A Sliding Window Based Closed Frequent Itemset Mining Algorithm Over Data Streams for Interestingness-Rich Association Rules

Ch.Ravikant, VCE, Hyderabad, India

Abstract

A data stream is a continuous sequence of data elements generated from a specified source. Mining frequent itemsets in dynamic databases and data streams encounters some challenges that make the mining task harder than static databases. In the existing work, an enhanced association rule mining algorithm was proposed to validate the rules utility and consistency by introducing new weightage validation. This method more accurately extracts the rules based on their utility and frequency. But this method has the drawback of memory usage and processing time. Because, in data streams data elements are arrive at a rapid rate. The incoming data is unbounded and probably infinite. Due to high speed and large amount of incoming data, frequent itemset mining algorithm must require a limited memory and processing time. To reduce this drawback in the existing method, a new algorithm is proposed in this paper.

Here, a new algorithm is named as CFIM is developed for mining closed frequent itemsets from the data streams based on their utility and frequency. The DataStream are accessed by the sliding window and in every sliding window the closed frequent itemsets is mined based on the utility and consistency validation process. During the closed frequent itemsets mining, a hash table is maintained to check whether the given itemset is closed or not. The computation of closed frequent itemsets from the data stream will minimize the memory usage and processing time. Since, set of frequent closed itemsets has smaller size rather than complete set of frequent patterns while it contains the same information. That is, the complete set of frequent itemsets can be induced by closed frequent itemsets. Therefore, closed itemset mining over data streams is more desirable than finding the complete set of frequent itemsets. Thus our proposed technique performance is analyzed by using the synthetic dataset and compared with the exiting closed frequent itemset mining techniques.

Keywords: Closed Frequent item, Motifs, Data Streams, mining algorithm, intelligent Data miner, high utility item, Doubly Linked Tree, E-Learning.
References


20) Kuthadi, V.M (2013). Detection of proficient and distinct motifs in sequence data sets using PSM. In Intelligent Systems and Control (ISCO), 2013 7th International Conference on, pp. 354-358. IEEE.


