

Resumo – Algoritmo Apriori

Aluno: Carlos Gabriel Santos Nunes.

Um dos principais algoritmos para a geração das regras de associação para a mineração de dados. Foi proposto em 1994 pela equipe de pesquisa do Projeto QUEST da IBM que deu origem ao *Software Intelligent Miner* [2]. Este algoritmo faz recursivas buscas no Banco de Dados à procura dos conjuntos frequentes.

As três fases que compõem o Apriori são:

- Geração dos conjuntos Candidatos;
- Poda dos conjuntos Candidatos;
- Contagem do Suporte (necessário visitar o *database*).

A este algoritmo é aplicada a propriedade de Antimonotonia da Relação ou Propriedade Apriori que é descrita a seguir: Se X está contido em Y e X não é frequente, logo Y também não é frequente.

Seu objetivo é gerar conjuntos de itens candidatos (padrões) de k elementos a partir de conjuntos de itens $k - 1$ elementos, padrões não frequentes são eliminados. Toda a base de dados é rastreada e os conjuntos de itens frequentes são obtidos a partir de itens candidatos. Segue abaixo uma imagem do algoritmo.

Algoritmo 1: Algoritmo Apriori

```
 $F_1 \leftarrow \{\text{Conjuntos de itens freqüentes de tamanho 1}\} \quad /* \text{ Na}$   
primeira passagem  $k = 1$  */  
1 para  $k = 2$ ;  $F_{k-1} \neq \text{vazio}$ ;  $k++$  faça  
    $/* \text{ Na segunda passagem } k = 2$  */  
2    $C_k \leftarrow \text{apriori-gen}(F_{k-1}) \quad /* \text{ Novos candidatos}$  */  
3   para todo transação  $t \in T$  faça  
4      $C_t \leftarrow \text{subconjunto}(C_k, t) \quad /* \text{ Candidatos contidos}$   
       em  $t$  */  
5     para todo candidato  $c \in C_t$  faça  
6        $c.\text{contagem}++$   
7     fim  
8      $F_k \leftarrow \{c \in C_k | c.\text{contagem} \geq \text{MinSup}\}$   
9   fim  
10 fim  
11 Resposta  $F \leftarrow \text{Reunião de todos os } F_k$ 
```

Fonte: (Vasconcelos; Carvalho, 2004)

Os níveis de confiança e suporte são essenciais para o funcionamento do algoritmo. Eles vão determinar diretamente tanto a quantidade como a qualidade das regras geradas.

Seu funcionamento consiste em selecionar todos os conjuntos que possuem itens frequentes, isto é, itens que satisfaçam o nível de suporte pré-estabelecido. Após isso o próximo passo é em obter as regras de associação ($X \Rightarrow Y$), para isso é necessário encontrar quais itens que satisfaçam o nível de confiança mínimo pré-estabelecido, ou seja, o percentual de vezes que o conjunto X aparece na base

Referências

[1] L. M. R. de Vasconcelos, C. L. de Carvalho. **Aplicação de Regras de Associação para Mineração de Dados na Web.** Disponível em: http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_004-04.pdf

[2] Agrawal, R., Srikant, R. **Fast Algorithms for Mining Association Rules.** Disponível em: <http://rakesh.agrawal-family.com/papers/vldb94apriori.pdf>