

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

Unex - Centro Universitário de Excelência - BA

Matheus Kaick - Henrique Estrela - Arthur Anjos

arthuraquino303@gmail.com - henriqueestrela2004@gmail.com - matheuskaick70@gmail.com

Resumo:

Este trabalho apresenta o desenvolvimento de um sistema de recomendação de filmes baseado no algoritmo Apriori de Regras de Associação. O sistema foi projetado para a plataforma fictícia de streaming *Manga Play*, com o objetivo de sugerir novos filmes a partir do histórico de preferências dos usuários. O modelo utiliza métricas como suporte, confiança e lift para gerar recomendações relevantes e personalizadas.

1. Introdução:

A arte, em suas múltiplas formas, constitui uma das mais ricas formas de expressão humana. O cinema, enquanto manifestação artística, proporciona experiências imersivas e emocionantes, conectando o público a histórias, culturas e emoções diversas. Com o avanço das tecnologias digitais, o consumo de conteúdo audiovisual passou por transformações significativas, especialmente com a popularização das plataformas de streaming.

Nesse novo cenário, a escolha de conteúdos tornou-se um desafio. Diante de um catálogo cada vez mais vasto, encontrar obras que atendam aos gostos individuais nem sempre é uma tarefa simples. Assim, surge a necessidade de sistemas de recomendação capazes de compreender as preferências dos usuários e sugerir conteúdos relevantes de forma eficiente. Este trabalho apresenta o desenvolvimento de um sistema de recomendação baseado em conteúdo para a plataforma Manga Play, utilizando Regras de Associação como heurística principal. O objetivo é recomendar filmes com base no histórico de títulos previamente assistidos e marcados como “gostados” pelos usuários.

O algoritmo Apriori é uma das técnicas mais conhecidas para a descoberta de padrões frequentes em grandes conjuntos de dados. Ele é amplamente utilizado na mineração de regras de associação, sendo eficiente para identificar relações entre itens em bases transacionais — neste caso, filmes assistidos pelos usuários. O algoritmo opera de forma

iterativa, gerando conjuntos frequentes de itens (itemsets) que satisfazem um valor mínimo de suporte. A partir desses itemsets, são extraídas regras de associação com base em métricas como suporte, confiança e lift.

2. Fundamentação Teórica:

O algoritmo Apriori é amplamente utilizado na área de mineração de dados para a identificação de padrões frequentes em grandes conjuntos de dados. Sua operação baseia-se em um processo iterativo, no qual são gerados conjuntos de itens que ocorrem com frequência superior a um limite mínimo pré-estabelecido, conhecido como suporte. A partir desses conjuntos frequentes, é possível extrair regras de associação fundamentadas em três métricas principais: suporte, confiança e lift.

O suporte indica a frequência com que um item ou conjunto de itens aparece na base de dados. A confiança expressa a probabilidade de ocorrência de um item, dado que outro já ocorreu. Já o lift avalia a relevância da regra ao comparar a confiança obtida com a frequência esperada do item consequente, permitindo identificar relações mais significativas entre os itens.

O processo tem início com a identificação dos itens individuais mais frequentes. Em seguida, gera-se iterativamente combinações de maior tamanho a partir dos conjuntos anteriores, mantendo apenas aquelas que continuam atendendo ao suporte mínimo. O algoritmo continua até que não seja mais possível formar novos conjuntos frequentes. Com base nesses resultados, as regras de associação são então geradas.

3. Metodologia:

Para desenvolver esse sistema de recomendação de filmes, eu comecei carregando dois conjuntos de dados: um com as avaliações dos usuários (**ratings_small.csv**) e outro com as informações dos filmes (**movies_metadata.csv**). Primeiro, fiz um pré-processamento nesses dados para manter apenas as colunas que realmente interessavam — como o ID do usuário, o ID do filme e a nota dada. Também limpei os dados de metadata, convertendo o ID do filme para o tipo numérico e tratando valores inválidos. Depois disso, fiz a junção dos dois conjuntos, usando o ID do filme como chave, e filtrei apenas os filmes que os usuários deram nota maior ou igual a 3, considerando esses como "curtidos".

A partir daí, criei um dicionário onde cada usuário tem uma lista com os filmes que curtiu, formando o que seria o nosso "dataset" de transações, como se fosse um carrinho de compras em um mercado — só que, nesse caso, cada carrinho tem os filmes que o usuário gostou.

Depois defini os parâmetros para o algoritmo Apriori: o suporte mínimo (quantas vezes um item aparece no total), a confiança mínima (a chance de, dado um filme curtido, outro também ser curtido), e o número máximo de níveis que o Apriori vai explorar — nesse caso até 4, para não gerar muitas combinações e sobrecarregar o tempo de execução. Também pedi para o usuário informar um ID de exemplo para gerar as recomendações.

Implementei manualmente as funções principais do algoritmo: uma função para gerar combinações de filmes (**generate_combinations**), outra para calcular o suporte (**count_support**), mais duas para confiança e lift (que é uma métrica que indica a força da regra em comparação ao acaso). A função principal do Apriori (**get_frequent_itemsets_by_level**) percorre essas combinações e retorna os itemsets frequentes por nível — primeiro os filmes individualmente, depois duplas, trios, etc., até o limite definido.

4. Resultados e Discussões:

Com a aplicação do algoritmo Apriori utilizando um suporte mínimo de 10% e confiança mínima de 50%, foi possível identificar uma quantidade significativa de padrões entre os filmes curtidos pelos usuários. O processo resultou em 4.141 itemsets frequentes e 5.151 regras de associação, o que evidencia a existência de gostos similares entre diferentes usuários e a recorrência de certas combinações de filmes. Um exemplo relevante foi a regra que indica que usuários que curtiram Back to the Future Part II têm 71% de chance de também curtir 48 Hrs., com um lift de 2,48 — valor que demonstra uma associação significativamente mais forte do que o esperado por acaso.

Avaliando o usuário de ID 5, que havia curtido 51 filmes, o sistema conseguiu gerar recomendações personalizadas como Monsoon Wedding, Terminator 3: Rise of the Machines, Sissi e Solaris. Essas sugestões surgiram de regras baseadas em múltiplos filmes do seu histórico, o que reforça a relevância das recomendações. No entanto, ao tentar recomendar com base apenas no último filme curtido, não foram encontradas regras fortes o suficiente, levando o sistema a recorrer ao fallback com sugestões de filmes populares que o usuário ainda não havia assistido, como Solaris e The 39 Steps.

5. Considerações Finais:

A aplicação do algoritmo Apriori para a construção de um sistema de recomendação de filmes apresentou resultados promissores. A partir do histórico de filmes curtidos pelos usuários, foi possível gerar regras de associação relevantes e utilizá-las para oferecer sugestões personalizadas. O uso de métricas como suporte, confiança e lift contribuiu para validar a qualidade das recomendações, garantindo que as associações identificadas fossem, de fato, significativas. A implementação também mostrou que é viável construir uma solução eficaz com base em técnicas simples e bem compreendidas.

Apesar dos resultados positivos, o projeto apresentou algumas limitações, especialmente em relação ao tempo de processamento, que tende a crescer com o aumento do número de usuários e itens no conjunto de dados. Como melhoria futura, recomenda-se otimizar a implementação do algoritmo e investigar abordagens mais escaláveis. Também seria interessante explorar técnicas mais avançadas, como modelos baseados em aprendizado de máquina, para enriquecer a capacidade preditiva do sistema. Além disso, incluir variáveis adicionais, como gêneros dos filmes, duração ou avaliações textuais, pode tornar as recomendações ainda mais precisas e personalizadas.

Links das Referências:

Agrawal, R. and Srikant, R. (1994) "Fast Algorithms for Mining Association Rules", In: *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB)*, Santiago, Chile, Morgan Kaufmann, pp. 487–499.

Han, J., Kamber, M. and Pei, J. (2011) *Data Mining: Concepts and Techniques* (3rd ed.), Morgan Kaufmann Publishers, USA.

Tan, P. N., Steinbach, M. and Kumar, V. (2018) *Introduction to Data Mining* (2nd ed.), Pearson Education, USA.

Ricci, F., Rokach, L., Shapira, B. and Kantor, P. B. (2011) *Recommender Systems Handbook*, Springer, New York, USA.

GroupLens Research (2016) "MovieLens Dataset", <https://grouplens.org/datasets/movielens/>, Accessed: May 2025.