

# Uma Análise de Interface para Construção do Perfil Inicial de Usuários em Sistemas de Recomendação Baseados em Conteúdo

Nathália Locatelli Cezar<sup>1</sup>, Isabela Gasparini<sup>1</sup>, Daniel Lichtnow<sup>2</sup>

<sup>1</sup>Universidade do Estado de Santa Catarina (UDESC)

<sup>2</sup>Universidade Federal de Santa Maria (UFSM)

czr.nathalia@gmail.com, isabela.gasparini@udesc.br, dlichtnow@gmail.com

**Abstract.** *A well-known problem in Recommender Systems is the Cold Start User, which refers to the fact that, when a user starts using a system, there is no information that allows generating good recommendations for this user. Some Recommender Systems seek to solve this by presenting items and asking users to evaluate these items. Here some questions involved are: (1) which items to select to be presented, (2) how many items are there and (3) how to present these items. This article discusses these issues a little and, in the case of the latter, presents an experiment and evaluation carried out with users.*

**Resumo.** *Um problema bastante conhecido em Sistemas de Recomendação é o chamado Cold Start User que refere-se ao fato de que, quando um usuário começa a utilizar um sistema não existem informações que permitam gerar boas recomendações para este usuário. Alguns Sistemas de Recomendação procuram resolver isto apresentando itens e pedindo que os usuários realizem a avaliação destes itens. Aqui algumas questões envolvidas são: (1) quais itens selecionar para serem apresentados, (2) qual a quantidade de itens e a (3) qual a forma de apresentar estes itens. O presente artigo, discute um pouco estas questões e no caso da última apresenta um experimento e avaliação feitas com usuários.*

## 1. Introdução

Sistemas de Recomendação procuram recomendar itens que sejam de interesse dos seus usuários. Os Sistemas de Recomendação são geralmente classificados em três grandes abordagens: (a) Recomendações Baseadas no Conteúdo, o usuário recebe recomendações de itens que são similares àqueles que ele preferiu no passado; (b) Recomendações baseadas na Filtragem Colaborativa, o sistema recomenda itens para o usuário em questão que foram avaliados por usuários com gostos semelhantes; e (c) Abordagens Híbridas, quando métodos de Recomendações Colaborativas e Baseadas no Conteúdo são combinados para recomendar itens [Consoni 2014].

Um dos problemas mais conhecidos em Sistemas de Recomendação, e que afeta as abordagens citadas, é o problema de *Cold Start*. O problema de *Cold Start* pode ser definido como a impossibilidade de gerar boas recomendações devido à falta de dados sobre um novo usuário ou um novo item [Monti et al. 2021]. Esse problema surge quando

um novo usuário (*Cold-Start User - Cold Start* do usuário) ou item (*Cold-Start Item - Cold Start* do Item) é adicionado ao Sistema de Recomendação.

O problema do novo usuário *Cold-Start User*, afeta todas abordagens citadas e está relacionado ao fato de que um usuário que interagiu pouco no sistema ou fez poucas avaliações, não receberá inicialmente recomendações de qualidade. Na abordagem denominada Filtragem Colaborativa existe ainda o problema do novo item (*Cold-Start Item*), itens que foram avaliados por poucos usuários ou que não foram avaliados não serão recomendados [Ricci et al. 2015].

Existem tentativas de minimizar o problema do *Cold-Start*, a abordagem Híbrida é uma delas. Além disto, alguns sistemas procuram solicitar ao usuário um determinado número de avaliação de itens de forma a construir um perfil inicial do usuário. Aqui algumas questões envolvidas são: (1) quais itens selecionar para serem apresentados, (2) qual a quantidade de itens e a (3) qual a forma de apresentar os itens a serem avaliados. O presente artigo, discute um pouco estas questões e no caso da (3) apresenta um experimento e avaliação feitas com usuários no sentido de obter parâmetros para definir isto.

O restante deste texto está organizado da seguinte forma: A Seção 2 apresenta técnicas utilizadas em trabalhos relacionados e resultados do Mapeamento Sistemático da Literatura realizado. A Seção 3 apresenta as definições sobre a interface e questões envolvidas nesse processo além de experimento realizado. Na Seção 4 são apresentadas as considerações finais.

## 2. Mitigação do *Cold Start* - Trabalhos Relacionados

Abordar o problema de *Cold Start* pode ser importante para o engajamento e retenção de um novo usuário recém chegado no sistema. Por isto vários trabalhos buscam minimizar este problema. Neste sentido, [Abdullah et al. 2021] apresentam um *survey* com o objetivo de apresentar várias abordagens e técnicas empregadas por pesquisadores e os desafios associados na obtenção de informações necessárias para recomendação de itens para usuários procurando minimizar o *Cold Start*. No trabalho de [Panda and Ray 2022] os autores indicam que pesquisas recentes sobre abordagens e algoritmos que tentam engajar novos usuários e promover novos itens.

Quando o sistema de recomendação entra na fase de adquirir informações sobre as preferências do usuário, são realizados os métodos de elicitación de preferências. Existem quatro métodos mais comuns que os Sistemas de Recomendação utilizam para adquirir essas preferências do usuário sendo **avaliação dos itens**, **pesos nos atributos**, **críticas ou comportamento explícito** e **comportamento implícito** [Ricci et al. 2015].

O método de avaliação de itens é o método de avaliação explícita mais utilizado. Os usuários utilizam escalas (normalmente via escala de *Likert* de 5 pontos ou escalas de até 10 pontos para atribuir notas de acordo com suas preferências. Outro tipo de escalas de avaliação são usadas como “concordo fortemente, concordo, neutro, discordo, discordo fortemente” ou ainda avaliações binárias quando o usuário é questionado simplesmente se o item é bom ou ruim [Goldberg et al. 2001, Schafer et al. 2007].

A elicitación de preferências via pesos nos atributos se origina da área de análise de decisão, onde a utilidade de multi-atributos é utilizada como um padrão para a tomada de decisão [Bettman et al. 1998]. Esses atributos normalmente são um conjunto de palavras

que descrevem os itens associados a um contexto numa escala indicando do melhor valor para o pior valor, como por exemplo, para o contexto de preço os atributos seriam “muito bom, bom, ruim, o pior, muito ruim” [Bettman et al. 1998]. Desta forma, cada atributo que o usuário interage possui um peso se baseando em gostos maiores que outros.

Outro método de elicitare as preferências do usuário é utilizar feedback detalhado nas recomendações, também conhecido como comportamento ou feedback explícito. Quando um usuário realiza uma avaliação por meio de uma crítica ou comentário no item, ele ajuda na melhora da confiança de decisão de outros usuários sobre aquele item.

O feedback implícito, isto é ações dos usuários, também pode ser considerado para obter dados sobre a preferência de um usuário e requer menor esforço do usuário [Ricci et al. 2015]. Por exemplo, em uma página Web/site, pode-se considerar a seleção de um anúncio como um feedback positivo, ou ainda ações de compra/consumo [Ricci et al. 2015]. Em Sistemas de Recomendação voltados para a recomendação de artigos, informações para construção do perfil do usuário podem ser extraídas a partir de ações anteriores do usuário, assim são utilizados dados do histórico do usuário como por exemplo *feedback* implícito, que inclui cliques e visualizações anteriores de artigos realizadas pelo usuário dentro do sistema [Ricci et al. 2015].

Algumas técnicas para minimizar o Cold Start são discutidas baseadas em domínios de aplicação específicos. O sistema do *MovieLens* [Jelassi et al. 2013] por exemplo, constrói o perfil do usuário a partir de suas preferências de filmes, com base na seleção de categorias de filmes seguido de avaliações (*ratings*) dos filmes recomendados. Já o *FOAFing* é outro Sistema de Recomendação que descobre, explora e recomenda músicas com abordagem Baseada em Conteúdo a partir do perfil do usuário construído com informações vindas de mídia social [Celma et al. 2005, Celma et al. 2006].

No trabalho de [Son 2016] as técnicas utilizadas em trabalhos que buscam minimizar o problema do Cold Start do novo usuário podem ser divididas em três principais categorias:

1. utilizar fontes de dados adicionais;
2. escolher os grupos mais proeminentes de usuários, e;
3. melhorar a previsão usando métodos híbridos.

A ideia de utilizar fontes de dados adicionais está frequentemente relacionada a dados demográficos do usuário ou as opiniões dos usuários extraídas de comentários nas redes sociais. A técnica de escolher os grupos mais proeminentes é relacionada a ideia de clusterizar os usuários por algum atributo, podendo ser inclusive pelos dados demográficos, e ter um conjunto de usuários ativos que representem boa parte dos outros usuários daquele grupo ou cluster. Por fim, as técnicas de utilizar métodos híbridos se refere a trabalhos que utilizam mais de uma técnica de recomendação.

Por outro lado, quando não existe nenhum histórico do usuário, nenhuma interação dentro do sistema, o problema de *Cold Start* é preciso definir como identificar as preferências e gostos do usuário, especialmente quando se trata de um dos primeiros usuários do sistema ou de um novo usuário, sem depender do histórico de interação de outros. A busca por formas de mitigar o *Cold Start* sem usar fontes externas motivou um mapeamento sistemático.

Com o intuito de verificar como Cold Start do usuário é minimizado, realizou-se

em [Cezar et al. 2024] o Mapeamento Sistemático da Literatura (MSL) para identificar técnicas e soluções para a construção inicial do perfil do usuário.

Basicamente em [Cezar et al. 2024] foram identificados 19 artigos, sendo que considerando a forma de criar o perfil inicial do usuário a partir da leitura dos trabalhos, foi identificar as seguintes possibilidades (P): P1 - Solicitar que o usuário responda perguntas/questionário; P2 - Solicitar a avaliação do usuário para um conjunto de itens, P3 - Solicitar que o usuário selecione itens; P4 - Solicitar para o usuário a seleção/cadastro de palavras-chave/tags; P5 - Usar itens populares entre usuários mais antigos (ou usuários considerados especialistas) do Sistema de Recomendação.

A técnica de perguntas/questionário (P1) consiste em realizar alguma(s) pergunta(s) ao usuário assim que este entra no sistema para identificar suas preferências. Quando é utilizada a avaliação, é solicitado ao usuário que avalie alguns itens no seu momento de entrada no sistema (P2) ou então que selecione itens de interesse dentre os apresentados (P3). A técnica de seleção/cadastro de palavras-chave/tags (P4) é quando o sistema solicita que o usuário selecione ou cadastre palavras que melhor descrevam seus interesses, e por fim, a técnica usando itens populares (P5), que são itens que foram melhor avaliados por usuários ativos ou chamados de especialistas.

### **3. Definições sobre Interface/Interação dos Usuários para Mitigar Cold Start**

Das formas de criar o perfil inicial do usuário a partir do Mapeamento Sistemático, a que foi considerada em [Cezar et al. 2024] para um trabalho futuro, envolvendo a implementação de um sistema de recomendação, foi a técnica de solicitar que o usuário selecione/avalie itens de forma a que seja possível construir um perfil inicial com as preferências deste usuário.

Aqui algumas questões envolvidas são: (1) quais itens selecionar para serem apresentados, (2) qual a quantidade de itens e a (3) qual a forma de apresentar os itens a serem avaliados. A forma de seleção dos itens a serem mostrados aos usuários, é discutida em outro trabalho, já para o presente trabalho as questões (2) e (3) são as discutidas.

No presente trabalho, está se pensando em definir a interface para seleção/avaliação de itens que permitirão definir o perfil inicial de um usuário, de um sistema de recomendação de artigos científicos que usa a abordagem Baseada em Conteúdo. Dado este contexto, duas opções de seleção/avaliação de itens do perfil do usuário via interface foram consideradas, devendo ser feita a escolha entre elas: (1) apresentar pares de itens para serem avaliados logo que o usuário acessa o sistema e (2) apresentar 10 itens simultaneamente logo que o usuário acessa o sistema.

Para embasar esta escolha, visando a implementação do sistema de recomendação, foram executados experimentos com alunos de graduação em Ciência da Computação e especialistas da área de Interação Humano-Computador para a definição da melhor interface, e do número ideal de itens a serem mostrados ao usuário para seleção/avaliação.

Para realização destes experimentos, foram criadas na ferramenta Figma uma simulação das interfaces de seleção/avaliação de itens perfil que seria apresentada ao usuário do sistema de recomendação, tão logo ele acessasse o sistema pela primeira vez.

Em ambas as alternativas, na primeira tela era possível visualizar uma breve descrição do trabalho, explicando o objetivo da escolha de interface. A interface com

10 alternativas foi apresentada com itens em duas colunas e o usuário conseguia visualizar todos os itens ao rolar a tela para baixo (caso estivesse usando um dispositivo como um celular), podendo selecionar itens ou não. Quando acessava esta interface o usuário conseguia visualizar os primeiros dois itens da lista porém ao rolar a tela para baixo conseguiria ver a lista de itens completa (Figura 1). Já na interface onde apenas 2 itens eram apresentados o usuário conseguia interagir visualizando um par por vez, selecionando um, dois ou nenhum item (ver Figura 2). Em ambas as interfaces, quando o usuário selecionava um item, este mudava de cor, identificando a seleção do item.

Em ambas interfaces, era solicitado que o usuário realizasse a seleção de pelo menos 5 itens. Isto foi considerado o mínimo de itens a serem avaliados/selecionados para criação de um perfil inicial - cabe observar que este número pode ter que ser maior, mas para este trabalho, que envolve a criação de um sistema de recomendação de artigos, foi considerado adequado a partir de testes preliminares realizados. Assim, no caso da interface que apresenta pares de itens para serem avaliados/selecionados o número de pares apresentados varia conforme o número de artigos selecionados (caso o usuário selecione sempre um dos itens apresentados pelo menos 5 pares são apresentados).

O experimento foi conduzido através de uma reunião online com um total de 6 participantes, os quais eram Graduandos do curso de Ciência da Computação e/ou alunos que realizaram a disciplina de IHC na UDESC, e após a demonstração das interfaces, os usuários foram convidados a responder um questionário sobre qual interface preferiram, além de sugestões.

No experimento, 83,3% dos *testers* utilizaram o computador para participar da reunião e avaliar a interface, os outros 16,7% utilizaram um *smartphone*. Independentemente do dispositivo utilizado, todos preferiram a interface que apresentava dois itens por vez pela interface em pares, sendo a interface selecionada para ser criada no sistema proposto. As características e especificidades que os fizeram escolher esta interface foram *“devido a organização, e a disposição ficar mais clara.”*, *“Mesmo que esse modelo seja mais lento que o outro, dá a impressão de que eu precisei pensar menos para efetuar as escolhas, acho é uma perca justa, perde no tempo e mas ganha na ”facilidade”de uso.”*, *“Gostei mais da interface em pares, pois a pessoa pode escolher entre apenas 2 opções e entendo que a carga de trabalho dele fique menor”*. Mais detalhes das respostas dadas por cada participante sobre a preferencia de interface, pode ser visualizada na Tabela 1.

No que diz respeito a quando os participantes foram questionados sobre sugestões das interfaces, muitas sugestões foram relacionadas ao *design* das telas, como por exemplo, *“Bordas arredondadas nos cartões”*, ou com relação a mensagem mostrada ao usuário na tela de seleção, como orientação do que o usuário precisa fazer na tela de seleção de itens ser algo como *“escolha os artigos de sua preferência: considerando além da area de interesse, a prefrenca por autores e palavras chave”*. Notou-se também que 50% dos participantes relataram a questão de deixar mais explícito a questão do usuário poder selecionar 1, 2 ou nenhum item, principalmente neste último caso que o usuário pode não selecionar nenhum item do par e ainda assim, visualizar o próximo par.

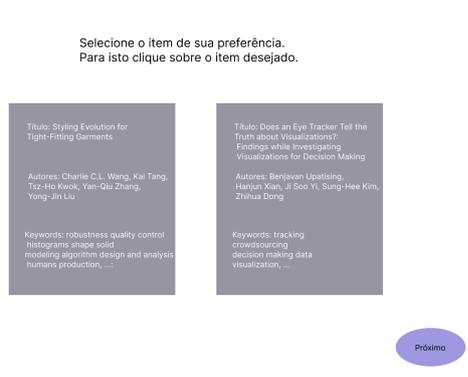
#### **4. Consideração Finais**

O *Cold Start* é um dos problemas ainda não solucionado em Sistemas de Recomendação, mesmo sendo um assunto sempre presente na área. Uma das formas de reduzir o problema

Figura 1. Interface com 10 itens



Figura 2. Interface com 2 itens



**Tabela 1. Respostas dos participantes sobre as interfaces apresentadas**

<b>Participante</b>	<b>Sobre as interfaces apresentadas na reunião com o grupo, qual você mais gostou? Quais as características e especificidades que fazem você ter escolhido tal interface?</b>
Participante 1	A interface em pares. Pois creio que o processo de escolha seja facilitado quando se precisa escolher entre duas opções ao invés de 10, também creio que essa opção seja a mais vantajosa quando o sistema for usado em telas menores, celulares principalmente, pois a quantidade de informações será bem menor e mais fácil de adaptar a displays menores.
Participante 2	O modelo em pares. A tela ficou menos poluída e consigo prestar mais atenção na escolha que estou fazendo.
Participante 3	Gostei mais da interface em pares, pois a pessoa pode escolher entre apenas 2 opções e entendo que a carga de trabalho dele fique menor, e a opção de visualizar todos os itens pode causar uma certa "preguiça" e fazer com que o mesmo acabe desistindo de responder.
Participante 4	Interface em Pares, tendo em vista a menor quantidade de esforço cognitivo exigido
Participante 5	a primeira devido a organização, e a disposição ficar mais clara.
Participante 6	Em pares, gosto da informação ser apresentada aos poucos e que a decisão seja apenas entre dois itens, acho que a ideia de continuar respondendo mesmo após a quantidade mínima fica mais clara no modelo em pares também. Mesmo que esse modelo seja mais lento que o outro, dá a impressão de que eu precisei pensar menos para efetuar as escolhas, acho é uma perda justa, perde no tempo e mas ganha na "facilidade" de uso. Também considerando meu comentário na reunião : o usuário tende a escolher a primeira opção que seja satisfatória, talvez entre os 10 apresentados no modelo lista vc não consiga captar os preferidos e sim apenas os primeiros que ele leu e achou que servem. Provavelmente o usuário não vai ler todos de uma vez só e em seguida avaliar quais são seus preferidos, a ordem mais "normal" seria ler até que o mínimo obrigatório seja atingido e depois não continuar a leitura procurando por algum que agrade mais. Semelhante ao comentário de outro participante sobre a preguiça de ler tudo. E isso pode atrapalhar o resultado final.

é pedir que o usuário avalie alguns itens, tão logo realize seu cadastro. No entanto quantos itens apresentar e a forma de realizar esta apresentação representa um desafio. No presente trabalho a discussão ateu-se a quantos itens mostrar ao mesmo tempo.

No presente trabalho, para que pudesse ser definida qual a melhor interface foi realizado um experimento com usuários a partir de uma reunião online. Com base nas respostas dos usuários participantes, o sistema desenvolvido apresenta no momento de criação do perfil do usuário pares de artigos onde este pode selecionar um, dois ou nenhum item do par. Como no presente trabalho a interface foi definida para um sistema de recomendação de artigos científicos, trabalhos futuros podem envolver a realização de experimentos similares para definição de interfaces em outros domínios envolvendo a recomendação de outros tipos de itens (e.g. filmes, música, etc.).

## **Agradecimentos**

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. Agradecemos o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) através do processo 302959/2023-8 (DT2) e da Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina (FAPESC) N° 48/2022 - Apoio à Infraestrutura para Grupos de Pesquisa da UDESC TO n°2023TR000245.

## Referências

- Abdullah, N. A., Rasheed, R. A., Nizam, M. H., and Rahman, M. M. (2021). Eliciting auxiliary information for cold start user recommendation: A survey. *Applied Sciences (Switzerland)*, 11(20).
- Bettman, J. R., Luce, M. F., and Payne, J. W. (1998). Constructive Consumer Choice Processes. *Journal of Consumer Research*, 25(3):187–217.
- Celma, O., Herrera, P., and Serra, X. (2006). Bridging the music semantic gap. *CEUR Workshop Proceedings*, 187:927–934.
- Celma, Ó., Ramírez, M., and Herrera, P. (2005). Foafing the Music: A music recommendation system based on RSS feeds and user preferences. *ISMIR 2005 - 6th International Conference on Music Information Retrieval*, pages 464–467.
- Cezar, N., Gasparini, I., Lichtnow, D., Lunardi, G., and de Oliveira, J. M. (2024). Exploring strategies to mitigate cold start in recommender systems: A systematic literature mapping. In *Proceedings of the 26th International Conference on Enterprise Information Systems - Volume 1: ICEIS*, pages 965–972. INSTICC, SciTePress.
- Consoni, G. (2014). Recuperação de informação em sistemas de recomendação : análise da interação mediada por computador e dos efeitos da filtragem colaborativa na seleção de itens no website da Amazon.com. *Tese - Programa de Pós Graduação em Comunicação e Informação da Universidade Federal do Rio Grande do Sul*.
- Goldberg, K., Roeder, T., Gupta, D., and Perkins, C. (2001). Eigentaste: A Constant Time Collaborative Filtering Algorithm. *Information Retrieval*, 4(2):133–151.
- Jelassi, M. N., Ben Yahia, S., and Nguifo, E. M. (2013). A personalized recommender system based on users' information in folksonomies. *WWW 2013 Companion - Proceedings of the 22nd International Conference on World Wide Web*, pages 1215–1223.
- Monti, D., Rizzo, G., and Morisio, M. (2021). *A systematic literature review of multicriteria recommender systems*, volume 54. Springer Netherlands.
- Panda, D. K. and Ray, S. (2022). Approaches and algorithms to mitigate cold start problems in recommender systems: a systematic literature review. *Journal of Intelligent Information Systems*, pages 341–366.
- Ricci, F., Rokach, L., and Shapira, B. (2015). *Recommender Systems*.
- Schafer, J. B., Frankowski, D., Herlocker, J., and Sen, S. (2007). *Collaborative Filtering Recommender Systems*, page 291–324. Springer-Verlag, Berlin, Heidelberg.
- Son, L. H. (2016). Dealing with the new user cold-start problem in recommender systems: A comparative review. *Information Systems*, 58:87–104.