**MOVIE RECOMMENDATION SYSTEM**

*Yash Kulkarni, Abhishek Kuber, Soham Kulthe*

*November 2020*

**Introduction**

A recommendation system generates a compiled list of items in which a user might be interested, in the reciprocity of their current selection of item(s). It expands users’ suggestions without any disturbance or monotony, and it does not recommend items that the user already knows.

have rated highly Recommendation System framework plays a vital role in today‟s internet surfing, be it buying a product from an e-commerce site or watching a movie on some video-on-demand service. In our everyday life, we depend on recommendations given by other people either by word of mouth or reviews of general surveys. People often use recommender systems over the web to make decisions for the items related to their choice. Recommendation systems are software tools and techniques whose goal is to make useful and sensible recommendations to a collection of users for items or products that might interest them. In other words, the recommender system or recommendation systems belongs to a class of information filtering system that aims at predicting the “preference” or “rating” given to an item. For instance, the Netflix recommendation system offers recommendations by matching and searching similar users' habits and suggesting movies that share characteristics with films that users

**About the dataset**

MovieLens data sets were collected by the GroupLens Research Project at the University of Minnesota.

This data set consists of:

\* 100,000 ratings (1-5) from 943 users on 1682 movies.

\* Each user has rated at least 20 movies.

\* Simple demographic info for the users (age, gender, occupation, zip)

The data was collected through the MovieLens web site (movielens.umn.edu) during the seven-month period from September 19th, 1997 through April 22nd, 1998. This data has been cleaned up - users who had less than 20 ratings or did not have complete demographic information were removed from this data set.

u.data -- The full u data set, 100000 ratings by 943 users on 1682 items.

Each user has rated at least 20 movies. Users and items are numbered consecutively from 1. The data is randomly ordered. This is a tab separated list of

user id | item id | rating | timestamp.

**IDE used:**

Spyder

**Recommendation Systems:**

Types of Recommendation Systems

There are two types of recommendation systems:

* Content filtering recommender systems.
* Collaborative filtering based recommender systems.

**Content filtering**

A content based recommender works with data that the user provides, either explicitly (rating) or implicitly (clicking on a link). Based on that data, a user profile is generated, which is then used to make suggestions to the user. As the user provides more inputs or takes actions on the recommendations, the engine becomes more and more accurate.

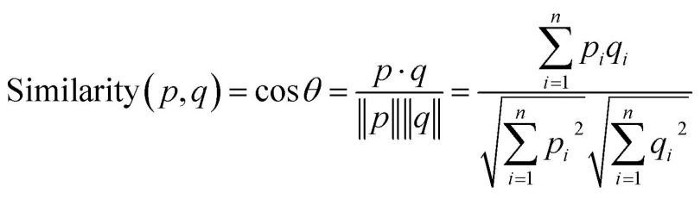
**Collaborative Filtering**

Collaborative filtering is a technique that can filter out items that a user might like on the basis of reactions by similar users.

It works by searching a large group of people and finding a smaller set of users with tastes similar to a particular user. It looks at the items they like and combines them to create a ranked list of suggestions.

**Cosine Similarity**

Cosine similarity is a method to measure the difference between two non zero vectors of an inner product space. It is a metric used to measure how similar two items are. Mathematically, it measures the cosine of the angle between two vectors projected in a multi-dimensional space. The output value ranges from 0–1. In which 0 means no similarity, where as 1 means that both the items are 100% similar.

****

**cosine** is an angle calculated between -1 to 1 where -1 denotes dissimilar items, and 1 shows items which are a correct match.

**cos p. q** — gives the dot product between the vectors.

**||p|| ||q||** — represents the product of vector’s magnitude

**K-Nearest Neighbors**

KNN algorithms use data and classify new data points based on similarity measures (e.g. distance function).

The KNN algorithm assumes that similar things exist in close proximity. In other words, similar things are near to each other.

This technique generates predictions based on similarities between different videos or movies or items.

Prediction for a user u and item i is composed of a weighted sum of the user u’s ratings for items most similar to i.



**Similar Movies Matrix**

It is a list of similar movies. We will use the movie\_index of the movie we have given as input movie\_user\_likes. The enumerate() method will add a counter to the iterable list cosine\_sim and return it in a form of a list similar\_movies with the similarity score of each index.

**Our Model**

We have illustrated the modelling of a movie recommendation system by making the use of content-based filtering in the movie recommendation system. The KNN algorithm is implemented in this model along with the principle of cosine similarity as it gives more accuracy than the other distance metrics and the complexity is comparatively too low.

Recommendations systems have become the most essential fount of a relevant and reliable source of information in the world of internet. Simple ones consider one or a few parameters while the more complex ones make use of more parameters to filter the results and make it more user friendly. With the inclusion of advanced deep learning and other filtering techniques like collaborative filtering and hybrid filtering a strong movie recommendation system can be built. This can be a major step towards the further development of this model as it will not only become more efficient to use but also increase the business value even further.

**Accuracy**

**References**

<https://www.kaggle.com/prajitdatta/movielens-100k-dataset>

<https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning>

<https://realpython.com/build-recommendation-engine-collaborative-filtering/#:~:text=Collaborative%20filtering%20is%20a%20family,type%20of%20collaborative%20filtering%20approach>

<http://www.grouplens.org/>