

Resumo do processamento de descoberta de conhecimento

Algoritmo A-priori

Disciplina: Banco de Dados II

Aluno: Albert Medeiros de Paula

Professor: Hiran Nonato

Resumo

Resumidamente, o algoritmo Apyori tem o objetivo de resolver problemas de regras de associação. As regras de associação por sua vez, servem para a partir de um banco de dados, retornar informações que podem, por exemplo, maximizar a venda de um estabelecimento. O algoritmo Apyori é utilizado para automatizar e realizar testes de regras de associação e retorná-las aos usuários.

Para encontrar regras relevantes, são utilizados 4 variáveis para análise, Support, Confidence, Lift, Conviction. O Support é a medida que indica a proporção de um objeto em relação aos outros. Confidence é o cálculo que mostra em porcentagem a chance que após um produto ser comprado, outro ser levado junto. O Lift refere-se a chance de um conjunto de produtos serem comprados levando em consideração a sua popularidade de venda. Por fim, o Conviction é o cálculo que mostra a frequência que um produto ocorre e outro não.

O algoritmo calcula o Support de cada item, salvando o conjunto de objetos que contém o valor maior do que o passado como o parâmetro. Assim sucessivamente com as combinações de itens.

Para este resumo foi utilizado o banco de dados de uma pizzaria. A extração de regras foi feita utilizando a linguagem Python juntamente com as plataformas Anaconda e Spyder. O código utilizado para a extração está disponibilizado no link <https://github.com/albert-medeiros/apriori>. Foram feitos vários testes utilizando vários parâmetros, porém, vale ressaltar o teste que utilizou como parâmetro para os valores de Support igual 0.3, Confidence com 0.8 e Lift a 1.

Com tais parâmetros, foram encontrados 7 valores, 2 de produtos sozinhos e 5 de combinação de 2 ou mais produtos juntos. Foram encontradas que, Sem borda, sem refrigerante, se é entrega então é sem borda, se o tempo de demora é de 33 a 44 então sem borda, se é pra buscar então sem refrigerante, se demora de 44 a 55 então refrigerante não, e por fim, se é sem borda e é pra buscar então sem refrigerante, são as regras descobertas pelo algoritmo devido aos parâmetros.

Portanto, tendo esses valores em mente, a pizzaria poderia ver opções para subir as vendas de bordas e refrigerantes, chamando a atenção de seus clientes, maximizando o lucro. Sem o uso de tais algoritmos de mineração de dados, o processo de descoberta seria lento e às vezes não eficaz, deste modo, seu uso é imprescindível para a época que mais gera dados na história.

