Journal of Theoretical and Applied Information Technology
30th September 2025. Vol.103. No.18
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

FedDNN-LDP: PRIVACY-PRESERVING FEDERATED RECOMMENDER SYSTEM USING DEEP NEURAL NETWORK AND LOCAL DIFFERENTIAL PRIVACY

THENMOZHI GANESAN[1*], PALANISAMY VELLAIYAN[2]

[1]Ph.D., Research Scholar, Department of Computer Applications, Alagappa University, Karaikudi, Tamilnadu, India
[2]Senior Professor and Head, Department of Computer Applications, Alagappa University, Karaikudi, Tamilnadu, India
E-mail: *[1]thenmozhiganesan23@gmail.com, [2]palanisamyv@alagappauniversity.ac.in

ABSTRACT

Federated recommender system utilizes federated learning to preserve user data in local and centralized system by training intermediate parameters instead of training raw user data. Currently, deep neural network gaining significant attention in recommender system due to its efficiency in processing in massive training samples and capturing intricate user-item interactions. However, global sharing of the entire user-item interaction in centralized network is prevented and limited for privacy concern. To address this gap, several integrated techniques utilizes the feature of pseudo-interaction of users and items in the neural network to recompense the missing values for each user and item, which adds random noise to the model and raise the privacy threatening in the network. This research proposes FedDNN-LDP, a novel federated deep neural recommendation system where centralized deep neural network is formed and preserved to enable the intricate user-item interactions on the server-end. On the client-end, to preserve the server from leaning user local data, pseudo-interaction is applied on the user data for data obfuscation. Furthermore, for privacy concern of local user-item interaction and obscure intermediate gradient parameters, integration of pseudo labeling and local differential privacy is exploited. Extensive experiments performed on two real world benchmark datasets namely MovieLens 100k and MovieLens 1M and experimental results show the outperformance of proposed model with existing competitive approaches ensuring privacy preservation by the following performance measures with the improvement: NDCG (15.47%), precision (30.06%) and recall (28.66%) for 100K dataset and NDCG (32.54%), precision (58.38%) and recall (40.86%) for 1M dataset. Further root mean squared error, mean absolute error, training and validation loss and hit ration are measured to exhibit the effectuality of the proposed model in terms of performance and privacy preservation.

Keywords: *Federated Recommender System, Deep Neural Network, Pseudo Labeling, Local Differential Privacy, Recommender System and Privacy Preservation.*

1. INTRODUCTION

In this digital era, massive volume of data is processed at every fraction of second. It is tedious to find the relevant information from huge data which causes information overload problem [1]. Recommender system has proven as an effectual solution to alleviate the information overloaded issue by predicting and suggesting user preferences based on their behaviors and past history [2]. To provide the personalized recommendations to the client, recommendation system utilizes relevant features and user interest, making the progress of finding most suitable items or services more effectively and conveniently with effortlessness [3]. Attaining this, conventional recommender system necessitates storing and processing tons of user's private data that made the user privacy at risk and raised the threats about user's privacy concerns [4].With the escalating alertness of privacy and the implementation of pertinent regulations such as the General Data Protection Regulation (GDPR).

To alleviate the key issue, federated learning is become a promising strategy in which user's intermediate gradients are exploited instead of sharing raw user data. Federated learning ensures the user privacy by localizing the data and then

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

transferring model gradients to the centralized server [5]. Integrating conventional recommendation system with the federated learning constitutes the thirst and novel research field to the researchers named federated recommender system. Federated recommender system is capable of aggregated and functioned with various extended area of federated collaborative filtering, matrix factorization and clustering, graph convolutional and neural network [6].

Deep learning techniques are the subset of machine learning capable of learn and extract relevant features from raw user data to construct such models to handle the complex and high dimensional datasets. This characteristic of deep learning model is similar to the functionality of recommender system where personalized recommendations are generated depends on the user preferences [7] . Furthermore, deep neural networks are composite in terms of several neural building blocks can be compiled into a single differentiable function involving more number of hidden layers with end-to-end training.

With this motivation and advancement in privacy preservation of federating learning and deep neural network, this research implements the collaborative approach of deep neural network with federated learning to formulate robust deep neural federated recommender system. Recent research on privacy concern federated recommender system demonstrated the sturdiness and feasibility of deep neural architectures in this domain. In this work, we integrated the deep neural network with local differential privacy and pseudo labeling approaches for ensured security of user data in the centralized network. In addition this research is extended into the federated learning environment which transmits the weights calculated by neural network model and model embeddings to the centralized server instead of raw data. Contributions of this research are as follows:

- We propose a deep neural recommender system incorporated with the federated learning paradigm which exploits intermediate gradient parameters abstain from user's raw data collection in centralized environment for privacy preservation.
- We exploit pseudo-interaction approach to preserve user-item interactions for each user by producing tuple of pseudo interaction erratically to prevent the precise inference of such user-item interaction record.
- Concurrently, local differential privacy (LDP) is implemented to assure the protection of all uploaded model gradients and embeddings which promote the privacy of federated learning paradigm.
- Efficacy of the proposed model is evaluated using two benchmark datasets and compared with existing approaches to validate the performance.

Rest of the paper organized as follows: Section 2 reviews the related works. Section 3 discussed the preliminaries. Proposed methodology is explained in section 4. Section 5 illustrated the experimental analysis. Section 6 demonstrated the result and discussion of this work and the conclusion is presented in section 7.

2. RELATED WORKS

Federated learning is initiated by Google to ensure user privacy and information security in the distributed environment. It further classified as vertical federated learning and horizontal federated learning. The collaboration of federalization of recommender system with deep neural network inevitably provide significant improvement in personal privacy and security concern in centralized environment studied by various authors discussed as follows.

Neural collaborative filtering (NCF) technique is studied in [8] which utilize the features of both matrix factorization and multi-layer perceptron to non-linear feature representations of higher-order feature to the potential of model description. For simple and non-complex architecture, the studied model is subdivided into three parts as GMF, MLP and NeuMF.

In [9], deep interest network is studied in click through rate prediction method where the sparse inputs of large scale dimensions are identified and mapped as low dimensional vectors. Then the low dimensional vectors are fixed into the cluster of fixed-length vectors to condense user features into fixed-length representation. Efficiency of the model is evaluated on Alibaba dataset and compared with existing deep neural approaches such as DeepFM and Wide & Deep algorithm.

Combination of wide and deep learning is proposed in [10] which train wide linear models and deep neural networks simultaneously to yield the dual benefits of generalization and memorization for the

personalized recommendation system. Author evaluated and illustrated the efficacy of the proposed model by comparing with competitive models involving deep only algorithm and wide only algorithm.

In [11], generic collaborative federated recommendation is proposed which merges the advancements of random sampling and hybrid filling into single prototype to achieve item recommendations without compromising private information of users. User averaging is utilized on explicit sparse dataset to refill them with virtual values. Proposed model is implemented on MovieLens 100K and MovieLens 1M datasets to evaluate the effectiveness and usefulness of the proposed strategy.

Fed4Rec technique is proposed in [12] to recommend the better page recommendation to both public and private users based on federated learning and model agnostic meta learning methods. In Fed4Rec, data collected from public users are trained on the servers and data from private users are trained on their local devices instead of sharing them to the central server to protect user information. Mean reciprocal rank, hit ratio and recommendation accuracy are taken into consideration to the outperformance of the proposed model.

The collaboration of federated learning and deep neural network named Fedfast is proposed in [13] to formulate distributed recommender system without violating user's privacy concerns. Fedfast utilizes the novel sampling strategy and aggregation techniques to reduce the convergence speed of training the models which effects in communication cost of an entire system. Experimental results show that the proposed model achieves 4 times and 20 times of improvement in baseline of NDCG for MovieLens 100K and TripAdvisor respectively.

In [14] a potential semi-supervised learning is proposed for animal identification exploiting computer vision that integrates pseudo labeling with deep neural networks to improve the predictive accuracy of Holstein cows. Through this method unlabeled images are labeled automatically which raises accuracy in 20.4% compared with manually labeled methods. Further the studied model attained accuracy of 92.7 when implemented on the independent test set.

Graph neural network based on the knowledge graph aware recommendation with pseudo labeling is proposed in [15] which supplements samples with labels assigned by pseudo labeling to resolve the cold start problem for both user and item. Hybridization of pseudo labeling with graph neural network technique brought the new items to the limelight by explicitly assumes the unrated instances as a positive instances. Performance of the model is evaluated on three real world datasets with parameters precision and recall.

Deep semi-supervised detection method depends on Meta pseudo label is proposed in [16] to overcome the malicious attack of providing fake review and rating for the item. This method enhances the predictive rate of deep semi-supervised detection by training samples with small labeled as well as unlabeled. To detect the attack on the recommendation system, this system employs group of students and experiment performed with classical, mixed and GSA-GANs.

In[17] federated recommender system with global knowledge graph called FedRKG is proposed to permit the higher-order user item interaction globally using information that are available in public server. FedRKG acquires increases in average accuracy of 4% comparatively by collaborating pseudo interaction and local differential privacy to preserve the interaction stored in local devices and to prevent model gradients.

Personalized federated recommender with self-supervised pre-training called PerFedRec++ is proposed in [18] in which federated learning mechanism is utilized to create bi-augmented graph used in self-supervised graph learning. The proposed model focused three key issues in recommendation system involving data heterogeneity, degradation of learning model and communication overhead by learning from graph neural network, clustering of likelihood users and user-item level samples.

Integration of federated learning with local differential privacy is studied in [19] to address the key issue of model parameters updates with high dimensional and sequent decimal points that affected the local differential privacy protocols. By presenting two methods such as repeatedly collecting model parameters over enormous individual users and implementing the selection and filtering approach of data perturbation in global

server. Privacy preservation ratio, accuracy and capability of model are considered as an evaluation protocol of the proposed model with various competitive models.

To maintain the balance between data privacy preservation and efficient model training, local differential based clustering is incorporated with federated recommender system in [20]. The proposed model exploited local differential privacy to alleviate the privacy leakage in convolutional recommender system and performed security aggregation for global training like progression in distributed environment to prevent bad training efficacy.

Asynchronous federated deep factorization machine algorithm is proposed in [21] to resolve the inference attack and poison attack in distributed system. Local differential privacy is employed with limited added noise to prevent the user's actual data on shared instances and anti-poisoning mechanism is implanted to refrain from malicious attacks of intruders. Authors evaluated the AFedDFM model on two real world datasets with analyses of recommendation quality and privacy preservation.

Federated deep recommendation approach is proposed in [22] which unite heterogeneous information network and matrix factorization to augment both user personal privacy preservation and to raise the prediction accuracy. This heterogeneous information is learned by each client through meta paths which merged to the matrix factorization to extract the essential features foe model learning. Latent features are then fed to deep neural network that are collected from various participants of the system. Without tolerating the user privacy factor, the proposed model achieves significant improvement in accuracy, speed of convergence and reduces communication overhead.

Various conventional works in recommender system are reviewed in this section. As a result of this analysis, it is required to ensure the user's privacy and security along with the accurate prediction of relevant items. To achieve this, hybrid techniques are capable to escalate the prediction performance and privacy preservation in recommender system. Hence, we have implemented deep neural network based federated recommender system with local differential privacy and pseudo interaction.

3. PRELIMINARIES

3.1 Federated Recommender System

Federated learning (FL) is the novel security and privacy preserving mechanism introduced by Google. In FL data from the multiple local clients are not stored and processed in the centralized server instead kept in the local data providers and shared only their gradient information. It further classified as vertical federated learning and horizontal federated learning. Collaboration of federated learning with recommender system offers the path for the new research domain federated recommender system to guarantee the user's privacy by training their data locally and shared the gradients to the global server. Global server creates the global model by aggregate these local updates from multiple clients. In an iterative manner, then the updated global model sent back to the local client for further training process.

3.2 Deep Neural Network

Deep neural network (DNN), a subset of machine learning consists of number of layers with interconnected neurons. In general DNN contains one input layer, one output layer and N number of hidden layers where inputs of the network received by the input layer and hidden layers process these input by performing non linear mapping functions. Finally output layer generates the predictive models. Figure 1 illustrates the general architecture of the deep neural network with associated weights and activation function. Weights (w) are the way of learning of deep neural network from the input data and it controls the strength or relationship between the connected neurons. Activation functions $(f)$are used to introduce the non-linearities into the model and decide whether the neuron should be activated or not by the weighted sum of inputs. As input data is processed through this multiple complex layers, accurate and abstracted features from the inputs are extracted.

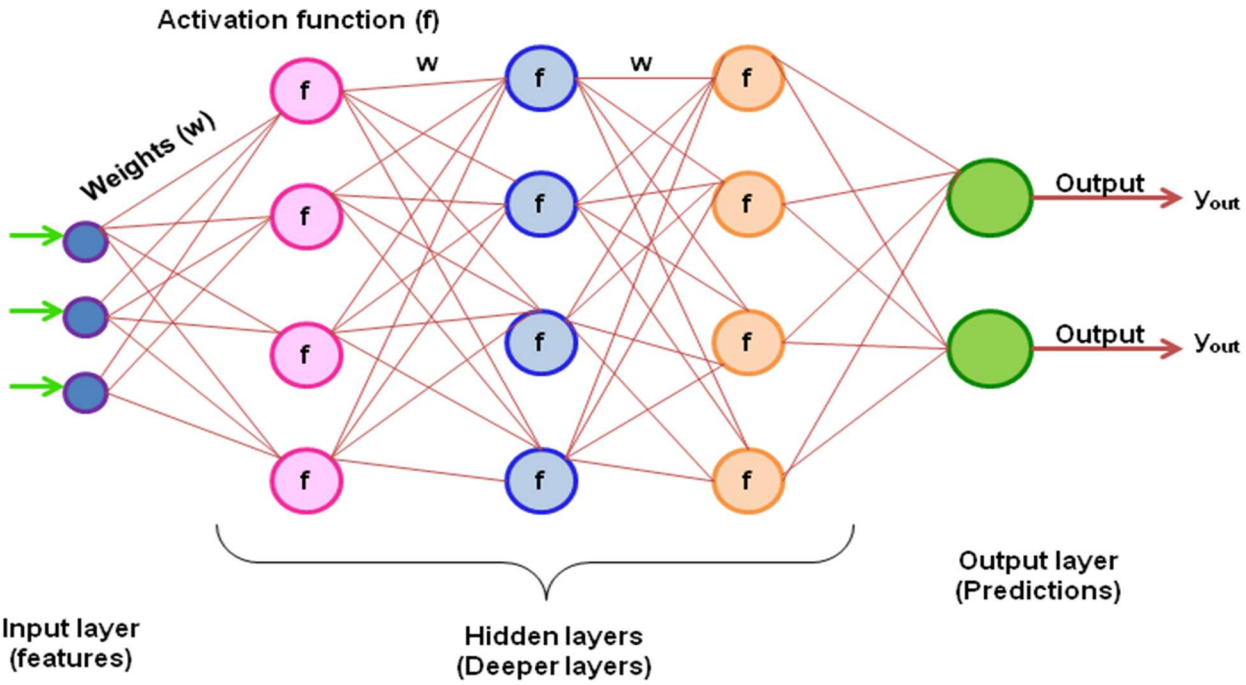

Figure 1: Deep neural network architecture with multiple layer and activation function

3.3 Local Differential Privacy

Local differential privacy is a state-of-art privacy preserving algorithm considers the data stealing or breaching of security and privacy concerns of users by the unauthorized server or aggregator in the distributed environment. In LDP, private and sensitive data of the client are perturbed in local device and the perturbed data are then sent to the global aggregator. To obtain the required statistical results, the data are reconstructed by the global aggregator. Laplace mechanism is one of the popular methods of local differential privacy algorithm which is the most appropriate for processing numerical value. Let consider the given function is f: $\mathbb{D} \rightarrow \mathbb{R}^d$ and then the Laplace mechanism can be written in equation 1.

$$M\,(f(x), \varepsilon) = f(x) + (Z_1, \ldots, Z_i) \qquad (1)$$

Where, Z_i are independent and identically distributed random variables depicted from Laplace distribution with scale $\Delta f/\varepsilon$, namely Lap $\Delta f/\varepsilon$. Here, f is the function, Δf is the function sensitivity and Δf can be written with the variable x and y shown in equation 2.

$$\Delta f = \max_{x,y \in \mathbb{D}} \|f(x) - f(y)\| \qquad (2)$$

4. PROPOSED METHODOLOGY

Predicting the user interest among massive volume of data without affecting the user privacy and security become the taunting task specifically in the centralized environment. To alleviate this issue, federated deep neural network with local differential privacy (FedDNN-LDP) method is proposed which incorporates federated learning with deep neural network and local differential privacy. Deep neural network is integrated with pseudo-interaction filling for predicting the user preference and local differential privacy is exploited on the trained model to ensure the user's privacy in the centralized architecture. Figure 2 demonstrated the architecture of the proposed FedDNN-LDP model.

The proposed architecture is consists of two phases namely client-end and server-end. The raw user data is collected from the users stored in the local device. Instead of sharing this raw user data to the centralized server, intermediate user gradients are shared to preserve the user private data. In local FedDNN model, user's input data are trained and output of gradients is achieved. Local differential privacy mechanism is applied on these gradient parameters by adding Laplace noise injection. The noise injected data are then aggregated to form the user clustering and the local weight updates are shared to the centralized server. Weight updates from the local devices are sent to the global server and in the server-end model parameter are generated and aggregated. The aggregated outcomes are then fed to the global model training to predict the global model. Global model updates are then transferred to the local device in the client-end.

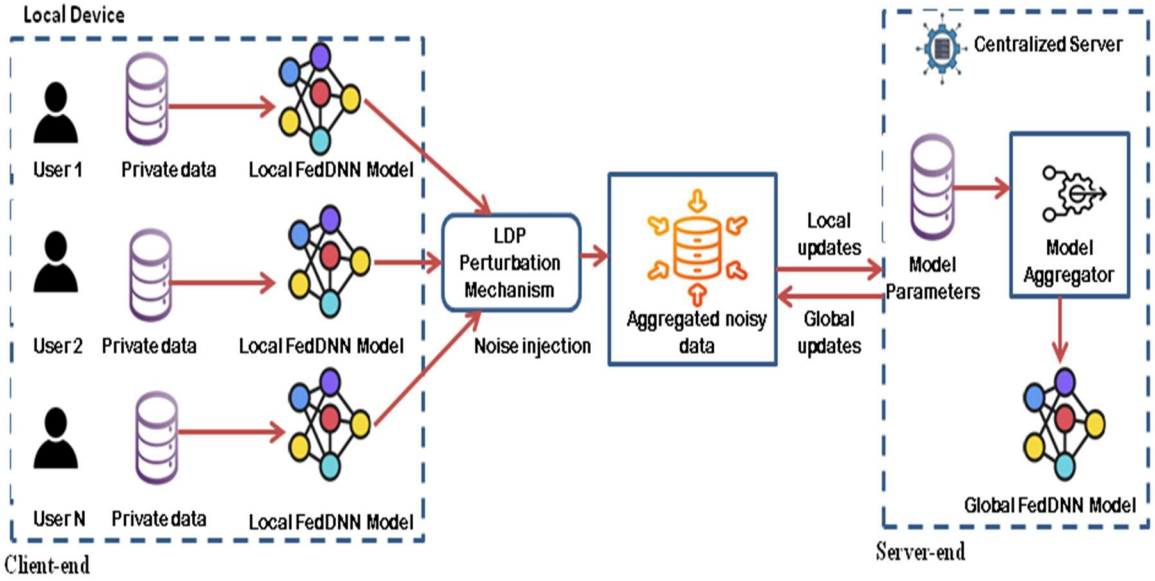

Figure 2: Architecture of proposed federated deep neural network with local differential privacy (FedDNN-LDP) technique

4.1 Algorithm and Processing Steps – FedDNN-LDP (Figure 2)

Input:

//Movie rating dataset
User rating data (user ID, movie ID, rating and timestamp), Learning rate η, Embedding d, Number of users U_n, number of items I_n, Count of pseudo interacted items ρ, Parameters of LDP δ, λ , Client model D_k.

Output:

Gradients and embeddings of the model Θ, top k recommendation (k = 20)

Steps:

Step 1: Begin:

Initialization of model parameter Θ.
Step 2: For epoch $\in \{1,\ldots,n\}$ do
Compute the gradient model. Update them.
Step 3: while not converge do
Train the federated model by selecting the client D_k using aggregator.
Step 4: Collection of intermediate parameters from the selected local clients.
$D_k = \text{ClientUpdation}\ (k, \Theta)$
Step 5: Intermediate parameters or gradients updation. Perform pseudo-interaction ρ
$\Theta - \eta . D_k \leftarrow \Theta$.
Step 6: Fn ClientUpdate (k, Θ):

Download Θ from the global server.

Step 7: Computation and prediction of top k recommendations.

Step 8: Incorporation of local differential privacy mechanism with parameters δ, λ by adding Laplace noise distribution to D_k.

$$D'_k = D_k + \text{Lap}\left(\frac{\Delta k}{\varepsilon}\right) \quad (2)$$

Step 9: return gradients D'_k.

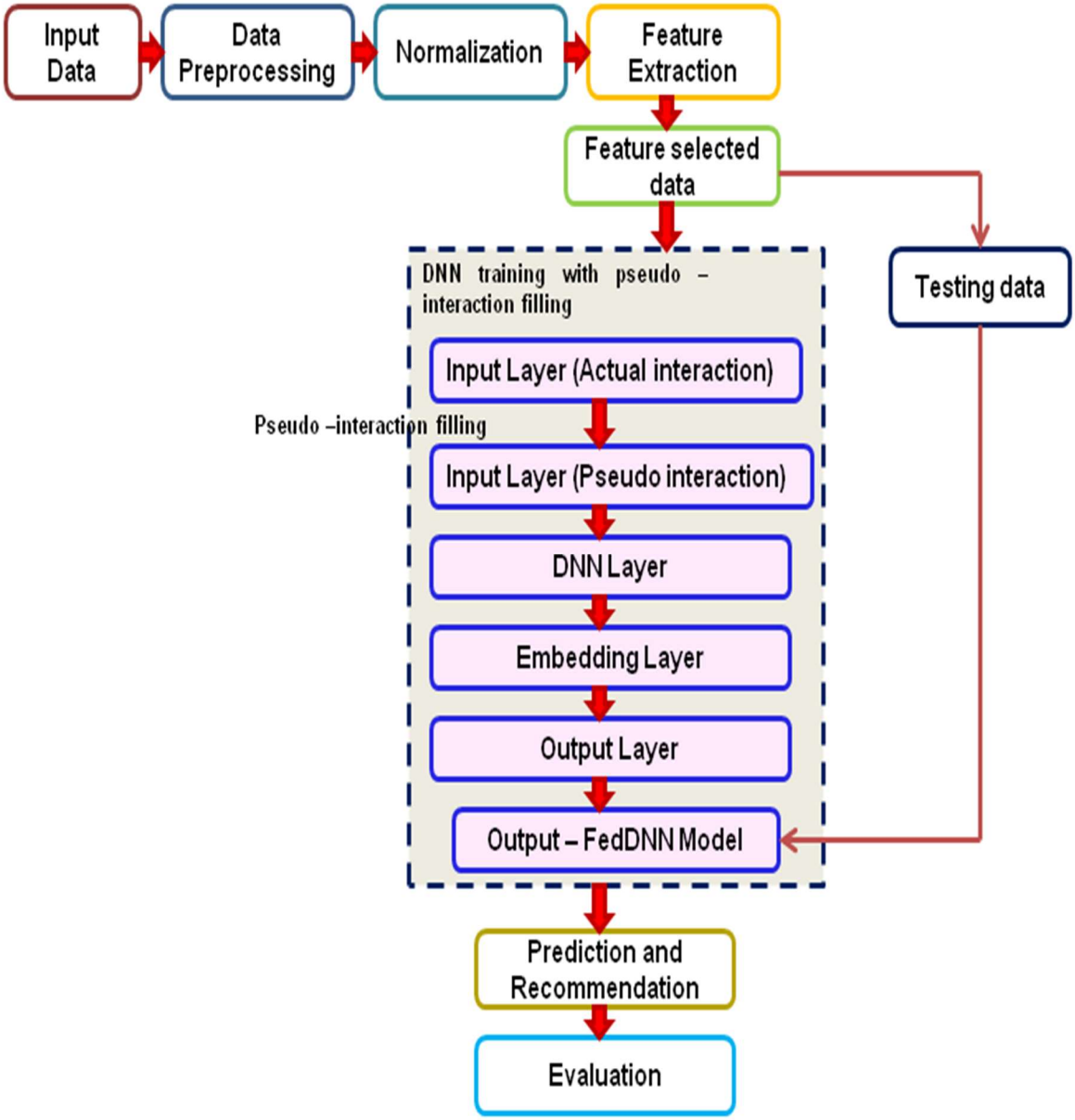

Figure 3: Training process of deep neural network with pseudo-interaction filling (FedDNN model)

Training process of deep neural network with pseudo-interaction filling (FedDNN) is depicted in figure 3. It involves the following phases: input data collection and loading, data preprocessing, normalization, feature selection, training the feature selected data with FedDNN model, prediction, recommendation and evaluation of the proposed model.

4.2 Workflow of the Proposed FedDNN-LDP Model

4.2.1 Data Collection and Loading

Input user movie rating data with user ID, movie ID, rating and timestamp.

4.2.2 Data Preprocessing

Checking missing records in the dataset and remove unwanted features using dimensionality reduction algorithms. // Removal of the field timestamp.

4.2.3 Data Normalization

To make sure the data consistency throughout the dataset D, normalizes the user rating into the general scale ranging from 0 to 1 using Min-Max scaling represented in equation (3).

$$D' = \frac{D - D_{min}}{D_{max} - D_{min}} \quad (3)$$

where D and D′ are the original and normalized datasets respectively.

4.2.4 Feature Selection

Select the relevant and essential features from the data to train them for modeling.

4.2.5 DNN Training

Selected features are then trained using deep neural network with pseudo-interaction features. During the model training, the embedding layer are only updated with the model trained features which makes the server can read the real data of the user from these features. Model training involves the conversion of sparse embeddings into dense embeddings in embedding layer and then updates them to the server. While updating these features, server directly infers the user's real data which leads the privacy leakage. To alleviate these issues pseudo-interaction filling is exploited in that the server becomes capable of find out probable values of features instead of infer the real data. This is written in equation 4.

$$D_{k_{train}} = D_{k_{real}} + \text{rand}(n, \mu, \text{inf}(x)) \quad (4)$$

Here, $D_{k_{train}}$ represents user data used for model training, $D_{k_{real}}$ indicates the raw user data, randomized data generated process denoted as rand(), n and μ indicated count of local data of the user and hyper parameter generated for pseudo-interaction filling respectively (for this study, the values of pseudo-interaction ranging from $\mu = 0.5$ to $\mu = 3.0$).

Here, $D_{k_{train}}$ represents user data used for model training, $D_{k_{real}}$ indicates the raw user data, randomized data generated process denoted as rand(), n and μ indicated count of local data of the user and hyper parameter generated for pseudo-interaction filling respectively (for this study, the values of pseudo-interaction ranging from $\mu = 0.5$ to $\mu = 3.0$).

These pseudo-interacted features are then fed to the deep neural network consists of n number of hidden layers. The outcome of the DNN are then updated to the embedding layer which is responsible for the conversion of sparse features such as user ID and item ID into dense embeddings or vectors. The initialization of embedding layer is achieved by pre-trained model to obtain the embeddings for user and item nodes. Output layer predict the most relevant items for the user based on the prediction score of embedding layer.

4.2.6 Prediction and Recommendation

From the outcome of the model training, this phase predicts and recommends decidedly relevant movies to the user. This recommendation can be done for three k values involving $k = 5$, $k = 10$ and $k = 20$. The model achieves visible accuracy for the value of $k = 20$ i.e. top 20 recommendations.

4.2.7 Model Evaluation

Efficacy of the studied model is evaluated by the performance measures such hit ratio, NDCG,

precision, recall, mean absolute error, root mean squared error and training and validation loss.

5. EXPERIMENTAL ANALYSIS

5.1 Software Description

The training and testing experiments were implemented and performed on a computer with Core i5, 8GB DDR 3200 MHz RAM, 512GB SSD with operating system Window 10 and implemented using Python 3.8 with Anaconda Navigator which is one of the IDEs of python programming. For the implementation, Python machine learning libraries such as Keras and TensorFlow were installed and utilized.

5.2 Dataset Description

To validate the performance of the proposed model, two publicly available real-world datasets were used from the GroupLens website namely MovieLens 100K and MovieLens 1M, contain explicit feedback of user rating ranges from 1 to 5 and implicit feedback involving user behavioral and demographical information. The statistical analysis of the experimental datasets is presented in Table 1.

Table 1: Statistical analysis of MovieLens 100K and MovieLens 1M datasets

S.No	Dataset	Interaction	User	Movie	Sparsity
1	MovieLens 100K	100,000	943	1682	93.69 %
2	MovieLens 1M	1,000,209	6,040	3952	95.53 %

5.3 Evaluation Metrics

Following are the evaluation metrics considered to evaluate the performance of proposed model.

5.3.1 Hit ratio (HR)

It is the ratio of correctly predicted item for the target user from the total number of prediction. i.e top k recommendation. 0 and 1 are the bound values of hit ratio (HR) where 0 indicates negative predictions and 1 represents the positive correct predictions.

5.3.2 Normalized discounted cumulative gain (NDCG)

Normalized discounted cumulative gain is the metric of predicting the relevance of the recommendation presented by the model. It is given in equation 5.

$$\mathrm{NDCG_k} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}}^{\mathrm{k}} \frac{\mathrm{In}\ (2)}{\mathrm{In}\ (\mathrm{i}+2)} \qquad (5)$$

where, N – total number of items to be evaluated; k – total number of top recommendations; i – relevant items in the top k list.

5.3.3 Precision@K

Precision measures the proportion of positive predictions in the predicted recommendation to evaluate the model accuracy by determining the relevance of items in the top k recommendation list.

5.3.4 Recall@K

Recall measures the proportion of relevant items in the predicted top k-recommendation list from the total number of relevant items for the target user.

5.3.5 Root mean squared error (RMSE)

Root mean squared error measures the average number of errors with quadratic evaluation function. It describes the square root of the standard squared deviations between the actual and noisy query results. Lower value of RMSE indicates the increase in privacy.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{\mathrm{i=1}}^{\mathrm{N}} = (\mathrm{u_{ai}} - \mathrm{u'_{ai}})}{\mathrm{N}}} \qquad (6)$$

In equation 6, $\mathrm{u_{ai}}$ and $\mathrm{u'_{ai}}$ are represent the actual and noisy response.

5.3.6 Mean absolute error (MAE)

Mean absolute error is the measurement of average number of errors. It is difference between the real and noise added response of the standard query results. A lower result in MAE represents the higher privacy and security of the model presented in equation 7..

$$MAE = \frac{1}{N}\sum_{N}^{i=2} u_{ai} - u'_{ai} \quad (7)$$

where u_{ai} and u'_{ai} are represent the actual and noisy response.

6. RESULT AND DISCUSSION

The proposed FedDNN-LDP is evaluated on two publicly available datasets namely MovieLens 100K and MovieLens 1M and FedRKG [17], FedDeepFM [23] and FedGCN [24] are the baselines considered for the comparative analysis of the proposed model. For the implementation process, both datasets are split as training and testing data with values 75% and 25% respectively.

Table 2: Performance of the proposed model at various k values on evaluation dataset

S. No	In this study	Dataset - MovieLens 100K				Dataset - MovieLens 1M			
		HR@ k	NDCG@ k	Precision@ k	Recall@ k	HR@ k	NDCG@ k	Precision@ k	Recall @k
1	K = 5	0.0732	0.1543	0.1212	0.0911	0.0715	0.0213	0.0325	0.0500
2	K = 10	0.0889	0.1712	0.1631	0.1290	0.0758	0.0479	0.0680	0.0732
3	K = 20	0.1121	0.1933	0.1856	0.1598	0.0842	0.0892	0.0708	0.0917

Table 2 presents hit ratio (HR), NDCG, Precision and recall of the proposed model for various top k recommendations including k = 5, k = 10 and k = 20. Figure 4 and figure 5 illustrated the performance comparison of studied model for various k values on MovieLens 100K and MovieLens 1M dataset respectively. For MovieLens 100K dataset, prediction of the top relevant item list with the count k = 20 achieved the highest accuracy for all the measures than for the predicted recommendation list with items k = 10 and k = 5. On the other hand, MovieLens 1M dataset also obtained the higher results for the proposed model yet it is comparatively lower than the results obtained for 100K datasets.

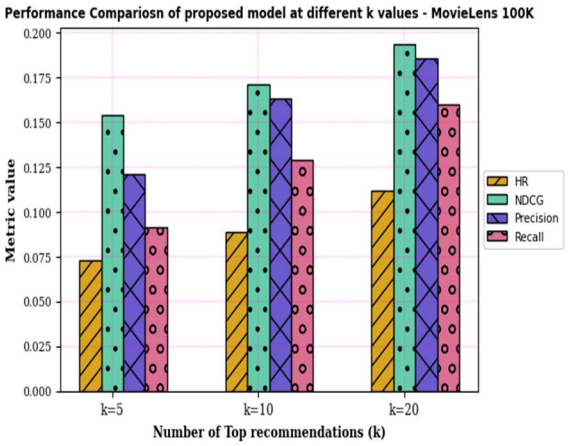

Figure 4: Performance of the proposed model at various k values for MovieLens 100K

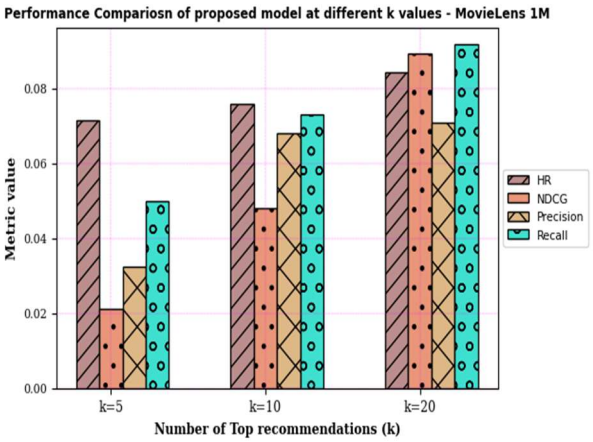

Figure 5: Performance of the proposed model at various k values for MovieLens 1M

Comparative analysis of the proposed model with competitive models is presented in table 3. Figure 6 demonstrated the performance comparison of the proposed FedDNN-LDP with existing conventional models at k = 20 for MovieLens 100K dataset whereas the figure 7 depicted for MovieLens 1M dataset. Normalized discounted cumulative gain, precision and recall are parameters considered for the performance comparison with the baselines FedRKG, FedDeepFM and FedGCN. From these three existing models, FedGCN obtained better results than the remaining two models. Even, the proposed

method surpassed the existing approaches by achieving NDCG = 0.1933, precision = 0.1856 and recall = 0.1598 which is improvement in 15.47% for NDCG, 30.06% for precision and 28.66% for recall on MovieLens 100K. For MovieLens 1M dataset, the obtained results and its percentage of improvement are as follows: NDCG = 0.0892 (32.54%), precision = 0.0708 (58.38%) and recall = 0.0917 (40.86%).

Table 3: Baseline comparison of the proposed model on evaluation dataset

S.No	Techniques	Dataset - MovieLens 100K			Dataset - MovieLens 1M		
		NDCG@k (k=20)	Precision@k (k=20)	Recall@k (k=20)	NDCG@k (k=20)	Precision@k (k=20)	Recall@k (k=20)
1	FedRKG	0.1578	0.1270	0.1112	0.0570	0.0352	0.0572
2	FedDeepFM	0.1644	0.1342	0.1263	0.0641	0.0407	0.0655
3	FedGCN	0.1801	0.1671	0.1352	0.0810	0.0582	0.0726
4	Proposed FedDNN-LDP	0.1933	0.1856	0.1598	0.0892	0.0708	0.0917
Improvement (%)		15.47%	30.06%	28.66%	32.54%	58.38%	40.86%

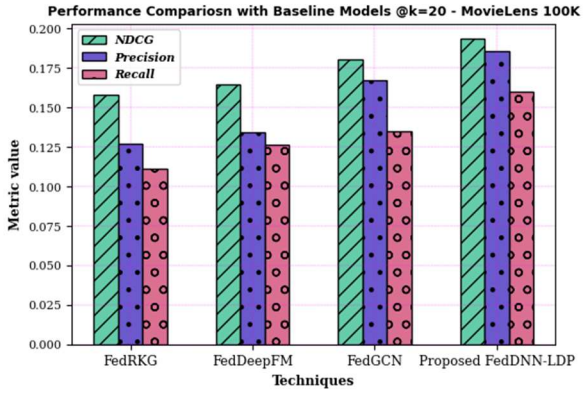

Figure 6: Baseline comparison of the proposed model at k=20 for MovieLens 1M

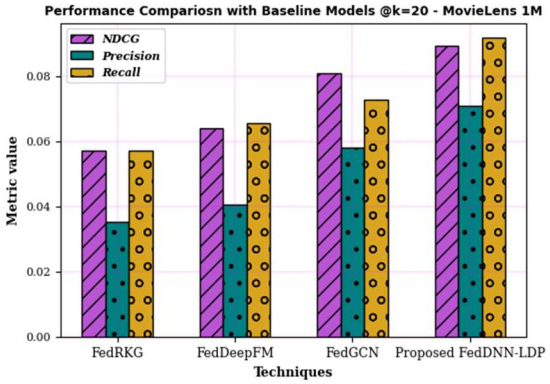

Figure 7: Baseline comparison of the proposed model at k=20 for MovieLens 1M

The proposed model achieved the better results for the top k recommendations than the conventional methods. It is tedious to predict the relevant recommendations without compromising the user privacy and security. But the hybrid FedDNN-LDP model predicts the most relevant items and suggested them to the users without affecting the user privacy on the distributed environment. To evaluate that, mean absolute error and root mean squared error are measured for various privacy budget (ε) values including ε = 0.5, ε = 1.0, ε = 1.5, ε = 2.0, ε = 2.5 and ε = 3.0. Table 4 illustrated the MAE and RMSE comparison of the proposed LDP with existing differential privacy (DP) for different privacy budgets on the evaluation datasets. For ε = 3.0, the proposed FedDNN-LDP model achieved higher accuracy than the privacy budget value 0.5 for both dataset. The decrease in the MAE and RMSE is the increase in the model accuracy and privacy.

Journal of Theoretical and Applied Information Technology
30th September 2025. Vol.103. No.18
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 4: MAE and RMSE comparison of proposed FedDNN-LDP with DP for various privacy budget (ε) on evaluation datasets

S.No	Privacy Budget (ε)	Dataset - Movielens 100K				Dataset - Movielens 1M			
		MAE		RMSE		MAE		RMSE	
		DP	Proposed FedDNN-LDP	DP	Proposed LDP	DP	Proposed FedDNN-LDP	DP	Proposed LDP
1	0.5	2.899	1.801	2.741	1.532	2.283	1.830	2.155	1.212
2	1.0	2.812	1.375	2.160	0.802	2.121	1.482	2.131	1.097
3	1.5	2.439	1.241	1.894	0.757	2.377	0.501	1.104	0.961
4	2.0	1.989	0.890	1.743	0.590	1.914	0.233	1.753	0.895
5	2.5	1.432	0.524	1.651	0.393	0.899	0.191	1.597	0.629
6	3.0	1.274	0.312	0.986	0.256	0.590	0.128	1.189	0.531

Figure 8 and figure 9 illustrated the training and validation loss over 200 epochs for MovieLens 100K and 1M datasets respectively. Both graph demonstrated the significant and swift loss reduction in both datasets for validation set over epochs which achieves significance in convergence of the proposed model. The outcome represents the robustness of the model with federated learning within the decentralized system with effectiveness and model correctness for error rate minimization.

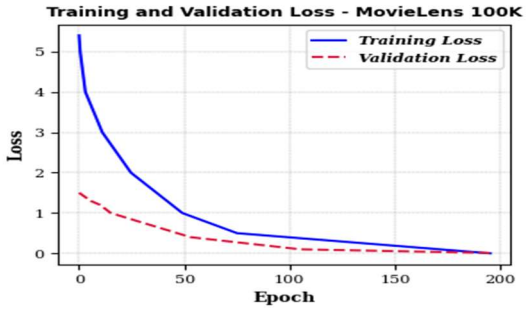

Figure 8: Training loss and validation loss of MovieLens 100K

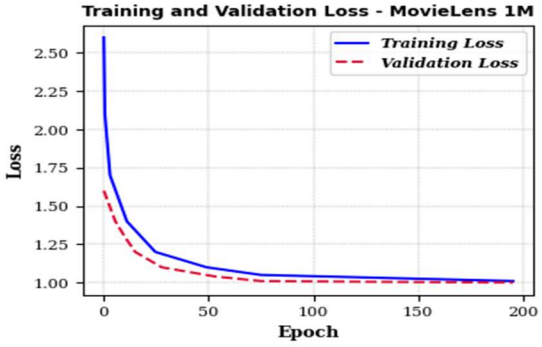

Figure 9: Training loss and validation loss of MovieLens 1M

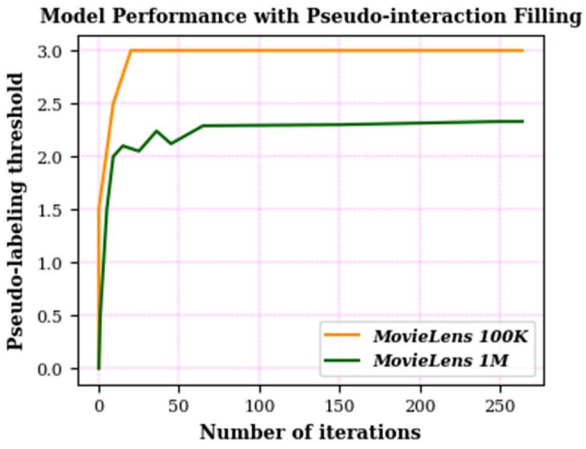

Figure 10: Accuracy of the proposed model with pseudo-interaction filling on evaluation datasets

The accuracy of the studied FedDNN-LDP model for number of iterations on the evaluations datasets is shown in figure 10. It is visible that performance of the model increased after single iteration of pseudo-labeling and remaining stable and steady after the number of iterations have performed on them. For the maximum of 250 iterations the model utilized best threshold values ranging from 0.5 to 3.0. When the highest threshold reached, the model remained stable for both datasets. For 100K the highest threshold is reached to 3.0 whereas the 1M reaches 2.0 with stability. The obtained results represents the simple pseudo-labeling interaction achieves visible and reliable improvement in prediction and performance when collaborated with neural network model even for the large datasets.

6. CONCLUSION

Predicting the user preferences without affecting their privacy is the major issue in recommendation system. Incorporation of federated learning and recommender system with local improved differential privacy is proposed in this research which is significantly enhances the performance of the recommender system and ensures the privacy concerns of the users. The proposed system exploits deep neural network with pseudo interaction filling to predict the top recommendations to the users. Model performance and accuracy is evaluated by the parameters such as hit ratio, normalized discounted cumulative gain, precision, recall, root mean squared error and mean absolute error. Comparative analysis described that the proposed FedDNN-LDP model outperformed the existing federated recommender model with the improvement of 15.47% in NDCG, 30.06% in precision and 28.66% in recall for MovieLens 100K dataset and 32.54% in NDCG, 58.38% in precision and 40.86% in recall. Furthermore, the graph of training and validation loss for both datasets illustrated that the gradual decrease in validation data than the training data. Privacy concern of the studied model is evaluated to define the efficacy of the model for various privacy budgets(ε). Even though the proposed model surpassed the existing conventional models still there is room for improvement. Future research will incorporate fuzzy logic to enhance the scalability and predictive accuracy of the recommender system.

REFERENCES

[1] T. Ganesan, R. A. Jothi, and P. Vellaiyan, "A comprehensive survey on recommender system techniques," *International Journal of Computational Systems Engineering*, vol. 7, no. 2/3/4, pp. 146–158, 2023, doi: 10.1504/IJCSYSE.2023.132915.

[2] D. Roy and M. Dutta, "A systematic review and research perspective on recommender systems", doi: 10.1186/s40537-022-00592-5.

[3] T. Ganesan and P. Vellaiyan, "An Enhanced Neural Network Collaborative Filtering (ENNCF) for Personalized Recommender System," *Lecture Notes in Electrical Engineering*, vol. 1194, pp. 183–195, 2024, doi: 10.1007/978-981-97-2839-8_13.

[4] T. Ganesan and P. Vellaiyan, "ETI-GCN Explicit to Implicit Graph Convolution Network for Personalized Recommender System in e-commerce and Healthcare," *Cybersecurity in Healthcare Applications*, pp. 281–295, Jan. 2025, doi: 10.1201/9781032711379-16/ETI-GCN-THENMOZHI-GANESAN-PALANISAMY-VELLAIYAN.

[5] T. Ganesan and P. Vellaiyan, "Optimized-FedRec: Optimized Privacy-Preserving Federated Recommender System with Elliptic Curve Cryptography based Homomorphic Encryption and Local Differential Privacy," *Procedia Comput Sci*, vol. 259, pp. 1602–1611, Jan. 2025, doi: 10.1016/J.PROCS.2025.04.115.

[6] Z. Sun *et al.*, "A Survey on Federated Recommendation Systems," *IEEE Trans Neural Netw Learn Syst*, 2024, doi: 10.1109/TNNLS.2024.3354924.

[7] Y. Liu, T. Lin, and X. Ye, "Federated recommender systems based on deep learning: The experimental comparisons of deep learning algorithms and federated learning aggregation strategies," *Expert Syst Appl*, vol. 239, p. 122440, Apr. 2024, doi: 10.1016/J.ESWA.2023.122440.

[8] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T. S. Chua, "Neural Collaborative Filtering," *26th International World Wide Web Conference, WWW 2017*, pp. 173–182, Aug. 2017, doi: 10.1145/3038912.3052569.

[9] G. Zhou *et al.*, "Deep Interest Network for Click-Through Rate Prediction," Jun. 2017, Accessed: Apr. 13, 2025. [Online]. Available: https://arxiv.org/abs/1706.06978v4

[10] H.-T. Cheng *et al.*, "Wide & Deep Learning for Recommender Systems," Jun. 2016, Accessed: Apr. 13, 2025. [Online]. Available: https://arxiv.org/abs/1606.07792v1

[11] G. Lin, F. Liang, W. Pan, and Z. Ming, "FedRec: Federated Recommendation with Explicit Feedback," *IEEE Intell Syst*, vol. 36, no. 5, pp. 21–30, 2021, doi: 10.1109/MIS.2020.3017205.

[12] S. Zhao, R. Bharati, C. Borcea, and Y. Chen, "Privacy-Aware Federated Learning for Page Recommendation".

[13] K. Muhammad *et al.*, "FedFast: Going beyond Average for Faster Training of Federated Recommender Systems," *Proceedings of the ACM SIGKDD*

International Conference on Knowledge Discovery and Data Mining, pp. 1234–1242, Aug. 2020, doi: 10.1145/3394486.3403176.

[14] R. E. P. Ferreira, Y. J. Lee, and J. R. R. Dórea, "Using pseudo-labeling to improve performance of deep neural networks for animal identification," *Sci Rep*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/S41598-023-40977-X.

[15] R. Togashi, M. Otani, and S. ' Ichi Satoh, "Alleviating Cold-Start Problems in Recommendation through Pseudo-Labelling over Knowledge Graph," 2020, Accessed: Apr. 14, 2025. [Online]. Available: https://doi.org/x

[16] Q. Zhou, K. Li, and L. Duan, "Recommendation attack detection based on improved Meta Pseudo Labels," *Knowl Based Syst*, vol. 279, p. 110931, Nov. 2023, doi: 10.1016/J.KNOSYS.2023.110931.

[17] D. Yao, T. Liu, Q. Cao, and H. Jin, "FedRKG: A Privacy-Preserving Federated Recommendation Framework via Knowledge Graph Enhancement," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 14504, pp. 81–96, 2024, doi: 10.1007/978-981-99-9896-8_6.

[18] LuoSichun, XiaoYuanzhang, ZhangXinyi, LiuYang, DingWenbo, and SongLinqi, "PerFedRec++: Enhancing Personalized Federated Recommendation with Self-Supervised Pre-Training," *ACM Trans Intell Syst Technol*, Nov. 2024, doi: 10.1145/3664927.

[19] S. Truex, L. Liu, K. H. Chow, M. E. Gursoy, and W. Wei, "LDP-Fed: Federated learning with local differential privacy," *EdgeSys 2020 - Proceedings of the 3rd ACM International Workshop on Edge Systems, Analytics and Networking, Part of EuroSys 2020*, pp. 61–66, Apr. 2020, doi: 10.1145/3378679.3394533.

[20] W. Li, H. Chen, R. Zhao, and C. Hu, "A Federated Recommendation System Based on Local Differential Privacy Clustering," *Proceedings - 2021 IEEE SmartWorld, Ubiquitous Intelligence and Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Internet of People, and Smart City Innovations, SmartWorld/ScalCom/UIC/ATC/IoP/SCI 2021*, pp. 364–369, 2021, doi: 10.1109/SWC50871.2021.00056.

[21] X. Zhao, X. Bai, G. Sun, and Z. Yan, "Asynchronous Federated Learning With Local Differential Privacy for Privacy-Enhanced Recommender Systems," *IEEE Internet Things J*, 2025, doi: 10.1109/JIOT.2025.3531117.

[22] X. Wang, S. Meng, Y. Chen, Q. Liu, R. Yuan, and Q. Li, "Federated Deep Recommendation System Based on Multi-View Feature Embedding," *Proceedings - 2022 IEEE 9th International Conference on Data Science and Advanced Analytics, DSAA 2022*, 2022, doi: 10.1109/DSAA54385.2022.10032411.

[23] Y. Wu, L. Su, L. Wu, and W. Xiong, "FedDeepFM: A Factorization Machine-Based Neural Network for Recommendation in Federated Learning," *IEEE Access*, vol. 11, pp. 74182–74190, 2023, doi: 10.1109/ACCESS.2023.3295894.

[24] S. Peng, S. Siet, I. Sadriddinov, D. Y. Kim, K. Park, and D. S. Park, "Integration of Federated Learning and Graph Convolutional Networks for Movie Recommendation Systems," *Computers, Materials & Continua*, vol. 83, no. 2, pp. 2041–2057, Apr. 2025, doi: 10.32604/CMC.2025.061166.