GALGOTIAS UNIVERSITY

(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

RECOMMENDATION SYSTEM

A Report for the Evaluation 3 of Project 2

Submitted by

SARTHAK MALHOTRA

(1613101630,16SCSE101185)

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING

Under the Supervision of

Dr. Kumar Dilip
Assistant Professor
Mr Dhamodharan D

(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

SCHOOL OF COMPUTING AND SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

Certified that this project report "RECOMMENDATION SYSTEM" is the bonafide work of "SARTHAK MALHOTRA (1613101630)" who carried out the project work under my supervision.

SIGNATURE OF HEAD

Dr. MUNISH SHABARWAL
PhD (Management), PhD (CS)
Professor & Dean,
School of Computing Science & Engineering

SIGNATURE OF SUPERVISOR

Dr. KUMAR DILIP
Assistant Professor
School of Computing Science & Engineering

TABLE OF CONTENTS

Abstract

On the web, where there exists a cyber-ocean of information, there is need to filter, prioritize and efficiently deliver relevant information. Here's where the concept of Recommendation Engine steps in. Our proposed model has the ability to recommend products to the individual users on the basis of the earlier purchase history and shopping experience The approach employed, contains the use of Restricted Boltzmann machine (RBM), our proposed approach have shown significant improvements over another baseline methods.

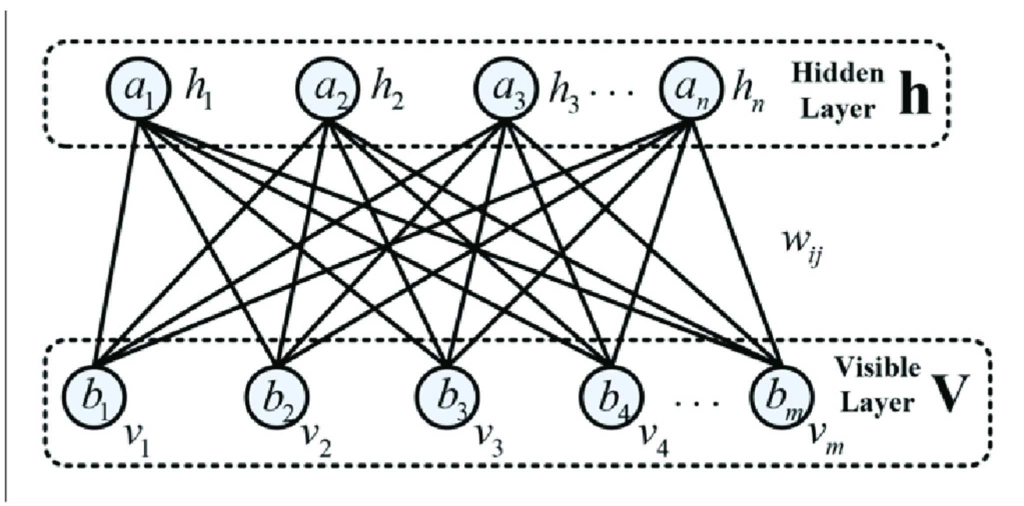


Fig-1 Architecture of Restricted Boltzmann Machine (RBM)

Introduction

In most general terms, **Recommendation Engines** are defined as a subclass of **information filtering system** that seeks to predict the "rating" or "preference" a user would give to an item. Thus, making the experienced more personalized and lucid. These items can be products, books, movies, restaurants etc. These preferences are being predicted using two approaches, first content-based approach which involves characteristics of an item and second **collaborative filtering approach** which takes into account user's past behavior to make choices.

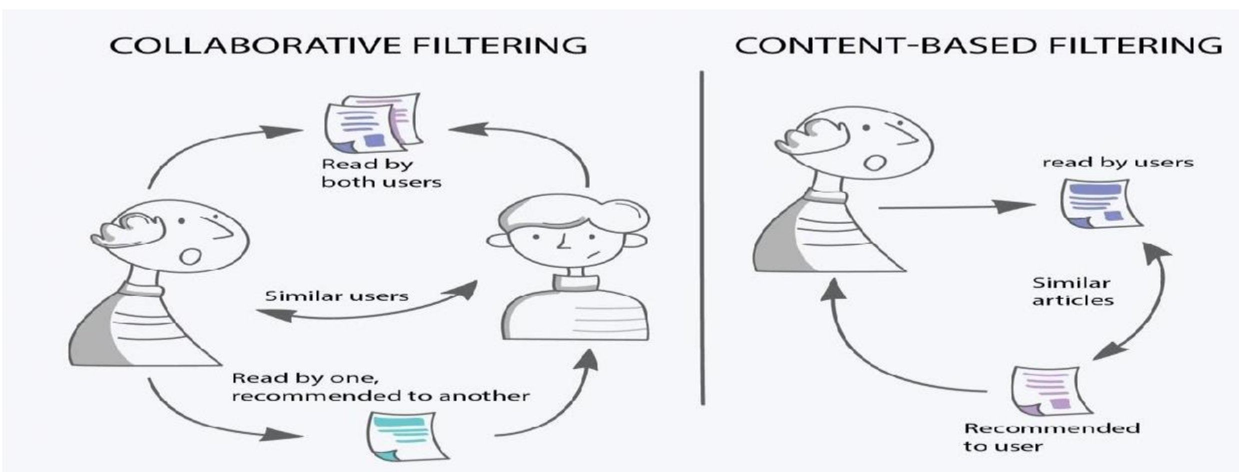


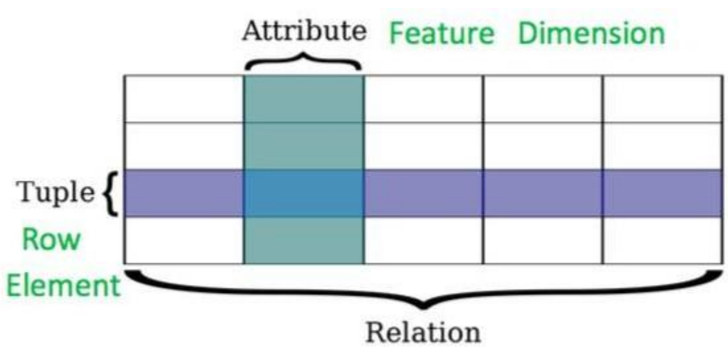
Fig-2. Collaborative and Content- Based filtering.

Purpose:-

The recommender system involves various techniques:- association mining, collaborative filtering and content filtering are the three widely employed methods for strong impact using search engines.

- **Restricted Boltzmann Machine (RBM):-** A restricted Boltzmann machine (RBM) is a type of artificial neural network. This type of generative network is useful for filtering, feature learning and classification, and it employs some types of dimensionality reduction to help tackle complicated inputs. With the help of RBM, we pass all the stock code of all the customers in the visible layer and then reconstruct the User-Item matrix and thus, find the recommended product for a particular user on the basis of the purchase history.

- **Association Rule Mining** is one of the ways to find patterns in data. It finds:
    - features (dimensions) which occur together
    - features (dimensions) which are “correlate



$$\mathbf{x_i} = (x_{i1}, x_{i2}, ..., x_{ip}) \in \mathbb{R}^p$$

$$X \in \mathbb{R}^{n \times p}$$

Proposed Model:-

- **Collaborative filtering (CF)** is a popular recommendation algorithm that bases its predictions and recommendations on the ratings or behavior of other users in the system. The fundamental assumption behind this method is that other users' opinions can be selected and aggregated in such a way as to provide a reasonable prediction of the active user's preference.
    - There are two categories of CF:
    - **User-based**: measure the similarity between target users and other users.
    - **Item-based**: measure the similarity between the items that target users rates/ interacts with and other items.

Assume there are *m* users and *n* items, we use a matrix with size *m*n* to denote the past behavior of users. Each cell in the matrix represents the associated opinion that a user holds. For instance, M{i, j} denotes how user i likes item j. Such matrix is called **utility matrix**.

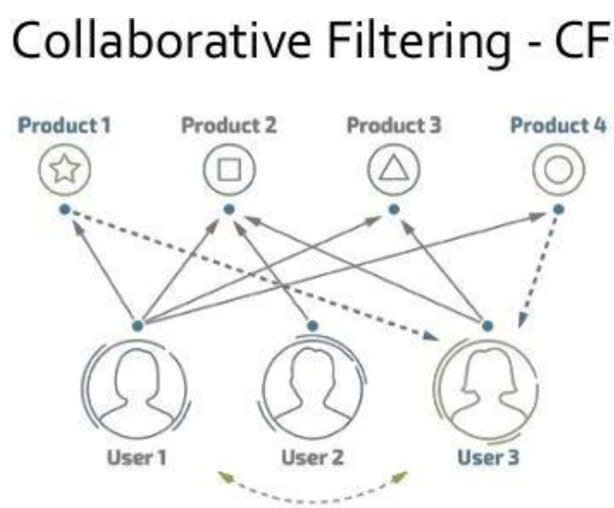


Fig3:-Collaborative Filtering

User-Based:-

There are two options, Pearson Correlation or cosine similarity. Let us{i, k} denotes the similarity between user i and user k and v_{i, j} denotes the rating that user i gives to item j with v_{i, j} = ? if the user has not rated that item. These two methods can be expressed as the followings:

$$u_{ik} = \frac{\sum_j (v_{ij} - v_i)(v_{kj} - v_k)}{\sqrt{\sum_j (v_{ij} - v_i)^2 \sum_j (v_{kj} - v_k)^2}}$$

Fig4:-Pearson Correlation

$$\cos(u_i, u_j) = \frac{\sum_{k=1}^{m} v_{ik} v_{jk}}{\sqrt{\sum_{k=1}^{m} v_{ik}^2 \sum_{k=1}^{m} v_{jk}^2}}$$

Fig5:-Cosine Similarity

Both measures are commonly used. The difference is that Pearson Correlation is invariant to adding a constant to all elements.

Item-Based:-

The item-based CF recommends items based on their similarity with the items that the target user rated. Likewise, the similarity can be computed with Pearson Correlation or Cosine Similarity.

This method is quite stable in itself as compared to User based collaborative filtering because the average item has a lot more ratings than the average user. So an individual rating doesn't impact as much.

To calculate similarity between two items, we looks into the set of items the target user has rated and computes how similar they are to the target item i and then selects k most similar items. Similarity between two items is calculated by taking the ratings of the users who have rated both the items and thereafter using the cosine similarity function mentioned below:

$$w_{ij} = \mathrm{sim}(i,j) = \frac{\sum_{u \in U_i \cap U_j} \hat{r}_{ui} \hat{r}_{uj}}{\sqrt{\sum \hat{r}_{uj}} \sqrt{\sum \hat{r}_{uj}}}$$

Fig6:- Cosine Similarity Function

User-based recommendations vs Item-based recommendation:-

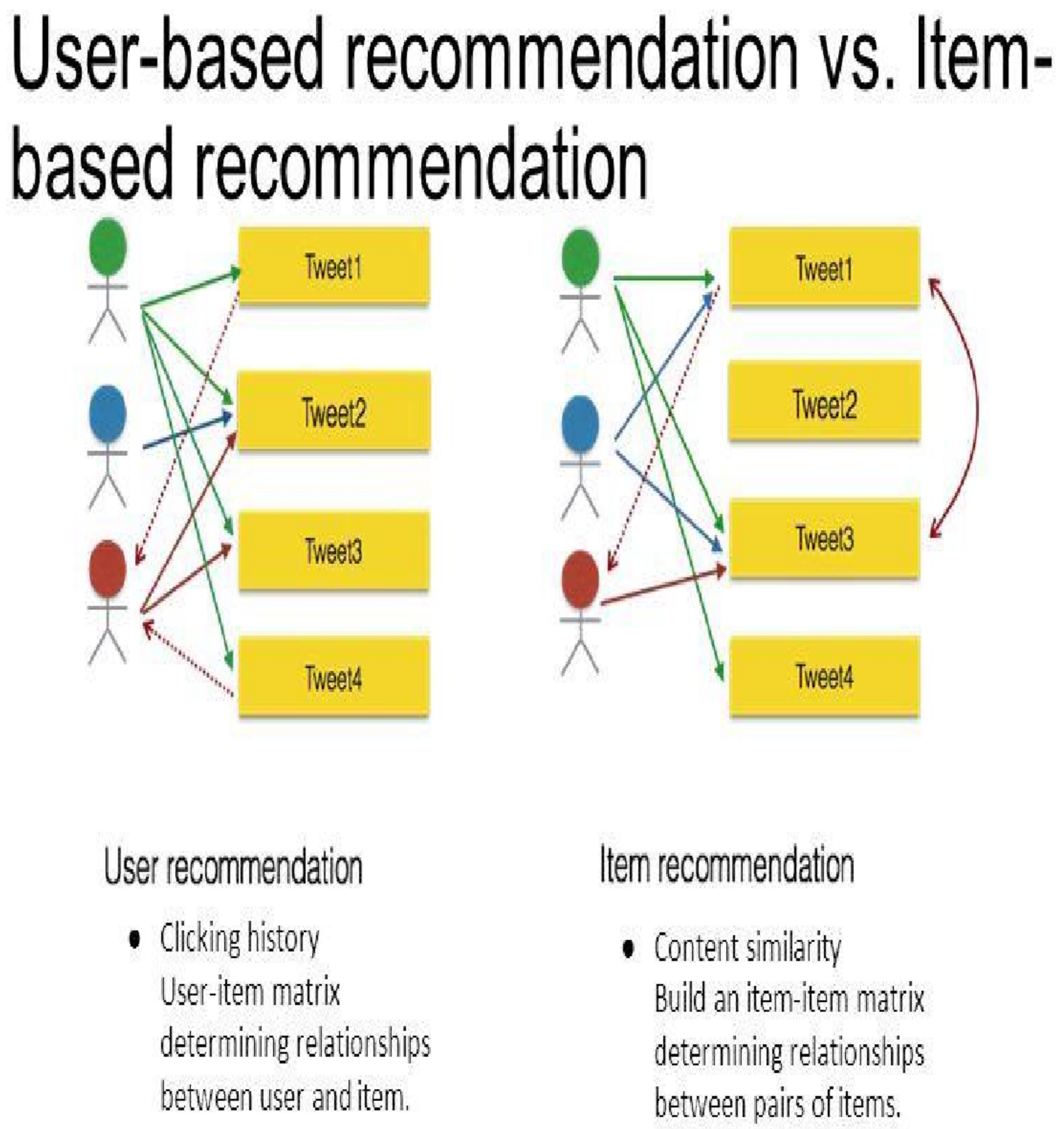


Fig7:- User-based recommendations vs Item-based recommendation:-

- **Content-Based Filtering:-**

Content-based filtering, also referred to as cognitive filtering, recommends items based on a comparison between the content of the items and a user profile. The content of each item is represented as a set of descriptors or terms, typically the words that occur in a document. The user profile is represented with the same terms and built up by analyzing the content of items which have been seen by the user.

Several issues have to be considered when implementing a content-based filtering system. First, terms can either be assigned automatically or manually. When terms are assigned automatically a method has to be chosen that can extract these terms from items. Second, the terms have to be represented such that both the user profile and the items can be compared in a meaningful way. Third, a learning algorithm has to be chosen that is able to learn the user profile based on seen items and can make recommendations based on this user profile.

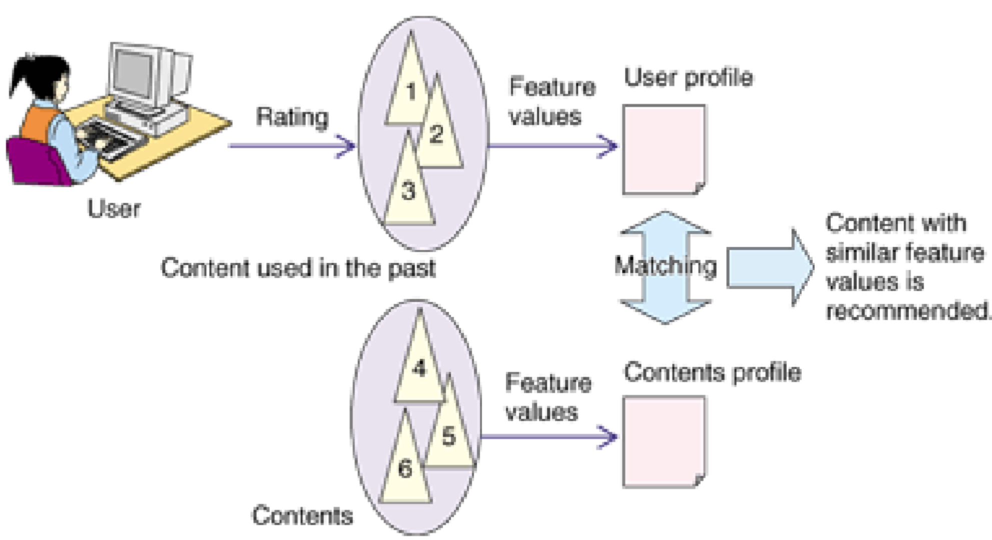


Fig8:-Content-Based Filtering

Restricted Boltzmann Machine (RBM)

A restricted Boltzmann machine (RBM) is a type of artificial neural network. This type of generative network is useful for filtering, feature learning and classification, and it employs some types of dimensionality reduction to help tackle complicated inputs. With the help of RBM, we pass all the stock code of all the customers in the visible layer and then reconstruct the User-Item matrix and thus, find the recommended product for a particular user on the basis of the purchase history.

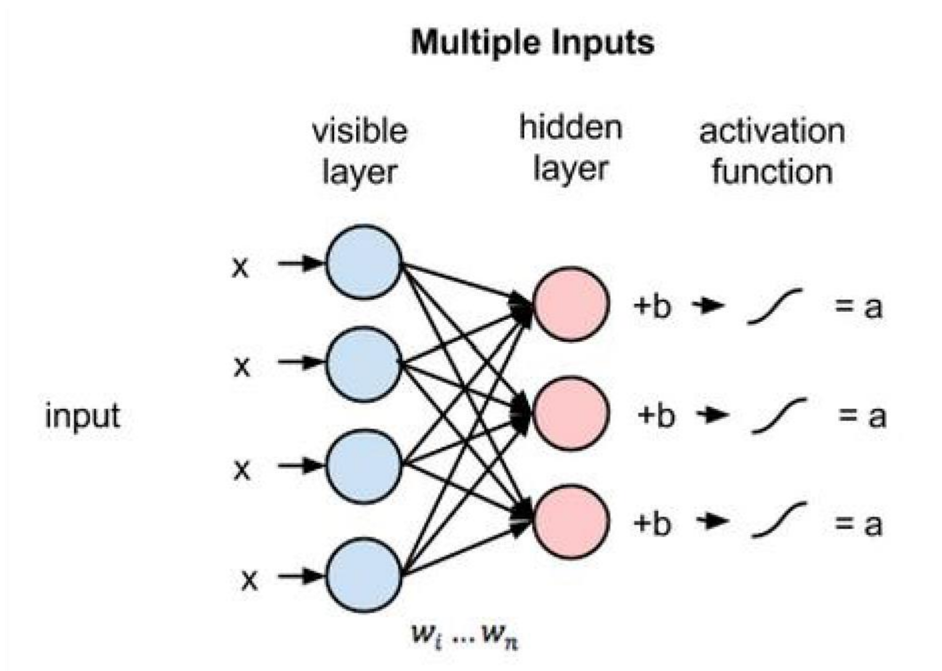


Fig9:- Architecture of Restricted Boltzmann Machine (RBM)

Data Sparsity

Data Sparsity:- Data sparsity is a crucial problem in recommender systems. Data sparsity plays an important role in recommendation systems. Data sparsity problem is discussed in collaborative filtering approach. In the work, it is concluded that the data sparsity negatively affects in the recommendations which is provided by collaborative filtering. For instance, in newspaper domain, some recent news is just rated by few people. In such condition, even if the news is highly important or crucial for people and have high rating only by few people, the chances of being recommended to other users is very slim.

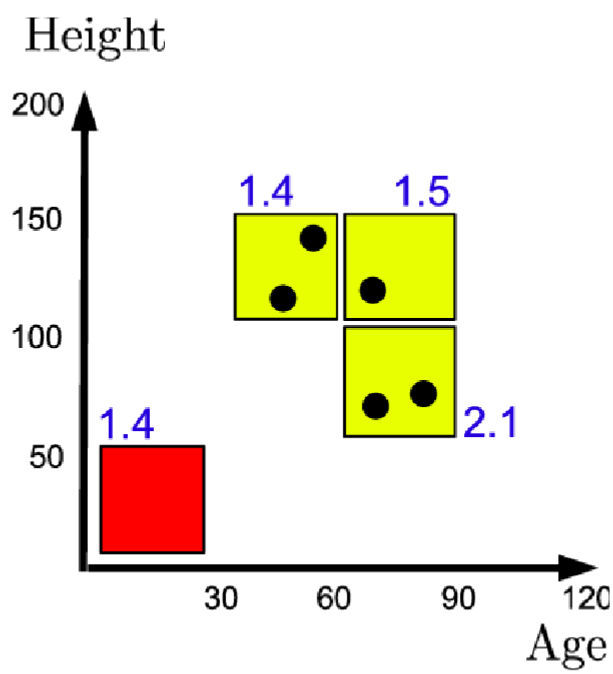


Fig10:- Data Sparsity

ARCHITECTURE DIAGRAM

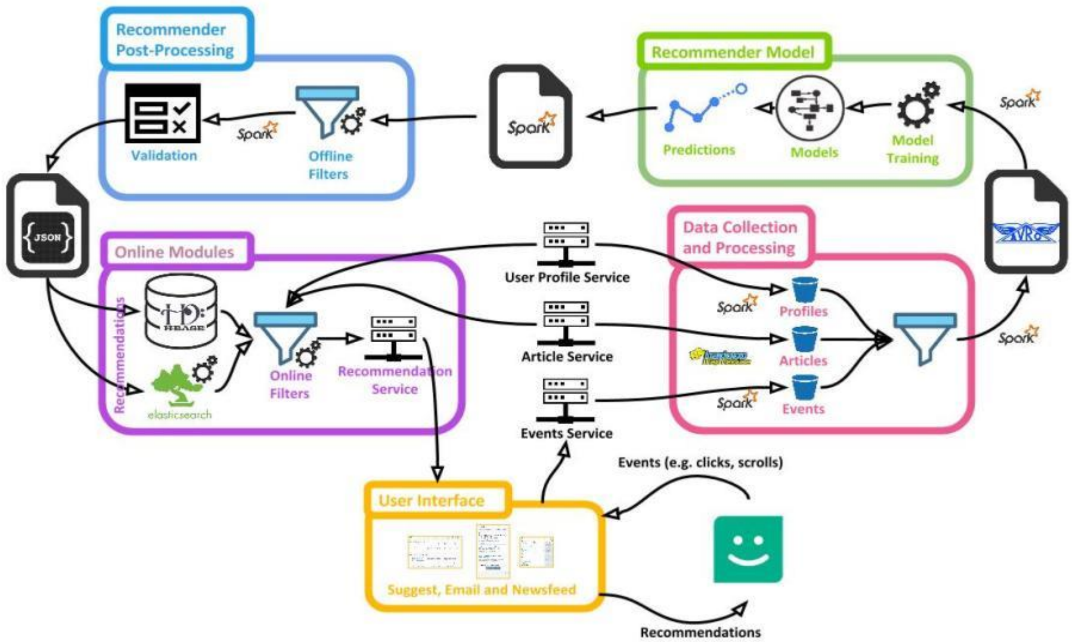


Fig 11:- Architecture of Recommendation System

Motivations and Scope

Many businesses nowadays embed recommendation systems in their web sites, in order to study the tastes of their customers, and achieve some business objectives. Among other objectives we have: the increase of traffic to their web site, the elaboration of marketing policies tailored to their customers' tastes, or simply the promotion of a given product.. Global world labor corp. since this is the business we are referring to, is an international company whose mission is to build a network between employment services worldwide, in order to promote the mobility of workers around the globe. The main services provided by the Global world labor corp. (GWLC) are the following: Information about living and working conditions in the countries that are part of its network. Assistance to workers and employers on practical, legal, and administrative matters. Recruitment / placement to the benefit of workers and employers. In the current economic situation, in which all the countries worldwide are recovering (or trying to recover) from the crisis, skilled, adaptable and mobile labor force is needed more than ever. Global world labor corp. intends to play a major role in that regard by providing innovative and adequate solutions, and services to the labor market through a new web portal.

Problem Statement

I summarize three ways of problem statement in the recommendation system: Rating Prediction, Sequence Prediction.

- Rating Prediction:- Given a user and his preference score on item 1 to item n, what is his preference score on item n+1? The preference score here can be either explicit feedback (ratings, reviews)or implicit feedbacks(click, views, purchases). Before modeling, quantify the feedbacks with some techniques like TFIDF for text, feature-importance for implicit events, or normalization and standardization. Machine Factorization (MF)is a classical type of models to solve this type of problem. Given a MxN matrix, where M is # unique users and N is # unique items, Rij is the ratings for user ion item j, MF can decompose the MxN matrix into two small matrix: MxK and KxN, where K is # dimension of latent factors. However, one possible disadvantage for this is that the matrix can be pretty sparse because one user can only interact with several items. Bonus: MF can be implemented as a Factorization Machines(FM) model. FM takes features' second-order interaction into consideration. You can consider MF as an FM model with only user and item features. However, you can add more features to FM models.

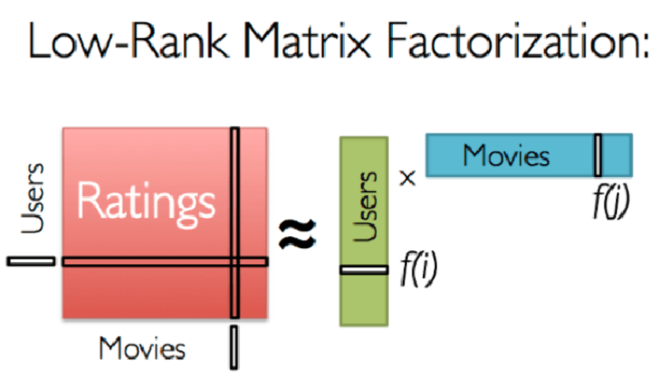


Fig12:-Low-Rank Matrix Factorization

Sequence Prediction

In real work, it performed very well compared with rating prediction-based models. The idea behind this type of problem is pretty straightforward, learning the evolution of user affinity's change in the timeline from the shopping sequence. The good news is that it's very similar to NLP tasks, and we can apply NLP techniques like embeddings, attentions into the models. Basically, the embeddings can better represent a user or item information in the model, and the attention can decide what in the previous sequence should play a more important role in the future sequence.

One naive approach is to use time-decay on the product embeddings to differentiate the importance of more recent events and more previous events. However, the RNN(LSTM/GRU) based models could do better if computing power is enough.

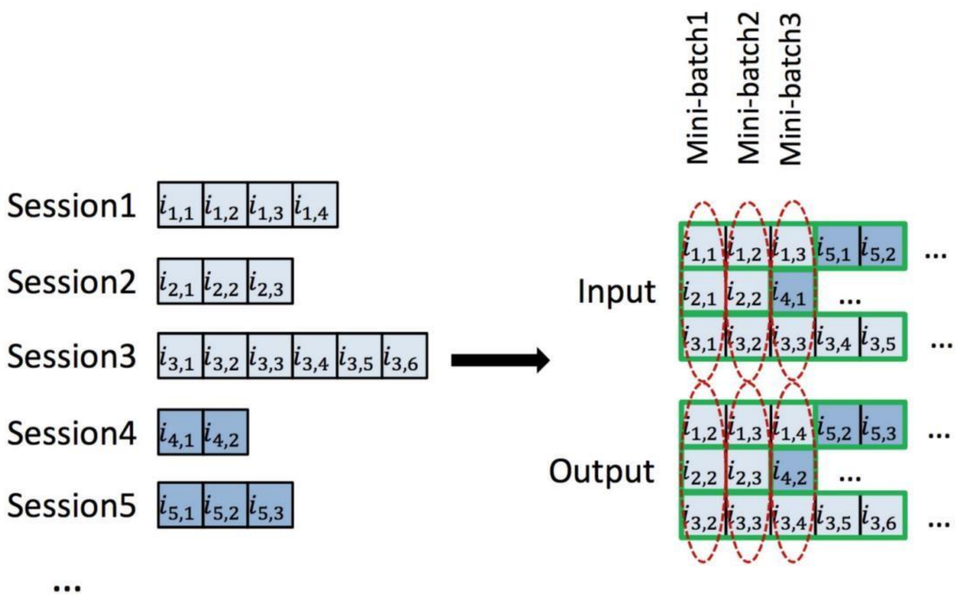


Fig13:- Sequence Prediction

Experimental Results and Discussion

The recommendation engine was built by implementation of RBM, which includes the following formulae-

$-(\partial \log p(v) / \partial W_{ij}) = E_v[p(h_i \mid v).v_j] - v^{(i)}_j . \, \mathit{sigm}(W_i \, . \, v^{(i)} + c_i)$ ---- eqn (1)

$-(\partial \log p(v) / \partial c_i) = E_v[p(h_i \mid v) - \mathrm{sigm}(W_i \, . \, v^{(i)})$ ---- eqn (2)

$-(\partial \log p(v) / \partial b_j) = E_v[p(v_j \mid h)] - v_j^{(i)}$ ---- eqn (3)

Result:-Movie Recommendation System

The movies dataset is used to show the process of recommendation system with the help of collaborative filtering and content based filtering to provide rating to the movies.

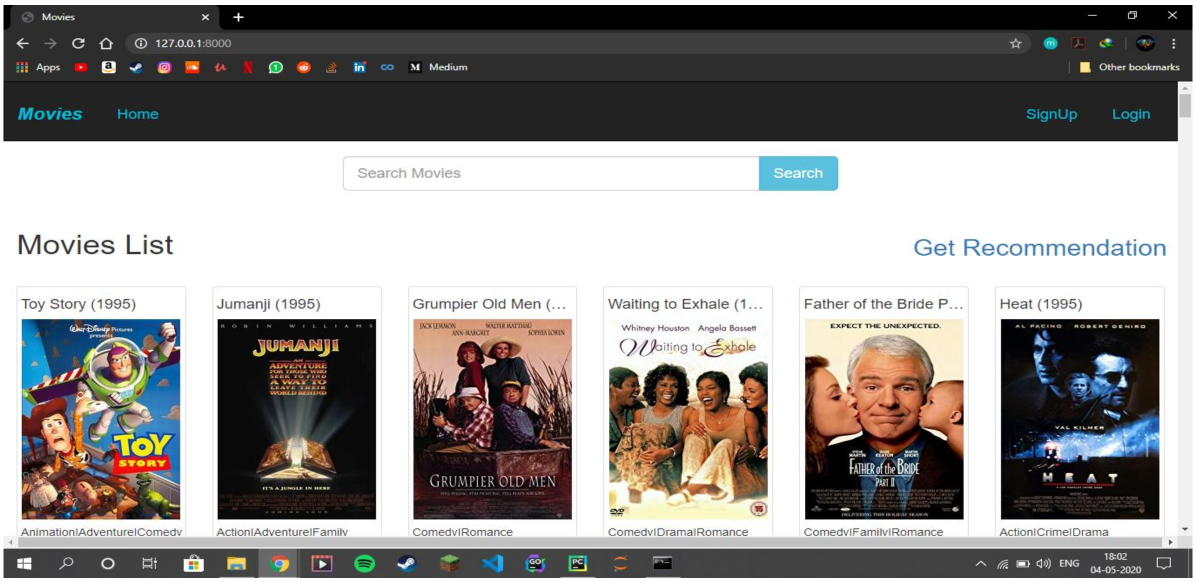


Fig14:-Screenshot of the Movie Recommendation System

Result:-Movie Recommendation System

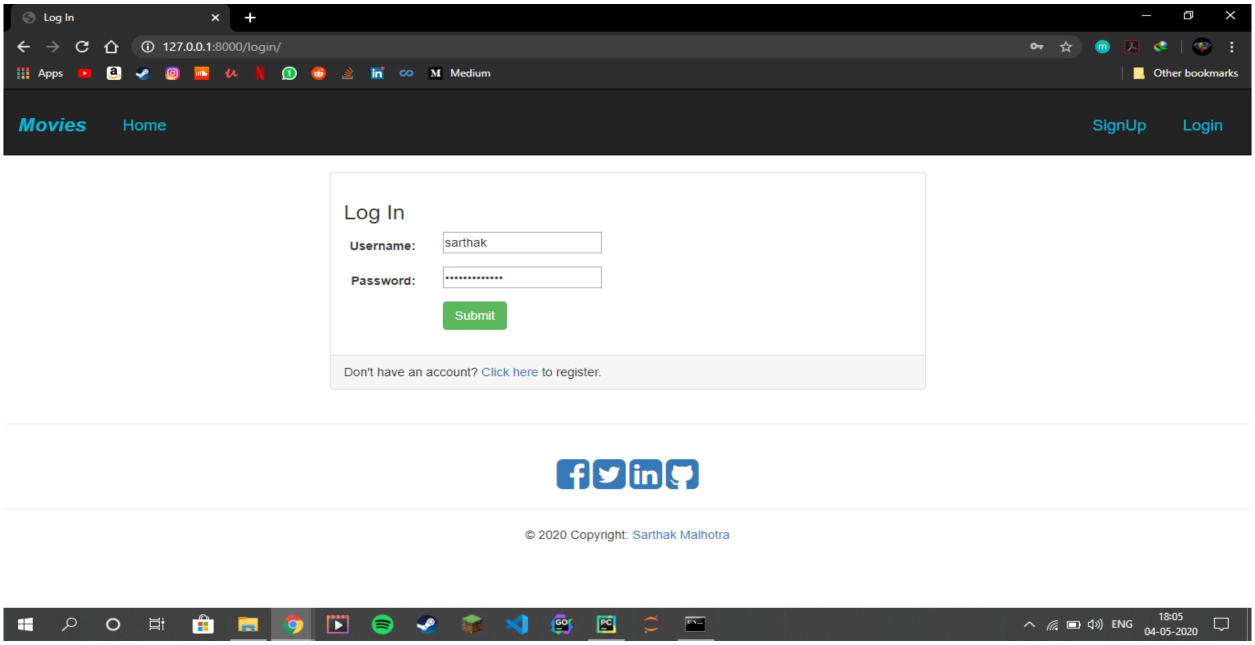


Fig15:-Login Process through the Homepage

Result:- Movie Recommendation System

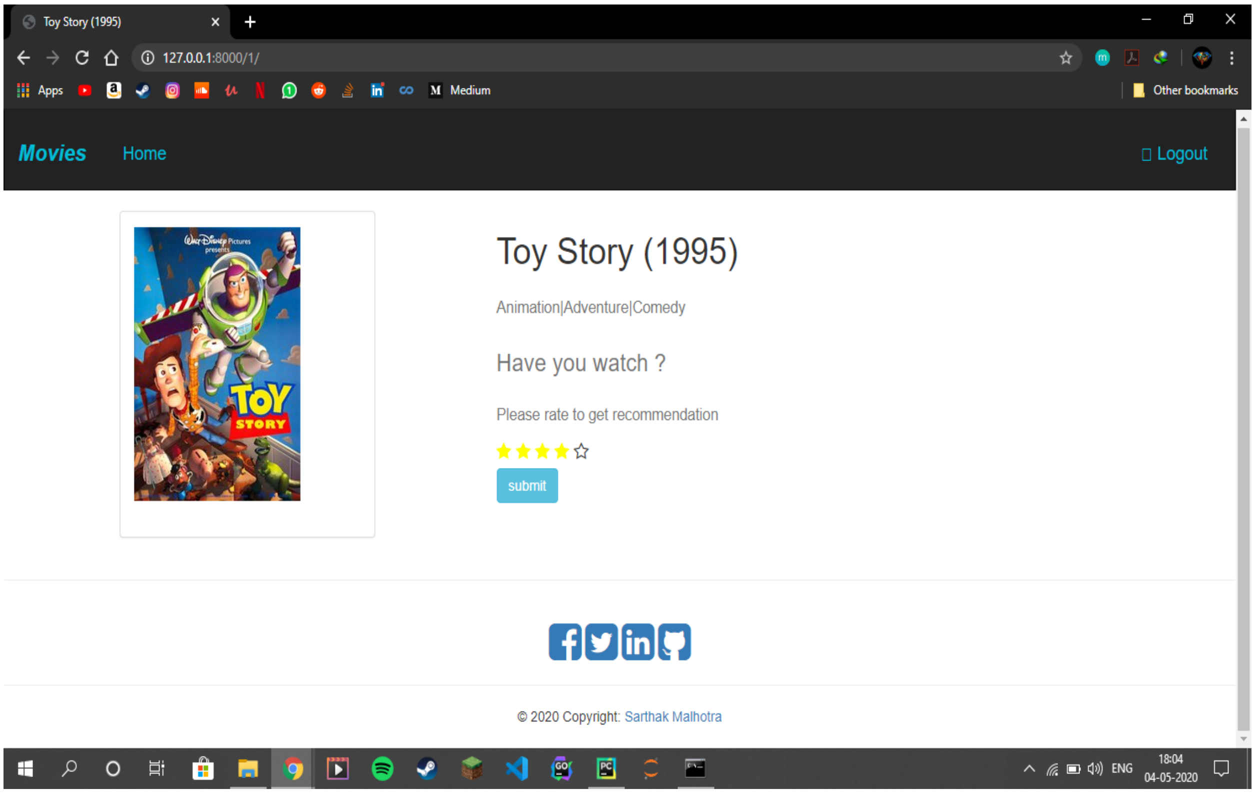


Fig16:- Process of rating to a particular movie after logging in process

Result:-Movie Recommendation System

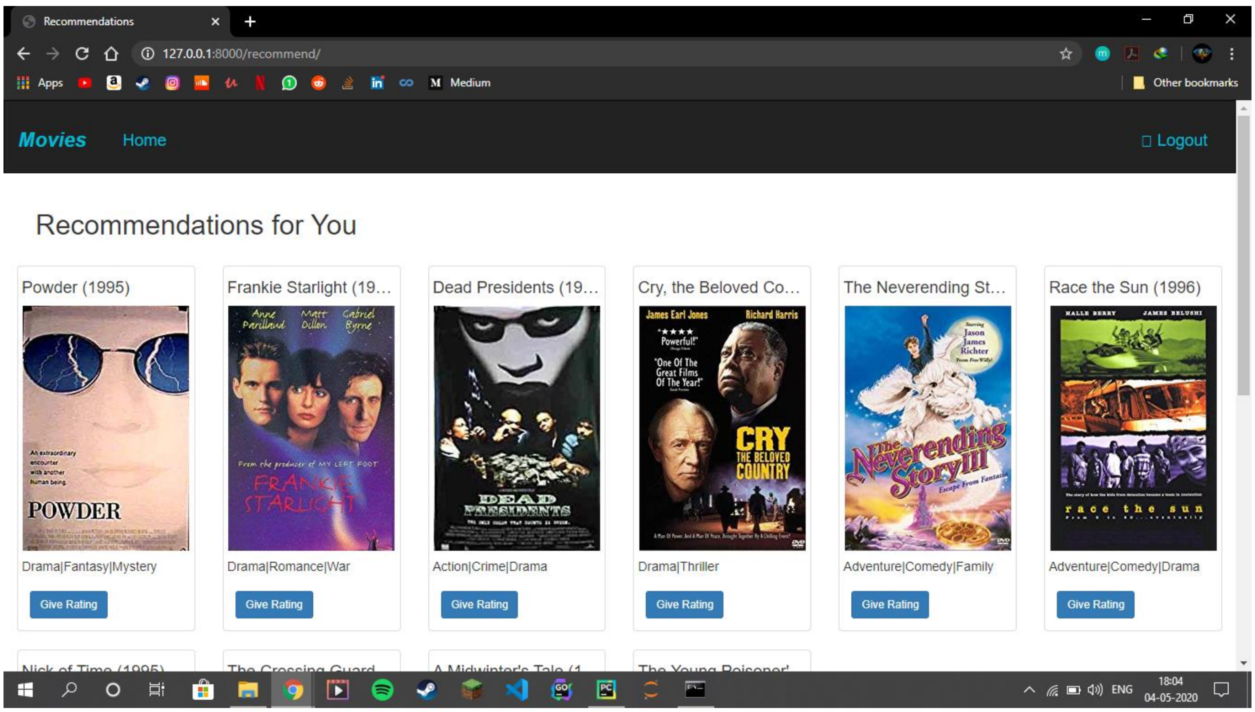


Fig17:-More Recommendations of the Movies after logged in process

Conclusion

- Recommender systems are rapidly becoming a crucial tool in E-commerce on the web for a personalized shopping experience. New technologies are needed that can dramatically improve the scalability of recommender systems.
- In this project, we presented and experimentally evaluated a recommender systems with the help of restricted Boltzmann machine (RBM).
- Our results show that RBM-led technique hold the potential to scale to large data sets and at the same time produce high-quality recommendations of items for a particular user.