Neural Adaptive Content-aware Internet Video Delivery

Title : Literature Review for “Neural Adaptive Content-aware Internet Video Delivery”
Name : Kiran Jojare
Date : Dec 3, 2022

The paper explains how the use of Content-Aware Adaptive Internet streaming is extremely useful in improving user quality experiences and reducing bandwidth constraints compared to conventional internet streaming methods, thanks to the use of Deep Neural Networks (DNN), which reduces dependencies for high quality video delivery.

The paper briefly describes the current video delivery system's limitations. The paper explains how, in today's video delivery system, video quality is heavily dependent on bandwidth availability. It also shows how underutilized the client's computing power is in terms of decoding video at the client side. The paper criticizes current video streaming techniques because their codecs are excellent for reducing redundancies in small GOPs but fail to do so in large GOPs.

The paper exploits this drawback in conventional internet streaming services and highlights how the use of a content adaptive deep neural network can overcome this drawback. Publishers highlight a few key observations in deep neural networks that can help in overcoming this requirement, which shows how DNN can use computing resources to improve video streaming quality, as well as how DNN are trained and operated on a large timescale for calculating a large number of GOPs, which is lacking in traditional internet video streaming services. The paper also highlights the use of SuperSolution DNN-based computing devices in achieving this above-highlighted system requirement for high-quality video streaming.

The paper very well argues and conceptualizes how the current limitations of convention streaming services can be overcome using DNN and how much QoE can be improved by implementing the system. This paper provides brief information about how this DNN-based content-aware system could be implemented by applying super-resolution DNN over bitrate adaptation. It shows how different video-on-demand services like Netflix and Hulu could use this in association with desktop-class GPUs in the real world.

The paper explains the key design principles inside NAS, which are Content-aware DNN, Multiple quality DNNs, Scalable DNN, Integrated ABR. It also discusses issues such as reliability in proving network quality, unpredictability in testing accuracy, implications for video encoding on the server side, enabling super real-time resolution on heterogeneous clients, and time required for utilisation.

Even after the above-mentioned challenges, I strongly agree with the publishers that implementing a DNN-based content-aware streaming algorithm could greatly improve the quality of experience and bandwidth constraints for clients. The paper also provides strong testing scenarios in agreement with this. However, I also believe that since DNN requires a lot of training, the reliability provided by DNN would not suffice in real-world scenarios. In the case of newly provided DNN content, this may result in unreliability. Also, I strongly believe that the use of computing power varies across systems and over time, and that researchers should have focused more on this aspect of computing power adaptation.

References :
[1] Hyunho Yeo, Youngmok Jung, Jaehong Kim, Jinwoo Shin, and Dongsu Han, "Neural Adaptive Content-aware Internet Video Delivery," Dec 5, 2023, at the 13th USENIX Symposium on Operating Systems Design and Implementation (OSDI ’18).