Journal of Network Intelligence ©2025 ISSN 2414-8105
Taiwan Ubiquitous Information Volume **10**, Number **2**, May **2025**

Recommended Online Learning Resources Based on Spatial Downscaling and Mean Shift Clustering

Tong Zhang
The Experimental College
The Open University of China, Beijing 100000, P. R. China
ouczhangt@126.com

Xin-Ling Dong*
Hebei Oriental University, Langfang 065001, P. R. China
dongchengjinxu@126.com

Cheng-Bo Jin
Faculty of Social Sciences and Humanities
Thammasat University, Bangkok 10330, Thailand
nb4830@163.com

*Corresponding author: Xin-Ling Dong
Received March 5, 2024, revised August 6, 2024, accepted December 27, 2024.

Abstract. For the purpose of addressing the issues of current recommendation algorithms in which the preference matrix is sparse and learners with similar preferences deviate from the center point resulting in the recommendation accuracy to be improved, this article suggests an online learning resource recommendation algorithm relied on spatial dimensionality reduction and Mean-Shift Clustering (MSC). Firstly, to deal with the issue of inaccurate clustering results of traditional MSC algorithm when addressing unevenly distributed datasets, the original mean-drift algorithm is optimized with semi-supervised learning, and the accuracy of the clustering results is enhanced by a priori information. Next, the high-dimensional sparse learner preference matrix is spatially downscaled to extract preference features, and learners are clustered relied on the downscaled data adopting an optimized mean drift algorithm. Then, the similarity between the target user and other learners in the cluster is calculated. Finally, the nearest-neighbor similarity weighting is used to derive the target user's rating for the unassessed items, and a recommendation list is generated on the ground of the ratings in descending order. By conducting comparison experiments on the MoocData dataset, from Recall, NDCG and HitRate metrics with the comparison methods, all achieved better performance and effectively enhanced the accuracy of recommendations.
Keywords: Spatial dimensionality reduction; Mean shift clustering; Learning resource recommendation; Semi-supervised learning; Nearest neighbor similarity

---

1. **Introduction.** As information technology rapidly developing, people's demand for education continues to improve, prompting the diversification of modern education methods. Among them, online education, which is mainly based on the Internet platform, has been recognized by the public and widely popularized [1,2]. The integration of big data, artificial intelligence and other emerging technologies "Internet and education" greatly reduces the cost of each person involved in learning. To provide students with suitable learning resources quickly and efficiently, e-learning platforms have taken some measures, including the use of keyword search [3], high-rating learning resources recommendation [4],

the latest learning resources recommendation [5] and other ways. Although the above methods can enable students to obtain some high-quality or the latest learning resources, these resources are not the most suitable for students and still cannot effectively solve the problem of students' access to effective learning resources [6]. Because each student's study habits, hobbies, learning ability and other aspects are different, even high-quality learning resources may not fully meet the needs of each student.

1.1. **Related work.** Guo and Cheng [7] suggested a learning diagnostic recommendation method that incorporates learners' interests to find the most appropriate courses based on their learning orientation. Chen et al. [8] mined learners' behavioral patterns to extract their preferences and achieve adaptive learning resources. Huang et al. [9] designed a recommendation technique for delivering e-learning content, integrating compatible legacy algorithms on learning objects, but the recommendation results were not accurate enough. Shanshan et al. [10] developed an online learning resource recommendation system based on learners' historical learning behaviors, but suffered from the sparse matrix problem. Shu et al. [11] applied a content-based recommendation method to an online learning resource platform and pushed the most suitable courses to the learners. Kushwaha et al. [12] adopted text mining tools to extract mathematical concepts and recommend large-scale e-textbook content.

Large-scale and high-dimensional resource volume poses a challenge to online learning resource recommendation. Nilashi et al. [13] suggested a recommendation framework integrating collaborative filtering and dimensionality reduction, adopting ontologies to describe semantic associations between knowledge concepts, but ignoring the high-dimensionality of the feature space. Ozkan and Koseler [14] build a user-resource recommendation model by considering the multidimensional feature differences between users and resources based on users' learning preferences. Jia and Zhao [15] combine Principal Component Analysis (PCA) with the k-nearest neighbor method to predict the learning resources as the dependent variable, but the computation consumes a lot of resources. Al-Sabaawi et al. [16] integrate multiple scoring modules with PCA to ensure recommendation accuracy. Zhou et al. [17] computed the degree of matching between learners and learning resources based on learners' needs and resource quality information. To enhance recommendation efficiency, Chen et al. [18] adopted fuzzy C-mean clustering; Li [19] used grey correlation models; Zahra et al. [20] employed K-means clustering; and Kausar et al. [21] adopted MSC with nearest-neighbor similarity, each addressing aspects of clustering or similarity but with various limitations.

1.2. **Contribution.** In this article, for the current recommendation algorithms existing sparse matrix and similar preferences of the user clustering effect is poor, this article suggests a spatial dimensionality reduction and MSC relied on online learning resources recommendation algorithm.

(1) To deal with the issue that the traditional MSC algorithm relies on the subjective selection of the bandwidth parameter, which leads to inaccurate clustering results, all the pairwise constraints are linearly transformed in the high-dimensional kernel space, and the accuracy of the clustering results is enhanced by the a priori information. Next, the high-dimensional sparse learner preference matrix is spatially downscaled to extract the preference features, and the learners are clustered based on the downscaled data using an optimized mean drift algorithm.

(2) Then the similarity between the target learner and other learners in the cluster is calculated. Finally, the nearest-neighbor similarity weighting is used to calculate the target user's rating of the unassessed items, and the recommendation list is generated in terms of the rating value from high to low.

2. Theoretical analysis.

2.1. Spatial downscaling techniques. The spatial downscaling technique [21] first requires an eigendecomposition of a matrix about the gradient of the model's output response, and then identifies the key directions of change affecting the output response based on the decomposed eigenvalues.

Assuming a model $y = f(x)$, $x \in X \subseteq R^n$, where x is a vector of all uncertainties and each component of x is a spatial dimension, the gradient of the output response y is as bellow.

$$\nabla_x f(x) = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right]^T \tag{1}$$

Define a $m \times m$ matrix A: $A = \int \nabla_x f (\nabla_x f)^T \rho\, dx$, ρ as the probability density function. Each element of A is the mean of the product of the partial derivatives of the output response f. Assuming A_{ij} is the (i, j)-th element of A, A_{ij} can be expressed as bellow.

$$\int \left(\tfrac{\partial f}{\partial x_i}\right)\left(\tfrac{\partial f}{\partial x_j}\right) \rho\, dx \tag{2}$$

The matrix A is allowed to be eigen-decomposed with $A = V \Lambda V^T$, where Λ is a diagonal distance matrix with non-negative eigenvalues arranged in descending order, and the j-th eigenvalue λ_j satisfies

$$\lambda_j = V_j^T A V_j = V_j^T \int (\nabla_x f)(\nabla_x f)^T \rho\, dx\; V_j \tag{3}$$

According to the descending order of the n eigenvalues, when there is a large gap between the m-th eigenvalue and the $m + 1$-th eigenvalue, the eigenvalues can be partitioned into two groups, and the partitioned eigenvalues and eigenvectors can be expressed as bellow.

$$\Lambda = \begin{bmatrix} \Lambda_1 & \\ & \Lambda_2 \end{bmatrix}, \quad V = [V_1\ V_2] \tag{4}$$

where $\Lambda_1 = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_m)$, $m < n$, V_1 are the feature vectors corresponding to Λ_1.

2.2. Mean shift clustering. MSC [23] is insensitive to the clustering parameter settings and the clustering results do not include no outliers, which is good compared to the clustering of algorithms such as k-means, fuzzy C-mean clustering, etc., as indicated in Figure 1. Assuming that the dataset $\{x_j, j = 1, 2, \ldots, m\}$ obeys the probability density function $f(x)$, the region of high density is considered to be locally maximal, and each point in the space is searched for peaks until convergence, and convergence to the same point is counted as a class. For $f(x)$, given m sample points R^c in a c-dimensional space $\{x_j\}, j = 1, 2, \ldots, m$, use the kernel function $K(x)$ to perform multivariate kernel density estimation, and obtain the kernel function estimate of $f(x)$ as bellow.

$$\hat{f}(x) = \frac{1}{m\, g^c} \sum_{j=1}^{m} K\Big(\frac{x - x_j}{g}\Big) \tag{5}$$

where g denotes the window length, and the kernel function $K(x)$ satisfies: $K(x) = d_{k,c}(\|x\|^2)$, where $d_{k,c}$ is a normalization constant that guarantees that $K(x)$ integrates to

1. Also define the negative derivative function $h(x)$ of $K(x)$, whose corresponding kernel function is $H(x) = d_{h,c}\, h(\|x\|^2)$.

Then the mean drift vector m_g of the probability density function $f(x)$ is as bellow.

$$m_g = \frac{\sum_{j=1}^{m} x_j\, h\left(\frac{x - x_j}{g}\right)^2}{\sum_{j=1}^{m} h\left(\frac{x - x_j}{g}\right)^2} \tag{6}$$

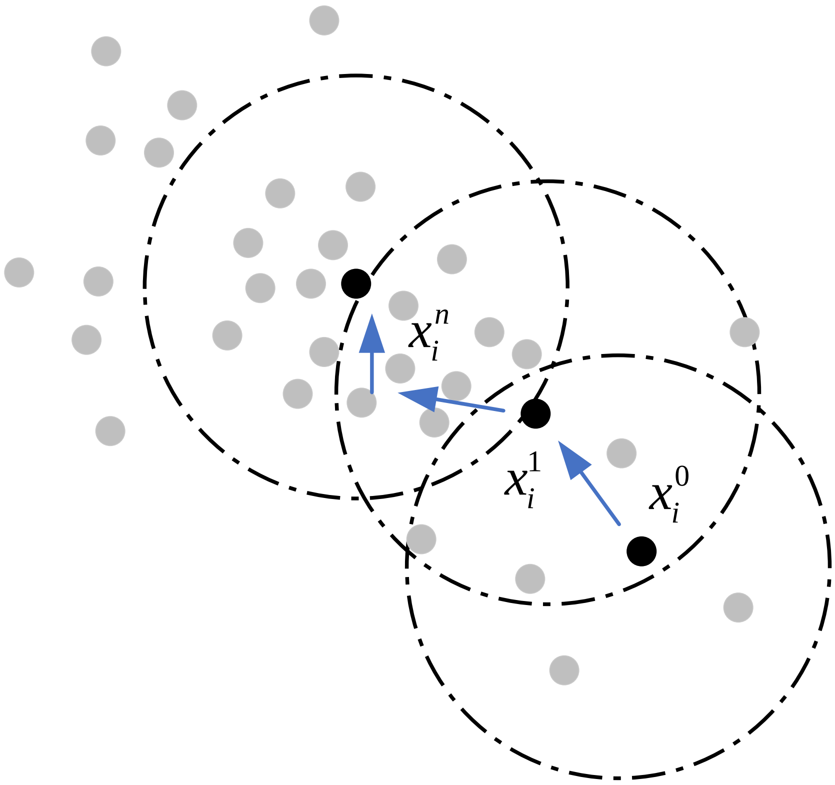

Figure 1. Mean shift clustering

3. **Optimization of mean shift clustering algorithm.** Intending to the issues that the clustering effect of MSC algorithm depends on the subjective selection of bandwidth parameter and the inaccuracy of the clustering results when dealing with unevenly distributed datasets, this article optimizes the original MSC algorithm by adopting semi-supervised learning [23], where all the pairwise constraints in the high-dimensional kernel space are linearly transformed, and then all the feature samples are projected to the zero space of constraint vectors, so that the ones whose distances between two samples conform to the distance target parameter are clustered into one class, and the accuracy of the clustering results is enhanced by the a-priori information.

Define (i_1, i_2) to be a pairwise constraint, and denote that i_1, i_2 are forced to be pairs, either must-link pairwise constraints or cannot-link pairwise constraints, with $\varphi(x_{i_1}) = \Phi e_{i_1}$, $\varphi(x_{i_2}) = \Phi e_{i_2}$. Given m_b pairwise constraints M_b, the d_φ-dimensional constraint vector can be denoted as $a_i = \Phi\big(e_{i_1} - e_{i_2}\big) = \Phi y_i$, where the m-dimensional vector y_i denotes the indicator vector of the i-th pairwise constraint, and then the constraint matrix $A = \Phi Y$ contains m_b constraint vectors, where $Y = [y_1, y_2, \ldots, y_{m_b}]$ is the indicator matrix of size $m \times m_b$.

Then the linear transformation matrix is $S = I_{d_\varphi} - t\,(aa^T)$, where t is a scaling factor for a. When $t = 1/(a^T a)$, the transformation becomes a zero-space projection from the feature space to the constraint vector a. When $0 < t < 2/(a^T a)$, the transformation decreases the distance between pairs of samples, and when $t < 0$ or $t > 2/(a^T a)$, the transformation increases the distance between pairs of samples.

Let the distance between pairs of samples be $d > 0$, then we have

$$S\Phi y_F^2 = y^T \hat{K} y = d \tag{7}$$

where the transformation kernel matrix is $\hat{K} = \Phi^T S^T S \Phi = \Phi^T S^2 \Phi$. Substituting $S = I_{d_\varphi} - t\,(aa^T)$ into Equation (7) yields

$$\hat{K} = k - 2t\,K\,y\,y^T K + t^2\,(y^T K\,y)\,K\,y\,y^T K \tag{8}$$

$T = 1/q\big(1 + \sqrt{d}/q\big)$, $\quad q = y^T K y > 0$. Finally, the expression for the transformation kernel matrix is achieved as bellow.

$$\hat{K} = K + \alpha\,K y y^T K,\ \alpha = \big(d/q^2 - 1/q\big) \tag{9}$$

Since the transformed kernel matrix needs to update the α value at each iteration, and the algorithm is sensitive to mislabeling if a linear transformation is used, this paper introduces the principle of Bregman scattering [24] to realize the kernel function updating, and the Bregman scattering is defined as bellow.

$$D_{ld}(X, Y) = \sum_{i,j \le r} \big(w_i^T v_j\big)^2 \big(\eta_i/\vartheta_j - \log \eta_j/\vartheta_j - 1\big) \tag{10}$$

Given n must-link pairwise constraint sets M and m cannot-link pairwise constraint sets B, there are $n + b = m_b$. The objective distance of the must-link constraints is d_n and the objective distance of the cannot-link constraints is d_b. The final problem of updating the kernel matrix is transformed into a problem of minimizing the Bregman scatter, i.e., the objective function is $\min D_{ld}(\hat{K}, K)$.

For a given pairwise constraint $(i_1, i_2) \in N \cup B$, the iterative formula for updating the kernel matrix can be similarly obtained according to Equation (10) as bellow.

$$\hat{K}_{s+1} = \hat{K}_s + \alpha_s\,\hat{K}_s\,(e_{i_1} - e_{i_2})\,(e_{i_1} - e_{i_2})^T\,\hat{K}_s \tag{11}$$

Since the first parameter X in the Bregman scattering is required to be convex, updating the kernel matrix with scattering minimization ensures that the algorithm converges to a globally optimal solution.

4. **Recommended online learning resources based on spatial downscaling and mean shift clustering.**

4.1. **Feature extraction based on spatial dimensionality reduction.** Focusing on the issues of sparse matrix and poor user clustering that exist in current recommendation algorithms, leading to low recommendation accuracy, this article suggests an online learning resource recommendation algorithm based on spatial dimensionality reduction and MSC. The algorithm adopts spatial dimensionality reduction to retain the dimension that best represents the user's interest, to alleviate the issue of sparse preference matrix and then adopts MSC algorithm to cluster users on the low-dimensional vector space after dimensionality reduction, to reduce the search range of the nearest neighbors of the target user. The model framework is indicated in Figure 2.

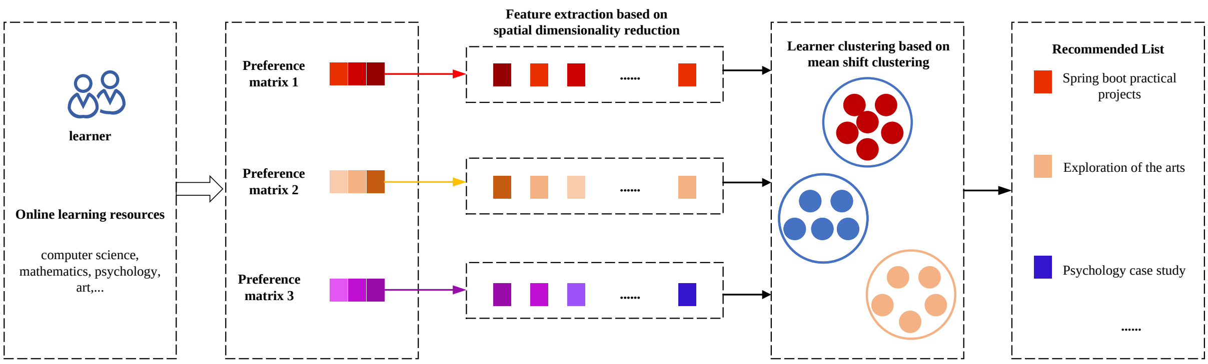

Figure 2. The entire framework of the suggested algorithm

Intending to the issue that the high dimensionality of the feature space of online learning platform generates redundant features, this paper adopts spatial downscaling and correlation embedding for feature extraction of learners and learning resources.

Firstly, relied on the learner's preference matrix $Y \in \{0,1\}^{m\times p}$, the preference correlation matrix $D \in \mathbb{R}^{p\times p}$ is obtained using the cosine similarity, where m denotes the number of instances in the feature space X and p denotes the number of learning resource categories. Each element d_{ij} represents the correlation between the i-th and j-th markers, and the calculation equation is indicated below.

$$d_{ij} = \sum_{g=1}^{m} Y_{gi} Y_{gj} \Big/ \sqrt{\sum_{g=1}^{m} Y_{gi}^2} \sqrt{\sum_{g=1}^{m} Y_{gj}^2} \tag{12}$$

where Y_{gi} represents the i-th preference value of the g-th instance in the preference matrix Y. $Y_{gi} = 1$ denotes that the g-th instance possesses preference i, and vice versa $Y_{gi} = 0$ denotes that it does not contain preference i.

Next, the preference correlation matrix D is embedded into the learning resource as the weighting matrix V as bellow.

$$\min_{V} \frac{1}{2}\|XV - Y\|_F^2 + \frac{\mu_1}{4}\|D - V^T V\|_F^2 + \mu_2\|V\|_{21} \tag{13}$$

where X is the learner's feature matrix; $V = [v_1, v_2, \ldots, v_p] \in \mathbb{R}^{c_1\times p}$; μ_1 and μ_2 are non-negative weight parameters [25]. l_{21}-paradigm regularization is selected to be applied to V to ensure its sparsity and to allow for the selection of discriminative features. Moreover, l_{21}-paradigm regularization has been shown to be robust to outliers and noise.

A multivariate linear regression model [26] is then adopted to build a probability density function $f(X, Q, V) = XQV$ from the dimensionally-approximated space XP to the original preference space Y, where $Q \in \mathbb{R}^{c\times c_1}$ is the projection matrix and $V \in \mathbb{R}^{c_1\times p}$ is the weight matrix. The equation is indicated bellow.

$$\min_{Q,V} \frac{1}{2}\|XQV - Y\|_F^2 \tag{14}$$

Finally a feature decomposition is performed on the learner's feature matrix X: $\hat{X}_l = X_i^T V \Sigma_m^{-1}$, where Σ is the feature dependency matrix. The learner's features are reduced from l to m features.

4.2. **Learner clustering based on mean shift clustering.** MSC of learners on the ground of the above spatially downscaled matrix X. A sampling set $X' = \{x_1', x_2', \ldots, x_m'\}$ is formed by randomly selecting m sampling samples (m is much smaller than n) from the input sample set of learning resources, and the kernel density function at point x is estimated as follows for sampling set X' in a given c-dimensional dimensional space $\mathbb{R}^c$ adopting Parzen window estimation method [27].

$$\hat{f}(x') = \frac{b_k}{m\, g^c} \sum_{j=1}^{m} \hat{K}\big(\|x' - x'_j\|/g\big) \tag{15}$$

where b_k is a constant, k is the kernel function, and g is the kernel width. Local density maxima are usually located in the gradient of the density function in the zero point, so the derivation of Equation (15) can be obtained from the probability density function $\hat{f}(x')$ of the local density maxima of the analytical formula $\nabla\hat{f}$ is estimated as bellow.

$$\nabla\hat{f}^2 = \frac{2d_k}{m\, g^{c+2}} \sum_{i=1}^{n} (x' - x'_j) \cdot k'\big(\|x' - x'_j\|^2/g\big) \tag{16}$$

Then the mean drift vector is indicated in Equation (17).

$$M_g(x') = \frac{\sum_{j=1}^{m} x_j\, g\big(\|(x' - x'_j)/g\|^2\big)}{\sum_{i=1}^{n} g\big(\|(x' - x'_j)/g\|^2\big)} - x' \tag{17}$$

Defining the subspace as $H_a = \operatorname{span}\big(K(x'_1, \cdot), K(x'_2, \cdot), \ldots, K(x'_m, \cdot)\big)$, the error sum of squares for mean drift clustering is indicated bellow.

$$\min J = \sum_{i=1}^{k} \sum_{j=1}^{n} M_g(x') \big|K(x'_j, \cdot) - d_i(\cdot)\big|^2_{H_a}, \quad d_i(\cdot) \in H_a \tag{18}$$

That is, the ultimate goal of mean drift clustering is to minimize J in Equation (18) by defining the cluster center as bellow.

$$d_i(\cdot) = \sum_{j=1}^{m} \beta_{ij} K(x'_j, \cdot), \quad i = 1, 2, \ldots, k \tag{19}$$

For the goal of minimizing J, β needs to satisfy $\beta = M_g(x') K_A \hat{K}^{-1}$, where K_A is the kernel similarity matrix between the original learning resource sample and the sampling sample, i.e. $K_A \in \mathbb{R}^{m \times n}$.

Then, minimizing J is transformed into minimizing the matrix trace, and the clustering of learners with the same preference can be obtained through Equation (20).

$$\min J = \min\{\operatorname{tr}(K) - \operatorname{tr}\big(M_g(x') K_A \hat{K}^{-1}\big)\} \tag{20}$$

4.3. **Recommended online learning resources based on spatial downscaling and mean shift clustering.** On the ground of learner feature extraction and MSC, an online learning resource recommendation model based on spatial downscaling and MSC is constructed. The steps in detail are indicated below.

(1) Convert the learner's preference data into an $n \times m$-order rating matrix R. Fill in the missing values in each row of the matrix with 0. Calculate the mean value of each dimension for the n m-dimensional samples $x_1, x_2, \ldots, x_n$ of the matrix R, and subtract the mean value of that dimension from the data in each dimension in the matrix R to achieve the matrix X.

(2) The eigenvectors of the preference matrix are listed in descending order of the eigenvalues, and the unit eigenvectors corresponding to the first t eigenvalues are taken to form matrix Q. The matrix $Y = QR^T$ that best represents the user's preference is

extracted, and the matrix Y consists of n t-dimensional samples $y_1, y_2, \ldots, y_n$, which is the data of R after downscaling to the t-dimension.

(3) Randomly mark a centroid e among the unclassified data points in Y. The bandwidth d is automatically estimated using the kernel function encapsulated in the MSC algorithm.

(4) Calculate the mean drift vector $M(e)$ from all data points in the cluster to this center point e using Equation (17).

(5) Move the center point e in the direction of $M(e)$ by the length of $M(e)$. The process of moving the center point is indicated in Equation (21).

$$e^{s+1} = M^s + e^s \tag{21}$$

where e^{s+1} is the center in state $s+1$, M^s is the mean drift vector obtained in state s, and e^s is the center in state s.

(6) Repeat steps (1) to (5) until the size of the drift vectors meets the set threshold, and record the centroid at this point. If the distance between the center point of the current cluster C_Q and the center point of the previously existing cluster C_i is less than the threshold at the time of convergence, then C_Q and C_i should belong to the same cluster and be merged into C_i. Otherwise, C_Q should be treated as a new cluster.

(7) The individual sample points are grouped into the cluster with the highest number of markers to produce the final cluster $\text{clus}_1, \text{clus}_2, \ldots, \text{clus}_k$.

(8) In the feature space, adopting Equation (22) to calculate the similarity of users in the cluster where the target learner u is located, select a number of users with the highest similarity to u as their nearest neighbors according to the need, and establish a list of scored items of nearest neighbor users.

$$\text{SIM}_{uv} = \frac{\sum_{i \in I} r_{ui} \cdot r_{vi}}{\sqrt{\sum_{i \in I} r_{ui}^2} \sqrt{\sum_{i \in I} r_{vi}^2}} \tag{22}$$

where $\text{SIM}_{u,v}$ is the similarity between u_i and u_j; i denotes a learning resource; I denotes the set of resources that have been jointly rated by learners u and v; r_{ui} is the rating of resource i by learner u; and r_{vi} is the rating of resource i by learner v.

(9) Predict the target learner u's rating of the unrated learning resources in the list adopting Equation (23).

$$P_{u,i} = \frac{\sum_{v \in N_u} \text{sim}(u,v)\, R_{vi}}{\sum_{v \in N_u} |\text{sim}(u,v)|} \tag{23}$$

where $P_{u,i}$ is the rating of u on the unrated resource i; N_u is the set of u's nearest neighbors; v is one of the nearest neighbors; $\text{sim}(u,v)$ is the similarity between learner u and v; and R_{vi} is the rating of learner v on the learning resource i.

(10) Sort the resources in descending order of predictive scores to form a list and select the Top N learning resources to recommend to learners.

5. Performance testing and analysis.

5.1. Effect analysis of entity reasoning.

To estimate the performance advantages of the suggested recommendation model, this article conducts comparative experiments among the suggested model and other current models, and all the experiments are carried out under Linux operating system and Python 3.9.7 programming environment. For the convenience of analysis, the literature [16] is denoted as EISRD, the literature [21] as IDMCA, the literature [29] as AMCRS, and the algorithm in this article as SDMSC.

The experimental data adopted in this article comes from the MoocData dataset [30], which contains 53927 learners, 41395 learning resources and 65284 learner–learning resource interactions. The training, validation and testing sets are divided in the ratio of 5:3:2. A total of 10 experiments are conducted to calculate the performance indexes of the above four algorithms, and the average of the 10 experiments is taken for comparison.

The goal of this article is to make Top-N recommendations for learners, so Recall@N (Rec@N), NDCG@N (NC@N), HitRate@N (HR@N), and MAE, which are commonly used in recommendation algorithms, are selected to evaluate the performance of recommendation algorithms, where N represents the length of the recommendation list. Before the experiment, it was measured that after dimensionality reduction, better results were achieved when the dimensionality was taken as the first 100 dimensions and the number of nearest neighbors was 20. Therefore, in the experiments, the dimensionality of the dimensionality reduction is taken as 50 and the number of nearest neighbors is taken as 20. Table 1 lists the comparisons of Rec@10, Rec@30, NC@10, NC@30, HR@10 and HR@30 of the four algorithms when the number of recommendations is 10 and 30.

Table 1. Comparison of recommendation performance metrics for different models

Model	Rec@10	Rec@30	NC@10	NC@30	HR@10	HR@30
EISRD	0.619	0.724	0.297	0.475	0.475	0.651
IDMCA	0.652	0.759	0.314	0.509	0.492	0.682
AMCRS	0.714	0.791	0.343	0.522	0.539	0.714
SDMSC	0.753	0.852	0.371	0.547	0.568	0.743

The suggested model SDMSC achieves the best performance in Rec@10, Rec@30, NC@10, NC@30, HR@10 and HR@30 metrics, which is 21.65%, 17.68%, 24.92%, 15.16%, 22.15%, 14.13% better than the EISRD method, and 15.49%, 12.25%, 18.15%, 7.47%, 15.45%, 8.94% better than the IDMCA method. It is 5.46%, 7.71%, 8.16%, 4.19%, 5.38%, 4.06% respectively over the AMCRS method. This is because the EISRD model only adopts the most popular data without considering user preferences, and its performance is much worse than the other three recommendation models. The AMCRS model is better than the IDMCA model in terms of evaluation metrics, due to the fact that the user's behavioral data in the IDMCA model is more sparse, while the AMCRS model has a higher accuracy of dimensionality reduction on the features of learning resources. The SDMSC model adopts spatial dimensionality reduction in achieving the user preference matrix, which obtains the user's behavioral interests with fewer parameters and faster training speed when the data volume is large, and at the same time, the improved MSC algorithm can be utilized to cluster the users on the low-dimensional vectors after dimensionality reduction, which ensures the improvement of the recommendation performance.

In addition to the above indicators, the MAE value is also an indicator for verifying the accuracy of the recommendation model. The MAE value is the average of the absolute error between the predicted score and the actual score of the learner, and the smaller the MAE value is, the closer the score of the recommendation model is to the actual value. The MAE values of each recommendation method are indicated in Figure 3.

When the length of the recommended list is 10, 30, and 50, the MAE is 0.462, 0.293, and 0.211 for the A method, 0.410, 0.261, and 0.175 for the B method, 0.373, 0.217, and 0.142 for the C method, and 0.326 and 0.152 for the AMCRS method, respectively. The SDMSC method outperforms the EISRD, IDMCA and AMCRS methods. Since spatial dimensionality reduction itself can effectively alleviate data sparsity, adding the

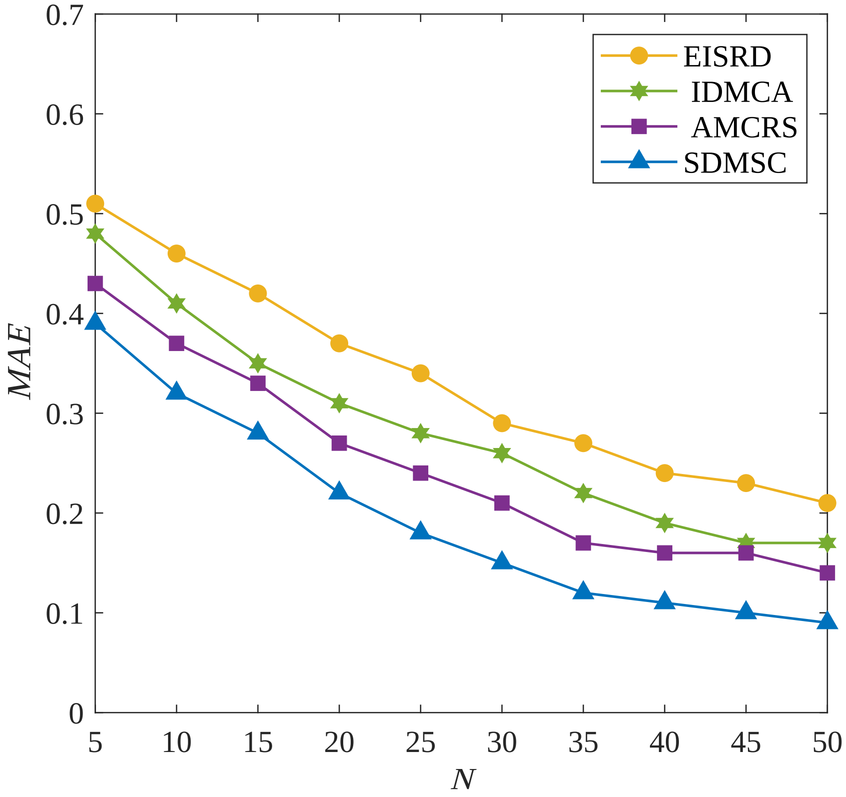

Figure 3. Comparison of MAE for different models

mean drift clustering module on top of it for learner clustering alleviates the issue of poor convergence of clustering and data sparsity that most recommendation algorithms may have.

5.2. **Ablation experiments and analysis of results.** To better estimate the impact of spatial dimensionality reduction module and MSC module in SDMSC, ablation experiments are conducted on MoocData dataset respectively, and two comparison models are designed for analysis. Evaluation metrics use top-N recommendation, click prediction rate CTR and MAE.

(1) Remove the spatial dimensionality reduction feature extraction module. The optimized MSC module is directly used to recommend learning resources, denoted as SDMSC-SD. (2) Remove MSC module. The preference matrix is downscaled using the spatial downscaling module without clustering the learners, denoted as SDMSC-MS.

Table 2. Results of ablation experiments with different modules

Method	CTR		Top-N	
	Accuracy	Precision	N@10	N@30
SDMSC-SD	0.819	0.791	0.193	0.351
SDMSC-MS	0.845	0.837	0.228	0.386
SDMSC	0.892	0.885	0.297	0.437

As can be seen from Table 2, when N takes the value of 10, the recommendation success rate of SDMSC-SD is 19.3%, SDMSC-MS is 22.8%, and SDMSC is 29.7%. In addition, the Accuracy of SDMSC is improved by 8.91% and 5.56% compared with that of SDMSC-SD and SDMSC-MS, and the Precision is improved by 11.88% and 5.73%, and the recommendation success rate of SDMSC is obviously better than that of SDMSC-SD and

SDMSC-MS, respectively. Secondly, it can be observed that there is a significant decrease in the recommendation performance of SDMSC after removing the spatial dimensionality reduction module, which proves the importance of the space of learners' preference matrix. Without clustering similar learners, the recommendation performance also decreases, and thus SDMSC with all the modules integrated achieves the best performance.

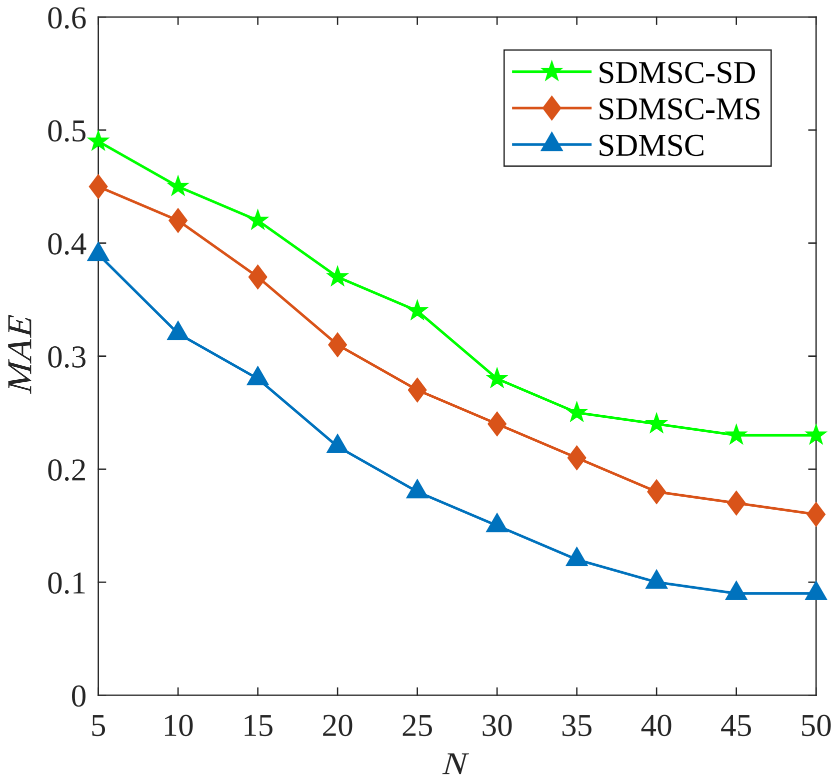

Figure 4. MAE after ablation experiments with different methods

The outcome of the ablation experiments is indicated in Figure 4. The MAE values of the three models first decrease gradually with the increase of the recommended list length and then stabilize, and the SDMSC recommended model always has a smaller MAE value. Meanwhile, when the N value is relatively small, the difference between the comparison models in terms of MAE value is not large, however, as the N value increases, SDMSC decreases at a higher rate than the other two models, which indicates that the prediction scoring error of SDMSC is smaller and proves that SDMSC has better performance.

6. **Conclusion.** Focusing on the current recommendation algorithms' issues of sparse matrix and poor user clustering, which results in low recommendation accuracy, this article suggests an online learning resource recommendation algorithm based on spatial downscaling and MSC. Firstly, all pairs of constraints are linearly transformed in the high-dimensional kernel space, and all feature samples are projected into the zero space of the constraint vectors, so as to enhance the accuracy of the clustering outcome through the a priori information. Then the sparse high-dimensional learner preference matrix is spatially downscaled to extract the preference features, and the optimized mean drift algorithm is adopted to cluster the learners relied on the downscaled data; then the nearest neighbors are searched in the clusters, and the ratings of the target users on the unassessed items are computed through the nearest-neighbor similarity weighting, and a recommendation list is generated based on the ratings from high to low. Finally, the

experimental outcome indicates that the proposed algorithm improves the Recall, NDCG and HitRate indexes compared with the comparison methods, which indicates that the suggested algorithm can be better applied to online learning resources recommendation.

REFERENCES

[1] L. Harasim, "Shift happens: Online education as a new paradigm in learning," *The Internet and Higher Education*, vol. 3, no. 1–2, pp. 41–61, 2000.

[2] J. Larreamendy-Joerns and G. Leinhardt, "Going the distance with online education," *Review of Educational Research*, vol. 76, no. 4, pp. 567–605, 2006.

[3] M. Hanna, "Data mining in the e-learning domain," *Campus-Wide Information Systems*, vol. 21, no. 1, pp. 29–34, 2004.

[4] T.-Y. Wu, H. Li, S. Kumari, and C.-M. Chen, "A Spectral Convolutional Neural Network Model Based on Adaptive Fick's Law for Hyperspectral Image Classification," *Computers, Materials & Continua*, vol. 79, no. 1, pp. 19–46, 2024.

[5] T.-Y. Wu, A. Shao, and J.-S. Pan, "CTOA: Toward a Chaotic-Based Tumbleweed Optimization Algorithm," *Mathematics*, vol. 11, no. 10, 2339, 2023.

[6] T.-Y. Wu, H. Li, and S.-C. Chu, "CPPE: An Improved Phasmatodea Population Evolution Algorithm with Chaotic Maps," *Mathematics*, vol. 11, no. 9, 1977, 2023.

[7] H. Guo and X. Cheng, "Individual recommendation method of college physical education resources based on cognitive diagnosis model," *EAI Endorsed Transactions on Scalable Information Systems*, vol. 9, no. 5, e8, 2022.

[8] H. Chen, C. Yin, R. Li, W. Rong, Z. Xiong, and B. David, "Enhanced learning resource recommendation based on online learning style model," *Tsinghua Science and Technology*, vol. 25, no. 3, pp. 348–356, 2019.

[9] M.-J. Huang, H.-S. Huang, and M.-Y. Chen, "Constructing a personalized e-learning system based on genetic algorithm and case-based reasoning approach," *Expert Systems with Applications*, vol. 33, no. 3, pp. 551–564, 2007.

[10] S. Shanshan, G. Mingjin, and L. Lijuan, "An improved hybrid ontology-based approach for online learning resource recommendations," *Educational Technology Research and Development*, vol. 69, pp. 2637–2661, 2021.

[11] J. Shu, X. Shen, H. Liu, B. Yi, and Z. Zhang, "A content-based recommendation algorithm for learning resources," *Multimedia Systems*, vol. 24, no. 2, pp. 163–173, 2018.

[12] R. C. Kushwaha, A. Singhal, and A. Biswas, "E-textbook enrichment using graph based e-content recommendation," *Journal of Computational and Theoretical Nanoscience*, vol. 17, no. 1, pp. 492–498, 2020.

[13] M. Nilashi, O. Ibrahim, and K. Bagherifard, "A recommender system based on collaborative filtering using ontology and dimensionality reduction techniques," *Expert Systems with Applications*, vol. 92, pp. 507–520, 2018.

[14] S. Ozkan and R. Koseler, "Multi-dimensional students' evaluation of e-learning systems in the higher education context: An empirical investigation," *Computers & Education*, vol. 53, no. 4, pp. 1285–1296, 2009.

[15] Y. Jia and Q. Zhao, "The learning behavior analysis of online vocational education students and learning resource recommendation based on big data," *International Journal of Emerging Technologies in Learning*, vol. 17, no. 20, 261, 2022.

[16] A. M. A. Al-Sabaawi, H. Karacan, and Y. E. Yenice, "Exploiting implicit social relationships via dimension reduction to improve recommendation system performance," *PloS One*, vol. 15, no. 4, e0231457, 2020.

[17] L. Zhou, F. Zhang, S. Zhang, and M. Xu, "Study on the personalized learning model of learner-learning resource matching," *International Journal of Information and Education Technology*, vol. 11, no. 3, pp. 143–147, 2021.

[18] J. Chen, G. Zhou, and Y. Zhong, "A personalized recommendation system for teaching resources in sports using fuzzy C-means clustering technique," *Soft Computing*, vol. 28, no. 1, pp. 703–720, 2024.

[19] Q. Li, "A New Methodology for Clustering of Online Learning Resources Based on Students' Learning Styles," *International Journal of Emerging Technologies in Learning*, vol. 18, no. 13, pp. 10–14, 2023.

[20] S. Zahra, M. A. Ghazanfar, A. Khalid, M. A. Azam, U. Naeem, and A. Prugel-Bennett, "Novel centroid selection approaches for KMeans-clustering based recommender systems," *Information Sciences*, vol. 320, pp. 156–189, 2015.

[21] S. Kausar, X. Huahu, I. Hussain, Z. Wenhao, and M. Zahid, "Integration of data mining clustering approach in the personalized E-learning system," *IEEE Access*, vol. 6, pp. 72724–72734, 2018.
[22] F. Zhang and A. P. Georgakakos, "Joint variable spatial downscaling," *Climatic Change*, vol. 111, pp. 945–972, 2012.
[23] K.-L. Wu and M.-S. Yang, "Mean shift-based clustering," *Pattern Recognition*, vol. 40, no. 11, pp. 3035–3052, 2007.
[24] J. Bagherzadeh and H. Asil, "A review of various semi-supervised learning models with a deep learning and memory approach," *Iran Journal of Computer Science*, vol. 2, pp. 65–80, 2019.
[25] H. Gao, H. Zhao, and S. Osher, "Bregman methods in quantitative photoacoustic tomography," *Cam Report*, vol. 30, no. 6, pp. 3043–3054, 2010.
[26] F. Nie, H. Huang, X. Cai, and C. Ding, "Efficient and robust feature selection via joint $\ell_{2,1}$-norms minimization," *Advances in Neural Information Processing Systems*, vol. 23, pp. 1813–1821, 2010.
[27] G. K. Uyanık and N. Güler, "A study on multiple linear regression analysis," *Procedia – Social and Behavioral Sciences*, vol. 106, pp. 234–240, 2013.
[28] Š. Raudys, "On the effectiveness of Parzen window classifier," *Informatica*, vol. 2, no. 3, pp. 434–454, 1991.
[29] M. Nilashi, O. B. Ibrahim, N. Ithnin, and R. Zakaria, "A multi-criteria recommendation system using dimensionality reduction and Neuro-Fuzzy techniques," *Soft Computing*, vol. 19, pp. 3173–3207, 2015.
[30] I. Uddin, A. S. Imran, K. Muhammad, N. Fayyaz, and M. Sajjad, "A systematic mapping review on MOOC recommender systems," *IEEE Access*, vol. 9, pp. 118379–118405, 2021.