# Generalized Cross-Multi Domains Representations using User Sentiments

Aabhaas Batra
aabhaas19001@iiitd.ac.in

Eeshaan Ravi Tivari
eeshaan19465@iiitd.ac.in

Jahnvi Kumari
jahnvi19469@iiitd.ac.in

Rishi Singhal
rishi19194@iiitd.ac.in

Soham Das
soham19477@iiitd.ac.in

Udit Narang
udit19120@iiitd.ac.in

## 1 Introduction

In today's day and age, people spend more time on social media and streaming apps like netflix, instagram, spotify, etc. than they do with their social circles. With over 5 billion users of e-commerce websites, social media websites and streaming platforms, providing the best user experience is key to these companies. Hence, to keep the user retention high, these companies need to continuously innovate and optimize their recommendation systems. The goal of a recommender system is to provide personalized recommendations to users by predicting what they are most likely to prefer among a large set of items. Recommendation systems often rely on user-item interactions to make recommendations. But data sparsity in recommendation systems is a common problem. It makes it challenging for the recommendation system to accurately predict the user's preferences and make relevant recommendations. This may arise due to a user interacting with an item (say, buying a dress) in a domain different from what they usually interact in (say, buying books). However, cross-domain representations can be used to transfer the knowledge of user-item interaction in interacted domains to a sparse domain. While work has been done on bi-domain models, generalizing them to k domains can help find the optimal number of domains to consider for representation learning. By considering multiple domains and user sentiments, the model can learn to generalize across a wider range of user-item interactions and make more informed recommendations. This method can help improve the accuracy and relevance of recommendations, even in sparser domains with limited user-item interaction data.

## 2 Problem Statement

The main objective of the project is to carry out the task of generating generalized multi-cross-domains representations while taking into account user reviews and sentiments.
Since most of the cross-domain work has been done across two domains. Thus, in the proposed task, given an adjacency list T of items (from multiple $K$ domains where $K \geq 2$) and users, and metadata M about the interactions such as comments, ratings, and item descriptions, we would generate cross-domain representations for recommender systems to capture the essence of user-item interaction in a better manner. To do so, we intend to make use of a weighted graph that would represent the user-item interactions along with their sentiment scores.
The problem statement we are trying to work on is novel since there's not much-existing literature on generalized multi-cross-domain representations. Also, adding user sentiments as a weighted graph would help provide a deeper insight into cross-domain recommendations.

## 3 Literature Survey

As part of existing work, various different techniques have been proposed to build an efficient cross-domain recommendation system based on user-item interactions.

Cao et al. [1] proposes a model to create disentangled cross-domain representations for the user given two different domains, a shared user set, and an adjacency list to determine user-item interaction. Their model, disenCDR, disentangles shared user information and domain specific information for the two domains. For this purpose, they use a variational bipartite graph encoder and carefully formulated mutual information based regularizers. The paper carries out its experiments on Amazon review dataset.
Liu et al. [4] have proposed a Joint Spectral Convolutional Network(JSCN) to capture the high-order connectivity information in the field of cross-domain recommendation. The proposed JSCN model, extracts higher-order connections by

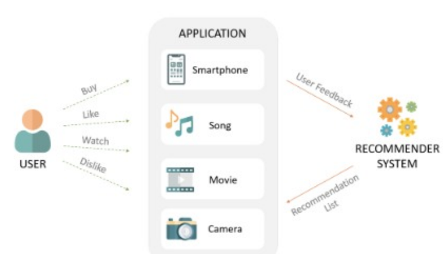


**Figure 1.** A basic recommender system's flowchart [2]

carrying out multi-layer spectral convolutions on different user-item bipartite graphs to transfer information across domains by learning a domain-invariant user representation. Finally, it is evaluated on Amazon Review Dataset to achieve state-of-the art performance in terms of precision and MAP scores.

Another paper, proposed by Wang et al. [6] leverages sentiment analysis to bridge the gap between the source and target domains. At first, latent features are extracted in the source domain. Then, the sentiment features are mapped to the target domain using a cross-domain mapping function which is learned using a transfer learning approach. The proposed model is then evaluated against the Amazon Review dataset and achieved low RMSE values.

The work proposed by Li et al. [3] leveraged the use of latent embeddings of features and user preferences across domains instead of explicit information between the source and the target domain to capture hidden complex interactions. It also focuses on dual transfer learning mechanism to enable learning in both source and target domains based on learnt knowledge instead of trivial unidirectional learning. They've shown RMSE, MAE, recall, precision scores using a large-scale anonymized dataset obtained from a European online recommendation service.

Lu et al. [5] proposed a model to capture the consistency of source and target domains in collaborative filtering settings. The smaller the variance of the empirical prediction error produced by their model, the more likely this user is consistent with those from other domains.. Thus, they embed this criterion into a boosting framework to perform selective knowledge transfer.

Zhao et al. [7] proposed a unified multi-task model through the construction of a cross-domain preference matrix. It models the interactions of different domains as a whole. They added a propagation layer to their model Preference Propagation GraphNet which captures how user preferences propagated in the graph. They also defined a joint objective for different domains.

# 4 Novelty Proposed

Our work will explore two new avenues. We will use a sentiment-weighted user-item interaction graph to train our model so that the feedback of the user is better captured in the representations learnt. We will also generalize cross-domain representation learning to K domains instead of being limited to just two. We shall propose methods to compare the performance of our multi domain models with differing number of source domains. This will aid in comparing our results with previous literature (predominantly on two domains).

Our research can help find a trend of performance with increasing number of source domains, as well as an optimal number of domains. It will help better identify possible roadblocks for research in multi-domain models for recommender systems.

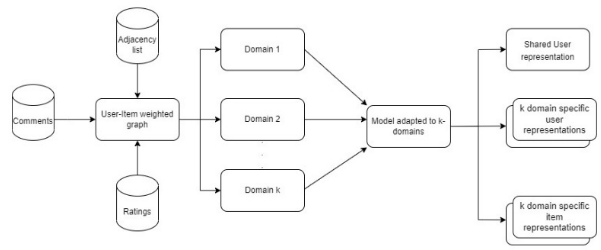


**Figure 2.** A flowchart of our proposed methodology

# 5 Methodology

## 5.1 Amazon Review Dataset

The Amazon review dataset is a collection of user reviews and ratings for products sold on Amazon. It contains information such as the product ID, user ID, review text, rating, and other metadata. It's commonly used in cross-domain recommendation systems as it provides a diverse set of products from various categories. It's a valuable resource for researchers and data scientists in the field of recommendation systems.

## 5.2 Solution Proposed

Figure[2] illustrates the methodology that we aim to follow. From the user-item pair in the dataset, a weighted bipartite graph containing two node classes User(U) and Item(i), is constructed by creating an edge u⟶i whenever user u interacts with the item i. The weight W(u->I) will represent the sentiment score S(u->i) of the user u to the item i. Thus incorporating this information in the form of graphs from different domains, we will finally get a shared cross-domain embedding containing three parts: shared user representation, domain-specific user representations, and domain-specific item representations.

## 5.3 Evaluation Metrics

To evaluate our learned representations, we will use HR (hit ratio) and NDCG (Normalized Discounted Cumulative Gain) to evaluate the quality of recommendations. To measure the amount of disentanglement achieved by our model, a metric using KL divergence between the shared and domain specific representations is suggested by [[1]] which we can use to determine if our method improves the disentanglement of the representations. Since we will work with k domains, we will have to compare our metrics with all pairwise cross domain representations, since our model's performance might be intermediate between some pairs for bi-cross-domain representation models.

## 6 Project Timeline

(1.) Data collection : Collecting data from the Amazon product reviews data. **[1 week]**
(2.) Data Processing : Cleaning and pre-processing the raw collected data from step 1 for further modelling purpose. **[1.5 weeks]**
(3.) Training for Sentiment Analysis: Training the sentiment analysis model over processed data to get sentiment scores that will be used in sentiment-weighted user-item interaction graph. **[1.5 weeks]**
(4.) Modelling k-domain : Modelling the cross-domain representation learning from 2 to K domains by tweaking loss function. **[2 weeks]**
(5.) Analysis : Implicit analysis and comparison of the wrestling representations. **[1 week]**

## References


[1] Jiangxia Cao, Xixun Lin, Xin Cong, Jing Ya, Tingwen Liu, and Bin Wang. 2022. DisenCDR: Learning Disentangled Representations for Cross-Domain Recommendation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 267–277.

[2] Diana Ferreira, Sofia Silva, António Abelha, and José Machado. 2020. Recommendation system using autoencoders. *Applied Sciences* 10, 16 (2020), 5510.

[3] Pan Li and Alexander Tuzhilin. 2020. Ddtcdr: Deep dual transfer cross domain recommendation. In *Proceedings of the 13th International Conference on Web Search and Data Mining*. 331–339.

[4] Zhiwei Liu, Lei Zheng, Jiawei Zhang, Jiayu Han, and S Yu Philip. 2019. JSCN: Joint spectral convolutional network for cross domain recommendation. In *2019 IEEE international conference on big data (big data)*. IEEE, 850–859.

[5] Zhongqi Lu, Erheng Zhong, Lili Zhao, W Xiang, Weike Pan, and Qiang Yang. 2012. Selective transfer learning for cross domain recommendation. Proceedings of the 2013 SIAM International Conference on Data Mining (SDM), pp. 641-649. (2012).

[6] Yongpeng Wang, Hong Yu, Guoyin Wang, and Yongfang Xie. 2020. Cross-domain recommendation based on sentiment analysis and latent feature mapping. *Entropy* 22, 4 (2020), 473.

[7] Cheng Zhao, Chenliang Li, and Cong Fu. 2019. Cross-domain recommendation via preference propagation graphnet. In *Proceedings of the 28th ACM international conference on information and knowledge management*. 2165–2168.