IJCSN - International Journal of Computer Science and Network, Volume 8, Issue 3, June 2019
ISSN (Online) : 2277-5420
www.IJCSN.org
Impact Factor: 1.5

Predicting the Top-N Popularity of YouTube and Twitter Video's Using Early View Data

[1] Christy Kunjumon; [2] Dr Radhakrishnan B

[1] CSE Department, BMCE
APJ Abdul Kalam Technological University, Sasthamcotta, Kerala, India

[2] CSE Department, BMCE
APJ Abdul Kalam Technological University, Sasthamcotta, Kerala, India

Abstract - In this paper, the proposed system jointly combines the properties of YouTube and Twitter. The linear regression method is used to predict the Top-N popular video's. This method is designed to improve the analysis of popular video's and their popularity trends. It is for understanding the video popularity characteristics and predicting the future video popularity. They have direct implications in various contexts such as service design, advertisement planning, network management and so on. The data are collected from YouTube and Twitter API's.

Keywords*- popularity prediction, YouTube, social media, Multilayer Perceptron, linear regression.*

1. Introduction

YouTube is the largest user-driven video content provider in the world. YouTube social network is widely used for uploading, viewing and sharing videos. We use two content sharing portals called YouTube and twitter. Twitter is a social textual stream-based platform and the information exposure is very much faster in Twitter. YouTube is an interactive media sharing platform. The most popular social media online platforms such as YouTube, Facebook, Twitter or Instagram, where users can easily share information's with other peoples and the peoples can interact very easily. In internet, a vast amount of video content has been generated but, while a few of videos become very popular. Hence, understanding the popularity characteristics of online videos and predicting the future popularity of individual videos plays a great importance. Predicting the future video popularity is useful in planning advertisements, service design etc. It conducts a large-scale analysis of statistical data of early viewers' attitude-related behavior and the long-term popularity of videos. The Viewers' attitudes can be reflected through related text, such as comments, and related behavior such as clicking "like" or "dislike" after he/she is watching. Here evaluate the proposed model using real-world data consisting of videos from YouTube and social network data from Twitter.

The main contributions of this paper can be summarized as follows:

- Here proposed a model for predicting the top- N popular video. In particular, the aim is to predict how the individual video features (through the meta-level features) and the social dynamics features contribute to the popularity of a video.
- The method used to predict the popularity is multilayer neural network and linear regression.

2. Related Works

William Hoiles et al [1] proposed that different machine learning methods to describe the popularity evolution of YouTube videos. we look into the sensitivity of the video meta-level features on the view counts of videos. The most important meta-level features is found that, it include: first day view count, the number of subscribers, disparity of the video thumbnail, Google hits, number of keywords, category of the video and title length. It can be illustrate that these meta-level features can be used to determine the popularity of a video. In particular, by optimizing the meta-level features after a video is posted it may boost up the video popularity. In social dynamics, to explore that there is a relationship between number of views to a channel and the number of subscribers. The popularity of the channel also depends on the social dynamics (channel interaction).

Peter Schultes et al [2] identified that User comments are the most popular but also extremely contentious form of communication on YouTube. The users may generally expected that most of the comments will be of little value or even in comprehensively bad taste. Even though, heaps of comments continue to be posted every day. In addition, user attitudes and behavior based on a new comment

classification approach which captures salient aspects of YouTube comments. Based on the new classification, it has been able to perform very fast lightweight semantic video analysis. In addition, results indicate that users' video perceptions (Likes and Dislikes) are indeed influenced by the dispersion of valuable and inferior comments.

Chenyu li et al [3] proposed a multivariate linear regression model to predict the video popularity based on early stage popularity evolution patterns and future popularity burst prediction. The data collected from a leading online video platform in China, namely Youku. Firstly, analyze the characteristics of Youku video popularity from four key aspects: long-term popularity of the video, video lifetime, popularity evolution pattern, and early stage popularity of the video. Then the challenge of future popularity prediction, by proposing a model that can capture the popularity dynamics based on early popularity evolution pattern and future popularity burst prediction.

Zhiyi Tan et al [4] proposed 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. By enhancing the discriminative power of early patterns for different popularity trends and by optimizing the utilization of multi-source data. The challenge for the task are as follows. First, popular and unpopular videos may have similar early view patterns. Second, prediction models that are overly dependent on early view patterns limit the effects of other features. The proposed model using real-world YouTube data.

Przemysław Rokita et al [5] proposed that predicting the popularity of an online video by using the regression method can be measured by its number of views. The method uses Support Vector Regression with Gaussian Radial Basis Functions. predicting popularity patterns with this method gives more stable and more precise prediction result. The accuracy of popularity prediction can be improved by combining early distribution patterns with social and visual features. The social features represent a much stronger signal in terms of video popularity prediction than the visual ones.

3. Design & Architecture

The proposed system jointly combines the properties of YouTube and twitter. For a given YouTube videos for a specific period of time. For each entry in the dataset it contains information about the likes, the dislikes, view count and the count of social media sharing and it can be achieved so far.

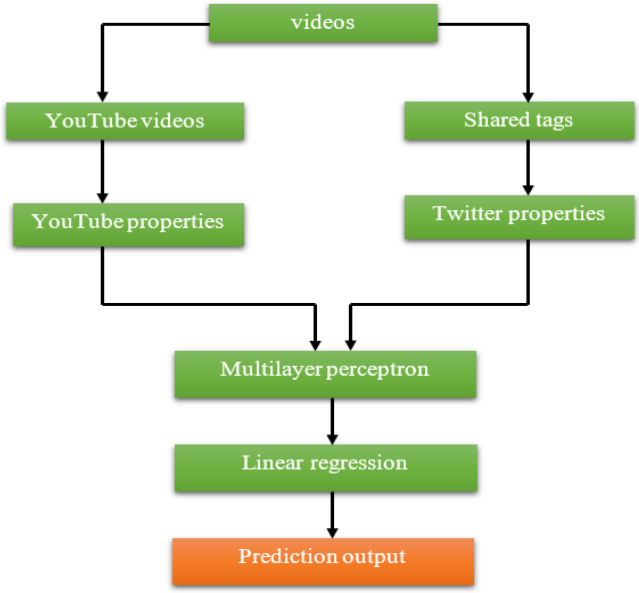

Fig.1.proposed System

Additionally, each entry also maintains information about its trending nature in twitter with its retweet and like count. That extracted data from the given set is first standardize and then the input data will be feed into a MLP (multilayer perceptron) with hidden layer and output layer. The MLP (multilayer perceptron) will be trained to achieve a saturation with more than 90% of accuracy and the trained pattern obtained from the MLP (multilayer perceptron) will be feed into a linear regression. The regressor evaluate the feedin'd data and optimizes the conclusions evolved from the MLP. finally, the data's will be tested against optimized regressor result.

3.1 Multi-layer perceptron

A multilayer perceptron (MLP) is a class of feedforward artificial neural network. Each neuron in the network includes a nonlinear activation function that is differentiable. MLP consists of at least three layers of nodes: an input layer, a hidden layer and an output layer.

The network exhibits a high degree of connectivity.

MLP training includes two phases: one is forward phase and another is backward phase.

In the forward phase, network of the weights are fixed and the input signal is propagated through the network, layer by layer, until it reaches the output.

In the backward phase, the error signal, which is produced by comparing the desired response and the output of the network, and the network is propagated through layer by layer, but in backward direction.

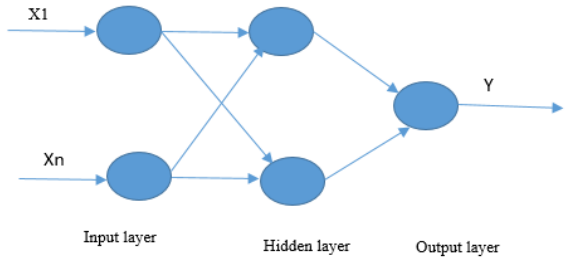

Fig.2.multilayer perceptron

3.2 Linear Regression

Linear regression is a linear approach to modeling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, the process is called multiple linear regression.

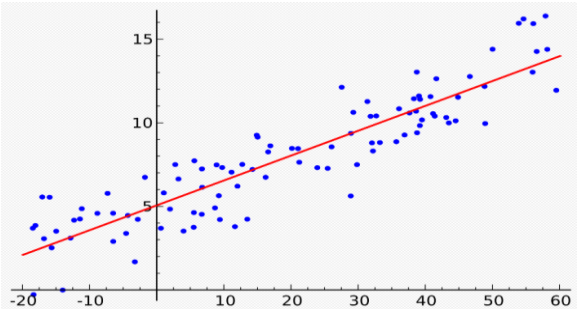

Fig.3. Regression Analysis

4. Experimental Result and Discussion

The experimental results of the popular videos prediction are obtained and discussed. The table below illustrates the comparison of the outcomes obtained by using previous methods and those by the present method. YouTube and twitter data is used for training. For training here use multilayer perceptron. In multilayer perceptron, Linear regression method is used for training the network. Recent results of YouTube and twitter dataset, by fixing the test data at 20%, the remaining 80% training data. The result shows that the proposed model is better as compared with the previous models and also it outperforms with an accuracy of 97.25%.

Table.1. Prediction accuracy of different methods

Model	*Accuracy*
MLP	97.25%
SVM	86.32%
ARMA	82.41%

5. Conclusion

In this paper, we have shown for a given YouTube videos for a specific period of time. The goal of the project is to use machine learning techniques to predict the popularity of YouTube videos and social feature network. Here the YouTube properties and twitter properties are considered as the input. Based on the properties, to predict the popularity of individual video. This study proposes a new prediction model based on multilayer perceptron and Linear Regression The result shows that the proposed model is better as compared with the previous models. Also, this model gives a consistent and accurate result.

References

[1] William Hoiles, Student Member, IEEE, Anup Aprem, and Vikram Krishnamurthy, Fellow, IEEE. Engagement dynamics and sensitivity analysis of YouTube videos.

[2] Peter Schultes, Verena Dorne and Franz Lehner. Leave a Comment! An In-Depth Analysis of User Comments on YouTube, University of Passau, Passau, Germany schult16@stud.uni-passau.de, {Verena. dorner, Franz. Lehner}@uni-passau.de

[3] Chenyu Li (student member, IEEE), Jun Liu (member, IEEE) and Shuxin Ouyang. Characterizing and Predicting the Popularity of Online Videos, Beijing Key Laboratory of Network System Architecture and Convergence, Center for Data Science, School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China

[4] Zhiyi Tan, Ya Zhang IEEE Predicting the Top-N Popular Videos via a Cross-Domain Hybrid Model.

[5] Tomasz Trzci´Nski and Przemysław Rokita, IEEE Transactions On Multimedia. predicting popularity of online videos using Support Vector Regression.

Authors profile

Christy kunjumon received her B.Tech (CSE) degree from University of Kerala in 2017. She is currently pursuing her Masters in Computer Science & Engineering from KTU.

Dr. Radhakrishnan B is working as the Head of CSE department. He has more than 14 years' experience in teaching and has published papers on data mining and image processing. His research interest include Image processing, Data mining, Image Mining .