

Recommender Vision: um sistema de recomendação para o comércio baseado em Visão Computacional

Malomar Alex Seminotti, Rafael Rieder
Programa de Pós-Graduação em Computação Aplicada
Universidade de Passo Fundo - UPF
Passo Fundo, RS, Brasil
E-mail: malomarbr@gmail.com, rieder@upf.br

Abstract—Recommender systems are useful to improve customer’s buying process by assisting them in decision-making processes. These systems usually use structured data, and unstructured data sets remain poorly explored. In this context, adding computer vision and artificial intelligence techniques can improve the result of the recommendations. This paper presents the progress of a recommender system that suggests products in real-time based on people’s behavior, considering video image analysis. The approach considers the user identity privacy, based on facial correspondence, without citing user identification. The sale suggestion inferences are based on the monitoring of images from a defined region on the store, considering products that users are viewing during a purchase.

Resumo—Sistemas de recomendação são úteis para melhorar a sugestão de compra aos clientes, ajudando-os em processos de tomada de decisão. Esses sistemas se utilizam com dados estruturados e, em virtude disso, conjuntos de dados não estruturados são pouco explorados. Nesse contexto, agregar técnicas de visão computacional e inteligência artificial pode melhorar o resultado das recomendações. Esse artigo apresenta o andamento da implementação de um sistema de recomendação que recomenda produtos em tempo real com base no comportamento das pessoas, considerando análise de imagens de vídeo. A abordagem considera a preservação de identidade dos usuários, a partir da correspondência facial, sem identificação nominal. As inferências para sugestões de vendas tomam por base o monitoramento de imagens de determinada região da loja, com produtos que os usuários estão observando durante uma compra.

I. INTRODUÇÃO

Empresas em geral buscam aumentar seu volume de vendas e de clientes para, conseqüentemente, aumentar seus lucros. Uma das formas de obter isso é melhorar as recomendações para seus produtos, oferecendo itens que realmente sejam de interesse dos consumidores. Este objetivo pode ser alcançado com a utilização de um sistema de recomendação (SR).

SRs fornecem sugestões de itens a serem recomendados para um usuário, ajudando-o a analisar mais informações sobre produtos ou serviços, e na tomada de decisão [1]. Esse tipo de solução tem contribuído no aumento de vendas on-line, principalmente para sites de comércio eletrônico [2]. Eles podem combinar várias técnicas de software para selecionar itens personalizados com base nos interesses dos usuários e conforme o contexto no qual estão inseridos.

Embora os SRs existentes sejam bem sucedidos na produção de recomendações, eles ainda enfrentam desafios de desempenho, como precisão, escalabilidade e inicialização a frio [3].

Para tanto, esses sistemas têm usado algoritmos de *deep learning* para melhorar a qualidade de suas recomendações [4].

Além disso, os SRs existentes geralmente são criados pensando em atender o comércio on-line, e não o físico [5]. As informações são coletadas e analisadas para permitir, *a posteriori*, a recomendação com base no histórico de compra das pessoas. Neste processo, perde-se a possibilidade de oferecer produtos de interesse do cliente em tempo real, ou seja, no momento em que o mesmo demonstrou o ímpeto de compra. Isso cria uma lacuna de SRs que poderiam se utilizar de dados não estruturados, por exemplo, para recomendar ações de vendas em lojas físicas.

A análise do comportamento do cliente em tempo real pode ajudar a atrair e reter clientes com mais eficiência e eficácia, e os varejistas podem obter vantagem competitiva sobre seus concorrentes [1]. O uso de um SR baseado em visão computacional (VC) pode mostrar a um comerciante os produtos que mais estão chamando atenção de seus clientes no momento da compra, mesmo sem a sua supervisão ou fora do alcance de um atendente. Este dados podem se tornar relevantes ao lojista, uma vez que, conhecendo o comportamento dos compradores, pode também reorganizar produtos nas prateleiras, deixando em destaque os itens mais vendidos ou com maior margem.

O sistema pode informar o interesse do cliente em um determinado item e, desta forma, ajudar na abordagem do vendedor, que poderá ter informações em tempo real sobre o interesse do comprador. O vendedor pode, também, sugerir produtos similares, com um custo mais atrativo. Isso tende a gerar um potencial maior na transação comercial, aumentando o valor da venda, além de fidelizar o cliente pela prestatividade e objetividade no atendimento [6].

Nesse contexto, este trabalho apresenta o andamento da implementação de um SR, baseado em conteúdo e utilizando técnicas de VC, para recomendar produtos de interesse de clientes em tempo real, considerando o comportamento das pessoas. As inferências para sugestões de vendas tomam por base o monitoramento de imagens de determinada região da loja, com produtos que os usuários estão observando durante uma compra, e também cruzando com histórico existente de interesses visuais anteriores. A abordagem considera a preservação de identidade dos usuários, a partir da correspondência facial, sem identificação nominal, considerando a Lei Geral de Proteção de Dados Pessoais (LGPD) [7].

II. TRABALHOS RELACIONADOS

Arapakis *et al.* [8] requisitaram a colaboração de trinta e quatro pessoas com perfis diferentes para participarem da avaliação. Eles utilizaram dois computadores para o teste: um deles contendo teclado, mouse e uma câmera; e o outro contendo o sistema de recomendação, o modelo SVM, o sistema de reconhecimento de expressões faciais (eMotion) e o software gravador de vídeo.

Para comparação, usaram duas versões do sistema proposto: na versão de linha de base, a técnica de criação de perfil integra informações que derivam apenas de ações do usuário (metadados e cliques); e, na versão multimodal, são integradas as informações afetivas (expressões faciais dos usuários), além dos dados de interação que estão sendo capturados. Como resultado do estudo, demonstraram que o recurso de interação multimodal é uma maneira promissora de melhorar o desempenho das recomendações.

Já o trabalho de Wu *et al.* [9] consistiu em um sistema de recomendação, integrado ao reconhecimento facial, para melhorar a identificação do perfil demográfico dos usuários (sexo e idade) em uma loja de varejo. Com o reconhecimento facial, eles conseguiram inferir o sexo e a idade de clientes que não possuíam informações armazenadas na base de dados (problema conhecido como inicialização a frio). Nesses casos, foi utilizada a filtragem demográfica para a recomendação de produtos. Para os clientes cujo perfil já era conhecido, foi utilizada a filtragem colaborativa para a sugestão de itens.

A criação do perfil dos clientes permitiu aos lojistas oferecer serviços personalizados a seus clientes, aprimorando o seu marketing e melhorando o resultado de suas vendas.

Por outro lado, Alfian *et al.* [1] desenvolveram um totem digital (DSOS - Digital Signage Online Store) e o instalaram em locais diferentes, para que os clientes pudessem ter a experiência de navegar e comprar um produto através de um menu em uma tela sensível ao toque, com imagens dos produtos em 3D para interação. O sistema de processamento de dados em tempo real coletou o histórico de navegação dos clientes e os dados de transações, conforme ocorriam. Além disso, os autores utilizaram a regra de associação para extrair informações úteis do comportamento dos usuários, de modo que ela pudesse ser usada pelos gerentes para aprimorar, de maneira eficiente, a qualidade do serviço.

O fornecimento de imagens 3D interativas aumentou a confiança e o prazer do cliente ao interagir com os produtos. O cliente pode ter uma visão melhor dos produtos, o que impactou positivamente seu comportamento de compra, aumentando, inclusive, a frequência das visitas aos estabelecimentos. Os resultados obtidos puderam ser usados tanto para a criação de promoções, quanto para a recomendação de produtos relevantes para os clientes DSOS.

Ademais, Zhang *et al.* [10] detalharam a evolução dos trabalhos relacionados aos sistemas de recomendação nos últimos anos. Fizeram uma revisão abrangente dos esforços de pesquisa recentes em SR baseados em *deep learning*, fornecendo alguns modelos para uso em trabalhos futuros.

Também relataram as tendências e as perspectivas de evolução dos estudos neste campo, unindo SR com redes neurais, aprendizagem de máquina e visão computacional associada ao processamento de linguagem natural.

O diferencial do estudo que está sendo apresentado é a recomendação em tempo real de produtos, com base na posição para onde o cliente está olhando, permitindo que o lojista perceba o interesse do comprador em um item e o ofereça no mesmo instante.

III. MATERIAIS E MÉTODOS

Os materiais e métodos apresentados nesta seção estão sendo aplicados para criar um sistema de recomendação que utiliza reconhecimento facial ou de padrões, como núcleo do processamento de informações, para fornecer sugestões de compra de produtos.

A. Software e Hardware

Para o desenvolvimento da aplicação, optou-se pelo uso da linguagem Python, com a interface gráfica criada com a IDE PySide2. As funcionalidades estão separadas em classes, com algoritmos paralelizados em *threads*.

Em relação ao uso de bibliotecas, destacam-se recursos da *OpenCV* (<http://www.opencv.org>) para algoritmos de visão de computacional; da Qt5 para interconexão da linguagem Python com a GUI; da *Face Recognition*, para a detecção da face da pessoa na imagem; e da *DLib* (<http://www.dlib.net>), para obter a direção para onde a pessoa está olhando.

Para a obtenção das imagens, considera-se o uso de apenas uma câmera com resolução HD ou superior para monitorar uma prateleira. Para este trabalho, adotou-se o uso do modelo Intelbras IP Dome HD VIP S4020 G2.

B. Funcionamento da Aplicação

A aplicação depende da captura das imagens através de um câmera de vídeo. Para tanto, a câmera deve ser instalada acima da prateleira que contém os produtos monitorados, de maneira que seu foco esteja em direção ao rosto dos clientes (Figura 1), permitindo o reconhecimento da face e a identificação do ângulo para onde o comprador está olhando, conforme exemplo na Figura 2. A partir do ângulo, pode ser obtido o produto de interesse.

Para a primeira versão da abordagem proposta, e para fins de validação, a aplicação considerará um cadastro com o mapeamento pré-definido dos produtos da prateleira em monitoramento.

A aplicação, de posse destas imagens, capta os quadros para verificar quanto tempo a pessoa passou analisando o produto (observando a mesma região da prateleira). O software considera, inicialmente, que o produto é objeto de interesse do cliente a partir de três segundos. De acordo com Gidlöf *et al.* [11], considerando os últimos cinco segundos anteriores a uma escolha, existe uma tendência maior no comportamento visual de uma pessoa cerca de dois segundos antes dela selecionar o produto que finalmente será escolhido.

A posição da pessoa no frame capturado é comparada com os frames seguintes, para confirmar o possível objeto

de interesse do cliente. O número de frames depende da configuração da câmera.

Como pode existir mais de uma pessoa na mesma cena, a aplicação deve reconhecê-las individualmente, segmentando a imagem. Ao mesmo tempo, também deve identificar os funcionários da loja e ignorá-los como objetos de interesse na imagem.

A partir disso, o reconhecimento e o registro do cliente é efetuado, sem que o mesmo seja identificado ou rotulado na base de dados. Nesse contexto, a recomendação apenas considera a exibição de seu rosto e a indicação do(s) item(ns) de interesse. Computacionalmente, existe apenas o registro por correspondência facial e o armazenamento dessas características no histórico do sistema, preservando sua identidade e considerando o art. 11, inciso 2º, alínea c da LGPD [7].

Após a obtenção do produto de interesse e da face da pessoa, estas informações são gravadas em uma base de dados e exibidas na tela da aplicação para o operador tomar a decisão de oferecer o produto ou seus similares ao cliente. O operador tem a opção de remover a face da base de dados, eliminando seu registro histórico. O funcionamento da aplicação está representado no diagrama da Figura 3.

Para validar o funcionamento da aplicação, a mesma será executada em duas lojas físicas, permitindo que sejam obtidas informações reais dos clientes em compra e, desta forma, verificando se as recomendações geradas pelo software estão de acordo com os interesses demonstrados pelos usuários.

IV. VISÃO DA ARQUITETURA DO SISTEMA

Este trabalho consiste no desenvolvimento de uma aplicação chamada Recommender Vision. Inicialmente, com o intuito de validar os recursos de software e hardware selecionados, a aplicação está identificando um conjunto restrito de pessoas e os objetos. A etapa também considerou a instalação do equipamento em uma posição que simula sua fixação em uma loja física (Figura 1).

Quando a aplicação é iniciada, ela faz uma varredura nas imagens previamente salvas em disco e analisa as características da face contida em cada arquivo. Posteriormente, é



Figura 1. Desenho da posição de instalação da câmera para monitoramento da prateleira. Fonte: o autor

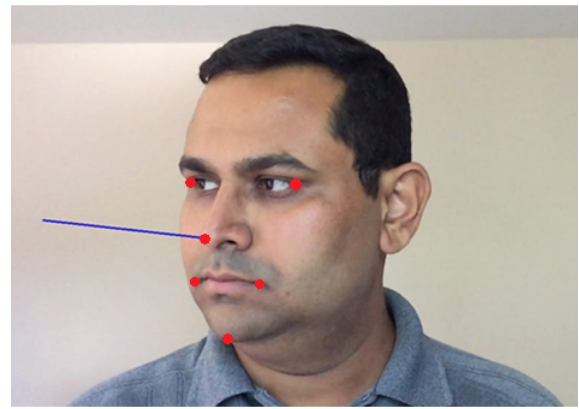


Figura 2. Exemplo da obtenção da posição estimada da cabeça. Fonte: <https://www.learnopencv.com/head-pose-estimation-using-opencv-and-dlib>

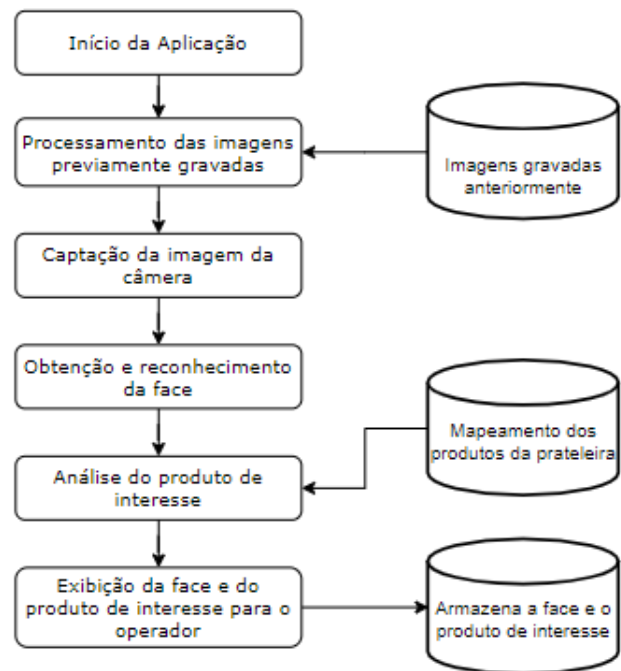


Figura 3. Diagrama esquemático da aplicação. Fonte: o autor

feita a comparação destas faces com as faces retornadas pela imagem da câmera. Assim, pode-se verificar se determinada pessoa não esteve no estabelecimento anteriormente e, desta forma, exibir ao lojista os produtos que o cliente demonstrou interesse naquelas visitas.

Concluída a etapa inicial, é feita a conexão com a câmera através da biblioteca OpenCV. Assim que a imagem é exibida, a biblioteca *Face Recognition* (FR) entra em ação para verificar se existe uma face na imagem que está sendo mostrada, executando em *loop* até a aplicação ser finalizada. A FR é uma biblioteca desenvolvida a partir de aprimoramentos da biblioteca DLib. Em virtude disso, compartilha de uma limitação no ângulo de visão, que varia de +30 a -30 graus.

Na etapa inicial, também é feita a análise das faces das pessoas que devem ser ignoradas, pois são funcionários da loja

e, portanto, não devem ser sugeridos produtos a estas pessoas.

Caso o rosto não seja de alguém que deva ser ignorado, utiliza-se recursos da biblioteca DLib para obter o ângulo em que a face está apontando, com base na detecção dos pontos faciais. Esse ângulo será utilizado para capturar o produto de interesse, calculando a posição do cliente em relação à câmera e, também, utilizando o mapeamento dos produtos da prateleira, configurado pelo lojista. O mapeamento da prateleira consiste em medir fisicamente o comprimento da mesma e os espaços ocupados por cada produto que será monitorado. São obtidas duas amostras a partir dos *frames* da imagem durante 3 segundos, comparando o ângulo de visão do usuário nelas. Se for detectado um movimento inferior a 5% no ângulo obtido, isto será considerado interesse do comprador no produto. Como a DLib retorna um *array* de faces detectadas, é possível tratar a imagem da câmera caso ela retorne mais de uma face na mesma cena (dois clientes, por exemplo).

De posse da face e do produto de interesse, a aplicação salva em base de dados a informação da face e do produto. Em seguida, como retorno visual, desenha-se uma borda ao redor da face para destacar ao operador que um rosto foi identificado, com informações adicionais sendo exibidas em uma lista na tela. Se a aplicação detectou que esta face já havia sido capturada em uma visita anterior, exibe-se, junto à imagem, o arquivo contendo a data e hora da última visita, conforme a Figura 4.

A lista terá uma opção para remover a face detectada, que também eliminará o arquivo em disco e as informações sobre o produto de interesse.



Figura 4. Exemplo de identificação da face previamente gravada. Fonte: o autor

V. RESULTADOS INICIAIS

Até o momento foram feitos testes com as bibliotecas e criada a estrutura básica da aplicação, com a utilização do PySide2 e a biblioteca Qt5. As imagens capturadas da câmera estão sendo salvas em disco, bem como já é realizada a identificação das faces e a comparação destas com as imagens gravadas anteriormente.

VI. CONCLUSÃO

Sistemas de recomendação são ferramentas que auxiliam o lojista a sugerir produtos de interesse do cliente, visando

umentar o valor das vendas e, consequentemente, o seu lucro. A união destes sistemas com visão computacional busca adicionar recursos adicionais para melhorar as sugestões. O desenvolvimento deste trabalho objetiva mesclar estas duas áreas para recomendar produtos com base no comportamento das pessoas em lojas físicas, através da análise de imagens de vídeo - uma vez que forma encontrados poucos trabalhos nesta área, existindo, portanto, uma lacuna a ser explorada.

A utilização de uma aplicação, como a que está sendo proposta, visa trazer um diferencial para os lojistas, fazendo com que mudem a abordagem ao cliente e a configuração do ponto de venda ou na forma de negócio. Produtos podem ser rearranjados nas prateleiras para chamar mais atenção ou os itens de maior interesse podem ser colocados em uma localização mais apropriada.

Como próximas etapas, pretende-se finalizar o desenvolvimento da aplicação, ajustando os cálculos da posição da cabeça em relação aos produtos da prateleira, e criando um algoritmo que seja assertivo na recomendação. Também serão considerados eventuais ajustes após testes em ambiente real (lojas físicas com lojistas e clientes), verificando se as sugestões geradas pelo software estão de acordo com os interesses demonstrados pelos compradores.

REFERÊNCIAS

- [1] G. Alfian, M. F. Ijaz, M. Syafrudin, M. A. Syaekhoni, N. L. Fitriyani, and J. Rhee, "Customer behavior analysis using real-time data processing," *Asia Pacific Journal of Marketing and Logistics*, p. 265–290, 2019.
- [2] L. Yu, F. Han, S. Huang, and Y. Luo, "A content-based goods image recommendation system," *Multimedia Tools and Applications*, p. 4155–4169, 2018.
- [3] I. Portugal, P. Alencar, and D. Cowan, "The use of machine learning algorithms in recommender systems: A systematic review," *Expert Systems with Applications*, pp. 205–227, 2018.
- [4] Z. Batmaz, A. Yurekli, A. Bilge, and C. Kaleli, "A review on deep learning for recommender systems: Challenges and remedies," *Artificial Intelligence Review*, p. 1–37, 6 2018.
- [5] G. Adomavicius and A. Tuzhilin, "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*, pp. 734–749, 2005.
- [6] K. Tarnowska, Z. W. Ras, and L. Daniel, *Recommender System for Improving Customer Loyalty*. Springer, 2020.
- [7] Brasil, "Lei nº 13.709, de 14 de agosto de 2018. Dispõe sobre a proteção de dados pessoais e altera a Lei nº 12.965, de 23 de abril de 2014 (Marco Civil da Internet)." *Diário Oficial da República Federativa do Brasil*, 14 ago. 2018. [Online]. Available: http://www.in.gov.br/material/-/asset/_publisher/Kujrw0TZC2Mb/content/id/36849373/do1-2018-08-15-lei-no-13-709-de-14-de-agosto-de-2018-36849337
- [8] I. Arapakis, Y. Moshfeghi, H. Joho, R. Ren, D. Hannah, and J. M. Jose, "Integrating facial expressions into user profiling for the improvement of a multimodal recommender system," in *2009 IEEE International Conference on Multimedia and Expo*. IEEE, 2009, pp. 1440–1443.
- [9] C.-C. Wu, Y.-C. Zeng, and M.-J. Shih, "Enhancing retailer marketing with an facial recognition integrated recommender system," *IEEE International Conference on Consumer Electronics-Taiwan*, pp. 25–26, 2015.
- [10] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep learning based recommender system: A survey and new perspectives," *ACM Computing Surveys (CSUR)*, vol. 52, no. 1, pp. 1–38, 2019.
- [11] K. Gidlöf, A. Wallin, R. Dewhurst, and K. Holmqvist, "Using eye tracking to trace a cognitive process: Gaze behaviour during decision making in a natural environment," *Journal of Eye Movement Research*, vol. 6, no. 1, Apr. 2013. [Online]. Available: <https://bop.unibe.ch/JEMR/article/view/2351>