Abstract

The processes of extracting knowledge structures for continuous, rapid records are known as the Data Stream Mining. The main issue in stream mining is handling streams of elements delivered rapidly which makes it infeasible to store everything in active storage. To overcome this problem of handling voluminous data we exposed a novel load shedding system using window based aggregate function of the data stream in which we accept those tuples in the stream that meet a criterion. Accepted tuples are conceded to another process as a stream, while further tuples are dropped. This proposed model conceivably segregates the data input stream into windows and probabilistically decides which tuple to drop based on the window function. The best window aggregate function used for dropping tuples is identified with the three prediction models used in data mining they are Decision Tree, Naïve Bayes and Logistic Regression. The result shows that the cumulative distance and density rank functions outperforms the remaining methods. Distinct to prior methods, our method preserves uniformity of windows all over a query plan, and constantly distributes subsets of the original query responds with insignificant denial in the excellence of the consequence.
Load Shedding using Window Aggregation Queries on Data Streams

- Shengliang Xu, Magdalena Balazinska.: Sensor Data Stream Exploration for monitoring applications: DMSN’2011.
- Department of Transporation Coordinated Highways Action Respose Team, Maryland, http://www.chart.state.md.us/default.asp.
- Mohamed Medhat Gaber, Shonali Krishnaswamy, Arkady Zaslavsky: Adaptive mining Techniques for Data Stream Using Algorithm Output Granularity.

- Y. Yao and J. Gehrke,. Query processing for sensor networks,. in First Conf. on Innovative Data Systems Research (CIDR), January 2003.

- Hua-Fu Li, Suh-Yin Lee,. Mining frequent itemsets over data streams using efficient Window sliding techniques: Expert Systems with Applications 36 (2009) page(s). 1466-

Index Terms

Computer Science Information Sciences

Keywords

Data stream mining Windows functions Load Shedding Scheme