Recomendação de Conteúdo e QoE: Quantificando o Papel da QoS nas Preferências por Vídeos

Felipe Assis[1], Mateus S. Nogueira[1], Daniel S. Menasché[1], João Ismael Pinheiro[1], Pavlos Sermpezis[2], Savvas Kastanakis[2], Thrasyvoulos Spyropoulos[3], Carla Delgado[1]

[1]Univ. Federal do Rio de Janeiro (UFRJ), [2] FORTH, Greece, [3] EURECOM, France

Abstract. *Recommender systems are increasingly present in Internet users' routine. Therefore, platforms like Youtube and Netflix seek to improve their recommendation systems, to provide a better experience for their users. However, the users' experience depends on a multitude of factors. In particular, caching systems have an important influence in the quality of experience (QoE), since they impact quality of service (QoS) metrics (such as the delay and the throughput) experienced by users. Our goal is to study the viability of a QoS-aware and QoE-friendly content recommendation system. To this aim, we conduct an experiment with real users, having different profiles. Each user is requested to evaluate different videos, which vary in their contents and in the corresponding QoS. Given our findings on QoS-QoR tradeoffs, we investigate their impact on the design of a recommendation system. A decision tree classifier reached accuracy of 77% using cross validation, which allows us to further understand the user's decision making process.*

Resumo. *Sistemas de recomendação estão cada vez mais presentes na rotina dos usuários da Internet. Portanto, plataformas como Youtube e Netflix procuram melhorar suas recomendações para fornecer uma melhor experiência aos seus usuários. No entanto, a experiência dos usuários depende de diversos fatores. Em particular, sistemas de cache têm uma grande influência na qualidade de experiência (QoE), visto que eles impactam métricas de qualidade de serviço (QoS), como atraso e vazão, experenciadas pelos usuários. Nosso objetivo é estudar a viabilidade de um sistema de recomendação sensível a QoS que maximize a QoE. Para tal, nós conduzimos experimentos com usuários reais com perfis distintos. Cada usuário é solicitado a avaliar diferentes vídeos, que variam em termos de conteúdo e de QoS. Dadas nossas descobertas quanto à relação entre QoS e QoR (quality of recommendation), investigamos seus impactos no modelo de sistema de recomendação. Um classificador baseado em árvore de decisão alcançou acurácia de 77% usando validação cruzada, e nos permite entender melhor o processo de tomada de decisão do usuário.*

1. Introdução

Motivação Sistemas de recomendação determinam uma parcela expressiva da demanda por conteúdos na Internet[Gomez-Uribe and Hunt 2016]. Assim, pode-se afirmar que esses sistemas são capazes de moldar as demandas dos usuários. Neste estudo, buscamos entender como utilizar essa capacidade para beneficiar os usuários do sistema, através de um sistema de recomendação consciente dos conteúdos já presentes em redes de *cache*, bem como os administradores do sistema, por meio de um aumento do desempenho das

caches, melhorando a qualidade de reprodução dos vídeos na máquina do usuário além de reduzir os custos de serviço. Para tal, analisamos o impacto que a qualidade de serviço (QoS) possui sobre o interesse dos usuários por conteúdos. Deste modo, redes de cache e QoS estão estreitamente relacionadas, no sentido de que uma influencia a outra. A análise do impacto da QoS e de recomendações de conteúdo apresentada neste artigo busca esclarecer os fatores que influenciam a qualidade de experiência (QoE) percebida por um usuário final.

Os experimentos apresentados neste trabalho visam a construção de um sistema de recomendação de conteúdos que levem em conta a QoS experimentada pelos usuários. Assim, espera-se que plataformas como Youtube e Netflix possam oferecer uma melhor QoE para seus usuários. Isto é, considerando também a qualidade com a qual determinado vídeo pode ser servido - além das preferências de categorias de conteúdo - no momento de fazer uma recomendação, espera-se que o sistema aumente a satisfação de seus usuários.

A Figura 1 ilustra a relação entre QoS, QoR e QoE considerada neste trabalho. Visamos compreender como a QoS (e.g., interrupções no vídeo) e a QoR (e.g., natureza do conteúdo recomendado, tipicamente medida em número de estrelas numa escala de 0 a 5) afetam a QoE (i.e., nível de satisfação geral do usuário).

Desafios Poucos estudos consideraram redes de cache e sistemas de recomendação conjuntamente, portanto a relação entre estes dois ainda não é bem compreendida. Dado que a experiência do usuário depende de forma integrada de uma série de fatores, compreender quais fatores (i.e, QoS ou natureza do conteúdo) mais influenciam os usuários, em cada situação, é uma tarefa não trivial visto que não é algo que possa ser simulado. Para entender como usuários se comportam, assim como quais são os fatores que influenciam suas decisões, é necessário coletar dados de usuários reais.

Objetivos Neste artigo, nosso objetivo é compreender o impacto dos interesses dos usuários por conteúdos, e a influência que distúrbios na qualidade de serviço (QoS) podem ter nestes interesses. Tal entendimento visa aprimorar o funcionamento de sistemas de recomendação que trabalhem, por exemplo, sobre redes de cache.

Pretendemos responder as seguintes perguntas:

1) Quais fatores afetam a sensibilidade do usuário com relação a natureza do conteúdo e com relação a QoS? É possível quantificá-los?

2) Considere a escolha entre dois conteúdos para se recomendar ao usuário. Um dos conteúdos está disponível em cache (alta QoS) mas não é tão interessante ao usuário, e o outro não está disponível em cache (baixa QoS) mas é fortemente relacionado aos interesses dos usuários. Qual dos dois conteúdos recomendar?

Métodos Para entender o impacto que a qualidade de serviço (QoS) possa ter sobre o interesse de usuários pela natureza dos conteúdos, construímos um experimento que consiste em uma série de vídeos a serem assistidos pela *Web* de forma remota. Os vídeos foram selecionados em grupos de dois, sendo eles relacionados entre si de acordo com a categoria do conteúdo deles. A cada par de vídeos, um deles possuía distúrbios de qualidade e sua ordem de exibição varia a cada grupo. Os distúrbios foram induzidos de forma metódica, considerando estudos anteriores sobre o impacto da qualidade de serviço (QoS) na qualidade de experiência (QoE) [Nam et al. 2016]. Os distúrbios uti-

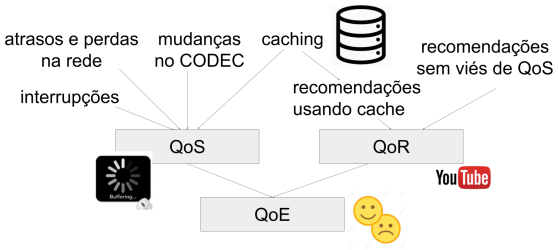


Figura 1. Ilustrando relação entre qualidade de serviço (QoS), qualidade de recomendação (QoR) e qualidade de experiência (QoE).

lizados foram: rebufferings e mudanças de bitrate. *Rebufferings* foram utilizados com mais frequência, visto que possuem um impacto muito maior na qualidade de experiência (QoE) [Nam et al. 2016].[1]

Contribuições Ao responder as perguntas propostas neste trabalho, fazemos as seguintes contribuições:

Experimento com usuários reais Realizamos experimentos com usuários reais, e levantamos preferências dos mesmos sobre diferentes pares de vídeos. Para cada par de vídeos apresentado, cada usuário informa aquele que lhe proporcionou maior satisfação global, levando em conta o conteúdo e a QoS, além de prover informações específicas sobre cada uma dessas dimensões.[2]

Entendimento da relação entre recomendação e QoS Analisando os dados coletados, é possível estabelecer algumas observações importantes. Analisando as respostas dos usuários, ganhamos um entendimento sobre a relação entre recomendação e *QoS*. Seja o interesse por um vídeo caracterizado pelo número de estrelas que o usuário atribui ao conteúdo do vídeo, conforme tipicamente feito em plataformas como *Netflix*. Definimos então a diferença de interesses como interesse no vídeo A, servido com QoS baixa, menos interesse no vídeo B, servido com *QoS* alta. Constatamos com nossos experimentos então que, numa escala de -5 a 5, se a diferença de interesses estiver entre 0 e 2, a *QoS* é um fator muito importante. Caso contrário, a *QoS* passa ter um papel secundário e o que determina o vídeo preferido é de fato o interesse.

Proposta de novo sistema de recomendação Com base no entendimento sobre a relação entre recomendação e *QoS*, criamos um modelo de classificação baseado em árvore de decisão com o qual foi possível alcançar uma acurácia (taxa de acertos) de 77%. É importante esclarecer que embora, em um primeiro momento, modelos baseados em árvores de decisão não tenham um poder preditivo tão forte quanto outros algoritmos, eles são facilmente interpretáveis, o que simplifica o entendimento do comportamento dos usuários.

Estrutura do artigo O restante deste artigo está estruturado da seguinte forma. Discutimos trabalhos relacionados na Seção 2. Na Seção 3 apresentamos a metodologia experimental adotada. Em seguida, na Seção 4 discutimos resultados preliminares diretamente obtidos a partir dos dados. A Seção 5 aborda mineração de texto e análise de sentimento. Na Seção 6 discutimos a exclusão de *outliers* e, em seguida, propomos um

[1]Convidamos os leitores a participarem: `http://www.experimentoqos.com/`.
[2]O *dataset* colhido está disponível sob requisição aos autores.

esboço de um sistema de recomendação de vídeos usando os dados coletados na Seção 7. As limitações e discussões acerca do trabalho estão na Seção 8. Finalmente, a Seção 9 apresenta a conclusão.

2. Trabalhos relacionados

A literatura sobre sistemas de recomendação é vasta: [Turnbull 2017], [Gomez-Uribe and Hunt 2016], [Véras et al. 2015]. Inúmeros trabalhos também estudam o impacto da *QoS* na retenção dos usuários [Krishnan and Sitaraman 2013], [Nam et al. 2016], [Mok et al. 2011]. Trabalhos relacionando caching e sistemas de recomendação também estão recebendo atenção da comunidade acadêmica: [Dehghan et al. 2016], [Sermpezis et al. 2017], [de Freitas et al. 2017], [Liu and Yang 2018]. Entretanto, nenhum desses trabalhos realizou um amplo estudo experimental com usuários reais, para compreender a relação entre QoS, QoR e QoE.

QoS e retenção dos usuários Em [Krishnan and Sitaraman 2013] foi mostrado que espectadores começam a abandonar um vídeo se ele demora mais de 2 segundos para começar. Em [Nam et al. 2016], os autores concluíram que 99% das sessões de vídeo possuíam menos que 4 *rebufferings*. Os autores também mostraram que o abandono aumenta de forma muito significativa a partir da segunda ocorrência de *rebuffering*. Tais resultados contribuíram com a indução de falhas nos vídeos do experimento descrito neste artigo.

Integrando sistemas de recomendação e QoS Alguns trabalhos já consideraram a integração de sistemas de recomendação e *caching*. Neste artigo, consideramos mais especificamente a relação entre sistemas de recomendação e QoS. Em [Chatzieleftheriou et al. 2017] os autores visam maximizar a taxa de *cache hit*, sujeitos a uma distorção máxima no sistema de recomendação. Um próximo passo natural ao trabalho de [Chatzieleftheriou et al. 2017] consiste em validar o algoritmo proposto utilizando *datasets* reais. O mesmo se aplica ao trabalho de [Liu and Yang 2018], também baseado em dados sintéticos. Com os experimentos realizados no nosso estudo, obtemos um *dataset* gerado por usuários reais, que pode ser utilizado para fins de parametrizar os resultados de [Liu and Yang 2018, Chatzieleftheriou et al. 2017].

Os trabalhos de integração entre sistemas de recomendação, *caching* e QoS buscam soluções que tornem os conteúdos recomendados mais amigáveis para caches, assim como para o usuário que demanda o conteúdo. Neste trabalho, contribuímos para a literatura que integra sistemas de recomendação, *caching* e QoS, estudando a sensibilidade dos usuários com relação a *QoS* do sistema e a natureza dos conteúdos recomendados (*QoR*).

3. Metodologia experimental

Propusemos e colocamos em prática um experimento para quantificar a sensibilidade dos usuários ao conteúdo de vídeos e à qualidade de transmissão. Ao final de mais de 10 meses de coleta de dados, obtivemos 1002 instâncias. Considerando que cada usuário responde em média 3 vezes, com alguns abandonando o experimento após 2 sessões, o número de participantes foi de 385.

3.1. Descrição do experimento

O experimento é dividido em 2 etapas:

Na primeira etapa o usuário assiste a um par de vídeos que possuem exatamente o mesmo conteúdo. A diferença está na qualidade de serviço: enquanto um dos vídeos é reproduzido sem nenhum problema o outro apresenta diversas interrupções e diminuição da taxa de *bit rate* ao longo de sua reprodução. Nossa intenção ao exibir 2 vídeos idênticos em conteúdo é aferir em que medida a qualidade de serviço afeta a experiência do usuário, de modo a podermos atribuir uma nota ao seu nível de estresse (em uma escala de 0 a 10).

Na segunda etapa do experimento o usuário tem 3 pares de vídeos sorteados dentro de um conjunto de 7 (sem repetições). Diferentemente da primeira etapa, os vídeos de cada par não são idênticos. Embora a categoria de cada vídeo do par seja a mesma (e.g., esporte ou música), os vídeos foram escolhidos de modo que um seja em média mais interessante do que o outro. Além disso, o vídeo supostamente mais interessante apresenta qualidade de serviço inferior ao passo que o vídeo menos interessante possui *QoS* mais elevada. A ordem em que o vídeo de baixa qualidade é exibida varia de acordo com cada sessão. *A ideia é analisar o quanto o usuário está disposto a comprometer a qualidade para assistir algo mais interessante ou vice-versa.*

Vídeo *LowQoS* vídeo que apresenta *QoS* inferior mas com conteúdo que, para a maioria dos usuários, é mais interessante (e.g., compilação de pequenos trechos de esportes radicais, que em geral têm maior número de visualizações do que esportes com menos ação).

Vídeo *HighQoS* vídeo com menor número de interrupções (maior *QoS*), mas com conteúdo que em geral é considerado menos interessante (e.g., trecho com uma pessoa praticando *slackline*. Trata-se de um esporte em que o indivíduo se equilibra em uma corda, em um lugar seguro).

O número de pares de vídeos que constitui a segunda parte do experimento é 7. O usuário é convidado a assistir a apenas 3 pares de vídeos pois foi observado experimentalmente que muitos indivíduos ficam cansados se tiverem que assistir a mais do que isso. Todos os vídeos foram programados para durarem aproximadamente 30 segundos para tornar o experimento curto e prático, e assim maximizar o número de participações remotas. Cabe destacar que o experimento foi disseminado majoritariamente num ambiente universitário, abrangendo alunos e funcionários de áreas diversas.

3.2. Dados coletados

A cada dois vídeos apresentados, foi solicitado que o usuário preenchesse um formulário com os seguintes campos:

Estresse: Denota o quanto o usuário se sentiu estressado com as falhas de qualidade no primeiro par de vídeo (conteúdos idênticos e qualidade diferente).

E, para cada um dos 3 pares de vídeos subsequentes:

Interesse pelo vídeo *lowQoS* (intLowQoS): Interesse pelo conteúdo do vídeo de *QoS* baixa: valor numérico de 0 a 5.

Interesse pelo vídeo *highQoS* (intHighQoS): Interesse pelo conteúdo do vídeo de *QoS* alta: valor numérico de 0 a 5.

Incômodo: Incômodo do usuário com relação a interrupções em um dos vídeos. O incômodo reflete a sensibilidade do usuário com relação às falhas de qualidade do vídeo

que continha perturbações: valor numérico de 0 a 5.

Preferência: Preferência por um dos dois vídeos: *lowQoS* ou *highQoS*. Esta é a variável alvo, tendo em vista que queremos saber qual vídeo o usuário vai preferir, considerando simultaneamente seu interesse pelos conteúdos e sua sensibilidade a falhas.

Comentários: Após cada sessão, o usuário é convidado a explicar, em formato de texto, a razão para ele ter escolhido determinado vídeo.

É importante deixar claro que quando se trata do interesse pelo vídeo *lowQoS* ou *highQoS*, enfatizamos ao usuário para informar apenas o seu interesse pelo conteúdo, abstraindo o fator qualidade neste momento. O estresse é uma variável atribuída uma única vez a cada usuário, quando o respondente assiste ao primeiro par de vídeos, que possui conteúdos idênticos. As demais variáveis são atribuídas uma vez a cada par de vídeos adicionais que o usuário assiste. Em geral, para cada usuário temos tipicamente uma amostra da variável estresse e três amostras de cada uma das demais variáveis listadas acima.

Em determinados momentos da análise de dados utilizamos a variável **Dif_Interesse**, que é basicamente a diferença entre o interesse pelo vídeo lowQoS e o interesse pelo vídeo highQoS. Sendo assim, quando a diferença é negativa significa que o interesse pelo vídeo de qualidade alta foi maior. Uma outra variável também utilizada posteriormente é a **métrica de polaridade (Polarity Score)**, uma métrica de análise de sentimento que denota o quanto um texto é positivo ou negativo. Esta métrica foi obtida através da informação textual fornecida pelos usuários (vide Seção 5). O valor obtido por este método é sempre um valor que pode variar continuamente entre -1 e +1.

3.3. Detalhes técnicos

Nossos experimentos foram realizados através da Internet. Os participantes acessaram o experimento através de um website, no qual foram expostos a pares de vídeo e questionários correspondentes. Observe que as interrupções e distúrbios de qualidades (perturbações de QoS) foram adicionados aos vídeos de maneira *off-line*, antes de esses vídeos serem apresentados aos usuários. A instabilidade da conectividade da Internet dos usuários poderia, em princípio, causar prejuízos adicionais à QoS. Para evitar isso, desenvolvemos um módulo de software que baixou os arquivos de vídeo inteiros em uma determinada sessão para um *cache* local na máquina do participante, antes do início do experimento. Ao fazer isso, conseguimos controlar as deficiências de QoS às quais cada usuário foi exposto. Uma lista detalhada das perturbações de QoS adicionadas a cada vídeo será disponibilizada junto com nossos conjuntos de dados. Para os propósitos deste artigo, o impacto dessas perturbações é totalmente capturado através das respostas que os usuários forneceram sobre como essas perturbações afetaram sua QoS.

3.4. Classificadores e definição de acurácia

Uma vez coletados os dados, construímos classificadores capazes de, dadas as características dos 2 vídeos e o histórico do usuário, gerar uma previsão quanto à preferência desse usuário por um dos 2 vídeos do par.

Para cada classificador aqui apresentado, reportamos suas respectivas: acurácia, precisão, *recall* e *f1 score*. A acurácia, que é repetidamente mencionada ao longo do artigo, é a taxa geral de acerto considerando o *dataset* inteiro. *Precision* indica a proporção

de registros de fato positivos, entre aqueles classificados como positivos, enquanto *recall* indica a fração dos registros de fato positivos que foram classificados corretamente. A média harmônica entre *recall* e *precision* é chamada de *f1 score* [Sammut and Webb 2017].

4. Resultados experimentais

Na Tabela 2 temos os valores das médias e respectivos desvios padrão para os 3 principais atributos. Podemos ver que a média do interesse pelo vídeo de QoS baixa foi de fato superior à do interesse pelo conteúdo do vídeo de *QoS* alta. É necessário ressaltar que embora o interesse pelo vídeo *QoS* alta tenha sido em média menor, a maioria dos usuários acabou preferindo o vídeo menos interessante, resultado que reforça a relevância da *QoS* na qualidade de experiência.

Tabela 1. Estatísticas das preferências dos pares de vídeos

Par de vídeo	Descrição do par	N. usários	Preferiu HighQoS	Preferiu LowQoS
-	Todos os pares de videos	1002	561	441
1	Futebol	132	75	57
2	Comédia	118	78	40
3	Animais	135	64	71
4	Esportes Radicais	158	84	74
5	Documentário animal	135	95	40
6	Documentário científico	141	55	86
7	Música	163	98	65

Das 1002 instâncias coletadas, os usuários preferiram os vídeos de QoS alta 561 vezes (i.e., 56,99%) e em 441 (44,01%) ocasiões os usuários preferiram o vídeo de QoS baixa porém mais interessante (Tabela 1).

Tabela 2. Média e desvio padrão das métricas por par de vídeo

	Incômodo (média $\pm$ dp.)	Interesse low QoS (média $\pm$ dp.)	Interesse high QoS (média $\pm$ dp.)
-	3.56 ± 1.33	3.31 ± 1.42	2.77 ± 1.51
1	3.83 ± 1.19	3.27 ± 1.64	2.50 ± 1.66
2	3.52 ± 1.42	3.26 ± 1.34	2.97 ± 1.50
3	3.32 ± 1.42	2.93 ± 1.52	2.13 ± 1.51
4	3.53 ± 1.24	3.16 ± 1.35	2.49 ± 1.35
5	3.58 ± 1.39	3.13 ± 1.38	3.08 ± 1.49
6	3.33 ± 1.33	3.94 ± 1.25	3.05 ± 1.43
7	3.8 ± 1.3	3.56 ± 1.22	3.18 ± 1.4

A Tabela 3 mostra a matriz de correlações entre as diferentes variáveis medidas. A partir desta matriz, é possível concluir que os participantes do experimento souberam responder o questionário corretamente, visto que há baixa correlação entre os interesses e o incômodo. Ou seja, os participantes souberam abstrair a QoS no ato de avaliar o interesse pelo conteúdo.

É interessante notar que não foi apenas na média geral que o interesse pelo vídeo de pior qualidade foi mais elevado que interesse pelo de melhor qualidade. Os vídeos de baixa QoS tiveram uma média de interesse superior aos de alta QoS em cada uma das 7 sessões (vide Tabela 2). No entanto, das 7 sessões, em 5 delas (futebol, comédia, esportes radicais, documentário sobre animais e música) a maioria dos usuários preferiu os vídeos de QoS alta. Nas outras 2 sessões (animais e documentário científico), em que o vídeo

Tabela 3. Matriz de correlação

	IntLowQoS	IntHighQoS	Stress	Annoyance	Sentiment
IntLowQoS	1	0.393	0.039	0.084	0.043
IntHighQoS	0.393	1	0.086	0.192	-0.019
Stress	0.039	0.086	1	0.313	-0.030
Annoyance	0.084	0.192	0.313	1	-0.160
Sentiment	0.043	-0.019	-0.030	-0.160	1

preferido pela maioria foi o de pior *QoS*, eram justamente aquelas em que houve maior diferença de interesses entre os vídeos do par. Nesses casos a diferença de interesses era superior a 0,8. Além disto, estas duas sessões foram as que apresentaram menor média de incômodo, o que significa que os resultados foram consistentes. Dentre todas as sessões, aquela que apresentou maior média de incômodo foi a da categoria "música" (vide Tabela 2). A categoria "música" é uma categoria importante, que analisaremos em mais detalhes na Seção 7 (Figura 6).

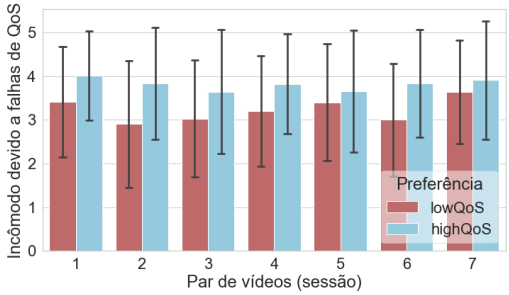


Figura 2. Incômodo por par de vídeos

Na Figura 3 temos no eixo das abscissas as diferenças entre interesse *lowQoS* e interesse *highQoS*. No eixo das ordenadas, representado por barras, temos o número de indivíduos que preferiu cada um dos vídeos. Na situação em que o interesse pelo vídeo *highQoS* é maior, o usuário quase sempre acaba escolhendo de fato o vídeo de maior qualidade (90% dos casos).

É possível observar também que quando a diferença de interesses é muito pequena ou nula, a qualidade passa a ser extremamente importante. Em particular, quando a diferença de interesses é 0, o vídeo QoS alta é escolhido em 71,9% dos casos. Nos outros 28,1% dos casos, o usuário escolhe o vídeo de pior qualidade, mesmo não havendo nenhum *trade-off* entre conteúdo e QoS. Atribuímos estas escolhas à granularidade da escala de avaliação escolhida. Isto é, em uma escala de 0 a 5 alguns usuários acabam não conseguindo expressar pequenas diferenças nos níveis de interesse e usam o campo textual para explicar a escolha (Seção 5).

Nas situações em que a diferença de interesses é igual a 1 ou 2, observamos que o vídeo de QoS baixa já passa a ser o preferido pela maioria. No entanto, notamos que há um segmento expressivo (30%) de usuários que continua escolhendo o vídeo de QoS alta apesar deste não apresentar seu conteúdo predileto (de menor grau de interesse).

5. Mineração de texto e análise de sentimento

Analisando a informação textual fornecida pelos usuários, podemos obter um entendimento mais profundo sobre como eles formulam seus processos de tomada de decisão.

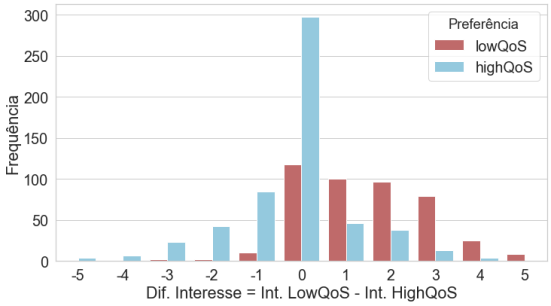


Figura 3. Preferência pelos vídeos em função dos interesses e QoS: quando a diferença de interesse pelos vídeos é menor ou igual a 2, usuários estão dispostos a consumir conteúdo menos interessante com maior QoS.

(a) Prefere vídeo highQoS (b) Prefere vídeo low QoS

Figura 4. Nuvem de palavras dos comentários em função das preferências.

Uma abordagem na análise do texto é contar a ocorrência de palavras ou termos específicos de modo a encontrar o que os usuários consideram mais importante. Para esta finalidade, as Figuras 4(a) e 4(b) apresentam as nuvens de palavras mais usadas nos comentários quando o video preferido foi o de alta / baixa QoS. Naturalmente, na nuvem de palavras referente à escolha do vídeo de *QoS* alta (Figura 4(a)) há uma presença muito mais forte de palavras relacionadas à qualidade de serviço, tais como: qualidade, imagem, interrupções, falha e travamento. É importante notar também a presença da preposição "sem", que quase sempre está associada a palavra "falha", para indicar "sem falha".

Na Tabela 4, que indica a frequência de termos usados nas justificativas, verifica-se que os usuários que escolheram o vídeo de *QoS* alta têm uma tendência a escrever algo relacionado a qualidade (e.g., "imagem", "interrupção", "qualidade"). Analogamente, para vídeos de baixa *QoS*, os participantes apresentaram uma tendência mais elevada a escrever algo relacionado ao conteúdo do vídeo (e.g., "interessante", "conteúdo"), conforme também indicado na nuvem de palavras da Figura 4(b). Em resumo, as nuvens de palavras ajudam-nos a compreender o processo de escolhas dos indivíduos que participaram dos experimentos, e corroboram o fato de que *QoS* e *QoR* têm papel fundamental na determinação da *QoE*.

Tabela 4. Frequência de palavras presentes na justificativa

Palavra	Frequência (highQoS)	Frequência (lowQoS)	Total
Vídeo	146	88	234
Qualidade	149	31	180
Falhas	104	31	135
Conteúdo	37	62	99
Interessante	35	58	93
Imagem	68	24	92
Interrupção	78	14	90

Análise de Sentimento A seguir, nós exploramos o relacionamento entre análise de sentimento e recomendação de vídeos. Em particular, nos concentramos em uma de nossas descobertas experimentais que relaciona (a) sentimento expresso pelos usuários em seus comentários sobre o experimento, com (b) sua escolha de vídeo preferido.

O VADER (*Valence Aware Dictionary and Sentiment Reasoner*) fornece um dicionário e uma ferramenta de análise de sentimentos baseada em regras. Dada uma entrada textual, o VADER retorna uma métrica de polaridade, que mede quão positivo ou negativo é um dado texto [Hutto and Gilbert 2014].

Usando este algoritmo, nós encontramos correlação entre os sentimentos expressos pelos usuários e suas escolhas de vídeos preferidos. Constatamos que usuários com sentimentos mais "positivos" tendem a preferir o vídeo de baixa QoS (eles são mais lenientes a falhas de QoS). *Por outro lado, usuários que escolhem o vídeo highQoS frequentemente reclamam da falta de qualidade naquele de baixa QoS. Desta forma, suas explicações contêm mais palavras negativas.*

A métrica de polaridade (*polarity score*) varia entre -1 e 1 e é calculada da seguinte maneira: $\frac{x}{\sqrt{x^2+\alpha}}$. Cada termo possui um valor de sentimento próprio que varia entre -4 e 4 e x é a soma dos valores dos termos presentes no texto. Utilizamos α, um parâmetro de normalização, fixado em 15. Nós empiricamente encontramos o limiar de 0.35 para separar os registros do nosso conjunto de dados. Utilizando este limiar, estabelecemos a seguinte regra,

$$\text{Preferência estimada} = \begin{cases} \text{high QoS}, & \text{se polaridade } < 0.35 \\ \text{low QoS}, & \text{caso contrário.} \end{cases} \quad (1)$$

Um achado interessante é que usando somente a métrica de polaridade e a simples regra acima, nós alcançamos uma taxa de acerto geral (acurácia) de 62.3%. A taxa de acerto para as respectivas classes (low e high QoS) foram de 59.1% e 64.1%. *Consideramos este um achado promissor para prever a preferência esperada dos usuários, a saber, pelos seus comentários prévios.*

6. Valores discrepantes (*outliers*)

A seguir, reportamos heurísticas para remover *outliers*. Nós definimos *outlier* como uma amostra que é inconsistente com uma regra implícita no domínio sob estudo. *Outliers* no nosso conjunto de dados podem ser decorrentes de um mal entendido em relação às questões propostas ou à granularidade das escalas consideradas (e.g., os usuários tinham que indicar seu interesse pelos vídeos em uma escala de números inteiros de 0 a 5).

Em particular, nós consideramos a seguinte verificação de consistência. Se um usuário atribuir o mesmo grau de interesse por ambos os vídeos de um par (Dif_interesse=0) e reportar um nível de incômodo alto (Incômodo > 3), o mais compreensível seria que ele escolhesse o vídeo de maior qualidade. Isto é, dado que o usuário se interessou igualmente por ambos os vídeos, não haveria uma troca entre qualidade de serviço e qualidade de conteúdo

De forma resumida, se as seguintes regras são satisfeitas simultaneamente, o registro é considerado um *outlier*: (i) Dif_interesse = 0; (ii) Incômodo > 3; (iii) VídeoPreferido = lowQoS. Nós indentificamos 59 registros que concomitantemente se enquadram

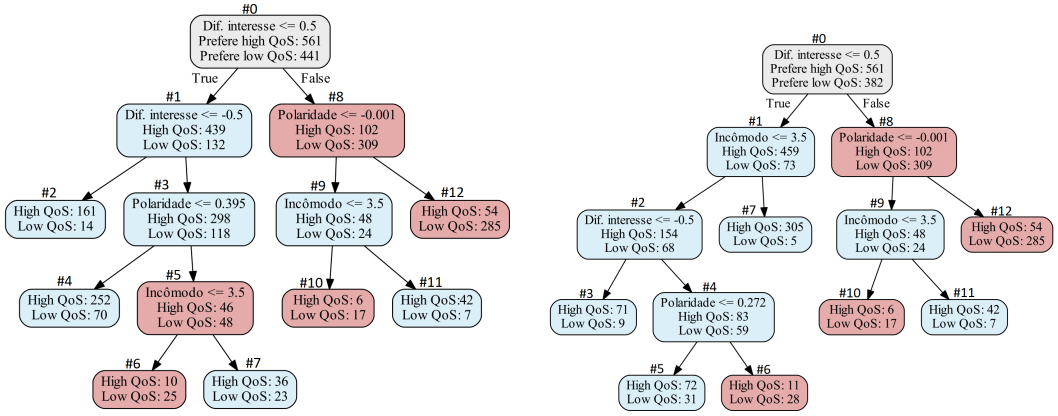
nas regras acima. Assim, todos estes foram classificados como *outliers* e removidos na análise seguinte.

Desempenho após remoção de *outliers* Aqui, reportamos resultados do desempenho de alguns classificadores depois de remover outliers (Tabela 5). Nós consideramos um número de distintos modelos de aprendizado de máquina, que são em geral mais complexos do que o modelo de árvore discutido até aqui. Os modelos de redes neurais e floresta aleatória foram os que tiveram melhor desempenho em todas as métricas consideradas. Os resultados apresentados aqui foram obtidas através de validação 10-cruzada (vide Seção 3.4).

7. Introduzindo *QoS-awareness* em modelos de recomendação

No início do artigo foi levantada a questão de como redes de *cache* estão relacionadas a sistemas de recomendação e sobre o quanto os usuário são sensíveis a problemas na *QoS*. Ao propormos um modelo de classificação que considera a qualidade de serviço e interesse pelo conteúdo, estamos propondo um esqueleto de um sistema de recomendação que prioriza conteúdos já em cache quando é sabido que o usuário é sensível em relação a interrupções e variações na taxa de transmissão dos vídeos.

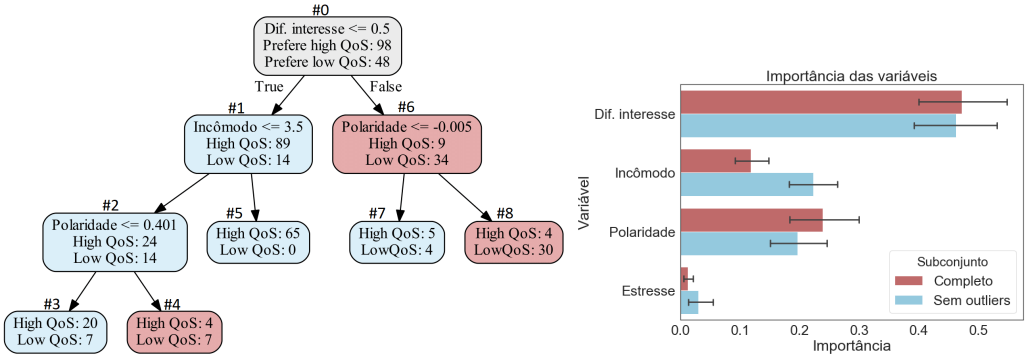
Neste sistemas de recomendação visamos prever que vídeo o usuário mais gostaria de assistir com base em seus interesses e sensibilidade a falhas de *QoS*. Usaremos primordialmente modelos baseados em árvores de decisão por terem alta capacidade explicativa, isto é, por serem facilmente interpretáveis.



(a) dataset completo (b) depois de remover outliers

Figura 5. Árvores de decisão que permitem interpretar as escolhas dos usuários

Na Figura 5 temos uma versão podada da árvore de decisão treinada, considerando o *dataset* completo (Fig. 5(a)) e removendo *outliers* (Fig. 5(b), vide também Seção 6). Cada nó da árvore, com a exceção das folhas, possui uma regra de divisão binária que separa o conjunto de dados. Além disso, cada nó também contém o número de registros que preferiram o vídeo de *high* QoS ou *low* QoS. O nó está em azul quando *high* QoS é maioria e em vermelho quando *lowQoS* é maioria. A árvore é construída usando a variável e limiar que fornece a melhor divisão do espaço a cada nó. Em particular, utilizamos a métrica Índice de Gini que pode ser obtida através da seguinte relação: $I_G(p) = 1 - \sum_{i=1}^{C}(p_i)^2$. Em que p_i é a probabilidade de um registro selecionado aleatoriamente pertencer à classe i e C é o número de classes, i.e. $i \in \{1, 2, ..., C\}$.



(a) Categoria música: muito sensível a QoS (b) Importância dos atributos

Figura 6. Árovre de decisão para categoria música e importância dos atributos.

.

A raiz da árvore na Figura 5 divide o conjunto de dados usando a diferença de interesses (*DifInterests*) entre os dois vídeos. Se DifInterests ≤ 0, os usuários tendem a escolher o vídeo de alta QoS. Caso contrário, o de baixa qualidade tenderia a ser preferido. Em ambos os casos, o incômodo (*Annoyance*) aparece logo em seguida em níveis inferiores na árvore, indicando a importância da QoS nas escolhas.

Seguindo o caminho dos nós #0, #1 e #2 é possível notar que, se DifInterests < 0, é extramemente provável que o vídeo escolhido seja o de alta qualidade. De modo análogo, *Polaridade* (nossa métrica de análise de sentimento) é útil em separar os dados quando a diferença de interesses é maior do que 0 (nós #0, #8 e #12).

A Figura 5(b) mostra a árvore de decisão com os *outliers* removidos (vide Seção 6). Ao remover os *outliers*, há naturalmente um aumento na acurácia além de obtermos uma árvore mais simples de ser interpretada. Agora, o incômodo (Annoyance) aparece no primeiro nível da árvore, logo abaixo da raiz. O caminho dos nós #0, #1 e #7 indica que uma fração significativa dos usuários seleciona os vídeos highQoS se DiffInterests < 0 e Annoyance > 3. O caminho através dos nós #0, #8 e #12 permanece o mesmo que já tínhamos na Fig. 5(a) e o restante da árvore é similar.

As árvores de decisão servem também para suportar o argumento de que comentários "positivos" indicam uma preferência em relação aos vídeos de low QoS. O caminho através dos nós #0, #1, #2 e #4 na Fig. 5(b) corrobora a discussão na Seção 5. O limiar de 0.272 (próximo do 0.35 escolhido na Seção 5) é selecionado para dividir os dados de modo que os registros com PolarityScore < 0.272 são classificados como os usuários que preferiram o vídeo highQoS.

Para ilustrar como árvores de decisão podem variar de acordo com as categorias, nós nos concentramos aqui na categoria música, que é uma das categorias de vídeos mais assistidas no Youtube (Figura 6(a)). De acordo com a Tabela 2, usuários foram menos tolerantes a falhas de QoS na categoria de música quando comparada com as demais. Incômodo (Annoyance) aparece no primeiro nível do processo de classificação e participa da separação de 70% dos registros.

Treinamos uma floresta aleatória, que apresentou os melhores resultados de acurácia quando comparada com outros métodos (vide Tabela 5). Outra vantagem deste modelo é permitir o cálculo das importâncias das variáveis que podem ser observadas na

Figura 6(b). Embora a diferença de interesses seja de longe a variável mais importante, o incômodo não é desprezível. As barras de erro representam um intervalo de confiança de 95%. *Estes resultados reforçam o papel conjunto da QoS e QoR na determinação da QoE.*

Cabe destacar que os resultados apresentados nesta seção são, do ponto de vista qualitativo, em parte intuitivos. Entretanto, a abordagem aqui apresentada faz-se necessária, do ponto de vista quantitativo, para determinação de limiares e implementação de recomendações automáticas. Na medida em que novos dados são recebidos sobre os usuários e vídeos, algoritmos de aprendizado por máquina (e.g., árvores de decisão) são chave para gerar recomendações sobre que vídeo indicar aos usuários, diretamente a partir dos dados históricos colhidos.

Tabela 5. Desempenho dos modelos de classificação

Modelos de Aprendizado de Máquina (com outliers removidos)				
Model	Acurácia	F1-Score	Recall	Precisão
Nearest Neighbors	0.8336	0.8291	0.8282	0.8301
Linear SVM	0.8122	0.8004	0.7957	0.8126
RBF SVM	0.8398	0.8346	0.8332	0.836
Random Forest	0.8557	0.8559	0.8588	0.8532
Neural Network	0.8547	0.8540	0.8558	0.8523
AdaBoost	0.8409	0.8350	0.8295	0.8406
Naive Bayes	0.8112	0.8029	0.7958	0.8103
QDA	0.8101	0.8023	0.797	0.8077

8. Limitações e discussões

A seguir, descrevemos quatro características de nosso estudo que merecem destaque. *Primeiro,* temos interdependência entre as observações porque um mesmo usuário assiste a mais de um par de vídeos. Estamos investigando métodos estatísticos específicos para lidar com tais dependências, que no momento não são explicitamente contempladas pelos métodos considerados. *Segundo,* temos um número limitado de temas (sete categorias) cobertos no experimento. Aumentar essa abrangência não é trivial, dado que identificamos que um usuário típico suporta assistir em torno de 4 pares de vídeos sem se cansar. *Terceiro,* a forma como a coleta de dados foi feita, baseada em escalas de preferência, necessariamente esbarra na subjetividade dos respondentes. Trabalhos futuros visarão identificar classes de usuários, e.g., baseado em seus interesses, para distinguir entre subjetividade inerente ao experimento e subjetividade relacionada aos interesses dos usuários. *Quarto,* a coleta como um todo demora cerca de 8 minutos. O experimento não envolve questões éticas ou de privacidade, por termos dado liberadade aos usuários para abandonar o experimento sempre que desejassem e a identificação dos usuários é voluntária.

9. Conclusão

A partir dos resultados coletados quantificamos como a QoS e QoR impactam a escolha dos usuários sobre seus vídeos preferidos (QoE). Os resultados mostram que para alcançar a satisfação do usuário, deve haver um equilíbrio entre o interesse pelo conteúdo e a qualidade do serviço. Possuindo dados das categorias de interesse de um usuário assim como o quanto o mesmo se incomoda com falhas de *QoS* como *rebufferings* e mudanças de *bitrate*, seria possível modelar um sistema de recomendação mais flexível e robusto.

Nossos resultados experimentais determinam como as preferências dos usuários dependem de diferentes fatores, como a categoria dos conteúdos em questão e a QoS da rede. Assim, os resultados nos permitem encontrar limiares para a tolerância a falhas em cada uma das situações. O uso de tais limiares para recomendar conteúdo, por sua vez, beneficia tanto os usuários (com aumento de *QoE*) quanto os administradores do sistema (menos custo para servir os usuários). Afinal, maior *cache hit* traduz-se em uma exibição com menos cortes (para os usuários) e menos consumo de banda no *backhaul*, que traduz-se em menores custos (para os administradores do sistema, que podem servir conteúdos em *cache*).

Referências

Chatzieleftheriou, L. E., Karaliopoulos, M., and Koutsopoulos, I. (2017). Caching-aware recommendations: Nudging user preferences towards better caching performance. In *INFOCOM*, pages 1–9. IEEE.

de Freitas, R. G., Menasché, D., Delgado, C., and Ziviani, A. (2017). Recomendação de conteúdo e desempenho de sistemas de cache.

Dehghan, M., Massoulie, L., Towsley, D., Menasche, D., and Tay, Y. C. (2016). A utility optimization approach to network cache design. In *INFOCOM 2016*, pages 1–9. IEEE.

Gomez-Uribe, C. A. and Hunt, N. (2016). The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS)*, 6(4):13.

Hutto, C. J. and Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *AAAI*.

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, D. and Yang, C. (2018). A learning-based approach to joint content caching and recommendation at base stations. *arXiv preprint arXiv:1802.01414*.

Mok, R. K., Chan, E. W., and Chang, R. K. (2011). Measuring the quality of experience of http video streaming. In *IFIP/IEEE Integrated Network Management*, pages 485–492.

Nam, H., Kim, K.-H., and Schulzrinne, H. (2016). Qoe matters more than qos: Why people stop watching cat videos. In *INFOCOM 2016*, pages 1–9. IEEE.

Sammut, C. and Webb, G. I. (2017). *Encyclopedia of Machine Learning and Data Mining*. Springer Publishing Company, Incorporated, 2nd edition.

Sermpezis, P., Spyropoulos, T., Vigneri, L., and Giannakas, T. (2017). Femto-caching with soft cache hits. In *GLOBECOM*, pages 1–7. IEEE.

Turnbull, D. (2017). High-quality recommendation systems with elasticsearch. `https://tinyurl.com/elasticc`.

Véras, D., Prudêncio, R., Ferraz, C., Bispo, A., and Prota, T. (2015). Context-aware techniques for cross-domain recommender systems. In *2015 Brazilian Conference on Intelligent Systems (BRACIS)*, pages 282–287. IEEE.