ABSTRACT
In this paper, we present a novel algorithm for mining complete frequent itemsets. This algorithm is referred to as the TM algorithm from hereon. In this algorithm, we employ the vertical representation of a database. Transaction ids of each itemset are mapped and compressed to continuous transaction intervals in a different space thus reducing the number of intersections. When the compression coefficient becomes smaller than the average number of comparisons for intervals intersection, the algorithm switches to transaction id intersection. We have evaluated the algorithm against two popular frequent itemset mining algorithms - FP-growth and dEclat using a variety of data sets with short and long frequent patterns. Experimental data show that the TM algorithm outperforms these two algorithms. In addition, we propose a new pruning approach using a 3-D bit array. In this paper we also introduce a simple probabilistic paradigm for the analysis of association rules mining algorithms.

1 This research has been supported in part by the NSF Grants CCR-9912395 and ITR-0326155.