修　士　論　文

A Research on Cross-Domain Recommendation under Non-Overlapping Settings by Integrating Gaussian Mixture User Preference Distributions and Optimal Transport

非重複設定下におけるガウス混合モデルユーザ嗜好分布と最適輸送を統合したクロスドメイン推薦に関する研究

指導教員　鈴村 豊太郎 教授

東京大学大学院 情報理工学系研究科
電子情報学専攻

氏 名　48-246453 肖　子吟

提 出 日　2026 年 1 月 22 日

Abstract

Recommendation systems are essential tools for helping users discover items of interest in various domains, such as e-commerce, streaming services, and social media. With the rapid growth of online platforms, the need for effective recommendation systems has become increasingly important.

However, traditional recommendation systems often face challenges such as data sparsity and cold-start problems, which can limit their effectiveness. Cross-domain recommendation has emerged as a promising approach to address these challenges by leveraging information from multiple domains to improve recommendation accuracy. Nevertheless, existing cross-domain recommendation methods often assume overlapping users or items between domains, which may not hold in many real-world scenarios. Also, they usually represent user preferences as point estimates, which may not capture the full complexity of user behavior.

To address these limitations, in this thesis, we propose a novel framework that models user preferences as Gaussian Mixture Models (GMMs) and integrates optimal transport techniques for cross-domain recommendation under non-overlapping settings. By representing user preferences as distributions, our approach can capture the uncertainty and diversity of user behavior more effectively and fine-grainedly. The use of optimal transport allows us to align user preference distributions across domains, facilitating knowledge transfer even in the absence of overlapping users or items.

We evaluate our proposed method on Amazon datasets across multiple domains, comparing its performance against state-of-the-art cross-domain recommendation techniques and traditional single-domain methods. We also conduct ablation studies to assess the effectiveness of representing user preferences as distributions and the impact of boosting target domain recommendation performance using source domain information. The experimental results demonstrate that our approach significantly outperforms existing methods in terms of recommendation accuracy, highlighting the effectiveness of modeling user preferences as distributions and leveraging optimal transport for cross-domain knowledge transfer.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

These days, information overload has become a significant challenge for users in various online platforms, such as e-commerce websites, streaming services, and social media[1, 2, 3]. Recommender systems play a crucial role in addressing this challenge by providing personalized guidance to users by selecting items with the highest predicted utility from a large space of options.[4, 5, 6]. By analyzing user-item interactions, recommender systems can help users discover relevant content, products, or services, thereby enhancing user satisfaction and engagement. Besides improving user experience, recommender systems also benefit businesses by increasing sales, customer retention, and overall platform activity[7]. As the volume of online content continues to grow, the importance of effective recommender systems becomes even more pronounced. Despite their significance, recommender systems face several challenges that can hinder their performance. One of the primary challenges is data sparsity, where users interact with only a small fraction of the available items, making it difficult to accurately model user preferences. This issue is particularly prevalent in new or niche platforms with limited user interactions. Another challenge is the cold-start problem, which occurs when new users or items are introduced to the system. In such cases, there is insufficient data to make reliable recommendations, leading to suboptimal user experiences. Both data sparsity and cold-start problems are caused by the limited availability of user-item interaction data, which can significantly impact the effectiveness of recommender systems. To address these challendges, cross-domain recommendation has emerged as a promising approach that leverages information from multiple domains to improve recommendation accuracy. Cross-domain recommendation aims to enhance the performance of recommender systems by transferring knowledge from a source domain to a target domain. This approach is particularly useful in scenarios where the target domain suffers from data sparsity or cold-start issues. By utilizing user-item interactions from a related source domain, cross-domain recommendation can help alleviate these challenges and provide more accurate recommendations. However, existing cross-domain recommendation methods often assume overlapping users or items between the source and target domains, and rely on these overlaps to build connections between the two domains in the training stage. However, in many real-world scenarios, such overlaps may not exist, limiting the applicability of these methods. Additionally, traditional recommender systems typically represent user preferences as point estimates, which may not capture the full complexity and uncertainty of user behavior. To overcome these limitations, there is a need for novel approaches that can effectively model user preferences and facilitate knowledge transfer between non-overlapping domains during the training process. In this thesis, we propose a novel framework that models user preferences as Gaussian Mixture Models (GMMs) and integrates optimal transport techniques for cross-domain recommendation under non-overlapping settings. By representing user preferences as distributions, our approach can capture the uncertainty and diversity of user behavior more effectively. The use of optimal transport allows us to align user preference distributions across domains, facilitating knowledge transfer even in the absence of overlapping users or items. The proposed framework consists of two main components: (1) modeling user preferences as GMMs to capture the complexity of user behavior, and (2) employing optimal transport to align these distributions between the source and target domains. This combination enables our method to leverage information from the source domain to enhance recommendation accuracy in the target domain, even when there are no overlapping users or items.

1.2 Contributions

The main contributions of this thesis are as follows:

- We introduce a novel framework for cross-domain recommendation under non-overlapping settings by modeling
- We propose representing user preferences as Gaussian Mixture Models (GMMs) to capture the uncertainty and diversity of user behavior more effectively.
- We integrate optimal transport techniques to align user preference distributions across domains, enabling knowledge transfer
- We conduct extensive experiments on Amazon datasets across multiple domains to evaluate the performance of our proposed method. The results demonstrate that our approach significantly outperforms existing cross-domain recommendation techniques and traditional single-domain methods in terms of recommendation accuracy.
- We perform ablation studies to assess the effectiveness of representing user preferences as distributions and the impact of boosting target domain recommendation performance using source domain information.
- We provide insights into the benefits of modeling user preferences as distributions and leveraging optimal transport for cross-domain knowledge transfer, highlighting the potential of our approach for addressing challenges in recommender systems.

1.3 Thesis Organization

The remainder of this thesis is organized as follows:

- In Chapter 2, we provide an overview of the fundamental concepts and techniques relevant to this research, including recommender systems, cross-domain recommendation, Gaussian Mixture Models, and optimal transport.
- In Chapter 3, we review related work on recommender systems, cross-domain recommendation, and optimal transport techniques.
- In Chapter 4, we present our proposed framework for cross-domain recommendation under non-overlapping settings, detailing the modeling of user preferences as GMMs and the integration of optimal transport for knowledge transfer.
- In Chapter 5, we describe the experimental setup, datasets, and evaluation metrics used to assess the performance of our proposed method.
- In Chapter 6, we present and discuss the experimental results, including comparisons with existing methods and ablation studies.
- Finally, in Chapter 7, we summarize the main findings of this thesis and discuss potential directions for future research.

Chapter 2

Preliminaries

2.1 Recommender Systems

Recommender systems can be mathmatically modeled as a function $R : U \times I \rightarrow S$, where U is the set of users, I is the set of items, and S is the set of possible scores or ratings. The goal of a recommender system is to predict the score $s \in S$ that a user $u \in U$ would give to an item $i \in I$. This prediction can be represented as $\hat{s} = R(u, i)$. The final task of recommender systems is to generate a ranked list of items for each user based on the predicted scores.

2.2 Cross-Domain Recommendation

Cross-domain recommendation aims to leverage user preferences and behaviors from one or more source domains to improve recommendation performance in a target domain. Formally, let D_s be the source domain with user set U_s and item set I_s, and D_t be the target domain with user set U_t and item set I_t. The objective is to learn a recommendation function $R_t : U_t \times I_t \rightarrow S$ for the target domain by utilizing information from the source domain(s) D_s. The infromation that can be transferred from the source domain to the target domain includes user-item interactions, user profiles, item attributes, and latent factors learned from the source domain. The challenge in cross-domain recommendation lies in effectively transferring knowledge while addressing issues such as domain heterogeneity, data sparsity, and cold-start problems.

2.3 Gaussian Mixture Models

A Gaussian Mixture Model (GMM) is a probabilistic model that assumes data is generated from a mixture of several Gaussian distributions. Formally, a GMM can be represented as:

$$p(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x|\mu_k, \Sigma_k) \tag{2.1}$$

where K is the number of Gaussian components, π_k are the mixture weights satisfying $\sum_{k=1}^{K} \pi_k = 1$, and $\mathcal{N}(x|\mu_k, \Sigma_k)$ is the Gaussian distribution with mean μ_k and covariance matrix Σ_k.

2.4 Optimal Transport

Optimal transport is a mathematical framework for comparing and transforming probability distributions. Given two probability distributions μ and ν defined on spaces X and Y, respectively, the optimal transport problem seeks to find a mapping $T : X \to Y$ that minimizes the cost of transporting mass from μ to ν. The cost function $c(x, y)$ quantifies the expense of moving mass from point $x \in X$ to point $y \in Y$. The optimal transport problem can be formulated as:

$$\min_T \int_X c(x, T(x)) d\mu(x) \tag{2.2}$$

subject to the constraint that the pushforward measure $T_{\#}\mu = \nu$. Optimal transport has been widely used in various applications, including image processing, machine learning, and economics, due to its ability to capture the geometric structure of probability distributions. In our proposed method, we are not using optimal transport in the standard sense above, but using a related concept called Wasserstein distance to measure the distance between two probability distributions. The Wasserstein distance is defined as:

$$W_p(\mu, \nu) = \left(\inf_{\gamma \in \Pi(\mu,\nu)} \int_{X \times Y} c(x, y)^p d\gamma(x, y) \right)^{1/p} \tag{2.3}$$

where $\Pi(\mu, \nu)$ is the set of all joint distributions (couplings) with marginals μ and ν, and $c(x, y)$ is the cost function. The Wasserstein distance provides a meaningful way to compare probability distributions, taking into account the underlying geometry of the data. With the distance between two probability distributions defined, we can replace the conventional pointwise distance (e.g., Euclidean distance) between vectors with the Wasserstein distance between the corresponding distributions in our proposed method.

Chapter 3

Related Work

3.1 Recommender Systems

This section reviews the existing literature on recommender systems, categorizing the related work into several key areas based on the techniques and approaches employed. Each subsection delves into specific methodologies and advancements within the field.

3.1.1 Traditional Recommender Systems

In the early stages of recommender systems, traditional techniques such as collaborative filtering, content-based filtering, and hybrid methods were predominantly used. Collaborative filtering (CF) is one of the foundational approaches in recommender systems, operating on the principle that users with similar preferences will like similar items, or that items liked by similar users will be preferred by a given user [8, 9].

CF can be divded into memory-based and model-based methods based on how recommendations are generated. Memory-based methods utilize user-item interaction data directly to compute similarities between users or items, while model-based methods employ machine learning algorithms to learn latent factors from the interaction data.

3.1.1.1 Memory-based CF

By user or item similarities, memory-based CF can be further categorized into user-based and item-based approaches. User-based CF recommends items to a user based on the preferences of similar users, while item-based CF suggests items similar to those the user has previously liked.

GroupLens [10] is one of the earliest and most influential memory-based CF systems, which introduced user-based collaborative filtering using Pearson correlation to compute user similarities. It demonstrated the effectiveness of CF in providing personalized recommendations and laid the groundwork for subsequent research in the field.

Sarwar et al. [11] points out that when the number of users and items is very large, user-based CF can be computationally expensive and the sparsity of the user-item interaction matrix can lead to poor recommendation quality. Noticing that users' preferences change quickly over time but items' characteristics are relatively stable, they proposed an item-based CF approach that computes item similarities based on user interactions. The merit of this method is that item similarities can be precomputed and stored, allowing for efficient recommendation generation. And the number of items is usually much smaller than the number of users, which helps alleviate the data sparsity issue. Also, one important attribution of this paper is that it proposed a new similarity measure called adjusted cosine similarity, which accounts for individual user rating biases when computing item similarities. This method has since become a standard technique in item-based CF and has been widely adopted in various recommender systems.

3.1.1.2 Model-based CF

Model-based CF methods utilize machine learning techniques to learn latent representations of users and items from interaction data, by fitting parametric models such as matrix factorization, probabilistic latent factor models, or neural networks. These models capture underlying preference patterns in a low-dimensional latent space, enabling generalization to unseen user – item pairs and alleviating data sparsity.

Breese et al. [12] conducted a comprehensive analysis of both memory-based and model-based CF methods. This paper is the first to systematically distinguish between memory-based and model-based CF approaches, providing a detailed comparison of their strengths and weaknesses. The model-based methods mentioned in this paper include Bayesian Clustering and Bayesian Networks. Bayesian Clustering groups users into clusters based on their preferences, while Bayesian Networks model the probabilistic relationships between users and items.

Ungar and Foster [13] pointed out that traditional clustering-based collaborative filtering methods suffer from instability and poor generalization when interaction data are highly sparse, as approaches based on KNN or simple K-means clustering rely heavily on local similarity patterns. To address this issue, they proposed Gibbs clustering for collaborative filtering, a probabilistic co-clustering approach that jointly clusters users and items into latent classes and models their interactions through class-level link probabilities. By explicitly formulating a generative model and employing Gibbs sampling for inference, their method enforces global consistency in user and item assignments and provides a principled alternative to heuristic clustering.

Koren et al. [14] systematically review matrix factorization techniques for recommender systems, demonstrating that latent factor models with bias, implicit feedback, and temporal dynamics achieve consistently superior accuracy and scalability over neighborhood-based methods, and establishing matrix factorization as a dominant model-based collaborative filtering paradigm.

Salakhutdinov and Mnih introduce Probabilistic Matrix Factorization as a scalable latent factor model for large, sparse recommender systems, and further extend it to a fully Bayesian framework using MCMC, significantly improving robustness and generalization—especially for infrequent users—while establishing PMF as a foundational model-based collaborative filtering paradigm [15, 16].

While most model-based collaborative filtering methods represent user preferences as point embeddings in a latent space, several approaches instead characterize user preferences in a probabilistic manner. Since our proposed method also adopts a distributional representation of user preferences, we briefly review related works along this line.

Hofmann [17] proposed probabilistic latent semantic analysis (PLSA) for collaborative filtering, formulating user – item interactions as a latent class mixture model. In PLSA, each user is associated with a probability distribution over latent topics (or communities), and each interaction is generated by first sampling a latent topic and then drawing an item conditioned on that topic. As a result, user preferences are represented as distributions over latent semantic factors, allowing different interactions of the same user to be explained by different latent causes, rather than being tied to a single latent representation.

Marlin [18] proposed the User Rating Profile (URP) model, a probabilistic latent variable approach for rating-based collaborative filtering that explicitly models uncertainty in user preferences. URP represents each user as a mixture over latent user attitudes, where the mixture proportions are drawn from a Dirichlet distribution. For each item, a latent attitude is sampled and the corresponding rating is generated according to an attitude-specific rating distribution. By modeling users as distributions over latent preference patterns rather than fixed point representations, URP enables different items rated by the same user to be explained by different latent factors and allows direct inference of rating distributions for unseen items.

Blei, Ng, and Jordan [19] proposed Latent Dirichlet Allocation (LDA), a hierarchical generative probabilistic model that represents each document as a mixture over latent topics, where the topic proportions are drawn from a Dirichlet prior. By introducing a document-level latent variable, LDA provides a fully generative framework that enables principled inference for previously unseen data. Although originally developed for text modeling, LDA has been extended to collaborative filtering by treating users as documents and items as words. Under this formulation, user preferences are modeled as probability distributions over latent topics, allowing each user – item interaction to be explained by different latent factors. This distributional representation enables LDA to capture heterogeneous user interests more flexibly than single-vector latent representations.

3.1.1.3 Content-based Filtering

Content-based filtering (CBF) recommends items to users by modeling user preferences from the attributes of items they have previously interacted with. Typically, both users and items are represented in a shared feature space, where recommendations are generated based on the similarity between user profiles and item representations. Since CBF relies solely on individual user history, it is less affected by user – user interaction sparsity but often suffers from limited diversity and difficulty in capturing evolving or complex user interests.

Salton et al. [20] proposed the vector space model for information retrieval, in which both documents and queries are represented as weighted term vectors. By assigning importance weights to terms (e.g., inverse document frequency) and computing similarity scores between query and document vectors, the model enables ranked retrieval based on relevance. This representation and similarity-matching paradigm laid the foundation for content-based recommender systems, where user profiles and item content are similarly modeled in a shared feature space.

Pazzani and Billsus [21] proposed a content-based recommender system that learns user profiles from explicit user feedback on item content. Their method represents items using content features and employs a naive Bayesian classifier to incrementally learn and revise user preference profiles, enabling the system to predict the interestingness of unseen items.

3.1.1.4 Hybrid Methods

Hybrid recommender systems combine both collaborative filtering and content-based filtering techniques to leverage the strengths of each approach and mitigate their respective weaknesses. By jointly exploiting user – item interaction patterns and item content information, hybrid methods can alleviate issues such as data sparsity and cold-start that commonly affect pure collaborative filtering models. These approaches typically integrate multiple signals at different stages of the recommendation pipeline, resulting in more robust and accurate predictions.

Burke [5] presented a comprehensive survey of hybrid recommender systems, systematically categorizing hybridization strategies such as weighted, switching, mixed, and feature combination approaches. The survey analyzed how different hybrid designs integrate multiple recommendation techniques to balance their respective strengths and weaknesses, demonstrating that hybrid methods can effectively improve recommendation accuracy and alleviate issues such as data sparsity and cold-start problems.

Pazzani [22] proposed a unified framework for integrating collaborative, content-based, and demographic filtering methods in recommender systems. By exploiting multiple sources of information, including user – item interactions, item content, and user profiles, the framework combines recommendations from different models to improve precision. Experimental results demonstrated that hybrid approaches within this framework consistently outperform single-method recommenders.

Melville et al. [23] proposed a content-boosted collaborative filtering (CBCF) framework that integrates content-based prediction into the collaborative filtering process. Specifically, a content-based predictor is first used to generate pseudo ratings for unrated items, producing a dense pseudo user – item matrix on which collaborative filtering is subsequently applied. By alleviating sparsity and the first-rater problem, this approach achieves significantly improved recommendation accuracy compared to pure collaborative, pure content-based, and naive hybrid methods.

Billsus et al. [24] developed a hybrid news recommender system for adaptive news access that integrates collaborative filtering and content-based filtering techniques. Their system learns personalized user models from both explicit and implicit user feedback, and combines short-term and long-term interest representations to adapt to users' evolving information needs. Deployed in a real-world news delivery environment, this work demonstrated the practical effectiveness of hybrid recommender systems in improving personalization quality without requiring additional user effort.

3.1.2 Supervised Machine Learning in Recommender Systems

Supervised machine learning techniques have also been widely applied to recommender systems to enhance recommendation accuracy. By formulating recommendation as a regression or classification problem, these methods learn predictive models from labeled user – item interaction data using features such as user demographics, item attributes, and contextual information. Algorithms including decision trees, support vector machines, and ensemble methods have been employed to capture complex relationships between features and user preferences. While effective in leveraging rich side information, supervised learning-based approaches often rely heavily on feature engineering and struggle to generalize under sparse interaction settings.

Basilico and Hofmann [25] proposed a unified supervised learning framework that integrates collaborative filtering and content-based filtering within a single prediction model. Their approach formulates recommendation as a learning problem over user – item pairs by designing joint feature representations and kernel functions that enable simultaneous generalization across both user and item dimensions. By incorporating user – item interaction data together with item and user attributes, the framework achieves improved recommendation accuracy compared to traditional collaborative or content-based methods.

Rendle [26] introduced Factorization Machines (FMs), a supervised learning model that generalizes matrix factorization by modeling pairwise feature interactions through factorized parameters. By representing user – item interactions, item attributes, and contextual information as sparse feature vectors, FMs can efficiently capture interactions in high-dimensional and highly sparse settings. This unified formulation subsumes several state-of-the-art factorization models and has demonstrated superior performance over traditional collaborative and content-based approaches in various recommendation tasks.

Burges et al. [27] proposed a learning-to-rank framework that directly optimizes the ordering of items rather than predicting absolute preference scores. Their approach formulates ranking as a pairwise learning problem and introduces RankNet, which models ranking preferences using a probabilistic cost function optimized via gradient descent. By focusing on ranking quality, this framework significantly improves recommendation effectiveness in scenarios where the relative order of items is more important than precise rating prediction.

Later, Burges [28] provided a comprehensive overview of learning-to-rank methods, including RankNet, LambdaRank, and LambdaMART. RankNet formulates ranking as a pairwise probabilistic learning problem optimized via gradient descent, while LambdaRank introduces the concept of lambda gradients to directly optimize ranking metrics such as NDCG. By combining LambdaRank with gradient-boosted decision trees, LambdaMART further improves ranking performance and has become a widely adopted approach in large-scale recommendation and information retrieval systems.

In industrial recommender systems, supervised learning techniques are widely adopted due to their strong predictive performance and flexibility in incorporating heterogeneous features. Among these methods, decision tree-based models and ensemble learning techniques are particularly popular, as they provide a good balance between interpretability and the ability to capture complex feature interactions.

He et al. [29] developed a large-scale recommender system for Facebook Ads based on gradient boosting decision trees. By modeling user preferences from rich user, item, and contextual features, their approach significantly improved ad targeting effectiveness and user engagement, demonstrating the practicality of supervised learning methods in real-world industrial recommendation scenarios.

Furthermore, advanced gradient boosting frameworks such as XGBoost [30] and LightGBM [31] have been widely applied in recommender systems to enhance both accuracy and scalability. These methods leverage efficient tree-based boosting strategies to model high-order feature interactions, making them particularly suitable for large-scale recommendation tasks with sparse and high-dimensional feature spaces.

3.1.3 Deep Learning-based Recommender Systems

Deep learning-based recommender systems have significantly advanced the field by enabling end-to-end representation learning and modeling complex, non-linear user – item interactions. Neural Collaborative Filtering (NCF) extends traditional matrix factorization by replacing fixed inner products with multi-layer perceptrons, allowing the model to learn more expressive interaction functions. Beyond interaction modeling, convolutional neural networks (CNNs) have been widely used to extract informative representations from unstructured item content such as images and text, thereby enriching item features for recommendation. Recurrent neural networks (RNNs) and their variants further incorporate temporal dynamics by modeling sequential user behaviors, enabling personalized recommendations that adapt to users' evolving preferences.

3.1.3.1 Neural Collaborative Filtering

Neural Collaborative Filtering (NCF) is a deep learning-based recommendation framework that replaces the fixed inner product used in matrix factorization with neural networks to model user – item interactions. By learning non-linear interaction functions through multi-layer perceptrons, NCF can capture more complex preference patterns than traditional collaborative filtering methods. The framework unifies several neural architectures, including generalized matrix factorization (GMF), multi-layer perceptron (MLP), and their hybrid variant NeuMF, which have demonstrated superior performance on various recommendation benchmarks.

He et al. [32] proposed the Neural Collaborative Filtering (NCF) framework, which formulates collaborative filtering as a neural interaction learning problem. Instead of relying on a fixed inner product as in traditional matrix factorization, NCF employs neural networks to learn flexible and non-linear user – item interaction functions directly from data. In this framework, users and items are embedded into low-dimensional latent spaces and their representations are combined through neural architectures—such as generalized matrix factorization (GMF), multi-layer perceptrons (MLP), and their fusion model NeuMF—to capture complex interaction patterns beyond linear similarity measures.

3.1.3.2 Sequential Recommender Systems

Sequential recommender systems exploit the sequential patterns in users' interaction histories to generate personalized recommendations. By explicitly modeling the order and temporal dependencies of user – item interactions, these methods capture the dynamic evolution of user preferences over time and have been widely applied in domains such as e-commerce, music streaming, and video platforms. Representative techniques for sequential recommendation include recurrent neural networks (RNNs) and their variants such as long short-term memory (LSTM) networks, as well as more recent Transformer-based architectures.

Hidasi et al. [33] proposed a session-based recommender system that applies recurrent neural networks (RNNs) to model user behavior within individual sessions. By representing a session as a sequence of item interactions and maintaining a recurrent hidden state, their approach captures both short-term and long-term dependencies in session data. Furthermore, the authors introduced ranking-oriented loss functions tailored to recommendation tasks, enabling the model to significantly outperform traditional item-to-item and neighborhood-based baselines.

The mainstream sequential recommender systems now are Transformer-based models. Kang and McAuley [34] proposed SASRec, a self-attentive sequential recommender system based on the Transformer encoder architecture. By employing self-attention mechanisms, SASRec adaptively weighs historical items in a user's interaction sequence, enabling the model to capture long-range dependencies while remaining efficient on sparse data. Unlike recurrent models that summarize sequences through a single hidden state, SASRec directly attends to relevant past interactions, leading to improved recommendation accuracy and scalability in sequential recommendation tasks.

Sun et al. [35] proposed BERT4Rec, a Transformer-based sequential recommender system that employs bidirectional self-attention to model user behavior sequences. Unlike unidirectional sequential models such as RNN-based methods and SASRec, BERT4Rec leverages bidirectional contextual information by predicting masked items within a sequence using a Cloze-style training objective. This design enables each item representation to incorporate both preceding and succeeding context, leading to more expressive sequence modeling and consistently improved recommendation performance across multiple benchmark datasets.

3.1.3.3 Graph-based Recommender Systems

Graph-based recommender systems model user – item interactions as graphs and apply graph neural networks (GNNs) to learn representations through neighborhood aggregation. By propagating information along graph edges, these methods can effectively capture high-order connectivity and collaborative signals that are difficult to model with point-wise interaction functions. Moreover, graph-based frameworks naturally support the integration of side information and heterogeneous relations, enabling richer modeling of user preferences and item characteristics.

Van den Berg et al. [36] proposed Graph Convolutional Matrix Completion (GCMC), which formulates collaborative filtering as a link prediction problem on a bipartite user – item interaction graph. By employing a graph convolutional auto-encoder architecture, GCMC learns user and item representations through message passing on the interaction graph and reconstructs ratings via a bilinear decoder. This approach effectively captures high-order collaborative signals and naturally incorporates side information, leading to improved recommendation performance on benchmark datasets.

Ying et al. [37] proposed PinSage, a graph-based recommender system designed for web-scale applications. PinSage combines graph neural networks with efficient random-walk-based neighborhood sampling to learn item representations that incorporate both graph structure and rich side information. By addressing the scalability limitations of conventional GCNs, PinSage was successfully deployed in large-scale industrial systems such as Pinterest, demonstrating the effectiveness of graph-based recommendation models in real-world production environments.

Wang et al. [38] proposed Neural Graph Collaborative Filtering (NGCF), which explicitly integrates graph neural networks into collaborative filtering by modeling user – item interactions as a bipartite graph. NGCF refines user and item embeddings through recursive message passing on the interaction graph, enabling the explicit modeling of high-order connectivity and collaborative signals, which leads to significant improvements in recommendation performance.

He et al. [39] introduced LightGCN, a simplified graph convolutional network tailored for recommender systems. Unlike prior GNN-based models that incorporate feature transformations and nonlinear activations, LightGCN argues that these components contribute little to collaborative filtering performance when only user and item IDs are available as input. Accordingly, LightGCN retains only the neighborhood aggregation operation to propagate embeddings over the user – item interaction graph, significantly simplifying the model architecture while preserving the ability to capture high-order connectivity. Extensive experiments demonstrate that LightGCN not only reduces computational complexity but also achieves superior recommendation accuracy compared to more complex

GNN-based methods such as NGCF, making it a widely adopted and strong baseline in graph-based recommender system research.

Wu et al. [40] proposed SR-GNN, a session-based recommender system that models user interaction sequences as graph-structured data and applies graph neural networks to learn item representations within sessions. By constructing a directed session graph for each interaction sequence, SR-GNN is able to capture complex transition patterns among items that go beyond simple sequential dependencies. This work represents an early attempt to integrate graph-based modeling with sequential recommendation, effectively combining the strengths of GNNs in capturing high-order relational information and sequential models in characterizing short-term user intent.

3.1.3.4 Generative Recommender Systems

Generative recommender systems have emerged as an important research direction that models user preferences and item characteristics from a probabilistic perspective. By leveraging generative models such as Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs), and Normalizing Flows, these approaches aim to learn the underlying distribution of user – item interactions rather than deterministic point estimates. Such distribution-aware modeling enables recommender systems to capture uncertainty in user preferences, generate diverse recommendations, and alleviate data sparsity issues. These advantages are closely related to the motivation of our proposed method, which also represents user preferences as probability distributions to support more expressive and robust recommendation.

Liang et al. [41] proposed a variational autoencoder (VAE)-based collaborative filtering framework for implicit feedback data. By modeling user preferences as latent random variables and employing a multinomial likelihood, this approach provides a probabilistic formulation that captures uncertainty and multi-modal structures in user behavior. Compared with deterministic autoencoder-based models, the VAE framework enables more expressive preference modeling and demonstrates strong empirical performance under sparse interaction settings.

Wang et al. [42] proposed IRGAN, a generative adversarial framework that unifies generative and discriminative models for information retrieval and recommendation tasks. By formulating the learning process as a minimax game, the generative model aims to approximate the underlying relevance distribution over items, while the discriminative model learns to distinguish relevant from non-relevant user – item pairs. Through adversarial training, IRGAN effectively improves recommendation performance, particularly under implicit feedback settings.

Chae et al. [43] proposed CFGAN, a generic collaborative filtering framework based on generative adversarial networks. Unlike prior GAN-based recommender systems that generate discrete item indices, CFGAN adopts vector-wise adversarial training, where the generator produces real-valued preference vectors and the discriminator distinguishes them from ground-truth interaction vectors. This design effectively stabilizes adversarial learning and improves recommendation accuracy, especially under sparse implicit feedback settings.

Beyond VAEs and GANs, diffusion models have recently been introduced to recommender systems as a new class of generative models. Wang et al. [44] proposed DiffRec, a diffusion-based recommender system that models the user – item interaction generation process through iterative denoising. By gradually corrupting user interaction histories and learning to recover the original interactions step by step, DiffRec provides a flexible and expressive framework for modeling complex preference distributions. This work demonstrates the potential of diffusion models to overcome the limitations of traditional generative approaches and further improve recommendation performance under noisy and sparse interaction settings.

3.2 Cross-Domain Recommender Systems

Despite the success of single-domain recommender systems, challenges such as data sparsity and cold-start users remain difficult to address when interaction data are limited. Cross-Domain Recommendation (CDR) tackles these issues by transferring knowledge from a source domain with abundant user – item interactions to a target domain where data are scarce.

The core objective of CDR is to leverage auxiliary information across domains to improve recommendation performance, under the assumption that user preferences or item characteristics exhibit certain transferable patterns. By exploiting correlations between domains, CDR methods aim to enhance recommendation accuracy and robustness, especially in cold-start and sparse-data scenarios.

Based on whether domains share common entities, existing CDR approaches can be broadly categorized into *overlapping* and *non-overlapping* settings. Overlapping CDR methods assume the existence of shared users or items across domains and utilize this overlap as a bridge for knowledge transfer in the training stage. Typical techniques include joint matrix factorization, co-clustering, and graph-based models that align user preferences or item representations across domains.

In contrast, non-overlapping CDR methods address more challenging scenarios where no users or items are shared between domains during training. These approaches generally rely on content features, latent representations, or learned mappings between domains to enable preference transfer without explicit entity overlap. Non-overlapping CDR is particularly relevant in practical applications, but remains challenging due to the lack of direct correspondence between domains.

3.2.1 Overlapping CDR

The core idea of overlapping CDR is to leverage the shared users or items between domains to facilitate knowledge transfer. By identifying commonalities in user preferences and item characteristics, these methods aim to enhance recommendation quality in the target domain. The techniques employed in overlapping CDR can be broadly categorized into three main approaches: representation alignment based on overlapping entities, collective matrix factorization, and adversarial learning for domain adaptation.

Singh et al. [45] proposed a multi-relational matrix factorization approach that jointly factorizes user-item interaction matrices from multiple domains. This method captures shared latent factors among overlapping users or items, enabling effective knowledge transfer across domains. This method is a classic work in overlapping CDR using collective matrix factorization. Li et al. [46] introduced a transfer learning framework for collaborative filtering that leverages overlapping users to enhance recommendation performance in the target domain. Their approach utilizes shared user preferences to facilitate knowledge transfer and improve recommendation accuracy. Pan et al. [47] proposed a transfer learning method that employs co-clustering techniques to align user-item interactions across domains. This approach effectively captures shared patterns among overlapping users or items, leading to improved recommendation quality.

3.2.2 Non-overlapping CDR

Non-overlapping CDR focuses on scenarios where there are no shared users or items between domains. These methods aim to bridge the gap between domains by leveraging content-based features, latent factor models, and deep learning techniques. Man et al. [48] proposed an embedding mapping approach that learns a transformation function to align user and item embeddings across domains. This method effectively captures cross-domain relationships, enabling knowledge transfer in non-overlapping scenarios.

3.3 Distributional User Preference Modeling

User preferences are inherently complex and multifaceted, often exhibiting diverse interests and varying degrees of uncertainty. Traditional recommender systems typically represent user preferences as fixed vectors in a latent space, which may not adequately capture the richness and variability of user behavior. To address these limitations, recent research has explored distributional user preference modeling, which represents user preferences as probability distributions rather than discrete points.

3.3.1 Multi-interest User Modeling

We first start with multi-interest user modeling, which aims to capture the diverse interests of users by representing their preferences with multiple vectors or components.

Blei et al. [19] proposed Latent Dirichlet Allocation (LDA), a generative probabilistic model that represents documents as mixtures of topics. Although originally designed for text modeling, LDA has been adapted for recommender systems to model user preferences as distributions over latent topics, providing a richer representation of user behavior. This work is one of the early attempts at multi-interest user modeling by capturing the diversity of user interests through topic mixtures

Hofmann [17] proposed Probabilistic Latent Semantic Analysis (PLSA) for recommender systems, which models user-item interactions using a mixture of latent topics. PLSA represents user preferences as probability distributions over latent topics, allowing for a more nuanced representation of user behavior. This work targeted the diversity of user interests by modeling them as mixtures of latent factors, thus can be seen as an early attempt at multi-interest user modeling.

Marlin [18] introduced a probabilistic approach to model user rating profiles, capturing the uncertainty in user preferences. This method utilized Gaussian distributions to represent user ratings, enabling more accurate recommendations by accounting for variability in user behavior. This work also addressed the diversity of user interests by modeling them as distributions, thus can be seen as an early attempt at multi-interest user modeling.

Zhou et al. [49] proposed Deep Interest Network (DIN), which captures users' diverse interests by employing an attention mechanism to dynamically aggregate multiple interest representations based on the target item. DIN effectively models the multifaceted nature of user preferences, leading to improved recommendation accuracy.

Li et al. [50] introduced Multi-Interest Network with Dynamic Routing (MIND), which represents user preferences using multiple interest vectors. MIND employs a dynamic routing mechanism to capture diverse user interests, enhancing recommendation performance. MIND is a industrial-level multi-interest user modeling method that has been deployed in large-scale recommender systems.

Cen et al. [51] proposed ComiRec, a comprehensive multi-interest recommender system that utilizes a capsule network to model diverse user interests. ComiRec can explicitly capture multiple interest representations and support controllable diversity in recommendations, leading to enhanced recommendation quality.

3.3.2 Probabilistic Preference Modeling

Beyond multi-interest user modeling, probabilistic preference modeling aims to capture the uncertainty and variability in user preferences by representing them as probability distributions. This approach allows for a more flexible and expressive representation of user behavior, accommodating the inherent uncertainty in user preferences.

The previously mentioned works by Marlin [18], Hofmann [17], and Blei et al. [19] are also modeling user preferences as distributions and their purpose is to capture the multi-faceted nature of user interests. Besides this target, uncertainty modeling is another important purpose of probabilistic preference modeling. Salakhutdinov and Mnih [16] introduced Bayesian Probabilistic Matrix Factorization (BPMF), which incorporated Bayesian inference into the MF framework. BPMF provided a principled way to handle uncertainty in user preferences and item characteristics. BPMF models user preferences as Gaussian distributions in the latent space, allowing for a more flexible representation that captures uncertainty in user behavior.

VAE-based methods also model uncertainty in user preferences. Liang et al. [41] introduced a VAE-based collaborative filtering approach that models user preferences as latent variables. This method captures uncertainty in user behavior, enhancing recommendation quality. By representing user preferences as distributions in the latent space, VAE-based methods can effectively model the variability and uncertainty inherent in user behavior.

Besides recommendation systems implementations, some works focus on the general representation learning of words or items as distributions. Vilnis and McCallum [52] proposed Gaussian Embeddings, which represent words as Gaussian distributions in the embedding space. This approach captures both the semantic meaning and uncertainty of word representations, providing a richer and more expressive representation compared to traditional point embeddings. Although this work is not specifically designed for recommender systems, the idea of representing entities as distributions can be extended to model user preferences and item characteristics in recommendation tasks.

3.4 Optimal Transport for Representation Alignment

Optimal Transport is a framework for finding an optimal transport plan between two probability distributions, minimizing the cost of transporting mass from one distribution to another. OT has been widely used in various machine learning tasks, including domain adaptation, generative modeling, and representation learning.

3.4.1 OT Basics in ML

Optimal Transport (OT) provides a mathematical framework for comparing and aligning probability distributions by finding the most efficient way to transport mass from one distribution to another. In machine learning, OT has been employed in various tasks, including domain adaptation, generative modeling, and representation learning.

Villani [53] provided a comprehensive overview of the mathematical foundations of OT, including the Kantorovich formulation and Wasserstein distances. This work laid the groundwork for applying OT in machine learning by establishing key theoretical concepts.

Peyré and Cuturi [?] presented a detailed survey of OT applications in machine learning, covering topics such as domain adaptation, generative modeling, and deep learning. This work highlighted the versatility of OT in addressing various machine learning challenges and provided practical algorithms for implementing OT-based methods.

Cuturi and Marco [54] introduced the Sinkhorn algorithm, which enables efficient computation of OT distances by adding an entropic regularization term. This method significantly reduces the computational complexity of OT, making it feasible for large-scale machine learning applications.

Genevay et al. [55] proposed the use of Sinkhorn divergences for training generative models. This work demonstrated the effectiveness of OT-based losses in improving the quality of generated samples, showcasing the potential of OT in generative modeling tasks.

3.4.2 OT in Domain Adaptation

Optimal Transport has been widely used in domain adaptation tasks to align feature distributions between source and target domains. By minimizing the transport cost between distributions, OT-based methods can effectively transfer knowledge from the source domain to the target domain, improving model performance in the target domain.

Courty et al. [56] proposed an OT-based domain adaptation method that aligns feature distributions between source and target domains. Their approach minimizes the Wasserstein distance between distributions, enabling effective knowledge transfer and improved performance in the target domain.

Frogner et al. [57] introduced a deep learning framework for domain adaptation using OT. Their method incorporates OT losses into deep neural networks, allowing for end-to-end training and effective alignment of feature distributions across domains.

With the development of efficient OT computation methods, such as the Sinkhorn algorithm, OT-based domain adaptation methods have become more practical for large-scale applications. These methods have demonstrated significant improvements in various domain adaptation tasks, showcasing the effectiveness of OT in addressing distributional shifts between domains.

Chapter 4

Proposed Method: Distributional Preference Modeling for Cross-Domain Recommendation

4.1 Background

The existing recommender systems have already achieved significant success in various applications. However, they often struggle with challenges such as data sparsity, cold start problems. Also, most traditional recommender systems represent user preferences as point estimates, which may not fully capture the uncertainty and multifaceted nature of user preferences. To address these issues, we propose a novel approach that models user preferences as probability distributions rather than single point estimates, and leverages cross-domain information to enhance recommendation accuracy. In existing cross-domain recommendation methods, it is common to assume that there are overlapping users or items between domains during the training phase to provide a bridge for knowledge transfer. However, we argue that this kind of assumption may not hold in many real-world scenarios, where users and items can be entirely distinct across different domains. Therefore, our proposed method focuses on scenarios where there are no overlapping users or items between domains during training, making it more applicable to a wider range of real-world applications. Based on this motivation, we introduce a distributional preference modeling framework for cross-domain recommendation without overlapping users or items. We call this method DUP-OT(Distributional User Preference with Optimal Transport). The key idea behind DUP-OT is to represent user preferences in each domain as probability distributions over items, capturing the uncertainty and diversity of user interests. By leveraging optimal transport theory, we can effectively align and transfer knowledge between the source and target domains, even in the absence of overlapping users or items. This alignment allows us to learn a shared latent space where user preferences from both domains can be compared and utilized for recommendation.

4.2 Methods

4.2.1 Overview

The overall architecture of the proposed DUP-OT framework is illustrated in Figure 4.1. The framework consists of three main components: (1) Shared Feature Extraction, (2) Distributional Preference Modeling, and (3) Optimal Transport-based Knowledge Transfer. The

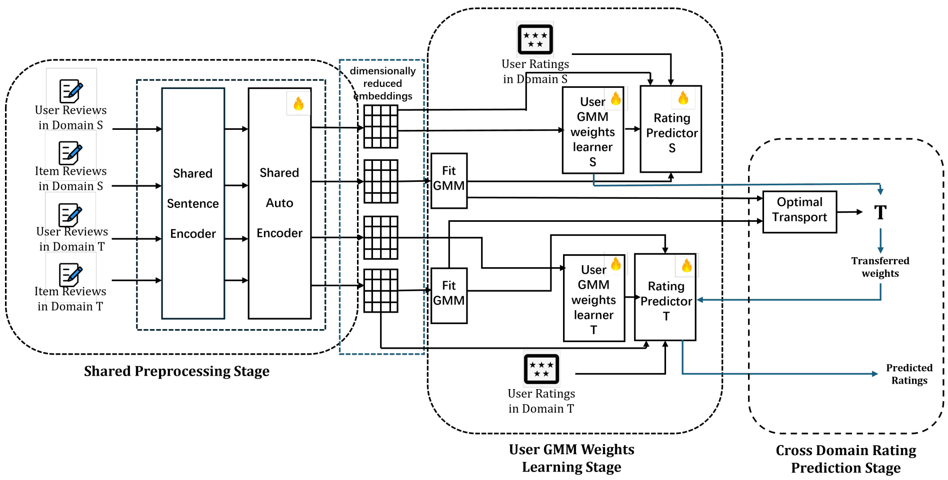

Figure 4.1: Architecture of the DUP-OT Framework

Chapter 5

Experiments

Chapter 6

Results

Chapter 7

Conclusion

References

[1] D. Bawden and L. Robinson. Information overload: An overview. In *Oxford Encyclopedia of Political Decision Making*. Oxford University Press, Oxford, June 2020. This is a draft of a chapter that has been published by Oxford University Press in the book Oxford Encyclopedia of Political Decision Making.

[2] Jun Lv and Xuan Liu. The impact of information overload of e-commerce platform on consumer return intention: Considering the moderating role of perceived environmental effectiveness. *International Journal of Environmental Research and Public Health*, Vol. 19, No. 13, 2022.

[3] Guihua Zhang, Junwei Cao, and Dong Liu. Examining the influence of information overload on consumers' purchase in live streaming: A heuristic-systematic model perspective. *PLOS ONE*, Vol. 18, No. 8, pp. 1–16, 08 2023.

[4] Paul Resnick and Hal R Varian. Recommender systems. *Communications of the ACM*, Vol. 40, No. 3, pp. 56–58, 1997.

[5] Robin Burke. Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, Vol. 12, No. 4, pp. 331–370, 2002.

[6] Robin Burke, Alexander Felfernig, and Mehmet H. Göker. Recommender systems: An overview. *AI Magazine*, Vol. 32, No. 3, pp. 13–18, Jun. 2011.

[7] Dietmar Jannach and Michael Jugovac. Measuring the business value of recommender systems. *ACM Trans. Manage. Inf. Syst.*, Vol. 10, No. 4, December 2019.

[8] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez. Recommender systems survey. *Knowledge-Based Systems*, Vol. 46, pp. 109–132, 2013.

[9] Xiaoyuan Su and Taghi M. Khoshgoftaar. A survey of collaborative filtering techniques. *Advances in Artificial Intelligence*, Vol. 2009, No. 1, p. 421425, 2009.

[10] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. Grouplens: an open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, CSCW '94, p. 175–186, New York, NY, USA, 1994. Association for Computing Machinery.

[11] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, pp. 285–295, 2001.

[12] John S. Breese, David Heckerman, and Carl Kadie. Empirical analysis of predictive algorithms for collaborative filtering, 2013.

[13] Lyle H Ungar and Dean P Foster. Clustering methods for collaborative filtering. In *AAAI workshop on recommendation systems*, Vol. 1, pp. 114–129. Menlo Park, CA, 1998.

[14] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, Vol. 42, No. 8, pp. 30–37, 2009.

[15] Andriy Mnih and Russ R Salakhutdinov. Probabilistic matrix factorization. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems*, Vol. 20. Curran Associates, Inc., 2007.

[16] Ruslan Salakhutdinov and Andriy Mnih. Bayesian probabilistic matrix factorization using markov chain monte carlo. In *Proceedings of the 25th International Conference on Machine Learning*, ICML '08, p. 880–887, New York, NY, USA, 2008. Association for Computing Machinery.

[17] Thomas Hofmann. Latent semantic models for collaborative filtering. *ACM Trans. Inf. Syst.*, Vol. 22, No. 1, p. 89–115, January 2004.

[18] Benjamin M Marlin. Modeling user rating profiles for collaborative filtering. In S. Thrun, L. Saul, and B. Schölkopf, editors, *Advances in Neural Information Processing Systems*, Vol. 16. MIT Press, 2003.

[19] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, Vol. 3, No. Jan, pp. 993–1022, 2003.

[20] Gerard Salton, Chris Buckley, and Edward A Fox. Automatic query formulations in information retrieval. *Journal of the American society for information science*, Vol. 34, No. 4, pp. 262–280, 1983.

[21] Michael Pazzani and Daniel Billsus. Learning and revising user profiles: The identification of interesting web sites. *Machine learning*, Vol. 27, No. 3, pp. 313–331, 1997.

[22] Michael J Pazzani. A framework for collaborative, content-based and demographic filtering. *Artificial intelligence review*, Vol. 13, No. 5, pp. 393–408, 1999.

[23] Prem Melville, Raymond J Mooney, Ramadass Nagarajan, et al. Content-boosted collaborative filtering for improved recommendations. *Aaai/iaai*, Vol. 23, pp. 187–192, 2002.

[24] Daniel Billsus and Michael J Pazzani. User modeling for adaptive news access. *User modeling and user-adapted interaction*, Vol. 10, No. 2, pp. 147–180, 2000.

[25] Justin Basilico and Thomas Hofmann. Unifying collaborative and content-based filtering. In *Proceedings of the twenty-first international conference on Machine learning*, p. 9, 2004.

[26] Steffen Rendle. Factorization machines. In *2010 IEEE International conference on data mining*, pp. 995–1000. IEEE, 2010.

[27] Chris Burges, Tal Shaked, Erin Renshaw, Ari Lazier, Matt Deeds, Nicole Hamilton, and Greg Hullender. Learning to rank using gradient descent. In *Proceedings of the 22nd international conference on Machine learning*, pp. 89–96, 2005.

[28] Christopher JC Burges. From ranknet to lambdarank to lambdamart: An overview. *Learning*, Vol. 11, No. 23-581, p. 81, 2010.

[29] Xinran He, Junfeng Pan, Ou Jin, Tianbing Xu, Bo Liu, Tao Xu, Yanxin Shi, Antoine Atallah, Ralf Herbrich, Stuart Bowers, et al. Practical lessons from predicting clicks on ads at facebook. In *Proceedings of the eighth international workshop on data mining for online advertising*, pp. 1–9, 2014.

[30] Tianqi Chen. Xgboost: A scalable tree boosting system. *Cornell University*, 2016.

[31] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, Vol. 30, , 2017.

[32] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web*, pp. 173–182, 2017.

[33] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. Session-based recommendations with recurrent neural networks. *arXiv preprint arXiv:1511.06939*, 2015.

[34] Wang-Cheng Kang and Julian McAuley. Self-attentive sequential recommendation. In *2018 IEEE international conference on data mining (ICDM)*, pp. 197–206. IEEE, 2018.

[35] Fei Sun, Jun Liu, Jian Wu, Changhua Pei, Xiao Lin, Wenwu Ou, and Peng Jiang. Bert4rec: Sequential recommendation with bidirectional encoder representations from transformer. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, CIKM '19, p. 1441–1450, New York, NY, USA, 2019. Association for Computing Machinery.

[36] Rianne Van Den Berg, N Kipf Thomas, and Max Welling. Graph convolutional matrix completion. *arXiv preprint arXiv:1706.02263*, Vol. 2, No. 8, p. 9, 2017.

[37] Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L Hamilton, and Jure Leskovec. Graph convolutional neural networks for web-scale recommender systems. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 974–983, 2018.

[38] Xiang Wang, Xiangnan He, Meng Wang, Fuli Feng, and Tat-Seng Chua. Neural graph collaborative filtering. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR'19, p. 165–174, New York, NY, USA, 2019. Association for Computing Machinery.

[39] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, YongDong Zhang, and Meng Wang. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '20, p. 639–648, New York, NY, USA, 2020. Association for Computing Machinery.

[40] Shu Wu, Yuyuan Tang, Yanqiao Zhu, Liang Wang, Xing Xie, and Tieniu Tan. Session-based recommendation with graph neural networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, No. 01, pp. 346–353, Jul. 2019.

[41] Dawen Liang, Rahul G. Krishnan, Matthew D. Hoffman, and Tony Jebara. Variational autoencoders for collaborative filtering. In *Proceedings of the 2018 World Wide Web Conference*, WWW '18, p. 689–698, Republic and Canton of Geneva, CHE, 2018. International World Wide Web Conferences Steering Committee.

[42] Jun Wang, Lantao Yu, Weinan Zhang, Yu Gong, Yinghui Xu, Benyou Wang, Peng Zhang, and Dell Zhang. Irgan: A minimax game for unifying generative and discriminative information retrieval models. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '17, p. 515–524, New York, NY, USA, 2017. Association for Computing Machinery.

[43] Dong-Kyu Chae, Jin-Soo Kang, Sang-Wook Kim, and Jung-Tae Lee. Cfgan: A generic collaborative filtering framework based on generative adversarial networks. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, CIKM '18, p. 137–146, New York, NY, USA, 2018. Association for Computing Machinery.

[44] Wenjie Wang, Yiyan Xu, Fuli Feng, Xinyu Lin, Xiangnan He, and Tat-Seng Chua. Diffusion recommender model. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '23, p. 832–841, New York, NY, USA, 2023. Association for Computing Machinery.

[45] Ajit P Singh and Geoffrey J Gordon. Relational learning via collective matrix factorization. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 650–658, 2008.

[46] Bin Li, Qiang Yang, and Xiangyang Xue. Transfer learning for collaborative filtering via a rating-matrix generative model. In *Proceedings of the 26th annual international conference on machine learning*, pp. 617–624, 2009.

[47] Weike Pan, Evan Xiang, Nathan Liu, and Qiang Yang. Transfer learning in collaborative filtering for sparsity reduction. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 24, pp. 230–235, 2010.

[48] Tong Man, Huawei Shen, Xiaolong Jin, and Xueqi Cheng. Cross-domain recommendation: An embedding and mapping approach. In *Ijcai*, Vol. 17, pp. 2464–2470, 2017.

[49] Guorui Zhou, Xiaoqiang Zhu, Chenru Song, Ying Fan, Han Zhu, Xiao Ma, Yanghui Yan, Junqi Jin, Han Li, and Kun Gai. Deep interest network for click-through rate prediction. In *Proceedings*

of the 24th ACM SIGKDD international conference on knowledge discovery & data mining, pp. 1059–1068, 2018.

[50] Chao Li, Zhiyuan Liu, Mengmeng Wu, Yuchi Xu, Huan Zhao, Pipei Huang, Guoliang Kang, Qiwei Chen, Wei Li, and Dik Lun Lee. Multi-interest network with dynamic routing for recommendation at tmall. In *Proceedings of the 28th ACM international conference on information and knowledge management*, pp. 2615–2623, 2019.

[51] Yukuo Cen, Jianwei Zhang, Xu Zou, Chang Zhou, Hongxia Yang, and Jie Tang. Controllable multi-interest framework for recommendation. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 2942–2951, 2020.

[52] Luke Vilnis and Andrew McCallum. Word representations via gaussian embedding, 2015.

[53] Cédric Villani, et al. *Optimal transport: old and new*, Vol. 338. Springer, 2008.

[54] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, Vol. 26, , 2013.

[55] Aude Genevay, Gabriel Peyré, and Marco Cuturi. Learning generative models with sinkhorn divergences. In *International Conference on Artificial Intelligence and Statistics*, pp. 1608–1617. PMLR, 2018.

[56] Nicolas Courty, Rémi Flamary, Devis Tuia, and Alain Rakotomamonjy. Optimal transport for domain adaptation. *IEEE transactions on pattern analysis and machine intelligence*, Vol. 39, No. 9, pp. 1853–1865, 2016.

[57] Charlie Frogner, Chiyuan Zhang, Hossein Mobahi, Mauricio Araya, and Tomaso A Poggio. Learning with a wasserstein loss. *Advances in neural information processing systems*, Vol. 28, , 2015.

Appendix

A Additional Experimental Results

Publications

- 肖　子吟 and 鈴村 豊太郎 WebDB 夏のワークショップ 2025

Acknowledgment

This work was supported by ...