A Collaborative Filtering Recommender System with Randomized Learning Rate and Regularized Parameter

M V V R Murali Krishna Rao
ASTROSAT Payload Operation Centre
Computer Centre, IUCAA
Pune, India
Email – mvvrmkr@gmail.com

***Abstract* - Recommender systems with the approach of collaborative filtering by using the algorithms of machine learning gives better optimized results. But selecting the appropriate learning rate and regularized parameter is not an easy task. RMSE changes from one set of these values to others. The best set of these parameters has to be selected so that the RMSE must be optimized. In this paper we proposed a method to resolve this problem. Our proposed system selects appropriate learning rate and regularized parameter for given data.**

Key words – Recommender Systems, Collaborative Filtering, Machine Learning, Learning Rate, Regularized Parameter, RMSE.

I INTRODUCTION

Now a day's e-commerce websites are playing great role in the world of business. Day by day the number of this type of sites is augmenting very rapidly. To be in race the system must be well organized and must be customer friendly. One of the best things used by organizations to attract customers is recommending the items to them which they like most, which is done best by using of recommender system. The main aim of the recommender system is to predict preference of item to user based on some principles.

Mainly two approaches are there for this, one is collaborative filtering and another one is content based filtering. The later one recommends the items to user which is similar to the items liked by user. Whereas the former one recommend the item to particular user based on finding the similar user to the selected user. There are two types in collaborative filtering, one is memory based, which recommend items to user based on his past data, another one is model based which implement some model to give recommendation item to user. Another hybrid method is also there which combine both collaborative filtering and content based filtering. This paper deals with collaborative filtering approach.

II LITERATURE SURVEY

Various papers and approaches of recommender systems are carefully studied. Mingrui Wu, who is working at Max Planck Institute for Biological Cybernetics published a paper on collaborative recommender systems which deals with all types of matrix factorizations such as Regularized Matrix Factorization (RMF), Maximum Margin Matrix Factorization (MMMF) and Negative Matrix Factorization[1].Takacs G, Pilaszy I, Nemeth B and Domonkos Tikk proposed several matrix factorizations approaches such as positive MF approach, momentum-based MF approach and hybrid MF neighbor based method[2].Xin Luo, Yunni Xia and Qingsheng Zhu did work in the regularized matrix factorization approach for collaborative filtering and simplified the training rule of the matrices. They mentioned that this yield a good results, which is illustrated by their paper in the ELSEVIER Knowledge-Based Systems journal[3].In the International Conference on Data Mining, which is held on 2012, Hsiang Fu Yu, Cho-Jui Hsieh Si si and Dhillon I, presented a paper regarding the matrix factorization and in that they mentioned that coordinate decent based methods have more efficient update rules than ALS and have more stable convergence than SGD[4].Xin Luo, in the year of 2012 with Yunni Xia and Qingsheng Zhu did a project in which they implemented the recommender system with adaptive learning rate techniques. Later they mentioned their work through ELSEVIER[5].Apart from individual recommender system, group recommender systems also plays very crucial role, Christensen I and Schiaffino S proposed a matrix factorization approach for group recommender systems[6].Oleksandra Krasnoshock and Yngve Lamo implemented a boosted matrix factorization algorithm to increase the prediction accuracy, which is the key point of recommender system. They published this paper on 18th International Conference on Knowledge-Base and Intelligent

978-1-5090-1936-6/16/$31.00 ©2016 IEEE

Information & Engineering Systems[7].The sparsity problem of collaborative filtering is attacked in the paper of Bu Sung Kim, Heera Kim, Jaedong Lee and Jee-Hyong Lee. In this paper they used user tag matrix to resolve the sparsity problem of CF[8].Another notable work in the field of matrix factorization was did by Jim Jing-Yan and Xin Gao, they had developed an algorithm for non-negative matrix factorization, which aim is to improve the discriminative ability of NMF. For that they worked by maximizing and minimizing the distance between the class pairs[9].The authors Bang Hai Le, Kien Quang Nguyen and Thawonmas R published a paper in of the IEEE conferences, in which they proposed a new matrix factorization approach for recommender system, which they called a bounded SVD[10].In 2015, Dheeraj Bokde and his crew presented a paper in the conference of International Conference on Advanced Computing, Communication and Control. In that they discussed various matrix factorization models, for instance Singular Value Decomposition (SVD), Principal Component Analysis (PCA) and Probabilistic Matrix Factorization (PMF) etc. And they also mentioned how to overcome the challenges of CF by using these matrix factorizations[11].Mohammed Wasid and Vibhor Kant published a paper in 11th Multi-Conference on Information Processing, in that paper they worked to increase the accuracy of memory based collaborative filtering and scalability of model based collaborative filtering. For that they used the fuzzy sets and Particle Swarm Optimization (PSO) algorithm[12]. The books, Data Mining : Concepts & Techniques by Jaiwei Han, Micheline Kamber, Jian Pei[13] and Artificial Intelligence A Modern Approach by Stuart J Russell, Peter Norvig[14] are studied.

III METHOD

The proposed system deals with collaborative filtering which is implemented by regularized matrix factorization. In this the matrix Y is the dataset which contain the information. In which rows are the users and columns are the items. For instance a Yij value is the rating of i^{th} user to j^{th} item. The aim is to find the U and V (where U and V are random matrices) such that U*V include predicted ratings with the original ratings.

The data loss equation used to decrease the loss of data while updating the matrix through machine learning algorithms. When the data loss is optimized that particular U and V become the required matrices.

Here λ is the regularized parmeter and τ is the learning rate.

Data loss Equation

$$\min_{U,V} \sum_{i,j \epsilon S} \left(y_{ij} - \boldsymbol{u_i}^T \boldsymbol{v_j}\right)^2 + \frac{\lambda}{2}(\|U\|^2 + \|V\|^2)$$
$$S = \{ij \mid y_{ij} > 0\}$$

Learning of U and V

$$\frac{\partial L}{\partial u_{il}} = \lambda u_{il} - 2 \sum_{j \mid i,j \,\epsilon\, S} \left(y_{ij} - \boldsymbol{u_i}^T \boldsymbol{v}_j\right) v_{lj}$$

$$\frac{\partial L}{\partial v_{lj}} = \lambda v_{lj} - 2 \sum_{i \mid i,j \,\epsilon\, S} \left(y_{ij} - \boldsymbol{u_i}^T \boldsymbol{v_j}\right) u_{il}$$

Updating of U and V

$$u_{il}^{(t+1)} = u_{il}^{(t)} - \tau \frac{\partial L}{\partial u_{il}^{(t)}}$$

$$v_{lj}^{(t+1)} = v_{lj}^{(t)} - \tau \frac{\partial L}{\partial v_{lj}^{(t)}}$$

But the learning rate and the regularized parameter plays very important role in this algorithm. One particular set of regularized parameter and learning rate will give efficient results than the other. But these set may vary from one particular data to another data and this also vary from one set of random U,V to other set of U,V . So we suggested a randomized learning rate(LR) and regularized parameter(RP) . At first we give some range to both RL and LR approach. In the very first step, the system fix one value let's say LR. For the given LR the system tries to optimize the RP value. After the RP values are optimized the system fixed with that RP and started to optimize LR. At final the system finished with optimized LR and RP set. This can also do in vice-versa.

The random RP and LR are picked based on linear interpolation.

$RP_r = RP_L+((RP_H-RP_L)\times(R_2-R_1))/(R_3-R_1)$

$LR_r = LR_L+((LR_H-LR_L)\times(R_2-R_1))/(R_3-R_1)$

where RP_r, LR_r is randomly picking RP and LR respectively

RP_H, RP_L, LR_H, LR_L are boundaries of the RP interval and LR interval respectively.

R_1,R_2,R_3 are random values which are in increasing order.

The updating of RP_L, RP_H and LR_L, LR_H

$RP_L = RP_r$ if RMES(RP_{k1}) < RMSE(RP_L)

$RP_H = RP_r$ if $RMES(RP_{k2}) < RMSE(RP_H)$

$LR_L = LR_r$ if $RMES(LR_{k1}) < RMSE(LR_L)$

$LR_H = LR_r$ if $RMES(LR_{k2}) < RMSE(LR_H)$

where $RMSE(RP_L)$ is the root mean square error with RP_L as regularized parameter.

RP_{k1} is the random value in the interval of (RP_L, RP_r)

RP_{k2} is the random value in the interval of (RP_r, RP_H)

Same is applicable to learning rate (LR)

In each iteration the interval shrinks towards the optimized value. When the interval shrinked to only one value, that value be the required optimized value. The error rate flows like a parabolic curve so each time either left boundary or right boundary modified.

IV RESULTS

The results of the system are shown in this section. We have worked with three type of datasets. Those are 10K, 50K and 100K represented with red, yellow and blue colours respectively in the graphs. In this graphical data we have shown the optimization of the Regularized Parameter for various initial learning rates. After that with the optimized regularized parameter, learning rate optimized is shown.

$$RMSE(Y^M,Y) = \sqrt{\frac{1}{|N|}\sum_{i,j\,\epsilon\,N}({y^m}_{ij} - y_{ij})^2}$$

where $Y^M = U^M * V^M$ where U^M and V^M are optimized matrices

Y is the given data matrix

The graphs are drawn in X axis the values of regularized parameter (from figure 1 to figure 4), learning rate (figure 5) and in Y axis RMSE values for the set of learning rate and regularized parameter. The graph is drawn until the global optimized value in a continuous flow with the random values.

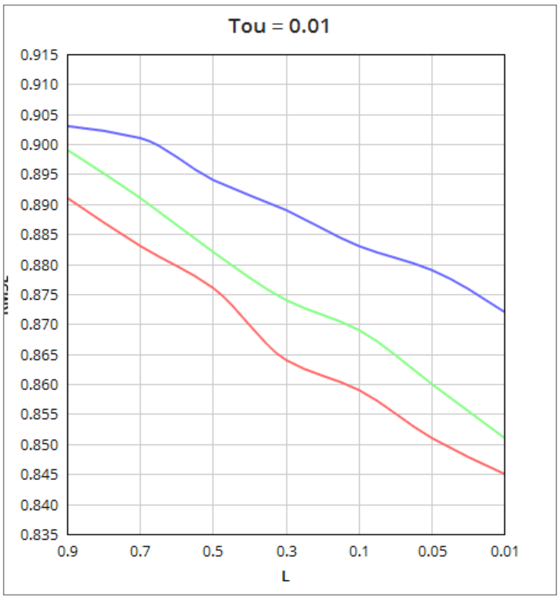

Figure 1 : Optimizing the regularized parameter if the learning rate is selected as 0.01

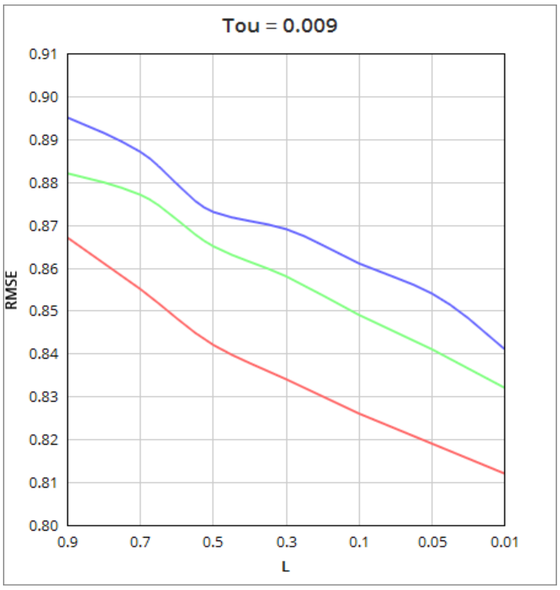

Figure 2 : Optimizing the regularized parameter if the learning rate is selected as 0.009

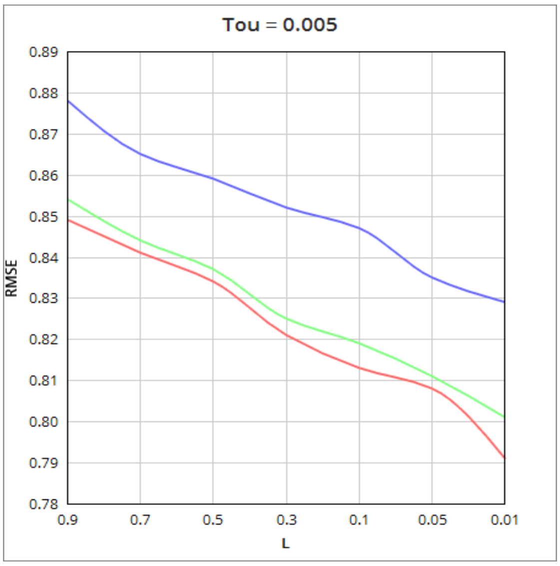

Figure 3 : Optimizing the regularized parameter if the learning rate is selected as 0.005

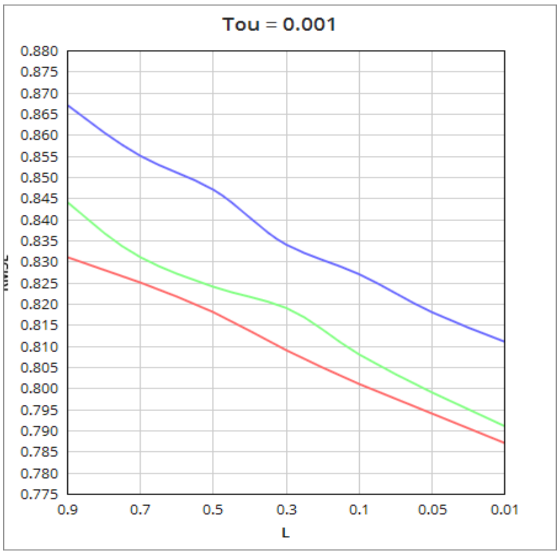

Figure 4 : Optimizing the regularized parameter if the learning rate is selected as 0.001

In general, selecting one learning rate sufficient to optimize the regularized parameter but here optimization of regularized parameters for various learning rates are shown to understand the process.

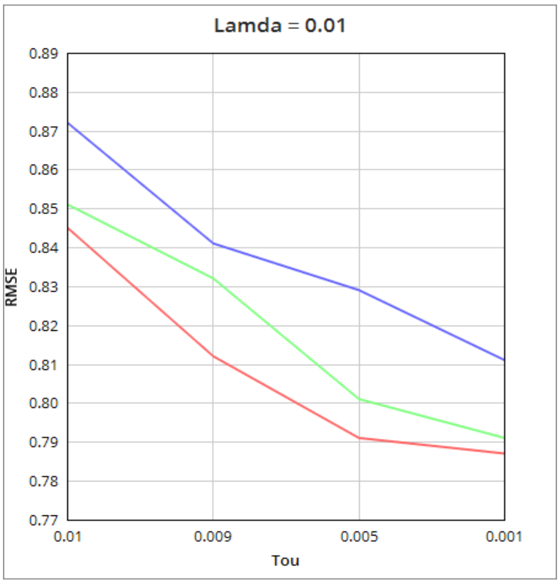

Figure 5 : Optimizing the learning rate at the optimized regularized parameter (which is 0.01 here)

V CONCLUSION & FUTURE WORK

In this papers we mentioned with a randomized function for finding the best regularized parameter and learning rate set for a given dataset. This function finds the suitable pair of LR and RP for the given data matrices. In our future work, we want to concentrate with implementing this work with maximum margin matrix factorization with the help of PSO approach. And this system is working with the assumption of flow of RMSE is without a local maximum (which is happen in most cases). Our future work is also deals with to resolve this problem.

REFERENCES

[1] Mingrui Wu "Collaborative Filtering via Ensembles of Matrix Factorizations" Max Plank Institute for Biological Cybernetics, Germany.

[2] Takacs G, Pilaszy I, Nemeth B, Domonkos Tikk "Investigation of Various Matrix Factorization Methods for Large Recommender Systems" International Conference on Data Mining Workshop (ICDMW) - 2008, IEEE, Pages 553-562

[3] Xin Luo, Yunni Xia, Qingsheng Zhu "Incremental Collaborative Filtering recommender based on Regularized Matrix Factorization" ELSEVIER Knowledge Based Systems Volume 27, March 2012, Pages 271–280

[4] Hsiang-Fu Yu, Cho-Jui Hsieh, Si si, Dhillon I "Scalable Coordinate Descent Approaches to Parallel Matrix Factorization for Recommender

Systems" International Conference on Data Mining (ICDM) - 2012, IEEE, Pages 765-774

[5] Xin Luo, Yunni Xia, Qingsheng Zhu "Applying the learning rate adaptation to the matrix factorization based collaborative filtering" ELSEVIER Knowledge-Based Systems Volume 37, January 2013, Pages 154–164

[6] Christensen I, Schiaffino S "Matrix Factorization in Social Group Recommender Systems" 12th Mexican International Conference on Artificial Intelligence (MICAI) - 2013, IEEE, Pages 10-16

[7] Oleksandr Krasnoshchok, Yngve Lamo "Extended content-boosted matrix factorization algorithm for recommender systems" 18th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES) - 2014

[8] Bu Sung Kim, Heera Kim, Jaedong Lee, Jee-Hyong Lee "Improving a recommender system by collective matrix factorization with tag information" 7th International Conference on and Advanced Intelligent Systems (SCIS) - 2014, IEEE, Pages 980-984

[9] Jim Jing-Yan Wang, Xin Gao "Max - min distance nonnegative matrix factorization" ELSEVIER Neural Networks Vol 61, January 2015, Pages 75-84

[10] Bang Hai Le, Kien Quang Nguyen, Thawonmas R "Bounded-SVD: A Matrix Factorization Method with Bound Constraints for Recommender Systems" Emerging Information Technology and Engineering Solutions (EITES) - 2015, IEEE, Pages 23-26

[11] Dheeraj Bokde, Sheetal Girase, Debajyoti Mukhopadhyay "Matrix Factorization Model in Collaborative Filtering Algorithms: A Survey" 4th International Conference on Advances in Computing, Communication and Control (ICAC3) - 2015

[12] Mohammed Wasid, Vibhor Kant "A Particle Swarm Approach to Collaborative Filtering based Recommender Systems through Fuzzy Features" 11th International Multi-Conference on Information Processing(IMCIP) - 2015

[13] Jiawei Han, Micheline Kamber, Jian Pei "Data Mining: Concepts and Techniques" Morgan Kaufmann Publishers ISBN-13: 978-0123814791, 3rd Revised edition edition (25 July 2011)

[14] Stuart J. Russell, Peter Norvig "Artificial Intelligence A Modern Approach" ISBN-13: 978-0136042594, Prentice Hall; 3 edition (1 December 2009)