Deep Learning-Based Anime and Movie Recommendation System

MSc Research Project
Data Analytics

Pramod Ramu
Student ID: x20205759

School of Computing
National College of Ireland

Supervisor: Prof. Dr Christian Horn

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Pramod Ramu
Student ID:	x20205759
Programme:	Data Analytics
Year:	2022
Module:	MSc Research Project
Supervisor:	Prof. Dr Christian Horn
Submission Due Date:	15/08/2022
Project Title:	Deep Learning-Based Anime and Movie Recommendation System
Word Count:	5277
Page Count:	19

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Pramod Ramu
Date:	17th September 2022

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Deep Learning-Based Anime and Movie Recommendation System

Pramod Ramu
x20205759

Abstract

In recent days it is observed that the over-the-top media platforms (OTT) ruling the entertainment sector which offers various contents such as films, tv series, animes and music directly to customers over internet. This created an interest to conduct a research in developing a recommendation system to generate top 'n' recommendations of movies and animes for the users. Various machine learning models such as the k-nearest neighbors, collaborative filtering, content-based method and autoencoders have been used in generating the recommendations. But these models face data sparsity issues. So, there is an opportunity to address this problem as well as to generate a quality recommendations. So, in this project recommendation system using model-based collaborative filtering through deep learning has been proposed through which recommendations have been generated based on the user's likes and interests on the anime and movies. Here the data will be preprocessed and then it is transformed by incorporating embedding techniques. The movie and anime dataset for the project is sourced from the open-source platform called Kaggle. The recommendations for the users are generated based on the ranking. The model is evaluated using Mean Squared Error(MSE) and Mean Absolute Error(MAE).The model successfully addressed the data sparsity problem by predicting the rating of the anime and movies as well as top 'n' anime and movie recommendations have been generated.

Keywords: Deep Learning, Data Sparsity, Recommendation System

1 Introduction

1.1 Background

Today one of the tedious task that a layman is facing is to get the relevant information from the huge amount of information through internet.There are plenty of web applications are available and it is challenge to filter out the useful and the useless information from it.Various search engines like Google and Bing failed to give the filtered as well as relevant content to the users which is based on their interests and likes.This drawback was the main reason for the development of the recommendation.Nowadays the recommendation system has gain more popularity.The main intention of building the recommendation system was to deliver the relevant content to the users.The recommendation system is incorporated in various domains for recommending music,movies,books,electronic stuff,clothes and many more.This system provides customer satisfaction by generating good recommendations to the customers which would help to increase the business value of the organization.Most recently the over the top (OTT) media platforms are ruling the entertainment sector.The main goal of these platforms is to deliver a quality entertainment content to their subscribers over internet.So,a variety of OTT platforms, including Netflix, Amazon Prime Video, Disney+, Spotify, and many others, are connected with the recommendation system to deliver the best products to the user based on their preferences. In order to build a recommendation system various techniques with respect to the machine learning models have been incorporated.However, more research has been done recently on the application of deep learning models.The data which is involved in processing encompasses various formats like text,images,audios and videos and also the size of the data is huge,so most of platforms employs deep learning techniques for building the recommendation system to process this data.So the usage of the deep learning models benefits in enhancing the predictive accuracy and to generate the quality recommendations.

1.2 Motivation

The main intention of building a recommendation system is to generate the quality recommendations by predicting future choice of the user based on his/her previous history. The result will be recorded as a list or as a rating in a numerical format. This outcome would be utilized for forecasting the user interests and to provide a good customer experience by various organizations. This will increase the organization's growth in terms of business and quality.Based on this a fact check process was conducted,it was observed that utilizing a recommendation algorithm, Amazon generates more than 35 percent of its revenue. MacKenzie et al. (2013). The products which belong to the different categories are recommended with an intention of increasing average order value. Also in recent days there is a huge demand for entertainment contents.If a film or series get released in a entertainment platforms it could reach more number of audience throughout the globe. Every year more more number of movies and tv series which belongs to different categories are getting released. It would be a tougher task for these platforms to personalize the content to their subscribers. Integration of a recommendation system could deliver the best content to the audience. According to business statistics, the market for anime was valued at USD 24.80 billion in 2021 and is projected to reach USD 43.73 billion by the year 2027. Additionally, a considerable amount of money is being made in the film production industry. The entertainment industry is booming nowadays, it would be interesting to conduct a research in this area and to develop a recommendation system using deep learning techniques with an objective of addressing the data sparsity problem and to generate quality recommendation to the audience.

1.3 Research Question and Objective

1.3.1 Research Question

How effectively does the model-based collaborative filtering recommendation system solve the data sparsity problem by incorporating the deep learning techniques ?

1.3.2 Research Objective

The main objective of this paper to address the data sparsity problem. A user within a particular system not able to watch and rate all the products under it. The idea here is to develop a model which predicts the rating of the anime and movie which was not rated by the user and generate top 'n' recommendations of the same.

1.4 Justification and Scope

In order to develop a recommendation system various machine learning models,content-based and collaborative filtering method and auto-encoders have been incorporated.But these models couldn't address the data sparsity issue effectively. In order to overcome this model based collaborative filtering approach using deep learning techniques have been employed in this research which predicts the rating of the anime and movies by utilizing the user's data.

2 Related Work

Before starting the development of the proposed research work. Various research papers were studied and analyzed in order to get a clear understanding of the recommendation system,the various types,techniques for the development of the recommendation system for the recommending different stuffs,research questions addressed and the results were studied and the overall summary of the literature was documented.

The first and foremost step to conduct this research was to understand the concept of the recommendation system and the applications of it. As per the paper published by Ko et al. (2022)recommendation system models and the various technologies incorporated in recommendation systems were studied and the direction of the research trends by year were analyzed. In this paper a research survey of various articles from the year 2010 to 2021 has been conducted.Also the various application area where recommendation systems were incorporated were classified and the number of researches involved with respect to the recommendation system models and techniques used in each areas were studied and analyzed. The author discuss the various types and techniques such as content-based method,collaborative filtering method, hybrid methods and the usage of techniques like text mining,clustering , K-nearest neighbour , matrix factorization and deep neural network. This study provided an overall summary of recommendation systems and also a detailed analysis of its various technologies and trends.From the study a key observation was made that recently the usage of deep learning in developing a recommendation system has gained more popularity .

On studying more about the usage of deep learning in the recommendation system it came to know the powerful recommendation system of the Youtube and the Amazon incorporated the deep learning techniques as the traditional techniques have become outdated.In the paper published by Zhang et al. (2019) discussed the necessity of the development deep learning based models.The traditional models have data sparsity,cold start and scalability issues. Also the data which is to be processed is huge which may be in text,images,audio,and video formats. The deep learning models could do well with huge data. The author provided a classification of the recommendation system models as well as discussed the various benefits and the drawbacks on employing deep learning techniques for the development of the recommendation system.

2.1 Convolution Neural Network-Based Recommendation system

With an intention of capturing complex user and item interaction and also to generate accurate recommendation Lee and Kim (2022) proposed recommendation system based on the convolution neural network using the outer product matrix of features and cross convolutional filters. As per the evaluation results the proposed model could capture higher level interactions between the users and the movies. The author incorporated the global average or max pooling instead of the fully connected layers to avoid overfitting issue. The dataset used for this research were MovieLens 1M and 100K.The evaluation metrics used were RMSE and MAE.The RMSE observed for the proposed method was 0.8989 and MAE observed was 0.7048. The proposed model outperformed the other models and also addressed the overfitting issue.

The main focus of the existing systems was on pairing the users and the items.Less importance was given to the connection between them. In order address this problem Chen et al. (2022) proposed a model using convolutional neural network which is the combination of Co-occurrence pattern and convolutional neural network with the implicit feedback data. The co-occurance pattern is used to capture the user–item and item–item information.The datasets used for this research were MovieLens100K,1M and Lastfm. The evaluation metrics used were Hit Ratio and Normalized Discounted Cumulative Gain .Even though the proposed model outperforms the traditional model in terms of evaluation metrics but it failed to address the objective of the research effectively.

2.2 Autoencoder-Based Recommendation system

On observing the data sparsity issue in the recommendation system which causes when a user item matrix is expanded in a multidimensional matrix,Jeong and Kim (2021) proposed deep learning-based context-aware recommendation system which focuses on the the contextual features. The model is the combination of neural networks and autoencoder to extract characteristics and predict the rating. The proposed model predicts user choices by focusing on the the feature of user, item and context.CARSKit, DePaulMovie, InCarMusic and Res-taurant dataset were incorporated for the research. The evaluation metrics used for the research were RMSE (Root Mean Square Error), MAE (Mean Average Error),precision and MAP (Mean Average Precision). The model had the higher precision of 0.01–0.05 than the traditional models. Even though the model outperformed the traditional models but it failed in addressing the research objective effectively.

The usage of the multi-criteria recommendation system is more beneficial than single-criterion recommendation systems in generating the accurate recommendations. Along with other areas including image processing, computer vision, pattern recognition, and natural language processing, the use of deep learning in recommendation systems has grown in prominence in recent years. Based on these facts Shambour (2021) proposed deep learning based algorithm for multi-criteria recommendation systems where autoencoder is incorporated for capturing the non-trivial, nonlinear and hidden relationship between users with respect to the multi-criteria choices and preferences and also to generate the accurate recommendations. The dataset used for this research were Yahoo Movies MC dataset and the TripAdvisor MC dataset. The evaluation metrics used were RMSE and MAE.The proposed model outperformed the baseline models and also addressed the objective effectively.

On incorporating the matrix factorization method most of the traditional collaborative filtering models faces the data sparsity issue. In order to address this issue as well as to capture latent factor on implementing the enhanced methodology Zhang et al. (2021) proposed a model by integrating stacked sparse auto-encoder into matrix factorization to predict the rating. Also the author used Stacked sparse autoencoder for capturing the item representation and SVD++ to extract the feedback information. The dataset used for this research were Ciao,Movielens-100k and 1M. Root Mean square Error(RMSE) and the Mean Absolute Error(MAE) were used as the evaluation metric to evaluate the performance of the proposed model. The model performed well on all the dataset and also it outperformed the baseline models.

2.3 Hybrid-Based Recommendation system

The increase in the available data in online platforms have become an overhead to filter out the relevant among that and also this process has become too slow due to the overload of the information. Also on observing the cold start,data sparsity and long-tail item issues in the existing single recommendation algorithms Ming et al. (2021) proposed hybrid recommendation algorithm based on deep learning IA-CN. The model incorporates integrated strategy to combine user-based and item-based collaborative filtering algorithms for the classification of the output. The model also captures the non-linear interactions between the users and the items. The dataset used for this research were Amazon product dataset which contains two parts product metadata and user comment data. The evaluation metrics which were used for the analysis of the model's performance were RMSE, Recall and Precision.Even though the model performed well in predicting the rating on the test set but the model partially addresses the research problem.

In another paper published by Khurana and Dhingra (2021) an enhancement on the existing hybrid and knowledge based recommendation system has been implemented by nesting the clustering method and content based filtering method,classification method and collaborative filtering method. The intention of this research is to handle the scalability and data sparsity issues. The scalability issue is handled using fuzzy clustering method and using the Bayesian network classifier the recommendations were predicted. The dataset used in this research was Movielens.RMSE and MAE were used to evaluate the performance of the model. The RMSE of the proposed model was 0.1987 which least than the existing models. The accuracy of the model was 75.79 percent which was more than the baseline models. Overall the model addressed the scalability and data sparsity issues effectively.

One of the important factor which is need to be considered on generating the recommendation is the reliability. By focusing on this factor Ortega et al. (2021) proposed factorization model called Bernoulli Matrix Factorization model which is a classification based model. This method encompasses Bernoulli distribution which is supported by the matrix factorization method. MovieLens,FilmTrust and MyAnimelist dataset were used for this research. As the process involved is the classification so the confusion matrix was used to verify the sensitivity and the specificitry of the model. The model performed well on all the dataset and also outperformed the baseline models and addressed the research objective effectively.

2.4 Model Based Collaborative Recommendation System

Nowadays most of the commercial websites incorporate a recommendation system. By implementing a personalize recommendation system the user's preferences or choices based on their past history such as ratings or clicks could be modelled. This technique could be termed as the collaborative filtering.However the traditional models which incorporate the collaborative filtering method experience the sparsity of matrix ratings and scalability issues. But the usage of deep learning techniques is more beneficial as it is gaining more popularity in recent days. Based on this fact Aljunid and Dh (2020) proposed deep learning based collaborative recommendation system. Here the author feds the input by modelling the latent factors of the users and then the rating is predicted using forward propogation technique. Movielens-1m and 100k were the dataset used for this research. The evaluation was conducted using RMSE.The RMSE values observed for the Movielens-100K and 1M were 0.917 and 0.903. The model outperformed the baseline models.

The main intention of developing a recommendation system is to provide personalize recommendation to the users by conducting an analysis of their choices or preferences.However the traditional collaborative filtering methods depends on the explicit feedback data. Also these models fails to tackle the data sparsity and the cold start problems. With an intention of addressing the cold start problem Feng et al. (2021) proposed collaborative filtering ranking model which is the combination of rating-oriented approach of Probabilistic Matrix Factorization and pairwise ranking-oriented approach of Bayesian Personalized Ranking. This model utilizes both implicit and the explicit feedback data. Movielens-100k,1M,FiLmTrust and Ciao were the dataset used for this research and the model was evaluated using Precision, Recall, Mean Average Precision and Mean Reciprocal Rank. The model addressed the cold start issue effectively but it has a drawback in particular case when the user rating is less than one.

3 Methodology

In order to carry out this project the Knowledge Discovery in Databases(KDD) methodology has been incorporated. The various steps involved in the designing, building the proposed model and it's evaluation process were data extraction, cleaning, pre-processing, model building and evaluation.

3.1 Data Extraction

The datasets for this project have been sourced from online opensource platforms.For this project two different datasets has been incorporated which are Movielens-20M [1] and Anime Recommendation Database-2020 [2] .Movielens-20M dataset is sourced from a website called grouplens and the Anime Recommendation Database-2020 dataset is sourced from kaggle.The Files named movies.csv, ratings.csv, tags.csv, links.csv, genome-tags.csv, and genome-scores.csv make up the MovieLens-20M dataset. The dataset encompasses 20 million ratings and 465,000 tag applications applied to 27,000 movies by 138,000 users.By using http requests, the Anime Recommendation Database-2020 dataset was extracted from the MyAnime.list website.This dataset consists of likes from 325772 different users

[1] `https://grouplens.org/datasets/movielens/20m/`
[2] `https://www.kaggle.com/datasets/hernan4444/anime-recommendation-database-2020`

and details on 17562 anime.It contains include watching-status.csv, rating-complete.csv, animelist.csv, and anime-with-summary files. There were various attributes in both the datasets and have incorporated only which are required for the project and it is shown in figure 1 and figure2

Attribute Name	**Type**	**Description**
Anime_Id	Integer	ID of the anime from MyAnimeList website
Name	String	The Full name of the anime
Genres	String	The Comma separated list of genres for the anime
English Name	String	The Full name of the anime in English
Type	String	The Type of anime such as movie, TV, OVA, etc
Episodes	Integer	Number of chapters for the Anime
Members	Integer	The number of community members who are in an anime group.
Synopsis	String	Text which provides the synopsis of the anime
User_id	Integer	Nonidentifiable randomly generated user id
Rating	Float	The rating that this user has assigned to an anime

Figure 1: Attributes of the Anime dataset

3.2 Data Cleaning

The data cleaning process mainly involved removing the null and missing values,dropping of the columns which are not necessary and merging of various datasets. Firstly the null values were removed from the anime dataset and there were 35 columns in the anime dataset. Among them the columns which are not required were dropped.The focus is on overall rating of the anime the average score column ('Score') was considered and the columns which contain the number of people who assigned different scores('score1' to 'score10') was dropped. The anime and and the anime with synopsis were two different datsets.It was merged into a single dataframe based on the anime-id in order to get the complete overview of the anime. The columns 'Score','English name','Type' and 'Episodes' had unknown values and those rows were removed from the dataset. Next the cleaning process was carried on the anime-rating dataset. There were no missing or null values found.Later the above process was repeated again for movielens-20m dataset for the movies and movie-rating datasets. There were no missing values or unknown values found. The 'timestamp' column in the movie-rating dataset was dropped as it was not necessary. Next the anime and anime-rating datasets as well as movie and movie-rating datasets were merged into two separate dataframes based on 'anime-id' and 'movieid' for the purpose of the exploratory data analysis.

Attribute Name	Type	Description
MovieId	Integer	The MovieId is an identifier for movies used by https://movielens.org.
Title	String	The name of the movie
Genres	String	The type of the genre that the movie belongs to.
UserId	Integer	Nonidentifiable randomly generated user id
Rating	Float	The rating that this user has assigned to a movie.

Figure 2: Attributes of the movie dataset

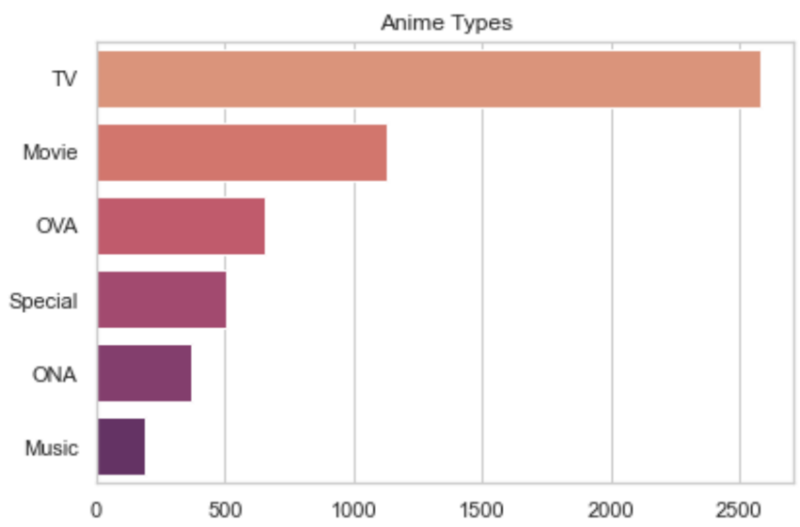

Figure 3: Plot of various types of anime

3.3 Exploratory Data Analysis

A barplot graph was plotted to check the various types of the anime.As observed in figure 3. As per the plot it seems that the anime belongs to the type tv-series is more and music is the least. The genres were the comma separated values it was pre-processed in such a way that a list with the count of different genres were obtained and the most common genres were identified that were comedy and action. Verified the details of the anime which encompasses the more number of members and this was checked for all the types of animes. The anime which belongs to tv-series type had more number of members which were more than 2 million as shown in the figure4. Also checked the

	Name	English name	Score	Members	Type	Genres	Synopsis
3446	Fullmetal Alchemist: Brotherhood	Fullmetal Alchemist:Brotherhood	9.19	2248456	TV	Action, Military, Adventure, Comedy, Drama, Ma...	"In order for something to be obtained, someth...
1336	Death Note	Death Note	8.63	2589552	TV	Mystery, Police, Psychological, Supernatural, ...	shinigami, as a god of death, can kill any per...
9394	One Punch Man	One Punch Man	8.57	2123866	TV	Action, Sci-Fi, Comedy, Parody, Super Power, S...	The seemingly ordinary and unimpressive Saitam...
6550	Shingeki no Kyojin	Attack on Titan	8.48	2531397	TV	Action, Military, Mystery, Super Power, Drama,...	Centuries ago, mankind was slaughtered to near...
5787	Sword Art Online	Sword Art Online	7.25	2214395	TV	Action, Game, Adventure, Romance, Fantasy	In the year 2022, virtual reality has progress...

Figure 4: Anime details which have more number of members

show which has highest number of episodes,the 'Doraemon' show had highest episodes of 1787. Also various plots different plots with the 'total ratings vs probability', 'mean ratings vs probability' and a joint plot of 'mean ratings vs total ratings' were created to understand the distribution of total ratings and mean ratings.It was deduced from all of the plots that the mean ratings range from 6 to 8.5 for a total of between 25,000 and 60,000 ratings. For the quality recommendation animes with rating greater than or equal to 6 are required. So the animes with the rating less than 6 were masked.The anime-rating dataset encompasses 57633278 rows with three columns 'user-id','anime-id' and 'rating'.

Next the movies and rating dataset with respect to movies were merged based on the 'movie-id'. Identified the various genres by pre-prprocessing the genres column to observe the count of various genres with respect to movies.As seen in figure 5 it can be inferred that comedy, drama and thriller are the top common genres. A combined plot between these two parameters was created in order to understand more clearly the distribution between the mean ratings and total ratings. From the graph as show in figure 6 it can be inferred that the mean ratings lie between 2.5 and 4.0 for the count of ratings ranging between 10,000 and 20,000.Also for movies created a bar plot was created to analyze the various ratings assigned for the movies. From the plot it was observed that 4.0 is the most assigned rating for the movies.

Figure 5: Various genres of movies

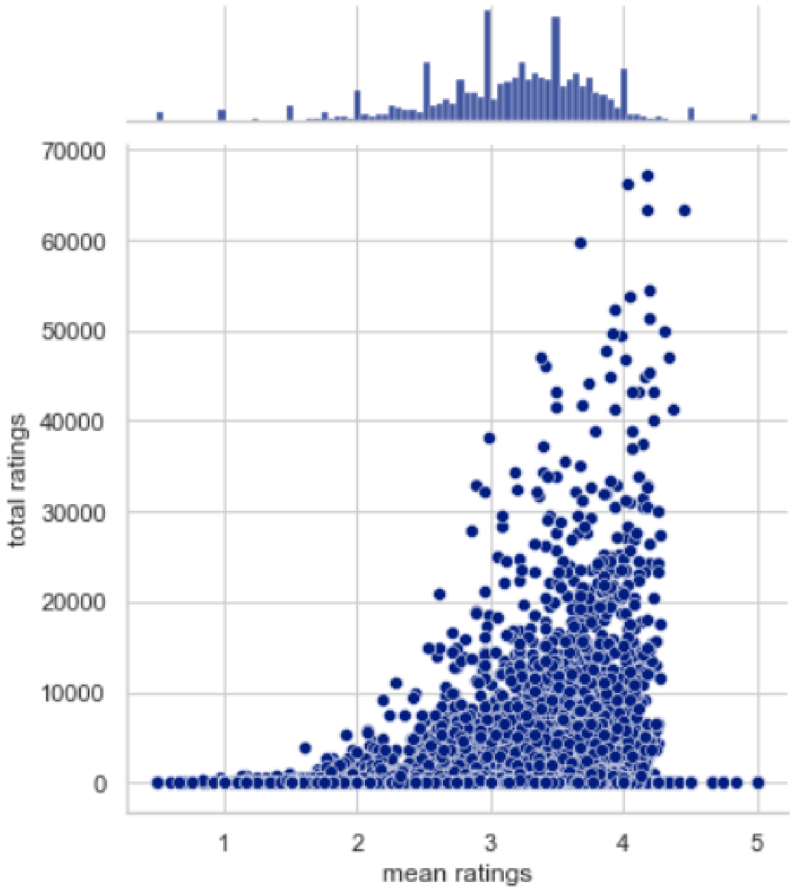

Figure 6: Plot of total ratings vs mean ratings for movies

3.4 Model Description

Model-based collaborative filtering using deep learning has been proposed in this study. In the proposed work the recommendation is based on an approach that suppose if the user A likes a product which has been liked by User B then there would be a probability that both the users A and B have similar preferences.While in the traditional system the recommendation was made on the collective information of users and items. In the proposed model the main focus is on predicting the user choices by collecting the data regarding their previous likes and preferences. There are two ways to achieve the collaborative filtering either by model based collaborative filtering or by memory based collaborative filtering.

3.4.1 Model based collaborative filtering

The model based collaborative filtering could be achieved by incorporating machine learning algorithms or various data mining techniques for predicting the unrated items by the user. Here the a user-item matrix will be computed which leads to the reduction of the dimensionality. The rows of the matrices is termed as embedded vectors.

3.4.2 Memory Based Collaborative Filtering

The memory based approach could be achieved by incorporating the available rating data from the users. This could be done in two ways one by user based collaborative filtering by computing similarity between users and item based collaborative filtering by computing the similarity between the items. The similarity could be computed either by person or cosine similarity methods.

3.4.3 Proposed Method using Model Based Collaborative Filtering using Deep Learning

On doing the study and analysis of various research papers it was observed that usage of the deep learning models is the state of art in the development of recommendation system. The working process of the proposed methodology is bit similar to the matrix factorization. So the matrix factorization works by computing a product of two rectangular matrices of lower dimension through the decomposition of user-item interaction matrix. So the first matrix encompasses users in rows and the second matrix encompasses items in columns. These rows and columns of the matrix are termed embeddings of users and items. In the proposed model instead of taking the product the matrix are concatenated for capturing the hidden interaction between the users and the items. The final output is to predict the ratings of the anime and movie which was not been rated by the user. The proposed model encompasses keras layers for predicting the rating and to make a quality recommendation.

3.5 Evaluation Metrics

Before generating the quality recommendation the proposed model is evaluated with usage of evaluation metrics. The evaluation metrics which are incorporated in this research in order to measure the performance of the model were Mean Square Error and Mean Absolute Error (MAE).

3.5.1 Mean Square Error

The Mean Square Error is defined as the squared difference of the actual and predicted values. In the research conducted the it will be squared difference of the actual and predicted ratings of the movie and anime. The mean Square Error is given by the formula

$$1/n * \sum (actual–predicted)^2$$

3.5.2 Mean Absolute Error

The mean absolute error is the difference between the actual value and the predicted value. In this research it is the difference between the actual and the predicted rating. The mean absolute error is given by the formula.

$$\sum |actual–predicted|$$

4 Design Specification

The figure 7 shows the architecture of the proposed model using model based collaborative filtering. The objective here is to predict the ratings of the anime and the movies which were not watched by the user and to generate the top 'n' recommendation of the movie and anime to the user. The input variables are the anime-id and the movie-id as well as the user-id of the anime and movie dataset are pre-processed and passed to the system.In the next step these variables are embedded into user-item matrix in the embedded layer. Once the input variable are embedded the next step is to flatten the user (user-id) and the item(anime-id or movie-id) embeddings and these flattened embeddings are concatenated. The similarity between the users could be captured using embeddings and concatenation helps to capture the hidden relationship between the users and the items. The next step is to fetch the outcome of the concatenation to a dense layer with activation function.This model encompasses ReLU(Rectified Linear Unit) activation function.The activation function mainly helps to overcome the gradient descent i problem. The outcome of the dense layer is passed to the loss function. The loss function incorporated for this model was 'Mean Square Error'.Now the outcome of the loss function is fed to the optimizer which is the Adam optimizer for the training of the data and finally the output can inferred from the model.

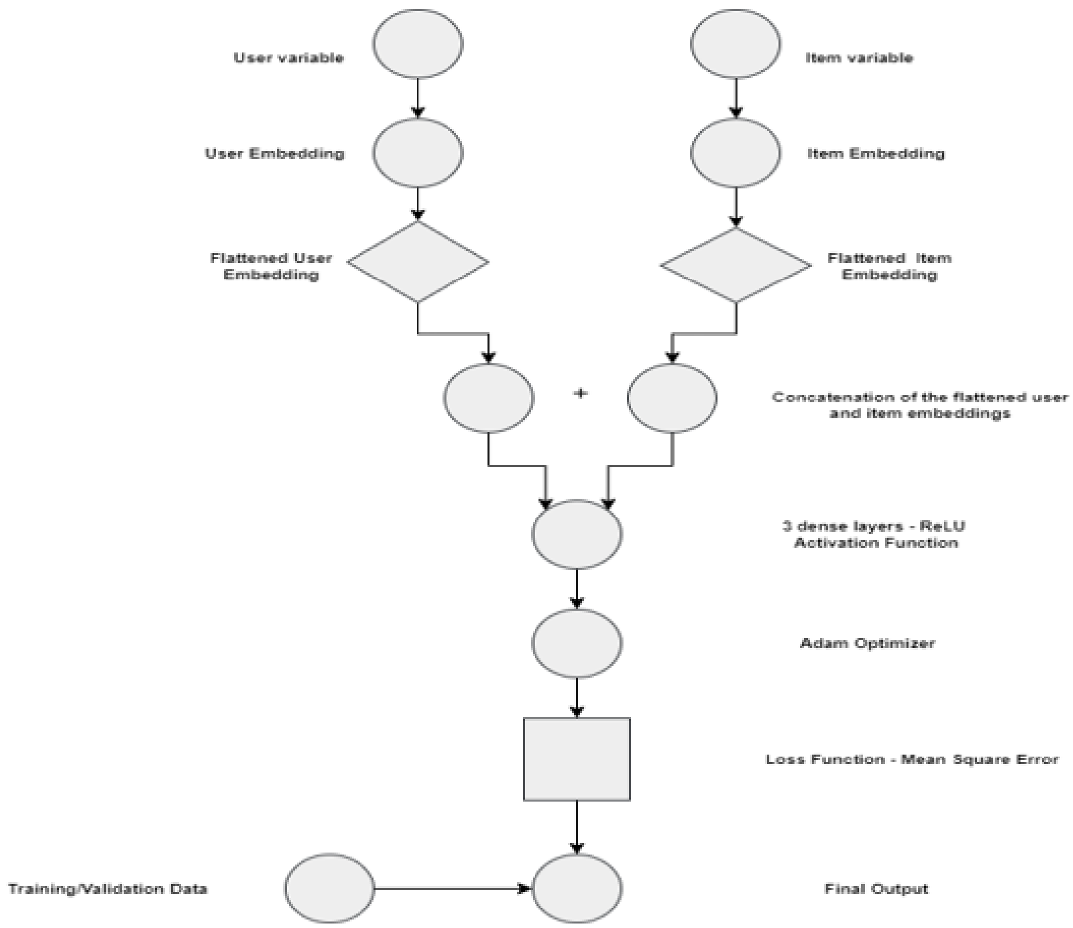

Figure 7: The proposed model for predicting rating of the anime and movie

5 Implementation

5.1 Model Setup

The model was implemented using a computer with 8gb ram, intel i5 processor and 64 bit windows operating system. The model was built and ran using jupyter notebook with the python version of 3.9.7. The major libraries required for model building were TensorFlow and Keras. For data pre-processing and exploratory data analysis process numpy , pandas, seaborn, matplotlib and WordCloud were used.

5.2 Building the model

As per the model design shown in figure 7 the model was implemented. Once the data is pre-processed then it is encoded using label encoder then unique user-ids and ratings are selected for both anime and movies.Then the user-ids and ratings are embedded into two matrices one for user and another one for items that are anime-id and movie-id and user-id.Later these embeddings are flattened individually. Then these flattened embeddings are concatenated. As explained in the design the main reason for the concatenation is to capture the hidden relationship between the user and the items. Then the outcome of the concatenation is fed to dense layer along with the ReLU activation function. The activation function mainly helps to overcome the vanishing gradient problem. Then the outcome of the first dense layer is fed to two hidden layers. Then the outcome from the hidden layer is fed to the output layer. Adam optimizer along with the mean square error as the loss function and mean absolute error as the metric is used for the training of the model. The overall model summary and the various variables used in the model is shown in figure 8.

5.3 Training and Validation Details

Once the data is pre-processed successfully the next step was to split the data into train and validation sets. The train set for anime and movie data encompasses 41045522 and 16000210 records. The validation set for the anime and movie encompasses 10261381 and 4000053 records. On running the model the train loss, validation loss and the metrics monitored and analyzed. Also the various parameters of the model were tuned for the enhancement of the output.

6 Evaluation

Once the proposed model has been implemented for the anime and movie the next step was to evaluate the model with different performance metrics. So here the model has been tested against the evaluation metrics such as Mean Absolute Error for metric and Mean Square Error. The details of the various parameters involved for the compiling and the evaluation of the model is shown in the figure9

The model was called twice,one for the recommendation of anime with anime rating train and validation set and another one for the recommendation of the movies with movie rating train and validation set. The overall results of both the dataset is shown in the figure 10

```
Model: "model"
__________________________________________________________________________________________________
 Layer (type)                   Output Shape         Param #     Connected to
==================================================================================================
 input_1 (InputLayer)           [(None, 1)]          0           []

 input_2 (InputLayer)           [(None, 1)]          0           []

 embedding (Embedding)          (None, 1, 5)         1548445     ['input_1[0][0]']

 embedding_1 (Embedding)        (None, 1, 5)         84215       ['input_2[0][0]']

 flatten (Flatten)              (None, 5)            0           ['embedding[0][0]']

 flatten_1 (Flatten)            (None, 5)            0           ['embedding_1[0][0]']

 concatenate (Concatenate)      (None, 10)           0           ['flatten[0][0]',
                                                                  'flatten_1[0][0]']

 dense (Dense)                  (None, 100)          1100        ['concatenate[0][0]']

 dense_1 (Dense)                (None, 32)           3232        ['dense[0][0]']

 dense_2 (Dense)                (None, 4)            132         ['dense_1[0][0]']

 dense_3 (Dense)                (None, 1)            5           ['dense_2[0][0]']

==================================================================================================
Total params: 1,637,129
Trainable params: 1,637,129
Non-trainable params: 0
__________________________________________________________________________________________________
```

Figure 8: Model Summary

Parameters	**Values**
Batch Size	10000
Optimizer	Adam
Regulizer	HeNormal
Loss	Mean Square Error
Metric	Mean Absolute Error
Activation	ReLU
Epochs	100,30

Figure 9: Parameters involved for the compiling and the evaluation of the model

Dataset	Metric	Training set	Validation Set
Anime Recommendation Database 2020	Mean Square Error (MSE)	0.8725	0.8872
Anime Recommendation Database 2020	Mean Absolute Error (MAE)	0.7431	0.7492
Movielense-20M	Mean Square Error (MSE)	0.7306	0.7417
Movielense-20M	Mean Absolute Error (MAE)	0.6556	0.6606

Figure 10: Results

6.1 Recommendation of the anime

For the recommendation of the anime, the anime rating dataset which was split into train and test set was fed to the model with initial number of epochs of were 20.Later on testing WITH different number of epochs it was fixed to 100.The model was ran till 100 epochs.The train loss observed was 0.8725 and the validation loss observed was 0.8872 .The MAE of the training set was 0.7431 and for the validation set was 0.7492.The loss curve for train and validation loss is shown in figure 11

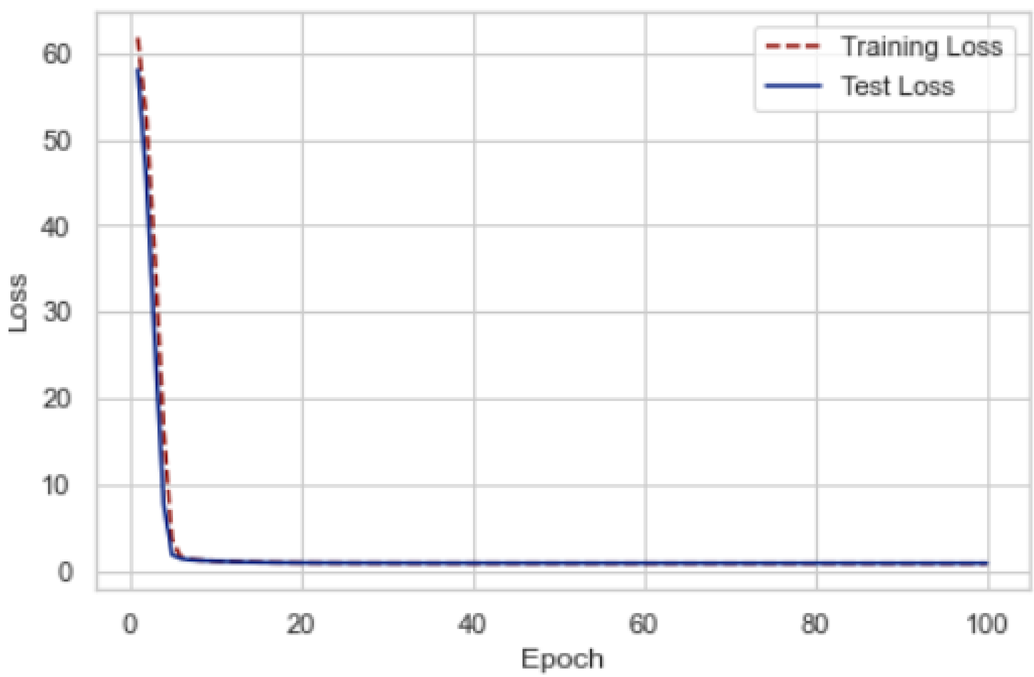

Figure 11: Loss curve for train and test sets for the anime dataset

As per the objective of this research the rating of the anime which was not rated by the user was predicted and and recommendations were generated. So here the ranking based recommendations were generated in two ways one the overall top 'n' anime recommendation irrespective of the type and the top 'n' anime recommendation particularly which belongs to the type 'Movie'. The figure 12 and figure 13 shows the predicted rating and the top 'n' anime recommendations for the random user within the system.

	anime_id	Name	English name	Type	Genres	Synopsis	predicted_rating
0	740	Bishoujo Senshi Sailor Moon R	Sailor Moon R	TV	Demons, Magic, Romance, Shoujo	Usagi Tsukino and her friends have been living...	9.666265
1	5673	Apfelland Monogatari	Appleland Story	Movie	Adventure, Historical, Drama	pfelland is a small but beautiful country surr...	9.296709
2	6463	Hoshi wo Katta Hi	The Day I Bought a Star	Movie	Sci-Fi, Kids	oung boy, tired of the city, escapes into the ...	9.272595
3	9863	SKET Dance	SKET Dance	TV	Comedy, School, Shounen	Kaimei High School there is a special club ded...	9.084867
4	427	Kaleido Star	Kaleido Star	TV	Comedy, Sports, Drama, Fantasy, Shoujo	The Kaleido Stage is known throughout the worl...	9.013369
5	238	Rekka no Honoo	Flame of Recca	TV	Action, Adventure, Martial Arts, Shounen, Supe...	ost people think that ninjas are a thing of th...	8.995009
6	1818	Claymore	Claymore	TV	Action, Adventure, Super Power, Demons, Supern...	hen a shapeshifting demon with a thirst for hu...	8.931616
7	25	Sunabouzu	Desert Punk	TV	Action, Adventure, Comedy, Ecchi, Sci-Fi, Shounen	The Great Kanto Desert, a sweltering wasteland...	8.926864
8	6571	Koume-chan ga Iku!	Here Comes Koume!	TV	Comedy, Romance, Slice of Life	Koume is a new Office Lady (OL), an employee o...	8.833509
9	313	Ace wo Nerae! (1979)	Aim for the Ace! (1979)	Movie	Sports, Drama, Romance, School, Shoujo	High school freshman Hiromi joins the tennis c...	8.825645

Figure 12: Top anime recommendations irrespective of the type

	anime_id	Name	English name	Type	Genres	Synopsis	predicted_rating
0	5673	Apfelland Monogatari	Appleland Story	Movie	Adventure, Historical, Drama	pfelland is a small but beautiful country surr...	9.296709
1	6463	Hoshi wo Katta Hi	The Day I Bought a Star	Movie	Sci-Fi, Kids	oung boy, tired of the city, escapes into the ...	9.272595
2	313	Ace wo Nerae! (1979)	Aim for the Ace! (1979)	Movie	Sports, Drama, Romance, School, Shoujo	High school freshman Hiromi joins the tennis c...	8.825645
3	791	Arion	Arion	Movie	Action, Adventure, Fantasy, Magic, Drama, Seinen	In Thrace, Arion is taken from his mother Deme...	8.747345
4	304	Aa! Megami-sama! Movie	Ah! My Goddess:The Movie	Movie	Comedy, Magic, Romance, Seinen, Supernatural	For centuries, a god named Celestin has been i...	8.666540
5	5096	Doraemon Movie 28: Nobita to Midori no Kyojin Den	Doraemon the Movie:Nobita and the Green Giant ...	Movie	Adventure, Comedy, Fantasy, Kids, Shounen	One day Nobita found a small sapling behind th...	8.655970
6	936	Naruto Movie 2: Dai Gekitotsu! Maboroshi no Ch...	Naruto the Movie 2:Legend of the Stone of Gelel	Movie	Adventure, Comedy, Drama, Fantasy, Shounen, Su...	Naruto, Shikamaru, and Sakura are executing th...	8.622528
7	437	Perfect Blue	Perfect Blue	Movie	Dementia, Drama, Horror, Psychological	J-pop idol group CHAM! has spent the last two ...	8.619439
8	16786	Bulsajo Robot Phoenix King	Defenders of Space	Movie	Action, Space, Mecha, Shounen	No synopsis information has been added to this...	8.605611
9	536	Slayers: The Motion Picture	Slayers:The Motion Picture	Movie	Adventure, Comedy, Magic, Fantasy, Shounen	In this prequel movie to the Slayers televison...	8.558195

Figure 13: Top anime recommendations particularly which belongs to the type 'Movie'

6.2 Recommendation of the movies

For the recommendation of the movie, the movie rating dataset which was split into train and test set was fed to the model with initial number of epochs of 20. After testing with different number of epochs the number of epochs fixed to 30. The model was ran till 30 epochs.The train loss observed was 0.7306 and the validation loss was observed was 0.7417. The MAE of the training set was 0.6556 and for the validation set was 0.6606. The loss curve for train and validation loss is shown in figure 14

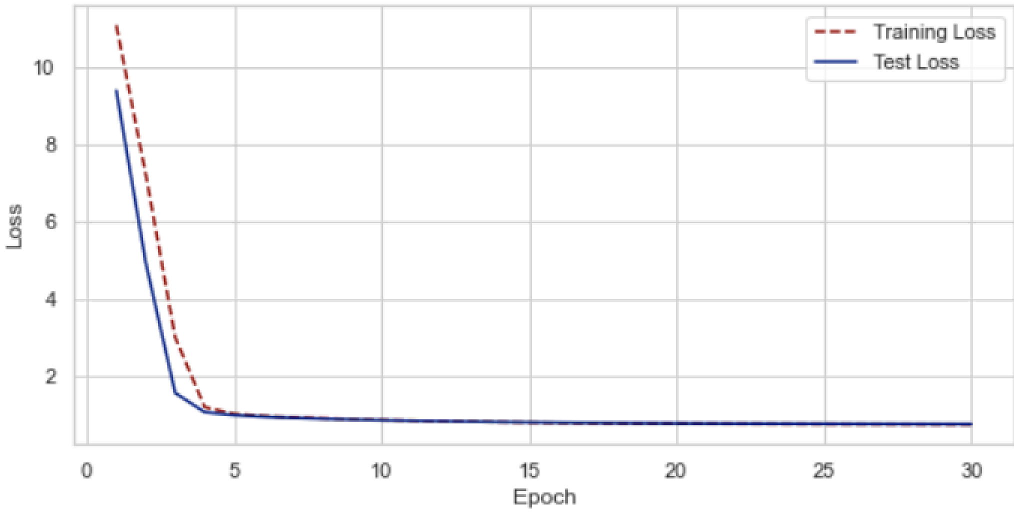

Figure 14: Loss curve for train and test sets for the movie dataset

Also as per the objective of this research the rating of the movie which was not rated by the user was predicted and the top 'n' recommendation of the movies was generated. The figure 15 shows the predicted rating and the top 'n' movie recommendations for the random user within the system.

	movieId	title	genres	predicted_rating
0	843	Lotto Land (1995)	Drama	4.035169
1	1935	How Green Was My Valley (1941)	Drama|Musical|Romance	3.967813
2	1155	Invitation, The (Zaproszenie) (1986)	Drama	3.953854
3	905	It Happened One Night (1934)	Comedy|Romance	3.950073
4	2943	Indochine (1992)	Drama|Romance	3.948989
5	1186	Sex, Lies, and Videotape (1989)	Drama	3.948775
6	7356	Night Crossing (1981)	Drama	3.939668
7	5917	Zoot Suit (1981)	Drama|Musical	3.933633
8	3346	Color Me Blood Red (1965)	Horror	3.931484
9	1232	Stalker (1979)	Drama|Mystery|Sci-Fi	3.918402

Figure 15: Top Movie Recommendation

6.3 Discussion

The proposed method incorporates model based collaborative filtering using deep learning which was evaluated on the anime and movie dataset. A single model has been developed for recommending both the anime and the movies. The activation function used was relu and on testing with different number of epochs and it would be beneficial for the enhancement of the results, so the number of epochs was increased from 20 to 100 for the anime and 20 to 30 for movie dataset. The 'he-normal' kernel initializer was used in order to avoid the overfitting. The Adam optimizer was used with learning rate of 0.00005.'Mean Square Error' was used as the loss function as the process was regression which is the prediction of the rating and the metric used was the 'Mean Absolute Error'. The model was also ran using 'sigmoid' and 'Softmax' activation functions. The quality of the results was better by using the 'ReLU' activation function.

7 Conclusion and Future Work

The overall objective of this research was to develop a recommendation system using model based collaborative filtering by incorporating deep learning techniques. The overall goal was to use this methodology to address the data sparsity issue as well as to generate the top 'n' quality recommendation of the anime and the movies. The objective was successfully addressed by the proposed methodology by predicting the rating of the anime and movies which was not rated by a particular user within the system. Also top 'n' recommendation of movies and the anime have been generated.

On observing the results the model has worked well with both the dataset. However, though the objective has been addressed successfully and the results are decent, still there is a room for the enhancement of the results. Also the intention behind this research to explore an another way to address the data sparsity issue by incorporating model based collaborative filtering methodology. In future the proposed model could be further improved by employing other attributes of the dataset and parameters for the enhancement of the recommendations which would be interesting to make a detailed observation of these results.

8 Acknowledgement

I would personally like to thank and express my gratitude to Prof. Dr Christian Horn for his guidance throughout this research project and also showing the right path to achieve the objectives of the research conducted. On every meeting with him ,would get more ideas and learning to carryout this research. He is one of the best project supervisor and his feedback were really valuable which would lead me to understand the concepts to the core.

References

Aljunid, M. F. and Dh, M. (2020). An efficient deep learning approach for collaborative filtering recommender system, *Procedia Computer Science* **171**: 829–836.

Chen, M., Ma, T. and Zhou, X. (2022). Cocnn: Co-occurrence cnn for recommendation, *Expert Systems with Applications* **195**: 116595.

Feng, J., Xia, Z., Feng, X. and Peng, J. (2021). Rbpr: A hybrid model for the new user cold start problem in recommender systems, *Knowledge-Based Systems* **214**: 106732.

Jeong, S.-Y. and Kim, Y.-K. (2021). Deep learning-based context-aware recommender system considering contextual features, *Applied Sciences* **12**(1): 45.

Khurana, D. and Dhingra, S. (2021). An improved hybrid and knowledge based recommender system for accurate prediction of movies, *2021 8th International Conference on Computing for Sustainable Global Development (INDIACom)*, IEEE, pp. 881–886.

Ko, H., Lee, S., Park, Y. and Choi, A. (2022). A survey of recommendation systems: recommendation models, techniques, and application fields, *Electronics* **11**(1): 141.

Lee, S. and Kim, D. (2022). Deep learning based recommender system using cross convolutional filters, *Information Sciences* **592**: 112–122.

MacKenzie, I., Meyer, C. and Noble, S. (2013). How retailers can keep up with consumers. mckinsey & company.

Ming, F., Tan, L. and Cheng, X. (2021). Hybrid recommendation scheme based on deep learning, *Mathematical Problems in Engineering* **2021**.

Ortega, F., Lara-Cabrera, R., González-Prieto, Á. and Bobadilla, J. (2021). Providing reliability in recommender systems through bernoulli matrix factorization, *Information Sciences* **553**: 110–128.

Shambour, Q. (2021). A deep learning based algorithm for multi-criteria recommender systems, *Knowledge-Based Systems* **211**: 106545.

Zhang, S., Yao, L., Sun, A. and Tay, Y. (2019). Deep learning based recommender system: A survey and new perspectives, *ACM Computing Surveys (CSUR)* **52**(1): 1–38.

Zhang, Y., Zhao, C., Chen, M. and Yuan, M. (2021). Integrating stacked sparse autoencoder into matrix factorization for rating prediction, *IEEE Access* **9**: 17641–17648.