Aplicação do Algoritmo Apriori em Sistemas de Recomendação Baseados em Avaliações de Usuário

Mickael Cedraz Alencar, Nalbert de Souza Santana

**Resumo.** Este meta-artigo descreve o estilo a ser usado na confecção de artigos e resumos de artigos para publicação nos anais das conferências organizadas pela SBC. É solicitada a escrita de resumo e abstract apenas para os artigos escritos em português. Artigos em inglês deverão apresentar apenas abstract. Nos dois casos, o autor deve tomar cuidado para que o resumo (e o abstract) não ultrapassem 10 linhas cada, sendo que ambos devem estar na primeira página do artigo.

1. Introdução

Os sistemas de recomendação se tornaram parte essencial da experiência digital contemporânea cada vez mais empresas como Netflix ou Spotify investem constantemente em algoritmos capazes de otimizar a navegação dos usuários, buscando produtos, serviços ou conteúdos que eles talvez tenham interesse com base em seus históricos de consumo, o que aumenta o engajamento e a satisfação do usuário.

Tendo isso em mente, o projeto tem como objetivo principal o desenvolvimento de um sistema de recomendação de filmes utilizando o algoritmo de associação Apriori. A implementação proposta considera uma base de dados de filmes contendo informações como título, avaliações de usuários e *tags* associadas ao filme, e aplica técnicas de pré-processamento para filtragem, integração e transformação desses dados em transações adequadas para a aplicação do algoritmo.

1. Fundamentação Teórica

O algoritmo Apriori é uma técnica clássica de mineração de dados utilizada com o intuito de encontrar grupos ou padrões frequentes de itens em grandes conjuntos transacionais.

2.1 Conceitos Fundamentais

Esses grupos podem ser usados para definir regras de associação, sendo útil em contextos como análise de padrões de compras, detecção de comportamentos, e, como será neste trabalho, sistemas de recomendação de produtos.

Desse modo, o algoritmo funciona através da busca de itemsets frequentes, que são os conjuntos de itens que correm juntos em um número mínimo de transações, definidos pelo parâmetro escolhido de suporte*.*

O suporte (*support*) é a métrica que mede a proporção de transações em que um determinado item ou conjunto de itens aparece, ou seja, indica o quão comum ele é no conjunto total de dados.

A confiança (*confidence*) é calculada com base no suporte, expressando a probabilidade da ocorrência de um item Y sendo que X já foi selecionado. Ela define a força de uma regra de associação entre dois itens.

O *lift* mede a importância de uma regra levando em conta a ocorrência independente dos itens, indicando a chance de Y ser selecionado, caso X tenha sido, ajustada para a popularidade geral de Y no conjunto.

2.2 Pseudocódigo

Apriori(D, min\_sup):

L1 = {itens frequentes de tamanho 1 em D com suporte ≥ min\_sup}

k = 2

enquanto L(k-1) não estiver vazio:

Ck = gerar\_combinacoes(L(k-1), k)

para cada transação t em D:

para cada candidato c em Ck:

se c ⊆ t então incrementar contador de c

Lk = {c em Ck com suporte ≥ min\_sup}

k = k + 1

retornar ∪ Lk

1. Metodologia

O desenvolvimento do sistema de recomendação baseado no algoritmo Apriori seguiu uma série de etapas, incluindo a seleção e a integração de dados, o pré-processamento de informações, e por fim, a aplicação do algoritmo para a geração de regras de associação.

Foi utilizado o banco de dados disponibilizado para o desenvolvimento do trabalho contendo as avaliações, informações detalhadas sobre os filmes, as correspondências entre IDs e outros dados importantes.

A primeira etapa consistiu na filtragem das avaliações, e integração ´de uma nova estrutura de dados na qual cada usuário está associado a uma lista de filmes que avaliou positivamente (acima ou igual à 4 estrelas), essas listas de filme por usuários foram transformadas em transações, de modo que cada transação representa uma “cesta” de itens.

Com as transformações preparadas, foi aplicada a implementação para identificar conjuntos frequentes, gerando combinações de tamanho 1 e 2 para os itens frequentes. As regras foram construídas com base no cálculo da confiança, considerando como válidas aquelas cujo valor fosse igual ou superior a 0.5.

1. Resultados e Discussões

O algoritmo gerou regras extraídas para o formato de item X para item Y, seguindo a lógica de que um usuário que gostou do filme X tem uma probabilidade considerável de que também goste do filme Y. Alguns dos exemplos de regras extraídas incluem:

* Se gostou de 'Twin Peaks: Fire Walk with Me', então pode gostar de 'Solaris' (Confiança: 0.60)
* Se gostou de 'Three Colors: Red', então pode gostar de 'Solaris' (Confiança: 0.80)

A abordagem trouxe vantagens, como simplicidade de interpretação, mas, ao mesmo tempo, trouxe limitações devido à restrição de volume de dados, adotada para permitir a execução local.

1. Considerações Finais

No trabalho foi desenvolvido um sistema de recomendações de filmes com base em regras de associação utilizando o algoritmo Apriori. A abordagem demonstrou eficácia para identificar os padrões de preferência entre usuários, levando em conta as suas avaliações, mesmo com um volume reduzido de dados.

A implementação foi rápida e simples, a claridade das regras geradas e da base teórica utilizada resultou em uma interpretação transparente e facilitada.