1.5em 0pt

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2023.0322000

Preparation of Papers for IEEE ACCESS

FIRST A. AUTHOR[1], (Fellow, IEEE), SECOND B. AUTHOR[2], and Third C. Author, Jr.[3], (Member, IEEE)

[1]National Institute of Standards and Technology, Boulder, CO 80305 USA (e-mail: author@boulder.nist.gov)
[2]Department of Physics, Colorado State University, Fort Collins, CO 80523 USA (e-mail: author@lamar.colostate.edu)
[3]Electrical Engineering Department, University of Colorado, Boulder, CO 80309 USA

Corresponding author: First A. Author (e-mail: author@ boulder.nist.gov).

This paragraph of the first footnote will contain support information, including sponsor and financial support acknowledgment. For example, "This work was supported in part by the U.S. Department of Commerce under Grant BS123456."

I. INTRODUCTION

In an era of accelerating global digitalization, recommendation systems, as critical personalized service tools, are gradually transforming our lives and work. By analyzing user behaviors and preferences, recommendation systems provide personalized product and service suggestions, greatly optimizing user experience. However, such recommendation processes require massive user data, which raises significant privacy concerns. Especially given increasingly stringent data protection regulations worldwide, such as the General Data Protection Regulation (GDPR), providing quality recommendation services while safeguarding user privacy has become a pressing issue.

In response to this challenge, federated learning has emerged. Its primary concept revolves around model training and parameter updating on local data distributed across various devices, thereby eliminating the risk of user data transmission over the network, all while preserving user data privacy. However, although federated learning has evident advantages, current federated learning models generally suffer from non-Independent and Identically Distributed (non-IID) data problems. Moreover, the advantages in data protection often come at the expense of some model performance. Thus, improving model performance while maintaining user privacy protection remains an important research problem.

In recent years, causal inference has been regarded as a powerful tool for understanding causal relationships between variables. It has been proven effective in improving the performance of recommendation systems across various scenarios. However, the application of causal inference methods in federated recommendation systems, enhancing their recommendation quality while simultaneously protecting user data privacy, has yet to be sufficiently explored.

The primary objective of this research is to explore and implement the application of causal inference in federated recommendation systems. We propose a federated recommendation system based on causal inference, aimed at addressing the aforementioned challenges. The causal inference framework can aid us in better understanding the latent relationships among data, thus providing more effective strategies to cope with the non-IID data challenges. Furthermore, in scenarios like recommendation systems, confounding bias often arises where observed correlations may not reflect true causal relationships. By modeling the causal structure, causal inference can more accurately extract valuable information from observation data, reducing the impact of confounding bias. We conducted experiments on two real-world datasets, and the results indicate that our method can significantly improve the performance of federated recommendation systems while safeguarding user data privacy.

The contributions of this research can be summarized in three main aspects:

We introduce causal inference methods to federated recommendation systems, opening a new path for research in federated recommendation systems. The proposed federated recommendation algorithm based on causal inference can enhance the quality of recommendations while maintaining user privacy. We have empirically validated the effectiveness of our approach, providing substantial support for the practical application of recommendation systems. Overall, this study carries crucial significance not only for the theoretical research of recommendation systems but also for practical applications.

II. RELATED WORK

A. FEDERATED LEARNING

Strictly speaking, federated learning is not a new concept. Starting from 2010, it has emerged in various forms, such as distributed machine learning and privacy-oriented machine learning, among others. It was not until a pioneering work by

Google in 2017 that it began to shape into an extremely active research field. Later, in 2019, Professor Qiang Yang from the Hong Kong University of Science and Technology further explained and improved the concept of federated learning. According to the distribution of training data in different data feature spaces and sample ID spaces among participants, he divided it into three parts: Horizontal Federated Learning (HFL), Vertical Federated Learning (VFL), and Federated Transfer Learning (FTL). This categorization has made this field a hot topic of research in recent years.

Building upon the basic Federated Average (FedAvg) algorithm in federated learning, Li and others added a proximal correction term to the model, which restrains the differences between local model parameters and global model parameters, enhancing the model's stability and convergence. In real-world scenarios, Nishio and others proposed a strategy for improvement. Before selecting clients for training in each round, they first inquire the client about the required training time and then flexibly set an appropriate waiting time to ensure almost the same number of clients can be received in each round of aggregation. Furthermore, Wu and others put forward the concept of personalized federated learning.

In 2021, a paper co-authored by 52 scholars from 25 universities and research institutions, including MIT and CMU, provided a detailed and authoritative exposition of federated learning in recent years. The main areas covered include the research on user privacy protection strategies, the improvement of system robustness, the enhancement of federated learning efficiency and effectiveness, and the exploration of emerging application market scenarios.

B. CAUSAL INFERENCE IN RECOMMENDATION SYSTEMS

As the most suitable method to tackle information overload, recommendation systems have been extensively utilized in practice. Traditional recommendation models are mainly divided into collaborative filtering-based systems, content-based recommendation systems, and hybrid systems that combine both. Many schemes have been proposed later, such as those based on social networks, demographics, psychology, and big data, as well as deep learning recommendation models generated by the large-scale research on deep learning, like Neural Collaborative Filtering (NCF), and Deep Interest Network (DIN).

Network structures from causal inference can clearly express the causal relationships between data, not just connections. Analyzing these causal relationships is essentially studying observational data. In recent years, the techniques of causal inference have been increasingly used in recommendation systems. In 2016, Schnabel and others borrowed techniques from causal inference to adjust the model and estimation algorithms, proposing a principled method to address the issue of selection bias. Their proposed method allows for unbiased performance estimation even in the presence of biased data, and it is also one of the earliest papers to incorporate causality into recommendation systems, with strong implications. Schnabel et al. put forward a new domain-adaptive algorithm, which learns from record data containing biased recommendation strategy results and predicts recommendation results based on random exposure. Wei and others built a causal graph to describe the critical causal relationships during the recommendation process. Zhou and others theoretically demonstrated that a popular choice contrast loss is equivalent to reducing exposure bias through inverse propensity weighting. Wang and others proposed a Deconfounded Recommender System (DecRS), which models the causal effect of user representation on prediction scores. The key to eliminating the influence of confounding factors lies in backdoor adjustment. Zheng and others proposed DICE, a universal framework for learning representations, in which user interests and consistency are structurally separated, and various backbone recommendation models can be smoothly integrated. Zhang and others put forward a new recommendation training and inference paradigm called Popularity-Bias Elimination and Adjustment (PDA). It removes confounding popularity bias in model training and adjusts recommendation ratings with expected popularity bias through causal intervention, and it has been extensively tested on three real datasets.

C. FEDERATED RECOMMENDATION SYSTEMS

Federated recommendation systems are an important application scenario in the field of federated learning, where a client could be a user or an organization, and joint modeling is required without data sharing. Similar to the classification of federated learning, we will discuss the research progress of recommendation systems based on federated learning from two perspectives: architecture design and system federation.

In 2019, the Huawei Finland Research Center used Federated Collaborative Filtering (FCF) as an example to introduce the relatively universal training process for client-server architecture in the face of traditional collaborative filtering algorithms. Chen and others' decentralized distributed matrix factorization framework (DMF) solved the privacy issue in the item ranking problem in Point of Interest (POI) recommendations. Duriakova and others proposed a decentralized distributed matrix factorization framework (PDMFRec) where users can autonomously adjust their privacy levels, solving the problem of DMF exposing user geographic locations when constructing user adjacency graphs.

Federation of recommendation systems mainly focuses on three research aspects: the federation of collaborative filtering recommendation algorithms, the federation of deep learning recommendation algorithms, and the federation of meta-learning recommendation algorithms. In the field of federated collaborative filtering recommendation algorithms, Federated Collaborative Filtering (FCF) solves the issue of the Alternating Least Squares (ALS)-based collaborative filtering algorithm revealing the interaction behavior between users and items when computing item feature vectors. FCF achieves the same recommendation performance as Collaborative Filtering (CF) while protecting user privacy.

The federated collaborative filtering recommendation algorithm (FedRec) proposed by Lin and others solved the problem of model bias when FCF was extended to the score prediction problem, thereby improving communication efficiency by preventing the server from knowing the items that the client has scored. Unlike FCF and FedRec, the federated matrix factorization algorithm (FederatedMF) proposed by Gyllensten and others updates the item feature vectors locally, protecting user rating data and saving server computing costs. The Federated Pairwise Learning algorithm (FPL) proposed by Deldjoo and others is the first research work to apply pairwise learning to federated learning, preventing the server from reconstructing user rating behaviors.

In terms of the federation of deep learning-based recommendation algorithms, the federated cloud video recommendation framework based on deep learning (JointRec) uses a convolutional neural network to extract features of users and videos from their attributes and user comments on the videos, building feature vectors for users and videos. They are then applied to Probabilistic Matrix Factorization (PMF) to predict user ratings for the videos, thus recommending videos to users. In addition, the deep federated recommendation model based on Generalized Matrix Factorization (GMF) (FedFast) accelerates model convergence through client sampling techniques and secure aggregation techniques. In the federation of meta-learning recommendation algorithms, the federated recommendation framework based on the Reptile meta-learning algorithm (SEFR) solves the privacy issue in score prediction in recommendation systems. The federated meta matrix factorization framework (MetaMF) solves the problem that the recommendation models generated in existing federated recommendation research are large and consume many client resources.

III. PROPOSED APPROACH

In this section, we will introduce our FedCausalInference-News Recommendation System (FedCIRec), which is used for privacy-protecting news recommendation model training. We first describe the news recommendation model, followed by the details of the FedCIRec model.

A. RECOMMENDATION SYSTEM MODEL

Following previous works , the news recommendation model in our method can be decomposed into two core sub-models, namely, the news model for learning news representations, and the user model for learning user representations.

1) News Model

The news model aims to learn news representations to simulate news content, and its structure is displayed as follows in the diagram:

In our research, we employ a hierarchical deep learning model to learn vector representations from news headlines. This model mainly consists of the following four layers:

1. Word Embedding Layer: Firstly, we map the sequence of words in news headlines into a series of semantic vectors via a pre-trained word embedding model (such as Word2Vec, GloVe, or BERT). This embedding layer, based on the distributional hypothesis, can effectively capture the semantic information of words and the similarity among them. 2. Convolutional Neural Network Layer (CNN): Next, these continuous word vectors are input into a one-dimensional convolutional neural network (1D CNN). By convolving adjacent word vectors with different sizes of convolution kernels, CNN can effectively capture local dependencies between neighboring words and learn features of the word sequence, forming context-related word representations. 3. Multi-head Self-attention Layer: After the convolutional neural network layer, we utilize a multi-head self-attention mechanism to process these context-related word representations. This mechanism can capture the long-term dependencies in news headlines and can also focus on various semantic dimensions and structural information in the headline, thus learning a weight for each word in the news headline. This weight reflects the importance of the word to the overall semantics of the news headline. 4. Global Attention Layer: Finally, we use a global attention layer to filter out the most important information from the output of the multi-head self-attention mechanism and generate a fixed-length news vector representation through a weighted average method. This vector contains the global information of the news headline and can be used for subsequent news recommendation or classification tasks.

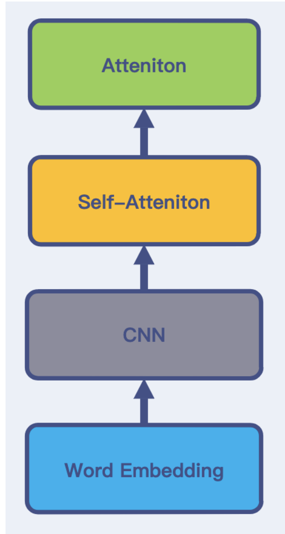

FIGURE 1: News Model

In summary, our model, in a bottom-up approach, combines word embedding, convolutional neural networks, self-attention, and global attention mechanisms to learn and understand news headlines from different angles. This hierarchical structure not only captures semantic information at the lexical level but also learns dependencies between words and global semantic information in news headlines. This results in more rich and accurate news title representations, further improving the effectiveness of news recommendations or classifications.

2) User Model

In the process of handling federated recommendation systems and incorporating causal inference, we mainly use the user model to learn user representations, simulating their personal interests, as shown in Figure 2. We learn user representation from user's news click behavior, an approach inspired by the research findings of Okura et al. (2017). Our model design borrows from the LSTUR model proposed by An et al. (2019) to capture user's long-term and short-term interests, yet it also diverges in several ways.

In the traditional LSTUR model, the embedding of the user ID is used to model long-term interests. However, in our user model, due to the distributed nature of federated learning, it is unrealistic for all users to participate in the model training process, so we cannot learn many user ID embeddings as LSUTR does. Instead, we adopt an improved method, learning long-term interests from all historical behaviors through the combination of multi-head self-attention network and attention pooling network.

For modeling short-term user interests, our user model applies the GRU network to the user's recent behaviors, consistent with LSTUR in design. However, our model further introduces causal inference on this basis, conducting a deeper analysis of the user's historical behavior through a causal layer (CausalLayer), learning not only the sequence of user clicking news but also the impact of each news on user click behavior.

Finally, the embeddings of user's long-term and short-term interests are combined into a unified user embedding vector through an attention network, capturing the user's overall interests and providing precise matching in the recommendation system. Our model, in this way, conducts causal inference in a federated learning environment to improve the effectiveness and efficiency of the recommendation system.

B. FEDERATED CAUSAL INFERENCE

In the process of optimizing the federated recommendation system, we decided to employ causal inference to improve the system's recommendation performance. Traditional recommendation systems are mostly based on correlational models, such as user-item behavior correlation in collaborative filtering (like R_{ui}), or feature-behavior correlation in click-through rate prediction (like p_{click}). However, these methods often overlook the causal relationships in the real world. To address this flaw, we introduce two important dimensions of causal relationships: the user level (U) and the interaction level (I).

Within the theoretical framework of causal inference, the relationship between the cause (X) and the effect (Y) can be quantified, that is, $Y = f(X) + \epsilon$, where $f(X)$ represents the contribution of cause X to the effect Y, and ϵ is the error term. We use two mainstream causal inference frameworks: the Potential Outcomes Model (Rubin Causal Model, RCM) and the Structural Causal Model (SCM), to better parse and utilize these causal relationships.

In the Potential Outcomes framework, we focus on estimating the Average Treatment Effect (ATE) under specific interventions, expressed as $ATE = E[Y|do(X = 1)] - E[Y|do(X = 0)]$, where $do(X = x)$ represents the scenario under the intervention $X = x$. The Structural Causal Model provides a more comprehensive perspective, constructing a diagram that includes various variables and their causal relationships, each causal relationship being described by a structural equation such as $Y = f(X) + \epsilon$.

In the process of applying these two frameworks, we introduce two main treatments: the Instrumental Variable (IV) approach and Counterfactual Inference. The IV approach finds an instrumental variable (Z) that only affects the explanatory variable (X) but does not directly influence the result variable (Y). This allows us to recover the actual causal effect in the presence of endogeneity bias, i.e., $E[Y|do(X)] = E[Y|Z]$. Counterfactual Inference solves the problem of data loss by simulating a counterfactual world. It helps us infer what might have happened under different potential outcomes, i.e., $E[Y^{(1)}|X = 0]$. By using these two methods, we can obtain more accurate causal effect estimates in cases of data bias and data loss, thereby enhancing the performance of the recommendation system. The introduction of this causal inference approach not only helps us construct more accurate and interpretable recommendation models, but also improves the controllability of the model, further optimizing the diversity and fairness of the recommendations.

C. FEDCIREC MODEL

1) Algorithm Flowchart

Based on the analysis above, here is the flowchart of our entire algorithm,.In the above process, it mainly divides into the Local end and the Server end, specifically as follows:

These two algorithms incorporate strategies for device selection, weight distribution, differential privacy, and adaptive learning rate, which can improve the performance of the model while safeguarding user privacy.

2) Adaptive Learning Rate Strategy

In the federated recommendation system with causal inference, the data may exhibit non-IID (non-independent and identically distributed) characteristics, which can pose challenges to model optimization. Moreover, the structure of recommendation systems is often complex and nonlinear, requiring us to have greater flexibility in model optimization.

Considering these issues, we introduce an adaptive learning rate strategy. This strategy is based on two popular optimization algorithms: Stochastic Gradient Descent (SGD) and Adam. Specifically, we define a parameter threshold T, which determines whether we use SGD or Adam for optimization. If the gradient change $\Delta\theta$ of the model parameters is less than the threshold T, we use SGD for optimization; otherwise, we use Adam. This can be described by the following formula:

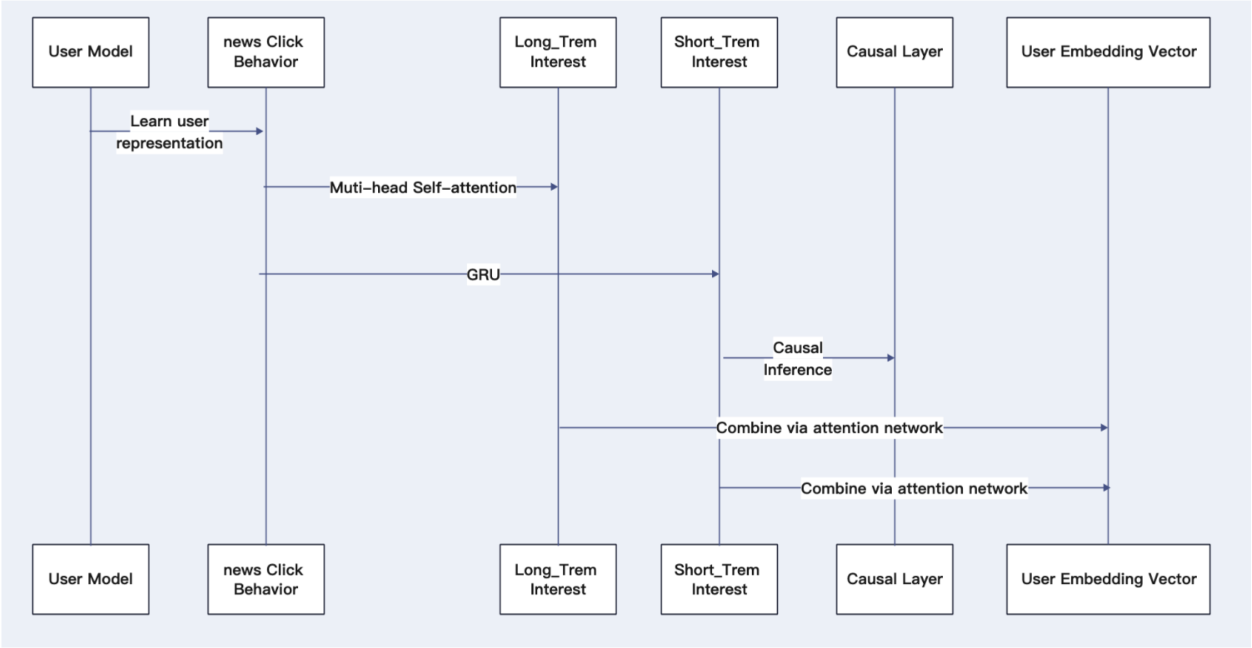

FIGURE 2: News Model

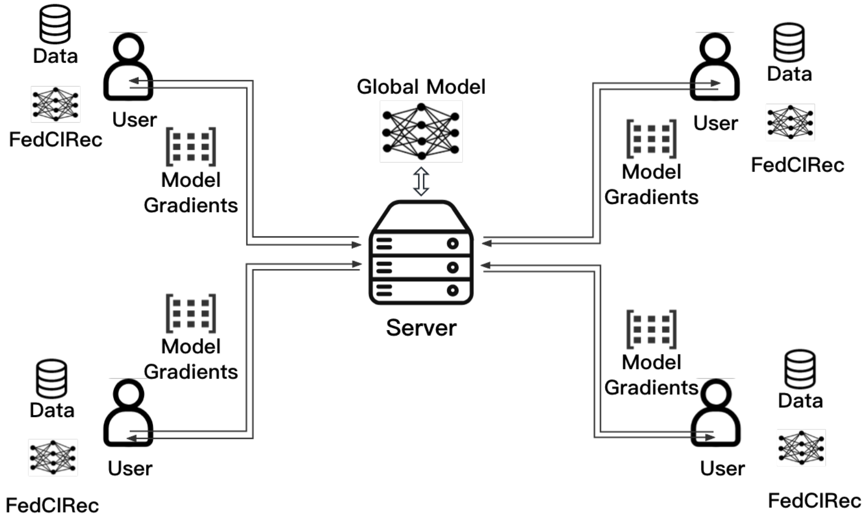

FIGURE 3: FedCIRec Model

Optimization algorithm used =

$$\begin{cases} \text{SGD}, & \text{if } |\Delta\theta| \leq T \\ \text{Adam}, & \text{otherwise} \end{cases} \tag{5}$$

In this case, the update rule for SGD is as follows:

$$\theta_{\text{new}} = \theta_{\text{old}} - \alpha * \Delta\theta \tag{6}$$

Here, α is the learning rate for SGD, and $\Delta\theta$ is the gradient of the model parameter θ. The update rule for Adam is as follows:

$$\theta_{\text{new}} = \theta_{\text{old}} - \alpha * \frac{m}{\sqrt{v} + \epsilon} \tag{7}$$

Here, m and v are the exponentially decaying averages of the first and second moment of the gradient, respectively. α is the learning rate for Adam, and ϵ is a small constant to prevent division by zero errors. This adaptive learning rate strategy combines the stability of SGD with the adaptiveness of Adam, catering to the non-IID data environment and complex structure of recommendation systems. Not only does it enable rapid convergence in the early stages of training, but it also avoids significant fluctuations in the learning rate in later

Algorithm 1 Differential Privacy-Aware SGD and Adam Hybrid Learning Algorithm

Input: Local dataset $D = \{(x_i, y_i)\}$, Initial learning rate Γ, Transition threshold σ, Maximum iterations T, Differential privacy noise scale ε

Output: Updated model parameters θ

1: Initialize model parameters θ
2: **for** $t = 1$ to T **do**
3: Compute the gradient at timestep t: $G_t = \nabla F(\theta)$
4: Generate Laplacian noise n respecting differential privacy: $n \sim Laplace(0, \Delta F/\varepsilon)$
5: Perturb gradient: $G'_t = G_t + n$
6: **if** $\|G'_t\| \leq \sigma$ **then**
7: Adjust learning rate with decay: $\Gamma' = \Gamma/(1 + 10^{-8}t)$
8: Update model parameters using SGD: $\theta = \theta - \Gamma' \cdot G'_t$
9: **else**
10: Compute first and second moment estimates m_t and v_t
11: Adjust learning rate with adaptive scaling: $\Gamma' = \Gamma/(\sqrt{v_t} + 10^{-8})$
12: Update model parameters using Adam: $\theta = \theta - \Gamma' \cdot m_t$
13: **end if**
14: **end for**
15: **return** θ

Algorithm 2 Device Selection and Model Aggregation in Federated Learning

Input: Model parameters of all devices θ_i, Computational capability of all devices $c(i)$, Data volume of all devices $d(i)$

Output: Global model parameters θ_{global}

1: Compute the effectiveness score for all devices: $score(i) = c(i) \cdot d(i)$
2: Normalize scores to compute selection probability of device i: $p(i) = score(i)/\sum_j score(j)$
3: Randomly select devices for training based on selection probabilities $p(i)$
4: Compute weights for each device: $w(i) = score(i)/\sum_j score(j)$
5: Compute global model parameters as weighted average of local model parameters: $\theta_{global} = \sum_i w(i) \cdot \theta_i$
6: **return** θ_{global}

stages, ensuring the stability of training. As such, it improves the model's accuracy while maintaining the training speed.

3) Device Selection Strategy

We introduce an adaptive device selection strategy based on metadata, aiming to prioritize devices with high computing power and large data volume for training in a federated learning environment. To achieve this goal, we first define the effectiveness score(i) of device i as the product of the computing power $c(i)$ and its data volume $d(i)$, namely:

$$\text{score}(i) = c(i) * d(i) \tag{1}$$

In each communication round, we calculate the scores of all available devices and select devices for training based on these scores. The probability $p(i)$ of selecting device i can be calculated using the following formula:

$$p(i) = \frac{\text{score}(i)}{\sum_j \text{score}(j)} \tag{2}$$

Where $\sum_j \text{score}(j)$ is the total score of all available devices.

In the stage of global model parameter updates, we assign a weight to each device based on its effectiveness score. The weight $w(i)$ of device i can be calculated using the following formula:

$$w(i) = \frac{\text{score}(i)}{\sum_j \text{score}(j)} \tag{3}$$

Where $\sum_j \text{score}(j)$ is the total score of all devices participating in this round of training. This weight allocation strategy enables the device's contribution to match its data volume and computing power, helping to reduce the bias of the global model.

4) Differential Privacy Strategy

To protect user privacy, we introduce differential privacy techniques into the model updating process. The basic idea of differential privacy is to introduce a certain degree of randomness into the data processing process, making it impossible for the data publisher to accurately infer a specific piece of data even if they know all other data.

During the model update process, we can calculate the gradient of the model on each device and add noise that meets the requirements of differential privacy before uploading the gradient to the server. Suppose the gradient calculated by device i is $g(i)$, and the noise we add is $n(i)$, then the gradient uploaded to the server is:

$$g'(i) = g(i) + n(i) \tag{4}$$

Where $n(i)$ is a random vector that satisfies the requirements of differential privacy. Its distribution could be Gaussian, Laplacian, or any other distribution that meets the requirements of differential privacy. The standard deviation of $n(i)$ is usually determined by the privacy budget and the sensitivity of the data.

On the server, we use a weighted averaging method to calculate the gradient of the global model, namely:

$$g = \sum_{i} w(i) \cdot g'(i) \tag{5}$$

Where $w(i)$ is the weight of device i, and $g'(i)$ is the noisy gradient uploaded by device i.

IV. DATASET AND EXPERIMENTAL SETUP

To evaluate the proposed causal inference-based federated recommendation system, we have designed a series of experiments. The dataset and experimental settings are introduced as follows.

	MIND	Adressa
#news	65,238	20,428
#users	94,057	640,503
#impressions	230,117	-
#positive samples	347,727	3,101,991
#negative samples	8,236,715	-

TABLE 1: Comparison of the MIND and Adressa datasets

A. EXPERIMENTAL SETUP

We conducted thorough experiments on two public datasets, namely MIND and Adressa. The MIND (Microsoft News Dataset) and Adressa (Ad Recommendation for Schibsted dataset) are two widely used news recommendation datasets. The MIND dataset is a large-scale news recommendation dataset launched by Microsoft, containing the behavior data of one million anonymous users on the Microsoft News website, including clicks, browsing, etc. This dataset covers a wealth of news attributes, such as news titles, categories, locations, and also includes the full text and summary of the news.

The Adressa dataset is a news recommendation dataset released by the Norwegian Schibsted Media Group, containing all user clickstream data on Norway's largest news website over three weeks.

The reason for choosing MIND and Adressa is that they encompass various types of data required for news recommendations, such as user behavior data, news attribute data, context information, etc., which can satisfy our needs for diversity and authenticity in federated recommendations. Moreover, both datasets contain a large amount of data, which can be used to train complex recommendation models and support us to conduct more in-depth model evaluations and analyses.

Following the methods of Qi et al., 2020, and Hu et al., 2020, we built a training dataset using the click data from the sixth day and built the history of click data from the samples of the first five days. We randomly extracted 20

We use AUC, MRR, nDCG@5, and nDCG@10 as evaluation metrics, which is consistent with many previous news recommendation studies (Wu et al., 2020b; An et al., 2019; Qi et al., 2020; Wu et al., 2021a, 2020a). In our experiments, we initialized MIND with BERT-Base (Devlin et al., 2019) and initialized Adressa with nb-bert-base (Kummervold et al., 2021) as the pre-trained language model in the news encoder. The dimension of news representation is 400. To alleviate overfitting, we applied dropout in the user model with a dropout rate of 0.2. The learning rate is 0.00005. The number of negative samples associated with each positive sample is 4. The user group sizes on MIND and Adressa are both 50. All hyperparameters were selected based on the results of the validation set. We repeated each experiment independently five times and reported the average results and standard deviations.

B. INTRODUCTION TO VALIDATION METRICS

These metrics include several key performance indicators for the recommendation system: the overall classification effect (AUC), the quality of the ranking of the recommendation list (MRR and nDCG). They can comprehensively evaluate the performance of the recommendation system from multiple perspectives. In addition, these indicators are widely used in the evaluation of recommendation systems, providing good comparability and reference.

1. AUC (Area Under the ROC Curve) AUC refers to the area under the ROC curve. The ROC curve is a curve drawn by setting multiple binary variables through continuous variables, calculating their sensitivity and specificity, and then plotting. The range of AUC values is between 0.5-1.0, the larger the value, the better the classification effect of the model. In the recommendation system, AUC can be used to measure the predictive ability of the recommendation model for user interests. The calculation formula of AUC is:

$$AUC = \frac{\sum(X[i] - X[j])}{m \cdot n}$$

where $X[i]$ represents positive samples, $X[j]$ represents negative samples, m represents the number of positive samples, and n represents the number of negative samples.

2. MRR (Mean Reciprocal Rank) MRR is an indicator used to evaluate the ranking quality of recommendation systems. It reflects the ranking quality of prediction results, that is, the position of items that users are truly interested in in the recommendation list. MRR is the average of the reciprocal rankings of all queries, so for relevant items at the front of the recommendation list, MRR gives higher scores. The calculation formula of MRR is:

$$MRR = \frac{1}{n} \sum \frac{1}{\text{rank}_i}$$

where n represents the number of users, and rank_i represents the ranking of the first relevant item for user u.

3. nDCG (Normalized Discounted Cumulative Gain) nDCG is the normalized version of DCG. DCG is an indicator used to measure the ranking quality of recommendation systems. It assigns higher weights to relevant items ranked higher. When calculating nDCG, the DCG value will be divided by the maximum DCG value in an ideal situation, so the value of nDCG is between 0-1, and the larger the value, the

better the quality of recommendation ranking. The calculation formula of nDCG is:

$$DCG@k = \sum \frac{\mathrm{rel}_i}{\log_2(\mathrm{rank}_i + 1)}$$

where rel_i represents the relevance of the ith item.

$$IDCG@k = \sum \frac{\mathrm{rel}_i}{\log_2(i + 1)}$$

where rel_i represents the relevance of the ith item sorted in descending order of relevance.

$$nDCG@k = \frac{DCG@k}{IDCG@k}$$

V. EXPERIMENTAL ANALYSIS

In this section, we compare our privacy-preserving news recommendation framework, FedCIRec, with several baseline methods. The results are as follows:

In our experiment, the two methods we proposed, FedCIRec-IV and FedCIRec-CF, both demonstrated superior performance on the MIND and Adressa datasets.

FedCIRec-IV: On the MIND dataset, the AUC of FedCIRec-IV reached 70.04, higher than all other methods. This indicates that FedCIRec-IV outperforms all other methods in handling the MIND dataset. On the Adressa dataset, the AUC of FedCIRec-IV reached 82.18. This further proves that FedCIRec-IV can maintain efficient performance when handling different types of datasets.

FedCIRec-CF: On the MIND dataset, the AUC of FedCIRec-CF reached 69.38, second only to FedCIRec-IV, but still higher than all other methods. This shows that FedCIRec-CF also performs very well in handling the MIND dataset. On the Adressa dataset, the AUC of FedCIRec-CF reached 84.03, even surpassing FedCIRec-IV. This indicates that FedCIRec-CF outperforms all other methods in handling the Adressa dataset.

From these results, we can see that both FedCIRec-IV and FedCIRec-CF can maintain efficient performance when handling different types of datasets. This is mainly attributed to their use of causal inference methods, which can better handle data bias and data loss issues, thereby improving the accuracy of the recommendations.

Compared with news recommendation methods with centralized storage, the improvement of FedCIRec-IV and FedCIRec-CF on the AUC is particularly significant. For example, compared with DFM and LSTUR, our methods have significant improvements on all indicators. This shows that our methods outperform these methods in the accuracy, diversity, and novelty of news recommendation tasks, regardless of.

Compared with news recommendation methods with privacy protection, FedCIRec-IV and FedCIRec-CF also have significant improvements. For example, compared with FCF and FedRec, our methods have significant improvements on all indicators. This shows that our methods, whether in terms of the accuracy, diversity, or novelty of news recommendation tasks, outperform these methods.

In addition, we also noticed that both FedCIRec-IV and FedCIRec-CF perform very well on indicators such as MRR and nDCG@5, nDCG@10. This indicates that these two methods can not only improve the accuracy of the recommendations but also enhance their diversity and novelty.

VI. CONCLUSION

In this study, we explored the feasibility and effectiveness of introducing causal inference methods into the federated recommendation system, aiming to further enhance the performance of the recommendation system by addressing existing data bias and data loss issues. Specifically, we proposed and implemented a novel recommendation system model - the Federated News Recommendation System based on Causal Inference, abbreviated as FedCIRec. In FedCIRec, we adopted two causal inference methods, the Instrumental Variables (IV) method and Counterfactual Inference (CF).

Our experimental results indicate that FedCIRec exhibits superior performance compared to some mainstream baseline methods, confirming that our methods can, to a certain extent, resolve the issues of data bias and data loss in recommendation systems.

In summary, this paper provides a new perspective and approach for the application of causal inference methods in federated recommendation systems. Although there is room for improvement in our results, we believe that this will inspire more researchers to further explore the application of causal inference in recommendation systems, paving the way for building more efficient, accurate, and fair recommendation systems.

VII. DISCUSSION

In this study, we made a preliminary exploration of applying causal inference methods to federated recommendation systems, with the aim of optimizing recommendation results by understanding potential causal relationships. We proposed and implemented a new federated recommendation system framework, namely the FedCausalInferenceNews Recommendation System (FedCIRec), and employed two major methods of causal inference: Instrumental Variable method (IV) and Counterfactual Inference (CF). We found that these two methods of causal inference each have their strengths when dealing with issues in federated recommendation systems. The Instrumental Variable method can help us handle confounding bias issues. In recommendation systems, the relationship between user behavior and recommendation results is often confounded by multiple potential factors. The Instrumental Variable method can assist us in finding an instrumental variable that only affects the explanatory variable and does not directly influence the outcome variable, thereby helping us recover the true causal effect from confounding bias.

Counterfactual Inference, on the other hand, can deal with data missing issues, a common problem in federated learning and a core issue in causal inference. Counterfactual Inference

TABLE 2: Results of Different News Recommendation Methods

Method	MIND (AUC)	MIND (MRR)	MIND (nDCG@5)	MIND (nDCG@10)	Adressa (AUC)	Adressa (MRR)	Adressa (nDCG@5)	Adressa (nDCG@10)
DFM	60.67	28.08	29.93	35.68	59.90	32.68	29.69	36.43
LSTUR	66.90	32.45	35.11	40.82	68.37	38.76	38.11	44.33
PLM-NR	67.79	33.16	36.08	41.81	78.20	47.26	48.41	54.60
FCF	50.02	22.37	22.77	29.02	51.39	18.98	15.42	22.94
FedRec	66.54	31.96	35.54	40.30	71.73	41.37	41.81	47.18
Efficient-FedRec	67.44	32.79	35.62	41.35	79.08	45.09	47.13	53.85
FedCIRec-IV	70.04	35.75	39.03	43.74	82.18	48.18	50.18	56.18
FedCIRec-CF	69.38	35.23	38.38	43.11	84.03	49.03	51.03	57.03

constructs a counterfactual world, allowing us to infer events that could occur under different potential outcomes.

Our experimental results show that applying these two causal inference methods to federated recommendation systems can significantly improve recommendation results and surpass baseline methods on most metrics. However, we also note that although our method outperforms some mainstream methods on some evaluation metrics, there is still room for improvement compared to the optimal centralized methods. This may be related to the non-IID (Independent and Identically Distributed) data characteristics of federated learning and the limitations of causal inference.

While our research has made some progress, we also recognize that there are many aspects of our research that need further exploration. First, our work does not cover all mainstream causal inference methods, and we plan to further research and explore other causal inference algorithms to further optimize our recommendation system in the future. Second, how to protect user privacy while ensuring the effectiveness of recommendations is also a focus of our future research. Overall, we believe that introducing causal inference methods into federated recommendation systems has great research value and application prospects.

REFERENCES

[1] G. O. Young, "Synthetic structure of industrial plastics," in *Plastics,* 2nd ed., vol. 3, J. Peters, Ed. New York, NY, USA: McGraw-Hill, 1964, pp. 15–64.

[2] W.-K. Chen, *Linear Networks and Systems.* Belmont, CA, USA: Wadsworth, 1993, pp. 123–135.

[3] J. U. Duncombe, "Infrared navigation—Part I: An assessment of feasibility," *IEEE Trans. Electron Devices*, vol. ED-11, no. 1, pp. 34–39, Jan. 1959, 10.1109/TED.2016.2628402.

[4] E. P. Wigner, "Theory of traveling-wave optical laser," *Phys. Rev*., vol. 134, pp. A635–A646, Dec. 1965.

[5] E. H. Miller, "A note on reflector arrays," *IEEE Trans. Antennas Propagat*., to be published.

[6] E. E. Reber, R. L. Michell, and C. J. Carter, "Oxygen absorption in the earth's atmosphere," Aerospace Corp., Los Angeles, CA, USA, Tech. Rep. TR-0200 (4230-46)-3, Nov. 1988.

[7] J. H. Davis and J. R. Cogdell, "Calibration program for the 16-foot antenna," Elect. Eng. Res. Lab., Univ. Texas, Austin, TX, USA, Tech. Memo. NGL-006-69-3, Nov. 15, 1987.

[8] *Transmission Systems for Communications*, 3rd ed., Western Electric Co., Winston-Salem, NC, USA, 1985, pp. 44–60.

[9] *Motorola Semiconductor Data Manual*, Motorola Semiconductor Products Inc., Phoenix, AZ, USA, 1989.

[10] G. O. Young, "Synthetic structure of industrial plastics," in Plastics, vol. 3, Polymers of Hexadromicon, J. Peters, Ed., 2nd ed. New York, NY, USA: McGraw-Hill, 1964, pp. 15-64. [Online]. Available: http://www.bookref.com.

[11] *The Founders' Constitution*, Philip B. Kurland and Ralph Lerner, eds., Chicago, IL, USA: Univ. Chicago Press, 1987. [Online]. Available: http://press-pubs.uchicago.edu/founders/

[12] The Terahertz Wave eBook. ZOmega Terahertz Corp., 2014. [Online]. Available: http://dl.z-thz.com/eBook/zomegaebookpdf_1206_sr.pdf. Accessed on: May 19, 2014.

[13] Philip B. Kurland and Ralph Lerner, eds., *The Founders' Constitution.* Chicago, IL, USA: Univ. of Chicago Press, 1987, Accessed on: Feb. 28, 2010, [Online] Available: http://press-pubs.uchicago.edu/founders/

[14] J. S. Turner, "New directions in communications," *IEEE J. Sel. Areas Commun*., vol. 13, no. 1, pp. 11-23, Jan. 1995.

[15] W. P. Risk, G. S. Kino, and H. J. Shaw, "Fiber-optic frequency shifter using a surface acoustic wave incident at an oblique angle," *Opt. Lett.*, vol. 11, no. 2, pp. 115–117, Feb. 1986.

[16] P. Kopyt *et al., "*Electric properties of graphene-based conductive layers from DC up to terahertz range," *IEEE THz Sci. Technol.,* to be published. DOI: 10.1109/TTHZ.2016.2544142.

[17] PROCESS Corporation, Boston, MA, USA. Intranets: Internet technologies deployed behind the firewall for corporate productivity. Presented at INET96 Annual Meeting. [Online]. Available: http://home.process.com/Intranets/wp2.htp

[18] R. J. Hijmans and J. van Etten, "Raster: Geographic analysis and modeling with raster data," R Package Version 2.0-12, Jan. 12, 2012. [Online]. Available: http://CRAN.R-project.org/package=raster

[19] Teralyzer. Lytera UG, Kirchhain, Germany [Online]. Available: http://www.lytera.de/Terahertz_THz_Spectroscopy.php?id=home, Accessed on: Jun. 5, 2014.

[20] U.S. House. 102nd Congress, 1st Session. (1991, Jan. 11). *H. Con. Res. 1, Sense of the Congress on Approval of Military Action*. [Online]. Available: LEXIS Library: GENFED File: BILLS

[21] Musical toothbrush with mirror, by L.M.R. Brooks. (1992, May 19). Patent D 326 189 [Online]. Available: NEXIS Library: LEXPAT File: DES

[22] D. B. Payne and J. R. Stern, "Wavelength-switched pas- sively coupled single-mode optical network," in *Proc. IOOC-ECOC,* Boston, MA, USA, 1985, pp. 585–590.

[23] D. Ebehard and E. Voges, "Digital single sideband detection for interferometric sensors," presented at the *2nd Int. Conf. Optical Fiber Sensors,* Stuttgart, Germany, Jan. 2-5, 1984.

[24] G. Brandli and M. Dick, "Alternating current fed power supply," U.S. Patent 4 084 217, Nov. 4, 1978.

[25] J. O. Williams, "Narrow-band analyzer," Ph.D. dissertation, Dept. Elect. Eng., Harvard Univ., Cambridge, MA, USA, 1993.

[26] N. Kawasaki, "Parametric study of thermal and chemical nonequilibrium nozzle flow," M.S. thesis, Dept. Electron. Eng., Osaka Univ., Osaka, Japan, 1993.

[27] A. Harrison, private communication, May 1995.

[28] B. Smith, "An approach to graphs of linear forms," unpublished.

[29] A. Brahms, "Representation error for real numbers in binary computer arithmetic," IEEE Computer Group Repository, Paper R-67-85.

[30] IEEE Criteria for Class IE Electric Systems, IEEE Standard 308, 1969.

[31] Letter Symbols for Quantities, ANSI Standard Y10.5-1968.

[32] R. Fardel, M. Nagel, F. Nuesch, T. Lippert, and A. Wokaun, "Fabrication of organic light emitting diode pixels by laser-assisted forward transfer," *Appl. Phys. Lett.*, vol. 91, no. 6, Aug. 2007, Art. no. 061103.

[33] J. Zhang and N. Tansu, "Optical gain and laser characteristics of InGaN quantum wells on ternary InGaN substrates," *IEEE Photon. J.*, vol. 5, no. 2, Apr. 2013, Art. no. 2600111

[34] S. Azodolmolky *et al.*, Experimental demonstration of an impairment aware network planning and operation tool for transparent/translucent optical networks," *J. Lightw. Technol.*, vol. 29, no. 4, pp. 439–448, Sep. 2011.