International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 2 Issue: 11 3428 – 3432

Hybrid based Collaborative Filtering with Temporal Dynamics

Çiğdem Bakır
Yildiz Technical University
Computer Engineering Department
İstanbul,Turkey
cigdem@ce.yildiz.edu.tr

***Abstract*—**Hybrid-based collaborative filters use some part or entire database relating to user preferences for making recommendations for new products and new users. In our time, it is of utmost importance to make recommendations in line with interests and demands of users by making their interest alive. However, although Hybrid-based collaborative filters are used in this area, changing of preferences of users in a time, emergence of new products and new users overshadow success of such systems. Traditional hybrid-based collaborative filtering (CF) technique become insufficient for responding interests and demands changing in a time. For this reason, temporal changes in recommendation systems become an important concept. Together with the study conducted, an appropriate and new method has been developed in line with changing pleasure and demands depending on time. In the recommended system, unlike traditional hybrid technique based CF technique, point given to the products depending on dates scored by users has been attempted to be estimated. In this study, process has been made over netflix data for measuring success of both traditional hybrid based CF technique and the recommended system. Quite successful and rewarding results have been obtained in the issue of accuracy of predicted points.

Keywords*-Recommendation System;, Data Mining; Temporal Dynamics.*

*****

I. INTRODUCTION

Collaborative filters are the systems applying techniques discovering information for making recommendations during interactions of products and services. Today, it is widely used in many areas. [3]. However, new product and services changing pleasures of the persons may arise while making recommendations. A number of factors changing perception and viewpoint of users may arise over time. For example, users may change type of film they prefer over time in film recommended systems. In addition, each user can undergo different variations. Or a person may change variation requirements occurred within a family structure or occurred within itself.

Giving accurate and reliable recommendations peculiar to each user may become highly difficult due to changing of demands of users depending on different factors. Hybrid-based CF technique may give recommendations by considering user and item similarity. However, hybrid-based CF technique may not make recommendations peculiar to pleasures of users for demands changing over time. Since these systems do not evaluate temporal variations, they suggest the same products to users in each time. If this is an e-commerce application, interest of customer will decrease since product in line with time-varying demand of the person and in this case, profitability of the firm drop accordingly. A system that will consider time variations for boosting firm profitability by keeping interest of users alive and that shall observe variations in behaviors of each user. In this study, user based and item based CF depending on the recommended time and temporal variations have been taken into consideration and prediction success has been attempted to be boosted. In addition to this, if a scored product is old, this means that that person has been scored by number of persons in this extent. In other words, older products have been seen and voted by more users in system. In prediction systems date of this product of score is as score as point of this product. Point scored in time close to our time is more valuable when compared with very old score of a product registered in the system. For this reason, ages of scores of users have been calculated in the recommended system and if score is young, it has been reinforced and if it is aged, it has been weakened. Evaluation of scores depending on their ages has significantly increased prediction success.

II. HYBRID APPROACH CF ALGORITHM

Since hybrid covers information deduction made according to both users and items, hybrid approach combines CF-U (user based) and CF-I (item based) techniques. If the($r_{7,6}$) score given for i_6 item by the u_7 user was attempted to be predicted by using user based CF technique in sample data sets in Schedule 2.4, u_4 , u_5 and u_6 users would be the most similar users with $u_7$users. In this case, it would determine the i_2 , i_4 and i_5 items that are preferences of the most similar users to u_7 user with the $r_{7,6}$ score value because these users show the same behaviors with active user. However, as seen in Figure 3, these items are related with football item and they are totally different from (i_6) English item requested to be found out. In this case, prediction to occur for its user by using only u_7 user based CF technique would not be accurate and reliable. On the other hand, when only item based CF technique is used, it is likely to encounter with the problems set out below. If the ($r_{7,6}$) point given for item by user was attempted to be predicted by

 ISSN: 2321-8169
3428 – 3432

using item based CF technique in sample data sets in Table 1, i_6 items would be the most similar items with i_1 and i_3 items. However, in this case, prediction will not be generated for the $u_7$user since u_4, u_5 and u_6 users have not provided any score for these items because there may be items that very few users score in CF data sets or there may be several items that are not scored by one user. This is a frequently encountered case in CF data sets. In this case, since it is not sufficient to use user based CF technique or only item based CF technique by itself, hybrid approach will be needed.

TABLE I
USER-ITEM MATRIX ON THE SCORES OF THE SAMPLE TABLE

User/ Item	i_1 English	i_2 Football	i_3 English	i_4 Football	i_5 Football	i_6 English
u_1	$r_{1,1}$=3	$r_{1,2}$=1	$r_{1,3}$=2	$r_{1,4}$=3	$r_{1,5}$=5	$r_{1,6}$=5
u_2	$r_{2,1}$=3	$r_{2,2}$=1	$r_{2,3}$=2	$r_{2,4}$=3	$r_{2,5}$=5	$r_{2,6}$=5
u_3	$r_{3,1}$=3	$r_{3,2}$=1	$r_{3,3}$=2	$r_{3,4}$=3	$r_{3,5}$=5	$r_{3,6}$=5
u_4	$r_{4,1}$=1	$r_{4,2}$=5	$r_{4,3}$=3	$r_{4,4}$=3	$r_{4,5}$=1	$r_{4,6}$=1
u_5	$r_{5,1}$=2	$r_{5,2}$=5	$r_{5,3}$= 2	$r_{5,4}$=3	$r_{5,5}$=2	$r_{5,6}$=1
u_6	$r_{6,1}$=3	$r_{6,2}$=5	$r_{6,3}$= 1	$r_{6,4}$=3	$r_{6,5}$=2	$r_{6,6}$=1
u_7	$r_{7,1}$=3	$r_{7,2}$=5	$r_{7,3}$=2	$r_{7,4}$=3	$r_{7,5}$=2	$r_{7,6}$=?

Hybrid algorithm steps are as follows:

- $w_{j,q}$ showing similarity between other items of target **j** items needed to be found out. This similarity is found out with Pearson correlation coefficient or other similarity methods. This statement shows that **j** item given by **u** user refers to average of the scores given to $\overline{r_j}$ **j** items and $r_{u,q}$ **u** users refers to **q** items and, $\overline{r_q}$ refers to average of the scores given **q** item. [1].

$$w_{j,q} = \frac{\sum_{u \in U} (r_{u,j} - \overline{r_j})(r_{u,q} - \overline{r_q})}{\sqrt{\sum_{u \in U} (r_{u,j} - \overline{r_j})^2} \sqrt{\sum_{u \in U} (r_{u,q} - \overline{r_q})^2}} (1)$$

- According to $w_{j,q}$ statement, similar items are identified for target item j -SI_j. There are two methods for identifying SI_j. The first method is to identify a specific a threshold value. If this threshold value is **s**, the items whose similarity value is greater than **s** will be taken as a *basis*. In the second method, the most similar items within a specific number will be taken from $w_{j,q}$ content. This value is represented with SI_j. [3].

- $w_{j\,(a,i)}$ Similarity of active user that is towards other users within SI_j is *calculated*. Here, **j** refers to the item, **a** refers to active user, **i**refers to other users within SI_j. It represents the items scored reciprocally for stating the value of similarities calculated towards other users of active user within $CSI_{j\,(a,i)} SI_j$.

$$w_{a,i} = \frac{\sum_{i=1}^{m} (r_{a,j} - \overline{r_a}) * (r_{i,j} - \overline{r_i})}{\sigma_a \sigma_i} (2)$$

In this statement, m refers to total number of items, $r_{a,j}$ **a** refers to the score given to **j** item of active user, $r_{i,j}$ refers to the score given to **j** item by **i**user, $\overline{r_a}$**a** refers to the average of the scores given to all items by active user, $\overline{r_i}$**i** refers to the average of the scores given to all items by its user, $\sigma_a ve \sigma_i$ refers to standard deviation of the points given to the items by **a** and **i**users. [1].

- According to active user, $w_{j\,(a,i)}$ similar neighbor users are identified. $Neighbor_{a,j}$ refers to similar neighbor users to active user. 2 methods are used for calculating $Neighbor_{a,j}$:

The first method is to select those greater than **w** by adjusting similarity threshold value of **w** absolute user. (**w**user correlation threshold value). Second method is the best-neighbor method. In this method, the most similar users are taken out of $w_{j\,(a,i)}$.

- After identifying $Neighbor_{a,j}$ users similar to active *user*, $P_{a,j}$ score given to **j** item of active user is calculated.

$$P_{a,j} = \overline{r_a} + k \sum_{i \varepsilon Neighbor_{a,j}} w_{j\,(a,i)} (r_{i,j} - \overline{r_i}) (3)$$

Similarity (for **j** item) of $w_{j\,(a,i)}$active user with other users refers to difference of average out of the score given to all items by the user than the score given to the item required to be predicted by $(r_{i,j} - \overline{r_i})$ user.

$$\frac{1}{k} = \sum_{i \varepsilon Neighbor_{a,j}} w_{j\,(a,i)} (4)$$

- Prediction *value* that intended for all items for $P_{a,j}$ (j ε I_{Pa}) active user is calculated. Recommendation is made to active user.

III. HYBRID BASED CF TECHNIQUE WITH TEMPORAL DYNAMICS

While similarity is calculated between users and items in traditional hybrid based CF technique, only the scores given for the products by the users are used. [1, 2]. The scores given to the product by the users are shown in user-product

 ISSN: 2321-8169
3428 – 3432

matrix form in Table 2. Score of active user is tried to be predicted according to user and item similarity calculated in these systems.

In this study, a method that has not been implemented previously has been developed hybrid based CF with temporal dynamics study. Unlike conventional hybrid based CF technique, age of the score given by the user has been added to evaluation system. In other words, user similarity and predicted score will vary according to age of the scores.

In hybrid based CF with temporal dynamics study, Age_{nm} **n** refers to the age of the score given to **m** product by **n** user. Let (r_{a3}) the score given to the product no. 3 by active user be test sample. In this recommended system, age of each score will be required while calculating similarity of one user or product. For example, while calculating age of r_{13}, it shall be looked in difference between score dates of r_{13} and test data. The date in which test data will be calculated is 28/01/2005 and since the difference date of r_{13} is 28/01/2003, it is 731 days, in other words, it is approximately 2 years. This period represents ages of r_{13} according to the date. (28/01/2005 testing date). Ages of the scores shown in Table 2 are dynamic information varying according to existing date.

TABLE II
USER-ITEM MATRIX ON THE SCORES OF THE SAMPLE TABLE WITH THE AGE OF SCORES GIVEN IN DAYS

Users/ Items	i_1	i_2		i_m
u₁	$r_{1,1}$ $Age_{1,1}$ *(353 days)* *10.02.2004*		$r_{1,3}$ $Age_{1,3}$ *(731 days)* *28.01.2003*	
u 2	$r_{2,1}$ $Age_{2,1}$ *(435 days)* *20.11.2003*			$r_{2,m}$ $Age_{2,m}$ *(27 days)* *01.01.2005*
a Active user	$r_{a,1}$ $Age_{a,1}$ *(469 days)* *17.10.2003*		***$r_{a,3}$=? $Age_{a,3}$ (28.01.2005) Test sample***	
uₙ	$r_{n,1}$ $Age_{n,1}$ *(136 days)* *14.09.2004*			$r_{n,m}$ $Age_{n,m}$ *(236 days)* *06.06.2004*

As seen in Table II, ages of the scores given are calculated on the basis of day according to testing date. In the system developed, ages of these scores are converted into year and used accordingly. Objective here is to reduce error in score prediction by using ages of the scores given in hybrid based CF technique. The scores given by the users in the method we have recommended have been used by weakening if ages of the scores are big (aged) or by reinforcing if ages of the scores are little (young). In this weakening or reinforcing procedure, various conversion functions have been tried and it has been attempted to find out the most appropriate conversion function. Conversion functions used for weighting the ages of the scores set out Fig.I are shown. This conversion function has been applied as set out in Table 2.

$$Age_{\mathrm{w}} = mAge_r + n \qquad (5)$$

In this function, Age_r refers to existing score age and Age_{w} refers to ages weighted of existing score. Since recommended method is executed by taking as a basis the scores given within two years to various films by Netflix customers, while Age_r is taken as a value in range of [0,…2], Age_{w} adopts the values calculated with conversion function shown in Table 2.Weighting result of Age_{w} existing score calculated with conversion function is used as shown in equation 5 in calculation of new value based on time.

In equation 6, r_{org} refers to existing score and r_{aged} refers to temporal dependent weighted score. In traditional methods described in Part 2 following this conversion procedure, recommendation system has become time dependent by using r_{aged} instead of r.

$$r_{aged} = r_{org} / Age_{\mathrm{w}} \qquad (6)$$

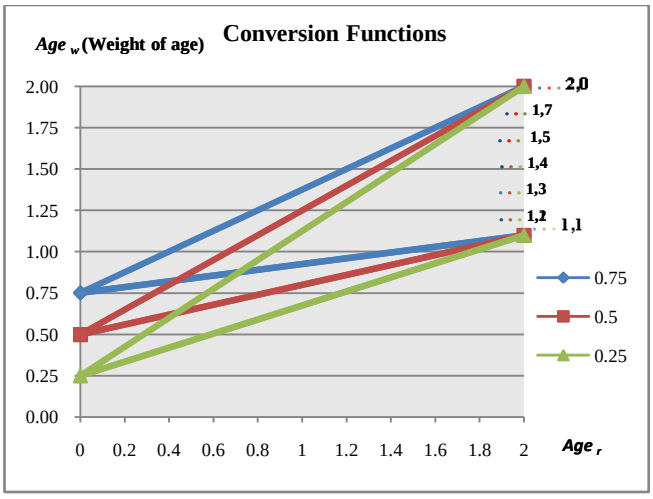

Fig.1. Conversion functions used for weighting the ages of scores

IV. EXPERIMENTAL STUDY AND RESULTS

Netflix data set borrowed from a company established for renting film and video in the USA has been used for measuring success of time dependent hybrid based IF technique. Netflix data set is composed of the scores given for 17770 films within the period elapsed from 1999 to 2005 by approximately half million customers. [6]. Results of the

 ISSN: 2321-8169
3428 – 3432

application performed have been calculated with RMSE (Room Mean Squared Error). RMSE is one of the evaluation criteria widely used in netflix data set.

$$\text{RMSE} \sqrt{\frac{1}{n}\sum_{\{i,j\}}(p_{ij} - r_{ij})2} \qquad (7)$$

In this equation, **n** refers to number of the scores prediction , the number of product scored by users in test set, p_{ij} refers to prediction score given to **j** item by **i**user, r_{ij} refers to real score given to **j** item by **i**user.

Error rate of traditional hybrid based CF technique is 31,86 % for the films group voted more and the users giving scores to more films. The results of hybrid based CF with temporal dynamics study are provided. In the curve given for 0.25 shown in green color, when weighted age rose from 1.1 to 2, it shows increase in error rate. When weighted age in the curve given for 0.5 shown in red color is 1.3, it is seen that error rate is the lowest value with 14 %.When traditional hybrid based CF technique is compared hybrid based CF technique with temporal dynamics, recommended system has yielded more approximately 17 % more successful outcome for the films groups voted more and the users giving scores to the film in greater number according to CF technique. Thanks to developed system, more accurate recommendations are offered to users by increasing prediction success and by considering age of existing scores as time dependent manner.

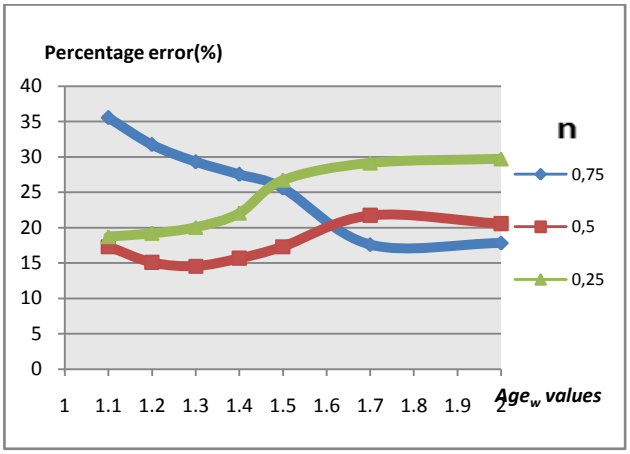

Fig.2-Hybrid CF with Temporal Dynamics of results

V. CONCLUSION

A score prediction is made by considering only scores of users of classical hybrid based CF technique. Thus, making recommendations peculiar to the person in line with demands, likes and requirements of the users varying over time becomes insufficient. With this study, deficient aspects of classical hybrid based CF technique has been attempted to be satisfied. Ages of existing scores have been reinforced by using various conversion functions and have been aged. In this way, more accurate and reliable system peculiar to the users has been designed. Product prediction success has been increased by developing a method considering time dependent variations of the users.

REFERENCES

[1] Herlocker J.L., Konstan J.A, Borchers A.& Rield J., "An Algorithmic Framework for Performing Collaborative Filtering , "ACM Conference on Research and Development Information Retrieval, pp.230-237, New York, 1999.

[2] Li Y., Lu L., Xuefeng L., "A hybrid collaborative method for multiple-interest and multiple-content recommendation in E-Commerce, " *International Journal of Expert System with Application*, pp.67-77,2005.

[3] Koren Y., "Collaborative Fİltering with Temporal Dynamics," *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.447-456, 2009.

[4] Sarwar B., Karypis G., Konstan J. and Rield J. , "Item-based Collaborative Filtering Recommendation Algorithms," *in WWW'01:International Conference on World Wide Web*, pp.285-295, 2001.

[5] Choi K., Yoo D., Kim G. ve Suh Y., "A hybrid online-product recommendation systems: Combining implicit rating-based collaborative filtering and sequential pattern analysis," *Elseiver Elektronic Commerce Research and Applications*, pp.309-317,2012.

[6] R.Burke, "Hybrid recommender systems: Survey and experiments, user modeling and user-adapted Interaction", 12,pp.331-370,2002.

[7] Su X., Khoshgoftaar T.M(2009). , "A Survey of Collaborative Filtering Techniques," *Advances in Artificial Intelligence*, no.Section 3, pp.1-20..

[8] Linden G., Smith B.&York J.(2003), "Amazon.com Recommendations Item-to-Item Collaborative Filtering, " IEEE Distributed Systems Online, pp.76-80,2003.

[9] Gao M., Wu Z., Jiang F., "Userrank for item-based Collabrative Filtering Recommendation", *Inf.Process.Lett.111*, vol.9, pp. 440-446, 2011.

[10] Linden G., Smith B. ve York J., "Amazon.com Recommendations Item-to-Item Collaborative Filtering, " *IEEE Distributed Systems Online*, pp.76-80, 2003.

[11] D.M.Pennock, H.Eric, S.Laurence ve C.L.Giles, "Collaborative filtering by Personality Diagnosis:A Hybrid Memory and Model-based approach", *in Processings of the 16th Conference on Uncertainty in Artificial Intelligence(UAI'00), pp.*473-480, 2000.

[12] D.Billsus ve M.Pazzani, "Learning collaborative information filters", in Processings of the 15th International Conference on Machine Learning(ICML'98), 1998.

[13] Wu J., Chen L., Feng Y., Zheng Z., "Prediction Quality of Service for Selection by Neighborhood-Based Collaborative Filtering, "*IEEE Transactions on Systems, Man, And Cybernetics:Systems*, vol.43, issue 2, 2013.

ISSN: 2321-8169
3428 – 3432

[14] Jia D., Zhang F., Liu S., "A Robust Collaborative Filtering Recommendation Algorithm on Multidimensional Trust Model, "*Journal of Software*, vol 8,no.1,pp.11-18, 2013.

[15] Zhang Z., Lin H., Liu K., "A hybrid fuzzy-based personalized recommender system for telecom products/services, "*Information Sciences*,vol.235,pp.117-129, 2013.

[16] Said A., Fields B., Jain B.J, Albayarak S., "User-Centric Evaluation of a K-Further Neighbor Collaborative Filtering Recommender Algorithm,"*CSCW 2013*, San Antonio, USA, 2013.

[17] Gong J.G., (2010). "A Collaborative Filtring Recommendation Algorithm Based on User Clustering and Item Clustering*," Academy Publisher*,pp.745-752,2010.

[18] Huang Z., Chen H., Zeng D., "Applying Associative Retrieval Techniques to Alleviate the Sparsity Problem in Collabrative Filtering",ACM Transactions on Information Systems, vol.22, issue 1,pp.116-142, 2004.