J. Electrical Systems 20-3 (2024): 739-749

[1]Shwetal Raipure

[2] S. Vairachilai

An Efficient Cross-Domain Recommendation System Based on Clustering Approach and User-Score

Journal of Electrical Systems

Abstract: - Growing consumers and online commercial products have put a great challenge on Recommender systems. Abundant browsing and online shopping across the world have provided immense opportunities to E-commercial giants to effectively learn the behaviour and attract users. However, due to the variety of available products and distinct consumers, crucial issues of data sparsity and cold start problems have been introduced. Single-domain recommendation systems are unable to handle such issues and research has been concentrated on developing more sophisticated cross-domain recommendation systems. Concerning other systems, non-overlapping domains form the most difficult part to solve using the cross-domain approach. This article focuses on partially overlapping domains and suggests a simple and efficient approach based on mapping users' interests by clustering books and movies. The cluster information from the auxiliary domain is transferred to the target domain using the user score and mapped properly to recommend books. The user and item scores in both domains are computed using the transfer frequency-inverse document frequency approach. The proposed recommender system offers low computational complexity and requires less time.

Keywords: Recommender systems, data sparsity, cold-start problems, cross-domain recommendation systems, non-overlapping domains, partially overlapping domains, transfer frequency, and inverse document frequency.

I. INTRODUCTION

Increasing consumers and the launch of a new variety of product items by the different manufacturers over online shops (E-commerce) have introduced a great scope of doing business over the net. It is a practically impossible and tedious job for a consumer to explore the variety of items from the E-store. On the other hand, the cost of advertising all the items for a single consumer on his browsing page is not feasible. Recommending related or unrelated items on users' purchases is an ongoing fashion to explore items and attract customers. Recommendation Systems (RSs) are used by commercial organizations to learn and attract consumers' minds by recommending items respective to past purchases and displaying new products in the vicinity of present discoveries (users' cognitive aspects such as personality, behaviour, and attitude).

RS develops a win-win platform or atmosphere over the browsing window for the seller and the buyer. RS invokes the desire of a user by frequently recommending new items whenever the user opens his/her account and somehow converts the desire into need. The convincing power and rate of a good RS should be higher to catch the customer and sell out an article or good even when the customer had hardly initiated thinking of buying the product. Several single and cross-domain RS have been already incorporated by different commercial giant companies. They are enhanced for higher accuracy, good novelty, and diversity. Content-based (Cb) approaches use information relating to users and items whereas another approach called collaborative filtering (Cf) decomposes the user-item rating matrix (RM) using matrix factorization (MF) and uses the latent equalities between item characteristics and user interests for recommendation. These two commonly used techniques provide a better win-win platform for the users and the manufacturers and can handle cold-start users. The performance of Cf based on Mf depends greatly on the quality of the user-item RM. However, due to the large number of items and user approach (blind eye and time) toward providing reviews for every item, the RM is usually very sparse. The sparsity of RM imposes a great challenge in analysing the latent features whenever users are required to be grouped according to similarity.

Cf-based work carried out in [1-5] used users' similar characteristics of purchasing items. Likewise, authors in [6] used a Mf-based approach to work out the similarity between item and user interaction matrices. The user-item matrix was decomposed using the conventional Mf method (a singular value decomposition (SVD)) in [7]. The work proposed in [8-9] used weighted Mf and extended SVD along with the decomposition technique. A Bayesian

[1] [1] Research Scholar, School of Computing Science and Engineering, VIT Bhopal University, Kothrikalan, Sehore, Madhya Pradesh-466114, shwetal.raipure2019@vitbhopal.ac.in

[2] School of Computing Science and Engineering, VIT Bhopal University, Kothrikalan, Sehore, Madhya Pradesh-466114 vairachilai2676@gmail.com

probability-based Mf was introduced in [10] interpreting the same for recommendation. The work in [10] was extended for optimization using regulated Mf in [11]. The user-item latent features were made to obey the Gaussian distribution with mean=0. The gap between the negative and positive samples was enlarged using Bayesian personalized ranking through optimization [12]. User preferences were predicted through Cf by integrating rich information as auxiliary features [13-15].

Cf-based work carried out in [1-5] used users' similar characteristics of purchasing items. Likewise, authors in [6] used a Mf-based approach to work out the similarity between item and user interaction matrices. The user-item matrix was decomposed using the conventional Mf method (a singular value decomposition (SVD)) in [7]. The work proposed in [8-9] used weighted Mf and extended SVD along with the decomposition technique. A Bayesian probability-based Mf was introduced in [10] interpreting the same for recommendation. The work in [10] was extended for optimization using regulated Mf in [11]. The user-item latent features were made to obey the Gaussian distribution with mean=0. The gap between the negative and positive samples was enlarged using Bayesian personalized ranking through optimization

Recommending suitable items for the cold start users and the problem of data sparsity of the RM was solved using the idea of an auxiliary domain and the target domain [16]. Recent work uses rich information from multiple auxiliary domains and transfers it to a single or more target domain using a transfer learning approach. Usually, the concept of recommending items to users is based on the hypothesis that there exists a strong correlation between item preferences and user interests. The correlation can be related to user-item interactions, the relationship between latent factors, the similarity between user-item interest and features, and so on. The correlation is useful in filling the unknown or missing values of the RM of the target domain. Exploiting the latent similarities, today's recommendation systems however are based on sharing knowledge across the domains. The knowledge sharing can be accomplished using reviews, and ratings. Ignoring the latent sentiments, however, weakens the partial user sentiments in the reviews. Therefore, it is required to extract rich information from the auxiliary domains and transfer it to the target domains so that the users and items in the latent domain are mapped to the users and items of the former domains.

The definition of the notion of domain is very confusing. For example, some authors treat movies and books as similar domains while others treat them as distinct domains or sub-categories [17-18]. The significant thing about defining the notion of the domain is that items are given importance while the users are usually ignored. According to [19], domains are defined from three perspectives which include content-level, user-level, and item-level relevance.

II. RELATED WORK

Recommender systems are derived from the development of cognitive science, the theory of prediction and approximations, and information extraction. Today, the rise of modern business is the outcome of the rapid evolution after they came into existence. They are the intermediates and act as catalysts of the present business economy [20]. Redundant information is filtered from a vast pool of databases and presented to the users to fulfill their latent needs. Most of the recent recommender systems are based on users' interest and their ratings awarded to items [21]. Past interactions of the users are analyzed efficiently and are used to uncover the user's present demands. Starting from the Cf and Cb systems which were respectively initiated in 1992 and 1999, the 21st century recommendation systems are developed using Deep Learning methods due to the rise in the social networks. These methods can handle large amounts of data and multimodal information even using a mobile device.

Weak correlation between the auxiliary and the target domain, misaligned items, and dependency of the auxiliary domain on the user side were the issues considered in [22]. The authors proposed a CCA-based framework for aligning the item side, to retain the user features of the destination domain. Their method not only intensified weak correlation between the domains but also prevented the destruction of user features between domains. Thus only significant rich information was allowed to travel to the target domain from the auxiliary domain. They evaluated the performance of the CCA-based model using three datasets including the MovieLens 10M, EachMovie, and NetFlix, and compared the metrics with RMGM, CBT, and Funk-SVD baseline model. CCA imposes extra computational overhead since it requires full RM.

The work in [23] is focused on enhancing the performance to mitigate the problem of cold-start users while recommending books. The authors used two different approaches viz. memory and model-based to compare the performance. In the former case, MF-Cf, and for the latter case KNN and SVD were used on 5189 records. They found that the SVD worked better than KNN with Cf. Their Cf-SVD model was successful in delivering recommendations, however, better results can be obtained using Cf with demographic recommender systems considering the implicit feedback to handle the cold-start problem.

The limitation of using a single auxiliary domain to transfer rich information was discussed in [24]. The use of CF-Mf with a single domain having sparse data RM limits to handle new users was addressed by using multiple auxiliary domains. The authors in their work discovered coherency between both the source and the target domain by optimizing the objective function by adopting an implicit stochastic gradient descent algorithm. They consolidated multiple RMs inside their approach and validated the results over Amazon Food and MovieLens datasets. The prediction accuracy and the computational overhead were balanced by transforming the complete problem into a sub-problem of smaller or lower dimensions.

The work introduced in [25] exploited the use of user and item information for transferring knowledge to the target domain. The authors investigated the real-world scenario of data sparsity associated with the auxiliary domain due to privacy concerns. They considered non-overlapping domains and used user-item feedback to bridge the gap between both domains. They uplifted the domain-shared and domain-specific information from the feedback and proposed a mixed heterogeneous factorization framework. The accumulative strategy was adopted capturing shared knowledge and domain-specific information which included preference and characteristic heterogeneity, and rating bias. Experimental evaluation over Amazon Books, Yahoo Music 10, MovieLens 20M, and NetFlix showed that the proposed model was able to transfer effective knowledge across the domains and performed well in ranking performance.vThe algorithm is required to complement the missing values and a more

efficient and effective technique is required in that case.

Most of the cross-domain recommender systems are usually single-target systems which do not benefit the auxiliary domain. The authors in [26] considered three different scenarios with dual-target, multi-target, cross-domain recommendation plus cross-system recommendation. They adopted a unified GA framework based on Graph embedding and Attention techniques. A personalized learning strategy was proposed to prevent the negative transfer and to minimize the embedding difference of similar entities. User-item embeddings were generated using separate heterogeneous graphs and further combined the embeddings of similar or common entities using an element-wise attention mechanism. Experiments conducted on four datasets proved to mitigate the negative transfer to some extent. Common and domain-specific factors were used to benefit the cross-domain recommendation system [27]. The rich information (common factors + domain-specific) was effectively used and the middle matrix between the user and item membership matrices was used to map and align unique factors to transfer the implicit similarities to the target domain. Their Mf-based strategy of knowledge transfer for non-coupled or isolated dimensions outperformed others twice concerning accuracy. Their evaluation was projected on three datasets which included data from the Amazon website, the Bureau of Crime and Research, and the Australian Bureau of Statistics, however, their cross-domain recommendation framework lacked generalization ability.

Recent work [28] on achieving similar probability distribution in both domains owing to the feature space is handled using Deep Networks. The source domain limited samples are trained using the classifier and directly applied to the target domain. The work proposed in [29] used two separate layers for feature mapping and prediction. The samples from both domains were incorporated and a regularization term was used aligning them in the feature space to measure the maximum mean discrepancy between the samples. Inspired by the min-max game used in Generative Adversarial Networks (GAN) the work introduced in [30] compared the feature mapping module with GAN and distinguished samples from the domains using a domain classifier. The richness of the information was evaluated depending on the domain classifier. If there was a failure, extracted features were considered to be waste and carried no domain information. The extension to dynamic distribution adaption was applied to DAN (Deep Adversarial Networks) by [31] where the disparities between marginal and conditional distribution were marked.

The work carried out in [32] highlighted the essence of transfer frequency and inverse document frequency (Tf-Idf) in capturing rich information and its potential use in refining recommendations. Advanced techniques for recommendation systems use textual information about items in terms of their reviews, descriptions, or metadata. The Tf-Idf term converts textual information into vectors that embed the relative significance of terms within documents. It enhances the item's distinctiveness owing to the term's unique priorities. The Tf-Idf technique rather than using user-item interactions gleans insight from the item reviews. This unveils the implicit preferences by introducing an external, data-driven angle to the user-item matrices [33].

The problem of partially overlapping domains considered in this article is shown in Figure 1 below. The common attributes (genres in our case) are matched in both domains and depending on user interest in the target domain, items (books) are recommended from the source domain. Only the common attributes are selected from both domains, especially the domain having fewer attributes are mapped to the domain having more attributes. The unmatched attributes of a domain having more attributes are ignored and not considered for computation, except where the meaning of attributes is the same or more than one attribute belongs to the same category. The rows in both domains are non-overlapping and do not show any similarity.

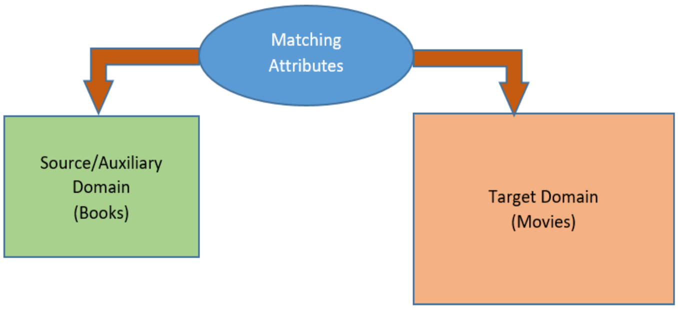

Figure. 1 – Problem of partial overlap in terms of attributes

The proposed cross-domain recommendation framework contributes in the following aspects:

1. It considers the case of domains partially overlapping across the attributes and suggests a simple and efficient approach for cross-domain recommender systems.
2. The framework uses a cluster-based approach to map user interest from the auxiliary domain to the target domain.
3. The knowledge transfer takes place through common attributes and is independent of the data sparsity of the target domain.
4. It offers low computational cost and requires less time to execute.

III. METHOD AND MATERIALS

The proposed cross-domain recommendation system based on common genres recommends books from the auxiliary domain to users who rated movies in the target domain. The source domain contains the movie-ID, title and year of the movie, and the genres. There are 9742 different movies and 20 genres associated with the target domain. The following Table 1 indicates the total genres associated with the movies. The last genre 'not genre listed' (Highlighted in Table 1) is ignored and only 19 genres are considered in the common domain. The release year of the movie was separated from the title and not taken into consideration in the recommendation process.

Table 1 – Genre available for the Movie domain (Alphabetically arranged)

S. N.	Movie Genre	S. N.	Movie Genre
1	Action	11	Horror
2	Adventure	12	IMAX
3	Animation	13	Musical
4	Children	14	Mystery
5	Comedy	15	Romance
6	Crime	16	Sci-Fi
7	Documentary	17	Thriller
8	Drama	18	War
9	Fantasy	19	Western
10	Film-Noir	20	**No genre listed**

The movie-ID's with respective titles were marked with the genres in binary form. The genres associated with the movies were marked with '1', otherwise a '0' was inserted. Another matrix consisting of user-ID, movie-ID, review ratings, and time stamp is attached with the movie data. The number of unique users who rated different movies is 610 where users have rated more than one movie (total movies being T = 9724). The real-valued scores of the ratings ranged from 1 to 5. The last column comprising the time stamp was ignored for this work. We constructed a rating matrix (I = 610 x T =9724) from the given movie user-rating information and filled the missing values by averaging the user-rating values. The missing values in each row were filled by computing the score mean of already-rated movies by the user. The following expression (1) represents the mean of scores by users.

$$\mathrm{UIRM}_{u,m} = \frac{1}{\alpha} \sum E_{x,y} \tag{1}$$

Where, 'u' are users and 'm' are movie entries such that m $\notin E_{x,y}$, and α are the number of movies already rated by the user. $E_{x,y}$ are the missing entries in the user ratings (row).

For each user, we find a list of movies (movie-ID) rated by him/her and subsequently form a separate independent rating matrix for the user. For example, the first user has rated 232 movies, and therefore a rating matrix of dimension 232x19 is constructed. To find the user's interest related to a specific genre, a sum of all columns is found (user genre frequency). The maximum value found in a particular genre slot identifies the user's interest relating to the genres. A value of zero corresponds to no user interest in the genre. The sum obtained for respective genres is added to get the total sum (user total genre value). Finally, we compute the user transfer frequency by dividing the user genre frequency by the user total genre value. The following expressions (2) to (4) explain the steps explained above.

User-specific genre matrix is obtained from the following expression (2).

$$U_I = MG_r \tag{2}$$

Where, r – are the movies (movie-ID) rated by the user. MG represents the movie-genre binary matrix.

Thus, U_i is the movie-genre matrix extracted for any user (from 610 users) from MG. It indicates the user interest concerning the preference given to the movies in terms of likeness towards a particular genre. The sum of column values (genre) reflects the user's interest. A higher sum means more likeness. The following expression (3) computes the user genre frequency and (4) calculates the total genre value corresponding to a user.

$$\mathrm{UGF}_k = \sum_{j=1}^{g} U_j \tag{3}$$

$$\mathrm{UTGV} = \sum UGF_k \tag{4}$$

Where, k is any user from 610 users, g – indicates 19 genres, while U_j is obtained using equation (2).

The user transfer frequency is finally obtained by (5).

$$\mathrm{UTF}_i = \frac{UGF_k}{UTGV} \tag{5}$$

The next step is to determine the Inverse Document Frequency (IDF) for the movie data. Here we consider the full movie-genre matrix and the frequency of each genre (Genre-Frequency) in all movies is found using the following expression (6 The same is obtained by column sum.

$$\mathrm{GF}_j = \sum_{j=1}^{g} MG_j \tag{6}$$

The IDF is given by the following expression (7) and the TF-IDF score is evaluated using expression (8)

$$\mathrm{IDFg} = \log_{10} \frac{T}{GF_g} \tag{7}$$

'g' takes the values from 1 to 19.

$$\text{TF-IDF-score\_m} = \mathrm{UTF}_i \times \mathrm{IDF}_g \tag{8}$$

The user transfer frequency (UTF) matrix has a dimension of 610x19 and the Inverse document frequency (IDF) is of size 1x19. The latter row matrix is repeated 610 times to perform the element-wise multiplication.

The Book data contains about 10635 books listed with 100 unique attributes. A book-attribute (genre) matrix was obtained in binary form setting value '1' for the specified listed genre and '0' for the absence of the genre. To map columns in movie data as indicated in Table 1 to book data, matching attributes in both domains were manually found. Some attributes of the book data were used in cooperation to match a single attribute in the movie data to reduce the sparsity of the book data matrix. The equivalent column was obtained by logical OR operation performed on multiple similar columns (attributes) having the same meaning. Table 2 shows the mapping of book genres with the movie genres. Therefore, 25 columns were used to map attributes in both domains from the book data.

Table 2 – Genre Mapping between Movies and Books (+ indicates logical OR along all books)

Movie Genre	Book Genres
Action	Action
Adventure	Adventure
Animation	Comics + Magic + Read for school + School + Teen
Children	Children
Comedy	Comedy
Crime	Crime
Documentary	Art + Short Stories
Drama	Drama
Fantasy	Fantasy + Urban Fantasy
Film-Noir	Film-Noir
Horror	Horror
IMAX	IMAX
Musical	Musical
Mystery	Mystery
Romance	Romance
Sci-Fi	Sci-Fi
Thriller	Thriller
War	War
Western	Western

After mapping the columns from the book data, the books having no genre corresponding to the genre in the movie data were eliminated. As a result, out of 10635 books, only 7641 books made it to the book genre matrix while 2994 books were eliminated. We converted the genre information from binary form to decimal values in a row-wise fashion (genre-based). This was done to cluster the books based on a common genre they carried. The number of groups or clusters based on a common genre was found to be 1318.However, not all groups from 1318 are significant due to the scarcity of group members (books) in most of the groups. The proposed framework considers clusters concerning the available genres. Thus, the work recommends books to the user based on the user's highest interest respective to the genre. The number of clusters can be improved by considering different combinations of the genre. For 19 genres, 2^{19} different clusters are possible. However, not all combinations exist, and significant clusters can be obtained using a threshold for the number of book members in each cluster. Too many clusters eventually weaken the cluster size and lower the recommendations. Also, with too few clusters, some clusters may remain unoccupied for the users in the auxiliary domain. The following Table 3 shows the number of books in each of the clusters formed concerning individual genres. There is a possibility of the presence of the same book in different clusters, therefore the sum of the second column in Table 3 will always exceed the total number of available books.

Table 3 – Clusters and their member book count

Book Genre	Number of Books
Action	214
Adventure	1286
Animation	1698
Children	1213
Comedy	367

Crime	481
Documentary	2206
Drama	969
Fantasy	1383
Film-Noir	116
Horror	479
IMAX	26
Musical	150
Mystery	1037
Romance	2149
Sci-Fi	4340
Thriller	519
War	501
Western	1260

Once the clusters/groups in the target domain are formed, the group transfer frequency and the inverse document frequency for the book domain are computed using the following equations (9) and (10) respectively.

Book-specific genre matrix is obtained from the following expression (9).

$$B_I = BG_r \tag{9}$$

Where, r – are the books (book-ID) corresponding to each of the genres. BG represents the book-genre binary matrix. Thus, B_i is the book-genre matrix extracted for any genre (from 19 genres) from BG. whereas, the transfer frequency for a genre in the book domain is computed using equation (10).

$$BTF_g = \frac{B_i}{\sum_g \sum_r B_i} \tag{10}$$

The inverse document frequency in the book domain is computed using the following expression (11). Even though the book-genre matrix is sparse, the inverse document frequency is a filled vector. That is, there are no missing elements in the book IDF.

$$BIDFg = \log_{10} \frac{S}{BGF_g} \tag{11}$$

S – represents books qualified for the common genres in both domains out of the total books in the book domain. 'g' takes the values from 1 to 19. BGF $_j = \sum_{j=1}^{g} MG_j$ is the book genre frequency computed by summing the columns for each genre. Finally, the TF-IDF-score in the target domain is given by expression (12).

$$\text{TF-IDF-score\_b} = BTF_g \times BIDF_g \tag{12}$$

The user transfer frequency (BTF) matrix has a dimension of 19x19 and the Inverse document frequency (BIDF) is of size 1x19. The row vector of BIDF is repeated 19 times to perform the element-wise multiplication.

To recommend related books for a user watching movies, the proposed framework accepts user input in terms of user-ID from the auxiliary domain (movie) out of 610 users. For a giver user-ID, the user score is computed using the following mathematical expression (13).

$$Uscore_{id} = E_{id}\,(\text{TF-IDF-score\_m}) \tag{13}$$

Where $Uscore_{id}$ is the user score for the user-ID input in the auxiliary domain. E_{id} represents the corresponding row entry in the TF-IDF-score matrix computed for the movie domain. The user score represents the user interest concerning all the movie genres and is used to find similar configurations in the book domain for recommendations. We have used the well-known cosine similarity (CS) metric to find the best cluster in the target domain. The CS is computed using the expression (14) which measures the distance between two points. The similarity decreases as the distance increases. That means, to recommend books for a user watching movies, the user score in the movie domain must have the minimum distance with the book score.

$$CS = \frac{\sum user\_score\ x\ book\_score}{\sqrt{user\_score^2\ x\ book\_score^2}} \tag{14}$$

Where $Uscore_{id}$ = user_score and TF-IDF-score_b = book_score. All the multiplication and division operations are performed element-wise. The CS is computed for genres respective to the user-ID, and the index of minimum value is found. The index represents the genre most liked by the user while watching the movies. The books corresponding to the genre (index) indicated in Table 3 are recommended to the user in the movie domain. Each cluster formed in the book domain carries a varying number of books. We recommend any 10 random books from the set/cluster.

IV. RESULTS AND DISSCUSSION

The proposed cross-domain recommendation approach uses genre information from both domains to recommend books to users in the movie domain. Since the size of the clusters in both domains was not restricted, clusters in the book and movie domains were of varying sizes. This was done to prevent the construction of small books and user clusters. The datasets have no common users and rely on partial overlap of the genres in both domains. The TF-IFD approach used for cross-domain recommendation is independent of the data sparsity of the target domain while the user ratings in the auxiliary domain were filled with mean values of already rated genres by the user. The timestamp field of the movie domain and the year of release field of the book were ignored. However, they can be effectively used for cold-start users. The following figures 2 to 4 show the books (first 10 entries from the cluster) recommended when different user-ID were provided to the proposed system.

For user-ID = 600, the first 10 recommendations were,

```
{[2498]}    {'the bobbsey twins at the'            }
{[3504]}    {'the triumphs of eugne valmont'       }
{[3505]}    {'the second jungle book'              }
{[3991]}    {'the colors of space'                 }
{[5471]}    {'white lies'                          }
{[5472]}    {'operation haystack'                  }
{[5473]}    {'final weapon'                        }
{[5474]}    {'spontaneous activity in education'}
{[5668]}    {'the sorrows of young werther'        }
{[5882]}    {'sea power'                           }
```

Figure 2 – First 10 recommendations for user-ID = 600 with book-ID and title

For user-ID = 5,

```
{[1978]}    {'the confessions of nat turner'}
{[1979]}    {'madame chrysantheme'          }
{[1980]}    {'australian history'           }
{[1981]}    {'joululahjat'                  }
{[1982]}    {'a little book for christmas'  }
{[1983]}    {'blown to bits'                }
{[1984]}    {'h g wells'                    }
{[1985]}    {'venereal diseases'            }
{[1986]}    {'venereal diseases'            }
{[1987]}    {'a first spanish reader'       }
```

Figure 3 – First 10 recommendations for user-ID = 5 with book-ID and title

For user-ID = 105,

```
{[1859]}    {'a briefe introduction to geography'  }
{[1860]}    {'platos republic'                     }
{[1861]}    {'readings in the history of education'}
{[1862]}    {'historic papers'                     }
{[1863]}    {'qvo vadis'                           }
{[1864]}    {'the keeper of the door'              }
{[1865]}    {'winnie childs'                       }
{[1866]}    {'tunnustus'                           }
{[1867]}    {'prose fancies'                       }
{[1868]}    {'selbstbetrachtungen'                 }
```

Figure 4 – First 10 recommendations for user-ID = 105 with book-ID and title

The minimum size group (26 samples) was obtained for the IMAX attribute while animation forms the largest group with 1698 samples. The performance of the recommendation system can be improved by increasing the groups to some extent and that can be achieved by considering genre combinations instead of using a single genre. Eventually, the computational cost will rise to find more groups from the book domain. As stated earlier, almost 1318 combinations (cluster/group) were possible from 2^{19} (Binary entries for genre) combinations. However, most of the combinations were not significant and occupied one or two book samples. However, an effective and efficient clustering algorithm can be used to cluster the book domain so that equal-sized clusters can be obtained.

V. CONCLUSION

The proposed cross-domain recommender framework is based on finding common features from the partially overlapped domains. The mechanism of knowledge transfer is built on user interest. The user’s interest in one domain is used to map interest in another domain. The strategy uses common factors from both domains to correlate the items. The proposed approach is simple and easy to implement. It can be used to solve problems for cold-start users considering other factors such as year and time stamp from the target and auxiliary domain respectively. The system is independent of the data sparsity in the target domain and the issue of missing values in the auxiliary domain can be solved using some other approach.

The performance of the system was tested and validated by manually comparing the actual user interest in the movie dataset and the book dataset respective to the genre field. However, the proposed framework has limitations regarding the manual mapping of genres from the book dataset corresponding to movie genres. Also, the cluster size is not limited, which results in varying size clusters in both domains. Eventually, the samples in both domains are not finely mapped between existing clusters. The cluster boundaries are not fine to classify the books properly due to few clusters and large clusters engulfing most of the samples with respect to other small-size clusters. Future work will be focused on these issues primarily on cluster dimensions, and matching common factors.

References

[1] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua, “Neural collaborative filtering,” in Proc. 26th Int. Conf. World Wide Web, , 2017, pp. 173–182.

[2] T. Man, H. Shen, X. Jin, and X. Cheng, “Cross-domain recommendation: An embedding and mapping approach.” in *Proc. 30th Int. Joint Conf. Artif. Intell.*, 2017, pp. 2464–2470.

[3] J. Lian, F. Zhang, X. Xie, and G. Sun, “CCCFNet: A content-boosted collaborative filtering neural network for cross-domain recommender systems,” in *Proc. 26th Int. Conf. World Wide Web Companion*, 2017, pp. 817–818.

[4] K. Georgiev and P. Nakov, “A Non-IID framework for collaborative filtering with restricted Boltzmann machines,” in *Proc. Int. Conf. Mach. Learn. Res.*, 2013, pp. 1148–1156.

[5] J. Bobadilla, A. Hernando, F. Ortega, and A. Gutiérrez, “Collaborative filtering based on significances,” *Inf. Sci.*, vol. 185, no. 1, pp. 1–17, 2012.

[6] Y. Koren and R. Bell, “Advances in collaborative filtering,” Recommender Systems Handbook, pp. 77–118, 2015.

[7] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems, *Comput"* vol. 42, no. 8, pp. 30–37, Aug. 2009.

[8] Y. Hu, Y. Koren, and C. Volinsky, "Collaborative filtering for implicit feedback datasets," in *Proc. 8th IEEE Int. Conf. Data Mining*, 2008, pp. 263–272.

[9] Y. Koren, "Factorization meets the neighborhood: A multifaceted collaborative filtering model," in *Proc. 14th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2008, pp. 426–434.

[10] R. Salakhutdinov and A. Mnih, "Bayesian probabilistic matrix factorization using Markov chain Monte Carlo," in *Proc. 25th Int. Conf. Mach. Learn.*, 2008, pp. 880–887

[11] A. Mnih and R. R. Salakhutdinov, "Probabilistic matrix factorization," *Adv. Neural Inf. Process. Syst.*, vol. 20, pp. 1257–1264, 2007.

[12] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, "BPR: Bayesian personalized ranking from implicit feedback," 2012, [Online]. Available: arXiv:1205.2618

[13] Z. Sun, Q. Guo, J. Yang, H. Fang, G. Guo, J. Zhang, and R. Burke, "Research commentary on recommendations with side information: A. survey and research directions," *Electron. Commerce Res. Appl.*, vol. 37, Art. no. 100879, 2019.

[14] F. Zhao, M. Xiao, and Y. Guo, "Predictive collaborative filtering with side information." in *Proc. 30th Int. Joint Conf. Artif. Intell.*, 2016, pp. 2385–2391.

[15] Y. Fang and L. Si, "Matrix co-factorization for recommendation with rich side information and implicit feedback," in *Proc. 2nd Int. Workshop Inf. Heterogeneity Fusion Recommender Syst.*, 2011, pp. 65–69.

[16] Cremonesi, P.; Tripodi, A.; Turrin, R. Cross-domain recommender systems. In Proceedings of the 11th International Conference on Data Mining Workshops, Vancouver, BC, Canada, 11 December 2011; pp. 496–503.

[17] Hu et al., 2013] Liang Hu, Jian Cao, Guandong Xu, Longbing Cao, Zhiping Gu, and Can Zhu. Personalized recommendation via cross-domain triadic factorization. In WWW, pages 595–606, 2013.

[18] Cao et al., 2010] Bin Cao, Nathan Nan Liu, and Qiang Yang. Transfer learning for collective link prediction in multiple heterogeneous domains. In ICML. Citeseer, 2010.

[19] Feng Zhu, Yan Wang, Chaochao Chen, Jun Zhou, Longfei Li, and Guanfeng Liu, "Cross-Domain Recommendation: Challenges, Progress, and Prospects," Proceedings of the 13th International Joint Conference on Artificial Intelligence (IJCAI-21), pp. 4721-4728.

[20] Zhang, S.; Yao, L.; Sun, A.; Tay, Y. Deep learning based recommender system: A survey and new perspectives. ACM Comput. Surv. (CSUR) 2019, 52, 1–38.

[21] Yang, Z.; Ding, M.; Zou, X.; Tang, J.; Xu, B.; Zhou, C.; Yang, H. Region or global a principle for negative sampling in graph-based recommendation. IEEE Trans. Knowl. Data Eng. 2022, 35, 6264–6277.

[23] Esmael Ahmed and Adane Letta, "Book Recommendation Using Collaborative Filtering Algorithm," Applied Computational Intelligence and Soft Computing, Volume 2023, 1514801.

[24] Nam D. Vo, Minsung Hong, and Jason J. Jung, "Implicit Stochastic Gradient Descent Method for Cross-Domain Recommendation System," Sensors, 2020, 2510.

[25] Ting Yu, Junpeng Guo, Wenhua Li, and Meng Lu, "A mixed heterogeneous factorization model for non-overlapping cross-domain recommendation," Decision Support Systems, Volume 151, December 2021, 113625.

[26] Feng Zhu, Yan Wang, Jun Zhou, Chaochao Chen, Longfei Li, and Guanfeng Liu, "A Unified Framework for Cross-Domain and Cross System Recommendations," IEEE Transactions on Knowledge and Data Engineering, vol. 35, No. 2, February 2023.

[27] Quan Do, Wei Liu, Jin Fan, and Dacheng Tao, "Unveiling Hidden Implicit Similarities for Cross-Domain Recommendation," IEEE Transactions on Knowledge and Data Engineering, vol. 33, No. 1, 2021.

[28] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84–90.

[29] Ghifary, M., Kleijn, W. B., & Zhang, M. (2014). Domain adaptive neural networks for object recognition. In Pacific Rim International Conference on Artificial Intelligence (pp. 898–904).

[30] Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., et al. (2016). Domain-adversarial training of neural networks. The Journal of Machine Learning Research, 17(1),1-35.

[31] Zhao, Y., Wang, K., Guo, G., & Wang, X. (2022). Learning compact yet accurate generative adversarial networks for recommender systems. Knowledge-Based Systems, 257, Article 109900.

[32] Manning, C. D., Raghavan, P., Schütze, H.: Introduction to Information Retrieval. Cambridge University Press, Cambridge (2008).

[33] B. H. Ahmad (ed.), Proceedings of the 2023 International Conference on Data Science, Advanced Algorithm and Intelligent Computing (DAI 2023), Advances in Intelligent Systems Research 180, https://doi.org/10.2991/978-94-6463-370-2_62 .