

# Association Rule in Recommendation to Reduce Scalability and Sparsity

Neha Sharma  
VNS Group of Institutions, Bhopal

Vivek Suryawanshi  
Faculty of Engineering  
VNS Group of Institutions

## ABSTRACT

In this Era of Internet, each and every people uses online websites for getting things done. Before purchasing any product users check the feedback /review related to that product on internet. Some system use information retrieval technique, so they will find the user tests and recommend the product to users. There are various recommendation technique are available. We proposed recommendation system for bike with the help of collaborative filtering technique. In which we are considering technical parameters for making dataset. Finding recommendation value Extract the parameters with threshold value. Also use text comments and apply association rules for finding recommendation bike in market. It gives better result by overcome scalability and sparsity problem.

## Keywords

Scalability, sparsity, dataset, apriori algorithm, association rule, hybrid recommendation

## 1. INTRODUCTION

A system which aids users who may not have adequate competence for arriving to a number of alternatives is termed as Recommender System. This system may also be defined in simpler words as data analysis programs which involves pattern matching which are complex in nature. Pattern matching of dataset is performed on predefined set of parameters which becomes efficient as the size of the content increases being inputted to the system.

In any areas, the data-set established through various characteristic or components enhance the recommendation[2]. This system provides assistant to targeted users in suggesting the significant products from large number of provided data through rigorous analysis of users behavior. But there is always a scope of improving collaborative filtering techniques. A parallel system which accepts the feedback and user reviews are considered from different inputs which when integrated with collaborative algorithms provides better recommendation results. However challenges and issues still persist in recommendation system like Scalability , Sparsity, gray sheep , cold start etc.

### 1.1 Content Based Recommendation:

Information of item content and ratings which is given to items by user is considered in this method[M]. In which recommendation is based on the content related to product. This is shown with example in figure 1.1 below. Suppose user like bike 1 as their design feature . So the system will recommend the bike which is best in design as per user test Hence Bike 3 is recommended to user 1.

Table. 1: Content Based Recommendation[8]

Bike ID	Feature
1	Design
2	Milage
3	Design
4	Speed

### 1.2 Collaborative Filtering:

This Filtering is depend on previous data of users. Two types of collaborative filtering

**User-based method:** It create group of users having same test and recommend the product to correlated user .This is shown with example in figure 2 below. Here user A like product 1,4,5 and user B likes product 2,4 . So product 4 is common in both the users , so they are correlated . Hence product 2 is also recommended for User A.

Table. 2: User Based Recommendation[8]

User name	Selected bike
User A	1,4,5
User B	2,4
User C	3

**Item-based method:** It based on correlated items. It means rated items given by user are mostly similar with the other user rated item[M]. From the figure 3, it shows that user B and user C likes Bike 1 and Bike 3. So, Bike 1 and Bike 3 are correlated to each other. And also user A likes Bike 1. So, Bike 3 recommend to user A.

Table. 3: Item Based Recommendation[8]

User Name	Selected Bikes
User A	1
User B	1,3
User C	1,3

Differnt technique gives various results . To improve the result accuracy we combine the recommender result.

## 2. EXISTING SYSTEM

Some existing recommendation technique used collaborative filtering with association rules. We study following papers for proposed recommender system

In paper [1] we studied different recommendation techniques.

Also different problems occurred in recommendation.

In paper[2], book recommendation system is proposed by author. This will give its users, the ability to view and search books, novels. Which will be used to draw out conclusion about the stream of a user the genre of the books liked by that user. Used these methods in other recommendation system to improve the result, as the fast recommendation base on groups.[4]

In paper [4] author presents a hybrid recommendation system. But the problem is scalability due to lack of information on same product.

In paper [15], author present the solution on scalability problem. This paper used association rule for to select best suitable group for any particular person. Association rule mining is to find out association rule that assure the predefined least support and assurance from a given database.

Above literature survey it is found that some common problems in current recommender system .

**1. Cold start:** It is difficult to gives recommendation to new users as his profile is almost empty and he has not rated any item. so his taste is unknown to the system.

**2.Sparsity:** If user has evaluated just few item then also its difficult to determine his taste. Mostly this problem due to lack of information .

**3. Scalability :** With the growth of number of users and items the system needs more resources for processing information and form result.

### 3. PROBLEM DEFINITION

Now a days, Before purchasing any product user searches online review . So, it is necessary to recommend the proper choice. The existing approaches for recommendation not gives the proper result . Sometimes sparsity problems arrive in recommendation due to the lack of information . Sometimes gray sheep problem arise because of inconsistency in user opinion. Such problem are mostly covered in proposed system We considered numeric review based on technical parameters and find the recommendation . Also use online review and apply association rule for recommending appropriate result. System gives less error value , so the predicted output gives faster and accurate result .

### 4. PROPOSED WORK

As per the existing systems 100 percent accurate result is essential. Mostly the challenges in recommendation is scalability and sparsity due to the lack of information. To Improve this problem we needs more resources for processing information and forming recommendation.

#### System Architecture:

##### Numeric Dataset:

Here we are created manually dataset in the numeric form. In which we are consider different parameters related to bike.

**Feature Extraction:** Here we are using threshold based feature extraction technique .There are many feature extraction technique but we are using rating in the form of numeric So the average calculation is suitable for our dataset .

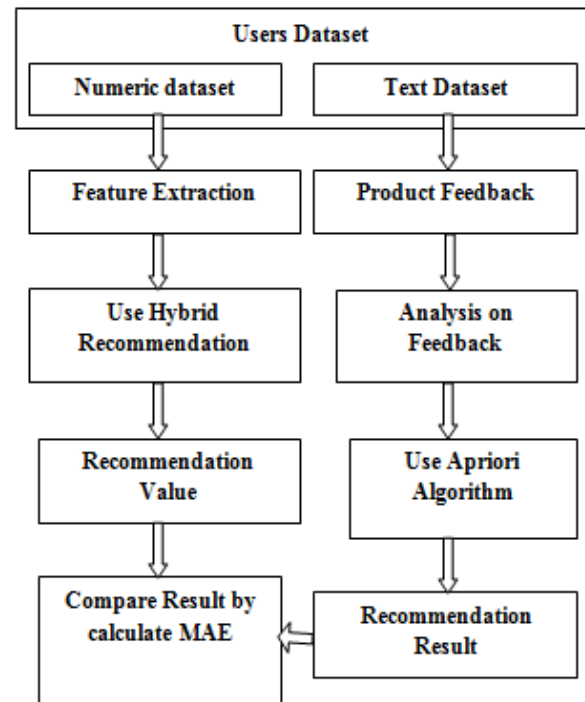
**Use Hybrid Recommendation :**Here we are using user based and Item based filtering , so the data will be filtered and recommend the bike to users . This result stored in system.

**Bike Feedback:** After recommendation we also check the text review related to the item on various forum.In order to optimize our sentiment it is necessary to gives review

perfectly. For that we use different dictionary like Positive, Negative ,Slang and Stop word.

**Analysis On Feedback:** From Item analysis and Review analysis we have predicted the final sentiment from users for n number of bikes of there interest ,but sentiment analysis is used here to recommend the required highest likely bikes to particular user.

Fig.1: System Architecture



#### Apply Apriori and Calculate recommendation value:

There are various algorithm for associating group of bikes. . Hence our proposed algorithm is used for checking the association between each bike in the input dataset .This is very helpful for finding out frequently purchase bikes. Thus it is used for forecasting on demand bike , ordering the bike which are in demand .This is also useful for manager and administrative person to order the product or item ,locate it for a business profit.

Mining the analysed data from extensive dataset which may be huge, uses Apriori Algorithm as a classic tool for data mining . This algorithm is quite popular in e-commerce application where targeted users are suggested to have products based on their past or present purchases. It ease the shopping experience of customers and also increases the sales.

Apriori algorithm may be used in other fields like healthcare also. Apriori algorithm produces a set of association rules which indicates all the best combinations based on input provided by user. Association rules are used to determine the relationship between items present in dataset. For example Lets assume we are buying few products like toothbrush and toothpaste together from an e-commerce website then based on our purchase and past purchases performed by other users it provides the best suggested products that we would most likely buy.

As explained in above example Apriori algorithm produces a set of association rules i.e it establishes the relation between the set of data. In this research we are using Apriori Algorithm:

Pass 1

1. Generate the candidate itemsets in C1

2. Save the frequent itemsets in L1

Pass k

1. Generate the candidate itemsets in Ck from the frequent itemsets in Lk-1

1. Join Lk-1p with Lk-1q as follow:

insert into Ck

select p.item1, p.item2, . . . , p.itemk-1, q.itemk-1

from Lk-1p , Lk-1q

where p.item1 = q.item1 , .....p.itemK-1=q.itemk-1

2. Generate all (k-1) subsets from the candidate itemsets in Ck

3. Prune all candidate itemsets from Ck where some (k-1) subset of the candidate itemset is not in the frequent itemset Lk-1

2. Scan the transaction database to determine the support for each candidate itemset in Ck

3. Save the frequent itemsets in Lk

Performance Evaluation:

There are two types of metrics used for recommendation system. One is Statistical Accuracy metrics. And other is Decision Support Accuracy metrics. In proposed system, all the evaluation of result is done using Statistical Accuracy Metrics. In this metrics, Root Mean Square Error and Mean Absolute Error is used for calculating result.[4].

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

where p=predicted ratings, q=actual ratings. The lower the MAE the more accurate the result.

## 5. EXPERIMENTAL RESULT

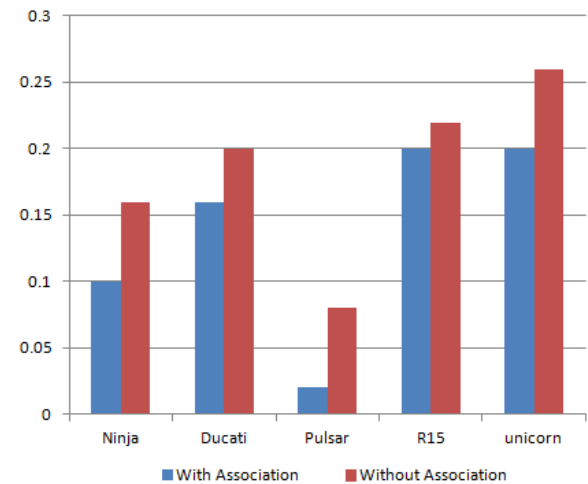
Existing method without ARM giving low output and another reason for this is lack of on demand information. So, for above Graph shows the result of existing method is low because of the above mention problem. For Overcoming this problem we are using our proposed system. It uses recommendation through online comments and numeric rating the experiment results indicated that the proposed approach can effectively alleviate the sparsity and scalability problem, have good coverage rate and recommendation quality.

**Table 4: MAE values With ARM**

Bikes	Pi value	qi value	MAE
Ninja	3.7	3.2	0.1
Ducati	3.5	2.7	0.16
Pulser	3.7	3.6	0.02
R15	3.1	2.1	0.2
Unicorn	4.2	3.2	0.2

**Table 5: Error values Without Association**

Bikes	Pi value	qi value	MAE
Ninja	4.0	3.2	0.16
Ducati	3.7	2.7	0.2
Pulser	4.0	3.6	0.08
R15	3.2	2.1	0.22
Unicorn	4.5	3.2	0.26



## 6. CONCLUSION

The goal of this research is to present recommendation system for bikes. In which first apply extraction technique with some threshold value and generate the result. Also external feedback data is scaled and collected. This data will overcome the sparsity problem as lack of data. We use multiple types of recommendation and analyze the result using association rule approach. It reduce the error value in result. So Our proposed system gives accurate result.

## 7. REFERENCES

- [1] Tejal Arekar, R.S. Sonar, Dr. N.J. Uke, "A survey on Recommendation System" IJIRAE ISSN:2349-2163 Volume 2 Issue1, Jan-2015.
- [2] Abhilasha sase, Kritika Varun, Prof. Deepali patil, "A proposed Book Recommendation system", IJARCC, Vol-4 Issue 2, February 2015.
- [3] Alejandro Baldominos, Yago Saez, Ignacio Marrero "An efficient and scalable recommender system for the smart web" 10.1109/INNOVATIONS. 2015.7381557 Dubai, United Arab Emirates
- [4] Kavinkumar.V, Rachamalla Rahul Reddy, Rohit Balasubramanian, Sridhar. M, Sridharan.K, Dr. D.Venkataraman "A Hybrid Approach for Recommendation System with Added Feedback Component" IEEE, 2015.
- [5] Nachiketa Sahoo, Ramayya Krishnan, George Duncan, Jamie Callan, Information Systems Research, Vol. 23, No. 1, pp. 231-246, 2012.
- [6] Yuan chun Jiang, Jennifer Shang, Yezheng Liu, Maximizing customer satisfaction through an online recommendation system: A novel associative classification model, Decision Support Systems, Vol. 48, No.3, pp 470-479, 2010.

- [7] Mustansar Ali Ghazanfar and Adam Prugel-Bennett, "Fulfilling the Needs of Gray-Sheep Users in Recommender Systems, A Clustering Solution" IEEE 11th International Conference on e-Business Engineering (ICEBE), pp. 77-91, 2014.
- [8] Fengkun Liu, Hong Joo Lee, Use of social network information to enhance collaborative filtering performance, *Expert Systems with Applications*, Vol. 37, pp 4772- 4778, 2010.
- [9] Mustansar Ali Ghazanfar and Adam Prugel-Bennett, "Fulfilling the Needs of Gray-Sheep Users in Recommender Systems, A Clustering Solution" IEEE 11th International Conference on e-Business Engineering (ICEBE), pp. 77-91, 2014.
- [10] Atisha Sachan and Vineet Richariya " A Survey on Recommender Systems based on Collaborative Filtering Technique" *IJJET*, Vol.2 Issue 2 April 2013.
- [11] P. N. Vijaya Kumar, Dr. V. Raghunatha Reddy PhD Research Scholar, Department of Computer Science and Technology, Sri Krishnadevaraya University, Anantapuramu, "Recommendation system and application" *IJIRCCE*-Vol. 2, Issue 8, August 2014.India.
- [12] On demand recommendation using association rule mining approach " ,*SCOPES*-2016
- [13] Maryam Habibi and Andrei Popescu-Belis " Keyword Extraction and Clustering for Document Recommendation in Conversations" *IEEE* ,2015.
- [14] Pijitra Iomsari , " Book Recommendation System for Digital Library Based On 'User Profiles ' by Using Association Rule" *IEEE*-2014
- [15] Solving the Sparsity Problem in Recommender Systems Using Association Retrieval YiBo Chen *JOURNAL OF COMPUTERS*, VOL. 6, NO. 9, SEPTEMBER 2011