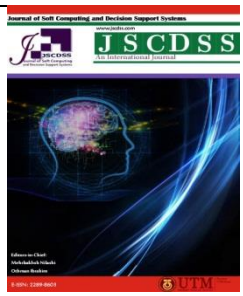
An International Journal
JSCDSS
http://www.jscdss.com
Vol.7 No.3 June 2020: 32-39
Article history:
Accepted 7 May 2020
Published online 7 May 2020

Journal of Soft Computing and Decision Support Systems



E-ISSN: 2289-8603

Distributed Frequent Itemset Mining with Bitwise Method and Using the Gossip-Based Protocol

Hoda Rafieipour [a], Azadeh Abdollah Zadeh [b,*], Mehrdad Mirzaei [c]

[a] Department of Computer Science, Memorial University of Newfoundland, NF, Canada, A1C 5S7
[b] Department of Computer and Electrical Engineering and Computer Science, Florida Atlantic University, FL, USA, 33431
[c] Department of Computer Science, University at Albany, NY, USA, 12222

*** Corresponding author email address**: aabdollahzad2016@fau.edu

Abstract

Nowadays, distributed systems are prevalent and practical in network environments. In distributed systems, pattern recognition help to extract information from network nodes. Meanwhile, data mining in such systems needs resource consideration in terms of storage and computational time. The primary requirement of these systems is a scalable mechanism to distribute the tasks on several databases. Moreover, to do a centralized process, relocating data from all nodes or partial nodes to a central node has confidential risks and traffic overhead. Therefore, distributed data mining in distributed environments needs systematic and structural techniques. In this paper, we propose a new algorithm to extract frequent itemsets in Wireless Sensor Networks. Through this algorithm, nodes frequent local itemsets are obtained with a Bitwise approach, and nodes are classified into clusters by using the Low Energy-Adaptive Clustering Hierarchy (LEACH) algorithm. Connecting the head cluster is performed by a Gossip-based protocol to achieve the values of global support, and it finally resulted in the extraction of frequent itemsets. The proposed algorithm has been simulated in various scenarios using Java software, and algorithm efficiency is evaluated in terms of execution time and average accuracy. Our algorithm is compared with a Gossip-based algorithm, and then some improvements in execution time have been presented.

Keywords: Frequent Itemset mining, distributed data mining, Gossip-based protocol, Bitwise approach

1. Introduction

Recently, different types of sensors are widely used, such as presser sensors, flow sensors, fluid sensors, and airflow sensors. Mohamed et al. (2019) utilized piezoelectric sensors for detecting external loading in structures fabricated by additive manufacturing. One piezoelectric element employed for exciting the part and another piezoelectric sensor is implemented for monitoring the response of the structure to different ultrasonic wave excitation. This method of Structural Health Monitoring (SHM) is called Surface Response to Excitation (SuRE) and has shown promising results for fault detection (Mohamed et al., 2019).

In structural health monitoring, the structure is monitored in predefined periods, and the results will be sent to a remote server for extra processing. Recently, compressive sampling is introduced as an efficient, fast, and linear method of data sampling. The simulation results show that this technique in recovery can highly improve the quality, while sensors can compress the signal in minimal size (Surakanti et al., 2019). In the case of mobile devices, it must be considered that the time for generating compressed samples is a crucial factor. Besides, the fact must be noted that data should be sent to the practitioner as soon as possible. On the other hand, the wearable ECG recorders that have restrained power, and may only be intelligent in doing simple algorithms. Izadi and his colleagues (Izadi et al., 2020) proposed a system architecture that can produce compressed ECG samples, in a linear method by CR 75%. They used sparsity of the ECG signal and suggested an approach based on Compressed Sensing (CS) that can compress ECG samples in real-time. In such cases, finding a DS attack is equivalent to finding the server, which has a frequency higher than a threshold.

The most crucial goal of distributed data mining is to obtain the result similar to the one received by centralized data mining so that it would not be necessary to transfer the original data from their primary storage location. One of the applications of identifying and extracting frequent itemsets in distributed networks is Denial of Service (DoS) attacks detection in the distributed model. In the distributed Denial of Service Attack, target attacks are made from different places in the network. In these attacks, several malicious nodes continuously send a great deal of traffic to the victim node, which is usually a server. Therefore, the