

Sistema de Recomendação com Regras de Associação para a Plataforma Manga Play

Gabriel A. S. S. Torres¹, Leandro V. S. Rocha², Vinicius C. Oliveira³

Departamento de Sistemas de Informação
Centro Universitário de Excelência (UNEX) – Feira de Santana, BA– Brasil

`gabriel.arlisson@hotmail.com, leandrovit.principal@gmail.com,`
`voliveira170@gmail.com`

Abstract. *This paper presents a movie recommendation approach based on the Apriori algorithm, widely used for association rule mining. The method aims to identify frequent patterns in user preferences to suggest movies based on past choices. Using a dataset containing user ratings, the model extracts relevant combinations of watched movies and generates personalized recommendations. The results show that this technique is effective in discovering meaningful associations, contributing to more accurate recommendation systems tailored to each user's individual profile.*

Resumo. *Este trabalho apresenta uma abordagem de recomendação de filmes utilizando o algoritmo Apriori, amplamente conhecido por sua aplicação em mineração de regras de associação. A proposta consiste em identificar padrões frequentes nos gostos de usuários para sugerir filmes com base em preferências anteriores. Utilizando um conjunto de dados com avaliações de usuários, o modelo extrai combinações relevantes de filmes assistidos e gera recomendações personalizadas. Os resultados demonstram que a técnica é eficiente na descoberta de associações significativas, contribuindo para sistemas de recomendação mais precisos e adaptados ao perfil individual de cada usuário.*

1. Introdução

A explosão do consumo de conteúdo audiovisual por meio de plataformas de streaming transformou radicalmente a maneira como assistimos filmes e séries.

O acesso instantâneo, a diversidade de títulos e a personalização da experiência do usuário são características essenciais para manter o público engajado.

Dentro desse contexto, a Manga Rosa Inc. propõe a criação da Manga Play, sua plataforma de streaming, com um diferencial competitivo: um sistema de recomendação de filmes baseado no conteúdo assistido e avaliado pelos usuários.

Este projeto, como parte de um processo seletivo para trainees da empresa, tem como objetivo desenvolver do zero um Sistema de Recomendação Baseado em Regras de Associação, utilizando o Algoritmo Apriori sem o uso de bibliotecas prontas para sua implementação. O sistema deverá sugerir filmes ao usuário com base no seu histórico de preferências, considerando os filmes que ele avaliou positivamente. A proposta abrange tanto recomendações baseadas no histórico geral quanto no último filme bem avaliado.

2. Fundamentação Teórica

O Algoritmo Apriori é uma técnica clássica de mineração de dados usada para descobrir regras de associação entre itens em grandes conjuntos de dados. Seu funcionamento baseia-se na identificação de itens frequentemente combinados em transações, possibilitando inferências preditivas. A aplicação desse algoritmo é ampla, incluindo sistemas de recomendação, análise de mercado e detecção de padrões.

Métricas principais:

- Suporte (support): indica a frequência com que um conjunto de itens ocorre no total de transações.

$$\text{support}(A \Rightarrow B) = (\text{n}^\circ \text{ de transações contendo } A \cup B) / (\text{n}^\circ \text{ total de transações})$$

- Confiança (confidence): mede a probabilidade de o item B ser selecionado dado que A já foi.

$$\text{confidence}(A \Rightarrow B) = \text{support}(A \cup B) / \text{support}(A)$$

- Lift: mede o quão mais provável é a ocorrência de B dado A em comparação com a ocorrência de B independente de A.

$$\text{lift}(A \Rightarrow B) = \text{confidence}(A \Rightarrow B) / \text{support}(B)$$

Pseudocódigo do algoritmo Apriori:

Apriori(D, min_sup):

L1 = {itens frequentes de tamanho 1 em D com suporte \geq min_sup}

k = 2

Enquanto L(k-1) \neq vazio:

Ck = candidatos gerados a partir de L(k-1)

Para cada transação t em D:

Ct = subconjuntos de Ck contidos em t

Para cada candidato c em Ct:

contar(c)++

Lk = candidatos em Ck com suporte \geq min_sup

k++

Retornar $\bigcup L_k$

3. Metodologia

A metodologia adotada neste trabalho compreende as seguintes etapas:

Unificação dos Dados: todos os conjuntos de dados relacionados aos usuários, avaliações e catálogo de filmes foram integrados em um único dataset estruturado. A unificação permite a análise cruzada das informações e facilita o uso do algoritmo Apriori para encontrar padrões relevantes. **Pré-processamento:** foram aplicadas as seguintes técnicas de pré-processamento:

- Filtragem de avaliações: apenas avaliações com nota superior a 3 foram consideradas.
- Conversão de dados: os dados foram transformados em uma estrutura transacional.
- Remoção de duplicatas e inconsistências.

Construção do Sistema de Recomendação: o sistema foi desenvolvido com base no Apriori, programado do zero em Python, utilizando apenas bibliotecas auxiliares como pandas e numpy. O processo consistiu em:

1. Geração dos conjuntos frequentes com suporte mínimo.
2. Extração das regras de associação com base na confiança e lift.
3. Recomendações personalizadas para cada usuário.

4.Resultados e Discussões

O sistema de recomendação foi avaliado utilizando duas abordagens distintas: uma baseada no histórico completo de avaliações positivas do usuário, e outra considerando apenas o último filme bem avaliado.

Na primeira abordagem, foram utilizados valores mínimos de suporte igual a 0,10 e confiança igual a 0,15, permitindo a geração de regras mais amplas e representativas do perfil geral do usuário. Essa estratégia proporcionou recomendações mais robustas, aproveitando o padrão de consumo acumulado.

Na segunda abordagem, foram utilizados suporte e confiança mínimos iguais a 0,10, com foco em gerar recomendações contextualizadas a partir da última interação positiva do usuário. Embora essa configuração gere regras mais específicas e menos abrangentes, mostrou-se eficiente para indicar conteúdos similares ao último item apreciado.

Ambas as abordagens foram capazes de gerar regras relevantes, e os valores de lift identificados indicaram que as recomendações apresentavam uma correlação significativa entre os itens, agregando valor à experiência do usuário na plataforma.

5.Considerações Finais

O desenvolvimento do sistema de recomendação baseado no algoritmo Apriori, implementado do zero, mostrou-se viável e eficaz para plataformas de streaming como a Manga Play. Destacam-se como pontos positivos a personalização das recomendações, a adaptação ao perfil do usuário e a independência de bibliotecas externas.

Como limitações, observou-se a necessidade de ajuste fino nos parâmetros de suporte e confiança, além do desempenho limitado em grandes volumes de dados.

Como melhorias futuras, propõe-se o uso de técnicas híbridas, unindo regras de associação a métodos baseados em conteúdo ou filtragem colaborativa, e a inclusão de feedback do usuário para refinar continuamente as sugestões.

Referência

- CAMILO, C. O.; SILVA, J. C. da. Mineração de dados: conceitos, tarefas, métodos e ferramentas. Universidade Federal de Goiás (UFG), 2009.
- ROMÃO, W. et al. Extração de regras de associação em C&T: o algoritmo Apriori. In: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO (ENEGEP), 19., 1999. Anais [...]. p. 37–39.