


Popularity Prediction of Video Content Over Cloud-Based CDN Using End User Interest

Rohit Kumar Gupta, Manipal University Jaipur, India

Shabbir Kurabadwala, Manipal University Jaipur, India

Pradeep Kumar Tiwari, Manipal University Jaipur, India*

 <https://orcid.org/0000-0003-0387-9236>

Ankit Mundra, Manipal University Jaipur, India

ABSTRACT

It is often believed that more is better, but that is not true in the case of data. As online data is increasing briskly, we are not able to handle such enormous data. With the increasing trends of speedy and uninterrupted access to data usage, CDNs have become quite popular in today's world. However, it has become difficult to store all the content on CDN servers. This paper aims towards optimizing one of the aspects of CDN's cached data that is video content. The authors propose a push-based caching approach by finding appropriate popular videos in accordance with a region to improve an end user's quality of experience. A semi-supervised machine learning approach has been implemented to classify videos as low, medium, or highly popular. Popularity prediction research has increased in energy lately. In any case, there has been little work done dependent on prior and significant video parameters for popularity prediction purposes. The experimental results show good accuracy, justifying the selection of parameters and the processing associated with them.

KEYWORDS

Cloud-Based Content Delivery Network, End User Interest, Popularity Prediction, Video Content

INTRODUCTION

A Content Delivery Network (CDN) is a network of server nodes distributed geographically around the world that are used for downloading resources (typically static content such as images and JavaScript), allowing for faster delivery of resources with lower latency. Over the years, CDNs have gained immense popularity. With the development of Cloud Computing technology and other utilities, it has become quite easy and efficient to rent bandwidth and storage resources and manage CDNs over the cloud. The petabytes of client produced content traffic moving through CDNs show that the substance index is enormous, and that there is a large amount of online content, and it is of extreme importance to spot popular data and cache it on these servers.

The basic building blocks of a CDN are - CDN PoPs (Points of Presence), Caching servers and SSD/HDD + RAM.

- CDN PoPs are smartly placed data centres which are responsible for transmitting data within their geographic region. Their principal aim is to scale back the total round-trip time by closing into the location of the user. Each PoP is normally equipped with various caching servers.

- The task of delivering cached files is done by the CDN Caching servers. They accelerate the website's load time and reduce the overall bandwidth utilization. Inside a caching server are several RAM assets and storage drives.
- Solid State Drives (SDD) and Hard Disk Drives (HDD) or RAM are used to store the cached files. More frequently requested files are hosted on faster memory. RAM is the fastest memory after SDDs and HDDs.

Predicting video popularity is of great importance as it would result in increased user experience and faster response time. The main focus of this paper is to use machine learning algorithms to predict the video popularity in advance, to utilize the resources of CDN caching servers efficiently.

The process of forecasting the happening of a scenario on a specific dataset, after some algorithm has been trained on it is known as Prediction. The machine learning algorithm generates the most anticipated values for some unknown variable for each tuple in the dataset, helping the model to identify what could be the most likely value of the unknown variable for that example. In this paper, video popularity has been predicted based on certain features which are available at the time of video upload and continuous evaluation is done to determine the dynamic change in its popularity. As mentioned earlier, this project is accomplished by a *semi-supervised* approach. The k-means clustering algorithm was used to assign labels to the dataset (which were not present in the occupied dataset), and three algorithms, namely Nave Bayes, K-Nearest Neighbour, and Support vector machine, were used to predict popularity after further pre-processing and vectorization of some features. The detailed stepwise process is further explained in the paper.

The following sections of this paper are described in detail: Section II focuses on related work, Section III explains the methodology and implementation of the proposed approach, Section IV analyses the results, and Section V concludes the work (Al-Abbasi et al., 2019; Wang et al., 2018)

RELATED WORK

Popularity Prediction

Due to an exponential increase of internet users and the enormous growth of UGC (user generated content), video popularity has gotten a great deal of consideration from research networks. Gloria et al. (2010) conducted a detailed study about the basic properties of videos on YouTube. The best interconnected features for predicting the video popularity were found to be the view count, comment count, favourite count, and rating. A Linear regression model was proposed to predict the view counts and, in conclusion, they characterized a small-world network by using the related videos. Zeng et al. (2013) used views, tags, category and title for predicting a traditional popularity score and fusing this with a social popularity metric extracted from the video content topic and developing a trend aware popularity which is then used to predict the final popularity. Chen et al. (2019) proposed an early popularity prediction model as most of the research work had been done on continuous prediction. Their approach is determined by three perspectives – related video, author, and keyword perspective. They used a revised TF-IDF model which was used to handle textual features in which they replaced document frequency with webpage frequency. They used 5 base classifiers and then constructed an ensemble model combining the features of all of them and compared the accuracy.

Nesrine et al. (2018) proposed a two-level prediction approach where prediction is performed by experts using their own method, and at the second level it is given to forecasters. They use various regression and moving average models. Tomasz et al. (2017) worked on a regression method using a support vector machine. They extracted visual features such as colour, face, scene dynamics, etc and converted them to a 1000-dimensional vector and combined it with the view count and social factors and then passed it to a support vector regression model. Nuno Moniz and Luís Torgo (2019) conducted a survey discussing the issues and open challenges regarding the prediction of popularity

11 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/popularity-prediction-of-video-content-over-cloud-based-cdn-using-end-user-interest/301227

Related Content

Topological Functioning Model as a CIM-Business Model

Erika Asninaand Janis Osis (2011). *Model-Driven Domain Analysis and Software Development: Architectures and Functions* (pp. 40-64).

www.irma-international.org/chapter/topological-functioning-model-cim-business/49153

Improved Fall Detection Model on GRU Using PoseNet

Hee-Yong Kang, Yoon-Kyu Kangand Jongbae Kim (2022). *International Journal of Software Innovation* (pp. 1-11).

www.irma-international.org/article/improved-fall-detection-model-on-gru-using-posenet/289600

Decreasing Service Coupling to Increase Enterprise Agility

Jose Carlos Martins Delgado (2015). *Achieving Enterprise Agility through Innovative Software Development* (pp. 225-261).

www.irma-international.org/chapter/decreasing-service-coupling-to-increase-enterprise-agility/135230

Fault Injection for On-Board ERTMS/ETCS Safety Assessment

Almir Villaro Arriola, Jon Mendizabal Samperand Juan Meléndez Lagunilla (2012). *Railway Safety, Reliability, and Security: Technologies and Systems Engineering* (pp. 128-150).

www.irma-international.org/chapter/fault-injection-board-ertms-etcs/66670

Development of APP-Controlled Hydraulic Damping Adjustment System for Bike Front Fork

Hsiung-Cheng Lin, Ming-Dian Wang, Jun-Wei Chenand Hong-Ming Chen (2022). *International Journal of Software Innovation* (pp. 1-16).

www.irma-international.org/article/development-of-app-controlled-hydraulic-damping-adjustment-system-for-bike-front-fork/309727