EAI Endorsed Transactions
on Context-aware Systems and Applications Research Article **EAI.EU**

Item-based recommendation with Shapley value

Tri Minh Huynh [1*] and Tai Huu Pham [3] and Vu The Tran [2] and Hiep Xuan Huynh [3]

[1] Kien Giang University, Viet Nam
[2] University of Science and Technology, Da Nang University, Viet Nam
[3] Can Tho University, Viet Nam

Abstract

Discovering knowledge in archival data is the goal of researchers. One of them is collaborative filtering recommender system is developing fastly today. It may be rather effective in sparse and "long tail" datasets. Calculating to make decision based on many criteria is really necessary. Relationships, interactions between criteria need to have been fully considered, decision will be more reliable and feasible. In this paper, we propose a new approach that builds a recommender decision-making model based on importance of item, set of items with Shapley value. This model also incorporates traditional techniques and some our new approaches and was tested, evaluated on multirecsys tool we develope from some available tools and uses standardized datasets to experiment. Experimental results show that the proposed model is always satisfactory and reliable. They can be applied in appropriate contexts to minimize limitations of recommender system today and is a research way next time.

Keywords: Collaborative Filtering (CF) Recommender System (RS), Multi-Criteria (MC), Interaction, Decision-Making (DM), importance, Shapley.

Received on 18 April 2019, accepted on 15 June 2019, published 26 June 2019

doi: 10.4108/eai.18-3-2019.159341.

*Corresponding author: Hmtri@vnkgu.edu.vn

1. Introduction

Nowadays, recommender system (RS) [1] [2] [4] [5] is more and more important in many areas of life. It responds quite well to needs of users about finding information in many forms and diversity variations. Traditional RS often base on historical factors, user preferences to recommend, it is very popular and always used in the past time. Today, the trend is based on the popularity, diversity of data in the system to recommend and be got attention more. That means there are many criteria in the system that are considered to select the desired information. Therefore, multi-criteria collaborative filtering recommender system will service well for this things and be the target to execute because it is very good effective current .

At present, there are many research about decision-making models for multi-criteria recommender system [3][5][6][17][18] is mainly based on collaborative filtering (CF on many criteria because if they only base on one criteria is too phantom to decide a problem, the result may be much misleading. Decision-making for RS is very important to satisfy the user requirements. So the decision-making model based on the criteria will be the good solution chosen by the researchers today. For the multi-criteria collaborative filtering recommender system, the objective is revieing information on the criteria and the number of criteria. Recommendation decisions depend on this information. There are many solutions proposed, but the importance thing for decision making is applicating appropriate operations to give the best results.

Most of the current consulting models do not think much about the relationships and the interaction between the criteria for making decisions. This may make the results of the consultancy unsustainable and don't give yet fully capable of the stored data.

In this paper, we present a solution for decision making in multi-criteria collaborative filtering recommender system based on importance of items in order to have the most appropriate result and in accordance with the requirements and characteristics of storage data. Usually in the intrinsic self of data there is a relationship, influence, reflection on each other, so the model needs to fully consider the interaction of the values of the criteria that make the recommendation model becomes more effective. The proposed model is built on multi-criteria collaborative filtering recommender system. The model was experimented on the multirecsys tool developed by us in the R language. The experiment data is standard datasets: the MovieLense, MSWeb and Jester5k. We maked two experiments to test and evaluate the model. The results of the model show that it is quite effective, can exploit information well on some data systems today.

The article is designed into five parts. The first part introduces an overview of the multi-criteria collaborative filtering recommender system, some current approaches. The second part presents about important degree of item. The third part presents about the designing of the multi-criteria collaborative filtering recommender system model and introduces decision making with Shapley operator for recommendation. The fourth part presents some experiment and evaluation of the model on the multirecsys tool. The last part is the conclusions.

2. Important degree of item, a set of items

2.1. Multi-criteria matrix

The matrix A (m × n) consists of m rows $u_1, u_2, \dots, u_m$ and n columns $i_1, i_2, \dots, i_n$. Each row of u_p $(p: 1..m)$ with each column i_q $(q: 1..n)$ determines the value r_{pq} as Table 1. Each row is a criterion. In contrast, for $\hat{r}(u, i)$, each $\hat{r}_q$ value is determined based on the set $R_q = \{r_{1q}, r_{2q}, \dots, r_{mq}\}$ $(q: 1..n)$ where r_{pq} is the value corresponding to u_p $(p: 1..m)$ and i_q.

Table 1. Multi-criteria matrix: rows and columns

rows/cols	i_1	i_2	…	i_n
u_1	1	3	…	5
u_2	2	2	…	3
…	…	…	…	…
u_m	4	1	…	2

2.2. Capacity of items

With set of items $I = \{i_1, i_2, \dots, i_n\}$. A capacity function μ [7-11] on I is function $\mu: \wp(I) \rightarrow [0,1]$, with $\mu(\emptyset) = 0, \mu(I) = 1$. A capacity function can be defined according to a principle or according to the characteristics or goals of the system when data is update in the system.

$$A \subseteq B \Longrightarrow \mu(A) \leq \mu(B), \qquad A, B \subseteq I$$

On I, define a vecto P with weights, $A \subseteq I$, a capacity function can be defined as follow:

$$\mu(A) = \sum_{a \in A} P(a), \sum_{i \in I} P(i) = 1,\ i \in I$$

The value of $\mu(A)$ can be changed depending on the criteria in A. With C_1 and C_2 is two criteria in A. The value of $\mu(C_1, C_2)$ can get the value as follow:

or $\mu(C_1, C_2) = \mu(C_1) + \mu(C_2)$

or $\mu(C_1, C_2) > \mu(C_1) + \mu(C_2)$

or $\mu(C_1, C_2) < \mu(C_1) + \mu(C_2)$

Example: A set has three criteria: Mathematics, Physics, Literature and the values of the capacity of each criterion is: $\mu(\{\text{Mathematics}\}) = 0.35$, $\mu(\{\text{Physics}\}) = 0.30$, $\mu(\{\text{Literature}\}) = 0.40$. The value of the capacity function of the criteria subset can be given as follow:

$\mu(\{\text{Mathematics, Physics}\}) = 0.55,$

$\mu(\{\text{Mathematics, Literature}\}) = 0.85$

$\mu(\{\text{Literature, Physics}\}) = 0.75$

$\mu(\{\text{Mathematics, Physics, Literature}\}) = 1$

2.3 Interaction and importance degree of item, importance degree of a set of items

As we presented above, the value of $\mu(C_1, C_2)$ and $\mu(C_1) + \mu(C_2)$ can be different. This shows that there is interaction between C_1 *and* C_2 when they come together. We call interaction degree $I(C_1, C_2)$ between C_1 *and* C_2 is a value in [-1,1]:

$$I(C_1, C_2) = \mu(C_1, C_2) - (\mu(C_1) + \mu(C_2))$$

or $\mu(C_1, C_2) = \mu(C_1) + \mu(C_2) + I(C_1, C_2)$

If two criteria: C_1, C_2 in a larger set $A \cup \{C_1, C_2\}$ [8]:

$$I(C_1, C_2) = \sum_{A \subseteq I \setminus \{C_1, C_2\}} \frac{(n-|A|-2)|A|!}{n!} [\mu(A \cup \{C_1, C_2\}) - \big(\mu(A \cup \{C_1\}) + \mu(A \cup \{C_2\})\big) + \mu(A)]$$

With a capacity function μ, the Shapley value [9 10 11] based on μ of $i_q \in I$ is defined by $\varphi_{i_q}(\mu)$:

$$\varphi_{i_q}(\mu) = \sum_{A \subseteq I \setminus \{i_q\}} \frac{(n-|A|-1)!|A|!}{n!} \big(\mu(A \cup \{i_q\}) - \mu(A)\big)$$
(1)

Here, we determine importance degree of item i_q depend on the value $\varphi_{i_q}(\mu)$. This value shows importance degree of item i_q in the criteria set which has i_q in that. We call item important degree i_q to be $\varphi_{i_q}(\mu)$ and in this model. Thereby, Shapley values also show the interaction between items together in the operation. When calculating the Shapley value of an item, it must be affected by other items.

In addition, this model, we call important degree of a set of items S is calculated by formula:

$$\varphi_{S\subseteq I} = \sum_{i_q \in S} \varphi_{i_q}(\mu) \qquad (2)$$

Example: $I_n = \{i_1, i_2, i_3\}$, n=3, to determine $\varphi_{i_1}(\mu)$, we need to depend on the value of: $\mu(\{i_1, i_2, i_3\}), \mu(\{i_2, i_3\})$ are given $\mu(\{i_1, i_2, i_3\}) = 1$, $\mu(\{i_2, i_3\}) = 0.55$

$$\varphi_{i_1}(\mu) = \frac{(n-|\{i_2,i_3\}|-1)!|\{i_2,i_3\}|!}{n!}(\mu(\{i_1, i_2, i_3\}) - \mu(\{i_2, i_3\}))$$

$$\varphi_{i_1}(\mu) = \frac{0!2!}{3!}(1 - 0.7) = 0.15$$

3. Multicriteria decision making with important degree of item

3.1. Rating matrix

Data applies to the model is as a table of values. It represents user's ratings for items. The value which item are not rated will be "?". Here, u_a is a consulted user. We need to determine the value of $\hat{r}$ function which give the result of the recommender system. That is the list of selected products.

Table 2. Data model with Rating matrix

	i_1	i_2	…	i_x	…	i_y	…	i_n
u_1	?	1	…	1	…	?	…	3
u_2	5	4	…	4	…	5	…	2
…	…	…	…	…	...	…	…	…
u_m	4	2	…	4	…		…	?
u_a	?	2	…	?	…	?	...	3
$\hat{r}$	?	-	…	?	…	?	…	-

3.2. Similarity

The model selects items based on collaborative filtering model with k nearest neighbors (kNN) [12]. In Table 2, kNN items are nearest neighbors to item i_q based on the similarity (or distance) between i_q (q:1..n) and each item in the system according by measures: cosine, pearson... Each item has weight separately. The Pearson measure [13] between two users are i_x and i_y $(x, y : 1..n)$ is defined:

$$\text{sim}(i_x, i_y) = \frac{\sum_{i\in I_{i_x,i_y}}(r_{i_x i} - \bar{r}_{i_x})(r_{i_y i} - \bar{r}_{i_y})}{\sqrt{\sum_{i\in I_{i_x,i_y}}(r_{i_x i} - \bar{r}_{i_x})^2}\sqrt{\sum_{i\in I_{i_x,i_y}}(r_{i_y i} - \bar{r}_{i_y})^2}} \qquad (3)$$

I_{i_x,i_y} is the set of data items evaluated by i_x, i_x; $\bar{r}_{i_x}$ is the average rating evaluation of i_x on all data items, $\bar{r}_{i_y}$ is the average rating evaluation of i_y on all data items. Then, the distance between two users is (1-r).

3.3. Determining item important degree

First, determine the value of the capacity function of each items. After finding the similar values set of each items S (sim), we determine the value of the capacity function $\mu(i_{qt}), k = kNN$, it is puted as follow:

$$\mu(i_{qt}) = w_{i_{qt}} = \frac{sim(i_t, i_q)}{\sum_{v=1}^{k} sim(i_v, i_q)}, \text{ t from 1 to k.} \qquad (4)$$

Next, determine $\varphi_{i_q}(\mu)$, the important degree of item i_q, we do two steps as below:

(a) Determining the value of the capacity function of a set $\mu(A)$, A is the subset in $S_{qk}, A \subseteq S_{qk}$, S_{qk} is the set k nearest neighbors items of i_q. We begin determining the value of the capacity function of the two items subset, the three items subset,... in A. We define a capacity function of a set of items as follow:

$$\mu(S_{qk}) = 1$$

$$\mu(A) = \sum_{B\in A} \mu(B), B \text{ is subsets in } A \qquad (5)$$

$\mu(A) = 1$ *if* $\mu(A) > 1$ and B has from two items or more.

First, we need determinate the two items set. To two items i and j, we put:

$$\mu(i,j) = \mu(i) + \mu(j) + w'_i + w'_j \qquad (6)$$

$$\mu(i,j) = 1 \text{ if } \mu(i,j) > 1$$

We call that $w'_i + w'_j$ is the interaction value between i and j in set of $\mu(i,j)$.

To be simple. we put $\mu(k) = w_{ik}[k]$. With w_{ik} is calculted by formula:

$$w_{ik} = \frac{w_i}{sum(w_i)} \text{ and } w_i = w[i, 1..k], \text{ k=kNN} \qquad (7)$$

We put $w = sim(n\ items)$ is the similarity values set of all items in system. This set is calculted by formula 2.

Next, we create a weights set w' from Rating matrix to support the interesting of each item. The formula as above:

$$w' = \frac{count(rating(i)\ in\ [4..5]) - count(rating(i)\ in\ [1..3])}{n} \qquad (8)$$

With n is a number of items in the system. We only get values in w' for items in the kNN set of item is ordering (i).

(b) Determining $\varphi_{i_q}(\mu)$: we depend on the fomulas: (5), (6) and (7).

$$\varphi_{i_{qt}}(\mu) = \sum_{A\subseteq S_{qk}\setminus\{i_{qt}\}} \frac{(k-|A|-1)!\,|A|!}{k!}\left(\mu(A\cup\{i_{qt}\}) - \mu(A)\right)$$

$$with\ t\ in\ [1:k] \qquad (9)$$

Example: There three items: i_1, i_2, i_3 and the similarity values between item i with these items and weights of item i is w' in the table above:

Table 3. Item important degree $\varphi_{i_{qt}}(\mu)$, $\varphi_{S\subseteq I}$

	i (w_i)	w'	$\varphi(\mu)$

i_1	0.23	0.17	0.21
i_2	0.12	-0.31	0.05
i_3	0.28	0.18	0.26
$\varphi_{\{i_1 i_2 i_3\}}(\mu)$			**0.52**

$$\varphi_{i_1}(\mu) = \frac{(3-2-1)!\,2!}{3!}\left(\mu(\{i_1,i_2,i_3\}) - \mu(\{i_2,i_3\})\right)$$

$$\mu(\{i_2,i_3\}) = \mu(\{i_2\}) + \mu(\{i_3\}) + w'_{i_2} + w'_{i_3}$$

Put $\mu(\{i\}) = w_i$ and w' is the interaction value of item join in items set.

$\mu(\{i_1\}) = 0.23,\ \ \mu(\{i_2\}) = 0.12,\ \ \mu(\{i_3\}) = 0.28$
$\mu(\{i_1,i_2\}) = 0.21$
$\mu(\{i_2,i_3\}) = 0.38$
$\mu(\{i_1,i_3\}) = 0.86$
$\mu(\{i_1,i_2,i_3\}) = \mu(\{i_1,i_2\}) + \mu(\{i_1,i_3\}) + \mu(\{i_2,i_3\})$
$\mu(\{i_1,i_2,i_3\}) = 1.45 \rightarrow \mu(\{i_1,i_2,i_3\}) = 1$
$\varphi_{i_1}(\mu) = 0.21, \varphi_{i_2}(\mu) = 0.5, \varphi_{i_3}(\mu) = 0.26$
and this model, we calculate important degree of a set of items is $\varphi_{\{i_1 i_2 i_3\}}(\mu) = 0.52$

3.4. Recommendation model

From Table 2, the model defines a table of similarities between items each orther. Next, the model determine the importance degree of each item based on items that are similarity with it and related factors as described above. This is the main issue for decision making in this paper. The model is shown below. The values of w is deminate by the fomulas: (3), (4), (7). The values of w' is deminate by the fomula (8).

Table 4. Proposed model

	i_1	i_2	i_3	i_4	i_5	i_6	w	w'
i_1	-	**0.5**	0.1	**0.8**	0.3	**0.4**	0.34	0.14
i_2	**0.5**	-	**0.3**	**0.4**	0.2	0.2	0.15	-0.02
i_3	0.1	**0.3**	-	0.1	**0.3**	**0.4**	0.18	-0.62
i_4	**0.8**	**0.4**	0.1	-	**0.4**	0.4	0.25	0.29
i_5	0.3	0.2	**0.3**	**0.4**	-	**0.5**	0.35	-0.13
i_6	**0.4**	0.2	**0.4**	0.4	**0.5**	-	0.25	0.47
u_a	**?**	5	**?**	**?**	4	**?**		
$\hat{r}\ (\varphi_{S\subseteq I})$	0.37	-	**0.77**	0.32	-	**0.62**		

Indentify results of RS

On the basis of development from traditional recommender models [4][5][15][16], first, we determine the similarity between the product i_q and each product in data, calculate weights w, w' as above, the results are as Table 4. Next, we calculate the values $\hat{r}_q$ at u_{aq} # "?". In this model, we put $\hat{r}_q = \varphi_{S\subseteq I}$, with S is the kNN items set. $\varphi_{S\subseteq I}$ is calculated by formula (2). At each $i_q, q: 1..n$, we take the similarity values of kNN (k highest values) are k nearest neighbors of i_q to calculate $\varphi_{i_{qt}}(\mu)$ by formula (9). After defining $\hat{r}_q$ values, rank these values in descending order, selecting the products corresponding to the high to low values to suggest to user. Suppose we choose two products to introduce to the user u_a: i_3 and i_6; $\hat{r}_3 = 0.77$ *and* $\hat{r}_6 = 0.62$. These are the two products with the highest $\hat{\boldsymbol{r}}$ value. With $\hat{\boldsymbol{r}}$ has values r_q is determined by (formula 6).

Evaluation recommendations

We evaluate recommendation model by the Receiver Operating Characteristic method (ROC) [5][14]. The method was developed for signal detection and goes back to the Swets model. The ROC-curve is a plot of the system's probability of detection (also called sensitivity or true positive rate TPR) by the probability of false alarm (also called false positive rate FPR). Evaluation for two systems can compare the size of the area under the ROC-curve, where a bigger area indicates better performance. The values need deminate: True Positives (TP), False Negatives (FN), True Negatives (TN), true Positive Rate (TPR): TPR=TP/(TP + FN), false Positive Rate (FPR): FPR= FP/(FP + TN). Deminate values and display the ROC curve, Recision/Recall to evaluate the effectiveness of models.

4. Experiment

4.1. Datasets used for experiments

The dataset used for experimentation on the proposed model is the MovieLens100K (sparse data) movie is available at http://grouplens.org/datasets/. The Movielens archive of 100,000 reviews performed by 943 users on a total of 1,682 films, each rated at least 20 movies and rated from 1 (bad) to 5 (good). The MSWeb was generated by sampling and processing the logs of www.microsoft.com in one week timeframe, episode stores information about the 98.653 rating made by 32.710 users on the number of 285 website (Vroot) with value of TRUE/1 (binary data). We also experimented on the Jester5k joke book (data is too thick) at address above, episode stores information about the 500,000 rating made by 5,000 users on the number of 100 jokes, with values from -10 to 10. Each user evaluates at least 36 jokes

4.2. Experimental tools

The model was experimented by multirecsys tool which we built, developed and installed applications on R [www.r-project.org]. We developed this tool based on the RecommenderLab package that researchers also are developing on it. Besides, we also inherited a number of open source tools of the world community that have built and developed. On the RecommenderLab package, we can display the results, calculate the time, evaluate the error and the effectiveness of the model.

4.3. Scenario 1: Experiment the model and compare it with some existing model

We tested the proposed model (IBCF_Shapley) on two datasets: Movielens100K (too sparse), MSWeb (too sparse and binary data) and Jester5k (too thick), and also on the three these datasets, we compared the results of the counseling with the existing models (IBCF, Random with item-based). Experimental results with kNN=10 gives 5 films on the Movielens100K, 5 websites on the MSWeb and 5 joke books on Jester5k showed that the results have some diffirence values and the results of the proposed model seem no change and change only when the data is updated new, while the result of the other models may change when new variables are restablished in information processing for new every test. The result of three model as follow:

Table 5. Five movies in MovieLense are consulted on three models

IBCF_Shapley	IBCF	RANDOM
[1,] "GoldenEye (1995)" [2,] "Four Rooms (1995)" [3,] "Get Shorty (1995)" [4,] "Copycat (1995)" [5,] "Shanghai Triad (Yao a yao yao dao waipo qiao) (1995)"	[1,] "Boys Life (1995)" [2,] "Ballad of Narayama, The (1958)" [3,] "No Escape (1994)" [4,] "Turning, The (1992) [5,] "Celestial Clockwork (1994)"	[1,] "Braveheart (1995)" [2,] "Free Willy 2: The A dventure Home (1995)" [3,] "Mad Love (1995)" [4,] "Clerks (1994)" [5,] "Crow, The (1994)"

Table 6. Five comics in MSWeb are consulted on three models

IBCF_Shapley	IBCF	RANDOM
[1,] Knowledge Base [2,] Microsoft.com Search [3,] Norway [4,] Misc [5,] International IE content	[1,] Office Free Stuff [2,] Knowledge Base [3,] isapi [4,] MS Office Info [5,] NT Server Support	[1,] South Africa [2,] Softlib [3,] Turkey [4,] Internet Service Providers [5,] Works Support

Table 7. The values of precision/recall of three models on MovieLense

IBCF_Shapley		IBCF		RANDOM	
precision	recall	precision	recall	precision	recall
0.13684	0.01678	0.00699	0.00011	0.03157	0.00215
0.04000	0.01822	0.00139	0.00011	0.04842	0.01326
0.04736	0.03663	0.00069	0.00011	0.02555	0.02555
0.05473	0.05783	0.00093	0.00028	0.03859	0.03542
0.04368	0.06631	0.00104	0.00045	0.03473	0.03871
0.04366	0.07794	0.00089	0.00045	0.04478	0.04478

Table 8. The values of precision/recall of three models on MSWeb

IBCF_Shapley		IBCF		RANDOM	
precision	recall	precision	recall	precision	recall
0.01819	0.00832	0.25279	0.20313	0.00749	0.00309
0.17052	0.17052	0.11366	0.37258	0.00724	0.01804
0.37324	0.37324	0.06832	0.42364	0.00704	0.03569
0.38419	0.38419	0.05217	0.45574	0.00681	0.05186
0.49785	0.49785	0.04345	0.46819	0.00693	0.06946
0.04066	0.58108	0.03871	0.48155	0.00703	0.09438

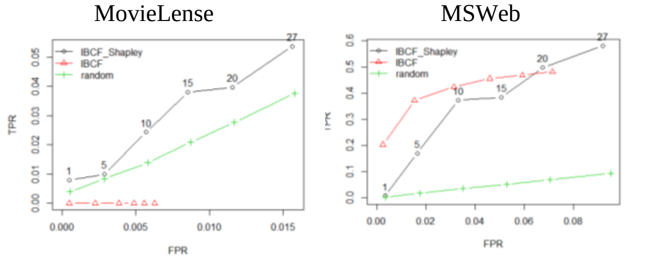

Figure 1. ROC curve of three models on MovieLense and MSWeb

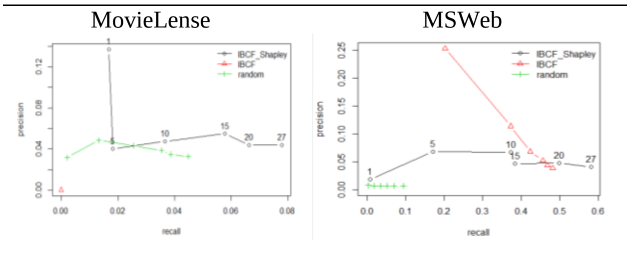

Figure 2. ROC curve of three models on MovieLense and MSWeb

The experiment result of three models on Jerter5k (too thick datasets): We do some experiments of models on Jester5k. This is the too thick dataset. The result of the proposed model is not so good because our solution in this model has a comparison of the number of user's ratings on each items, so there is no meaning for too thick datasets.

Table 9. Five joke books in Jester5k are consulted on three models

IBCF_Shapley	IBCF	RANDOM
"j71"	"j85"	"j71"
"j72"	"j86"	"j76"
"j73"	"j71"	"j80"
"j74"	"j81"	"j83"
"j75"	"j84"	"j84"

Table 10. The values of precision/recall of three models

IBCF_Shapley		IBCF		RANDOM	
precision	recall	precision	recall	precision	recall
0.17200	0.00848	0.24649	0.01547	0.17500	0.00892
0.15240	0.04213	0.26372	0.08663	0.05000	0.05000
0.16580	0.09692	0.27160	0.17424	0.17920	0.09772
0.17040	0.15407	0.25654	0.24308	0.17866	0.15271
0.16390	0.19616	0.24079	0.29387	0.18090	0.20203
0.17981	0.28565	0.22985	0.33377	0.17911	0.27030

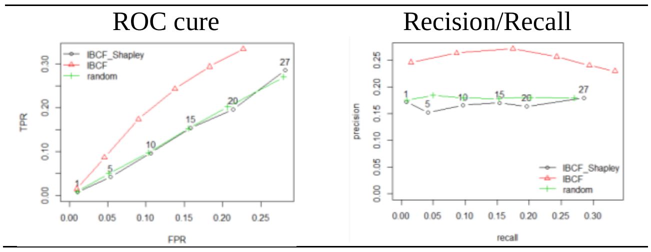

Figure 3. ROC curve and Recision/Recall of three models on Jester5k

Based on the ROC Curve and precision/recall of models, they have showed that IBCF_Shapley is given the pretty good result on all datasets. Proposed model always seem has hight effective on sparse datasets more than thick datasets with item-based collaborative filtering, especially with sparse and non-binary datasets. We can fully belive it is applied to the current recommendation system.

4.5. Scenario 2: Experiment to evaluate the model on two datasets: Movielens100K and MSWeb with some different kNN values (test the model with a number of criteria increase)

4.5.1. Experiment to evaluate the model on two datasets with k=25

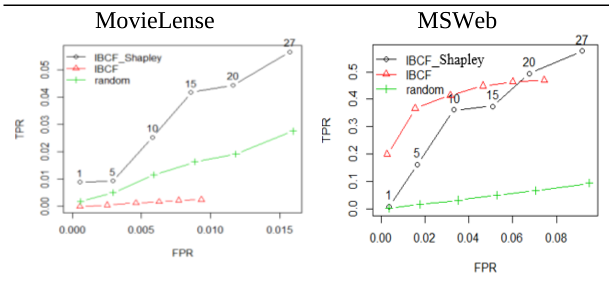

Figure 4. ROC curve of three models with k=25

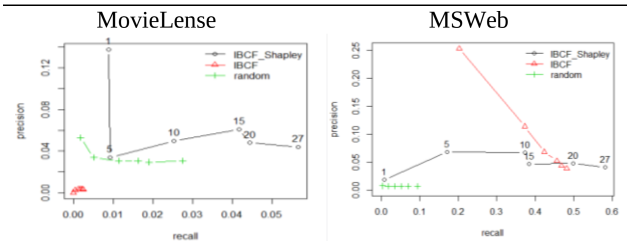

Figure 5. Recision/Recall of three models with k=25

4.5.2. Experiment to evaluate the model on two datasets with k=35

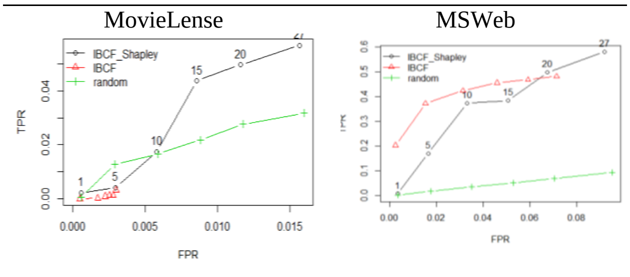

Figure 6. ROC curve of three models

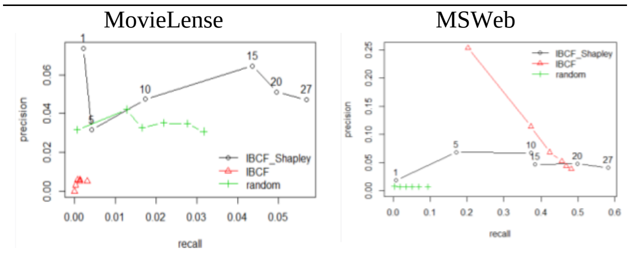

Figure 7. Recision/Recall of three models

4.5.3. Experiment to evaluate the model on two datasets with k=45

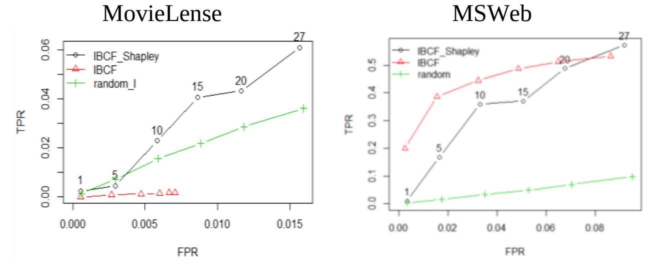

Figure 8. ROC curve of three models

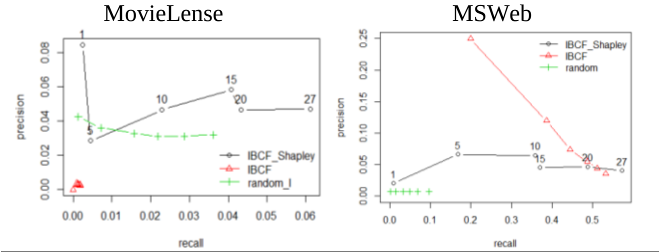

Figure 9. Recision/Recall of three models

Experimental results with many difference values of k, also show that IBCF_Shapley model always give quite good results and quite effective on all sparse datasets and give the best result on sparse and non-binary datasets.

5. Conclusions

Any recommender model can give a good results if it is placed in the appropriate context and characteristics of the archived data. Our proposal model, item-based collaborative filtering multi-criteria recommender system with Shapley operator was built based on interaction, ability and importance of the criterions in the system. This helps to give the consultant decision to support well the requirements of the counsed user. The model is developed on the basis of traditional consulting systems and exploits tools and datasets on the RecommenderLab package. We set the formulas to calculate capacity fonction. Since then, the value of Shapley is calculated to serve as a consulting decision. We do two main experiments to evaluate the proposed model. The results show that the proposed model satisfies quite well the requirement.

This model shows the coherence, interactions of the criteria, improvement of the results with discrete information, lack of information and mutation of data. The paper provides a method of counseling with the weighting of criteria and get relationship values for decision making. The proposed model can be applied on many datasets and the results will be reliable, especially on sparse dataset and non-binary. Although the execution time of program is still long as lost time to make weighted and interaction values, calculate Shapley value and ordered rankings, but the results are more responsive. In the coming time, we will continue to research and improve the algorithm more to shorten the time of consulting to promote better model.

Acknowledgements.

This model was built and developed on the basis of past studies, inheriting innovative results and tools to implement the model. This helps our new proposal be made easy, saving a lot of effort and time. We always recognize this and will join together with the scientific community to develop research in the future.

References

1. Michael D. Ekstrand, John T. Riedl and Joseph A. Konstan, *Collaborative Filtering Recommender Systems*, 2011, Human–Computer Interaction, Vol. 4, No. 2 (2010) 81–173DOI: 10.1561/1100000009
2. Gediminas Adomavicius, Alexander Tuzhilin, *Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions*, IEEE Transactions On Knowledge And Data Engineering, Vol. 17, No. 6, June 2005
3. Tri Huynh Minh, Vu The Tran, Hung Huu Huynh, Hiep Xuan Huynh, *Decision making operations: arithmetic mean, geometric mean, harmonic mean in user-based collaborative filtering of multi-criteria recommender systems*, The 6th national scientific conference on information technology and applications in various fields, Da Nang, Viet Nam, 2017.
4. Badrul Sarwar, George Karypis, *Joseph Konstan, and John Riedl, Item-based Collaborative Filtering Recommendation Algorithms,* Appears in WWW10, 2001, Hong Kong.
5. Michael Hahsler, *Recommenderlab: A Framework for Developing and Testing Recommendation Algorithms*, R package version 0.2-2, 2017, http://lyle.smu.edu/IDA/recommenderlab
6. Tri Huynh Minh, Vu The Tran, Hung Huu Huynh, Hiep Xuan Huynh, *Collaborative filtering recommender system base on the interaction multi-criteria decision with ordered weighted averaging operator*, Proceedings of 2018 The 2nd International Conference on Machine Learning and Soft Computing ICMLSC 2018, Vietnam February 2-4, 2018 ACM ISBN: 978-1-4503-6336-5, Pp: 57-61.
7. Eiichiro Takahagi, *Lamda Fuzzy Measure Identification Methods using Lamda and Weights*, Visiting Fellow, University of Bristol, UK - Until August 2005.
8. Anath Rau Krishnana, Maznah Mat Kasimb, Engku Muhammad Nazri Engku Abu Bakar, *A Short Survey on the Usage of Choquet Integral and Its Associated Fuzzy Measure in Multiple Attribute Analysis*, International Conference on Computer Science and Computational Intelligence-ICCSCI 2015, Procedia Computer Science 59. 2015, Pp 427 – 434.
9. Michel Grabisch, Christophe Labreuche, Fuzzy measures and integrals in MCDA. *Multiple Criteria Decision Analysis, Kluwer Academic Publishers*, pp.563-608, 2004. <halshs-00268985>
10. Michel Grabisch, Ivan Kojadinovic, Patrick Meyer, A review of methods for capacity identification in Choquet integral based multi-attribute utility theory: Applications of the Kappalab R package, *European Journal of Operational Research, Elsevier*, 2008, 186 (2), pp.766-785.
11. Art Owen, Clémentine Prieur, On Shapley value for measuring importance of dependent inputs, *SIAM/ASA Journal on Uncertainty Quantification*, ASA, American

Statistical Association, HAL Id: hal-01379188 https://hal.archives-ouvertes.fr/hal-01379188v3, 2017
12. László Kozma, Lkozma@cis.hut.fi, *k Nearest Neighbors algorithm (kNN)*, T-61.6020 Special Course in Computer and Information Science, Helsinki University of Technology, 2008.
13. Sung-Hyuk Cha, Comprehensive Survey on Distance, Similarity Measures between Probability Density Functions, *international journal of mathematical models and methods in applied sciences*, Issue 4, Volume 1, 2007.
14. Guy Shani and Asela Gunawardana, *Evaluating Recommendation Systems*, www.Netflix.com, www.amazon.com
15. Sheetal R. Thakare, Multi-Criteria Based Recommender System Scalability Optimization: The Approach Based on Clustering of Users, *IJSRD - International Journal for Scientific Research and Development*, Vol. 3, Issue 02, 2015. ISSN, online: 2321-0613.
16. Ferdaous Hdioud, Bouchra bouchra Frikh, Brahim Ouhbi, *Multi-Criteria Recommender Systems based on Multi-Attribute Decision Making*, iiWAS2013, 2-4 December, 2013, Vienna, Austria. Copyright 2013 ACM 978-1-4503-2113-6/13/12.
17. Tri Huynh Minh, Vu The Tran, Hung Huu Huynh, Hiep Xuan Huynh, Solution for Ordered Weighted Averaging Operator for Making in The Interaction Multi-Criteria Decision in User-Based Collaborative Filtering Recommender System, *The International Journal of Machine Learning and Computing (IJMLC)*, Volume 8 Number 4 (Aug. 2018).
18. Tri Huynh Minh, Vu The Tran, Hung Huu Huynh, Hiep Xuan Huynh, Item-Based Collaborative Filtering In The Multi-Criteria Recommender System With Ordered Weighted Averaging Operator On Sparse Datasets, *JP Journal of Heat and Mass Transfer, Special Volume, Issue II, Advances in Mechanical System and ICT-convergence*, Pp: 183-194, July 2018, http://dx.doi.org/10.17654/HMSI218183.