Research on Energy Management System Integration and Energy Saving Optimization

Qiang Zhao[1], and Chenxiao Zhi[2,a,*]

[1]North China Electric Power University,Changping District,Beijing,China

[2]North China Electric Power University,Changping District,Beijing,China

[a]18810692826@163.com

*Chenxiao Zhi

Keywords: Energy management, Wireless sensor networ, Zigbee, Flume, Kafka, Storm, Redis, SVM, Spark.

Abstract. With the rapid development of modern society, an increasing number of people start to realize the importance of the electric power consumption of buildings. Based on this, the article explore the effective management method of energy consumption of buildings: First of all, through the DHT11(temperature and humidity sensors), on-site temperature and humidity data, together with the energy consumption data are collected and sent back to the linux file through the zigbee technology. The real-time data is saved to a linux file through executing Python program. Then applying the Flume+Kafka+Storm+Redis real-time analysis structure: As the producer of Kafka, Flume monitors whether the file has generated new data. As the consumer of Kafka, Storm cleans and organizes the data, monitors and predicts the energy consumption of the buildings. The data is analyzed periodically by using the Support Vector Machine (SVM) algorithm in spark MLlib and establish the forecasting model. At least, the data is stored in Redis and offline analysis is conducted periodically.

1. Introduction

With the rapid development of construction industry and the expansion of construction scale, the growth of energy consumption of buildings becomes increasingly prominent which means it is more important to manage the energy consumption of large buildings effectively in order to improve the utilization efficiency of energy. Generally speaking, the energy consumption of buildings includes water, electricity, coal gas, natural gas and other energy resources. According to the carbon dioxide emissions and consumption situation, the energy consumption of electricity occupies the principal position. Therefore, the management of energy consumption in this paper refers to the management of electric energy in buildings.

2. Data collection, transmission, prediction

2.1 Zigbee Technology

2.1.1 Introduction
Considering that the energy consumption varies greatly under different environmental conditions, therefore, we take two major environmental factors including temperature and humidity into consideration. In this paper, we collect the temperature and humidity of different buildings (by using DHT11 sensor) and electric power consumption data which is transmitted back to the Linux server by applying the wireless communication technology. At present, the wireless sensor networks of wireless communication technologies include ZigBee technology, Bluetooth, WiFi and infrared technology etc. ZigBee[1] is a bi-directional wireless communication and network technology that can be applied to short distance, low complexity, low power consumption, low data rate, low-cost etc. ZigBee[2] is a set of communication network technology which is developed based on the IEEE802.15.4 wireless standard, security and application software. Therefore, ZigBee is suitable for

the field of automatic control and remote control which can be embedded in various devices. Based on the above advantages, this paper applies ZigBee wireless communication technology. Texas Instruments (TI) launched CC2530[3] chip for ZigBee's on chip system (SOC) solution, also CC2530 can be equipped with a network protocol stack (Z-stack) which is provided by TI in order to simplify development.

2.1.2 types of logical devices

There are three types of logical devices in the ZigBee Network including the Coordinator, Router and End-Device. ZigBee network consists of a Coordinator, multiple Router and multiple End-Device. The coordinator is responsible for starting the entire network. The main function of the router is to allow other devices to join the network, communication with multi-hop routing and assistance End-Device. End-Device is mainly responsible for collecting data through sensors. ZigBee network support three network topologies: star, tree, and mesh.

2.2 Real-time data writing to Linux

The data collected by the End-Device is transmitted to the Coordinator through the router, and the coordinator transfers the data to the Linux by serial port. Then by using the serial port communication tool Minicom in Linux, the real-time received data can be displayed. We also write the python script to record the received data and the current real-time to a file.

Part of the Python code is shown as follows:

```
myUartPort = serial.tools.list_ports.comports()[1][0]
myUartSer = serial.Serial(myUartPort,115200)
if myUartSer.inWaiting():
  UartStr = myUartSer.readline()
  Tolalstr='%s  %s' %(datetime.datetime.now().strftime('%H-%M-%S.%f'),UartStr)
  try:
    f = open("/usr/local/data/flume_sources/data-1", 'a+')
    f.flush()
    f.write(Tolalstr)
  finally:
    f.close()
```

2.3 The structure of real-time analytic system: Flume, Kafka, Storm, Redis

The structure of the real-time analysis system is the use of Flume[4] to monitor /usr/local/data/flume_sources/data-1 if new data is generated, and every log information is collected in real time and saved in the Kafka message system which is then consumed by Storm system. Meanwhile, the consumption record is based on the Zookeeper cluster management which means last time consumption record can be find even if Kafka is down after the restart and continue to consumption from Kafka Broker[5]. Because of the non-atomic operations including consumption before record or record before consumption, few data loss or repeat consumption problems will occur when Kafka is down or similar problems happened at the time message is not recorded to Zookeeper after consumption. The solution of the problem is to arrange the Kafka Broker and Zookeeper on the same machine and apply the defined Storm[6] Topology to analyze the log information and export to Redis cache database (which can also be persisted). The reason to build up Kafka message system between Flume and Storm is that the data may growth spurt under high concurrent conditions. If the speed of Storm consumption[7] (The real time computing ability of Storm is one of the fastest, but with some exceptions. And the open source real-time computing framework Heron Twitter is faster than Storm) is lower than the speed log generates and Flume's own limitations, this will inevitably lead to the lag and loss of a large amount of data. Therefore, Kafka message system is included as a data buffer. In addition, Kafka is based on the log file message system which means message will be persistent on the hard disk and Kafka make effective use of the characteristics of Linux I/O which

provides considerable handling throughput. Redis[8] is also applied in the structure as database because it also improves the read-write speed in real time environment.

2.4 Support Vector Machine

2.4.1 Introduction

The characteristics of time series data are noise, instability, random and nonlinear, while the neural network plays an important role in nonlinear modeling which can complete complex prediction without mathematical model. The limitation of neural network algorithm is that it applies the empirical risk minimization principle which means it is easy to fall into the local minimum point and the convergence speed is slow etc. that greatly limit the practical application of the method. Support vector machine (SVM)[9] applies the structural risk minimization principle which means the solving process has been transformed to a two-convex quadratic programming problem and the solution is global optimal and unique. As SVM overcomes some shortcomings of neural network, it has achieved good performance in the practical application.

2.4.2 Feature

Support vector machines (SVM) have three main characteristics as follows[10]:

(1) Based on the principle of structural risk minimization, the upper bound of the actual risk is given to ensure that the learning machine has good generalization ability.

(2) The algorithm is finally transformed into a convex optimization problem with linear constraints, which guarantees the global optimality and the uniqueness of the algorithm.

(3) The linear non-separable problem of input space is transformed into linear separable problem in feature space by applying nuclear technology.

2.4.3 SVM vs Neural network

Compared with the SVM, the neural network[11] has the following deficiencies:

(1) the network structure needs to be specified in advance or applies the heuristic correction algorithm in the training process. The heuristic algorithm is difficult to guarantee the optimization of network structure. For multilayer networks, it is a complex combinatorial problem.

(2) The adjustment method of network weight coefficients is limited. Specifically, the training may end prematurely and results in the weight loss.

(3) The neural network is easy to fall into local optimum, and some training algorithms can not even get the optimal solution.

(4) Over reliance on learning samples, that is, the performance of the model depends too much on the quantity and quality of the sample data in the training process. In most cases, sample data is limited. In addition, the input space in many practical problems is high-dimensional, and the sample data is only the sparse distribution in the input space. Even if high quality training data can be obtained, the amount of data is bound to be large. A large amount of sample data will greatly increase the training time of the algorithm.

(5) At present, there is no theoretical support to quantitatively analyze the convergence speed of the training process of neural networks, determine the conditions of the convergence speed and control them.

(6) The optimization goal of neural network method is to minimize the risk based on experience which can only guarantee the minimum of the estimation error of sample points. In practice, the error achieves the minimum for all possible points, i.e. the best generalization performance. The neural network method avoids some important problems[12], such as whether the empirical risk can converge to the actual risk and the convergence conditions etc., while SVM has strict theoretical and mathematical basis, and can effectively overcome the shortcomings of the neural network.

2.5 The Stages of Data Analysis

Flume is used for real-time monitoring of /usr/local/data/flume_sources/data-1, to determine whether a new type of data generates. The data in the files includes four fields: current time, temperature, humidity and electric power consumption. Flume source is set to spooldir. The sink type is set to

org.apache.flume.sink.kafka.KafkaSink.As the producer of Kafka, sink the data to Kafka for consumption. As the consumer of Kafka, consume the related topic data in the Storm. Storm obtains the corresponding topic data through kafkaSpout and passes it to the Bolt method to complete related processes and the data is written into Redis for offline data analysis eventually. Regularly (such as monthly) applying the SVMWithSGD algorithm in spark MLlib on the historical data of the training model, load the model in the Bolt method of the Storm and predicts the data of the next moment. If the actual value is larger than the predicted value, an alarm is raised, and relevant information will be written to the log file.

References

[1] Fukunaga, S. (2007). ZigBee—Wireless PAN Technology—. IEICE Communications Society Magazine, 2007(2), pp.2_62-2_73.

[2] Md Riyasat, A. and Vijayshree A More, V. (2011). Zigbee in Wireless Networking. Indian Journal of Applied Research, 3(3), pp.127-129.

[3] Kim, B. (2012). Implementation of IEEE 802.15.4 Protocol Stack on the Texas Instrument CC2530. Journal of Digital Contents Society, 13(3), pp.353-358.

[4] Kang, Y. (2014). Design of a Framework of a System for Handling Streaming Data by Using Apache Flume. The Journal of Korean Institute of Information Technology, 12(11).

[5] Wang, G., Koshy, J., Subramanian, S., Paramasivam, K., Zadeh, M., Narkhede, N., Rao, J., Kreps, J. and Stein, J. (2015). Building a replicated logging system with Apache Kafka. Proceedings of the VLDB Endowment, 8(12), pp.1654-1655.

[6] Zhou, L., Chen, N. and Chen, Z. (2017). Efficient Streaming Mass Spatio-Temporal Vehicle Data Access in Urban Sensor Networks Based on Apache Storm. Sensors, 17(4), p.815.

[7] Hussain Iqbal, M. and Rahim Soomro, T. (2015). Big Data Analysis: Apache Storm Perspective. International Journal of Computer Trends and Technology, 19(1), pp.9-14.

[8] Choi, J., Jeong, D., Yoon, J. and Lee, S. (2016). Digital Forensics Investigation of Redis Database. KIPS Transactions on Computer and Communication Systems, 5(5), pp.117-126.

[9] Fearn, T. (2004). Support vector machines II: The support vector classifier with overlapping groups. NIR news, 15(1), p.8.

[10] L.J.Cao,Q.M.Gu. Dynamic support vector machines for non-stationary time series forecasting. Intelligent Data Analysis,2002,6(1):1-16.

[11] Yayik, A. and Kutlu, Y. (2014). Neural Network Based Cryptography. Neural Network World, pp.177-192.

[12] Yalkin, C. and Korkmaz, E. (2012). Neural network world: A neural network based selection method for genetic algorithms. Neural Network World, 22(6), pp.495-510.