Realization of Smart Home Information System with High-accuracy Data Import Based on Big Data

Gang Li1*

1 Xi’an Fanyi University, Xi’an 710105, Shaanxi, China

*Corresponding author e-mail: 284584835@qq.com

Abstract. To solve the problem of low data import accuracy of intelligent Internet of Things (IoT) information system in the existing smart home application scenarios, this article develops a scheme of smart home information system based on big data with high-accuracy data import. The scheme is based on the system architecture, combined with the related hardware system of smart home, which can successfully complete the accurate import and stable data preprocessing of a large number of data Manage and classify data. By comparing the existing smart home data import performance, big data simulation and comparative experiments are carried out. The experimental results show that the smart home information system can ensure the accuracy and reliability of information import when importing a large number of data. The validity of the scheme is verified. The smart home information system realized by the scheme can operate accurately and reliably in the big data environment, which also shows that the scheme can be extended to the whole society.

Keywords: High-Accuracy, Big Data, Smart Home, Information System

1 Introduction

With the development of IoT industry, perceptual devices and terminals in various fields gather people together through the network. After the IoT transforms a large number of data information into electrical signals and transmits them to the upper application system through the network, this information transmission method has become a development trend, which also leads to the rapid growth of massive unstructured data. The growth of big data is nonlinear[1]. With the increase of the number of sensors, the data shows an irresistible linear growth trend. In the era of big data, the value of IoT big data has been maximized. No matter in Intelligent City, intelligent community, intelligent family, intelligent education, scientific research service and other fields, big data can be collected, mined, useful information can be formed, and analyzed with intelligent hardware.

With the rapid development of broadband technology and smart appliances, the smart home market has begun to enter a period of rapid development[2]. At present, smart home systems at home and abroad mainly involve consumer electronics, it industry, construction and communication industries. Intelligent interaction, health, data security and rapid response have gradually become a huge driving force for the overall technology update of smart home information system. Smart home applications include many, such as automated transportation, intelligent closed circuit television (CCTV), energy
management system applications, network architecture, mobile applications, security applications and environmental monitoring[3]. The application of smart home information system should also have the ability to detect and identify the "health status" of residents, and collect information to support residents' health, so as to help the owners cope in advance[4]. The real challenge of smart home information system is big data processing, which needs the help of big data platform[5]. With the application of big data, it also poses an unprecedented challenge to protect the internal privacy of smart home[6]. Therefore, the scheme of smart home information system based on big data can effectively improve the effectiveness of information processing of smart home equipment under big data. There are many existing smart home information systems at home and abroad, which can basically meet the requirements of non emergency, reliability of data collection and processing and low data response speed. The accuracy of data import is low, the response time is long, and the accuracy of data processing is uncertain.

In view of the problems of low precision and complex implementation of data and information import in the current international and domestic smart home information system, a simple, universal and operable solution is proposed. Compared with the existing conventional system, the smart home information system implemented by this scheme has the characteristics of high-accuracy data import, fast response speed, accurate processing and analysis results.

2 Smart Home Information System Architecture

According to the characteristics of smart home information processing in the context of big data, its architecture is shown in Fig 1.

The Smart home information system needs to be able to meet the requirements of continuous collection and processing of a large number of detailed data, and remote access to smart devices, so that families can be more controllable [7]. The system architecture mainly includes data import, data preprocessing, and data classification. Among them, data import mainly supports offline and online data acquisition methods, as well as host address filtering settings and protocol feature filtering settings; data preprocessing mainly includes protocol feature filtering, combination filtering settings, original data and destination address filtering settings, and provides Friendly UI display, flexible IDC control, and task parameter information extraction.

![Fig 1. Smart home information system architecture](image_url)

3 Smart Home Information System Hardware

The Smart home system hardware integrates digital sensing and communication equipment [8]. To improve the overall system performance and the effectiveness of the overall data processing, the system hardware selection is aimed at the portability of the system on different platforms.

A computer capable of importing a database is selected as the host, and the host parameter is 802.11b / g / n / ac Wireless LAN Mini-PCI Express Adapter II network adapter, 500G hard drive, Intel (R) Core (TM) i3 processor, and ATIMobility Radeon HD 5154 graphics card. The big data integration platform applied in the system needs to have the characteristics of low cost, short delay, low power consumption, low speed, high security, high capacity, and high reliability. Therefore, the system hardware module chooses ZIGBEE to download and debug the template, ZIGBEE (CC2431) module and BLK-MD-BC04-B serial module, and realizes big data collection through SHTXO sensor. In the ZIGBEE module, the 8051 microcontroller is used, which can ensure the overall performance of the system under low power consumption conditions, and the current consumption in the system standby and hibernation modes is only 0.9 μA and 0.6 μA [9]. At the same time, it has a 2.4 GHz
wireless transceiver and supports multiple protocols USART. The overall resolution of the hardware is 0.1 or 1. The power supply voltage ranges from 2.0 to 3.6 V. Its power consumption is not greater than 20 W. It supports master-slave mode and software / hardware setting master-slave mode. Based on the system hardware, the hardware interface is designed, and the ZIGBEE module is designed to facilitate debugging. Two sets of LED indicators are designed to connect with the two IO pins of the ZIGBEE module P1_0 and P1_1, respectively, to ensure the smooth implementation of each function.

4 Smart Home Information System Software

4.1 System Database Construction
As the core part of the whole system, the database mainly meets the needs as the main purpose, and can improve the data query efficiency based on the reasonable expression of the relationship between the data structure design and the data. In the process of database construction, we must fully consider the security, integrity, consistency, recoverability, and scalability of the database. Therefore, under the operating conditions of the hardware PC, based on the Myeclipse 10 project, MySQL is used to establish the database.

After analyzing the process of database requirements, considering the difference between data collection speed and processing speed, different data application and import technologies are used. Data import has the characteristics of real-time, but because the system will not write the data that is not processed in time, it will need to solve the database memory problem. The system database uses Kafka, which enables the system to better support online applications [10].

After the data collection is completed, the distributed data is exchanged on the distributed platform by the pre-processing system, while the offline data processing is (HDFS) o Flume NG for distributed file processing. (HDFS) o Flume NG is reliable and distributed. It can efficiently collect big data from different data sources, and finally store it in a unified and centralized database.

4.2 System Data Import
To facilitate the system's processing of big data, the system's big data acquisition can obtain data as an observer without affecting the normal operation of the data and damaging the network topology, and supports both online and offline data acquisition [11]. The system data import implementation is shown in Table 1.

| Mode  | Realization Way                                | Application Scenario                                                                 |
|-------|------------------------------------------------|---------------------------------------------------------------------------------------|
| Off-line | Save data as binary files through packet capture tools such as Wneshark, Tcpdump | Verify the system flow after analyzing the task data in the later stage of system integration joint test and integration experiment |
| On-line | Get in real time with Wneshark                    | Use the system integration joint test phase and the integration experiment phase to verify whether the information exchanges data according to the task format and timing |

The online data import of the system mainly focuses on the real-time analysis of task data to quickly obtain the parameters during the operation of the system. At the same time, the integration test of real-time working status confirmation and rapid positioning is carried out to determine the ability of rapid positioning integration test. The offline data import of the system is mainly to obtain the task data generated by the system in a certain working time. After the task is completed, the data is analyzed to determine whether the system workflow can meet the task requirements, and the data accident status is diagnosed according to the dynamic response.
4.3 Data Preprocessing

Data preprocessing mainly filters the data imported in the database to extract the data required for the task. The data rarely presents itself in a ordered, clean way[12]. Various objectives may exist in data processing, which needs to provide sufficient operational space. So data preprocessing is required. In addition to the host IP space and public protocol filtering functions, the system needs to provide multiple filtering methods based on internal protocol attributes, as shown in Table 2.

Table 2. Internal protocol data filtering

| Realization Way | Description |
|-----------------|-------------|
| ProtoField string ( abbr, name, base, ) | Protocol feature field with "channel" as keyword |
| ProtoField uint32 ( abbr,name, base, [valuestring] ) | Protocol feature field with "txTime" as keyword |
| rotoField uint16 ( abbr, name, base, [valuestring] ) | Protocol feature field with "code" and "length" as keywords |
| ProtoField uint8 ( abbr, name, base, [valuestring] ) | Protocol feature field with "priority" as keyword |
| Proto new(name, desc) | Protocol feature field with "myProto" as keyword |

4.4 Data Classification

Binary data stream processing is performed on the processed data to realize the classification of the overall data. After classifying the system ICD, the data is entered into the parser according to the interface control definition format, and the task data is parsed according to the agreed format, as shown in Table 3. According to Table 3, the data is added to the Wireshar protocol tree at the level defined by the ICD to achieve overall data classification.

Table 3. Data classification and analysis

| Realization Way | Description |
|-----------------|-------------|
| Distantable try | Dynamically matching formulation data in the protocol list |
| Treeitem append_text | Add parameter description on the protocol tree |
| Treeitem set_text | Added parameter description on the protocol tree |

5 Simulation Experiment

To test the effectiveness of the smart home information system based on big data, a comparative experiment is designed in this article.

5.1 Experimental Process

To ensure that the experimental data and calculation results are true and valid, all experimental data are from professional statistical software, maintaining the same experimental environment and status, including network bandwidth, data capacity, prerequisites for processing data, and storage media. Define and input different data parameters to ensure the smoothness and accuracy of the experiment. Compared with common information systems, record and analyze the accuracy of information import.

5.2 Experimental Results

The accuracy of information import can measure the big data processing ability of smart home information system. The higher the accuracy of information import, the stronger the data processing ability. The experimental results are shown in Fig 2.
According to Fig 2, the information import accuracy of smart home information system based on big data has been stable above 90 with the increase of big data information. However, with the increase of big data information in the common system, the accuracy of information import decreases significantly, and the trend is linear, lacking of processing capacity. The system developed in this article has the advantages of strong data processing ability and high use value.

6 Conclusion
In the process of realizing of smart home information system with high-accuracy data import based on big data, according to the characteristics of importing big data information, preprocessing the big data collected by intelligent hardware system, and realizing large-scale and high-accuracy import of various data in smart home through intelligent identification processing, classification of data and other processing. Finally, the goal of improving the accuracy of information import and data processing efficiency of the system is achieved.

Acknowledgements
This work was supported by the Scientific Research Project of Shaanxi Provincial Department of Education, grant number: 18JK1004.

References
[1] Melis Onel, Chris A. Kieslich, Yannis A. Guzman, Christodoulos A. Floudas, Efstratios N. Pistikopoulos: Big Data Approach to Batch Process Monitoring: Simultaneous Fault Detection and Diagnosis Using Nonlinear Support Vector Machine-based Feature Selection. Computers & Chemical Engineering. 115 (7), 46-63 (2018).
[2] Zhang H: Construction of Smart Home Standard System. TV Technology. 2014 (4), 86-88 (2014). (in Chinese).
[3] Mussab Alaa, A.A. Zaidan, B.B. Zaidan, Mohammed Talal, M.L.M. Kiah: A Review of Smart Home Applications Based on Internet of Things. Journal of Network and Computer Applications. 97 (21), 48-65 (2017).
[4] Sarah J. Darby: Smart Technology in the Home: Time for More Clarity. Building Research & Information. 46 (1), 140-147 (2018).
[5] Biljana L. Risteska Stojkoska, Kire V.Trivodaliev: A Review of Internet of Things for Smart Home: Challenges and Solutions. Journal of Cleaner Production. 140 (3), 1454-1464 (2017).
[6] Noah Apthorpe, Danny Yuxing Huang, Dillon Reisman, Arvind Narayanan, Nick Feamster: Keeping the Smart Home Private with Smart(er) IoT Traffic Shaping. 2019 (3), 128-148 (2019).
[7] L Nicholls, Y Strengers, J Sadowski: Social Impacts and Control in the Smart Home. Nature Energy. 5, 80-182 (2020).
[8] Kirsten Gram-Hanssen, Sarah J. Darby: “Home Is Where the Smart Is”? Evaluating Smart Home Research and approaches against the concept of home. Energy Research & Social Science. 37 (3), 94-101 (2018).
[9] Song L X, Zhou Z M: Design and Implementation of Internet of Things Information service System of Tea Industry Chain Under the Background of Big Data. Fujian Tea. 40 (11), 25-27 (2018). (in Chinese).

[10] Zhang L: Framework Design of Ship Traffic Management Information System Based on Intelligent Internet of Things Platform. Ship Science and Technology. 19 (2), 58-60 (2018). (in Chinese).

[11] Dong Z Y: Vineyard Information Acquisition and Intelligent Irrigation System Design Based on the Internet of Things. Agricultural Mechanization Research. 40 (4), 212-215 (2018). (in Chinese).

[12] Ignacio Cordón, Julián Luengo, Salvador García, Francisco Herrera, Francisco Charte: Smartdata: Data preprocessing to achieve smart data in R. Neurocomputing. 360 (37), 1-13 (2019).