QoE driven CDN resource allocation approach: literature review

Hanyu Li

February 7, 2020

1 The method that is being used now by industry

The most commonly used method is randomly assign new users to currently available CDN(Content Delivery Networks), usually according to the hashed id of users. Apparently, this kind of round robin approach gives rise to a set of problems. For example, the dynamic CDN status is neglected. Furthermore, the QoE(Quality of Experience) of users are affected by various factors which may contain factors other than QoS(Quality of Service).

2 State of art works

2.1 Prediction based

2.1.1 confounding factors(2013)[1]

It aims to develop a predictive model of engagement that accounts for the complex relationship between quality and engagement, the interdependencies between quality metrics(e.g., buffering, join time), and the confounding factors(e.g., device, genre) that impact different aspects of this learning. The dataset has 40 million video viewing sessions collected over 3 months spanning two popular video content providers in US. The model

input are quality metrics and confounding factors, and it will output the prediction of the video engagement(QoE).

Its main observation is that decision tree is expressive enough for the quality metrics, while not for confounding factors. Thus they captured several most important confounding factors and proposed a predictive model where a decision tree is trained for every combination of confounding factors.

Its accuracy on *** testbed is

Its pros and cons are

2.1.2 user mapping stradegies(2016)[2]

Zhu et al. proposed a prediction model considering QoS metrics and the preference on videos of users. Afterwards they came up with a heuristic CDN mapping stradegy. The input when testtime is the video requests consist of (user, video), and a hyperparameter K. The output will be the mapping of (user, CDN).

The used a simple but effective method to model the user preference: the cosine similarity between videos based on the categorial vector, and users based on weighted historical sum of videos vectors. Then they feed the QoS factors and this preference factor in to a decision tree to get the predicted QoE. The mapping stradegy is also straightfoward. They simulate the assignment of users randomly and pick K users who has the largest QoE improvement and change their CDN in reality.

Its accuracy on *** testbed is

One of the disadvantages is that we can not change CDN for existing users frequently in live streaming scenarios, which is not compatible with the optimization part of this stradegy.

2.1.3 CFA(2016)[3]

It aims to capture the complex relationships between session features and video quality. Furthermore, it tries to address the need of fresh updates. Also being a predictive model, CFA takes feautres of sessions under prediction as input, and outputs their assigned CDNs based on the prediction of QoE on every CDN.

Its main observation is that few critical features is able to determine the QoE for a video session. Thus they proposed the prediction model which is formed by a critical feature learner and a quality estimator based on the sessions with the same critical features in a period of time. They also observed that critical features will remain for a relatively long period of time. Therefore CFA only updates the critical feature when there triggered(e.g., when the performence drop to a certain threshold). The quality estimator on the other hand is updated every ten seconds.

Its accuracy on *** testbed is

The QoE metric used to evaluate the model is a simple weighted summation of several QoS metrics, which is homogeneous for users, so there is still space to alter the assignment of CDN to achieve better QoE by considering user heterogeneity.

2.1.4 k-NN VCP(2019)[4]

This system monitors the QoE delivered by any of the CDN belonging to its pool and selects the most performing one when a new video request is received. The VCP takes QoE feedbacks of old users and ISP(Internet Service Provider) from all users as inputs and output the assigned CDN for old users with bad QoE and new users.

The authors performs a k-NN regression on the prediction of QoE by dividing users in to groups regarding CDN and ISP and taking means of the QoE. When a new user comes in, the CDN ranker ranks the candidate CDN based on the QoE prediction and selects the best.

Its accuracy on *** testbed is

The old users will be affected by new users and if their QoE declines they will be new users as well. Thus the greedy assignment for every users is not perfect.

2.1.5 crowdsourced live streaming(2019)[5]

This approach aims to minimize access delay and video stall costs by carefully choosing which edge server to allocate the broadcast vedio and how

to migrate between servers. Input: session features. Output: prediction of the video quality.

The predictive part will predict the number of viewers expected near each cloud site. Based on the predicted results, the optimization part will allocate live videos replicas across the geo-distributed cloud sites near the viewers proximity by integral programming.

Its accuracy on *** testbed is

Its pros and cons are

2.1.6 E2E(2019)[6]

E2E is the first resource allocation system that embraces user heterogeneity to allocate server-side resources in a QoE-aware manner. It is still a prediction-based system, consisting of the prediction of QoE and the optimization of the total QoE. Like its predecessors, the input is the requests and it outputs the assignment.

They made simple assumptions on the prediction part where the QoE is a single variable function of the sum of client and server side delay because they used database request as an example. However, considering the user heterogeneity means that the optimizing step will be much more difficult because they can not simply choose the best CDN or database for every user in a random order. On contrast, they have to take all users in the same time chunk in to consideration to make decisions. They managed to transform the complicated and coupled problem into two steps, and solved them seperately.

Its accuracy on *** testbed is

The prediction part can be modified to a more expressive model to fit in the video session scenario.

2.2 E2 based

2.2.1 pytheas(2017)[7]

The authors believe that data-driven QoE optimization should instead be cast as a real-time exploration and exploitation(E2) process rather than

as a prediction problem. Input: session features. Output: prediction of the video quality. The inputs includes historical measurements and the measurements of the requested session. The output will be the decision.

They Their argue that the methods based on prediction suffers from the prediction bias and slow reaction. Thus they proposed pytheas which formulates measurement collection(exploration) and decision making(exploitation) as a joint process with real-time QoE measurements using E2 process. They group the sessions based on the factors on which their QoE and best decisions depend, and run discounted UCB(Upper Confidence Bound) algorithm on each group.

Its accuracy on *** testbed is

Its pros and cons are

3 User heterogeneity: a new way to improve QoE

4 How we mitigate those drawbacks

Reference

[1] Athula Balachandran, Vyas Sekar, Aditya Akella, Srinivasan Seshan, Ion Stoica, and Hui Zhang. Developing a predictive model of quality of experience for internet video. In *ACM SIGCOMM 2013 Conference, SIGCOMM'13, Hong Kong, China, August 12-16, 2013*, pages 339–350, 2013.

[2] Guowei Zhu, Chou Mo, Zhi Wang, and Wenwu Zhu. User mapping strategies in multi-cloud streaming: A data-driven approach. In *2016 IEEE Global Communications Conference, GLOBECOM 2016, Washington, DC, USA, December 4-8, 2016*, pages 1–6, 2016.

[3] Junchen Jiang, Vyas Sekar, Henry Milner, Davis Shepherd, Ion Stoica, and Hui Zhang. CFA: A practical prediction system for video qoe optimization. In *13th USENIX Symposium on Networked Systems Design*

and Implementation, NSDI 2016, Santa Clara, CA, USA, March 16-18, 2016, pages 137–150, 2016.

[4] Luca De Cicco, Saverio Mascolo, and Vittorio Palmisano. Qoe-driven resource allocation for massive video distribution. *Ad Hoc Networks*, 89:170–176, 2019.

[5] Fatima Haouari, Emna Baccour, Aiman Erbad, Amr Mohamed, and Mohsen Guizani. Qoe-aware resource allocation for crowdsourced live streaming: A machine learning approach. In *2019 IEEE International Conference on Communications, ICC 2019, Shanghai, China, May 20-24, 2019*, pages 1–6, 2019.

[6] Xu Zhang, Siddhartha Sen, Daniar Kurniawan, Haryadi Gunawi, and Junchen Jiang. E2E: embracing user heterogeneity to improve quality of experience on the web. In *Proceedings of the ACM Special Interest Group on Data Communication, SIGCOMM 2019, Beijing, China, August 19-23, 2019*, pages 289–302, 2019.

[7] Junchen Jiang, Shijie Sun, Vyas Sekar, and Hui Zhang. Pytheas: Enabling data-driven quality of experience optimization using group-based exploration-exploitation. In *14th USENIX Symposium on Networked Systems Design and Implementation (NSDI 17)*, pages 393–406, Boston, MA, March 2017. USENIX Association.