

Sistema de Recomendação de Filmes: Descrição do Problema e Metodologia

Alunos: Gabriel Gomes Marchesan, Rafael Nalin Alves.

Descrição do Problema

Com o crescimento exponencial das plataformas de streaming, os usuários enfrentam um dilema: entre essa vasta quantidade de títulos para assistir, como saber se é bom ou ruim? Essa sobrecarga de opções pode resultar em frustração e insatisfação, levando os usuários a não explorarem novas produções. Levando em consideração esse problema, um sistema de recomendação eficaz se torna essencial para personalizar a experiência do usuário, ajudando-o a encontrar filmes que atendam às suas preferências. O problema central a ser tratado nesta proposta é como prever as avaliações de filmes que um usuário ainda não assistiu, utilizando dados de classificações anteriores de maneira eficiente.

Dados Utilizados

Para o desenvolvimento do sistema, serão utilizados dados provenientes de quatro fontes principais:

1. **links.csv**: Contém identificadores dos filmes e links relevantes para IMDb e TMDb, facilitando o acesso a informações adicionais.
2. **movies.csv**: Apresenta informações sobre os filmes, incluindo títulos, gêneros e ano de lançamento, que serão fundamentais para contextualizar as recomendações.
3. **ratings.csv**: Registra as avaliações feitas pelos usuários, permitindo entender as preferências e comportamentos de cada um.
4. **tags.csv**: Inclui tags atribuídas pelos usuários aos filmes, fornecendo uma dimensão adicional sobre o que os filmes representam.

Esses dados serão processados e integrados em um único DataFrame, possibilitando uma análise mais rica e uma estrutura robusta para a matriz de avaliações.

Métodos a Serem Explorados

1. Carregamento e Preparação dos Dados

Os dados serão carregados utilizando a biblioteca Pandas, permitindo a organização e limpeza das informações. O primeiro passo consiste na combinação dos DataFrames, resultando em uma estrutura que une informações sobre classificações e características dos filmes.

2. Criação da Matriz de Avaliações

A matriz de avaliações (user-item matrix) será criada, onde as linhas representam usuários e as colunas representam filmes. As células conterão as avaliações feitas pelos usuários, e as células correspondentes a filmes não avaliados serão preenchidas com valores ausentes (NaN).

3. Preenchimento de Valores Ausentes

Para lidar com os valores ausentes, será adotada a estratégia de preencher essas lacunas com a média das avaliações de cada filme. Essa abordagem garantirá que a matriz esteja completa, permitindo que o modelo de aprendizado de máquina opere de maneira eficaz.

4. Decomposição de Valor Singular (SVD)

A técnica de Decomposição de Valor Singular (SVD) será aplicada para decompor a matriz de avaliações em três matrizes menores: U (usuários), S (valores singulares) e V^T (filmes). Essa técnica permitirá identificar padrões nas avaliações, sendo crucial para prever as classificações que um usuário atribuiria a filmes ainda não assistidos.

5. Reconstrução da Matriz de Avaliações

Após a decomposição, a matriz de avaliações será reconstruída, preenchendo as células não avaliadas com previsões do modelo. Isso permitirá que o sistema produza recomendações personalizadas.

6. Previsão de Avaliações e Avaliação do Modelo

As avaliações para filmes não assistidos serão previstas, e a eficácia do modelo será avaliada utilizando o RMSE (Root Mean Squared Error). Este método quantificará a precisão das previsões em relação às avaliações reais, validando a performance do sistema.

7. Recomendação de Filmes

Por fim, o sistema recomendará filmes aos usuários com base nas previsões geradas, classificando-os segundo as avaliações previstas. Essa funcionalidade permitirá oferecer sugestões personalizadas, enriquecendo a experiência do usuário.

Conclusão

Este sistema de recomendação, baseado em filtragem colaborativa e na técnica SVD, representa uma abordagem nova para personalizar a experiência do usuário em plataformas de streaming. A proposta não apenas busca prever avaliações de filmes, mas também aprimorar a satisfação do usuário. Embora o sistema enfrente desafios, como o problema de "cold start" e a necessidade de dados substanciais, a aplicação prática de técnicas de aprendizado de máquina aqui apresentada demonstra um caminho promissor para transformar a maneira como os usuários interagem com a grande quantidade de títulos dos dias de hoje.