

Um Modelo Sensível a Adaptação para Previsão de QoE na Transmissão de Vídeos ao Vivo

Thiago Guarnieri¹, Alex Vieira², Jussara Almeida¹

¹Departamento de Ciência da Computação – Universidade Federal de Minas Gerais (UFMG)

²Departamento de Ciência da Computação – Universidade Federal de Juiz de Fora (UFJF)

{thiagoguarnieri, jussara}@dcc.ufmg.br, alex.borges@ufjf.edu.br

Abstract. *Today, video streaming is a well-established application on the internet. Providers are capable of transmitting video globally, in large-scale, although there are still some challenges, such as offering high quality of experience (QoE) for all clients. To face this problem, the content providers try to predict low QoE based on historical system performance to prevent the occurrence of new cases of low quality. In other words, the provider monitors the session performance (the average bitrate and rebuffering rate, for example) and applies methods to correlate this to a metric that expresses QoE, such as user engagement. This method allows the detection of the aspects that contribute to early session abandonment. However, the success of this approach depends on the accuracy of the QoE prediction: past works report an accuracy up to 70% using the pre-mentioned correlation between performance and engagement. In this work, we propose a new method for engagement prediction through the use of the client adaptation regime in replacement of the pre-mentioned performance metrics. We show that this new approach is capable of achieving 81% of prediction accuracy. We also present a case study for the CDN control plan that uses QoE prediction. We found that its use can lead to potential gains in user engagement in comparison to the default provider approach.*

Resumo. *As transmissões de vídeo ao vivo já romperam a fronteira da escala global e agora enfrentam um novo desafio: oferecer alta qualidade de experiência (QoE) de maneira uniforme. Para atacar esse problema, diversas abordagens utilizam informações históricas relacionadas ao desempenho de transmissão para inferir a qualidade de experiência (QoE) de suas sessões. Prever a QoE de maneira acurada permite que provedores de serviço, proativamente, estimem e aloquem recursos a cada um de seus clientes. No entanto, o sucesso dessa alocação de recursos customizada depende de uma previsão acurada de QoE: tipicamente, a previsão é feita por meio de métricas de desempenho como taxa de interrupções e bitrate médio e alcançam uma acurácia abaixo de 70%. Neste trabalho, apresentamos uma nova abordagem, que correlaciona a QoE do cliente com o perfil de adaptação da sua mídia. Nós mostramos que o novo conjunto de métricas proposto aumenta a acurácia de previsão para 81%. Também apresentamos um estudo de caso para alocação de clientes usando o previsor e demonstramos que existe um potencial de duplicação da QoE geral se comparado à alocação padrão empregada pelo provedor.*

1. Introdução

Atualmente, plataformas de transmissão de vídeo já são capazes de oferecer conteúdo ao vivo em escala global. No entanto, existe ainda uma parcela significativa de clientes que sofre interrupções frequentes na reprodução e baixa qualidade. Por exemplo, 23% dos clientes experimentam *bitrate* abaixo de 500 Kbps, um valor insuficiente para o padrão de alta definição das transmissões atuais [Guarnieri et al. 2017].

A impossibilidade de oferecer qualidade de maneira uniforme decorre das limitações naturais de transmissões ao vivo. No cliente, os algoritmos de adaptação reagem às oscilações de banda escolhendo o *bitrate* mais adequado, mas têm eficiência limitada devido ao tamanho reduzido do *buffer* de recepção. Já no servidor, a intensa taxa de chegadas de clientes pode demandar uma redução artificial de qualidade, a fim de evitar a interrupção de serviço para alguns de seus requisitantes.

Além disso, as redes de distribuição de conteúdo (*Content Distribution Networks* ou CDNs) não priorizam por padrão os requisitos de qualidade de experiência (*QoE*) de seus clientes, provocando uma alocação sub-ótima de seus recursos. Por exemplo, para o cliente, uma baixa taxa de interrupções na reprodução é mais importante do que um alto *bitrate* [Dobrian et al. 2011]. Notadamente, grupos de clientes diferentes têm margens de tolerância diferentes a quedas de desempenho. Por exemplo, clientes em dispositivos móveis, com conexão lenta, podem ser mais tolerantes a baixa *QoE* quando comparados a clientes em banda larga residencial. Assim, um servidor que seja capaz de estimar a *QoE* percebida por seus clientes a partir das características de transmissão pode priorizar aquelas que são mais importantes, alocando de forma mais racional os recursos disponíveis.

Existem iniciativas na literatura para quantificar *QoE*, relacionando-a a características da transmissão. No caso em particular de métricas para quantificar *QoE*, temos a popular *Mean Opinion Score (MOS)*. Nela os usuários dão notas a aspectos relacionados à qualidade do vídeo assistido. *MOS* é usada, por exemplo, por [da Costa Filho et al. 2016], que propõe um previsor para estimar o impacto de diversas métricas de qualidade de serviço (*QoS*) no valor de *MOS*. Apesar de confiável, a métrica *MOS* não é escalável, visto que a obtenção de um conjunto de referência representativo requer contribuição de uma grande parcela de usuários, o que nem sempre é trivial. Por consequência, o uso de *MOS* pode levar a uma visão enviesada da qualidade provida.

O *engajamento*, que nada mais é do que o *tempo total de permanência do usuário no sistema*, tem surgido como uma alternativa à tradicional métrica *MOS* como uma medida para traduzir *QoE*. A motivação para seu uso é que clientes que experimentam boa qualidade têm permanência mais alta. Nesse sentido [Balachandran et al. 2013], por exemplo, propõem correlacionar o engajamento com desempenho (i.e. taxa de interrupções e *bitrate*), partindo da premissa de que sessões com alto desempenho têm maior engajamento [Dobrian et al. 2011]. Os autores utilizam essa correlação em um previsor de engajamento que, a partir das métricas de desempenho, consegue estimar o engajamento com uma acurácia de até de 68%.

No presente trabalho, propomos uma nova abordagem para previsão de *QoE* (i.e. engajamento), a partir de informações da adaptação de mídia que um cliente faz. Ao invés das métricas de desempenho, nós exploramos a dinâmica de adaptação do cliente

para efetuar a previsão de engajamento. Mais precisamente, a transmissão de um vídeo ao vivo é, na maior parte do tempo, adaptativa (isto é, muda de *bitrate* de acordo com as condições de rede do cliente). Nós modelamos essa dinâmica de adaptação através de uma *Cadeia de Markov* e a utilizamos como entrada para um algoritmo de previsão, que aprende a relação entre este fluxo e o engajamento.

A vantagem dessa abordagem é que (I) o engajamento como medida de *QoE* proporciona uma interpretação mais direta, que pode ser usada inclusive em setores não técnicos, para planejamento de publicidade ao longo do vídeo; (II) cadeias de Markov proporcionam uma perspectiva *dinâmica* da qualidade de sessão. Ou seja, é possível avaliar como ela evolui ao longo do tempo e mapear quedas de capacidade (e.g. banda) em comportamentos específicos do fluxo de adaptação. Assumimos que isto é válido para qualquer algoritmo adaptativo, pois é esperado que diferentes condições de rede desencadeiem diferentes regimes de trocas, seja qual for a lógica adotada.

Com base nessa premissa, nós modelamos o problema da previsão de engajamento usando técnicas de aprendizado de máquina como classificação, onde o modelo é treinado para prever classes de engajamento, e regressão, em que a previsão é feita para valores contínuos. Através dessas abordagens, nós mostramos que nossa proposta aumenta a precisão de previsão em relação a abordagem tradicional, com um salto de 68% para 81% na acurácia de classificação e um erro médio de 3% na regressão. Também apresentamos um caso de uso do previsor num cenário simulado de alocação de clientes em *CDN*. Neste cenário, o previsor é usado para avaliar diversos fluxos de adaptação para atribuir o que proporciona o maior potencial de engajamento para o cliente. Com o uso desse esquema nós constatamos que existe um potencial aumento no engajamento geral dos clientes.

2. Cenário considerado e conjunto de dados

Nesta seção apresentamos as características do provedor de vídeo ao vivo assim como o conjunto de dados analisados neste estudo.

2.1. Sistema de Transmissão de Vídeo ao Vivo

Para este trabalho, nós utilizamos os registros dos *logs* das requisições HTTP atendidas pelos servidores de mídia ao vivo da Globo.com, coletados durante a Copa do Mundo de Futebol FIFA de 2014. A Globo distribui seu conteúdo a partir de dois centros de distribuição, localizados em São Paulo e Rio de Janeiro, onde tem conexões de parceria (*peering*) em pontos de troca de tráfego com uma série de ISP's. Os centros de distribuição usam *anycast* e os clientes se conectam a um dos centros de forma transparente.

Os servidores da Globo.com utilizam *HTTP Live Streaming* (HLS)¹ como abordagem de transmissão de vídeo. Como mostra a Figura 1, o vídeo é dividido em segmentos de tamanho fixo. Cada segmento é então codificado em diferentes *representações*² (taxas de vídeo ou *bitrates*). Quando um cliente deseja iniciar a reprodução, ele requisita um arquivo de metadados (extensão *m3u8*) que lista as representações disponíveis para a transmissão. Após escolher uma, o cliente requisita outro arquivo de metadados contendo uma lista de segmentos (extensão *ts*) para serem baixados. O processo de requisitar novas listas e seus respectivos segmentos (etapas 2 e

3) é feito periodicamente ao longo de toda a sessão.

Um cliente pode, a qualquer momento, trocar de representação, sendo essa troca gerenciada pelo *algoritmo de adaptação* do cliente. Geralmente o adaptador escolhe uma representação que minimiza a possibilidade de esvaziamento de *buffer*, fenômeno que leva à interrupção da reprodução ou *stall*, visto que sua ocorrência gera um impacto mais negativo na QoE do que a baixa qualidade em si [Seufert et al. 2015, Guarnieri et al. 2017]. Várias propostas de algoritmos de adaptação já foram apresentadas [Sun et al. 2016, Mao et al. 2017], e estudos já demonstraram o benefício da adaptação para aumento de *QoE* [Dobrian et al. 2011, Krishnan and Sitaraman 2013].

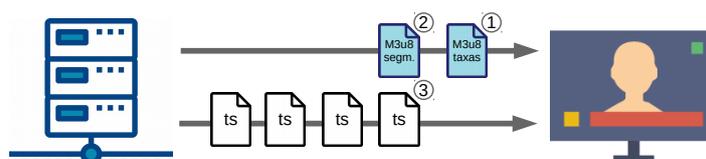


Figura 1. Abordagem adaptativa

Apesar das melhorias proporcionadas pelos algoritmos de adaptação, ainda existe uma parcela de clientes que sofre com baixa *QoE*, sobretudo em conteúdo ao vivo. Isso se deve a características inerentes dessas transmissões, como o tamanho limitado de *buffer*, que reduz o tempo de reação do adaptador, e o ingresso repentino de uma grande quantidade de usuários, que pode sobrecarregar o sistema de transmissão. Nesse sentido, o provedor poderia adotar um papel mais proativo, orientando o fluxo de adaptação do cliente ou escolhendo sua CDN, com base no conhecimento que tem da quantidade de recursos (e.g. banda) disponíveis globalmente. Essa abordagem é particularmente útil para problemas detectados antes da última milha, em servidores ou regiões específicas.

2.2. Conjunto de dados

Nos servidores do provedor estudado, cada requisição HTTP para um segmento de vídeo é registrada em servidores nginx, no formato padrão. Estes registros contêm a data e a hora da requisição, IP do cliente, URL requisitada, *status* HTTP, total de *bytes* enviados e a identificação do navegador (*user agent*). O descritor do navegador nos permite identificar o tipo de dispositivo (e.g., *PC*, *tablet* ou *smartphone*). Identificamos clientes combinando o descritor do agente de navegação e o endereço IP do cliente.

Para capturar o fluxo de adaptação de um cliente, agrupamos suas requisições em sessões. Para isso, observamos que os clientes solicitam, em 98,9% dos casos, segmentos consecutivos de mídia em um intervalo médio de 3 segundos. Dado esse comportamento, definimos de forma conservadora uma sessão como sendo uma sequência de requisições de um mesmo cliente limitadas por um intervalo máximo entre requisições consecutivas de 120 s. Diferentes limiares (de 30 s a 180 s) levaram a resultados qualitativamente similares.

Muitos usuários acessam a Internet através de NAT, o que pode dificultar a identificação de múltiplos clientes advindos de uma mesma rede. Optamos pela exclusão

²<https://tools.ietf.org/html/rfc8216>

²Os *bitrates* disponíveis em kilobits/s. são 264, 464, 750, 1264, 1864, 2564

das sessões com possível influência de NAT, ignorando sessões em que foi observado recebimento duplicado de mais de 5% do total de segmentos transmitidos ou uma taxa de envio de segmentos superior a 1 pacote a cada 2 segundos. Todavia, após a identificação de clientes por IP e *user-agent*, menos de 5% das sessões apresentaram um desses tipos de comportamento.

No geral, observamos até 1,1 milhão de clientes únicos em um único dia (com 4 partidas) e até 470 mil sessões simultâneas durante uma única partida. Ao analisar os descritores de agentes de navegação, identificamos que 81% das sessões são iniciadas a partir de *PCs*, 12% de *smartphones* e 6% de *tablets*. Algumas sessões de tipos de dispositivos não identificados e robôs completam o conjunto.

3. Correlacionando engajamento e aspectos de sessão

O provedor pode tirar proveito das características de transmissão de sessões passadas para avaliar se o vídeo que está sendo transmitido atende aos requisitos de qualidade de seus usuários. Ter essa informação permite ao servidor uma alocação mais inteligente de seus recursos. Por essa razão, o provedor precisa saber especificamente quais características são relevantes para melhor explicar *QoE*.

Muitos trabalhos têm surgido na última década com o objetivo de explorar a influência das características de transmissão na *QoE*. Por exemplo, [Seufert et al. 2015] demonstram que a ocorrência de mais de um *stall* por minuto já provoca insatisfação por parte dos usuários. Como base nessa informação, um servidor pode oferecer uma taxa de vídeo menor, mas que reduza os *stalls*. Note que, diferentes contextos apresentam diferentes definições de qualidade: um cliente de dispositivo móvel, numa conexão 3G, provavelmente tem expectativas de desempenho inferiores se comparado a um usuário *desktop* com conexão de banda larga. Nesse sentido, percebemos que as reações tomadas para aumentar a satisfação dos clientes vão muito além de simplesmente oferecer um *bitrate* elevado, devendo-se levar em conta também as complexas relações entre as características de uma sessão.

Apesar de trabalhos de caracterização terem avançado no entendimento das correlações entre aspectos da transmissão e *QoE*, ainda não existe um consenso sobre qual conjunto de métricas pode melhor explicá-la, nem sobre qual abordagem utilizar para traçar tais correlações. Sendo assim, o objetivo deste trabalho é apresentar um novo conjunto de métricas capaz de melhorar a correlação entre as características de sessão e *QoE*. Na Seção 3.1 explicamos detalhadamente os trabalhos atuais e na Seção 3.2 introduzimos nossa proposta para o conjunto de métricas que visa alcançar o objetivo proposto.

3.1. Correlacionando engajamento e desempenho

Como mencionado anteriormente, diversos trabalhos têm se dedicado a quantificar e modelar a correlação existente entre desempenho e *QoE*. Na presente década, tem havido uma mudança no conjunto de métricas que medem desempenho de transmissão: medidas voltadas para o conteúdo como *Peak Signal-to-Noise Ratio (PSNR)* têm sido substituídas por outras, centradas na percepção de qualidade pelo usuário, como taxa de *stalls*, e *bitrate* médio. A *QoE*, por sua vez, também tem sido representada de maneiras diferentes como, por exemplo, pelo *engajamento*.

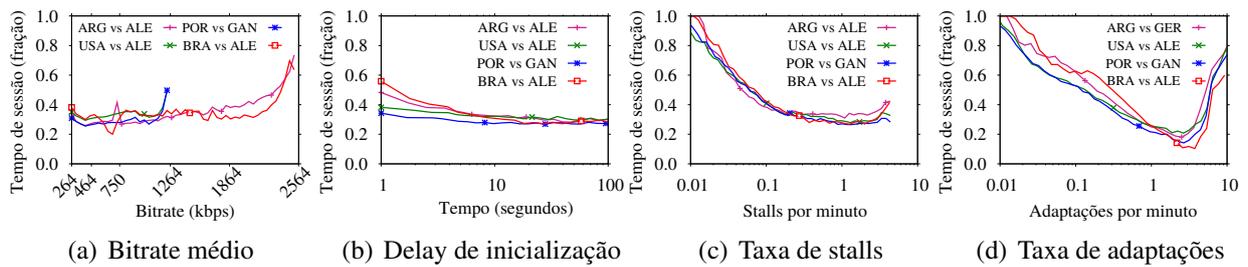


Figura 2. Impacto de desempenho no engajamento [Guarnieri et al. 2017].

A partir dessas novas métricas, diversos trabalhos têm investigado as correlações existentes entre as métricas de desempenho e o engajamento [Dobrian et al. 2011, Krishnan and Sitaraman 2013]. Em nosso trabalho anterior [Guarnieri et al. 2017] nós exploramos a correlação entre a duração de sessão (i.e. engajamento), normalizada pelo tempo restante de exibição, e as métricas mais comuns de desempenho (*delay* de inicialização, *bitrate* médio, taxa de *stalls* e taxa de adaptações) para 4 jogos do mesmo conjunto de dados apresentado na Seção 2.2. A Figura 2, extraída do referido trabalho, mostra que sessões com uma taxa de 1 *stall* por minuto já têm um engajamento menos de 35%. Além disso, sessões com *bitrates* mais altos possuem maior engajamento, permanecendo acima de 60% do tempo restante de transmissão.

Apesar da Figura 2 explicitar que o desempenho explica em boa parte o engajamento, há características conhecidas como *fatores de confusão* podem interferir na eficiência de esquemas de alocação de recursos baseados em *QoE*. Os fatores de confusão são características pertencentes ao contexto de transmissão, mas que não podem ser medidas facilmente. Como exemplo podemos citar o *interesse do usuário no conteúdo*, o *compartilhamento de banda do usuário* ou mesmo o *tipo de dispositivo do cliente*. Assim, é necessário explorar outras formas de representação de uma sessão para poder explicar o engajamento com maior precisão.

3.2. Correlacionando engajamento e fluxo de adaptação

Em abordagens adaptativas de transmissão, o algoritmo de adaptação do cliente reage de acordo com suas condições de rede, podendo dar pistas sobre o nível de qualidade de serviço disponível. Por essa razão, nós passamos a observar o fluxo de adaptação do cliente e constatamos que ele provê informações relevantes, difíceis de serem extraídas das métricas tradicionais de desempenho. Por exemplo, quando o cliente requisita trocas de *bitrate* muito frequentemente, isso pode indicar instabilidade na sua largura de banda. Além disso, quedas muito amplas de qualidade (e.g., de 2564 kbps para 264 kbps) podem indicar a ocorrência de *stalls*. Assim optamos por desenvolver um *modelo de adaptação* que visa explicar a dinâmica de trocas de *bitrate* do cliente e extrair informação que possa ser correlacionada com seu engajamento.

Para modelar o fluxo de adaptações empregado pelos clientes, nós utilizamos *Cadeias de Markov* [Norris 1998]. Neste formalismo, um *bitrate* corresponde a um *estado*, e as *transições* representam as probabilidades de se passar de um estado (*bitrate*) a outro. Um cliente também pode permanecer no mesmo *bitrate* sendo isso representado por uma transição cíclica. Para cada estado, as probabilidades de suas transições de saída devem somar 1. A partir de cada estado, um cliente pode transicionar para qualquer outro,

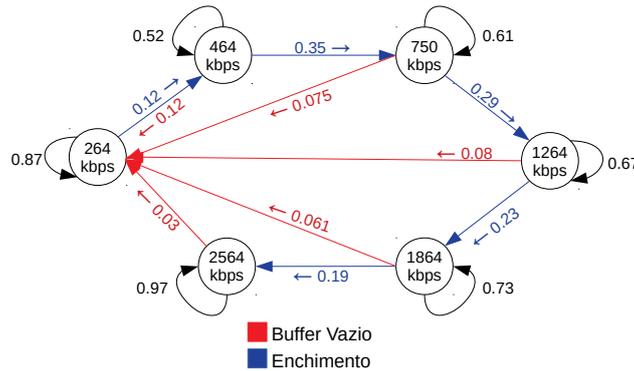


Figura 3. Fluxo de adaptação médio do sistemas

desde que a probabilidade para a transição desejada seja maior que zero. A Figura 3 apresenta o comportamento médio das sessões analisadas em nosso trabalho². Podemos observar que o caminho padrão consiste em chegar na resolução máxima suportada pela banda do cliente (caminho azul). Também há uma alta probabilidade de permanecer nos extremos das representações (264 e 2564 kbps). Além disso, não foi observada uma taxa significativa de trocas entre bitrates não adjacentes. Por fim, observa-se uma significativa probabilidade de ocorrerem transições de retorno ao menor *bitrate* (caminho vermelho). Tais transições são relacionadas a cenários de *stalls*.

4. Processo de previsão de engajamento

A previsão de engajamento é feita explorando um conjunto de atributos monitorados e coletados durante a sessão do cliente, sejam eles atributos de desempenho (como em técnicas da literatura [Balachandran et al. 2013, Ahmed et al. 2017]) ou mesmo atributos extraídos do fluxo de adaptação (nossa proposta). Em ambos os casos, um *algoritmo de previsão* é empregado para aprender como combinar as contribuições de cada atributo, de cada sessão, para o engajamento em um modelo \mathcal{M} . A aprendizagem é feita de maneira supervisionada, isto é, parte das sessões é utilizada como *conjunto de treinamento*, afim de inferir as correlações entre as *variáveis independentes* (atributos de entrada) e a *variável dependente* (engajamento) e assim aprender o modelo de previsão \mathcal{M} . A eficácia do modelo aprendido é posteriormente avaliada no restante das sessões, considerado *conjunto de teste*, quando é possível verificar a capacidade do modelo de generalizar para outros conjuntos de sessões. As seções a seguir apresentam a nossa formalização da tarefa de previsão de engajamento, o algoritmo de previsão utilizado assim como os atributos explorados por esse algoritmo.

4.1. Tarefas de previsão de engajamento

A previsão do engajamento pode ser formalizada de duas maneiras, cada uma com soluções diferentes. Como uma tarefa de *classificação*, o objetivo é obter uma função f que mapeie um conjunto de variáveis independentes X em uma variável dependente Y *discreta*. Em outras palavras, a variável dependente (engajamento) é discretizada em classes e objetiva-se prever a qual classe uma sessão correntemente pertence. Neste caso, a eficácia do modelo esta relacionada ao *número de vezes em que o predictor*

²Nós ocultamos as transições com probabilidade menor que 2% para facilitar visualização.

prevê corretamente a classe de engajamento de uma sessão. As classes são definidas considerando o engajamento normalizado pelo tempo total de transmissão. Por exemplo, para uma classificação considerando 4 classes, o engajamento de uma sessão pode estar nos intervalos 0-25%, 26-50%, 51-75% e 76-100%. Ou seja, uma sessão com permanência de 30 minutos em uma transmissão de 2 horas tem engajamento de 25%.

Como uma tarefa de *regressão*, o objetivo é obter uma função f que mapeie um conjunto de variáveis independentes X em uma variável dependente Y *contínua*. Ou seja, a previsão é feita sobre os dados contínuos do engajamento normalizados pelo tempo total de transmissão. Em outras palavras, deseja-se prever o tempo que o cliente permanecerá no sistema na sua sessão atual como uma fração do tempo total de transmissão. Neste caso, um bom modelo é aquele que consegue prever o engajamento com o *menor erro possível entre valores previsto e real*.

4.2. Algoritmo de previsão

Nós investigamos vários algoritmos considerados estado-da-arte tanto em classificação quanto em regressão. Em experimentos preliminares nós observamos que os melhores resultados foram obtidos com árvores de decisão (*Classification and Regression Trees* (CART) [Breiman et al. 1984]). Neste algoritmo, diversas árvores de decisão são geradas e seleciona-se a que proporciona maior ganho de informação, estimado pela entropia. Além disso, para aumentar a precisão do modelo, empregamos uma técnica de aprendizagem em conjunto chamada *Random Subspace* [Ho 1998]. Nela, diversas instâncias de previsão são criadas, cada qual com uma amostra aleatória das métricas. As previsões das instâncias são posteriormente combinadas através de, por exemplo, média ou votação. Essa técnica permite que haja uma redução do sobre-ajuste ou *overfitting*, fenômeno no qual os resultados alcançados na etapa de treinamento, apesar de muito acurados, não generalizam adequadamente para conjuntos desconhecidos de teste.

4.3. Conjunto de atributos

Esta seção apresenta as variáveis independentes utilizadas nas sessões para efetuar a previsão de engajamento. Nós consideramos tanto uma abordagem tradicional (Seção 4.3.1) que explora apenas métricas de desempenho, quanto uma nova estratégia que explora atributos extraídos do fluxo de adaptação (Seção 4.3.2). A abordagem tradicional é amplamente empregada na literatura e é utilizada aqui para fins de comparação com o modelo adaptativo, proposto neste trabalho.

4.3.1. Abordagem tradicional

Nesta abordagem, nós tomamos como base as métricas de desempenho mais comuns na literatura [Seufert et al. 2015]. São elas:

- **Atraso de inicialização:** Tempo necessário para preencher o *buffer* de reprodução do cliente no início da reprodução;
- **Bitrate médio:** cada segmento possui em sua URL o valor do *bitrate*. Assim o *bitrate* médio é computado pela média dos valores reportados;
- **Taxa de interrupções (stalls):** Quantidade de *stalls* dividido pela duração da sessão;

- **Taxa de adaptações:** a quantidade de trocas de *bitrate* dividida pela duração da sessão. Nós separamos as adaptações de aumento de qualidade e de queda, visto que seu impacto é diferente para o engajamento;
- **Dispositivo:** as demandas de qualidade dependem das capacidades do dispositivo do cliente. Assim, essa informação é útil para definir QoE de forma mais precisa.

4.3.2. Abordagem baseada no fluxo de adaptação

Para a abordagem que explora o fluxo de adaptação, as seguintes métricas foram consideradas como atributos:

- **Cadeia de adaptação:** como apresentado na Seção 3.2, nós modelamos o fluxo de adaptação de cada sessão dos clientes através de uma Cadeia de Markov, que descreve as probabilidades de transição para cada *bitrate* disponível na transmissão. Mais especificamente, a cadeia é representada internamente por uma matriz $M_{n \times n}$ onde n é igual ao número de *bitrates* disponíveis. Qualquer elemento a_{ij} armazena a probabilidade de transição do *bitrate* i para j .
- **Permanência relativa em *bitrate*:** nós também armazenamos o tempo de permanência em cada qualidade disponível, ou seja, o número de segmentos de cada *bitrate* dividido pelo total de segmentos recebidos;
- **Taxa de adaptações:** assim como nas métricas de desempenho, nós armazenamos a quantidade de trocas de *bitrate* dividido pelo tempo de sessão do cliente;
- **Dispositivo:** de forma análoga a abordagem anterior, os dados de dispositivo também foram armazenados.

5. Metodologia de Avaliação

Nesta seção, descrevemos a metodologia de divisão dos dados entre conjuntos de treinamento e de teste bem como as métricas de avaliação dos modelos de regressão e classificação. Ressalta-se também que para a tarefa de classificação, nós consideramos cenários diversos com número de classes variando entre 2 e 10 classes de engajamento normalizado pelo tempo total de transmissão.

5.1. Conjuntos de treinamento e de teste

Tanto para classificação quanto para regressão, a divisão das sessões entre conjuntos de treinamento e teste é etapa essencial para eficácia do modelo. Uma metodologia amplamente adotada é conhecida como *validação cruzada*, onde os dados são divididos em n fatias (geralmente 5 ou 10); o treinamento é feito em $n - 1$ fatias e o teste na fatia restante. Este processo é repetido n vezes, de forma a utilizar cada fatia uma vez. A vantagem da validação cruzada é que ela é capaz de reduzir a interferência de dados enviesados na análise de precisão, principalmente em conjunto de dados reduzido. Para efetuar o processo de verificação de precisão nós utilizamos a validação cruzada com 5 fatias.

5.2. Métricas de Avaliação

Para a tarefa de classificação, o modelo desenvolvido é avaliado segundo as seguintes métricas, comumente usadas neste contexto:

- **Acurácia:** é a fração de previsões certas em relação a todas as previsões.
- **Área sob a curva ROC:** apresenta uma curva com base na taxa de verdadeiros e falsos positivos. O valor da área abaixo da curva varia entre 0 e 1 e indica o quão bem o modelo faz distinção entre classes. Além disso, valores próximos de 0.5 indicam que o modelo não é melhor que uma previsão puramente aleatória;
- **F1 score:** é a média harmônica de precisão (P) e revocação (R), dado por $\frac{2*P*R}{P+R}$. A precisão P é a fração de acertos dentre todas as previsões. Já a revocação R é a fração de acertos dentre todos os acertos possíveis. Os valores de F1 variam entre 0 e 1.

Já na regressão, os valores preditos são contínuos. Desta forma, a precisão é medida pelo erro entre o valor previsto e o valor real, que pode ser estimado por:

- **Erro médio absoluto:** média das diferenças absolutas entre valor previsto e valor real. É dada pela fórmula $1/n \sum_{i=1}^n |y_i - \hat{y}_i|$, onde y_i é o valor previsto e \hat{y}_i o valor real.
- **Erro médio quadrático:** média do quadrado das diferenças entre valor previsto e valor real. É dada pela fórmula $1/n \sum_{i=1}^n (y_i - \hat{y}_i)^2$, onde y_i é o valor previsto e \hat{y}_i o valor real.

6. Análise de precisão dos previsores de engajamento

Nesta seção apresentamos os resultados de precisão do predictor tanto para a abordagem padrão, usando métricas de desempenho, quanto para a abordagem adaptativa, com fluxo de adaptação. A tabela 1 apresenta a precisão para classificação de engajamento. Como é possível observar, a precisão da abordagem adaptativa é superior em todos os cenários. Para 5 classes, observamos um ganho relativo de acurácia de 14.5% da abordagem adaptativa em relação à tradicional. Além disso a perda de acurácia de um problema de 2 classes para um de 10 classes é maior na abordagem tradicional, indo de 86,129% para 67,074%.

Já a medida de área sob a curva ROC, dá um indicativo da frequência em que um predictor erra a classe de engajamento de uma sessão. Tanto a abordagem tradicional quanto a de fluxo de adaptação apresentam bons resultados. No entanto o valor inferior da abordagem tradicional pode indicar que existe uma maior taxa de falsos positivos em sua classificação.

O *F1 score* também tem bom desempenho na abordagem adaptativa, ficando acima de 0.82 para todas as variações de classe apresentadas. Ter um valor alto nessa medida é importante, sobretudo porque nossa base de dados é *desbalanceada*, ou seja, possui muito mais sessões curtas do que longas. Em bases dessa natureza, um baixo valor de *F1 score* indica que o predictor acerta simplesmente porque está enviesado pela classe mais representativa. Por exemplo, para 10 classes, a abordagem tradicional tem um *F1 score* de 0,687. Isso ocorre porque 72% dos dados estão na classe de engajamento {0-10%}. Assim, o poder do predictor nesse caso não é significativamente superior a uma atribuição aleatória de classes. Já na abordagem com fluxo de adaptação, é possível observar que a previsão tem boa acurácia também para as classes menos representativas, compostas de sessões com maior engajamento.

Já a tabela 2 mostra o erro para a tarefa de previsão por regressão. Note que o valor do engajamento é relativo, portanto o erro é apresentado em forma de porcentagem de

classes	Tradicional			Fluxo de adaptação		
	Acurácia	AUC ROC	F1 score	Acurácia	AUC ROC	F1 score
2	86,129	0,896	0,894	91,630	0,972	0,938
3	83,139	0,894	0,856	90,410	0,969	0,921
4	79,647	0,889	0,819	88,689	0,967	0,901
5	76,378	0,885	0,786	87,325	0,966	0,886
6	73,976	0,880	0,760	86,342	0,966	0,874
7	72,109	0,885	0,741	85,220	0,965	0,862
8	69,768	0,875	0,716	83,739	0,962	0,848
9	67,986	0,876	0,700	82,783	0,961	0,837
10	67,074	0,874	0,687	81,444	0,957	0,823

Tabela 1. Avaliação de precisão (classificação)

Tradicional		Fluxo de adaptação	
Mean abs. error	Mean sqr. error	Mean abs. error	Mean sqr. error
0,0701	0,116	0,0291	0,059

Tabela 2. Avaliação de precisão (regressão)

engajamento. Assim, é possível observar que o erro absoluto para a abordagem adaptativa é menor que 3% representando uma redução de 2.4 vezes em relação ao erro da abordagem tradicional. O erro quadrático por sua vez fica abaixo dos 6%, indicando uma significativa proximidade dos valores preditos e reais. Tomando um exemplo absoluto, consideremos uma sessão que iniciou no início da transmissão e teve engajamento de 90%. Neste caso, sua duração foi de 162 minutos e o erro de previsão será de aproximadamente 5 minutos para a abordagem adaptativa e 12 minutos para a tradicional.

7. Estudo de caso: seleção automática de fluxo de adaptação

Nesta seção, nós apresentamos um estudo de caso para o previsor de engajamento. Nosso objetivo é demonstrar a viabilidade de um uso prático do modelo. Assim, nós propomos um *seletor de fluxo de adaptação* para clientes ingressantes em um sistema de distribuição de vídeo. Nesse esquema, o seletor define uma cadeia de adaptação para um novo cliente C com base em uma medição de qualidade geral extraída de clientes similares a C . Este estudo de caso é baseado em propostas apresentadas pela primeira vez em [Balachandran et al. 2013, Liu et al. 2012]

Mais especificamente, quando C ingressa no sistema, o seletor prevê o engajamento, via regressão, de uma série de cadeias de adaptação hipotéticas e sugere para o cliente aquela de alcança o maior engajamento potencial. Para escolher as cadeias de adaptação hipotéticas de C , o seletor extrai uma qualidade média Q_{S_C} (taxa de permanência nos bitrates + média de adaptações) observada para um conjunto S_C de clientes como mesmo *dispositivo, provedor e cidade* de C . Em seguida, uma amostra de segmentos é gerada e permutada para gerar diversas versões de cadeias, onde todas produzem qualidade compatível com Q_{S_C} . Para garantir que Q_{S_C} representa a qualidade atual da região, nós dividimos a transmissão em janelas fixas de 20 minutos (com base na duração média de sessão de 16 minutos) e incluímos em S_C apenas as sessões finalizadas na janela corrente ou imediatamente anterior. Por exemplo, para um cliente C ingressante na janela 5, a extração de Q_{S_C} se dará para um S_C com sessões das janelas 5 e 4. Por fim, no caso de não haverem sessões com mesmo *dispositivo, provedor e cidade*, nós escolhemos sessões com menor granularidade de semelhança, na ordem $\{cidade, provedor, dispositivo\} \rightarrow \{cidade, provedor\} \rightarrow \{cidade\} \rightarrow original$. O

funcionamento está resumido na Figura 4.

O resultado esperado da aplicação deste modelo é que, mesmo em situações onde a banda disponível é limitada, o seletor vai sugerir um regime de adaptação que maximiza o engajamento para a quantidade de recursos disponíveis. Nós acreditamos que o seletor pode funcionar de maneira complementar ao algoritmo de adaptação do cliente visto que, enquanto este adapta o fluxo em resposta à variações na última milha, aquele oferece um *feedback* para o cliente sobre as condições no núcleo da rede, podendo ser usado para, por exemplo, calibrar o mecanismo de adaptação.

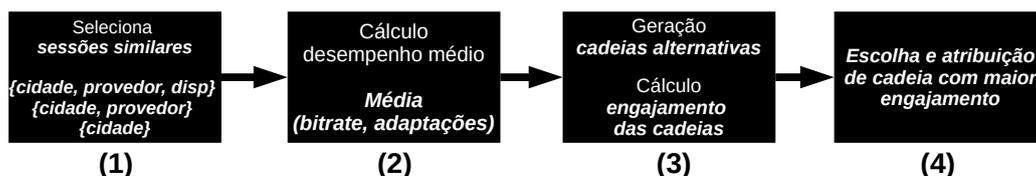


Figura 4. Seletor de fluxo de adaptação

7.1. Resultados da aplicação do seletor de fluxos

A Figura 5 mostra o engajamento quando os clientes seguem o fluxo de adaptação sugerido pelo seletor, bem como o engajamento real registrado nos *logs*. Podemos observar que, em todas as fatias de todos os jogos, existe um potencial de aumento de engajamento. Na partida 1 por exemplo (Figura 5(a)), o aumento de engajamento é de 2.3 vezes em média. As outras duas partidas seguem essa tendência, com um ganho médio de 2.9 e 2.4 vezes, respectivamente (Figuras 5(b) e 5(c)). Além disso, com o uso do seletor, o engajamento nunca foi menor do que 5%, evidenciando que este esquema pode ser útil na redução da incidência de abandono precoce de clientes.

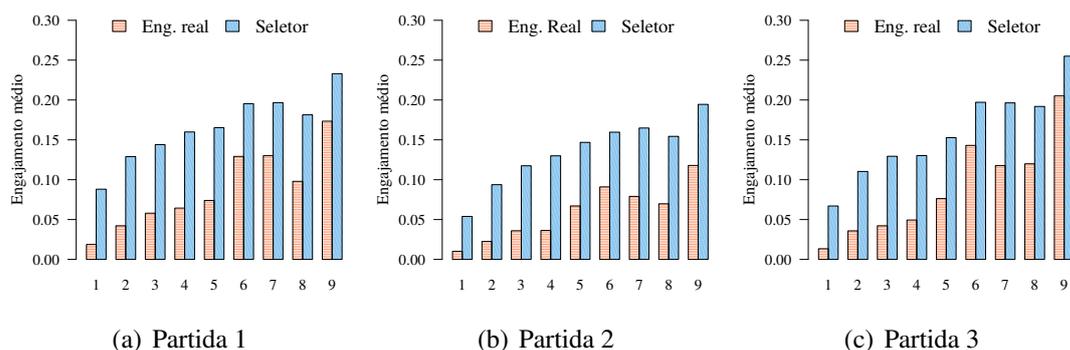


Figura 5. Engajamento real versus seletor de fluxos

8. Trabalhos relacionados

O uso do engajamento como medida de *QoE* tem ganhado atenção nos últimos anos. [Dobrian et al. 2011] demonstram que existe uma forte correlação de engajamento e métricas de desempenho como taxa de *stalls* e *delay* de inicialização, tanto para vídeos ao vivo quanto sob demanda. Já [Krishnan and Sitaraman 2013] avançam propondo uma abordagem que prova, além da correlação, a causalidade entre as métricas de desempenho e o engajamento. Os autores mostram que uma *delay* de inicialização de mais de 2

segundos já influencia no abandono de clientes. Além, disso, cada segundo adicional provoca um aumento de 5.8% na taxa de abandono de clientes.

Os resultados das caracterizações de engajamento citadas tem sido corroborados por uma infinidade de trabalhos posteriores. Em nosso recente trabalho [Guarnieri et al. 2017], nós mostramos que 1 interrupção por minuto já provoca a redução de engajamento para menos de 40% do tempo restante de transmissão em clientes *desktop* e 20% em clientes móveis. Existem também trabalhos, como [Seufert et al. 2015], que se dedicam a analisar exaustivamente todas as métricas de desempenho e suas relações com QoE, evidenciando que, por exemplo, não somente a taxa de interrupções diminui engajamento, como também sua duração e intervalo entre ocorrências sucessivas.

Por outro lado, ainda existem poucas soluções práticas que usam a correlação entre desempenho e engajamento. [Balachandran et al. 2013], por exemplo desenvolvem um previsor de engajamento a partir de desempenho e alcançam uma acurácia de classificação de 68%. Os autores também demonstram uma aplicação prática do previsor para uma alocação de clientes em CDN. [Ahmed et al. 2017] utiliza análise de componentes principais (PCA) para detecção de anomalias na transmissão que possam interferir no engajamento. Outros estudos como [Jiang et al. 2017] focam na melhoria de métricas de desempenho diretamente, sem levar em consideração suas complexas relações com o engajamento.

Com base nos trabalhos aqui descritos e em nossa revisão sistemática, nós acreditamos que o presente estudo é uma das primeiras iniciativas no sentido de correlacionar *QoE* com a cadeia de adaptação de uma sessão. Nós demonstramos que essa abordagem explica mais acuradamente o engajamento, podendo servir de base para soluções que visem a melhoria de qualidade de experiência.

9. Conclusão

Neste trabalho nós apresentamos uma nova abordagem para descrever engajamento a partir de características das sessões de clientes. Ao invés de utilizarmos métricas de desempenho tradicionais como taxa de *stalls* e *delay* de inicialização, nós modelamos o fluxo de adaptação das sessões dos cliente através de *cadeias de Markov* e, em conjunto com técnicas de aprendizado de máquina, demonstramos que é uma alternativa mais precisa para previsão de engajamento.

Nós também propusemos uma aplicação prática do previsor, capaz de avaliar o melhor fluxo de adaptação para as condições a que um cliente está submetido. Através desse esquema, nós constatamos que existe um potencial melhoramento de engajamento, no caso do cliente adotar o fluxo sugerido pelo servidor. A confiança de que o fluxo escolhido pode ser o melhor disponível para o cenário do cliente é embasada na alta acurácia do previsor, que possui um erro absoluto médio de menos de 3%. Como trabalho futuro, nós planejamos implementar o seletor de fluxos numa transmissão real, afim confirmar as potencialidades observadas em nossos testes.

10. Agradecimentos

Os autores agradecem a CAPES, FAPEMIG, CNPq e FAPESP pelo aporte financeiro e ao time da Globo.com pelos *logs* de transmissão da Copa do Mundo FIFA de 2014.

Referências

- Ahmed, A., Shafiq, Z., Bedi, H., and Khakpour, A. (2017). Suffering from buffering? detecting QoE impairments in live video streams. In *2017 IEEE 25th International Conference on Network Protocols (ICNP)*, pages 1–10.
- Balachandran, A., Sekar, V., Akella, A., Seshan, S., Stoica, I., and Zhang, H. (2013). Developing a predictive model of quality of experience for internet video. *SIGCOMM Comput. Commun. Rev.*, 43(4):339–350.
- Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984). *Classification and Regression Trees*. Wadsworth and Brooks, Monterey, CA.
- da Costa Filho, R. I. T., Lautenschlager, W., Kagami, N., Roesler, V., and Gaspar, L. P. (2016). Network fortune cookie: Using network measurements to predict video streaming performance and qoe. In *2016 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6.
- Dobrian, F., Sekar, V., Awan, A., Stoica, I., Joseph, D., Ganjam, A., Zhan, J., and Zhang, H. (2011). Understanding the impact of video quality on user engagement. In *Proceedings of the ACM SIGCOMM 2011 Conference*, SIGCOMM '11.
- Guarnieri, T., Ítalo Cunha, Almeida, J., Drago, I., and Vieira, A. B. (2017). Characterizing QoE in large-scale live streaming. In *Proc. of the IEEE GLOBECOM*.
- Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):832–844.
- Jiang, J., Sun, S., Sekar, V., and Zhang, H. (2017). 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. USENIX Association.
- Krishnan, S. S. and Sitaraman, R. K. (2013). Video stream quality impacts viewer behavior: Inferring causality using quasi-experimental designs. *IEEE/ACM Transactions on Networking*, 21(6):2001–2014.
- Liu, X., Dobrian, F., Milner, H., Jiang, J., Sekar, V., Stoica, I., and Zhang, H. (2012). A case for a coordinated internet video control plane. *ACM SIGCOMM Computer Communication Review*, 42.
- Mao, H., Netravali, R., and Alizadeh, M. (2017). Neural adaptive video streaming with pensieve. In *Proceedings of the Conference of the ACM Special Interest Group on Data Communication*, SIGCOMM '17, pages 197–210.
- Norris, J. R. (1998). *Markov chains*. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press.
- Seufert, M., Egger, S., Slanina, M., Zinner, T., Hoßfeld, T., and Tran-Gia, P. (2015). A survey on quality of experience of http adaptive streaming. *IEEE Communications Surveys Tutorials*, 17(1):469–492.
- Sun, Y., Yin, X., Jiang, J., Sekar, V., Lin, F., Wang, N., Liu, T., and Sinopoli, B. (2016). Cs2p: Improving video bitrate selection and adaptation with data-driven throughput prediction. In *Proceedings of the 2016 ACM SIGCOMM Conference*, SIGCOMM '16, pages 272–285. ACM.