MINERAÇÃO DE PADRÕES FREQUENTES EM BASES DE DADOS COM ALTA DIMENSIONALIDADE

LAURO ANGELO GONCALVES DE MORAES (Orientador), Joubert de Castro Lima (Orientador)

Neste trabalho é apresentado um estudo descritivo e comparativo de alguns dos métodos de mineração de conjuntos de itens frequentes fechados, um tipo de padrão frequente importante na área de mineração de dados. Trata-se de um conjunto de itens que representa, sem perda de informação ou generalidade, todo o conjunto de itens frequentes de uma base de dados. Alguns algoritmos foram propostos para realizar esta tarefa de mineração, tais como o AprioriClose (Pasquier et al., 1999), AprioriTIDClose (Pasquier et al., 1999), LCM (Uno et al., 2005), DCI Closed (Luccheseand Orlando, 2004) e FPClose (Grahne, 2005). São descritos os paradigmas que guiam estes algoritmos, suas propriedades, vantagens e limitações. O algoritmo TD-Close foi devidamente implementado seguindo as orientações presentes em Liu et al. (2009). Os códigos de todos os algoritmos citados foram testados e avaliados, utilizando as métricas de consumo de tempo de execução e consumo de memória, possibilitando uma comparação analítica entre os resultados destes em relação a bases de dados reais e sintéticas com variações relativas à dimensionalidade, número de tuplas e cardinalidade de suas dimensões. Como mostrado em Liu et al. (2009) o TD-Close se mostrou eficiente em encontrar conjuntos de itens frequentes com alta dimensionalidade em bases de dados que representam microarranjos. Este tipo de base possui muitas dimensões e reduzido número de tuplas.

Instituição de Ensino: Universidade Federal de Ouro Preto

ISSN: 21763410