International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

A Survey of Content Aware Video based Social Recommendation System

Pankaj Chaudhary[1], A. A. Deshmukh[2]

[1]P. G. Student, Dept. of Computer Engineering, Smt. Kashibai Navale College of Engineering , Savitribai Phule University, Pune, India

[2]Professor, Department of Computer Engineering, Smt. Kashibai Navale College of Engineering, Savitribai Phule University, Pune

Abstract: ***Collaborative Filtering (CF) has achieved widespread success in recommender systems, which automatically aggregate and predict preferred products of a user using known preferences of other users from large scale SRSs. But on the other hand, a large portion of them cannot manage the cold-start issue that indicates a circumstance that social media sites neglect to draw suggestion for new things, users or both; hence accurate and scalable recommendations are difficult to generate. This supposition is against the way that low-level ratings help little to recommending things that are liable to be of enthusiasm of users. To this end, we propose a system using bi-clustering and fusion, a recently modeled scheme for the cold-start issue focused around the system procedures under a distributed computing setting. To lessen the dimensionality of the rating matrix, the system influences the bi-clustering method. To defeat the information exiguity and rating differences, it utilizes the smoothing and fusion strategy. At long last, the system proposes content aware video based social media substance from both thing and user bunches. Finally our experiments result will be shown that our method generates better recommendations.***

Keywords: Social Recommendation System (SRSs), cold-start problem, social media sites, ratings, bi-clustering and fusion, rating matrix.

1. Introduction

Recommender systems have been extensively used within e-commerce and communities where items like movies, music and articles are recommended. A key challenge for building an effective recommended system is the well-known cold-start problem—how to provide recommendations to new user, items or both? [1].

The cold-start problem is the fundamental issue that should be addressed in the SRSs and most of the CF alterations fail with that. The cold-start problem can be stated as the situation faced by the social media site while recommending the items for a new item, a new user or in cases, both of them, because the system knows very little about those item and user in term of their preferences. This problem occurs due to the fact that very few ratings are available for the new elements in the matrices [2].

To bridge the gaps between the indications of existing and the newly arrived users or items are spontaneous, incorporating the external information like, content information might have helped the CF. various aspects of the items and users can be involved in the newly added information. In [3], for predictions, Zhou has integrated the user's initial review and the matrix factorization. In [4], generative probabilistic model was proposed by Schein. This model was augmented with the actors and directors, the raw movie ratings.

Thus, in this paper, we have proposed a novel scheme, which enable the social media site to improve the influence of the cold start problem in recommendation system. A novel scheme to solving the cold start problem is bi-clustering technique with smoothing and fusion strategy.

This system could be further enhanced in the near future. The remaining paper can be summarized as follows: The section II contains briefly review existing studies for the similar field or similar topics, briefly. Section III introduces the Similarity Analysis. Section IV introduces the proposed system .Section V concludes our work with the future directions.

2. Literature Review

(a) Bi-Clustering

Bi-Clustering technique (also referred to as two mode clustering) simultaneously clusters both item and user in item-user matrix. the Bi-clustering technique performs better than one-way cluster technique to deal with sparse and high dimensional recommendation matrices [5][6].

Survey of Bi-Clustering Analysis:

Approach	*Algorithms/Type*
Distance-based biclustering [5][6]:- Distance-based biclustering typically uses some distance-based variance metric to measure the quality of the bi- clusters, and performs an iterative search for the bi- clusters by minimizing the Residual sum of squares cost.	Direct Clustering/Constant $SSQ_k = \sum_{i \in X_k, j \in Y_k} (a_{ij} - a_{X_k Y_k})^2$ Where $a_{X_k Y_k}$ = average value in the bicluster. **;** FLOC (Flexible overlapped biclustering) /Coherent Values $H(X,Y) = \frac{1}{\lvert X \rvert \lvert Y \rvert} \sum_{i \in X, j \in Y} (a_{ij} - a_{iY} - a_{Xj} + a_{XY})^2$ Where a_{iY} , a_{Xj} , a_{XY} = row mean, column mean, and the mean in the sub matrix. **;**

	δ-bicluster/Coherent Values; δ-pCluster/Coherent Values;
Factorization-based biclustering [5][6]:- Factorization-based biclustering algorithm uses spectral decomposition technique to Uncover "natural" substructures that are related to the main patterns of the data matrix	Spectral(Singular Value Decomposition(SVD)) /Coherent Values;
Probabilistic based biclustering [5][6]:- The biclustering method in this category typically assumes a probabilistic model of biclusters and applies statistical parameter estimation techniques to search for the biclusters.	Plaid Models /Coherent Values $a_{ij} = \mu_0 + \sum_{k=1}^{k} \theta_{ijk} \rho_{ik} K_{jk} + \varepsilon_{ij}$ Gibbs Sampling/Constant Columns; Probabilistic Relational Models (PRMs)/Coherent Values
Biclustering for coherent evolution [5][6]:- In this , the elements of a data matrix are usually of numeric values but they can alsobe transformed into symbols that reflect trends in the data. The symbols can be purely nominal or encode positive and negative changes relative to a normal value.	Order-Preserving Submatrix (OPSMs) /Coherent Evolution; OP(order preserving)-Clusters /Coherent Evolution SAMBA(Statistical-Algorithmic Method for Bicluster Analysis) /Coherent Evolution;
Geometric-based biclustering [5][6]:- In this bicluster, the geometric viewpoint makes and it provides a unified Mathematical formulation for simultaneous detection of different types of linear biclusters (constant, additive, multiplicative, and mixed additive and multiplicative)	

(b) Collaborative Filtering

Collaborative Filtering (CF) is a process that automatically aggregates and predicts user preferences from large-scale item-user matrices. It is based on the idea that people who agreed in their evaluation of certain items in the past are likely to agree again in the future [7].

CF Categories

Approach	*Explanation*
Memory-based CF [7] [8]	This mechanism uses user rating data to compute similarity between users or items. They predict user preferences by identifying similar item or like-minded users over the entire item-user matrix.
Model-based CF[7][8]	These are used to make predictions for real data. They cluster all items or users into classes.Then they use these classes to predict unrated data for users.

(c) Collaborative Filtering with the Bi-Clustering Technique:

A considerable lot of the CF plots principally utilize one-dimension clustering to cluster items or users exclusively. In any case, the one-dimension clustering system typically disregards the valuable data in the inverse dimension. To this end, Bi-clustering method at the same time clusters both item dimension and user dimension in the item-user matrix. Correspondingly, the Bi-clustering procedure performs a ton better than one-way cluster strategy to manage scanty and high dimensional recommendation matrices [8], [9].

[10] The framework permits clients to impart and search videos, connect to companions, remark and semantically comment videos at frame level, share and browse related videos and clients through recommendation, clustering and semantic similarity, and to make personal profiles of hobbies in a semi-automatic manner. The framework consists of two primary parts that are nearly interconnected: it gives a recommendation engine of videos and comparable clients, perceptible in the personal home page of the social network, and it additionally offers the programmed creation of an open profile of diversions for which the framework adventures clustering and semantic separations to make recommendations and suggestions of assets (videos, persons, Wikipedia articles) that match those investments.

Lee [11] has proposed a personalized DTV program recommendation system. This system refined the channels, selecting different processes, and for satisfying the requirements of the customer. Lai [12] has proposed a system called CPRS (A cloud-based program recommendation system). This system was used to recommend the programs to the customers of digital TV platforms. A blog personality recommender system was proposed by Jiang [13]. This system recommended various high quality and personalized blogs for the users. This method was based on the cloud computing environment.

Table1: Comparison of CF and Bi-Clustering

Collaborative Filtering	*Bi-Clustering*
One-mode clustering	Two-mode clustering
Cold start problem	No Cold start problem

The most of CF schemes mainly employ one-mode clustering to cluster items or users individually. However, the one-dimension clustering technique usually ignores the useful information in the opposite dimension. Bi-clustering technique is two-mode clustering to cluster simultaneously both item and user in the item-user matrix [5].

A Collaborative Filtering has to rely on the past experiences of the items, users or both, it will have problem when assessing new item or user that have not yet been seen. This problem of unseen or almost unseen users and items is generally referred as the cold-start problem. The bi-clustering technique deal with this problem [7].

3. Similarity Analysis

Similarity

The similarity of items can be measured by Pearson Correlation Coefficient (PCC). PCC often achieves better performance [14].

$$Sim(i,j) = \frac{\sum_u (r_{u,i} - \overline{r_i}) * (r_{u,j} - \overline{r_j})}{\sqrt{\sum_u (r_{u,i} - \overline{r_i})^2} * \sqrt{\sum_u (r_{u,j} - \overline{r_j})^2}}$$

Where

$\overline{r_i}$ and $\overline{r_j}$ = value of the average rating calculated from item i and j respectively.
$r_{u,i}$ = rating given by user u to item i.
$r_{u,j}$ = rating given by user u to item j.
User similarity is often calculated by Pearson Correlation Coefficient (PCC).

Evaluation Metrics

Two general methods for evaluating the performance. One is Mean Absolute Error (MAE) and the other metric is Root Mean Squared Error (RMSE)[14].

$$MAE = \sum_{u \in T} |P_{u,i} - r_{u,i}| / |T|$$

$$RMSE = \sqrt{\sum_{u \in T} (P_{u,i} - r_{u,i})^2 / |T|}$$

Where
T = test set.
|T| = size of the test set.
$P_{u,i}$ = prediction.
$r_{u,i}$ = actual value.

4. Proposed System

[15] Proposed a novel methodology for content-based music recommendation. The primary innovation of the proposed procedure consists of a similarity function that, as opposed to considering whole songs or their thumbnail representations, breaks down audio similitude between semantic sections from diverse audio tracks

To this end, we proposed a content aware video based social recommendation system. The recommendation system will be fed the browsing activity of users in a community/group and will build an interest profile for each participating user. It will build an interest profile by storing the metadata (title, rating, description, comments, view count etc.) of the audio and video that a user listen or watch. The system will use this interest profile to find other users/audio , videos with matching interests/content and use that knowledge to recommend it.

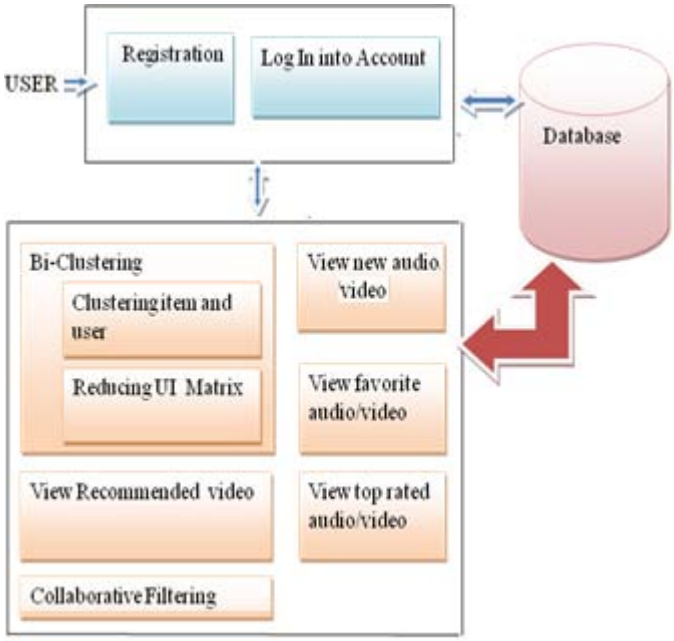

Figure 1: architecture of Recommendation System

5. Conclusion

The cold start recommendation problem is the most basic and important problem that the social networks faces. Cold start recommendation problem is that, when a user joins a social network, the system needs to recommend him with something, like, videos, audios, other users, etc. But all these facts are useless if the system knows nothing about the user, as it is easier to recommend things to user if the user history is available. Thus, in this paper, we have proposed a novel scheme, which enable the social media site to improve the influence of the cold start problem in recommendation system. Firstly, we have used the trivial ratings, which are not useful for current social recommendations. An identifying mechanism is used further. It influences the Bi-clustering technique for items and users as well. This method will be useful in recommending the things for the new users, or those users, having no history in the network. At long last, the system proposes content aware video based social media substance from both thing and user bunches. Though we have provided a new solution for the cold-start recommendation problem, we do not state that, this is the complete solution for this problem; as, there is always a room for improvement

References

[1] P. Victor, C. Cornelis, A. Teredesai, and M. De Cock, “Whom should I trust?: The impact of key figures on cold start recommendations”, in Proc. ACM Symp. Appl. Comput., 2008.
[2] D. Zhang, Q. Zou, and H. Xiong, “CRUC: Cold-start recommendations using collaborative filtering in internet of things”,Energy Proc.,2011.
[3] K. Zhou, S.-H. Yang, H. Zha, “Functional matrix factorizations for cold-start recommendation”, in *Proc. 34th Int. ACM SIGIR*, 2011.
[4] A. Schein, A. Popescul, L. Ungar, and D. Pennock, “Generative models for cold-start recommendations”, in Proc. SIGIR Workshop, 2001.

[5] Hongya Zhao and Hong Yan et. al. "Biclustering Analysis for Pattern Discovery: CurrentTechniques, Comparative Studies and Applications".
[6] Sara C. Madeira and Arlindo L. Oliveira "Biclustering Algorithms for Biological Data Analysis: A Survey" JAN 2004
[7] D. Zhang, J. Cao, M. Guo, J. Zhou, and R. Vaskar, "An efficient collaborative filtering approach using smoothing and fusing", in Proc. ICPP, 2009.
[8] K. W.-T. Leung and W.-C. Lee, "CLR: A collaborative location recommendation framework based on co-clustering", in Proc.34th Int. ACM SIGIR Conf., 2011.
[9] T. George and S. Merugu, "A scalable collaborative filtering framework based on co-clustering", in Proc. 5th IEEE ICDM, Nov. 2005.
[10] M. Bertini, A. Del Bimbo, A. Ferracani, F. Gelli, "A novel framework for collaborative video recommendation, interest discovery and friendship suggestion based on semantic profiling", In Proceedings of the 21st ACM international conference on Multimedia, 2013.
[11] S. Lee, D. Lee, and S. Lee, "Personalized DTV program recommendation system under a cloud computing environment", IEEE Trans. Consumer Electron., 2010.
[12] C.-F. Lai, J.-H. Chang, C.-C. Hu, Y.-M. Huang, "CPRS: A cloud-based program recommendation system for digital TV platforms", Future Generat. Comput. Syst., 2011.
[13] J. Jiang, W. Pang, Y. Deng, K. He, and Z. Gu, "A blog personality recommender system based on cloud computing infrastructure," in Proc.IJCSS, 2012.
[14] Zhang et. al., "Cold Start Recommendation using Bi-Clustering and Fusion for large Scale Social Recommender Systems"IEEE Transactions on Consumer Electronics, June 2014.
[15] A. Bozzon, G. Prandi,"A Music Recommendation Sys. Based on Semantic Audio Segments Similarity",IASTED Int. Conf. on Internet ,Multimedia Sys. and App. (EuroIMSA 2008).