JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A COMPREHENSIVE ANALYSIS ON CHARACTERIZING YOUTUBE VIDEOS: A Survey

[1]Mr.A.R.KARTHEKEYAN, Ph.D Research Scholar, Department of Computer Science, Kongu Arts and Science College (Autonomous), Erode – 638107

[2]Dr.K.G.SANTHIYA, Associate Professor, Department of Computer Science (UG), Kongu Arts and Science College (Autonomous), Erode – 638107

Abstract

All viewers of YouTube video are not seeing a video simultaneously. As a result, viewers have been categorized into viewer groups based on the behavioral measure of the viewers who moderately watches the movies sooner than others. Viewer classification is critical and must be established, since viewers may assist in identifying prospects for an online video and forecast the film's continuing sharing. The YouTube video-sharing community is a new and successful phenomena that offers an expressive depiction of a social network. Despite its rapid development, a comprehensive analysis of YouTube's topology has yet to be published. This article presents a review of results on the features of YouTube and similar sites from both video and user viewpoints, as well as the outstanding research topics. This kind of research is essential for comprehending the use of YouTube and other comparable user-generated content websites.

Keywords: YOUTUBE, Searching, Videos, YouTube Users, survey

I. INTRODUCTION

YouTube has over 1 milliard active members and hundreds of millions watch videos which create thousands of views daily on YouTube. It offers a wide range of contents over a billion videos. YouTube is not only an entertainment centre; it also offers a plethora of information from many professionals who freely share their expertise. YouTube gives students the chance to acquire new skills from professionals and experts worldwide [1]. The video medium enables pupils to comprehend and maintain their understanding [6]. It offers the freedom to study at one's own speed. But the interesting material on YouTube may easily encourage pupils to divert themselves. A highly focused and effective platform should thus be developed, where visitors may explore instructional material without being distracted [2]. In the classification of videos most methods utilize either video or video processing but not both. The premise of the study is to take video and metadata into consideration to categorize videos and create a rapid and accurate classification. The visual processing model uses the video and the keywords are processed by the text processing pattern, one of the metadata of the movies. Both models utilize the cumulative result in the classification of the videos [8]. This classifies the videos as educational and non-educational. Popular websites like YouTube or MetaCafe request the uploaded to provide written contents to accompany every one of the millions of video uploaded each day. Although up loaders may offer text information in several formats, including a title, a description and a series of free-form tags, metadata deserts remain — films that are too short on their titles or which are either short or absent in their descriptions and tags. Furthermore, the real semantic connections of the uploaded video are often only partly reflected by the submitted text [9]. While YouTube mostly includes video or audio material produced by users, this kind of service has been transcended. Videos (movies) are licensed and professionally made and many social networking functions are now accessible. These include a sophisticated reference and video system linked to other social media sites, channels, ratings and relationships. Nevertheless, the studies examined may assist in the development of comparable new systems, capacity planning, network management and advertising strategies [12].

The rest of this paper is organized as follows. Section II analyzes the research that differentiates behavior from the individual user perspective. Section III the work done on analyzing YouTube videos in terms of video characteristics. Open research issues in this area are addressed in Section IV. Finally, Section V contains our conclusions.

II. BACKGROUND STUDY

Cao, L. et al. [3] the authors examine the problem of the grouping of videos in "themes" on line in an on-line survey scenario to enable effective viewing of YouTube videos where either video material or tags are completely confused for direct use.

Chtouki, Y. et al. [4] a good learning experience requires students to have an interest in the topic and to intuitively convey the information. Because kids' interest in reading decreases, instructors must employ other techniques to guarantee that they still learn. In this article, the writers showed the findings of a research carried out for non-major students on an introduction to computer courses. The goal of every school was to improve the productive and success percentage of pupils. This research focuses on one of many current online solutions that may improve student learning experience.

Figueiredo, F. et al. [7] the author’s distinguished video popularity trends in the presently best-known video sharing programmed, YouTube. The authors examined the popularity of specific videos since the video was published and the various kinds of references that most frequently bring the viewers to the videos, using newly supplied data from the application.

Juasiripukdee, P. et al. [10] the authors propose that a clustering technique enables users to discover what they are interested in on UGC websites simply and rapidly. The system analyses the search results' similarity and then automatically aggregates related material. A very lengthy list of results must not be drawn up by the user. He/she just chooses desired content groups from the cluster list. The result was that the user may quickly discover interesting information and minimize their search time.

Richier, C. et al. [11] the authors have developed an automated method of categorizing the dynamic development of the video count into six biologically inspired mathematical models. The authors demonstrated that majority of the films in our data sets are linked to a model with a mean error rate below 0.05, indicating the perfect matching mathematical model with data sets observed. The authors presented various results using our categorization. The writers first were able to identify the main characteristics of videos as virality and population expansion potential. Secondly, two of the six models frequently came to the fore, and both models anticipate most of them to undergo an immigration process in which the potential population would increase over the course of time. Thirdly, the authors have created a stringent model that allows us to anticipate an overview over a window period.

Saxena, V. et al. [13] the availability of video materials did not improve the anatomy and radiology performance of students. However, among individuals who view the videos at least once compared to the control group, a small 3.4% impact on the performance of anatomic has been seen in the analysis by usage level. This indicates that video may successfully help at least some students attain their performance especially if large-scale instruction is required.

Toderici, G. et al. [14] the authors proposed a method for YouTube uploads automated tag recommendation. The authors also provide a framework for integrating category web information with video-learned visual tag information. By utilizing video tags, the suggested approach detects significant video categories. In a human experiment the tags were assessed and the significance of their initial super-concept was approximately the same as for the user tags provided.

III. CHARACTERIZING YOUTUBE VIDEOS

A. YouTube Video Characteristics

[15] Conducts a thorough examination of the features of YouTube videos. Following the associated video connections of certain famous YouTube videos, information from about 2.6 million videos was gathered. At the time, the projected number of posted videos on YouTube was approximately 42.5 million. The following factors were taken into account: video length, genre, active life duration, and connection to other videos. The uploading rate of YouTube videos may be fitted with a power law curve, and 3 Music and Entertainment videos were discovered to be the most often posted out of 15 categories. In terms of video length, almost 98 percent of the videos were determined to be shorter than 600 seconds in length. This may be because YouTube set a video length restriction in 2006, which is why films for television programmes and movies posted at the time were discovered in parts. There is no

connection between video duration and video popularity.

B. Popularity Growth Pattern of YouTube Videos

[17] Used the Most Recent standard feed supplied by the YouTube API to gather information on 29,791 YouTube videos in order to monitor the time-varying popularity of YouTube videos (essential for effective object caching). Their data collecting method was sufficient to provide an impartial dataset; the Most Recent standard feed gives information on a random selection of movies that were posted lately, independent of the amount of views. According to their findings, the majority of the videos reach their peak popularity in less than six weeks after being uploaded. Furthermore, in order to determine whether or not a video's present popularity is an indication of future popularity, Pearson's correlation coefficient was determined between additional views at weekly snapshots.

C. Content Aliasing in YouTube

[18] Examined YouTube content duplication and overlap. Content-based copy detection techniques (CBCR) were utilized to identify duplicate scenes across various movies. Sets of graphs were created in such a way that the graph's edges reflect highly linked films on YouTube. The fingerprint-based CBCR's components are divided into three steps: the fingerprint creation module, the reference content database, and the search module. All movies are converted into a series of points in the fingerprint feature space in the fingerprint creation module. The reference content database is a collection of known fingerprints that may be created via supervised training sessions. Finally, the fingerprints for all incoming video streams are checked with the reference content database in the search module phase. A pilot experiment was performed for a known database to assess the efficacy of the CBCR method, which confirmed 90 percent accuracy of CBCR.

D. Playback Quality Concerns/Potential Solutions

[16] Depicts the dissatisfying experience of YouTube viewers when viewing videos, as well as a potential remedy. Initially, an experiment was carried out to assess user experience when viewing YouTube videos—the frequency and length of pauses during video playing. The pause frequency data was obtained automatically by analyzing video download traces. 12 volunteers from 12 distinct settings (with various network access methods) were requested to record YouTube traffic using the Wireshark network protocol analyzer5. To estimate the amount of pauses in playback, a model was created. The sample dataset revealed that 10 of the 12 settings had playbacks with stops, and 41 of the 117 playbacks had pauses, accounting for roughly 35% of the total playbacks. This finding shows that YouTube viewers are subjected to loud playbacks, which may be more pronounced for higher-quality films. As high quality films grow more popular on YouTube, this issue may become unbearable.

F. Regional Popularity of YouTube

[19] Examined the connection between location and YouTube video popularity. The amount of daily views for over 20 million videos was compiled. This study examined 250 distinct areas for the studies, including formal states and smaller territories. Approximately 40% of YouTube videos get at least 80% of their views from a single location. This data suggests that YouTube videos

gain popularity in a localized area rather than a worldwide zone. Not unexpectedly, various categories showed distinct patterns of global and local popularity. As a result, it seems that several variables contribute to video attracting viewers from all over the globe, which is unusual. Similarly, there is a significant connection between the location of a video's uploader and its geographical popularity. For example, due to comparable interests, movies posted from the United States are popular in the United Kingdom, Mexico, and Canada. Videos posted in Japan and Brazil, on the other hand, get 90 percent of their views in their posting area alone.

IV. COMPARATIVE ANALYSIS OF SURVEY

Author Name	Methodology	Limitations
Chowdhury, S. A., & Makaroff, D. [2]	improved P2P caching method	limits the details of the viewing pattern
Chtouki, Y. et al. [4]	Students Education	the limitation of how much was covered in introductory courses, and finally the time limitation doesn’t allow the instructor to elaborate or go into much specific detail on a given topic or concept.
Camm, C. F. et al. [5]	Videos searched on Health Oriented	Its effectiveness was limited.
Figueiredo, F. et al. [7]	patterns of video popularity	Each popularity growth curve was registered with at most 100 points, regardless of the age of the video.
Juasiripukdee, P. et al. [10]	metadata refining, related content grouping, and group clustering	spending time to sifting through a long list of search results until they can find all the content for which they were actually looking.

Table 1: Evaluation on various authors views.

IV DISCUSSION

Because of the flexibility in video posting, scalability is regarded as one of the most significant problems in YouTube. Despite its potential answer to the scaling issue, peer-to-peer methods cannot be implemented without proper changes, particularly to their incentive mechanisms. Imposing download speed limits, for example, as used by incentive systems, may have a detrimental impact on YouTube's and other sites' present popularity. This aspect of video dissemination needs further investigation. Other well-known video distribution methods, such as batching and patching, may be explored alongside P2P strategies. For example, it would be interesting to examine the performance of batching for YouTube live streaming; batching has been shown to be a viable option for improving playback quality for these types of video distributions. Multilayer caching rules for caching mechanisms may reduce buffering time. Different methods to local caching may be created based on the geographical popularity of videos. In the event of a local cache miss, the request will be sent to the central server, where movies will be cached in a manner that reflects the worldwide popularity of videos. Another source of worry for online marketers is content aliasing on YouTube, which affects the popularity of original videos. Considering the kinds of video objects while evaluating the time-varying popularity may help to better understand the video growth pattern. Some areas, such as news and sports, may have the majority of their opinions at a young age. This kind of study may not only enhance the caching process, but it may also be useful in developing suitable advertising strategies.

V. CONCLUSION

In this article, we review the features of YouTube that have previously been investigated in the literature. Other video distribution sites were compared in the literature. One of our objectives was to identify elements of the site's use for video data streaming and upload, as well as metadata operations like ratings and comments. The majority of academic research relied on YouTube behavior to generate workload for prototype systems or simulations of current or new content distribution/media caching methods. The technique used to measure request traffic and/or crawl the YouTube server site using the available API affected the results that might be reached. We verified that certain problems, such as the presence of a heavy-tailed distribution of request frequency, produced consistent outcomes. On the contrary, several of the findings contradict each other, suggesting that there are problems that need to be investigated further. A framework for uniform data collecting technique is required in order to effectively compare solutions for improved user experience and server efficiency.

VI. REFERENCES

[1] Aggrawal, N., Arora, A., & Anand, A. (2018). Modeling and characterizing viewers of You Tube videos. International Journal of System Assurance Engineering and Management, 9(2), 539–546. doi:10.1007/s13198-018-0700-6

[2] Chowdhury, S. A., & Makaroff, D. (2012). Characterizing Videos and Users in YouTube: A Survey. 2012 Seventh International Conference on Broadband, Wireless Computing, Communication and Applications. doi:10.1109/bwcca.2012.47

[3] Cao, L., Ji, R., Liu, W., Gao, Y., Duan, L.-Y., & Men, C. (2012). Weakly supervised topic grouping of YouTube search results. 2012 19th

IEEE International Conference on Image Processing. doi:10.1109/icip.2012.6467502

[4] Chtouki, Y., Harroud, H., Khalidi, M., & Bennani, S. (2012). The impact of YouTube videos on the student's learning. 2012 International Conference on Information Technology Based Higher Education and Training (ITHET). doi:10.1109/ithet.2012.6246045

[5] Camm, C. F., Sunderland, N., & Camm, A. J. (2012). A Quality Assessment of Cardiac Auscultation Material on YouTube. Clinical Cardiology, 36(2), 77–81. doi:10.1002/clc.22080

[6] Esen, E., Ozkan, S., & Atil, I. (2016). Fast video search on recurring segments. 2016 24th Signal Processing and Communication Application Conference (SIU). doi:10.1109/siu.2016.7496057

[7] Figueiredo, F., Benevenuto, F., & Almeida, J. M. (2011). The tube over time. Proceedings of the Fourth ACM International Conference on Web Search and Data Mining - WSDM '11. doi:10.1145/1935826.1935925

[8] France, S. L., Vaghefi, M. S., & Zhao, H. (2016). Characterizing viral videos: Methodology and applications. Electronic Commerce Research and Applications, 19, 19–32. doi:10.1016/j.elerap.2016.07.002

[9] Jaffar, A. A. (2012). YouTube: An emerging tool in anatomy education. Anatomical Sciences Education, 5(3), 158–164. doi:10.1002/ase.1268

[10] Juasiripukdee, P., Wiyartanti, L., & Kim, L. (2010). Clustering search results of non-text user generated content. 2010 Fifth International Conference on Digital Information Management (ICDIM). doi:10.1109/icdim.2010.5663293

[11] Richier, C., Altman, E., Elazouzi, R., Jimenez, T., Linares, G., & Portilla, Y. (2014). Bio-inspired models for characterizing YouTube viewcout. 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014). doi:10.1109/asonam.2014.6921600

[12] Ramesh, I., Sivakumar, I., Ramesh, K., Venkatesh, V. P. P., & Vetriselvi, V. (2020). Categorization of YouTube Videos by Video Sampling and Keyword Processing. 2020 International Conference on Communication and Signal Processing (ICCSP). doi:10.1109/iccsp48568.2020.9182158

[13] Saxena, V., Natarajan, P., O'Sullivan, P. S., & Jain, S. (2008). Effect of the use of instructional anatomy videos on student performance. Anatomical Sciences Education, 1(4), 159–165. doi:10.1002/ase.38

[14] Toderici, G., Aradhye, H., Pasca, M., Sbaiz, L., & Yagnik, J. (2010). Finding meaning on YouTube: Tag recommendation and category discovery. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. doi:10.1109/cvpr.2010.5539985

[15] X. Cheng, C. Dale, and J. Liu, "Understanding the characteristics of internet short video sharing: YouTube as a case study," Cornell University, arXiv e-prints, Tech. Rep., Jul. 2007.

[16] S. Khemmarat, R. Zhou, L. Gao, and M. Zink, "Watching user generated videos with prefetching," in *MMSYS 2011*, San Jose, CA, Feb. 2011, pp. 187–198.

[17] Y. Borghol, S. Mitra, S. Ardon, N. Carlsson, D. Eager, and A. Mahanti, "Characterizing and

modelling popularity of user-generated videos," *Performance Evaluation*, vol. 68, pp. 1037–1055, Nov. 2011.

[18] J. Pedro, S. Siersdorfer, and M. Sanderson, "Content redundancy in YouTube and its application to video tagging," *ACM Trans. on Info. Sys.*, vol. 29, no. 3, pp. 13:1–13:31, Jul. 2011.

[19] A. Brodersen, S. Scellato, and M. Wattenhofer, "YouTube around the world: Geographic popularity of videos," in *WWW 2012*, Lyon, France, Apr. 2012, pp. 241–250.