Hybrid Recommender System with Temporal Information

Farman Ullah, Ghulam Sarwar, Sung Chang Lee, Yun Kyung Park*, Kyeong Deok Moon*, Jin Tae Kim*
School of Information and Communications
Korea Aerospace University, Republic of Korea, *ETRI
{farman, sarwar, sclee}@kau.ac.kr, *{parkyk, kdmoon, jtkim}@etri.re.kr

Abstract: In the last few years, many recommender systems have been proposed but most of them suffer from scalability, sparsity and cold start issues. The existing recommender systems don't consider contextual information in term of user current device, location, company and time etc. In this paper, we proposed Hybrid Recommender System that accounts item attributes similarity, user rating similarity, user demographic similarity and the temporal information to do recommendation. The proposed algorithm will produce better results as it uses temporal information in computing and uses hybrid structure, model-based and memory-based system to improve system scalability and accuracy simultaneously. It uses the temporal information in the recommendation process to make recommendation for user at specific time.

Keywords: Sparsity; Relative feature score time; Static temporal information; Demographic information; Hybrid model with temporal information

1. INTRODUCTION

In past few years, the advent of power computational and broadband network technologies spurred much interest in multimedia research. Multimedia applications in the areas of education, business and entertainment are gaining more popularity in the internet community. As the density of available data, contents or items are growing exponentially; User Preferences and Recommendation Systems have gained much popularity and are implemented in a number of applications like widely used in E – Commerce, Automatic IPTV content recommendation, Media sharing, Net Surfing and Web Service selection etc. Collaborative filtering recommender systems are used widely and do recommendations on bases of similar user preferences. Collaborative filtering can be divided into two types, memory-based [6] and model-based [7]. Memory-based collaborative filtering uses all data with high recommendation accuracy; however it is costly in computing with bad scalability. Model-Based collaborative filtering first constructs a model on all data offline, and then generates the recommendation online on the basis of the model. It improves system scalability at the cost of accuracy.

In this paper, we use method which first constructs a user model offline by combining Rating Similarity, Attribute Similarity, temporal Information and User Demographic Information. Do neighbor selection and recommendation on the temporal aware hybrid model, rating matrix and some dynamic contextual information i.e. current time. The improved recommender system will have accuracy of memory-based systems and scalability of model- based system. The computational times of our proposed system will also less as compared to previous approaches because we are considering contents that user like at specific time. For example a user may like a funny movie and game with friends in working hours i.e. at school time and job time, however the same user like to see drama with family members and suspense type movie at personal room. So our recommender system considers a portion of rating matrix, which improves system scalability by minimizing the number of items to be considered.

The paper is organized as follow: Section-I briefly describe about hybrid recommender system using temporal information. Section-II explains the offline construction of temporal aware hybrid user model. Section-III describes collaborative filter recommendation. Section-IV describes our simulation and results, and last section is about conclusions and future work.

II. TEMPORAL AWARE HYBRID RECOMMENDER SYSTEM

Conventional recommender systems use only user-rating matrix for recommendation [1], some improved recommender systems use attribute similarity [2] and demographic similarity [10]. Some proposed recommender system used all the three types of similarity [4][5], and some combine two type similarity to produce prediction [8]. As the number of users and contents increases, the complexity and size of rating matrix increases, so it becomes difficult for conventional systems to find accurate neighbor within user-bearable time. Conventional systems recommend most liked contents for a user, don't consider whether user like such content at that specific time or not. In our proposed Temporal Aware hybrid recommender system the recommendation will be made from the set of content that user like at that time.

The workflow of Hybrid User Model with Temporal information is shown in Figure.1. First of all, the Clustering Engine classifies the contents based on their attribute similarity into some classification [2] to fill the vacant spaces in user rating matrix. For new content or user, no rating is available so the clustering engine uses demographic information to fill the vacant spaces in the rating matrix to avoid sparsity and improve accuracy [10]. Then it constructs temporal aware hybrid model offline on the basis of Relative Feature-Score-Time matrix and demographic information. The model is much smaller than user-item rating matrix and tries to improve system scalability. Secondly generates the neighbor set by using model-based collaborative filtering on temporal aware hybrid model. It requires less computational time for neighbor selection as the recommender system considers only those contents that user like at that time. Finally, make recommendation on whole rating matrix, temporal information and neighbors set on memory based collaborative filtering algorithm.

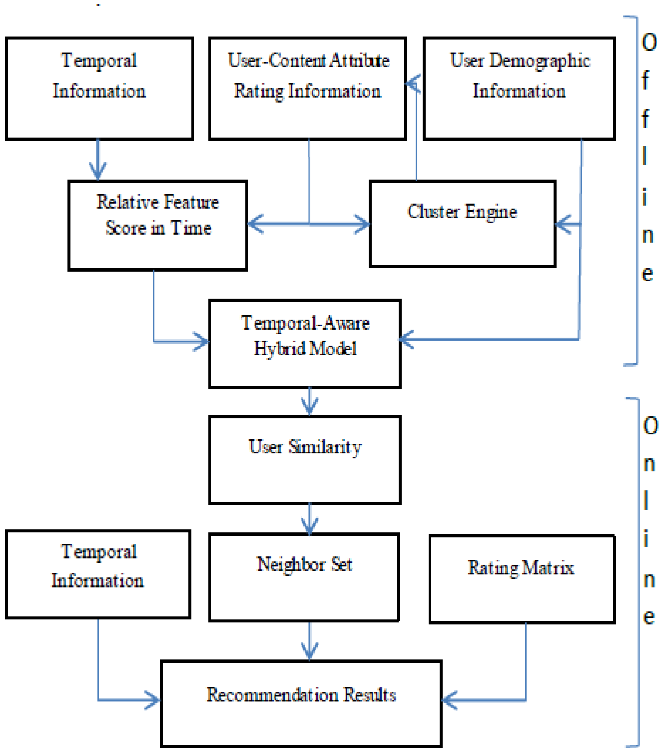

Fig.1 Workflow of Hybrid Recommender System with Temporal Information

III. TEMPORAL-AWARE HYBRID USER MODEL

In this paper, we have combined the model and memory based concepts to have better scalability and accuracy. The recommender system constructs the temporal-aware hybrid user model offline by using the Relative Feature Scoring Time in time t (RFST) and User Demographic Information, and does recommendation online by using user similarity to find the neighbor set, temporal information and rating matrix. If a user have not rate for item attribute, we apply clustering engine using user content attribute rating information and demographic information to dense the user content attribute rating information matrix. Actually, our proposed scheme makes recommendation by:

- Finding the proportion of likeness of content feature by aggregating all the rating for all the content attributes rated by the user in that specific time range.
- Then add demographic similarity to make context-aware hybrid model to be used to find the neighbors and to make the final recommendation.

A. *Related concepts and definitions*

Definition 1 Effective Total Rating: the sum of effective ratings R that user 'i' grades for all the item attributes in that time range, which is expressed as ETR (i,t).

$$\mathrm{ETR}\ (\mathrm{I},\mathrm{t}) = \sum_{\mathrm{j}\in \mathrm{Ti},\mathrm{R}\ge \frac{\mathrm{v}}{2}} \mathrm{R}(\mathrm{I},\mathrm{j},\mathrm{t}) \tag{1}$$

Here Ti stands for the set composed of items have been graded by user i. v is the maximum rating value by user i in time t.

Definition 2 Feature Rating: the sum of effective ratings that user i grades for the feature k in that time range, which is expressed as FR(i,k,t).

$$\mathrm{FR}\ (\mathrm{i},\mathrm{k},\mathrm{t}) = \sum_{\mathrm{k}\in \mathrm{Fk}\subset \mathrm{Ti},\mathrm{R}\ge \frac{\mathrm{v}}{2}} \mathrm{R}(\mathrm{i},\mathrm{k},\mathrm{t}) \tag{2}$$

Definition 3 Feature Frequency: the total occurrences of effective ratings that user i grades for Feature k in the time range, which is expressed as FF (i,k,t).

$$FF(i,k,t)=\sum_{j\in Fk\subset Ti} \delta p(Ri,j,t) \quad \mathrm{p}\in \left\{ x \,\middle|\, x \ge \frac{v}{2} \right\} \tag{3}$$

$$\delta p(i,j,t) = \begin{cases} 1, P = Ri,j,t \\ 0, P \neq Ri,j,t \end{cases}$$

Definition 4 Relative Feature Rating: relative feature rating that user i grades for feature k in the time range t, expressed as RFR(i,k,t), is defined as the ratio of Feature Rating FR(i,k,t) and Effective Total Rating ETR(i,t),

$$\text{RFR}\ (i,k,t) = \frac{\text{FR(i,k,t)}}{\text{ETR(i,t)}} \tag{4}$$

Definition 5 Relative Feature Frequency: relative feature frequency, is the relative number that feature k has been graded by user i in time t. expressed as RFF(i,k,t), is defined as the ratio of Feature Frequency FF(i,k,t) and ETF(i,t).

$$\text{RFF}\ (i,k,t) = \frac{\text{FF(i, k, t)}}{\text{ETF(i, t)}} \tag{5}$$

ETF(i,t) is the total occurrences of effective ratings graded by user i in time t.

B. Example

In this section we are going to explain the concept defined in section A, on some arbitrary data. We consider 30 movies Mi, graded by three users Ui, in four time ranges, Ti. We consider four attributes of each movie, which are shown in Table.1. The column identifier Ci denotes the content attribute category, Ui shows the rating that each user graded for the corresponding movie, and Ti represents the time range in which the user watched and graded the movie. We computed RFR and RFF separately with formulas defined in Definition 4 & 5 respectively, and the results are shown in Table.II.

Table. I Feature Rating by users at Time (Ti)

Movie	User Rating Information												Time
	User1				User2				User3				
	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4	Ti
M1	4	2	3	5	3	4	4	2	2	4	1	4	T1
M2	2	3	2	4	4	3	2	3	4	2	2	2	T1
M3	3	4	3	3	3	4	5	2	3	4	4	3	T1
M4	2	2	4	2	2	2	3	5	1	3	2	3	T1
M5	3	2	5	1	2	1	1	1	5	2	4	4	T1
M6	5	5	3	4	4	5	2	3	3	4	3	5	T1
M7	2	2	2	3	2	3	3	4	2	1	1	0	T1
M8	3	5	4	2	3	2	0	5	4	5	5	5	T1
M9	3	5	3	4	2	3	2	4	4	5	3	3	T1
M10	4	3	2	3	3	2	4	2	3	3	4	2	T2
M11	2	2	4	3	4	5	3	4	5	2	5	4	T2
M12	5	4	3	5	5	4	3	3	2	2	2	3	T2
M13	1	2	4	2	1	3	3	3	3	4	3	2	T2
M14	0	4	5	4	4	4	2	5	4	3	1	2	T2
M15	4	0	2	5	2	2	4	5	3	2	1	1	T2
…..	…..	…	..	.	…	..	.	..	.	.	.	.	.
M28	1	3	4	4	5	1	3	3	4	2	3	3	T4
M29	5	2	3	1	1	3	4	2	5	3	2	2	T4
M30	2	5	5	2	2	4	1	5	1	4	3	2	T4

Definition 6 Relative Feature Score in Time t: the Relative Feature Score in Time t can be obtained by using the values of RFF and RFR in the following formula;

$$RFST\,(\text{i, k, t}) = 10 * \frac{(\text{RFR(i,k,t)+RFF(i,k,t)})}{2} \tag{6}$$

Table II RELATIVE FEATURES RATE & FREQUENCY

FST	USER	Time1				Time2			
		C1	C2	C3	C4	C1	C2	C3	C4
RFR	U1	0.23	0.22	0.2	0.3	0.19	0.23	0.27	0.2
	U2	0.23	0.26	0.2	0.2	0.21	0.25	0.22	0.3
	U3	0.2	0.24	0.2	0.2	0.33	0.22	0.28	0.1
RFF	U1	0.2	0.2	0.3	0.2	0.19	0.23	0.28	0.2
	U2	0.2	0.2	0.2	0.2	0.2	0.25	0.25	0.3
	U3	0.25	0.25	0.2	0.3	0.35	0.23	0.23	0.1

The values of RFST are calculated using Eq. 6 and its results are shown in Table. III.

Table III RFST IN RESPECTIVE TIME RANGES

User	RFST in Time 1				RFST in Time 2			
	C1	C2	C3	C4	C1	C2	C3	C4
U1	2.43	2.11	2.94	2.5	1.94	2.36	2.78	2.91
U2	2.51	2.65	2.10	2.72	2.05	2.5	2.36	3.07
U3	2.42	2.48	2.11	2.98	3.43	2.31	2.61	1.63

C. Temporal-Aware Hybrid Model

The Temporal-Aware Hybrid model is a three dimensional matrix contains relative feature scoring, demographic information and time ranges for all users as shown in Table. IV. In this model item features are obtained from RFST and user features are extracted from the demographic information.

Table IV TEMPORAL-AWARE HYBRID USER MODEL

User	Features					
	Item Features/Attributes			User Demographic Features		
	F1	….	Fp	D1	…..	Dq
U1	H11t	…	H1pt	H1(p+1)t	…	H1(p+q)t
…	…	…	…	…	…	…
Um	Hm1t	…	Hmpt	Hm(p+1)t	…	Hm(p+q)t

Where m, p and q are the number of users, RFST and user demographic features respectively. H11t represents the relative score value for user 1 for item feature 1 in time ti, where i=1, 2,3,………

The Temporal-Aware Hybrid Model has the following advantages.

1. The model is constructed offline to make the recommendation system to recommend item for a user within user bearable time.
2. In calculating each entry in this model, it considers only effective user rating in recommendation for a specific time range, so minimize the computational time.
3. The size of the Temporal Hybrid Model is much smaller as compared to the original rating matrix, so finding similar user requires less time.
4. This model solves scalability problem, it considers the content that user have rated in that time range.

III. COLLABORATIVE FILTERING RECOMMENDATION

After using the content classification and demographic information to pre-produce prediction where necessary, we have dense three-dimensional matrix of users, content attributes and time domain. We calculate the RFST and made temporal-aware hybrid model. In this section, firstly we identify the neighbors of the target user in time t, and then use neighbors set in User-Based Collaborative Filtering Recommendation algorithm to produce recommendation.

Measuring the Neighbor Set: finding the nearest neighbor is to compute the similarity of users. There are several similarity algorithms that have been used: Pearson Correlation, cosine vector similarity, adjusted cosine vector similarity, relevant similarity, mean-squared difference and Spearman correlation. Now on the basis of temporal-aware hybrid user model, we can calculate the similarity of feature vector of user 'a' and user 'b' in that specific time range. We have used relevant similarity for computing the similarity between user "a" and "b" at time t, the formula is as follows:

$$sim(a,b,t) = \frac{\sum_{k=1}^{p+q} |(Hakt - \overline{Hat})\ X\ (Hbkt - \overline{Hbt})|}{\sqrt{\sum_{k=1}^{p+q} (Hakt - \overline{Hat})^2 X \sum_{k=1}^{p+q} (Hbkt - \overline{Hbt})^2}} \qquad (7)$$

Where Hakt and Hbkt is the relative interest score for feature k in time t for user a and b respectively. $\overline{Hat}$ and $\overline{Hbt}$ is the average rating of User a and b in time t respectively.

User Based Collaborative filtering to produce results: After obtaining the similarity between target user and the neighbors at a specific time t, the rating can be predicted that user grades for item:

$$P(a,j,t) = \left\{ R_{at} + \frac{\sum_{i \in U} Sim(a,i,t)\ X\ (Rijt - \overline{Rit})}{\sum_{i \in U} sim(a,i,t)} \right\} \qquad (8)$$

Where:

P(a,j,t) is the predicted rating that user a grades for item j at time t, and U is the nearest neighbor set of user a calculated at that time. Rijt is rating of neighbor j for the content i in that time range. Rat is the rating to user a in time t.

The Top–N items with the highest predicted rating at that time will the final recommendation.

IV. EXPERIMENT AND RESULTS

We could not find such a database that has time information with rating information for movies. So we take the movies rating information from Movie Lens work group (www.grouplens.org) and add the time ranges. We consider 24 times ranges, 200 users and 1000 movies with 4 features in each time range. We also include the user demographic information having 4 features.

Many statistical measures can be used to measure the accuracy of recommendation. The most common method used for accuracy measurement is the Mean Square Error (MAE). To test the effectiveness of our Algorithm we consider the MAE and Recommendation time that a Recommender system takes for recommending an item, we consider only the online process processing time. In MAE we consider three different algorithms and for recommendation time we consider two algorithms. The results of MAE are shown in Figure.2 and Recommendation time is shown in Figure.3.

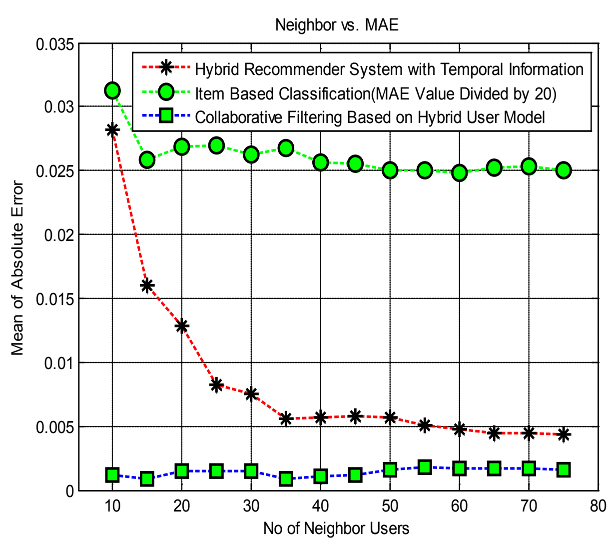

Fig. II RESULTS FOR MEAN OF ABSOLUTE ERROR

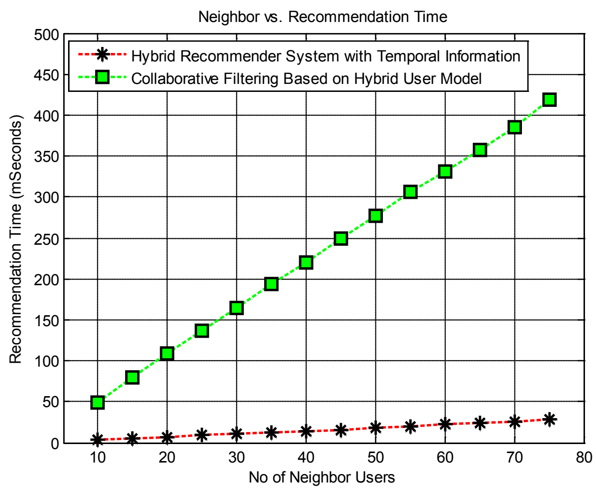

Fig. III EXPERIMENT RESULT FOR RECOMMENDATION TIME

V. CONCLUSION AND FUTURE WORK

Traditional Collaborative filtering recommender systems don't use the temporal information. The algorithm proposed in the paper is basically for recommendation for user at current time, but, in general can be placed instead of traditional approaches. It solves the problems like scalability, sparsity and cold start issue. Our algorithm account the relative feature score in specific time to make the final recommendation quickly and accurately. It does recommendation based on user preference at that time, which is highly desired. In future, we are working on the architecture, adding user geographic information, improving time stamping and implementation of the recommendation system for N-Screen service.

REFRENCES

[1] "PuWANG", "HongWu YE" A Personalized Recommendation Algorithm combining Slope One Scheme and User Based Collaborative Filtering, 2009 International Conference on Industrial and Information Systems

[2] "HengSong Tan", "HongWu Ye" A Collaborative Filtering Algorithm Based on Item Classification, 2009 Pacific-Asia Conference on Circuits and Communications and System

[3] "Xiaohui Li", "Tomohiro Murata", Customizing Knowledge-based Recommender System by Tracking Analysis of User Behavior, 2010 International conference.

[4] "K. Palanivel", "R. Sivakumar", Fuzzy Multicriteria Decision–making Approach for Collaborative Recommender Systems, International Journal of Computer Theory and Engineering, Vol. 2, No. 1 February, 2010.

[5] "Qain Wang", "Xianhu Yuan", "Min Sun", Collaborative Filtering Recommendation Algorithm based on Hybrid User Model, 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery.

[6] Shardanand, U., & Maes, P. (1995). Social information filtering: Algorithms for automating 'Word of Mouth'. In proceedings of the conference on human factors in computing systems,

[7] Shahabi, C., Banaei-Kashani, F. Chen, Y., & Mcleod, D. (2001). Yoda: accurate and scalable web-based recommendation systems, In the preceeding of the sixth international conference on cooperative information systems (coopIS 2001), Trento Italy.

[8] "Mastansar Ali Ghazanfar" and "Adam Prugel-Bennett", United Kingdom: A Scalable, Accurate Hybrid Recommender System, 2010 Third International Conference on Knowledge Discovery and Data Mining

[9] "Shunichi SEKO", "Manabu Motcgi", "Takashi Yagi", "Shinyo Muto", Video Content Recommendation for Group Based on Viewing History and Viewer Preferences, 2011 IEEE International conference on Consumer Electronics.

[10] "YaE Dai", "Hong Wu Ye", Personalized Recommendation Algorithm using User Demographic Information, Second International Workshop on Knowledge Discovery and Data Mining.