

Manga Play: Um Sistema de Recomendação Baseado em Regras de Associação para uma Plataforma de Streaming

1. Introdução

Com o crescimento de plataformas de streaming, oferecer recomendações personalizadas se tornou essencial para retenção de usuários. O projeto *Manga Play*, desenvolvido para a empresa fictícia Manga Rosa, propõe uma solução baseada no algoritmo Apriori para sugerir filmes a partir das preferências dos usuários.

O sistema oferece dois tipos de recomendação: (1) com base no histórico de filmes que o usuário avaliou positivamente, e (2) com base no último filme assistido e bem avaliado. O objetivo é proporcionar uma experiência personalizada e relevante, incentivando maior engajamento com o catálogo da plataforma.

2. Fundamentação Teórica

O algoritmo **Apriori** é uma técnica clássica de *data mining* utilizada para gerar regras de associação a partir de conjuntos de transações. Seu objetivo é encontrar padrões frequentes que ocorrem juntos em um grande número de casos.

As principais métricas utilizadas são:

- **Suporte**: frequência relativa de um itemset no conjunto de transações.

$$\text{suporte}(A) = \frac{\text{ocorrências de } A}{\text{total de transações}}$$

- **Confiança**: probabilidade de ocorrência do consequente dado o antecedente.

$$\text{confiança}(A \rightarrow B) = \frac{\text{suporte}(A \cup B)}{\text{suporte}(A)}$$

- **Lift**: medida da importância da regra.

$$\text{lift}(A \rightarrow B) = \frac{\text{confiança}(A \rightarrow B)}{\text{suporte}(B)}$$

Pseudocódigo do Algoritmo Apriori

Input: Transações T, suporte mínimo s, confiança mínima c

Output: Regras de associação fortes

1. Gerar itemsets frequentes de tamanho 1 com suporte $\geq s$
2. Para $k = 2$ até não gerar novos itemsets:
 - Gerar candidatos de tamanho k a partir de combinações dos itemsets de tamanho k-1
 - Filtrar os candidatos com suporte $\geq s$
3. Para cada itemset frequente:
 - Gerar todas as regras possíveis $A \rightarrow B$
 - Calcular suporte, confiança e lift
 - Manter as regras com confiança $\geq c$

3. Metodologia

3.1 Unificação dos Dados

Foram utilizados dois arquivos CSV: movies.csv contendo o catálogo de filmes, e dataset.csv contendo as avaliações dos usuários. As bases foram unificadas utilizando pandas.

3.2 Pré-processamento

Apenas avaliações com nota superior a 3 foram consideradas como "gostou". Cada transação corresponde a uma lista de filmes que um usuário avaliou positivamente.

3.3 Implementação do Apriori

O algoritmo foi implementado do zero, utilizando apenas itertools e collections. Todos os itemsets frequentes foram gerados com base em um suporte mínimo definido, e as regras com confiança suficiente foram extraídas.

3.4 Sistema de Recomendação

Foram desenvolvidos dois modos:

- **Histórico de gostos:** recomendações baseadas em todas as avaliações positivas do usuário.

- **Último filme gostado:** recomendações com base exclusivamente no último filme com nota > 3.

4. Resultados e Discussões

4.1 Configurações Utilizadas

- Suporte mínimo: 0.2 (20%)
- Confiança mínima: 0.6 (60%)

- **4.2 Regras**

Antecedente	Consequente	Suporte	Confiança	Lift
[O Castelo Animado]	[Akira]	0.4	0.66	1.10
[Akira]	[Princesa Mononoke]	0.4	0.75	1.25
[Princesa Mononoke]	[Your Name]	0.4	0.66	1.10

As regras mostraram boa relevância com lift > 1, indicando que os filmes recomendados tendem a estar correlacionados com os interesses dos usuários.

4.3 Discussão

A abordagem se mostrou eficaz em ambientes com base de usuários e catálogo modestos. No entanto, conforme o número de usuários e filmes aumenta, a performance do algoritmo pode se tornar um gargalo. O modelo é interpretável e permite fácil auditoria das recomendações.

5. Considerações Finais

O projeto *Manga Play* demonstrou que é possível construir um sistema de recomendação interpretável e funcional sem o uso de bibliotecas externas específicas, utilizando o algoritmo Apriori. A abordagem baseada em regras oferece transparência e personalização ao usuário.

Limitações:

- Escalabilidade limitada com grandes volumes de dados.
- Não considera metadados dos filmes (gênero, diretor, etc.).

Melhorias Futuras:

- Integrar técnicas híbridas (Apriori + filtragem colaborativa).

- Otimização de desempenho para bases maiores.
 - Introdução de filtros contextuais (tempo, humor, etc.).
-

6. Referências Bibliográficas

- Agrawal, R., & Srikant, R. (1994). *Fast Algorithms for Mining Association Rules*. VLDB. Site: <https://snap.stanford.edu/class/cs224w-readings/Agrawal94AssosiationRule.pdf>
- Han, J., Pei, J., & Kamber, M. (2011). *Data Mining: Concepts and Techniques*. Morgan Kaufmann. – site: <https://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf>
- SBC. (2024). Disponível em: <https://www.sbc.org.br/wp-content/uploads/2024/07/2024-Manual-de-Publicacoes.pdf>