K-Means+ID3 Method for Detecting Anomalous Activities via Large Multi-dimensional Data Sets

(ROI # 2007-02)

Description

- K-Means+ID3 is an effective method for detecting anomalous activities via large multi-dimensional data sets collected from dynamic systems like computer networks or electronic circuits. The K-Means+ID3 method learns the normal and anomalous activities of a system by combining the k-Means clustering method and the ID3 decision trees. The k-Means clustering method partitions the data into 'k' distinct data groups. On each data group, an ID3 decision tree is built to learn the subgroup-structure within the data group. To obtain a final decision on anomaly detection, the decisions of k-Means and ID3 decision tree are combined using the nearest-neighbor rule or the nearest consensus rule.

Advantages

- Very fast and efficient detection of anomalous activities.
- The K-Means+ID3 method effectively detects rare anomalous activities
- The K-Means+ID3 method is data-driven method and therefore adapts to the changes in normal activities of a system.
- The K-Means+ID3 method provides a tunable threshold parameter for automatic enhancement or relaxation of decisions that alert anomalous activities.

Performance Summary: The K-Means+ID3 anomaly detection method was tested on multi-dimensional data sets collected by monitoring three systems: (1) a large computer network, (2) an electronic circuit implementing forced Duffing equation, and (3) a mechanical mass-beam structure subjected to fatigue crack damage. The detection accuracy of the K-Means+ID3 method was as high as 96.24 percent at a low false positive rate of 0.03 percent on network data; the total accuracy was as high as 80.01 percent on mass-beam structure data and 79.9 percent on electronic circuit data.

Areas of Application

- Detecting attacks on computer systems and networks; Detecting spurious traffic bursts at servers, routers, and other network devices; Detecting anomalies in complex electronic circuits; Detecting fatigue-generated cracks in mechanical structures; Detecting spurious events via data gathered by surveillance sensor networks.

Patent Status

- US 7,792,770