Bayesian K-means clustering based quality of experience aware multimedia video streaming

Manjunatha Peddareddygari Bayya Reddy, Sheshappa Shagathur Narayanappa
Department of Information Science and Engineering, Sir M Visvesvaraya Institute of Technology, Bengaluru, India

Article Info

Article history:

Received Oct 12, 2023
Revised Nov 17, 2023
Accepted Nov 21, 2023

Keywords:

Amazon prime video
Bayesian K-means clustering
From patches to pictures
Multimedia streaming
'Patching Up' the video quality problem
Quality of experience

ABSTRACT

Media streaming is an essential approach for delivering multimedia information from the source distributor to the end-user through the Internet. Along with the development of more number of users and the spread of mobile devices, the availability and diversity of multimedia applications has also increased. Multimedia users primarily prioritize quality of experience (QoE), as they seek to access multimedia content with high availability and enjoy smooth video streaming in the shortest possible time. The impact of video delivery plays a significant role in QoE, which is efficiently made by delivering the content through a specialized content delivery network architecture. In this research, a Bayesian K-means clustering algorithm is proposed for the identification of QoE in multimedia video streaming. In this multimedia video streaming, the Amazon Prime video dataset is utilized for determining the performance of the proposed model. The proposed method is developed from the 'Patching Up' the video quality problem (PatchVQ) model, the from patches to pictures (PaQ-2-PiQ) model is utilized for the spatial feature extraction, and 3D ResNet-18 is utilized for temporal feature extraction. The proposed Bayesian K-means achieved a QoE reward function of 5,237.42 and 5841.36 as well as a fairness reward function of 5,841.36 and 8,732.08 at the speed of 1,500 kB/s and 2,000 kB/s respectively.

Corresponding Author:

Manjunatha Peddareddygari Bayya Reddy
Department of Information Science and Engineering, Sir M Visvesvaraya Institute of Technology
Bengaluru, India
Email: manjupb.reddy@gmail.com

1. INTRODUCTION

The With the massive development of the internet, multimedia services are becoming increasingly popular among the internet users worldwide [1]. Media streaming is a significant approach for transmitting multimedia data from the source provider to an end-user through the internet [2]. Due to media streaming, the utilization of smartphones and other such portable devices are increasing day by day [3]. Multimedia applications are commonly used in scenarios wherein multiple vehicles need to access similar type of data, by switching from one base station to another. The short transmission of infrastructure components like road side units (RSU) and base station (BS) provides inefficient solutions for vehicle-to-infrastrucure (V2I) communication [4]. Now, video streaming is identified to be a significant internet traffic source, and its pursual is developing at a faster rate. Video sharing and calling is the activity of most occurrence in this epoch that enables communication among users in various regions across the globe which is effectively run by hypertext transfer protocol (HTTP) [5]. This protocol gives a tool for adaptive streaming called hypertext transfer protocol adaptive streaming (HAS). This tool has become popular for utilizing HTTP as its major approach for transportation via the internet medium [6]. The quality of experience (QoE) is the service distributors' major objective in the present-day network, due to its closest accessibility to consumer choices [7]. These choices create a customer bond between them and the service, thus facilitating greater quality of service, in the minimum duration [8]. A multiview video system contains multiple viewpoints which shows the scenes from various viewing directions, collected by a group of videos cameras. The users select their viewing preference by navigating through the available viewpoints [9]. Hence, QoE has been embraced by the scientific community as a significant approach to estimating video streaming services' performance [10]. A better QoE for an end user is potentially possible by utilizing multiple constituents like initial delay, quality switch and stall frame [11], [12]. Real-time multimedia streams critically need to be efficient for an end user's quality of view. The video coding tool is based on majorly identifying a parameter of optimal coding. QoE determines user satisfaction and is quantified as a number in the interval, which is based on the application and its evolving state [13]–[15]. A number of researches, regard security as the main goal rather than it being just a precondition to optimizing performance of the network. However, this often depends on the screening of a large number of historical data, and the improvement of performance depends too much on the accuracy of cache. In machine learning (ML)-based multimedia video streaming, the prediction accuracy is majorly dependent on the parameter setting as well as false results that are produced.

Naresh *et al.* [16] implemented an asynchronous actor-critic (A3C) with actor-learner architecture for the development of adaptive bit rates for video streaming in the environment of internet of things (IoT). This method combined the follow then forage exploration (FFE) and average A3C algorithms to solve the high variance in value estimates. This method determined the video streaming benefits over various network constraints and also in terms of various alternatives to QoE metrics. This method enhanced the QoE by utilizing the A3C algorithm rather than the adaptive bit rate (ABR) algorithm as well as reduced the computation time while training. However this approach did not control the long-term bandwidth fluctuation problems. Wu *et al.* [17] developed a plan for QoE-aware video delivery in multimedia IoT networks by potential eavesdroppers. In this method, the selection of joint video quality as well as multicast group beamforming plan was developed to enhance a QoE network at the time of removing the eavesdropping as much as possible. Successive convex approximation (SCA) method was utilized to optimize multicast group beamforming and minimize multicast video eavesdropping. This method effectively enhanced the QoE by reducing the delay time in the wireless transmission process. But this method still had some fluctuations and also required more space to store the infrequent content. Martinez-Caro and Cano [18] implemented prediction, detection, classification, and measurement of stalling events' duration utilizing the data from network and transport layers. This method utilized the two approaches to solve the problems in video streaming with the framework of dynamic adaptive streaming over HTTP (DASH). The first was intended to detect the stalling events. For simplicity, it was based on the behavior of the transport layer and also capable of classifying an internet protocol (IP) packet as belonging or not belonging to the stalling events. The second approach was able to predict when an IP packet in the multimedia stream belonged to the event with long short-term memory. This improved the accuracy and performance of the presented model by utilizing the QoE parameter. However, this model provided better accuracy only for the balanced datasets.

Mane and Khot [19] introduced an enhanced cross-layer method to increase the transmitted video's end-to-end delay and peak signal to noise ratio in the long-term evolution (LTE) mobile network. The suggested method's solution involved performing two modifications over tradtional QoE, namely, resource allocation and adaptive modulation, for an improvised and efficient QoE. Then the method individually determined the transmission of up and down links and for determination, a number of nodes were attached to the system which had video and non-video traffic. This method enhanced the quality of the video playback even for random channel fluctuations and also provided effortless playback experience to the user. The LTE network still had problems with certain security issues and it consumed a high computation cost. Tanjung *et al.* [20] developed a multiview video priority (MVP) with the quality-based (ABR) for the Multiview video by identifying the viewpoint lead as well as targeting the viewpoint quality of QoE. This viewpoint majorly chose the level of lower bit rate than rate-based ABR and it worked efficiently in the cognitive quality of video, particularly when the viewpoint switched repeatedly. This method also recognized the different delay viewpoint types that were, transition and quality delay. This method efficiently removed the transition delay as well as minimized the delay in the quality of viewpoint by the utilization of MVP algorithm. However, this method does not provide a high-quality video when watching the significant viewpoint. Rahman and Huh [21] implemented a context-aware hybrid multi-access edge computing (MEC)-aided quality-adaptation approach for achieving multiple objectives in mobile streaming. This method utilized the characteristics of video content, application-layer data, and HAS client settings to jointly adapt to multiple clients' bit rates. For long and short-duration videos, this method created two different strategies to optimize the performance. It enhanced the bandwidth frequency efficiency, fairness, and quality of the video. However, it consumed greater processing power and also required more memory to execute the model.

Taha *et al.* [22] presented a model for QoE evaluation in adaptive video streaming through wireless networks. This method was based on various frameworks related to video characteristics, switch strategy, segment length, video stalls, initial delay, and quality of service (QoS), for the creation of an optimized evaluation. This presented model delivered a positive pearson correlation value and this result provided a better performance than mean opinion score (MOS) and structural similarity (SSIM), for all video samples. This model did not utilize deep reinforcement learning by double Q-values against end users. Waheed *et al.* [23] implemented an efficient architecture for an adaptive bitrate video, that authorized efficient and effective streaming, caching, and distribution of the multimedia content. The experiments were carried out for factors impacting QoE and the outcomes were further evaluated and analyzed. This method was then evaluated based on throughput and response time attained from different segment sizes in adaptive bitrate video streaming. It achieved a minimum response time as well as maximum performance by utilizing by a content delivery network. However, it consumed high energy and was computationally complex in terms of its hardware. Wang *et al.* [24] implemented a multi-agent policy gradient for finite time horizon (MAPG-finite) for a nonlinear optimization achievement approach of the multiple agents rewards through the finite time boundary. This model comprised a bandwidth task for optimizing QoE as well as fairness achievements in multi-user, through online video streaming. It achieved a better result on QoE and fairness reward function but had a high computational complexity.

From the overall analysis, the existing methods contain some limitations such as, the inability to control the long-term bandwidth fluctuation problems, requirement of more space to store the infrequent content, provided better accuracy results only for the balanced datasets, consumed more processing power and also required larger memory to execute the model, high energy consumption and computational complexity in hardware. To address these complications, this research proposed a Bayesian K-means clustering algorithm for the identification of QoE in multimedia video streaming. The proposed method is developed by the 'Patching Up' the video quality problem (PatchVQ) model, the from patches to pictures (PaQ-2-PiQ) model is utilized for the spatial feature extraction, and 3D ResNet-18 is utilized for temporal feature extraction. The main contributions of this research are as follows:

- This research mainly focuses on introducing a new method for the identification of QoE in multimedia video streaming using the Bayesian K-means clustering algorithm.
- The deep neural network (DNN) is utilized for training and specifying the suggested model on the evaluation of QoE.
- The PaQ-2-PiQ is utilized for spatial feature extraction and the 3D ResNet-18 model is utilized for temporal features extraction.

The remaining sections are arranged as follows: section 2 describes the proposed methodology. Section 3 explains the Bayesian K-means clustering. Section 4 demonstrates the results and discussion and finally, section 5 provides the conclusion.

2. PROPOSED METHOD

This research proposes a QoE estimation framework that is carried out on a content distribution server (CDS). The objective of this framework is to provide real-time tracking of achieved QoE and suggest approapriate adaptation, that is, reduction in bitrate. The client-side QoE estimation methods are not appropriate for the real-time aspect of video streaming, since the information reported back to the server is always late. Figure 1 depicts a block diagram of the proposed method which denotes an interaction type among client and server.

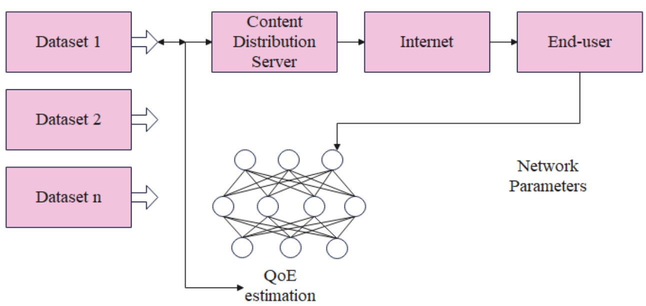

Figure 1. Block diagram of server-side QoE evaluation framework

The proposed approach seeks to capture the distinct phenomenon of video streaming QoE, for example, the observation of minimum startup delays having some effect on QoE, or identifying rebuffering events that seriously influence the QoE. As a result, the proposed method is designed such that, it is capable of seemless execution under real-time scenarios and performance as a proactive operating model for prediction. It is built to examine multiple operations of the CDS, the internet or network, and the client.

A CDS has entrée to binary input sources such as visual information encoded in multiple sorts of similar videos as well as parameters of contemporary networks. Regarding visual information, a server has entrée of multiple sorts is designed by utilizing the various condensation parameters on the H.264 standard. For condensed stream generation, traditional videos are condensed by utilizing the FFmpeg. A condensed parameter consists of rebuffering duration, stalling events, and throughput. The server is entrée to multiple versions of condensed videos and is as good as real-time information of network states, for evaluating the user's QoE.

For visual features, the proposed method comprises PatchVQ [25] model, which contains multiple stages that are spatiotemporal feature extraction, feature pooling, and the regression of the temporal. The spatio-temporal feature extraction happens by considering the four scales for the sequence of every video: the full video, sv-patch, tv-patch, and stv-patch. Consequently, the extraction of the feature is performed by working with DNN-based architecture [26], particularly the region-based convolutional neural network (R-CNN) and residual network (ResNet) models. Spatio-temporal features are extracted by utilizing the architecture of 3D ResNet-18. The feature pooling is executed for reduction of various trainable parameters and also to enable the network to highlight the correct RoI. A faster R-CNN is developed for spatial as well as temporal domains to extract features, and it implements the region approach process to choose a suitable region [27], [28].

2.1. Training of DNN for QoE evaluation

Subsequently, the training dataset must be thoroughly utilized to train the network in order to optimize or improve its performance. Three major factors namely loss function, optimization algorithm, and backpropagation, are utilized for efficiently optimizing a neural network. Using these factors, an iterative method is developed for training and later established by sustaining training data in the network when randomly initializing the weight of each layer. The outcome generated from this process is then analyzed alongside the expected label, using the support of a loss function. This estimates the network error which is accurately classified as an input. Now, the training network's main task is to reduce the error. Providing an input set x, DNN generates the recognized label $\hat{y}$, which is estimated as opposed to the ground truth y utilizing the $\mathcal{L}1$ error metric as shown in (1).

$$\mathcal{L}(x, y) = \sum_i |y_i - \hat{y}_i| \quad (1)$$

Providing a loss function output and a set of error gradients according to every network weight is accomplished by utilizing the backpropagation algorithm. Then, the gradients are used for optimization, generally imitative of gradient descent, which fine-tunes every weight for reducing an error and for identifying a local optimum.

2.2. Workflow of compressed video segment and associated metadata

The proposed model is developed based on the models of p.1203 and PatchVQ. It consists of multiple processes that are, feature extraction, spatiotemporal pooling and temporal regression. The PaQ-2-PiQ is implemented for the spatial feature extraction and 3D ResNet-18 is for the temporal features. The spatiotemporal pooling utilizes the spatial pool, which is followed through the segment of interest. The temporal pool method is then applied, from which the obtained outcomes are provided to the Inception Time model. The PatchVQ has the effect of incorporating contemplation of semantic data of video features. To add those features, the model has to modify an Inception Time component of PatchVQ. Figure 2 Shows the block diagram of the compressed video segments and associated metadata, which generates a MOS.

An inception time architecture contains multiple inception blocks, each of which consists of many parallel convolutions in various filter lengths, followed by the concatenation layer to integrate outcomes of the parallel convolutions. The inception blocks are arranged one after the other, that makes up for the whole architecture of inception time. Outcomes of inception time (IT) block network are accompanied by global average pooling (GAP) as well as fully connected (FC) layers through the softmax activation function. In this approach, the FC layer of an IT constituent is changed into the outcome of 20 values instead of one. Later, the 20 values are integrated into a vector consisting of video features and streaming data consisting the recent throughput of the network. Moreover, this model extends two FC layers, FC1 and FC2, to authorize the model to deduce the relationships between MOS and the given streaming metadata.

Process of training model: For the process of training, this method is performed with the $L1$ loss function and an Adam optimization for size 128. The considered data for implementation, is the Amazon Prime video dataset, which comprises of 420 video streams from 15 various uncompressed videos, which have been encoded based on given dataset information. From these 15 original videos, 13 are considered for training and the rest 2 are for testing. For every video, the applicable streaming versions are determined during the training as well as the validation of the model.

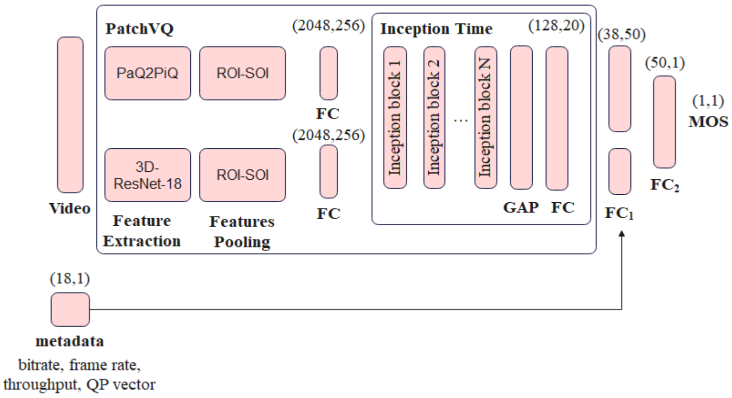

Figure 2. Block diagram of the compressed video segments and associated metadata

3. BAYESIAN K-MEANS CLUSTERING

A novel methodology of Bayesian K-means clustering is proposed for the identification of QoE in multimedia video streaming. The clustering technique is helpful to acquire basic insights and information of streaming data as well as automatic recovery from detection failures without user intervention. Clustering is a process of dividing the unlabeled data points into a number of grouped data points, that are similar to the other data points within the group, and dissimilar to data points in other groups. Clustering is basically the inherent grouping between unlabeled data and there is no criteria for good clustering. It is fully based on users and the criteria they use which satisfy their needs. It is a procedure of creating a group of conceptual features into similar feature classes. A cluster of data objects can be treated as one group. While performing a cluster analysis, a set of data is initially partitioned into groups according to the data similarity after which labels are assigned to the groups. The K-means clustering is a non-hierarchical cluster evaluation process to partition existing objects into single or multiple clusters based on their features. Every object with similar features is formed into a similar cluster and objects with non similar features are grouped into other clusters. The k-means clustering approach tries to classify the multimedia streaming information contained into various groups. An average K-means algorithm is a simple and popular unsupervised learning approach. It is an enlargement of a vector quantization approach used in the processing of signals. With various evaluation approaches, various clustering classification algorithms are naturally derived, which are produced based on various foundations. The splitting or aggregation is chosen for screening and after, a suitable cluster is chosen from an outcome as an outcome and an average algorithm's formula is expressed in (2) and (3) as;

$$\sqrt{\sum_{j=1}^{k}\left(X_i - C_j\right)^2} \tag{2}$$

$$J = \sum_{j=1}^{k}\sum_{i=1}^{n}\left\|X_i^{(j)} - C_j\right\|^2 \tag{3}$$

The mean of data points μ_j within each cluster is estimated to update a centroid in each iteration. This process continues until there is no modification in the values of the centroids or the number of maximum iterations is accomplished. The data points similarity in every cluster is maximum and by minimizing the objective function, the identified centroid value is expressed in (4) to (6) as;

$$c^{(i)} = arg_j min\left\|x_i^{(j)} - \mu_j^2\right\| \tag{4}$$

$$\frac{\partial J}{\partial \mu_j} = -2\sum_{i=1}^{m}\left(x_i^{(j)} - \mu_j\right) = 0 \tag{5}$$

$$\mu_j = \frac{\sum_{i=1}^{m} x_i^{(j)}}{\sum_{i=1}^{m}(1)} \tag{6}$$

where, j and $arg_j min$ is the objective function, k is the number of clusters, X is the case i, C is the centroid for cluster j, n is number of cases, X_i and C_j are the parts in dimensional Euclidean space, $\|X_i^{(j)} - C_j\|$ is Euclidean distance among X_i and C_j, and $c^{(i)}$ is an output cluster vector of the K-means.

The approach of K-means is according to a spherical cluster, in which the data points converge near the centroid of a cluster. The K-means clustering algorithm's steps are as follows: $X = \{x1, x2, x3, \dots xn\}$ is the set of data points, $C = \{c1, c2, c3, \dots cn\}$ where $k \leq N$ is the set of selected data centers. Initially, "j" centers of the cluster are randomly selected, and after allocating each data point to every cluster center, the distances among each data point and the cluster centers are evaluated, out of which the minimum distance is selected. The re-evaluation of the new cluster center is expressed in (7).

$$J_i i = \left(\frac{1}{C_i}\right) \sum_{j=1}^{C_i} Xi \tag{7}$$

The K-means average algorithm is set to various clusters to be grouped at inception and afterwards, a grouping motive is achieved by repeatedly modifying until the group center is stable. The K-means algorithm is based on a summation of the squared difference of the distances among every point in every cluster and the center of the cluster to which it exists. The K-means average algorithm is to set the number of clusters to be grouped at the beginning and, then, achieve the purpose of grouping by repeatedly modifying until the group center is stable and there is no change. K-means averaging algorithm is based on the sum of the squared difference of the distance between each point in each cluster and the cluster center to which it belongs, as small as possible.

Through K-means, the average approach has features of greater calculation speed as well as real-time update of a group cluster. It examines, classifies and recognizes the frequency of the computerized numerical control (CNC) milling machine by various drilling approaches and various materials. Figure 3 depicts the block diagram of the K-means grouping cluster. In this method, the evaluation outcomes from samples of sound detection on a process of video are taken ten times.

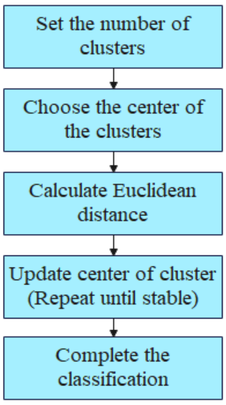

Figure 3. Block diagram of the K-means grouping cluster

3.1. QoE level identification by using Bayesian clustering

This section describes the identification of the observation points by utilizing the non-parametric inference with the collapsed gibbs (CG) and this method utilized the clustering model for the modeling of the data. This method required to identify the vector of the cluster which is explained as $\vec{C} = [c_1, c_2, \dots, c_N]$, where c_i illustrates the cluster to which $\vec{x}_i$ belongs. For evaluation of the parameters, the infinite gaussian mixture model is majorly utilized to perform the channel clustering by supporting the various algorithms. For clustering, this method utilizes the markov chain monte carlo (MCMC) algorithm for obtaining accuracy in the clustering method. Moreover, a CG algorithm is used to derive an inference about the dataset and that is an understanding of the MCMC algorithm. For the sampling of CG, a random sample sequence is needed for evaluating the secret parameters, which are acquired by an integrated probability of distribution of the number of random variables with less complexity. Identification of $\vec{C}$, is computed as $P\left(c_i = k|\vec{X}, \vec{C}_{-1}, \alpha{:}\vec{H}\right)$, rather than identifying an integrated probability of distribution because of the difficult integration. The Bayesian rule $P\left(c_i = k|\vec{X}, \vec{C}_{-1}, \alpha{:}\vec{H}\right)$ is provided in (8) and (9).

$$P\left(c_i = k|\vec{X}, \vec{C}_{-1}, \alpha : \vec{H}\right) = P\left(c_i = k|\vec{X} - i, \alpha, \vec{\theta}_k, \vec{H}, \vec{x}_i\right) \tag{8}$$

$$P\left(c_i = k|\vec{X}, \vec{C}_{-1}, \alpha : \vec{H}\right) \propto P\left(\vec{x}_i|\vec{X} - i; \vec{H}\right)P\left(c_i = k|\vec{C}_{-1}, \alpha\right) \tag{9}$$

where, $\vec{H}$ is hyper-parameters by (MS-t) distribution, which is utilized to identify a normally distributed population mean with secret standard deviation as well as minimum sample size.

4. RESULTS AND DISCUSSION

In this research, a Bayesian K-means clustering algorithm is proposed to identify a QoE in multimedia video streaming. The proposed Bayesian K-means clustering method is implemented on the testbed utilizing Python 3.5. The outcomes for the QoE, and fairness reward function with their policies, are shown in Table 1 and Table 2. Figure 4 represents the grapgical representation of QoE with policies. Figure 5 Shows the graphical representation of fairness with the policies.

Table 1. Reward breakdown for QoE function on 1500 kB/s and 2000 kB/s download link

Policies	Total reward (QoE)	
	1500 kB/s	2000 kB/s
Even	3877.05	4935.52
Adaptive	1773.30	4292.47
SARSA	3799.09	4936.76
MPAG-finite	4781.88	5690.77
K-means	5105.22	5764.1
Bayesian K-means	5237.42	5841.36

Table 2. Reward breakdown for fairness function on 1500 kB/s and 2000 kB/s download link

Policies	Total reward (Fairness)	
	1500 kB/s	2000 kB/s
Even	7053.77	7796.12
Adaptive	6927.24	8251.66
SARSA	6037.97	6486.17
MPAG-finite	7394.76	8441.75
K-means	7472.18	8562.99
Bayesian K-means	7561.82	8732.08

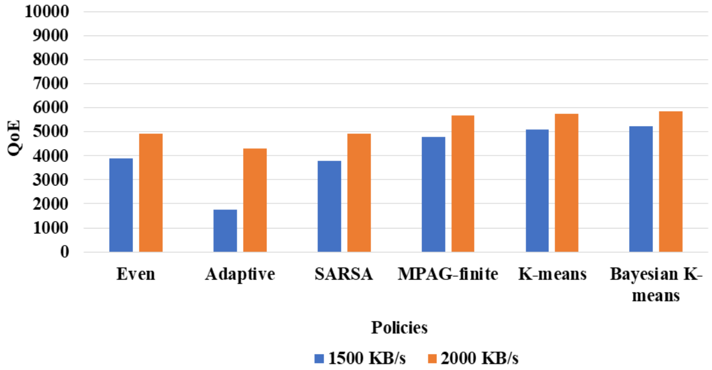

Figure 4. Graphical representation of QoE with the policies

Table 1 and Figure 4 represent the outcomes of the QoE reward function with the 1500 kB/s and 2000 kB/s download link. The existing policies namely even, adaptive, SARSA, MPAG-finite, K-means are compared with the proposed Bayesian K-means. The proposed method obtained the QoE reward by the download links 1500 kB/s and 2000 kB/s of 5237.42 and 5841.36 respectively. The proposed method achieved better QoE than the other policies.

Table 1 and Figure 5 represent the outcomes of the fairness reward function with the 1500 KB/s and 2000 kB/s download link. The existing policies even, adaptive, SARSA, MPAG-finite, K-means are compared with the proposed Bayesian K-means. The proposed method obtained the Fairness reward function of download links 1500 kB/s and 2000 kB/s respectively at 6561.82 and 8732.08. The proposed method achieved better fairness than the other policies.

Table 3 and Figure 6 represent the users video bitrate and users download rate distribution decided by a Bayesian K-means to a data rate (Mbps). Table 3 depicts an average download bandwidth for the tested users according to Bayesian K-means strategy with fairness function, on 2500 kB/s download link. Table 4 and Figure 7 represent the achieved QoEs with Policies for without and with Bayesian. The proposed Bayesian K-means achieved better results compared to other methods. The policies with Bayesian achieved a higher QoE than those without Bayesian.

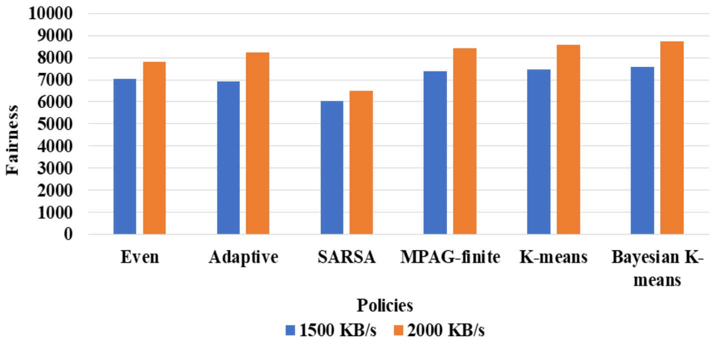

Figure 5. Graphical representation of fairness with the policies

Table 3. Analysis of average video bitrate and download rate with the users

Users	Data rate (Mbps)	
	Video bitrate	Download rate
1	8	7
2	8	7
3	8	7
4	6	5
5	5	4

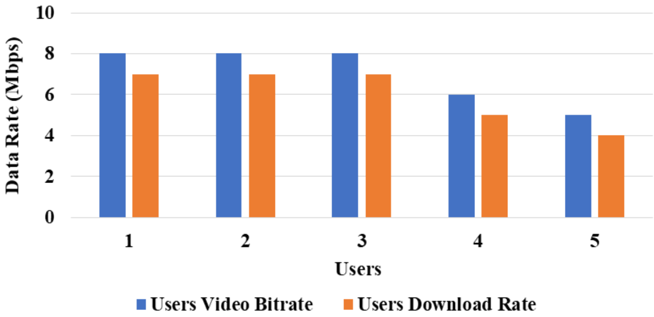

Figure 6. Graphical representation of data rate with the users

Table 4. QoE reward comparisons without Bayesian feature and with Bayesian feature

Policies	Without Bayesian	With Bayesian
Even	4935.52	4921.96
Adaptive	4292.46	4279.49
SARSA	4936.47	4921.91
MPAG-finite	5690.77	5680.03
Bayesian K-means	5805.22	5837.42

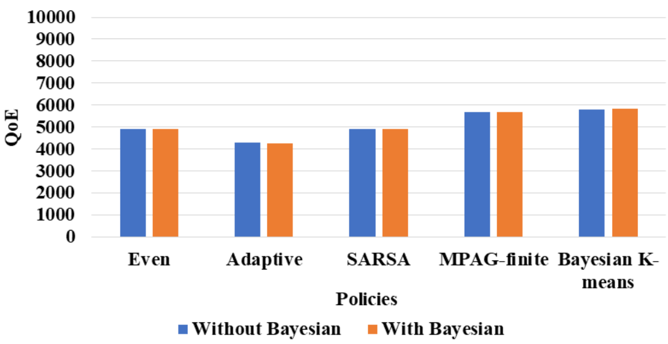

Figure 7. Graphical representation of the QoE with the policies for the without and with Bayesian

4.1. Comparative analysis

This section shows the comparative analysis of proposed Bayesian K-means clustering alongside few previously developed models. The metrics were QoE and fairness reward function, with several bitrates, as depicted in Table 5. The AL-FFEA3C [13] method utilized the 12 Mbps bitrate and achieved only 1055.40 in its QoE. The context-aware MEC [18] method utilized the 3450 Kbps bit rate and gains a QoE of 1186.91 with fairness of 0.92. The MPAG-finite utilized two bit rates of 1500 kB/s and 2000 kB/s and achieved QoEs of 4781.88 and 5690.77, correspondingly and fairness of 7394.76 and 8441.75, respectively. The proposed Bayesian K-means achieved better QoE reward functions of 5237.42 and 5841.36 with 1500 kB/s and 2000 kB/s bitrates. Furthermore, it gains better fairness reward functions of 5841.36 and 8732.08 with 1500 kB/s and 2000 kB/s bitrates respectively.

Table 5. Comparative Analysis of proposed method with existing methods

Author	Methods	Bit rate	QoE	Fairness
Naresh *et al.* [16]	AL-FFEA3C	12 Mbps	1055.40	N/A
Rahman and Huh [21]	context-aware MEC	3450 Kbps	1186.91	0.92
Wang *et al.* [24]	MPAG-finite	1500 kB/s	4781.88	7394.76
		2000 kB/s	5690.77	8441.75
Proposed	Bayesian K-means	1500 kB/s	5237.42	5841.36
		2000 kB/s	5841.36	8732.08

5. CONCLUSION

The multiview video has become a popular phenomenon in multimedia streaming services as it provides an immersive as well as interactive user experience throughout the Internet. In this research, a Bayesian K-means clustering is proposed for identifying the better QoE in multimedia streaming. In this multimedia video streaming, the Amazon Prime video dataset is utilized for evaluating the performance of a model. The proposed method is developed by the PatchVQ model, PaQ-2-PiQ model is utilized for the spatial feature extraction and 3D ResNet-18 is utilized for temporal feature extraction. The outcomes of the proposed method showcases its efficiency in terms of its higher QoEs and fairness reward functions at different download link bitrates. The proposed Bayesian K-means achieved QoE reward functions of 5237.42 and 5841.36 for bitrates of 1500 kB/s and 2000 kB/s, along with better fairness reward functions of 5841.36 and 8732.08 for bitrates of 1500 kB/s and 2000 kB/s respectively.

The proposed method contains certain limitations which are, the QoS parameters are affected the performance as well as the model had achived only the minimum latency. In the future, the proposed multimedia streaming will extend to improve the latency of this method. Moreover, the proposed method will extend to performed for the large-scale experiments with network scale users.

REFERENCES

[1] M. O. Elbasheer, A. Aldegheishem, N. Alrajeh, and J. Lloret, "Video streaming adaptive QoS routing with resource reservation (VQoSRR) model for SDN networks," *Electronics*, vol. 11, no. 8, p. 1252, Apr. 2022, doi: 10.3390/electronics11081252.

[2] P. Dai, F. Song, K. Liu, Y. Dai, P. Zhou, and S. Guo, "Edge Intelligence for adaptive multimedia streaming in heterogeneous internet of vehicles," *IEEE Transactions on Mobile Computing*, vol. 22, no. 3, pp. 1464–1478, 2021, doi: 10.1109/TMC.2021.3106147.

[3] E. Palacios Jara, A. M. Mezher, M. A. Igartua, R. P. D. Redondo, and A. Fernández-Vilas, "QSMVM: QoS-aware and social-aware multimetric routing protocol for video-streaming services over MANETS," *Sensors*, vol. 21, no. 3, p. 901, Jan. 2021, doi: 10.3390/s21030901.

[4] Q. Cheng, H. Shan, W. Zhuang, L. Yu, Z. Zhang, and T. Q. S. Quek, "Design and analysis of MEC- and proactive caching-based 360º mobile vr video streaming," *IEEE Transactions on Multimedia*, vol. 24, pp. 1529–1544, 2022, doi: 10.1109/TMM.2021.3067205.

[5] A. Erfanian, F. Tashtarian, A. Zabrovskiy, C. Timmerer, and H. Hellwagner, "OSCAR: on optimizing resource utilization in live video streaming," *IEEE Transactions on Network and Service Management*, vol. 18, no. 1, pp. 552–569, Mar. 2021, doi: 10.1109/TNSM.2021.3051950.

[6] W. G. de Morais, C. E. M. Santos, and C. M. Pedroso, "Application of active queue management for real-time adaptive video streaming," *Telecommunication Systems*, vol. 79, no. 2, pp. 261–270, Feb. 2022, doi: 10.1007/s11235-021-00848-0.

[7] A. Zhang *et al.*, "Video Super-resolution and caching—an edge-assisted adaptive video streaming solution," *IEEE Transactions on Broadcasting*, vol. 67, no. 4, pp. 799–812, Dec. 2021, doi: 10.1109/TBC.2021.3071010.

[8] S. A. Ansari, K. Pal, M. C. Govil, M. Ahmed, T. Chawla, and A. Choudhary, "Score-based incentive mechanism (SIM) for live multimedia streaming in peer-to-peer network," *Multimedia Tools and Applications*, vol. 80, no. 13, pp. 19263–19290, May 2021, doi: 10.1007/s11042-021-10709-2.

[9] S. Felici-Castell, M. García-Pineda, J. Segura-Garcia, R. Fayos-Jordan, and J. Lopez-Ballester, "Adaptive live video streaming on low-cost wireless multihop networks for road traffic surveillance in smart cities," *Future Generation Computer Systems*, vol. 115, pp. 741–755, Feb. 2021, doi: 10.1016/j.future.2020.10.010.

[10] D. Gupta, S. Rani, S. H. Ahmed, S. Garg, M. Jalil Piran, and M. Alrashoud, "ICN-based enhanced cooperative caching for multimedia streaming in resource constrained vehicular environment," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 7, pp. 4588–4600, Jul. 2021, doi: 10.1109/TITS.2020.3043593.
[11] R. Bhattacharyya *et al.*, "QFlow: a learning approach to high QoE video streaming at the wireless edge," *IEEE/ACM Transactions on Networking*, vol. 30, no. 1, pp. 32–46, Feb. 2022, doi: 10.1109/TNET.2021.3106675.
[12] A. A. Ahmed, S. J. Malebary, W. Ali, and O. M. Barukab, "Smart traffic shaping based on distributed reinforcement learning for multimedia streaming over 5G-VANET communication technology," *Mathematics*, vol. 11, no. 3, p. 700, Jan. 2023, doi: 10.3390/math11030700.
[13] A. T. Kalil Al-Khayyat and S. A. Mahmood, "Peer-to-peer media streaming with HTML5," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 2, pp. 2356–2362, Apr. 2023, doi: 10.11591/ijece.v13i2.pp2356-2362.
[14] N. Edan and S. A. Mahmood, "Multi-user media streaming service for e-learning based web real-time communication technology," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 1, pp. 567–574, Feb. 2021, doi: 10.11591/ijece.v11i1.pp567-574.
[15] M. A. Jasim and T. S. Atia, "An IoT-fuzzy based password checker system for wireless video surveillance system," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 6, pp. 3441–3449, Dec. 2022, doi: 10.11591/eei.v11i6.4375.
[16] M. Naresh, V. Das, P. Saxena, and M. Gupta, "Deep reinforcement learning based QoE-aware actor-learner architectures for video streaming in IoT environments," *Computing*, vol. 104, no. 7, pp. 1527–1550, Jul. 2022, doi: 10.1007/s00607-021-01046-1.
[17] D. Wu *et al.*, "QoE-aware video delivery in multimedia IoT network with multiple eavesdroppers," *Security and Communication Networks*, vol. 2022, pp. 1–11, Aug. 2022, doi: 10.1155/2022/4126494.
[18] J.-M. Martinez-Caro and M.-D. Cano, "On the identification and prediction of stalling events to improve QoE in video streaming," *Electronics*, vol. 10, no. 6, p. 753, Mar. 2021, doi: 10.3390/electronics10060753.
[19] A. D. Mane and U. P. Khot, "QOE improvement for dynamic adaptive streaming of multimedia in LTE cellular network using cross-layer communication," *International Journal of Wireless and Microwave Technologies*, vol. 12, no. 2, pp. 51–59, Apr. 2022, doi: 10.5815/ijwmt.2022.02.05.
[20] D. Tanjung, J.-D. Kim, D.-H. Kim, J. Lee, S. Kim, and J.-Y. Jung, "QoE optimization in DASH-based multiview video streaming," *IEEE Access*, vol. 11, pp. 83603–83614, 2023, doi: 10.1109/ACCESS.2023.3300380.
[21] W. Rahman and E.-N. Huh, "Content-aware QoE optimization in MEC-assisted mobile video streaming," *Multimedia Tools and Applications*, vol. 82, no. 27, pp. 42053–42085, Nov. 2023, doi: 10.1007/s11042-023-15163-w.
[22] M. Taha, A. Ali, J. Lloret, P. R. L. Gondim, and A. Canovas, "An automated model for the assessment of QoE of adaptive video streaming over wireless networks," *Multimedia Tools and Applications*, vol. 80, no. 17, pp. 26833–26854, Jul. 2021, doi: 10.1007/s11042-021-10934-9.
[23] M. H. Bin Waheed *et al.*, "A new efficient architecture for adaptive bit-rate video streaming," *Sustainability*, vol. 13, no. 8, p. 4541, Apr. 2021, doi: 10.3390/su13084541.
[24] Y. Wang, M. Agarwal, T. Lan, and V. Aggarwal, "Learning-based online QoE optimization in multi-agent video streaming," *Algorithms*, vol. 15, no. 7, p. 227, Jun. 2022, doi: 10.3390/a15070227.
[25] G. Margetis, G. Tsagkatakis, S. Stamou, and C. Stephanidis, "Integrating visual and network data with deep learning for streaming video quality assessment," *Sensors*, vol. 23, no. 8, p. 3998, Apr. 2023, doi: 10.3390/s23083998.
[26] S. Kim, U. E. Akpudo, and J.-W. Hur, "A cost-aware DNN-based FDI technology for solenoid pumps," *Electronics*, vol. 10, no. 19, p. 2323, Sep. 2021, doi: 10.3390/electronics10192323.
[27] D. Tsourounis, D. Kastaniotis, and S. Fotopoulos, "Lip reading by alternating between spatiotemporal and spatial convolutions," *Journal of Imaging*, vol. 7, no. 5, p. 91, May 2021, doi: 10.3390/jimaging7050091.
[28] P. Liu *et al.*, "A research on landslides automatic extraction model based on the improved mask R-CNN," *ISPRS International Journal of Geo-Information*, vol. 10, no. 3, p. 168, Mar. 2021, doi: 10.3390/ijgi10030168.

BIOGRAPHIES OF AUTHORS

Manjunatha Peddareddygari Bayya Reddy is currently pursuing Ph.D. in Visvesvaraya Technological University (VTU), he received M.E degree from Bangalore University (UVCE) and Bachelor of Engineering (B.E.) degree from Visvesvaraya Technological University (VTU) and he presently holds the position of assistant professor in the Department of Artificial Intelligence and Machine Learning, Bangalore Institute of Technology. He has having 13 years of experience in teaching and research. His research interests include networking, machine learning, and data analytics. He can be contacted at email: manjupb.reddy@gmail.com.

Sheshappa Shagathur Narayanappa is a Associate Professor in the Department of Information Science and Engineering, Sir M Visvesvaraya Institute of Technology, Bangalore, Karnataka, India. The author has completed Ph.D. from GITAM University, Andhra Pradesh. He has having 23 years of experience in teaching and research. His research interests include computer networking, machine learning, and data analytics. He has many publications and is presently workings on many more papers. He can be contacted at email: sheshappa_is@sirmvit.edu.