A Collaborative Deep Learning Model for Recommending e- Learning Resources for Intelligent Tutoring Systems

Manish Jhangiani[1], Dr.Rajesh Kumar Chakrawarti[2]

Abstract: Big Data Analytics has been shown to be a powerful tool for curating personalized recommendation systems in diverse fields such as e-learning, movie and music etc. Due to the proliferation of e-learning applications and their associated advantages, it has become a difficult task to optimize e-learning resources according to user requirements, mostly due to the sheer volume of resources available. Individuals have varying preferences for learning resources, particularly depending on factors such as age, country, area, and specific needs. There is no definitive method to customize the recommendation system in order to assist customers in accessing the most optimal resources in the shortest possible timeframe. Collaborative learning has emerged as a successful method in machine learning approaches for developing recommendation systems. In this paper, collaborative filtering based deep learning (CFDL) approach with RMSProp has been developed for recommending e-learning resources. It has been shown that the proposed approach attains low Mean Absolute Error even a reasonably less iterations making it suitable for large dataset based recommendation systems.

Keywords: Big Data, Machine Learning, Collaborative Filtering, Recommendation Systems, e-Learning, RMS Prop, MAE.

I. INTRODUCTION

Big data analytics and machine learning has seen a wide domain of application in recommendation systems, one of which happens to be the e-learning platform. E-learning has revolutionized the way education is delivered, providing flexibility and accessibility to learners worldwide [1]. However, the vast amount of data generated from e-learning platforms necessitates advanced analytical tools to derive meaningful insights. Big data analytics and machine learning are pivotal in enhancing e-learning experiences, enabling personalized learning, predicting learner outcomes, and improving educational content [2]. E-learning platforms generate massive amounts of data from various sources, including user interactions, assessments, discussion forums, and multimedia content. Big data analytics involves processing and analyzing these large datasets to uncover patterns and trends. By leveraging big data, educators and institutions can gain insights into learner behavior, preferences, and performance [3]. This information is crucial for developing targeted interventions, improving curriculum design, and enhancing overall learning outcomes [4].

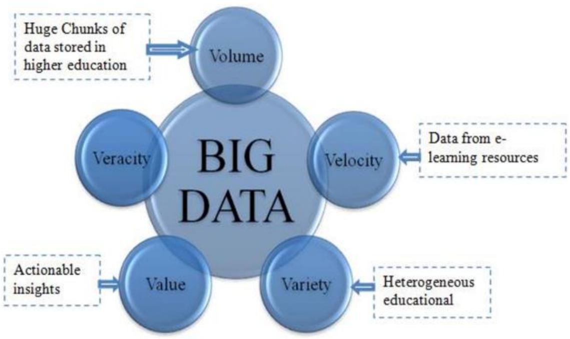

Fig.1 Role of Big Data Analytics in e-Learning Applications [1]

In e-learning, ML algorithms can analyze learner data to predict academic performance, recommend personalized learning paths, and identify at-risk students [5] . For instance, recommendation systems can suggest relevant courses or resources based on a learner's past behavior and performance [6]. Additionally, predictive analytics can help educators intervene early to support students who may be struggling [7]. One of the most significant benefits of big data and machine learning in e-learning is the ability to provide personalized and adaptive learning experiences [8]-[9]. Personalization involves tailoring educational content and experiences to meet individual learner needs and preferences [10]. Adaptive learning systems use ML algorithms to dynamically adjust the difficulty and type of content based on real-time learner performance and engagement. This approach ensures that learners receive the appropriate level of challenge and support, enhancing their motivation and learning efficiency [11]. Predictive analytics in e-learning involves using historical data to forecast future outcomes [12]. By analyzing patterns in learner behavior and performance, ML models can predict which students are likely to succeed or struggle in a course [13]. This predictive capability allows educators to provide timely interventions, such as additional resources or personalized support, to help at-risk students improve their performance. Predictive analytics can also inform curriculum design and teaching strategies by identifying areas where learners commonly face difficulties [14].

II. COLLABORATIVE FILTERING BASED DEEP LEARNING (CFDL)

Collaborative filtering has become a cornerstone technique in recommendation systems, playing a critical role in platforms ranging from e-commerce to content streaming services. With the advent of deep learning, the capabilities of collaborative filtering have been significantly enhanced [15]. Deep learning-based collaborative filtering techniques can analyze vast amounts of data, capture complex patterns, and provide highly accurate recommendations. The training of deep neural networks, one must confront the challenges of general nonconvex optimization problems. Local gradient descent methods that most deep learning systems rely on, such as variants of stochastic gradient descent (SGD), have no guarantee that the optimization algorithm will converge to a global minimum. It is known that an ensemble of multiple instances of a target neural network trained with different random seeds generally yields better predictions than a single trained instance. However, an ensemble may be computationally expensive at inference time. The framework of collaborative learning consists of three major parts: the generation of a population of classifier heads in the training graph, the formulation of the learning objective, and optimization [16].

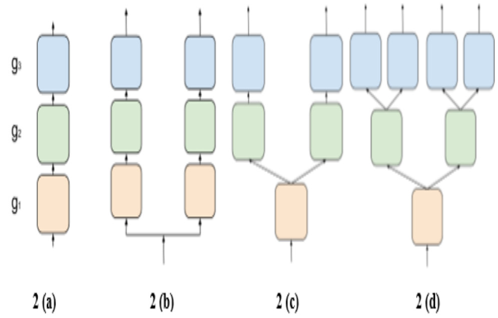

**Fig.2 (a) Target Network (b) Multiple Instances (c) Simple ILR Sharing
(d) Hierarchical ILRs haring**

The figure above depicts the sub-categories of training. Similar to auxiliary training [8], we add several new classifier heads into the original network graph during training time. At inference time, only the original network is kept and all added parts are discarded. Unlike auxiliary training, each classifier head here has an identical network to the original one in terms of graph structure. This approach leads to advantages over auxiliary training in terms of engineering effort minimization. First, it does not require to design additional networks for the auxiliary classifiers. Second, the structure symmetry for all heads does not require additional different weights associated with loss functions to well balance injected backpropagation error flows, because an equal weight for each head's objective is optimal for training. Mathematically [17]:

If the target network to be trained in figure 2(a) is given by:

$$\boldsymbol{z} = \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{\theta}) \qquad \textbf{(1)}$$

Here,

g is determined by the graph architecture
θ represents the network parameters.
The term g can also be represented as the cascade of the following sub-nets, given mathematically by:

$$\boldsymbol{g(x,\theta) = g3(g2(g1(x1,\theta1),\theta2),\theta3)} \quad \textbf{(2)}$$

The cascade of the network is often termed as **Ensemble Neural Network (ENN).**

Here,

$$\boldsymbol{\theta = [\theta1,\theta2,\theta3]} \quad \textbf{(3)}$$

In general, it is observed that that the training memory size is roughly proportional to the number of layers/operations. With the multi-instance pattern, the number of parameters in the whole training graph is proportional to the number of heads. Obviously, ILR sharing can proportionally reduce the memory consumption and speed up training, compared to multiple instances without sharing.

The feature learning can be expressed as:

Feature learning is the process of estimating the features based on:

Given $\theta_1, \theta_2, \ldots\ldots\ldots\ldots\ldots\ldots\ldots.\theta_{nu}$, to learn the value of features x_i

The following cost function is to be minimized:

$$\boldsymbol{\min\left[\sum_{j:r(i,j)=1}^{n} {[\theta_j^T x^i - y^{i,j}]^2}/{n}\right] + \frac{\lambda}{2}\sum_{i=1}^{n}[x_i]^2} \quad \textbf{(4)}$$

Here,

i is the number of features

j is the number of features which are labelled

$\sum_{j:r(i,j)=1}^{n} {[\theta_j^T x^i - y^{i,j}]^2}/{n}$ is the mean square error (mse) of actual class and predicted class.

$\frac{\lambda}{2}\sum_{i=1}^{n}[x_i]^2$ is the regularizing term to limit the number of features to become too large, to avoid overfitting.

Collaborative filtering is a method used to predict user preferences based on past interactions and the preferences of similar users. It can be divided into two main types: user-based and item-based collaborative filtering. User-based collaborative filtering identifies users with similar tastes and recommends items they liked. Item-based collaborative filtering, on the other hand, finds items that are similar to those a user has enjoyed in the past. Despite its effectiveness, traditional collaborative filtering often struggles with sparse data and scalability issues. Deep learning, with its ability to model complex relationships, offers a powerful extension to traditional collaborative filtering. By leveraging neural networks, deep learning-based collaborative filtering can automatically learn feature representations from raw data, such as user-item interaction [18].

III. RMS Prop for CFDL

RMSProp, which stands for Root Mean Square Propagation, is an adaptive learning rate optimization algorithm designed to address some of the issues encountered with the stochastic gradient descent (SGD) method in training deep neural networks. The RMSProp algorithm has been found to be an improved version of the stochastic gradient descent (SGD) [19]. The SGD is a widely used optimization technique for training machine learning models, particularly deep neural networks. However, SGD has some limitations, especially when dealing with complex optimization landscapes. One significant challenge is the choice of a global learning rate for the model's parameters. If the learning rate is too high, the model may overshoot minima, and if it's too low, training can become extremely slow and may get stuck in local minima or saddle points [20].
RMSProp addresses the issue of a global learning rate by maintaining a moving average of the squares of gradients for each weight and dividing the learning rate by this average. This ensures that the learning

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 07 | July - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

rate is adapted for each weight in the model, allowing for more nuanced updates. The general idea is to dampen the oscillations in directions with steep gradients while allowing for faster movement in flat regions of the loss landscape. The RMSProp update adjusts the Adagrad method to reduce its aggressive, monotonically decreasing learning rate. Instead of accumulating all past squared gradients, RMSProp uses an exponential decay that discards history from the extreme past so that it can converge rapidly after finding a convex bowl, as if it were an Adagrad with a fresh start. The RMSProp for CFDL can be expressed as:

Algorithm:

Start.

Step.1: Set maximum number of iterations to Maxitr

Step.2: Initialize weights randomly.

Step.3: Define cost function: $\boldsymbol{J}(\boldsymbol{w}, \boldsymbol{\theta})$

Step.4: Compute gradient as:

$$\boldsymbol{g} = \boldsymbol{\nabla}(\boldsymbol{w}, \boldsymbol{\theta})$$

Step.5: Compute squared gradient as:

$$\boldsymbol{E}\left(\boldsymbol{g}_t^2\right) = \boldsymbol{\alpha} \boldsymbol{g}_{t-1}^2 + (\mathbf{1} - \boldsymbol{\alpha}) \boldsymbol{g}_t^2$$

Step.6: Compute learning rate as:

$$\chi_{t+1} = \frac{\chi_t}{\boldsymbol{E}\left(\boldsymbol{g}_t^2\right) + \epsilon}$$

Step.7: Update weights as [21]:

$$\boldsymbol{w}_{t+1} = \boldsymbol{w}_t - \chi_t g_t$$

$$\boldsymbol{if}(\boldsymbol{J}\ \boldsymbol{converges}\ \boldsymbol{or}\ \boldsymbol{t} = \boldsymbol{Maxitr}$$

$$\boldsymbol{stop}\ \boldsymbol{training}$$

else

iterate over step 7.

Stop.

RMSProp offers several advantages over standard SGD:

Adaptive Learning Rates: By adjusting the learning rate for each parameter, RMSProp can handle different scales of data and varying curvatures of loss functions.

Convergence Speed: RMSProp can converge faster than SGD with momentum, especially in scenarios with noisy or sparse gradients.

Stability: The method avoids the diminishing learning rates found in Adagrad, which can stall the training process in the later stages.

IV. RESULTS

The dataset used in this study is the Book Crossing Dataset from Kaggle. https://www.kaggle.com/datasets/syedjaferk/book-crossing-dataset.

The system has been designed for recommending e-books as learning resources but can be extended to other learning resources as well. The deep neural network designed bas 100 units of neurons in each of the hidden layers. The RMSprop has been used as the training function which is computationally more efficient compared to the conventional gradient descent.

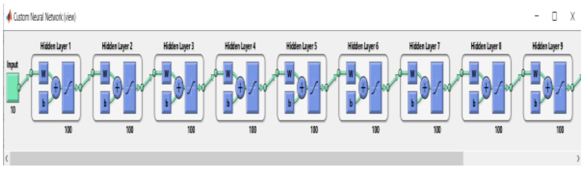

Fig.3 Network Architecture

Figure 3 depicts the deep learning architecture with 10 hidden units of 100 neurons each stacked through the CFDL approach.

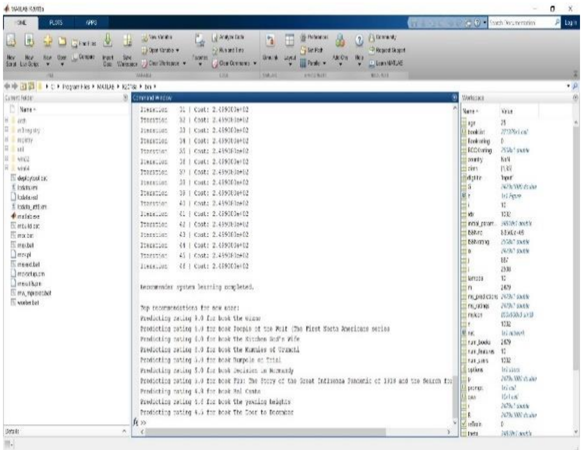

Fig.4 Training the CFDL model with RMSProp

Figure 4 depicts the CFDL model being trained through the collaborative learning appraoch.

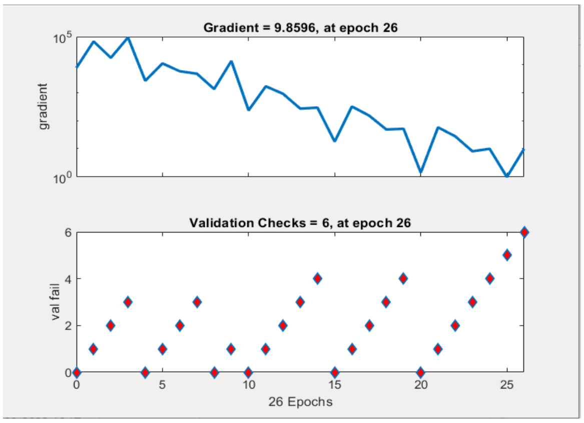

Fig.5 Training States

Figure 5 depicts the variation in the gradient and the validation check status of the designed CFDL model

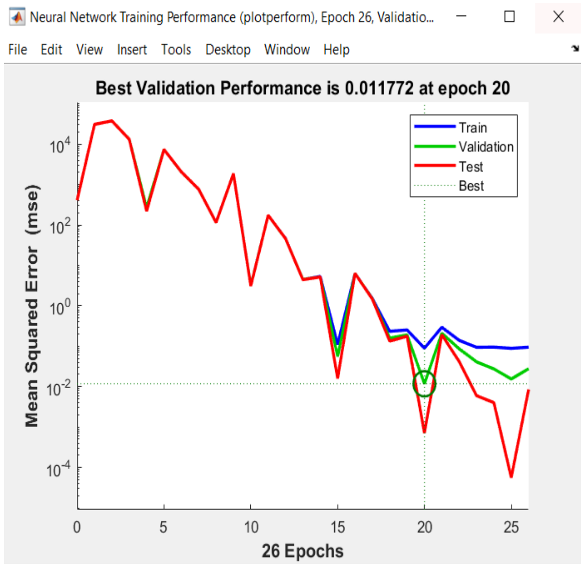

Fig.6 Training to Convergence

Figure 6 depicts the training to convergence for the CFDL model.

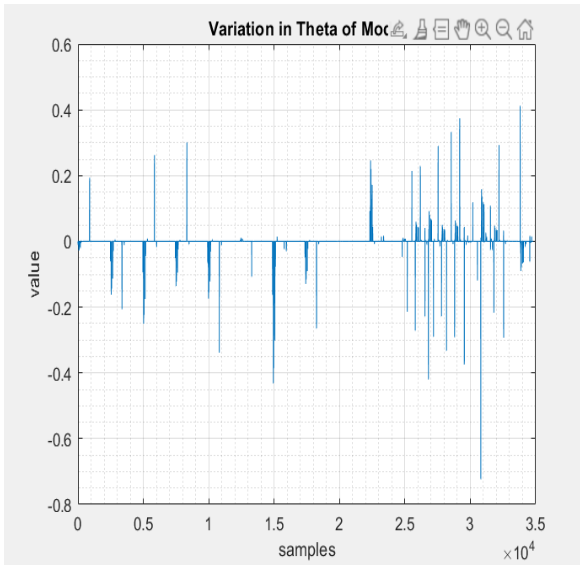

Fig.7 Variation in Theta

Figure 7 depicts the variation in Theta for the CFDL model.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 07 | July - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

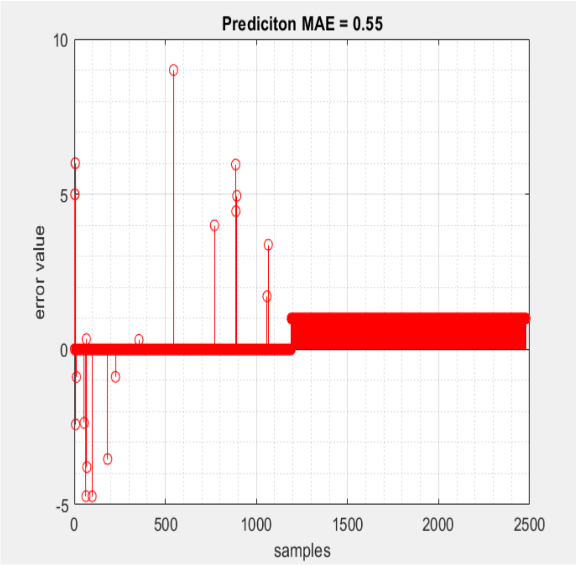

Fig.8 Error Plot for Predictions

Figure 8 depicts the prediction error for the proposed system. The MAE obtained in this case is 0.55 only compared to best case MAE of 5 in previous work [2].

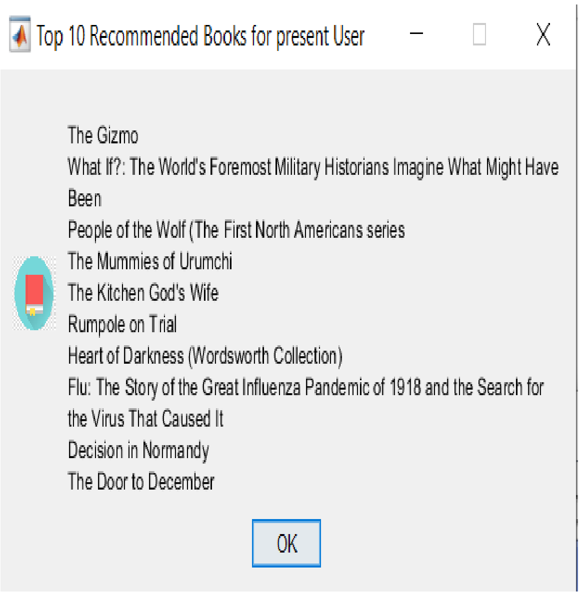

Fig.9 Recommended Resources

Figure 9 depicts the actual learning resource recommendations by the CFDL model. A similar approach can be used to predict or recommend other learning resources or intelligent tutoring systems.
Table 1 summarizes the results obtained through the designed model

Table.1 Summary of Results

S.No.	Parameter	Value
1	Dataset	https://www.kaggle.com/datasets/syedjaferk/book-crossing-dataset
2	ML Model	Collaborative Filtering Based Deep Learning (CFDL)
3	Training Rule	RMS Prop
4	Neurons in Hidden Units	100
5	Iterations	26
6	Cost Function	0.0886
7	MAE	0.55
8	MAE of Previous Work [2]	5

It can be observed that the proposed approach comprising of CFDL and RMS Prop attains a very low convergence MSE of just 0.0886 at 26 iterations. This is particularly important for large recommendation systems which need to churn out data for a multitude of users details. Moreover, the approach attains an MAE of just 0.55 outperforming existing work in the domain [2].

CONCLUSION

It can be concluded from the previous discussions that with more candidates opting for e-learning applications due to its specific advantages, it has become necessary to design an optimized recommendation system for learning resources. One of the major challenges is however the plethora of resources to choose from. This paper presents a CFDL based approach trained with the RMSProp algorithm which happens to be more computationally effective compared to standard SGD models in terms of stability and error convergence. It has been shown that the proposed approach attains an MAE of just

0.55 clearly outperforming existing contemporary work in the domain.

References

1. Tulsi B., Suchitra R, "Big Data Analytics And E Learning In Higher Education", International Journal on Cybernetics & Informatics, 2016, vol. 5, no. 1,pp.81-85.
2. S Bhaskaran, R Marappan, "Design and analysis of an efficient machine learning based hybrid recommendation system with enhanced density-based spatial clustering for digital e-learning applications", Complex & Intelligent Systems, Springer, 2023, vol.9, pp. 3517–3533.
3. H Oubalahcen, L Tamym, "The Use of AI in E-Learning Recommender Systems: A Comprehensive Survey", Procedia Computer Science, Elsevier 2023, vol.224, pp. 437-442.
4. G. S. Hukkeri and R. H. Goudar, "Machine Learning-Based Personalized Recommendation System for E-Learners," 2022 Third International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), Bengaluru, India, 2022, pp. 1-6
5. S. M. Aslam, A. K. Jilani, J. Sultana and L. Almutairi, "Feature Evaluation of Emerging E-Learning Systems Using Machine Learning: An Extensive Survey," in IEEE Access, vol. 9, pp. 69573-69587, 2021
6. SS Khanal, PWC Prasad, A Alsadoon, "A systematic review: machine learning based recommendation systems for e-learning", Education and Information Technologies, Springer 2020, vol.25, pp. 2635–2664.
7. G. Srivastav and S. Kant, "Review on e-Learning Environment Development and context aware recommendation systems using Deep Learning," 2019 3rd International Conference on Recent Developments in Control, Automation & Power Engineering (RDCAPE), 2019, pp. 615-621
8. D. F. Murad, Y. Heryadi, B. D. Wijanarko, S. M. Isa and W. Budiharto, "Recommendation System for Smart LMS Using Machine Learning: A Literature Review," 2018 International Conference on Computing, Engineering, and Design (ICCED), 2018, pp. 113-118.
9. X. Wang, Y. Zhang, S. Yu, X. Liu, Y. Yuan and F. -Y. Wang, "E-learning recommendation framework based on deep learning," 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Banff, AB, Canada, 2017, pp. 455-460.
10. D. Herath and L. Jayaratne, "A personalized web content recommendation system for E-learners in E-learning environment," 2017 National Information Technology Conference (NITC), Colombo, Sri Lanka, 2017, pp. 89-95.
11. M. Moreno, ., Segrera, S., López, V.F., Muñoz, M.D., Sánchez, Á.L.: Web mining based framework for solving usual problems in recommender systems. A case study for movies′ recommendation. Neurocomputing Springer, 2016, 176, 72-80.
12. Wang, Y., Shang, W.: Personalized news recommendation based on consumers' click behavior. In: Fuzzy Systems and Knowledge Discovery (FSKD), 2015 12th International Conference on, 2015. IEEE, pp 634–638
13. Shi, B., Ifrim, G.: Hurley N Learning-to-Rank for Real-Time High-Precision Hashtag Recommendation for Streaming News. In: Proceedings of the 25th International Conference on World Wide Web, 2016. International World Wide Web Conferences Steering Committee, pp 1191–1202
14. MK Khribi, M Jemni, O Nasraoui, "Recommendation systems for personalized technology-enhanced learning", Ubiquitous Learning Environments and Technologies, Springer 2015, pp.159–180.
15. M. Gupta, A. Thakkar, Aashish, V. Gupta and D. P. S. Rathore, "Movie Recommender System Using Collaborative Filtering," 2020 International Conference on Electronics and Sustainable Communication Systems

(ICESC), Coimbatore, India, 2020, pp. 415-420.

16. K Mao, J Zhu, J Wang, Q Dai, Z Dong, X Xia, "SimpleX: A simple and strong baseline for collaborative filtering", Proceedings of the 30th ACM International Conference on Information & Knowledge Management, pp.1243-1252.
17. L. Wu, X. He, X. Wang, K. Zhang and M. Wang, "A Survey on Accuracy-Oriented Neural Recommendation: From Collaborative Filtering to Information-Rich Recommendation," in IEEE Transactions on Knowledge and Data Engineering, 2023, vol. 35, no. 5, pp. 4425-4445.
18. A. Iftikhar, M. A. Ghazanfar, M. Ayub, Z. Mehmood and M. Maqsood, "An Improved Product Recommendation Method for Collaborative Filtering," in IEEE Access, vol. 8, pp. 123841-123857, 2020.
19. T. Lu, J. Sun, K. Wu and Z. Yang, "High-Speed Channel Modeling With Machine Learning Methods for Signal Integrity Analysis," in IEEE Transactions on Electromagnetic Compatibility, vol. 60, no. 6, pp. 1957-1964.
20. R. Zaheer and H. Shaziya, "A Study of the Optimization Algorithms in Deep Learning," 2019 Third International Conference on Inventive Systems and Control (ICISC), Coimbatore, India, 2019, pp. 536-539.
21. F Zou, L Shen, Z Jie, W Zhang, "A sufficient condition for convergences of adam and rmsprop", roceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 11127-11135.