

Article

Not peer-reviewed version

Denoising Diffusion Model-driven Adaptive Estimation of Distribution Algorithm Integrating Multi-modal Data

[Lin Bao](#) , Lina Wang , [Biao Xu](#) * , Hang Yang , Yumeng Peng

Posted Date: 23 September 2025

doi: 10.20944/preprints202509.1882.v1

Keywords: multi-modal data; personalized search; denoising diffusion model; estimation of distribution algorithm; surrogate model



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Denoising Diffusion Model-Driven Adaptive Estimation of Distribution Algorithm Integrating Multi-Modal Data

Lin Bao ¹, Lina Wang ¹, Biao Xu ^{2,*}, Hang Yang ¹ and Yumeng Peng ¹

¹ School of Electronics and Information, Jiangsu University of Science and Technology, Zhenjiang, China

² College of Engineering, Shantou University, Shantou, China

* Correspondence: xubiao@stu.edu.cn

Abstract

Personalized search and recommendation algorithms for multi-modal data have attracted widespread attention. In view of the insufficient ability of multi-source information fusion and global search in complex optimization problems, this paper proposed a denoising diffusion model-driven adaptive estimation of distribution algorithm integrating multi-modal data. Firstly, multi-modal information is extensively collected from various sources. By introducing diffusion model, a user interest preference model based on the denoising diffusion model is established to extract the user preference features, which fits users' cognitive experiences and behavioral patterns. Meanwhile, in the framework of an estimation of distribution algorithm, some estimation of distribution strategies are designed to enhance the global exploration and local development ability of the optimization algorithm. Finally, according to new data information, a dynamic model update mechanism is established to promptly update the user interest preference model and related models, which tracks the changes of users' interest preferences for actual scenarios. It will help users quickly filter out items that match their interest preferences from a vast amount of information. The feasibility, effectiveness and superiority of the proposed algorithm were verified through many experiments on general public datasets. It improved the search efficiency and recommendation effect of the personalized recommendation algorithm.

Keywords: multi-modal data; personalized search; denoising diffusion model; estimation of distribution algorithm; surrogate model

MSC:

1. Introduction

With the rapid development of information technology, Internet, big data and other technologies, network information is growing explosively. In the digital age, it is almost impossible to control the generation and dissemination of information. The sources, types and forms of data information have become increasingly diverse and differentiated. At the same time, the continuous growth of the number of users and their generated data led to the emergence of new ways of connection, collaboration and sharing in the network. The scale and complexity of data have reached an unprecedented level, giving rise to the problem of information overload. In numerous fields such as e-commerce, social media, news information and video platforms, a vast amount of product information is constantly emerging and overwhelming. Users' available information has far exceeded their processing capacity, making it difficult to search for effective information or the content they are interested in. The application platform hopes to find the items that users are interested in from massive amounts of information, thereby enhancing users' experience and improving commercial benefits. Therefore, personalized search and recommendation systems are born at the right moment,

attracting extensive attention from researchers and practitioners^[1-5]. Personalized search and recommendation methods utilize multi-source and multi-modal data information, including tags, text, images, audio, video and other information with users' preferences. They mine feature information and its potential correlations to capture users' intentions to predict users' interest preferences in the decision-making process for personalized recommendation services. It has become a research hotspot in the field of artificial intelligence and its practical applications^[6-9].

With the development of machine learning and deep learning, many personalized recommendation methods have been proposed, such as: Collaborative Filtering (CF)^[10-11], Bayesian Personalized Ranking (BPR)^[12-13], matrix factorization^[14-15], Neural-CF^[16], etc., to achieve good recommendation effects. However, users' explicit feedback in the network requires more decision-making costs, and users need to think carefully before providing quantifiable feedback. Therefore, recommendation system considers implicit feedback, such as clicking, liking, following, purchase, and various operation behaviors, to comprehensively predict user's potential intentions and interest preferences for personalized recommendation tasks. Although these recommendation methods have achieved remarkable results, they still face severe challenges such as data sparsity, cold start problems, and dynamic preference modeling. Deep generative models can generate new data samples by learning the latent distribution of data in situations such as missing data, noise, and incomplete information. They possess powerful feature extraction and data modeling capabilities. They have gradually become an important method for handling complex multi-source heterogeneous data and have been successfully applied in tasks such as image generation and text generation. However, traditional generative models still face the challenges of how to integrate multimodal data to effectively extract feature information to promote practical applications. In recent years, diffusion model is an emerging generative model. It simulates the propagation process of information among graph nodes to effectively capture the similarities and correlations between nodes to achieve the feature extraction of multi-source heterogeneous data and the information fusion across data modalities. Diffusion model is applied to recommendation system, which will bring new opportunities to the research fields of personalized search and recommendation.

In practical application scenarios, the needs and interests of different users vary for the same task (such as purchasing books, searching for movies, etc.), and those of the same user at different time periods may also be different. In addition, users' knowledge experiences, potential needs, interest preferences and behavioral motivations may undergo dynamic changes with the influence of various factors such as the passage of time, environmental migration or increasing information. When dealing with complex and dynamical user personalized demands, the coarse-grained modeling of users' preferences limits the performance improvement of personalized recommendation methods. Due to the difficulty in quantitatively representing users' preferences and their dynamic changing characteristics, it is necessary to design an appropriate personalized search and recommendation framework. It models users' potential intentions and interest preferences, and provides the feedback based on users' information. The model can accurately capture feature information to predict users' preferences and dynamic trends. This further guides the model to adaptively optimize and adjust the dissemination mechanism in the subsequent iterative process to search for items that users are interested in for personalized recommendation tasks. It will improve the search efficiency, recommendation effect and robustness of personalized search and recommendation algorithms. However, personalized search and recommendation methods still face the following significant challenges. Users' interactions are scarce, and there are problems such as data sparsity, cold start, and long-tail item recommendation. It will lead to underfitting in model training. It is difficult to accurately capture users' intention representation by using supervisory signals, resulting in poor modeling of dynamic user interest preferences.

To address the aforementioned issues and challenges, this paper proposed a denoising diffusion model-driven adaptive estimation of distribution algorithm integrating multi-modal data. User-generated contents and related information are extensively collected to explore multi-modal user-generated contents. Multi-source multi-modal data is effectively integrated by diffusion model. A

user interest preference model based on denoising diffusion model is established to efficiently extract the deep-seated potential interest preference features of users and the development evolution pattern of their interests. In the framework of estimation of distribution algorithm, adaptive estimation of distribution strategies and a surrogate model based on user preferences in multi-modal data fusion mode are designed to simulate users' cognitive experiences and behavioral patterns to guide the direction of evolutionary optimization search to timely track users' interest preferences. A dynamic management mechanism for the model is designed to dynamically update the user interest preference model and related models based on new information. It helps users filter out items that match their interest preferences from a vast amount of information for personalized search and recommendation tasks. The feasibility, effectiveness and superiority of the proposed algorithm have been verified through a large number of experiments on actual multi-domain public datasets. It enhances the global exploration and local development capabilities of the search algorithm, which improves users' personalized search experience and satisfaction on recommendation system platforms. It has good scalability and adaptability.

The contributions of this paper mainly include three aspects. (1) For dynamic personalized search and recommendation tasks, considering multimodal information fusion and cross-modal alignment representation, a user interest preference model based on the denoising diffusion model is constructed to understand user interests and multimodal content and obtain users' preference features. (2) The adaptive estimation of distribution operators and strategies based on the user interest preference model are designed in the framework of the estimation of distribution algorithm. It refines users' intention representation and interest selection tendency from a micro perspective to generate new individuals with the user preference information to fit the dynamic change process of user interest preferences. (3) A denoising diffusion model-driven user interest preference surrogate model is established to estimate the fitness of individuals to track users' interest preferences for guiding the forward direction of the personalized evolutionary search. It helps efficiently complete personalized search and recommendation tasks.

The remainder of this study is organized as follows. Section 2 introduces the notations of our study and related work. In Section 3, the proposed algorithm is described in detail. Section 4 presents comparative experiment results and corresponding analysis. Finally, the conclusion is followed.

2. Related Work

2.1. Mathematical Description for Personalized Search Problems with User-Generated Contents

Personalized search tasks for UGCs involve searching for optimization targets that meet users' potential needs and personalized interest preferences in a dynamic evolving search space of massive multi-source heterogeneous user-generated data, which recommended personalized item (product or solution) lists to users. Essentially, it is a kind of complex dynamic qualitative index optimization problem. In the process of personalized search, users conduct qualitative analysis, evaluation and decision-making on the searching items based on their cognitive experience and interest preferences. However, users' cognitive experience and interest preferences are diverse, ambiguous, uncertain and progressive. The definition of user satisfaction solutions is very subjective and varies from person to person, and the search results and recommendation effects are completely determined by users' subjective preferences. Here, the objective function $f_u(\mathbf{x})$ for the personalized search problem with UGCs can be defined as follows:

$$\begin{cases} Y = f(X_{u,v}; \theta_f) \\ \text{s.t. } u \in U, v \in V \end{cases} \quad (1)$$

where $U = \{u_1, u_2, \dots, u_{|U|}\}$ is the user set, $|U|$ indicates the number of users; $V = \{v_1, v_2, \dots, v_{|V|}\}$ is the set of items (the feasible solution space), usually the feasible solution space X is large and sparse, $|X|$ represents the number of items. Each item (solution) \mathbf{x}_i contains n decision variables

representing the characteristics of the item, expressed as $\mathbf{x}_i = \{x_{i1}, x_{i2}, \dots, x_{in}\}$; The user's preference for the item is a model function $f(X_{u,v})$ with learnable parameters θ_f . $X_{u,v}$ represents a set of features associated with the user u and the item v .

When users conduct personalized searches, they will quickly search for a list of recommended items $TopN$ that meet their needs and may be of interest to them in the feasible solution space, that is, the set of N items with higher value $f_u(\mathbf{x})$. Effective item recommendations will be made to complete personalized search and recommendation tasks to stimulate user exploration and enhance users' experience and satisfaction.

2.2. Recommendation Algorithms Integrating Multi-Modal Data

In recommendation systems, various extractable and analyzable multi-modal data, including text, images, audio and other information, is often used as important additional information for users' interaction behavior data, which enriched the representation features of users and items to a certain extent alleviate the data sparsity and cold start. The multi-modal data and collaborative information of users and items are deeply mined to effectively integrate the single-modal representations and multi-modal features to maintain the integrity and diversity of the fused modal information. It is the key to the recommendation algorithms integrating multi-modal data that captures comprehensive and accurate user preferences and item representations.

In early work, He et al.^[17] extracted items' image features through Convolutional Neural Networks (CNN) and used matrix factorization to predict users' preferences. The Visual Bayesian Personalized Ranking (VBPR) model was proposed to use Bayesian personalized ranking to train the model to alleviate the cold start problem. Kim et al.^[18] combined CNN and Probabilistic Matrix Factorization (PMF) to capture the context information of documents, and proposed a convolutional Matrix factorization model (ConvMF) to improve the recommendation accuracy. Chen et al.^[19] utilized the ResNet-152 model to extract modal information in image and video and proposed an attentive collaborative filtering (ACF) method. Wei et al.^[20] utilized multi-modal information of vision, audio and text to construct a user-short video bipartite graph. A multi-modal graph convolution network (MMGCN) is proposed by the topological structure of neighboring nodes to enrich the representation of each node. It can learn high-order features from the user-short video bipartite graph to improve the recommendation performance. Wang et al.^[21] used a pre-trained word embedding model to represent text features and utilized convolutional neural networks to obtain different single-level visual features from pooling layers of different depths respectively. A movie recommendation system based on Visual Recurrent Convolutional Matrix Factorization (VRConvMF) is proposed to improve the accuracy of the recommendation system. Yang et al.^[22] designed the multi-modal module, attention module and multi-head residual network module to extract the image features of video cover to expand the learned feature set. The multi-head multi-modal deep interest network (MMDIN) is proposed to enhance the representational ability, predictive performance and recommendation performance. Deng et al.^[23] proposed a recommendation model based on multi-modal fusion and behavior expansion. A learning-query multi-modal fusion module is designed to perceive the dynamic content of flow fragments and handle complex multi-modal interactions. A graph-guided interest expansion method is presented to learn the representations of users and information flows in multimodal attribute large-scale graph. Yan et al.^[24] presented a pre-trained model that can extract high-quality multimodal embedding representations. A content interest-aware supervised fine-tuning is designed to guide the alignment of user preference embedding representations through users' behavior signals to bridge the semantic gap between contents and user interests. A multi-modal content interest modeling paradigm for user behavior modeling is proposed by integrating multi-modal embedding and ID-based collaborative filtering signals into a unified framework.

The above-mentioned methods conduct the feature extraction and information fusion of multi-modal data from multiple perspectives, which to a certain extent improves the performance of

recommendation system. However, the existing methods still face the problem of relying on commodity-related information to enrich items' representation, incomplete mined multimodal semantic information, and ignoring the correlations and differences among different modalities. It will lack the ability to model modal-specific user preferences, resulting in an insufficient understanding of users' potential needs and deep interest preferences. Meanwhile, a large amount of noise is introduced when conducting multi-modal information fusion. There are deficiencies in feature fusion and semantic association modeling. It will affect the accuracy and recommendation effect of recommendation system.

2.3. *Diffusion Models*

In recent years, Diffusion Models have made breakthroughs in fields, such as computer vision, natural language processing and other fields, due to their powerful data generation ability, representation learning ability and sequence modeling ability. It has become an emerging research hotspot^[25-29]. Diffusion model is a deep learning method based on probabilistic generation. Its core idea is the forward diffusion process of gradually adding Gaussian noise to the data and the data reverse generation process of learning reverse denoising. Diffusion models are introduced into the research of recommendation systems to provide a new generative modeling paradigm, which can effectively alleviate data sparsity and support the recommendation integrating multi-modal information^[30-33]. Li et al.^[34] proposed a sequence recommendation based on diffusion model. Items are represented as a model distribution that adaptively reflects users' multiple interests and various aspects of items. The target item is embedded into the Gaussian distribution by adding noise, which is applied to the sequential item distribution representation generation and uncertainty injection. Based on users' historical interaction behaviors, Gaussian noise is reverse-transformed into the representation of target items. Zhao et al.^[35] proposed a denoising diffusion recommendation model. It utilizes multi-step denoising process of diffusion model to inject controllable Gaussian noise in the forward process and iteratively remove noise in the reverse denoising process to robust the embedding representations of users and items. Jiang et al.^[36] proposed a knowledge graph diffusion model for recommendation. By integrating generated diffusion models with data augmentation paradigms, it achieves robust knowledge graph representation learning and promotes the collaboration between knowledge-aware item semantic and collaborative relationship modeling. A collaborative knowledge graph convolution mechanism with collaborative signals reflecting user-project interaction patterns was introduced to guide the knowledge graph diffusion process. Cui et al.^[37] utilized context information to generate reasonable enhanced views to propose an enhancing sequential recommendation with context-aware diffusion-based contrastive learning. Xia et al.^[38] proposed an anisotropic diffusion model for collaborative filtering in spectral domain. It mapped the user interaction vector to spectral domain and parameterized the diffusion noise to align with graph frequency. These anisotropic diffuses retained significant low-frequency components and maintained a high signal-to-noise ratio. Further, the conditional denoising network is adopted to encode users' interactions to restore the true preferences from noise data. However, diffusion models still face some challenges such as computational efficiency and interpretability. With model optimization and cross-domain integration, the research field of the recommendation algorithms based on diffusion models will focus on efficient reasoning, causal modeling and other directions, promoting their application in actual business.

3. Denoising Diffusion Model-Driven Adaptive Estimation of Distribution Algorithm Integrating Multi-Modal Data

3.1. *The Proposed Algorithm Framework*

The framework of the proposed denoising diffusion model-driven adaptive estimation of distribution algorithm integrating multi-modal data (DDM-AEDA) is shown in Figure 1.

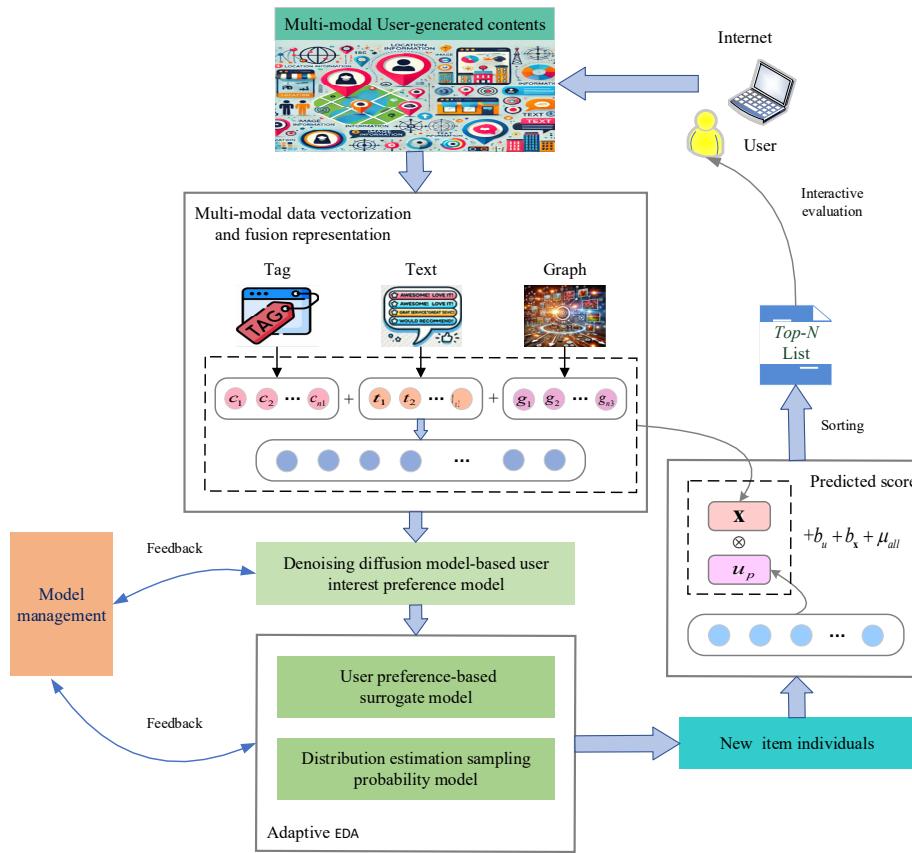


Figure 1. Algorithmic Framework.

The proposed algorithm mainly consists of four parts: (1) Multi-modal data information processing, including data acquisition and vectorized representation of multi-modal data; (2) User interest preference model based on denoising diffusion model; (3) Adaptive estimation of distribution algorithm based on surrogate model; (4) A model management mechanism for dynamically tracking the evolution of users' interest preferences.

3.2. Multi-Modal Data Information Processing and Its Fusion Learning Representation

In network, a large amount of multi-modal user-generated contents are widely collected, including rich historical user interaction behaviors (such as: users' ratings for items, users' text comments on items, etc.), items' content information (such as: items' category tags, items' brief descriptions, items' images, etc.), and social network relationships or other information. The above data contains numerous explicit and implicit interest preference information of users. The above-mentioned useful knowledge information that reflects users' interest preferences from different perspectives is fully utilized to help alleviate the problem of data sparsity in a big data environment to effectively improve the comprehensive performance of personalized search and recommendation algorithms.

The following presents the preprocessing and learning representation methods for multi-modal user-generated contents:

(1) Users' ratings: User's ratings on items are represented as a user rating matrix $\mathbf{R} = [r_{ij}]_{U \times V}$, where r_{ij} represents the rating of user u_i for item x_j . The larger the value, the more the user u_i likes the item x_j . Users' ratings explicitly express the degree of users' preferences for items.

(2) Items' category tags: Items' category tags briefly describe items' specific contents or feature information. To a certain extent, they reflect users' interest preferences. Here, multi-hot encoding is adopted. Based on the discrete values of the limited item category tags, the category tags in the item

individual \mathbf{x} are vectorized as $\mathbf{c} = [c_1, c_2, \dots, c_i, \dots, c_{n_1}]$, where c_i is the category tag of item \mathbf{x} . If $c_i = 1$, it indicates that the item \mathbf{x} contains the i th category tag; otherwise, it indicates that the item \mathbf{x} does not contain the i th category tag. n_1 is the total number of category tags for all items in the feasible solution space.

(3) Users' text comments: Text comments contain a large amount of users' implicit preference information. Users express their latent needs and interest preferences through emotional tendencies and semantic information in text comments. Based on natural language processing methods and text vectorization representation techniques, the users' text comments obtained by items are collected and organized to carry out natural language data preprocessing. A corpus based on the dataset is constructed to train a unsupervised sentence vector doc2vec model [329], which obtains a feature representation model that precisely expresses the latent semantic information of users' text comments.

Considering the interrelationships among text context, word order, semantics, etc., the doc2vec model compresses high-dimensional sparse word vectors into low-dimensional dense feature vectors and learns the fixed-length feature representations from variable-length text content (such as: sentences, paragraphs, documents, etc.). A user text comment vectorized representation matrix with items' ID index is generated to denote as $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_i, \dots, \mathbf{t}_{|X|}]^T$, $\mathbf{T} \in \mathbb{R}^{|X| \times n_2}$, where $\mathbf{t}_i = [t_{i1}, t_{i2}, \dots, t_{in_2}]$ represents the text comment vectorized representation of the item \mathbf{x}_i , and n_2 is the length of the user text comment vectorized representation.

(4) Social network relationships: Social network relationships express the friendship between users or the degree of similarity in their interest preferences. Usually, neighboring users have similar interests and hobbies, so these social network relationships imply a large amount of users' interest preference information. Here, Pearson correlation will be adopted to calculate the Pearson similarity coefficient $Sim(u_i, u_j)$ between users. The neighboring users with higher correlation coefficients are selected. Those neighboring users and their related information are utilized to assist the personalized search algorithm for deducing the items or content that users might like.

(5) Image information: A pre-trained ResNet model is utilized to extract high-dimensional visual feature vectors $g_i \in \mathbb{R}^{d_v}$ of Items' image. The image tensor of the item set is expressed as $G = [g_1, \dots, g_{n_3}] \in \mathbb{R}^{n \times d_v}$.

(6) Multi-modal information fusion and cross-modal alignment representation: Through the fully connected layer, the vector representations of items' tags, text comments and images are mapped to a shared low-dimensional space.

$$\mathbf{g}_{emb} = \mathbf{W}_g \mathbf{g} + \mathbf{b}_g \in \mathbb{R}^{n_3} \quad (2)$$

$$\mathbf{t}_{emb} = \mathbf{W}_t \mathbf{t} + \mathbf{b}_t \in \mathbb{R}^{n_2} \quad (3)$$

$$\mathbf{c}_{emb} = \mathbf{W}_c \mathbf{c} + \mathbf{b}_c \in \mathbb{R}^{n_1} \quad (4)$$

The mapped features are concatenated into a consistent learning representation:

$$\mathbf{x} = \text{Concat}(\mathbf{c}_{emb}, \mathbf{t}_{emb}, \mathbf{g}_{emb}) \in \mathbb{R}^{n_1 + n_2 + n_3} \quad (5)$$

The above process achieves the fusion representation of multi-modal data information through embedding technology to obtain the genotype representation $\mathbf{x} = [\mathbf{c}, \mathbf{t}, \mathbf{g}] = [c_1, c_2, \dots, c_{n_1}, t_1, t_2, \dots, t_{n_2}, g_1, g_2, \dots, g_{n_3}]$ of the item individual \mathbf{x} .

3.3. User Interest Preference Model Based on Denoising Diffusion Model

The collection of items that users like is screened to establish an advantageous population containing users' positive preference information. A training dataset $T_a = \{\mathbf{x}_i | i = 1, 2, \dots, |P_a|\}$ is formed through combining the genotype representation of individual items. A user interest

preference model based on denoising diffusion model is constructed and trained by introducing a denoising diffusion model to model users' interest preferences. The user interest preference model based on denoising diffusion model is shown in Figure 2.

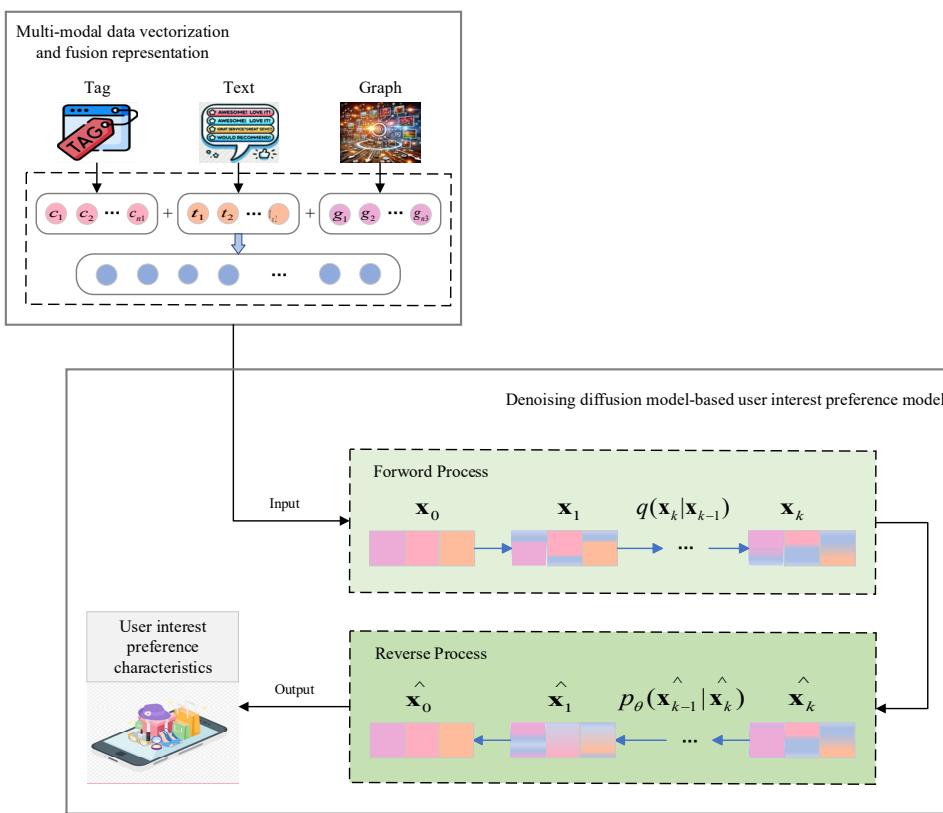


Figure 2. User Interest Preference Model based on Denoising Diffusion Model.

The training dataset is fed into the user interest preference model based on denoising diffusion model. The training process mainly includes two steps: the forward process and reverse process. In the forward process, Gaussian noise is gradually added to original data. \mathbf{x}_0 represents the original user behaviors, which simulates Markov Chain to obtain a series of intermediate states $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$, where k is the size of time step and the distribution of each step is a conditional Gaussian distribution. The forward process simulates that the data is contaminated by noise (such as: accidental clicks, random browsing, etc.). Given the current state \mathbf{x}_k , the next state \mathbf{x}_{k+1} is obtained:

$$\mathbf{x}_{k+1} = \sqrt{1 - \beta_k} \mathbf{x}_k + \sqrt{\beta_k} \varepsilon_k \quad (6)$$

where β_k is the noise variance, controlling the noise intensity at each step; $\varepsilon_k \sim N(0, I)$ is the noise of standard normal distribution.

The reverse process is also modeled as a Markov process, starting from the noise sample $\mathbf{x}_k \sim N(0, I)$ to recover $\hat{\mathbf{x}}_0$ by gradually removing the noise. The denoising distribution at each step is:

$$p_\theta(\hat{\mathbf{x}}_{k-1} | \hat{\mathbf{x}}_k) = N(\hat{\mathbf{x}}_{k-1}; \mu_\theta(\hat{\mathbf{x}}_k, k), \sum_\theta(\hat{\mathbf{x}}_k, k)) \quad (7)$$

where $\mu_\theta(\hat{\mathbf{x}}_k, k)$ and $\sum_\theta(\hat{\mathbf{x}}_k, k)$ are the parameters learned through the diffusion model.

The reverse process will reverse the noise addition process in the forward process as much as possible to restore the original data $\hat{\mathbf{x}}_0$:

$$\hat{\mathbf{x}}_0 = f_{DDPM}(\mathbf{x}_0) \quad (8)$$

The true distribution $q(\hat{\mathbf{x}}_{k-1} | \hat{\mathbf{x}}_k, \hat{\mathbf{x}}_0)$ is derived from the forward process, representing the transfer of $\hat{\mathbf{x}}_k$ to $\hat{\mathbf{x}}_{k-1}$. Given \mathbf{x}_k and \mathbf{x}_0 , the true distribution $q(\hat{\mathbf{x}}_{k-1} | \hat{\mathbf{x}}_k, \hat{\mathbf{x}}_0)$ is a Gaussian distribution:

$$q(\hat{\mathbf{x}}_{k-1} | \hat{\mathbf{x}}_k, \hat{\mathbf{x}}_0) = N(\hat{\mathbf{x}}_{k-1}; \mu_t(\hat{\mathbf{x}}_k, \hat{\mathbf{x}}_0), \beta_k I) \quad (9)$$

The negative log-likelihood of the forward process is minimized by adjusting the parameters θ . During the training process, the following variational lower bound is used for optimization:

$$L(\theta) = E_q \left[\sum_{k=1}^K D_{KL} \left(q(\hat{\mathbf{x}}_{k-1} | \hat{\mathbf{x}}_k, \hat{\mathbf{x}}_0) \| p_\theta(\hat{\mathbf{x}}_{k-1} | \hat{\mathbf{x}}_k) \right) \right] \quad (10)$$

where, D_{KL} is the Kullback-Leibler divergence, which is used to measure the difference between the true distribution $q(\hat{\mathbf{x}}_{k-1} | \hat{\mathbf{x}}_k, \hat{\mathbf{x}}_0)$ and the generated distribution $p_\theta(\hat{\mathbf{x}}_{k-1} | \hat{\mathbf{x}}_k)$.

Through the above process, the user interest preference model based on denoising diffusion model learns to recover data from noise and gradually approaches the true distribution of the data during the optimization process.

3.4. Surrogate-Driven Adaptive Estimation of Distribution Algorithm

The implicit representations of users and items are learned through matrix factorization to design a surrogate model based on the user preferences to estimate the fitness values of individual items in the population. Meanwhile, by integrating an estimation of distribution algorithm, a distributive estimation sampling probability model and an adaptive distributive estimation strategy are established to enhance the data utility from a personalized perspective to guide the personalized search process for the recommendation algorithm.

The representation $\mathbf{x} = [\mathbf{c}, \mathbf{t}, \mathbf{g}] = [c_1, c_2, \dots, c_{n_1}, t_1, t_2, \dots, t_{n_2}, g_1, g_2, \dots, g_{n_3}]$, $\mathbf{x} \in P_a$ of the item individual is feed into the user interest preference model based on denoising diffusion model that has been trained in Section 3.3 to obtain the user implicit representation u_p with the current user interest preferences:

$$u_p = f_{DDPM_p}(\mathbf{x}) \quad (11)$$

By using matrix factorization model, the predicted preference value $f_u(\mathbf{x})$ of the current user for the item individual is estimated through user's implicit representation u_p and the vector representation of the item individual \mathbf{x} :

$$f_u(\mathbf{x}) = \hat{R}_{DDPM_p} = u_p \cdot \mathbf{x} + b_u + b_x + \mu_{all} \quad (12)$$

where, b_u and b_x are respectively user bias and item bias; μ_{all} is the average rating of all samples.

In the framework of an estimation of distribution algorithm, $f_u(\mathbf{x})$ is served as the fitness function of the estimation of distribution algorithm. Its estimated value is used as the fitness value of the evolutionary individual to participate in the personalized evolutionary search process.

The dominant population P_a is taken as the initial population $P(t=0)$. The activation probability v_i of the neural unit of the denoising diffusion probability model is obtained through the user interest preference model based on denoising diffusion model.

$$P_{\gamma_{DDPM_p}}(v_i) = \mu' = \tanh(\sum_j W_{ij} h_j + b_i) \quad (13)$$

The population is fed into the denoising diffusion probability model to obtain the reconstructed representation of item individuals. Then, the mean $\bar{\mu}_i$ and covariance matrix element C of the

reconstructed representation of item individuals are calculated to establish the distribution estimation sampling probability model.

$$G(\mathbf{x}) = \frac{2\pi^{-n/2}}{(\det C)^{1/2}} \exp(-(\mathbf{x} - \bar{\mu})^T (C)^{-1} (\mathbf{x} - \bar{\mu})/2) \quad (14)$$

According to the elite selection strategy and the fitness function, the evolutionary individuals with higher fitness in the population are selected to generate a subpopulation $P_{sr}(t) = sr * P(t)$, where sr is the selection ratio. The maximum likelihood estimation of $P_{sr}(t)$ is calculated to obtain the mean $\bar{\mu}(t)$.

$$\bar{\mu}(t) = \frac{1}{|P_{sr}(t)|} \sum_{i=1}^{|P_{sr}(t)|} P_{sr}(t)_i \quad (15)$$

The mean $\bar{\mu}(t)$ mainly controls the center of sampling offspring. An overly large selection ratio may cause the mean to be too far from the optimal fitness region, which increases the search diversity and slows down the convergence in the later stage. A too small sr will bring the mean closer to the optimal fitness region, which is suitable for local development and may lead to premature convergence. Here, an adaptive strategy of sr is designed as follows:

$$sr = sr_{max} - (sr_{max} - sr_{min}) \left(\frac{FES}{FES_{max}} \right)^{0.1} \quad (16)$$

where sr_{max} and sr_{min} represent the maximum and minimum selection ratios; FES_{max} indicates the maximum number of evaluating the fitness function; FES represents the number of evaluating the fitness function up to the current generation.

According to $sc = cs * P(t)$, a subpopulation $P_{sc}(t)$ is formed, where cs is the covariance scaling parameter, $cs > sr$. An archive set $A(t) = P_{sc}(t-1) \cup P_{sc}(t-2) \dots \cup P_{sc}(t-l)$ is obtained by storing each generation's subpopulation $P_{sc}(t)$, where l is the set length of the archive set. Combining the current subpopulation, $H(t) = A(t) \cup P_{sc}(t)$ is obtained to estimate the covariance C to select more individuals with higher fitness to participate in the covariance estimation:

$$C(t) = \frac{1}{|H(t)|} \sum_{i=1}^{|H(t)|} (H(t)_i - \bar{\mu}(t)) (H(t)_i - \bar{\mu}(t))^T \quad (17)$$

An overly large cs here will lead to a wide sampling range, which is beneficial for diversity search in the early stage but not conducive to centralized search in the later stage. However, a cs that is too small will result in an overly narrow sampling range, which is suitable for local development in the later stage but may lead to early convergence. Therefore, during the evolution of the estimation of distribution algorithm, cs will be dynamically adjusted and the adaptive strategy of its cs is as follows:

$$cs = 1 - (1 - sr_{min}) \left(\frac{FES}{FES_{max}} \right)^2 \quad (18)$$

A framework of the adaptive estimation of distribution algorithm is constructed by dynamically adjusting the selection ratio and covariance scaling parameter and combining with the historical information of the archive set $A(t)$, which can maintain the global exploration ability in the early stage of evolution to precisely converge to the optimal region corresponding to users' interest preferences in the later stage for achieving a dynamic balance in the personalized search process.

New evolved individuals are generated by the distributable estimation sampling probability model. The similarity between new individuals and the real items in feasible solution space is calculated according to the Pearson similarity criterion. Items with higher similarity are selected to replace the evolutionary individuals to construct the item set s_u to be recommended. Subsequently, the surrogate model is used to estimate the individual fitness. The top N items are selected through the elite selection strategy and recommended to the current user to complete an interactive personalized recommendation.

At the same time, the population $P(t) \rightarrow P(t+1)$ is updated based on users' feedback information to carry out the next round of interactive personalized evolutionary search process.

3.5. Model Dynamic Management Mechanism

A dynamic model management mechanism is designed in view of the diversity of users' actual demands and the time-varying interest preferences in a complex network environment. The dynamic update of the models in dynamic scenarios and the continuous evolution of optimization strategies are achieved through a closed-loop feedback mechanism. When the prediction accuracy of the models drops to the preset threshold with the dynamic changes of the environment, it will trigger the collaborative updating optimization process of the user preference model and its related models. Based on new user-generated contents and related information, the above models and parameters will be dynamically updated to promptly track users' interest preferences and potential demands and monitor the prediction accuracy. It guides the personalized search process to gradually find users' satisfied solutions for completing the personalized recommendation task.

3.6. Algorithm Implementation and Computational Complexity Analysis

The specific implementation steps of the proposed algorithm are pseudocode as follows:

Algorithm: DDM-AEDA

Input: Multi-modal UGCs information

Output: Item recommendation list $TopN$

Start

1. Data preprocessing: The training dataset is constructed according to the method in Section 3.2;
2. Pre-trained models: User-generated contents are extensively collected to train pre-trained doc2vec and ResNet models;
3. Initialization: In the search space, a set of items with users' preferences is filtered based on user-generated data to form the initial dominant group P_a ;
4. Social network: According to the method in Section 3.2, neighboring users are screened out to establish users' social network groups;

Do while (The algorithm termination condition has not been met)

5. User preference model: According to the method in Section 3.3, a user preference model based on denoising diffusion model is constructed and trained to extract users' preference features;
6. Probability model: A distributed estimation sampling probability model is designed by using the method in Section 3.4;
7. Surrogate model: The user preference surrogate model is designed using the method in Section 3.4;
8. Population update: New individuals with user preferences are generated through a probabilistic model to form a set of items to be recommended, and the individual fitness values are estimated by using the surrogate model;
9. Recommendation list: Based on the elite selection strategy, N item individuals with higher rating are selected to generate an item recommendation list that user may be interested in;
10. Interactive evaluations: The item recommendation list is submitted to the current user for interactive evaluations. It will determine whether the algorithm termination

condition has been met. If so, the algorithm ends and outputs the result. Otherwise, the dominant group is updated according to new multi-modal user-generated contents;

11. Model Management: The accuracy of the user preference surrogate model is evaluated. If the average accuracy of the surrogate model is lower than the threshold, proceed to step 5 to timely track user preferences. Otherwise, proceed to step 8 to implement evolutionary iterative search.

End Do

End

The computational complexity of the denoising diffusion model-driven adaptive estimation of distribution algorithm integrating multi-modal data mainly consists of five parts: the vectorized representation of users' text comments, the vectorized representation of visual image features, training the user interest preference model, screening the recommended item set, and predicting items' scores. The vectorized representation of text comments and visual images are trained for offline computing. The computational complexity of training the user interest preference model is $O(|P_a| \times (n_1 + n_2 + n_3) \times m)$. The time cost of filtering the recommended item set is $O(s_u \times |X|)$, where $|X|$ is the total number of individual items in the feasible solution space. The time consumption for predicting items' scores is $O(s_u)$. Therefore, the computational complexity of the proposed algorithm is $O(|P_a| \times (n_1 + n_2 + n_3) \times m + s_u \times |X|)$.

4. Experimental Results and Analysis

4.1. Experimental Environment

To demonstrate the comprehensive performance of the proposed algorithm, a series of experiments were conducted by using the Amazon dataset^[39] from the research team of Professor Julian McAuley at the University of California, San Diego, and the common public dataset Yelp^[40] on the review website in the United States. The statistical information of each dataset is described in Table 1.

Table 1. Statistical Information of Datasets.

Dataset	#Users	#Items	#Interactions	Sparsity
Amazon- Beauty	22,363	12,101	198,502	99.927%
Amazon-Sports	35,598	18,357	296,337	99.955%
Yelp	31,668	38,048	1,561,406	99.870%

The experimental environment is configured with an Intel Core i5-4590 CPU at 3.30GHz and 4 GB RAM. The experimental platform is developed using Python 3.11. In the experiment, some evaluation indicators, such as: Root Mean Square Error (RMSE), Hits Ratio (HR), mean Average Precision (*mAP*) and Normalized Discounted Cumulative Gain (NDCG), were used to evaluate the performance of personalized search and recommendation algorithms.

4.2. Comprehensive Performance Comparison Experiments

To verify the effectiveness of the proposed algorithm in this paper, some comparative experiments are conducted with TruthSR^[41], MMSSL^[42] and TiM4Rec^[43]. The brief introduction of these comparative algorithms is as follows:

- TruthSR: Capturing the consistency of user-generated content and the complementarity of reducing noise interference, the credibility of prediction is dynamically evaluated by combining subjective and objective perspectives to conduct personalized recommendations.

- MMSSL: An interaction structure between user-item collaborative view and multi-modal semantic view is constructed through anti-perturbation enhanced data. Meanwhile, cross-modal contrastive learning is utilized to capture users' preferences and semantic commonalities for achieving the diversity of preferences.
- TiM4Rec: By introducing a time-aware structured mask matrix, the time information is efficiently integrated into a state space dual framework to conduct a time-aware Mamba recommendation algorithm.

In the experiment, some evaluation metrics, such as: $RMSE$, $HR@10$, $mAP@10$ and $NDCG@10$, are used to measure the comprehensive performance of algorithms. Each algorithm was independently run 10 times. The statistical average experimental results are shown in Table 2.

Table 2. Comparative Experimental Results (The optimal results are bolded, and Improve-x represents the performance comparison between DDM-AEDA and suboptimal algorithms.).

Algorithm	TruthSR	MMSSL	TiM4Rec	DDM-AEDA	Improve-x
Amazon- Beauty	<i>RMSE</i>	1.789	1.567	1.615	1.120 -28.53%
	<i>HR@10</i>	0.0652	0.0709	0.0838	0.0878 4.77%
	<i>mAP@10</i>	0.708	0.765	0.812	0.843 3.82%
	<i>NDCG@10</i>	0.0479	0.0487	0.0446	0.0554 13.76%
Amazon- Sports	<i>RMSE</i>	1.533	1.489	1.388	1.477 6.41%
	<i>HR@10</i>	0.0678	0.0689	0.0587	0.0786 14.08%
	<i>mAP@10</i>	0.627	0.731	0.822	0.864 5.11%
	<i>NDCG@10</i>	0.0342	0.0371	0.0401	0.0423 5.49%
Yelp	<i>RMSE</i>	1.654	1.358	1.874	1.107 -18.48%
	<i>HR@10</i>	0.0186	0.0197	0.0241	0.0297 23.24%
	<i>mAP@10</i>	0.694	0.758	0.863	0.896 3.82%
	<i>NDCG@10</i>	0.0479	0.0423	0.0342	0.0542 13.15%

By observing the experimental results, the following conclusions are drawn:

(1) In the personalized search and recommendation algorithms, the proposed DDM-AEDA algorithm has generally achieved excellent prediction accuracy and recommendation effect. The $RMSE$ values of DDM-AEDA are mostly superior to those of other comparison algorithms. For example: The $RMSE$ value of DDM-AEDA achieved the optimal value of 1.120 in Amazon-Beauty, which was 28.53% lower than that of the suboptimal MMSSL algorithm, 37.40% lower than that of TruthSR, and 30.65% lower than that of TiM4Rec. The results of these comparative experiments demonstrate that the proposed algorithm has excellent prediction accuracy. It indicates that DDM-AEDA can effectively capture users' dynamic preferences by combining the powerful modeling ability of DDPM and multi-step generative modeling. It reduces multimodal noise interference to model complex data distributions more fully to fit users' preference behaviors more accurately for guiding personalized search tasks.

(2) The proposed DDM-AEDA algorithm can sort the items to be searched better, which the arrangement order is more in line with users' interest preferences. It places the items that users are interested in at the front of the recommendation list for providing users with good search and browsing experience and achieving a better hit rate and average accuracy rate. For example, the DDM-AEDA algorithm achieved the optimal values of $HR@10$, $mAP@10$ and $NDCG@10$ in the Yelp dataset experiment by comparing with other algorithms. Among them, the $HR@10$ value of DDM-AEDA is 23.24% higher than that of TiM4Rec, which is significantly better than the suboptimal algorithm. It is 59.68% higher than TruthSR, and 50.76% higher than MMSSL. The $mAP@10$ value of DDM-AEDA is 3.82% better than that of the suboptimal TiM4Rec algorithm. It is 29.11% higher than that of TruthSR, and 18.21% higher than that of MMSSL. The $NDCG@10$ value of DDM-AEDA is 13.15% higher than that of TruthSR and is superior to the suboptimal algorithm. It is 28.13% higher than that of MMSSL, and 58.48% higher than that of TiM4Rec. Overall, DDM-AEDA integrates multi-

modal information, such as: users' historical behaviors, text, tag, image and other data, to more accurately capture the characteristics of multi-modal data when modeling users' preferences. It has enabled highly relevant items to be ranked higher to enhance the ability of personalized recommendation for achieving excellent recommendation results and user satisfaction.

In order to better demonstrate the comprehensive performance of the proposed DDM-AEDA algorithm in personalized search and recommendation tasks, DDM-AEDA was compared with the other IECs algorithms on various datasets, such as: multi-layer perceptron-driven IEDA, RBM-MSH-assisted IEDA^[44], and enhanced interactive estimation of distribution algorithm driven by dual sparse variational autoencoders integrating user-generated content (DSVAE-IEDA)^[45]. In the experiments, the personalized search and recommendation algorithms were independently run 10 times each. The average evaluation indicators are calculated to measure the comprehensive performance of each algorithm. The average experimental results are shown in Table 3.

Table 3. Comparative Experimental Results of DDM-AEDA with other IECs (The optimal results are bolded, and Improve-y represents the performance comparison between DDM-AEDA and suboptimal algorithms.).

Algorithm	MLPIEDA	RIEDA-MsH	DSVAEIEDA	DDM-AEDA	Improve-y
Amazon- Beauty	RMSE	2.200	1.584	1.261	1.120
	HR@10	0.0603	0.0750	0.0781	0.0878
	mAP@10	0.650	0.763	0.860	0.843
	NDCG@10	0.0398	0.0330	0.0458	0.0554
Amazon- Sports	RMSE	1.585	1.394	1.751	1.477
	HR@10	0.0705	0.0569	0.0766	0.0786
	mAP@10	0.669	0.711	0.894	0.864
	NDCG@10	0.0361	0.0356	0.0388	0.0423
Yelp	RMSE	1.532	1.723	1.152	1.107
	HR@10	0.0138	0.0176	0.0209	0.0297
	mAP@10	0.657	0.723	0.854	0.896
	NDCG@10	0.0407	0.0396	0.0498	0.0542

By observing the above experimental results, the following conclusions are drawn:

(1) The proposed DDM-AEDA algorithm makes full use of multi-modal user-generated contents to build user preference model and surrogate model based on users' preferences. It can generate new evolutionary individuals with users' interest preferences and estimates the fitness values of new item individuals, which guides the personalized search and recommendation process under the interactive evolutionary optimization framework. In the personalized search experiments on various datasets, DDM-AEDA achieved excellent prediction accuracy and recommendation effect by comparing with other algorithms. For example: The average evaluation indicator of DDM-AEDA achieved the optimal value in Yelp. Among them, the average RMSE of DDM-AEDA is 3.91% better than that of the suboptimal algorithm (DSVAEIEDA), and respectively 27.74% and 35.75% lower than those of MLPIEDA and RIEDA-MsH. The average HR@10 of DDM-AEDA is 42.11% higher than that of the suboptimal algorithm (DSVAEIEDA), and respectively 115.22% and 68.75% higher than those of MLPIEDA and RIEDA-MsH. The average mAP@10 of DDM-AEDA is 4.92% higher than that of the suboptimal algorithm (DSVAEIEDA), and respectively 36.38% and 23.93% higher than those of MLPIEDA and RIEDA-MsH. The average NDCC@10 of DDM-AEDA is 8.84% higher than that of the suboptimal algorithm (DSVAEIEDA), and respectively 33.17% and 36.87% higher than those of MLPIEDA and RIEDA-MsH. Although the proposed algorithm did not achieve the optimal values on some evaluation indicators of certain datasets, it still obtained good comprehensive search effects and recommendation results.

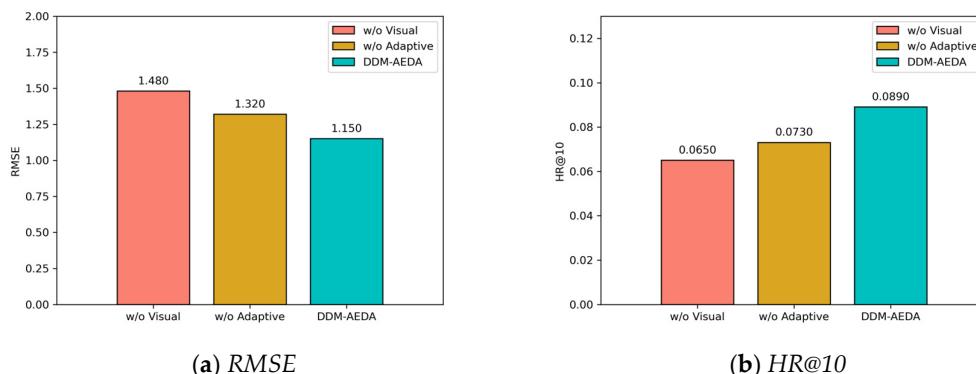
(2) In the comparative experiments of various datasets, DDM-AEDA is generally superior to other algorithms. It demonstrates the feasibility and effectiveness of the proposed algorithm, and it has excellent prediction accuracy, search efficiency and recommendation accuracy. These

experimental results show that the proposed algorithm efficiently integrates different information sources through an interactive adaptive optimization strategy in the IEDA evolutionary optimization framework to accurately models users' interest preferences. It is conducive to the generation of excellent evolutionary individuals to ensure the relative ranking relationship of dominant groups. The extracted knowledge information of excellent solutions is fully utilized to generate new items that meet users' needs and preferences to select excellent individuals and successfully retain them for the next generation of populations. It will guide the personalized evolutionary search process to avoid getting trapped in local optima during the optimization process. At the same time, the surrogate model based on user preferences is used to predict items' ratings to rank the excellent solutions preferred by users at the top of the recommendation list. The items that users like are selected to conduct item recommendations for quickly searching user satisfaction solutions. It improves the rating prediction ability and recommendation effect of personalized search and recommendation algorithms.

In summary, in personalized search and recommendation tasks, DDM-AEDA fully exploits multi-modal user-generated contents to construct a user interest preference model based on denoising diffusion model to extract users' preferences, which effectively explores the deep-seated potential user preference features and their development evolution pattern to improve the fitting accuracy of the user interest preference model. An adaptive EDA probabilistic model and a surrogate model based on the user preferences are established to utilize the IECs evolutionary optimization framework to enhance the interactive personalized evolutionary search process, which improves the prediction accuracy of the user evaluation surrogate model. Meanwhile, it guides the forward direction of personalized evolutionary search based on objective assessment indicators such as user experience and feedback evaluation, which helps users search for their satisfactory solutions to achieve a good balance between recommendation effect and search efficiency. It improves the optimization ability, search efficiency and recommendation effectiveness of personalized search algorithms, and has good stability and scalability. It is adapted to the actual needs of users' personalized search and recommendation in complex multi-modal data scenarios.

4.3. Ablation Experiments

The proposed algorithm DDM-AEDA in this paper contains several important modules, such as: multimodal feature fusion and adaptive estimation of distribution strategy. Ablation experiments were conducted here to verify the effectiveness of these modules. The experimental results on Amazon- Beauty are shown in Figure 3. Among them, the removed visual feature input module is denoted as w/o Visual, and the removed adaptive strategy module is recorded as w/o Adaptive.



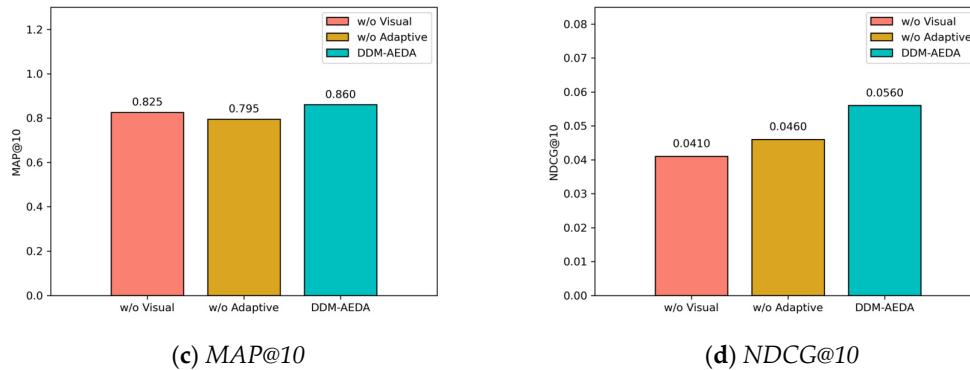


Figure 3. Results of the ablation experiments.

The above experimental results show that it significantly reduces the comprehensive performance of personalized search and recommendation algorithms after DDM-AEDA removes the visual feature input module. The visual feature input section of the multi-modal information processing module in DDM-AEDA enhances the richness of the item representation by extracting the visual information of items' image (such as: color, texture, shape, etc.). It helps the user preference model capture users' visual preferences, which is particularly crucial in scenarios with strong visual dependence (such as: beauty, clothing, etc.). However, the visual feature information was removed in the w/o Visual experiment, which causes the model to be unable to distinguish items with similar appearances (for example: lipsticks of different shades or similar sports shoes). It will significantly reduce the recommendation hit rate and ranking quality. Therefore, the visual feature input part in DDM-AEDA plays an important role in improving the recommendation accuracy and user satisfaction of personalized search and recommendation algorithms.

On the other hand, it also reduces the comprehensive performance of personalized search and recommendation algorithms after DDM-AEDA removes the adaptive distribution estimation strategy module. In the interactive estimation of distribution framework, DDM-AEDA can dynamically adjust the selection ratio (sr) and covariance scaling parameter (cs) through adaptive strategies, and combines the historical information of the archive set to achieve dynamic balance in the searching process. In the w/o Adaptive experiment, the selection ratio and covariance scaling parameters were fixed and the archive set update mechanism was turned off, which led to the model being unable to dynamically adjust the search direction based on users' feedback. It significantly reduces the diversity and convergence speed of the recommendation results. Therefore, the adaptive estimation of distribution strategy part in DDM-AEDA can improve the search efficiency and recommendation effect of personalized search and recommendation algorithms.

4.4. Hyperparameter Sensitivity Analysis

Some hyperparameters in the proposed DDM-AEDA algorithm have a crucial impact on the search efficiency and recommendation performance of the personalized search and recommendation algorithms. Therefore, some hyperparameter sensitivity analysis experiments are conducted here to evaluate the important hyperparameters in the proposed algorithm, such as: selection ratio (sr) and covariance scaling parameter (cs).

In experiments, sr is set as $sr \in \{0.1, 0.2, 0.3, 0.4\}$. cs is set as $cs \in \{0.3, 0.6, 0.9\}$ when $sr = 0.1$; cs is set as $cs \in \{0.2, 0.4, 0.6, 0.8\}$ when $sr = 0.2$; cs is set as $cs \in \{0.3, 0.6, 0.9\}$ when $sr = 0.3$; cs is set as $cs \in \{0.4, 0.8\}$ when $sr = 0.4$. The evaluation metrics HR and $NDCG$ are used. The experimental results of hyperparameter sensitivity analysis experiments on Amazon-Beauty are shown in Figure 4.

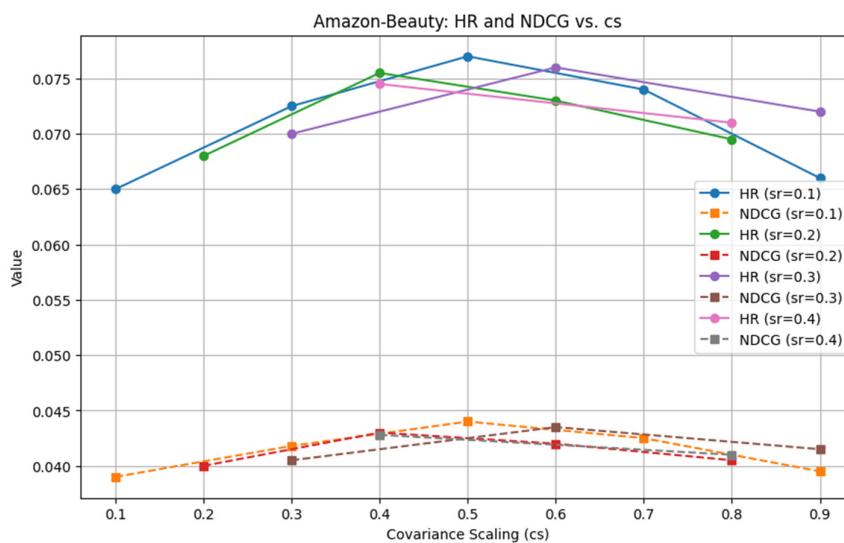


Figure 4. Hyperparameter sensitivity analysis experiments in Amazon- Beauty.

It can be seen from the experimental results that when sr is fixed, HR and $NDCG$ first rise and then fall with the increasing of cs . It shows a smooth curve trend. It indicates that a moderate covariance scaling parameter can bring the optimal recommendation effect. The large cs has a wide sampling range, which is conducive to diversity search in the early stage but not to concentrated search in the later stage to obtain the optimal solutions. Too small cs has a small sampling range. Although it is conducive to centralized search in the later stage, it is prone to fall into local optimum in the early stage for resulting in premature convergence. Therefore, $sr=0.2$ and $cs=0.4$ are selected in the proposed DDM-AEDA algorithm.

5. Conclusions

This paper focuses on personalized search and recommendation tasks in complex network environments. A denoising diffusion model-driven adaptive estimation of distribution algorithm integrating multi-modal data is proposed by integrating user interest preference modeling and the IECs algorithm based on surrogate model. Comprehensively considering multi-modal user-generated contents, a user interest preference model based on denoising diffusion model is constructed to extract users' preference features. A surrogate model based on user preferences is established in the framework of the estimation of distribution algorithm, and adaptive estimation of distribution operators and strategies are designed to guide the personalized evolutionary search process. At the same time, based on new user-generated contents and model dynamic management mechanism, the constructed relevant models are updated dynamically to track the changes in users' interest preferences in a timely manner for efficiently completing personalized search tasks. Through many experiments and systematic demonstrations, it has been verified that the proposed algorithm in this paper shows significant advantages in functional completeness, prediction accuracy and decision transparency. In future research work, it is planned to explore issues, such as: the effective utilization of multi-modal data, computational efficiency in personalized recommendation algorithms, system security, and user privacy protection. It will enhance the comprehensive performance of personalized search and recommendation algorithms to provide all-round services that are intelligent, exclusive, and secure.

Author Contributions: Conceptualization, Lin Bao and Hang Yang; methodology, Lin Bao and Hang Yang; software, Lina Wang and Hang Yang; validation, Lina Wang and Hang Yang; formal analysis, Lina Wang and Yumeng Peng; investigation, Biao Xu; resources, Biao Xu; data curation, Yumeng Peng; writing—original draft preparation, Lin Bao and Hang Yang; writing—review and editing, Lin

Bao and Lina Wang; visualization, Lin Bao and Lina Wang; supervision, Lin Bao and Biao Xu; project administration, Lin Bao; funding acquisition, Biao Xu. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Natural Science Foundation of Guangdong Province of China (2023B1515120020, 2024A1515012450). This work was supported by National Natural Science Foundation of China under grants No. 61876184.

References

1. Zou F, Chen D, Xu Q, et al. A two-stage personalized recommendation based on multi-objective teaching-learning-based optimization with decomposition [J]. *Neurocomputing*, 2021, 452: 716-727.
2. Raza S, Ding C. News recommender system: a review of recent progress, challenges, and opportunities [J]. *Artificial Intelligence Review*, 2022, 55(1): 749-800.
3. Qi L, Lin W, Zhang X, et al. A correlation graph based approach for personalized and compatible web apis recommendation in mobile app development [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2022, 35(6): 5444-5457.
4. Vullam N, Vellela S S, Reddy V, et al. Multi-agent personalized recommendation system in e-commerce based on user [C]. In *Proceedings of the 2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, Salem, India, 2023: 1194-1199.
5. Liang K, Liu H, Shan M, et al. Enhancing scenic recommendation and tour route personalization in tourism using UGC text mining [J]. *Applied Intelligence*, 2024, 54(1): 1063-1098.
6. Cai D, Qian S, Fang Q, et al. Heterogeneous graph contrastive learning network for personalized micro-video recommendation [J]. *IEEE Transactions on Multimedia*, 2022, 25: 2761-2773.
7. El-Kishky A, Markovich T, Park S, et al. Twhin: Embedding the twitter heterogeneous information network for personalized recommendation [C]. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*. 2022: 2842-2850.
8. Hui B, Zhang L, Zhou X, et al. Personalized recommendation system based on knowledge embedding and historical behavior [J]. *Applied Intelligence*, 2022, 52(1): 954-966.
9. Mu Y, Wu Y. Multimodal movie recommendation system using deep learning [J]. *Mathematics*, 11 (4): 895[EB/OL].(2023)
10. Linden G, Smith B, York J. Amazon. com recommendations: Item-to-item collaborative filtering [J]. *IEEE Internet computing*, 2003, 7(1): 76-80.
11. Wang J, De Vries A P, Reinders M J T. Unifying user-based and item-based collaborative filtering approaches by similarity fusion [C]. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, ACM, 2006: 501-508.
12. Rendle S, Freudenthaler C, Gantner Z, et al. BPR: Bayesian personalized ranking from implicit feedback [C]. In *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence*, 2009: 452-461.
13. Qiu H, Liu Y, Guo G, et al. BPRH: Bayesian personalized ranking for heterogeneous implicit feedback [J]. *Information Sciences*, 2018, 453: 80-98.
14. Rendle S. Factorization machines with libFM [J]. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2012, 3(3): 1-22.
15. Pang G, Wang X, Hao F, et al. ACNN-FM: A novel recommender with attention-based convolutional neural network and factorization machines [J]. *Knowledge-Based Systems*, 2019, 181: 104786.
16. He X, Liao L, Zhang H, et al. Neural collaborative filtering [C]. In *Proceedings of the 26th international conference on world wide web*. 2017: 173-182.
17. He R, McAuley J. VBPR: visual bayesian personalized ranking from implicit feedback [C]. In *Proceedings of the 30th AAAI conference on artificial intelligence*. 2016, 30(1): 144-150.
18. Kim D, Park C, Oh J, et al. Convolutional matrix factorization for document context-aware recommendation [C]. In *Proceedings of the 10th ACM conference on recommender systems*. 2016: 233-240.
19. Chen J, Zhang H, He X, et al. Attentive collaborative filtering: Multimedia recommendation with item-and component-level attention [C]. In *Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval*, Shinjuku Tokyo: Association for Computing Machinery, 2017: 335-344.

20. Wei Y, Wang X, Nie L, et al. MMGCN: Multi-modal graph convolution network for personalized recommendation of micro-video [C]. In *Proceedings of the 27th ACM international conference on multimedia. Nice: Association for Computing Machinery*, 2019: 1437-1445.
21. Wang Z, Chen H, Li Z, et al. VRConvMF: Visual recurrent convolutional matrix factorization for movie recommendation [J]. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2022, 6(3): 519-529.
22. Yang M, Zhou P, Li S, et al. Multi-head multimodal deep interest recommendation network [J]. *Knowledge-Based Systems*, 2023, 276: 110689.
23. Deng J, Wang S, Wang Y, et al. MMBe: Live Streaming Gift-Sending Recommendations via Multi-Modal Fusion and Behaviour Expansion [C]. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2024: 4896-4905.
24. Yan B, Chen S, Jia S, et al. MIM: Multi-modal Content Interest Modeling Paradigm for User Behavior Modeling [J]. arxiv preprint arxiv:2502.00321, 2025.
25. Ho J, Salimans T, Gritsenko A, et al. Video diffusion models [J]. *Advances in neural information processing systems*, 2022, 35: 8633-8646.
26. Croitoru F A, Hondu V, Ionescu R T, et al. Diffusion models in vision: A survey [J]. *IEEE transactions on pattern analysis and machine intelligence*, 2023, 45(9): 10850-10869.
27. Chen S, Sun P, Song Y, et al. Diffusiondet: Diffusion model for object detection [C]. In *Proceedings of the IEEE/CVF international conference on computer vision*. 2023: 19830-19843.
28. Cao H, Tan C, Gao Z, et al. A survey on generative diffusion models [J]. *IEEE transactions on knowledge and data engineering*, 2024, 36(7): 2814-2830.
29. Chen H, Zhang Y, Cun X, et al. Videocrafter2: Overcoming data limitations for high-quality video diffusion models [C]. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024: 7310-7320.
30. Wang W, Xu Y, Feng F, et al. Diffusion recommender model [C]. In *Proceedings of the 46th international ACM SIGIR conference on research and development in information retrieval*. 2023: 832-841.
31. Yang Z, Wu J, Wang Z, et al. Generate what you prefer: Reshaping sequential recommendation via guided diffusion [J]. *Advances in Neural Information Processing Systems*, 2023, 36: 24247-24261.
32. Li Z, Xia L, Huang C. Recdiff: Diffusion model for social recommendation [C]. In *Proceedings of the 33rd ACM international conference on information and knowledge management*. 2024: 1346-1355.
33. Wei T R, Fang Y. Diffusion Models in Recommendation Systems: A Survey [J]. arxiv preprint arxiv:2501.10548, 2025.
34. Li Z, Sun A, Li C. Diffurec: A diffusion model for sequential recommendation [J]. *ACM Transactions on Information Systems*, 2023, 42(3): 1-28.
35. Zhao J, Wang W, Xu Y, et al. Denoising diffusion recommender model [C]. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '24)*. Association for Computing Machinery, New York, NY, USA, 2024: 1370-1379.
36. Jiang Y, Yang Y, Xia L, et al. DiffKG: Knowledge graph diffusion model for recommendation [C]. In *Proceedings of the 17th ACM international conference on web search and data mining (WSDM '24)*. Association for Computing Machinery, New York, NY, USA, 2024: 313-321.
37. Cui Z, Wu H, He B, et al. Context Matters: Enhancing Sequential Recommendation with Context-aware Diffusion-based Contrastive Learning [C]. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*. 2024: 404-414.
38. Xia R, Cheng Y, Tang Y, et al. S-diff: An anisotropic diffusion model for collaborative filtering in spectral domain [C]. In *Proceedings of the Eighteenth ACM International Conference on Web Search and Data Mining (WSDM '25)*. Association for Computing Machinery, New York, NY, USA, 2025: 70-78.
39. Yue Z, Wang Y, He Z, et al. Linear recurrent units for sequential recommendation [C]. In *Proceedings of the 17th ACM international conference on web search and data mining*. 2024: 930-938.
40. Wu J, Wang X, Feng F, et al. Self-supervised graph learning for recommendation [C]. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*. 2021: 726-735.

41. Yan M, Huang H, Liu Y, et al. Truthsr: trustworthy sequential recommender systems via user-generated multimodal content [C]. In *International Conference on Database Systems for Advanced Applications*. Singapore: Springer Nature Singapore, 2024: 180-195.
42. Wei W, Huang C, Xia L, et al. Multi-modal self-supervised learning for recommendation [C]. In *Proceedings of the ACM web conference 2023*. 2023: 790-800.
43. Fan H, Zhu M, Hu Y, et al. TiM4Rec: An efficient sequential recommendation model based on time-aware structured state space duality model [J]. *Neurocomputing*, 2025: 131270.
44. Bao L, Sun X, Gong D, et al. Multisource heterogeneous user-generated contents-driven interactive estimation of distribution algorithms for personalized search [J]. *IEEE Transactions on Evolutionary Computation*, 2021, 26(5): 844-858.
45. Yang H, Bao L, Sun X, et al. Dual Sparse Variational Autoencoder-driven Interactive Estimation of Distribution Algorithm with User Generated Contents [C]. In *Proceedings of the 2024 International Conference on New Trends in Computational Intelligence (NTCI)*, Qingdao, China, 2024: 172-177.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.