Recomendação de Conteúdo e QoE: Um Experimento Quantificando o Impacto da QoS nas Preferências por Conteúdos

Mateus S. Nogueira, Daniel S. Menasché, Carla Delgado

[1]Depto. de Ciência da Computação
Universidade Federal do Rio de Janeiro (UFRJ)

mateussnogs@gmail.com, sadoc@dcc.ufrj.br, carla@dcc.ufrj.br

Abstract. *Recommender systems are increasingly present in Internet users's routine. Therefore, platforms like Youtube and Netflix seek to improve their recommendation systems, to provide a better experience for their users. The experience of users, however, 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.*
In this paper, our goal is to devise a QoE-friendly recommender system. To this aim, we conduct an experiment with real users, having different profiles. Each users is requested to evaluate different movies, which vary in their contents and in the corresponding QoE. Our results provide a novel perspective on the relationships between recommender systems and caching systems. With decision trees, we propose a recommender system which accounts for QoE and content nature simultaneously. The proposed system reached a precision of 83% using our preliminary dataset which counts with hundreds of user ratings.

Resumo. *Sistemas de recomendação estão cada vez mais presentes na rotina de usuários na Internet. Portanto, plataformas como Youtube e Netflix buscam aprimorar seus sistemas de recomendação tendo em vista uma melhor experiência para os seus usuários. Porém, a experiência dos usuários depende de inúmeros fatores não apenas relacionados a natureza do conteúdo, mas também a forma como o conteúdo é entregue. Sistemas de cache, por exemplo, também possuem uma grande influência na qualidade de experiência (QoE) dos usuários, já que podem determinar a qualidade de serviço (QoS) dos conteúdos.*
Neste artigo, estabelecemos uma relação entre sistemas de recomendação e qualidade de serviço (QoS) tendo em base dados coletados a partir de um experimento realizado de forma remota com participações de diversos usuários, com diferentes perfis. Acreditamos que nossos resultados possibilitam abordagens inovadoras no estudo de sistemas de recomendação e algoritmos de caching de forma conjunta. Utilizando árvores de decisão, propomos um sistema de recomendação que leva em conta conjuntamente QoS e a natureza dos conteúdos. Em resultados preliminares, alcançamos uma precisão de 83% em decisões de recomendação baseadas em dados fornecidos por usuários reais que participaram dos experimentos propostos.

1. Introdução

Motivação Sistemas de recomendação determinam uma parcela expressiva da demanda por conteúdos na Internet. 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 redes de cache, analisando o impacto que a qualidade de serviço (QoS) possui sobre o interesse dos usuários por conteúdos. Como redes de cache e QoS estão estreitamente relacionadas no sentido que um determina o outro, respectivamente, a análise do impacto da QoS feita neste artigo também pode ser utilizada por outros estudos que considerem redes de cache e suas especificidades.

Beneficiando redes de cache por sistemas de recomendação, espera-se que plataformas como Youtube e Netflix possam oferecer uma melhor qualidade de experiência (QoE) para seus usuários.

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 (e.g., QoS ou natureza do conteúdo) mais influenciam os usuários, em cada situação, é uma tarefa não trivial.

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?
2. considere a escolha entre dois conteúdos para se recomendar ao usuário. Um deles disponível em cache (alta QoS) mas não 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 consistia 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 grupo de dois vídeos, um deles possuía distúrbios de qualidade. 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 utilizados 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]

[1]O experimento realizado para este trabalho pode ser acessado via Internet pelo link: `http://xexperiencement.ddns.net/` Convidamos os leitores que acessem o link e participem do experimento.

Estado da arte Em [Dehghan et al. 2016] os autores estudaram a relação entre sistemas de recomendação e *caching*. Em [de Freitas et al. 2017], estendeu-se o trabalho de recomendação de conteúdo levando em conta redes de *caching*. Entretanto, em nenhum desses trabalhos executou-se experimentos reais com usuários para compreender a sensibilidade dos mesmos com relação à QoS e a natureza dos conteúdos. O objetivo deste trabalho é preencher esta lacuna.

Contribuições Ao responder as perguntas propostas nesse trabalho, fazemos as seguintes contribuições.

Experimento com usuários reais realizamos experimentos com usuários reais, e colhemos preferências dos mesmos sobre diferentes pares de vídeos. Para cada par de vídeos apresentado, cada usuário informa aquele que lhe proporcinou 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 as respostas dos usuários, ganhamos um entendimento sobre a relação entre recomendação e QoS. Em particular, aprendemos que, numa escala de 0 a 10, se a indiferença dos usuários com relação aos conteúdos é menor que 0.5, a QoS é fator determinante da satisfação do usuário. Caso contrário, a QoS é irrelevante, e pode-se servir o usuário com qualquer um dos dois conteúdos garantindo o mesmo nível de satisfação;

Proposta de novo sistema de recomendação baseado no entendimento sobre a relação entre recomendação e QoS, propomos um novo sistema de recomendação de conteúdos, usando uma árvore de decisão cuja precisão, para o conjunto preliminar de dados considerados, foi de 83%.

Estrutura do artigo O restante deste artigo está estruturado da seguinte forma. Na Seção 2 apresentamos a metodologia experimental adotada. Em seguida, na Seção 3 discutimos resultados preliminares diretamente obtidos a partir dos dados. Na Seção 4 propomos um novo sistema de recomendação usando os dados colhidos, fazendo uso de uma árvore de decisão. Finalmente, as Seções 5 e 6 apresentam trabalhos futuros e conclusão.

2. Metodologia experimental

2.1. Visão geral

Buscamos simular situações reais que usuários de plataformas como *Youtube* poderiam se deparar para obtermos um *dataset* coerente com a realidade.

Até o momento deste artigo, um *dataset* com 56 instâncias foi construído, porém sendo uma quantidade totalmente passível de ser aumentada. Em cada instância temos valores numéricos de 0 a 10 que representam: interesse do usuário pelo conteúdo do primeiro vídeo de uma determinada categoria, interesse pelo segundo vídeo dessa mesma categoria, e o incômodo com as falhas de qualidade em um desses vídeos. As falhas eram presentes sempre em apenas um dos vídeos de cada categoria. Além disso, a instância

[2]O *dataset* colhido está disponível sob requisição aos autores.

também possui a preferência do usuário por um dos dois vídeos da categoria, e a justificativa dele. Com os parâmetros destas instâncias analisamos se um usuário poderia preferir um vídeo em que ele teria menos interesse pelo conteúdo, porém sem distúrbios na qualidade de serviço (QoS), ou se ele optaria por um vídeo com falhas, mas com o conteúdo que despertaria um maior interesse.

2.2. Descrição do experimento

A ordem de cada dois vídeos apresentados ao usuário foi escolhida de forma arbitrária. Entretanto, para fins de descrição de nosso conjunto de dados, reordenamos os vídeos, e denotamos os vídeos de cada par por vídeo 1 e vídeo 2, de tal forma que:

vídeo 1 vídeo com maior número de interrupções (menor QoS), mas com conteúdo que, para a maioria dos usuários, é mais interessante

vídeo 2 vídeo com menor número de interrupções (maior QoS), mas com conteúdo que, para a maioria dos usuários, é menos interessante

Para fins de ilustração, considere o seguinte exemplo:

vídeo 1 segmento da vitória da Alemanha sobre o Brasil na copa do mundo. O vídeo não possui nenhuma interrupção (QoS excelente) mas espera-se que para os espectadores brasileiros o interesse/gosto seja baixo;

vídeo 2 segmento de um trecho de um gol do Brasil, mas com interrupções (QoS baixa). Espera-se que os espectadores brasileiros tenham interesse nesse vídeo, mas que fiquem incomodados com a QoS.

Em versões iniciais, o experimento era composto por 14 vídeos. No entanto, o número de vídeos exibido por participante foi reduzido a 7, visto que muitos participantes abandonavam o experimento após realizarem metade dele. Todos os vídeos foram designados a durarem apenas 30 segundos para tornar o experimento curto e prático, a fim de maximizar o número de participações remotas. Para mais detalhes sobre os vídeos exibidos, e as falhas induzidas neles, o *link* `https://goo.gl/tTmEbn` pode ser consultado.

2.3. *Features* coletadas

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

- **Interesse_Vídeo1:** Interesse pelo conteúdo do primeiro vídeo: Valor numérico de 0 a 10.
- **Interesse_Vídeo2:** Interesse pelo conteúdo do segundo vídeo: Valor numérico de 0 a 10.
- **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 10.
- **Preferência:** Preferência por um dos dois vídeos: primeiro ou segundo. Essa é nossa variável alvo, também denotada por *class* (classe da preferência).

Para cada par de vídeos, também colhemos uma justificativa textual do usuário pela sua preferência.

3. Resultados experimentais

A partir das instâncias coletadas, algumas estatísticas foram realizadas. Em média os usuários relataram um incômodo numérico de 6 quanto as falhas de qualidade dos vídeos. 75% relataram um incômodo maior que 8, e 20% um incômodo de 4. A Figura 1 apresenta esses resultados. Dos 56 dados, em 27 a preferência foi pelo vídeo que tinha a melhor qualidade, porém, com conteúdo provavelmente desinteressante, e 29 escolheram o vídeo com o conteúdo mais interessante, porém de pior qualidade. Essa divisão quase que simétrica de preferências demonstra como pode ser desafiador para sistemas de recomendação tomarem decisões considerando redes de cache.

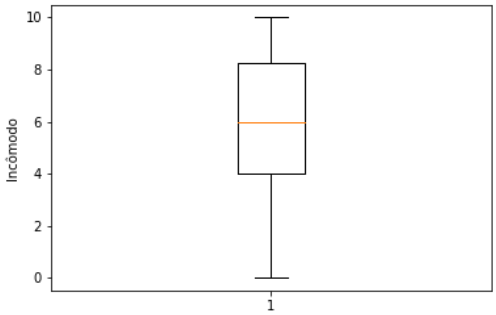

Figura 1. Boxplot dos incômodos

4. Um novo sistema de recomendação sensível à QoS

Como a nossa variável alvo é a preferência de um determinado usuário pelo primeiro ou segundo vídeo em função dos interesses pelos conteúdos, e o incômodo com as falhas de qualidade, o *dataset* pode ser utilizado por algoritmos de *machine learning* voltados para classificação.

No entanto, grande parte do interesse do estudo é analisar sistematicamente os fatores que determinam as preferências dos usuários. Para esse fim, utilizamos árvore de decisão. Com ela, evidenciamos o peso de cada campo para a classificação dos dados.

Árvores de decisão classificam instâncias testando uma determinada característica por nó, de forma que a classificação final se dê nas folhas da árvore. É importante notar que a raiz da árvore testa a característica de maior relevância para a classificação conforme o uso de métricas como impureza de Gini e ganho de informação. Com os dados do experimento, a característica que se situou na raiz foi a diferença entre os interesses pelos conteúdos. A árvore é exibida na Figura 2.

4.1. Análise dos resultados

A Figura 2 apresenta a árvore de decisão construída a partir dos dados coletados.

Se o usuário é indiferente sobre os conteúdos, a QoS importa Pelo caminho de sequência (1, 2, 6), a diferença entre os interesses sendo muito pequena(<= 0.5), o incômodo com a perturbação da qualidade não sendo pequeno(> 4.5), a decisão é considerar que o usuário irá preferir o segundo vídeo, que é o que possui a melhor qualidade.

Se o usuário tem forte preferência por um conteúdo, a QoS não importa Pelo caminho (1, 7, 9, 12), a diferença entre os interesses é considerável(> 0.5), o interesse pelo conteúdo do primeiro vídeo (de pior qualidade) é grande(> 5.5), e o incômodo com as interrupções não é tão grande(<=6.5), então a preferência será pelo primeiro vídeo, mesmo tendo uma qualidade inferior à opção alternativa.

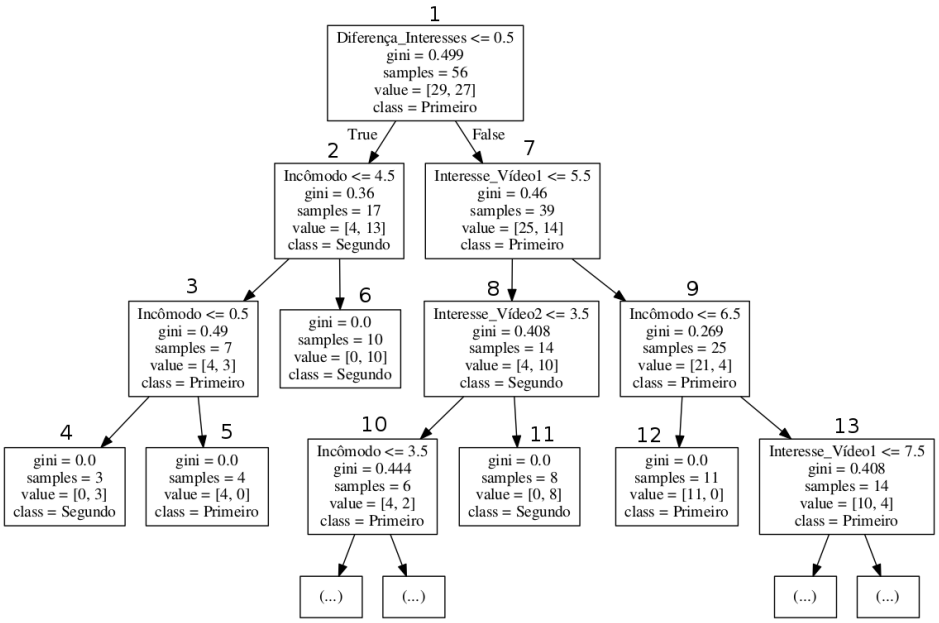

Figura 2. Árvore de decisão

Cabe destacar que, embora as observações acima possam parecer intuitivas, o nosso trabalho permite quantificar preferências. Essa quantificação é de fundamental importância para tomada de decisões automática. Assim, nossa contribuição consiste em estabelecer limiares quantitativos de tolerância dos usuários sobre a natureza do conteúdo, e sobre a QoE, para melhorar a qualidade dos sistemas de recomendação.

4.2. Precisão

Adicionalmente às nossas interpretações da árvore, medimos a precisão do algoritmo empregado por ela. No nosso contexto de classificação, é o número de verdadeiros positivos dividido pela soma de verdadeiros positivos e falsos positivos.

Para saber o número de verdadeiros positivos e falsos positivos, é preciso analisar o campo *values* contido nos nós da árvore. Esse campo informa a quantidade de instâncias que estão classificadas como Primeiro e Segundo. No nó 3, por exemplo, *values* = [4, 3]. Isso significa que há 4 instâncias classificadas como Primeiro e 3 como Segundo. Na folha de número 5, pode-se observar que as 4 instâncias que estavam classificadas no *dataset* como Primeiro foram previstas como sendo da classe Primeiro, resultando em 4 verdadeiros positivos.

Pelos caminhos (1, 2, 3, 5), (1, 7, 8, 10), (1, 7, 9, 12) e (1, 7, 9, 13) concluímos que o número de verdadeiros positivos é igual a 29. Pelos caminho (1, 7, 8, 10) identificamos 6 falsos positivos. Portanto, a precisão é igual a: 29/(29+6) = 29/35 = 83%.

5. Trabalhos relacionados

A literatura sobre sistemas de recomendação é vasta: [Turnbull 2017], [Gomez-Uribe and Hunt 2016]. 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].

5.1. 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], a partir de um dataset com uma quantidade considerável de dados colhido pela experiência de usuários no Youtube, concluíram que 99% das sessões de vídeo possuíam menos que 4 rebufferings. Também mostraram que múltiplos rebuferrings causam um abandono muito maior do que apenas uma ocorrência. Tais resultados contribuíram com a indução de falhas nos vídeos do experimento descrito neste artigo.

5.2. 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] é proposto um algoritmo para solucionar as direções opostas que sistemas de recomendação e redes caches seguem. O primeiro procura determinar a demanda dos usuários de acordo com seus interesses por conteúdos, ao passo que o segundo busca disponibilizar conteúdos já contidos na cache. Portanto, o problema consiste em maximizar a taxa de *cache-hit*, porém sujeito a uma distorção máxima no sistema de recomendação. Um próximo passo desse artigo seria validar o algoritmo proposto utilizando datasets reais. No experimento realizado no nosso estudo, um dos nossos objetivos principais é exatamente obter um dataset que reproduz a realidade da melhor forma possível.

Em [Liu and Yang 2018] um outro algoritmo é proposto para resolver o mesmo problema: maximizar a taxa de hit em caches, perturbando a recomendação de conteúdos mais interessantes para usuários o menos possível. Porém, nesse artigo dados hipotéticos foram considerados, inclusive sobre o possível perfil de usuários. Com o dataset ainda em expansão do nosso experimento, é possível traçar perfil de usuários reais, e contribuir com dados reais a proposta do algoritmo desse artigo citado.

Todos estes trabalhos de integração buscam o objetivo em comum de encontrar soluções que tornem os conteúdos recomendados mais amigáveis para caches, assim como para o usuário que demanda o conteúdo, da melhor forma possível. O que resultaria em uma QoE otimizada disponível por plataformas de multimídia online.

Neste trabalho, contribuímos para a literatura de sistemas de recomendação, estudando a sensibilidade dos usuários com relação a QoS do sistema e a natureza dos conteúdos.

6. Conclusão

Com nossos resultados preliminares foi possível obter um vislumbre sobre como a QoS pode impactar no interesse dos usuários. Os resultados mostraram que possuindo dados sobre o interesse de um usuário por determinado conteúdo, e sabendo o quanto ele se incomoda com falhas de qualidade como *rebufferings* e mudanças de *bitrate*, um sistema de recomendação pode se adaptar de uma melhor forma, não necessariamente recomendando conteúdos que despertarão o maior interesse, pois podem ter falhas de qualidade, ou também podendo recomendar conteúdos que tenham uma melhor qualidade, mas não sejam do maior interesse do usuário. Neste último caso, as redes de cache poderiam ser beneficiadas, visto que muitos conteúdos que possuem chance de apresentar menos falhas são aqueles que já estariam contidos em cache.

Link para o experimento e convite aos leitores

O experimento pode ser acessado via Internet pelo link: `http://xexperiencement.ddns.net/`

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.

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 *Integrated Network Management (IM), 2011 IFIP/IEEE International Symposium on*, pages 485–492. IEEE.

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.

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

Turnbull, D. (2017). High-quality recommendation systems with elasticsearch. `https://dzone.com/articles/high-quality-recommendation-systems-with-elastic-2`.