

Anomaly Detection Based Fuzzy K-Mode Clustering for Categorical Data

Authors : Murat Yazici

Abstract : Anomalies are irregularities found in data that do not adhere to a well-defined standard of normal behavior. The identification of outliers or anomalies in data has been a subject of study within the statistics field since the 1800s. Over time, a variety of anomaly detection techniques have been developed in several research communities. The cluster analysis can be used to detect anomalies. It is the process of associating data with clusters that are as similar as possible while dissimilar clusters are associated with each other. Many of the traditional cluster algorithms have limitations in dealing with data sets containing categorical properties. To detect anomalies in categorical data, fuzzy clustering approach can be used with its advantages. The fuzzy k-Mode (FKM) clustering algorithm, which is one of the fuzzy clustering approaches, by extension to the k-means algorithm, is reported for clustering datasets with categorical values. It is a form of clustering: each point can be associated with more than one cluster. In this paper, anomaly detection is performed on two simulated data by using the FKM cluster algorithm. As a significance of the study, the FKM cluster algorithm allows to determine anomalies with their abnormality degree in contrast to numerous anomaly detection algorithms. According to the results, the FKM cluster algorithm illustrated good performance in the anomaly detection of data, including both one anomaly and more than one anomaly.

Keywords : fuzzy k-mode clustering, anomaly detection, noise, categorical data

Conference Title : ICPRL 2024 : International Conference on Pattern Recognition and Machine Learning

Conference Location : Paris, France

Conference Dates : May 16-17, 2024