Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 198 (2022) 460–465

Procedia Computer Science

www.elsevier.com/locate/procedia

International Workshop on Internet of Smart Things (IST 2021)
November 1-4, 2021, Leuven, Belgium

The research of Wireless Edge Computing Gateway with Anomaly Detection

Yunxia Chu[a,b,c], Siyu Li[a,b,c], Zhangjun[a,b,c]*

[a] *Hebei Province Internet of Things Intelligent Perception and Application Technology Innovation Center, Shijiazhuang, Hebei 050035, China*
[b] *Hebei Province Key Laboratory of IOT Blockchain Integration, Shijiazhuang, Hebei 050035, China*
[c] *Shijiazhuang University,Shijiazhuang, Hebei, 050035, China*

Abstract: To the superiority of low cost and convenience, large-scale Internet of things has been deployed for various monitoring pervasively. Typically cloud platform is utilized as a remote data and control center. However, it is a challenge to the bandwidth and real-time feedback to the server because of mass data with uploading and processing. In this paper, to cope with the above problems, we design a gateway based on edge computing. The gateway not only retains the function of connecting multiple sensors and the interaction between devices and cloud, but also can provide real-time local services for terminal devices .It also can drop abnormal data through sensors' correlation analysis and reduce the amount of data upload by diverting the computing tasks of a server. In this paper, an anomaly detection method based on the bag of words model is introduced. At first, the gateway makes a bag of words representation by the data collected from each sensor, including normal and abnormal data. Then it aggregates these class predictions using the bag of words to calculate the final prediction. Through the simulation analysis based on the data such as humidity, temperature, and lighting monitored by sensors in agricultural greenhouses, the results show that the bag of words model can detect abnormal sensor data.

Keywords: Bag-of-words model；Internet of things；Edge computing；Abnormal data detection；

1. Introduction

Nowadays, sensor-based detection network has been widely used in various environmental monitoring and control, such as greenhouse monitoring, forest monitoring, security monitoring and so on. Plenty of online IoT terminal devices are connected to the system. At the same time, the increasing number of node connections also put forward higher requirements for the overall system of IoT. Moreover, as the network environment used by devices becomes much harsher, the requirements for the IoT network agreement selection and the load requirements for the platform are also higher. For better information services, large-scale data storage and processing bring new

* Corresponding author. Tel.: +0-086-6661-7198 ; fax: +0-086-6661-7198 .
E-mail address: zhang72jun@163.com

10.1016/j.procs.2021.12.270

challenges to the IoT system, such as bandwidth load, real-time feedback and network connection stability[1]. In order to solve the above problems, edge computing technology is integrated into the large-scale IoT monitoring system.

In this paper, first of all, the framework of an IoT monitoring system based on edge computing is proposed [2]. The framework is mainly composed of wireless sensor monitoring nodes, edge computing gateway, MQTT server. Moreover, the function of each component and the communication process among them are described in detail [3]. Secondly, based on this system framework, a data anomaly detection method based on a bag of words is introduced [2]. In the large-scale IoT monitoring system, data anomalies often occur. Data anomaly detection is a typical classification problem in the field of machine learning; this is a new two-step superposition integration method: the first step is that input data is multi-sensor data, and the second one is that classification result is abnormal or normal. Superposition and Integration is a method that takes the prediction of sub-models as input, and then tries to find how to combine the inputs best to make better output prediction. There are many reasons for abnormal data, such as emergency events in the monitoring environment and sensor faulty. It is necessary to detect the abnormal data timely and accurately [3]. Considering the ability and position of the edge computing devices in the system, we propose an anomaly detection method based on edge computing. The spatial correlation of monitoring data is analyzed by the word bag model. The reason for using the word bag is mainly to consider the complexity of the monitored environment [3]. Although the collected data has strong regularity, it is also possible to present a non-linearity correlation [2]. Bag of words, as a kind of neural network, can deal with complex data models. It can model and analyze linear correlation data, so do nonlinear one.

Because of the complexity of neural network model training, the processing of the new technology is deployed in the MQTT server. Moreover, the obtained model and parameters will be sent back to the gateway for subsequent anomaly detection. In addition, the existing problems and potential research directions of the IoT system based on edge computing are briefly discussed at the end of this paper [2].

2. IoT monitoring system based on edge computing framework

The IoT monitoring system based on edge computing comprises a wireless sensor monitoring node, edge computing gateway and cloud computing platform [4].

2.1. Wireless sensor monitoring node

The whole online monitoring system of IoT consists of numerous wireless sensor nodes. In the monitoring area, a large number of wireless sensor nodes are put out randomly or organized to use for collecting real-time environmental data. Each node contains sensors with different parameters to monitor the surrounding environment and collect real-time detection data. In the framework, wireless sensor nodes mainly communicate with the gateway.

The whole detection system uses LoRa wireless transmission. The interface type connecting the sensor is RS485 interface that the communication distance can be up to 1200 meters. Imported probes are used for sensor equipment, which has the advantages of high precision, wide range, and good consistency. It is also easy to install because of a wireless feature. This system adopts the standard Modus RTU protocol. The Wireless sensor monitoring system has super stability, anti-interference ability, strong protection performance and first-class anti-thunder protection.

2.2. Edge computing gateway

In the framework, the edge computing gateway as the middle layer provides local services to the wireless sensor and can improve the quality of service and real-time information feedback [6]. As a result, the delay caused by the remote interaction with the MQTT server will be reduced. It also can provide the initial processing of data and undertake part of the calculation task of the MQTT server in order to reduce the amount of data uploaded to the MQTT server and lighten the bandwidth load of the backbone link [5].

The gateway plays a vital role in the anomaly detection method mentioned above. At ordinary times, the gateway is used as a relay device to upload the monitoring data collected by the wireless sensor gateway to the MQTT server. At the same time, anomaly detection is carried out for the collected data. Once anomaly data is detected, the edge computing gateway will immediately report to the MQTT server and drive the underlying controller to provide emergency response solutions [5]. The function of the edge computing gateway is shown in Figure 1.

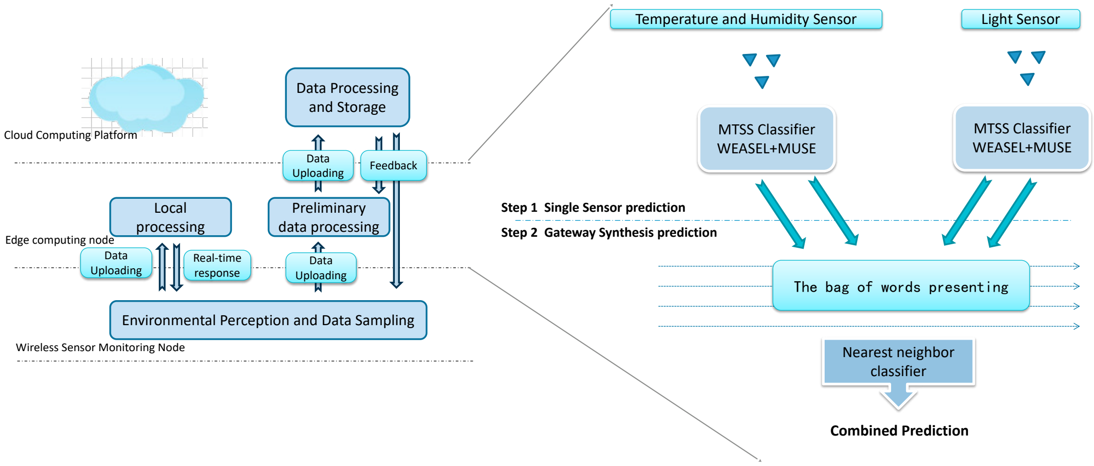

Fig.1.the function of the edge computing gateway

The gateway adopts Raspberry Pi with high-cost performance. The system uses the Ubuntu system. Lora Wan SX1302 is used for gateway data receiving; The transmission module uses 4G or 5G; In the detection algorithm, an outlier algorithm or self- coding neural network is used to detect abnormal data.

2.3. Cloud computing platform

A cloud computing platform is a center of storage data, control center, and process data. Regarding the anomaly detection method mentioned above, a bag of words model is used to capture and utilize the spatial correlation of monitoring data. The model training process is deployed in the MQTT server. The generated model parameters are sent back to the gateway for real-time anomaly detection. On account of progressive and developing network environment supporting devices, the requirements for protocol selection and platform load of the IoT are also higher and higher. MQTT protocol based on TCP (Transmission Control Protocol) is selected as the main communication protocol, and EMQX Message Server is used as the access layer of the IoT platform. Based on these schemes, we can establish a complete set of the IoT systems, which has an extremely practical significance that meets the various needs of small IoT platforms or individual developers.

The system takes the B/S systematic structure of front-end and back-end separation and cross-platform application. The front-end mainly uses the Vue framework to build the users' interface, the Axios obtaining data and rendering the page. The back-end mainly uses the Django web framework and ORM for data operation. Meanwhile, the installation of Rest Framework library is used to build the Rest API ,only providing JSON data to front-end pages. EMQX middleware connects the IoT devices with the platform for the data exchange between the terminal devices and the platform. App for users utilizes cross-platform technology to reduce costs for development and maintenance.

3. Anomaly detection method based on bag of words

Edge Computing Gateway is designed to use distributed network facilities to perform exception judgment for reducing the pressure on cloud servers. As close as possible to the data source, part of the abnormal data is processed directly on the Edge Computing Gateway.

3.1. Multivariate Time Series Classification

Time series is multivariate when existing a sequence of measurements from multiple variables [6]. The data collected by different sensors at different times form multivariable time series (MTSS), and the output results have two types: normal data and abnormal data. Therefore, multi-sensor data transmission can be expressed as an MTSS classification problem.

Regarding the analysis on characteristics of each type of MTSS classifiers and the practical results, we ultimately have chosen the best-in-class category from similarity-based methods, feature-based methods, and the methods based deep learning) shown by Table 1. They were DTWD, DTW I, WEASEL+MUSE, and MLSTM-FCN classifiers, as mentioned in the Fauvel's article [4].

Table 1. The introduction of MTSS classifiers

MTSS classifiers	similarity-based	dependent models(DTWD)	
		independent models(DTWI)	
	feature-based	bag-of-words(BoW)models	WEASEL+MUSE
	deep learning based	LSTM and CNN	MLSTM-FCN

3.2. Targeting a Distributed Cyber Infrastructure

In this section, we introduce the distributed multi-sensor gateway algorithm（DMSDAW）. It is a new two-step superposition integration method. The super position integration method is that the gateway takes the predictions of the sub-model as input and tries to learn how to better combine the inputs to make output predictions. DMSDAW makes predictions (normal, abnormal) based on the data collected by each sensor, then using a bag of words to aggregate these predictions for calculating the final prediction. The rest of this section explains the steps of this algorithm.

Step 1-Predicting the MTSS Category at the Sensor-Level: We have multiple types of sensors ,and we train an MTSS classifier for each type of sensor. The classifier uses a set of data composed of a fixed length of time (which can be set in the gateway according to the time of data collection) for training. We illustrate the first step of our method in the top half of Figure 1.

In order to predict the anomaly category at the individual sensor level, we employ the WEASEL+MUSE MTS classifier. Firstly, WEASEL+MUSE creates a symbolic representation of the MTSS (a Symbolic Fourier Approximation - SFA) on each dimension. Then generates a set of features showed by multiple window lengths, unigrams, bigrams, dimension identification. Finally, it performs the classification based on a heat encoding representation of the MTSS by bag-of-words and feature selection.

Step 2- Detecting abnormal data by combining a series of sensors: We collect the class predictions from the different sensors, and perform a bag-of-word representation. Each class-predicted is seen as a word and the relative frequency vector of each abnormal words is used to classify. This relative frequency vector is obtained by normalizing the number of instances (per abnormal data of MTSS from sensors) [6]. Then, the whole spectrum of potential destructive anomalies is described by combining the text packets of each sensor anomaly. Finally, we train a classifier to perform the combined class prediction.

A distributed gateway composed of geographically distributed data sources is introduced. The function of it is to support the processing of a large amount of data collected by various sensors and meet the real-time requirements of applications. In addition, in disaster situations such as partial network shutdown, this type of architecture allows data generated at the sensor level to be reposition to other processing data centers. All in all, the gateway has two main levels: sensor level and cloud data centers. The sensor level consists of sensor devices with limited computing power. The middle level ,such as cloud data centers consists of well-configured computing systems to meet large-scale computing needs. This kind of architecture is based on edge computing; edge computing is an emerging paradigm in the IoT. Due to the limitation of space, we have left a deep discussion on the gateway in our future work. In this paper, we focus on the distributed algorithms for analyzing the abnormal data generated by sensors and their interactions with the main layers of the infrastructure.

The distributed algorithm needs to be implemented in two steps. The first step of the algorithm is to run the MTSS classifier on each sensor at the sensor level of the infrastructure to generate the sensor level prediction according to the data generated by each sensor. Then, the output of the MTSS classifier from each sensor is transmitted to the cloud data center part of the gateway through the network. Next, the second part of the algorithm is executed at the cloud data center layer. That is to say, firstly, all class predictions from sensors are collected, and the machine learning method is used to form the final class prediction.

3.3. Anomaly detection method based on bag of words

The potential topic model named probabilistic Latent Semantic Analysis （pLSA） applies to v-sized vocabularies. Supposing that in the training data consolidation, we consider the set of descriptors corresponding to all spatial-temporal data detected by sensors to learn the vocabulary of spatial-temporal words. This vocabulary or codebook is constructed by clusters forming by the k-means algorithm and Euclidean distance. The center of each resulting cluster is defined to a spatial-temporal word. Therefore, each sensor can find a set of spatial-temporal words corresponding to it from the codebook. In the codebook, the potential theme Model PLSA relies on M sensors [7,8].

The pLSA is a theoretical method to interpret a long text by acquiring and expressing the associated meaning of words. In the following, we put the pLSA into practice. Supposing we have a set of M (j = 1, ..., M) sensor sequences containing spatial-temporal words from a vocabulary of size V (i = 1,...,V). The corpus of the sensor is summarized in a symbIoTic table $\bar{M}$=V×M. Among it, m (w_i, d_j) stores the number of occurrences of a spatial-temporal word w_i in sensor d_j. In addition, there is a latent topic variable z_k associating with each occurrence of a spatial-temporal word w_i in a sensor d_j. Each topic corresponds to a type of sensor, such as Temperature and Humidity Sensor, Light Sensor.

Our algorithm that has learned the sensor category models, the new task is to categorize new sensor sequences. We have already obtained the sensor-category-specific abnormal-word distributions $P(w_i|z_k)$ from a different set of training sequences. When receiving a new sensor, the unseen sensor is 'projected' on the simplex spanned by the learned $P(w\,|\,z)$. We need to find the mixing coefficients $P(z_k\,|\,d_{test})$ such that the KL divergence between the measured empirical distribution $\tilde{P}(w\,|\,d_{test})$ and $P(w\,|\,d_{test}) = \sum_{k=1}^{K} P(z_k\,|\,d_{test}) P(w\,|\,z_k)$ is minimized. Similar to the learning scenario, we apply an EM algorithm to find the solution. Thus, we make the categorization decision by selecting the action categories explaining the observation best, that is:

$$Action\ \ Category = \arg\ \max_{k} P(z_k, d_{test}) \tag{1}$$

Furthermore, we are also committed to research localizing multiple actions in a single sensor sequence. Though our bag of spatial-temporal words model itself does not explicitly represent the spatial or temporal relationships of the local sensor regions, it is capable enough of localizing different motions within each sensor.

4. Results and Discussions

In this section, we conducted simulation experiments on the data of agricultural greenhouse. We selected an area of 250m*200m to deploy 30 sets of wireless sensor nodes. The node MCU we selected is STM32. The transmission chip is Lora 1278, and the sensors include temperature, humidity, light, and carbon dioxide sensors; Two sets of raspberry gateways are placed, equipped with SX1302 chip, and the WAN port is RJ45 wired transmission. One has a built-in anomaly detection algorithm based on word bag, and the other is equipped with a transparent gateway; Deploy MQTT server is deployed on Windows 10 system. The abnormal learning training based on word bag is the abnormal training of a single node and a single temperature sensor. When the gateway detects abnormal data, it will not send messages to MQTT. All experiments are tested separately in this environment，.

Experiment 1: By artificially suddenly changing the value of a single sensor at a single node (Suddenly irradiating the fixed illumination sensor with a flashlight), we check whether the edge computing gateway can detect abnormal values.

Experiment 2: By artificially changing the values of multiple sensors at a single node (ignite at the edge of the fixed sensor), we check whether the edge computing gateway can detect abnormal values.

Experiment 3: By artificially suddenly changing the value of a single sensor at multiple nodes （suddenly irradiating multiple illumination sensors with a flashlight）, we check whether the edge computing gateway can detect abnormal values.

Experiment 4: By artificially suddenly changing multiple nodes and sensors (ignite multiple nearby sensors), we check whether the edge computing gateway can detect abnormal values.

The experiment was set to the 5-minute interval of collecting and sending data. The experiment lasted 2 hours. The occurrence times of each above experimental method based on artificial control is 50 times. Based on the number of messages sent by the transparent gateway, the message will not be sent after an exception is detected by the gateway with a word packet. The difference between the two is the number of exceptions detected by the gateway. The experimental results are shown in Table 2.

Table 2. The experimental results

	Single-node single-sensor	Single-node Multi-sensor	Multi-node single-sensor	Multi-node multi-sensor
Gateway	120	119	120	120
Gateway with algorithm	72	80	75	84
Recognition rate	96%	79.5%	90%	72%

The above experiments show that the gateway of abnormal data detection based on word packet can effectively detect abnormal data, playing a certain role in reducing the server pressure and reducing the unnecessary traffic off the gateway. Due to the problem of the training set and method, the gateway has good effects on abnormal data detection of single sensor, but the anomaly recognition of multi-node and multi-sensor anomaly needs to be improved, and the relevant sensor anomaly need to be studied.

Acknowledgments

This research is supported by the project of Hebei University Science and technology research project (ZD2021421), Hebei's key R & D plan project(20321801D), and Shijiazhuang high tech Zone Industry University research cooperation project(20420002C).

Reference：

[1] ZHANG Y, MERATNIA N, HAVING A P J M. Outlier detection techniques for wireless sensor networks: a survey[J]. IEEE Communications Surveys and Tutorials, 2010, 12(2):159-170.

[2] Zhiyong Cao, Xiaobo Cai, Haipeng Li, Wei Huang, Xiujuan Yang, Linli Tao, Xi Zhang, Huihui Wang. "A Network Anomaly Monitoring Method Based on Edge Computing for CPS/IOT," 2019 IEEE International Conference on Industrial Internet (ICII), 2019.

[3] Yu, Tianqi. "Data Analytics and Performance Enhancement in Edge-Cloud Collaborative Internet of Things Systems, "Scholarship@Western, 2019.

[4] Kevin Fauvel, Daniel Balouek-Thomert, etc. A Distributed Multi-Sensor Machine Learning Approach to Anomaly EarlyWarning[C]//The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20),2020：403-411.

[5] Musong Gu,Kuan-Ching Li,Zhongwen Li,Qiyi Han,Wenjie Fan, Recognition of Crop Diseases Based on Depthwise Separable Convolution in Edge Computing: Sensors (Basel). 2020 Jul 22; 20(15):4091.

[6] Fauvel, Kevin et al. "Local Cascade Ensemble for Multivariate Data Classification." ArXivabs/2005.03645 (2020).

[7] Juan Carlos Niebles, Hongcheng Wang, Li Fei-Fei. Unsupervised Learning of Human Action Categories Using Spatial-Temporal Words [J]. Int J Comput Vis (2008) 79: 299–318.

[8] Ziqiang Cheng, Yang Yang, Wei Wang, Wenjie Hu, Yueting Zhuang, Guojie Song. Time2Graph: Revisiting Time Series Modeling with Dynamic Shapelets[C]// The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20),2020：3617-3624.