RECOMMENDATION SYSTEM FOR E-COMMERCE USING MEMORY-MODEL BASED COLLABORATIVE FILTERING

BY

ABDULLAH AL-REDWAN
ID: 171-15-9538

TARIQ AL JAMIL
ID: 171-15-9534

AND
RAZWAN SHAHRIAR KHAN
ID: 171-15-9525

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By
MS. FAHMIDA AFRIN
Lecturer
Department of CSE
Daffodil International University

Co-Supervised By
Mr. MASUD RABBANI
Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

APPROVAL

This Project/internship titled **"Recommendation System for E-Commerce Using Memory Model Based Collaborative Filtering"**, submitted by Name: Tariq Al Jamil ID No: Student ID: 171-15-9534, Name: Abdullah Al Redwan Student ID: 171-15-9538, Name: Razwan Shahriar Khan Student ID: 171-15-9525 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 01.06.2021.

BOARD OF EXAMINERS

Chairman

_____

Dr. Touhid Bhuiyan

Professor and Head

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University

Internal Examiner

_____

Md. Tarek Habib

Assistant Professor

i

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University

Internal Examiner

_____

Saiful Islam

Senior Lecturer

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University

External Examiner

_____

Dr. Md Arshad Ali

Associate Professor

Department of Computer Science and Engineering

Hajee Mohammad Danesh Science and Technology University

ACKNOWLEDGEMENT

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Fahmida Afrin, Lecturer, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Ms. Fahmida Afrin
Lecturer
Department of CSE
Daffodil International University

Submitted by:

Abdullah Al Redwan
ID: -171-15-9534
Department of CSE
Daffodil International University

Tariq Al Jamil
ID: -171-15-9534
Department of CSE

Razwan

Razwan Shahriar Khan
ID: -171-15-9525
Department of CSE

ABSTRACT

In today's technical era, every startup or a company attempt to establish a better sort of communication between their products and the users, and for that purpose, they require a type of methods which can promote their product effectively, and here the recommender system serves this motive with positivity. we all know numerous attempts have been made for expanding the accuracy in recommendation system, but somehow recommendation scenarios are much more complex and most of the cases have limited rating system for their items. Here, we present a demonstrative approach that will show how memory and model-based collaborative filtering enhance the accuracy and efficiently in our proposed recommendation system. As we know recommendation systems are used in many various areas' like music, movie, news, books, social media platform. It is a filtering system that tries to predict and show the items that a user would like to purchase. In this paper we are using Memory based (Item to Item) and Model based (Item to User) collaborative filtering to solve various problem like cold start, grey sheep, data sparsity using Utility Matrix method which will help to find user item relationship from previous purchase history and cluster the Item to item relationship using K-means Algorithm which will solve the problem of cold start, grey sheep and spared the data problem.

TABLE OF CONTENTS

CONTENTS	PAGE

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Introduction

To increase customer experience and boost sales, product recommendations system is applied in most e-commerce's websites. The recommendations are based on customers' purchasing habit and their past orders, as well as inferred references of other same-interest's purchasers. The core of the product recommendation system is the combination of artificial intelligence, algorithms and data structures. A recommendation for a specific product is based on the most relevance to a particular user in a given context.

A recommendation system speeds up searches and make easier for users to find relatable content they are interested into and they usually discover related word they never think of. So they start to feel known and curious about the next content. By utilizing the pattern of users shopping experience a company gains huge competitive advantages and they gain huge amount of customers for their better search capabilities and experiences.

1.2 Motivation

Many Researchers use collaborative filtering to give user and company a better shopping experience and with the solve of cold start problem, data sparsity problem, Grey sheep problem etc. But we find that there was a problem of efficiency and accuracy in those papers. For this purpose, we have decided to do a research on this topic and we think this research knowledge will help us in future to make a better and accurate recommendation system which will help our people and all over the world.

The ultimate goal of our research is to develop a new approach to recommend product to the user which will solve cold start problem, Grey sheep problem and spared the data problem with better efficiency and accuracy. During this thesis, the key technological

innovation is the Utility Matrix with decomposing and correlation to find relationship between user and item, K-means algorithm to find relationship between many items.

From this research, we hope that customer satisfaction will be ensured through solving this essential problem of a recommendation system and service provider will be enhancing their revenues using this system.

1.3 Rationale of the Study

In our thesis, we used Memory and Model based collaborative filtering with Utility Matrix and K-means Algorithm over the 20 thousands amazon dataset including the feature user id, product id, product rating and product description.

Research Questions

- How does a person choose the best topic that may suits best to the person's ability?
- How can we involve students with new research trending topics?
- Do students know about the Data Science/Revolution?

1.4 Expected Outcome

Many students on their thesis period struggle to find research paper which is related to their content or what they want to do. From this research, universities and research institutions can discover the papers most pertinent to their thesis or research projects. So, searching for the exact papers to read becomes a very significant part of their academic studies and introducing with trending and upcoming research topic impacts on their thought. 'Recommendation System for E-commerce using Memory-Model Based Collaborative Filtering' help these people to find out the most pertinent papers and saving their valuable time.

In addition, it can helpful for those people who interested in Data Science, a hot topic in recent times, which motivates them to work with new problems, finding new solutions, and help the students who are confused about their field of interest. In computer vision, there are many algorithms and techniques to extract the best information and knowledge from a

massive number of collected data. For this research, we are using here Two model for our data set and find out the best result but this is the prediction and a continuous process and result depends on the data set and the features.

We wish that our work will be more effective than we thought and it will be helpful for many people who are interested in this field.

1.5 Report Layout

In chapter 1, Introduction is discussed about what is recommendation system and its area of usage, motivation, rationale of the study, research questions and the result we expect from this thesis and it is followed by the report layout later on.

The rest of the report arrangement is as follows:

In chapter two, the background of our research topic is discussed. This chapter also deals with the related works, compare previous work with our work, scope of the problem and the challenges we face.

In chapter three, we will explain the methodologies that are related in our study, how we collect our data, statistical analysis from our data, what we proposed and how we implement our requirements.

In chapter four, we will explain the experiment results and discussion about the result.

In chapter five, we will explain how our project will be positive impact on our society and environment, its ethical aspects and sustainability plan.

Finally, we will conclude our work with some future work scopes and recommendations in chapter six.

CHAPTER 2

BACKGROUND

2.1 Terminologies

Model based collaborative filtering is basically used to finds similarity between the products based on their context. Customers previous history is also taken for find similar products the customer may like.

For example, if a customer like grocery items then the customer also has a chance to buy kitchen gadgets or if a customer buys some clothes then the customer also has the possibility to buy shoe or other accessories.

The data used in this filtering is divided in two categories: the users and the items. Each user likes certain items, and the rating value r_{ij} (from 1 to 5) is the data associated with each user **i** and item **j** and represents how much the user appreciates the item. These rating values are collected in matrix, called utility matrix **R**, in which each row **i** represents the list of rated items for user **i** while each column **j** lists all the users who have rated item **j.**

	Item 1	Item 2	Item 3	Item 4	Item 5
User 1	1	0	0	1	0
User 2	0	0	0	0	1
User 3	0	0	0	1	0
User 4	0	0	1	0	0

Utility Matrix

Figure I: Utility matrix relation between user and item

In other Hand, Memory based collaborative filtering find the relationship between two items based on their item description which is like if we search a keyword in an e-commerce site then it will show us related meaning keyword as recommendation apart from what we searching.

For example, if a customer search "cutting tool" then it will be showing the recommendation of word like blade, metal, grip, saw etc.

For this filtering, the data we used is item description. From this description we will find different clusters and we will then recommend the word by key search. K defines the number of pre-defined clusters that need to be created in the process, as if K=2, there will be two clusters, and for K=3, there will be three clusters and so on. The algorithm takes the unlabeled dataset as input, divides the dataset into k-number of clusters, and repeats the process until it does not find the best clusters. The value of k should be predetermined in this algorithm.

Given a set of observations ($\mathbf{x}_1$, $\mathbf{x}_2$, ..., $\mathbf{x}_n$), where each observation is a d-dimensional real vector, k-means clustering aims to partition the n observations into k ($\leq n$) sets $\mathbf{S} = \{S_1, S_2, ..., S_k\}$ so as to minimize the within-cluster sum of squares (WCSS) (i.e. variance). Formally, the objective is to find:

$$\underset{\mathbf{S}}{\arg\min} \sum_{i=1}^{k} \sum_{\mathbf{x} \in S_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2 = \underset{\mathbf{S}}{\arg\min} \sum_{i=1}^{k} |S_i| \operatorname{Var} S_i$$

where $\boldsymbol{\mu}_i$ is the mean of points in S_i. This is equivalent to minimizing the pairwise squared deviations of points in the same cluster:

$$\underset{\mathbf{S}}{\arg\min} \sum_{i=1}^{k} \frac{1}{2|S_i|} \sum_{\mathbf{x},\mathbf{y} \in S_i} \|\mathbf{x} - \mathbf{y}\|^2$$

The equivalence can be deduced from identity . Because the total variance is constant, this is equivalent to maximizing the sum of squared deviations between points in different clusters (between-cluster sum of squares, BCSS), which follows from the variance.

The k-means clustering algorithm perform two main tasks:

- Determines the best value for K center points or centroids by an iterative process.
- Assigns each data point to its closest k-center. Those data points which are near to the particular k-center, create a cluster.
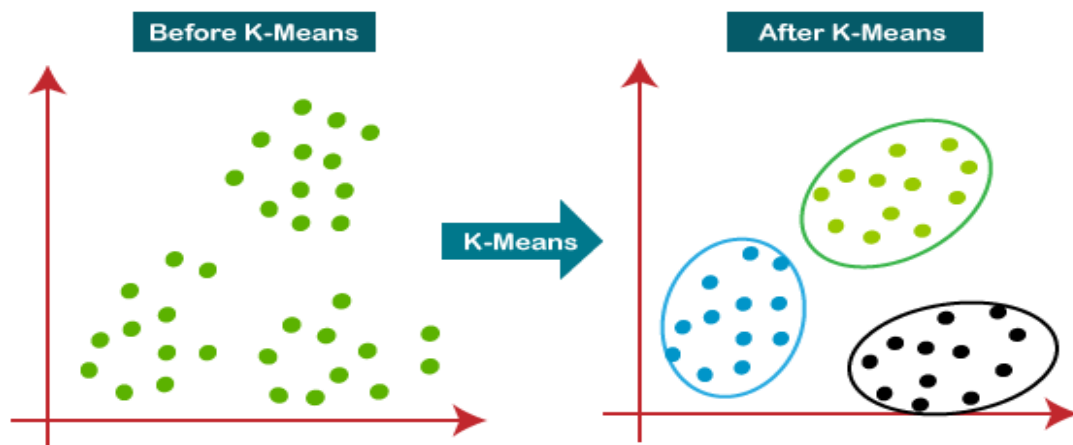

Figure II: How k-means cluster work

Each cluster has some data points common, that is why it is away from other clusters.

2.2 Related Works

Research and development of recommendation system have been made in the field of computer vision for many years now even we read a paper which is from 1995. So it's been clear that recommendation systems are widely popular for its time consuming and better experience from the customer's end. However, many work solve the cold start, data

6

©Daffodil International University

sparsity problem and many of them solve the gray sheep problem but none of them combined those two problems into a solution.

One of the work by K. Ravikanth, K. Chandrashekar, K. Sreekanth and P. Santhosh Kumar from department of CSE, IIIT India used Memory-based approach where they find similarities through Jaccard–Cosine similarity, for the missing values they used Pearson correlation. For Model-based approach they use Matrix Factorization then they combined both approach into hybrid RS. For this model they used amazon product dataset.

Another work by Ayad R. Abbas and Shaymaa Ashor from University of Technology, Baghdad-Iraq where they build a model of Content and Model based CF using Cosine Similarity and Pearson correlation.

Rong Hu and Pearl Pu from Swiss Federal Institute of Technology in Lausanne, Switzerland worked on CF approach with Personality Information using Cascade mechanism where they solve cold-start problem and for prediction they used Mean Absolute Error (MAE).

A. Felfering and R. Burke from college of computing and digital media, Depaul University Chicago, USA describe how many constraints are there in RS and research issues. They describe different type of recommendation approach and different types of relation between item and a customer or user.

2.3 Comparative Analysis and Summary

From doing our research, we studied several research papers, articles, book, and conference paper. In this section, we explained on the fore thought of other research and their outcome. Also, we discussed about the scope of the problem and challenges and an overview of the background. In computer vision, many researchers applied the same algorithms on the same dataset. Sometimes the same dataset and same algorithms can give different types of result. Therefore, it is very confusing.

Many researchers used memory-model based CF on their work and solve the problem they wanted to but there have been many loop holes into their work like it's not much efficient and fast responsive and some authors, in paper they did not write enough information about their method and algorithm they used. It creates a critical situation for re-implement or re-used the algorithm. Later, different embodiments are used in the same method that might create varieties in output. A little change in datasets, methods or user inputs prophesies creates a bigger change in the representation of the methods. Therefore, selecting the perfect approaches is a significant task.

2.4 Scope of the Problem

While doing our research, we found that selecting the actual research field from the huge ocean of knowledge fields is quite difficult. There are many knowledge fields where a researcher wants to do research. However, they could not able to choose most of the cases. Therefore, in colleges, universities and research institutions, professors, graduate students, and other researchers got to find the papers that are most pertinent to their research projects. In their academic lives, looking for the exact papers to read becomes a very significant part.

This research paper 'Recommendation System for E-commerce using Memory-Model Based Collaborative Filtering' will help these people in utilizing their precious time and the most pertinent papers. There are many related works. However, I hope our work will be more descriptive and easy to understand and it will be helpful for our running society.

2.5 Challenges

Recommend accurate item to a customer is an interesting challenge over the years now from different aspects. Specifically, recommendation system is becoming more significant in today's world in the revolution of e-commerce in every single country. In this paper, we address the problem of low accuracy and inefficient recommendation from the perspective of E-commerce popularity in this pandemic. Practically a research paper/field like recommendation system has to use a big dataset. After collecting the dataset, our main

challenge is to implement our methodology where we solve complex problems like cold start, data sparsity and gray-sheep.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Research Subject and Instrumentation

In this chapter, we are going to describe the method used for this project in the implementation requirements part. We will also discuss the phases of our working process and data collection method. In addition, the chapter elaborates the statistical analysis. Research subject and instrumentation are also included in this chapter and working process will also present in this section.

This research is based on memory-model based CF where we used Utility Matrix and K-means clustering algorithm.

Utility matrix has some advantages:

- Simpler and faster.
- Reduction of time.
- Reliable.

K-means clustering advantages:

- Guarantees convergence.
- Can warm-start.
- Can cluster different shapes and sizes.

Instrument-

For this research, we are using 3.70GHz AMD Ryzen 5, 16GB memory, 64-bit operating system and x-64 based processor.

3.2 Data Collection Procedure

Dataset

Bangladesh has several e-commerce site which is very popular like Daraz.com, Bikroy.com, Chaldal.com, Evaly etc. We want to work on our country dataset for our thesis but we couldn't collect data from our country e-commerce platform even though we sent them email several times. After that we collected our thesis dataset from amazon. Dataset snippet are added below:

	A	B	C	D
1	UserId	ProductId	Rating	Timestamp
2	A39HTATAQ9V7YF	205616461	5	1369699200
3	A3JM6GV9MNOF9X	558925278	3	1355443200
4	A1Z513UWSAAO0F	558925278	5	1404691200
5	A1WMRR494NWEWV	733001998	4	1382572800
6	A3IAAVS479H7M7	737104473	1	1274227200
7	AKJHHD5VEH7VG	762451459	5	1404518400
8	A1BG8QW55XHN6U	1304139212	5	1371945600
9	A22VW0P4VZHDE3	1304139220	5	1373068800
10	A3V3RE4132GKRO	130414089X	5	1401840000
11	A327B0I7CYTEJC	130414643X	4	1389052800
12	A1BG8QW55XHN6U	130414643X	5	1372032000
13	AIFAAVTUYHEHB	130414643X	4	1378252800
14	AVOGV98AYOFG2	1304146537	5	1372118400
15	A22VW0P4VZHDE3	130414674X	5	1371686400
16	AVOGV98AYOFG2	1304168522	5	1372118400
17	A6R426V4J7AOM	1304168522	5	1373414400
18	A22VW0P4VZHDE3	1304174778	5	1372896000
19	AKGB62WGF35J8	1304174778	5	1372896000
20	A22VW0P4VZHDE3	1304174867	5	1373068800
21	A1BG8QW55XHN6U	1304174867	5	1372291200
22	A1BG8QW55XHN6U	1304174905	5	1372291200
23	A22VW0P4VZHDE3	1304196046	5	1372896000
24	A22VW0P4VZHDE3	1304196062	5	1372896000
25	A3A4C2K3TWDAO5	1304196070	1	1378425600
26	A3FV2Q7WPZMQPV	1304196135	5	1375488000
27	A22VW0P4VZHDE3	1304196135	5	1372896000
28	A1RXI3A1E99112	1304351475	5	1405296000
29	A26QL1FBQO9C0E	1304351475	5	1392076800
30	A19KEEVZYO1KO6	1304351475	4	1388707200
31	A1F6HRR58TQ2R0	1304351475	5	1394928000
32	AVQP6PO1NK0IJ	1304351475	1	1389657600
33	A32IWHU4TUEQZR	1304351475	4	1383523200
34	ASPJCWRXUOYYY	1304351475	5	1392681600
35	A1LAW520BT5AM0	1304351475	1	1392422400
36	A3SWQ2QQ7JBPFA	1304351475	5	1402790400
37	A2D7CPCZH5VFVZ	1304351475	1	1396742400
38	A274NIJWOQWE30	1304351475	5	1385251200
39	A2QMUJ0R0VM5O9I	1304351475	1	1397606400

Figure III: A portion of our working amazon dataset

Data pre-processing:

Data pre-processing is a technique which is transform raw data into an understandable format. We know that real world data is incomplete, inconsistent and contain many lacks. So we transform our data into an understandable format and delete many features which we don't need and tidying up our data from process of deal missing values and we also divide out dataset into two for better work flow.

3.3 Statistical Analysis

Statistical analysis is a component of data analytics. For recommendation system we gathered 210334 data from many different customers with their purchased item. Each customer has their user id, product id which they bought and the product rating with timestamp. We find most popular item by their rating.

12
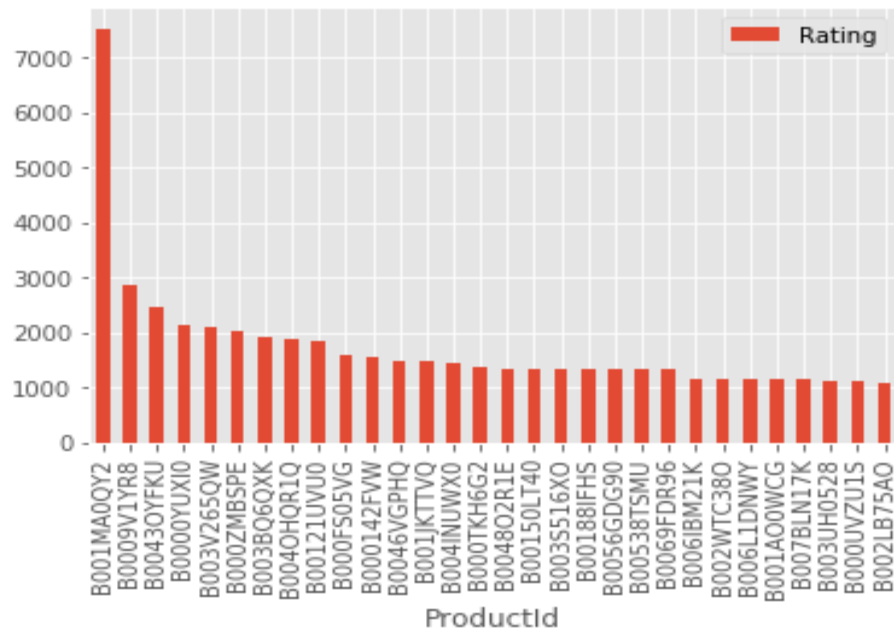
Figure IV: Most popular item by their rating

3.4 Proposed Methodology

We used utility matrix and k-means clustering in a specific way so that we can find a better and efficient model. The following diagrammatic representation explains the methodology of our proposed models:

13
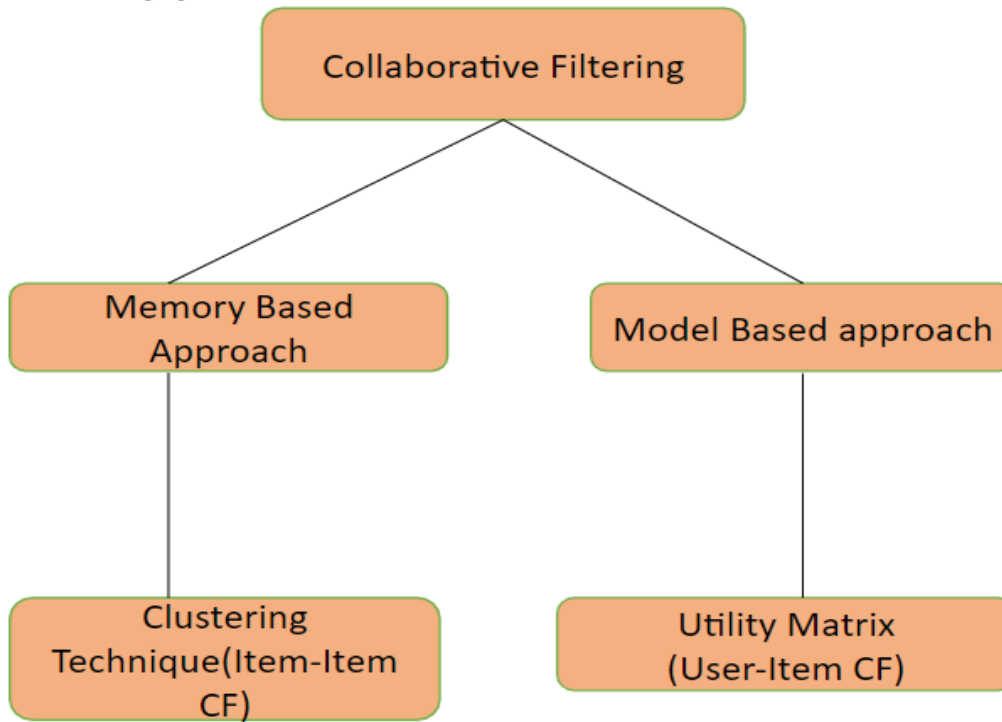
Figure V: Flowchart of the proposed model

The whole working process is something like this, where we used clustering technique for item to item CF and utility matrix for user to item CF.

3.5 Implementation Requirements

Model installation

For the implementation, first we have downloaded and installed pandas, numpy, matplotlib and sklearn libraries on conda environment where pandas, numpy libraries help in data preprocessing, mat[lotlib library help in data analysis period and sklearn package help in modeling and clustering.

14

Model Dissection

To build and run the model we need to use some library and modules of "Python" such as "pandas", "numpy", "matplotlib", "sklearn" with some additional modules of "sklearn". The use of those library and module on the python code are shown on the figure given below:


```
Import Necessary libraries

In [3]:  import numpy as np
         import pandas as pd
         import matplotlib.pyplot as plt
         %matplotlib inline
         plt.style.use("ggplot")

         import sklearn
         from sklearn.decomposition import TruncatedSVD
         from sklearn.feature_extraction.text import TfidfVectorizer, CountVectorizer
         from sklearn.neighbors import NearestNeighbors
         from sklearn.cluster import KMeans
         from sklearn.metrics import adjusted_rand_score
```

Figure VI: Used libraries and modules

After that we import our dataset as required, then we perform data preprocessing section. Then we use a matrix that show of our utility matrix use then we decompose the matrix and find correlation between user and item. After that we select a specific user id that will show us the recommend product.

15

Loading the Dataset

```
product_data = pd.read_csv('amazon.csv')
product_data = product_data.dropna()
product_data.head()
#product_data.shape
```

Figure VII: Load the dataset

For Model Based Collaborative Filtering Using Utility Matrix based on products sold and user reviews

```
#subset of the data
product_data1 = product_data.head(10000)

#create utility matrix
ratings_utility_matrix = product_data1.pivot_table(values='Rating', index='UserId', columns='ProductId', fill_value=0)
ratings_utility_matrix.head()
#ratings_utility_matrix.shape

#Transposing the matrix
A = ratings_utility_matrix.T
A.shape

#decomposing the matrix
SVD = TruncatedSVD(n_components=10)
matrix_decompose = SVD.fit_transform(A)
matrix_decompose.shape

#find correlation
corr_matrix = np.corrcoef(matrix_decompose)
corr_matrix.shape

#after that we find recommendate product
```

Figure VIII: Create utility matrix and correlation for recommendation

For clustering we use description of the products then we use clustering model to predict the clusters which show us the recommendation for our searched keyword.

Loading the data

```python
#Load the data
product_description = pd.read_csv('product_des.csv')
product_description.shape

#deal with missing values
product_description = product_description.dropna()
product_description.shape

#subset
product_description1 = product_description.head(500)

#extract the features
#Converting the text in product description into numerical data for analysis
vectorizer = TfidfVectorizer(stop_words='english')
A1 = vectorizer.fit_transform(product_description1["product_description"])
A1

# Fitting K-Means to the dataset
A=A1
k_means = KMeans(n_clusters = 10, init = 'k-means++')
y_kmeans = k_means.fit_predict(A)
plt.plot(y_kmeans, ".")
plt.show()


#Define a function to show the clusters
def print_cluster(a):
    print("Cluster %d:" % a),
    for i in order_centroids[a, :10]:
        print(' %s' % terms[i])


#Top words in each cluster based on product description
true_k = 10
model = KMeans(n_clusters=true_k, init='k-means++', max_iter=100, n_init=1)
model.fit(A1)
print("Top words per cluster:")
order_centroids = model.cluster_centers_.argsort()[:, ::-1]
terms = vectorizer.get_feature_names()
for i in range(true_k):
    print_cluster(i)


#after that we find the suggestion keyword from the searched keyword
```

Figure IX: Use k-means cluster to show recommendation for searched keyword

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Experimental Setup

We have implemented an efficient memory-model based collaborative filtering approach and achieved fast recommendation approach. In the end of this chapter, one will understand the reason behind choosing this proposed model and its function to this project. We can find out the best model with the best accurate result by using different types of datasets. I hope future researchers will follow and perform their research on new data.

4.2 Experimental Results and Analysis

The result of the recommendations from memory and model based CF approach are shown in Figure X, Figure XI and Figure XII respectively.

```
['0205616461',
 '0762451459',
 '1304139212',
 '130414089X',
 '130414643X',
 '1304146537',
 '1304168522',
 '1304174867',
 '1304174905']
```

In [28]: show_recommendations("spray paint")

```
Cluster 2:
 concrete
 paint
 wood
 vary
 product
 ft
 coating
 stake
 use
 water
```

Figure X: User selected item recommendation   Figure XI: Search keyword recommendation

```
Top words per cluster:
Cluster 0:
 cutting
 blade
 pruner
 saw
 tree
 pole
 blades
 head
 pruning
 easy
Cluster 1:
 power
 volt
 free
 lbs
 unit
 tool
 led
```

Figure XII: Clusters of k-means algorithm

4.3 Discussion

At the end of the experiment, we can decide that the utility matrix and clustering technique are more time consuming and better. And hopefully, this research paper will help the student and the researchers who wants to research and know more on this topic.

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

5.1 Impact on Society

We all know that recommendation system is huge popular now and it has been used in variety of areas' not only just e-commerce but also in the field of music, movie, social media platforms, news, books etc. Recommendation systems are also popular for restaurants and online dating and it have also been popular for explore research articles and experts, collaborators and financial services. Recommendation systems use an algorithm that help users to discover items they might have like.

In recommendation system, recommender algorithm use demographic data like user name, search history, user age, name of the website they visit, for their purchase algorithm might guess their income also that is very stressful but as we think of its positive site, we now live in 21'st century and the era of artificial intelligence where a machine can think like a human and make our life simple. So, it is considered that we need to advanced our life in this way where a machine can recommend what we like next.

Recommendation system also paved the way in our country where e-commerce site popularity is over the sky. So an efficient, fast and accurate recommender system will help those websites to gain its relevance from our country people and also gain huge revenue which lead our country to economic success.

5.2 Impact on Environment

We all know that recommendation system is huge popular now and it has been used in variety of areas' not only just e-commerce but also in the field of music, movie, social media platforms, news, books etc. Recommendation systems are also popular for restaurants and online dating and it have also been popular for explore research articles and

experts, collaborators and financial services. Recommendation systems use an algorithm that help users to discover items they might have like.

In our daily life, everything we use surrounded by recommendation system, when we use social media platform like Facebook, Instagram, LinkedIn etc. This platform suggests us many videos and post which we like previously or view the videos or post. They only suggest those videos or post which we watch daily now and also suggest friends where we have some common friend with suggested friends.

In music recommendation system, they suggest us music as per our listening history and like history. That help us without any doubt. In movie also recommender system recommends us movie through our watch history, key search history and trending movies where we can easily find the movies we like.

In news and books recommendation system, they also use our search history and reading history for suggest us any news and books. Which helps us discover new news and books or sorting through the large choice list.

Recommendation system also helps us with choose restaurants, showing popular research articles, their collaborations and many financial services which we needed.

There was a trend "rich get richer" which is somehow true for recommendation system where popular products get much more popularity after recommendation and sometimes recommendation systems unable to recognize novel items that have not been discovered which is really a bad impact.

5.3 Ethical Aspects

Privacy Intrusion

If we see in recommendation system, algorithm used user demographic data to perform recommendation for the user which is ethically not right because these data are a user's

personal information. But again if we think of it then it's basically help us to choose an item from a list of things.

Transparency

Recommendation system used personal information to done their filtering but the information about how their filtering is done is not made publically available. Due to lack of the transparency it is impossible to gain full understanding in how the data is gathered and used. A framework named RRI suggests that the internet research community should make user-friendly recommendation system testing kits available so that the general public is evaluate for themselves if their profile personalization used by the RS or not.

Behavior Manipulation

Concern about lack of transparency and use of "filter bubbles" is the question of how much user behavior is manipulated by the use of personalize filtering. Behavior manipulation is unavoidable since people will choose the first items from a list more often further down. It is impossible to place all item at the top of the list; it is the act of behavior manipulation.

5.4 Sustainability Plan

We often think that we will be launched a startup in e-commerce. When we thinking about what we should choose as our thesis topic then somehow this topic came to our mind. So we have to make a sustainable plan for our business that will be equipped our business and work up to date and efficient.

First we learn about what sustainability is all about. How can we use our resource wisely, how can we manage our financial, social and environmental impact and opportunities, how can we set our business in long run and we have to forget about old talking?

Secondly we study about issues like global warming, energy, fuel crisis and ecosystem for set up our goal. Then we talk with some of our friends so that we can improve our plan.

Our vision is to use this recommendation system in our own start-up business so that we can provide a better, fast and efficient RS to the user and improve it as much as we can so that user will happy. We will develop a review option so that we can ask user about our service and we can change it accordingly.

CHAPTER 6

SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH

6.1 Summary of the Study

In this paper, we studied how to apply utility matrix, find correlation and recommendation between the common user to item relationship and also find the suggestion words for searched keywords using k-means algorithm where we find clusters of different words. Many other methods have been introduced in recommendation systems for recommend items. Furthermore, utility matrix and k-means algorithm was really helpful.

6.2 Recommendations

It is recommended:

- that appropriate approach selection is a significant part of any work.
- that if any people wanted to know deeply about RS system then they need to understand all the approached clearly because some of the approaches are confusing.
- that an increase of data diversity will help to find better result.
- that avoid behavior manipulation and privacy intrusion in the recommendation systems.
- that set up a goal and sustainable plan about the work which will helpful for people and country.

6.3 Conclusion

E-commerce websites are evolving and RS are playing major role in the e-commerce in our day to day life. At the same time recommendation system are facing many problems such as cold start, data sparsity, gray sheep which can affect the accuracy and efficiency of

the RS. In future, e-commerce will increase its revenue from customer and help to gain economic success and it also make customer life easy. The proposed system helps the e-commerce websites to avoid the problem like cold start, gray sheep and it also helps the company to get the profits for their company. So we can say that using memory-model based CF contributes to overcoming the drawbacks such as new customer, new product and sparsity problem.

6.4 Implication for Future Research

We proposed this model based on utility matrix and clustering technique for amazon dataset. But we want to use other techniques like content based, constraint based and hybrid RS approach so that we can compare this work approaches with other approaches and we want to use our own country data for research purpose because our main goal is to use this models in our own start-up. Using our country data will help our start-up to give accurate and efficient recommendation.

REFERENCES

[1]Recommendation system Wikipedia, available at << https://en.wikipedia.org/wiki/Recommender_system>>, last accessed on 01-05-2021 at 3:00pm

[2]K-means cluster algorithm, available at<<https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1>> , last accessed on 01-05-2021 at 2:00pm

[3]Utility matrix, available at<<https://towardsdatascience.com/math-for-data-science-collaborative-filtering-on-utility-matrices-e62fa9badaab >>, last accessed on 01-05-2021 at 2:00pm

[4] K. RaviKanth1, K. ChandraShekar2, K. Sreekanth3 and P. Santhosh Kumar4," Recommendation System for E-Commerce by Memory Based and Model Based Collaborative Filtering", ICSCPR (August 2020)

[5] Ayad R. Abbas1 and Shaymaa Ashor2," Designing Personalized Recommendation in E-Commerce Site Based on Content-Based and Collaborative Filtering", Journal of Al-Nahrain University, Vol.20 (2), (June 2017) pp.142-152

[6] Rong Hu and Pearl Pu," Enhancing Collaborative Filtering Systems with Personality Information", RECSYS (October 2011)

[7] A. Felfernig and R. Burke," Constraint-based Recommender Systems:Technologies and Research Issues", ICEC (June 2008)

[8] Pallavi Paunikar, Payal Modake, Shamali Wagh, Prof.Prerna Mishra," Implementation of User Based Collaborative Algorithm in Real Estate Property", IJIRCCE (January 2017) Vol. 5

[9] Yong Soo Kim, Bong-Jin Yum*, Junehwa Song, Su Myeon Kim," Development of a recommender system based on navigational and behavioral patterns of customers in e-commerce sites", ELSEVIER (2005), Expert Systems with Applications 28 (2005) 381–393

[10] JohnBordrick Nnadozie1," Collaborative Filtering Algorithm for Recommender Systems", Researchgate (January 2021)

[11] Shivani Saini(B), Sunil Saumya(B), and Jyoti Prakash Singh(B)," Sequential Purchase Recommendation System for E-Commerce Sites", CISIM (January 2017), Springer International Publishing pp. 366–375

[12] Sanjeevan Sivapalan1, Alireza Sadeghian2, Hossein Rahanam3 and Asad M. Madni," Recommender Systems in E-Commerce", Researchgate (August 2014)

©Daffodil International University

[13] Ssvr Kumar Addagarla, Anthoniraj Amalanathan," a visual similar product recommender system for E-commerce", Indonesian Journal of Electrical Engineering and Computer Science Vol. 22, (April 2021), pp. 563-570, ISSN: 2502-4752

PLAGARISM REPORT

turnitin **Turnitin Originality Report**
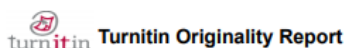
thesis by Tariq - Al - Jamil

From aaa (Spring 2021 Defense)

Processed on 26-May-2021 22:52 +06
ID: 1594690486
Word Count: 4528

Similarity Index	Similarity by Source	
22%	Internet Sources:	14%
	Publications:	9%
	Student Papers:	14%

sources:

1   3% match (student papers from 30-Jun-2020)
Submitted to Daffodil International University on 2020-06-30

2   2% match (student papers from 15-May-2020)
Submitted to University of Ibadan on 2020-05-15

3   2% match (publications)
"Ethics of Personalized Information Filtering", Lecture Notes in Computer Science, 2015.

4   2% match (Internet from 22-Nov-2020)
https://www.javatpoint.com/k-means-clustering-algorithm-in-machine-learning

5   1% match (Internet from 25-Mar-2021)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/5250/162-15-
7955%20%287_%29.pdf?isAllowed=y&sequence=1

6   1% match (Internet from 18-Jul-2020)

http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/3994/P15461%20%2823_%29_.pdf?
isAllowed=y&sequence=1

7   1% match (student papers from 04-Apr-2018)
Submitted to Daffodil International University on 2018-04-04

8   1% match (publications)
Jatin Sharma, Kartikay Sharma, Kaustubh Garg, Avinash Kumar Sharma. "Product
Recommendation System a Comprehensive Review", IOP Conference Series: Materials
Science and Engineering, 2021

29