

Sistema de Recomendação de Filmes utilizando o Algoritmo Apriori na Plataforma Manga Play

Rodrigo S. Silva, Monyc Luisa A. Cerqueira, Pedro César P. Jesus

Curso de Sistemas de Informação – Centro Universitário de Excelência (UNEX)
Av. Artêmia Pires Freitas, s/n – Sim – 44085-370 – Feira de Santana – BA – Brasil

{Silva, Rodrigo} monyc_luisa@hotmail.com, pdrocesar78@gmail.com

Resumo: Este artigo descreve a implementação de um Sistema de Recomendação de Filmes utilizando o algoritmo de Regras de Associação Apriori na plataforma de streaming Manga Play. A plataforma visa sugerir filmes aos usuários com base em seu histórico de preferências e no filme mais recente assistido. O algoritmo Apriori é utilizado para gerar regras de associação entre filmes que compartilham características semelhantes, promovendo uma recomendação personalizada. Apresentamos o processo de desenvolvimento, incluindo a unificação dos dados, pré-processamento e a construção do sistema de recomendação, além dos resultados obtidos.

Introdução: A experiência cinematográfica é enriquecida pela imersão proporcionada pela qualidade do som e da imagem, oferecendo ao espectador uma vivência única. Com o advento do streaming, a forma como consumimos conteúdos audiovisuais foi transformada, permitindo acesso instantâneo a uma vasta gama de filmes e séries. No entanto, um problema comum é a dificuldade em encontrar novos filmes de qualidade, semelhantes aos já apreciados pelo usuário. O presente trabalho propõe a construção de um Sistema de Recomendação baseado no algoritmo Apriori, visando sugerir filmes de forma personalizada na plataforma de streaming Manga Play. O objetivo é fornecer uma recomendação eficiente com base nos gostos passados dos usuários, explorando as regras de associação entre filmes.

Fundamentação Teórica

O algoritmo Apriori é uma heurística clássica utilizada para descobrir padrões frequentes em grandes bases de dados, especialmente aplicável em sistemas de recomendação. A principal característica do algoritmo é a exploração de regras de associação, que descrevem relações entre diferentes itens em um conjunto de dados. Regras de associação são relações do tipo $X \Rightarrow Y$, em que X e Y são subconjuntos de um conjunto de itens I e não possuem elementos em comum. Nesse contexto, X é chamado de antecedente, e Y , de consequente da regra.

O Apriori trabalha com itens frequentes, ou seja, itens cuja frequência (suporte) é maior que um determinado limiar. A partir desses itens frequentes, gera-se um conjunto de regras de associação, avaliadas por métricas como suporte, confiança e lift.

- **O suporte:** mede a proporção de transações nas quais um item (ou conjunto de itens) aparece.

$$\text{Suporte}(\text{Itemset}) = \frac{\text{Número de transações que contêm o itemset}}{\text{Número total de transações}}$$

- **A confiança:** avalia a probabilidade de ocorrência de Y, dado que X já ocorreu.

$$\text{Confiança}(X \Rightarrow Y) = \frac{\text{Suporte}(X \cup Y)}{\text{Suporte}(X)}$$

- **Lift:** compara a confiança da regra com a expectativa de ocorrência de Y, considerando sua popularidade

$$\text{Lift}(X \rightarrow Y) = \frac{\text{Confiança}(X \rightarrow Y)}{\text{Suporte}(Y)}$$

Com base nesses conceitos, o algoritmo opera em etapas: inicia com itens individuais, calcula seus suportes, filtra os que atendem ao limite mínimo, gera combinações entre eles, e repete o processo até que não haja mais conjuntos frequentes. Em seguida, gera regras e as avalia pelas métricas mencionadas.

Metodologia

O desenvolvimento do sistema de recomendação *Manga Play* envolveu diversas etapas, desde a unificação de diferentes fontes de dados até a implementação manual do algoritmo Apriori e a definição de dois modos de recomendação distintos.

- **Unificação dos Dados:** A base de dados utilizada no projeto foi construída a partir de múltiplos arquivos CSV disponibilizados em um diretório compartilhado pelo professor. Foi necessário percorrer todos os arquivos de usuários para extrair e consolidar os dados de avaliações. Essa etapa de unificação foi realizada com o uso da biblioteca pandas.
- **Pré-processamento:** O primeiro passo do pré-processamento foi filtrar apenas as avaliações com nota superior a 3, considerando-as como indicações de que o usuário gostou do filme. Cada transação corresponde a um conjunto de títulos de filmes avaliados positivamente por um mesmo usuário, e esses conjuntos foram utilizados como entrada para o algoritmo de mineração de padrões frequentes.

- **Implementação do Algoritmo Apriori:** A Apriori foi feita do zero. O código procura por filmes que aparecem juntos nas listas dos usuários, começando por pares e depois aumentando a combinação. As combinações são criadas com uma função própria, sem usar bibliotecas prontas. Só os conjuntos que aparecem bastante são mantidos. Depois, o código monta regras do tipo "se gostou disso, talvez goste daquilo", com base na frequência dessas combinações.
- **Sistema de Recomendação:** O Sistema de Recomendação no seu código utiliza o algoritmo Apriori para identificar padrões de gosto dos usuários a partir das avaliações de filmes. Ele começa filtrando as avaliações positivas e gerando itemsets frequentes, ou seja, conjuntos de filmes que são frequentemente avaliados juntos. Com esses conjuntos, o sistema cria regras de associação que indicam, por exemplo, que "se o usuário gostou de A, ele provavelmente gostará de B".

Quando o usuário avalia positivamente um filme, o sistema busca regras em que esse filme aparece e recomenda outros filmes baseados nessas regras, priorizando aqueles com maior **confiança e lift**. Assim, o sistema sugere filmes que o usuário provavelmente gostará, com base em padrões de preferências de outros usuários.

Resultados e Discussões

Utilizando o algoritmo Apriori com suporte mínimo de 0.3 e confiança mínima de 0.4, foram identificados 10 conjuntos frequentes de filmes, incluindo combinações como (*Solaris*, *Terminator 3: Rise of the Machines*) com suporte de 0.32. Os filmes isolados com maior suporte foram *Terminator 3* (0.47), *The Million Dollar Hotel* (0.45) e *Solaris* (0.43).

Com base nesses dados, foram geradas 4 regras de associação. Entre elas, destacam-se:

- Se o usuário assistiu *Solaris*, há alta chance de também gostar de *Terminator 3* (confiança: 0.74, lift: 1.58).
- Se assistiu *The Million Dollar Hotel*, pode gostar de *Terminator 3* (confiança: 0.69, lift: 1.47).

Essas regras mostram padrões de co-preferência entre os filmes. No entanto, para o usuário 1, cujo último filme curtido foi *American Pie*, nenhuma recomendação foi possível, pois esse título não aparece entre os itens frequentes analisados.

```

C:\Users\pdroc\PycharmProjects\MangaPlay\.venv\Scripts\python.exe C:\Users\pdroc\PycharmProjects\MangaPlay\main.py
Itens Frequentes Descobertos:
  Nivel      Itemset  Suporte
0      1      (Three Colors: Red,)  0.31
1      1      (Terminator 3: Rise of the Machines,)  0.47
2      1      (Sissi,)  0.31
3      1      (Monsoon Wedding,)  0.37
4      1      (Once Were Warriors,)  0.35
5      1      (Solaris,)  0.43
6      1      (The Million Dollar Hotel,)  0.45
7      1      (The 39 Steps,)  0.41
8      1      (Men in Black II,)  0.31
9      2      (Solaris, Terminator 3: Rise of the Machines)  0.32
10     2      (Terminator 3: Rise of the Machines, The Milli...  0.31

Regras de Associação:
  Se tiver ... Lift
0      (Solaris,) ... 1.58
1      (Terminator 3: Rise of the Machines,) ... 1.58
2      (Terminator 3: Rise of the Machines,) ... 1.47
3      (The Million Dollar Hotel,) ... 1.47

[4 rows x 4 columns]

Usuário: 1
Último filme curtido: American Pie

Nenhuma recomendação encontrada para o último filme assistido.

```

Considerações Finais

O trabalho apresentou de forma eficiente a aplicação do algoritmo Apriori, identificando itens frequentes e gerando regras de associação com base em dados de preferência de filmes. As métricas de suporte, confiança e lift foram bem empregadas, revelando padrões relevantes. Como ponto positivo, destaca-se a capacidade de extrair relações úteis entre filmes com boa clareza. No entanto, houve limitações, como a ausência de recomendações para alguns casos devido à base de dados restrita. Como melhoria futura, sugere-se ampliar o conjunto de dados e combinar a abordagem com outros métodos de recomendação, como filtragem colaborativa, para cobrir mais situações e usuários.

Referências Bibliográficas

Alves, A. (2021). "Algoritmo Apriori e Data Mining". Medium. Disponível em: https://www.computersciencemaster.com.br/como-funciona-o-algoritmo-apriori/#google_vignette

IBM. (2024). "What is the Apriori algorithm?". IBM Knowledge Center. Disponível em: <https://www.ibm.com>.

Santos, V. (2018). "Como funciona o algoritmo Apriori". Disponível em: https://www.computersciencemaster.com.br/como-funciona-o-algoritmo-apriori/#google_vignette