A NOVEL RECOMMENDER ALGORITHM FOR MOBILE ALERT SERVICES

Asoka Korale - akorale@sltnet.lk

ABSTRACT

Recommender Systems are part of our daily experience as the number of choices facing consumers increase rapidly to levels most individuals find difficult to evaluate without recourse to an automated system of reference. In the past when consumer choices were limited, one had the ability to research all available options or make a decision utilizing suggestions from ones social network. In this paper we develop a novel recommender algorithm for a Mobile Telecommunications service application in which customers subscribe to certain types of Short Message Service (SMS) alerts depending on their interests.

A Mobile Network Operator (MNO) has access to considerable demographic and user preference information in addition to the Mobile Telecommunications Service usage and calling patterns of its subscribers. We develop an algorithm where we use this information to supplement our understanding of the subscriber and build a profile that can be used to more effectively compare one subscriber with another. Thus both behavioral and demographic characteristics of the subscribers are utilized in predicting preferences. The additional attributes so introduced are used to group the users using cluster analysis. Clustering reduces the overall number of variables needed to describe the grouped users as each group is represented by a Cluster ID, allowing the introduction of a great many profiling attributes without increasing the dimensionality of the enhanced ratings space aiding in the more accurate and efficient similarity calculation

The algorithm also uses item-item similarity in addition to user-user similarity in predicting customer preferences. The item-item similarity computation is enhanced by the use of a classification of the items in to product categories with the aim of improving the accuracy of the prediction. This categorization of items or products is based on a broad classification of customer preferences or interests that are mapped to the type of product. Thus we are able to incorporate a customer's "interest factor" in to the products purchased by a mobile subscriber further enriching the ratings space. The proposed algorithm employs a hybrid scheme linearly combining the results of user-user and item-item predictors that are each determined independently.

The algorithm is validated on past ratings by a process of back testing that predicts ratings for product categories already rated by using part of the historical ratings record of a portion of the subscriber base. We demonstrate that even highly sparse ratings spaces can be effectively modeled and user preferences predicted when sufficient ancillary information is used together with dimension reducing transformation techniques.

1. INTRODUCTION

Recommender Systems play a crucial role in informing choices to consumers particularly in the case of new products that the customer has no prior experience of. As the number and variety of choices available to a customer increases exponentially to levels never before experienced by consumers the need for an automated system of reference becomes essential. When the available choices are too numerous and varied it may not even be possible to rely on ones acquaintances or social network to make recommendations to satisfy the very individual preferences of a consumer. This factor coupled with the ability to customize a product or service to meet ones needs only adds more complexity to the available choices confronting a consumer.

In this context the online retailer or in this instance the Mobile Network Operator has the complex task of matching available product choices with their

individual customers. Hence businesses have a critical need to deploy systems that utilize predictive algorithms to accomplish this task. These systems also increasingly determine the success or failure of many online market places as customers are not likely to rely on or return to a service that does not provide suggestions that meet their desires and expectations.

2. FORMULATION OF RECOMMENDER LGORITHM

Mobile Telecommunications Operators are increasingly marketing their products electronically, making recommender algorithms play an important role in growing sales and awareness of the product portfolio among the subscriber base. Recommender systems though based on machine learning and statistical techniques essentially endeavor to determine similarities between customers for whom the recommendations are being made and also between the different product categories that are being recommended. The recommendation is then made utilizing the premise that "similar customers" will have preference for "similar items" or product choices while a product similar in some way to the products in the portfolio consumed by a particular customer is more likely be preferred by the consumer in question.

As a categorization of items is used in this model the correlation between items (categories of items) may not be prevalent to the extent that it would in the case of other data sets. In fact the item-wise categorization utilized will tend to make item-item similarity have lower correlation as the particular categorizations utilized would tend to make the categorized items more independent of each other.

We present a recommender algorithm that was developed to operate in an environment that has a great degree of sparsity in the ratings that under normal circumstances makes similarity computations difficult and as a consequence the recommendation unreliable. This high sparsity is the result of modeling customer preferences for certain Value Added Service (VAS) products that have a very non-uniform popularity and therefore low level of incidence among the subscriber base. The level of sparsity is further increased as relatively few customers simultaneously subscribe to more than one VAS product at a time. Thus only 3-4 ratings may be available for each subscriber. This level of the "lack of knowledge" about a particular customer's preferences would ordinarily make the suggestion of potential new items (product choices) quite difficult. Traditional recommender systems employed in movie databases and other online market places usually have several dozen choices or histories for which the customer has rated products and expressed an opinion, and so considerable information is available to build a profile of customer preferences.

In light of this we utilize ancillary data drawn from the demographic information about the subscriber that is stored within the Mobile Network Operator's customer information data bases and characteristics of the mobile usage pattern of the customer. Both types of information which are not directly related to the rated products are used in order to build a profile of the customer and use that information in identifying groups of similar users which are in this case taken to mean users who have similar interests or preferences.

The paper shows that the similarity computation and thereby the recommender performance is greatly improved by factoring in mobile subscriber profiling information. This ancillary customer data helps mitigate the lack of correlation that is otherwise an issue due to the sparsity in the ratings matrix. Thus while the ratings matrix alone cannot provide an accurate assessment of user similarity the introduction of ancillary data greatly improves the ability to correlate between subscribers and increase overall predictive accuracy. Thus we enhance the ratings matrix by combining it with the ancillary data resulting in a new more comprehensive ratings space.

Singular Value Decomposition (SVD) is employed in this implementation to transform the ratings space to facilitate a more accurate determination of those neighbors closest to a particular subscriber by considering their relative positions in the transformed space. This transformation is then restricted to operate at a lower dimension that is able to more closely represent the relationships between subscribers. This lower dimensional

representation makes the similarity computation also more efficient but is also able to retain most of the important information that was contained in the original data (ratings space). By carrying out the similarity computation in the transformed space with reduced dimensions one can also eliminate redundant information.

In the proposed model, K–Means clustering is used to group mobile use subscriber data in to non overlapping clusters in an attempt to reduce the number of variables or attributes that would need to be decomposed via the SVD and accounted for in the similarity computation stage. The clustering also succeeds in reducing the range of the attributes that need modeling as it's a cluster index number that is used in the subsequent analysis instead of a group of interval variables.

Finally the ratings are predicted employing a hybrid collaborative filtering method that accounts for both user and item based similarity predictors.

3. RECOMENDER APPROACHES [1]

Recommender Systems which are also broadly known as Collaborative Filtering algorithms feature a wide variety of approaches to accomplish the task of recommending the next best product choices for a user.

Among them co-relational approaches are in widespread use [1]. In this approach users similar to the one for which ratings are to be predicted are identified by correlating that user's ratings vector with the ratings vector of its nearest neighbors. A typical recommender using the correlational approach to predict a rating r_{pred} takes the form given by equation (1).

$$r_{pred} = \frac{\sum_i w_i r_i}{\sum_j |w_j|} \qquad (1)$$

Where w_i are the correlational weights associated between the user of interest and the users in his neighborhood and r_i are the corresponding ratings of the users in the same neighborhood. This user-user correlational approach is also used in item-item predictors where correlations between other items similar to the items rated by the user in question are utilized to suggest a rating for an unknown item for that user. In fact both user and item based approaches are usually combined to obtain hybrid recommender system algorithms that combine the decision from the two approaches to arrive at the final rating.

In Content based approaches the goal is to identify common characteristics of items that have been rated highly by the user in question and then determine new items that share the same characteristics for recommendation to that user. This approach is not particularly suitable to the problem at hand as the categories (of items) do not share many common traits. In fact the categories have been defined with a view to demarcating the items in to groups with meanings that are clearly separable.

Further, graph theoretic approaches have also been developed for the recommender problem. In this approach the user-item matrix is represented as a bipartite graph where user nodes are mapped to item nodes. The main advantage of this approach is that it allows transitivity of influence by allowing nodes not directly connected to each to influence each other as through indirect paths through the network. In one application the items to be recommended to a user are those that are closest to it in the graph. In a another application users similar to each other may be found by their closeness in the graph domain and these similar users are then employed in one of the neighborhood based techniques discussed earlier. The similarity or closeness in the graph domain is usually measured in terms of the number of paths that separate the nodes in question and several distance measures are defined for this purpose.

4. SINGULAR VALUE DECOMPOSITION [3]

Singular value decomposition enables the decomposition of a user (subscriber) vs. item ratings matrix in to a product of three matrices, corresponding to a user vs. concept matrix (U), a diagonal matrix with "singular values" that indicate the strength of each concept ($\sum$) and a concept vs. item matrix (T).

This decomposition allows one to extract the hidden user vs. concept relationships which are useful as the correlations between users in this transformed space are indicative of the underlying relationships between subscribers that are not normally visible when directly correlating subscribers using their attribute values (item rating values).

Consider the Diagonalization of a matrix A

$$T_D^{-1} A T_D = \sum \qquad (2)$$

Where the columns of T_D are the Ortho-Normal Eigen vectors of A and

$$\sum_D = \begin{pmatrix} \lambda_1 & & \\ & & \\ & & \lambda_n \end{pmatrix} \qquad (3)$$

where $\sum_D$ is the diagonal matrix with corresponding Eigen values, which are ordered in descending order of magnitude . It follows then that

$$T^{-1} A^T A T = \sum^2 \qquad (4)$$

Where T contains the Eigen vectors of A^TA. Pre and Post multiplying by the inverse then yields

$$\sum^{-1} T^{-1} A^T A T \sum^{-1} = I$$

and setting

$$U = AT \sum^{-1} \qquad (5)$$

by pre and post multiplying appropriately gives

$$U \sum T^{-1} = A$$

As T is composed of Ortho-Normal Eigen vectors

$$T^T = T^{-1}$$

leading to the decomposition

$$U \sum T^T = A$$

Thus it is possible to obtain a lower dimensional approximation to the original matrix A by considering the largest "k" Eigen values corresponding to selecting the first "k" columns of the matrices. This result is used in the transformation of coordinates representing the user attributes for whom ratings have to be predicted and where the transformed coordinates are correlated with the transformed coordinates of the other subscribers for whom ratings are available and those closest in terms of the cosine similarity chosen as neighbors for use in the user-user collaborative filtering.

Thus the Singular Value Decomposition in addition to allowing the lower dimensional representation of the original data allows one to extract hidden correlations and find groups of similar users.

4.1. **K-Means Clustering [2]**

Clustering allows a reduction in dimensionality by grouping subscribers based on the values of their profiling attributes. The object of clustering is to group objects or in this case subscribers in such a way that members of a cluster are more similar to each other than they are similar to members of other clusters. The similarity in this context is taken to be a "distance criteria" from the cluster centroids. This process also lends to finding commonalities between subscribers, as in the similarity computation of a recommender system when viewed in the context of the distance measure that is employed.

The pattern matrix X represents subscribers as rows and their corresponding attribute values in columns.

$$X = \begin{pmatrix} x_{1,1} & x_{1,2} & & x_{1,n} \\ & & & \\ & & & \\ x_{N,1} & x_{N,2} & & x_{N,n} \end{pmatrix}$$

The K-Means algorithm takes as input the number of clusters and allocates subscribers to each cluster to which they are closest measured in terms of a distance measure. This "distance" can take many

forms including "Euclidian", "Citiblock" and "Cosine" among others. The centers are then recalculated as the average of the positions of the members of each cluster in "n" dimensional space and the entire process repeated until no subscriber is seen to move between clusters.

In this application of clustering, the user attributes corresponding to a subscribers VAS usage, Mobile Network related data and some demographic information are clustered. By assigning each subscriber to a cluster based on distance-similarity measure we are able to both rationalize the number of variables needed to describe each subscriber and also avoid being directly impacted by the range of each interval attribute.

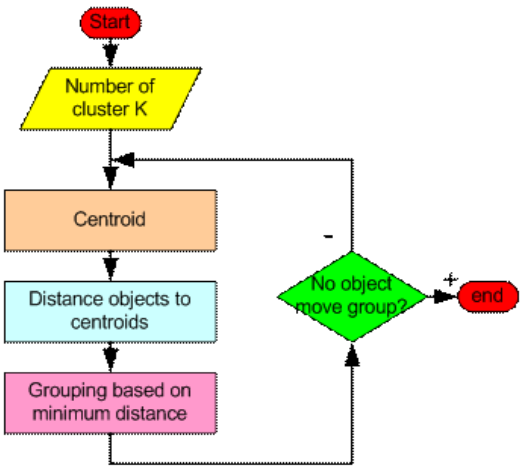

Figure 1: K-Means Clustering Algorithm [2]

5. THE DATA MODEL

The SMS alerts for which recommendations are to be found fall in to fourteen broad categories which are given the following subjective titles.

1. 'Business', 2. 'Mobile Apps', 3. 'Education', 4. 'Entertainment', 5. 'Fun & Jokes', 6. 'Health', 7. 'Information', 8. 'News', 9. 'Other', 10. 'Social', 11. 'Sports', 12. 'Utilities', 13. 'sdp', 14. 'soltura'

Thus each subscriber can have one or more of the fourteen product categories in their subscription portfolio. Additionally since we are dealing with product categories a subscriber may have more than one occurrence of a particular product category (if for example a subscriber has subscribed to two different news alerts). These events are accounted for in the ratings matrix by entering the number of incidences of each product category in the corresponding entry.

A Mobile Operator has access to a vast amount of mobile usage, demographic, and network related customer data. Of these only those that are most relevant to the problem at hand which is the measurement of the propensity of a subscriber to select a particular "SMS Alert" product category are relevant. Thus attributes such as the length of stay as a customer with the network (network stay), age, gender, mobile payment plan type (Pre Paid or Post Paid), use of International Direct Dial (IDD) Service, use of Value Added Services (VAS), average spend on VAS, average spend on Mobile Voice service and average spend on Data services together with the information on type of phone (Feature Phone or Smart Phone) used by the customer are considered important.

There is of course a tradeoff between selecting a large number of attributes to describe a subscriber which will increase the dimensions that are added to the ratings space (when the ratings matrix and ancillary data matrix are combined) and impact the complexity/speed of the algorithm which will also eventually result in negative returns as extraneous variables that do not impact the modeled user preferences will increases sparseness and decrease predictor accuracy. It may also introduce noise and spurious correlations not related to the problem of modeling the SMS Alert preferences and will impair performance.

The profiling attributes available for each subscriber has to be transformed in to a form that is amenable for incorporation in an extended ratings matrix/ ratings space. Thus the interval variables of Network Stay, Age, Average Mobile Revenue, Average spend on VAS services and Average spend on Data service were clustered using the K-Means clustering algorithm in to seven clusters. The cluster IDs that each user was assigned was then used to extend the ratings matrix by placing a "1" in the appropriate column of the extended ratings matrix that corresponds to the respective cluster ID number.

The categorical variables of gender (which can take values Male, Female, Corporate, and

Passport) were similarly assigned a value "1" in the appropriate column corresponding to each type of gender. Likewise the variables of payment plan type, whether an IDD user, and whether customer uses a Smart or Feature phone are all binary variables and represented in the extended ratings matrix by a "1" indicating its presence and "0" indicating its absence.

6. THE MODEL / ALGORITHM PROCEDURE

The ratings matrix for the subscribers is obtained by allocating for each subscriber the corresponding number of product types (categories) in his subscription.

The ancillary data used to profile the subscriber is transformed in to binary values {0,1} and is used to extend the ratings matrix as described in section 5. This includes the results from the K-Means clustering that grouped the customer's revenues, demographic and VAS related data. This extended ratings matrix has 31 columns that includes 14 items (product categories) and 17 profiling attributes.

This extended ratings matrix is then subject to the Singular Value Decomposition procedure described in section 4. Where estimates for U, $\sum$, and T are obtained. We then select 25 dimensions as it accounts for over 90% of the information contained in the original data, and truncate the decomposed matrices as described in section 4.

The reduced matrices are then used to predict the preferences of some members of the subscriber base who have had their ratings entries decremented by one (each user selected for prediction / validation had their ratings entry decremented by one). The ratings matrix chosen for this validation had at most 4 entries (or ratings) for the 14 product categories. This prediction takes the form given by equation (6).

$$U_C = A^{pred} T_k \sum_k^{-1} \qquad (6)$$

where

$$\sum_k^{-1} = \begin{pmatrix} 1/\lambda_1' & & & \\ & 1/\lambda_2' & & \\ & & \ldots & \\ & & & 1/\lambda_k' \end{pmatrix}$$

is the inverse of the diagonal matrix containing the first "k" singular values of $A^T A$. and T_k is a matrix containing the first "k" columns of T. Uc are the coordinates (in the transformed space) of the users with ratings matrix A^{pred} that are to have their ratings predicted for the choice of next best product.

Next the set of neighbors that are most highly correlated with each of the users identified in A^{pred} are found by computing a measure of correlation between each of the coordinates (U_c) of the selected users with that of the coordinates (U_i) of the

other users for whom ratings are available.

$$corr = \frac{U_C . U_i}{\sigma_{uc} . \sigma_{ui}} \qquad (7)$$

where σ_{uc} and σ_{ui} are the magnitudes of the vectors U_c and U_i respectively and "corr" in equation (7) is then the cosine similarity measure. This correlation is carried out in the transformed user vs. concept space (U). In this way a set of neighbors most similar to each of the users in A^{pred} are determined.

In the final stage the ratings for all items of a particular user are predicted by utilizing the correlation between each user (for whom ratings are to be predicted) with that of the users in his neighborhood determined in the previous stage. Only those neighbors that exhibit a level of correlation above a threshold are selected to be used in the prediction. The ratings of these selected users (r_j) are then scaled by the correlation level and summed to obtain the final ratings vector ($rpred_i$) for all products. This correlation and prediction are however carried out in the ratings space.

$$rpred_i = \sum_{j \in N_i} \frac{r_i \bullet r_j}{\|r_r\| \|r_j\|} r_j \quad (8)$$

where N_i is the neighborhood of users similar to the user for whom the ratings are predicted calculated in the previous step.

The item-item predictor is analogous to the user-user correlational prediction scheme in that it uses the correlation between the ratings given by other users to other items rated by the user in question. The similarity between the ratings for those other items given by other users to an item for which the user in question requires a prediction is what determines the item-item rating prediction.

The results of the user-user and item-item predictions are linearly combined to arrive at a hybrid predictor for each rating. We then determine the product with the highest rating for which the user has not yet given a rating. This then is the next best product recommendation for that user.

7. RECOMMENDER RESULTS FROM A MOBILE COMMUNICATIONS NETWORK

The next best item ratings for a random sample of 50 subscribers were predicted utilizing the algorithm described in section 6.

Figure 2, depicts the number of subscribers in a cluster, when the demographic, network and revenue variable attributes were clustered using K-Means. It indicates the existence of some well defined segments.

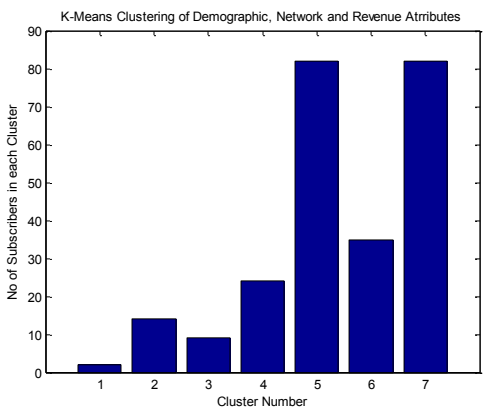

Figure 2: Size of K-Means Clusters

It was noted that using clustering to identify similar subscribers by clustering (coordinates) in the user vs. concept matrix (U) instead of the correlational approach we used did not give accurate results for creating valid neighbor lists for the users for whom ratings were to be predicted. This is likely to be because clustering produces groups around centroids and is not necessarily superior to the correlational approach in indicating the degree to which two users are similar to each other, even though the members of a particular cluster do share similar characteristics from the point of view of the similarity measure employed.

Figure 3, depicts the normalized cumulative sum of the singular values obtained when the augmented ratings matrix (ratings + subscriber profile information) was decomposed via the SVD procedure. It is clear that a dimension of 25 accounts for more than 90% of the cumulative sum of the normalized singular values and is therefore sufficient to represent the data instead of using all 31 dimensions.

In validating the predictive capability of the algorithm we compared the ratings suggested for the next best choice for a random sample of users for whom their ratings had already been decremented by one. It was seen that in the majority of the cases the algorithm correctly identified the rating that had been decremented and indicated that item as the recommended product category for that user. It was also possible to correctly identify with a higher majority (66%) the decremented rating as both the first or second preference for a given user.

We note that the identification of the decremented rating as the first or second preference does not preclude the scenario where the suggested rating for a third item may still be relevant to that user.

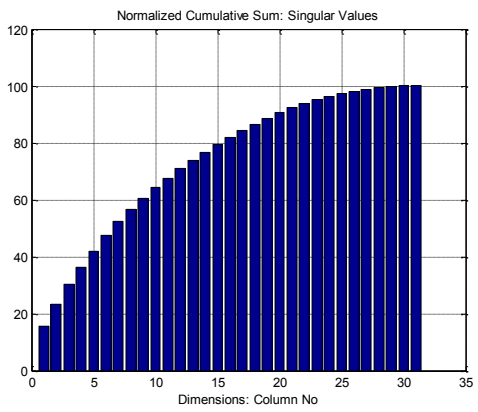

Figure 3: Normalized Cumulative Sum Singular Values

Additionally we note that due to the categorization of the items in to independent groupings, item-item similarity computations were less suitable than the user-user similarity procedure for prediction of ratings. In spite of this the combined effect of the hybrid user-user and item-item predictor outperformed the individual predictors.

It was also observed that the item-item predictor provided additional rating weight to those cases where the user-user predictor did not perform adequately, indicating that the item-item predictor plays an important role in the hybrid system compensating for lack of accuracy in a user-user only predictive system.

8. CONCLUSION

This paper developed a novel algorithm for recommending subscription SMS services to groups of subscribers of a mobile telecommunications operator. We proposed and validated a method utilizing both user-user and item-item recommenders combined in a hybrid configuration where each of the two predictors employed novel approaches. In the user based predictor customer profiling information was clustered to arrive at a condensed view of a customer's attributes while a SVD procedure was used to determine suitable neighborhoods to be used in the correlation based prediction. The item based predictor utilized a interest based categorization of items in to independent groups to rationalize the available choices and reduce sparsity in the ratings matrix.

We demonstrated that the utilization of ancillary customer data such as demographics, mobile usage patterns, and network related information and subscriber revenues greatly improves the overall performance of the recommender system. Through our study we were able to identify key demographic and mobile telecommunications specific attributes that had a bearing on modeling the individual user's preferences for mobile alert services.

We also demonstrated that the ratings space can be enriched by the use clustering without greatly increasing the dimensionality of the enhanced ratings space though the use of cluster IDs to represent groups of profiling variables.

In fact given the sparseness of the original ratings matrix and the categorization of the items in to more or less independent item-categories would otherwise make the recommendation task impractical. Thus we demonstrated that in instances when ratings are sparse and item-item similarities weak, utilization of customer profile information can greatly improve the user-user similarity calculation and the overall recommender system performance.

We also demonstrated that similarities computed in the user vs. concept space as opposed to directly in the ratings space utilizing all the information available about a subscriber helped identify correlations which were otherwise not discernible.

The determination of optimal weights for combining the user-user and item-item recommender results each derived independently using the existing algorithm in a hybrid configuration is left as future work. In this regard it is believed a dynamic weight may be employed for each user for whom ratings are predicted depending on the degree of correlations estimated and neighborhood sizes calculated in the user-user based predictor stage.

Taking in to account the current findings with respect to the modeling done thus far it is believed that a graph theoretic approach to aid in the similarity calculation and neighborhood determination may help further improve performance. Future work in this direction can

consider using path lengths between users in the similarity calculation in the instances where the correlational approach does not perform as well.

9. ACKNOWLEDGEMENTS

Shafraz Rahim, BSc. (Hons) MIS, Engineer Product Service Innovation, for providing business intelligence and detailed SMS alert data records from a live Mobile Telecommunication Network.

10. REFERENCES

1) F. Ricci, L. Rokach, B Shapira, Recommender Systems Handbook, Springer, 2010, pp 107-153.

2) http://home.dei.polimi.it/mattteucc/Clustering , extracted 5 May 2015.

3) V. Castelli, A. Thomas, CSVD: Cluster and Singular Value Decomposition for Approximate Similarity Searches in High Dimensional Spaces, IBM Research Report, May 2000.

4) M. Gu, C. Ding, H. Zha, Bipartite graph partitioning and data clustering, CIKM, 2001.

5) J. Riedl etal, Application of dimensionality reduction in recommender system – A case study, Dept of Computer Sceience University of Minnesota, July 2014..

6) G. Karypis, etal, Analysis of recommender algorithms for ecommerce, dept of COmouter Sceience University of Minnesota, 2000.

7) K. Margaritis, etal, analysis of recommnder system algorithms, Dept of Applied Informatics, University of Macedonia, (2000).

8) W. Du, etal, SVD collaborative filtering with privacy,2005.

9) C. Zhen, etal, RBRA: As simple and efficient rating-based recommender algorithm to cope with sparisity in recommender systems, Beijing University of Posts and Telecommunications.

10) Brin, S., Page, L., The anatomy of a large scale hypertextual web search engine., Proceedings of 7th International WWW Conference, (1998).

11) S. Spiegel, A hybrid aprach to recommender systems based on matrix factorization, Technical University Berlin, New product diffusion models in marketing: A review and directions for research., Journal of marketing, (1990).