

UNIVERSIDADE DO VALE DO RIO DOS SINOS - UNISINOS
UNIDADE ACADÊMICA DE EDUCAÇÃO CONTINUADA
ESPECIALIZAÇÃO EM BIG DATA, DATA SCIENCE E DATA ANALYTICS

Cleber B. Pizzato

SISTEMAS DE RECOMENDAÇÃO COM AUXÍLIO DE
PROCESSAMENTO DE IMAGENS DE REDES SOCIAIS

São Leopoldo

2020

UNIVERSIDADE DO VALE DO RIO DOS SINOS - UNISINOS
UNIDADE ACADÊMICA DE EDUCAÇÃO CONTINUADA
ESPECIALIZAÇÃO EM BIG DATA, DATA SCIENCE E DATA ANALYTICS

Cleber B. Pizzato

SISTEMAS DE RECOMENDAÇÃO COM AUXÍLIO DE
PROCESSAMENTO DE IMAGENS DE REDES SOCIAIS

Trabalho de Conclusão de Curso de Especialização apresentado como requisito parcial para obtenção do título de Especialista em BigData, Data Science e Data Analytics, pelo Curso de Especialização em Big Data, Data Science e Data Analytics da Universidade do Vale do Rio dos Sinos - UNISINOS.

Orientador: Prof. Dr. **Sandro José Rigo**

São Leopoldo

2020

Sistemas de Recomendação com Auxílio de Processamento de Imagens de Redes Sociais

Cleber B. Pizzato

¹Universidade do Vale do Rio dos Sinos (UNISINOS)
Av. Unisinos, 950, Bairro Cristo Rei, São Leopoldo, RS – Brasil

Abstract. *Recommendation Systems are of great relevance for systems that want to deliver more value to the user. They provide personalized results by the use of particular algorithms. Among the various recommendation techniques the Image Recommendation Systems have become extremely useful since the visual aspect of the items can influence the user's decision. A recommendation model can be designed to benefit from Social Network data and their interconnection aspects to improve their results. The experiments carried out demonstrate an approach capable of using images from social networks in the construction of a recommendation model and, among other results, finding alternatives to mitigate the cold-start problem.*

Resumo. *Os Sistemas de Recomendação são de grande relevância para sistemas que pretendam entregar mais valor ao utilizador. Eles fornecem resultados personalizados concebidos pela utilização de algoritmos. Dentre as várias técnicas de recomendação os Sistemas de Recomendação de Imagens tem se tornado muito úteis uma vez que o aspecto visual dos itens podem influenciar o usuário. Um modelo de recomendação pode beneficiar-se das Redes Sociais e de seus aspectos de interconexão para aprimorar seus resultados. Os experimentos realizados demonstram uma abordagem capaz de utilizar imagens das redes sociais na construção de um modelo de recomendação e, dentre outros resultados, conseguir alternativas para mitigar o problema cold-start.*

1. Introdução

Sistemas de Recomendação são utilizados atualmente em diversas áreas e com inúmeros formatos de conteúdo, sua aplicação vai desde sua utilização para recomendação de produtos em um *e-commerce* até sugestões de filmes e músicas ou ainda pessoas conhecidas em uma rede social. Por definição sistemas de recomendação podem ser descritos como um conjunto de ferramentas de software e técnicas que apresentam sugestões de itens de interesse para os usuários de um sistema computacional [Ricci et al. 2011].

Os sistemas de recomendação precisam conhecer dados relevantes ao contexto onde estão atuando para poder prover recomendações aos usuários, dessa forma, o aumento exponencial de dados web, boa parte deles provenientes de redes sociais, pode ser visto como um aliado para geração de recomendações. Segundo [Fan e Bifet 2013] o aumento da aceitação dos usuários para com as redes sociais permitindo aos mesmos criar conteúdo de forma rápida e fácil, juntamente com a popularização dos dispositivos móveis que potencializaram a troca de informações em tempo real, são fatores que contribuíram diretamente para o aumento de dados na web e para a realidade atual de *Big Data*.

[Ricci et al. 2011] afirma que um dos principais problemas na área de sistemas de recomendação é conhecido como *cold-start*, estando esse relacionado a incapacidade do sistema fornecer uma recomendação devido a falta de informação que o ele possui sobre um determinado usuário, essencialmente quando esse é um novo usuário dentro de um determinado contexto.

Este trabalho propõe criar um modelo de recomendação através do processamento de imagens de redes sociais, visando utilizar-se do grande volume de dados de imagens para aprimorar recomendações, e dessa forma também auxiliar na solução de problemas *cold-start*. De acordo com [Prando 2016] a utilização de dados de redes sociais na solução desse tipo de problema se mostra como sendo uma possibilidade real. No que diz respeito a possibilidade de utilização de imagens, sendo esses os dados das redes sociais a serem analisados como entrada de informação, percebe-se que a preferência por um produto pode ser influenciada pela aparência visual de uma imagem e por isso os sistemas de recomendação de produtos que estão associados a imagens tornam-se imprescindíveis para ajudar os usuários com recomendações consideradas agradáveis ou úteis para eles [Melo et al. 2016].

Ainda sobre utilização de técnicas para a geração de recomendações [Pathan et al. 2015] conclui que a melhor experiência para um usuário é ter sempre recomendações da forma mais personalizada possível e assim técnicas híbridas na concepção de recomendações parecem atender a esses anseios. É com a utilização de uma técnica híbrida que esse trabalho pretende apresentar como a análise de dados de imagens de redes sociais, através da influência de tendências, pode moldar sistemas de recomendações de produtos e garantir ao usuário uma melhor experiência. Isso baseado no que está em evidência no momento e com a possibilidade de proporcionar melhores resultados aos sistemas de recomendação com a atualização constante do que está em visualização nas redes sociais.

No desenvolvimento do trabalho é utilizado o método de pesquisa *Design Science Research*, pois existe a intenção de resolver um problema do mundo real, bem como realizar uma contribuição científica que tenha relevância. O processo que guiou esse projeto teve início através da revisão bibliográfica de trabalhos relacionados ao problema apresentado, a seleção de artigos para a base da pesquisa, a definição do diferencial para pesquisa aplicada, a contextualização dos conceitos envolvidos, o desenvolvimento do protótipo para resolução do problema proposto e posteriormente a coleta, avaliação e apresentação dos resultados.

Todo desenvolvimento foi guiado para apresentar um modelo de recomendação baseado na análise do processamento de imagens de redes sociais e demonstrar a aplicabilidade do mesmo, comparando resultados, com o intuito de avaliar a influência dos aspectos da recomendação social por imagens no aprimoramento dos sistemas de recomendação e fornecer possibilidades para sanar o problema *cold-start*. Aliado a esse objetivo principal é possível avaliar modelos de processamento e aprendizagem para obtenção de dados de similaridade entre imagens, avaliar comparativamente o modelo com uso de recomendação social frente a outro sistema de recomendação já existente, verificar a implementação de técnicas para extração de dados de imagens das redes sociais e de áreas específicas dessas imagens e indicar possíveis nichos de utilização para um sistema de recomendação com processamento de imagens de redes sociais.

O artigo foi estruturado de forma a apresentar nos primeiros tópicos uma contextualização teórica sobre sistemas de recomendação, mais especificamente sistemas de recomendação de imagens, técnicas de recomendação por similaridade e obtenção de informações de redes sociais. Na sequência ocorre a explicação dos experimentos propostos seguida pela análise de resultados, conclusões e trabalhos futuros.

2. Sistemas de Recomendação

Sistemas de recomendação (SR) são de grande relevância porque geram valor tanto para quem os concebe e os vincula a sistemas de computação, passando a fornecer resultados com maiores níveis de envolvimento dentro de um grande conjunto de itens, como para quem utiliza sistemas e aplicativos e passa por diminuir interações de navegação ou pesquisa através de resultados mais personalizados.

Diminuir o esforço do usuário na busca de produtos ou serviços, principalmente quando sistemas oferecem uma grande quantidade destes, oferecer resultados cada vez mais relevantes atingindo melhores resultados financeiros, por parte das empresas, e maiores níveis de fidelização e satisfação, por parte dos usuários são alguns dos aspectos onde, conforme [Ricci et al. 2011], é possível vislumbrar a contribuição dos sistemas de recomendação.

O conceito de sistemas de recomendação já foi abordado por diversos autores ao longo do tempo e algumas definições são importantes para clarificar o tema. Por exemplo, [Schafer et al. 2001] indica que os SR guiam clientes na complicada tarefa de encontrar produtos relevantes através da utilização do conhecimento disponibilizado ou minerado do comportamento dos clientes. Da mesma forma tem-se o conceito de que os sistemas de recomendação baseiam-se em atributos do utilizador, preferências implícitas e explícitas de um ou mais usuários e características dos itens para sugerir novos elementos de interesse [Schein et al. 2005].

2.1. Visão geral

Os SR, de uma forma geral, podem ser representados por uma matriz que, por sua vez, é composta de linhas que representam os utilizadores de um sistema, colunas representando os itens que participam do processo de recomendação e células contendo um valor de avaliação correspondente à relação usuário x item [Melo et al. 2016, Pereira 2016].

Com base nessa contextualização, a matriz a ser observada na Tabela 1 pode ainda ser representada e explicada por uma notação que de acordo com [Melo et al. 2016] traduz-se em $\mathcal{R} = \{r_{ui} \mid 1 \leq u \leq |\mathcal{U}|, 1 \leq i \leq |\mathcal{I}|\}$ tendo em $|\mathcal{U}|$ e $|\mathcal{I}|$ respectivamente o total de usuários e itens e onde u representa o u -ésimo usuário enquanto i representa o i -ésimo item.

Existem diversas abordagens que tratam de características específicas para a concepção de Sistemas de Recomendação, dentre essas diversas abordagens autores bem conceituados no tema, como [Adomavicius e Tuzhilin 2005, Burke 2007, Ricci et al. 2011], referenciam três delas como sendo as principais:

- **Recomendações Baseadas em Conteúdo:** itens recomendados tem relação direta com a preferência que o utilizador demonstrou anteriormente.

Tabela 1. Matriz de representatividade de um SR.

u \ i	1	2	3	4	5
1	5	3	*	2	3
2	2	1	4	*	4
3	5	3	*	2	1
4	3	*	2	1	*
5	*	*	*	1	3

- **Recomendações Colaborativas:** também nomeada de filtragem colaborativa, recomendam ao usuário itens que outros usuários com preferências semelhantes mostraram interesse no passado.
- **Recomendações Híbridas:** tenta unir os melhores aspectos das duas abordagens anteriores e minimizar potenciais problemas relacionados com a utilização individual das mesmas.

Para além das abordagens mais amplamente discutidas convém mencionar algumas que aparecem nos estudos de [Burke 2002] [Burke 2007], que no caso são:

- **Recomendações Demográficas:** utiliza dados do perfil do utilizador para fornecer recomendações com base em classes demográficas das quais o usuário faz parte.
- **Recomendações com Foco em Utilidade:** através de uma função, que necessita da interação do utilizador no cadastro de preferências, calcula a utilidade de um item para um usuário.
- **Recomendações com Foco em Conhecimento:** utiliza técnicas de sistemas baseados em conhecimento para determinar como um item atende das necessidades específicas de um usuário.

Ainda duas outras formas de conceber SR trazem consigo um envolvimento mais amplo e menos centrado em um único usuário, são elas:

- **Recomendações Sociais:** também conhecida como recomendações baseadas em comunidade, tem se tornado cada vez mais tópico de estudos e conforme [Ricci et al. 2011] é uma maneira de construir recomendações através de preferências de indivíduos conhecidos, mais próximos do utilizador, e que por isso acabam tendo suas influências melhor aceitas do que quando o usuário não conhece a fonte das recomendações.
- **Recomendações Baseadas em Confiança:** mais além do que é premissa das Recomendações Sociais essa abordagem inclui segundo [Prando 2016] que deva existir uma certa relação de confiança explícita, entre os usuários, para que dados sejam utilizados na geração de recomendações visto que usuários preferem itens indicados por pessoas nas quais eles confiam.

2.2. Técnicas e Algoritmos

As técnicas e algoritmos utilizados nos SR estão sempre em constante evolução, podem ser utilizadas em conjunto umas com as outras e na maioria das vezes estão ligadas as características, explicadas na seção 2.1. Para chegar a resultados de recomendações [Adomavicius e Tuzhilin 2005, Ricci et al. 2011] indicam que técnicas de mineração de

dados e aprendizagem de máquina devem ser consideradas como possíveis partes do processo.

[Prando 2016] dentro de sua pesquisa apresenta uma amostragem de algoritmos de aprendizagem de máquina, subdivididos por etapas dentro de um processo de geração de recomendações, conforme pode ser visualizado na tabela 2.

Tabela 2. Algoritmos utilizados em SR, adaptada de [Prando 2016].

Etapa	Método	Objetivo	Algoritmos/Técnicas
Pré-Processamento de Dados - Relacionado com a limpeza, filtro ou transformação, preparando os dados para os algoritmos utilizados na etapa de análise	Medidas de Similaridade	Determina a semelhança entre duas entidades.	Euclidean Distance
			Minkowski Distance
			Mahalanobis
			Similaridade de Cossenos
	Amostragem	Seleciona um conjunto de dados relevantes em uma base de dados considerada grande.	Pearson Correlation
			Random Sampling
			Stratified Sampling
			Without Replacement 80/20 (WR)
	Redução de Dimensionalidade	Reduz a dimensão de um conjunto de dados ou informações.	Leance-one-out (LOO)
			Principal Component Analysis (PCA)
	Denoising	Remove efeitos indesejados nos dados, maximizando a informação contida neles.	Matrix Factorization (MF)
			Singular Value Decomposition (SVD)
Análise de Dados - Relacionado com o processamento e obtenção das informações para o domínio do negócio	Classificação	Determinar corretamente a classe de exemplos ainda não rotulados através de exemplos já rotulados.	Denoising Approach
			Nearest Neighbor Classifier (KNN)
			Árvores de decisão
			Naive Bayes
			Support Vector Machines (SVM)
	Agrupamento	Analisa exemplos e determina grupos semelhantes.	Artificial Neural Networks (ANN)
			k-Means

2.3. Problemas

O problema mais recorrente e conseqüentemente o que é mais abordado pela literatura é conhecido como *cold-start* e conforme [Pereira 2016] ocorre sempre que não é possível estabelecer confiabilidade nas recomendações devido ao fato do não haver histórico suficiente de informações sobre um determinado novo item ou usuário. Também pode ocorrer com uma certa porção dos utilizadores que já estão no sistema, mas por algum motivo não realizaram as avaliações necessárias, não tiveram tempo para persistir essas informações ou por algum motivo se negaram, de alguma forma, a conceder as informações [Ricci et al. 2011].

Tratando mais especificamente de SR de Filtragem Colaborativa um dos possíveis problemas que pode ocorrer está relacionado a quão esparsa é a matriz de utilidade que o sistema tem disponível no momento. [Melo et al. 2016] enfatiza que caso os utilizadores não interajam com o sistema afim de fornecer preferências de forma explícita ou implícita, e principalmente com o crescimento desse sistema e o aumento na quantidade de itens, mais difícil será para o SR chegar a boas recomendações.

A escalabilidade, fortemente relacionada ao exponencial aumento de itens e usuário nos sistemas, independente da característica dos SR pode ser um problema pela capacidade computacional e tempo gastos para chegar a um resultado. Para mitigar

possíveis problemas relacionados a escalabilidade técnicas de agrupamento podem ser empregadas para reduzir o número de similaridades ou ainda técnicas baseadas em modelo, com por exemplo SVD indicado na tabela 2, para diminuir a dimensionalidade dos dados [Yi et al. 2015].

[Menk dos Santos 2015] aborda ainda o possível problema de superespecialização onde usuários com preferências muito centralizadas em uma categoria específica podem sempre vir a receber as mesmas recomendações e o problema de análise de conteúdo limitado, onde ocorre uma possível ausência de características suficientemente abrangentes nos itens. Como possível solução para a superespecialização pode-se sugerir uma certa aleatoriedade coerente nas respostas de recomendações enquanto que para análise de conteúdo limitado deve-se sempre tentar ser eficaz nas informações obtidas para os itens.

2.4. Métricas de Avaliação

A avaliação de SR ocorre muitas vezes de forma *offline*, sem interação de usuários finais e com a ajuda de bases de dados específicas para esse tema a fim de possibilitar uma análise comparativa pós aplicação de um novo método ou algoritmo [Melo et al. 2016, Pereira 2016].

Para [Herlocker et al. 2002] avaliar Sistemas de Recomendação não é trivial porque algoritmos podem ser interferidos pelo conjunto de dados e também pelo objetivo final da avaliação, uma vez que as métricas podem ser divididas em métricas de acurácia na predição e métricas de acurácia de ranqueamento.

2.4.1. Acurácia na Predição

[Burke 2002] explica que para verificar a acurácia das predições \hat{r}_{ui} deve-se dividir as avaliações já conhecidas em dois grupos: um deles é elencado como conjunto de treinamento, representando as avaliações previamente realizadas pelos utilizadores r_{ui} e o outro é o conjunto de teste $|\mathbb{T}|$ onde o SR trabalhará na geração das predições.

A qualidade da predição é obtida através da utilização de métricas que capturam a acurácia dos resultados das recomendações, sendo a mais utilizada *Mean Absolute Error* (MAE) representada pela equação 1. Também pode ser avaliada, conforme [Herlocker et al. 2002, Melo et al. 2016], por *Root Mean Squared Error* (RMSE) que é uma variação da MAE, descrita na equação 2.

$$MAE = \sqrt{\frac{1}{|\mathbb{T}|} \sum_{(u,i) \in \mathbb{T}} (\hat{r}_{ui} - r_{ui})} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{|\mathbb{T}|} \sum_{(u,i) \in \mathbb{T}} (\hat{r}_{ui} - r_{ui})^2} \quad (2)$$

2.4.2. Acurácia de Ranqueamento

Tem o objetivo de apresentar ao utilizador as recomendações que ele considera mais relevante primeiro. Através das preferências do usuário os SR devem ser capazes de ordenar os resultados e assim conseguir verificar se há uma maior interação nos primeiros itens apresentados [Melo et al. 2016].

Conforme abordado por [Pereira 2016] a avaliação das métricas ainda pode ser observada por duas óticas distintas: avaliação do conjunto ou da lista de recomendações.

- Medidas que avaliam o conjunto de recomendações: utiliza das métricas Precisão (P), razão entre itens relevantes recomendados e o total de itens, e Revocação (R), razão entre itens relevantes recomendados e o total de itens relevantes disponíveis [Melo et al. 2016]. Ainda conforme explanação de [Pereira 2016] essa ótica visa a eficácia dos SR ao invés de considerar apenas a acurácia na predição e para isso utiliza-se da fórmula 3 onde F1 é a média ponderada de P e R.

$$F1 = 2 \times \frac{P \times R}{P + R} \quad (3)$$

- Avaliação da lista de recomendações gerada: representa, segundo abordagem realizada por [Pereira 2016] o maior desejo dos utilizadores que é o de verificar que o resultado das recomendações está ordenado com base nos seus interesses, ou seja, itens de maior interesse são recomendados primeiro ao utilizador. [Melo et al. 2016] que cita esse método de avaliação como sendo conhecido também por *Top-N* descreve que *Normalized Discount Cumulative Gain* (NDCG) e *Distance-based Performance Measure* (NDPM) são alternativas que da mesma forma consideram a relevância da ordem da lista na avaliação dos SR.

$$R_u = \sum_i \frac{\max(\hat{r}_{ui} - d, 0)}{2^{(k-1)/(\alpha-1)}} \quad (4)$$

$$R = \frac{\sum_u R_u}{\sum_u R_u^{max}} \quad (5)$$

A equação 4 e consequentemente a equação 5 são segundo [Pereira 2016] formas de chegar a valores para a avaliação da lista de recomendações e se decompostas são representadas por: \hat{r}_{ui} predição calculada pelo SR, d valor *default* neutro de avaliação, k posto do item i na lista ordenada, α especifica que a probabilidade do usuário u visualizar o item da lista com posto igual a α seja de 0.5 e R_u^{max} a equação normalizada pela máxima utilização alcançada.

3. Sistemas de Recomendação de Imagens

As preferências do utilizador podem ser influenciadas pela aparência visual de um item e podem ajudar na recomendação de itens representados por imagens, da mesma forma essas recomendações de imagens podem ser utilizadas nas mais variadas situações [Melo et al. 2016].

Da mesma forma que os SR têm particularidades em cada uma de suas características, conforme abordado na seção 2.1, os Sistemas de Recomendação de Imagens

(SRI) na perspectiva de [Melo et al. 2016] quando baseados em conteúdo focam em características que descrevem uma imagem e em como identificar similaridade entre imagens e quando baseados em filtragem colaborativa usam o comportamento passado dos usuário, conforme matriz de utilidade, sem necessitar de características das imagens.

Os SRI tem ganhado bastante notoriedade e tem influenciado o surgimento de inúmeros estudos relacionados também porque, como destaca [Zhou et al. 2018], áreas como o comércio eletrônico de vestuário e suas funcionalidades para recomendações automáticas de itens tem crescido em larga escala.

Alguns estudos relacionados com Sistemas de Recomendação de Imagens [Jagadeesh et al. 2014, Kalantidis et al. 2013, Liu et al. 2012], e nesse caso também vinculados com recomendações de vestuário, como [Jagadeesh et al. 2014, Liu et al. 2012], propõem realizar a extração completa das características do vestimento a partir de uma única imagem. No caso, [Jagadeesh et al. 2014] em seu estudo construindo um conjunto de dados de imagens categorizadas por *'fashionistas'* e [Kalantidis et al. 2013] construindo algoritmos para segmentar e categorizar itens de vestuário através de relações como a ocasião mais apropriada para determinada peça de roupa.

Ainda sobre trabalhos relacionados que abordaram temas como similaridade no âmbito de vestuário ou recomendação visual podemos elencar um estudo de *Functional Pairwise Interaction Tensor Factorization* (FPITF) realizado por [Hu et al. 2015] que foca em prever melhores correspondências de roupas para determinados trajes completos e um estudo de correspondência personalizada de itens com base em recursos visuais de [McAuley et al. 2015].

De certa forma mais alinhado com os objetivos que levaram a realização desse estudo, [McAuley et al. 2015] considera importante a ideia de não modelar a aparência dos objetos individualmente, mas sim verificar como uma imagem pode influenciar atributos visuais desejáveis em outras, indicando mais do que uma simples semelhança visual, mas um modelo de relacionamento de nível superior entre objetos.

Um SRI que vislumbra modelar preferências visuais humanas, recomendando itens com base no interesse aparente de um utilizador em um outro, deve utilizar-se de características comuns de sistemas de recomendação, porém deve-o fazer com base na utilização das próprias imagens ao invés de metadados [McAuley et al. 2015]. Outra forma de atingir preferências visuais humanas para recomendação de itens seria a utilização de atenção visual humana como foi apresentado por [Melo et al. 2016], sendo que uma das vantagens dessa abordagem seria que os dados da atenção visual podem ser obtidos sem esforço adicional do utilizador ou do responsável pelo sistema uma vez que esse já está em funcionamento.

[Liu et al. 2017] em um estudo que abrange Sistemas de Recomendações de Imagens conclui que um item visual é composto de estilo e categoria e portanto pode-se obter características de estilo de um item eliminando variáveis categóricas correspondentes através de um método, que foi denominado no trabalho como *DeepStyle*, que utiliza de *Convolutional Neural Network* (CNN) juntamente com um *framework Bayesian Personalized Ranking* (BPR) amplamente utilizado para recomendações personalizadas abordado em mais detalhes por [Rendle et al. 2009].

4. Informações de Redes Sociais

Uma das definições aceitas para o termo Rede Social descreve o termo como sendo uma forma de conectar indivíduos por algum determinado vínculo, seja ele de ordem genética, afetiva ou profissional [Albalooshi et al. 2012]. A interação dos indivíduos com redes sociais tem evoluído e se potencializado ao longo do tempo e da evolução tecnológica.

As informações provenientes de redes sociais tem alterado a forma como determinados sistemas são pensados e projetados, [Prando 2016] cita o exponencial aumento de redes como o *Facebook* como um dos mais importantes fatores que alavancaram o atual momento de armazenamento, utilização e notoriedade de grandes volumes de dados, também nomeado como era do *Big Data*.

As redes sociais tem, para além das características de conexão interpessoal citadas anteriormente, forte apelo comercial, de modo que empresas ou influenciadores também acabam por utilizar-se desse recurso para apresentar ideias ou produtos, muitas vezes através da utilização de imagens. Elas são com frequência um meio de ligar autoridades, influenciadores ou especialistas em determinados assuntos com utilizadores, e por isso suas informações podem ser úteis em SR já que [Adomavicius e Tuzhilin 2005] aborda o próprio processo de recomendação como sendo uma extensão do processo social, com influências de vínculos e interações humanas. Da mesma forma [Amatriain 2013] visualiza aspectos positivos na utilização de dados de redes sociais em SR e complementa indicando que essas informações devem ser combinadas técnicas e abordagens tradicionais.

Na área de recomendações voltadas ao vestuário, por exemplo, pessoas estão mais propícias a receber e considerar sugestões de roupas por parte de pessoas que estão fortemente vinculadas ao mundo da moda e são consideradas autoridades no assunto [Zhou et al. 2018].

A utilização de informações de redes sociais em SR pode contribuir para solução de problemas *cold-start*, uma vez que conforme [Amatriain 2013] elas podem ser utilizadas, na seleção de itens mais úteis para novos utilizadores mesmo que em certo momento a recomendação através de preferências de amigos diretos possa ter sua utilidade questionada.

5. Abordagem Proposta

Nessa seção será apresentada a abordagem proposta e o experimento que guiou esse trabalho com o intuito de construir um Sistema de Recomendação que possa utilizar de imagens de redes sociais para aprimorar os seus resultados.

5.1. Modelo de Recomendação Proposto

A visão geral da arquitetura foi segmentada em 3 grandes etapas (Figura 1) com o intuito de facilitar o entendimento do modelo, sua construção e posteriormente a análise.

As etapas foram nomeadas de acordo com a sua ordem de execução dentro do experimento, assim sendo: a Etapa 1 corresponde a criação de um SR de filtragem colaborativa, a Etapa 2 compreende a utilização de imagens das redes sociais e a utilização de algoritmos que realizem o cálculo de similaridade entre as *features* das figuras dos itens e das redes sociais e finalmente última etapa, ou Etapa 3, corresponde a inclusão dos dados

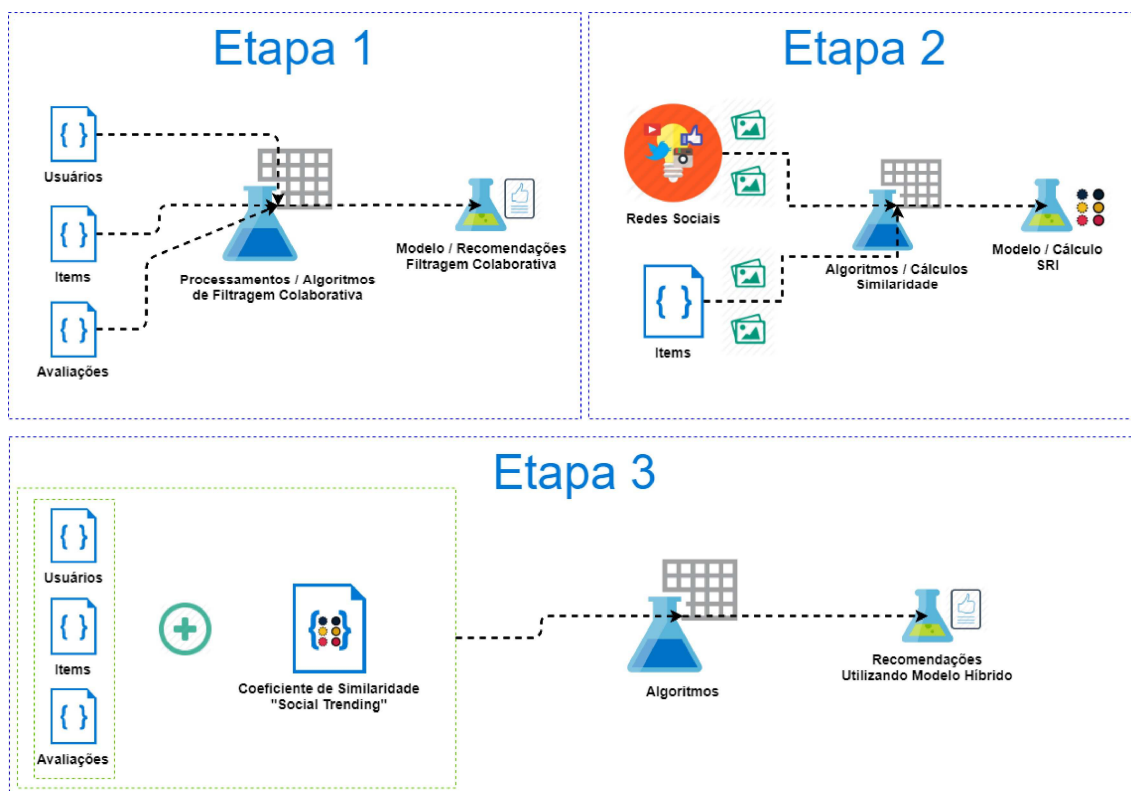


Figura 1. Visão Geral da Abordagem proposta

obtido na Etapa 2 para construção de um SR com características híbridas utilizando-se de tudo que foi construído na primeira etapa e adicionando o fator referente a similaridade das imagens das redes sociais obtido na segunda etapa.

5.1.1. Tecnologias e Ferramentas Utilizadas

A ambiente de trabalho utilizado no desenvolvimento do modelo proposto foi o *Azure Machine Learning* que conforme descrito em [MicrosoftDocs 2020] traz uma área de trabalho personalizada para estudos de *machine learning* e *data science*. Dentre suas funcionalidades destaca-se locais específicos para armazenamento dos conjuntos de dados utilizados, máquinas virtuais configuradas e específicas para essa finalidade e até serviços personalizados para disponibilização e teste dos modelos criados.

No decorrer do trabalho e para realização das três etapas descritas foi utilizada uma máquina virtual da categoria *Standard NC6* com as seguintes especificações: 6 vCPU, 56GiB de memória, 340GiB de armazenamento SSD e 1 GPU de 12GiB de memória dedicada. É importante destacar que as máquinas virtuais criadas dentro do ambiente de trabalho do *Azure Machine Learning* já são concebidas para conter um vasto conjunto de ferramentas para análise de dados e *machine learning*.

Dentro das ferramentas disponíveis na máquina virtual foram utilizadas durante o desenvolvimento: *TensorFlow*, *Python*, *Anaconda*, *JupyterLab*, *Jupyter Notebook Server* e *Azure Storage* para além das outras bibliotecas em *Python* que serão mencionadas nas suas respectivas etapas.

5.2. Obtenção dos Dados

Os dados utilizados no desenvolvimento desse projeto, disponíveis em <http://jmcauley.ucsd.edu/data/amazon/index.html> e elaborados por [He e McAuley 2016, McAuley et al. 2015], fazem parte de um conjunto de informações relativas a metadados de produtos (descrições, informações de categoria, preço, marcas e imagens) e *reviews* (avaliações, comentários) da *Amazon* no período de Maio de 1996 a Julho de 2014. Dentro dos *datasets* disponíveis no repositório foi escolhido o que contém dados referentes a produtos das categorias *Clothing, Shoes and Jewelry*.

As razões pela escolha desse conjunto de dados são principalmente três: a confiabilidade e qualidade do mesmo, já tendo sido esse referenciado em diversos artigos, o fato das imagens referentes aos produtos ainda estarem disponíveis para download possibilitando assim a realização das análises e a relação de influência e visibilidade que itens de vestuário tem nas redes sociais.

Para o melhor entendimento das referências que forem realizadas aos dados é importante caracterizar alguns tipos que serão amplamente utilizados, são eles:

- *overall*: representa o valor decimal da avaliação de um usuário sobre um determinado item;
- *reviewerID*: representa o identificador único do usuário;
- *asin*: identificador único do produto.

Especificamente para atender as necessidades desse estudo e por uma questão de dimensionalidade dos dados e capacidade computacional para realizar o experimento foi utilizado a partir do conjunto de dados original de avaliações dos usuários um subconjunto que compreende as informações relativas somente a produtos das categorias, *Shoes e Fashion Sneakers*, e que será referenciado no restante do estudo como *Dataset de Origem*. A Figura 2 mostra uma descrição estatística da coluna *overall* do *dataset* original comparada com o *subset* de dados que serve de ponto de partida para a análise dos dados.

count	5748920.00000	count	1520846.00000
mean	4.14468	mean	4.22214
std	1.23310	std	1.15419
min	1.00000	min	1.00000
25%	4.00000	25%	4.00000
50%	5.00000	50%	5.00000
75%	5.00000	75%	5.00000
max	5.00000	max	5.00000
Name: overall, dtype: float64		Name: overall, dtype: float64	

Dataset de Referência Dataset de Origem

Figura 2. Dataset de Referência x Dataset de Origem do experimento

5.3. Metodologia Utilizada

A utilização do método *Design Science Research* é adequada para áreas da ciência da informação e conforme [Bax 2017] ainda mais especificamente na subárea de sistemas de informação e tecnologia onde mostra-se capaz de tratar de problemas de pesquisa sem se distanciar da relevância e aplicabilidade ou ainda do rigor imprescindível para geração de novo conhecimento.

A pesquisa desenvolvida é de caráter exploratório, já que no decorrer do trabalho ocorrem a construção de hipóteses buscando criar um modelo de sistema de recomendação satisfatório através da utilização de imagens das redes sociais, e quantitativa com o uso de técnicas de matemática e estatística para apresentação dos resultados obtidos.

Conforme descrito na seção 5.1 e apresentado na Figura 1, após a obtenção dos dados, a realização do estudo compreende a execução de três etapas que devem ser executadas respeitando a ordem descrita afim de conseguir reproduzir o experimento e realizar uma posterior análise dos resultados.

5.3.1. Etapa 1

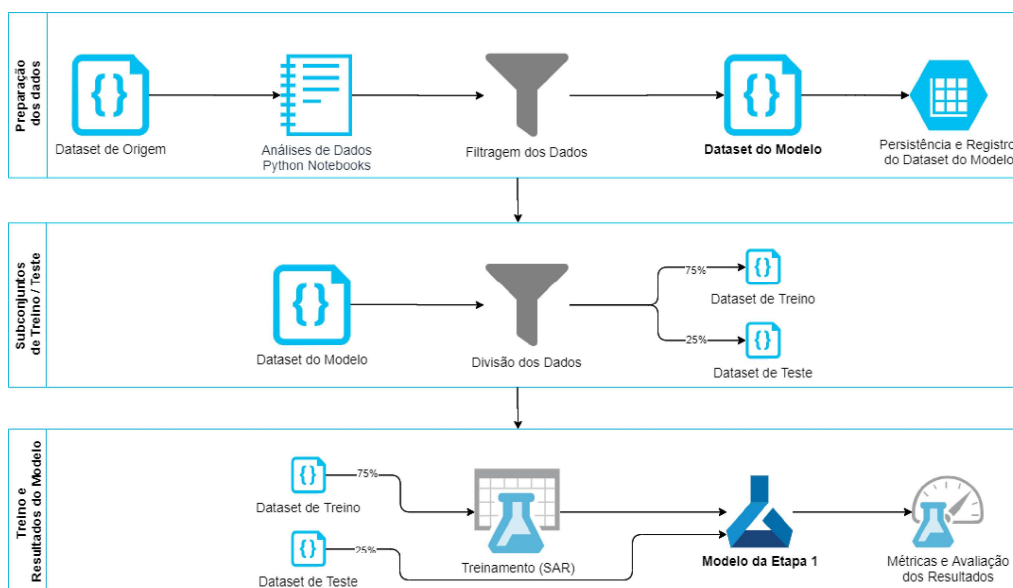


Figura 3. Representação detalhada da Etapa 1

De acordo com o que foi brevemente descrito na seção 5.1 a Etapa 1 visa utilizar um algoritmo de Filtragem Colaborativa para servir de base para o restante do experimento e ser referência em termos de métricas a serem observadas e comparadas. Uma visão detalhada dessa etapa pode ser visualizada na Figura 3.

Na escolha do algoritmo a ser utilizado nessa etapa levou-se em conta o fato de que ele deveria ser baseado em transações ou registros de eventos que pudessem ser incrementados com o objetivo de que dados obtidos posteriormente pudessem evoluir o modelo e gerar novos resultados na etapa final. Assim, após a pesquisa por possíveis abordagens a serem utilizadas, optou-se por utilizar uma implementação da *Microsoft* para Filtragem Colaborativa com um algoritmo chamado *Simple Algorithm for Recommendation (SAR)*. Uma representação visual do conceito utilizado pelo SAR pode ser vista na Figura 4.

Em [Microsoft 2020] o algoritmo SAR é descrito como sendo adaptativo e escalonável, capaz de lidar com transações de usuários para gerar recomendações através da multiplicação de matrizes de similaridade de itens e afinidade dos usuários à esses mesmos itens. A descrição do algoritmo contida em sua documentação, indicando que ele foi

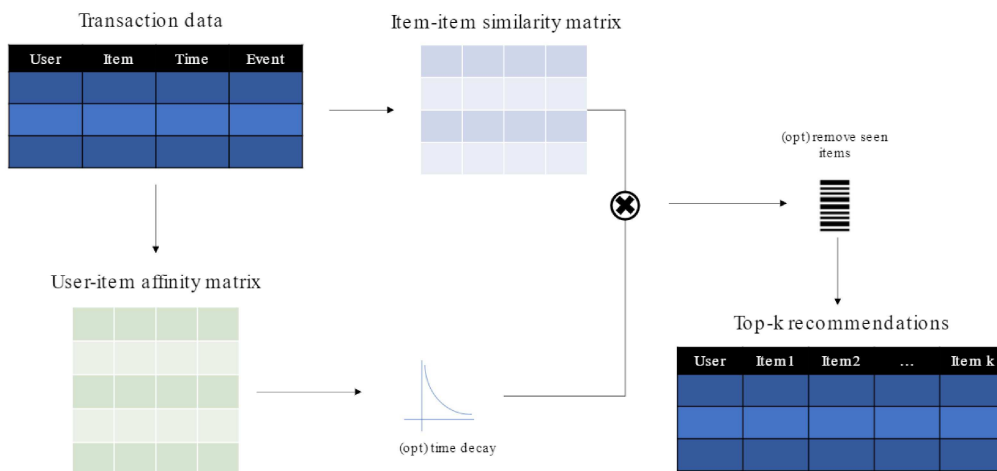


Figura 4. [Microsoft 2020] Representação do algoritmo SAR em alto nível

	overall	reviewerID_total_reviews	asin_total_reviews
count	70185.00000	70185.00000	70185.00000
mean	3.93098	2.47490	41.78319
std	1.17322	1.11547	38.27564
min	1.00000	2.00000	1.00000
25%	3.00000	2.00000	16.00000
50%	4.00000	2.00000	28.00000
75%	5.00000	3.00000	54.00000
max	5.00000	26.00000	214.00000

Figura 5. Dataset de Modelo a ser dividido em subconjunto de treino e teste

pensado para trabalhar com outros eventos para além das avaliações fornecidas para os produtos, foi determinante para a sua escolha, uma vez que nas etapas seguintes técnicas serão utilizadas para extrair da imagens da redes sociais dados que serão incrementados aos *ratings* para concepção do modelo final.

De posse do *Dataset de Origem* uma série de análises foram realizadas aos dados para identificar quais quais seriam os dados que realmente seriam utilizados para obtenção dos subconjuntos de treino e teste do modelo SAR. A análise teve 3 focos principais, sendo eles: a comprovação de que os *ratings* dos itens pelos usuários tinham uma distribuição normal dentro do conjunto de dados, a análise da dimensionalidade e da relação de *reviews* por produto e *reviews* por usuário.

Após a análise, e de acordo com [Microsoft 2020], verificou-se que devido ao SAR gerar recomendações baseadas nas preferências dos usuários, todos que estão no subconjunto de testes também devem pertencer ao subconjunto de treinamento. Para tanto filtragens foram realizadas aos dados e após o tratamento dos mesmos originou-

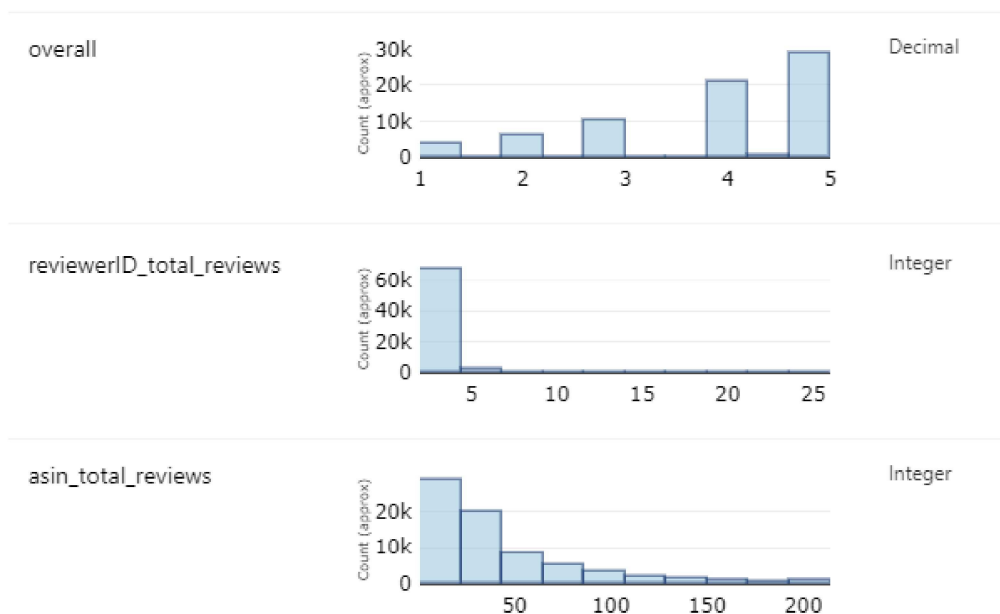


Figura 6. Representação gráfica do Dataset de Modelo

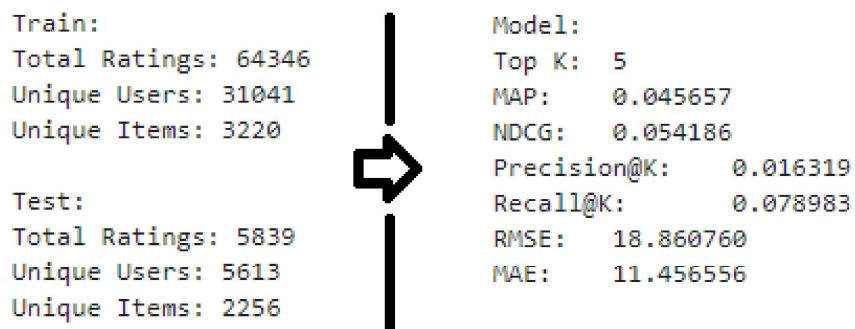


Figura 7. Métricas de avaliação do Dataset de Modelo

se o subconjunto, com características de acordo com a Figura 5 e Figura 6, que passa a ser chamado para fins de estudo de *Dataset do Modelo*.

Com a divisão dos dados em treino (setenta e cinco por cento) e teste (vinte e cinco por cento) procedeu-se com a realização da etapa de treino do modelo onde obteve-se as métricas, abordadas em mais detalhes na seção 2.4, que podem ser visualizadas na Figura 7 e que representam a base de avaliação do experimento proposto. Ainda, segue como referência uma lista de 5 produtos que o modelo recomendou para o usuário *A00475741WVLQL5CVWK2Z* (Figura 8).

5.3.2. Etapa 2

É nessa etapa que é adicionado ao experimento um fator externo as transações de avaliações de itens por usuários, passando a considerar um elemento totalmente novo, e que conforme a seção 4 pode potencializar e influenciar relações de recomendação. O



Figura 8. Recomendações para o usuário A00475741WVLQL5CVWK2Z

método escolhido para correlacionar as imagens do *Dataset do Modelo* com as imagens das redes sociais é a geração de um coeficiente de similaridade de imagens, através de algoritmos e técnicas próprias de SRI, abordadas na seção 3. Para facilitar o entendimento e o acompanhamento, na Figura 9 pode-se identificar visualmente as sub-tarefas da etapa.

Dessa forma, as tarefas necessárias para conclusão dessa etapa iniciam-se por no *Dataset do Modelo* identificar os itens únicos e as imagens associadas a esses itens, nesse caso totalizando 3220 imagens únicas de produtos. Paralelamente a esse processo também realizou-se a obtenção de imagens para cada uma das *hashtags* (*fashionsneakers, sneakers, sneakerspic, sneakernews, shoesshoesshoes, menshoes, womenshoes, shoeslovers, shoeshopping, shoes*) da rede social *Instagram*, levando em consideração publicações consideradas relevantes pelo algoritmo *Top Posts* da própria rede social e nesse caso obtendo um total de 114 imagens.

Com o objetivo de automatizar o processo de obtenção das imagens das redes sociais, evitar interferência manual no processo e facilitar a adição dessa tarefa na *pipeline* do modelo de recomendações foi utilizada uma técnica de *scraper*, através da utilização da biblioteca *Instaloder* disponível em <https://github.com/instaloder/instaloder>.

Posteriormente à obtenção desses dois conjuntos de imagens é possível iniciar a geração do coeficiente de similaridade e para tal optou-se por utilizar a biblioteca *Annoy* (*Approximate Nearest Neighbors Oh Yeah*) e um modelo de CNN pré-treinada para geração de vetores de valores numéricos (*image features*) de todas as imagens, disponível no repositório de modelos do *Tensorflow Hub* e denominado *mobilenet-v2-140-224*.

A *mobilenet-v2-140-224* conforme [Sandler et al. 2018] é uma nova arquitetura

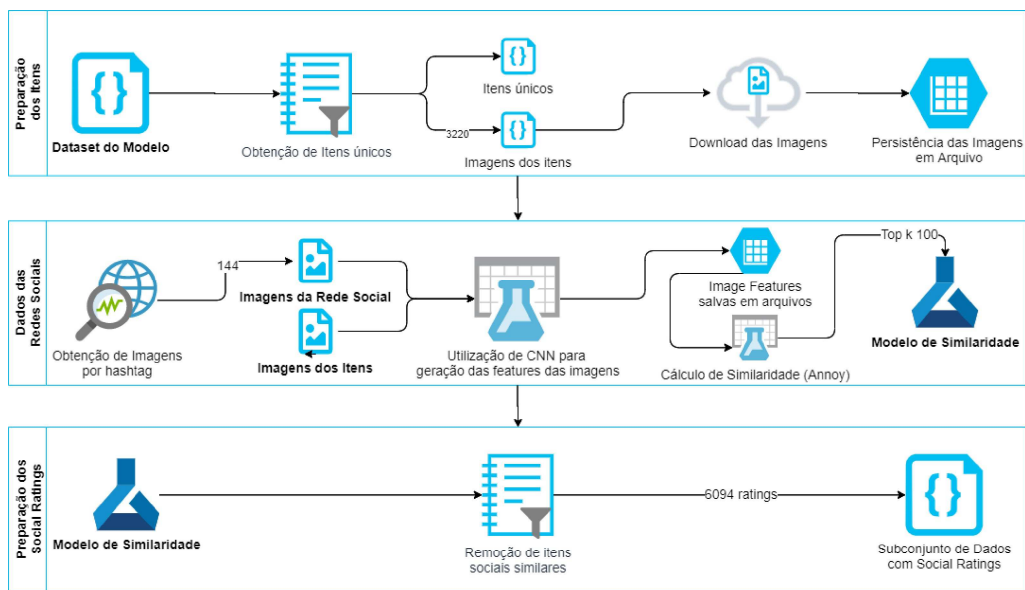


Figura 9. Representação detalhada da Etapa 2

de rede neural adequada para aplicativos móveis e/ou com recursos limitados e, de acordo com o que seus autores demonstram, é caracterizada por conseguir bons resultados no processamento de redes convolucionais com eficiência na utilização de recursos computacionais. Essa arquitetura de rede neural é utilizada para obter as *image features*, tanto dos produtos do conjunto de dados quanto das imagens que foram obtidas através das redes sociais e assim preparar esses vetores de informações para serem processados com o objetivo de obter o coeficiente de similaridade.

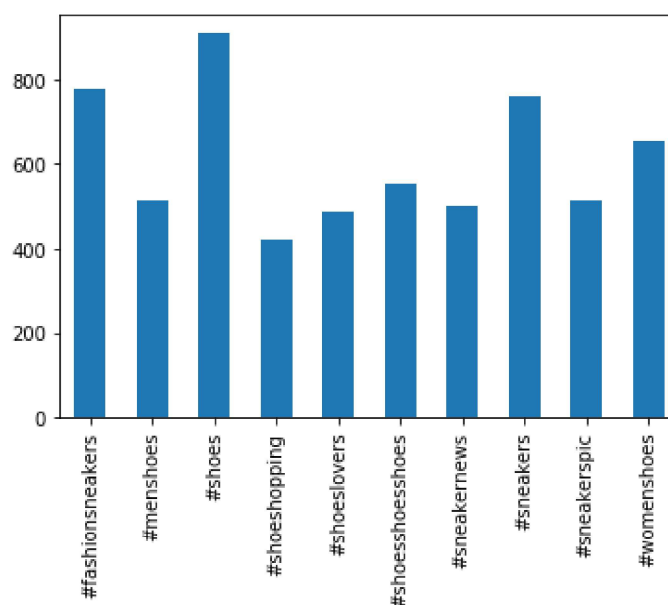


Figura 10. 6094 social ratings por hashtag

Por sua vez a biblioteca *Annoy* é definida por [Spotify 2020] como sendo parte do algoritmo utilizado para gerar recomendações de músicas do *Spotify*, tendo como diferen-

cial a rapidez com que efetua esse processamento de busca pelo vizinho mais próximo e também o fato de permitir o compartilhamento de arquivos estáticos de índices através de diferentes processos.

O algoritmo *Annoy* vai utilizar as informações processadas pelas CNN, através da utilização das imagens, e chegar ao coeficiente de similaridade usando de uma implementação que na sua essência possui conceitos de KNN, um dos algoritmos citados na seção 2 como fazendo parte dos principais algoritmos utilizados em sistemas de recomendações.

Ao final dessa etapa, após a geração dos vetores numéricos que representam as *features* de cada uma das imagens, obtém-se através do algoritmo *Annoy* as 100 imagens mais similares. Juntamente da listagem está o coeficiente de similaridade para cada uma das imagens provenientes das requisições a Rede Social e a referência para o item similar dentro do conjunto total de imagens que engloba as imagens distintas dos itens obtidos na Etapa 1. Da listagem gerada utilizando técnicas de SRI removeu-se os itens similares que apontavam para as próprias imagens do conjunto de dados do *Instagram* e a partir dessa filtragem, representada na Figura 10, chegou-se aos novos 6094 *social ratings*, ou avaliações obtidas através da Rede Social, que serão utilizadas na próxima etapa.

5.3.3. Etapa 3

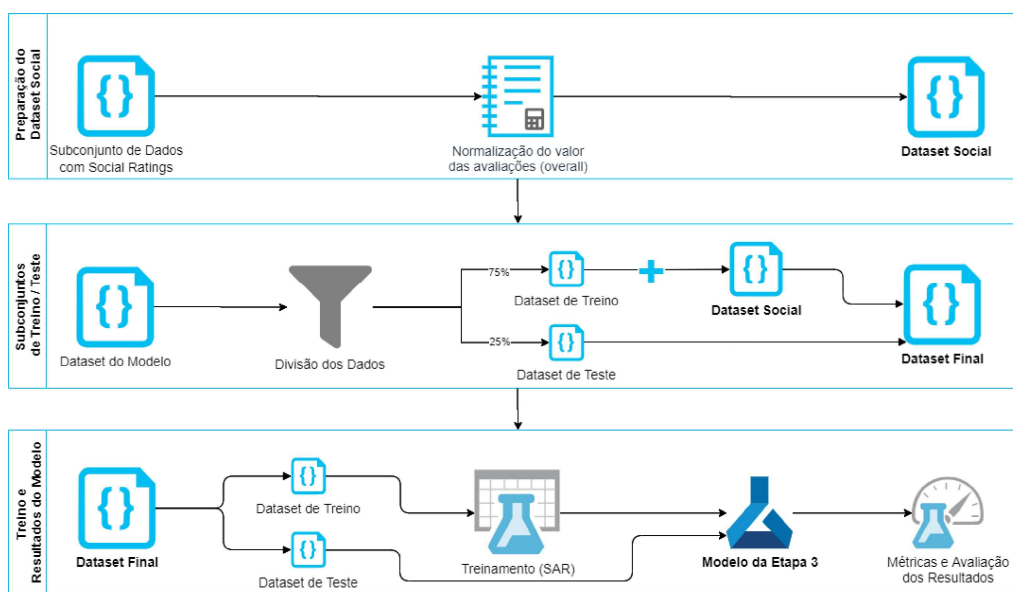


Figura 11. Representação detalhada da Etapa 3

Num primeiro momento, dentro das atividades necessárias para essa etapa (Figura 11) e na preparação do conjunto de dados a ser utilizado nela, multiplica-se o coeficiente de similaridade por 5 com o objetivo de manter a escala da coluna *overall* de 1 a 5, uma vez que o coeficiente de similaridade apresenta valores entre 0 e 1. A Figura 12 mostra a representação desse conjunto de dados denominado *Dataset Social* que será concatenado ao *Dataset do Modelo* para realizar o treinamento de um novo modelo e consequentemente gerar novos resultados.

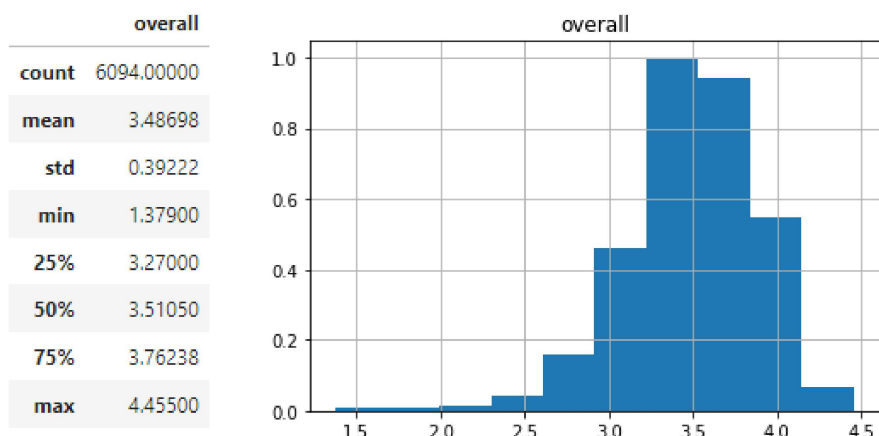


Figura 12. Dataset Social de avaliações de itens

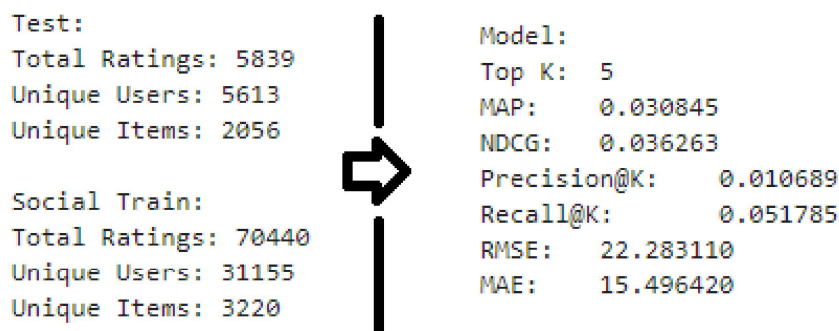


Figura 13. Métricas de avaliação do modelo do Dataset Final

A última etapa do experimento proposto visa treinar um novo modelo SAR, novamente com os dados do *Dataset do Modelo*, porém dessa vez adicionado os dados referentes ao *Dataset Social* na forma de registros de transações adicionais. Nesse caso a coluna *ReviewerID* será preenchida com o identificador único da imagem da rede social, o *asin* continua sendo o identificador único do produto e o *overall* passa a ser o coeficiente de similaridade já com a conversão para a escala das avaliações dos usuários.

O conjunto de dados a ser utilizado nesse novo modelo passa a ser chamado de *Dataset Final*. Para manter o conjunto de teste do modelo com os mesmos itens da Etapa 1 os dados do *Dataset Social* são inseridos apenas no conjunto de treinamento. As métricas de avaliação do novo modelo podem ser visualizadas na Figura 13.

5.4. Análise dos Resultados

Os resultados podem ser comparados, em um primeiro momento, utilizando as métricas obtidas através da avaliação dos modelos da Etapa 1 e da Etapa 3 sendo essas apresentadas respectivamente na Figura 7 e 13. Através das métricas para avaliação da predição e acurácia, detalhadas nas seções 2.4.2 e 2.4.1, podemos perceber que o modelo de recomendação gerado na Etapa 3, com o auxílio de elementos de recomendação social, apresenta suave queda de desempenho em todas as métricas avaliadas se comparado com o modelo da Etapa 1, embora em valores absolutos essa diferença não seja significativa.

Ao avaliar recomendações para o mesmo usuário *A00475741WVLQL5CVWK2Z*, obtidas através do novo modelo, percebe-se que houve alteração de ranqueamento e também modificação de uma das recomendações na lista de cinco, como pode-se visualizar na Figura 14 se comparado com a Figura 8.

A alteração na lista de recomendações para o mesmo usuário é um claro reflexo da adição dos dados das redes sociais obtidos através da Etapa 2. Nesse exemplo o fato de só ter ocorrido uma modificação na lista de cinco, aliado com o que pode ser visto nas métricas indicando uma pequena variação entre um modelo e outro, indica que os dados das redes sociais causaram influência nas recomendações, como era o objetivo, porém sem causar uma mudança radical no que são as recomendações baseadas nas avaliações dos usuários.



Figura 14. Recomendações para o usuário *A00475741WVLQL5CVWK2Z* através do modelo da Etapa 3

6. Conclusões e Trabalhos Futuros

O experimento chegou ao objetivo proposto que era de criar um Modelo de Recomendação Híbrido que utiliza-se de informações de Redes Sociais para complementar um modelo de Filtragem Colaborativa. Embora as métricas não tenham alcançado valores relevantes nos quesitos ranqueamento e predição a abordagem de criação da *pipeline* do modelo de recomendação se mostrou viável e a adição de fatores externos, no caso a similaridade com as imagens obtidas das Redes Sociais, não causou danos impactantes as métricas obtidas no primeiro modelo.

O fato do *Dataset do Modelo* possuir um percentil de 75 por cento de usuários tendo avaliado apenas até 3 produtos, como observa-se na Figura 5, pode explicar parte do desempenho do modelo, uma vez que isso acaba por aumentar a variabilidade de usuários e reduzir o número de avaliações disponíveis para validar o conjunto de teste para o mesmo usuário.

Com o novo modelo de recomendações surge uma nova possibilidade para minimizar ou até eliminar o problema *cold-start* presente na Etapa 1 e abordado na seção 2.3. Para tal pode-se fornecer temporariamente a um novo usuário, que não possui dados dentro do conjunto de dados de avaliações, o *ReviewerID* de um *social rating*. Dessa forma um novo usuário a utilizar o sistema, mesmo antes de realizar qualquer interação com o mesmo, pode receber uma lista de recomendações que terão influência da etapa que considera os aspectos das Redes Sociais. Para fins ilustrativos é exibido na Figura 15 uma demonstração do que seria fornecer para um novo usuário um *ReviewerID*, nesse caso o *shoeshopping-CEsh3VBD4Wv*, obtido através de *social ratings* e referente a imagem obtida do *Instagram* e disponível no endereço <https://www.instagram.com/p/CEsh3VBD4Wv/>.



Figura 15. Recomendações obtidas através de uma das imagens da hashtag *shoeshopping* fazendo uso do modelo da Etapa 3

No decorrer do trabalho identificou-se processos candidatos a melhorias e possibilidades de trabalhos futuros que poderiam ser realizados para expandir a pesquisa e abordar novas opções, são eles:

- Novos algoritmos poderiam ser testados com o intuito de substituir os utilizados atualmente (*SAR*, *mobilenet-v2-140-224* e *Annoy*) nas várias etapas da arquitetura do modelo.
- O experimento poderia ser repetido com conjuntos de dados mais normalizados, com mais avaliações por usuários, ou com um conjunto de dados que já tivesse tido a performance avaliada anteriormente através de outros experimentos com modelos de recomendação.
- Para além das métricas estatísticas usadas no processo de avaliação *offline* poderiam ser realizados estudos para que os próprios usuários voltassem a avaliar as recomendações provenientes do modelo.
- Considerando um experimento onde os usuários avaliassem posteriormente as novas recomendações a etapa de obtenção dos dados sociais poderia ser utilizada para relacionar a pesquisa das imagens sociais com o próprio usuário ou com um conjunto deles. Dessa forma, poderiam ser avaliadas apenas algumas imagens para um determinado conjunto de usuários, por exemplo, determinadas imagens de acordo com a geolocalização dos usuários.

Referências

- Adomavicius, G. e Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge & Data Engineering*, (6):734–749.
- Albalooshi, N., Mavridis, N., e Al-Qirim, N. (2012). A survey on social networks and organization development. In *2012 International Conference on Collaboration Technologies and Systems (CTS)*, pages 539–545. IEEE.
- Amatriain, X. (2013). Mining large streams of user data for personalized recommendations. *ACM SIGKDD Explorations Newsletter*, 14(2):37–48.
- Bax, M. P. (2017). Design science: filosofia da pesquisa em ciência da informação e tecnologia.
- Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, 12(4):331–370.
- Burke, R. (2007). Hybrid web recommender systems. In *The adaptive web*, pages 377–408. Springer.
- Fan, W. e Bifet, A. (2013). Mining big data: current status, and forecast to the future. *ACM SIGKDD Explorations Newsletter*, 14(2):1–5.
- He, R. e McAuley, J. (2016). Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In *proceedings of the 25th international conference on world wide web*, pages 507–517. International World Wide Web Conferences Steering Committee.
- Herlocker, J., Konstan, J. A., e Riedl, J. (2002). An empirical analysis of design choices in neighborhood-based collaborative filtering algorithms. *Information retrieval*, 5(4):287–310.
- Hu, Y., Yi, X., e Davis, L. S. (2015). Collaborative fashion recommendation: A functional tensor factorization approach. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 129–138. ACM.
- Jagadeesh, V., Piramuthu, R., Bhardwaj, A., Di, W., e Sundaresan, N. (2014). Large scale visual recommendations from street fashion images. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1925–1934. ACM.
- Kalantidis, Y., Kennedy, L., e Li, L.-J. (2013). Getting the look: clothing recognition and segmentation for automatic product suggestions in everyday photos. In *Proceedings of the 3rd ACM conference on International conference on multimedia retrieval*, pages 105–112. ACM.
- Liu, Q., Wu, S., e Wang, L. (2017). Deepstyle: Learning user preferences for visual recommendation. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 841–844. ACM.
- Liu, S., Feng, J., Song, Z., Zhang, T., Lu, H., Xu, C., e Yan, S. (2012). Hi, magic closet, tell me what to wear! In *Proceedings of the 20th ACM international conference on Multimedia*, pages 619–628. ACM.

- McAuley, J., Targett, C., Shi, Q., e Van Den Hengel, A. (2015). Image-based recommendations on styles and substitutes. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 43–52. ACM.
- Melo, E. V. d. et al. (2016). Sistema de recomendação de imagens baseado em atenção visual.
- Menk dos Santos, A. (2015). A hybrid recommendation system based on human curiosity. In *Proceedings of the 9th ACM Conference on Recommender Systems*, pages 367–370. ACM.
- Microsoft (2020). Microsoft recommenders. https://github.com/microsoft/recommenders/blob/master/examples/02_model_collaborative_filtering/sar_deep_dive.ipynb.
- MicrosoftDocs (2020). How azure machine learning works: Architecture and concepts. <https://docs.microsoft.com/en-us/azure/machine-learning/concept-azure-machine-learning-architecture>.
- Pathan, S., Panjwani, K., Yadav, N., Lokhande, S., e Thakare, B. (2015). Product recommendations system survey. *International Journal of Computer Applications*, 975:8887.
- Pereira, A. L. V. (2016). *Agrupamento de dados baseado em predições de modelos de regressão: desenvolvimentos e aplicações em sistemas de recomendação*. PhD thesis, Universidade de São Paulo.
- Prando, A. V. (2016). Um sistema de recomendação para ecommerce utilizando redes sociais para solução de cold-start.
- Rendle, S., Freudenthaler, C., Gantner, Z., e Schmidt-Thieme, L. (2009). Bpr: Bayesian personalized ranking from implicit feedback. In *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence*, pages 452–461. AUAI Press.
- Ricci, F., Rokach, L., e Shapira, B. (2011). Introduction to recommender systems handbook. In *Recommender systems handbook*, pages 1–35. Springer.
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., e Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520.
- Schafer, J. B., Konstan, J. A., e Riedl, J. (2001). E-commerce recommendation applications. *Data mining and knowledge discovery*, 5(1-2):115–153.
- Schein, A. I., Popescul, A., Ungar, L. H., e Pennock, D. M. (2005). Croc: a new evaluation criterion for recommender systems. *Electronic Commerce Research*, 5(1):51–74.
- Spotify (2020). Spotify annoy. <https://github.com/spotify/annoy>.
- Yi, M.-Y., Lee, O.-J., e Jung, J. J. (2015). Mbti-based collaborative recommendation system: A case study of webtoon contents. In *ICCASA*, pages 101–110. Springer.
- Zhou, Z., Di, X., Zhou, W., e Zhang, L. (2018). Fashion sensitive clothing recommendation using hierarchical collocation model. In *2018 ACM Multimedia Conference on Multimedia Conference*, pages 1119–1127. ACM.