A

Mini Project

On

**PREDICTING THE TOP-N POPULAR VIDEOS VIA CROSS-DOMAIN**

**HYBRID MODEL**

(Submitted in partial fulfillment of the requirements for the award of Degree)

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

By

ROHIT MOON (177R1A05G4)

Under the Guidance of

**SANJANA NAZARE**

(Assistant Professor)



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

# CMR TECHNICAL CAMPUS

# UGC AUTONOMOUS

(Accredited by NAAC, NBA, Permanently Affiliated to JNTUH, Approved by AICTE, New Delhi) Recognized Under Section 2(f) & 12(B) of the UGCAct.1956, Kandlakoya (V), Medchal Road, Hyderabad-501401.

**2019-2023**

**ABSTRACT:**

Predicting the performance of popular videos and their future views for a large batch of newly uploaded videos is of great commercial value to online video services (OVSs). Although many attempts have been made on video popularity prediction, the existing models has a much lower performance in predicting the popular videos than that of the entire video set. The reason for this phenomenon is that most videos in an OVS system are unpopular, so models preferentially learn the popularity trends of unpopular videos to improve their performance on the entire video set. However, in most cases, it is critical to predict the performance on the popular videos which is the focus of this study. The challenge for the task are as follows. First, popular and unpopular videos may have similarity view patterns. Second, prediction models that are overly dependent on early view patterns limit the effects of other features. To address these challenges, we propose a novel multifactor differential influence (MFDI) prediction model based on multivariate linear regression (MLR). The model is designed to improve the discovery of popular videos and their popularity trends are learnt by enhancing the discriminative power of early patterns for different popularity trends and by optimizing the utilization of multisource data. We evaluate the proposed model using real-world YouTube data, and extensive experiments have demonstrated the effectiveness of our model.

**INTRODUCTION:**

Since the popularity of online videos has been proven to be predictable through the statistical analysis of large-scale YouTube data, numerous related studies have been conducted. Szabo and Huberman (S-H) proposed a content-scaling (CS) model based on log-transformed relations between a video’s long-term popularity and its early popularity Their conclusion is one of the most important foundations of popularity prediction research and has been succeed by many related works. All of the approaches cited above achieved initial success, but their shortcomings has been uncovered by subsequent research. Popularity prediction of online videos, especially the prediction of the top-N popular videos is of great importance to support the development of online video services (OVSs).

**EXISTING SYSTEM:**

Popularity prediction of online videos, especially the prediction of the top-N popular videos is of great importance to support the development of online video services (OVSs). From the perspective of better user experience, the ability to identify the top-N popular videos is beneficial to video services, such as caching and recommendation. From the perspective of commercialization, identifying the top-N popular videos helps the video service providers to maximize their profits, as advertisers are more likely to pay more for popular videos. Although many attempts have been made on popularity prediction of online videos, because most of the videos in an OVS system are unpopular; consequently, models preferentially learn the popularity trends of these unpopular videos to achieve better performance on the video set as a hole.

**DISADVANTAGES:**

* First, popular and unpopular videos may have similar early view patterns, and this similarity limits the performance benefit of video classification based on early view patterns.
* Second, existing studies show that the strong correlation between early views and long-term popularity dominates the training of the prediction models.

**PROPOSED SYSTEM:**

We propose a model for predicting the top-N popular videos .By enhancing the ability of early patterns to distinguish among popularity trends. We evaluate the proposed model using real-world data consisting of videos form YouTube and social network data from twitter. The most video in OVS system is unpopular. So models learns the popularity trend of unpopular videos to improve their performance.

**ADVANTAGES:**

* Proposed model outperforms state of art models, there by confirming the benefits of our efforts to improve the prediction performance for the top-N popular videos.
* The model is designed to improve discovery of popular videos and their popularity trends.

**HARDWARE REQUIREMENTS:**

* + System : Minimum of Intel core i3 processor.
  + Hard disk : 4 GB
  + Monitor : 14’ Colour Monitor.
  + Mouse : Optical Mouse.
  + RAM : 4 GB

**SOFTWARE REQUIREMENTS:**

* + Operating system : Windows 7 Ultimate.
  + Coding Language : Python.
  + Designing : HTML, CSS and JavaScript.
  + Data Base : MySQL

**CONCLUSION:**

In this project, we have investigated the problem of top-N popular video prediction and have proposed a novel MFDI Prediction model. The proposed model predicts the top-N popular videos by enhancing the ability of early patterns to identify different popularity trends and by optimizing the model’s utilization of multi-source data. Experimental results obtained using real-world data demonstrate that the proposed model outperforms other models, including the state-of-the-art model.

**NOVELTY:**

The Internet has become one of the primary means of communication and information in the world With the popularization of the Internet, streaming video services, such as YouTube, Netflix, GloboPlay, and Amazon Prime, have also grown. The Internet’s popularization has increased the amount of content produced and consumed on the web. To take advantage of this new market, major content producers such as Netflix and Amazon Prime have emerged, focusing on video streaming services. However, despite the large number and diversity of videos made available by these content providers, few of them attract the attention of most users. For example, in the data explored in this article, only 6% of the most popular videos account for 85% of total views. Finding out in advance which videos will be popular is not trivial, especially given many influencing variables. Nevertheless, a tool with this ability would be of great value to help dimension network infrastructure and properly recommend new content to users. In this way, this manuscript examines the machine learning-based approaches that have been proposed to solve the prediction of web content popularity. To this end, we first survey the literature and elaborate a taxonomy that classifies models according to predictive features and describes state of-the-art features and techniques used to solve this task. While analyzing previous works, we saw an opportunity to use textual features for video prediction. Thus, additionally, we propose a case study that combines features acquired through attribute engineering and word embedding to predict the popularity of a videos.

To, Date: 30-07-2022

The PRC Incharge,

J. Narsimha Rao,

Associate Professor,

CMR Technical Campus

Subject: Requesting Letter of Acceptance for the mini project.

Respected sir,

I am Rohit Moon from IV- year, department of Computer science and engineering , I have explained the project idea on “PREDICTING THE TOP-N POPULAR VIDEOS VIA CROSS-DOMAIN HYBRID MODEL”. My kind request you to provide your acceptance for the mini project.

Thanking You,

Guide: Yours Obediently,

Sanjana Nazare 177R1A05G4