International Journal of Engineering & Technology, 7 (4.39) (2018) 942-946

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Effectual Recommendations Using Concealed Feature Method

Noopur Saxena[1*], Shashank Awasthi[2], Arun Pratap Srivastav[3], Raj Gaurang Tiwari[4]

[1]Research Scholar, SRM University, Lucknow, INDIA
[23]G. L. Bajaj Institute of Technology & Management, INDIA
[4]SRMGPC, Lucknow, INDIA
**Corresponding author E-mail: noopur.saxena123@gmail.com*

Abstract

In the Collaborative Filtering, for the product recommendation, we not only consider the silhouette of the lively user but also consider the neighborhood of the lively consumer with analogous inclinations. In the approach of Collaborative filtering, we collaborate to assist each other in filtering the files they access, through using their reactions/comments. The recommender systems are exploited by massive researchers to improve the internet search. Content based filtering is another approach of recommender systems. In this paper, we concentrate on user's conduct rather than product/ object information. We determine the concealed characteristic of the product due to which product is highly/poorly rated by user. We estimate the missing rankings of unrated products by way of thinking about concealed characteristic and by using exploiting collaborative suggestion is performed.

Keywords: *Rating prediction; Web recommendation; web mining; usage mining*

1. Introduction

In the Collaborative Filtering approach, not only the profile of the active user is considered but also other users with similar preferences, referred to as the active user's vicinity is also considered for recommending items.
Now-a-days, collaborative filtering approach has become quite popular towards Personalization. This form of collaborative filtering based recommendation systems undergoes from following three tribulations:

1) Ascendable: As the time complexity of executing the nearest-neighbor algorithm increases linearly when quantity of items and users increases. Therefore, the recommendation device can't keep large-scale. Thus some approaches like dimensionality reduction, clustering and Bayesian Network, are exploited to tackle such problem.

2) Sparsity: As the nature of Profile matrix is supposed to be sparse due to colossal quantity of items and consumer disinclination to rate the items. As a result, the system cannot offer suggestions for various users, and the generated suggestions are no longer perfect[1].

3) Synonymy: Since contents of the items are definitely ignored, dormant affiliation between items is not viewed for recommendations. Consequently, provided that novel items are not rated, they are not recommended; so, fake negatives are initiated.

1.1 Content Based Recommendation

When we generate recommendations by comparing illustrations of content contained in an item with illustrations of content about which the user is fascinated, such approach is known as content-based recommendation. In this approach, we primarily develop a model of user ratings. In this model building process we exploit three diverse machine learning algorithms that is Clustering[2] Bayesian network and Rule-based models[3].
Followings are the weakness of the Content-based filtering systems:

1) Content constraint: IR strategies can solely be practical to a few sorts of substance, such as textual content and image, and the mined facets can only detain certain aspects of the substance.
2) Over-specialization: Content-based suggestion scheme offers recommendations truely based on user profiles. So, users have no prospect of exploring novel objects that are now not analogous to those objects covered in their profiles.

2. Literature Review

Chin-Chih Chang et al. [4] proposed a web service configuration method which was based on user ratings as well as collaborative filtering. They took the value of the web services, the reaction of users and the similarity between users in mind to choose web services. The proposed method was confirmed by a case study of the information system and the Mean Average Precision (MAP) is then estimated by the experiments.
Antonio Hernando et al. Allah [5] given the Recommendation systems based on cooperative nomination and visualization of element trees. It provides users with a quick and wonderful way to understand recommendations. This type of visualization provides users with useful information about the reliability of recommendations and the importance of user assessments, which may help users determine which recommendation to choose.
Bo Wang [6] used the concept of oncology and proposed a welcoming approach to the personal recommendation for liquidation. It is recommendation approach to traditional recommended problems, such as matrix variability and cold start problems.
To alleviate the problem of inequality and cold start, confidence is incorporated into collaborative filtering approaches while encouraging experimental results. Such collaborative filtering based on

Ttrust was proposed by Sung, William et al. Allah [7]. Their approach generates and spreads trust in a social network. They applied this method to measure the level of confidence in user hotel ratings, and demonstrated their usefulness by comparing test results with traditional collaborative filtering methods.

In the new approach suggested by Jun Zhang et. Allah [8], the similarity of the user is calculated on the basis of the weighted binary network and the principle of resource allocation for the cooperative nomination recommendation. They calculated the asymmetric balanced user matrix and translated it into a similarity matrix with the user. They conducted extensive experiments on the Movilens data set and demonstrated that the proposed approach could result in better recommendation accuracy and in part could alleviate the problem of interruptions.

Based on the navigation table of the active user, Samuel Nowakowski and Anne Boyer [9] presented a innovative approach in generating appropriate recommendations. Their main idea to solve this obscurity is to consider that users who browse web pages or web content can be seen as things moving along paths in the web space. With this hypothesis, they published the appropriate description of the so-called recommendation area to suggest a mathematical model describing the behavior of users / targets in the web / along the paths within the recommendation area. The second major hypothesis can be expressed as follows: If they are able to track users / targets along their paths, they can predict potential locations in the sub-spaces of the recommendation area, that is, they were able to devise a new technique for web recommendation and behavior monitoring. To achieve these goals, they used the theory of dynamic state estimation and more specifically the Kalman theory. They determine the appropriate model of target tracking and they derive repeated wording of the filter. They then propose a new system for bidders to form formulas as control. They have proven their approach to data extracted from online video and devised to user-monitoring approach. Conclusions and perspectives were derived from the investigation of the findings and focus on the formulation of the topology of the Recommendation area.

3. Proposed Concealed Feature Method

In this work we build a more accurate combined model by merging the features of product and neighborhood models.

3.1 Objective

In the Recommendations system, there is a group of users and a set of elements. If each user rates some elements in the system, we would like to know how users evaluate the elements that have not yet been evaluated, in order to create recommendations for users. In this case, all the information we have about the current evaluations can be represented in a matrix. Let's say now that we have 5 users and 10 elements, and the estimates are integers ranging from 1 to 5, the matrix can look like this (the script means that the user has not yet categorized the element):

	D1	D2	D3	D4
U1	5	3	-	1
U2	4	-	-	1
U3	1	1	-	5
U4	1	-	-	4
U5	-	1	5	4

Thus, the task of visualizing missing valuations can be considered as filling the spaces (the hyphens in the matrix), so that the values are consistent with the classifications in the matrix.

3.2 Idea

The intuition behind this narrow solution is that there must be some hidden features that determine how the user classifies an item. For example, two users will give high ratings for a particular movie if they both like the actors / actresses in the movie, or if the movie is a motion picture, a favorite genre for each user. Therefore, if we are able to verify these hidden features, we must be able to predict a specific classification for a specific user and element, because user-related features must match the attributes associated with the element.

3.3 Approach

Let U be set of users, D be set of items and R be the matrix of size $|U| \times |D|$ that contains all the ratings the users have provided to the different items. Also, we presume that we would like to discover K concealed features. Then our major task is to find two matrics matrices **P** (a $|U| \times K$ matrix) and **Q** (a $|D| \times K$ matrix) such that their product approximates **R**:

$$R \approx P \times Q^T = \hat{R}$$

In this manner, each row of matrix **P** would signify the strength of the associations between a user and features. Correspondingly, each row of matrix **Q** would characterize the potency of the associations between an item and the features. To get the prediction of a rating of an item d_j by u_i, we can calculate the dot product of the two vectors corresponding to u_i and d_j:

$$\hat{r}_{ij} = p_i^T q_j = \sum_{k=1}^{k} p_{ik} q_{kj}$$

Now, we have to determine a way to get P and Q matrices. One spectacular way to address this hitch is to first initialize the both matrices with some random values, calculate how different your product is from R and then try to iteratively minimize this disparity. This method is called a gradient slope, with the objective of finding a local minimum of the difference. This divergence is also termed as error between the actual rating and the estimated rating. It is calculated using the following equation for each pair of user elements:

$$e_{ij}^2 = (r_{ij} - \hat{r}_{ij})^2 = (r_{ij} - \sum_{k=1}^{K} p_{ik} q_{kj})^2$$

As the estimated rating can be either greater or smaller than the real rating, we have taken the squared error.

To diminish the error, we have to know in which direction we have to alter the values of $\mathbf{p_{ik}}$ and $\mathbf{q_{kj}}$. In other words, we must have the knowledge of gradient at the current values, and therefore we need to differentiate the above equation with respect to $\mathbf{p_{ik}}$ and $\mathbf{q_{kj}}$ separately:

$$\frac{\partial}{\partial p_{ik}} e_{ij}^2 = -2(r_{ij} - \hat{r}_{ij})(q_{kj}) = -2 e_{ij} q_{kj}$$
$$\frac{\partial}{\partial q_{ik}} e_{ij}^2 = -2(r_{ij} - \hat{r}_{ij})(p_{ik}) = -2 e_{ij} p_{ik}$$

After determination of gradient we formulate the update rules for both $\mathbf{p_{ik}}$ and $\mathbf{q_{kj}}$ as follows:

$$p'_{ik} = p_{ik} + \alpha \frac{\partial}{\partial p_{ik}} e_{ij}^2 = p_{ik} + 2\alpha e_{ij} q_{kj}$$
$$q'_{kj} = q_{kj} + \alpha \frac{\partial}{\partial q_{kj}} e_{ij}^2 = q_{kj} + 2\alpha e_{ij} p_{ik}$$

Here, α is a constant whose value determines the rate of approaching the minimum. Usually we will choose a small value for α, say 0.0002.

3.4 Proposed Algorithm

Step 1: Initialize P and Q with random small number

Step 2: For step until max_step

For Row Col in R

If R[Row][Col] > 0

Compute squared error as follows:

$$e_{ij}^2 = (r_{ij} - \hat{r}_{ij})^2 = (r_{ij} - \sum_{k=1}^{K} p_{ik} q_{kj})^2$$

Compute gradient from error as follows:

$$\frac{\partial}{\partial p_{ik}} e_{ij}^2 = -2(r_{ij} - \hat{r}_{ij})(q_{kj}) = -2 e_{ij} q_{kj}$$
$$\frac{\partial}{\partial q_{ik}} e_{ij}^2 = -2(r_{ij} - \hat{r}_{ij})(p_{ik}) = -2 e_{ij} p_{ik}$$

Update P and Q with new entry as follows:

$$p'_{ik} = p_{ik} + \alpha \frac{\partial}{\partial p_{ik}} e_{ij}^2 = p_{ik} + 2\alpha e_{ij} q_{kj}$$
$$q'_{kj} = q_{kj} + \alpha \frac{\partial}{\partial q_{kj}} e_{ij}^2 = q_{kj} + 2\alpha e_{ij} p_{ik}$$

Compute total error

If error < threshold

Break

Step 3: Return P, Q

4. Illustration

Suppose following is User Item rating matrix R

users

items

1		3			5			5		4	
		5	4			4			2	1	3
2	4		1	2		3		4	3	5	
	2	4		5			4			2	
		4	3	4	2					2	5
1		3		3			2			4	

- Find Q and P such that $\mathbf{R} \approx \boldsymbol{Q} \cdot \boldsymbol{P}^T$
- For now let's assume we can approximate the rating matrix $\boldsymbol{R}$ as a product of $\boldsymbol{Q} \cdot \mathbf{P}^{\mathrm{T}}$
- $\boldsymbol{R}$ has missing entries but let's ignore that for now!
- Basically, we will want the reconstruction error to be small on known ratings and we don't care about the values on the missing ones

Concealed feature

Q =

.1	-.4	.2
-.5	.6	.5
-.2	.3	.5
1.1	2.1	.3
-.7	2.1	-2
-1	.7	.3

items

$\mathbf{P}^{\mathrm{T}}$ =

1.1	-.2	.3	.5	-2	-.5	.8	-.4	.3	1.4	2.4	-.9
-.8	.7	.5	1.4	.3	-1	1.4	2.9	-.7	1.2	-.1	1.3
2.1	-.4	.6	1.7	2.4	.9	-.3	.4	.8	.7	-.6	.1

users

Estimation of the missing rating of user x for item i

$$\hat{r}_{ij} = p_i^T q_j = \sum_{k=1}^{k} p_{ik} q_{kj}$$

If we have to predict the following missing rating

Users

1		3			5			5		4	
		5	4	?		4			2	1	3
2	4		1	2		3		4	3	5	
	2	4		5			4			2	
		4	3	4	2					2	5
1		3		3			2			4	

Items

On applying the dot product predicted value is 2.38

.1	-.4	.2
-.5	.6	.5
-.2	.3	.5
1.1	2.1	.3
-.7	2.1	-2
-1	.7	.3

•

1.1	-.2	.3	.5	-2	-.5	.8	-.4	.3	1.4	2.4	-.9
-.8	.7	.5	1.4	.3	-1	1.4	2.9	-.7	1.2	-.1	1.3
2.1	-.4	.6	1.7	2.4	.9	-.3	.4	.8	.7	-.6	.1

5. Experimental Evaluation

To understand the operation of the proposed algorithm, the Concealed function method and the existing global average, user average, article average, greater popularity, UserKNN and Article KNN, we implemented the java application in the use of the eclipse IDE. We run all our experiments on a Windows-based PC with Intel Pentium III processor with a speed of 2.1 MHz and 2GB of RAM.

To apply the proposed and existing algorithms in the real recommendation scenario and to test the performance of the system, we use a FilmTrust data set consisting of 35497 classifications of elements in format: user ID, movie ID, movie rating.

The snapshots of the developed framework are the following:

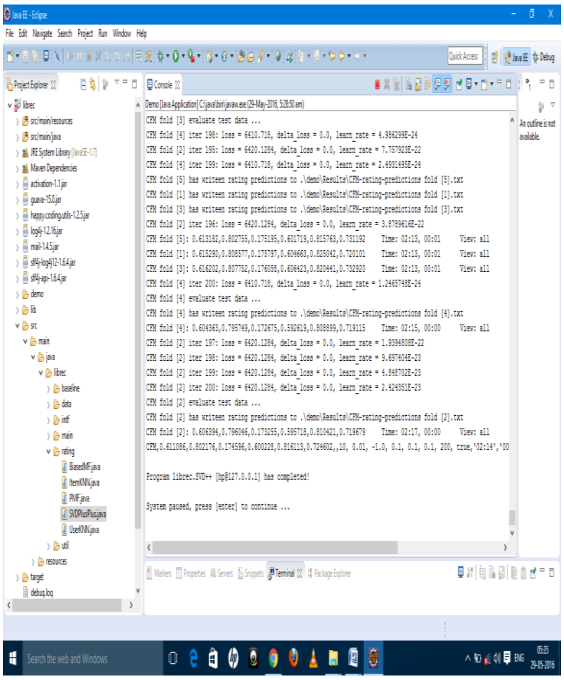

Fig. 1: Snapshot displaying Concealed Feature Method prediction results

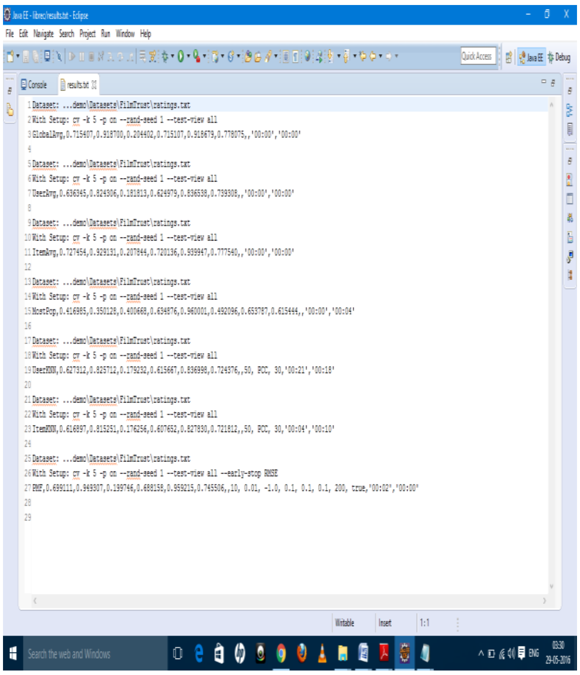

Fig. 2: Snapshot of all the results at one glance

6. Performance Evaluation

Recommender systems research has used numerous kinds of measures for assessing the quality of a recommender system. Following evaluation criteria have been used in this study:

Mean Absolute Error (MAE): MAE is measure of the divergence of recommendations from their accurate user-specified values. For each ratings prediction pair < pi,qi> this metric treats the absolute error between them i.e., |pi − qi | equally. For calculation of MAE we first sum these absolute errors of the N corresponding ratings-prediction pairs and then compute the average. Formally,

$$MAE = \frac{\sum_{i=1}^{N} |p_i - q_i|}{N}$$

The lower the MAE, the more precisely the recommendation engine predicts user ratings.

Root Mean Square Error(RMSE): It is well known measure of the divergence between values envisaged by a model and the values actually observed from the environment that is being modeled.

The RMSE is defined as the square root of the mean squared error:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (X_{obs,i} - X_{model,i})^2}{n}}$$

where X_{obs} is observed values and X_{model} is modelled values at time/place i.

MAE and RMSE values of the algorithm used in this study is as follows.

The apparent conclusion from following figure is that our proposed approach is superior.

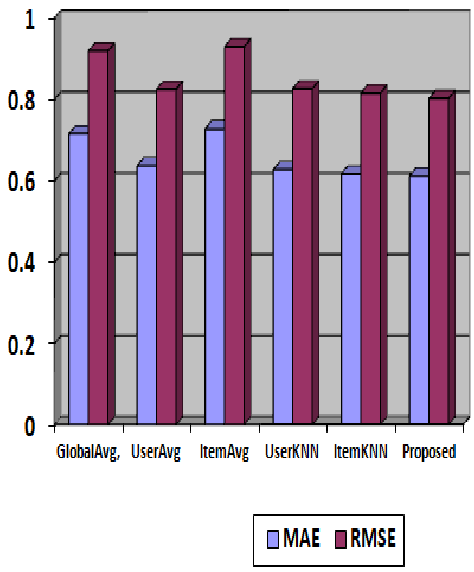

Fig. 3: Comparison

7. Conclusion

In this research, the ideas of combining the dynamic user profile and collaborative filtering have generated great interest for IR. We proposed a recommendation approach using collaborative filtering.

It is analyzed from experiments that this approach does not mislead users, since it obtains implicit representations of them, transforming the results of the appropriate recommendations into effective results.

References

[1] Melville P and Raymond J Mooney and Ramadass N (2002)., Content- Boosted Collaborative Filtering for Improved Recommendations, Department of Computer Sciences, University of Texas, Austin, TX 78712. (AAAI-2002), Edmonton, Canada, pp.187-192,

[2] Gauch S, Speretta M, Chandramouli A and Micarelli A (2007), User profiles for personalized information access, Lecture Notes in Computer Science, 4321, pp. 54-60,

[3] Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001l), Item-based collaborative filtering recommendation algorithms. In Proceedings of the 10th international conference on World Wide Web , pp. 285-295.

[4] Chin-Chih Chang, Chu-Yen Kuo (2013), A Web Service Selection Mechanism Based on User Ratings and Collaborative Filtering, Smart Innovation, Systems and Technologies, Vol. 20, pp.439-449.

[5] Antonio Hernando (2013), Trees for explaining recommendations made through collaborative filtering, Information Sciences, Vol. 239, pp. 1–17.

[6] Pu Wang(2012), “An Ontology-Based Collaborative Filtering Personalized Recommendation”, Applied Mechanics and Materials, Vol. 267, pp. pp.79-82.

[7] Song, W. W., Wu, Q., Forsman, A., & Yu, Z. (2013), A computational model for trust-based collaborative filtering: an empirical study of hotel recommendations. In 26th European Conference on Operational Research, Rome, Vol. 8182, pp. 266-279.

[8] Jun Zhang et. al.(2012), “A Novel Similarity Measure Based on Weighted Bipartite Network for Collaborative Filtering Recommendation”, Applied Mechanics and Materials, Volumes 263 - 266, pp.1834-1837.

[9] Samuel Nowakowski, Anne Boyer(2013), Automatic tracking and control for web recommendation New approaches for web recommendation, International Journal On Advances in Intelligent Systems..