VOL. 10, NO 20, NOVEMBER, 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences
©2006-2015 Asian Research Publishing Network (ARPN). All rights reserved.

www.arpnjournals.com

IMPACT OF ONLINE STREAM CLUSTERING IN BANDWIDTH-CONSTRAINED MOBILE VIDEO ENVIRONMENT

P. M. Arun Kumar[1] and S. Chandramathi[2]
[1]Department of IT, Hindusthan College of Engineering and Technology, Coimbatore, Tamil Nadu, India
[2]Department of Electronics and Communication Engineering, Sri Krishna College of Technology, Coimbatore, Tamil Nadu, India
E-Mail: arunkumarpm@gmail.com

ABSTRACT

Mobile Video Streaming is becoming increasingly popular in today's Multimedia community. Various adaptive streaming techniques have been proposed by multimedia researchers to dynamically vary the video quality according to the available bandwidth. However, the deployment of best video adaptation techniques in real time is highly challenging due to critical QoE (Quality of Experience) requirements in wireless multimedia streaming. Resource constrained wireless multimedia networks demands better perception on the behavior of critical factors such as bandwidth in varying geographic milieu. In this paper, Machine Learning based online stream clustering is adopted to study the bandwidth impact in a streaming environment using 3G wireless video dataset. Massive Online Analysis (MOA) software framework is used to infer the results using algorithms such as CluStream and DenStream. The experimental result shows the effect of stream clustering based on unsupervised study.The measures such as Sum Square Error (SSQ) and Silhouette coefficient are deployed to perform cluster analysis. The results demonstrate the efficiency of CluStream with K means algorithm over density based streaming algorithm. The proposed framework justifies the scope of context aware computing applications in the broader areas of wireless multimedia.

Keywords: mobile video streaming, stream clustering, massive online analysis, bandwidth, machine learning.

INTRODUCTION

In recent years, video streaming over wireless networks is becoming a hot research area. Due to the advent and usage of various mobile devices on a large scale, demand for better video streaming services with high user expectations is inevitable. The quality of video content transmitted over wireless devices is critically affected by bandwidth fluctuations due to varying network conditions. The improvement in Video QoE can be achieved by adopting advanced adaptive streaming techniques. Streaming can be of three types namely Traditional Streaming, Progressive download and Adaptive Video Streaming [1]. Though dynamic nature of adaptive streaming is effective for today's mobile broadband networks, the huge volume of video streams evolving in real time impose stringent time/space constraints on the algorithms that process them.

Adapting the video quality effectively to the available network bandwidth is an important focus point in Video quality assessment process. Henceforth behavior of real time bandwidth fluctuations in various error prone networks needs extensive study to explore the possible impairments that affect end users' Quality of Experience. The paper depicts the usage of Machine learning algorithms in evaluating certain data from mobile video stream output through online clustering schemes. The real world bandwidth logs observed while streaming video over HTTP is taken for our experiments and evaluated using Massive Online Analysis (MOA) framework [2]. MOA supports WEKA machine learning workbench and it consists of online and offline algorithms for Classification and Clustering. The MOA software environment enables Massive Data Mining (MDM) and it provides online learning from evolving data streams. The results of the work suggest the relevance of bandwidth in 3G mobile video streaming using video stream clustering inferences. The cluster formation can be visualized on real time and the cluster analysis is carried out by various performance metrics.

The rest of the paper is organized in the following manner. Section 2 describes the related work in mobile video streaming. Section 3 outlines the characteristics of dataset and specifications of MOA. Section 4 describes the working of proposed model. Section 5 provides insights on Experimental evaluation and results. Finally, concluding remarks are highlighted in Section 6.

RELATED WORK

Since the advent of smart phones and tablets, video traffic remains the fastest multimedia traffic in mobile broadband networks. Various multimedia streaming strategies are deployed in real time and such techniques should bridge the gap between QoS and QoE dimensions so as to provide better efficiency for variety of applications. Real Time Streaming Protocol (RTSP) is a viable traditional streaming protocol with state-full nature in which Realtime Transport Protocol (RTP) is used for data transmission [1]. Relying on push based streaming such as RTSP demands specialized servers and hence such traditional streaming results in cost inefficiency. Also usage of UDP in such schemes results in large blocking

©2006-2015 Asian Research Publishing Network (ARPN). All rights reserved.

probabilities when moving through firewalls in network. In progressive download method, the client initiates the media to be played back while the multimedia file download is still in progress. This method is less flexible as the bit rate is constant throughout the session irrespective of bandwidth levels. Also here, bandwidth scarcity causes interruptions in video playback due to buffer underflow and non-viewing of fully downloaded video results in bandwidth wastage.

In recent years, Adaptive multimedia streaming provides effective measures to combat network impairments in wireless domain. In Adaptive technique, the video is encoded at multiple bitrates and using a rate adaptation criterion, the user with highest bandwidth watches the multimedia content with best quality and vice-versa. Rate adaptation in video streaming depends on critical factors such as bandwidth availability, device CPU availability and playback buffer size [3]. Hence the bandwidth plays a vital role in determining Video QoE of mobile streaming applications. The Multimedia industry makes use of variety of proprietary schemes that goes in line with the principles of adaptive video streaming. Smooth Streaming (Microsoft), HTTP Live Streaming (Apple), HTTP Dynamic Streaming (Adobe) are few examples [1, 4, 5]. Also, a new method named MPEG-Dynamic Adaptive Streaming over HTTP(MPEG-DASH) is endorsed by ISO as a standard for hybrid streaming applications [6]. In general, HTTP based streaming acts as pull based streaming and here client takes the critical role in imparting video adaptation. Market-friendly nature of HTTP and TCP/IP protocols, effective usage of existing HTTP servers instead of specialized servers provide better scalability and cost effectiveness in HTTP Adaptive Streaming (HAS) [11].

SPECIFICATIONS OF DATA SET AND SOFTWARE USED

Bandwidth data log

Real world data measurement of Bandwidth in a wireless scenario can be used effectively to study the inherent properties and behavior of such networks. The dataset considered in this paper is a bandwidth log data collected by applying Adaptive HTTP streaming over 3G networks using handheld devices [7]. The whole data is observed at application layer through HTTP based media streaming client using 3G network connectivity (UMTS and HSDPA). The study is conducted in a particular time period in Oslo, Norway and bandwidth measurement is done by considering different routes and different transportation means so as to observe the fluctuations. Each log entry contains fields such as Unix time stamp, monotonically increasing time stamp, GPS coordinates, number of bytes received and number of milliseconds elapsed [8].

MOA software environment

Massive Online Analysis (MOA) is a "software environment for implementing algorithms and running experiments for online learning from evolving data streams" [9]. MOA is a popular open source software environment written in Java and it encompasses collection of machine learning algorithms and tools for evaluation. MOA is applied extensively in Data stream mining and is released under GNU GPL license. MOA deals with a group of offline and online algorithms addressing both classification and clustering. The stream learning algorithms should process the incoming example with memory and time restrictions and it should portray the output at any time.

The general process of MOA framework aims at choosing a data feed followed by setting of learning algorithm that can be of type stream classification or stream clustering. The last stage deals with the evaluation techniques to infer results.

For stream classification, typical stream generators used are SEA concepts generator, STAGGER concepts generator, Random RBF generator, Function generator, e.t.c. Some of the important classifiers adopted by MOA are Naïve Bayes, Decision stump, Hoeffding tree, Hoeffding option tree, Bagging and Boosting [9].

In stream Clustering process, after the initial data feed, typical stream clustering algorithms provide evaluation measure as shown in Figure-1.

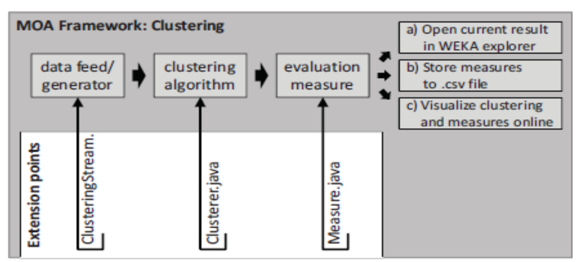

Figure-1. Clustering in MOA framework [9].

Stream clustering uses File stream or Random RBF data generator. Different parameter setting for kernel radius,speed,noise level,decay horizon is performed. Typical Stream Clustering algorithms used are StreamKM++, CluStream, ClusTree, Den-Stream, D-Stream and CobWeb.Various Internal and external evaluation measures are deployed to assess the performance of clustering output. The visualization component portrays stream visualization along with clustering results [10].

PROPOSED FRAMEWORK

The proposed model aims to study the behavior of bandwidth in adaptive streaming applied over wireless scenario. Machine learning based stream clustering is

©2006-2015 Asian Research Publishing Network (ARPN). All rights reserved.

applied to the log data collected by real world measurement and an unsupervised identification of clusters is done. Cluster analysis of the system provides critical insights on the effect of bandwidth fluctuations in mobile video streaming. The proposed framework is divided in to four major phases as shown in Figure-2.

The first phase is to set up a mobile video streaming test bed to record the bandwidth measurement in various routes over a period of time. The bandwidth log is observed and collected while streaming video over HTTP in real time. Mobile video receiver system with GPS and 3G connectivity records the GPS coordinates, timestamps and bytes received in a periodical fashion. Different routes with various transportation means such as car, bus, ferry are used for log measurement. The real world measurements accessed from [7] is used as dataset for our study.

The second phase does the accumulation of bandwidth log collected in the previous phase. Machine learning approach with unsupervised method is to be adopted by the system. Henceforth, the collected data is to be preprocessed according to the need and relevance. Data preparation process such as preprocessing and transformation of data is done. Appropriate preprocessing steps such as formatting and cleaning helps to fine tune the input data for better results while applying for machine learning systems. The dataset is processed through such steps and stored as ARFF (Attribute Relation File Format).

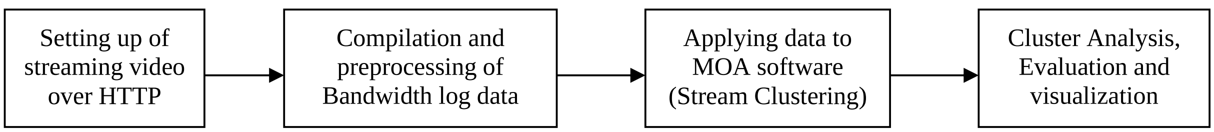

Figure-2. Phases of the proposed model.

The third phase of the proposed work aims to apply MOA software environment for the processed data using Stream Clustering. In general, clustering aims to extract homogeneous sub groups from the observed data. The dynamic flow of data from various applications has paved the way for exploration of novel data mining technique called Data stream mining. Data stream processing algorithms have properties such as single pass, finite storage and real time output generation. Hence, traditional clustering techniques are unsuitable for streaming applications. Data stream clustering algorithms provide an online-offline setup in which summarization of incoming data points as micro clusters is done in online process and applied to the offline component. Selection of data stream model, algorithm selection and performance evaluation are the three major steps followed by MOA framework.

In the final phase, Cluster analysis provides visualization of results and modifications of performance metrics over time. Cluster algorithm settings are given according to the need.The dataset is given as stream input. CluStream with k means and DenStream with DBSCAN (Density-Based Spatial Clustering of Applications with Noise) are the algorithms chosen for the analysis [12, 13].The static data is made to run in online stream clustering environment and the resulting clusters are visualized. Also by selecting a particular internal measure, cluster analysis is done. The evaluation inferences with real time graphical representations are observed.

EXPERIMENTAL EVALUATION

Massive Online Analysis (MOA) is used for the experimental evaluation and result analysis. The setting provided for capturing the results is briefed in Table-1.

Table-1. Parameter setting details for experimental evaluation.

Software used	**MOA (stream clustering framework)**
Input dataset source	Bandwidth log obtained through adaptive HTTP streaming [7].
Selection of data stream	ARFF formatted file with 47200 Log entries with each entry having 6 attributes.
Stream clustering algorithms considered for study	Algorithm 1 : DenStream with DBSCAN Algorithm 2 : CluStream with K means
Evaluation measures selected	Sum Square Error (SSQ) and Silhouette Coefficient

MOA graphical user interface allows selecting two different stream clustering algorithms to be operated on the above said dataset. CluStream algorithm observes statistical relevance about the data using micro clusters and DenStream algorithm uses core micro clusters to summarize the clusters. An Extensive study on various stream clustering algorithms is elucidated in [14] and critical features of density based stream clustering algorithm is surveyed in [15]. Internal validation criteria

ARPN Journal of Engineering and Applied Sciences
©2006-2015 Asian Research Publishing Network (ARPN). All rights reserved.

validate the cluster quality using only the clustering results gathered through the clustering algorithms. Sum Square Error (SSQ) measures the total of the squared distances from data points to their respective cluster centers. Better cluster formation is justified if lower SSQ values are observed. Silhouette Coefficient determines the compactness and separation of clusters. This value has a range between -1 and +1. Generally a value closer to +1 denotes good clustering quality [16, 17].

Figure-3 shows the MOA Graphical User Interface screen shot with the cluster algorithm setup where the input stream and two algorithms of DenStream and CluStream types are selected. The evaluation measures can be selected by the user and the stream clustering is done. The visualization of the cluster formation for both the algorithms is observed. The clusters generated from the input clearly reflect the underlying structure of the data.

Figure-4 and Figure-5 shows the evaluations of stream clustering algorithms based on SSQ (Sum Squared Error) and Silhouette Coefficient respectively. Both internal measures are used for cluster validation. The X axis is the data log values taken and Y axis is the chosen evaluation measure.

Figure-4 portrays SSQ evaluation analysis after running the experiment for 47200 data with respect to the two algorithms considered. For the given SSQ setting, first algorithm Den-stream with DBSCAN takes current value as 11.75 and mean value as 5.91 whereas CluStream with k means algorithm takes 1.76 as current value with mean value 1.35. The graph demonstrates a visible high peak fluctuating values of SSQ after 25000 data points for DenStream algorithm whereas its counterpart CluStream algorithm maintains low values comparatively. Hence lower SSQ values justify the superiority of CluStream algorithm used here.

Similarly Figure-5 shows the behavior of Silhouette Coefficient for the given datastream input. The Mean values with respect to Silhouette Coefficient for DenStream and CluStream algorithms are 0.68 and 0.80 respectively. Due to the inherent properties of density based stream algorithms Silhouette value reaches 1 consistently. But larger fluctuations are noted for the same during the running process and hence its mean value of 0.68 is marginally lesser than CluStream algorithm's mean value of 0.80.

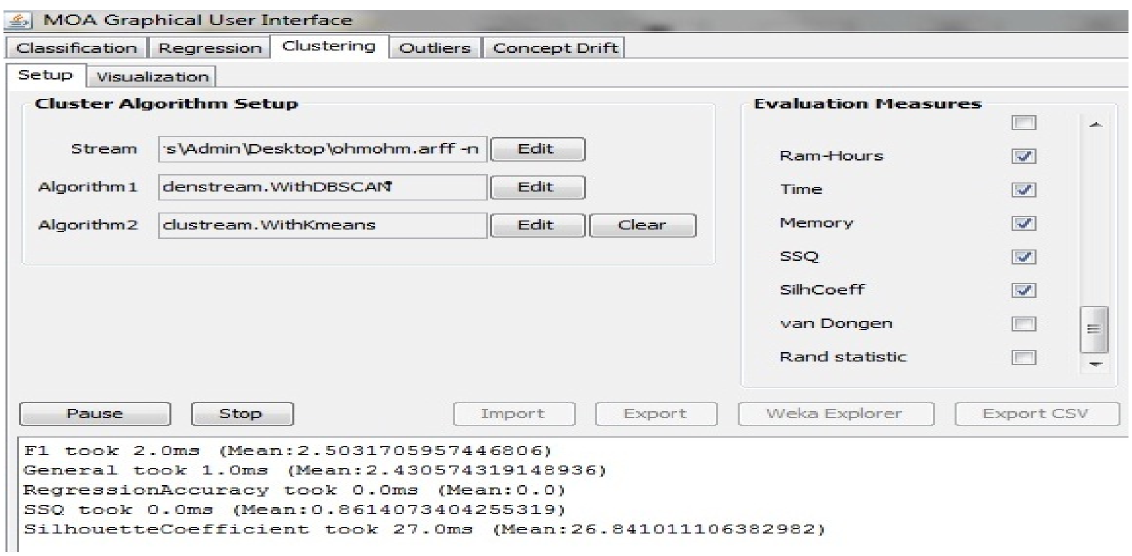

Figure-3. Snapshot of MOA user interface.

ARPN Journal of Engineering and Applied Sciences
©2006-2015 Asian Research Publishing Network (ARPN). All rights reserved.
www.arpnjournals.com

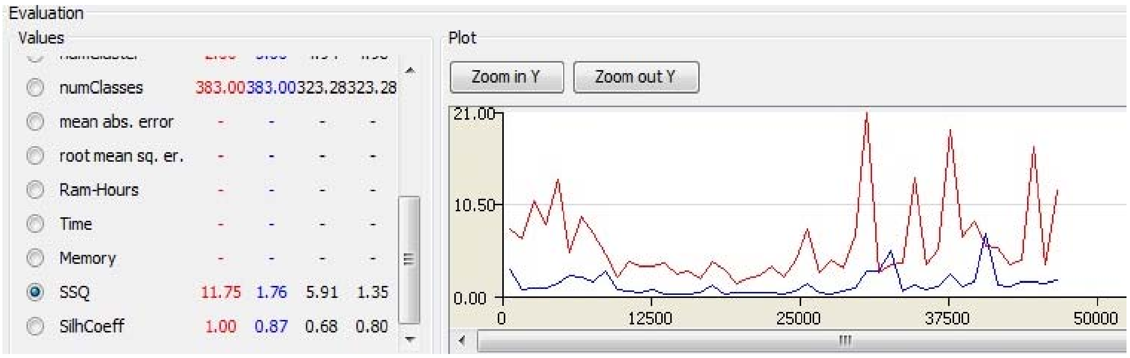

Figure-4. Screenshot of visualizing data (x axis) vsSSQ measure (y axis).

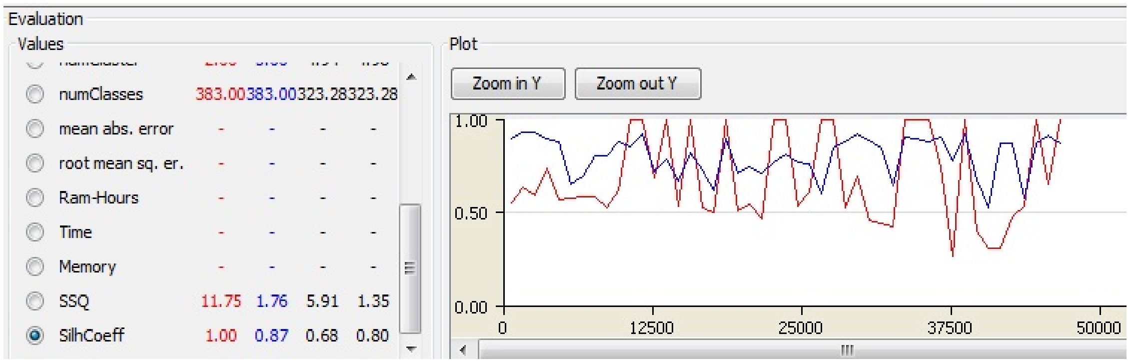

Figure-5. Screenshot of visualizing data (x axis) vs Silhouette Coefficient measure (y axis).

CluStream with k means algorithm fairs better than the DenStream algorithm during Silhouette evaluation. As the bandwidth log entries are collected by adaptive video streaming by various transport means at different geographical locations, the unsupervised cluster analysis of such data help in drawing better conclusion on reflecting the need for better throughput provisioning in certain areas that are vulnerable to wireless impairments. This will in turn help to improve the multimedia quality over mobile networks in real time adaptive environments.

CONCLUSIONS

The mobile clients in any multimedia network are often subjected to fluctuating network impairments and the effect of bandwidth in particular is noteworthy factor in achieving best end user experience. Video streaming applications in error-prone networks hardly achieves best QoE as end result due to the inherent limitations in mobile multimedia framework. Henceforth,better prediction on the bandwidth variations in mobile video streaming will provide a realistic benchmark for improving video QoE. The paper deploys machine learning based stream clustering method using MOA software environment for a real world Bandwidth log data and the cluster analysis is performed. The result highlights the behavior of the stream data in real time processing and it shows CluStream with k means algorithm outperforms the DenStream algorithm substantiated by cluster validation measures such as SSQ and Silhouette Coefficient. The stream clustering results and visualizations explored in the proposed work portrays the need for studying network impairment such as bandwidth in mobile video streaming and such explorationcan be extended to the frontiers of context aware mobile multimedia.

REFERENCES

[1] Zambelli A. 2009. IIS Smooth Streaming Technical Overview. Whitepaper. Microsoft Corporation.

[2] Bifet A., Holmes G., Kirkby R. and Pfahringer B. 2010. MOA: Massive Online Analysis.http://sourceforge.net/projects/moa-datastream/. Journal of Machine Learning Research (JMLR).

[3] Begen A., Akgul T. and Baugher M. 2011. Watching Video over the Web: Part 1: Streaming Protocols. Internet Computing. IEEE. 15: 54-63.

©2006-2015 Asian Research Publishing Network (ARPN). All rights reserved.

[4] Adobe. 2010. HTTP dynamic streaming on the Adobe Flash platform. https://bugbase.adobe.com/index.cfm?event=file.view&id=2943064&seqNum=6&name=httpdynamicstreaming_wp_ue.pdf.

[5] Pantos R., Batson J., Biderman D., May B. and Tseng A. 2010. HTTP live streaming. http://tools.ietf.org/html/draft-pantoshttp-live-streaming-04.

[6] Stockhammer T. 2011. Dynamic adaptive streaming over HTTP:standards and design principles. Proceeding of the 2ndAnnual ACM Multimedia Systems Conference (MMSys '11). New York, USA. pp. 133-144.

[7] Riiser H., Vigmostad P., Griwodz C. and Halvorsen P. 2012. DATASET: HSDPA bandwidth logs for mobile HTTP streaming scenarios. http://home.ifi.uio.no/paalh/dataset/hsdpa-tcp-logs/.

[8] HaakonRiiser. Paul Vigmostad. CarstenGriwodz and Pal Halvorsenn. 2013. Commute Path Bandwidth Traces from 3G Networks: Analysis and Applications. Proceedings of the International Conference on Multimedia Systems (MMSys). Oslo. Norway. pp. 114-118

[9] Bifet A., Holmes G., Pfahringer B., Kranen P., Kremer H., Jansen T. and Seidl T. 2010. MOA: Massive online analysis, a framework for stream classification and clustering. Journal of Machine Learning Research (JMLR) Workshop and Conference Proceedings. Workshop on Applications of Pattern Analysis. 11: 44-50.

[10] Kranen P., Kremer H., Jansen T., Seidl T., Bifet A., Holmes G. and Pfahringer B. 2010. Clustering Performance on Evolving Data Streams: Assessing Algorithms and Evaluation Measures within MOA. IEEE International Conference on Data Mining Workshops (ICDMW). pp. 1400-1403.

[11] Oyman O. and Singh S. 2012. Quality of experience for HTTP adaptive streaming services. IEEE Communications Magazine. 50(4): 20-27.

[12] Charu C., Aggarwal, Jiawei Han, Jianyong Wang and Philip S. Yu. 2003. A framework for clustering evolving data streams. In: Proceedings of the 29th international conference on Very large data bases. VLDB Endowment. 29: 81-92.

[13] FengCao, Martin Ester, WeiningQian and Aoying Zhou. 2006. Density-based clustering over an evolving data stream with noise. In: 2006 SIAM Conference on Data Mining. pp. 328-339.

[14] Shifei Ding, Fulin Wu, Jun Qian, HongjieJia and Fengxiang Jin. 2015. Research on data stream clustering algorithms. Artificial Intelligence Review. Kluwer Academic PublishersNorwell. MA. USA. 43(4): 593-600.

[15] Amini A., Wah T. Y. and Saboohi H. 2014. On density-based data streams clustering algorithms: A survey. Journal of Computer Science and Technology. 29(1): 116-141. DOI 10.1007/s11390-013-1416-3.

[16] Hassani M., Seidl T. 2015. Internal Clustering Evaluation of Data Streams. The fourth Quality issues, measures of interestingness and evaluation of data mining model workshop (QIMIE’15). Springer.

[17] http://jwijffels.github.io/RMOA/MOA_2014_04/doc/pdf Tutorial3.pdf.