Journal of Engineering, Project, and Production Management
2026, 16(4), 2025-291

An Improved CF Algorithm Personalized Recommendation Model for E-Commerce Cold-Start Problems

Wenwei Chen
Associate Professor, School of Business Management, Hangzhou Polytechnic, Hangzhou, 311402, China, E-mail: chenwenwei1979@163.com

Project Management
Received November 26, 2025; revised February 2, 2026; accepted February 9, 2026
Available online May 29, 2026

Abstract: Given the low recommendation accuracy of traditional algorithms in cold-start scenarios in e-commerce, this study proposes a new personalized recommendation algorithm. First, the independent meta-learning model is used to improve the collaborative filtering algorithm, enabling fast parameter adaptation for new users. Multimodal content encoding is combined to introduce product-related multimodal information to alleviate semantic sparsity. Second, a population-society dual regularization constraint is introduced to address the vector offset problem during fine-tuning in Model-Agnostic Meta-Learning (MAML). Ultimately, a personalized recommendation model is constructed for cold-start scenarios. The model was validated on the Amazon Electronics Dataset (AED) and AlicCP datasets. In the experiment, Hit Rate at 5 (HR@5), which measures how often the correct item appears among the top five recommendations, improved by 62.99% on AED when compared to the baseline. In the AlicCP dataset, its HR@5 and Novelty at 5 (Novelty@5) increased by an average of 26.02% and 15.23% compared to other methods. The research model can effectively achieve accurate recommendations in cold-start scenarios, bring users a better shopping experience, and improve the platform's first purchase conversion rate and next day retention rate.

Keywords: Collaborative filtering, e-commerce, cold-start recommendation, multimodal content encoding, meta learning.

DOI 10.32738/JEPPM-2025-291

1. Introduction

In an era where the traffic dividend of e-commerce platforms is gradually diminishing, precise, personalized recommendations can improve consumer decision-making efficiency and drive high user stickiness and sustained growth in total product transactions on the platform (Sharma et al., 2022). However, new users are trapped in a dilemma because no personalized recommendations are available due to zero interaction records, resulting in sharp drops in the platform's first-purchase conversion rate and next-day retention rate (Panda and Ray, 2022). Therefore, how to achieve personalized recommendations in Cold-Start Scenarios (CSSs) has become a key challenge for improving the performance of E-Commerce Recommendation Systems (ECRSs). Many scholars have researched this (Gheisari et al., 2023). Raji et al. (2024) developed an AI-driven personalized recommendation technology analysis framework to address the difficulty of grasping dynamic changes in consumer behavior in the e-commerce field, thereby improving consumer experience and platform operational efficiency. Liao and Sundar (2022) proposed a dual-factor influence model of "content matching+group effect" to address differences in recommendations between subjective and objective quality products, thereby improving the accuracy and user acceptance of ECRS. Multi-Modal Feature (MMF) extraction and user interest modeling in e-commerce recommendation suffer from noise interference caused by inconsistent targets. Liu et al. (2023) proposed a multi-interest evolutionary graph collaborative filtering model that integrates semantic entities and graph convolutional networks to optimize matching between MMFs and user preferences. Wu and Chi (2023) proposed a differentiated recommendation scheme to address cold-start issues for new users and changes in demand for existing users, thereby optimizing user retention and conversion effects. In e-commerce interactive recommendation, there are issues with user cold-start, and existing deep reinforcement learning methods neglect multi-hop social relationships. Ma et al. (2023) proposed a social graph neural recommendation framework based on social graph neural networks, which improved the recommendation performance of cold-start users. The Collaborative Filtering (CF) algorithm is a recommendation technique based on swarm intelligence that uses group behavior similarity for preference inference. Collaborative Filtering

(CF) is widely used in personalized recommendations, such as e-commerce, because it does not require domain knowledge and can automatically discover potential associations (Ahamed et al., 2024). Nahta et al. (2021) proposed a multi-modal embedding dynamic CF model. This model integrated user portraits and product category metadata, addressing the problems of traditional recommendation models that ignore metadata, leading to poor cold-start performance and high prediction errors, thereby improving recommendation accuracy in Cold-Start Scenarios (CSS). Xu et al. (2021) proposed a multi-domain CF model based on MMF fusion to address cold-start and low optimization efficiency in hash recommendation, thereby improving recommendation performance in CSS. Anwar et al. (2022) proposed an improved method based on a K-nearest neighbor baseline to address cold-start and data sparsity in CF, thereby reducing the recommendation error rate. Kannout et al. (2023) proposed a personalized recommendation framework built on feature aggregation that integrates CF, content filtering, and frequent item mining to address item cold start in recommendation systems, thereby improving the accuracy of new item recommendations. Wang et al. (2021) proposed an improved algorithm that integrates matrix filling and temporal context to address low accuracy and poor scalability in traditional CF algorithms, thereby improving recommendation accuracy and system scalability. In summary, existing research can be divided into three categories: user profile-based methods, social relationship-based methods, and MFF-based methods. Specifically, the user profile-based approach constructs initial profiles for new users by introducing static attributes such as age and gender. Although it can achieve preliminary matching, the profile information is often sparse and static, making it difficult to capture dynamically changing interests. In contrast, the social relationship-based method primarily leverages user's social networks for trust transmission and interest diffusion. However, its effectiveness depends on dense social networks, thus limiting its efficacy for isolated new users or privacy-sensitive scenarios. MFF-based methods compensate for the lack of interaction data by mining content features such as product text and images. Nevertheless, they still have shortcomings in MFF alignment and deep semantic fusion and have not been effectively integrated with meta-learning mechanisms that rapidly adapt to new users. A comprehensive analysis shows that although existing research has explored the cold-start problem in ECRS from multiple dimensions, personalized recommendations for new users in CSSs still face challenges, including low recommendation accuracy and poor real-time performance. In particular, the dynamic evolution process of user interests in the "zero-interaction" stage cannot be effectively modeled. To this end, this study proposes an improved CF algorithm using Model-Agnostic Meta-Learning (MAML) and a user Cold-Start Personalized Recommendation (CSPR) model combined with Multi-modal Content Encoding (MCE). This model aims to achieve real-time, accurate, and compliant personalized recommendations for new users. Given the limited accuracy of personalized customer recommendations during the cold-start stage, this study developed an improved model to deliver real-time, accurate recommendations under zero-interaction conditions, thereby improving the platform's first-purchase conversion rate and next-day retention rate. Based on the above related research, the research topics, main indicator methods, and deficiencies are summarized in Table 1.

The above literature shows that existing methods have made significant progress in general recommendation scenarios. However, their performance is still sub-optimal when facing the strict cold-start problem, especially for new users with no interaction records. Furthermore, semantic sparsity, vector drift during fine-tuning, and inadequate modeling of dynamic user interests have not been fully addressed. Therefore, this study innovatively integrates MAML with MCE and introduces a population-social dual regularization strategy to stabilize user vector learning. A federated learning deployment is also incorporated to ensure privacy protection, thereby constructing a robust and adaptive personalized recommendation model for cold-start scenarios.

2. Methods and Materials

This study first introduces the CF algorithm, clarifies its shortcomings, and then uses MAML to optimize the algorithm. Subsequently, an MCE module embedded in the CF framework is introduced, and an MAML+MCF model is designed.

2.1. Improved CF Algorithm based on MAML

The CF algorithm is one of the most representative technologies in recommendation systems. In the e-commerce scenario, the CF algorithm analyzes user browsing, bookmarking, purchasing, and other behavioral data to construct a user-product interaction matrix. Then, it identifies similar user groups or associated product sets, ultimately generating personalized recommendation lists that meet individual preferences (Papadakis et al., 2022). Traditional CF achieves systematic integration of user behavior data, precise quantification of user-product associations, and efficient generation of recommendation results through a three-layer progressive structure design that includes behavior data collection, matrix construction, similarity calculation, and recommendation generation. Its structure is shown in Fig. 1.

In Fig. 1, the CF algorithm generates a recommendation list by first collecting user behavior data on the platform, which can be divided into explicit and implicit behavior. The collected behavioral data are cleaned and transformed into a matrix structure. The entries in the user-product interaction matrix reflect the strength of user interaction with the product, and the matrix is then used to calculate user similarity. Finally, a recommendation list is generated based on similar calculation results. Among them, the similarity calculation is shown in Eq. (1).

$$sim(u,v)=\frac{\sum_{i\in I_{uv}}(R_{ui}-R_u)(R_{vi}-R_v)}{\sqrt{\sum_{i\in I_{uv}}(R_{ui}-\overline{R_u})^2}\sqrt{\sum_{i\in I_{uv}}(R_{vi}-\overline{R_v})^2}} \tag{1}$$

In Eq. (1), u is the target user, v is the user ID. I_{uv} is the collection of products that u and v have interacted with together. R_{ui} is u's rating for product i. $\overline{R_u}$ is the average rating of u. Cosine similarity only relies on the product sum of non-zero dimensions of vectors, and has immunity to non-interacting terms, avoiding the misleading calculation results caused by zero value differences in traditional Euclidean distance. Meanwhile, its computation only requires multiplication and summation of non-zero dimensions, with a much lower complexity than Euclidean distance.

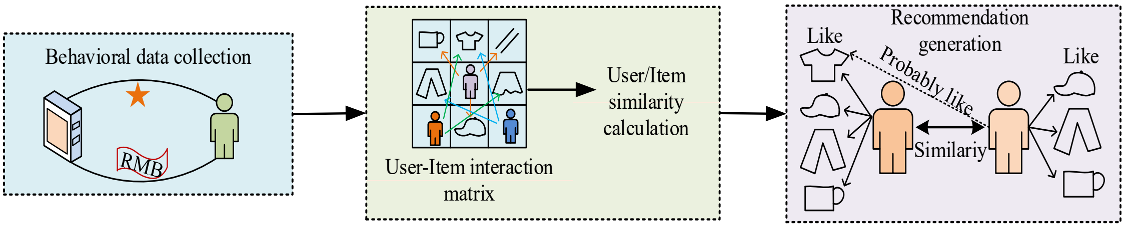


Fig. 1. CF structure diagram

Therefore, for e-commerce systems with millions of users, it can reduce the resources required for real-time similarity calculation. Subsequently, based on the similarity calculation results, interpretable recommendation results can be generated, but it requires the use of a rating prediction model to predict user's preference ratings for non-interactive products, and generate the top N recommendation items according to the rating sequence. The calculation of the rating prediction model is shown in Eq. (2).

$$\begin{cases} \overrightarrow{R_{uj}} = \overline{R_u} + \dfrac{\sum_{v \in N(u)} sim(u,v) \cdot (R_{vj} - \overline{R_v})}{\sum_{v \in N(u)} \left| sim(u,v) \right|} \\ \overrightarrow{R_{uj}} = \overline{R_u} + \dfrac{\sum_{k \in N(j)} sim(j,k) \cdot R_{uk}}{\sum_{k \in N(j)} \left| sim(j,k) \right|} \end{cases} \tag{2}$$

In Eq. (2), j is the target product ID, and $N(u)$ is the nearest neighbor set of u. $\overrightarrow{R_{uj}}$ is the predicted rating of the product j by u. $N(j)$ is the nearest neighboring set of product j. $sim(j,k)$ is the similarity between product j and k. Due to its reliance on explicit rating data in generating recommendations, the CF algorithm still has inherent limitations, such as its inability to handle sparse matrices and e-commerce CSS. Specifically, the model is unable to estimate user vectors and generate effective personalized recommendations when encountering zero interactions from new users. Therefore, to generate high-quality personalized recommendations in the CSS of e-commerce, this study introduces the MAML architecture. MAML is a type of meta-learning algorithm that pre-trains the model on multiple tasks to learn a good initial parameter state, enabling it to achieve high performance with only a small number of gradient updates when facing new tasks (Mo et al., 2023). In the pre-training stage, it learns a universal recommendation model from a massive amount of historical user behavior, enabling new users to quickly adapt to their personal preferences. Its specific structure is shown in Fig. 2. In Fig. 2, the MAML framework operates on the core idea of learning a universal model initialization parameter across many tasks, enabling the model to adapt to new tasks rapidly. Within this framework, each task is defined as a user group with distinctive preference patterns, i.e., a user category. Initially, a batch of tasks Ti is randomly sampled from the task distribution, with all tasks sharing a set of initial parameters θ. Subsequently, the model uses the support set corresponding to each task (composed of a small amount of historical interaction data of the target user group) to simulate CSS and perform rapid inner-loop adaptation. Through k-step gradient descent, task-specific adapted parameters $\theta i'$ are generated. The generalization capability is then evaluated using the task's query set, which consists of data from the same user group that does not appear in the support set. The losses computed from the query sets are aggregated into a meta-loss, which drives the outer-loop update of the shared initial parameters. Eventually, when a new user arrives, it is identified as belonging to a certain user category. Using pre-trained initial parameters and based on its minimal initial interaction data, personalized adaptation is quickly achieved, thereby achieving accurate cold start recommendations. The update of task-specific parameters is shown in Eq. (3).

$$\theta_i^{'} = \theta - \alpha \nabla_\theta L_{Ti}(f_\theta) \tag{3}$$

In Eq. (3), α is the inner loop learning rate, and L_{Ti} is the loss function of task Ti. In the pre-training stage, for each task, the loss gradient is calculated based on the initial parameter θ, and task-specific parameters are generated through one or more gradient descent steps. This step simulates the rapid adaptation process of the model on a new task. The calculation of the meta loss is given by Eq. (4).

$$L_{meta} = \sum_{Ti \sim p(T)} L_{Ti}(f_{\theta_i^{'}}) \tag{4}$$

In Eq. (4), L_{meta} is the meta loss and $p(T)$ is the task distribution. The total loss of all sampling tasks under the adapted parameters is called meta loss, because optimizing the initial parameters can enable it to quickly adapt to various tasks after a small number of gradient updates. MAML is based on pre-training a universal initialization parameter for existing users, allowing new users to fine tune their personal vectors with only a small amount of data, achieving zero rating interaction in e-commerce. However, due to the extremely small number of fine-tuning samples, the vectors are prone to fitting or drifting. Therefore, to compensate for the fitting and drift risks during MAML fine-tuning, this study adds population social regularization to the loss function, as given by Eq. (5).

Table 1. Summary of relevant information of relevant studies

	Method	Dateset	Result	Reported results	Shortcomings
Raji et al. (2024)	AI-powered personalization strategies (e.g., machine learning, predictive analytics) and market trend analysis.	/	Provides a comprehensive review of how AI enhances customer engagement and shapes market trends.	/	Provides a high-level overview but does not address the cold-start problem specifically.
Liao et al. (2022)	A user study comparing Content-Based Filtering and CF algorithms	User experiment (N=469)	CBF preferred for experience goods, CF for search goods.	Significant preference shifts (p<0.05)	Lab study, lacks scalable cold-start algorithm.
Liu et al. (2023)	MEGCF (Graph model with semantic entities)	Multiple public datasets	Improves recommendati on accuracy.	HR@20: 0.3952, NDCG@20: 0.2215	Requires existing user interactions, fails for zero-interaction cold-start.
Wu et al. (2023)	Three-tiered strategy for different customer stages.	/	Improves customer engagement and retention.	/	Strategy-focused, lacks algorithmic innovation for zero-interaction new users.
Ma et al. (2023)	SGNR (Graph Neural Network with multi-hop social relations for DRL)	Two real-world datasets	Alleviates cold-start and improves long-term engagement.	/	Performance relies on rich social network data, ineffective for isolated new users.
Nahta et al. (2021)	MEDCF (fusion of GMF, MLP and NeuMF) with metadata	Movie Lens, Amazon Movies	Improves rating prediction accuracy	Significantly reduces RMSE/MAE vs baselines	Relies on static metadata, lacks fast adaptation to truly zero-interaction users.
Xu et al. (2021)	MDCF (multi-modal hashing with discrete optimization)	Large-scale datasets	Improves cold-start recommendati on efficiency	Outperforms SOTA hashing methods	Requires multi-modal auxiliary data, ineffective when such data is incomplete.
Anwar et al. (2022)	Comparison of multiple CF methods (KNN variants, SVD, etc.)	MovieTrust	KNNBaseline performs best in cold-start scenarios	Reduces prediction error rate	Traditional CF requires historical interactions; limited scalability for pure cold-start.
Kannout et al. (2023)	FPRS (hybrid of CF, content filtering and frequent pattern mining)	Three benchmark datasets	Mitigates item cold-start problem	/	Focuses on item cold-start; ineffective for new user cold-start with zero interactions.
Wang et al. (2021)	Improved CF with matrix filling and temporal context on Hadoop platform	E-commerce dataset	Improves recommendati on accuracy and scalability	Lower MAE vs UBCF/IBCF; improved scalability	Focuses on algorithm efficiency, not cold-start scenarios with zero-interaction users.
This study	MAML+MCF with population-social regularization	Amazon Electronics, AliCCP	Effectively addresses user cold-start problem.	HR@5: +62.99%; Novelty@5: +12.72% (vs. baseline CF)	/

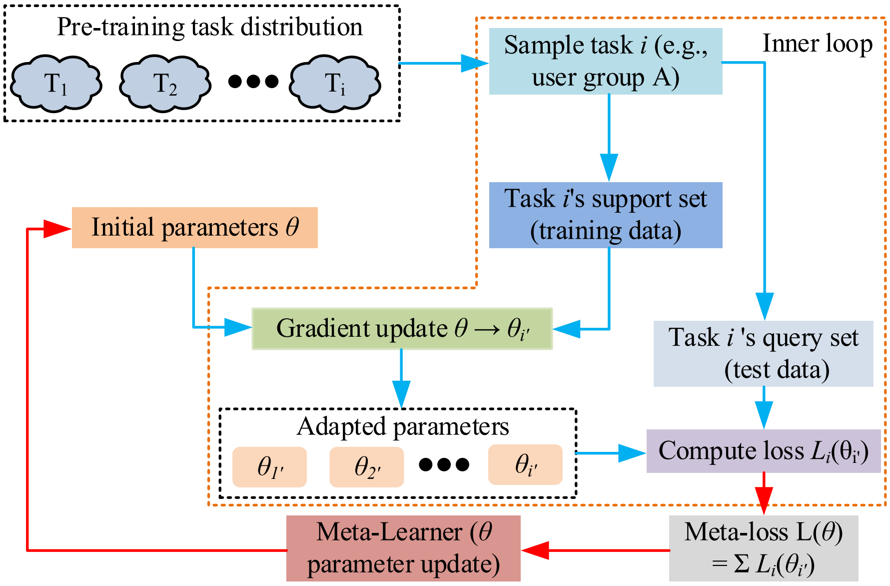


Fig. 2. MAML structure diagram

$$C = \lambda_1 \|V\_u\text{-}\phi\_demo\|_2^2 + \lambda_2 \|V\_u\text{-}V\_friend\_avg\|_2^2 \tag{5}$$

In Eq. (5), V_u is the new user cold-start vector, and ϕ_demo is the demographic center vector. V_friend_avg is the social center vector, and λ_1 and λ_2 are weights. The demographic-social regularization graph is shown in Fig. 3.

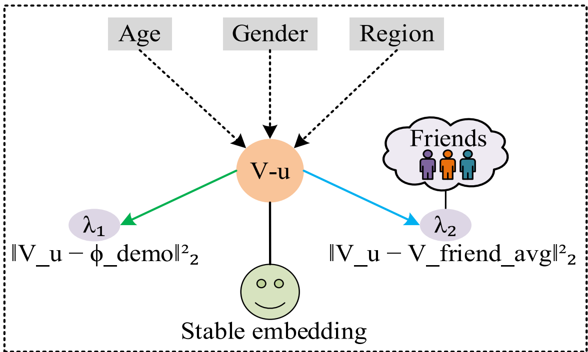


Fig. 3. Demographic-Social regularization schematic diagram

In Fig. 3, the basic demographic attributes, including age, gender, and region, are input and jointly applied to the middle "V_u" node. This indicates that the impact of basic demographic factors on the final generated results will be considered when constructing model-related representations. The regularization coefficient λ_1 and constraint term $\|V\_u\text{-}\phi\text{-}demo\|_2^2$ is the pulling force of demographic information on V_u. λ_2 and $\|V\_u\text{-}V\_friend\_avg\|_2^2$ are the pull-back effects of social relationships, such as friends. The dual role of population and society can constrain the cold-start vector from deviating from the reasonable region, thereby improving the stability and accuracy of recommendation results, and generating a robust user representation Stable embedding after dual regularization.

2.2. CSPR Model based on Improved CF Combined with MCE

This study improves the CF algorithm based on MAML. It adds population social regularization to the loss function, which can achieve zero interaction and generate stable user vectors in e-commerce. However, the improved CF only uses ID vectors and cannot generate diverse recommendation results using rich information such as text and images. This leads to severe homogenization of recommended content, making it challenging to match users' potential interests accurately, reducing long-term user stickiness, and triggering cross-platform migration behavior. This study further introduces the MCE module and constructs a CSPR model that combines MAML with Multi-modal CF (MCF) to solve the two major challenges of new product non-interaction and semantic sparsity. MCE refers to the process of converting heterogeneous multi-source information related to products or users, such as text descriptions, images, and structured attributes, into a unified continuous vector representation through modality-specific encoders in ECRS and aligning and fusing them in a shared semantic space (Feng et al., 2023; Wang et al., 2023). Its structure is shown in Fig. 4.

In Fig. 4, the text title, main image information, and product information are extracted by the MCE structure. Then, it will enter three different network structures: Bidirectional Encoder Representation from Transformers (BERT), Residual Neural Network (ResNet), and One-Hot Encoding (One Hot) for feature extraction, and obtain feature vectors with different attributes. To achieve semantic alignment of MFF, a shared semantic space mapping strategy is employed. The feature vectors from different sources are concatenated and then non-linearly transformed via a shared fully-connected layer, i.e., a mapping matrix W_s, which projects them into a unified 512-dimensional common semantic space. This process has the dual purpose of dimensionality reduction and making text, visual, and attribute features co-adapt in this space according to the recommendation task goals, thereby achieving implicit semantic alignment and laying the foundation for subsequent CF calculations. Ultimately, the 512-dimensional vector obtained through this fully-connected layer constitutes the final multi-modal embedding vector. The encoding of different modal information is specifically shown in Eq. (6).

$$\begin{cases} h_t = BERT(X_j) \in R^{768} \\ h_v = \mathrm{Re}\,sNet50(I_j) \in R^{2048} \\ h_a = Embedding(A_j) \in R^{64} \end{cases} \tag{6}$$

In Eq. (6), h_t is text encoding, h_v is image encoding, and h_a is attribute encoding. X_j, I_j, and A_j are the title text, main image, and structured attributes of product j. The encoding vectors of different modalities are concatenated to form a comprehensive multi-modal vector, which is calculated as shown in Eq. (7).

$$m_j = \mathrm{Re}\,LU(W\left[h_t; h_v; h_a\right] + b \in R^{512} \tag{7}$$

In Eq. (7), m_j is the multi-modal fusion vector of product j, b is the bias term, and $W \in R^{512\times(768+2048+64)}$ is the trainable weight matrix. Multi-modal encoding vectors need to be processed through FCLs before they can be mapped to lower dimensional spaces, thereby reducing the number of features and lowering the computational complexity and overfitting risk. MCE encodes the features of the product content itself into a vector and calculates its similarity with the user vector to enter the recommendation candidate pool and improve the missing information of user behavior data. This module also integrates semantically rich vectors into ECRS through fine-grained semantic mining of product text and images, thereby recommending products that meet the needs of users. However, MMFs often contain sensitive information, and centralized training contains a lot of personal information, which can easily infringe on personal privacy. Therefore, federated learning deployment has been carried out to reduce the risk of privacy leakage and ensure the security of user information. The model deployment process is shown in Fig. 5.

In Fig. 5, each platform first registers its identity and submits a public key. Subsequently, all participating parties specify their local data feature dimensions (e.g., e-commerce platforms possess user behavior data, while payment platforms hold payment data). In practice, data distributions among participants are often non-independent and identically distributed (non-IID), meaning data across platforms exhibit significant heterogeneity. The process is as follows: After confirming sample consistency, each platform uses its local data to calculate intermediate model results and uploads them to the coordination server. For sensitive features, homomorphic encryption is applied without transmitting the raw data to ensure privacy protection. The server collects encrypted gradients from all parties, decrypts and aggregates them. To mitigate the potential negative impact of non-IID data distribution on model performance, the framework adopts a weighted average strategy on the server side for gradient aggregation. It dynamically assigns weights based on each participant's data volume or data quality, thereby optimizing the update direction of global model parameters. The updated global parameters are then encrypted and distributed to each platform to initialize their local models. Finally, the local inference results from all platforms are fused via weighted aggregation to generate the final recommendation score. Overall, this study improves the CF algorithm by introducing two architectures, MAML and MCE, to address the cold-start problem in e-commerce. It uses population social regularization and federated learning deployment to optimize recommendation results, thereby constructing an MAML+MCF CSPR model, as shown in Fig.6. In Fig.6, in the e-commerce system, the basic information of new users and the tokens of initially browsed items are fed into the model as parallel inputs. Following the user-side data flow indicated by the blue arrows, the basic information of new users flows into the MAML framework, where it undergoes rapid adaptation via meta-learning to generate an initial user feature vector. This vector is subsequently constrained and refined by the population-society dual regularization module, resulting in a more stable and rational robust user vector V_u. Meanwhile, the item-side data flow marked by the green arrows shows that the item tokens enter the MCE module, where they are encoded into a unified multi-modal item vector v through feature extraction and fusion layers. Thereafter, u_{final} and v are jointly fed into the MCF layer for interactive computation to generate recommendation scores. Finally, to ensure user privacy, the recommendation results undergo secure aggregation processing through the federated learning layer before outputting the personalized recommendation list. It is noteworthy that the red dashed arrows further indicate the backward parameter update path during training, which propagates from the MCF layer back to both the MAML and MCE modules. The collaborative processing mechanism of the federated learning layer is specifically defined by Eq. (8).

$$\begin{cases} \nabla_k^{enc} = Enc_{pk}(\nabla L_k) \\ \nabla_{global} = Dec_{sk}(\sum_{k=1}^{K} \nabla_k^{enc}) \end{cases} \tag{8}$$

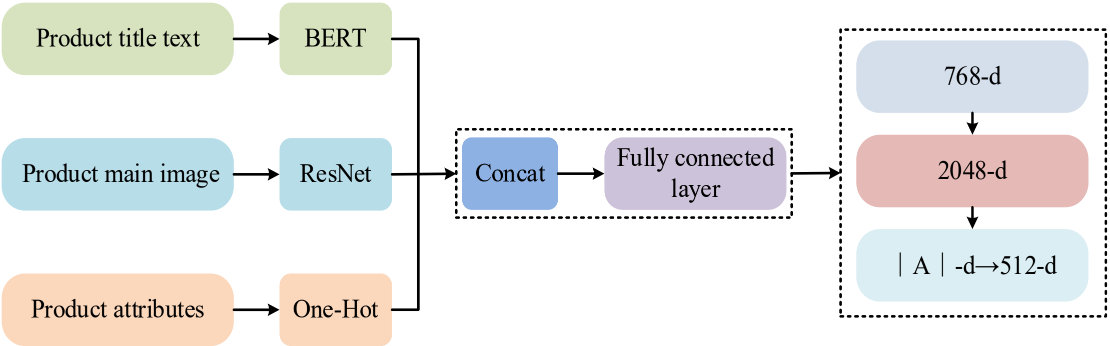


Fig. 4. MCE structure diagram

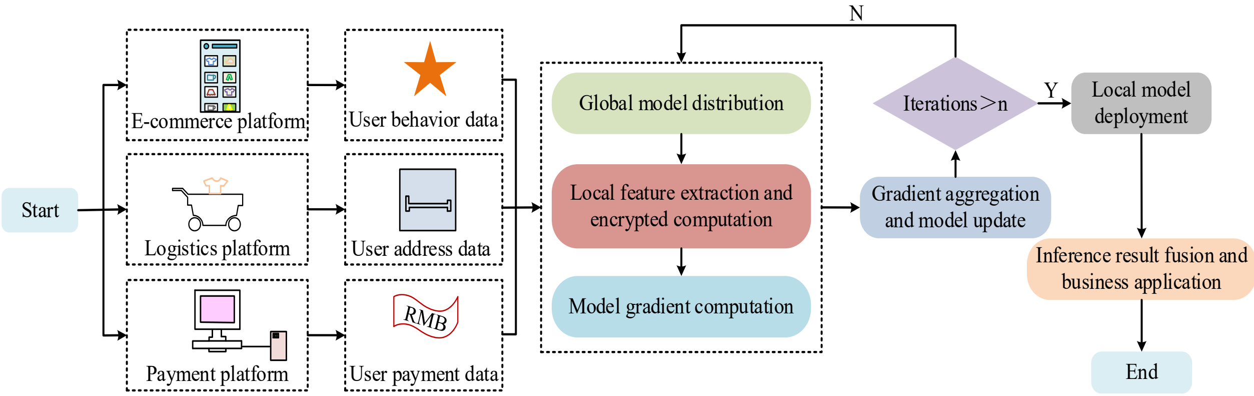


Fig. 5. Federated learning deployment flowchart

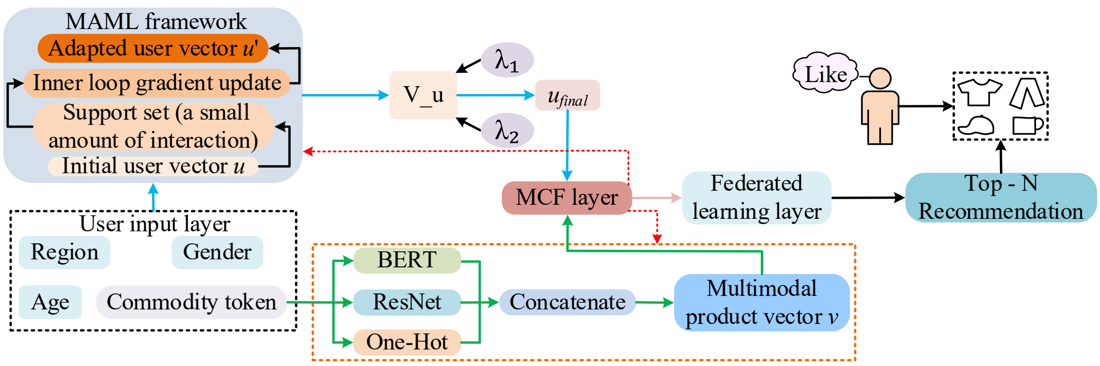


Fig. 6. MAML+MCF structure diagram

In Eq. (8), L_k is the local loss of the k -th participant. pk is the public key of the coordination server. sk is the private key of the coordination server. In the model testing phase, further verification is needed for the stability of the cold-start vector $\mathrm{V\_u}$, and the verification process is shown in Eq. (9).

$$L_{test} = \left\| V_u^{pred} - V_u^{ture} \right\|_2^2 + \lambda_1 \left\| V_u - \phi\_demo \right\|_2^2 + \lambda_2 \left\| V_u - V\_friend \right\|_2^2 \tag{9}$$

In Eq. (9), L_{test} is the loss function used during the testing phase to verify the stability of the cold-start vector. V_u^{pred}

is the predicted value of the vector representation of a cold-start user u by the model. λ_1 and λ_2 is the regularization parameter. V_u^{ture} is the true value represented by the u-vector of the cold-start user. ϕ-demo is a vector related to demographic information. V_friend is a vector that may be related to the user's social relationships. $\|\cdot\|_2^2$ is the square of the Euclidean norm.

3. Results and Discussion

This study first introduced the experimental setup and related parameters and analyzed the recommended performance of models with different combination structures in ablation experiments. Secondly, the recommendation performance before and after the model improvement was verified, and finally, other personalized recommendation models were introduced for performance comparison.

3.1. Recommendation Performance Evaluation of MAML+MCF Model in CSS

To verify the effectiveness of the MAML+MCF model in CSS, this study conducted performance validation based on the Amazon Electronics Dataset (AED) and AlicCP datasets. As mentioned earlier, the AED comprised user behavior logs and product information snapshots from real e-commerce platforms (Hamza et al., 2024). AlicCP was released by Alibaba Group in 2018, with data sourced from real user logs after anonymization on Taobao and Tmall platforms (Zhou et al., 2023). In this study, users in the AED and AliCCP datasets with the number of interactive items ≤ 2 and registration time < 7 days were defined as cold-start users. To comprehensively evaluate the overall performance of the proposed model in CSSs, multiple groups of evaluation metrics covering the three dimensions of accuracy, novelty, and diversity were selected. Among them, Hit Rate at 5 (HR@5) measures how often the correct item appears among the top five recommendations, and Recall@5 measures how many of the items a user truly prefers are captured in the top 5 recommendations. They focus on assessing the basic accuracy of the recommendation list in capturing user's true interests. Normalized Discounted Cumulative Gain (NDCG@5) measured the ranking quality of the recommendations, giving higher scores to systems that place items users are most likely to prefer at the top of the list. F1@5 provided a single-dimensional comprehensive accuracy assessment, while Novelty@5 quantified the extent to which non-popular items are discovered in the recommendation list. Its calculation could be expressed as $Novelty@5=\sum_{i\in Top5}\frac{1}{popularity}$, where a higher value indicates that the recommendations are better at guiding users to explore long-tail items. Coverage@5 evaluated the range of items within the entire product catalog that the recommendation system can uncover, reflecting the system's exploration capability. Popularity@5 recorded the average popularity of the recommended items, serving as an inverse reference indicator for novelty. The entire experiment was conducted on a Windows 11 system with an Intel Core i7-10875H @ 2.30 GHz processor, and the software was written in Python 3.6.13 and PyTorch 1.2.0 + torchvision 0.4.0. Table 2 shows other parameter configurations.

Table 2. Experimental settings

Name	Setting
Memory	32 GB DDR4-3200
Storage	1 TB NVMe PCIe 3.0 SSD
Optimizer	Adam optimizer
Embedding dimension d	512
Batch size	128
Learning rate	0.001 (outer) / 0.01 (MAML inner loop)
Early-stopping patience	20 epochs
Loss function	Categorical crossentropy loss
λ1, λ2	0.1, 0.05

In Table 2, the population social regularization weights λ1 and λ2 were set to 0.1 and 0.05, which were used to stabilize the cold-start user vector. If there was no improvement after 20 rounds of validation, an early stop would be performed. The values of the two weights were determined as optimal via Bayesian optimization. To further investigate their sensitivity, a parameter analysis experiment was conducted, with the results summarized in Table 3.

In Table 3, when λ2 is fixed at 0.05 while λ1 is adjusted, the model maintained optimal and stable performance within the λ1 range of [0.05, 0.2]. Similarly, when λ1 was fixed at 0.1 with λ2 being adjusted, the model also demonstrated its best performance within the λ2 range of [0.05, 0.1]. These results indicated that the selected optimal parameter combination (λ1=0.1, λ2=0.05) lay within a flat region of the performance landscape, where the model was insensitive to minor parameter perturbations within this range.

This study first conducted ablation experiments with the following settings: Model A=CF, Model B=MAML+CF, Model C=MCF, and Model D=MAML+MCF. To more systematically decompose the contributions of individual modules, the demographic regularization was removed from Model D, yielding Model E. Additionally, the regularization was removed from Model B to investigate whether it remains important in the absence of MCE. The validation results for the four models on the AED are shown in Fig. 7.

In Fig.7(a), as the proportion of the dataset increased, the HR@5 of all models gradually increased, indicating that multi-data can improve the hit rate of model recommendations. However, Model D demonstrated the best performance. When the dataset ratio reached 100%, its hit rate increased by 62.99%, 35.29%, 32.65%, 21.21%, and 48.15% compared to Models A, B, C, E, and F, respectively. In Fig. 7(b), Model D continued to outperform all other models, with average increases of 11.67% in Coverage@5 and 15.60% in Novelty@5 compared to the other models. Adding MAML and MCE network structures could effectively improve the hit rate of model recommendations, and Model D had the highest Novelty@5 value, indicating that its recommendation results have the strongest novelty and can bring users a richer user experience. Furthermore, it also demonstrated that the absence of regularization compromised stability, while MCE could provide rich multi-modal information. To verify the generalization of the proposed model, this study further analyzed the recommendation performance before and after improvement on the AliCCP dataset. The evaluation indicators are HR@5, NDCG@5, Recall@5, F1@5, Popularity@5, Novelty@5, Coverage@5, Precision@5, and Diversity@5. The verification results are shown in Fig.8.

Table 3. Regularization parameter sensitivity analysis results

With λ2 fixed at 0.05, λ1 is adjusted.			With λ1 fixed at 0.05, λ2 is adjusted.		
λ1	HR@5	NDCG@5	λ2	HR@5	NDCG@5
0.01	0.175	0.158	0.01	0.185	0.168
0.05	0.195	0.178	0.05	0.200	0.182
0.1	0.200	0.182	0.1	0.198	0.180
0.2	0.198	0.179	0.2	0.192	0.175
0.5	0.180	0.165	0.5	0.178	0.162
1.0	0.172	0.156	1.0	0.170	0.155

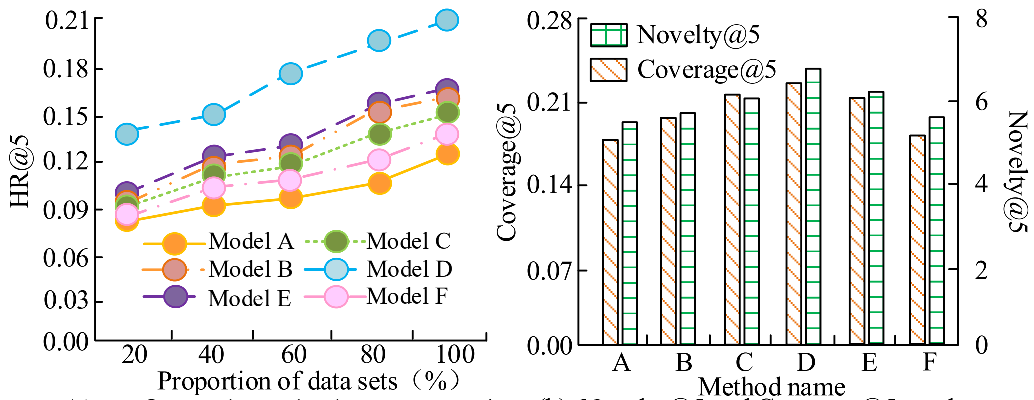
In Fig.8(a), Model D performed better than Model A in NDCG@5, Recall@5, and F1@5, with increases of 52.57%, 55.05%, and 53.76% compared to Model A. This indicated that the improved model could discover more items that users are interested in in the recommendation results and also provided users with a more appropriate ranking of recommendation results, thus meeting their needs. In Fig.8(b), Model D demonstrated superior performance over Model A in both Novelty@5 and HR@5, achieving improvements of 12.72% and 39.43%, respectively. However, its performance in Popularity@5 was inferior to Model A, showing a decrease of 12.05%. This observed decline in popularity bias was a concomitant phenomenon of successfully enhancing recommendation diversity and novelty, indicating the model's effective reduction in reliance on popular items while shifting focus toward discovering long-tail items to satisfy user's potential interests better. Commercially, this popularity desensitization mechanism helped break the information cocoon, stimulate users' willingness to explore, and enhance the ecological health of the platform, thus having a positive impact on long-term user retention and lifetime value. Excessive deviation from popular items could impact short-term conversion rates. Therefore, in practical deployment, introducing an adjustable popularity smoothing factor into the loss function could establish a dynamic balance between exploration and exploitation, ultimately aiming to maximize long-term GMV. In Fig. 8(c), Model D performed better than Model A in Diversity@5, Recall@5, Diversity@5, and Coverage@5, with increases of 16.17%, 32.39%, and 33.15% compared to Model A. The MAML+MCF model could more accurately understand users' interests, provide users with recommendation results that better meet their needs, and perform better in ranking and comprehensiveness of recommendations. To systematically evaluate the statistical significance of performance differences between models, this study employed paired t-tests for all key comparisons, with the results summarized in Table 4.

In Table 4, the results demonstrated that Model D achieved statistically significant superiority across all key evaluation metrics (p<0.05). Specifically, its improvement in HR@5 over the baseline Model A was highly statistically significant (p<0.001), confirming the effectiveness of the complete model architecture. Furthermore, Model D maintained significant advantages over the ablated models (Models E, B, and C; p<0.05), which systematically validated the independent contributions of the MAML framework, MCE module, and regularization mechanism to the overall performance enhancement.

Table 4. Paired t-test results of performance differences between model D and comparative models

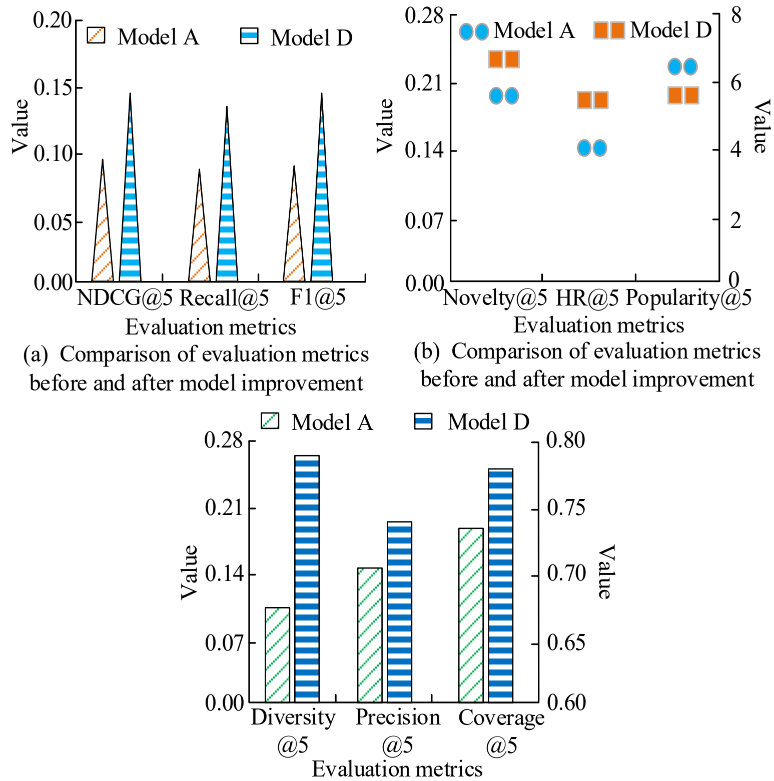
Model comparison	HR@5 *p*-value	NDCG@5 *p*-value	Significance grouping
Model D vs Model A	< 0.001	< 0.001	***
Model D vs Model B	0.003	0.005	**
Model D vs Model C	0.007	0.009	**
Model D vs Model E	0.018	0.022	*
Model D vs Model F	0.002	0.004	**

Note: ***p < 0.001, **p < 0.01, *p < 0.05



(a) HR@5 results as the dataset proportion varies (b) Novelty@5 and Coverage@5 results on the Amazon Electronics dataset

Fig. 7. Ablation study results on the AED



(a) Comparison of evaluation metrics before and after model improvement (b) Comparison of evaluation metrics before and after model improvement (c) Comparison of evaluation metrics before and after model improvement

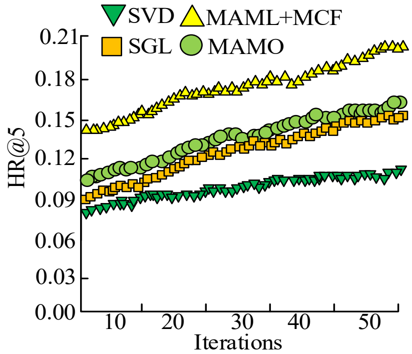
Fig. 8. Performance comparison before and after model improvement on the AliCCP dataset

3.2. Performance Comparison of Different Personalized Recommendation Models in CSS

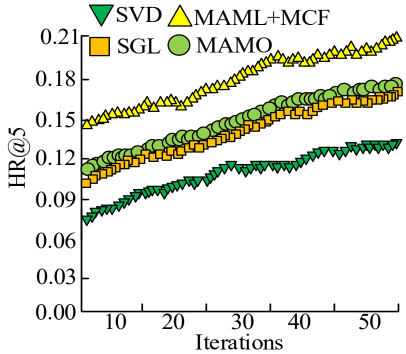
To further demonstrate the effectiveness of the MAML+MCF model in CSS, this study introduced other recommendation models for validation on the AED and AlicCP datasets. The study primarily compared three mainstream technical paradigms for addressing the cold-start problem: Singular Value Decomposition (SVD) represents traditional matrix factorization methods, modeling user-item interactions through latent vectors; Self-Supervised Graph Learning (SGL) represents modern graph neural network methods, enhancing topological representations of user-item graphs through contrastive learning. Meta-Learning with Adaptive Memory Optimization (MAMO) is a meta-learning approach specifically designed for CSSs. Although more recent models, such as LightGCN and SimGCL, existed, they shared fundamental conceptual similarities with SGL within the graph self-supervised learning paradigm. As a typical representative of graph self-supervised recommendation models, SGL's comparative results could provide valuable insights into the relative performance of the proposed method. Therefore, the current baseline model configuration constituted a comprehensive and reasonable frame of reference. The comparison of HR@5 validation of different models in AED and

AlicCP is shown in Fig. 9.

In Fig. 9(a), the HR@5 performance of different models on AED showed an increasing trend with the increase of iteration rounds. When the iteration round was 50, the HR@5 value of MAML+MCF increased by 9.5%, 6%, and 5.1% compared to SVD, SGL, and MAMO models. In the AlicCP dataset in Fig. 9(b), the HR@5 of MAML+MCF was consistently ahead of other models, with an average increase of 26.02% compared to the HR@5 of the three models when the iteration round was 50. Combining Figs. 9(a) and (b), on both datasets, the HR@5 value of the MAML+MCF model was higher than the other four models in each iteration round, indicating that this model has strong advantages in recommendation system tasks. This might be because it combines technologies such as MAML and MCE, which can better capture the complex relationship between users and items and improve recommendation effectiveness. The validation results of Novelty@5 and NDCG@5 for different models in AED and AlicCP are shown in Fig. 10.



(a) Changes HR@5 of different teaching effectiveness evaluation methods in the Amazon Electronics dataset



(b) Changes HR@5 of different teaching effectiveness evaluation methods in the AliCCP dataset

Fig. 9. HR@5 validation results of different models on AED and AliCCP dataset

In Fig. 10(a), the Novelty@5 values of all models showed an increasing trend with the proportion of the AED. When the data volume increased to 100%, MAML+MCF models increased by 25.33%, 14.56%, and 10.87% compared to SVD, SGL, and MAMO models. In Fig. 10(b), MAML+MCF performed better in AliCCP than other models. When the data volume increased to 100%, it increased by an average of 15.23% compared to the Novelty@5 values of the three models. In Fig. 10(c), when the data volume of AED increased to 100%, the NDGG@5 of the proposed model improved by an average of 51.23% compared to other models. In the AlicCP of Fig. 10(d), when the data volume increased to 100%, all models reached their maximum values, and the NDGG@5 value of MAML+MCF was consistently higher than other models, increasing by 56.34% compared to the average. The novelty exploration ability of MAML+MCF was the strongest, and it could still maintain its optimal position in recommendation scenarios in different fields, indicating that it has a certain degree of cross domain stability.

To comprehensively and deeply evaluate the performance of the model, this study also introduced evaluation metrics with a cutoff length of @10, including HR@10, NDCG@10, Recall@10, F1@10, Popularity@10, and Novelty@10, for comparison. Table 5 shows the validation results from the AED and AlicCP datasets.

In Table 5, in AED, MAML+MCF had the highest proportion of hitting the target item among the top 10 recommendation results, with an average increase of 25.97% compared to the other three models. It performed the best in Novelty@10, with growth rates of 19.89%, 10.29%, and 7.48% compared to the SVD, SGL, and MAMO models. Its performance was the worst in Popularity@10, and the high popularity in the model's recommendation results indicated that its recommendation relies on popular items. However, for personalized recommendations, the recommendation results should have diversity. The performance trend of different models in AlicCP was similar to that of AED. The performance of the research model in HR@10, NDCG@10, and Novelty@10 increased by an average of 35.13%, 18.78%, and 13.65% compared to other models. The model showed an average increase of 38.89% and 33.01% in Recall@10 and F@110 compared to the comparison model in two datasets, indicating that MAML+MCF has a significant advantage in personalized recommendation in CSS.

Given that real-time ECRS must handle high-concurrency requests and deliver millisecond-level responses, the paper conducted inference latency tests, primarily measuring the single-user recommendation time consumption for each model. The specific results are presented in Table 6.

In Table 6, the proposed MAML+MCF model demonstrated superior inference efficiency on both datasets, with significantly lower latency compared to both SGL and MAMO. This advantage stemmed from the MAML framework's ability to rapidly adapt to users through a single forward propagation after pre-training, thereby avoiding the computationally intensive graph topological computations in SGL and the complex memory read-write operations in MAMO. Additionally, MFFs were pre-encoded during the offline phase, requiring only efficient vector similarity calculations during online inference. However, its inference efficiency was lower than that of SVD, which achieved minimal latency due to its lightweight structure and simple computations. Overall, the proposed model achieved the best comprehensive performance when considering both accuracy and efficiency.

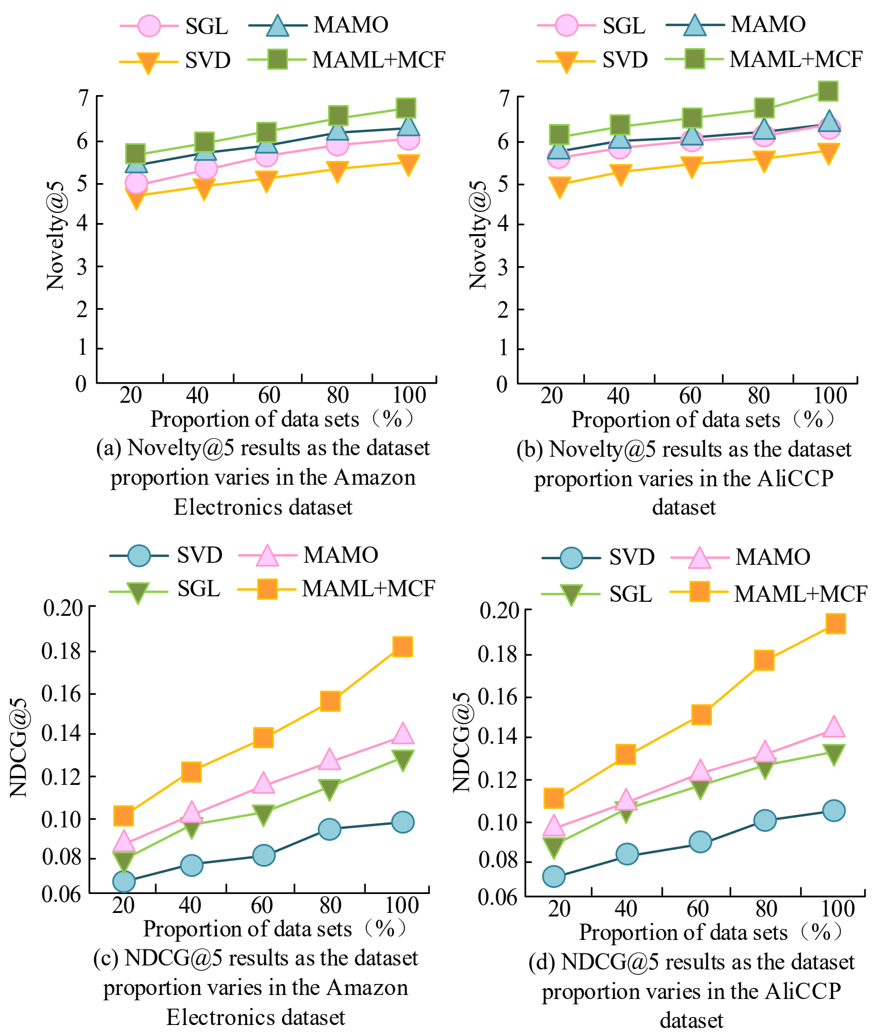


Fig. 10. Novelty@5 and NDCG@5 validation results of different models on AED and AliCCP datasets

Table 5. Top-K recommendation evaluation metrics of different models on AED and AliCCP datasets

Data sets name	Method name	HR@10	NDCG@10	Recall@10	F1@10	Popularity@10	Novelty@10
AED	SVD	0.195	0.098	0.122	0.108	0.312	5.63
	SGL	0.243	0.128	0.156	0.140	0.258	6.12
	MAMO	0.255	0.137	0.165	0.150	0.241	6.28
	MAML+MCF	0.291	0.183	0.186	0.176	0.219	6.75
AliCCP	SVD	0.207	0.105	0.129	0.115	0.298	5.88
	SGL	0.251	0.132	0.163	0.146	0.256	6.05
	MAMO	0.264	0.141	0.172	0.155	0.241	6.31
	MAML+MCF	0.325	0.197	0.234	0.185	0.220	6.91

3.3. Privacy-Performance Trade-Off Analysis in Federated Learning Deployment

To evaluate the comprehensive performance in practical deployment, this study conducted a quantitative analysis of efficiency and performance within the federated learning framework from the perspective of privacy protection. Tests demonstrated that after introducing homomorphic encryption, the MAML+MCF model maintained a single-user recommendation latency of 18.5±1.8ms, meeting real-time service requirements. To further investigate the impact of privacy protection intensity on the system, different security levels were simulated by adjusting the key length of homomorphic encryption, with the results shown in Table 7.

Table 6. Inference latency comparison of different models on two datasets

Model name	AED (ms)	AliCCP (ms)
SVD	10.2±0.9	8.5±0.7
SGL	38.5±3.1	35.8±2.8
MAMO	25.7±2.2	22.1±1.9
MAML+MCF	16.3±1.4	14.1±1.2

Table 7. Privacy-performance trade-off under different encryption strengths

Encryption strength	Key length	HR@5	Single round global training time
Baseline (plaintext)	/	0.200	1.0x
Low	1024-bit	0.199	3.8x
Middle	2048-bit	0.197	9.5x
High	4096-bit	0.194	29.7x

The results in Table 7 demonstrated that even under strong encryption settings, HR@5 only decreased by 3.0%, confirming the model's ability to maintain excellent performance while ensuring privacy protection. However, as the encryption intensity increased, the training time grew significantly, revealing a clear trade-off between privacy preservation and computational efficiency. This finding provided crucial insights for selecting appropriate encryption schemes based on specific privacy requirements in practical applications.

4. Conclusion

In response to the dilemma of insufficient accuracy of traditional recommendations in CSS in e-commerce, this study innovatively proposed a personalized recommendation model that utilized MAML to improve the CF algorithm and combined it with MCE. Ablation experiments were conducted in AED and AlicCP. In the AED, MAML+MCF increased by 62.99% compared to the HR@5 of the unimproved model. In the AlicCP dataset, it increased by 52.57%, 55.05%, 53.76%, 12.72%, and 39.43% compared to the unimproved models NDCG@5, Recall@5, F1@5, Novelty@5, and HR@5. In comparative experiments, the HR@5 of the model in AED increased by 9.5%, 6%, and 5.1% compared to SVD, SGL, and MAMO, while Novelty@5 increased by 25.33%, 14.56%, and 10.87%. In the AliCCP dataset, the proposed MAML+MCF model achieved average improvements of 26.02% in HR@5 and 15.23% in Novelty@5 over the other three models. Research has shown that MAML+MCF can effectively achieve accurate recommendations in CSS. However, the model still faces challenges, including relatively high computational complexity and limited dynamic interest capture capabilities. Future research will focus on optimizing computational efficiency and enhancing temporal modeling capabilities, while incorporating explainability techniques such as SHAP to quantify the decision contributions of user attributes and MFFs, thereby improving model transparency. In addition, efforts will be made to promote A/B testing cooperation with industry partners at the application level and verify the actual improvement effect of the model on core business indicators (such as "next-day retention rate") through online testing. Ultimately, the purpose is to promote the effective transformation of academic innovation into tangible business value.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

Declaration of Artificial Intelligence (AI) Tools

The author confirms that no AI tools were used in the preparation of this manuscript.

References

Ahamed, J., Noori, N., and Ahmed, M. (2024). Matrix factorization and cosine similarity based recommendation system for cold-start problem in e-commerce industries. *International Journal of Computing and Digital System,* 15(1), 775-787. DOI:10.12785/ijcds/150156s

Anwar, T., Uma, V., Hussain, I., and Pantula, M. (2022). Collaborative filtering and kNN based recommendation to overcome cold-start and sparsity issues: A comparative analysis. *Multimedia Tools and Applications,* 81(25), 35693-35711. DOI:10.1007/s11042-021-11883-z

Feng, D., He, X., and Peng, Y. (2023) MKVSE: Multimodal knowledge enhanced visual-semantic embedding for image-text retrieval. *ACM Transactions on Multimedia Computing Communications and Applications,* 19(5), 1-21. DOI:10.1145/3580501

Gheisari, M., Hamidpour, H., Liu, Y., Saedi, P., Raza, A., Jalili, A., Rokhsati, H., and Amin, R. (2023). Data Mining Techniques for Web Mining: A Survey. *Artificial Intelligence and Applications,* 2023, 1(1), 3-10. DOI:10.47852/bonviewAIA2202290

Hamza, A., Majeed, K.,B., Rashad, M., and Jaffar, A. (2024). An Integrated Approach for Amazon Electronic Products Reviews by Using Sentiment Analysis. *Bulletin of Business and Economics,* 13(2), 142-153. DOI:10.61506/01.00309

Kannout, E., Grodzki, M., and Grzegorowski, M. (2023). Towards addressing item cold-start problem in collaborative filtering by embedding agglomerative clustering and FP-growth into the recommendation system. *Computer Science and Information Systems,* 20(4), 1343-1366. DOI:10.2298/CSIS221116052K

Liao, M., and Sundar, SS. (2022). When e-commerce personalization systems show and tell: Investigating the relative persuasive appeal of content-based versus collaborative filtering. *Journal of Advertising,* 51(2), 256-267. DOI:10.1080/00913367.2021.1887013

Liu, K., Xue, F., Guo, D., Wu, L., Li, S., and Hong, R. (2023). MEGCF: Multimodal entity graph collaborative filtering for personalized recommendation. *ACM Transactions on Information Systems,* 41(2), 1-27. DOI:10.1145/3544106

Ma, D., Wang, Y., Ma, J., and Jin, Q. (2023). SGNR: A social graph neural network based interactive recommendation scheme for e-commerce. *Tsinghua Science and Technology,* 28(4), 786-798. DOI:10.26599/TST.2022.9010050

Mo, Y., Li, L., Huang, B., and Li, X. (2023). Few-shot RUL estimation based on model-agnostic meta-learning. *Journal of Intelligent Manufacturing,* 34(5), 2359-2372. DOI:10.1007/s10845-022-01929-w

Nahta, R., Meena, Y., K., Gopalani, D., and Chauhan, G., S. (2021). Embedding metadata using deep collaborative filtering to address the cold-start problem for the rating prediction task. *Multimedia Tools and Applications,* 80(12), 18553-18581. DOI:10.1007/s11042-021-10529-4

Panda, D., K., and Ray, S. (2022). Approaches and algorithms to mitigate cold-start problems in recommender systems: a systematic literature review. *Journal of Intelligent Information Systems,* 59(2), 341-366. DOI:10.1007/s10844-022-00698-5

Papadakis, H., Papagrigoriou, A., Panagiotakis, C., Kosmas, E., and Fragopoulou, P. (2022). Collaborative filtering recommender systems taxonomy. *Knowledge and Information Systems,* 64(1), 35-74. DOI:10.1007/s10115-021-01628-7

Raji, M., A., Olodo, H., B., Oke, T., T., Addy, W., A., Ofodile, O., C., and Oyewole, A., T. (2024). E-commerce and consumer behavior: A review of AI-powered personalization and market trends. *GSC Advanced Research and Reviews,* 18(3), 66-77. DOI:10.30574/gscarr

Sharma, S., Rana, V., and Malhotra, M. (2022). Automatic recommendation system based on hybrid filtering algorithm. *Education and Information Technologies,* 27(2), 1523-1538. DOI:10.1007/s10639-021-10643-8

Wang, F., Tian, S., Yu, L., Liu, J., Wang, J., Li, K., and Wang, Y. (2023). TEDT: transformer-based encoding–decoding translation network for multimodal sentiment analysis. *Cognitive Computation*, 15(1), 289-303. DOI:10.1007/s12559-022-10073-9

Wang, J., and Zhang, Y. (2021). Using cloud computing platform of 6G IoT in e-commerce personalized recommendation. *International Journal of System Assurance Engineering and Management,* 12(4), 654-666. DOI:10.1007/s13198-021-01059-1

Wu, K., and Chi, K. (2023). Enhanced e-commerce customer engagement: A comprehensive three-tiered recommendation system. *Journal of Knowledge Learning and Science Technology,* 2(3), 348-359. DOI:10.60087/jklst.vol2.n2.p359

Xu, Y., Zhu, L., Cheng, Z., Li, J., Zhang, Z., and Zhang, H. (2021). Multi-modal discrete collaborative filtering for efficient cold-start recommendation. *IEEE Transactions on Knowledge and Data Engineering,* 35(1), 741-755. DOI:10.1109/TKDE.2021.3079581

Zhou, J., Yu, Q., Luo, C., and Zhang, J. (2023). Feature decomposition for reducing negative transfer: a novel multi-task learning method for recommender system. *Proceedings of the AAAI Conference on Artificial Intelligence,* 37(13), 16390-16391. DOI:10.1609/aaai.v37i13.27055

Wenwei Chen, currently serves as an Associate Professor and the Program Leader of Cross-Border E-Commerce at Hangzhou Polytechnic. He teaches courses in E-Commerce, Cross-Border E-Commerce Visual Design, and Fundamentals of Artificial Intelligence Application. His research interests include e-commerce algorithms, e-commerce trust, and image processing.