

Elaboração de um serviço de Recomendação Híbrido Ponderado e Misto implantado em Webservice RESTful

Herik S. Lorenção¹, Rafael V. M. Santos¹

¹Campus Cachoeiro de Itapemirim – Instituto Federal do Espírito Santo (IFES)
Caixa Postal 527 – 29.300-970 – Cachoeiro de Itapemirim – ES – Brazil

heriksantoslorencao@gmail.com, rafaelv@ifes.edu.br

Abstract. *Recommendation systems aim to suggest items, products or information to their users according to their interests, making their use increasingly widespread, both in the market and in decision-making processes. Based on the context described, this work seeks to provide a hybrid recommendation service implemented on the structure of a textit WebService RESTful, focused on structuring the results obtained in recommendation algorithms based on collaborative, content and hybrid filtering (using the approaches weighted and mixed). A case study was developed based on music evaluations of the most varied genres. The weighted hybrid filtering algorithm obtained the best results with 81.4 % of correctness in the recommendations, as well as the average of the recommendations statistically equal to the average of evaluations, based on the T test.*

Resumo. *Os sistemas de recomendação possuem o objetivo de sugerir aos seus usuário itens, produtos ou informações de acordo com seus interesses, fazendo com que sua utilização seja cada vez mais difundida, tanto no mercado como em processos de tomada de decisão. Com base no contexto descrito, esse trabalho busca disponibilizar um serviço de recomendação híbrido implantado sobre a estrutura de um Webservice RESTful, focado na estruturação dos resultados obtidos em algoritmos de recomendação baseados em filtragem colaborativa, conteúdo e híbrida (utilizando as abordagens ponderada e mista). Foi elaborado um estudo de caso baseado em avaliações de músicas dos mais variados gêneros. O algoritmo de filtragem híbrida ponderada obteve os melhores resultados com 81,4% de acerto nas recomendações, bem como, média das recomendações estatisticamente iguais as médias de avaliações, com base no teste de T.*

1. Introdução

Em um contexto onde a quantidade e disponibilidade de informações vem sendo cada vez maior, as pessoas vem se deparando com uma enorme diversidade de opções [Cazella et al. 2010]. Deste modo, o excesso de informação, muitas vezes, acaba tornando o processo de tomada de decisão mais complexo para o utilizador [Centeno et al. 2018].

Como uma alternativa ao problema levantado, têm-se os sistemas de recomendação, que, segundo [Cazella et al. 2010], podem ser definidos como um sistema que a partir de dados recomendados por usuários, consegue agregar e direcionar a recomendação gerada para os potenciais indivíduos interessados nelas.

Segundo estudos de [Pathak et al. 2010], a utilização de sistemas de recomendação de produtos pode acabar gerando aumento nas vendas e lucratividade em diversos setores como o comércio eletrônico, fazendo com que a utilização desses sistemas gere vantagem competitiva para seus utilizadores. Como exemplo de sucesso nesse setor, pode-se citar a Amazon, que como apresentado nos trabalhos de [Smith and Linden 2017], vem utilizando algoritmos de recomendação há mais de duas décadas.

O uso de sistemas de recomendação também ocorre em diversas outras áreas, como na educação por exemplo, onde a utilização de recomendações pode auxiliar no desenvolvimento de sistemas educacionais mais eficazes [Costa et al. 2013]. Áreas como a da saúde também podem ser beneficiadas, uma vez que a utilização desses sistemas pode auxiliar na identificação de fatores e problemas relacionados a saúde, permitindo a identificação e utilização do perfil e características do paciente para recomendações [Weitzel and Oliveira 2010].

2. Objetivos

Esse trabalho tem como objetivo geral a elaboração de um serviço de recomendação híbrido focado na estruturação dos resultados de algoritmos baseados em filtragem colaborativa e conteúdo, podendo ser utilizado em diferentes domínios de problemas, devido a sua modelagem com proposta mais genérica. Além disso, foram desenvolvidas aplicações clientes para administração dos projetos como também para avaliação dos itens, por parte dos usuários. Para recomendação híbrida, foram selecionadas para desenvolvimento as abordagens ponderada e mista.

Além disso, esse trabalho tem como objetivo continuar os estudos apresentados por [Mendes 2018] e [Bertolaci 2019], buscando integrar os sistemas desenvolvidos nesses projetos e amplificar os resultados gerados. Para avaliação dos resultados, será realizado um estudo de caso baseado em um contexto cultural de recomendação de músicas.

3. Referencial Teórico

3.1. Sistemas de Recomendação

O sistema de recomendação é definido como uma estratégia de tomada de decisão para usuários em ambientes de informação complexos [Rashid et al. 2002]. Desse modo, a utilização de sistemas de recomendação permite que o usuário consiga lidar com grandes quantidades de informação, provendo recomendações personalizadas e exclusivas do conteúdo analisado [Isinkaye et al. 2015].

Para desenvolvimento de um sistema de recomendação, diversos fatores devem ser levados em consideração. De acordo com [Falk 2019], os seguintes componentes podem ser definidos como a estrutura principal de um sistema de recomendação:

- **Domínio:** Refere-se ao tipo de conteúdo recomendado. Com base nesse dado é possível entender, com mais facilidade, o que fazer com as recomendações geradas.
- **Objetivo:** Define qual será o foco e estratégia utilizado em cima da recomendação gerada.

- **Contexto:** Refere-se ao ambiente no qual o consumidor utilizará o sistema de recomendação.
- **Níveis de personalização:** As recomendações podem apresentar diversos níveis de personalização para seu usuário final, sendo que essas personalizações podem ser divididas em três níveis:
 - *Não-personalizada:* apresenta as recomendações de forma padronizada para todos os usuários do sistema.
 - *Semi-personalizada ou dividida em segmento:* nesse modelo, o sistema de recomendação procura agrupar os usuários de forma a gerar recomendações voltadas aos interesses de cada grupo.
 - *Personalizada:* nesta abordagem o sistema busca realizar recomendações de acordo com interações passadas do usuário, gerando resultados exclusivos para cada um deles.
- **Quem recomenda:** Em casos específicos, a opinião de especialistas pode apresentar relevância no momento da construção da recomendação, nestes casos, é necessário que esse item seja levado em consideração no desenvolvimento do sistema.
- **Privacidade e confiabilidade:** A forma como lidamos com os dados informados e gerados dos usuários é outra questão bastante relevante a ser analisada nos sistemas de recomendação, principalmente em contextos onde informações sigilosas e sensíveis são manipuladas. A confiabilidade refere-se ao quanto o usuário confia nas recomendações ao invés de considerá-las como propagandas ou tentativas de manipulação.
- **Interface:** Refere-se aos métodos utilizados para comunicar o usuário ao sistema de recomendação, sendo dividido em *input*: entrada de dados do usuário (de forma explícita ou implícita) e *output*: retorno da recomendação pelo sistema.
- **Algoritmos:** Existem diversos algoritmos que podem ser utilizados na construção de um sistema de recomendação. Esses algoritmos costumam ser divididos em dois grupos: **filtragem colaborativa** e **baseada em conteúdo**, sendo que a escolha do tipo de algoritmo dependerá do tipo de dado utilizado para construção das recomendações. Uma abordagem **híbrida** também pode ser utilizada, mesclando técnicas dos dois grupos apresentados.

3.2. Filtragem Colaborativa

Os algoritmos de filtragem colaborativa têm como objetivo explorar informações a partir das experiências dos usuários para recomendação de itens [Sedhain et al. 2015]. Para isso, esses algoritmos utilizam técnicas como a fatoração matricial como apresentado por [Koren et al. 2009], [Lee et al. 2013] e modelos de vizinhança como definido por [Sarwar et al. 2001], de forma a recomendar itens parecidos para usuários com certo nível de proximidade [Falk 2019].

Um dos problemas ao utilizar-se dessa abordagem é a dependência existente entre a recomendação e os dados dos demais usuários, fazendo com que as recomendações geradas na fase inicial desses sistemas possam apresentar desvios nos resultados gerados [Wei et al. 2017] e [Choi and Han 2010].

Esse modelo de recomendação é bastante conhecido, sendo um dos mais utilizados nos sistemas de recomendação [Alyari and Jafari Navimipour 2018].

3.3. Filtragem baseada em Conteúdo

A filtragem baseada em conteúdo, como definido por [Falk 2019], é um modelo de algoritmo que utiliza de metadados para gerar as recomendações dos itens. Nesses sistemas, é buscada a criação de um perfil para os usuários baseado nos itens do sistema, de forma com que seja possível comparar esse perfil aos itens para geração das recomendações.

Como principal problema dessa abordagem de recomendação têm-se o momento em que um novo usuário acaba de ser criado, uma vez que o sistema não consegue inferir quais são as características dos itens (assuntos ou tags) de maior relevância para esse usuário, reduzindo a taxa de acertos nas recomendações iniciais [Khusro et al. 2016].

3.4. Recomendação Híbrida

Abordagens de recomendação colaborativas e baseadas em conteúdo apresentam pontos positivos e negativos [Falk 2019]. Portanto, com intuito de melhorar a qualidade da recomendação e como forma de mitigar alguns dos problemas identificados, foi proposto o modelo de filtragem híbrida, que busca a partir da combinação das técnicas de filtragem, aumentar o desempenho e precisão dos sistemas de recomendação [Göksedef and Gündüz-Öğüdücü 2010]. Com essa abordagem, busca-se aproveitar os pontos fortes vistos em cada uma das abordagens, como também, nivelar suas fraquezas [Al-Shamri and Bharadwaj 2008].

Diversas estratégias podem ser utilizadas na recomendação híbrida, onde as principais, levando em conta a forma como os componentes serão combinados para gerar as recomendações, são as seguintes [Barbosa 2014]:

- **Ponderada:** nesse modelo de abordagem, as filtragens colaborativa e baseada em conteúdo são aplicadas de forma separada, sendo que, após a geração das recomendações, é realizado um processo de combinação linear, utilizando os resultados gerados.
- **Mista:** nesse modelo as recomendações geradas são mescladas para geração do resultado final. Desse modo, o resultado apresentado ao usuário será uma lista dos dados gerados na recomendação colaborativa e baseada em conteúdo.
- **Combinação sequencial:** nesse modelo temos a criação de um perfil do usuário, em um primeiro momento, a partir da recomendação baseada em conteúdo. A partir dos perfis criados, é realizada a recomendação colaborativa que gerará os resultados finais dessa abordagem.
- **Comutação:** para essa abordagem, é necessário a utilização de algum critério para avaliação, que pode ser definido de acordo com a natureza do sistema. Esse critério é utilizado para comutar ou chavear os resultados obtidos na recomendação colaborativa e baseada em conteúdo, gerando o resultado final.

3.5. Web Services, REST (*Representational State Transfer*) e RESTful

Os *web services* podem ser classificados como softwares fornecidos por uma rede (como a internet, por exemplo). São entidades executáveis que funcionam de maneira modular e independente, sendo acessadas e publicadas em uma rede [Falter et al. 2009].

Para arquitetura de web services pode-se utilizar o modelo REST (*Representational State Transfer*, definido por [Fielding and Taylor 2000], que acabaram defendendo o uso do modelo no protocolo da internet. Essa arquitetura acaba permitindo a implantação de uma comunicação padronizada e semântica entre os componentes do sistema.

4. Metodologia

Para elaboração do sistema proposto, foi utilizado uma arquitetura baseada em sistemas distribuídos, que a partir do protocolo HTTP e RESTful comunica-se entre seus componentes. Cada componente apresenta uma função no sistema, como mostrado abaixo:

- **Backend:**¹ responsável pelo processamento e gerenciamento das recomendações e de todos os recursos existentes no sistema. A aplicação disponibiliza as seguintes abordagens de recomendação: colaborativa, baseada em conteúdo, híbrida ponderada e híbrida mista.
- **Cliente Administrativo:**² aplicação gráfica para acesso e manipulação dos recursos pelo administrador do sistema.
- **Cliente de Recomendação:**³ aplicação gráfica para acesso e manipulação dos recursos pelos avaliadores do sistema.

Para validação das recomendações foi realizado um estudo de caso a partir de uma aplicação cliente de recomendação disponibilizada e divulgada na internet. Como escopo da avaliação, foi utilizado uma lista de dez músicas, previamente selecionadas.

Como forma de realização dos testes de comparação das recomendações, foram retiradas, aleatoriamente, um conjunto de registros das notas avaliadas, sendo possível rodar o algoritmo de recomendação e comparar as notas geradas com as avaliadas.

5. Resultados

Para que a coleta dos dados fosse possível, disponibilizou-se um cliente para avaliação dos itens na internet. A figura 1 exibe a tela inicial para avaliação das músicas.

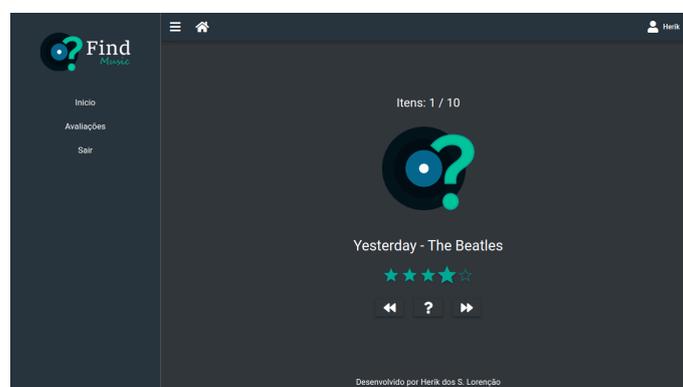


Figura 1. Tela Inicial da aplicação cliente de recomendação

Conforme observado na Figura 1, as avaliações utilizadas pelo sistema foram coletadas a partir de uma seleção de 10 músicas previamente selecionadas, onde os avaliadores poderiam definir uma nota em uma escala entre 0 e 5.

A partir das avaliações coletadas foi possível realizar o estudo de caso voltado para o contexto musical. De maneira geral, os dados de coleta e cadastros no sistema podem ser resumidos na tabela 1:

¹Repositório do código: <https://github.com/herikLorencao/srh-backend>

²Repositório do código: <https://github.com/herikLorencao/srh-client-admin>

³Repositório do código: <https://github.com/herikLorencao/srh-findbymusic>

Itens	Quantidade
Número de avaliadores	60
Número de músicas	10
Número de avaliações	600
Número de tags	6
Número de recomendações	141

Tabela 1. Dados do estudo de caso

5.1. Análise Descritiva

Utilizando-se da análise descritiva foi possível definir a taxa de acerto do sistema sobre as recomendações que seriam aceitas pelos usuários. Nesse processo, foram comparadas as notas que seriam recomendadas pelo sistema de recomendação (notas em uma escala de 0 a 5 que foram superiores à 4) com as avaliações dos usuários (que também apresentaram notas superiores à 4). Vale ressaltar, que somente os valores que seriam definidos como recomendados foram contabilizados, fazendo com que os valores totais mudem de acordo com a abordagem de recomendação.

Com isso, foi possível obter os resultados apresentados na figura 2. Como observado nas células marcadas na cor verde, as abordagens híbridas de recomendação apresentaram resultados superiores as abordagens colaborativa e baseada em conteúdo, quando utilizadas de maneira isolada.

	Colaborativa	Baseada em Conteúdo	Híbrida (Ponderada)	Híbrida (Mista)
SIM	23	49	35	14
NÃO	14	16	8	3
TOTAL	37	65	43	17
% Acertos	62,16	75,38	81,4	82,35

Figura 2. Resultados da Estatística Descritiva

5.2. Teste T de Student

Como outra forma de avaliação estatística das recomendações utilizou-se o Teste t de Student, buscando-se analisar a diferença entre as recomendações calculadas pelo sistema de recomendação e as notas avaliadas pelos avaliadores.

A partir da avaliação do coeficiente **bi-caudal** de todas as médias das abordagens de recomendações comparadas as avaliações dos usuários ⁴, foi possível definir os seguintes resultados apresentados na tabela 2:

Tipo de Recomendação	P(t<=t) bi-caudal	Resumo
Colaborativa	0,606	Médias estatisticamente iguais
Baseada em Conteúdo	0,912	Médias estatisticamente iguais
Híbrida Ponderada	0,827	Médias estatisticamente iguais
Híbrida Mista	0,001	Médias estatisticamente diferentes

Tabela 2. Resultados do Teste T de Student

⁴Para definição da divergência é analisado se o coeficiente bi-caudal apresenta valor inferior a 0,05.

6. Conclusão

O trabalho realizado propiciou o desenvolvimento de um sistema de recomendação híbrido, prosseguindo com os estudos dos trabalhos de [Bertolaci 2019] e [Mendes 2018].

Com o desenvolvimento e disponibilização do sistema de recomendação, juntamente de aplicações para visualização e geração das recomendações, torna-se possível a utilização e melhoria do sistema por qualquer pessoa que possua interesse.

A partir do estudo de caso aplicado, foi possível observar resultados superiores na utilização das abordagens híbridas em relação à colaborativa e baseada em conteúdo. Nesse caso, em específico, é importante visualizar que os dados da abordagem mista divergem estatisticamente das avaliações dos usuários, fazendo com que os resultados obtidos na abordagem ponderada sejam mais confiáveis.

Como trabalhos futuros podem ser adicionados novas abordagens de recomendação, como a combinação sequencial e comutação, por exemplo, possibilitando o estudo e aplicação de novas técnicas. Novos sistemas de consulta e visualização de dados também podem ser desenvolvidos buscando atingir nichos ou plataformas diferentes.

Referências

- Al-Shamri, M. Y. H. and Bharadwaj, K. K. (2008). Fuzzy-genetic approach to recommender systems based on a novel hybrid user model. *Expert systems with applications*, 35(3):1386–1399.
- Alyari, F. and Jafari Navimipour, N. (2018). Recommender systems: a systematic review of the state of the art literature and suggestions for future research. *Kybernetes*, 47(5):985–1017.
- Barbosa, C. E. M. (2014). Estudo de técnicas de filtragem híbrida em sistemas de recomendação de produtos. *Monografia. Centro de Informática, Ciência da Computação, UFPE*.
- Bertolaci, L. M. (2019). Elaboração de um serviço de recomendação com filtragem baseada em conteúdo. Monografia (Bacharel em Informática), IFES (Instituto Federal do Espírito Santo), Cachoeiro de Itapemirim, Brazil.
- Cazella, S. C., Nunes, M., and Reategui, E. (2010). A ciência da opinião: Estado da arte em sistemas de recomendação. *André Ponce de Leon F. de Carvalho; Tomasz Kowaltowski..(Org.). Jornada de Atualização de Informática-JAI*, pages 161–216.
- Centeno, R., Fresno, V., and Chaquet, J. (2018). From textual reviews to individual reputation rankings: Leaving ratings aside solving mpc task. *Expert Systems with Applications*, 114:1–14.
- Choi, S.-M. and Han, Y.-S. (2010). A content recommendation system based on category correlations. In *2010 Fifth International Multi-conference on Computing in the Global Information Technology*, pages 66–70. IEEE.
- Costa, E., Aguiar, J., and Magalhães, J. (2013). Sistemas de recomendação de recursos educacionais: conceitos, técnicas e aplicações. *Jornada de Atualização em Informática na Educação*, 1(1).
- Falk, K. (2019). *Practical Recommender Systems*. Manning Publications.

- Falter, T., Bender, J., Huvar, M., Savchenko, V., and Wiechers, V. (2009). System and method for a web service definition. US Patent 7,620,934.
- Fielding, R. T. and Taylor, R. N. (2000). *Architectural styles and the design of network-based software architectures*, volume 7. University of California, Irvine Irvine.
- Göksedef, M. and Gündüz-Öğüdücü, Ş. (2010). Combination of web page recommender systems. *Expert Systems with Applications*, 37(4):2911–2922.
- Isinkaye, F., Folajimi, Y., and Ojokoh, B. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*, 16(3):261–273.
- Khusro, S., Ali, Z., and Ullah, I. (2016). *Recommender Systems: Issues, Challenges, and Research Opportunities*, pages 1179–1189.
- Koren, Y., Bell, R., and Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, (8):30–37.
- Lee, J., Kim, S., Lebanon, G., and Singer, Y. (2013). Local low-rank matrix approximation. In *International conference on machine learning*, pages 82–90.
- Mendes, T. M. (2018). Elaboração de um serviço de recomendação colaborativa baseada em memória. Monografia (Bacharel em Informática), IFES (Instituto Federal do Espírito Santo), Cachoeiro de Itapemirim, Brazil.
- Pathak, B., Garfinkel, R., Gopal, R. D., Venkatesan, R., and Yin, F. (2010). Empirical analysis of the impact of recommender systems on sales. *Journal of Management Information Systems*, 27(2):159–188.
- Rashid, A. M., Albert, I., Cosley, D., Lam, S. K., McNee, S. M., Konstan, J. A., and Riedl, J. (2002). Getting to know you: learning new user preferences in recommender systems. In *Proceedings of the 7th international conference on Intelligent user interfaces*, pages 127–134. ACM.
- Sarwar, B. M., Karypis, G., Konstan, J. A., Riedl, J., et al. (2001). Item-based collaborative filtering recommendation algorithms. *WWW*, 1:285–295.
- Sedhain, S., Menon, A. K., Sanner, S., and Xie, L. (2015). Autorec: Autoencoders meet collaborative filtering. In *Proceedings of the 24th International Conference on World Wide Web*, pages 111–112. ACM.
- Smith, B. and Linden, G. (2017). Two decades of recommender systems at amazon. com. *Ieee internet computing*, 21(3):12–18.
- Wei, J., He, J., Chen, K., Zhou, Y., and Tang, Z. (2017). Collaborative filtering and deep learning based recommendation system for cold start items. *Expert Systems with Applications*, 69:29–39.
- Weitzel, L. and Oliveira, J. (2010). Sistemas de recomendação de informação em saúde baseado no perfil do usuário. In *Anais do XII Congresso Brasileiro de Informática em Saúde*, pages 18–22.