iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** V **Month of publication:** May 2022

DOI: https://doi.org/10.22214/ijraset.2022.43051

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue V May 2022- Available at www.ijraset.com

A Micro Video Recommendation System Based on Big Data

Maddukuri Manideep[1], Mrs. Subhita[2]
[1, 2]*Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab*

Abstract: ***With the development of the Internet and social networking service, the micro-video is becoming more popular, especially for youngers. However, for many users, they spend a lot of time to get their favorite micro-videos from amounts videos on the Internet; for the micro-video producers, they do not know what kinds of viewers like their products. Therefore, we proposes a micro-video recommendation system. The recommendation algorithms are the core of this system. Traditional recommendation algorithms include recommendation algorithms, and so on. At the Big Data times, the challenges what we meet are data scale, performance of computing, and other aspects. Thus, we improves the traditional recommendation algorithms, using the popular parallel computing framework to process the Big Data. Slope one recommendation algorithm is a parallel computing algorithm based on MapReduce and Hadoop framework which is a high performance parallel computing platform. The other aspect of this system is data visualization. Only an intuitive, accurate visualization interface, the viewers and producers can find what they need through the micro-video recommendation system.***
Keywords: micro-video; recommendation system; Slope one; data visualization

I. INTRODUCTION

Video is a new form of information media. With the development of the Internet, 3G (the 3rd Generation mobile communication technology), and 4G (the 4th Generation mobile communication technology) network, the bandwidth and speed of network become faster and faster. These technologies provide conditions for dissemination of information media. Video is a short time video, which lasts for 30 seconds to 300 seconds. The short time micro-videos are popular with young people, because the teenagers prefer to watch the micro-video on their comfortable time through mobile devices. For micro-video producers, the problem is they do not know how many people like their products, and do not know how many times their video have been watched. Therefore, this proposes a video recommendation system (MRS). One of the purposes is an overview of videos for the producer. In this way, the producer knows how many users love their video, and how many times their videos are on- demand. Another purpose is for users. The system can analyze the users' favorites and watching history, automatically push appropriate video to the users. It is becoming more popular with the internet technology development, which means the data sets whose size is beyond the ability
of current technology, method and theory to capture, manage, and process the data within a tolerable elapsed time. In order to enhance the MRS accuracy, we need to collect large volume data sets about who and when watched the micro-video, how many times the micro- video on demanded, and how many people love the micro-video. Therefore, the MRS, proposing in this paper, use technology to process the collected data sets. Data sets are the foundation of the recommendation system. The first step of video 1recommendation is to collect data as far as possible from the Internet. We download data from video websites, video forum, video online chat websites, and so on. Web crawlers, one of basic data sets collection, can download resources from Internet. The web crawlers originally used for search engine. In this paper, the results of crawler directly affect the accuracy of recommendation system.Figure 1 Dig Data There are many successful practices of the application of big data technology in the video industry. In the early stages of the production of the video, there are a lot of data mining and data analysis. First, obtain massive video data from the internet. Then, through the analysis of these data, obtain many video information that the user groups are interested in, like themes, actors, songs. In the end, it is possible for video producers to create videos that can get a large amount of video viewing times according to the information obtained from the analysis.

This system gets a large amount of video data from large social platforms, video sites, search engines and other ways. Then, combines massive video information and business requirements to provide more valuable creative information for companies' video creation, so as to make the video more in line with the needs of users, and enhance the number of video viewing. 2Transmission capacity can reach 500-600 set of programs. It meets the needs of the various users, but users are difficult to find their favorite programs in so many programs. Also, users want to see more of their favorite programs, but the current source is single, so we need to consider how to get more program. Some ways of presenting recommendations affect the ex- planation more than others. In fact, some ways of offering recommendations, such as the organizational structure we will describe shortly (See Section 4.5), can be seen as an explanation in itself.

©IJRASET: All Rights are Reserved

The way a recommendation is presented may also show how good or relevant the item is considered to be. Relevance can be represented by the order in which recommendations are given. In a list, the best items are at the top. When a single item is recommended, it tends to be the best one available. Relevance can also be visualized using e.g. different colors and font sizes, or shown via ratings. Ratings can use different scales and different symbols such as numbers or stars. With the constant increase of Internet data and content, people attach greater importance to the function of application of recommended engines in Internet. It is conceivable that due to too much data on the Internet, it is very difficult for the user to the information he needs, it is far from enough to solve this problem only by providing search function. Recommendation system predict user preferences by the analysis of user behavior to make the users more easily to find the potential information they need. 3Data sets are the foundation of the recommendation system. The first step of video recommendation is to collect data as far as possible from the Internet. We download data from video websites, video forum, video online chat websites, and so on. Web crawlers, one of basic data sets collection, can download resources from Internet. The web crawlers originally used for search engine. In this paper, the results of crawler directly affect the accuracy of recommendation system.

II. SYSTEM DESIGN

Focus on "HOW" the data can be analyse using 3V's of big data. Break the large data into smaller pieces.Decide how each component works and how they Work together. The data can be combined by data deduplication To eliminate the similar data. By this process the large data can be compressed And get valuable and creative data information

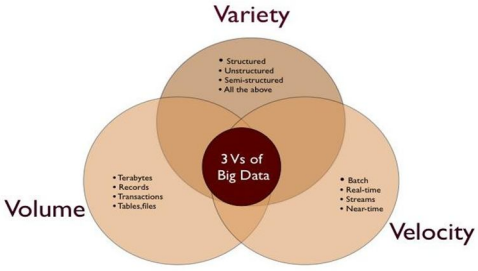


Figure 2: 3V's of big data

A. *3V's of Big Data*

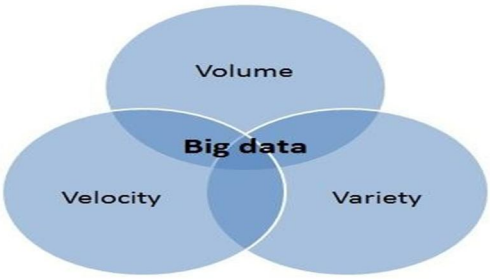


Figure 3: Volume, Velocity, Variety

1) *Volume:* The most obvious one is where we'll start. Big data is about volume. Volumes of data that can reach unprecedented heights in fact. It's estimated that 2.5 quintillion bytes of data is created each day, and as a result, there will be 40 zettabytes of data created by 2020 – which highlights an increase of 300 times from 2005. As a result, it is now not uncommon for large companies to have Terabytes – and even Petabytes – of data in storage devices and on servers. This data helps to shape the future of a company and its actions, all while tracking progress.
2) *Velocity:* The growth of data, and the resulting importance of it, has changed the way we see data. There once was a time when we didn't see the importance of data in the corporate world, but with the change of how we gather it, we've come to rely on it day to day. Velocity essentially measures how fast the data is coming in. Some data will come in in real-time, 13whereas other will come in fits and starts, sent to us in batches. And as not all platforms will experience the incoming data at the same pace, it's important not to generalise, discount, or jump to conclusions without having all the facts and figures.

©IJRASET: All Rights are Reserved

3) *Variety:* Data was once collected from one place and delivered in one format. Once taking the shape of database files - such as, excel, csv and access - it is now being presented in non- traditional forms, like video, text, pdf, and graphics on social media, as well as via tech such as wearable devices. Although this data is extremely useful to us, it does create more work and require more analytical skills to decipher this incoming data, make it manageable and allow it to work. Big Data is much more than simply 'lots of data'. It is a way of providing opportunities to utilise new and existing data, and discovering fresh ways of capturing future data to really make a difference to business operatives and make it more agile.

B. Data De-Duplication

Data deduplication refers to a technique for eliminating redundant data in a data set. In the process of deduplication, extra copies of the same data are deleted, leaving only one copy to be stored. Data is analysed to identify duplicate byte patterns to ensure the single instance is indeed the single file. Then, duplicates are replaced with a reference that points to the stored chunk. Given that the same byte pattern may occur dozens, hundreds, or even thousands of times think about the number of times you make only small changes to a PowerPoint file or share another important business asset the amount of duplicate data can be significant. In some companies, 80% of corporate data is duplicated across the organization. Reducing the amount of data to transmit across the network can save significant money in terms of storage costs and backup speed in some cases, savings up to 90%. While data deduplication is a common concept, not all deduplication techniques are the same. Early breakthroughs in data deduplication were designed for the challenge of the time: reducing storage capacity required and bringing more reliability to data backup to servers and tape. One example is Quantum's use of **file-based or fixed-block-based** 14**storage** which focused on reducing storage costs. Appliance vendors like Data Domain further improved on storage savings by using **target-based-** and variable-block based techniques that only required backing up changed data segments rather than all segments, providing yet another layer of efficiency to maximize storage savings.

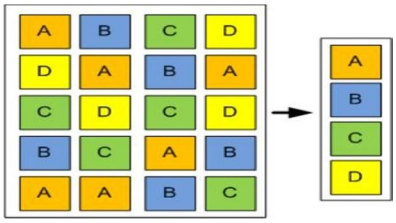


Figure 4: Data De-duplication

Used to improve storage utilization and can also be applied to network data transfers to reduce the number of bytes that must be sent. In the deduplication process, unique chunks of data, or byte patterns, are identified and stored during a process of analysis. As the analysis continues, other chunks are compared to the stored copy and whenever a match occurs, the redundant chunk is replaced with a small reference that points to the stored chunk. Given that the same byte pattern may occur dozens, hundreds, or even thousands of times (the match frequency is dependent on the chunk size), the amount of data that must be stored or transferred can be greatly reduced.

III. PROPOSED METHODOLOGY

A. System Architecture

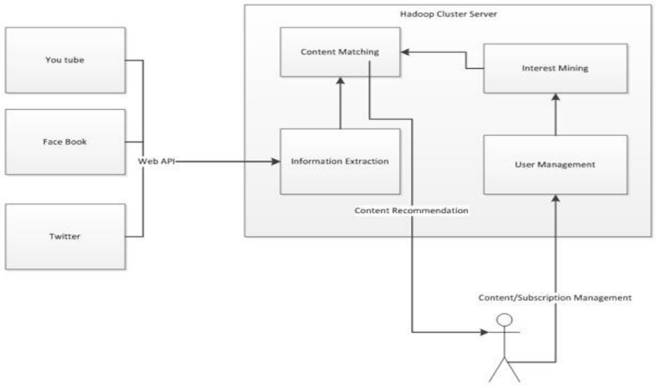
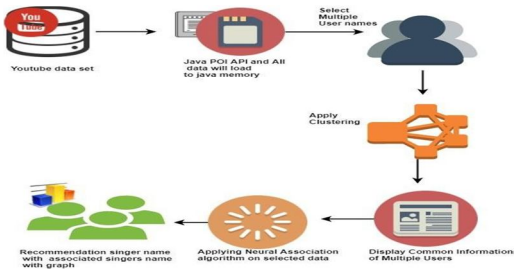


Figure 5: System architecture

1) Collect and organize information on users and products You need to know who your users are and what they are using. In our case, it was Klips, the data visualizations that drive engagement with data that Klipfolio users connect to in the product.
2) Compare User A to all other users Using those standard forms, you next design a function that compares User A to all other users Create a function that finds products that User A has not used, but which similar users have 16 Rank and recommend and the reviews given by the users, by this the recommendation can be justified Valuate and test the reviews and likes the recommended can goes on. WEB API where the data will be collected from the social medias or the data content store by the system. Information extraction will takes place by the system. Content matching will check the browse history or the reviews and likes given by the user .Then the user management gets the recommendation through the above process .

B. System Flow Diagram

Data source is the basic of MRS, since all the data what we need are download from the web sites. The video websites are the main data source from which we can download, such as youku, iQiyi, and so on. The contents that we need download include the video ID, brief description, click rate, ranking list and so on. We also download video comments from social networking service web sites, for example, Weibo, We chat, and micro-video forum.



MODULES

1) *Reading Data Set:* In the reading dataset first we need to read the dataset of you tube video links 17using Java POI API after successful reading of dataset we need to load all the data into Java memory
2) *Selecting Multiple Users:* After successful reading of dataset we need to select the multiple users. After successful selection of the multiple users the corresponding video links will bee fetched from the java memory.
3) *Clustering:* After fetching the video links of the corresponding user the following steps we need to follow to form the cluster

C. Structure Of System

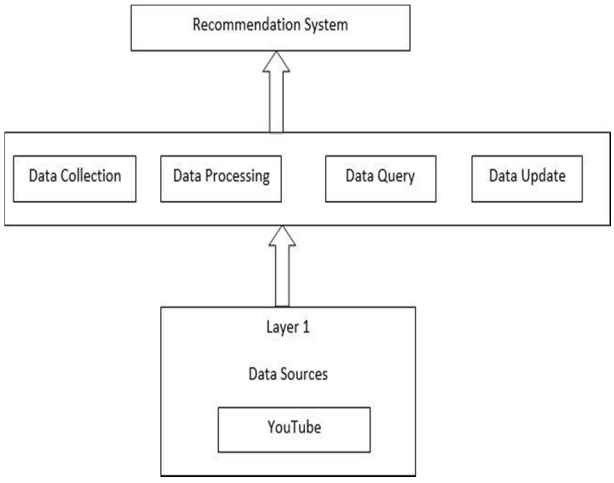


Figure 7: Structure of system

With the development of the new generation of information network technology represented by mobile Internet, big data, cloud storage, the traditional advertising micro video creation has a fundamental change. This project studies the design of the IT architecture and the division of the system service level in the era of big data, and designs the structure of the system.

D. Mathematical Model

Let S be the system,
S= {I, O, F}
Let I be the input to the system,(Dataset created based on below parameter)
I= {I1, I2, I3, I4, I5}
I1= Mouse Click
I2= Mouse Select
I3= Particular video playing time
I4= Mouse Hover
I5= Genres Search
O = Output of the Filtering Function that Gives Video Recommendation Let F be the functionality
F = {F1, F2, F3}
F1 = Generation of Factor Table
F2 = Calculation of Pearson Co-efficient
F3 = Apply Collaborative Filtering
F2 is the Calculation of Pearson Co-efficient, F2 F1
F3 is Appling collaborative filtering On Parameter getting from Pearson Co-efficient F3 F2 F3 is dependent on F2.
F3 = O O is the output of Function F3 i.e. Result matching to provide video recommendation. .
Mathematical representation of the video recommendation system.

E. UML Representation

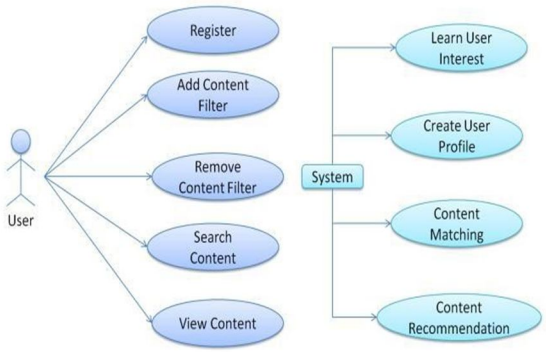


Figure 8: Use Case

By the users revies and the likes the recommendations starts by the user watching history or the previous browse history The content taken by the user will be consider for the recommendation perpose The system will get this data by the user as the IP address of an uses or the E Mail Address .By all this the user will be provide the related data content video by the recommendation system. Therefore the recommended video will be comprised na dit only contain the valuable information in that. It is very use full for all teenagers and young stars it will reduce time efficient and data consuming will be less

IV. CONCLUSION

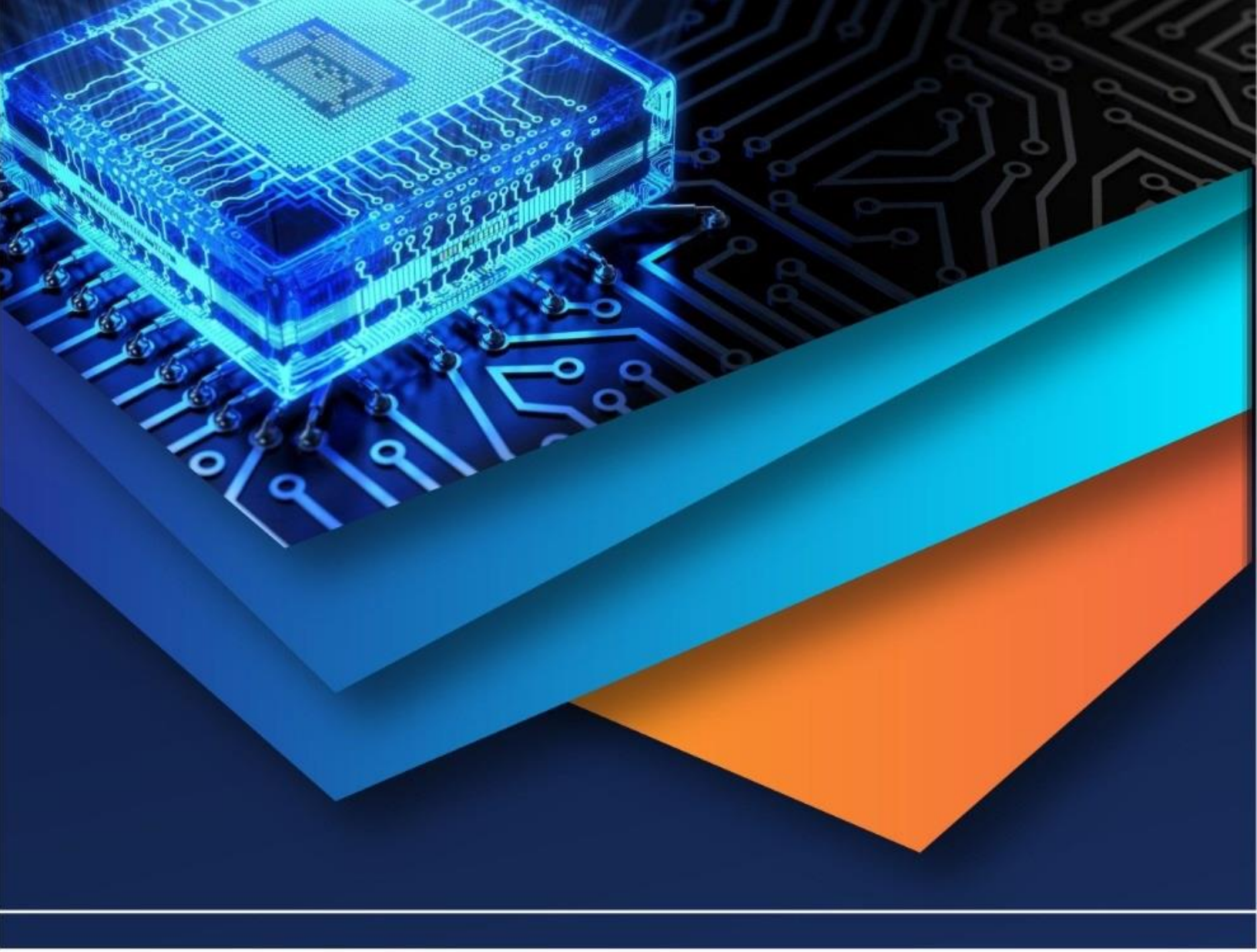
With the Development, the micro-video are increasingly common, especially young teenagers are likely to watch videos on mobile device According to the viewers browsing or watching history, this system can recommend the videos to the viewers. Big Data has just started, in which Data deluge is going to keep on increasing throughout the next years, and each data scientist will have to handle much more quantity of data every year. This data is becoming more diverse, larger, and faster. The core function of the system is the recommendation algorithms. The commonly recommendation algorithms are suite for tradition date sets, such as collaboration, and so on. However, with the development of Big Data, the recommendation algorithms should have the ability to deal with the Big Data. The Slope one algorithm is a Big Data micro video topic recommendation system, including data layer, interface layer, core layer, service layer, business layer. Then research on the key algorithms involved in the system, and pick up a

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

suitable algorithm. Finally, introduce the implementation of the system. video recommender system, i.e., when a new coming video is added to the library, the recommender system has to bootstrap the video relevance score with very few user behavior with respect to the newly added video. To solve this problem, we proposed a content-based video recommendation approach by taking the advantage of deep convolutional neural networks. As per the watcher's looking at or watching history, genres of videos, mouse movements, recommendation will be generated. With the help of big data, we are going to handle huge amount of video data set. This structure collects the above mentioned data and gives suggestions to the video maker. 26Specifically, we extract frame-level features using state-of- the-art deep models and propose to utilize synthetic anchor points to bridge the gap between training data and test data and cope with the serious data incompleteness. In addition to the vision feature, we also conducted extensive evaluation on video meta-data, and audio features. In summary, withthe supportfrom deep convolutional neural networks, frame-level. The presented results increased our understanding of how viewers react to videos. We gained insights from using physically noninvasive measurements, which are relevant for our long-term goal of their use in an affective video recommendation system. However, our contributions are not limited to recommendation systems. Our approach and the affective graph visualization may also be meaningful in the film-making industry for evaluating effectiveness of certain scenes. This work could also be applicable in online learning environments that lack face-to-face interactions of traditional learning contexts. The inconveniences what we require at show is the strategies by which to locate the most venerated video. Something special, for video makers, what they control to is the thing that number of watchers like what sort of annals. In context of this see, this paper proposes a video suggestion structure. As per the watchers' looking at or watching history, this structure can support the most revered annals to the watchers. Obviously, this structure can collect the responses and give a few suggestion for video makers with what number of watchers like the video. The center furthest reaches of the framework is the recommendation figures. The all things considered Neural Association Cluster estimations are suite for custom date sets, for example, content-based suggestion, encourage. Notwithstanding, with the movement of Big Data, the suggestion estimations ought to have the ability to manage the Big Data. Plus, our framework will give suggestion for related vocalists names and records.

REFERENCES

[1] Y. Z. Li, T. Gao, and X. Y. Li, "Design of video recommender system based on cloud computing", Journal on Communications, Vol. 34, No. Z2, pp. 138-140, 147, 2013.
[2] Y. Li, "Development mode of micro-video communication", Academic Exchange, Vol. 248, pp.177-181, 2014.
[3] S. M. Meng, W. C. Dou, X. Y. Zhang, et al., "KASR: a keyword-aware service recommendation method on MapReduce for Big Data application", IEEE Transactions on parallel and distributed system, Vol. 25, No. 12, 2014.
[4] D. M. Zhou, Z. J. Li, "Survey of high-performance web crawler", Computer Science, Vol. 36, No. 8, pp.26-29, 53, 2009.
[5] G. Y. Su, J. H. Li, and Y. H. Ma, et al., "New focused crawling algorithm", Journal of Systems Engineering and Electronics, Vol. 16, No. 1, pp.199-203, 2005.
[6] X. W. Meng, X. Hu, and L. C. Wang, et al., "Mobile recommender system and their applications", Journal of Software, Vol. 24, No.1, pp. 91-108, 2013.
[7] G. X. Wang, H. P. Liu, "Survey of personal recommendation system", Computer Engineering and Applications", Vol. 48, No. 7, pp. 66-76, 2012.
[8] G. F. Sun, L. Wu, and Q. Liu, et al., " Recommendations based on collaborative filtering by exploiting sequential behaviors", Journal of Software, Vol. 24 No. 11, pp.2721-2733, 2013.
[9] L. Guo, J. Ma, and Z. M. Chen, et al., "Incorporating item relations for social recommendation", Chinese Journal of Computers, Vol. 37, No. 1, pp. 218-228, 2014.
[10] Y. R. Wang, M. Chen, and H. H. Wang, et al., "A content-based filtering algorithm for scientific literature recommendation", Computer Technology and Development, Vol .21, No. 2, pp. 66-69, 2011.
[11] W. F. Liu, J. Z. Gu, and X. Lin, et al., " A Big Data management and analysis system based on Hadoop and Mahout", Computer Application and Software, Vol. 32, No. 1, pp. 47-50, 2015.
[12] E. Jain, S. K. Jain, "Twitter users on the basis of their interests using Hadoop/Mahout platform", Industrial and Information Systems (ICIIS), 2014 9th International Conference, pp. 15-17, 2014. 28
[13] B. Fan, C. Jia, "Visual framework for big data in d3.js", Electronics, Computer and Applications, pp.47-50, 2014.
[14] Y. Z. Wang, S. Mao, "A blocks placement strategy in HDFS", Computer Technology and Development, Vol. 23, No. 5, pp. 90-92,96, 2013.
[15] X. S. Zhang, H. L. Wang, "AJAX Crawling Scheme Based on Document Object Model", Computational and Information Sciences (ICCIS), pp. 1198-1201, 2012.
[16] C. Qian, X. L. Yang, " Design and Realization of the Web Crawler Supporting Ajax Framework", Computer & Digital Engineering, Vol. 40, No. 4, pp. 69-71, 98, 2012.
[17] http://www.seleniumhq.org/ M. Lan, C. L. Tan, J. Su, and Y. Lu, "Supervised and Traditional Term Weighting Methods for Automatic Text Categorization", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 31, No. 4, pp. 721-735, 2009.
[18] S. J. Gong, H. W. Ye, and H. S. Tan, "Combining Memory-Based and Model- Based Collaborative Filtering in Recommender System", 2009 Pacific-Asia Conference on Circuits, Communications and System, pp. 690-693, 2009.
[19] G. R. Bam note, S. S. Agrawal, "Evaluating and Implementing Collaborative Filtering Systems Using Apache Mahout", 2015 International Conference on Computing Communication Control and Automation, pp. 858-862, 2015.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)