

عنوان مقاله:

Summarization Algorithm for Data Stream to Speed up Outlier Data Detection

محل انتشار:

مجله محاسبات و امنیت, دوره 10, شماره 1 (سال: 1402)

تعداد صفحات اصل مقاله: 12

نویسندگان:

.Hadid Mollashahi - Faculty of Electrical & Computer Engineering, University of Birjand, Birjand, Iran

.Hamid Saadatfar - Faculty of Electrical & Computer Engineering, University of Birjand, Birjand, Iran

.Hamed Vahdatnejad - Faculty of Electrical & Computer Engineering, University of Birjand, Birjand, Iran

خلاصه مقاله:

Outlier detection in data streams is an essential issue in data processing. Today, due to the massive growth of streaming data generated by the spread of the Internet of Things, outlier detection has become a significant challenge. Much progress has been made in outlier detection based on local outlier detection algorithms, such as density-based local outlier factor algorithms, suitable for static data. The incremental version of these algorithms is used to detect the local outliers in streaming data. However, outlier detection in streaming data faces the challenges of limited memory capacity, high execution time, inaccessibility of all data at one time, and changes in data distribution (increasing and decreasing input rates, uncertainty, etc.). In this paper, we propose a density-based summarization algorithm, which summarizes data, every time the buffer is filled. The proposed algorithm maintains the desired shape of the clusters, with a low computational cost. To this end, larger clusters are selected and the data of their dense areas are reduced so that the shape of the old clusters is not lost. The proposed summarization algorithm reduces execution time and increases precision, recall, and F1 score compared with the evaluated algorithms.

کلمات کلیدی:

Outlier detection, data stream, Machine Learning, Clustering, IoT

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/1947289>

