

Movie Recommendation System

190428996

FYP Topic Proposal

Theme: CM3015 Machine Learning and Neural Network

Project Idea 2: Deep Learning on a public dataset

Business Pain Point



Incorrect movie recommendation leads to a bad user experience



Customer might lose interest in the movie streaming websites



Unsubscribe to their streaming services or they will avoid the website



Leads to a decrease in user retention rate and revenue loss for company

What is needed

Vision

A Deep Learning Model that can accurately recommend movies to the users is needed.

Value

Having an efficient movie recommendation system → movie streaming websites will have a competitive advantage over the other company → enhance user experience → This policy of recommendation system is really helpful in giving optimum results to an application profitability and to make the organisation more connected.



Project Overview



Machine Learning

Deep Learning on a given dataset



Dataset Used

Using the MovieLens

Data contains ratings, userID, movie ID, movie tags, movie title, movie genre and timestamp



Objective

Create an efficient movie recommendation model using deep neural network (DNN)



Metrics

Evaluation metrics

Root Mean Square Error (RMSE) will be used to evaluate the performance of the recommendation model.

Movie recommendation with Deep Learning

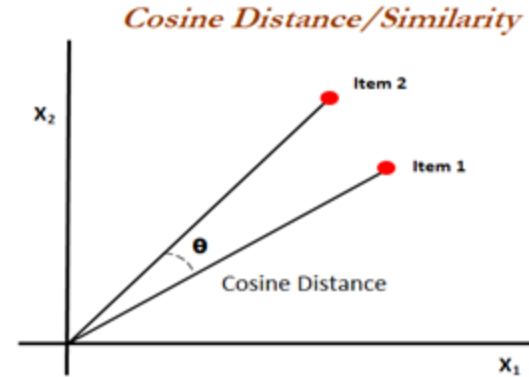
Why use this?

Deep learning-based recommender systems outperform traditional ones due to their capability to process non-linear data. Non-linear transformation, representation learning, sequence modeling, and flexibility are the principal benefits of applying DL for recommendations.

State-of-the-art Review 1: Content-based filtering with cosine similarity

Cosine similarity will be used to find the similarity between 2 non-zero vectors of an inner product space that measures the cosine of the angle between them.

Smaller the angle, higher the similarity between the two movies. Higher chances for it to get recommended.



Advantages

- Able to capture the angle of data objects
- Even if 2 data objects are far apart by Euclidean distance, they can still have a small angle.



Disadvantages

- Magnitude of vectors is not taken into account, just the direction.

References:

<https://www.oreilly.com/library/view/statistics-for-machine/9781788295758/eb9cd609-e44a-40a2-9c3a-f16fc4f5289a.xhtml>

<https://www.geeksforgeeks.org/cosine-similarity/#:~:text=Advantages%20%3A,the%20angle%2C%20higher%20the%20similarity.>

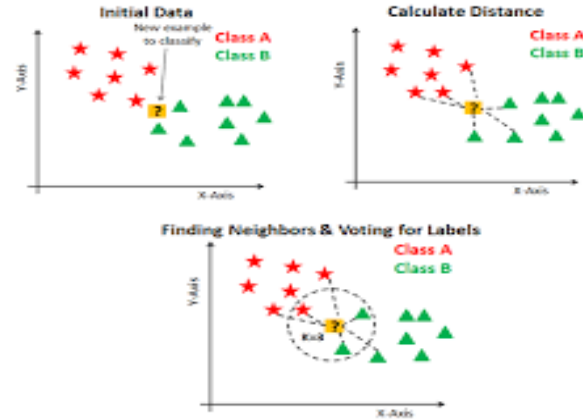
State-of-the-art Review 2: Collaborative-based filtering with K Nearest Neighbours

What is KNN?

KNN is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point.

How it works?

KNN will calculate the “distance” between the target movie and other movies in database and ranks its distances to return the top K nearest neighbor movies as the most similar movie recommendations



Advantages

- Easy to implement
- Most common method
- Data can be easily added



Disadvantages

- Does not work well with large dataset
- Sensitive to noisy data

References:

<https://www.ibm.com/sg-en/topics/knn>

<https://medium.com/@anuuz.soni/advantages-and-disadvantages-of-knn-ee06599b9336>

<https://www.datacamp.com/tutorial/k-nearest-neighbor-classification-sikit-learn>

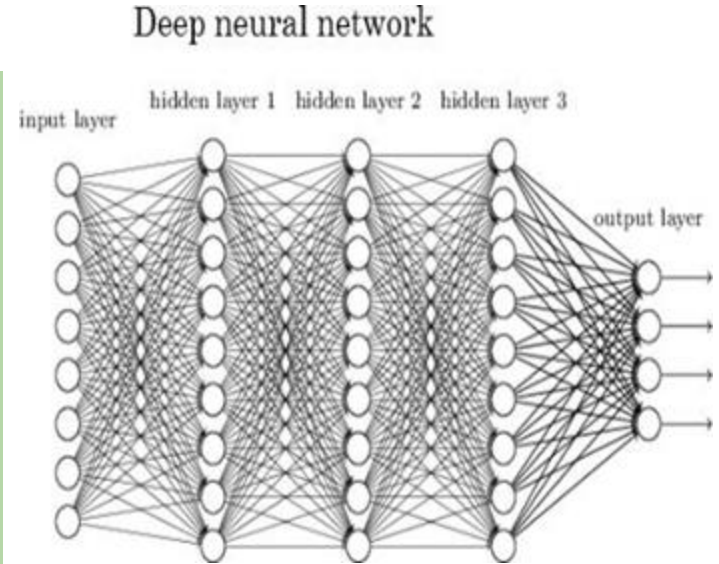
Proposed FYP Model: Deep Neural Network (DNN)

What is it?

DNNs are typically feedforward networks in which data flows from the input layer to the output layer without looping back

How it works

DNN transforms data at each layer, producing a new representation as output and divide datas in a classification problem, enhancing this action layer by layer until it reaches an output layer.



Advantages

- Able to process non-linear data

Disadvantages

- Deep learning works only with large amounts of data.

References:

<https://www.kdnuggets.com/2020/02/deep-neural-networks.html>

Example of Applications using Deep Neural Network(DNN)



References: <https://www.creativebloq.com/news/youtube-logo-history>

Methodology

1

Literature Review

Researched and reviewed similar related recommendation system models

2

Data Curation

Movie Lens Dataset

3

Exploratory Data Analysis

Visualising the data set to find valuable data from dataset and transform data into a format suitable for ML development

4

Evaluation Metrics

Defining appropriate metrics to measure model performance like Root Mean Square Error (RMSE)

5

Creating Baseline Models

Using movielens data, training and testing will be done to the existing models as they are the baselines for performance evaluation

6

Creating a DNN

Training and testing the proposed solution. Exploration of different DNN algorithms and architectures for the proposed solution

7

Performance Evaluation

Performance Evaluation for both the baseline model and proposed solution

8

Deployment : AI as a service

PoC of the developed model as AlaaS

DATASET



- Contains 25M movie ratings
- 25 million ratings and one million tag applications applied to 62,000 movies by 162,000 users.
- Includes tag genome data with 15 million relevance scores across 1,129 tags
- Released 12/2019
- The data are contained in the files genome-scores.csv, genome-tags.csv, links.csv, movies.csv, ratings.csv and tags.csv.

References : <https://grouplens.org/datasets/movielens/25m/>

Summary: FYP Proposal Recap

Improve accuracy of movie recommendation system using the Deep Neural Network (DNN) to improve user experience and retain more users to a website.