WSN Spatio-temporal Correlation Data Fusion Method for Dairy Cow

Huaji Zhu, Yisheng Miao$^{(\boxtimes)}$, Huarui Wu
National Engineering Research Center for Information Technology in Agriculture, Beijing, China,
Beijing Academy of Agriculture and Forestry Sciences, Beijing, China
`myth1101@126.com`

Abstract—The cowshed environment has significant impacts on the yield, diseases and behaviors of dairy cows. Heat stress, in particular, has a great impact on yield. The cowshed environment monitoring system based on wireless sensor network can accurately sense the temperature and other environmental parameters in real time and provide basis for manual environmental intervention and control. Energy constraint is one of the important problems that affect the long-term stable monitoring by the dairy cow wireless sensor network. So, the weighted Markov chain method is used to predict the time series of cowshed temperature. Replacing the actual values with the predicted values at the cluster head can effectively reduce data traffic in the cluster, thereby reducing network power consumption. Test data show that, the average variance of the cowshed environment temperature predicted by the method proposed in this paper is 0.185, and the average power consumption is reduced by about 40% when the compression ratio is 0.3, which effectively prolongs the network lifetime. In addition to that, the cowshed environment prediction can also help make prejudgments for environmental control, reduce or avoid the heat stress of dairy cows after the environmental parameters exceed the thresholds and provide the basis for the multi-source data fusion for dairy cow.

Keywords—internet of things in dairy farming, wireless sensor network, data fusion, time series prediction, weighted Markov chain

1 Introduction

Since 2010, dairy farming in China has shifted from family free-ranging to intensive breeding. The increasing scale of dairy farming shows the necessity of automated production and information management in the dairy farming industry. An appropriate environment is the basic condition to maintain normal growth and production of dairy cows because cows will have stress against environmental changes, which will affect their production, behaviors and epidemic incidence. In North China, the most common heat stress causes huge production losses every year. Therefore, in a semi-open cowshed area, how to accurately monitor and control the environment in real time is very important to scientific breeding of dairy cows [1-2].

In recent years, scholars at home and abroad have conducted studies on the dairy information acquisition, monitoring and integration from different angles and provided solutions to the breeding environment feedback regulation and early warning of animal diseases [3-9]. Neng Benhai et al. from the Institute of Animal Sciences, the Chinese Academy of Agricultural Sciences developed a livestock and poultry breeding environment monitoring IoT system, which uses sensors to remotely monitor environmental parameters and transmits them to the data server through mobile network. It reminds users to make appropriate environmental regulation and control according to thresholds [10-12]. Considering the poor mobility, complex operations and high-power consumption of the current remote monitoring system for cowshed environment, Li Minzan et al. from China Agricultural University developed a remote monitoring system based on Ethernet and WeChat platform, which achieves stable and reliable monitoring on the environmental information of dairy farm. Shandong Agricultural University developed a universal intelligent terminal that can be applied in environmental monitoring. The terminal and temperature-humidity, gas particle concentration, air velocity, ammonia gas, light and carbon dioxide concentration sensors form an intelligent monitoring network through wireless transmission, which achieves fuzzy automatic control of environmental parameters [13,14]. At present, studies on environmental monitoring of dairy farming focuses on the design of sensor or system and give little attention to wireless sensor network (WSN) transmission technology for dairy farming environment. Data fusion is an important part of the transport layer. Since the dairy farming environment gradually changes on a continuous and regular basis, the data are highly redundant. The data fusion method can reduce the data traffic between nodes, thus reducing the power consumption and prolonging the network lifetime. The LEACH algorithm, which first proposed the concept of clustering, suggested carrying out data fusion at the cluster head to reduce the amount of remote data transmission. Wang et al. proposed a multi-parameter greenhouse WSN data fusion method in 2014, which achieves multi-parameter fusion at the cluster head through inter-parameter correlation analysis [15]. Gao Hongju et al. proposed a cluster head data fusion method based on the K-means method in 2015, which reduces the amount of data transmission by clustering and preserves data differences while eliminating redundancy [16].

In summary, currently there is little research on WSN transmission fusion in dairy farming environment, and the existing WSN data fusion methods mainly focus on data fusion at the cluster head. Between sensing nodes and the cluster head, data are still completely transmitted. Therefore, this paper proposes a WSN spatio-temporal correlation data fusion for dairy cow (STCDF-DC) from the perspective of cowshed temperature time series prediction, which provides basis for the long-term stable operation of dairy cow WSN, early warning against abnormal farming environment and subsequent multi-source data fusion research.

2 Cowshed temperature prediction model based on weighted Markov chain

Environment temperature is an important parameter of the dairy farming environment. Abnormal temperature change can cause stress responses of dairy cows, affecting milk production, and being in an unsuitable environment for a long time can also lead to diseases. This paper, based on the traditional cowshed WSN monitoring system, constructs a cowshed temperature prediction model with the weighted Markov chain, and reduces the actual data volume sent by sensing nodes to the cluster head through model prediction so as to realize time correlation fusion. The environment temperature prediction can also give early warning against abnormal environmental changes and improve the effects of environmental control by management personnel to ensure the appropriate growth of livestock and poultry. This prediction model also be extended and applied to other environmental parameters that continuously vary over time.

This paper converts the cowshed temperature data into a time series. The specific steps to construct the prediction model based on weighted Markov chain are as follows:

(1) Divide the state space into N grades (S={1, 2, …, N}) according to the state value of Markov chain;

The number of state spaces affect the final precision of the prediction results. After repeated tests, it is decided that N=5.

(2) According to the space grading standards, determine the grade of each state in the time series and then calculate the Markov transition probability matrix with different lag time, as shown below:

$$P^{(k)} = \begin{bmatrix} p_{11}^{(k)} & p_{12}^{(k)} & \text{L} & p_{1N}^{(k)} \\ p_{21}^{(k)} & p_{22}^{(k)} & \text{L} & p_{2N}^{(k)} \\ \text{M} & \text{M} & \text{O} & \text{M} \\ p_{N1}^{(k)} & p_{N2}^{(k)} & \text{L} & p_{NN}^{(k)} \end{bmatrix}$$

(1)

Where, k is lag time, k=1,2,…,n; n is the maximum order of lag time in the calculation; $P^{(k)}$ stands for the Markov transition probability matrix when the lag time is k; $p_{ij}^{(k)}$, i,j=1,2,…,N, stands for the one-step transition probability matrix from state i to state j when the lag time is k.

(3) Use the test statistic χ^2 to test the Markov property of the discrete time series. f_{ij} is the frequency of the time series one step from state i to state j. Then,

$$p_{ij} = f_{ij} \left/ \sum_{j=1}^{N} f_{ij} \right.$$

(2)

A marginal probability of transition probability is denoted as $p_{\cdot j}$:

$$p_{\cdot j} = \sum_{i=1}^{N} f_{ij} \Big/ \sum_{i=1}^{N} \sum_{j=1}^{N} f_{ij}$$

(3)

The test statistic χ^2:

$$\chi^2 = 2 \sum_{i=1}^{N} \sum_{j=1}^{N} f_{ij} \left| \log \frac{p_{ij}}{p_{\cdot j}} \right|$$

(4)

obeys the chi-square distribution with a free degree of $(N-1)^2$. Calculate the value of the test statistic χ^2 for the time series. If $\chi^2 > \chi_\alpha^2 (N-1)^2$, and $\alpha=0.05$, then it is believed that the original time series has the Markov property.

(4) According to the weighted sum result calculated based on the correlation between the previous and current time periods, predict the state the system is in at the current interval.

First, calculate the autocorrelation coefficient r_k of every order:

$$r_k = \sum_{i=1}^{m-k} \left(X_i - \bar{X} \right) \left(X_{i+k} - \bar{X} \right) \Big/ \sum_{i=1}^{m} \left(X_i - \bar{X} \right)^2$$

(5)

where, X_i is the i-th state value in the time series; $\bar{X}$ is the mean value of the time series; and m is the length of the time series.

Then, according to the autocorrelation coefficient, calculate the weight coefficient of the Markov chain model at each lag time ω_k:

$$\omega_k = \left| r_k \right| \Big/ \sum_{k=1}^{n} \left| r_k \right|$$

(6)

Then, according to the initial value of each state grade and the transition probability matrix from the i-th state grade to other states, obtain the probability $p_i^{(k)}$ at the same state grade in the prediction time period. Assign the weight ω_k to $p_i^{(k)}$ to obtain the predicted probability of the time series being in this state.

$$p_i = \sum_{k=1}^{n} \omega_k p_i^{(k)}$$

(7)

At last, take the state value at each grade in the time series as the initial value and obtain the predicted state value based on the predicted probability.

3 Dairy cow WSN data fusion

The dairy cow WSN data fusion proposed in this paper consists of two parts – the within-cluster fusion based on the data time series prediction model and the cluster head data fusion based on the spatial correlation of nodes.

In the within-cluster data fusion, the sensing node first calculates the predicted value T* according to the time series prediction model, compares it with the measured value T and calculates the variance. If the predicted variance is smaller than the threshold T0, it only transmits the instruction MESSAGE_ES, indicating using the predicted value to replace the measured value. If the predicted variance is greater than the threshold, it gives the instruction MESSAGE_TR_VAR, which contains the information on variance and deviating direction.

The cowshed environment is generally highly consistent, and the network coverage itself also contains a considerable degree of spatial redundancy information, but the acquisition nodes themselves can only acquire the environment information within their coverage and cannot consider the spatial correlation between different nodes in data fusion, so the acquisition nodes report their own temporal-correlated fusion results to the cluster head and the cluster head will carry out spatial-correlated data fusion according to the relationships between nodes. In order to achieve the maximum degree of cluster head fusion, this paper uses the "Within-Cluster Variance" method to aggregate the node data in the cluster so that the variance of the node data in each class with respect to their mean value is less than the threshold. The algorithm process is as follows:

1. Randomly select k nodes out of n ones in the cluster as the original class
2. For each of the remaining nodes, calculate its variance with the mean value of each class. If it is less than the threshold, then group it to the class and recalculate the mean of each class that has been obtained;
3. If none of them satisfies the requirement, they become a new class;
4. Iterates Step 2~3 until the number of classes becomes stable.

The final classification results should satisfy:

$$\mathrm{var}(i, \overline{D_j}) < TH, (\forall i \in D_j) \tag{8}$$

where, *TH* represents the intra-class variance threshold, and D_j is any class in the classification results.

For each class of node data, the arithmetic mean is used to represent the environmental data of all points to realize the data fusion and compression at the cluster head. In order to ensure the consistency of the degree of system information loss, the intra-class variance and the predicted variance threshold are both set at T0. In this way, the data fusion ratio can be changed by changing the threshold T0, so as to improve network power consumption and lifetime. This paper obtains the optimal threshold parameter through simulation test and comparison.

4   Simulation test and analysis

You may mention here granted financial support or acknowledge the help you got from others during your research work.

This paper validates the model and method proposed through simulation test. In order to verify the effectiveness of the prediction model, this paper carried out an experiment on No.3 cowshed of Hebei Fengning Ranch. In the experiment, the cowshed environment temperature was taken as the test data, collected on an hourly basis and three times in total. On July 4th, 2017, the measured and predicted cowshed environment temperatures are shown in Figure 2. On that day, measured or predicted temperature was collected 72 times in total. As can be seen from Figure 1, the predicted temperature curve is generally consistent with the measured one, indicative of good prediction accuracy. The points with large differences between the predicted and measured values are mainly distributed in the vicinity of the curve inflection points, or the area where the curve trend changes greatly. When there is a big difference, the predicted curve can quickly adjust itself, showing the good convergence of the model. The variances of the predicted and measured temperatures are shown in Figure 2. It can be seen that the minimum predicted variance of temperature is 0.00003 and the maximum variance is 1.84, with an average variance of 0.185. Considering the local average temperature was 22.7 °C on that day, if the average deviation is within ±0.5 °C, i.e. the estimation deviation is less than 2%, it can be deemed that this model realizes accurate prediction of environment temperature.
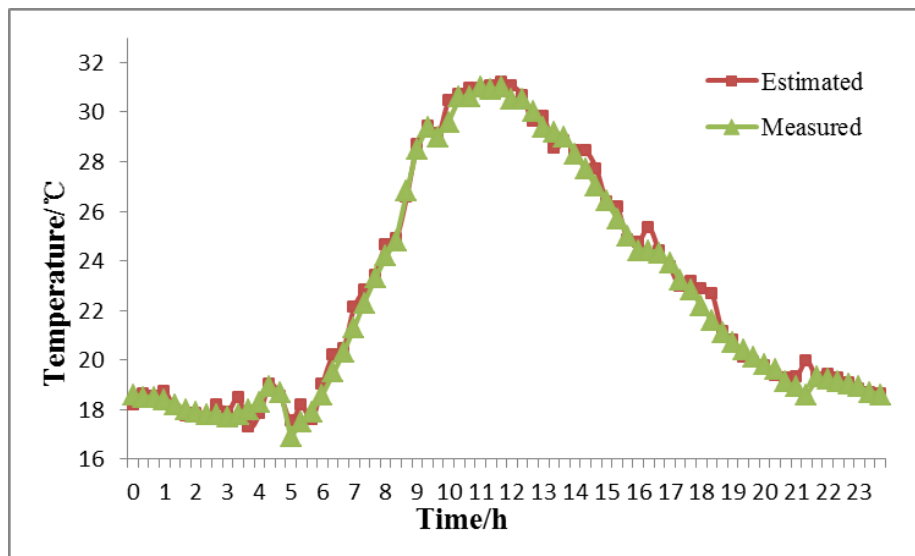

Fig. 1. Comparison between the predicted value and measured value of cowshed environment temperature
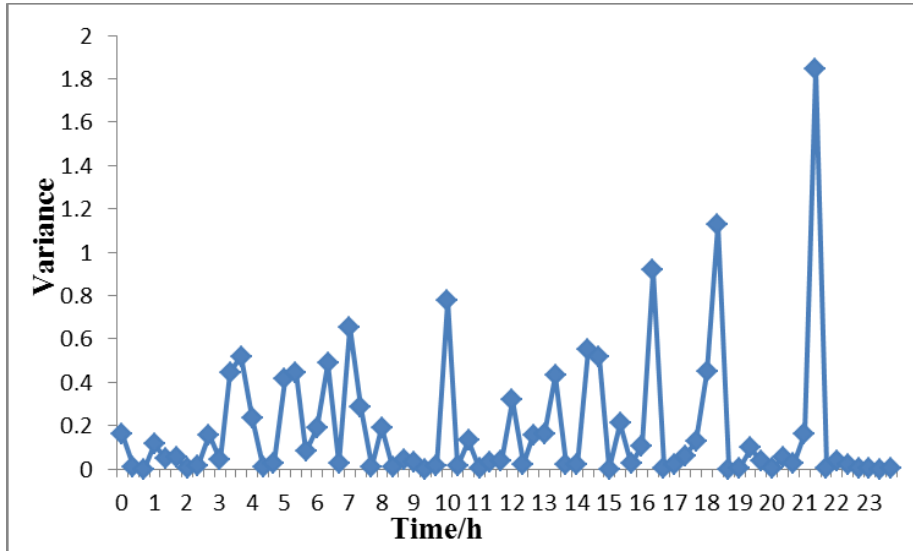
Fig. 2. Variances of the predicted value and measured value of cowshed temperature

In order to verify the performance of the proposed method, the cluster-head k-means data fusion algorithm is used for comparison. The analysis focuses on data traffic volume and power consumption. MATLAB 7.0 is used as the simulation environment and the wireless-signal first-order power consumption model is used as the network power consumption model [17-19]. System parameters are shown in Table 1.

As shown in Figure 3, with the compression ratio increasing, the average power consumption of both algorithms in a single round is on decline, because the high compression ratio reduces the amount of network data transmission, thereby effectively reducing the power consumption by network transmission. It can also be seen from the figure that the power consumption curve of the proposed STCDF-DC algorithm declines faster than the k-means algorithm, and under the same compression ratio, the network power consumption is significantly lower than that by the k-means algorithm, because the k-means algorithm only performs data fusion at the cluster head, while the STCDF-DC algorithm reduces the amount of within-cluster data transmission through time series prediction, resulting in better fusion efficiency and lower power consumption.

Under different compression ratios, the mean square errors of data in the two algorithms are shown in Figure 4. With the compression ratio increasing, the mean square errors in the two algorithms both increase rapidly. The mean square error of the k-means algorithm is about 0.1°C when the compression ratio is 0.2, and reaches the maximum value – about 3.2 °C when the compression ratio is 0.65. The mean square error of the STCDF-DC algorithm is about 0.25 °C when the compression ratio is 0.2, and reaches maximum value - about 3 °C when the compression ratio is 0.65. Overall, the mean square error of the k-means algorithm is lower than that of the STCDF-DC algorithm. The reason is that the k-means algorithm transmits the original data in the

Table 1. Simulation Parameters

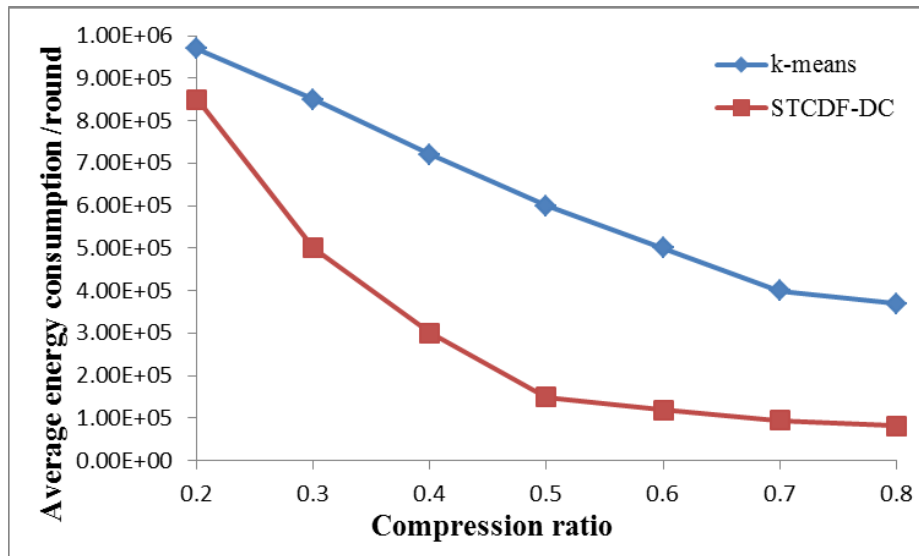| Parameter	Value	Parameter	Value
Monitoring area	100×100	ε_f	10pJ/b•m^{-2}
Number of nodes	100	E_{tx}, E_{rx}	50 nJ/b
E_0	0.5J	l	400bit
Sink location	(200, 200)		

Fig. 3. Average power consumption per single round under different compression ratios
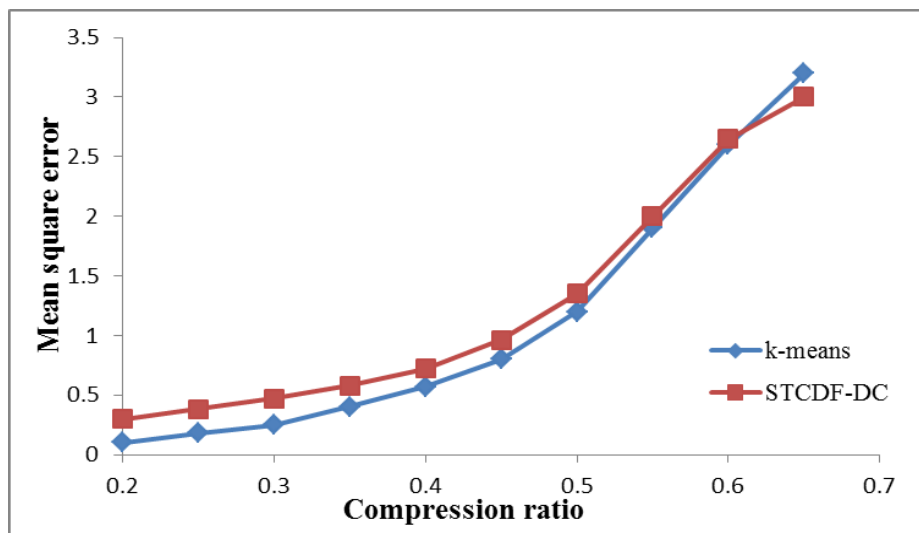

Fig. 4. Mean square errors under different compression ratios

collection of cluster data, while the STCDF-DC algorithm adopts the time-series-based prediction fusion method in the cluster data collection, which slightly increases the overall mean square error. But with the increase of the compression ratio, the mean square errors of the two algorithms are getting closer. When the compression ratio is 0.65, the mean square error of the STCDF-DC algorithm is even lower than that of the k-means algorithm, probably because with the compression ratio increasing, the contribution of the cluster head fusion to the mean square error is gradually increasing, which reduces the deviation effect brought by the within-cluster prediction fusion in the STCDF-DC algorithm. This also shows that the STCDF-DC algorithm is effective in controlling the interclass variance during cluster head fusion.

From Figure 3 and 4, it can be found that, when the compression ratio is 0.3, despite the increment of less than 0.1 °C in the mean square error, the time series prediction fusion method used by the STCDF-DC algorithm can reduce the average power consumption per single round from about 850uJ to about 500uJ, a decrease of about 40%, compared with the k-means algorithm. The two algorithms experience the biggest difference in relative power consumption when the compression ratio is 0.5, but since the mean square errors of both algorithms reach about 1.5 °C, which cannot fully meet the requirement for accurate warning about dairy production environment, it is suggested that setting the compression ratio within the range of 0.2~0.3 in actual practice to achieve better results.

5 Conclusion

In light of the transmission and early warning problems in the environmental monitoring IoT for dairy cow, based on the cluster head data fusion in the cluster-structure network, this paper proposes a dairy cow WSN data fusion algorithm based on time series prediction to address the redundancy in inter-cluster data transmission. The method uses the weighted Markov chain to predict the time series of the perceived data and replaces the actual value with the predicted value at the cluster head to effectively reduce the within-cluster data traffic and realize temporal-correlated fusion; it classifies the within-cluster node data at the cluster head and optimize the classification result through the intra-class variance to achieve spatial-correlated fusion. The simulation test results show that the predicted mean variance is 0.185, indicating it can achieve good prediction results. By means of spatio-temporal data fusion, this method not only greatly reduces the average power consumption by communication, but also ensures good performance in mean square error, which provides the basis for long-term and stable network monitoring. In the future, in addition to temporal and spatial correlation analysis, subsequent studies may also consider deep fusion from the perspective of multi-parameter correlation analysis to further improve fusion efficiency.

6 Acknowledgements

7 References

[1] Zhu, J., Ma, S. S., Bi, Y. G., Cui, H. M. (2010). The development status and trend of agriculture and husbandry informatization in our country. *Journal of Agricultural Mechanization Research*, 32(4), 199-204.

[2] Yao, L. J. (2014). Present situation, mode and prospect of animal husbandry informatization. *Technical Advisor for Animal Husbandry*, (7), 278.

[3] Gillespie, J. R. (2009). Modern livestock & poultry production. *Boston: Cengage Learning*.

[4] He, D. J., Liu, D., Zhao, K. X. (2016). Review of perceiving animal information and behavior in precision livestock farming. *Transactions of the Chinese Society for Agricultural Machinery*, 47(5), 231-244.

[5] Pomar, C., Hauschild, L., Zhang, G. H. (2009). Applying precision feeding techniques in growing-finishing pig operations. *Revista Brasileira de Zootecnia*, 38(SPE), 226-237. https://doi.org/10.1590/S1516-35982009001300023

[6] Fu, R., Zhang, Q., Yu, F. (2014). Information technology direction and problem analysis of pig breeding. *Journal of Henan Agricultural Sciences*, 43(12), 169-172.

[7] Zhang, Q. D., Tian, L. Y. (2006). Problems and countermeasures in the development of livestock and poultry breeding areas in China. *Transactions of the Chinese Society of Agricultural Engineering*, 22(S2), 68-70.

[8] Cornou, C. (2007). Automated monitoring methods for group housed sows, *Copenhagen: University of Copenhagen*.

[9] Lu, M. Z., Shen, M. X., Ding, Y. Q., Yang, X. J., Zhou, B., Wang, Z. G. (2012). Review on the intelligent technology for animal husbandry information monitoring. *Scientia Agricultura Sinica*, 45(14), 2939-2947.

[10] Song, F. Z., Zhang, D. J., Chen, D. M., Li, S. Q. (2013). A systematic warning approach for the livestock-poultry's environmental system. *Journal of Chongqing University (Natural Science Edition)*, 36(5), 130-137.

[11] Mei, S. Q., Peng, X. W. (2009). Research progress on the healthy breeding of pigs. *Journal of Anhui Agricultural Sciences*, (2), 602-604.

[12] Xiong, B. H., Yang, Z. G., Yang, L., Pan, X. H. (2015). Review on application of Internet of Things technology in animal husbandry in China. *Transactions of the Chinese Society of Agricultural Engineering*, 31(S1), 237-246.

[13] Wen, S. H. (2014). Wen group explores the use of Internet of things in pig farms, *Investigation report on rural informatization demand in China, discussion on sustainable development model of rural informatization and experience exchange meeting*.

[14] Xue, X. Y., Liu, P. Z. (2008). Design of A control and information collecting system for livestock breeding farms. *Agricultural Equipment & Technology*, 34(5), 22-24.

[15] Gao, H. J., Liu, Y. Z., Chen, S. (2015). Data fusion of WSN cluster head nodes based on improved K-means algorithm. *Journal of agricultural machinery*, (s1), 162-167.

[16] Wang, X. H., Xu, L. H., Wei, R. H. (2014). A new fusion structure model on greenhouse environment data and a new fusion algorithm of sunlight. *International Conference on Wireless Communication and Sensor Network. IEEE*, 418-424. https://doi.org/10.1109/WCSN.2014.91

[17] Qiu, L. D., Liu, T. J., Lin N. (2014). Data aggregation in wireless sensor network based on deep learning model. *Chinese Journal of Sensors and Actuators*, (12), 1704-1709.

[18] Xiang, M., Shi, W. R. (2010). Intra cluster data management algorithm for wireless sensor networks based on data association. *Journal of Automation*, 36(9), 1343-1350.

[19] Davankov, V. A., Rogozhin, S. V., Kurganov, A. A. (2015). Analysis of stochastic - coverage and connectivity in sensor networks with boundary deployment. *IEEE Transactions on Intelligent Transportation Systems*, 16(4), 1-11.

8   Authors

Huaji Zhu is an associate research fellow in National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China. His current research interests include Aquaculture informatization, visualization of agricultural spatial data and wireless sensor networks (e-mail: zhuhj@nercita.org.cn).

Yisheng Miao is an assistant research fellow at National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China. He is mainly engaged in wireless sensor networks and intelligent systems in agriculture. (e-mail: myth1101@126.com).

Huarui Wu is a research fellow at National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China. He is mainly focused on artificial intelligence and Internet of things applications in agriculture. (e-mail: wuhuarui1975@163.com).