ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645
Indian Journal of Science and Technology, *Vol 10(29), DOI: 10.17485/ijst/2017/v10i29/115863, August 2017*

Collaborative Recommender Systems using User-item's Multiclass Co-clustering

Mugdha Adivarekar* and Vina Lomte

Department of Computer Engineering, RMDSSOE, Pune – 411502, Maharashtra, India;
mugdha.adivarekar@gmail.com, vinalomte.rmdssoe@sinhgad.edu

Abstract

Recommender Systems are playing very crucial and vital role in day today's life. People are very active on e-commerce sites as they get whatever they want at home. Some financial recommender sites like Money Control are getting popular due to their variety of sectors. These systems actually work on basis of Collaborative filtering model and apply knowledge discovery techniques for live interaction with person. E-commerce sites also provide top-N recommendations to users when they log in to system based on their previous shopping or surfing or interests. Hence collaborative filtering is most used technique over the decade. **Objectives**: 1. To find meaningful subgroups, formulate the Multiclass Co-Clustering (MCoC) algorithm and propose an effective solution to it. 2. Applying traditional and easily adoptable CF algorithm to merge Top-N recommendation results. **Methods/Analysis:** Collaborative filtering methods have been applied to different data like Sensing and monitoring data, financial data, and Electronic commerce and web applications. **Findings**: Simplicity, efficiency and Classification accuracy are most important feature provided by CFA. It is more natural to provide recommendations based on correlated user groups but it's not mandatory that one user must have interest in other things that are liked by people in group. **Novelty /Improvement**: This approach can be considered as an extension of traditional clustering CF models. Multiclass co-clustering technique proposes to generate top-N recommendations by maintaining user-item interaction matrix and making clusters by matrix factorization.

Keywords: Collaborative Filtering, Information Retrieval, Multiclass Co-clustering, Recommender System, User Profile

1. Introduction

This category of research falls under the domain data mining and information retrieval. Here one need to have some important concepts very clear.

Information retrieval - Information retrieval is process or act of getting information and data resources relevant to needed information from a collection of information resources. Full-text or other content-based indexing techniques can be used for search.

Information filtering - Information filtering system is a system that removes redundant and unwanted information from an information stream before presenting it to a user. Its main goal is the management of the information overload on web and increment of the semantic signal-to-noise ratio.

**Author for correspondence*

Recommender system - Recommendations are the filtered information that comes in form of the suggestions. Recommender systems or recommendation systems are a subclass of information filtering system that find to judge the "rating" or "preference" that a user would give to an item.

User profile - A user profile is a visual display of personal data associated with logged in user, or a customized environment by user. A profile is digital representation of a person's identity about his interests or likings.

1.1 Collaborative Filtering

Collaborative filtering is method of information filtering and it associate user with like-minded people or user group. It is used to create recommendation system. For ex. Amazon shopping site, music site. These systems are inspired by users surfing trend like some users prefers electronics, clothing, jewelry, kitchenware, etc. This includes explicit and implicit user-item interaction. Implicit interactions are cookies and sessions of browser and explicit interactions are provided rating or feedback etc.

In recommender systems, CF algorithms are mainly classified into two types: Memory-based algorithms and model based algorithms. In memory-based CF algorithms, the entire user-item rating matrix is directly used to predict unknown ratings for each user. User-based and item-based CF algorithms are the well-known methods that fall into this category. User-based CF methods finds few nearest neighbors with high similarities for each user and make predictions based on the weighted average ratings of his or her neighbors. The neighbors can be determined by different similarity measures like Pearson correlation coefficient and cosine similarity in rating space. Similarly, item-based CF methods find the nearest neighbors for each item. The similarity computation processes computationally expensive for large data sets and neighbors cannot be found accurately in highly sparse data. In model based CF algorithms, a predictive model is trained from observed ratings in advance. Latent factor models can reduce data sparsity through dimensionality reduction and usually generate more accurate recommendations than memory-based CF algorithms. When numbers of existing users and items grow tremendously, traditional CF algorithms like memory based or model based, will suffer a serious scalability problem and computational cost goes beyond practical or acceptable levels. Clustering CF models address the scalability problem by making recommendations within smaller clusters instead of the entire database, demonstrating promising performance in trade-off between scalability and recommendation accuracy.

2. Research Methods

Different survey/research papers and their results oriented study for understanding user-item matrix representation and traditional CFA are very useful. Observed facts, realistic scenarios and new techniques and their advantages – disadvantages are as below:

[1]Proposed primitive and simplest technique for implementation of User-based recommendation system and demonstrated its simplicity, efficiency in comparative manner with help of Pearson's correlation coefficient.

To overcome some drawbacks of User-User based recommendations Item based Collaborative Filtering Recommendation Algorithms[2] were introduced. Item based CF algorithms analyze different item based recommendation algorithms where item - item similarities are computed with the help of cosine similarity, correlation similarity.

[3]Improved Collaborative Recommender Systems via User-Item subgroups. This is the first research in user-item subgroups technique and proposes multiclass co-clustering of user-itemmatrix.[4]A survey on Parallel Hybrid Multigroup co-clustering using Collaborative Filtering Model to deal with heterogeneous sources of information where hybrid clustering can be used. Also

solution to overcome data sparsity and scalability problems is proposed.

In research of Recommendation system there are diversified enhancement that came into picture because of need of the time and growth of E-commerce. Another approach defined[5] was Sparse Linear Methods for Top-N Recommender Systems which uses Sparse Linear method to generate top-N recommendations by aggregating from user purchase/rating profiles. Sparse aggregation coefficient matrix is learned from SLIM by solving l1 and l2 normalization.

[6]Addressed new and improvised approach for Collaborative Filtering via Scalable User-Item Co-clustering. This research supports same techniques of user-tem matrix factorization for making clusters is used along with traditional CF algorithms. This approach has advantages like scalability, flexibility, interpret ability and extensibility.

[7] Focuses Item/User Representation for Recommender systems based on Bloom filters, with which user-item relation is represented in a bloom filter vector and authors have proposed a method to compute bitwise AND and XNOR similarity by using those bloom filters. Bloom filter is bit structure that allows to represent a set of elements in very low space so this technique is used in case of low data structures.

[8]Generation of Interpretable Recommendations via overlapping co-clusters is major area of work and proposed an algorithm which uses matrix factorization technique to identify co-clusters and recommends client-product pair because of its membership in one or more client-product co-cluster. This is applicable to large datasets. This approach is capable of offering textually and visually interpretable recommendations.

A Survey of Collaborative Filtering Based Recommender Systems for Mobile Internet Applications[9] was useful to implement recommendation systems with Mobile applications by altogether comparing various collaborative filtering methods, categorizing into pros and cons and they are classified into memory based and content based approaches.

[10] Theme or technique that deals with Collaborative filtering algorithms which is implemented by regularized matrix factorization. Here matrix is nothing but User-item interaction matrix based on which 2 new matrices are generated which will include predicted ratings with original ratings. Data loss equation is used to avoid data loss while updating matrix through machine learning algorithm.

A New Recommender System Using Context Clustering Based on Matrix Factorization Techniques[11] has been based on matrix factorization technique to get high accuracy results. This approach uses k-mode algorithm to reduce complexity of matrix and increase relevance of user-item matrix.

[12]Advanced Hybrid Multi Co-Clustering framework which can cluster users and items into multiple groups simultaneously with different information resources. And then apply conventional CF algorithms in each cluster to make predictions. By merging these predictions top-N recommendations are given.

2.1 Existing Methodologies

Following techniques are used for current recommender systems:

2.1.1 Baseline Predictors

These methods are useful for launching non-personalized baselines against which custom-made algorithms can be compared, as well as for pre-processing and normalizing data for use with more refined algorithms.

Baseline algorithms that do not depend on the user's ratings can also be advantageous for providing estimates for new users. If an item or user is new and therefore has no ratings, its baseline can be set to 0, effectively assuming that it is an average user or item.

2.1.2 User-user Collaborative Filtering

User - user collaborative filtering also known as k-NN collaborative filtering, was the first of the automated CF methods. User-user CF is a straightforward algorithmic interpretation of the core principle of collaborative filtering: find other users whose past rating behavior is similar to that of the current user and use their ratings on other items to predict what the current user will like.

2.1.3 Item-item Collaborative Filtering

User-user collaborative filtering is effective but suffers from scalability problems as the user base grows. Searching for the neighbors of a user is an O(N) operation. To extend collaborative filtering to large user bases and facilitate deployment on e-commerce sites, it was necessary to develop more scalable algorithms.

Item - item collaborative filtering, also called item-based collaborative filtering, takes a foremost step in this direction. Rather than using resemblances between users rating behavior to predict preferences, item - item CF uses resemblances between the rating patterns of items. If two items tend to have the same users like and dislike them, then they are similar and users are expected to have similar preferences for similar items. As a user rates and re-rates items, their rating vector will change along with their similarity to other users.

2.1.4 Dimensionality Reduction

In both of the traditional collaborative filtering algorithms user-user and item-item, there are suggestions of viewing the user- item ratings domain as a vector space. With this view, however, the vectors are of extremely high dimension; further, there is redundancy in these dimensions, as both users and items will typically be divisible into groups with similar preference profiles. Latent semantic analysis, a technique established in information retrieval, provides a way to do this decomposition using only the rating data.

2.1.5 Probabilistic Methods

Several probabilistic formulations of collaborative filtering have been suggested and these methods generally aim to build probabilistic models of user behavior and use those models to expect future behavior. Personality judgment is a probabilistic user model that assumes that a users' ratings are a combination of their preference and Gaussian noise.

2.1.6 Hybrid Recommenders

It is expected to consider combining several different recommender algorithms into a hybrid recommender system. In some applications, hybrids of various types have been found to beat individual algorithms. Hybrids can be particularly advantageous when the algorithms involved cover different use cases or different features of the data set.

There are 3 major steps in existing systems: object data collections and representations, similarity decisions, and recommendation computations.

Maintaining/building relationships in new individual and the existing data is very difficult to avoid cold-start. These systems need to improve correlation-based collaborative filtering and performing clustering on item ratings from users.

2.1.7 Drawbacks

Since existing recommender systems are not covering all 3interactions user-user, user-item and item-item simultaneously, this proposed technique will assure any user guaranteed top-N recommendations.

2.2 Problem Analysis

We find that two users with similar tastes on one item subset may have totally different tastes on another set. It is more natural to make preference predictions for a user via the correlated subgroups than the entire user-item matrix.CF based on only user-item interaction and co-clustering; but cannot determine relation of User-to-item, Item-to-item, User-to-user simultaneously.

ex. 2 people like comedy genre of movies but it's not necessary that anyone of them would not like action genre. Hence recommendation system should not treat users in traditional pattern.

3. Proposed System

After careful analysis the system has been identified to have the following modules:

1)Information Filtering Module- As per growing information on web it has become necessary to filter out necessary information resources from large set of sources.

2)Collaborative Filtering Module- Collaborative filtering algorithms helps to determine relative information from some of the shortlisted sources.

3)Recommendation System Module- Recommendations are generated on basis of current and historical data that can be represented as User-item matrix. There is more attention on meaningful subgroup discovering and less attention on tradition CF algorithms.

This also involves

i) Generate missing data by measuring similarity

ii) Create new algorithm to form MCoC subgroups.

iii) Compare results with existing CFA.

User-item matrix can be showcased as in Figure 1.

User interacts with e-commerce system which internally keeps track of user's surfing trend and gives recommendations based on user's interest.

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
User 1	0	3	0	3	0	0
User 2	4	0	0	2	0	0
User 3	0	0	3	0	0	5
User 4	0	0	0	0	3	0
User 5	4	0	0	4	0	0

Figure 1. User item matrix.

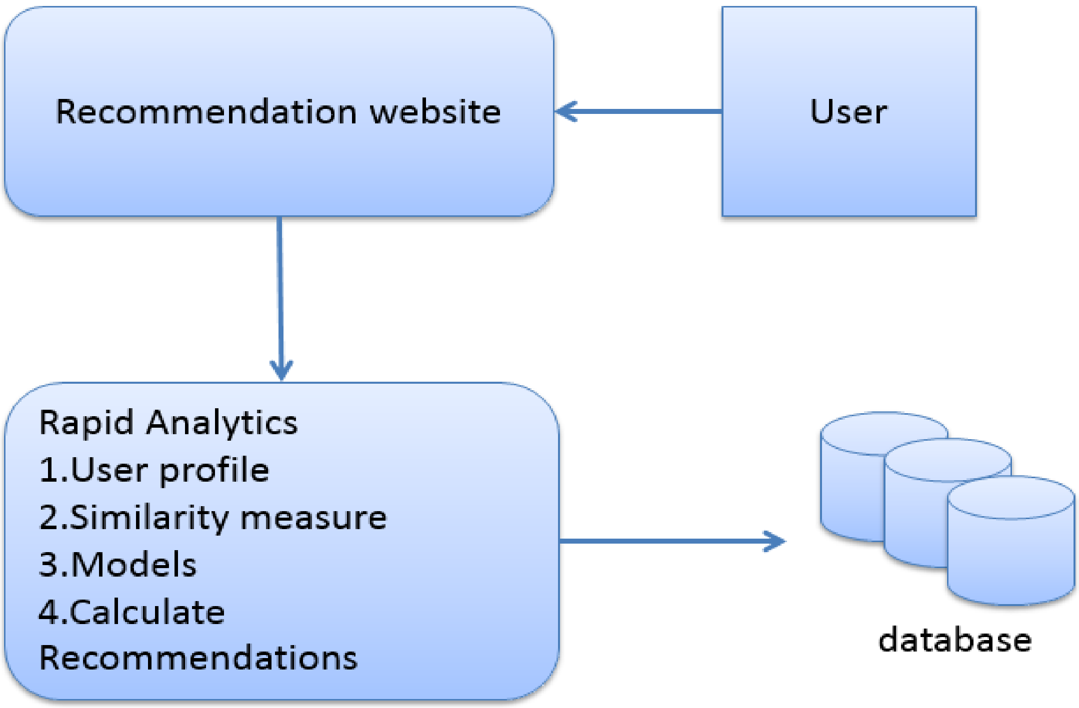

Figure 2. Architecture.

These recommendations are computed with the help of user profiles and these user profiles are maintained in the form of user-item matrix as shown in Figure 1.

As shown in Figure 2 Architecture- Modules of the System can be represented as multifunction block.

3.1 Mathematical Model

Pearson's correlation formula is used for finding missing data.

This is useful when data center is failed to update user-item matrix and some values are missing in it.

We have this formula below:

$$sim(a,b) = \frac{\sum_{p\in P}(r_{a,p} - \vec{r_a})(r_{b,p} - \vec{r_b})}{\sqrt{\sum_{p\in P}(r_{a,p} - \vec{r_a})^2}\sqrt{\sum_{p\in P}(r_{b,p} - \vec{r_b})^2}}$$

Where,

a,b : users

$r_{a,p}$:rating of user a for item p

P: set of items, rated both by a and b (Possible similarity values between 1 and 1);

$\vec{r_a}$ And $\vec{r_b}$: User's average ratings

Now R be the Rating matrix for m users and n items.

Rows of rating matrix are represented by Users and columns are represented by Items.

L be the no.of Subgroups i.e clusters. Hence $Z \in R^{(m+n)\times r}$

Subgroups of this Rating matrix can be denoted as V_k and for group V_k sub matrix of rating matrix is defined as

$R_k \in R^{(mk \times nk)}$

Where k = 1, L.

m_k and n_k are the number of users and items in that group, respectively

4. System Analysis

In recommendation engine matrix dimensionality reduction, matrix factorization, applying conventional CF algorithm is performed to generate recommendations.

4.1 Algorithm

Step1: Dimensionality Reduction of User-Item Matrix
Step2: Find subgroups i.e. Partition matrix
Step3: Get Top-N recommendations using traditional CF algorithm.

4.2 Advantages

User can have functionality like sign up, surfing for interested products, viewing recommendations based on previous searches.

Existing recommender systems are not covering all 3 interactions user-user, user-item and item-item simultaneously, this system analyses those and will assure any user guaranteed top-N recommendations.

5. Proposed Algorithm

Following algorithm for Top-N recommendation can be proposed once User-Item matrix representation is updated. This is available in[12]

***ALGORITHM**: Top-n Recommendation Algorithm*

***Input** : Rating Matrix* $R \in \mathbb{R}^{m\times n}$ *, all the groups* $\{V_1, \ldots, V_L\}$*, a chosen CF method, and the number of items in recommendation list N*

***Output** : Recommendation list for each user*

for** k ← 1 **to** L **do

Extract submatrix R_k *from rating matrix R with users and items belonging to group* V_k *;*

Apply CF recommendation method with R_k *as input and predict missing scoresr*(u_i, i_j, k).

end

for** i ← 1 **to** m **do

for** j ← 1 **to** n **do

if** Rij is missing **then

Find group index $k = \{\max Y_{ik} \mid Y_{ik} \neq 0 \text{ and } Y_{jk} \neq 0\}$

If** k is **null then

Set $\widehat{Rij} = 0$

end

else

Set $\widehat{Rij} = r(u_i, i_j, k)$

end

end

end

Generate top-N recommendation list for user u_i *according to the decreasing order of the predicted score.*

end

6. Experimental Setup and Results

Proposed recommender system is implemented in architecture framework with INTEL 2.3 GHz i3 processor and 4 GB RAM and .NET (Visual Studio 2013).

We use following parameters for comparative analysis:

1. Storage cost - Storage will be less comparatively because data related to user is maintained for limited period and then archived with MS SQL archiving policies.
2. Computing cost - Representing stored data in user-matrix form will not require more time than computing one-by-one user profile data.
3. Accuracy of results - Comparing recommendations given by traditional approach to proposed approach and validating them.

7. Applications

- Creating Web log pattern applications
- Recommender systems for shopping sites like Amazon, flip kart etc.
- User preference arranger for financial applications.
- Monitoring and sensing the data to provide predictions
- Providing advertises of product over web as per user's interests.
- Travel/tourism recommendations

8. Conclusion

As compared to traditional CF techniques of Content based model and memory based model, User-item interaction matrix is more precise solution for recommender systems.

Combining old approach of similarity detection for filling up missing values could be helpful to give more than 90 percent accuracy.

Combining correlated groups as multiclass co-clustering would reduce data sparsity and produce diversified recommendations.

9. Future Scope

In future combining all the positive features of Traditional recommender system to design Customized algorithm for recommender system along with code optimization will be the main challenge. As we saw existing system has 6 variants hence for avoiding cold-start and producing diversified recommendations, they have some beneficial features so we can combine it all together to design powerful recommendation engine.

10. References

1. Babu MSP, Kumar BRS. An implementation of user-based collaborative filtering algorithm. International Journal of Computer Science and Information Technologies. 2011; 2:1283–6.
2. Sarwar B, Karypis G, Konstan J, Riedl J. Item based collaborative filtering recommendation algorithms. Group Lens Research Group/Army HPC Research Center; 2001. p. 285–95. Crossref.
3. Xu B, Bu JJ, Chen C. Improving collaborative recommender systems via user-item subgroups. IEEE Transactions on Knowledge and Data Engineering. 2016; 28:2363–74. Crossref.
4. Kale P, Patil MR. Survey on parallel hybrid multigroupco-clustering using collaborative filtering model. International Journal of Advanced Research in Computer and Communication Engineering. 2015; 4(12):536–8.
5. Ning X, Karypis G. SLIM: Sparse Linear Methods for Top-N recommender systems. Proceedings of the 2011 IEEE 11th

International Conference on Data Mining; 2011. p. 497–506. Crossref. PMid:21364493

6. Wu Y, Liu X, Xie M, Ester M, Yang Q. CCCF: Improving collaborative filtering via scalable user-item co-clustering. WSDM 16 San Francisco CA USA Ninth ACM International Conference onWeb Search and Data Mining; 2016. p. 73–82.
7. Pozo M, Chiky R, Meziane F, Metais E. An item/user representation for Recommender systems based on Bloom filters. Research Challenges in Information Science (RCIS) IEEE Tenth International Conference. 2016; 8:978–90.
8. Hackel R, Vlachos M. Scalable and interpretable product recommendations via overlapping co-clustering; 2016.
9. Yang Z, Wu B, Zheng K, Wang X, Lei L. A survey of collaborative filtering based recommender systems for mobile internet applications. IEEE Translations and Content Mining. 2016; 21:389–97.
10. Rao MVVRMK. A Collaborative Filtering Recommender System with Randomized Learning Rate and Regularized Parameter. Conference: 2016 IEEE International Conference on Current Trends in Advanced; 2016. p. 1–5.
11. Xiaoyao Z, long LY, Liping S, Fulong C. A new recommender system using context clustering based on matrix factorization techniques. Chinese Journal of Electronics. 2016; 25(2):334–40. Crossref.
12. Huang SS, Ma J, Wang S. A hybrid multi group co clustering recommendation framework based on information fusion. ACM Transactions on Intelligent Systems and Technology; 2015. p. 1–22. Crossref.