

Frequent Itemset Mining Using Rough-Sets

Authors : Usman Qamar, Younus Javed

Abstract : Frequent pattern mining is the process of finding a pattern (a set of items, subsequences, substructures, etc.) that occurs frequently in a data set. It was proposed in the context of frequent itemsets and association rule mining. Frequent pattern mining is used to find inherent regularities in data. What products were often purchased together? Its applications include basket data analysis, cross-marketing, catalog design, sale campaign analysis, Web log (click stream) analysis, and DNA sequence analysis. However, one of the bottlenecks of frequent itemset mining is that as the data increase the amount of time and resources required to mining the data increases at an exponential rate. In this investigation a new algorithm is proposed which can be uses as a pre-processor for frequent itemset mining. FASTER (FeAture SelecTion using Entropy and Rough sets) is a hybrid pre-processor algorithm which utilizes entropy and rough-sets to carry out record reduction and feature (attribute) selection respectively. FASTER for frequent itemset mining can produce a speed up of 3.1 times when compared to original algorithm while maintaining an accuracy of 71%.

Keywords : rough-sets, classification, feature selection, entropy, outliers, frequent itemset mining

Conference Title : ICCSIM 2014 : International Conference on Computer Science and Information Management

Conference Location : London, United Kingdom

Conference Dates : October 24-25, 2014