

Sistema de sugestão de produtos para e-commerce utilizando Inteligência Artificial*

Gabriel C. Ullmann¹, Cristiano A. Künas¹, Leandro P. Heck¹, Edson L. Padoin¹

¹Universidade Regional do Noroeste do Estado do Rio Grande do Sul (UNIJUI)
Santa Rosa – RS – Brasil

{gabriel.cavalheiro, cristiano.kunas, leandro.h}@sou.unijui.edu.br,
padoin@unijui.edu.br

Resumo. *O trabalho trata da implementação de um sistema de recomendação baseado em filtragem colaborativa, utilizando técnicas de Machine Learning e Deep Learning. Esse sistema, diferentemente de outros do tipo, não depende de avaliações de clientes e infere a popularidade dos produtos com base em seu volume e recorrência de venda. Nesse estudo, as recomendações geradas são analisadas a fim de obter insights sobre as lojas virtuais das quais se originaram.*

1. Introdução

Atualmente, vem se tornando popular a utilização de abordagens de Inteligência Artificial (IA) e *Machine Learning* (ML) como meios para, através da análise de hábitos de compra e características de produtos, determinar quais produtos podem interessar a um determinado cliente ou grupo de clientes em serviços de *e-commerce* [Rheude 2019]. Nesse contexto, este trabalho descreve o desenvolvimento de um sistema de recomendação capaz de gerar resultados através da análise dos históricos de compra dos clientes de uma loja virtual, levando em conta um critério para categorização dos mesmos.

O artigo está estruturado da seguinte maneira: a Seção 2 trata dos diferenciais das abordagens de recomendação propostas nesse trabalho em relação a outras similares. A Seção 3 detalha os métodos de avaliação e *datasets*, enquanto a Seção 4 descreve a estrutura de rede neural utilizada. A Seção 5 trata da análise das recomendações produzidas e *insights* em relação aos *datasets*. Finalmente, a Seção 6 traz a conclusão e propostas de trabalhos futuros.

2. Trabalhos Relacionados

No contexto de *e-commerce*, há trabalhos acadêmicos descrevendo diferentes abordagens de geração de recomendações, tais como sistemas de filtragem colaborativa (FC) que sugerem itens "de acordo com o comportamento de usuários similares" [Fressato 2019]. A análise e correlação do histórico de compra de clientes pode ser feito através de técnicas como a fatoração de matrizes, que permite inferir se um usuário se interessa por determinado produto com base em avaliações numéricas dadas por ele e usuários similares [He et al. 2017].

Outros tipos de avaliação podem ser também considerados, tais como resenhas escritas por usuários [Shoja and Tabrizi 2019], críticas do usuário em relação às recomendações geradas [Burke 2002], bem como a relação desses *feedbacks* com sazonalidade e intenção de compra [Hwangbo et al. 2018]. Neste trabalho, em vez de trabalhar com avaliações dadas pelos

*Trabalho desenvolvido com recursos do edital MCTIC/CNPq - Universal 28/2018 sob número 436339/2018-8 e do edital da VRPGPE bolsa PIBIC/UNIJUI.

usuários, as recomendações serão produzidas com base somente na frequência da ocorrência de pedidos para um produto em um determinado segmento de clientes. Isso permite que a geração ocorra mesmo em aplicações com cadastros simplificados ou que não coletam dados sobre a satisfação do consumidor no contexto pós-venda.

3. Proposta

Primeiramente, foram selecionados 3 *datasets* de pedidos: *Online Retail 1 e 2* (ambos provenientes do repositório UCI ML) e *Brazilian E-commerce* (proveniente do Kaggle). Todos os conjuntos escolhidos contêm mais de 100 mil registros; nenhum deles contém avaliações de clientes, somente dados básicos sobre o produto comprado e o comprador. O estado/país dos clientes é utilizado como critério de categorização.

Foram definidos 2 métodos para inferir as avaliações de usuários, ambos baseados na relação entre volume de venda de um produto e o histórico dos compradores: FC Geral e FC por Categoria (Tabela 1). Ambos levam em consideração os 10 produtos mais pedidos em um dado conjunto, porém, enquanto o método por Categoria considera cada estado/país como um agrupamento separado, o método Geral contempla o *dataset* como um todo.

A lógica desses métodos foi representada em *scripts Python*, a fim de avaliar cada registro do conjunto em relação as condições definidas na Tabela 1, designando a este uma nota de 1 a 4. Os conjuntos foram então processados por esse *script*, resultando em um terceiro *dataset* no qual estão relacionados os produtos com as avaliações inferidas para cada cliente.

Os métodos de avaliação foram definidos de forma a dar evidência à recomendação de produtos que já sejam populares com uma parcela dos usuários, mas não ainda com todos (nota 4). Ao mesmo tempo, produtos pouco vendidos ganham também certo destaque, embora secundário (nota 3). Essa estratégia é aplicada como maneira de equilibrar a visibilidade dos produtos, visto que em uma loja com um vasto catálogo os usuários visualizarão apenas uma fração dos itens disponíveis na maior parte do tempo.

Tabela 1. Lógica dos métodos de avaliação

Produto/Cliente	Já comprou	Nunca comprou
É popular na categoria	2	4
Não é popular na categoria	1	3

Finalmente, o conjunto resultante foi utilizado para treinar uma Rede Neural Artificial (RNA). As recomendações geradas por esse modelo foram analisadas no que se refere a sua distribuição de frequência e densidade de probabilidade, de forma a determinar se o processo de avaliação foi consistente, bem como visando extrair *insights* dos *datasets* estudados.

4. Metodologia

Os métodos de avaliação, descritos em *scripts Python*, foram aplicados sobre os registros dos *datasets* e o conjunto de avaliações resultante foi utilizado para alimentar uma RNA criada com a biblioteca *Keras* com base na estrutura proposta por Tanner (2018). Esta é constituída por sete camadas, sendo quatro delas utilizadas para pré-processamento.

A primeira camada é do tipo Input, na qual são definidos hiperparâmetros como dimensões e *batch size*. As entradas são transformadas em *embeddings* na camada seguinte. A

terceira, do tipo Flatten, planifica a matriz de *embeddings*, transformando-a em um vetor. Esse processo é repetido para cada conjunto que deseja-se utilizar como entrada, nesse caso, para códigos de clientes e produtos.

O pré-processamento é concluído por uma camada do tipo Concatenate, que combina os vetores de *embeddings*. O conjunto resultante é então designado como entrada para uma camada do tipo Dense, que por sua vez é seguida por mais duas do mesmo tipo.

Os tempos de treinamento foram muito similares, independente do método de avaliação aplicado. Para o *dataset Online Retail 2*, que possui por volta de 1 milhão de registros, o tempo total foi de aproximadamente 10 minutos utilizando uma CPU Intel Core i5-8300H com *clock* de 2,3 GHz. Também para ambos os métodos, a perda média do modelo para 10 treinamentos com quatro épocas foi de 0,0116.

5. Resultados

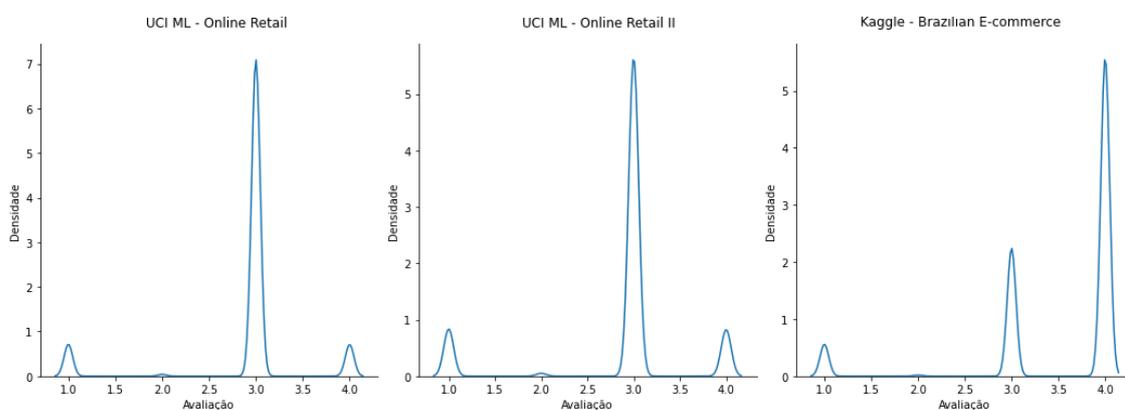


Figura 1. Densidade de avaliações para os 3 *datasets* analisados

Os gráficos da Figura 1 mostram a densidade de avaliações para os três *datasets* selecionados. A altura dos picos representa, não a quantidade absoluta de ocorrências, mas sim a densidade de probabilidade, ou seja, a probabilidade relativa de que um determinado valor na distribuição ocorra no intervalo dado. Embora os gráficos aqui mostrados representem as avaliações geradas através do método de FC Geral, os resultados por categoria seguiram a mesma tendência.

Em todos os *datasets*, as avaliações com valores iguais a 3 e 4 são as mais frequentes. Isso indica que o alto volume de venda dos produtos populares é resultado dos pedidos de uma pequena parcela dos clientes. Portanto, pode-se concluir que em todos os conjuntos a maior parte dos usuários compra com pouca frequência e recorrência. Esse comportamento pode ter diversas causas, como o tipo de produto vendido pela loja, o período ou perfil de clientes considerado. Contudo, há pouca informação que identifique o cliente nos conjuntos e por esse motivo não é possível chegar a uma conclusão quanto à causa da baixa recorrência.

Além disso, a alta incidência de avaliações com valor igual a 3 revela que, assim como os produtos populares, muitos dos produtos não-populares também venderam pouco. Entretanto, diferente do que ocorre para os *datasets Online Retail*, na base *Brazilian E-commerce* há um maior equilíbrio entre as proporções de venda dos produtos populares e não-populares, visto que a diferença de altura entre as curvas não é tão acentuada.

6. Conclusões e trabalhos futuros

Este trabalho apresentou o desenvolvimento de um sistema de recomendação de produtos para *e-commerce* com utilização de técnicas de IA, ML e análise estatística. O estudo alcançou o objetivo de gerar sugestões e permitir a compreensão do comportamento dos consumidores nos *datasets* analisados.

Em trabalhos futuros, pretende-se validar esse modelo através de sua implementação em um *e-commerce* em produção, avaliando as interações de usuários com o sistema e posteriores avaliações dos mesmos quanto ao pedido realizado. Nesse contexto, torna-se possível testar diferentes abordagens para geração de avaliações e recomendações, e sistemas de filtragem baseados em conteúdo podem ser utilizados a fim de complementar ou refinar os resultados produzidos pelos métodos descritos neste estudo.

Referências

- Burke, R. (2002). Interactive critiquing for catalog navigation in e-commerce. *Artificial Intelligence Review*, 18(1):245–267.
- Fressato, E. P. (2019). *Incorporação de metadados semânticos para recomendação no cenário de partida fria*. PhD thesis, Universidade de Sao Paulo.
- He, X., Liao, L., Zhang, H., Nie, L., Hu, X., and Chua, T.-S. (2017). Neural collaborative filtering. In *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee.
- Hwangbo, H., Kim, Y. S., and Cha, K. J. (2018). Recommendation system development for fashion retail e-commerce. *Electronic Commerce Research and Applications*, 28:94–101.
- Rheude, J. (2019). Will e-commerce benefit from machine learning or face a new threat? <https://neilpatel.com/blog/will-e-commerce-benefit-from-machine-learning>.
- Shoja, B. and Tabrizi, N. (2019). Customer reviews analysis with deep neural networks for e-commerce recommender systems. *IEEE Access*, 7:119121–119130.
- Tanner, G. (2018). Keras tutorial #10 - book recommendation system. <https://www.youtube.com/watch?v=4vwNkHFuZBk>.