

Instituto de Pesquisas Tecnológicas do Estado de São Paulo

Fabio Longatto Janes

**Recomendação baseada em imagens e textos de redes
sociais para tratamento do problema de Cold-Start de usuário em E-
commerce**

São Paulo

2019

Fabio Longatto Janes

Recomendação baseada em imagens e textos de redes sociais para tratamento do problema de Cold-Start de usuário em E-commerce

Dissertação de Mestrado apresentada ao Instituto de Pesquisas Tecnológicas do Estado de São Paulo - IPT, como parte dos requisitos para a obtenção do título de Mestre em Engenharia de Computação.

Data da aprovação ____/____/____

Prof. Dr. Marcelo Novaes de Rezende
(Orientador)
Mestrado Engenharia de Computação

Membros da Banca Examinadora:

Prof. Dr. Marcelo Novaes de Rezende (Orientador)
Mestrado Engenharia de Computação

Prof. Dr. Fábio Silva Lopes (Membro)
Mestrado Engenharia de Computação

Prof. Dr. Plínio Roberto Souza Vilela (Membro)
UNICAMP – Universidade Estadual de Campinas

Fabio Longatto Janes

Recomendação baseada em imagens e textos de redes sociais
para tratamento do problema de Cold-Start de usuário em E-commerce

Dissertação de Mestrado apresentado ao Instituto de Pesquisas Tecnológicas do Estado de São Paulo - IPT, como parte dos requisitos para a obtenção do título de Mestre em Engenharia de Computação.

Área de concentração: Engenharia de Software

Orientador: Prof. Dr. Marcelo Novaes de Rezende

São Paulo
Março/2019

Ficha Catalográfica

Elaborada pelo Departamento de Acervo e Informação Tecnológica – DAIT
do Instituto de Pesquisas Tecnológicas do Estado de São Paulo - IPT

J33r **Janes, Fabio Longatto**
Recomendação baseada em imagens e textos de redes sociais para tratamento do problema de Cold-Start de usuário em e-commerce. / Fabio Longatto Janes. São Paulo, 2019.
176p.
Dissertação (Mestrado em Engenharia de Computação) - Instituto de Pesquisas Tecnológicas do Estado de São Paulo. Área de concentração: Engenharia de Software.
Orientador: Prof. Dr. Marcelo Novaes de Rezende
1. Sistema de recomendação 2. Mineração de dados 3. Problema cold-start 4. Redes sociais 5. E-commerce 6. Tese I. Rezende, Marcelo Novaes de, orient. II. IPT. Coordenadoria de Ensino Tecnológico III. Título
19-27 CDU 004.658(043)

DEDICATÓRIA

Agradeço a Deus por todas as conquistas que tem me permitido.

A minha esposa, Luciana, pelo apoio incondicional.

A meus pais, Marli e Claudinei que, com muito carinho e apoio, não mediram esforços para que eu chegasse até esta etapa de minha vida.

AGRADECIMENTOS

A meu orientador Prof. Marcelo Rezende, pelos ensinamentos e incentivo durante todo o processo de desenvolvimento deste trabalho.

Aos membros da banca examinadora dessa pesquisa, o Prof. Dr. Fábio e o Prof. Dr. Plínio, pelas brilhantes considerações que guiaram a confecção final deste trabalho.

Ao meu colega Morgany Mendes, agradeço por todas as discussões e ensinamentos sobre Aprendizado de Máquina, Arquitetura, Big Data e hardware.

A minha tia Claudete, pela colaboração.

A todos os funcionários do IPT, pelo apoio e suporte em todos os momentos.

Aos meus colegas e amigos que de alguma forma contribuíram na realização deste trabalho.

RESUMO

Os sistemas de recomendação têm atraído grande interesse de negócios digitais por proporcionar experiência personalizada aos usuários por meio de sugestões relacionadas às suas preferências. O problema de *cold-start* de usuário, abordado por diversos autores na literatura, é muito comum em sistemas de recomendação. Este problema é caracterizado pela falta de informação (ex.: avaliações de itens, compras efetuadas, marcas favoritas, categorias de interesse) do usuário. A falta de informação causa a diminuição da assertividade das recomendações efetuadas por esses sistemas. Este trabalho apresenta uma abordagem baseada em conteúdo social multimodal com intuito de atenuar o problema de *cold-start*. São selecionados algoritmos de Aprendizado de Máquina utilizados no processamento das recomendações, que são baseadas em informações de imagens e textos provenientes de redes sociais. A validação da proposta se dá pelo atendimento aos requisitos funcionais e não funcionais oriundos da literatura. Entre os requisitos não funcionais, define-se: (1) comportamento em relação ao tempo e (2) utilidade das recomendações, validada por meio das métricas *Root Mean Square Error* (RMSE) para *feedback* explícito e *Click-Through Rate* (CTR) para *feedback* implícito.

Palavras-chave: Sistemas de Recomendação; Mineração de dados em redes sociais; Recomendação Social; *Cold-start* de Usuário; *Deep Learning*.

ABSTRACT

E-commerce recommendation based on images and texts from Social Networks for Cold-start Users

Recommendation systems have attracted great interest from digital businesses by providing user with personalized experience through suggestions related to their preferences. Several authors in the literature have addressed the user cold-start problem, which is very common in recommendation systems. This problem is characterized by the lack of information (e.g., item evaluations, purchases made, favorite brands, and categories of interest) about the user. The lack of information causes a decrease in the assertiveness of the recommendations made by these systems. This work presents an approach based on multimodal social content in order to reduce the cold-start problem. Machine Learning algorithms are used to process the recommendations, which are based on information from images and texts from social networks. The validation of the proposal is based on the fulfillment of functional and non-functional requirements from the literature. Among non-functional requirements are: (1) time-behavior and (2) utility of recommendations, validated using the metrics Root Mean Square Error (RMSE) for explicit feedback and Click-Through Rate (CTR) for implicit feedback.

Key-words: Recommender Systems; Social Network Data Mining; Social Recommendation; User *cold-start*; Deep Learning.

Lista de Abreviaturas e Siglas

ANN	Artificial Neural Networks
API	Application Programming Interface
BOW	Bag of Words
CNN	Convolutional Neural Network
CTR	Click-Through Rate
DBaaS	Database as a Service
GPU	Graphics Processing Unit
IaaS	Infrastructure as a Service
KNN	K-Nearest Neighborhood
LOO	Leave-one-out
MAE	Mean Absolute Error
MF	Matrix Factorization
ML	Machine Learning
NoSQL	Not only SQL
PCA	Principal Component Analysis
RBF	Radial Basis Function
ReLU	Rectified Linear Unit
Rest	Representational State Transfer
RMSE	Root Mean Square Error
SIFT	Scale-Invariant Feature Transform
SR	Sistema de Recomendação
SVD	Singular Value Decomposition
SVM	Support Vector Machine
TF-IDF	Term Frequency / Inverse Document Frequency
URL	Uniform Resource Locator
VSM	Vector Space Model
WR	Without Replacement

SUMÁRIO

1.1	Motivação	11
1.2	Objetivo	12
1.3	Contribuição.....	12
1.3.1	Contribuição Acadêmica.....	12
1.3.2	Contribuição para a Indústria.....	14
1.4	Método de trabalho.....	14
1.4.1	Levantamento e análise das referências bibliográficas	14
1.5	Organização do trabalho.....	17
6	CONCLUSÃO E TRABALHOS FUTUROS.....	18
6.1	Introdução	18
6.2	Avaliação dos resultados.....	18
6.3	Contribuições.....	19
6.4	Conclusões.....	20
6.5	Possíveis limitações.....	20
6.6	Sugestões para trabalhos futuros.....	21
	REFERÊNCIAS.....	22

INTRODUÇÃO

1.1 Motivação

Na era do comércio digital, devido ao crescimento acelerado na utilização de *e-commerces* a tarefa de efetuar recomendações eficientes de produtos para um grande número de usuários se torna cada vez mais complexa. Devido a isso sistemas de recomendação desempenham o papel de realizar recomendações priorizando os produtos de interesse do usuário, minimizando a oferta de produtos possivelmente irrelevantes, maximizando a probabilidade do usuário utilizar a recomendação e consequentemente aumentar a satisfação por parte deste usuário em relação ao *e-commerce*.

Casos de sucesso que representam a importância dos sistemas de recomendação são as recomendações de filme da Netflix, recomendações de música do Pandora e recomendação de produtos da Amazon. Essas recomendações representam grande impacto econômico para estas aplicações de negócio.

No caso específico da Amazon, 20% das vendas são originadas por sistemas de recomendação. Estes sistemas têm sido utilizados com sucesso para resolver o alto volume de informações envolvidas e apoiar as decisões de compra dos usuários (QIAN et al., 2014). Uma pesquisa realizada de forma *on-line* com varejistas do mercado identificou que mais de 50% deles planejam ampliar os recursos de recomendação em seus sites (ENG; HUANG; CHU, 2014).

Um problema bastante comum em sistemas de recomendação caracterizado pela falta de informação (ex.: avaliações de itens, compras efetuadas, marcas favoritas, categorias de interesse) do novo usuário é chamado de *cold-start*. Esta falta de informação é uma preocupação contínua, visto que a entrada de novos usuários acontece a todo instante ocasionando a diminuição da eficiência das recomendações.

Atualmente existem diversas pesquisas relacionadas ao problema de *cold-start* de usuário, são propostas diferentes técnicas e fontes adicionais de informação para mitigar o problema (GONZALEZ CAMACHO; ALVES-SOUZA, 2018). Estas fontes adicionais geralmente são divididas em: informações explícitas (fornecidas pelo usuário) e informações de fontes alternativas. As informações de fontes alternativas podem ser obtidas por exemplo de dados obtidos de redes sociais,

informações de localização do usuário, informações relacionadas ao contexto e comportamento de cliques no *website*.

As recomendações de produtos personalizadas com base em dados sociais do usuário são consideradas uma grande oportunidade de negócio no *e-commerce*, levando em consideração o fato de que as redes sociais estão em constante crescimento. Estima-se que o número de usuários ativos de mídia social no mundo alcance cerca de 3,02 bilhões até 2021 (ANASTACIO, 2018).

Grande parte dos *e-commerces* já propiciam a funcionalidade de login social, os quais possibilitam o acesso aos dados sociais dos usuários como: dados demográficos, fotos, vídeos, postagens, compartilhamentos e *likes*.

As redes sociais Facebook e Twitter, foco do presente trabalho, apresentam enorme crescimento na última década sendo o Facebook a rede social mais utilizada na atualidade (QUBA, 2015). Atualmente existem mais de dois bilhões de usuários conectados a pelo menos uma rede social e estima-se que são criadas quatro milhões de postagens por minuto (OLIVEIRA et al., 2016). A média diária de compartilhamento de fotos no Instagram é 60 milhões (YOU; BHATIA; LUO, 2016).

A utilização conjunta de textos e imagens provenientes de diferentes elementos das redes sociais pode agregar mais informações para identificar o interesse do usuário e melhorar a qualidade da recomendação o que resultaria em recomendações mais eficientes, as quais beneficiam ambas as partes, o usuário e o *e-commerce*.

1.2 Objetivo

O objetivo do trabalho é apresentar uma abordagem alternativa para o tratamento do problema de *cold-start* de usuário. A abordagem parte da mineração de dados multimodais (textos e imagens) provenientes de redes sociais e realiza recomendações com base em técnicas e algoritmos de *Machine Learning* (ML).

1.3 Contribuição

1.3.1 Contribuição Acadêmica

Diversos trabalhos fazem uso de dados sociais para efetuar recomendações (QUBA, 2015) (PRANDO, 2016) (ZHANG; PENNACCHIOTTI, 2013a e 2013b),

porém foram encontradas poucas pesquisas e experimentos que evidenciam resultados sobre a utilização conjunta de textos e imagens compartilhados em redes sociais (OLIVEIRA et al., 2016), (CINAR; ZOGHBI; MOENS, 2015) para inferir preferências do usuário ou efetuar recomendações. Desta forma, a principal contribuição deste trabalho é avaliar o uso de textos e imagens obtidas de redes sociais combinados como uma alternativa para tratamento do problema de *cold-start*. É investigada a extração e associação de diferentes elementos das redes sociais para identificar o interesse do usuário por produtos de um *e-commerce* utilizando as redes sociais Facebook e Twitter.

O presente trabalho estende os seguintes estudos:

- 1) Prando (2016) - Um Sistema de Recomendação para *E-commerce* Utilizando Redes Sociais para solução de *cold-start*

O estudo apresenta uma alternativa para tratamento do problema de *cold-start* com o uso de ações diretas de usuário na rede social (ex.: postagem, compartilhamento e *likes* em conteúdo). O autor se aprofunda no uso de tecnologias Big Data, para efetuar processamento das recomendações com eficiência para *e-commerces*, levando em consideração a grande quantidade de produtos e dados presentes nas redes sociais. São utilizadas técnicas de mineração de dados e mineração de texto.

É utilizado como base o processo de recomendação, arquitetura e tecnologias propostas pelo autor.

- 2) Cinar, Zoghbi e Moens (2015) - *Inferring User Interests on Social Media from Text and Images*

No estudo são utilizados *pins* (posts) obtidos da rede social Pinterest para inferir as preferências do usuário por categorias de interesse existentes na rede social. O autor apresenta uma visão geral comparando diferentes técnicas de ML utilizando somente texto, somente imagens e a combinação de texto e imagens. A combinação de dados de natureza multimodal apresentou os melhores resultados na inferência da preferência do usuário.

Neste trabalho são utilizadas técnicas apresentadas pelo autor para pré-processamento e representação de dados de natureza multimodal e junção de diferentes formas de representação.

1.3.2 Contribuição para a Indústria

A abordagem pode ser utilizada como alternativa para o tratamento do problema de *cold-start* de usuário e possivelmente obter um interesse mais personalizado de novos usuários utilizando a combinação de dados multimodais originados a partir de diferentes ações diretas do usuário efetuadas em redes sociais. A implementação é facilitada para casos em que o *e-commerce* já possua a funcionalidade de *login* social e permissões para acesso aos dados do usuário (ex.: postagens, compartilhamentos, *likes* em conteúdo).

1.4 Método de trabalho

O método de trabalho é composto pelas seguintes macro atividades:

1.4.1 Levantamento e análise das referências bibliográficas

O trabalho foi iniciado com levantamento das principais referências bibliográficas relacionadas ao processo de mineração de dados de redes sociais aplicados em SRs (Sistemas de Recomendação).

Os autores Li et al. (2015), Zhang e Pennacchiotti (2013a, 2013b), Isinkaye, Forlajimi e Ojokoh (2015) e Reshma, Ambikesh e Thilagam (2016) abordam o problema de *cold-start* de usuário ocasionado pela falta de informações sobre o usuário.

Em seguida levantam-se hipóteses com foco na alternativa para tratamento do problema de *cold-start* utilizando a combinação de dados de natureza multimodal compartilhados pelo usuário em redes sociais (postagem, compartilhamento de imagem e *likes* em conteúdo). Para isso, são utilizadas as pesquisas de Cinar, Zoghbi e Moens (2015) e Oliveira (2016).

Na sequência é utilizada a pesquisa de Prando (2016) como base para definição de tecnologias, arquitetura e processo de recomendação.

Posteriormente é efetuado o levantamento e análise de técnicas e algoritmos de ML dentro do processo de mineração de dados, textos e imagens de redes sociais. Para isso são utilizados os trabalhos de Amatriain et al. (2011), Zhang e Pennacchiotti (2013a, 2013b). Também são utilizados e os trabalhos de Hu e Liu (2012), Chen, Yang e Yang (2017), Cinar et al. (2015) e Oliveira (2016). A finalidade da análise é selecionar as técnicas e algoritmo de ML que devem ser utilizados no sistema a ser construído.

1) Análise de características das redes sociais

Neste passo, foram analisadas as principais redes sociais levando em consideração as seguintes características (PRANDO, 2016): (1) propósito geral (QUBA, 2015); (2) quantidade de usuários (JIN et al., 2013) (QUBA, 2015); (3) acesso aos dados por meio de *Application Programming Interfaces* (APIs) (QUBA, 2015) (PRANDO, 2016). Foram adotadas as redes sociais Facebook e Twitter, pois se adequam às necessidades do trabalho.

2) Levantamento de dados

A amostra de dados selecionada se divide em: (1) um conjunto de dados coletados de um *e-commerce* do tipo genérico que comercializa produtos de diferentes categorias e subcategorias. Foi feita a coleta de dados por meio de chamadas a APIs disponibilizadas pelo *e-commerce* e de um *webcrawler*, programa que navega pela web de forma automatizada; (2) um conjunto de avaliações (*feedback*) de usuários para os produtos a serem recomendados pelo SR obtidas por meio do formulário disponibilizado na web que é apresentado ao usuário após efetuar o *login*; (3) um conjunto de dados extraídos por meio de chamadas as API das redes sociais Facebook e Twitter, estes dados representam interações do usuário nestas redes sociais. As permissões para acesso a estes dados, no caso do Facebook, são obtidas por meio de solicitação de acesso às informações durante o *login* social do usuário.

3) Elaboração da Proposta

A partir da análise dos conceitos e trabalhos relacionados, foi desenvolvida a proposta deste trabalho. A proposta é composta principalmente por (A) método de mineração de dados sociais adaptado para conteúdo multimodal e (B) processo de recomendação adaptado para conteúdo multimodal e análise de sentimentos.

4) Implementação da aplicação social

Neste passo, foi construída a aplicação do Facebook que solicita ao usuário autorização para acesso a postagens, compartilhamentos e *likes* em conteúdo. Para o Twitter, não é necessário desenvolvimento de uma aplicação, pois a rede social não restringe acesso a postagens públicas do usuário, porém deve ser criada uma

aplicação no portal de desenvolvedor da rede social para obter acesso as chaves para utilização das APIs do Twitter.

5) Implementação de Referência do SR (Modelagem)

A finalidade deste passo é apresentar os detalhes em relação ao escopo, funções e requisitos do sistema de acordo com a abordagem proposta. São utilizados como base os estudos de Prando (2016) e Amatriain (2012).

6) Implementação de Referência do SR (Construção)

Neste passo, são apresentados os passos seguidos na implementação referencial. São detalhados os dados utilizados e obtenção destes dados. Além disso, são apresentados detalhes de implementação das tarefas do processo de recomendação proposto, algoritmos selecionados e detalhes da comparação destes algoritmos.

7) Avaliação das Recomendações

No passo de avaliação das recomendações, são utilizadas a avaliação *online* - utiliza opiniões de usuários reais - conforme proposta de Quba (2015). Para avaliação do *feedback* explícito do usuário é utilizada a medida *Root Mean Square Error* (RMSE), também utilizada nas propostas de Zhang e Pennacchiotti (2013a, 2013b) e Prando (2016). Para avaliação de *feedback* implícito é levada em consideração a taxa *Click-Through Rate* (CTR), conforme apresentado por Quba (2015).

8) Ambiente de Testes

Para o passo de testes do SR, é utilizado hardware virtualizado no modelo Infraestrutura como Serviço (IaaS) e banco de dados no modelo Banco de Dados como Serviço (DBaaS). O SR é disponibilizado na web e deve permitir o login social, apresentar os itens recomendados e solicitar a avaliação destas recomendações.

9) Análise de Resultados

Neste último passo, é efetuada a validação da proposta de acordo com os requisitos funcionais e não funcionais provenientes da literatura. Para avaliação das recomendações, são utilizados os resultados obtidos por meio de (1) *feedback* explícito dos usuários que permitiram acesso aos dados de sua conta no Facebook

e (2) *feedback* implícito de usuários do Twitter selecionados para receber recomendações. A análise dos resultados tem como objetivo apresentar o atendimento aos requisitos do SR proposto. A precisão da recomendação é obtida por meio da comparação das recomendações efetuadas com o *feedback* do usuário.

1.5 Organização do trabalho

A Seção 2, **Revisão Bibliográfica**, está dividida em conceitos básicos e estado da arte. Nos conceitos básicos, é efetuada a apresentação de conceitos relacionados SRs e problemas identificados. Em seguida são apresentados métodos relacionados a representação multimodal (texto e imagem), conceitos sobre redes sociais, análise de sentimentos e técnicas e algoritmos de ML utilizados no processamento das recomendações. No estado da arte, são apresentados os principais estudos relacionados que utilizam dados de redes sociais em sistemas de recomendação, bem como os eventuais vínculos com a abordagem proposta.

A Seção 3, **Proposta de abordagem baseada em conteúdo social multimodal**, apresenta detalhes sobre a abordagem proposta. São apresentadas e justificadas as técnicas empregadas na análise dos dados sociais e algoritmos de ML utilizados no processamento das recomendações relacionadas a abordagem proposta.

A Seção 4, **Implementação de Referência**, desenvolve a aplicação da abordagem proposta por meio da aplicação referencial. São apresentados detalhes do escopo, funções, requisitos, amostras de dados selecionadas, detalhes da implementação, tipos e métricas de avaliação e ambiente utilizados.

A Seção 5, **Análise dos Resultados**, apresenta a validação dos requisitos e a avaliação do SR. As recomendações são avaliadas de acordo com o critério de utilidade das recomendações por meio das métricas: (1) *Root Mean Square Error* (RMSE) para *feedback* explícito do usuário para recomendações efetuadas na rede social Facebook e (2) *Click-Through Rate* (CTR) para *feedback* implícito do usuário para recomendações efetuadas na rede social Twitter.

Na Seção 6, **Conclusão**, apresenta a análise final do trabalho, descreve limitações do trabalho e sugestões para trabalhos futuros.

6 CONCLUSÃO E TRABALHOS FUTUROS

6.1 Introdução

Nesta seção é apresentada a análise final do trabalho contendo o resumo e avaliação dos principais resultados obtidos, contribuições, possíveis limitações e sugestões para trabalhos futuros.

A proposta deste trabalho é composta por uma abordagem a qual parte da mineração de dados multimodais (textos e imagens) provenientes de redes sociais com base em técnicas e algoritmos de ML com intuito de atenuar o problema de *cold-start*. Houve um cuidado para tornar a proposta genérica possibilitando o uso de diferentes redes sociais.

Na Seção 3, foi apresentada a abordagem proposta com detalhamento do método de mineração de dados sociais adaptado para conteúdo multimodal e do processo de recomendação adaptado para conteúdo multimodal e análise de sentimentos.

Para mostrar a abordagem de forma detalhada e prática foi efetuada uma implementação referencial, na qual foram detalhados escopo, requisitos, dados utilizados e a sequência das atividades.

Em seguida foi efetuada a análise dos resultados, apresentando a validação dos requisitos previamente definidos na subseção **Erro! Fonte de referência não encontrada.**, indicando que o uso de dados sociais de natureza multimodal se mostra propício no contexto de SRs.

Outra alternativa para avaliar a abordagem proposta seria realizar a comparação do comportamento de compra dos usuários e dados obtidos de redes sociais, porém para ter acesso a estes dados, seria necessário a utilização de um *e-commerce* real, o que não foi aplicado neste trabalho. Contudo, fica como proposta para trabalhos futuros.

6.2 Avaliação dos resultados

A implementação referencial da abordagem proposta apresentou um resultado similar ao apresentado no estudo de Prando (2016) utilizando a métrica RMSE e resultado inferior ao apresentado por Quba (2015) utilizando a métrica CTR.

O valor do RMSE obtido foi de 1.76. Este valor é considerado um bom resultado, levando em consideração a utilização de somente dados não estruturados de natureza multimodal (textos e imagens) extraídos de alguns dos elementos presentes nas redes sociais (postagem, compartilhamento de imagem e *likes* em conteúdo) para identificar o interesse do usuário *cold-start*.

A relação entre notas do sistema e avaliações dos usuários apresentam resultados razoáveis. Apesar de a correspondência exata entre as notas do sistema e avaliações dos usuários ser igual a 21,69%, os valores menores ou igual a 1 somam 59,84%.

Foi apresentado um bom índice (60%) de interesse positivo dos usuários pelas recomendações por meio do agrupamento das avaliações dos usuários. Não foi possível efetuar a comparação com o estudo de Prando (2016) de acordo com o interesse positivo do usuário recomendados, pois o estudo de Prando (2016) não apresentou estas informações.

O valor do CTR obtido foi de 3,08%. Apesar do valor ser menor que o valor apresentado por Quba (2016), o valor é um pouco maior que o CTR do Twitter, 2,8%. Uma possível justificativa para os melhores resultados obtidos por Quba (2106) é a utilização de “robôs” que simulam pessoas com diferentes personalidades e o uso de frases como “super oferta” nas recomendações.

O comportamento em relação ao tempo foi de 134 segundos. Apesar de ser maior que o esperado (90 segundos), o valor pode ser considerado razoável, levando em consideração as máquinas utilizadas (sem GPU), a utilização de bibliotecas (ex.: Tensorflow) sem efetuar a compilação para o hardware específico da máquina, o que traria o benefício das otimizações efetuadas na biblioteca pelos fabricantes de hardware.

Assim, por meio da implementação referencial, considera-se que a abordagem proposta apresenta razoável assertividade e se apresenta como uma boa abordagem para um cenário de *cold-start* de usuário e que pode ser aperfeiçoada com o uso de outros elementos da rede social.

6.3 Contribuições

- Evidenciou-se o uso de dados multimodais oriundos de redes sociais para minimizar o problema *cold-start*, explorando técnicas de mineração de texto, mineração de imagens e análise de sentimentos.

- Analisou-se, utilizando dados reais, os classificadores SVM, *Naïve Bayes*, Árvores de decisão e ANN com base nas métricas *F1 score*, *precision* e *recall* nas tarefas de classificação de texto em classes de produtos e classificação de polaridade do texto.
- Foi disponibilizado o código fonte da implementação de referência da abordagem permitindo que trabalhos futuros façam uso do código.

6.4 Conclusões

Os resultados obtidos por meio da implementação de referência, mostram a viabilidade da utilização da abordagem proposta. O uso de diferentes elementos das redes sociais de natureza multimodal podem contribuir na formação do perfil do usuário *cold-start* e conseqüentemente atenuar o problema de *cold-start* de usuário na recomendação de produtos.

A abordagem proposta, em princípio, pode ser utilizada por diferentes tipos de *e-commerces* mesmo que estes já possuam SR, com alteração somente no processamento de recomendações para cenários de usuários *cold-start*.

Em conseqüência da proposta ser baseada em conteúdo, também pode ser utilizada com o intuito de minimizar o problema de *cold-start* de itens

6.5 Possíveis limitações

- Os *datasets*, em relação ao tamanho e tipo dos dados utilizados na implementação referencial configuram uma limitação, já que foram utilizadas 13 categorias do produto, 25 subcategorias do produto e 27 categorias de terceiro nível, 162 pessoas participaram da avaliação das opiniões implícitas do usuário e 64 pessoas participaram da avaliação das opiniões explícitas do usuário.
- A aquisição e uso de diferentes elementos das redes sociais para um sistema de produção pode ser mais difícil, dependendo de questões como: comprovação da empresa, aprovação da rede social, necessidade de estar de acordo com as regulações de proteção de dados do país ou região e restrições dos usuários em conceder permissão de acesso aos dados sociais.
- Sensibilidade da proposta no que se refere aos parâmetros ajustáveis, pois este ajuste é dado de acordo com o contexto da aplicação, devendo ser levada em consideração as formas de representação utilizadas.

- A implementação referencial apresentou um RMSE bom de acordo com a escala na interpretação dos resultados. Porém, é arriscado supor que os resultados obtidos nesta proposta podem ser generalizados em outros contextos de acordo com tipos de *e-commerce* e redes sociais.

6.6 Sugestões para trabalhos futuros

- Utilizar outras formas de representação de texto, como por exemplo, baseadas em *word embedding* com aprofundamento das técnicas de mineração de texto.
- Avaliar o uso de outras técnicas para mineração de imagens, voltadas a imagens muito complexas e com objetos diversos.
- Utilização de outros dados de natureza multimodal, como outros elementos das redes sociais (ex.: locais nos quais uma pessoa foi marcada) e diferentes dados relacionados aos elementos das redes sociais (ex.: vídeos, imagens que possuem conteúdo textual).
- Contribuir com a abordagem proposta, explorando a utilização de técnicas de extração de conteúdo textual das imagens.
- Utilizar a combinação de dados estruturados presentes nas redes sociais (ex.: dados demográficos, preferência por categorias pré-definidas na rede social) em conjunto com dados de natureza multimodal (texto e imagens) apresentados neste trabalho.
- Apresentar a comparação de diferentes valores para o parâmetro ajustável do cálculo da similaridade ponderada intragrupo.
- Apresentar uma alternativa para automatizar as tarefas manuais utilizadas na seleção das mensagens da rede social Twitter.
- Efetuar a comparação do uso de somente texto, somente imagens e dados de natureza multimodal (texto e imagens) na recomendação de produtos.
- Analisar aspectos de segurança e privacidade do usuário.

REFERÊNCIAS

ADOMAVICIUS, G.; TUZHILIN, A. ***Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions.*** IEEE Transactions on Knowledge and Data Engineering, V.17, n.6, p.734-749, June 2005.

AMATRIAIN, X.; JAIMES, A.; OLIVER, N.; PUJOL, J. M. ***Data Mining Methods for Recommender Systems.*** In: RICCI, F.; ROKACH, L.; SHAPIRA, B. **Recommender Systems Handbook.** Springer, 2011. p.39-71.

_____. ***Mining large streams of user data for personalized recommendations.*** ACM SIGKDD Explorations Newsletter, V.14, n.2, p.37-48, Dez. 2013.

AMATRIAIN, X.; BASILICO, J. ***Past, Present, and Future of Recommender Systems: An Industry Perspective.*** ACM Press, 2016.

AMATRIAIN, X. ***Building industrial-scale real-world recommender systems.*** [s.l.]: ACM Press, 2012.

ANASTACIO, R. S. **Utilização de informações de redes sociais em algoritmos de recomendação baseados em filtragem colaborativa.** 2018. 104 f. Dissertação (Mestrado) – Instituto de Pesquisas Tecnológicas do Estado de São Paulo, São Paulo, Brasil, 2018.

APPLE vImage Programming Guide. **Consulta geral a página.** Disponível em: <<https://developer.apple.com/library/archive/documentation/Performance/Conceptual/vImage/ConvolutionOperations/ConvolutionOperations.html>>. Acesso em: 20/nov./17. il. color.

BAMBINI, R.; CREMONESI, P.; TURRIN, R. ***A Recommender System for an IPTV Service Provider: a Real Large-Scale Production Environment.*** In: RICCI, F. et

al. (Orgs.). *Recommender Systems Handbook*. Boston, MA: Springer US, 2011. p. 299–331.

BELLO-ORGAZ, G.; JUNG, J. J.; CAMACHO, D. ***Social big data: Recent achievements and new challenges***. *Information Fusion*, [s.l.], v. 28, p. 45–59, 2016.

BENGIO, Y. ***Learning Deep Architectures for AI***. *Foundations and Trends® in Machine Learning*, [s.l.], v. 2, n° 1, p. 1–127, 2009.

BURKE, R. ***Hybrid recommender systems: Survey and experiments***. *User modeling and user-adapted interaction*, [s.l.], v. 12, n° 4, p. 331–370, 2002.

_____. ***Hybrid Web Recommender Systems***. [s.l.]: CTI Technical Report 06-012. 2006.(Available at <http://www.cs.depaul.edu/research/technical.asp>), 2006.

CANZIANI, A.; PASZKE, A.; CULURCIELLO, E. ***An Analysis of Deep Neural Network Models for Practical Applications***. *arXiv:1605.07678 [cs]*, [s.l.], 2016.

CHEN, L.; YANG, F.; YANG, H. ***Image-based Product Recommendation System with Convolutional Neural Networks***. [s.l.], 2017.

CINAR, Y. G.; ZOGHBI, S.; MOENS, M.-F. ***Inferring User Interests on Social Media from Text and Images***. [s.l.]: IEEE, 2015.

CONTRATRES, F. G. et al. ***Sentiment Analysis of Social Network Data for Cold-Start Relief in Recommender Systems***. In: ROCHA, Á. et al. (Orgs.). *Trends and Advances in Information Systems and Technologies*. Cham: Springer International Publishing, 2018. v. 746, p. 122–132.

CORRÊA JR, E. A. et al. ***PELESent: Cross-domain polarity classification using distant supervision***. *arXiv:1707.02657 [cs]*, [s.l.], 2017.

ENG, S.; HUANG, L.; XU, G. ***Social network-based service recommendation with trust enhancement***. Expert Systems With Applications, China, v. 41, n. 18, p.8075-8084, dez. 2014. Elsevier BV.

FACEBOOK GRAPH API. **Consulta geral a homepage oficial**. Disponível em: <<https://developers.facebook.com/docs/graph-api>>. Acesso em: 03 jan. 2017.

KAGGLE CDISCOUNT CHALLENGE. **Consulta geral a homepage oficial**. Disponível em: <<https://www.kaggle.com/c/cdiscount-image-classification-challenge>>. Acesso em: 03 fev. 2018.

FACEBOOK LOGIN. **Consulta geral a homepage oficial**. Disponível em: <<https://developers.facebook.com/docs/facebook-login/web/login-button>>. Acesso em: 03 mar. 2017.

FAN, W.; BIFET, A. ***Mining big data: current status, and forecast to the future***. ACM SIGKDD Explorations Newsletter, [s.l.], v. 14, nº 2, p. 1–5, 2013.

GONZALEZ CAMACHO, L. A.; ALVES-SOUZA, S. N. ***Social network data to alleviate cold-start in recommender system: A systematic review***. Information Processing & Management, [s.l.], v. 54, nº 4, p. 529–544, 2018.

GUELLIL, I.; BOUKHALFA, K. ***Social big data mining: A survey focused on opinion mining and sentiments analysis***. [s.l.]: IEEE, 2015.

HAN, J.; KAMBER, M. ***Data mining: concepts and techniques***. 3rd ed ed. Burlington, MA: Elsevier, 2011. 703 p.

HILL, W. et al. ***Recommending and evaluating choices in a virtual community of use***. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. [s.l.]: ACM Press/Addison-Wesley Publishing Co., 1995.

HU, M.; LIU, B. ***Mining Opinion Features in Customer Reviews***. [s.l.], p. 6, 2004.

HU, X.; LIU, H. **Text Analytics in Social Media**. In: AGGARWAL, C. C.; ZHAI, C. (Orgs.). *Mining Text Data*. Boston, MA: Springer US, 2012. p. 385–414.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO). ISO/IEC 25010. Systems and software engineering. systems and Software Quality Requirements and Evaluation (SQuaRE) — System and software quality models. Geneva, 2011.

ISINKAYE, F. O.; FOLAJIMI, Y. O.; OJOKOH, B. A. **Recommendation systems: Principles, methods and evaluation**. *Egyptian Informatics Journal*, [s.l.], v. 16, n° 3, p. 261–273, 2015.

JIN, L. et al. **Understanding user behavior in online social networks: A survey**. *IEEE Communications Magazine*, [s.l.], v. 51, n° 9, p. 144–150, 2013.

JSON SCHEMA. **Consulta geral a homepage oficial**. Disponível em: <<https://json-schema.org/>>. Acesso em: 01 out. 2018.

KARPATHY, A.; LI, F.; JOHNSON, J. **CS231n Convolutional Neural Network for Visual Recognition**, 2016. Disponível em: <<http://cs231n.github.io/>>. Acesso em: 12/jun./17. il. color.

KONSTAN, J. A.; RIEDL, J. **Recommender systems: from algorithms to user experience**. *User Modeling and User-Adapted Interaction*, [s.l.], v. 22, n° 1–2, p. 101–123, 2012.

LAMPROPOULOS, A. S.; TSIHRINTZIS, G. A. **A Survey of Approaches to Designing Recommender Systems**. In: TSIHRINTZIS, G. A.; VIRVOU, M.; JAIN, L. C. (Orgs.). *Multimedia Services in Intelligent Environments*. Heidelberg: Springer International Publishing, 2013. v. 24, p. 7–30.

LECUN, Y. et al. **Object Recognition with Gradient-Based Learning**. In: *Shape, Contour and Grouping in Computer Vision*. London, UK, UK: Springer-Verlag, 1999.

LECUN, Y.; BENGIO, Y.; HINTON, G. **Deep learning**. *Nature*, [s.l.], v. 521, n° 7553, p. 436–444, 2015.

LI, C. et al. **Exploring social network information for solving cold start in product recommendation**. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, [s.l.], v. 9419, p. 276–283, 2015.

LIU, J.; DOLAN, P.; PEDERSEN, E. R. **Personalized news recommendation based on click behavior**. [s.l.]: ACM Press, 2010.

NGUYEN, K. et al. **Iris Recognition With Off-the-Shelf CNN Features: A Deep Learning Perspective**. *IEEE Access*, [s.l.], v. 6, p. 18848–18855, 2018.

NIAZ, U.; MERALDO, B. **Fusion methods for multi-modal indexing of web data**. In: *Image Analysis for Multimedia Interactive Services (WIAMIS), 2013 14th International Workshop on*. [s.l.]: IEEE, 2013.

NOGUEIRA, R. F.; ALENCAR LOTUFO, R. DE; CAMPOS MACHADO, R. **Fingerprint Liveness Detection Using Convolutional Neural Networks**. *IEEE Transactions on Information Forensics and Security*, [s.l.], v. 11, n° 6, p. 1206–1213, 2016.

OLIVEIRA, E. et al. **Recommending Groups to Users based Both on Their Textual and Image Posts**: [s.l.]: SCITEPRESS - Science and Technology Publications, 2016.

PEEMEN, M.; MESMAN, B.; CORPORAAL, H. **Efficiency Optimization of Trainable Feature Extractors for a Consumer Platform**. In: BLANC-TALON, J. et al. (Orgs.). *Advanced Concepts for Intelligent Vision Systems*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011. v. 6915, p. 293–304. il. color.

PRANDO, A. V. ***Um Sistema de Recomendação para E-commerce Utilizando Redes Sociais para solução de cold-start***. 2016. 120 f. Dissertação (Mestrado) – Instituto de Pesquisas Tecnológicas do Estado de São Paulo, São Paulo, Brasil, 2016.

QIAN, X.; et al. ***Personalized Recommendation Combining User Interest and Social Circle***. **IEEE Trans. Knowl. Data Eng.**, China, v. 26, n. 7, p.1763-1777, jul. 2014. Institute of Electrical and Electronics Engineers (IEEE).

QUBA, R. C. A. ***On enhancing recommender systems by utilizing general social networks combined with users goals and contextual awareness***. 2015. Tese (Doutorado) – Université Claude Bernard-Lyon I, Lyon, França, 2015.

RESHMA, R.; AMBIKESH, G.; THILAGAM, P. S. ***Alleviating data sparsity and cold start in recommender systems using social behaviour***. In: *2016 International Conference on Recent Trends in Information Technology (ICRTIT)*. [s.l.]: [s.n.], 2016.

RESNICK, P. et al. ***GroupLens: an open architecture for collaborative filtering of netnews***. In: *Proceedings of the 1994 ACM conference on Computer supported cooperative work*. [s.l.]: ACM, 1994.

RESNICK, P.; VARIAN, H. R. ***Recommender systems***. *Communications of the ACM*, [s.l.], v. 40, n° 3, p. 56–58, 1997.

RICCI, F.; ROKACH, L.; SHAPIRA, B. ***Introduction to Recommender Systems Handbook***. In: RICCI, F. et al. (Orgs.). *Recommender Systems Handbook*. Boston, MA: Springer US, 2011. p. 1–35.

RIVAS-SÁNCHEZ, M. et al. ***Using Deep Learning for Image Similarity in Product Matching***. In: ROJAS, I.; JOYA, G.; CATALA, A. (Orgs.). *Advances in Computational Intelligence*. Cham: Springer International Publishing, 2017. v. 10305, p. 281–290.

SHANKAR, D. et al. ***Deep Learning based Large Scale Visual Recommendation and Search for E-Commerce***. *arXiv preprint arXiv:1703.02344*, [s.l.], 2017.

SALLOUM, S. A.; AL-EMRAN, M.; SHAALAN, K. ***Mining Social Media Text: Extracting Knowledge from Facebook***. *International Journal of Computing and Digital Systems*, [s.l.], v. 6, n° 2, p. 73–81, 2017.

SHARDANAND, U.; MAES, P. ***Social information filtering: algorithms for automating “word of mouth”***. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. [s.l.]: ACM Press/Addison-Wesley Publishing Co., 1995.

NASCIMENTO E SILVA, R. G. ***Sistema de Recomendação baseado em conteúdo textual: avaliação e comparação***. 2014. 123 f. Dissertação (Mestrado) – Universidade Federal da Bahia (UFBA) e a Universidade Estadual de Feira de Santana (UEFS), Bahia, Brasil, 2014.

SOUZA, M.; VIEIRA, R. ***Sentiment Analysis on Twitter Data for Portuguese Language***. In: CASELI, H. et al. (Orgs.). *Computational Processing of the Portuguese Language*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. v. 7243, p. 241–247.

TWITTER REST API. ***Consulta geral a homepage oficial***. Disponível em: <<https://developer.twitter.com/>>. Acesso em: 03 jan. 2017.

VERMA, J. P.; PATEL, B.; PATEL, A. ***Big Data Analysis: Recommendation System with Hadoop Framework***. [s.l.]: IEEE, 2015.

WOHLGENANT, G.; MINIC, F. ***Using word2vec to Build a Simple Ontology Learning System***. In: *International Semantic Web Conference (Posters & Demos)*. [s.l.], 2016.

YIN, H. et al. **LCARS: A Location-content-aware Recommender System**. In: *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: ACM, 2013.

_____. **LCARS: A Spatial Item Recommender System**. *ACM Trans. Inf. Syst.*, [s.l.], v. 32, n° 3, p. 11:1–11:37, 2014.

YOSINSKI, J. et al. **How transferable are features in deep neural networks?** [s.l.], p. 9, 2014.

YOU, Q.; BHATIA, S.; LUO, J. **A picture tells a thousand words About you! User interest profiling from user generated visual content**. *Signal Processing*, [s.l.], v. 124, p. 45–53, 2016.

ZHANG, X. et al. **Classify social image by integrating multi-modal content**. *Multimedia Tools and Applications*, [s.l.], 2017.

ZHANG, Y.; PENNACCHIOTTI, M. **Predicting purchase behaviors from social media**. In: *Proceedings of the 22nd international conference on World Wide Web*. [s.l.]: ACM, 2013a.

_____. **Recommending branded products from social media**. [s.l.]: ACM Press, 2013b.

REFERÊNCIAS CONSULTADAS

ALPAYDIN, E. ***Introduction to machine learning***. Third edition ed. Cambridge, Massachusetts: The MIT Press, 2014.

CARON, S.; BHAGAT, S. ***Mixing Bandits: A Recipe for Improved Cold-start Recommendations in a Social Network***. In: *Proceedings of the 7th Workshop on Social Network Mining and Analysis*. New York, NY, USA: ACM, 2013.

DOU, Y.; YANG, H.; DENG, X. ***A Survey of Collaborative Filtering Algorithms for Social Recommender Systems***. In: *2016 12th International Conference on Semantics, Knowledge and Grids (SKG)*. [s.l.]: [s.n.], 2016.

FACELI, K. et al. ***Inteligência artificial: uma abordagem de aprendizado de máquina***. [s.l.]: Grupo Gen - LTC, 2011.

MA, T. et al. ***Social Network and Tag Sources Based Augmenting Collaborative Recommender System***. *IEICE TRANSACTIONS ON INFORMATION AND SYSTEMS*, [s.l.], v. E98D, n° 4, p. 902–910, 2015.

MINGXUAN SUN; FEI LI; JIAN ZHANG. ***A Multi-Modality Deep Network for Cold-Start Recommendation***. *Big Data and Cognitive Computing*, [s.l.], v. 2, n° 1, p. 7, 2018.

POZZI, F. A. et al. ***Challenges of Sentiment Analysis in Social Networks: An Overview***. In: POZZI, Federico Alberto et al. (Orgs.). *Sentiment Analysis in Social Networks*. Boston: Morgan Kaufmann, 2017. p. 1–11.

RUSSAKOVSKY, O. et al. ***ImageNet Large Scale Visual Recognition Challenge***. *International Journal of Computer Vision*, [s.l.], v. 115, n° 3, p. 211–252, 2015.

SINGH, S.; SINGH, N. ***Big Data analytics***. In: *2012 International Conference on Communication, Information Computing Technology (ICCICT)*. [s.l.]: [s.n.], 2012.

ZHAO, W. X. et al. ***Connecting Social Media to E-Commerce: Cold-Start Product Recommendation Using Microblogging Information***. *IEEE Transactions on Knowledge and Data Engineering*, [s.l.], v. 28, n° 5, p. 1147–1159, 2016.