Movie Recommender System

B.Supraja-17MIS0360
VIT University, Vellore
bondu.supraja2017@vitstudent.ac.in

R.Sai Kiran-17MIS0422
VIT University, Vellore
ragesai.kiran2017@vitstudent.ac.in

K.Ramya-17MIS0441
VIT University,Vellore
ramya.kovi2017@vitstudent.ac.in

G.Harshavardhan-17MIS0454
VIT University, Vellore
harshavardhan.2017b@vitstudent.ac.in

T.Bhargavi-17MIS0469
VIT University,Vellore
thanam.bhargavi2017@vitstudent.ac.in

Abstract— **This project is about movie analysis based on user ratings by concept of correlation. The content based filtering algorithm is implemented in the system. In today's digital world where there is an endless variety of content to be consumed like books, videos, articles, movies, etc., finding the content of one's liking has become a complicated task. On the other hand, digital content providers want to engage as many users on their service as possible for the maximum time. This is where analysis system comes into picture where the content providers recommend users the content according to the users' liking. The objective of Movie Vender is to provide accurate movie analysis to users. Usually the basic analysis systems consider one of the following factors for generating analysis; the preference of user (i.e. content based filtering).**

***Keywords*- Content based filtering, Correlation**

INTRODUCTION

We have taken two datasets from movies Department, one contains user Id, ratings, Item Id and another dataset contains Item Id and Movie names. If we watch any movie in any medium the next movie recommended based on user rating by concept of correlation. analysis systems help users find and select items from the huge number available on the web. Given a large set of items and a description of the user's needs, they present to the user a small set of the items that are well suited to the description. Similarly, a movie analysis system provides a level of comfort and personalization that helps the user interact better with the system and watch movies that cater to his needs. The chief purpose of our system is to recommend movies to its users based on their viewing history and ratings that reviewer provide.

OBJECTIVE

1. The background of analysis of a good movie to the user.
2. These challenges, researchers have been putting a lot of efforts to evolve analysis platform which can predict the best movies that can be recommended to the end user.
3. Mostly analysis system implements knowledge discovery technique to provide accurate analysis.
4. These analysis systems generate the list of analysis on the basis of two types of filtering process i.e. Collaborative filtering, content based filtering and hybrid based filtering.

LITERATURE SURVEY

Predicting and ranking box office revenue of movies based on big data

Author: ZhaoyuanWanga ,JunboZhanga, Shenggong Ji , ChuishiMeng ,Tianrui Li , Yu Zheng

Year:2020

Predicting box office revenue (BOR) of movies before releasing on big screens successfully becomes an emerging need, as it informs investment decisions on the stock market, the design of promotion strategies by advertisement companies, movie scheduling by cinemas, etc. However, the task is very challenging as it is affected by a lot of complex factors. In this paper, we first provide a strategic investigation of these influential factors. Then, we put forward a novel framework to predict a movie's BOR by modeling these factors using big data. Specifically, the framework consists of a series of feature learning models and a prediction and ranking model. In particular, there are two models devised for learning features: (1) a novel dynamic heterogeneous network embedding model to simultaneously learn latent representations of actors, directors, and companies, capable of capturing their cooperation relationship collectively; (2) a deep neural network-based model designed to uncover high level representations of movie quality from trailers. Based on the learned features, we train a mutually-enhanced prediction and ranking model to obtain the BOR prediction results. Finally, we apply the framework to the Chinese film market and conduct a comprehensive performance evaluation using real-world data. Experimental results demonstrate the superior performance of both extracted knowledge and the prediction results.[10]

Typical opinions mining based on Douban film comments in animated movies

Author: Ting Wu, FeiHao, Mijin Kim

Year: 2021

The film comments data contains a huge amount of mining research value, and text mining analysis of the animated film's comments can objectively reflect the quality of the animated film presentation and the problems generally expressed by the audience. However, these film comments are often mixed. The existing well-known film reviews websites have not excavated typical reviews on the users film comment text, so neither the audience nor the animation creators can analyze and apply the comments.This paper presents a general framework for mining typical opinions of film comments and uses crawler technology to obtain network review data, extract comment keywords based on the TF-IDF algorithm, and convert comments segmentation into word vectors trained by a neural network through Word2Vec. Then, using certain extraction rules and the K-means algorithm, the typical opinions with the same semantics but different expressions are aggregated together, and the typical opinions of the animation review of "Monkey King: Hero Is Back" are excavated. From the excavated information, we find out the production problems of the animation, so as to provide a certain reference to the creation of animated movies.[20]

Poster-Based Multiple Movie Genre Classification Using Inter-Channel Features

Author:JEONG A. WI, (Student Member, IEEE), SOOJIN JANG , (Student Member, IEEE), YOUNGBIN KIM , (Member, IEEE)

Year:6th April,2020

As the scale of the film industry grows, the demand for well-established movie databases is also growing. The genre of a movie supplies information on its overall content and has multiple values. Therefore, it should be well classified utilizing the characteristics of movies, without omissions in the database. In this study, we extract the optimal information and characteristics from movie posters to aid the classification of movies into genres and propose the use of a Gram layer in a convolutional neural network (CNN). The Gram layer first extracts style features by applying the Gram matrix to produce a feature map of a poster image. Using this as a style weight, the existing feature map is merged with style information to perform the genre classification task. The proposed Gram layer performed multi-genre classification tasks with higher efficiency than a residual neural network (ResNet), which is the current CNN model used for such tasks. We compared the activation map with the Squeeze-and-Excitation network, which gives weight to the image, and we confirmed that the introduction of the Gram layer actually focuses on the style of the movie poster. To classify the movie genres, we reconstructed the poster dataset into 12 multi-genres that emphasized the characteristics of each poster.[19]

Bayesian Approach to Users Perspective on Movie Genres

Author: Artem A. Lenskiy* and Eric Makita

Year: 2020

Over the last decade, movie genre prediction has been successfully applied in various domains such as social networking, online movie viewing websites, and ecommerce. Thus far, most of the techniques that have been proposed and reported in the literature do not take into account the perception of the users about the content genre information. In this work, we proposed an approach that expands the traditional movie classification algorithms by predicting the genre of a movie under evaluation instead of using user ratings of the watched movies. This approach can be easily generalized from movies to other items and their corresponding categories. To show the correctness of our approach, we conducted an experimental study using the Movie Lens 1M dataset. The experimental results showed that predicting the genre of a movie under evaluation can achieve an accuracy of more than 50% on the basis of only 1% of the training set of the users' combined ratings and of 83.8% when 80% of the whole set was taken as the training set. This finding is deemed valuable in many applications in practice. For instance, it can complement the genres given by experts.[4]

Calibrated Recommendations

Author: Harald Steck

Year: sep,2018

In this paper, we showed that recommender systems that are trained toward accuracy in the typical offline-setting may generate unbalanced recommendations, especially when the available training data are limited or noisy. We motivated the importance of calibration as an additional objective besides recommendation-accuracy. We outlined established metrics for quantifying the degree of calibration. It is desirable that they are particularly sensitive to discrepancies regarding the lesser areas of interest of a user, especially when such an area of interest is completely missing from the recommended list. Moreover, we presented a simple yet effective greedy algorithm, and outlined an optimality-guarantee due to sub modular functions. These approaches can be applied for post-processing the recommendation-lists generated by recommender systems. We also discussed the difference to diversity in its typical sense of minimal similarity or redundancy among the recommended items. Given that calibration is a property of the entire recommended list, future improvements may be achieved by integrating calibration in the objective of list wise learning-to-rank approaches, and by going beyond the typical offline-setting of training and testing recommender systems. This paper took a user-centric view, i.e., the recommendations were calibrated for each user. The complementary perspective is the item-centric view, which we leave for future work: as to calibrate the recommendations with respect to each item, one may consider, for instance, whether an item that is recommended twice as often as another item, is also consumed twice as often (across all the users).[15]

Movie Recommendation via Markovian Factorization of Matrix Processes

Author: RICHONG ZHANG,AND YONGYI MAO

Year: February 6, 2019

The achievement of the probabilistic network factorization (PMF) model has motivated the quick advancement of community oriented separating calculations, among which timeSVD++ has exhibited extraordinary execution advantage in tackling the film rating forecast issue. Permitting the model to develop over time, timeSVD++ represents "idea float" in community oriented separating by heuristically altering the quadratic advancement issue got from the PMF model. Thusly, timeSVD++ does not convey anymore any probabilistic translation. This absence of systems makes the speculation of timeSVD++ to other community oriented separating issues rather troublesome. This paper presents another model family named Markovian factorization of grid measure (MFMP). On one hand, MFMP models, like timeSVD++, are skilled of catching the worldly elements in the dataset, and then again, they additionally have clean probabilistic plans, permitting them to adjust to a wide range of community oriented sifting issues. Two straightforward model models in this family are presented for the expectation of film evaluations utilizing time-stepped rating information. The trial study utilizing MovieLens dataset shows that the two models, albeit straightforward furthermore, crude, as of now have equivalent or shockingly better execution than timeSVD++ and a standard tensor factorization model. [22]

Social-Aware Movie Recommendation via Multimodal Network Learning

Author: Zhou Zhao ,Qifan Yang, Hanqing Lu, Tim Weninger, Deng Cai, Xiaofei He, and Yueting Zhuang

Year: February 2018

With the fast improvement of Internet film industry, social-mindful film proposal frameworks (SMRs) have become a famous online web administration that give important film suggestions to clients. In this exertion, many existing film suggestion approaches become familiar with a client positioning model from client criticism concerning the film's substance. Lamentably, this methodology experiences the sparsity issue inherent in SMRdata . In the present work, we address the sparsity issue by learning a multimodal network portrayal for positioning film suggestions. We build up a heterogeneous SMR network for film proposal that abuses the printed depiction and film banner picture of every film, just as client ratings and socialrelationships. With this multimodal data,we then present a heterogeneous data network learning system called SMR-multimodal network portrayal learning (MNRL) for film proposal. To take in a positioning measurement from the heterogeneous data network we additionally built up a multimodal neural organization model. We assessed this model on a huge scope dataset from a genuine SMR Web website, and we find thatSMR-MNRL accomplishes better performance than other state-of- the-craftsmanship answers for the issue. [24]

Personalized Real-Time Movie Recommendation System: Practical Prototype and Evaluation

Author: Jiang Zhang, YufengWang ,Zhiyuan Yuan, and QunJin

Year: April 2020

With the emission of large information, down to earth suggestion plans are currently vital in different fields, including internet business, informal organizations, and various electronic administrations. These days, there exist many customized film proposal plans using freely accessible film datasets (e.g., MovieLensalso, Netflix), and returning improved execution measurements (e.g., Root-Mean-Square Error (RMSE)). Nonetheless, two crucial issues looked by film suggestion frameworks are as yet dismissed: first, versatility, and second, down to earth use criticism and check dependent on genuine usage. Specifically, Collaborative Filtering (CF) is one of the major winning strategies for actualizing proposal frameworks. Nonetheless, customary CF plans experience the ill effects of a period intricacy issue, which makes them terrible possibility for genuine suggestion frameworks. In this paper, we address these two issues. Initially, a basic however high-productive suggestion calculation is proposed, which misuses clients' profile ascribes to segment them into a few bunches. For each bunch, a virtual assessment pioneer is considered to address the entire group, with the end goal that the element of the first client thing framework can be essentially diminished, at that point a Weighted Slope One-VU technique is planned and applied to the virtual assessment pioneer thing framework to acquire the proposal results. Contrasted with customary grouping based CF proposal plots, our technique can essentially lessen the time intricacy, while accomplishing tantamount suggestion execution. Besides, we have developed a genuine customized online film suggestion framework, MovieWatch, opened it to the general population, gathered client criticism on proposals, also, assessed the attainability and precision of our framework dependent on this true information.[21]

Exploiting Aesthetic Features in Visual Contents for Movie Recommendation

Author:XIAOJIE CHEN , PENGPENG ZHAO , YANCHI LIU , LEI ZHAO , JUNHUA FANG ,VICTOR S. SHENG AND ZHIMING CUI

Year: April 2019

As perhaps the most broadly utilized recommender frameworks, film suggestion plays an significant part in our life. Be that as it may, the information sparsity issue seriously ruins the viability of individual alized film proposal, which requires more rich substance data to be used. Banners and still casings, which straightforwardly show the visual substance of films, have huge effects on film proposal. They uncover rich information for understanding motion pictures as well as helpful for under- standing client inclinations. Be that as it may, existing suggestion techniques infrequently think about tasteful highlights, which tell how the film looks and feels, separated from these photos for the film proposal. To this end, in this paper, we propose a tasteful mindful brought together visual substance grid factorization (called UVMFAES) to integrate visualfeature learning and recommendation into aunified framework. Specifically,

we initially incorporate the convolutional neural organization (CNN) highlights and tasteful highlights into probabilistic grid factorization. At that point we set up a brought together streamlining system with these highlights for the film suggestion. The test results on two genuine world datasets show that our proposed strategy UVMF-AES is fundamentally better than the best in class techniques on film suggestion.[3]

Movie Recommendation System Using Sentiment Analysis From Microblogging Data

Author: SudhanshuKumar ,Kanjar De, and ParthaPratim Roy

Year: AUGUST 2020

Suggestion frameworks (RSs) have accumulated enormous interest for applications in web based business and advanced media. Customary methodologies in RSs incorporate, for example, collabora- tive sifting (CF) and substance based separating (CBF) through these approaches that have certain constraints, like the need of earlier client history and propensities for playing out the assignment of proposal. To limit the impact of such constraint, this article proposes a cross breed RS for the motion pictures that influence the best of ideas utilized from CF and CBF alongside assumption examination of tweets from microblogging destinations. The reason to utilize film tweets is to comprehend the latest things, public assessment, and client reaction of the film. RSs are a significant mode of data separating frameworks in the advanced age, where the huge measure of information is promptly accessible. In this article, we have proposed a film RS that utilizes estimation investigation information from Twitter, alongside film metadata and a social diagram to suggest films. Slant examination gives data about how the crowd is react to a specific film and how this data is seen to be helpful. The proposed framework utilized weighted score combination to improve the suggestions. In view of our trials, the normal exactness in Top-5 furthermore, Top-10 for feeling comparability, mixture, and proposed model are 0.54 and 1.04, 1.86 and 3.31, and 2.54 and 4.97, individually. Trials led on the general population data set have yielded promising outcomes. [12]

DeepStar: Detecting Starring Characters in Movies

Author: IJAZ UL HAQ, KHAN MUHAMMAD, AMIN ULLAH, AND SUNG WOOK BAIK

Year:2019

Recent advances in the film industry have given rise to exponential growth in movie/drama production and adaptation of the Big Data concept. Automatic identification and classification of movie characters have received tremendous attention from researchers due to its applications in video semantic analysis, video summarization, and personalized video retrieval for which several methods have been recently presented. However, these methods cannot detect main characters properly due to their variation in pose and style in different scenes of a movie. To address this problem we present DeepStar, a novel framework for starring character identification based on deep high-level robust features. The proposed framework is threefold: the extraction of shots with clear faces from the input video; face clustering using discriminative deep features; and the occurrence matrix generation, helping in the selection of starring characters. The promising results obtained using representative Hollywood movies demonstrate the effectiveness of our method in detecting starring characters over the state-of-the-art methods.[8]

Estimating Audience Engagement to Predict Movie Ratings

Author: RajithaNavarathna, Peter Car, Patrick Lucey, and Iain Matthews

Year:2019

While watching movies, audience members exhibit both subtle and coarse gestures (e.g., smiles, head-pose change, fidgeting, stretching) which convey sentiment (i.e., engaged or disengaged) during feature length movies. Noticing these behaviors using computer vision systems is a very challenging problem—especially in a movie theatre environment. The environment is dark and contains views of people at different scales and viewpoints. Feature length movies typically run 80-120 minutes, and tracking people uninterrupted for this duration is still an unsolved problem. Facial expressions of audience members are subtle, short, and sparse; making it difficult to detect and recognize activities.

Finally, annotating audience sentiment at the frame-level is prohibitively time consuming. To circumvent these issues, we use an infraredilluminated test-bed to obtain a visually uniform input of audiences watching feature length movies. We present a method which can automatically detect the change in behavior (key-gestures) using "key-frames", which can convey audience sentiment. As the number of key-frames are many orders of magnitudes lower than the number of frames, the annotation problem is reduced to assigning a sentiment label for each key-frame. Using these discovered key-gestures, we create a movie rating classifier from crowd-sourced ratings and demonstrate its predictive capability. Our dataset consists of over 50 hours of audience behavior collected across 237 subjects.[25]

An Attention-Based Unsupervised Adversarial Model for Movie Review Spam Detection

Author: Yuan Gao, MaoguoGong , Yu Xie , and A. K. Qin

Year:2021

With the prevalence of the Internet, online reviews have become a valuable information resource for people. However, the authenticity of online reviews remains a concern, and deceptive reviews have become one of the most urgent network security problems to be solved. Review spams will mislead users into making suboptimal choices and inflict their trust in online reviews. Most existing research manually extracted features and labeled training samples, which are usually complicated and time-consuming. This paper focuses primarily on a neglected emerging domain - movie review, and develops a novel unsupervised spam detection model with an attention mechanism. By extracting the statistical features of reviews, it is revealed that users will express their sentiments on different aspects of movies in reviews. An attention mechanism is introduced in the review embedding, and the conditional generative adversarial network is exploited to learn users' review style for different genres of movies. The proposed model is evaluated on movie reviews crawled from Douban, a Chinese online community where people could express their feelings about movies. The experimental results demonstrate the superior performance of the proposed approach.[6]

Movie Question Answering via Textual Memory and Plot Graph

Author: Yahong Han, Bo Wang, Richang Hong, and Fei Wu

Movies provide us with a mass of visual content as well as attracting stories. Existing methods have illustrated that understanding movie stories through only visual content is still a hard problem. In this paper, for answering questions about movies, we introduce a new dataset called Plot Graphs, as external knowledge. The dataset contains massive graph-based information of movies. In addition, we put forward a model that can utilize movie clip, subtitle, and graph-based external knowledge. The model contains two main parts: a layered memory network (LMN) and a plot graph representation network (PGRN). In particular, the LMN can represent frame-level and clip-level movie content by the fixed word memory module and the adaptive subtitle memory module, respectively. And the plot graph representation network can represent the entire graph. We first extract words and sentences from the training movie subtitles and then the hierarchically formed movie representations, which are learned from LMN. At the same time, the PGRN can represent the semantic information and the relationships in the graph. We conduct extensive experiments on the Movie QA dataset and the Plot Graphs dataset. With only visual content as inputs, the LMN with frame-level representation obtains a large performance improvement. When incorporating subtitles into LMN to form the clip-level representation, we achieve the state-of-the-art performance on the online evaluation task of "Video + Subtitles." After the integration of external knowledge, the performance of the model consisting of the LMN and the PGRN is further improved. The good performance successfully demonstrates that the external knowledge and the proposed model are effective for movie understanding.[7]

Unsupervised Discovery of Character Dictionaries in Animation Movies

Author: Krishna Somandepalli, Naveen Kumar, TanayaGuha, and Shrikanth S. Narayanan

Year: 2018

Automatic content analysis of animation movies can enable an objective understanding of character (actor) representations and their portrayals. It can also help illuminate potential markers of unconscious biases and
their impact. However, multimedia analysis of movie content has predominantly focused on live-action

features. A dearth of multimedia research in this field is because of the complexity and heterogeneity in the design of animated characters—an extremely challenging problem to be generalized by a single method or model. In this paper, we address the problem of automatically discovering characters in animation movies as a first step toward automatic character labeling in these media. Movie-specific character dictionaries can act as a powerful first step for subsequent content analysis at scale. We propose an unsupervised approach which requires no prior information about the characters in a movie. We first use a deep neural networkbased object detector that is trained on natural images to identify a set of initial character candidates. These candidates are further pruned using saliency constraints and visual object tracking. A character dictionary per movie is then generated from exemplars obtained by clustering these candidates. We are able to identify both anthropomorphic and nonanthropomorphic characters in a dataset of 46 animation movies with varying composition and character design. Our results indicate high precision and recall of the automatically detected characters compared to human annotated ground truth, demonstrating the generalizability of our approach.[16]

Sentiment Analysis of Movie Reviews using Machine Learning Techniques

Authors: PalakBaid, Apoorva Gupta, NeelamChaplot

Year: 2017

Sentiment Analysis the investigation of feelings and conclusions from any type of text. Sentiment Analysis is additionally named as opinion mining. Sentiment Analysis of the information is helpful to communicate the assessment of the mass or gathering or any person. This method is utilized to discover the assumption of the individual concerning a given source of substance. Social media and other online platforms contain a large amount of the information as tweets, websites, and updates on the status, posts, and so on. In this paper, authors have examined the Movie surveys utilizing different procedures like Naive Bayes, K-Nearest Neighbor and Random Forest. The best results were given by Naive Bayes classifier. The Naive Bayes classifier achieved 81.45% accuracy, Random Forest classifier we achieved 78.65% accuracy, K-Nearest Neighbour classifier achieved 55.30% accuracy. As just couple of calculations were tried , it is needed to test other calculations or make hybrid strategies so that accuracy of the results can be expanded.[12]

Sentiment Analysis of Movie Ratings System

Authors: SagarChavan ,AkashMorwal, ShivamPatanwala, PrachiJanrao

Year: 2017

Sentiment Analysis likewise term as to the utilization of natural language processing, text examination and computational language to distinguish and remove abstract data in source materials. Lately, Opinion mining is an area of interest in the field of natural language processing, and it is likewise a difficult issue Sentiment examination is broadly applied to audits and social media for a variety of utilizations, going from showcasing to client assistance. Film ratings are a significant method to measure the performance of a film.

The goal of this paper is to separate highlights from the item audits and characterize surveys into positive, negative. In order to perform sentiment analysis, data have to be prepared in order to obtain a data set which is called the training set. The training data is performed preprocessing techniques if required. Then, such a data set is involved in the learning step, which uses Machine Learning algorithm yields a trained classifier. After training the classifier has to be tested on a different data set – namely the test set. [18]

Movie Rating Prediction by Opinion Mining

Authors: Malini R, Dr. Sunitha M.R, Mrs. Arpitha C.N

Year: 2019

Film ratings is critical occupation in movie fields, for instance, customer movie proposals, checking the association between client's submitted reviews and assessments, etc. The limit is to predict the rating of a film would be significant contemplating these points. In this work, first thing, we propose techniques to gathering and anticipate the film rating reliant on its studies. In order to achieve this, we consider the connected film reviews as well while anticipating the rating. Authors have used the overview based assumption close by the summation in order to anticipate the rating even more exactly since the assessment

gets a lot of fundamental information that can help rating prediction. In this paper, authors isolated new features that firmly influence finding the limit of the film reviews. The essential target of this paper is to aggregate the sentences according to its idea by using SVM course of action system. To perform tweets orders, there are a couple of AI classifiers. NB and SVM execute extraordinary and moreover give most raised accuracy as in the outcomes. When comparing to NB classifier, the SVM classifier exhibits higher precision results. [14]

Forecasting Movie Rating Through Data Analytics

Authors: LatikaKharb , Deepak Chahal , and Vagisha

Year: 2019

Film forecast is a significant method to predict film income and performance .Through information investigation, we can locate the most well known genres, performance in recent years and how it affects the standing of the upcoming film. As film creation brings about immense expense and work, authors goal is to predict the chances of success, so creation could be managed as needed. In the paper, authors have examined about the various manners by which the information examination utilized by the film gives an exact plan to every creation about the best or on the other hand most exceedingly awful odds of progress as well as disappointment. In this paper, authors goal is to anticipate profitability of a film to support film investment decisions in the beginning phases of film creation. The film makers and directors can make use of the proposed model in different manners like: change the film measures for turning out to be blockbusters, dispatch film at specific time span to get most extreme benefit, predict the fan following to get a blockbuster, etc..[17]

Movie Rating System Based on Sentiment Analysis

Authors: Rakshanda Mulay ,Shivangi Pandey , NiranjanPatil , MehulPatil , and Mrs.SeemaMandlik

Year: 2020

Manual reading of contents of film audits is a time consuming process for the film watchers. So they can't make proper sentiments/judgment to watch films or not. Subsequently, authors have build up a automated movie rating system based on sentiment analysis. We have two sentiments which we are showing utilizing emoticons- Happy for positive sentiments and Sad for negative sentiments. Sentiment analysis on arrangement of film surveys which is given by film watchers which measures the attitude of the film

watchers towards the film. For example regardless of whether it is positive or negative or attempt to understand what was the general response to the film was by them, for example in the event that they loved the film or they hated it. First we perform data pre-processing then we perform feature extraction after we perform classification of movie review into positive or negative. This proposed System will have two kinds of clients, registered client or unregistered client. Unregistered client can see the trailer of film and can just can only read movie review of movie viewers. The registered client can see trailers, give review about film and see feeling of the film watchers is shown after analyzing the review given by him.[9]

A hybrid recommender system for recommending relevant movies using an expert system

Author: Bogdan Walek , Vladimir Fojtik

Year:2020

Currently, the Internet contains a large amount of information, which must then be filtered to determine suitability for certain users. Recommender systems are a very suitable tool for this purpose. In this paper, we propose a monolithic hybrid recommender system called Predictory, which combines a recommender module composed of a collaborative filtering system (using the SVD algorithm), a content-based system, and a fuzzy expert system. The proposed system serves to recommend suitable movies. The system works with favorite and unpopular genres of the user, while the final list of recommended movies is determined using a fuzzy expert system, which evaluates the importance of the movies. The expert system works with several parameters – average movie rating, number of ratings, and the level of similarity between already rated movies. Therefore, our system achieves better results than traditional approaches, such as collaborative filtering systems, content-based systems, and weighted hybrid systems. The system verification based on standard metrics (precision, recall, F1-measure) achieves results over 80%. The main contribution is the

creation of a complex hybrid system in the area of movie recommendation, which has been verified on a group of users using the MovieLens dataset and compared with other traditional recommender systems.[2]

Leveraging Long and Short-Term Information in Content-Aware Movie Recommendation via Adversarial Training

Author: Wei Zhao, Benyou Wang, Min Yang ,Jianbo Ye, Zhou Zhao , Xiaojun Chen , and Ying Shen

Year:2020

Movie recommendation systems provide users with ranked lists of movies based on individual's preferences and constraints. Two types of models are commonly used to generate ranking results: 1) long-term models and 2) session-based models. The long-term-based models represent the interactions between users and movies that are supposed to change slowly across time, while the session-based models encode the information of users' interests and changing dynamics of movies' attributes in short terms. In this paper, we propose the LSIC model, leveraging long and short-term information for content-aware movie recommendation using adversarial training. In the adversarial process, we train a generator as an agent of reinforcement learning which recommends the next movie to a user sequentially. We also train a discriminator which attempts to distinguish the generated list of movies from the real records. The poster information of movies is integrated to further improve the performance of movie recommendation, which is specifically essential when few ratings are available. The experiments demonstrate that the proposed model has robust superiority over competitors and achieves the state-of-the-art results.[23]

Analysis of Movie Recommendation Systems; with and without considering the low rated movies

Author: Muppana Mahesh Reddy, R.SujithraKanmani ,B.Surendiran

Year:2020

Movie recommendation system is one of the top research areas, currently. Due to the impact of high internet speeds, multimedia has become one of the best entertainments. Recommendation system has its applications like movie recommendations, course recommendations, e-commerce etc..Movie recommendation system scope is not limited to entertainment, but also in information sharing. Movie recommendation systems suffer from problems like Cold-start problem, Sparsity, Long-tail problem, Grey sheep problem etc.. Some of these problems can be solved or at least be minimized if we take the right decisions on what kind of movies to ignore, what movies to consider. This paper examines the recommendations that are obtained with and without considering the movies that have never got an above-average rating, where average rating is defined here as the mid-value between 0 and maximum rating used, for example, 2.5 in 1 to 5 rating scale. The technique used is “collaborative filtering” and the similarity measure used is the “Pearson correlation coefficient”. Dataset considered is Movie-Lens-100k. This experiment result shows that low rated movies are not significant in finding the movie predictions. So it's suggestable to ignore them while calculating movie predictions.[13]

Research on Movie Rating Prediction Algorithms

Author: Xiaovue Li, Haonan Zhao, Zhuo Wang, Zhezhou Yu

Year:2020

In the era of data explosion of movie and video websites, “informahon overload” and “information labyrinth” have brought serious troubles, but movie recommender systems can efficiently solve such problems. In order to solve these problems, we propose the RF that uses the users' activity and rating to select suitable experimental data and proved this method can efficiently reduce RMSE and MAE of various recommendation algorithms. On this basis, we propose the MCBF-SVD, which is a movie rating prediction algorithm based on explicit data and implicit data from movie datasets to predict the future ratings of movies from users with a certain degree of activity. Firstly, the MCBF-SVD uses weighting factors to discuss the impact of movie categories on predicting future rating behavior of users, and also improves the filtering method based on movie categories. Finally, the MCBF is combined with SVD algorithm which has good performance in CF. Compared with several existing algorithms, our MCBF-SVD greatly enhanced the accuracy of rating prediction, and improved the scalability and efficiency of personalized recommender systems.[5]

Movie Success Prediction Using ML

Author: Maria S. Pera ,Yiu-Kai Ng

Year:2019

Movies continue to be a major source of entertainment in any country. However, this industry also incurs a lot of losses when the movie does not perform at the Box Office. Our solution will try to predict the success

rate of a movie by doing predictive analysis on the various features of the movie. Our model will predict the Success, based on different attributes / features of the movie. i.e. Movie crew (including director producer, music director), Movie plot

(Storyline), Box-Office revenue, Audience and Critics reviews / ratings. In this paper a detailed study of machine learning algorithms such as Random Forest, DecisionTree, K-NearestNeighbours (KNN), NLP, XGBoost Classifier and Deep Neural Network were done and were implemented on IMDB dataset for predicting Success of movies. Based on the results, XGBoost Classifier gave best accuracy.[1]

IV. ALGORITHMS USED

1.Content Based Filtering Algorithm: -

This algorithm recommends products which are similar to the ones that a user has liked in the past. The system save all the information related to each user in a vector form. This vector contains the past behavior of the user, i.e. the movies liked/disliked by the user and the ratings given by them. This vector is known as the *profile vector*. All the information related to movies is stored in another vector called the *item vector*. Item vector contains the details of each movie, like genre, cast, director, etc.
The content-based filtering algorithm finds the cosine of the angle between the profile vector and item vector, i.e. **cosine similarity**. Suppose A is the profile vector and B is the item vector, then the similarity between them can be calculated as:

$$sim(A,B) = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|}$$

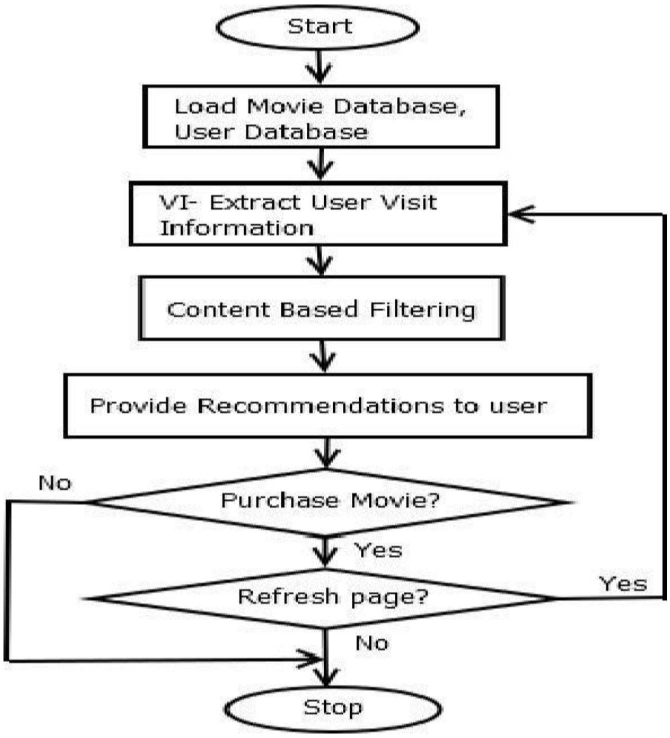
Based on the cosine value, which ranges between -1 to 1, the movies are arranged in descending order and one of the two below approaches is used for analysiss:

1. Top-n approach: where the top n movies are recommended (Here n can be decided by the business).
2. Rating scale approach: Where a threshold is set and all the movies above that threshold are recommended.

PROPOSED WORK:

These Movie Recommender Ststem analysis the movies based on user ratings by using concept of correleation and content-Based filtering algorithm. These movie recommender system provides movies according to the user liking and the main objective is to provide accurate movie analysis to the users. We used python technology for coding and executed in Spyder IDE.

Architecture: -



Software and Hardware Requirements:

Software

- Anaconda Tool
- Jupyter Notebook
- Windows 10

Hardware

- 1 TB Hard Disk
- 8 GB RAM

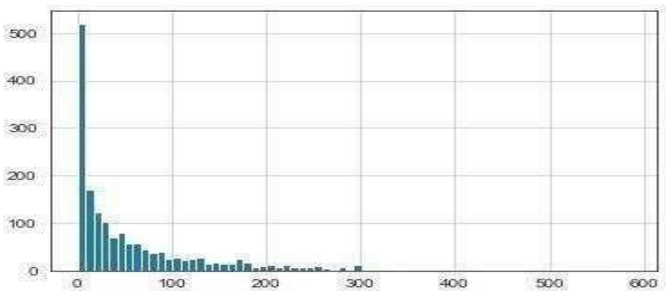
Sample Code: -

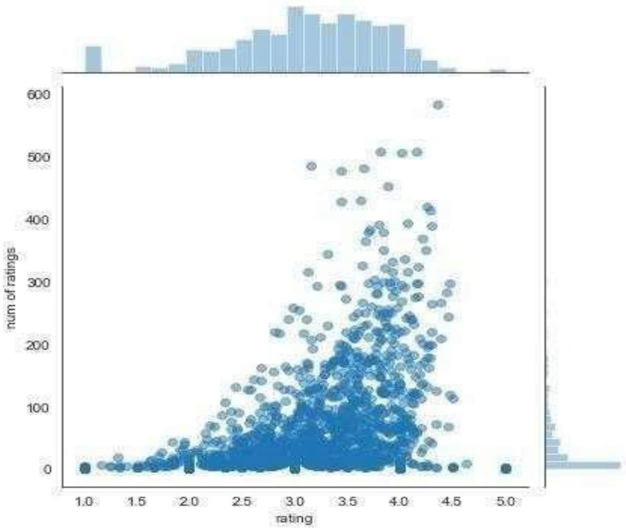
```
import pandas as pd

r_cols = ['user_id', 'movie_id', 'rating']
ratings = pd.read_csv('ml-100k/u.data', sep='\t', names=r_cols, usecols=range(3), encoding="ISO-8859-1")

m_cols = ['movie_id', 'title']
movies = pd.read_csv('ml-100k/u.item', sep='|', names=m_cols, usecols=range(2), encoding="ISO-8859-1")

ratings = pd.merge(movies, ratings)
print(ratings.head())

movieRatings = ratings.pivot_table(index=['user_id'],columns=['title'],values='rating')
print(movieRatings.head())

starWarsRatings = movieRatings['Star Wars (1977)']
starWarsRatings.head()

similarMovies = movieRatings.corrwith(starWarsRatings) # pairwise correlation of Star Wars vector of user rating with every other movie
similarMovies = similarMovies.dropna() # Drop any results that have no data
df = pd.DataFrame(similarMovies) # Construct a new Dataframe of movies and their correlation score to Star Wars
print(df.head(10))

print(similarMovies.sort_values(ascending=False))


import numpy as np
movieStats = ratings.groupby('title').agg({'rating': [np.size, np.mean]})
print(movieStats.head())

popularMovies = movieStats['rating']['size'] >= 100 # Ignore movies rated by less than 100 people
print(movieStats[popularMovies].sort_values([('rating', 'mean')], ascending=False)[:15])

df = movieStats[popularMovies].join(pd.DataFrame(similarMovies, columns=['similarity']))
print(df.head())

print(df.sort_values(['similarity'], ascending=False)[:15])
```

Results

Title	Correlation	Number of ratings
'Til There Was You (1997)	0.872872	9
1-900 (1994)	-0.645497	5
101 Dalmatians (1996)	0.211132	109
12 Angry Men (1957)	0.184289	125
187 (1997)	0.027398	41

Title	Correlation	Number of ratings
Star Wars (1977)	1.000000	584
Empire Strikes Back, The (1980)	0.748353	368
Return of the Jedi (1983)	0.672556	507
Raiders of the Lost Ark (1981)	0.536117	420
Austin Powers: International Man of Mystery (1997)	0.377433	130





Conclusion

- Here the accuracy depends on correlation result. +1 is perfect positive and-1 is perfect negative. In between values are correctly predicted value
- For the film star wars (1977), The recommended movies are Return of the jedi, Return ofthe jedi, Raiders of the Lost Ark, Austin powers, International man of mystery.
- So the accuracy is given correctly. and movies predicted based on pearson correlation.

Future Work

We plan to improve our recommendation system in the following two ways: first, make the latest films available for users instead of having outdated movies in the dataset; and second, optimize the selection of virtual users to better represent the entire set of real users in a cluster, by using for instance fuzzy C-means , which could further improve the recommendation accuracy.

References

[1] Darapaneni, Narayana, and Christopher Bellarmine. 2018. “Movie Success Prediction Using ML.” : 870–75.

[2] Farashah, Mohammadsadegh Vahidi, Akbar Etebarian, Reza Azmi, and Reza Ebrahimzadeh Dastjerdi. 2021. “A Hybrid Recommender System Based - on Link Prediction for Movie Baskets Analysis.” Journal of Big Data. https://doi.org/10.1186/s40537-021-00422-0.

[3] Chen, Xiaojie et al. 2019. “Exploiting Aesthetic Features in Visual Contents for Movie Recommendation.” IEEE Access 7: 49813–21.

[4] Lenskiy, Artem A, and Eric Makita. 2017. “Bayesian Approach to U Sers ’ Perspective on Movie Genres.” 15(1): 43–48.

[5] Li, Xiaoyue, Haonan Zhao, Zhuo Wang, and Zhezhou Yu. 2020. “Research on Movie Rating Prediction Algorithms.” : 121–25.

[6] Gao, Yuan et al. 2021. “An Attention-Based Unsupervised Adversarial Model for Movie Review Spam Detection.” 23: 784–96.

[7] Han, Yahong, Bo Wang, Richang Hong, and Fei Wu. 2020. “Movie Question Answering via Textual Memory and Plot Graph.” 30(3): 875–87.

[8] Haq, Ijaz U L, Student Member, and Khan Muhammad. 2019. “DeepStar : Detecting Starring Characters in Movies.” IEEE Access 7: 9265–72.

[9] Mulay, Rakshanda et al. 2020. “Movie Rating System Based on Sentiment Analysis.” 1(March).

[10] Khan, Umair A L I et al. 2020. “Movie Tags Prediction and Segmentation Using Deep Learning.” 8: 6071–86.

[11] Navarathna, Rajitha, Peter Carr, Patrick Lucey, and Iain Matthews. 2019. “To Predict Movie Ratings.” 10(1): 48–59.

[12] Kumar, Sudhanshu, Kanjar De, and Partha Pratim Roy. 2020. "Movie Recommendation System Using Sentiment Analysis From Microblogging Data." 7(4): 915–23.

[13] Reddy, Muppana Mahesh. 2020. "Analysis of Movie Recommendation Systems ; with and without Considering the Low Rated Movies." : 1–4.

[14] Singh, Vibhor, Priyansh Saxena, Siddharth Singh, and S Rajendran. 2017. "Opinion Mining and Analysis of Movie Reviews." 10(May).

[15] Steck, Harald, and Harald Steck. "Calibrated Recommendations." : 154–62.

[16] Somandepalli, Krishna, Naveen Kumar, and Tanaya Guha. 2018. "Unsupervised Discovery of Character Dictionaries in Animation Movies." 20(3): 539–51.

[17] Kharb, Latika. 2020. 1 Forecasting Movie Rating Through Data Analytics. Springer Singapore. http://dx.doi.org/10.1007/978-981-15-5830-6_21.

[18] Chavan, Sagar, Akash Morwal, Shivam Patanwala, and Prachi Janrao. "Sentiment Analysis of Movie Ratings System." : 43–47.

[19] Wi, Jeong A, Student Member, Soojin Jang, and Student Member. 2020. "Poster-Based Multiple Movie Genre Classification Using Inter-Channel Features." IEEE Access 8: 66615–24.

[20] Wu, Ting, Fei Hao, and Mijin Kim. 2021. "Typical Opinions Mining Based on Douban Film Comments in Animated Movies." Entertainment Computing 36(December 2019): 100391.

[21] Zhang, Jiang, Yufeng Wang, Zhiyuan Yuan, and Qun Jin. 2020. "Personalized Real-Time Movie Recommendation System : Practical Prototype and Evaluation." : 180–91.

[22] Zhang, Richong, and Yongyi Mao. 2019. "Movie Recommendation via Markovian Factorization of Matrix Processes." IEEE Access 7: 13189–99.

[23] Zhao, Wei et al. 2020. "Leveraging Long and Short-Term Information in Content-Aware Movie Recommendation via Adversarial Training." 50(11): 4680–93.

[24] Zhao, Zhou et al. 2018. "Social-Aware Movie Recommendation via Multimodal Network Learning." 20(2): 430–40.

[25] Navarathna, Rajitha, Peter Carr, Patrick Lucey, and Iain Matthews. 2017. "To Predict Movie Ratings." X(SEPTEMBER 2014): 1–11.