Probabilistic Static Load-Balancing of Parallel Mining of Frequent Sequences

Frequent sequence mining is well known and well-studied problem in data mining. The output of the algorithm is used in many other areas like bioinformatics, chemistry, and market basket analysis. Unfortunately, the frequent sequence mining is computationally quite expensive. In this paper, we present a novel parallel algorithm for mining of frequent sequences based on a static load-balancing. The static load-balancing is done by measuring the computational time using a probabilistic algorithm. For reasonable size of instance, the algorithms achieve speedups up to where is the number of processors. In the experimental evaluation, we show that our method performs significantly better than the current state-of-the-art methods. The presented approach is very universal: it can be used for static load-balancing of other pattern mining algorithms such as item set/tree/graph mining algorithms.