

Mineração de Dados - Trabalho Extra

Luis Felipe de Deus¹, Nathanael Luchetta¹, Tiago Knorst¹, Yuri Oliveira¹

¹Curso de Engenharia de Computação – Universidade Federal de Santa Maria

{felipe.deus@ecomp.ufsm.br, naaatha_@outlook.com, tiago.knorst@ecomp.ufsm.br,
yuri.alves@ecomp.ufsm.br}

1. REGRAS DE ASSOCIAÇÃO:

1.1. Algoritmo Eclat

Eclat é um algoritmo para mineração de dados de itens frequentes, consiste em um método de mineração de dados que foi desenvolvido para analisar uma cesta de mercado. A mineração frequente de um conjunto de itens, visava encontrar regularidades no comportamento de compra dos clientes de um supermercado ou lojas on-line. Particularmente, ele tenta identificar um conjunto de itens que são frequentemente comprados juntos. E uma vez identificados, esses conjuntos de produtos associados podem ser utilizados para otimizar a organização dos produtos em uma prateleira de supermercado, ou na página inicial de lojas virtuais, possibilitando dicas de quais produtos convenientemente poderiam ser agrupados, ou também facilitar para um cliente já conhecido, onde se encontram os itens que ele compra com mais frequência.

No entanto, a mineração frequente de um conjunto de itens pode ser utilizada para uma variedade muito maior de dados, desde que compartilhem o interesse de encontrar regularidades entre variáveis desta unidade de dados. este algoritmo está sendo desenvolvido e aprimorado desde 2002, baseado em um esquema básico desenvolvido por volta de 1997.

1.2. Comparação com algoritmo Apriori

Comparado com o algoritmo Apriori, que sempre escaneia todo o dataset para calcular o suporte, o Eclat escaneia somente o dataset útil no momento da mineração, diminuindo o tempo de cálculo do suporte de regras que será utilizado.

Enquanto o algoritmo Apriori trabalha com uma exploração do dataset horizontalmente, como um algoritmo de busca em largura, o algoritmo Eclat trabalha

de maneira vertical, assim como um algoritmo de busca em profundidade. Isso significa, na prática, que ao invés de gerar os valores de suporte e confiança explorando item a item, como ocorre no Apriori, a geração é feita abrindo-se nodos nas a partir das transações.

Desse modo, o algoritmo Eclat corta ramificações infrequentes já antes de criá-las, o que lhe dá a vantagem de precisar analisar um dado nodo uma única vez durante sua execução. Outra vantagem inerente ao algoritmo é o menor requerimento de memória, uma vez que a execução é feita verticalmente de maneira recursiva.

2. RESULTADOS

Com o objetivo de obter os jogadores, duplas e equipes mais e menos vitoriosos, foi executado o pré-processamento dos dados no *dataset*, efetuando a limpeza e transformação dos dados em um *data frame* no formato de lista de presença.

A mineração de dados foi efetuada gerando regras de associação por meio de algoritmo Eclat, que foi descrito na seção anterior. A primeira inferência constatada, foi de que a regra de que a jogadora Ana era a mais vitoriosa, tinha um suporte de 30% e uma confiança de 66%. Seguindo foi então usado como parâmetro Ana para determinar a melhor dupla, encontrando Ágata, a regra gerada para vitórias de que juntas, Ana e Ágata conseguiram teve suporte de 14% e confiança de 67%. Por fim, o terceiro integrante da equipe, onde Ana e Ágata estavam, foi Fábio, com um suporte de 2,66% e uma confiança de 80%. Os resultados podem ser visualizados na Tabela 1.

Como o objetivo era encontrar tanto os melhores, quanto os piores jogadores, foi então gerado novas regras, procurando agora com o resultado de derrota, assim foi encontrado Fábio como sendo o pior jogador, com um suporte de 22% e uma confiança de 52%, pode-se destacar que tanto Bárbara, quanto Shelda tiveram um suporte entre 13% e 14%, com uma confiança maior que 60 % e poderiam ser resultados válidos.

Usando Fábio como parâmetro, encontra-se a pior dupla sendo Fábio e Bárbara com um suporte de 12% e confiança de 54%, e por último, a pior equipe é composta por Fábio, Bárbara e Ricardo, com 1,3% de suporte e 66% confiança, existia uma regra onde o terceiro integrante era Fábio, com um suporte de 2,6%, entretanto a confiança era de 40%. Os resultados podem ser visualizados na Tabela 2.

Tabela 1 - Resultados de melhores jogadores

Melhor Jogador	Suporte (%)	Confiança (%)
Ana	30	66
Melhor Dupla	Suporte	Confiança
Ana, Ágata	14	67
Melhor Trio	Suporte	Confiança
Ana, Ágata, Fábio	2,66	80

Tabela 2 - Resultados de piores jogadores

Pior Jogador	Suporte (%)	Confiança (%)
Fábio	22	52
Pior Dupla	Suporte	Confiança
Fábio, Bárbara	12	54
Pior Trio	Suporte	Confiança
Fábio, Bárbara, Ricardo	1,3	66