Data Stream Mining in Fog Computing Environment with Feature Selection Using Ensemble ofSwarm Search Algorithms

ABSTRACT

- Fog computing emerged as a contemporary strategy to process big streaming data efficiently. It is designed as a distributed computing platform for supporting the data analytics for Internet of Things (IoT) applications that pushes the data analytics from Cloud server to the far edge of a sensor network.

- As the name suggests, ubiquitous data which is collected from the sensors are processed locally rather than on the central servers.

- Fog computing helps avoid performance bottleneck at the center point and relieves raw data from overwhelming towards the center of the network.

- However, suitable data analysis algorithms suchas those of data stream mining that are consist of learning and recognizing patterns from the incoming data streams must be fast and accurate enough for supporting Fog computing. This paper reports about a computer simulation of running data stream mining algorithms in Fog environment.

- Furthermore, feature selection that is powered by s warm search is used as a pre- processing method for improving the accuracy and speed of local Fog data analytics. Through the experiment, the results reveal which algorithms are the best choice to deliver edge intelligence in Fog computing environment.

EXISTING SYSTEM

- FOG computing is designed as a distributed computing platform for supporting the data analytics for Internet of Things (IoT) applications that pushes the data analytics from Cloud server to the far edge of a sensor network.

- As the name suggests, ubiquitous data which is collected from the sensors are processed locally rather than on the central servers.

- Fog computing helps avoid performance bottleneck at the center point and relieves raw data from overwhelming towards the center of the network.

- However, suitable data analysis algorithms such as those of data stream mining that are consist of learning and recognizing patterns from the incoming data streams must be fast and accurate enough for supporting Fog computing.

DISADVANTAGE

- Fog computing helps avoid performance bottleneck at the center point and relieves raw data from overwhelming towards the center of the network.

- However, suitable data analysis algorithms such as those of data stream mining that are consist of learning and recognizing patterns from the incoming data streams must be fast and accurate enough for supporting Fog computing.

- In Fog computing, the edge nodes are mainly responsible in data preprocessing and analyzing patterns fro m the incoming data streams.

- Speed, efficiency and accuracy are required fro m data mining algorithms for data mining big streaming data which may amount to infinity.

- For supporting edge intelligence in Fog computing, to find a proper data mining algorithm(s) is essential.

PROPOSED SYSTEM

▸ This paper mainly focuses on analyzing about the feasibility of traditional data mining and data stream mining algorithms and compare the m in a Fog computing scenario.

▸ The data mining experimentation is on a classification problem where air/gas samples are collected fro m sensors and the model that is built by the algorithm(s) in the form of decision tree would decide what type of gas it is.

▸ Decision tree is a kind of non-black-box machine learning model, which is a flow-chart-like structure for decision making. The tree branches could be extracted into useful predicate-type of decision rules.

▸ The rules are simple to understand and interpret by both human and machine where they could be coded as logics into embedded devices.

▸ Moreover, combining correlation-based feature selection algorithm, traditional search methods with ensemble of swarm search methods are to be integrated into the data mining algorithm as pre-processing mechanism.

ADVANTAGE

- The simulation of this experiment is concerning: in case of an IoT environment, emergency service takes over in priority.

- It will take into account of real-time constraint and capability requirements. Through experiment, we for search proper data mining algorithms coupled with an appropriate pre-processing mechanism.

HARDWARE REQUIREMENTS

- Processor – Pentium –III
- Speed – 1.1 Ghz
- RAM – 256 MB(min)
- Hard Disk – 20 GB
- Floppy Drive – 1.44 MB
- Key Board – Standard Windows Keyboard
- Mouse – Two or Three Button Mouse
- Monitor – SVGA

MICANS INFOTECH

SOFTWARE REQUIREMENTS

- Operating System   :   Windows 8
- Front End   :   Java /DOTNET
- Database   :   Mysql/HEIDISQL

MICANS INFOTECH

CONCLUSION

- Fog computing provides advantages of bringing analytics to the edge intelligence. The paper shows a simulation experiment comparing two classification algorithms C4.5 and HT respectively.

- C4.5 classification rules have a high accuracy which used to apply in Cloud platform. HT is a popular choice of data stream min ing algorithm which could be well used for Fog computing. The simulation experiment is dedicated to IoT eme rgency services.

- Through collecting a large amount of data fro m gas sensor data to analyze all kinds of gas and then measure air quality. As a consequence of the experiment,

- C4.5 potentially gets high accuracy if the whole data are tra ined. But in the Fog computing environment, the data are streaming in large a mount nonstop into the data stream mining model.

- So, the model must be able to handle incremental learn ing from seeing only a portion of the data stream at a time . And it updates itself quickly each time fresh data is seen. Real-time latency and accuracy are required in IoT environment especially in Fog environment;

- the experiment concludes that FS would have a slightly greater impact on C4.5. However, FS contribute, to ameliorate the performance of HT in Fog environment. Moreover, Harmony search is an effective search method to strengthen the accuracy, time require ment and time cost for HT model in the data stream mining environment. Fog computing using HT coupled with FS- Harmony could have a good accuracy, low latency and reasonable data scalability.

REFERENCE

[1] Rabindia K. Bairk, Harishhandra Dubey, Arun B. Samddar, Rajan D. Gupat a and I.N. Prakash K. Ray, " FogGIS: Fog Comput ing for Geospatial Big Dat a Analytics" IEEE UPCON pp. 613-618, Dec 9-11 2016.

[2] Ivan Kholod, Mikhail Kuprianov, llya Pet ukhov, " Distribut ed Data Mining Based on actors for Int ernet ofThings" MECO 2016 pp. 480-484, 2016

[3] Trans. AMS, 39:472-482 (1936) [17] Kholod, I.: Framework for multi threads execut ion of data mining algorithms. In: Proceeding of 2015 IEEE North West Russia Sect ion Young Researchers in Electrical and Electronic Engineering Conference. (2015 ElConRusW), pp. 74-80.IEEE Xplore (2015)

[4] Sat yanarayanan, Mahadev, et al., Edgeanalytics in t he internet of things, Pervasive Comput ing, IEEE 14.2 pp. 24-31, 2015.

[5] P. Spachos, L. Song, and D. Hat zinakos, "Prototypes of opportunistic wireless sensor net works supporting indoor air qualit y monitoring," IEEEConsumer Communications and Net working Conference (CCNC) 2013, Jan. 2013