A Fault Data Capture Method for Water Quality Monitoring Equipment Based on Structural Pattern Recognition

Hao Yang, Daoliang Li, Yong Liang

A Fault Data Capture Method for Water Quality Monitoring Equipment Based On Structural Pattern Recognition

Hao Yang[1,2,3,4,5], Daoliang Li[1,2,3,4,6,*], and Yong Liang[5]

[1]Key Laboratory of Agricultural Information Acquisition Technology, Ministry of Agriculture, Beijing 100083, P.R. China
[2]Beijing Engineering and Technology Research Center for Internet of Things in Agriculture, Beijing 100083, P.R. China
[3]China-EU Center for Information and Communication Technologies in Agriculture, China Agricultural University, Beijing 100083, P.R. China
[4]Beijing Engineering Center for Advanced Sensors in Agriculture, Beijing 100083, P.R. China
[6]College of Information and Electrical Engineering, China Agricultural University, Beijing, 100083, P.R China
[5]Institute of Informatics Science and Engineering, Shandong Agricultural University, Tai'an, 271018, P.R China

Abstract. To capture equipment fault in real time and automate fault diagnosis, a pattern recognition method, based on data eigenvector and TCP transport protocol, was proposed to capture Water Quality Monitoring equipment's fault information. Fault data eigenvector was designed after analyzing the equipment fault feature and capture strategy, structural pattern recognition strategy was confirmed and specific data frame was designed in response to the fault data eigenvector, by integrating the data frame design into the equipment's communication protocol, data related to different fault compiled into fault data frames by transmitters or communication module of equipment's different components, the remote sever captures equipment fault on transport via fault data frames according to the structural pattern recognition strategy. With 7 months of practical application in Taihu aquaculture project and research center of agricultural information technology, combining with historical fault data and contrast with artificial recognition result, the simulate experiment shows this method has higher response rate and process rate with a nice accurate.

Keywords: fault data capture, data eigenvector, transport layer, fault filter.

1. Introduction

Water quality parameters includes DO (dissolved oxygen), EC (electrical conductivity), SAL(solidity), WT (water temperature), PH and TUR (turbidity), all of these parameters effect deeply to aquaculture management and breeding decision, so collecting water quality parameters is an extremely important work to aquaculture. The Water quality monitoring equipment in the paper is made to collect water quality parameters in time and offer water quality optimization services include SMS warning,

* Corresponding author: Daoliang Li. Email: dliangl@cau.edu.cn

oxygenation and other services. Equipment have been deployed in maricultural breeding bases and industrial aquaculture in Beijing, Tianjin, Hebei, Shandong, Jiangsu and Guangdong. Nevertheless, equipment fault prevents farmers from increasing more reliance on the equipment, researches must be done to develop fault diagnosis technology and form an integrated strategy.

ANNs requires large amount of sample data for training, as its process is pure mathematical procedure [1-13].There are many fault detection methods applied the support vector machine [14].

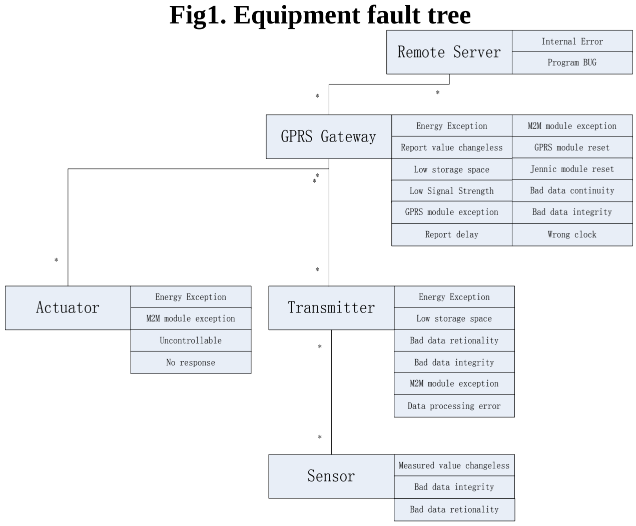
Fault detection should be an independent step in fault diagnosis process, fault diagnosis should base on fault data captured by fault detection step, and serious redundancy calculations exist in diagnosis process.

2. Data Eigenvector analysis

2.1 Equipment Faults Description

Equipment faults are divided into five types due to component types. Namely sensor faults, transmitter faults, actuator faults, GPRS gateway faults and remote server faults, see fig1.

Fig1. Equipment fault tree



When combined with some high-class arithmetic like support vector machine to solve fault capture task in every node, long training time cost is not acceptable.

2.2 Data Eigenvector Related to Different Equipment Faults

To an object represents a specific fault, recognition arithmetic need to judge whether it contains fault information according data eigenvector, for each fault type, there is a specific eigenvector and type record; these records are listed in Form 1.

Form1. Data type and eigenvector corresponding to different faults

Recognition target	Data type	Eigenvectors
Uncontrollable	Actuator state	Response data and executive result
No response	Actuator state	Response data
Energy Exception	M2M state	Power supply mode & nominal voltage & measured value
Low storage space	Device information	Residual capacity calculated value
Jennic module exception	M2M state	Jennic module reset times
M2M module exception	M2M state	M2M signal value and emission frequency
Data process error	M2M state	Conversion formula
Wrong clock	GPRS state	GPRS gateway clock
GPRS module exception	GPRS state	State code & SIM card state & heart beat interval
GPRS module exception	GPRS state	Reset times
Low Signal strength	GPRS state/Device state	Signal emission frequency
Internal error	Message from server	Error code
Program BUG	None	Server crash
Bad data integrity	Record	Actual record count divide theoretical record count & sensor state
Bad data continuity	Record	Actual record interval time
Report delay	Record	Report time minus collected time
Report value changeless	Record	value & sensor collect voltage value
Bad data rationality	Record	Value change rate & value

3 Method and Capture Strategy

3.1 Data Transfer Process on Transport Layer

Data frame format was proposed in the following passage. Form.2 is the transmitting data frame format.

Form2. Transmitting Data Frame Format

Length	2 bytes	1 byte s	1 bytes	1 bytes	2 bytes	M bytes	N bytes	2 bytes	2 bytes
Content	Start segment	Frame type	Frame length	Frame code	Control code	Address segment	Data segment	CRC check	Terminal
Description	Head						Eigenvector data	Tail	

		frame

The eigenvector data frames and fault filter introduced in following passages are based on the data frame format and this protocol.

3.2 Define Eigenvector Data Frames and Frame Filter

The design of eigenvector data frame is shown in form 3 below. Eigenvector type is corresponding to the frame code; form 4 is the detail design of eigenvector data frame model.

Form3. Eigenvector data frame format

	Type	GPRS gateway clock/Report time	Eigenvector type	Eigenvector data
Length	2 bytes	6 bytes	1 bytes	N

Form4. Detail Design of Eigenvector Data Frame Model

Eigenvector source	**Frame code**	**Type code**	**Eigenvetor type**	**Description**
GPRS gateway state	0x01	Link heartbeat:[0x01][0x00];	Device heart beat interval time	Hexadecimal value
			Reporting cycle	Hexadecimal value
		State heartbeat:	Acquisition cycle	Hexadecimal value
		long connection: [0x02][0x01]	Storage cycle	Hexadecimal value
		Short connection:[0x02][0x02]	Signal strength	Hexadecimal value
		no connection: [0x02][0x03]	Not reported data count	Hexadecimal value
			Gateway voltage	Hexadecimal value
M2M state	0x02	Power on: [0x01][0x00]	Energy type	0x00:battery
		Pause: [0x02][0x00] Power off:[0x03][0x00]		0x01:electric supply
				0x02:solar pannel
			Device reset time	Hexadecimal value
			Not reported data count	Hexadecimal value
			Deivce voltage	Hexadecimal value
Record	0x03	Channel	Channel	Hexadecimal value

		count:Generated by sending device, equals the amount of channels belong to the device	number	
			channel type	DO:0x00;
				EC:0x01;
				WT:0x02;
				PH:0x03;
			channel value	Hexadecimal value
Actuator state	0x04	Power on: [0x01][0x00]	Channels corresponding to control action	Hexadecimal value
		Pause: [0x02][0x00]	Control trigger	0x00:unknown
		Power off: [0x03][0x00]		0x01:timing
		No action: [0x04][0x00]		0x02:automatic
				0x03:remote control
				0x04:manual operation
				0x05 exception protect
				0x06:Illeagal control
				0x07:SMS control
Actuator state	0x04		Control type	0x00:power on
				0x01:pause
				0x02:power off
				0x03:no action
			State feedback type	0x00:poer on
				0x01:pause
				0x02:power off
				0x03:running
			State feedback channels	Hexadecimal value
			Original channel value	Hexadecimal value

A frame filter is designed here to filter eigenvector information from data frames with a high rate. The filter design is showed in Form 5.

Form5. Data filter design

Process	Byte number	Mask code	Description
1	1	0xFF	Extract the first byte, whose low three bits maybe the high three bits of the frame length.
2	4	0xFF	Extract the length of this frame
3	5	0xFF	Extract the frame code.
4	6	0xF0	Address type
4	6	0x0F	Address type

4	7	0xF0	Address type
4	7	0x0F	Address type
5	8	0xF0	Address description
6	9~9+M	0xF…F	Device address
7	9+M~11+M	0xFFFF	Data type
8	11+M~17+M	0xF…F	Report time/ GPRS Gateway clock
9	17+M~18+M	0xFF	Eigenvector type
10	18+M~18+M+N	0xF…F	Eigenvector value

3.3 A method to Capture Fault by Filtering Data Frame

Form 6 is the recognize rule.To explain the process more clearly and detail, the filtering process is shown in the following fig2. Fig3 introduces recognition strategy respectively for GPRS gateway fault, because of space limit, device fault, record fault and actuator fault recognition strategy is not introduced here.

Form6. Recognize rule

Data type	**Recognize parameter**	**Recognize result**
GPRS Gateway state	No GPRS connection	GPRS offline
GPRS Gateway state	GPRS signal strength<16	Communication error: low signal strength
GPRS Gateway state	SIM card service stopped or Reset times in 24 hours>1 or Heart beat interval time>10minutes	GPRS module exception
GPRS Gateway state	Reporting cycle> heart beat interval time Device voltage<5 GPRS clock time difference>5 minutes	Bad data integrity Low device voltage Wrong GPRS clock
Device state	Device power off	Device shut down
Device state	Device voltage as the high byte and energy type as the low byte	<0x0502, solar pannel exception
Device state	Device voltage as the high byte and energy type as the low byte	<0x0501,electric supply exception
Device state	Device voltage as the high byte and energy type as the low byte	<0x0500,battery exception
Device state	Device reset time>1 in 24 hours	M2M module exception
Device state	Signal strength<50	Communication error:Low signal strength
Device state	Report time difference>5minutes	Report delay
Device state	Not reported data count>200 in 24 hours	Bad data integrity
Record	Report data value changeless	M2M communication exception

Record	Report data value out the threshold	Bad data rationality
Record	Unnormal report data value change rate	Bad data rationality
Record	Report time difference>5minutes	Report delay
Record	(Data count/heart beat count)/(Heart beat interval time/data collection cycle)	>1,GPRS connection fault
Record	(Data count/heart beat count)/(Heart beat interval time/data collection cycle)	<1,Bad data integrity
Actuator state	Actuator state code equals [0x02 0x00] or [0x03 0x00]	Uncontrollable
Actuator state	Actuator state code equals [0x04 0x00]	No response
Actuator state	Report delay	Report time difference>5minutes

Fig2. Filtering process

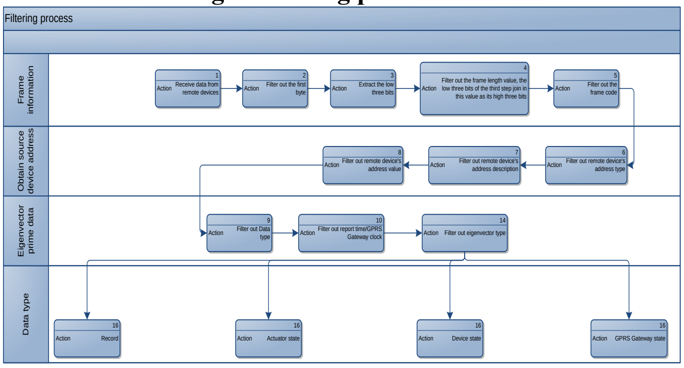
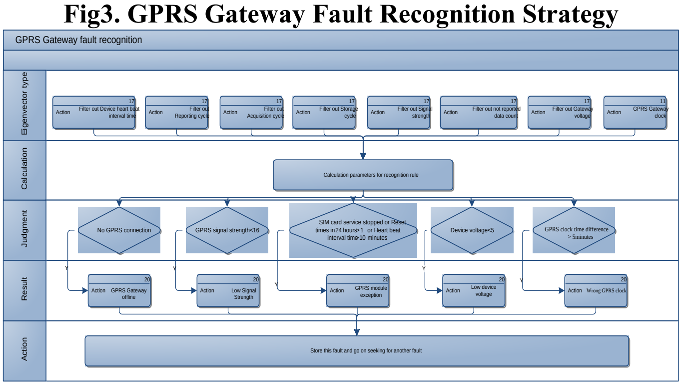


Fig3. GPRS Gateway Fault Recognition Strategy

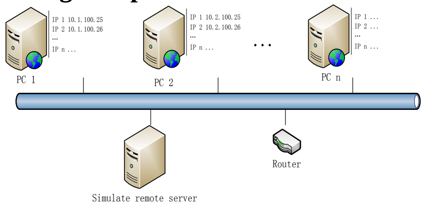


4 Simulation Experiment and Discussion

4.1 Simulation strategy and build simulation LAN

To verify the strategy's processing rate and accuracy, different personal computers with simulating IP addresses are regarded as the GPRS gateways to report collected data to remote server. The experiment LAN structure is shown in fig4.

Fig4. Experiment LAN Structure



The router simulate the GPRS service, simulate remote server is another PC, simulate remote server receive the data frame via TCP/IP protocol and telnet service, then decode and filter the data frames with programs based on the fault recognition strategy in chapter 3.3. The program are written in java and based on apache mina framework.

4.2 Experiment result and Evaluation

1290 records of fault are found from working log of operation and maintenance team is used in the experiment. The program was divided into two executable jar file, namely Simulation Server and Simulation Gateway. Fig 5 to 6 shows simulation experiment process and result.

Fig5. Handling process

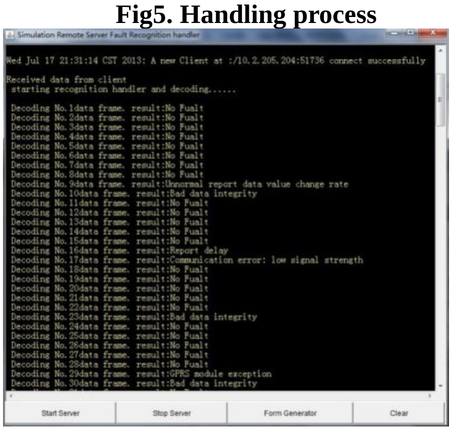
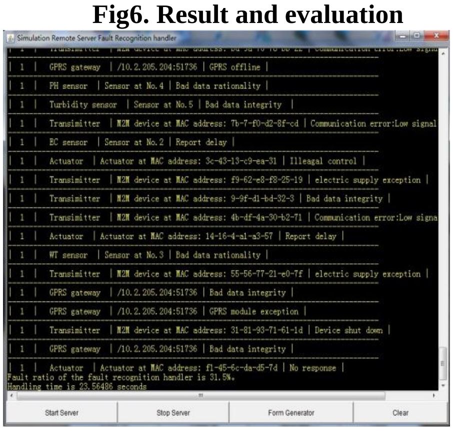


Fig6. Result and evaluation



The first experiment is to get the accuracy of the recognition method under different actual fault record ratio, and total records' count is fixed to 200. The experiment result is shown in fig7. The second experiment is about the accuracy under different total record count, the result is shown in fig8.

The last experiment is to test the time-cost under different fault record count, as the strategy time-cost is mainly on recognized fault records' handling, so this experiment change the fault record count and take the average fault records' handling time as the time-cost evaluation parameter, total record count is fixed to 400. The average fault records' handling time equals total handling time divides 400. The result proves the handling time is proportional to fault record count, as shown in fig .22.

Fig7. Accuracy under Different Actual Fault Record Ratio

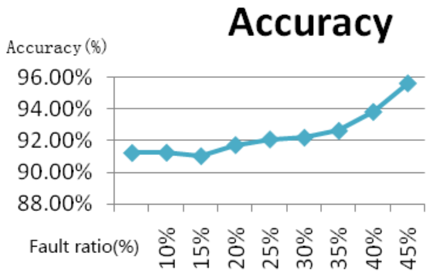


Fig8. Accuracy under Different Total Record Count

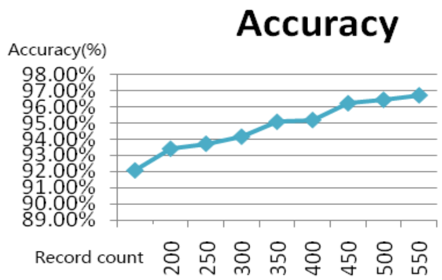
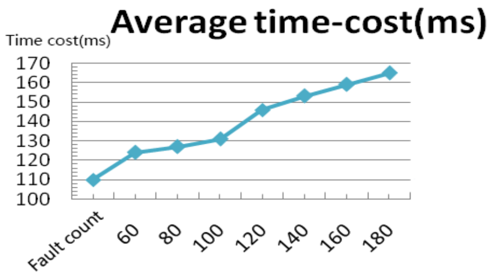


Fig9. Average time-cost



4.3 Application effect

The recognition strategy was applied in the Water Quality Monitoring system developed by China Agricultural University, the program has been deployed on the remote server of the system since December, 2012, by now it recognized 246586 records, and is still running on the server. Its high efficient, low time-cost and nice accuracy received praises of maintenance team.

5 Conclusions

The proposed pattern recognition method, based on data eigenvector and TCP transport protocol analyses the Water Quality Monitoring equipment comprehensively. The author designed the data transport protocol based on transmit rule, based on this transport protocol, the data filter is proposed. According to actual diagnosis experiment and knowledge, added threshold to the data filter, as a result, the filter is able to filter out data that beyond the threshold. A novel recognition method is proposed to handle fault data with high efficiency and low time-cost, the method is a combination of four strategies, the strategy is specifically, the method manage these strategies to cooperation more efficient.

After 7 months' application, program based on the proposed method performed well on Water Quality Monitoring system. Work logs and praises from maintenance team prove the method feasible and efficient.

Nevertheless, because of the shortage of eigenvector analysis and filter design, recognition misses and errors exist, the method has potential to be improved better, further research should focus on the reduction of the recognition misses and errors. Moreover, relative diagnosis algorithm is necessary to be designed.

Acknowledgement

Financial support from state science and technology support program (2011BAD21B01). Data, diagnosis knowledge and device support from device maintenance team of Research Center of Agricultural information technology. China Agricultural University is gratefully acknowledged.

References

[1]. Wu, K., D. Choudhury and H. Matsumoto, Wireless Power Transmission, Technology, and Applications. PROCEEDINGS OF THE IEEE, 2013. 101(6SI): p. 1271-1275.
[2]. Xiaodan, P., et al., Uplink transmission in the W-band (75-110 GHz) for hybrid optical fiber-wireless access networks. Microwave and Optical Technology Letters, 2013. 55(5): p. 1033-1036.
[3]. Lakshmanan, S. and R. Sivakumar, Proteus: Multiflow Diversity Routing for Wireless Networks with Cooperative Transmissions. IEEE TRANSACTIONS ON MOBILE COMPUTING, 2013. 12(6): p. 1146-1159.
[4]. Cheng, S.T., et al., Hierarchical Distributed Source Coding Scheme and Optimal Transmission Scheduling for Wireless Sensor Networks. WIRELESS PERSONAL COMMUNICATIONS, 2013. 70(2): p. 847-868.
[5]. Tahir, Y.H., et al., Transmission of Visual Data over Wireless Fading Channel in Real-Time Systems Based on Superposition Coding Scheme. ARABIAN JOURNAL FOR SCIENCE AND ENGINEERING, 2013. 38(6): p. 1459-1469.
[6]. Chunlai, Y., et al., Study on Wireless Energy and Data Transmission for Long-Range Projectile. IEEE Transactions on Plasma Science, 2013. 41(5): p. 1370-1375.
[7]. Refaat, S.S., et al., ANN-based for detection, diagnosis the bearing fault for three phase induction motors using current signal. 2013 IEEE International Conference on Industrial Technology (ICIT 2013), 2013: p. 253-258.
[8]. Talebi, H.A. and K. Khorasani, A Neural Network-Based Multiplicative Actuator Fault Detection and Isolation of Nonlinear Systems. IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY, 2013. 21(3): p. 842-851.
[9]. Jettanasen, C. and A. Ngaopitakkul, A novel probabilistic neural network-based algorithm for classifying internal fault in transformer windings. IEEJ TRANSACTIONS ON ELECTRICAL AND ELECTRONIC ENGINEERING, 2013. 8(2): p. 123-131.
[10]. Prieto, M.D., et al., Bearing Fault Detection by a Novel Condition-Monitoring Scheme Based on Statistical-Time Features and Neural Networks. IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, 2013. 60(8): p. 3398-3407.
[11]. Li, B., et al., Applying the fuzzy lattice neurocomputing (FLN) classifier model to gear fault diagnosis. NEURAL COMPUTING & APPLICATIONS, 2013. 22(3-4SI): p. 627-636.
[12]. Liu, H., et al., Adaptive neuro-fuzzy inference system based faulty sensor monitoring of indoor air quality in a subway station. KOREAN JOURNAL OF CHEMICAL ENGINEERING, 2013. 30(3): p. 528-539.
[13]. Hashemi, M. and M.S. Safizadeh, Design of a fuzzy model based on vibration signal analysis to auto-detect the gear faults. INDUSTRIAL LUBRICATION AND TRIBOLOGY, 2013. 65(3): p. 194-201.
[14]. Keskes, H., A. Braham and Z. Lachiri, Broken rotor bar diagnosis in induction machines through stationary wavelet packet transform and multiclass wavelet SVM. ELECTRIC POWER SYSTEMS RESEARCH, 2013. 97: p. 151-157.