Decision-making Recommender System using Machine Learning Collaborative Filtering

Pallavi Vyas[1], Anuj Bhardwaj[2], Dr. Ravindra Kumar Vishwakarma[3]

[1]Assistant Professor, Lovely Professional University, Jalandhar, Punjab
[2]Professor, Chandigarh University, Punjab
[3]Professor, Motherhood University, Roorkee, Uttarakhand
[1]pallavi.18751@lpu.co.in

Article Info
Page Number: 3813-3820
Publication Issue:
Vol. 71 No. 4 (2022)

Article History
Article Received: 25 March 2022
Revised: 30 April 2022
Accepted: 15 June 2022
Publication: 19 August 2022

Abstract
In this digital era, a personalized recommender system can play a vital role in selecting suitable items to fulfill user requirements. These types of recommender system also help students and universities in various ways, with these types of recommender system student is able to find suitable program and courses. The recommender system helps in increasing the average result of a university. A recommender system is a platform that can be helpful in reducing the searching cost and time of users and growing effectiveness while selecting programs and courses. The core technology implemented behind this recommender system comprises content-based, collaborative filtering, and hybrid recommender systems. This paper basically focuses on the principles of the collaborative filtering recommender system. Later on, the paper explains the working of a collaborative filtering algorithm. Also, its application in various areas.

Keywords: Recommender system, Machine learning, Collaborative filtering, Support Vector Machine

Introduction:

Recommender System provides a personalized system where users can find items of their choice. With a collaborative filtering technique, we can develop a system that can understand users' choices. Recommender System is capable of producing recommendations for individuals or a group[1]. The interface of the Recommender System is capable of smartly and intelligently processing the user information and providing genuine suggestions. The following figure shows the function of the recommender system.

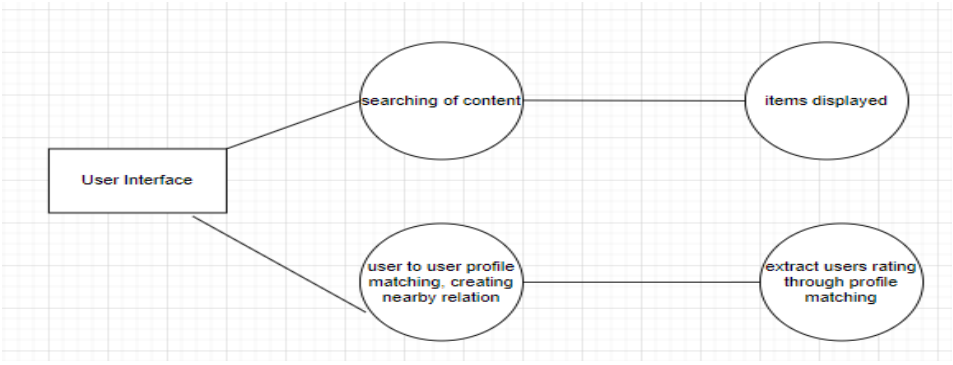

Fig 1: Function of Recommender System

Recommender System implements knowledge exploration techniques to create a personalized recommendation for the problem, product, or service[9]. This technique has achieved great success on the internet. The first recommender system application was Grouplens 1992 and firefly 1994. Later Yahoo and Amazon.com introduced the collaborative filtering idea.
Collaborative Filtering is the most effective technique to build an intelligent recommender system that is able to learn how to give better recommendations[2][4]. Many websites like AMAZON, youtube, and Netflix use collaborative filtering. This technique can be used to give suggestions to the user in order to show items based on his or her likes or dislikes[3].
Majorly collaborative filtering works on filtering information to create a recommender system. Recommender System helps in solving the challenges of how to produce recommendations of user interest[8].

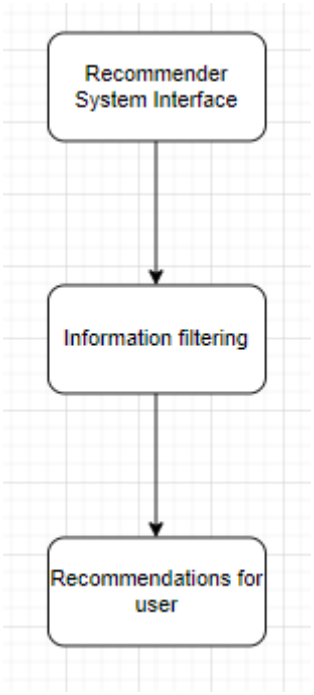

Fig 2: Filtering process

Basically, there are three forms of collaborative filtering technique[5]:

1. Model-Based Collaborative filtering.
2. Content-Based Collaborative filtering.
3. Hybrid Collaborative filtering.

Model-based collaborative filtering, provides item-based recommendations based on user ratings. This approach uses a machine learning-based algorithm. Many varieties of algorithms are available in this group.
On the other hand, content-based collaborative filtering provides user-based recommendations of a product. A content-based collaborative filtering system is to know about user and item both [10]. It compares and identifies user and items profile.
Hybrid filtering is a combination of the above technique.

Recommender system built on an interface that is capable of smartly and intelligently process the user information. After gathering user information, it processes and provide recommendations or suggestions to user.

Another school of thought is to find ratings from the user. Memory-based and model-based content filtering, Explicit and implicit ratings, and passive and active systems.
To make predictions memory-based algorithm apply whole user item database. The recommender system implements statistical methods to source users (neighbours) having the same interest in the past. Firstly, a neighbourhood of a user needs to be generated. This technique is known as nearest neighbours or user-based collaborative filtering. This technique is widely used. Secondly, model-based collaborative filtering provides recommendations on the basis of user rating. To implement this probabilistic approach needs to be implemented to find the expected value of users' actions.

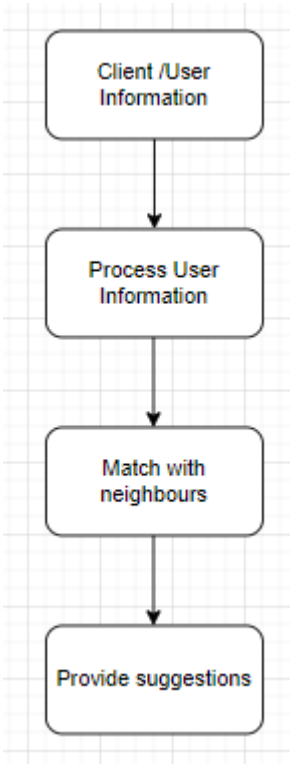

Fig 3: Principle of Recommender system

Different machine learning algorithm such as Bayesian network and clustering algorithm applied by model-based collaborative filtering
The explicit rating takes place when the user provides feedback on any item in the form of telling what he thinks about it. This is sometimes hard to verify.
Implicit rating is information or feedback gathered from user behaviour and then a conclusion drawn for further recommendations. This type of implicit rating is much more authentic than explicit rating.

Algorithm available for Collaborative filtering:

Collaborative filtering finds out a rating based on two methods one is item-based the other one is user-based. Various types of implementation algorithm are available with Collaborative filtering [6][7]. Some of the most popular are as follows:

One crucial and typical task in machine learning applications is to group the objects according to their similarities. Several modifications to the original k-means clustering approach, such as k-means ++ and kernel k-means, have been developed to improve machine learning applications [11]. These methods are among the many clustering techniques that have been developed. K-means is a non-linear methodology, whereas kernel k-means is a linear clustering method that separates the items into linearly separable groups.

[12][13] In order to answer a problem with any given uncertainty, the Bayesian Belief Network (BBN) is a vital framework technology. It deals with probabilistic occurrences. Through a directed graph without directed cycles, a probabilistic graphical model visually depicts variables and their particular dependencies. The BBN displays conditional dependencies between two or more entirely arbitrary variables in general. These probabilistic networks are only used to identify probable anomalies. This makes them very applicable for machine learning, which primarily relies on anomaly detection. Its application area includes prediction based on past data, providing automating understandings, logical reasoning of AI machines and so many others.

[14] The idea of similarity measure is the most crucial stage in any clustering since it determines how to accurately calculate how similar two components are. As a result, certain elements may be close to one another according to one distance and farther apart according to another, which will affect how the group is represented. Finding dense areas and estimating cluster assignment of new data objects both need numerous distance/similarity computations. In order to select the ideal one, it is crucial to comprehend the efficiency of various methods. Many other similarities are being used for measuring pairwise similarity between products such as cosine similarity and Jaccard coefficient.

[15] The test statistic used to assess the association or statistical link between two continuous variables is called Pearson's correlation coefficient. Due to the fact that it is based on the principle of covariance, it is regarded as the best technique for determining the relationship between variables of interest. Being based on the method of covariance, it is regarded as the best method for determining the relationship between variables of interest.

[16] Regression is a statistical technique for simulating the correlation between dependent (target) and independent (predictor) variables using one or more independent variables. Precisely, when other independent variables are maintained constant, regression analysis enables us to comprehend how the value of the dependent variable changes in relation to an independent variable. Regression analysis predicts real values like age, temperature, grade, price, etc.

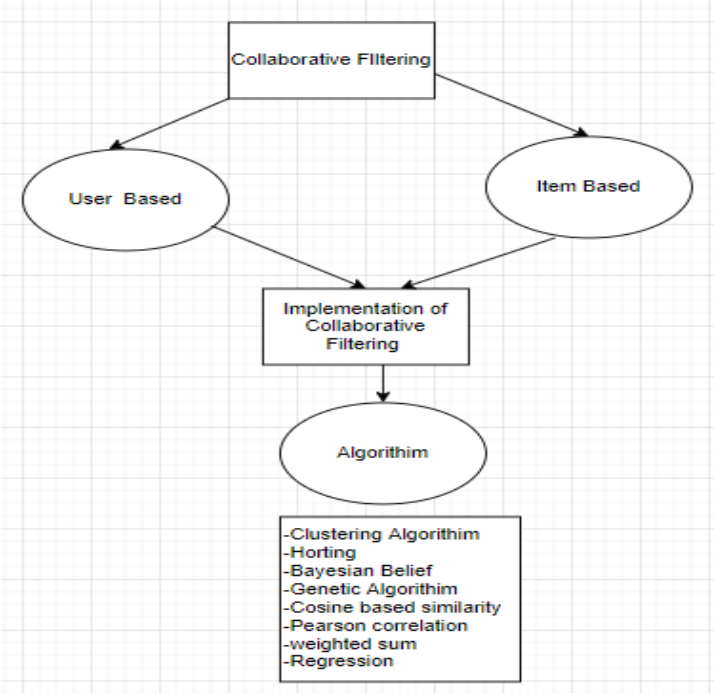

Fig 4: Classification of Collaborative filtering

Implementation of Collaborative filtering:

[17][18] One of the powerful and traditional machine learning methods that can still be used to help with data categorization issues is the Support Vector Machine(SVM). The support vector machine is mathematically complex and challenging to implement. It is most likely used in supervised learning algorithms to solve classification problem and regression problem.

In order to swiftly categorise new data points in the future, the SVM algorithm aims to determine the optimum line or decision boundary that can divide n-dimensional space into classes.
In SVM algorithm hyperplanes are created by chosen extreme points which is called support vector.
Two different variants of SVM are available:
1. Linear SVM: The term "linearly separable data" refers to data that can be divided into two groups using only a single straight line.
2. Non-Linear SVM: When a dataset cannot be identified using a straight line, it is said to be non-linear.

Hyperplane and Support Vector:

In SVM, a hyperplane is a decision border that distinguishes between the two classes. Depending on which side of the hyperplane a data point falls, it can belong to a different class. Hyperplane dimensions depend on the number of inputs given in the dataset. If 2 inputs are there it will create a straight line, if it is 3 it will be a two-dimensional plane.

A support vector is a point that can have an impact on hyperplane position it is a point near to hyperplane. Usually, select a point from which the distance between the support vector and hyperplane is maximum.

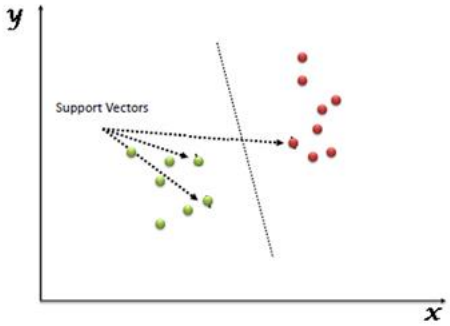

Fig 5: Hyperplane and support vector

SVM is implemented in python using the scikit library can be implement SVM to predict grade of students.

Step1: Import library

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%Matplotlib inline
```

Step2: Data pre-processing

It includes splitting data into attributes and labels and dividing data set into training and test sets.

```
X = dataset1.drop('Class', axis=1)
y = dataset1['Class']
```

split data into test and training set.

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.20)
```

Step3: Training and making prediction.

```
from sklearn.tree import DecisionTreeClassifier
classifier = DecisionTreeClassifier()
classifier.fit(X_train, y_train)
```

```
y_pred = classifier.predict(X_test)
```

Step4: Evaluating the algorithm.

```
from sklearn.metrics import classification_report, confusion_matrix
print(confusion_matrix(y_test, y_pred))
print(classification_report(y_test, y_pred))
```

Above code will produce the evaluation and returns the accuracy.

Future scope:

Further SVM can be implemented on students results, train the data set and will be able to design a model which is able to predict accurate grade. Past data of students will help. Through this, we can create a model for the prediction of programme which will be more beneficial for a student to select options for higher education. Decesion tree can also be implemented as it is a classical model for prediction.

References:

1. Building Switching Hybrid Recommender System Using Machine Learning Classifiers and Collaborative Filtering Mustansar Ali Ghazanfar and Adam Pr¨ugel-Bennett
2. P. Resnick and H. R. Varian, “Recommender systems,” Commun. ACM, vol. 40, no. 3, pp. 56–58, 1997.
3. J. Y. G Linden, B Smith, “Amazon.com recommendations: item-to-item collaborative filtering,” in IEEE, Internet Computing, vol. 7, 2003, pp. 76–80.
4. Collaborative filtering adapted to recommender systems of e-learning J.Bobadilla F.Serradilla A.Hernando MovieLens
5. Recommender systems to support learners’ Agency in a Learning Context: a systematic review Michelle Deschênes International Journal of Educational Technology in Higher Education volume 17, Article number: 50 (2020) Cite this article.
6. Collaborative Filtering Recommender Systems J. Ben Schafer, Dan Frankowski, Jon Herlocker & Shilad Sen.
7. A New Collaborative Filtering Recommendation Method Based on Transductive SVM and Active Learning Xibin Wang ,[1,2] Zhenyu Dai,[3] Hui Li ,[3] and Jianfeng Yang[1]

8 . H. Liu, Z. Hu, A. Mian, H. Tian, and X. Zhu, “A new user similarity model to improve the accuracy of collaborative filtering,” Knowledge-Based Systems, vol. 56, pp. 156–166, 2014.

9. A Survey of Collaborative Filtering-Based Recommender Systems: from Traditional Methods to Hybrid Methods Based on Social Networks Rui Chen1,2 , Qingyi Hua*,1, Yan-Shuo Chang*,3 , Bo Wang1,4 , Lei Zhang5 , and Xiangjie Kong*,6

10. Item-Based Collaborative Filtering Recommendation Algorithms Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl fsarwar, karypis, konstan, riedlg@cs.umn.edu GroupLens Research Group/Army HPC Research Center Department of Computer Science and Engineering University of Minnesota, Minneapolis, MN 55455

11. Silhouette Analysis for Performance Evaluation in Machine Learning with Applications to Clustering Meshal Shutaywi 1 and Nezamoddin N. Kachouie .

12. Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Mateo (CA)

13. Learning Bayesian networks from data: An information-theory based approach JieCheng[a]RussellGreiner[a]JonathanKelly[a]DavidBell[b]WeiruLiu[b]

14 Yonggang. Lu, Xiaoli. Hou and Xurong. Chen, A novel travel-time based similarity measure for hierarchical clustering, Elsevier, August 2015.

15. A guide to appropriate use of Correlation coefficient in medical research MM Mukaka[1,2,3]

16. A comparative analysis on linear regression and support vector regression. Kavitha S; Varuna S; Ramya R

17. M. A. Hearst, S. T. Dumais, E. Osman, J. Platt, and B. Scholkopf. “Support vector machines.” Intelligent Systems and their Applications, IEEE, 13(4), pp. 18–28, 1998.

18. An improved training algorithm for support vector machines E. Osuna; R. Freund; F. Girosi, IEEEplore