

An Enhanced Hidden Semi-Markov model for Outlier Detection in Multivariate Datasets

G Manoharan¹ and K Sivakumar¹

¹Sathyabama Institute of Science and Technology

July 21, 2021

Abstract

Outlier detection in data mining is an important arena where detection models are developed to discover the objects that do not confirm the expected behavior. The generation of huge data in real time applications makes the outlier detection process into more crucial and challenging. Traditional detection techniques based on mean and covariance are not suitable to handle large amount of data and the results are affected by outliers. So it is essential to develop an efficient outlier detection model to detect outliers in the large dataset. The objective of this research work is to develop an efficient outlier detection model for multivariate data employing the enhanced Hidden Semi-Markov Model (HSMM). It is an extension of conventional Hidden Markov Model (HMM) where the proposed model allows arbitrary time distribution in its states to detect outliers. Experimental results demonstrate the better performance of proposed model in terms of detection accuracy, detection rate. Compared to conventional Hidden Markov Model based outlier detection the detection accuracy of proposed model is obtained as 98.62% which is significantly better for large multivariate datasets.

Hosted file

Paper.pdf available at <https://authorea.com/users/426775/articles/531167-an-enhanced-hidden-semi-markov-model-for-outlier-detection-in-multivariate-datasets>