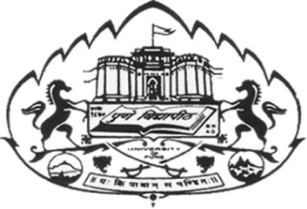
**Create a Movie Recommendation System.**

Mini Project Report submitted to Savitribai Phule Pune University, Pune



In partial Fulfillment for the awards of Degree of Engineering in

Computer Engineering BE(Computer)

**Submitted by**

Mrs. Roshani Dhule (Roll No. 21)

Mrs. Amruta Bhore(Roll No.10)

Mrs. Jyotirani Sahoo(Roll no.60)

**Under the Guidance of**

**Prof. Rupali Wagh**



# Department of Computer Engineering

Parvatibai Genba Moze College Of Engineering, Wagholi, Pune

(2022-23)

# Certificate



This is to certify that, Roshani, Amruta, Jyotirani (Roll No. 21,10,60), have successfully completed the Mini project entitled “**Movie Recommendation System.’’** under my guidance in partial fulfillment of the requirements for the Third Year of Engineering in Computer Engineering under the Savitribai Phule Pune University during the academic year 2022- 2023

|  |  |
| --- | --- |
| **Prof. Rupali Wagh** | **Prof. Shrikant Dhamdhere** |
| **Project Guide** | **Head Of Department** |

**Date : / /**

**Place :** Pune



This is to certify that, Mr Omkar Panchal (Roll No. 51), have successfully completed the Mini project entitled “**Create a dApp (Decentralized app) for E- Voting System.’’** under my guidance in partial fulfillment of the requirements for the Third Year of Engineering in Computer Engineering under the Savitribai Phule Pune University during the academic year 2022- 2023

|  |  |
| --- | --- |
| **Prof. Rupali Wagh** | **Prof. Shrikant Dhamdhere** |
| **Project Guide** | **Head Of Department** |

**Date : / /**

**Place :** Pune

# Acknowledgements

With deep sense of gratitude we would like to thank all the people who have lit our path with their kind guidance. We are very grateful to these intellectuals who did their best to help during our project work.

It is our proud privilege to express a deep sense of gratitude to **Prof. Dr. M . G. Jadhav,** Principal of Parvatibai Genba Moze College Of Engineering, Wagholi, Pune for his comments and kind permission to complete this project. We remain indebted to **Prof. Shrikant Dhamdhere,** H.O.D. Computer Engineering Department for his timely suggestion and valuable guidance. The special gratitude goes to (**Prof. Rupali wagh**) excellent and precious guidance in completion of this work .We thanks to all the colleagues for their appreciable help for our working project. With various industry owners or lab technicians to help, it has been our endeavor throughout our work to cover the entire project work.

We are also thankful to our parents who provided their wishful support for our project completion successfully .And lastly we thank our all friends and the people who are directly or indirectly related to our project work.

Mrs. Roshani Dhule (Roll No. 21)

Mrs. Amruta Bhore (Roll No.10)

Mrs. Jyotirani Sahoo (Roll No .60)

# Abstract

Recommender System is a tool helping users find content and overcome information overload. It predicts interests of users and makes recommendation according to the interest model of users. The original content-based recommender system is the continuation and development of collaborative filtering, which doesn’t need the user’s evaluation for items. Instead, the similarity is calculated based on the information of items that are chose by users, and then make the recommendation accordingly. With the improvement of machine learning, current content-based recommender system can build profile for users and products respectively. Building or updating the profile according to the analysis of items that are bought or visited by users. The system can compare the user and the profile of items and then recommend the most similar products. So this recommender method that compare user and product directly cannot be brought into collaborative filtering model. The foundation of content-based algorithm is acquisition and quantitative analysis of the content. As the research of acquisition and filtering of text information are mature, many current content-based recommender systems make recommendation according to the analysis of text information. This paper introduces content-based recommender system for the movie website of VionLabs. There are a lot of features extracted from the movie, they are diversity and unique, which is also the difference from other recommender systems. We use these features to construct movie model and calculate similarity. We introduce a new approach for setting weight of features, which improves the representative of movies. Finally we evaluate the approach to illustrate the improvement.

**Table of Contents**

**Acknowledgements ..........................................................................................** 3

**Abstract .............................................................................................................** 4

**Chapter 1....................................................................................................**

**Introduction...................................................................................................**

1.1Background**..................................................................................**

1.2ProblemStatement **......................................................................**

1.3Goals**................................................................................................**

1.4Methodology**..................................................................................**

1.5Ethics **............................................................................................**

1.6 Delimitations**.................................................................................**

1.7 Outline**..........................................................................................**

**Chapter 2**

**Overview of Related Work**

2.1 Content Recommendation System………………………………

2.2 Collaborative – Filtering Recommendation**……………………..**

2.2.1 User-based collaboration -filtering**……………………………**

2.3 Famous Recommender System**………………………………..**

2.3.1 Movie and Video website**……………………………………..**

2.3.2 E-Commerce**……………………………………………………**

2.3.3 Internet Radio**………………………………………………….**

**Chapter 3**

**3.1 Advantages……………………………………………………………..**

**3.2 Disadvantages………………………………………………………….**

**Chapter 4**

**Software Requirement Specification**

**4.1 Hardware Requirement……………………………………..**

**4.2 software Requirements ………………………………………….**

**Chapter 5**

**Summary……………………………………………………………**

**Chapter 6**

**6.1 conclusion…………………………………………………..**

**6.2 scope………………………………………………………..**

# References………………………………………………………….

**Chapter 1**

**Introduction**

As the development of information technology and Internet, people enter the era of information overload from that of information deficiency gradually[17]. The report is the result of the Master Thesis in Information and Communication Technology School at Royal Institute of Technology in Stockholm, Sweden. The project is provided in VionLabs AB that is a media-tech company providing media content such as movie information for customers. The recommender system I implemented is for the movie website Vionel.com.

* 1. **Background**

In the era of information overload, it is very difficult for users to get information that they are really interested in. And for the content provider, it is also very hard for them to make their content stand out from the crowd. That is why many researchers and companies develop Recommender System to solve the contradiction. The mission of Recommender System is to connect users and information, which in one way helps users to find information valuable to them and in another way push the information to specific users. This is the win-win situation for both customers and content providers. VionLabs is a media-tech startup company. The company provides a new way on how consumers are given access to good and suitable content. The mission of VionLabs is to increase needs of its digital user base. Vionel is the movie website developed by VionLabs, which is a place for people who love movies can gather all the information about films in one place[5]. This thesis report will present a more practical recommendation method that can be used on a movie website that does not have enough users.

* 1. **Problem Statement**

For building a recommender system from scratch, we face several different problems. Currently there are a lot of recommender systems based on the user information, so what should we do if the website has not gotten enough users. After that, we will solve the representation of a movie, which is how a system can understand a movie. That is the precondition for comparing similarity between two movies. Movie features such as genre, actor and director is a way that can categorize movies. But for each feature of the movie, there should be different weight for them and each of them plays a different role for recommendation. So we get these questions:

• How to recommend movies when there are no user information.

• What kind of movie features can be used for the recommender system.

• How to calculate the similarity between two movies.

• Is it possible to set weight for each feature

**1.3 Goals**

The goals of this thesis project is to do the research of Recommender Systems and find a suitable way to implement it for Vionel.com. There are many kinds of Recommender Systems but not all of them are suitable for one specific problem and situation. Our goal is to find a new way to improve the classification of movies, which is the requirement of improving content-based recommender systems.

**1.4 Methodology**

In order to achieve the goal of the project, the first process is to do enough background study, so the literature study will be conducted. The whole project is based on a big amount of movie data so that we choose quantitative research method. For philosophical assumption, positivism is selected because the project is experimental and testing character. The research approach is deductive approach as the improvement of our research will be tested by deducing and testing a theory. Ex post facto research is our research strategy, the movie data is already collected and we don’t change the independent variables. We use experiments to collect movie data. Computational mathematics is used data analysis because the result is based on improvement of algorithm. For the quality assurance, we have a detail explanation of algorithm to ensure test validity. The similar results will be generated when we run the same data multiple times, which is for reliability. We ensure the same data leading to same result by different researchers.

**1.5 Ethics**

Movie information is the only part that may have ethics problem. However, all the information we get for research is from public database such as Wikipedia and our own movie database. So there are no data confidentiality and user privacy problems.

**1.6 Delimitations**

In this project, we will not deep in how to collect and generate the data, which is the business of other teams in VionLabs. In the validation part, the data is divided into training and testing parts. But how to split data will not presented in detail neither. It is important to build a real recommender system for a business company and in the end I have built a successful demo which helped our company to get the opportunity of cooperation with other companies. But in the thesis report, I will focus on the improvement of movie representation and explain why our new approach is more helpful to recommender system. I will not talk about the detail of implementation of the system because of confidentiality agreement.

**1.7 Outline**

In the first chapter, background of the project is introduced. After that, the layout of the report is as follows: Chapter 2 provides an overview of related work on recommender systems. I will introduce the mainstream approaches and some famous business recommender systems. Chapter 3 shows the basic principles of our recommender system and how we improve it. Chapter 4 I will illustrate how recommendation works by a specific movie. Chapter 5 is the validation part, I will present our test result to illustrate the improvement of our approach. In the end, Chapter 6 and 7 is the conclusion and possible future work respectively

# Chapter2

**Overview of Related Work**

Recommender system is a very hot research topic in recent years. Many researchers raised a lot of different recommendation approaches. The most famous category of these approaches is:

• Content-based Recommendation.

• Collaborative-filtering Recommendation.

• Hybrid Recommendation.

**2.1 Content-based Recommendation**

Content-based recommendation is an important approach in recommender systems. The basic idea is to recommend items that are similar with what user liked before[21]. The core mission of content-based recommender system is to calculate the similarity between items. There are a lot of methods to model item and the most famous one is Vector Space Model[2]. The model extracts keywords of the item and calculate the weight by TF-IDF. For example, set ki as the ith keyword of item dj , wij is the weight of ki for dj , then the content of dj can be defined as:

Content(dj ) = {w1j , w2j , ...}

As we talked before, content-based recommender system recommends items that are similar with what user liked before. So the tastes of a user can be modeled according to the history of what the user liked. Consider ContentBasedP rof ile(u) as the preference vector of user u, the definition is:

ContentBasedP rof ile(u) = 1 |N(u)| sumd∈N(u)Content(d)

N(u) is what the user u liked before. After calculating content vector Content(.) and content preference vector ContentBasedP rof ile(.) of all users, given any user u and an item d, how the user like the item is defined as the similarity between ContentBasedP rof ile(u) and Content(d):

p(u, d) = sim(ContentBasedP rof ile(u), Content(d))

Using keywords to model item is an important step for many recommender systems. But extracting keywords of an item is also a difficult problem, especially in media field, because it is very hard to extract text keywords from a video. For solving this kind of problem, there are two main ways. One is letting experts tag the items and another one is letting users tag them. The representative of experttagged systems are Pandora for music and Jinni for movies. Let’s take Jinni as an example, the researchers of Jinni defined more than 900 tags as movie gene, and they let movie experts to make tags for them. These tags belong to different categories, including movie genre, plot, time, location and cast. Figure 2.1 is from Jinni, which are the tags for movie Kung Fu Panda. As we can see from the figure, the tags of Kung Fu Panda are divided into ten categories totally, Mood, Plot, Genres, Time, Place, Audience, Praise, Style, Attitudes and Look. These tags contain all aspects of movie information, which can describe a movie very accurately.

Compared with expert-tagged system, user-tagged system is applied more widely. The representative websites are Delicious and Flickr. The feature of user-tagged system is the tags are more diversity than that of expert-tagged system. But the weakness is that the tags are of lower quality, even there are a lot of wrong tags. So in the user-tagged systems, there are two main problems, one is tag recommendation[34], which means when a user tags an item, the system can recommend some relative tags for him to choose. The purpose is first to be convenience for users and second it can increase the quality of tags. Another question is how to recommend items based on tags(tag-based recommendation[33]). After items are tagged, the simplest recommendation approach is to use tags as keywords of the item, and recommend by the content-based algorithm.

**2.2 Collaborative-filtering Recommendation**

Collaborative-filtering recommendation is the most famous algorithm in recommender systems. This algorithm models user’s taste according to the history of user behavior. GroupLens published the first paper[31] about collaborative filtering and the paper raised user-based collaborative filtering. In 2000, Amazon came up with item-based collaborative filtering in their paper[20]. These two algorithms are very famous in business recommender systems.

**2.2.1 User-based collaborative-filtering**

In user-based collaborative filtering, it is considered that a user will like the items that are liked by users with whom have similar taste. So the first step of user-based collaborative-filtering is to find users with similar taste. In collaborative filtering, the users are considered similar when they like similar items. Simply speaking, given user u and v, N(u) and N(v) are items set liked by u and v respectively. So the similarity of u and v can be simply defined as:

suv = |N(u) ∩ N(v)| |N(u) ∪ N(v)|

There are a lot of similarity algorithm, Equation 2.4 is one of them. User u’s likeability for item i can be calculated by:

pui = X v∈S(u,k)∩N(i) suvpvi (2.5)

**2.3.Famous Recommender Systems**

What is the difference between recommender system and search engine is that recommender system is based on the behaviors of user. There are a lot of websites using recommender system in the world. Personalized recommender system analyzes a huge amount of user behavior data and provides personalized content to different users, which improves the click rate and conversions of the website[36]. The fields that widely use recommender system are e-commerce, movie, video, music, social network, reading, local based service, personalized email and advertisement.

**2.3.1 Movie and Video website**

Personalized recommender system is a very important application for movie and video website, which can help users to find what they really like among the vast of videos. Netflix is the most successful company in this field[24]. Amazon and it are the two most representative companies in recommender systems .

**2.3.2 E-Commerce**

The most famous e-commerce website, Amazon, is the active application and promoter of recommender system. The recommender system of Amazon reaches deeper into all kinds of products[32]. Figure 2.2 is the recommendation list of Amazon. Apart from personalized recommendation list, another important application of recommender system is relevant recommendation list. When you buy something in Amazon, the relevant goods will be shown below [15]. Amazon has two kinds of relevant recommendation, one is customers who bought this item also bought in Figure 2.3. Another is what other items do customers buy after viewing this item in Figure 2.4. The difference between the two recommendations is the calculation of the different user behaviors. The most important application of relevant recommendation is cross selling. When you are buying something, Amazon will tell you what other customers who bought this item also bought and let you decide whether buy it at the same time[7]. If you do, the goods will be packed and provide a certain discount.

**2.3.3 Internet Radio**

The successful application of personalized recommender system needs two requirements, one is information overload because if users can find what they like easily, there is no reason to use recommender systems. The second

is the user doesn’t have clear requirements. Because they will use search engine directly if they do.

Under the two requirements, recommender system is very suitable for personalized Internet radio. First of all, people cannot listen to all the music in the world and find which one they like. Secondly, people often do not want to listen to a specific music, they wish to listen to whatever musics that match their mood at that moment

**Chapter 3**

**3.1 Advantages**

* Recommendation result is intuitive and easy to interpret;
* No need for users? access history data;
* No new item problem and no sparsity problem
* Supported by the mature technology of classification learning

* No need for professional knowledge;
* Performance improving as the increasing of the user number;
* Automatic; Easy to find user?s new interesting point;
* Complex unstructured item can be processed. eg. Music, Video, etc.

**3.2 Disadvantages**

# Limited by the features extraction methods;

# New user problem;

# The training of classifier needs massive data;

# Poor scalability

* Sparsity problem;
* Poor scalability;
* New user and new item problem;
* The recommendation quality limited by the history data set.

# Chapter 4

**Software Requirement Specification**

**4.1 Hardware Requirement**

* Higher Processor of 2.4 GHZ speed
* 8GB RAM maximum
* 80GB maximum disk space

**4.2 software Requirements**

* Operating System (Windows 10 or later)
* Visual Studio
* Colab

# Chapter 5

**Summary**

A recommendation system or recommendation engine is a model used for information filtering where it tries to predict the preferences of a user and provide suggests based on these preferences. These systems have become increasingly popular nowadays and are widely used today in areas such as movies, music, books, videos, clothing, restaurants, food, places and other utilities. These systems collect information about a user's preferences and behaviour, and then use this information to improve their suggestions in the future.

Movies are a part and parcel of life. There are different types of movies like some for entertainment, some for educational purposes, some are animated movies for children, and some are horror movies or action films. Movies can be easily differentiated through their genres like comedy, thriller, animation, action etc. Other way to distinguish among movies can be either by releasing year, language, director etc. Watching movies online, there are a number of movies to search in our most liked movies . Movie Recommendation Systems helps us to search our preferred movies among all of these different types of movies and hence reduce the trouble of spending a lot of time searching our favourable movies. So, it requires that the movie recommendation system should be very reliable and should provide us with the recommendation of movies which are exactly same or most matched with our preferences.

A large number of companies are making use of recommendation systems to increase user interaction and enrich a user's shopping experience. Recommendation systems have several benefits, the most important being customer satisfaction and revenue. Movie Recommendation system is very powerful and important system. But, due to the problems associated with pure collaborative approach, movie recommendation systems also suffers with poor recommendation quality and scalability issues.

**Chapter 6**

# 

# Conclusion and Future Scope

**6.1 Conclusion**

Conclusion Recommender system has become more and more important because of the information overload. For content-based recommender system specifically, we attempt to find a new way to improve the accuracy of the representative of the movie. For the problems we mentioned at beginning, firstly, we use content-based recommender algorithm which means there is no cold start problem. In Section 4.1, we list all the features in our recommender system. Some of them are from other research team in the company, so the features are diversity and more accurate than others. Then we introduced the cosine similarity which is commonly used in industry. For the weight of features, we introduced TF-IIDF-DC which improve the representative of the movie. This master thesis introduces a content-based recommender system for the movie website of VionLabs. The features used in the system are extracted from various aspects of the movie, which are diversity and unique. We introduce a new approach for setting weight for these features, the movie can be represented more accurately by TF-IIDF-DC which is the key point of our research. In the end of the project, we use k-NN and various metrics to evaluate the improvement of the new approach. It is illustrated that the new approach contributes positively according to the evaluation.

**6.2 Scope**

Future Work Recommender system has developed for many years, which ever entered a low point. In the past few years, the development of machine learning, large-scale network and high performance computing is promoting new development in this field. We will consider the following aspects in future work.

• Use collaborative filtering recommendation. After getting enough user data, collaborative filtering recommendation will be introduced. As we discussed in Section 2.2, collaborative filtering is based on the social information of users, which will be analyzed in the future research.[37]

• Introduce more precise and proper features of movie.[1] Typical collaborative filtering recommendation use the rating instead of object features. In the future we should extract features such as color and subtitle from movie which can provide a more accurate description for movie.

• Introduce user dislike movie list. The user data is always useful in recommender systems. In the future we will collect more user data and add user dislike movie list. We will input dislike movie list into the recommender system as well and generate scores that will be added to previous result. By this way we can improve the result of recommender system.

• Introduce machine learning. For future study, dynamic parameters will be introduced into recommender system, we will use machine learning to adjust the weight of each feature automatically and find the most suitable weights.

• Make the recommender system as an internal service. In the future, the recommender system is no longer a external website that will be just for testing. We will make it as an internal APIs for developers to invoke. Some movie lists in the website will be sorted by recommendation.

**References**

[1] Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. Knowledge and Data Engineering, IEEE Transactions on, 17(6):734–749, 2005.

[2] Ricardo Baeza-Yates, Berthier Ribeiro-Neto, et al. Modern information retrieval, volume 463. ACM press New York, 1999.

[3] Shumeet Baluja, Rohan Seth, D Sivakumar, Yushi Jing, Jay Yagnik, Shankar Kumar, Deepak Ravichandran, and Mohamed Aly. Video suggestion and discovery for youtube: taking random walks through the view graph. In Proceedings of the 17th international conference on World Wide Web, pages 895–904. ACM, 2008.

[4] Robert Bell, Yehuda Koren, and Chris Volinsky. Modeling relationships at multiple scales to improve accuracy of large recommender systems. In Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 95–104. ACM, 2007.

[5] Suvir Bhargav. Efficient features for movie recommendation systems. 2014.

[6] Robin Burke. Hybrid recommender systems: Survey and experiments. User modeling and user-adapted interaction, 12(4):331–370, 2002.

[7] Robin Burke. Hybrid web recommender systems. In The adaptive web, pages 377–408. Springer, 2007.

[8] James Davidson, Benjamin Liebald, Junning Liu, Palash Nandy, Taylor Van Vleet, Ullas Gargi, Sujoy Gupta, Yu He, Mike Lambert, Blake Livingston, et al. The youtube video recommendation system. In Proceedings of the fourth ACM conference on Recommender systems, pages 293–296. ACM, 2010.

[9] Scott C. Deerwester, Susan T Dumais, Thomas K. Landauer, George W. Furnas, and Richard A. Harshman. Indexing by latent semantic analysis. JAsIs, 41(6):391–407, 1990.

[10] Francois Fouss, Alain Pirotte, Jean-Michel Renders, and Marco Saerens. Random-walk computation of similarities between nodes of a graph with ap39 BIBLIOGRAPHY plication to collaborative recommendation. Knowledge and data engineering, ieee transactions on, 19(3):355–369, 2007.