Anytime Itemset Mining with Definite Guarantees

Qiong Hu
Rutgers University

5/13/2016 at 11:00 am
CoRE B (305)

Abstract

Mining very wide data sets with hundreds of thousands of attributes may require contiguous computation with no real time bounds. Usually, the itemset mining algorithms produce outputs only at the completion (either run to completion or provide no useful results) and are not amenable to real-time decision-making need. One solution is to turn the basic data mining methods into anytime algorithms. In literature, there are numerous anytime algorithms for classification, regression, clustering, etc. However, the direction of anytime association mining is largely unexplored.

We defined the anytime itemset mining and proposed the ALPINE algorithm, which proceeds in the defined anytime mining manner and can be interrupted at any time but offer intermediate meaningful and complete results with definite guarantees. ALPINE is, to our knowledge, the first interruptible anytime algorithm to mine frequent itemsets and closed frequent itemsets. It guarantees that all itemsets with support exceeding the current checkpoint’s support have been found before it proceeds further. This ANYTIME feature is the most important contribution of ALPINE, which is also fast but not necessarily the fastest algorithm around. ALPINE runs literally "forever" and without apriori decided minsup value.

Examination Committee: Prof. Tomasz Imielinski (Chair), Prof. Alex Borgida, Prof. Shan Muthukrishnan and Prof. Mubbasir Kapadia