International Journal of Hyperconnectivity and the Internet of Things
Volume 7 • Issue 1

An Efficient Machine Learning-Based Cluster Analysis Mechanism for IoT Data

Sivadi Balakrishna, Vignan's Foundation for Science, Technology, and Research, India*

ABSTRACT

The prevailing developments in internet of things (IoT) and other sensor technologies such as cyber physical systems (CPS) and wireless sensor networks (WSNs), the huge amount of sensor data has been generating from various IoT devices and protocols. Making predictions and finding density patterns over such data is a challenging task. In order to find the density patterns and make analysis over real-time dynamic data, the machine learning (ML) based algorithms are widely used to deal with the IoT data. In this article, the authors proposed an efficient ML-based cluster analysis mechanism for finding density patterns in IoT dynamic data effectively. In this proposed mechanism, the k-means and GMM models are used for clustering data analysis. The proposed mechanism has been implemented on ThingSpeak Cloud platform for analysing the data efficiently on daily and weekly basis. Finally, the proposed mechanism acquired superior results than the existing benchmarked mechanisms over all the performance evaluation metrics used for analysis over IoT dynamic data.

KEYWORDS

data analysis, Internet of Things (IoT), Machine Learning (ML), sensor data, ThingSpeak

1. INTRODUCTION

In recent years, the urban population growth is increased tremendously. As per the recent statistics of the World Health Organization (WHO), From 2015 through 2020, the global urban population will expand by 1.86 percent annually. This growth is expected to be 1.63% between 2020 and 2025 and 1,44% between 2025 and 2030. In urban areas, a considerable proportion of cars are owned by a single home, and at least two cars are owned by a single home. Private cars are becoming increasingly popular with urban traffic. This means that transport in metropolitan areas all over the world is becoming one of the biggest concerns (Shafiq et al. 2020). The large majority of individuals trafficking in urban areas leads to congestion, loss of property, waste of time, damage to the environment, and occasionally to the next level of human mortality. As a result, there is a significant need for smart traffic monitoring and strategies for reduction in cities (Zong et al., 2020). The IoT and ML approaches are the best way to overcome this challenge. It ushers in a new era of intelligent traffic control by effectively aggregating travel times.

DOI: 10.4018/IJHIoT.330680
*Corresponding Author

This paper aims to utilize IoT, cloud computing, Raspberry Pi, and ML approaches to enhance collection and data analysis. This research is based on existing scenarios: information generated by IoT device data gathered from roads and gates is accessible to all passengers and users (Liu et al., 2020). The system will be able to detect existing traffic, traffic and anticipate future traffic to urban areas when it collects real-time sensor data using unsupervised learning techniques. After that, sensor data monitoring and sensor data detection have been measured for analyzing and visualizing the acquired data. Drivers can use the system-generated data for the optimum route selection. As a result, the system is dynamically administrated, controlling, and monitoring moving cars. The boundary time and conditions of different traveler times may be vague significantly (D. Singh et al., 2017).

The unsupervised learning-based clustering techniques are significantly applied in transport research areas for identifying travel patterns (Pyykonen et al., 2013; Yu et al., 2012). The k-means is a crisp clustering algorithm based on partitioning and the Gaussian Mixture Model (GMM) is a fuzzy-based clustering technique, which is widely used for grouping transport patterns during peak and off-peak hours. To measure the number of clusters necessary for optimal transportation data clustering, this research uses the silhouette coefficient and elbow method.

The rest of this paper is organized as follows: Section 2 shows the related work of this study. Section 3 describes the proposed IoT data acquisition sensor and cluster analysis framework, Section 4 illustrates the outcome evaluation and the outcomes of tests and analysis of the proposed work. Section 5 finally concludes this paper and other improvements are highlighted.

2. LITERATURE SURVEY

In this section, the related works included over the recent years discussed and inferred the limitations found over it. (Nallaperuma et al., 2019) have addressed various traffic management system limitations and proposed a Smart Traffic Management Platform (STMP) for making quicker decisions, traffic flow forecasting, concept-drift detection, and optimized traffic control decisions by effective utilization of Artificial Intelligence (AI) techniques over the IoT big data streams. The STMP potentially integrates the various IoT sensor data and is able to predict traffic patterns quickly. However, this platform does not address data acquisition and analysis of dynamic data environments. (Puschmann et al., 2017) have proposed a method for the adaptive adjustment of clusters for IoT dynamic data streams. Which gain useful insights from various data sources and make real-time observations and predictive decisions. Moreover, this method significantly detects drifts by dynamically adjusting clusters over the various data segments. They applied their proposed method in a use-case scenario by considering real-time traffic data through potential utilization clustering algorithms. (Lu et al., 2018) have propounded a parallel approach for clustering large traffic data streams with the utilization of moving object density estimation. In addition, they intend to improve cluster efficiency in data streams, they proposed a modern parallel computation framework with high volume, high-speed traffic stream, and minimum delay with optimal clusters. (Nandurge et al., 2017) intend to overcome the limitations of road traffic accidents data analysis, they used a k-means unsupervised clustering algorithm for segmentation of road accident data. Further, the association rule mining technique has widely applied to discover the predictive traffic data patterns occurred over heterogeneous nature of road accident data.

Similarly, in the studies of traffic analysis over network edge, (Hafeez et al., 2020) have proposed anomaly detection mechanism to predict data analysis at edge level. The IoT-KEEPER enforces malicious network activity detection by utilizing fuzzy c-means clustering algorithm. They evaluated their proposed mechanism by utilizing parameters accuracy and false positive rate with comprehensive dataset. (Dommaraju et al., 2020) have introduced a Deep Learning (DL) based technique for traffic prediction accuracy in big data environment. They are processed and analyzed data through multiple perceptron layers such as input, hidden and output. Further, the activation applied on output layer to predict network traffic based on similarity measurement functions. (Shridevi et al., 2019) have