Academic Journal of Computing & Information Science
ISSN 2616-5775 Vol. 3, Issue 3: 70-82, DOI: 10.25236/AJCIS.030308

Research on Association Rule Recommender System Combining Time and Rating Information

Wenjing Yan[1, *]

[1] School of Information engineering, Nanjing University of Finance and Economics, Nanjing 210023, China
[*]Corresponding author e-mail: 844569289@qq.com

Abstract. *In this paper, when using Apriori algorithm to mine association rules, it often appears that the antecedent of the rule is the same, but the recommended consequent is more than one. In order to solve the problem, this paper proposes a weighted association rule recommendation algorithm which combines time and rating information. First of all, we describe the user's personalization and interest from the perspective of time and rating. From the point of view of time, the user's viewing data is sorted according to time to generate orderly user number, movie type and corresponding rating data. Time weight is introduced to add time weight to the user's rating on the sequence, so as to describe the importance of time. From the perspective of scoring, this paper filters out the low scoring and studies whether the scoring will affect the recommendation effect. Then, time and rating weighting are introduced to get the weighted value of association rules. Then, the weights and association rules are combined to select association rules whose confidence and support meet the threshold to generate strong association rules. Finally, strong association rules are used to recommend users, so as to improve the efficiency of the recommendation system.*

Keywords: *time, rating, weighted association rules, recommendation system*

1. Introduction

In recent years, with the rapid development of network information technology, the amount of information has shown an explosive growth. According to the China Internet Network Information Center, China's Internet users show a growing trend year by year. China's Internet users have increased from 298 million in 2008 to 828.51 million in 2018 [1]. In such an era of information explosion, it is difficult for people to meet their own needs. The main function of recommender system is to balance the contradiction between users and information producers, mine the items that users are interested in from massive data, and recommend the users, so as to improve the trust degree of users to the system and enhance the user experience [2]. In today's society, recommender system has been applied to many fields, which facilitates all aspects of people's life. The main research content of this paper is the recommendation of association rules, mainly from two aspects: user's time information and user's rating information. By adding weights to association rules, redundant association rules are filtered and strong association rules are generated.

2. Related works

The purpose of this paper is to improve the accuracy of the algorithm based on association rules. By weighting association rules, mining strong association rules in time series and filtering redundant association rules. There are many researches on association rules recommendation and weighted association rules at home and abroad. The following paper will elaborate on several methods.

Yu et al. completely described the relationship between attributes and preferences by using the concept of attribute rules [3]. Chen et al. changed borrowing records of different majors into multi-dimensional attributes and carried out association rule mining [4]. Wei Quanbin et al. used a matrix to represent the user's historical behavior, thus optimizing the efficiency of the recommendation algorithm [5]. He et al. reevaluated the frequent itemsets mining problem of Apriori algorithm for analyzing the balance and effective processing of cold and hot data [6]. Scholars at home and abroad also have some research on time series. The paper [7] solves the problem of complex feature extraction of time series, and proposes a time series classification model based on BP and naive Bayes. Reference [8] used Varma (vector autoregressive moving average) and LSTM (long-term and long-term memory network) algorithms to model and analyze the traffic congestion data of 57 roads near the capital airport. On this basis, the core idea of LSTM in dealing with multiple time series is added to the multiple regression algorithm, so that the multiple regression algorithm has the ability to deal with multiple time series. By observing the research at home and abroad, it is found that the previous research on weighted association rules focuses more on the association rules itself, and seldom considers the value of data itself, neglecting the data personalized mining. However, the data itself often plays an important role in the mining of weighted association rules, which can effectively filter the association rules in the rule base and generate strong association rules.

3. Weighted association rule algorithm combining time and rating information

Generally speaking, we divide Apriori algorithm to mining association rules into two main steps. The first step is to mine frequent itemsets, that is, to keep frequent itemsets whose support is greater than or equal to the minimum support threshold. The second step is to select rules whose confidence level is greater than or equal to the minimum confidence threshold from frequent sets, and then recommend users according to the association rule base. It is found that there may be more than one consequent rule in association rule recommendation. In this paper, taking Movielens dataset as an example, in the process of mining association rules by using Apriori algorithm, the situation as shown in Table 1, The antecedent of the rule 'Crime|Drama''Comedy' 'Drama|Mystery' has two kinds of rule consequent, which are 'Comedy' and 'Drama'. The association rules mined by Apriori algorithm are redundant and low value, and the recommendation accuracy of Apriori algorithm is not high. In this case, how to select association rules is the main problem of this paper.

Table 1 Rule antecedent and consequent

Number	Antecedents of rules	Consequent of rules
1	'Crime|Drama''Comedy''Drama|Mystery'	'Comedy'
2	'Crime|Drama''Comedy''Drama|Mystery'	'Drama'
3	'Drama''Action''Drama'	'Comedy|Romance'
4	'Drama''Action''Drama'	'Crime|Film-Noir'
5	'Drama''Action|Adventure''Comedy|Horror'	'Comedy'
6	'Drama''Action|Adventure''Comedy|Horror'	'Crime|Drama’

After observing and reading the literature, we consider to make a specific analysis of each user sequence, analyze the differences between different user sequences and how to show the differences between them. Finally, we find that the corresponding time stamp and rating data of each user sequence are different, so we can make specific analysis on them, so as to decide how to recommend association rules to users. Therefore, this paper studies whether time and rating will affect the recommendation effect, and proposes a weighted association rule recommendation algorithm combining time and rating information.

3.1 Time weight

Using Apriori algorithm to mine association rules will produce a lot of redundant association rules, in order to solve this problem, a weighted association rule algorithm combining time and rating information is proposed.

Generally speaking, the rating represents the user's preference. If the user gives a high rating on an item, it indicates that the user is interested in the item. Users' preferences often change with the change of time, so a user's recent rating can better reflect the user's interest, and is in a stable state in a short period of time. Taking the movie data of Movielens as an example, if a user scores a movie 5 points on October 1 and another movie 5 points on October 7, we tend to focus on the rating on October 7 when recommending association rules for users. Because the score on October 7th is closer to the current time than the score on October 1st, which can reflect the current interest of users. Therefore, the time weight can be added to the score. The weight should be related to the time length from the current time. The closer the distance is to the current time, the greater the weight is. The time weight formula is defined as shown in Formula (1):

ISSN 2616-5775 Vol. 3, Issue 3: 70-82, DOI: 10.25236/AJCIS.030308

$$NR = R \times \frac{T}{MT} \tag{1}$$

Where NR is the score plus time weight, R is the user's rating of the movie, T is the time corresponding to the score, and MT is the time of the current system. The algorithm flow of time weight is shown in Table 2:

Table 2 Flow chart of time weight algorithm

Algorithm: time weighting
Input: user rating R
Output: time weighted rating NR
Process: (1) Sort user rating series by time
(2) Calculate the time weight of the rating
(3) Multiply the rating with its corresponding time weight and output NR

3.2 Time plus rating weight

As mentioned above, ratings can represent a user's preferences. In other words, when a user gives a rating of 1 for a comedy movie, but a rating of 5 for an action movie, it means that he likes to watch action movies. The recommender system aims to recommend the movies that the user is interested in, and at the same time, it weakens the part of the movie that the user is not interested in or considers unimportant, then using the method of setting weights can achieve this purpose. We design the rating data corresponding to user time series as shown in Formula (2):

$$Rating = \overline{R} \tag{2}$$

In other words, the average rating of the project is taken, and then we combine the time weight with the average rating to obtain the final weight as shown in Formula (3):

$$\left(Weight = \overline{R} \times \frac{T}{MT} \right)^{*} \tag{3}$$

The algorithm is described in Table 3

Table 3 Weighted association rule algorithm combining time and rating information

Algorithm: weighted association rule algorithm combining time and Rating information
Input: user viewing sequence list L1, rating R, time stamp T
Output: recommended list L2
Process:
(1) Sort the user sequence L1 by time
(2) For the rating R, bring it into Formula (1) to calculate the time weight NR of the rating
(3) For the rating R, the average rating is calculated by using Formula (2)

(4) Take the above data into Formula (3) to get the weight of time plus rating

3.3 Design of weighted association rules

In the design of time plus rating weight mentioned above, we fuse the weight into association rules to form weighted association rules. The existing experimental data are shown in Figure 1:

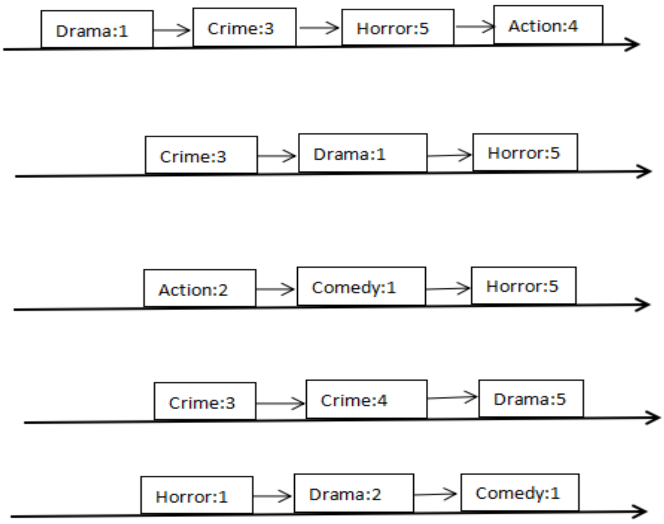

Figure. 1 Film types and ratings

The five long arrows in the Figure 1 represent the viewing sequence of five users, and each rectangle represents the type of movie users watch and the corresponding rating. The average time weights of the five users are 0.12, 0.21, 0.18, 0.16 and 0.26, respectively. According to Formula (2) and Formula (3), the calculation results are shown in Table 4:

Table 4 Average rating and final weight of type items

Type	Average rating	Final weight
Drama	(1+1+5+2)/4=2.25	0.12×2.25=0.27
Crime	(3+1+3+4)/4=2.75	0.21×2.75=0.58
Action	(4+2+4)/3=3.34	0.18×3.34=0.60
Horror	(5+5+5+1)/4=4	0.16×4=0.64
Comedy	(1+1)/2=1.00	0.26×1.00=0.26

ISSN 2616-5775 Vol. 3, Issue 3: 70-82, DOI: 10.25236/AJCIS.030308

Therefore, the contents shown in Table 5 can be obtained according to Table 4. The calculation method is to find the average value of the sequence corresponding to each user type sequence.

Table 5 Average weight of user sequence

User	Type	Calculation process	Average weight of user sequence
1	Drama,Crime Horror,Action	(0.27+0.58+0.64+0.60)/4	0.52
2	Crime,Drama Horror	(0.58+0.27+0.64)/3	0.50
3	Action,Comedy,Horror	(0.6+0.26+0.64)/3	0.50
4	Action,Crime Drama,	(0.6+0.58+0.27)/3	0.48
5	Horror,Drama Comedy	(0.64+0.27+0.26)/3	0.39
sum			2.39

After calculating the average weight of the user sequence, the weight is applied to the association rules. Assuming that the rule 'crime' horror '- > drama has been mined out, the weighted association rule of' crime 'horror' = (0.52 + 0.50) / 2.39 = 0.43 (the 'crime' horror 'appears in the sequence of user 1 and user 2 respectively). The specific algorithm flow is shown in Figure 2:

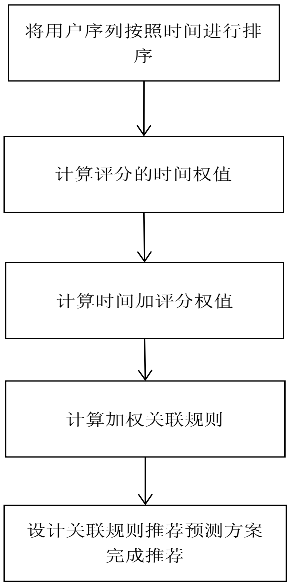

Figure. 2 Flow chart of time weighted association rule recommendation algorithm

ISSN 2616-5775 Vol. 3, Issue 3: 70-82, DOI: 10.25236/AJCIS.030308

4. Experiment

This section will design a variety of experiments to prove the algorithm proposed in this paper. Verify the feasibility of the algorithm, and through experimental comparison, to analyze the advantages of the algorithm. In general, the following experiments were completed.

(1) Compared with the traditional Apriori algorithm for mining association rules, whether the time will affect the recommendation is studied.

(2) Compared with the traditional Apriori algorithm for mining association rules, this paper studies whether the rating will affect the recommendation.

(3) Time and rating weights are added to the association rules to verify the effect of weighted association rules on the recommendation system.

4.1 Apriori algorithm mining association rules with temporal constraints

Apriori algorithm is used to mine temporal constrained unweighted association rules in time series, and then non weighted association rules are used to predict the recommendation. The recommendation strategy is that once the leader of the rule matches a certain sequence in the time series, the neighbor nodes of the sequence are used as the prediction range, and the same is the hit. Finally, the accuracy and coverage of the experimental prediction are calculated. The experimental results are shown in Table 6 and Table 7:

Table 6 The accuracy table of each interval obtained by apriori algorithm

Support / Confidence	0.2-0.25	0.25-0.3	0.3-0.35	0.35-0.4
0.5-0.6	0.0357	0.0325	0.0461	0.0216
0.6-0.7	0.0423	0.0413	0.0543	0.0432
0.7-0.8	0.0654	0.0734	0.0876	0.0798
0.8-0.9	0.1644	0.1756	0.1344	0.0834
0.9-1	0.0765	0.0568	0.0165	0.0143

Table 7 the Coverage Table of Each Interval obtained by Apriori Algorithm

Support / Confidence	0.2-0.25	0.25-0.3	0.3-0.35	0.35-0.4
0.5-0.6	0.0023	0.0012	0.0008	0.0006
0.6-0.7	0.0054	0.0026	0.0018	0.0011
0.7-0.8	0.0133	0.0042	0.0032	0.0027
0.8-0.9	0.0361	0.0148	0.0041	0.0032
0.9-1	0.0534	0.0124	0.0027	0.0021

ISSN 2616-5775 Vol. 3, Issue 3: 70-82, DOI: 10.25236/AJCIS.030308

From the above Table 6 and Table 7, when the minimum support is in the range of [0.25,0.3] and the minimum confidence is in the interval of [0.8.0.9], the accuracy reaches the maximum of 17.56%. When the minimum support is in the range of [0.35,0.4] and the minimum confidence is in the interval of [0.9.1], the accuracy reaches the minimum of 1.43%. When the minimum support is in the range of [0.2,0.25] and the minimum confidence is in the interval of [0.9.1], the coverage reaches the maximum of 5.34%. When the minimum support is in the range of [0.35,0.4] and the minimum confidence is in the interval of [0.5.0.6], the coverage reaches the minimum of 0.06%. Therefore, it is not ideal to use the association rules mined by the traditional Apriori algorithm to directly recommend the association rules. The reason is that the traditional rules are not suitable for combination types.

4.2 Time information

In the user time series data, we move forward 10 units, 20 units, 30 units (one unit has 13 user viewing data), and then mine association rules on these time-series data, and then make recommendations. Finally, the accuracy and coverage of recommendation are calculated. (the minimum support is set to 0.2, and the minimum confidence level is set to 0.8)

The experimental results are shown in Figure 3 and Figure 4:

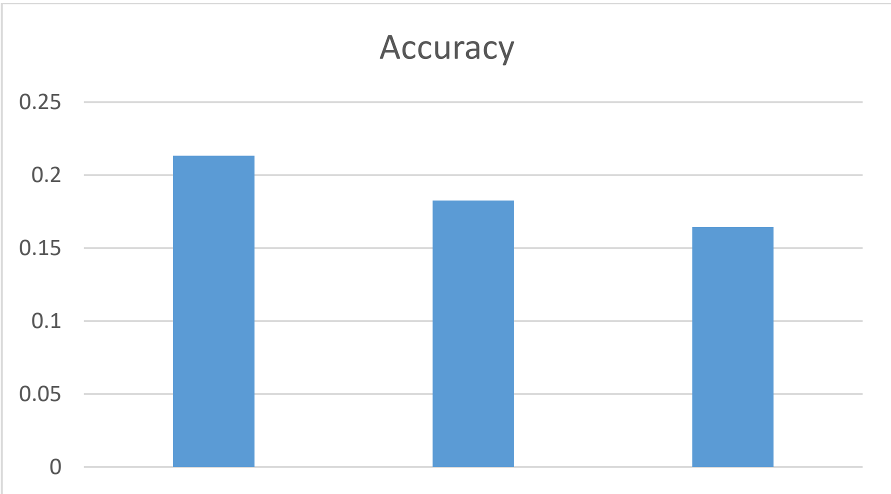

Figure. 3 Comparison chart of time accuracy

ISSN 2616-5775 Vol. 3, Issue 3: 70-82, DOI: 10.25236/AJCIS.030308

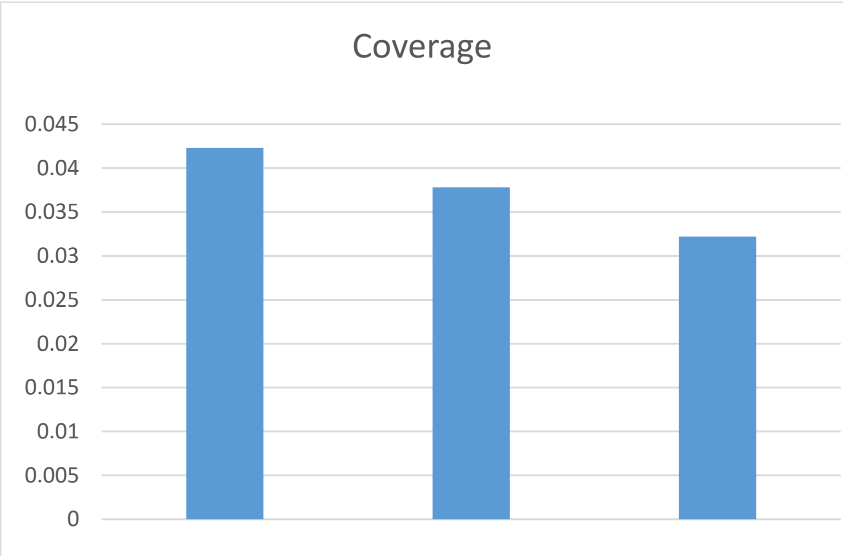

Figure. 4 Comparison chart of time coverage

As can be seen from the above figures, the accuracy and coverage of the recommender system are 21.33% and 4.23% respectively when moving forward 10 units, 18.26% and 3.78% respectively when moving forward 20 units. The accuracy and coverage of the recommendation system are 16.45% and 3.22% respectively when moving forward 30 units. Through comparison, it can be found that the recommendation effect of moving forward 10 units is better than that of 20 units, and the recommendation effect of moving forward 20 units is better than that of 30 units. Therefore, it can be concluded that the distance of time does have an impact on the recommendation effect. The closer the user sequence is, the more valuable it is for association rule mining and has a more positive impact on the recommendation effect.

4.3 Score information

The rating data of a user often represents a user's preference. The movie sequence with higher rating is more interested by the user, while the one with lower rating is not interested by the user. The recommender system aims to recommend the movie of interest to the user. Therefore, we can filter out the movie data whose rating is lower than 3.5, and then divide the data set into 80% training set and 20% test set. Taking 13 data as a time period, Apriori algorithm is used to mine frequent itemsets, and then make recommendations. Finally, the accuracy and coverage of experimental prediction are calculated. The accuracy and coverage of the recommender system are shown in Table 8 and Table 9:

ISSN 2616-5775 Vol. 3, Issue 3: 70-82, DOI: 10.25236/AJCIS.030308

Table 8 the accuracy table of each interval obtained by filtering score

Confidence \ Support	0.2-0.25	0.25-0.3	0.3-0.35	0.35-0.4
0.5-0.6	0.23	0.17	0.17	0.21
0.6-0.7	0.45	0.25	0.12	0.13
0.7-0.8	0.37	0.19	0.24	0.32
0.8-0.9	0.41	0.33	0.18	0.18
0.9-1	0.27	0.42	0.16	0.33

Table 9 The coverage table of each interval obtained by filtering score

Confidence \ Support	0.2-0.25	0.25-0.3	0.3-0.35	0.35-0.4
0.5-0.6	0.0024	0.0017	0.0009	0.0008
0.6-0.7	0.0063	0.0027	0.0022	0.0013
0.7-0.8	0.0135	0.0043	0.0043	0.0029
0.8-0.9	0.0367	0.0157	0.0052	0.0042
0.9-1	0.0546	0.0127	0.0028	0.0031

As can be seen from Table 8 and Table 9, compared with Apriori algorithm for association rule recommendation, the recommendation effect of this experiment has been greatly improved in both accuracy and coverage. It can be seen that higher rating information will have a positive impact on the recommendation effect.

4.4 The association rules are weighted by rating and time

From the above experiments, we can see that time and rating do have an impact on the recommendation effect, that is, the closer to the current recommendation and the higher the rating, the greater the weight. Design weights according to the formula mentioned in this paper and the association rules are weighted to get the result of the rules to recommend to users. Finally, the accuracy and coverage of the recommendation algorithm are calculated.

The experimental results are shown in Table 10 and Table 11:

Table 10 the accuracy table of weighted association rules

Confidence \ Support	0.2-0.25	0.25-0.3	0.3-0.35	0.35-0.4
0.5-0.6	0.51	0.43	0.32	0.14
0.6-0.7	0.71	0.54	0.63	0.36
0.7-0.8	0.35	0.57	0.17	0.12
0.8-0.9	0.49	0.11	0.37	0.18
0.9-1	0.27	0.42	0.15	0.33

Table 11 The coverage table of weighted association rules

Support / Confidence	0.2-0.25	0.25-0.3	0.3-0.35	0.35-0.4
0.5-0.6	0.23	0.17	0.17	0.21
0.6-0.7	0.45	0.25	0.12	0.13
0.7-0.8	0.37	0.19	0.24	0.32
0.8-0.9	0.41	0.33	0.18	0.18
0.9-1	0.27	0.42	0.16	0.33

Through the observation of Table 10 and Table 11, it can be found that when the minimum support is in the interval of [0.2,0.25], the minimum confidence is in the interval of [0.6,0.7], the accuracy can reach up to 70%; when the minimum support is in the interval of [0.2,0.25], the minimum confidence is in the interval of [0.6,0.7], the maximum coverage can reach 45%. Compared with the traditional Apriori algorithm, the accuracy and coverage rate of Apriori algorithm are greatly improved no matter which interval of support and confidence is.

4.5 Comparative experiment

The weighted association rule algorithm combined with time and rating information proposed in this paper is compared with some existing association rule recommendation algorithms on Movielens, Flixster and Epinions datasets respectively, and the accuracy and coverage of each algorithm are calculated and compared. Among them, "AP" is Apriori association rule recommendation, "FP" is FP Tree Association Rule recommendation, "PW" is probability weighted association rule recommendation, "SEW" is sequential expected weight association rule recommendation, "TRW" is time rating weighted association rule recommendation proposed in this paper.

The experimental results are shown in Figure 5 and Figure 6:

ISSN 2616-5775 Vol. 3, Issue 3: 70-82, DOI: 10.25236/AJCIS.030308

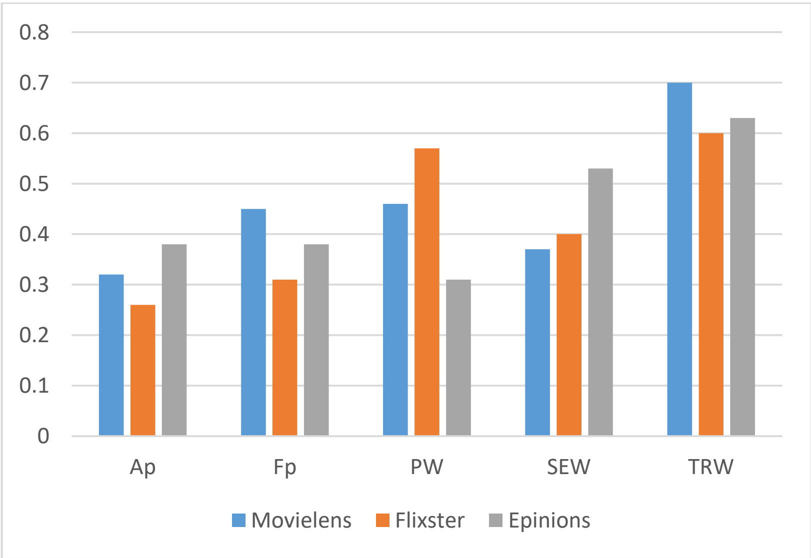

Figure. 5 Accuracy comparison chart

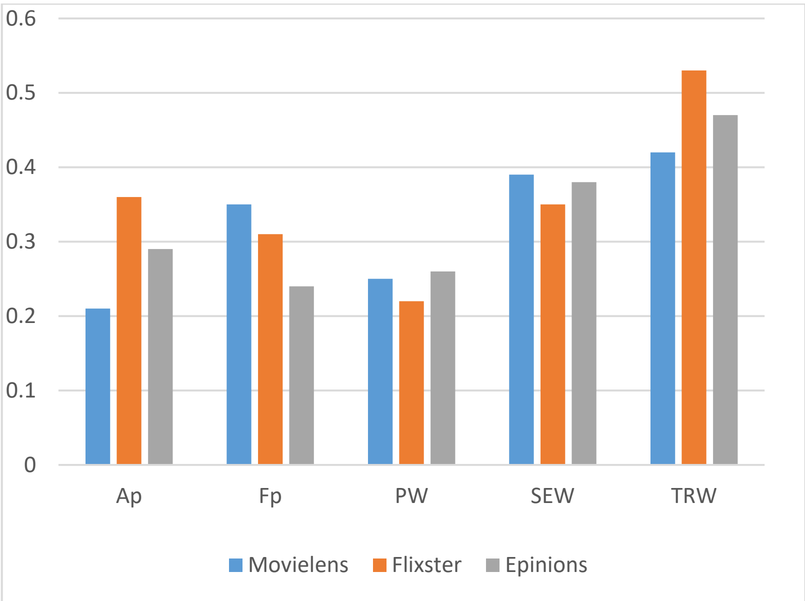

Figure. 6 Coverage comparison chart

ISSN 2616-5775 Vol. 3, Issue 3: 70-82, DOI: 10.25236/AJCIS.030308

By observing the data in Figure 5 and Figure 6, it can be found that the time rating weighted association rule recommender algorithm proposed in this paper has a greater improvement in accuracy and coverage compared with other algorithms on Movielens, Flixster and Epinions datasets. It can effectively improve the efficiency of the recommendation system.

5. Conclusion

The research of association rule recommendation based on time and rating information proposed in this paper can greatly improve the accuracy of recommendation, and use time and rating information to better describe the user's interest. However, there are still some areas worthy of further study. The next research focuses on the following aspects: (1) on the basis of this paper, combined with social network to study its impact on rating; (2) studying the user's interest drift to improve the accuracy of recommendation; (3) combining with content-based recommendation system can be used to solve the cold start of items; (4) on the basis of this paper, combining with social network to study its impact on rating There are a lot of association rules in mining association rules. In the future, the mining algorithm will be improved to filter out more redundant association rules.

References

[1] China Internet Network Information Center. The 43rd Statistical Report on Internet Development in China [R]. Beijing: China Internet Network Information Center, 2019.

[2] Kalgotra P, Sharda R. BIARAM: A Process for Analyzing Correlated Brain Regions Using Association Rule Mining [J]. Computer Methods & Programs in Biomedicine, 2018, (162):99-108.

[3] Yu Pengcheng, Li Ye. Design of recommendation model based on attribute rules and association rules [J]. Electronic technology, 2017, 30 (3): 26-29.

[4] Chen Shuying, Xu Jianying, Liu Yuwei, et al. Book recommendation service of University Library under the application of association rules [J]. Library forum, 2018, 38 (2): 97-102.

[5] Research on hybrid recommendation algorithm based on collaborative filtering and attribute association rules [D]. Southwest Jiaotong University, 2018.

[6] He Ming, Liu Weishi, Zhang Jiang, et al. Association rule recommendation algorithm supporting non null rate [J]. Journal of communications, 2017 (10): 18-25.

[7] Bai Yingying. Time series classification model based on BP and naive Bayes [J]. Computer application research, 2019 (8).

[8] Song Jiayi, Li GUANCHEN, et al. Intelligent transportation system based on multivariate time series prediction [J]. Modern information technology, 2019, 3 (12).