ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

Survey: Collaborative Recommender Systems Using Multiclass Co-Clustering

Mugdha Adivarekar, Vina Lomte

M.E Student, Dept. of Computer Engg., RMD Sinhgad School of Engineering, S. P.Pune University, Pune, India

Asst. Professor &HOD, Dept. of Computer Engg., RMD Sinhgad School of Engineering, S. P. Pune University, Pune, India

ABSTRACT: Recommender Systems are playing very crucial and vital role in day today's life. 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 user when he log in to system based on his previous shopping or surfing or interests. Hence collaborative filtering is getting more popular. 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. Hence, multiclass co-clustering technique proposes to generate top-N recommendations by maintaining user-item interactionmatrix and making clusters by matrix factorization.

KEYWORDS: Collaborative Filtering, Information Retrieval, Multiclass co-clustering, Recommender Systems, User Profiles.

I. INTRODUCTION

This category of research falls under the domain data mining and information retrieval. I like Observing peoples style and their interest and Community forming because of same interest. Also I am habitual of visiting shopping sites and getting surprised about advertises of visited items and keen to learn information retrieval and information filtering process behind them. Here one need to have some important Concepts very clear -

What is Information retrieval?

Information retrieval is the act of getting information resources relevant to an information need from a collection of information resources. Searches can be based on full-text or other content-based indexing.

What is Information filtering?

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

What is 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 predict the "rating" or "preference" that a user would give to an item.

What is User profile?

A user profile is a visual display of personal data associated with a specific user, or a customized environment by user. A profile refers to the explicit digital representation of a person's identity.

Collaborative Filtering:

Collaborative filtering is method of information filtering and it associate user with like-minded people group.It is used to create recommendation system. ex. Amazon shopping site, music site

These systems are inspired by users surfing trend like some users prefers Electronics, clothing, jewellery, kitchenware, etc.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

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 neighbours with high similarities for each user and make predictions based on the weighted average ratings of his or her neighbours. The neighbours 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 neighbours for each item. The similarity computation process is computationally expensive for large data sets and neighbours 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.
Collaborative filtering methods have been applied to different kinds of data including:
Sensing and monitoring data - mineral exploration, environmental sensing over large areas or multiple sensors;
Financial data - financial service institutions that integrate many financial sources;
Electronic commerce and web applications where the focusis on user data, etc. But these applications follow traditional CF which can produce predictive result also.
So objectives ofthis research are-
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. This approach can be considered as an extension of traditional clustering CF models.

II. RELATED WORK

Maddali Surendra [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 overcomes some drawbacks of User-User based recommendations Badrul Sarwar Et.al[2] Item based Collaborative Filtering Recommendation Algorithms 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. Bin Xu Et.al[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-item matrix. Pramod Kale[4] has conducted 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. Yao Wu Et.al [6]addressed new and improvised approach for Collaborative Filtering via Scalable User-Item Co-clustering. This research supports my base paper. 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.
Manuel Pozo Et.al [7] worked on Item/User Representation for Recommender systems based on Bloom filters, with which user-item relation is represented in a bloom filter vector an dauthors 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 invery low space so this technique is used in case of low data structures. Reinhard Hackel, Michail Vlachos[8] worked on Generation of Interpretable Recommendations via overlapping co-clusters and proposed an algorithm which uses matrix factorization technique to identify co-clusters and recommends client-product pair because of its

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

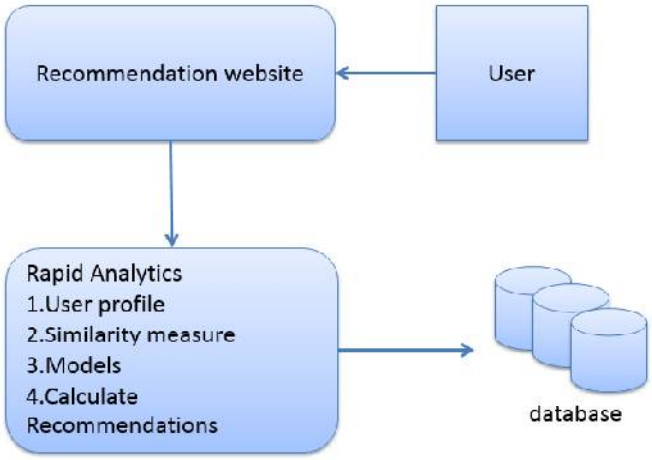
Vol. 5, Issue 1, January 2017

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 New Recommender System Using Context Clustering Basedon Matrix Factorization Techniques by Zheng Xiaoyao Etal.[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. Shanshan Huang[12] proposed an advanced HMCoC 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. In research of Recommendation system there are diversified enhancement that came into picture because of need of the time and growth ofE-commerce. Another one approach defined by Xia Ning[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. A Survey of Collaborative Filtering Based Recommender Systems for Mobile Internet Applications [9] by Zhe Yang was useful to implement recommendation systems with Mobile applications by altogether comparing various collaborative filtering methods, Categorizing into pros and cons and they areclassified into memory based and content based approaches. Murali Krishna Rao [10] worked on a 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 which2 new matrices are generated which will include predicte dratings with original ratings. Data loss equation is used to avoid data loss while updating matrix through machine learning algorithm.

III. PROPOSED ALGORITHM

A. *DESIGN CONSIDERATIONS:*

After careful analysis the system has been identified to have the following modules as shown in figure below:
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.



Pearson's correlation formula is used for finding missing data. This is useful when data centre is failed to update user-item matrix and some values are missing in it.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

We have this formula below-

$$sim(a,b) = \frac{\sum_{p \in P}(r_{a,p} - \bar{r}_a)(r_{b,p} - \bar{r}_b)}{\sqrt{\sum_{p \in P}(r_{a,p} - \bar{r}_a)^2} \sqrt{\sum_{p \in P}(r_{b,p} - \bar{r}_b)^2}}$$

Where
a,b : users
$r_{a,p}$:rating of user a for item p
P: set of items,rated both bya 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 $\epsilon R^{(m+n)\times r}$
Subgroups of this Rating matrix can be denoted as V_k and forgroup V_k sub matrix of rating matrix is defined as
$R_k \epsilon R^{(mk \times nk)}$
Wherek = 1,…., L.
m_k and n_k are the number of users anditems in that group, respectively

B. *Description of the Proposed Algorithm:*

Aim of the proposed algorithm is to maximize the accuracy of recommendations by operating available data in in systematic manner. The proposed algorithm is consists of three main steps.

Step1: Dimensionality Reduction of User-Item Matrix

Step2: Finding subgroups i.e. Partition matrix

Step3: Get Top-N recommendations using traditional CF algorithm.

IV. PSEUDO CODE

Pseudo code for top-N recommendation is available in [12]:

Input: Rating matrix $R \in R^{(m \times N)}$, all the groups {*V*1, . . . , *VL*}, a chosen CF method, and the number of items in recommendation list *N*.

Output: Recommendation list for each user.

Step 1: for k←1 to L do
Extract sub matrix$R_k$from rating matrix R with users and items belonging to group V_k;
Apply the CF recommendation method with $R_k$as input and predict missing scores r(u_i,i_j, k).
end
Step 2: for i←1 to mdo
Step 3:for j ←1 to n do
if R_{ij} is missing then
Find group index k = {maxkY_{ik}|Y_{ik} = 0andY_{jk} = 0};
if k is null then
Set R_{ij} = 0;
end
else
Set R_{ij} = r (u_i , i_j , k);
end
end
end
Step4:Generate top-n recommendation list for user uiaccording to the decreasing order of thepredicted scores.
Step5:End

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

V. ADVANTAGES AND APPLICATIONS

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.

Applications –
1. Creating Web log pattern applications
2. Recommender systems for shopping sites like Amazon, flip kartetc.
3. User preference arranger for financial applications.
4. Monitoring and sensing the data to provide
5. Providing advertises of product over web as per user's interests.
6. Travel/tourism recommendations

VI. CONCLUSION AND FUTURE WORK

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.

VII. ACKNOWLEDGMENT

I take this opportunity to express my heartfelt gratitude to my guide and head of department, Prof.Vina M Lomte, Department of Computer Engineering, RMDSSOE, Savitribai Phule Pune University, for her kind cooperation, constant Encouragement and suggestions and capable guidance during the research, without which it would have been difficult to proceed with.

REFERENCES

1. MaddaliSurendra Prasad Babu,Boddu Raja Sarath Kumar, "An Implementation Of User-based Collaborative Filtering Algorithm", IJCSIT,Vol.2,pp.1283-1286,2011
2. BadrulSarwar,GeorgeKarypis,JosephKonstan,Johnriedl,"Item based Collaborative Filtering Recommendation Algorithms",GroupLens Research Group/Army HPC Research Center , Vol.5,pp.285-300,2011
3. Bin Xu,JiajunBu,Chun Chen,"Improving Collaborative Recommender Systems via User-Item subgroups",IEEE Transactions on Knowledge and Data Engineering,Vol.28,pp.2363-2374,2016
4. PramodKale,Prof.M.R.Patil,"Survey on Parallel Hybrid Multigroup co-clustering using Collaborative Filtering Model",IJARCCE, vol. 4, no.12, pp-536-538, 2015
5. Xia Ning,GeorgeKarypis,"SLIM: Sparse Linear Methods For Top-N Recommeder Systems",Proceedings of the 2011 IEEE 11th International Conference on Data Mining, pp.497-506, December 11-14, 2011.
6. Yao Wu,XudongLiu,MinXie,MartinEster,Qing Yang,"CCCF:Improving Collaborative Filtering via Scalable User-Item Co-clustering", WSDM 16,San Francisco,CA,USA,Ninth ACM International Conference on Web Search and Data Mining,pp.73-82,2016.
7. Manuel Pozo,RajaChiky,FaridMeziane,ElisabethMetais,"An Item/User Representation for Recommender systems based on Bloom filters," Research Challenges in Information Science (RCIS), 2016 IEEE Tenth International Conference., vol. 8, pp.978-990, 2016.
8. ReinhardHackel,Michail Vlachos,"Interpretable Recommendations viaoverlapping co-clusters," IDR Information Retrieval, vol. 1, no. 7,.2016.
9. ZheYang,BingWu,KanZheng,XianbinWang,Lei Lei,"A Survey OfCollaborative Filtering Based Recommender Systems for Mobile InternetApplications,"IEEE Translations and content mining,vol.21,pp.389-397,2016.
10. M V V R Murali Krishna Rao,"A Collaborative Filtering RecommenderSystem with Randomized Learning Rate and Regularized Parameter,"Astrosat Payload Operation Centre,IUCAA,Pune,2016.
11. Zheng Xiaoyao,LuoYonglong,SunLiping,ChenFulong,"A NewRecommender System Using Context Clustering Based on MatrixFactorization Techniques," Chinese Journal of Electronics,vol. 25, no.2,2016.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

12. ShanshanHuang,JunMa,Shuaiqiang Wang,"A Hybrid MultigroupCoClustering Recommendation Framework Based On InformationFusion," ACM Transactions on Intelligent Systems and Technology,vol.6,no.2,Article 27,2016

BIOGRAPHY

Mugdha Adivarekar is a Student in the Computer Engineering Department, RMD Sinhgad School of Engineering, Warje, Pune University. She is pursuing Master of Computer engineering degree in. Her research interests are Information Retrieval, Data mining, Algorithms, etc.