International Journal of Research in Engineering
www.engineeringpaper.net
Online ISSN: 2664-8784, Print ISSN: 2664-8776
Received: 13-03-2022, Accepted: 28-03-2022, Published: 13-04-2022
Volume 4, Issue 1, 2022, Page No. 13-18

Enhancing item based collaborative filtering

Kush Patel[1], Ronak Patel[2]
[1] Department of Computer Science, Charusat University, Changa, Anand, Gujarat, India
[2] Assistant Professor, Department of Computer Engineering, Charusat University, Changa, Anand, Gujarat, India

DOI: https://doi.org/10.33545/26648776.2022.v4.i1a.40

Abstract
Recommendations play an important role in this contemporary era. Most use online platforms like Amazon, Net-flix, etc. to buy daily necessities and entertainment. Moreover, the referral system helps both consumers and customers regardless of age group and financial background. There are mainly three types of referral systems namely Content-Based, Collaborative and Hybrid. Collaborative systems have many problems and some of them are solved to some extent, but the cold start problem is hardly solved by the researcher. Traditional recommendation algorithms such as matrix factorization and collaborative filtering perform poorly when they lack information on the interaction between the user and the product, known as the user cold-start problem, which can cause reduce the revenue of the e-commerce platform. Cold - star problem is divided into two types User cold star and item cold star. In this search, we have selected cross- domain to solve item coldstar and it works fine for cold non-star issues as well. We choose the Amazon dataset for deployment where one is considered the source domain and the other as the target domain.

Keywords: recommendation, Cross-domain, Accuracy mea-sures, cold star, collaborative filtering, transfer learning

Introduction
Recommend systems have become an important area of research since the first articles on collaborative filtering appeared in the mid-1990s [9]. RS is that the subclass of data filtering within which system suggest relevent items or prod- ucts to the users. There has been much work done both within the industry and academia on developing new approaches to recommender systems over the last decade as this method plays a crucial role in contemporary era. Content-based rec-ommendation, collaborative filtering-based recommendation, and hybrid recommendation methods are the three current research directions in recommender systems [4]. In addition, recommendation systems can predict user ratings before the user has provided one, making them a useful tool. Generally, recommendation systems process data through the following four phases: Collection, Storage, Analysis, and Filtering.

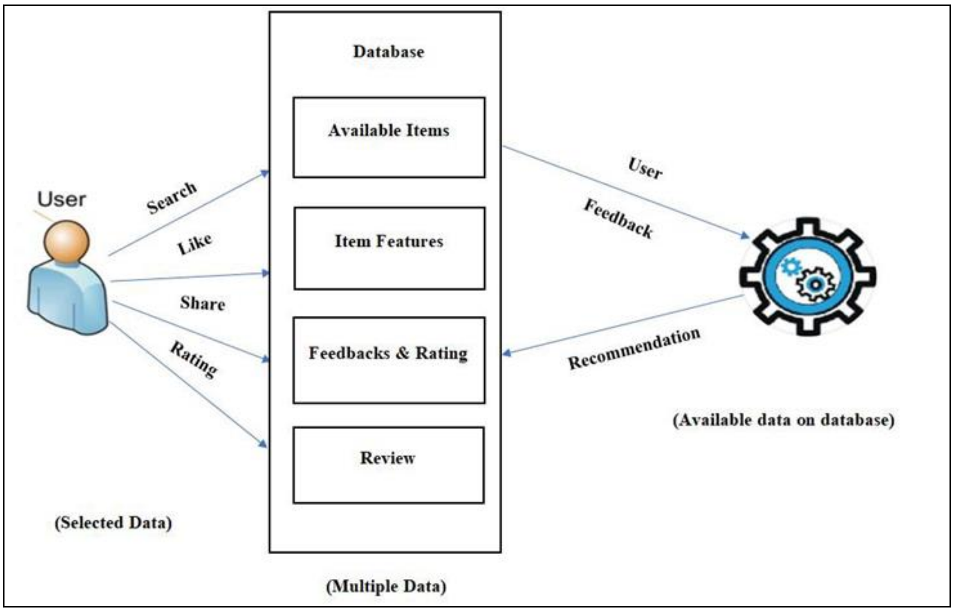

Fig 1: Process of Recommendation System

1. Content-Based Recommender Systems
Content material-primarily based recommender systems make use of the contents of items and finds the similarities among them. After studying sufficient numbers of gadgets that one person has already shown favor to, the consumer hobbies profile is mounted. Then the recommender device ought to search the database and pick proper gadgets in keeping with this profile [18].

2. Collaborative Filtering (Cf)Recommender
Systems: Collaborative filtering (CF) has emerged as one of the most influential recommendation algorithms. Unlike, the content-based approaches, CF simplest is predicated on the object ratings from every user. it is based totally on the assumption that users who have rated the identical items with similar rankings are possibly to have similar options. Collaborative filtering recommends objects primarily based on the interest of different like-minded users or identifies items similar to the ones previously rated utilizing the goal user. It makes use of statistical strategies to discover the similarity among the user or item vector. CF strategies may be categorized into two categories memory-based totally and model-based totally [8].

3. Hybrid Recommender Systems
Hybrid advice systems are divided into monolithic hybrid advice, parallel hybrid recommendation, and pipeline hybrid recommendation. The monolithic hybrid recommendation is a hybrid advice method that integrates numerous pieces of advice into one set of rules [10]. The last hybrid recommendations require at the least specific advice strategies after which integrate them. In step with the input, the parallel hybrid recommendation operates independently of every different, respectively gener- ating a recommendation list, after which the output records is combined into the very last advice set.
All CF recommendation methods ambitions to acquire as least prediction mistakes as viable. But, the prediction step relies specifically on the similarity computation step among customers/objects. Two fundamental issues, bloodless-begin, and sparsity affect the best of the advice and prediction accuracy. Those two important problems face the CF because of the whole dependency on the ancient user-item score matrix. The sparsity trouble is a result of that most customers do not fee maximum of the items or handiest rate some gadgets [1]. A cold-start difficulty faces the CF whilst there aren't always sufficient facts available approximately the AU as he/she is a brand new user or whilst a brand new item is introduced.

Related Work
Make recommendations to users based on historical ratings. There are many ways to generate recommendations [3], but in recent years, collaborative filtering is one of the most well-known methods. It comes in two types: memory-based and model-based. The first type of method is purely based on memory space. This means that the complete data set is loaded instantly and you can use statistical tools to calculate predictions. Filtering-based cooperation based on three steps: 1) Calculation of similarities, 2) Choosing neighbors and 3) Predict points that cannot be observed. Similarities can be calculated between users (cfuu) or between elements (CFIIs). Here we will only focus on filtering cooperation with CFUU. On the other hand, decomposing the matrix, a basic method modeled, is one of the most famous methods to divide the matrix to assess the user's user's user or matrix element And two extra matches for latent factors [17]. Both methods are based on past symbols and cannot provide recommenda- tions if these symbols are limited.The work of cross-domain recommendations has been elaborated in different proposed scenarios, such as single-target CDR, multi-domain proposal, dual-target CDR, and multi-target CDR [24]. It has been suggested that CDRs use relatively more complete information in more affluent regions to improve implementation of recom- mendations in sparsely populated areas.Several authors [15, 2] have proposed how to exploit source domain knowledge in source domain. You can overlap user/item overlap, i.e. you can overlap and categorize four types [6] of non-overlapping categories with item or item without overlapping item, user, and element. User overlap: the authors analyze the correlation of different fields by item category, and are very small in both user item and user item and user item and user item and overlap of [21]. It's too hard to find common objects in the field. Non-overlapping user or item-based methods for trans- ferring knowledge across domains. The author has extended the work, including the probabilistic model [13]. Both meth- ods use cluster-level ranking model to link domains.Cross- domain recommendation systems [20, 5], which combine correlative information from different domains to improve recommendation efficiency, have attracted the attention of researchers in recent years. However, most methods [14, 23] ignore sequential information and predict user behavior based on long-term preferences. While some sequence-based approaches take into account sequence features, most use fixed pop-ups [12] to simplify sequence processing or to model sequence dynamics without user intervention [19]. If the two domains are explicitly or implicitly connected, this can be used to troubleshoot Cold Stores [22]. In addition, general transfer and filtering enables knowledge extraction with enough data to extract knowledge from the source domain, in order to improve the accuracy of recommendations in the target domain. Current CDR methods for these options do not provide a seamless reconciliation of these options between the two domains using common information for users and items, even without a direct domain [16]. Current CDR methods for these options do not provide a seamless reconciliation of these options between the two domains using common information for users and items, even without a direct domain.Because the main limitations of traditional RS are not available because there is no information that cannot be used, there is no problem with the Cold Start Cold start, so it is a problem with the entire curved user user because there is no

interactive information on the user and the product. However, you can first resolve it using three steps [11]. Review is dynamically selected using the adjacent matrix. Second, the vector and domain preferences of the domain vector and domain preferences of the domain preferences are extracted from the text generated from the text generated from the text generated from the text generated from the view text using the encoder and the decoder, and the preferences of the Cold Start User Combined to predict the rating (target domain) [7]. This method does not incorrectly determine the problem to troubleshoot the problem.

Praposed Approach

In the literature, the researchers examined distinct concepts of domain. For example, some considered items such as movies and books as belonging to different domains, while another considered items such as action movies and comedy films as belonging to different domains. To our knowledge, in the study of recommendation systems, no attempt was made to define the domain concept. Here we distinguish several domain concepts by suggested item types and attributes. More precisely, we assume that the domain can be defined at four levels:

1. (Item) Attribute level.
2. (Item) Level of type.
3. Item level.

System level. Where I have selected, the item levels in this Recommended are not of the same type, differing in most, if not all, of their attributes. For example, movies and books belong to different fields, even if they have certain common attributes (title, release / published year).

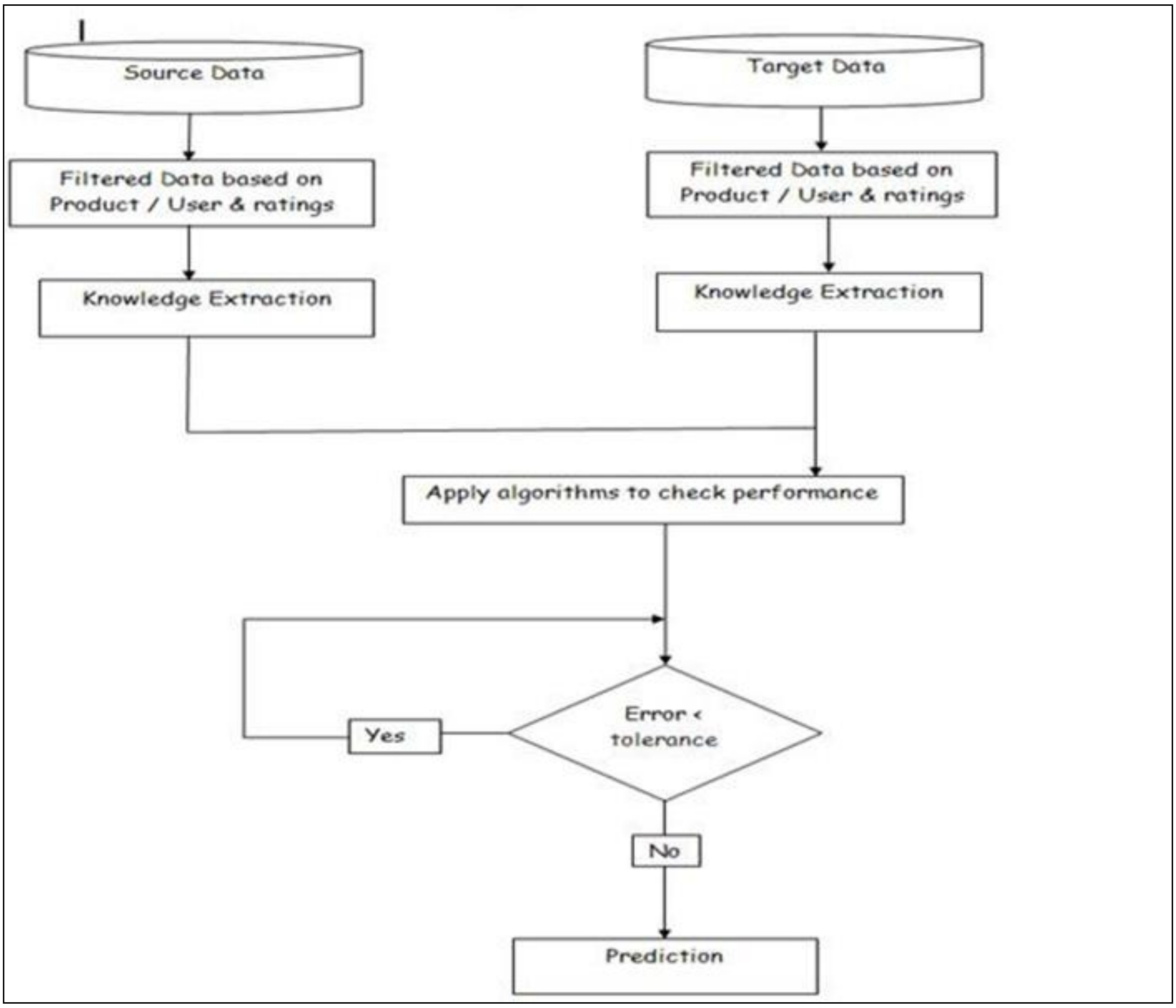

Fig 2: The architecture of proposed approach

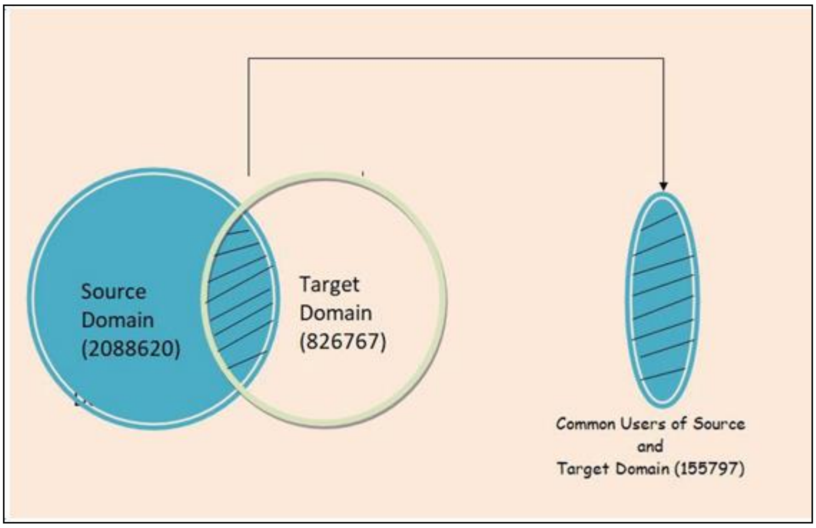

Fig 3: Fetch Common User

The present work shows both source and subdomains. Most researchers work with a small dataset but in my work i choose Amazon realtime dataset, more than ten lakh. According to the figure, we first take two datasets and filter them based on user ID. In the next step, we extract knowledge, apply different algorithms and check if condition error is below the allowed level. Then the items are suggested, otherwise the process is started from the original steps.

Dataset

The video games are the targeted domain and Movie and television are the source domains used for the same purposes but to very different degrees.

Table 1: Analysis of Original Dataset

Attributes	Videogame	Movie & TV
Original size	1324753	4607047
Unique Users	826767	2088620
Unique Products	50210	200941

For implementation purposes, We have taken two real-world datasets from amazon which are particularly rating based and it is open source so everyone can access them easily. As shown in the table, an analysis of both datasets is given in which mentions core three attributes namely original size, unique userId, and unique productId. Both datasets have four columns as userId, productId, rating and timestamp. Time. The last column means that the timestamp is no longer needed, so we dropped it.

Evalution Measure and Result Analysis

As per the above discussion, we have two datasets after uploading them on Google drive and utilizing them in Google collab through the mounting process. Furthermore, We found some unique users from both datasets. Apart from that applying the intersection method to unique users of both datasets to extract common users which are shown in the

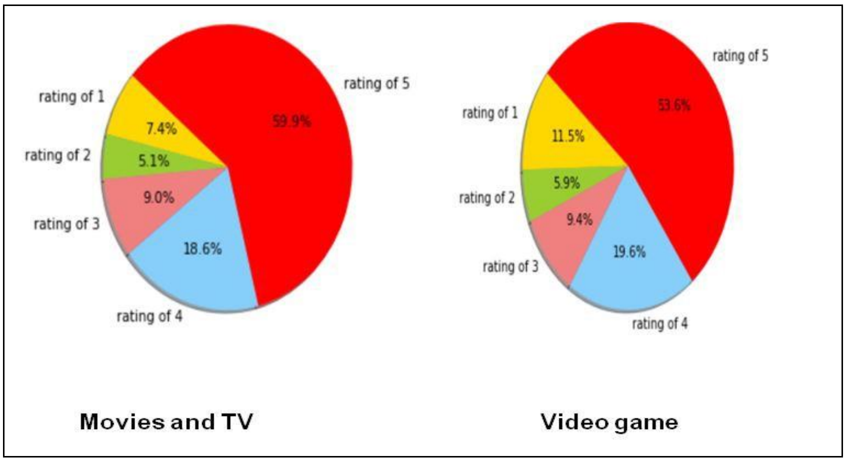

Fig 4: Visual Representation of Datasets

Figure 4 indicates visual representation of ratings dataset in percentage. The recommendation system involves providing customers with personalized recommendations when choosing an item from a product set (movies, books, etc.). Collaborative filtering is the most widely used technique for recommendation systems. Algorithms recommend predictive accuracy based on measurements of mean absolute error (MAE) and mean square error (RMSE).

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}$$

Where, MAE= mean absolute error yi = prediction
Xi = true value
n = total number of data points

$$\mathrm{RMSD} = \sqrt{\frac{\sum_{i=1}^{N} \left(x_i - \hat{x}_i\right)^2}{N}}$$

RMSD = root-mean-square deviation i = variable i
n = number of non-missing data points xi = actual observations time series

Table 2: Comparative Rmse & Mae of Algorithms

Sr. No.	Algorithms	RMSE	MAE
1	SVD	0.98	0.73
2	KNN with mean	1.18	1.02
3	Gridsearch	1.08	0.86

As shown in the above table we have evaluated RSME and MAE for the mentioned algorithm. As per the research, Singular Value Decomposition is the best for cross-domain Recommendation.

Conclusion

The recommendation system is the most remarkable concept in this modern era. It can be categorized into three types in which collaborative is the most important research area as there are many crux in it. Cross-domain recommendation aims to solve the problems in the recommendation system caused by sparse data by combining user preference data from multiple domains to improve the performance of the recommendation system. Crossdomains are also useful for dealing with cold stars in recommendation systems. During the implementation, we solved the item Cold-star problem by applying the intersection method and different algorithms. In the future, we will use more than one source domain to handle the same problem because it has more advantages.

Acknowledgment

The authors would like to thank the all people for their comments and suggestions to improve this paper. We also wish to express our sincere thanks to all faculty members of the Computre Science Department for their Cooperation and valuable suggestions in each stage, which lead to proceed of our work.

References

1. Bushra Alhijawi, Nadim Obeid, Arafat Awajan, Sara Tedmori. Improving collaborative filtering recommender systems using semantic information. In 2018 9th International Conference on Information and Communication Systems (ICICS), IEEE, 2018, 127-132.
2. Shlomo Berkovsky, Tsvi Kuflik, Francesco Ricci. Distributed collaborative filtering with domain specialization. In Proceedings of the 2007 ACM conference on Recommender systems, 2007, 33-40.
3. Laurent Candillier, Frank Meyer, Marc Boulle´. Comparing state- of-the-art collaborative filtering systems. In International Workshop on Machine Learning and Data Mining in Pattern Recognition Springer, 2007, 548-562.
4. Erion C¸ ano, Maurizio Morisio. Hybrid recommender systems: A systematic literature review. Intelligent Data Analysis,2017:21(6):1487-1524.
5. Iva´n Cantador, Ignacio Ferna´ndez-Tob´ıas, Shlomo Berkovsky, Paolo Cremonesi. Cross-domain recommender systems. In Recommender systems handbook, Springer, 2015, 919-959.
6. Paolo Cremonesi, Antonio Tripodi, Roberto Turrin. Cross-domain recommender systems. In 2011 IEEE 11th International Conference on Data Mining Workshops, Ieee, 2011, 496-503.
7. Ali Mamdouh Elkahky, Yang Song, Xiaodong He. A multi-view deep learning approach for cross domain user modeling in recommen- dation systems. In Proceedings of the 24th international conference on World Wide Web, 2015, 278-288.

8. Song Jie Gong, HongWu Ye, Heng Song Tan. Combining memory- based and model-based collaborative filtering in recommender system. In 2009 Pacific-Asia Conference on Circuits, Communications and Systems, IEEE, 2009, 690-693.
9. Will Hill, Larry Stead, Mark Rosenstein, George Furnas. Rec- ommending and evaluating choices in a virtual community of use. In Proceedings of the SIGCHI conference on Human factors in computing systems, pages, 1995, 194-201.
10. Tim Hussein, Timm Linder, Werner Gaulke, Ju¨rgen Ziegler. Hy- breed: A software framework for developing context-aware hybrid recommender systems. User Modeling and User-Adapted Interaction,2014:24(1):121-174.
11. Yaru Jin, Shoubin Dong, Yong Cai, Jinlong Hu. Racrec: review aware cross-domain recommendation for fully-cold-start user. IEEE Access,2020:8:55032-55041.
12. Masahiro Kazama, Istva´n Varga. Multi cross domain recommenda- tion using item embedding and canonical correlation analysis. In RecSys Posters, 2017.
13. Bin Li, Qiang Yang, Xiangyang Xue. Transfer learning for collab- orative filtering via a rating-matrix generative model. In Proceedings of the 26th annual international conference on machine learning, 2009, 617-624.
14. Chung-Yi Li, Shou-De Lin. Matching users and items across domains to improve the recommendation quality. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, 2014, 801-810.
15. Haifeng Liu, Zheng Hu, Ahmad Mian, Hui Tian, Xuzhen Zhu. A new user similarity model to improve the accuracy of collaborative filtering. Knowledge-based systems,2014:56:156-166.
16. Jie Lu, Vahid Behbood, Peng Hao, Hua Zuo, Shan Xue, Guangquan Zhang. Transfer learning using computational intelligence: A survey. Knowledge-Based Systems,2015:80:14-23.
17. Ashish K Sahu, Pragya Dwivedi. Matrix factorization in cross- domain recommendations framework by shared users latent factors. Procedia computer science,2018:143:387-394.
18. Ben Schafer J, Joseph Konstan, John Riedl. Recommender systems in e-commerce. In Proceedings of the 1st ACM conference on Electronic commerce, 2019, 158-166.
19. Yaqing Wang, Chunyan Feng, Caili Guo, Yunfei Chu, Jenq-Neng Hwang. Solving the sparsity problem in recommendations via cross- domain item embedding based on co-clustering. In Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, 2019, 717-725.
20. Yaqing Wang, Caili Guo, Yunfei Chu, Jenq-Neng Hwang, Chunyan Feng. A cross-domain hierarchical recurrent model for personalized session-based recommendations. Neurocomputing,2020:380:271-284.
21. Pinata Winoto, Tiffany Tang. If you like the devil wears prada the book, will you also enjoy the devil wears prada the movie? a study of cross-domain recommendations. New Generation Computing,2008:26(3):209-225.
22. Qian Zhang, Dianshuang Wu, Jie Lu, Feng Liu, Guangquan Zhang. A cross-domain recommender system with consistent information trans- fer. Decision Support Systems,2017:104:49-63.
23. Zihan Zhang, Xiaoming Jin, Lianghao Li, Guiguang Ding, Qiang Yang. Multi-domain active learning for recommendation. In Thirtieth AAAI Conference on Artificial Intelligence, 2016.
24. Feng Zhu, Yan Wang, Chaochao Chen, Jun Zhou, Longfei Li, Guanfeng Liu. Cross-domain recommendation: challenges, progress, and prospects. arXiv preprint arXiv:2103.01696, 2021.