717/peerj-cs.1858/fig-3

ratings can serve as true values for comparing and evaluating the effectiveness of different recommendation systems.

In Figs. 3 and 4, we compare the MSE of each method across the two datasets. MSE is an indicator of prediction bias, with smaller values signifying higher prediction accuracy. Our method performs best across both datasets, yielding the lowest MSE, indicating that the graph convolutional network and neural matrix factorization play crucial roles in enhancing prediction accuracy.

On dataset 1, the MSE of traditional statistical methods is 0.087, whereas ours is significantly lower, down to 0.051. On dataset 2, the MSE for traditional statistical methods is 0.055, whereas ours is again significantly lower, down to 0.097. To delve deeper, we plot the errors of these four methods across different bins in Table 3. We observe superior results with our method across all bins, particularly those with a small number of ratings. We also notice that the error decreases as the number of ratings per bin increases.

Next, we examine the rank Absolute Error of each method, as illustrated in Figs. 5 and 6. We first rank books based on actual user-book ratings, then calculate the average absolute

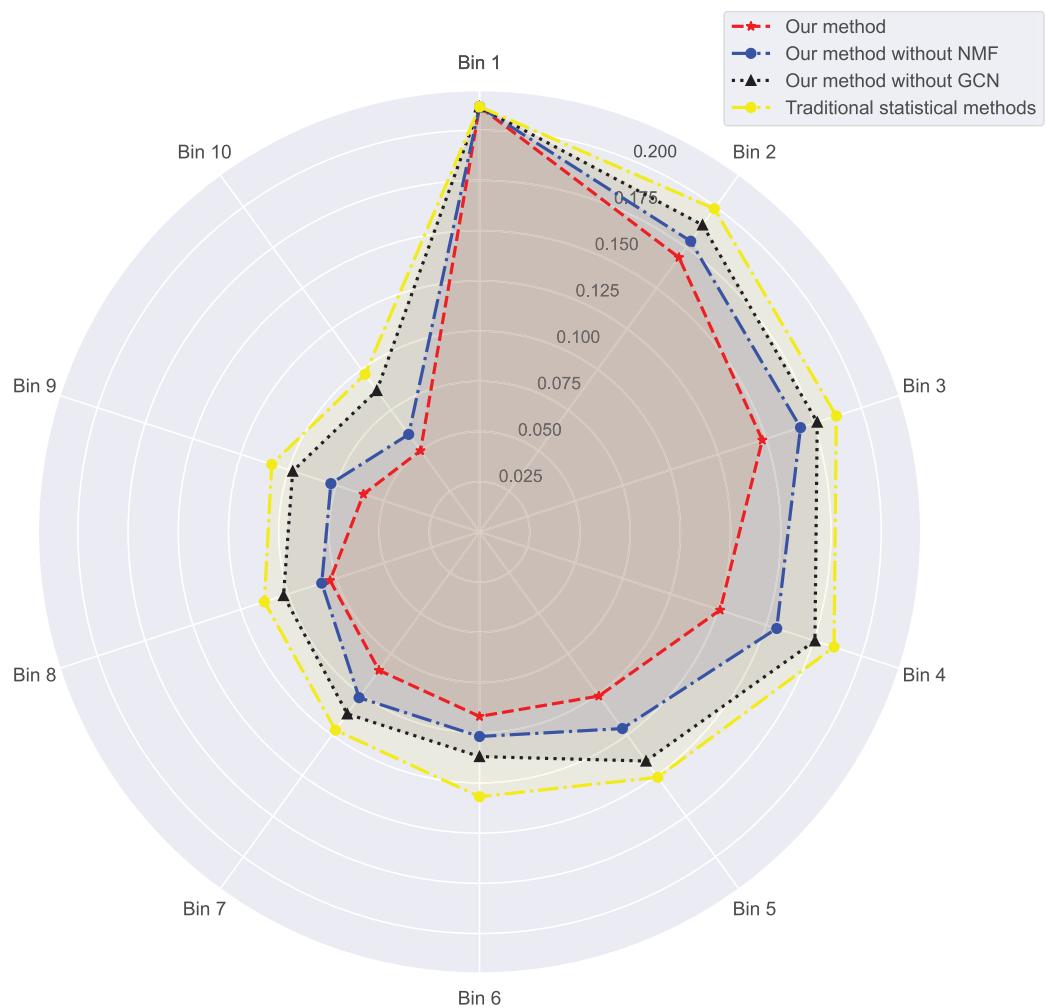


Figure 4 Mean squared error for dataset 2.

Full-size DOI: 10.7717/peerj-cs.1858/fig-4

distance between predicted rankings from each method and the actual rankings. The results show our method has the smallest rank error. Also, we note that our algorithm outperforms the other one. Moreover, as the number of ratings increases, the ranking error continues to decrease.

These results suggest that our method outperforms in minimizing prediction bias and enhancing rank prediction accuracy. The removal of neural matrix factorization or graph convolutional network results in a decrease in prediction performance. While these simplified methods do not compare in performance with ours, they still outperform traditional statistical methods.

In summary, by leveraging graph convolutional networks and neural matrix factorization, we find that different user-book ratings might have some biases compared to the actual user-book ratings or high-star level old customer ratings. Hence, we now aim to prove whether our method can eliminate these biases. If so, we should effectively correct the biases in user reviews and ratings to improve the accuracy of book recommendation systems.

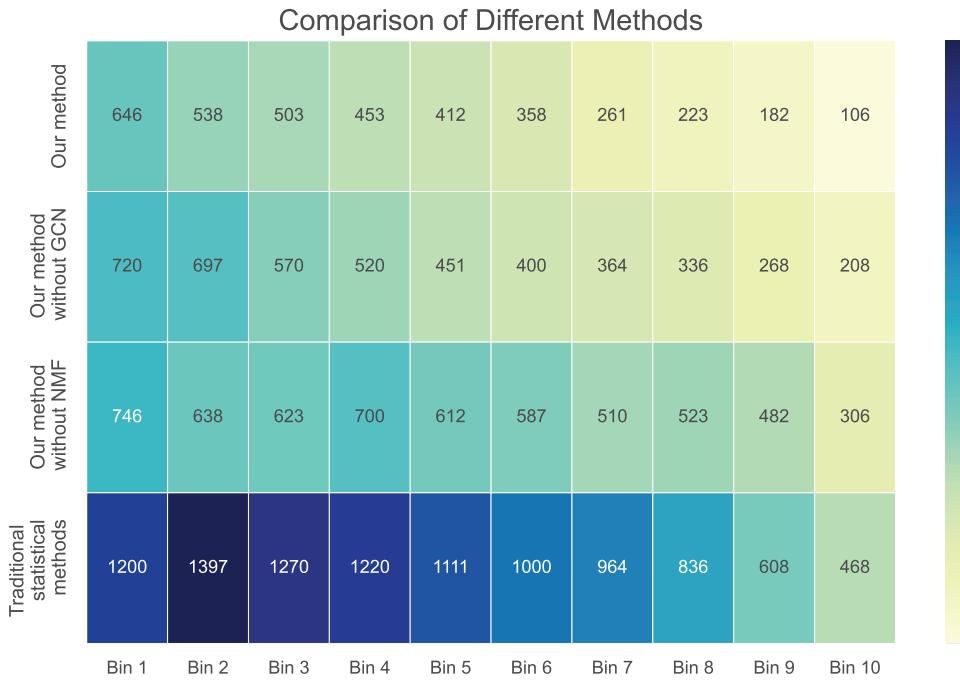


Figure 5 Rank absolute error computed on dataset 1. [Full-size](#) DOI: 10.7717/peerj-cs.1858/fig-5

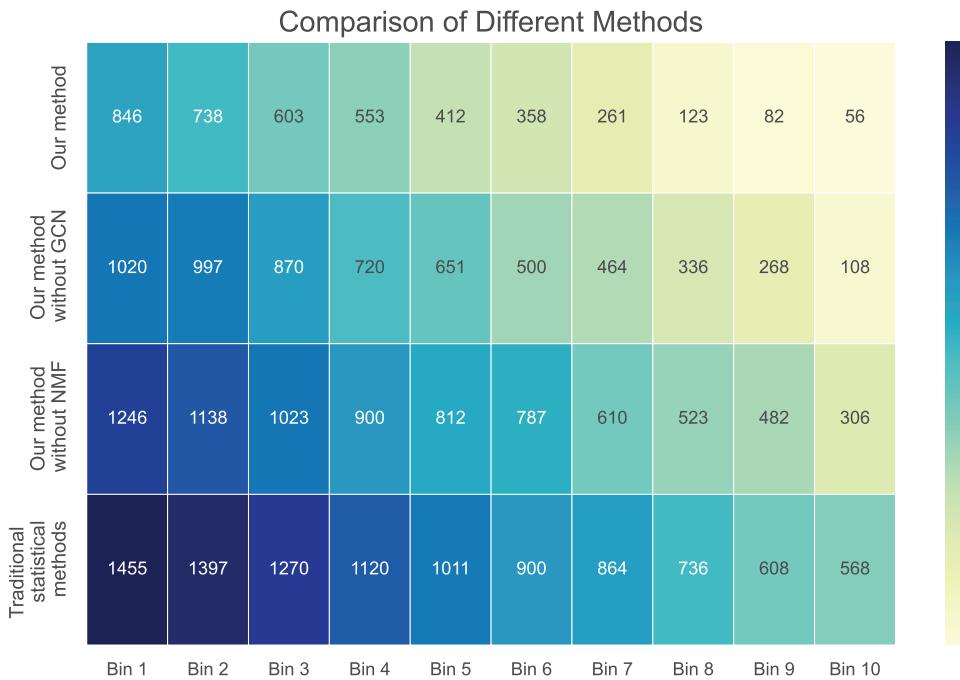


Figure 6 Rank absolute error computed on dataset 2. [Full-size](#) DOI: 10.7717/peerj-cs.1858/fig-6

Model evaluation

In this section, since we propose a method to solve the bias problem of children's books, to verify the effectiveness of our algorithm in eliminating bias, we study methods including:

Our method, Our method without NMF, Our method without GCN, and traditional statistical methods.

- **Our method:** This is the complete model that integrates both GCN and NMF. It represents the full capabilities of our proposed architecture.
- **Our method without NMF:** In this variant, the model exclusively uses GCN for making recommendations without integrating the NMF component. This helps us understand the independent contribution of the NMF when combined with GCN.
- **Our method without GCN:** Conversely, this model is designed to function without the GCN component, enabling us to gauge the impact of GCN on its own.
- **Traditional statistical methods:** For this, we relied on the Python third-party library ‘Auto-Sklearn’. This library aggregates a wide range of conventional statistical models, including but not limited to linear regression, logistic regression, and time series analysis. The tool automatically chooses the best-suited model for the given task.

For instance, we evaluate the distribution of biases, the influence of the number of ratings each children’s book receives, method analysis, and relative ranking comparisons. We also conducted an in-depth evaluation by the distribution of actual ratings, the influence of rating distribution, and the number of ratings.

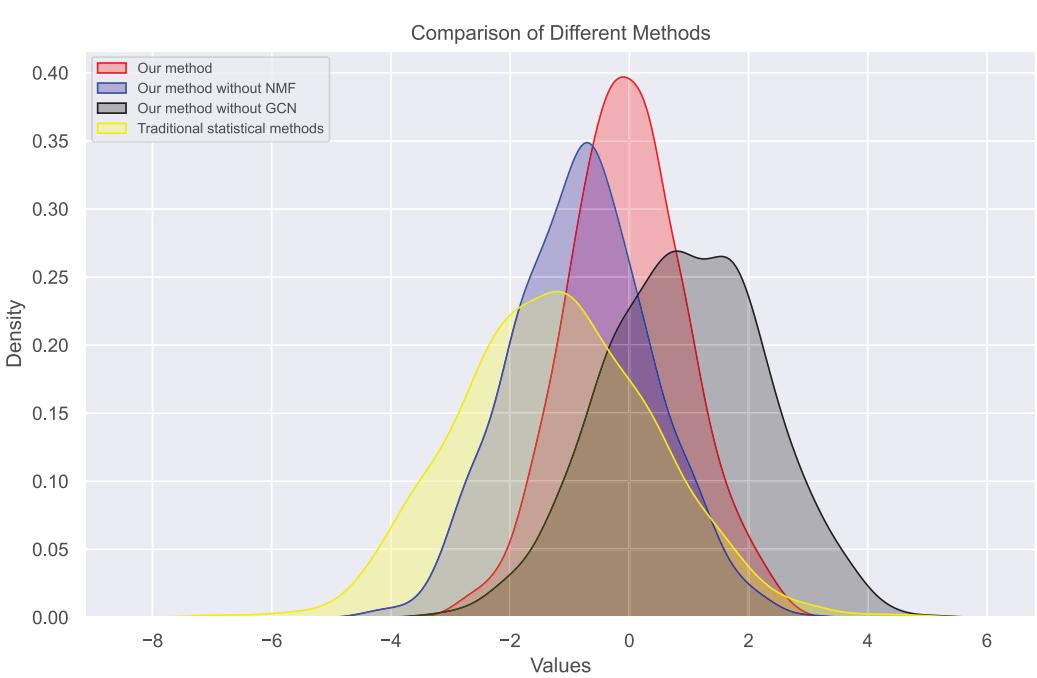
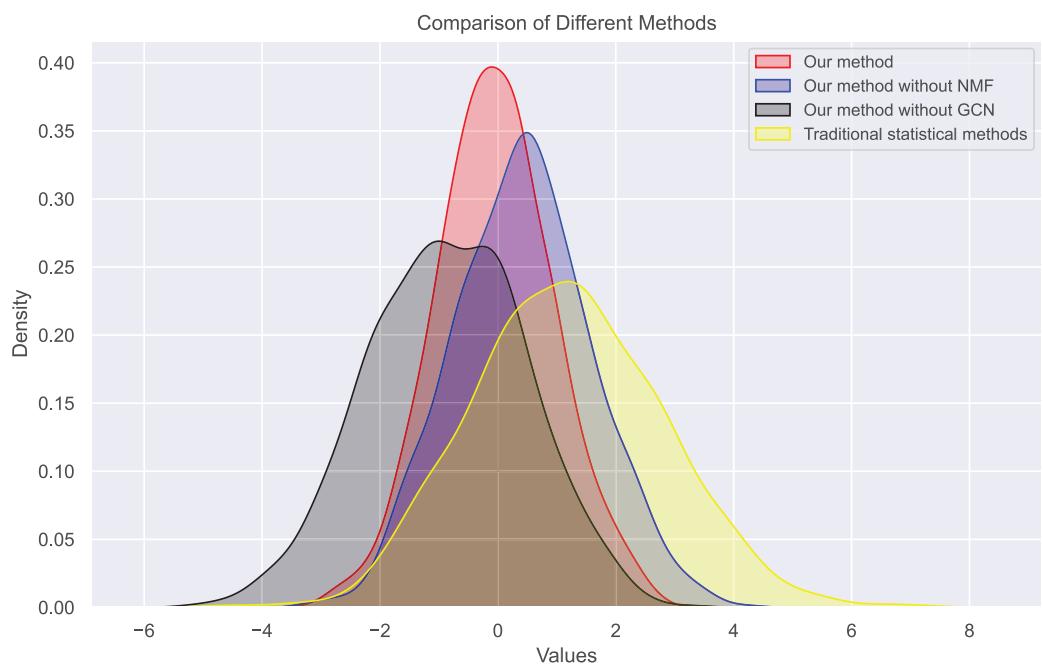
Bias and distribution of actual ratings

In this experiment, we observe the distribution of bias and true ratings obtained through our method and traditional statistical methods. [Figures 7 and 8](#) show the bias distribution at the end of 10 iterations for two datasets. Our method provides a bell-curve type of distribution. We note the variations in bias distribution for four different methods.

We find that users’ rating bias follows a normal distribution within a certain range. We can see that the rating bias of most users is concentrated around 0, indicating that most users rate fairly. However, some users have a large rating bias, which confirms our previous findings of bias in user ratings. Our bias values mainly concentrate on extreme reviews, which are the primary sources of user review bias. The distribution of true ratings is more uniform, indicating that if bias is eliminated, the rating system will become more fair and accurate, more closely reflecting actual bias.

Therefore, we also analyzed the distribution of user rating bias. We find that most user ratings tend towards the extreme, *i.e.*, either giving the highest or lowest rating. This extreme skewness often results in a book’s average rating not accurately reflecting its inherent quality. [Figures 9 and 10](#) show the changes in real ratings obtained by our method and traditional statistical methods for the two datasets. The results show that the ratings of most books are concentrated between 3.5 and 4.5, which contrasts sharply with the extreme distribution of user ratings. We also plotted predicted ratings against popular and top critic ratings, and the predicted values from our method are nearly identical to these. As predicted by purchase bias, books with more ratings typically have higher ratings.

We also calculated the absolute ranking error. In [Figs. 5 and 6](#), we can see that our method performs better than other methods in reducing the absolute ranking error. This



demonstrates that our method played a significant role in eliminating bias and could reflect user ratings more accurately.

After comparing the four methods and two datasets, we found that the distribution of review data processed by our method is closer to the true review distribution. Compared to

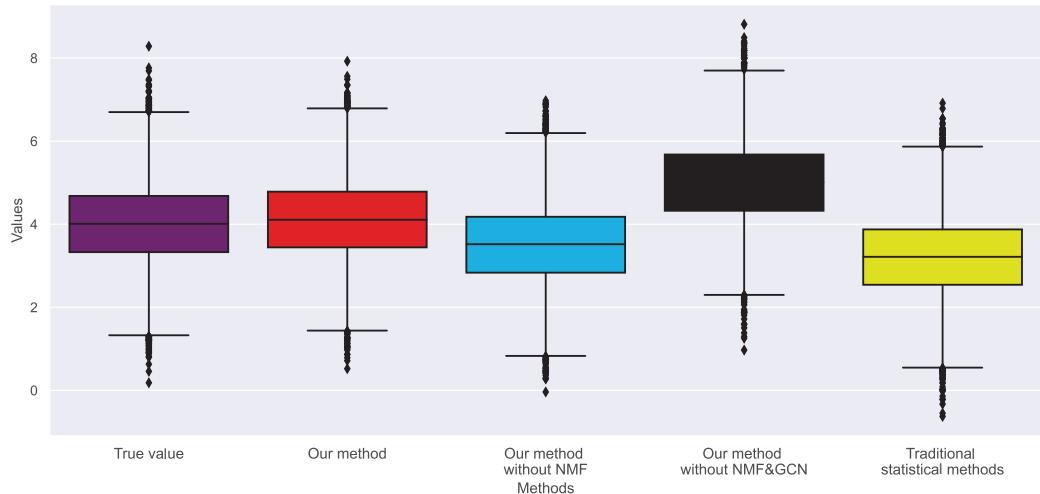


Figure 9 Rating changes in dataset 1.

Full-size DOI: 10.7717/peerj-cs.1858/fig-9

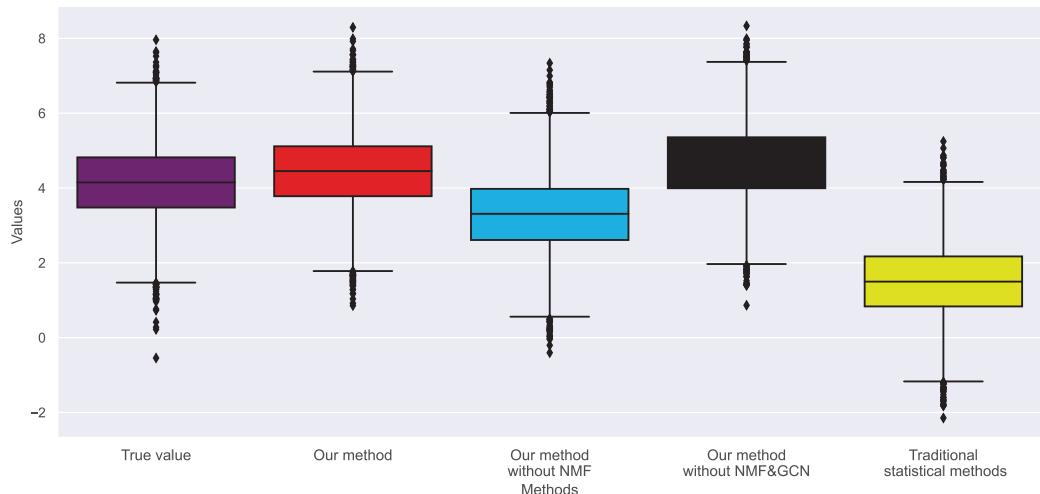


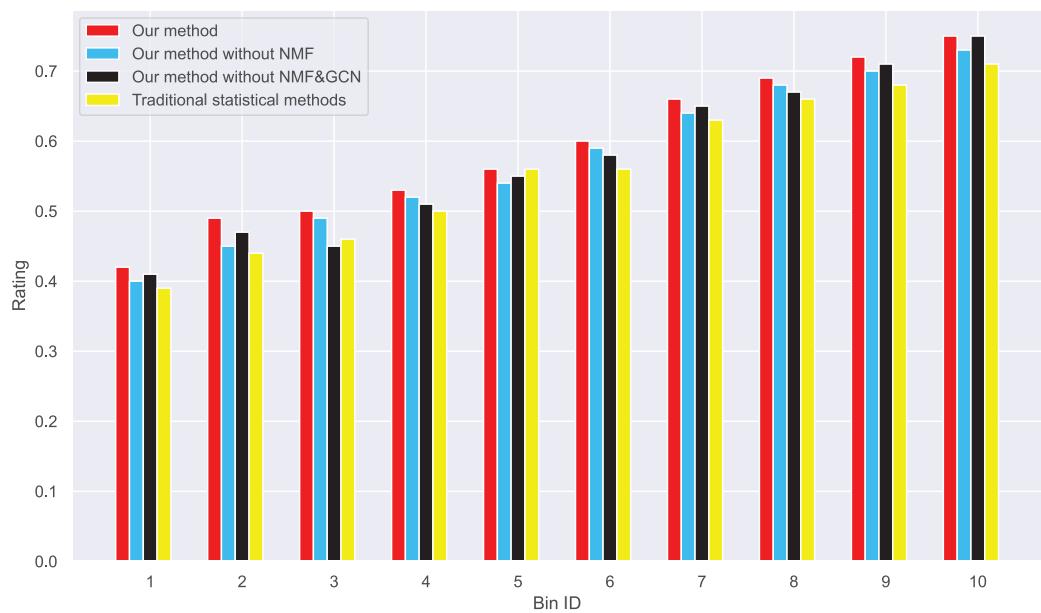
Figure 10 Rating changes in dataset 2.

Full-size DOI: 10.7717/peerj-cs.1858/fig-10

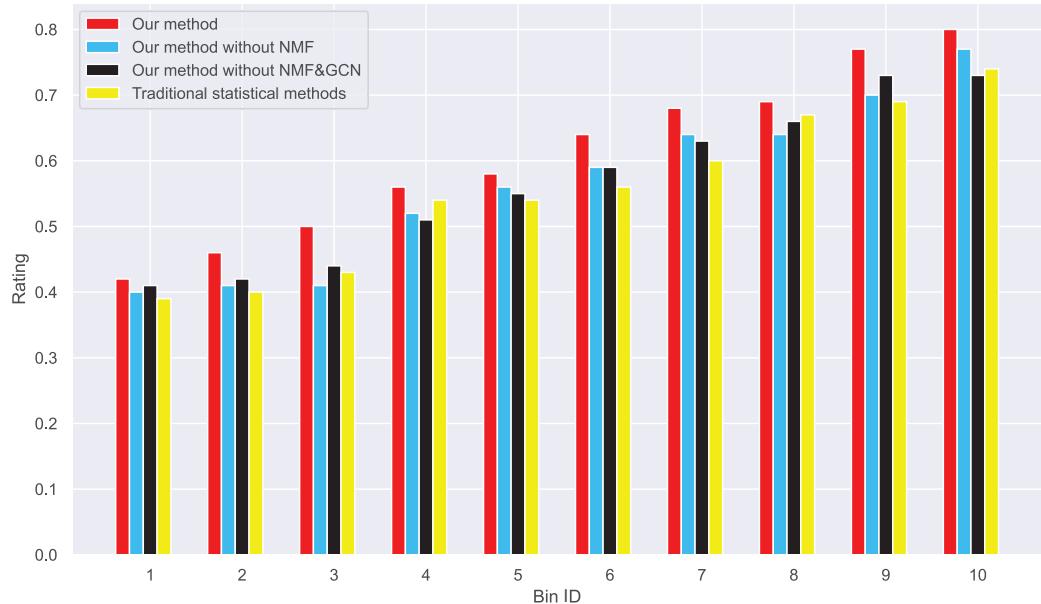
the true review distribution, the review distributions obtained by the other three methods show larger biases. This is because these three methods do not accurately calculate bias and cannot effectively detect and correct review bias.

Impact of the number of ratings

As previously mentioned, many children's books have relatively few ratings. There are many reasons for having fewer ratings. A book marked as "poor" by users may be read by fewer people, resulting in fewer ratings. On the other hand, when a book is rated as "excellent", more people will read or buy it, and the book receives more ratings. This behavior is roughly termed as *purchase bias*. Thus, books with many reviews tend to have high ratings, while books rated by only a few users generally have lower ratings. The experiment also revealed this behavior.

**Figure 11** Changes in real ratings and original average ratings in dataset 1.

Full-size DOI: 10.7717/peerj-cs.1858/fig-11

**Figure 12** Changes in real ratings and original average ratings in dataset 2.

Full-size DOI: 10.7717/peerj-cs.1858/fig-12

In Figs. 11 and 12, the real ratings and changes in original average ratings were obtained. As predicted by purchase bias, books with more ratings typically have higher ratings, but because they are very high and out of range, their absolute values do not signify anything. Instead, we plan to compare the relative rankings, which we will do later. Our method

Table 4 Top-ranked books and their IDs sorted by average rating.

Book ID	Mean rating	Traditional method	Our method			Target func	global min
			NMF No GCN	No NMF GCN	NMF GCN		
3280	1	4	12	101	1,417	4	
3233	2	2	3	3	4	2,021	
1830	3	6	7	7	9	19	
3881	4	7	9	10	33	1,641	
3656	5	5	8	8	15	2,688	
787	6	4	5	5	5	1,152	
3607	7	5	6	6	10	10	
3172	8	3	4	4	6	3,456	
3382	9	1	1	1	1	1	
989	10	1	2	2	3	3,369	

performs better than traditional statistical methods in terms of the relative rankings of recommended children's books.

Next, to be more persuasive, we compare the rankings of children's books obtained using various algorithms and our algorithm. In **Table 4**, books are ranked according to the average ratings they originally received. When our method is very close to the ranking by average rating, we do not expect a book with a high average rating to become very poor after bias is removed. There may be some differences, but we do not expect dramatic changes. We compared the relative rankings of books derived from our method and the traditional statistical method. Overall, we found that our method performed better than other methods in recommending the relative rankings of children's books.

In the following table, as evidenced by **Table 5** and mentioned earlier, many children's books have a very small number of ratings. Books with a large number of reviews tend to receive high ratings, while books rated by only a few users usually have lower ratings. As expected, these books have poor ratings, whether based on the average rating or our method. However, our algorithm shows some inconsistent results and ranks some poorly rated books very high. Our method shows more meaningful and stable rankings.

Lastly, when comparing the NDCG of the four methods, as illustrated in **Tables 6** and **7**, we found that our method consistently achieves the highest NDCG value across all rounds. Specifically, our method can maintain a high NDCG value even when dealing with many reviews. Even in situations with a high concentration of extreme ratings, this method effectively adjusts these extreme values, resulting in an NDCG distribution that more closely aligns with the true NDCG distribution. While the NDCG values for our method without NMF and our method without NMF & GCN are slightly decreased, they still generally outperform the traditional statistical methods. The NDCG value for traditional statistical methods is the lowest across all rounds, further proving the superiority of our method in terms of NDCG.

Table 5 Books that received only a 0–1 rating.

Book ID	Traditional statistical method	Our method			Mean	Objective func. global minima
		NMF No GCN	No NMF GCN	NMF GCN		
127	3,694	3,674	3,634	3,225	3,678	3,686
133	3,690	3,670	3,544	2,355	3,679	3,696
139	390	450	529	709	374	3,555
142	3,701	3,696	3,691	3,687	3,684	3,688
226	3,521	3,529	3,536	3,573	3,520	3,551
286	2,639	2,369	2,145	1,727	2,464	94
311	2,398	2,358	2,112	1,652	2,466	3,477
396	401	440	503	654	376	38
398	378	366	333	286	377	3,380
402	2,491	2,400	2,216	1,857	2,543	37

Table 6 Dataset 1 performed on NDCG comparison of four methods.

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	Bin 7	Bin 8	Bin 9	Bin 10
Our method	0.62	0.73	0.78	0.82	0.86	0.90	0.92	0.93	0.94	0.95
Our method without NMF	62	0.72	0.75	0.78	0.82	0.84	0.88	0.90	0.91	0.92
Our method without GCN	62	0.71	0.73	0.76	0.78	0.80	0.82	0.83	0.84	0.84
Traditional statistical methods	62	0.69	0.70	0.72	0.75	0.76	0.76	0.77	0.775	0.778
	Bin 11	Bin 12	Bin 13	Bin 14	Bin 15	Bin 16	Bin 17	Bin 18	Bin 19	Bin 20
Our method	0.96	0.97	0.98	0.98	0.98	0.985	0.985	0.986	0.986	0.986
Our method without NMF	0.93	0.94	0.95	0.95	0.955	0.96	0.96	0.962	0.964	0.966
Our method without GCN	0.84	0.85	0.85	0.855	0.86	0.862	0.864	0.866	0.868	0.87
Traditional statistical methods	0.783	0.788,	0.793	0.797	0.801	0.805	0.807	0.81	0.813	0.817

Table 7 Dataset 2 performed on NDCG comparison of four methods.

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	Bin 7	Bin 8	Bin 9	Bin 10
Our method	0.52	0.63	0.74	0.82	0.86	0.90	0.92	0.93	0.94	0.95
Our method without NMF	0.52	0.60	0.64	0.67	0.71	0.77	0.83	0.85	0.86	0.87
Our method without GCN	0.52	0.59	0.62	0.65	0.69	0.74	0.77	0.78	0.80	0.81
Traditional statistical methods	0.52	0.58	0.61	0.61	0.64	0.65	0.66	0.67	0.70	0.72
	Bin 11	Bin 12	Bin 13	Bin 14	Bin 15	Bin 16	Bin 17	Bin 18	Bin 19	Bin 20
Our method	0.96	0.97	0.97	0.975	0.976	0.977	0.978	0.98	0.982	0.984
Our method without NMF	0.88	0.89	0.90	0.905	0.91	0.915	0.92	0.922	0.924	0.926
Our method without GCN	0.82	0.825	0.83	0.835	0.84	0.842	0.844	0.846	0.848	0.85
Traditional statistical methods	0.73	0.735	0.74	0.742	0.744	0.746	0.748	0.75	0.752	0.754

Table 8 Overall model comparison.

Method	Handling large ratings	Handling extreme ratings	Bias removal	Accuracy
Our method	Efficient	Efficient	Yes	High
Our method without NMF	Medium	Below medium	Weak	Medium
Our method without GCN	Medium	Below medium	Weak	Low
Traditional statistical methods	Inefficient	Inefficient	No	Low

In summary, our method outperforms the other three methods in dealing with user review bias, particularly in situations involving large numbers and extreme reviews. Through various model comparisons, the accuracy of our method in book recommendation systems can be significantly improved, thereby helping parents make better decisions when choosing children's books. Therefore, we concluded that our method is effective in eliminating and addressing children's book bias and is an effective method for handling user review bias.

Performance of our method in recommendation

As evidenced earlier, we concluded that our method effectively mitigates and addresses bias in children's books. Therefore, we have also proposed a ranking solution to evaluate the recommendation performance of our model.

As mentioned before, our method outperforms the other three methods in dealing with user rating bias, particularly in the face of massive and extreme ratings. Firstly, our algorithm preliminarily screens and categorizes user ratings, effectively eliminating the influence of extreme and irrelevant ratings. Furthermore, we introduce our method, which deeply analyzes the sentiment of user ratings and conducts a deep analysis based on the books each user has bought previously. By our model, it can recommend a suitable children's book by inputting each user's previously purchased books, the type of books purchased, the child's age, the type of user's needs, educational level, *etc.* This would assist parents in making better decisions when choosing children's books. Parents no longer need to be troubled by a plethora of rating information or worry about making wrong choices due to extreme ratings. They can focus more on their children's needs and select truly appropriate books.

In general, the main considerations are the analysis of massive and extreme ratings and accuracy analysis for understanding users' true feelings about books, improving recommendation accuracy. As shown in the experimental result **Table 8**, the performance of our method is compared with the other three methods in a tabular format, clearly demonstrating that our method surpasses the other three methods in all aspects.

Our method effectively eliminates book bias for massive ratings, and it also plays a significant role in recommendations. All these notably improve the accuracy of the book recommendation system. In conclusion, our method successfully handles user rating bias and proves its superiority in experiments. This method of recommending children's books can assist parents in making better choices among a myriad of books, thereby effectively

meeting children's reading needs. We will continue to optimize this method to further enhance recommendation accuracy and provide parents with more practical book recommendation services.

CONCLUSION

This research introduces an innovative algorithm to quantitatively reveal user bias and eliminate it. Given that everyone carries subjectivity when evaluating a product, identifying and quantifying such bias is both necessary and challenging. Our algorithm successfully achieves this goal, thus unveiling the true rating of products. Another advantage of our algorithm is that by calculating bias and actual ratings, we can establish a clear connection with the ratings provided by users. This provides a practical, direct tool to quantitatively understand user feedback. In experimental evaluations, our method demonstrated excellent consistency and high-quality results, further validating its effectiveness. Our algorithm is also able to accurately recommend suitable books based on user needs.

In the future, we plan to further deepen our research, exploring how user bias evolves over time and how these changes affect product ratings. We believe that by delving deeper into and understanding the dynamics of user bias, we will be able to reveal a product's true value more accurately and provide more personalized and precise product recommendations.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The authors received no funding for this work.

Competing Interests

The authors declare that they have no competing interests. Jiang Liping is an employee of Guangxi Suhang Firefighting Technology Co., Ltd. This declaration does not imply a conflict of interest but is made in the interest of full transparency, to avoid any potential or perceived conflict of interest.

Author Contributions

- Lijuan Shen performed the experiments, performed the computation work, prepared figures and/or tables, and approved the final draft.
- Liping Jiang conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The data is available in Kaggle:

- MODHI ALMANNA. (2021, December). 1000+ Children Books, Version 1. Retrieved December 1, 2021 from <https://www.kaggle.com/datasets/modhiibrahimalmanna/1000-children-books-on-amazon>.

- ROHINI KRISHNAMOORTHY. (2022, December). Goodreads Children's book Dataset, Version 1. Retrieved December 3, 2022 from <https://www.kaggle.com/datasets/rohinikrishnamoorthy/goodreads-childrens-book-dataset>.

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj-cs.1858#supplemental-information>.

REFERENCES

- Almannaa M. 2021. 1000+ children books, version 1. Available at <https://www.kaggle.com/datasets/modhiibrahimalmannaa/1000-childrenbooks-on-amazon>.
- Chen J, Dong H, Wang X, Feng F, Wang M, He X. 2023a. Bias and debias in recommender system: a survey and future directions. *ACM Transactions on Information Systems* **41**(3):1–39 DOI [10.1145/3564284](https://doi.org/10.1145/3564284).
- Chen J, Wu J, Chen J, Xin X, Li Y, He X. 2023b. How graph convolutions amplify popularity bias for recommendation? ArXiv preprint DOI [10.48550/arXiv.2305.14886](https://doi.org/10.48550/arXiv.2305.14886).
- Chen J, Yu J, Lu W, Qian Y, Li P. 2021. IR-Rec: an interpretive rules-guided recommendation over knowledge graph. *Information Sciences* **563**(10):326–341 DOI [10.1016/j.ins.2021.03.004](https://doi.org/10.1016/j.ins.2021.03.004).
- Chen C, Zhang M, Zhang Y, Liu Y, Ma S. 2020. Efficient neural matrix factorization without sampling for recommendation. *ACM Transactions on Information Systems (TOIS)* **38**(2):1–28 DOI [10.1145/3373807](https://doi.org/10.1145/3373807).
- Curcic D. 2023. Children's Book Sales Statistics. *WordsRated*. Available at <https://wordsrated.com/children-book-sales/>.
- Daniels D, Salley B, Walker C, Bridges M. 2022. Parent book choices: how do parents select books to share with infants and toddlers with language impairment? *Journal of Early Childhood Literacy* **22**(2):279–307 DOI [10.1177/1468798420985668](https://doi.org/10.1177/1468798420985668).
- De Bondt M, Willenberg IA, Bus AG. 2020. Do book giveaway programs promote the home literacy environment and children's literacy-related behavior and skills? *Review of Educational Research* **90**(3):349–375 DOI [10.3102/0034654320922140](https://doi.org/10.3102/0034654320922140).
- Du Y, Jin X, Yan R, Yan J. 2023. Sentiment enhanced answer generation and information fusing for product-related question answering. *Information Sciences* **627**:205–219 DOI [10.1016/j.ins.2023.01.098](https://doi.org/10.1016/j.ins.2023.01.098).
- Ece Demir-Lira Ö, Applebaum LR, Goldin-Meadow S, Levine SC. 2019. Parents' early book reading to children: relation to children's later language and literacy outcomes controlling for other parent language input. *Developmental Science* **22**(3):e12764 DOI [10.1111/desc.12764](https://doi.org/10.1111/desc.12764).
- Garner PW, Parker TS. 2018. Young children's picture-books as a forum for the socialization of emotion. *Journal of Early Childhood Research* **16**(3):291–304 DOI [10.1177/1476718X18775760](https://doi.org/10.1177/1476718X18775760).
- He X, Deng K, Wang X, Li Y, Zhang Y, Wang M. 2020. LightGCN: simplifying and powering graph convolution network for recommendation. In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York: ACM, 639–648.
- Heidari N, Iosifidis A. 2021. Progressive graph convolutional networks for semi-supervised node classification. *IEEE Access* **9**:81957–81968 DOI [10.1109/ACCESS.2021.3085905](https://doi.org/10.1109/ACCESS.2021.3085905).
- Hennecke M, Czikmantori T, Brandstätter V. 2019. Doing despite disliking: self-regulatory strategies in everyday aversive activities. *European Journal of Personality* **33**(1):104–128 DOI [10.1002/per.2182](https://doi.org/10.1002/per.2182).

- Jin R, Xia T, Liu X, Murata T, Kim K-S. 2021.** Predicting emergency medical service demand with bipartite graph convolutional networks. *IEEE Access* **9**:9903–9915 DOI [10.1109/ACCESS.2021.3050607](https://doi.org/10.1109/ACCESS.2021.3050607).
- Juneja P, Mitra T. 2021.** Auditing e-commerce platforms for algorithmically curated vaccine misinformation. In: *Proceedings of the 2021 Chi Conference on Human Factors in Computing Systems*. 1–27.
- Kalkan İ.E, Şahin C. 2023.** Evaluating cross-selling opportunities with recurrent neural networks on retail marketing. *Neural Computing and Applications* **35**(8):6247–6263 DOI [10.1007/s00521-022-08019-1](https://doi.org/10.1007/s00521-022-08019-1).
- Kim J, Lee H, Hong S-E, Park S-H. 2022.** A bipartite graph neural network approach for scalable beamforming optimization. *IEEE Transactions on Wireless Communications* **22**(1):333–347 DOI [10.1109/TWC.2022.3193138](https://doi.org/10.1109/TWC.2022.3193138).
- Krishnamoorthy R. 2022.** Goodreads children's book dataset, version 1. Available at <https://www.kaggle.com/datasets/rohinikrishnamoorthy/goodreads-childrens-book-dataset>.
- Liu H, Lin H, Fan W, Ren Y, Xu B, Zhang X, Wen D, Zhao N, Lin Y, Yang L. 2022.** Self-supervised learning for fair recommender systems. *Applied Soft Computing* **125**(8):109126 DOI [10.1016/j.asoc.2022.109126](https://doi.org/10.1016/j.asoc.2022.109126).
- Liu Z, Qin C-X, Zhang Y-J. 2021.** Mining product competitiveness by fusing multisource online information. *Decision Support Systems* **143**(5):113477 DOI [10.1016/j.dss.2020.113477](https://doi.org/10.1016/j.dss.2020.113477).
- Mao Z, Du Z, Yuan R, Miao Q. 2022.** Short-term or long-term cooperation between retailer and MCN? New launched products sales strategies in live streaming e-commerce. *Journal of Retailing and Consumer Services* **67**(2):102996 DOI [10.1016/j.jretconser.2022.102996](https://doi.org/10.1016/j.jretconser.2022.102996).
- Mishra A. 2016.** An approach towards debiasing user ratings. ArXiv preprint DOI [10.48550/arXiv.1603.04482](https://arxiv.org/abs/1603.04482).
- Safeer AA, He Y, Abrar M. 2021.** The influence of brand experience on brand authenticity and brand love: an empirical study from Asian consumers' perspective. *Asia Pacific Journal of Marketing and Logistics* **33**(5):1123–1138 DOI [10.1108/APJML-02-2020-0123](https://doi.org/10.1108/APJML-02-2020-0123).
- Sarridis I, Kotropoulos C. 2021.** Neural factorization applied to interaction matrix for recommendation. In: *2021 29th European Signal Processing Conference (EUSIPCO)*. Piscataway: IEEE, 1336–1340.
- Smith A, Anderson M. 2016.** *2 Online Reviews*. Pew Research Center: Internet, Science & Tech, Pew Research Center.
- Statista. 2024.** Ebooks-worldwide: statista market forecast. Available at <https://www.statista.com/outlook/dmo/digital-media/epublishing/ebooks/worldwide> [Accessed 21 February 2024].
- Shahani R, Ahmed KA. 2022.** Investigating the moderating roles of basic and advanced financial literacy between behavioural biases and stock buying decisions. *Journal of Organisational Studies & Innovation* **9**(1):1–19 DOI [10.51659/josi.21-141](https://doi.org/10.51659/josi.21-141).
- Singh H, Chakrabarti S, Utkarsh. 2023.** How do gratifications to read reviews and perceived reviewers' credibility impact behavioural intentions in fashion e-commerce? A mediating-moderating perspective. *Computers in Human Behavior* **143**(4):107677 DOI [10.1016/j.chb.2023.107677](https://doi.org/10.1016/j.chb.2023.107677).
- Sochneva EN, Malakhova AA, Dobretsov GB, Rusakov AG, Zyablikov DV, Dudina SP, Faida VV, Malimonov IV, Kravtsov DI. 2022.** Efficiency of public procurement in Russia. In: *Computer Science On-line Conference*. Cham: Springer, 452–468.
- Sugiarto HS. 2022.** Graph convolution approach for labor market analysis. *Journal of Computational Science* **64**(2):101855 DOI [10.1016/j.jocs.2022.101855](https://doi.org/10.1016/j.jocs.2022.101855).

- Sun J, Zhang Y, Ma C, Coates M, Guo H, Tang R, He X.** 2019. Multi-graph convolutional collaborative filtering. In: *2019 IEEE International Conference on Data Mining (ICDM)*. Piscataway: IEEE, 1306–1311.
- Thanapalasingam T, van Berkel L, Bloem P, Groth P.** 2022. Relational graph convolutional networks: a closer look. *PeerJ Computer Science* 8(5):e1073 DOI [10.7717/peerj.cs.1073](https://doi.org/10.7717/peerj.cs.1073).
- Tran HNT, Thomas JJ, Malim NHAH.** 2022. DeepNC: a framework for drug-target interaction prediction with graph neural networks. *PeerJ* 10(11):e13163 DOI [10.7717/peerj.13163](https://doi.org/10.7717/peerj.13163).
- Trotzke P, Müller A, Brand M, Starcke K, Steins-Loeber S.** 2020. Buying despite negative consequences: interaction of craving, implicit cognitive processes, and inhibitory control in the context of buying-shopping disorder. *Addictive Behaviors* 110(3):106523 DOI [10.1016/j.addbeh.2020.106523](https://doi.org/10.1016/j.addbeh.2020.106523).
- Wang Y, Zhao Y, Zhang Y, Derr T.** 2023. Collaboration-aware graph convolutional network for recommender systems. In: *Proceedings of the ACM Web Conference 2023*. New York: ACM, 91–101.
- Wei Q, Hu G.** 2022. Evaluating graph neural networks under graph sampling scenarios. *PeerJ Computer Science* 8(2):e901 DOI [10.7717/peerj.cs.901](https://doi.org/10.7717/peerj.cs.901).
- Wu S, Sun F, Zhang W, Xie X, Cui B.** 2022. Graph neural networks in recommender systems: a survey. *ACM Computing Surveys* 55(5):1–37 DOI [10.1145/3535101](https://doi.org/10.1145/3535101).
- Xiao T, Shen H.** 2019. Neural variational matrix factorization for collaborative filtering in recommendation systems. *Applied Intelligence* 49(10):3558–3569 DOI [10.1007/s10489-019-01469-6](https://doi.org/10.1007/s10489-019-01469-6).
- Zhang R, Volkow ND.** 2019. Brain default-mode network dysfunction in addiction. *NeuroImage* 200:313–331 DOI [10.1016/j.neuroimage.2019.06.036](https://doi.org/10.1016/j.neuroimage.2019.06.036).