Efficient Frequent Itemset Mining Methods over Real-Time Spatial Big Data

Authors : Hamdi Sana, Emna Bouazizi, Sami Faiz

Abstract : In recent years, there is a huge increase in the use of spatio-temporal applications where data and queries are continuously moving. As a result, the need to process real-time spatio-temporal data seems clear and real-time stream data management becomes a hot topic. Sliding window model and frequent itemset mining over dynamic data are the most important problems in the context of data mining. Thus, sliding window model for frequent itemset mining is a widely used model for data stream mining due to its emphasis on recent data and its bounded memory requirement. These methods use the traditional transaction-based sliding window model where the window size is based on a fixed number of transactions. Actually, this model supposes that all transactions have a constant rate which is not suited for real-time applications. And the use of this model in such applications endangers their performance. Based on these observations, this paper relaxes the notion of window size and proposes the use of a timestamp-based sliding window model. In our proposed frequent itemset mining algorithm, support conditions are used to differentiate frequents and infrequent patterns. Thereafter, a tree is developed to incrementally maintain the essential information. We evaluate our contribution. The preliminary results are quite promising.