

Sistema de Recomendação Personalizado para Viagens nos Estados Unidos

LUIZ F. SCHULZ CORIA and JOÃO BRASILEIRO M. DA SILVA, Universidade de São Paulo, Brazil



Fig. 1. New York City

Explorar diferentes culturas e ambientes é um aspecto fundamental da experiência humana, enriquecendo nosso conhecimento e compreensão do mundo. A ampla gama de oportunidades de viagem oferecidas pelos Estados Unidos representa um cenário empolgante, mas ao mesmo tempo desafiador, devido à sua vasta diversidade em termos geográficos, culturais e gastronômicos. Frequentemente, os viajantes se encontram diante de um leque extenso de opções, o que pode gerar uma sensação de sobrecarga na decisão de escolha de destinos.

Diante deste cenário, o objetivo do presente estudo é desenvolver um algoritmo de recomendação para viagens nos Estados Unidos. A proposta do algoritmo é considerar as preferências individuais dos viajantes, como destinos de interesse e culinária favorita, ao mesmo tempo em que leva em conta as características únicas de diversas cidades e estados. Através deste método, o algoritmo busca fornecer recomendações personalizadas que refletem a rica diversidade geográfica, cultural e gastronômica encontrada nos Estados Unidos

Vídeo: https://drive.google.com/file/d/1_CVfaO_H_-GoBz8-Udfl39KVd1HedlxV/view?usp=drive_link

1 INTRODUÇÃO

1.1 Contexto

Os sistemas de recomendação têm se tornado uma ferramenta tecnológica crucial em diversos setores, especialmente na era da informação, onde a quantidade de opções disponíveis pode ser esmagadora. Esses sistemas utilizam algoritmos sofisticados para filtrar, prever e sugerir itens que são mais relevantes para os usuários, com base em suas preferências e comportamentos anteriores. No campo do turismo, os sistemas de recomendação adquirem uma importância particular,

Authors' address: Luiz F. Schulz Coria, luizschulz@usp.br; João Brasileiro M. da Silva, Joabrasileiro@usp.br, Universidade de São Paulo, Av. Trab. São Carlsense, 400, São Carlos, São Paulo, Brazil, 13566-590.

2023. Manuscript submitted to ACM

Manuscript submitted to ACM

pois têm o potencial de transformar a maneira como os viajantes descobrem e planejam suas experiências. Com a crescente demanda por experiências de viagem personalizadas, estes sistemas ajudam a navegar através da vasta gama de opções de destinos, atividades e experiências culinárias, proporcionando recomendações que melhor se alinham com os interesses individuais dos usuários. A adaptação e inovação contínua nesses sistemas são essenciais para atender às expectativas dinâmicas dos viajantes modernos.

1.2 Propósito de Estudo

A proposta deste projeto é desenvolver um sistema de recomendação sofisticado, que tem como objetivo otimizar a experiência de viagem dos usuários. A funcionalidade central do sistema é baseada na coleta de dados específicos fornecidos pelo usuário, incluindo a duração da viagem (em dias), o destino desejado, preferências de tipos de restaurante (como estabelecimentos Michelin, restaurantes convencionais e opções de fast-food), e gostos culinários particulares. Utilizando essas informações, o algoritmo de recomendação personalizado do projeto sugere acomodações hoteleiras adequadas. Além disso, para cada dia da viagem, o sistema propõe atividades relevantes na localidade selecionada e recomenda restaurantes que se alinham com as preferências culinárias do usuário. Um aspecto adicional do sistema é a integração com o Google Maps, que oferece aos usuários rotas otimizadas para os locais de interesse, facilitando a logística de deslocamento durante a viagem. Este projeto visa, portanto, proporcionar uma experiência de viagem personalizada e eficiente, levando em conta as preferências individuais e necessidades logísticas dos usuários.

2 METODOLOGIA E MATERIAIS

2.1 Datasets

Para a concretização deste projeto, serão empregados cinco conjuntos de dados obtidos no Kaggle: Hotel Reviews[3], Michelin Restaurants[5], Fast Food Restaurant [4], TripAdvisor Restaurants data [6] e US Travel Check-ins[7], além de um dataset auxiliar US zipCodes Dataset[8].

- **Hotel Reviews Dataset:**

- Contém informações sobre hotéis, incluindo nome, localização e avaliações dos hóspedes.

- **Michelin Restaurants Dataset:**

- Compreende dados como nome do restaurante, localização geográfica, tipo de culinária servida e classificação em estrelas.

- **US Travel Check-ins Dataset:**

- Contém registros de locais de viagem nos Estados Unidos, frequência de visitas, além de informações temporais, incluindo datas e estações do ano das visitas.

- **Tripadvisor Dataset:**

- Contém registros de restaurantes nos Estados Unidos, incluindo nome, localização, tipo de culinária e avaliações dos usuários.

- **Fast Food Dataset:**

- Inclui registros de Fast Foods nos Estados Unidos, com informações como nome e localização.

- **Us zips Dataset**

- Inclui registros de todos ZIP Codes dos Estados Unidos e suas respectivas coordenadas

2.2 Bibliotecas e APIs

- **Gmaps[1]:**
 - *Descrição:* Biblioteca em Python destinada a facilitar a interação com a API do Google.
 - *Utilização:* Procura informações sobre cidades desejadas pelos usuários e visualização de rotas entre locais nas recomendações finais.
- **Pandas[9]:**
 - *Descrição:* Biblioteca em Python especializada na organização e manipulação de arquivos .csv.
 - *Utilização:* Armazenamento e manipulação de datasets.
- **Geopy[2]:**
 - *Descrição:* Biblioteca em Python especializada em localização de coordenadas de endereços, cidades e países, além de outras funcionalidades com tais coordenadas.
 - *Utilização:* Cálculo da distância entre diferentes localidades utilizando as coordenadas de cada.

3 DESENVOLVIMENTO E IMPLEMENTAÇÃO

Nesta seção, a fim de fornecer uma compreensão mais aprofundada do projeto, será dividida a explicação em duas partes distintas. Primeiramente, será abordado a estrutura e a modularização do código, uma vez que cada conjunto de dados apresenta suas particularidades e, portanto, exigiu abordagens separadas em termos de estruturação. Posteriormente, será entrado em detalhes sobre cada componente e a progressão do algoritmo

3.1 Classes

- **Classe TripInformation:** A classe 'Trip Information' foi desenvolvida para gerir e obter informações cruciais das viagens dos usuários, incluindo o nome da cidade, duração da estadia e preferências de restaurantes.
- **Classe Hotéis:** A classe 'Hotels' foi projetada para realizar a leitura, o pré-processamento dos dados do dataset Hotel-Reviews e a elaboração do ranking para recomendação.
- **Classe Locations:** A classe 'Locations' foi estruturada para executar a leitura e o pré-processamento de dados do dataset US Travel Check-ins, culminando no ranking dos locais para recomendação.
- **Classe Restaurants:** A classe 'Restaurants' atua como uma superclasse, encapsulando outras três subclasses: 'FastFood', 'Michelin' e 'Normal', guardando algumas informações e funções utilizadas por suas subclasses. Além de também ter uma função de input do usuário, para conseguir as culinárias preferidas deste.
 - **Classe Michelin:** A classe 'Michelin' foi desenvolvida para gerir e obter os dados do dataset Michelin e gerar um ranking de recomendação.
 - **Classe FastFood:** A classe 'FastFood' foi desenvolvida para gerir e obter os dados do dataset Fast Food restaurants e elaborar uma lista de estabelecimentos próximos da cidade.
 - **Classe Normal:** A classe 'Normal' foi desenvolvida com o propósito de realizar a leitura e o pré-processamento dos dados provenientes do conjunto de dados do TripAdvisor Restaurant Reviews, bem como a geração de um ranking para recomendação.
- **Classe Recommendation:** A Classe 'Recommendation' foi criada para juntar todos os rankings provenientes das outras classes e fazer uma recomendação final do projeto.
- **Classe Plot:** A classe 'Plot' serve para visualizar as recomendações, possibilita imprimir o resultado da recomendação e visualizar as rotas das atividades de cada dia recomendado.

3.2 Funcionamento do Algoritmo

(1) **Entrada do Usuário:** O algoritmo começa solicitando informações ao usuário, incluindo o local da viagem, a duração da viagem e as preferências de tipos de restaurantes (Michelin, FastFood, Normal ou todos). Para facilitar o tratamento de erros de digitação ou diferenças de idioma, o programa utiliza a API do Google Maps para obter o nome correto da cidade em inglês, o estado, a latitude e a longitude.

(2) **Pré-Processamento dos Dados:**

- **Hotels:** Inicialmente, o algoritmo realiza a leitura do CSV de hotéis e calcula a distância entre cada hotel e o centro da cidade, com um limite de 15 km.
- **Locations:** Para a categoria Locations, o processo envolve a leitura do CSV, a limpeza de registros que contêm apenas o nome da cidade/estado e não são relevantes para a recomendação, e o cálculo da distância de cada local até o centro da cidade, com um limite de 20 km.
- **Michelin:** Na categoria Michelin, o algoritmo lê três arquivos CSV (um para cada estrela), combina-os e calcula a distância de cada restaurante para o centro da cidade, com um limite de 15 km.
- **FastFood:** Para a categoria FastFood, é realizada a leitura do arquivo CSV e o cálculo da distância dos estabelecimentos em relação ao centro da cidade, com um limite de 15 km.
- **Normal:** Na categoria Normal, o algoritmo inicia a leitura do arquivo CSV, converte o código postal de cada estabelecimento em latitude e longitude usando um arquivo CSV auxiliar e, em seguida, calcula a distância até o centro da cidade, com um limite de 15 km.

(3) **Filtragem:**

- **Hotels:** Utiliza um método de filtragem colaborativa não personalizada para recomendação de hotéis, considerando as avaliações feitas pelos usuários para cada hotel. O algoritmo calcula uma nota ponderada levando em conta a média das avaliações e sua quantidade.
- **Locais:** Aplica um método de filtragem colaborativa não personalizada e implícita. O algoritmo utiliza a frequência com que um determinado local aparece no dataset de Check-Ins para avaliá-lo como relevante.
- **Fast Food:** Utiliza apenas a filtragem de distância mencionada anteriormente. Os estabelecimentos finais são escolhidos de forma aleatória.
- **Normal:** Emprega uma abordagem de filtragem colaborativa e baseada em conhecimento em duas etapas. Inicialmente, é realizada uma filtragem colaborativa, na qual os restaurantes são classificados com base em uma nota ponderada (número de avaliações multiplicado pela média delas). Em seguida, o algoritmo consulta o usuário sobre suas preferências culinárias para realizar uma filtragem baseada em conhecimento.
- **Michelin:** Utiliza uma abordagem de filtragem baseada em conhecimento, escolhendo o restaurante com base nas preferências culinárias do usuário.

*Observação: As filtrações para hotéis, locais e fast food produzem rankings com mais opções do que as que serão recomendadas no final.

(4) **Recomendação:** Para a recomendação final, é criada uma matriz para cada categoria de restaurante (Michelin, Fast Food, Normal), na qual é calculada a distância de cada restaurante para cada combinação de local e hotel considerada. No final, são selecionados os N (número de dias)trios (hotel + local + restaurante) com a menor distância total.

Se o usuário optar por visitar os três tipos de restaurantes, a distribuição será a seguinte: 60% dos dias serão reservados para restaurantes normais, enquanto 20% serão destinados a Fast Food e 20% a restaurantes Michelin, a fim de equilibrar as recomendações.

- (5) **Plot:** Para concluir, as recomendações são exibidas no terminal, incluindo informações sobre o hotel e as atividades de cada dia, juntamente com explicações sobre a recomendação de cada restaurante (às culinária escolhida). O programa também dá a opção para o usuário de trocar um local de visita de um determinado dia, melhorando ainda mais a interatividade e personalização do programa. Finalmente, o usuário tem a opção de visualizar as rotas das atividades diárias no Google Maps(utilizando a API do Google) e utilizar o Street View para explorar os locais que serão visitados.

4 MÉTODO DE AVALIAÇÃO

O sistema de recomendação desenvolvido para este projeto não dispõe de um método quantitativo para avaliação de desempenho. Portanto, foi adotado um método de avaliação baseado em estudos com usuários. Nesse sentido, solicitei a participação de 10 colegas distintos para utilizarem o programa e fornecerem avaliações e opiniões sobre sua experiência.

- **Usabilidade:**

- **Nota média:** 9.2
- **Comentário Principal:** Extremamente fácil de utilizar, todas as interações com o algoritmo foram bem intuitivas, simples e diretas.

- **Efetividade das Recomendações:**

- **Nota média:** 8.5
- **Comentário Principal:** Para grandes cidades (Nova Iorque, São Francisco) o algoritmo funcionava muito bem e dava boas recomendações, porém para cidades de pequeno/médio porte o algoritmo não funcionava.

- **Personalização:**

- **Nota média:** 8.0
- **Comentário Principal:** A personalização é muito boa, principalmente na parte dos restaurantes e conseguir decidir que tipo de restaurante ir e que culinária é a preferida, porém senti falta da personalização nos locais e nos hotéis, a possibilidade de conseguir trocar os locais na recomendação final é boa, mas seria mais interessante conseguir decidir que tipo de atividade (museus, natureza, praias) e uma customização maior nos hotéis, como preço.

- **Sugestão para Melhorias:**

- Melhorar a recomendação para cidades menores.
- Criar um plataforma/site para melhorar a experiência do usuário
- Melhorar a personalização para hotéis e locais

- **FeedBack Geral:**

- **Nota média:** 9.1
- **Comentário Principal:** Sistema muito bom e intuitivo. Porém utilizar no terminal não é algo muito interessante e a personalização poderia ser um pouco melhor.

5 CONCLUSÃO

A implementação do algoritmo representou um desafio significativo, dada a complexidade inerente à utilização de cinco diferentes conjuntos de dados, cada um com suas peculiaridades distintas. Isso exigiu um processo de pré-processamento e filtragem específico para cada conjunto, além do desenvolvimento de um algoritmo capaz de integrar todas essas informações de forma coesa. Adicionalmente, a tarefa de tornar o sistema intuitivo e interativo apresentou seus próprios obstáculos, exigindo a adaptação a diferentes tipos de entrada em momentos diversos do fluxo do algoritmo, bem como a integração da Google API para aprimorar a experiência do usuário.

No entanto, os resultados obtidos com este projeto são extremamente satisfatórios, como evidenciado pela pesquisa realizada com os usuários. O esforço dedicado à superação dos desafios técnicos e de usabilidade se traduz em um sistema eficaz e altamente funcional para a recomendação de itinerários de viagem.

6 PRÓXIMOS PASSOS

Conforme destacado durante a avaliação com os usuários, a principal prioridade consiste em adquirir uma quantidade maior de dados. Isso implica na obtenção de informações mais detalhadas sobre locais e hotéis, com o objetivo de aprimorar a personalização das recomendações para cada caso específico. Além disso, planeja-se expandir a base de dados para incluir informações sobre cidades e localidades de pequeno e médio porte, visando abranger um espectro mais amplo de regiões e oferecer recomendações de maior qualidade.

Por fim, seria interessante desenvolver uma plataforma/site dedicado para o sistema. Isso melhoraria a experiência do usuário, contribuindo significativamente para a usabilidade do sistema.

REFERENCES

- [1] Gmaps development team. 2018. *Gmaps Python Library*. <https://github.com/pbunghion/gmaps>
- [2] Geopy development team. 2023. *Geopy Python Library*. <https://github.com/geopy/geopy>
- [3] Kaggle. 2018. *Hotel-Reviews Dataset*. Retrieved September 24, 2023 from https://www.kaggle.com/datasets/datafiniti/hotel-reviews?select=Datafiniti_Hotel_Reviews.csv
- [4] Kaggle. 2020. *Fast Food Restaurants Dataset*. Retrieved December 27, 2023 from <https://www.kaggle.com/datasets/datafiniti/fast-food-restaurants>
- [5] Kaggle. 2020. *Michelin Restaurants Dataset*. Retrieved September 24, 2023 from <https://www.kaggle.com/datasets/jackywang529/michelin-restaurants>
- [6] Kaggle. 2020. *TripAdvisor Restaurant Dataset*. Retrieved December 27, 2023 from <https://www.kaggle.com/datasets/siddharthmandgi/tripadvisor-restaurant-recommendation-data-usa>
- [7] Kaggle. 2020. *US Travel Check-Ins Dataset*. Retrieved September 24, 2023 from <https://www.kaggle.com/datasets/thedevastator/us-travel-check-ins-analysis?select=IG.csv>
- [8] Meekohi. 2019. *US zips Dataset*. Retrieved September 28, 2023 from <https://gist.github.com/Meekohi/1f220f5a1d36feb9bd3cf5f59e518643>
- [9] The pandas development team. 2020. *pandas-dev/pandas: Pandas*. <https://doi.org/10.5281/zenodo.3509134>