International Journal of Research
Available at https://www.internationaljournalofresearch.org/
ISSN: 2236-6124
Volume 09 Issue 01
Jan-June 2019

Prediction Popularity of Online Videos

PARRIPATI V R ANUSHA SRIPRIYA **V KESAVA KUMAR**

B. Tech Student, Department of CSE, Sri Mittapalli College of Engineering, Tummalapalem, NH-16, Guntur, Andhra Pradesh, India.

Assistant Professor, Department of CSE, Sri Mittapalli College of Engineering, Tummalapalem, NH-16, Guntur, Andhra Pradesh, India.

Abstract **-** Big data is becoming an trending concept in today's world. The predictive analysis is the future growth of organization. Predictive analysis focus on 3 main things historical data, current data and user statistics. Predictive analytics use machine learning techniques to identify the future outcomes based on historical data. Watching YouTube videos become an main part in our daily lives. A large amount of video content has been created on the Internet, and user interests among those videos is assign in an unequal way, with the vast majority of users interests noticed while a few of videos become very popular. Predicting videos popularity is of great importance for advertising and recommendation systems.

The proposed work is to predict YouTube videos popularity by using machine learning techniques logistic regression and decision tree algorithms based on video popularity parameters age, length, views, rate, ratings and comments. In the model popularity is dependent variable and the parameters are independent variables.

Keywords *:- Online, prediction, videos, you tube data.*

I INTRODUCTION

With the ubiquitous access of Internet and the emergence of Web 2.0 services, an enormousand ever growing amount of online content has been brought into the digital world. Content producers now can reach audiences in inconceivable numbers that are unmatched through conventional channels. Among the various kinds of online content, online videos are currently a dominating component of the Internet. In terms of bytes, the video traffic accounts for around 64% of all the Internet traffic in 2014, and will be up to 80% by 2019 [1]. This explosive growth intensifies the online competition for user attention, and contributes to a ''winner-take-all'' online video ecosystem: a small fraction of videos attract most of the user interest, whereas the vast majority of videos are of limited views [2], [3]. Given the huge amount of video content and the high variability of user attention, it is of utmost importance for a number of tasks to understand the characteristics of online video popularity and further predict the popularity of individual videos. For service providers, the video popularity dynamics and prediction results can greatly benefit their future design of the content filtering, video ranking, and recommendation schemes, which help users to find videos with more potential values more easily[4] .For advertisers in the

online marketing, prediction of the next rising star of the Internetcanmaximizetheirrevenuesthroughbettera dvertisingplacement[3].With the extrapolation of video popularity, network operators can proactively manage the bandwidth requirement and deploy the cache servers in the content delivery network (CDN) for hot videos in advance[5]. In addition, in a quite different context, the video popularity will be of great interest in the opportunistic communications among mobile devices [6]. In such resource-constrained environments (e.g. limited bandwidth and storage space), predicting hot videos is helpful for the content delivering, caching and replicating on the device end. In this paper, we study the video popularity of Youku (www.youku.com), a leading online video service provider in China. Our work is based on the data of 33,359 videos crawled from Youku website for 30 consecutive days. With these data, we analyze in-depth how the popularity of online video content evolves overtime, and how to predict the future popularity of an individual video. The main contributions of our work are s ummarized as follows:

1) We provide a detailed characterization of the popularity dynamics of online videos. In particular, we provide insights into the popularity evolution patterns of the individual videos.

2) We tackle the problem of popularity prediction by proposing a model that can capture the popularity evolution of an individual video. To the best of our knowledge, the proposed method is the first to specialize models by popularity evolution patterns in the popularity prediction.

3) We evaluate our model on a real-world dataset and compare the prediction performance with two state of-the-art online video popularity prediction models. Our approach leads to significant reductions in prediction errors of32.73%and11.28% over the base line models respectively.

4) We analyze the potentials and limitations of different detectors and model parameters in the prediction. We shed light on the importance of each feature and each feature group used in the burst prediction.

.II RELATED WORK

A. ANALYSIS OF GROUP POPULARITY The beginning research of video popularity can be found in the early studies on user access patterns. Gill et al. [7] analyzed the user access patterns, file properties, video view count distribution and referencing behaviors of YouTube, based on the traffic collected at a campus. Zink et al. [8] also captured traffic from a university campus and analyzed the duration, data rate, population and access patterns of YouTube videos. Both of them discovered that the video requests of YouTube followed a Zipf-like distribution. Chaetal. [4]analyzed the popularity distribution, popularity evolution and content characteristics of YouTube and a popular Korean video sharing service. They modeled the group video popularity as a power law with an exponential cut-off. They also investigated different mechanisms, such as caching and P2P, to improve the video distribution. Cheng et al. [9] performed a long-term crawling of YouTube and studied the static properties, access pattern, popularity distribution, popularity trend and social networking of YouTube videos.These previous studies provide important insights into the traffic and content popularity of

online video service. However, they focused mostly on the overall popularity distribution of a group of videos, and they collected data at either a single or several static time points (i.e. snapshots). Our study complements these existing works by further characterizing the video popularity at a more

fine-grained individual level besides the group level. Moreover, we track the popularity of each video every day since their publications, and analyze the popularity growth over time during the whole

observation period. We also further tackle the challenge of future popularity prediction in our paper.

B. CHARACTERIZATION OF POPULARITY EVOLUTION Some studies perform temporal analysis of how video popularity evolves over time, especially on the peak days, and clustervideoswhichhavesimilarpopularityevoluti ontrends. Crane and Sornette [10] first proposed epidemic models to describe the popularity evolution of YouTube videos. They distinguished four different evolution patterns based on the fraction of views on the peak day and explained the sepatterns in terms of endogenous and exogenous effects. Yang et al.
[11] studied how the popularity of online content grew and faded over time, and proposed a clustering algorithm based on the k-means method to identify the temporal patterns of video popularity. Figueiredo et al. [12]–[14] characterized the popularity evolution patterns of YouTube videos based on the classification method in[10],and studied the impacts of different types of referrers on such patterns. Ahmedetal.[15] identified the patterns of temporal evolution for distinct types of data overtime and predicted the evolution pattern of popularity in user generated content. Those patterns proposed by the previous works can well describe the evolution of video popularity. However, they focus more on the popularity growth shapes near the (single) peak day. In our study, we complement the definition of popularity evolution pattern, by considering thenumber and temporal locations of the popularity bursts throughout the whole observation period. In particular, we analyze the impacts of different popularity evolution patterns on the long-term video popularity.

III DESIGN OF THE WORKFLOW:

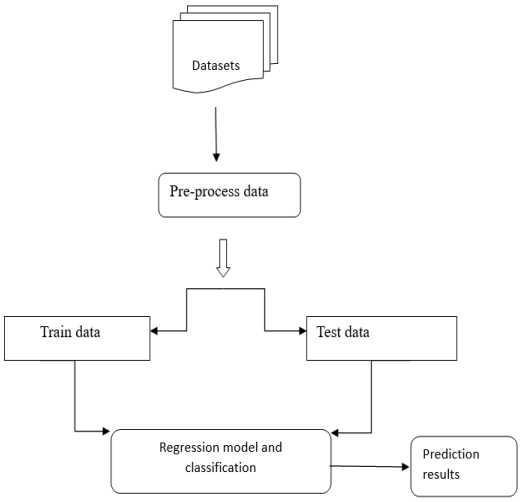


Figure 3.1 System architecture of proposed system

ISSN: 2236-6124
Volume 09 Issue 01
Jan-June 2019

Algorithm 1: Logistic regression outcome variable is a discrete variable. It measures the relationship between categorical dependent variable and one or more predictor variables. If we want to work with more than 2 variables then we use logistic regression model.

Algorithm 2: Decision tree is a type of supervised learning algorithm (having a pre-defined target variable) that is mostly used in classification problems. It works for both categorical and continuous input and output variables.

IV METHODOLOGY

1. Data collection

The YouTube data is downloaded from the website http://netsg.cs.sfu.ca/youtubedata/ and saved in the form of .csv file, which contains

50,713 records of video meta data. This data is used for predicting the video popularity.

2. Data preprocessing

In this module the unwanted data is cleaned from the dataset and labeling column names (i.e; video ID, up loader, category, length, age, rate, length, views, comments) are video parameters of YouTube data. These parameters are used for finding the video popularity.

3. Popularity prediction

The new popularity function is designed as shown in below Eq. 3.12, this is the Threshold value. YouTube data is split into train data and test data. For train data giving some parameters measures using glm model and predicting for a new data weather the video is popular or not. Next to the test data prediction is done that which video is getting popular and not popular using logistic and decision tree. Finally, the test data is compared with actual data to find that video popularity prediction is done correct or not.

$$Prpop = \alpha * mean(nbviews) + \beta * mean(nbcomments) + \gamma * mean(nbrating) + \delta * mean(nbrate) \quad \text{Eq. 3.12}$$

4. Prediction Results

The prediction results are which video is getting popular and not popular. Valid the results with using sensitivity and accuracy for logistic regression and decision tree models with all video parameters

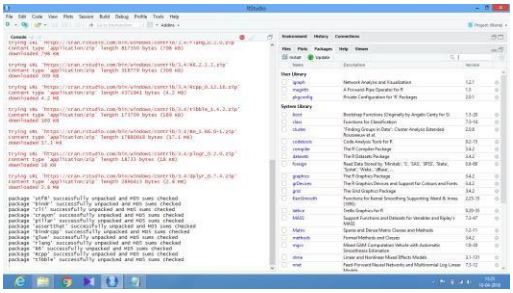
5. Generate the reports or Graphs

We will generate the graphs or reports using prediction results.

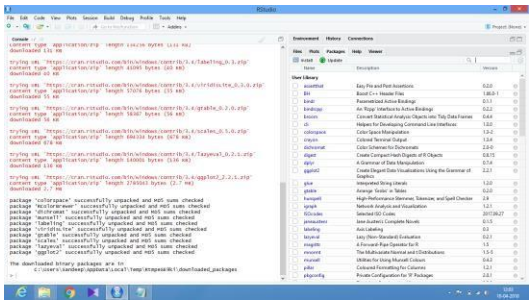
International Journal of Research
Available at
https://www.internationaljournalofresearch.org/

ISSN: 2236-6124
Volume 09 Issue 01
Jan-June 2019

V EXPERIMENT RESULTS



Data to console



Data Processing

VI CONCLUSION

The Data collection of YouTube video meta data which contains 5452 records, followed by preprocessing and labeling with the column names (i.e. Video parameters) age, rate, length, views, rate, ratings and comments. Video parameters used to predict the video popularity. In next step, apply logistic regression glm model and decision tree classification to the YouTube data for finding video popularity. The popularity is considered as the dependent variable in model and all the remaining parameters of the dataset are considered as the independent variables.

The result analysis is changing by combination of parameters. If all combination are used, the accuracy will be 0.71, if length, views, ratings are taken, the accuracy will be 0.85.

VII FUTURE ENHANCEMENT

We can use our model with any fields like popular songs prediction, popular Games prediction etc.

VIII REFERENCES

[1] Cisco. Cisco Visual Networking Index: Forecast and Methodology, 2014–2019 White Paper. [Online]. Available: http://www. cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-nextgeneration-network/white_ paper_c11-481360.html

[2] F.WuandB. A.Huberman, ‘‘Noveltyandcollectiveattention,’’ Proc.Nat. Acad. Sci. USA, vol. 104, no. 45, pp. 17599–17601, 2007.

[3] G. Szabo and B. A. Huberman, ‘‘Predicting the popularity of online content,’’ Commun. ACM, vol. 53, no. 8, pp. 80–88, 2010.

[4] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon, ‘‘Analyzing the video popularity characteristics of large-scale user generated content systems,’’ IEEE/ACM Trans. Netw., vol. 17, no. 5, pp. 1357–1370, Oct. 2009.

[5] Y. Zhou, L. Chen, C. Yang, and D. M. Chiu, ‘‘Video popularity dynamics and its implication for replication, ’’IEEE Trans.Mul timedia,vol.17,no.8, pp. 1273–1285, Aug. 2015.

[6] B.Han,P.Hui,V.S.A.Kumar,M.V.Marathe,J.S hao,andA.Srinivasan, ‘‘Mobile data offl oading through opportunisticcom munications and social participation,’’ IEEE Trans. Mobile Comput., vol. 11, no. 5, pp. 821–834, May 2012.

[7] P.Gill,M.Arlitt,Z.Li, and A.Mahanti,‘‘YouTub etraffiCcharacterization: Aviewfromtheedge,’’inProc.7thACMSIGCOM MConf.InternetMeas., 2007, pp. 15–28.

[8] M. Zink, K. Suh, Y. Gu, and J. Kurose, ‘‘Characteristics of YouTube network traffic at a campus network—Measurements, models, and implications,’’ Comput. Netw., vol. 53, no. 4, pp. 501–514, 2009.

[9] X. Cheng, J. Liu, and C. Dale, ‘‘Understanding the characteristics of Internet short video sharing: AYouTube-based measurement study, ’’IEEE Trans. Multimedia, vol. 15, no. 5, pp. 1184–1194, Aug. 2013.

[10] R.CraneandD.Sornette,‘‘Robustdynamiccla ssesrevealedbymeasuring the response function of a social system,’’ Proc. Nat. Acad. Sci. USA, vol. 105, no. 41, pp. 15649–15653, 2008.

[11] J.YangandJ.Leskovec, ‘‘Pattern softemporalv ariationinonlinemedia,’’ in Proc. 4th ACM Int. Conf. Web Search Data Mining, 2011, pp. 177–186.

[12] F. Figueiredo, F. Benevenuto, and J. M. Almeida, ‘‘The tube over time: Characterizing popularity growth of YouTube videos,’’ in Proc. 4th ACM Int. Conf. Web Search Data Mining, 2011, pp. 745–754.

[13] F. Figueiredo, ‘‘On the prediction of popularity of trends and hits for user generated videos,’’ in Proc. 6th ACM Int. Conf. Web Search Data Mining, 2013, pp. 741–746.

[14] F. Figueiredo, J. M. Almeida, M. A. Gonçalves, and F. Benevenuto, ‘‘On the dynamics of social media popularity: A YouTube case study,’’ ACM Trans. Internet Technol., vol. 14, no. 4, 2014, Art. no. 24.

[15] M. Ahmed, S. Spagna, F. Huici, and S. Niccolini, ‘‘A peek into the future: Predicting

the evolution of popularity in user generated content,’’ in Proc. 6th ACM Int. Conf. Web Search Data Mining, 2013, pp. 607–616.

AUTHORS:

PARRIPATI V R ANUSHA SRIPRIYA
B.Tech Student, Department of CSE,
Sri Mittapalli College of Engineering, Tummalapalem, NH-16, Guntur,
Andhra Pradesh, India.

PAVULURI SAI PRATHYUSHA
B. Tech Student, Department of CSE, Sri Mittapalli College of Engineering, Tummalapalem, NH-16, Guntur, Andhra Pradesh, India. **YANAMADURTI V SAI BHAVANI**

B. Tech Student, Department of CSE, Sri Mittapalli College of Engineering, Tummalapalem, NH-16, Guntur, Andhra Pradesh, India.

PACHA PAVANI
B. Tech Student, Department of CSE,
Sri Mittapalli College of Engineering, Tummalapalem, NH-16, Guntur,
Andhra Pradesh, India.

MOHAMMAD ABDUL NADEEM
B. Tech Student, Department of CSE,
Sri Mittapalli College of Engineering, Tummalapalem, NH-16, Guntur,
Andhra Pradesh, India.