CrossMark

Distributed crowd filtering mechanism based on heterogeneous network delay and data packet loss constraint

Xiang-jun Zhou

Abstract

The differences among heterogeneous networks, network delay, and data packet error could decline network performance. In order to solve the above problems, we study a kind of distributed crowd-filtering mechanism. This mechanism can effectively control the heterogeneous network diversity delay by updating the coordination signal matrix, the aggregation signal matrix, and the control signal matrix. The mechanism can effectively improve network reliability based on the packet loss constraint. Considering the influence of heterogeneous network latency on the reliability of data mining results, we studied the crowd-filtering scheme for eliminating the effect of packet loss on the efficiency of data filtering. Experimental results show that, compared with the ant colony algorithm, the distributed crowd filter can accurately predict the data position, reduce the filtering error, and improve the classification accuracy and the data characteristics.

Keywords: Crowd filter, Heterogeneous network, Diversity delay, Packet loss constraint

1 Introduction

Heterogeneous wireless networks, which have the advantages of mobility, easy deployment, distribution [1], and real-time, are widely used in the fields of environmental monitoring, traffic control, and medical care. Due to the differences of heterogeneous terminals [2], limited wireless bandwidth and energy supply [3], and other issues, a heterogeneous network [4] is difficult to effectively solve the data packet loss and diversity network delay [5].

A personalized fine-grained filtering scheme with privacy preservation was proposed by Zhang [6] in mobile social networks. The authors also developed a social-assisted filter distribution scheme where the filter creators send filters to their social friends.

The novel distributed consensus filter was proposed in order to solve the target tracking in the article by Zhu et al. [7]. The novel hybrid particle/finite impulse response filtering algorithm was proposed by Pak et al. [8] for improving reliability of particle filter-based localization schemes under harsh conditions causing sample impoverishment.

The information weighted consensus-based distributed particle filter was presented by Tang et al. [9] to solve the problem of information fusion estimation for large-scale sparse wireless sensor networks.

The improved particle filter was proposed by Yoo et al. [10] by incorporating semi-supervised machine learning for location estimation in mobile sensor networks.

The novel recurrent neural network-based vector control method was proposed by Fu and Li [11] for a single-phase inverter with an LCL filter, which is trained based on adaptive dynamic programming principle.

A new tractable Bernoulli filter was proposed by Abdulkadir et al. based on the random matrix framework to track an extended target in an ultra-wideband (UWB) sensor network [12].

However, how to carry on the network time delay management and the reliability guarantee under limited resources in the heterogeneous network has become the hot issue of domestic and international research at present. The rest of the paper is organized as follows. Section 2 describes the Heterogeneous network delay and data packet loss constraint scheme. The distributed crowd filtering mechanism was discussed in section 3. In Section 4, we

Correspondence: xajunzhou@163.com
Department of Information Technology, Guangdong Teachers College of
Foreign Language and Arts, Guangzhou, China

proposed the analysis results of crowd Filter algorithm. Finally, the conclusions is given in Section 5.

2 Heterogeneous network delay and data packet loss constraint scheme

In the heterogeneous sensor network, due to the burst and random error, there are the transmission delay, convergence delay, the upstream and the downstream delay etc. because of the burst and random error, as shown in Fig. 1. In this model, the sensor nodes are controlled independently and the signals are received or transmitted according to the change of environment. The coordinator is driven by the session requirement. Convergence nodes are often in periodic operation state or transient dormancy state. In a dormant period, the coordinator can be activated by a session or a server request. The bit error rate is random and burst. The bit error rate affects the data flow between the coordinator and the sensor. The bit error affects the data flow between the sensor and the sink node. The transmission delay is the network delay between coordinator and sensor. The convergence delay is the network delay between the sensor and the sink node. The uplink communication delay between the sink node and the server is the uplink delay. Downlink delay is the delay of the controller and the sink node.

Assuming that network error will not result in the packet error. In a continuous network operating cycle T, a linear description of the network system is shown in Eq. 1.

$$
\begin{cases}
x(t) = C_O(t)x + C_G(t)x + C_T(t)x \\
\tau = a_1\tau_{TD} + a_2\tau_{CD} + a_3\tau_{DLD} + a_4\tau_{ULD} \\
1 = a_1 + a_2 + a_3 + a_4 \\
\tau \in [\alpha T, \quad \beta T]
\end{cases}
\tag{1}
$$

In Eq. 1, we use the coordinate signal matrix X (T) CO, the convergence signal matrix CG (T) x, and the control signal matrix CT (T) x computing network using the signal sequence x (t). We can obtain the system network delay according to the weight coefficient a1, a2, a3 and a4. Here, τ_{TD} is transmission delay. τ_{CD} is delay said protector. τ_{DLD} is downlink delay. τ_{ULD} is uplink delay. In order to maintain the consistency of network delay and the nature of signal attenuation, the sum of a1, a2, a3, and a4 must be 1. α is the lower weight of the working cycle of the network delay. β is the upper bound weight for the work cycle of network delay.

In heterogeneous networks, (1) x (T) is synchronized with the distributed feedback. Further enhancement of the linear properties of heterogeneous networks, as shown in Eq. 2.

$$
\begin{cases}
x(\tau t) = e^{\alpha+\beta}x(t) + f(t) \\
f(t) = \int_{\alpha T}^{\beta T} e^{\beta T+\tau}(C_O(t) + C_G(t) + C_T(t))d\tau
\end{cases}
\tag{2}
$$

In the optimization Eqs. 1 and 2, the delay model of the heterogeneous network is evolved into a closed loop system. Heterogeneous network time delay control logic matrix $C_L(\tau)$ is as shown in Eq. 3.

$$
C_L(\tau) = \begin{bmatrix}
\phi(\tau) + \sum_{i=1}^{N} g(t_i)\tau_i & \alpha\sum_{i=1}^{N} g(t_i)\tau_i & \alpha \\
\phi(\tau) & \phi(\tau) & \phi(\tau) \\
\sum_{i=1}^{N} g(t_i)\tau_i & \beta\sum_{i=1}^{N} g(t_i) & \beta\sum_{i=1}^{N} \tau_i
\end{bmatrix}
\tag{3}
$$

Here, $\phi(\tau)$ denotes the network latency control overhead. N is the number of signal samples. $g(t_i)$ denotes the closed loop feedback weight.
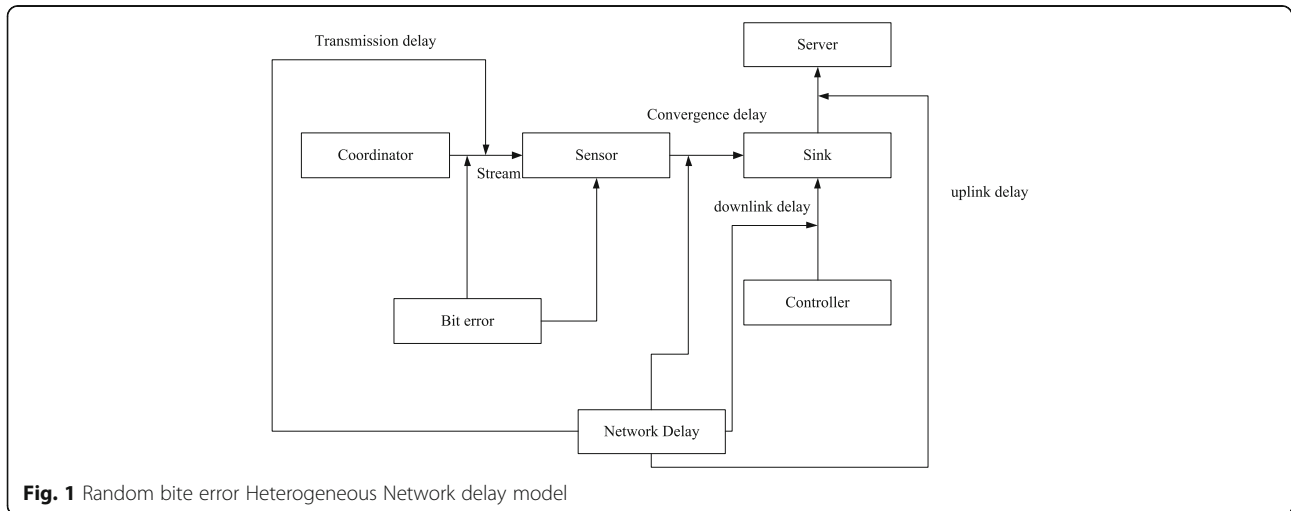

Fig. 1 Random bite error Heterogeneous Network delay model

When the error caused the packet error or loss, the packet loss constraints would be carried out. The improvement of Fig. 2 structure are shown in Fig. 2.

Heterogeneous networks take the controller as the core. The real-time processing of data packet dropout is met by the connection and closing of the constraint logic switch. If the data packet loss, the control logic would send $\phi(\tau)$ and $\sum_{i=1}^{N} g(t_i)$ to delay module. The delay module sends the control signal to the control module. The constraint logic module switches on the updated control logic. Based on data stream $C_L(\tau)$, the controller can get the bit error rate and packet loss rate. If data packet is lost in transmission process, the controller cannot obtain directly the measured values. Only through the constraint module, the reliability maybe guaranteed. The matrix $C_L(\tau)$ of formula (3) would be the $\overline{C}_L(\tau)$ of formula (4) after the packet loss constraint.

Equation 3 shows the matrix $C_L(\tau)$ of Eq. 4 after the packet loss constraint.

$$\overline{C}_L(\tau) = \begin{bmatrix} 0 & \sum_{i=1}^{N} g(t_i)\tau_i & C_{ST} \\ \int \phi(\tau)d\tau & C_{ST} & 0 \\ 0 & C_{ST}\int \phi(\tau)d\tau & 0 \end{bmatrix} \quad (4)$$

Here, C_{ST} is the data packet loss constraint weight.

In the heterogeneous network transmission process, through the collaboration coordinator and controller, the combination of network delay and data packet dropout, delayed constrained data flow, and transport and network distributed rate provide accurate and reliable network architecture.

3 Distributed crowd-filtering mechanism

In heterogeneous networks, it is a key problem to obtain the optimal mining results from the diversity of data. Some scholars have applied artificial intelligence technology in heterogeneous networks and improved the accuracy and efficiency of data mining. However, this scheme ignores the impact of heterogeneous network latency on the reliability of data mining results. It is difficult to deal with the impact of data packet loss on the efficiency of data filtering. Therefore, we apply the theory of crowd to heterogeneous network data mining. The distributed crowd heterogeneous network architecture is shown in Fig. 3. The integrated module of collection, transport and reservation of Fig. 3 is given by Fig. 4.

(1) Capture network latency
    If the network time delay is greater than the upper limit value, the network time delay is based on the data characteristics. Filtering data stream based on network characteristics and time delay. The most serious work cycle of data packet dropout is taken as a constraint feature subset.
(2) Construct a crowd compound set
    The network delay characteristic set and the restraint special collection would be processed with the crowd method.
(3) The excitation of crowd efficiency
    Based on the data size and filtering requirements, different incentive schemes are set up to speed up the convergence rate of the swarm intelligence. The dynamic adaptive scheme can be used in the crowd processing mechanism, which not only expands the existing incentive strategy but also can provide better protection for the crowd.
(4) Processing and filtering the features of datasets
    Data set feature processing scheme includes feature collection, feature conversion and reservation. The
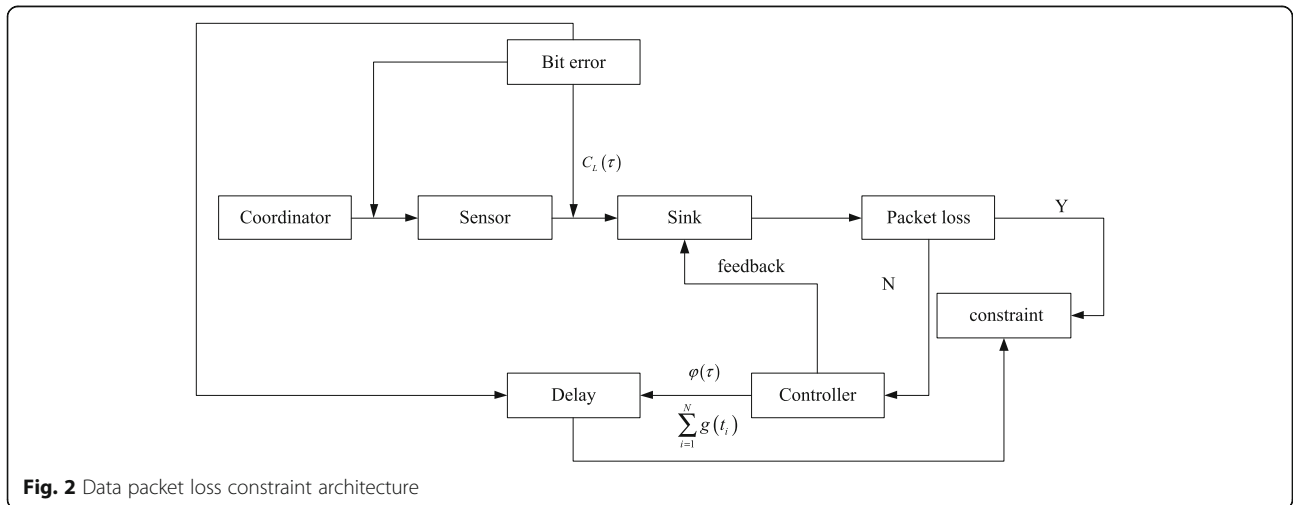

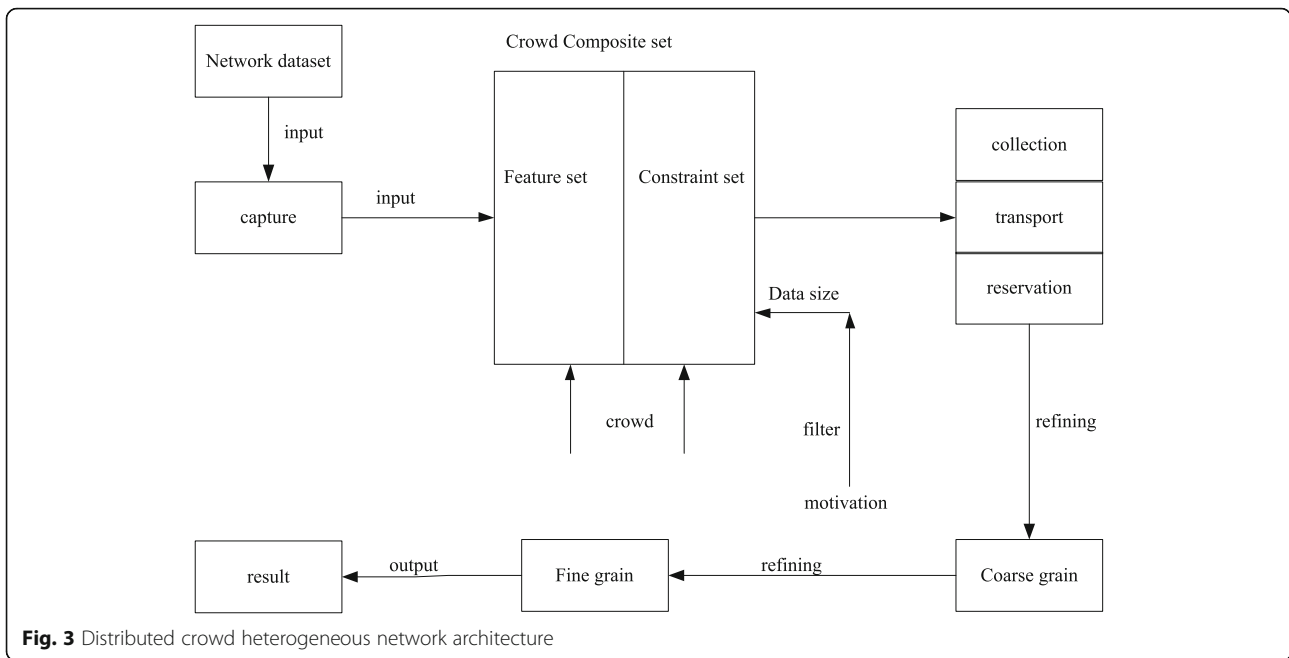**Fig. 2** Data packet loss constraint architecture

Fig. 3 Distributed crowd heterogeneous network architecture

filter feature of data set would be pre-fetched. The classification characteristics of data set could be obtained based on the classification accuracy of data set. According to the characteristics of the data set, the crowd filter would select the appropriate crowd incentive strategy. Based on the complexity characteristics of the data set classification, the transformation of the subset of the data set would be completed.

(5) Refine the coarse grain

Based on the coarse particle size feature subset, the global traversal is carried out from the highest complexity subset of data. Characteristic treatment of coarse grain size in the reverse order of incentive restriction. The data features of each data subset have smaller feature subset, higher classification accuracy, and better clustering.

(6) Refine the fine grain

According to the filtering principle and precision, the fine granularity classification is performed on several approximate coarse grain feature subsets. The difference between the maximum fine grained data subset generates a fine grain refining combination. The results from the combination of XOR and single feature subset of data, as the final result of refining.
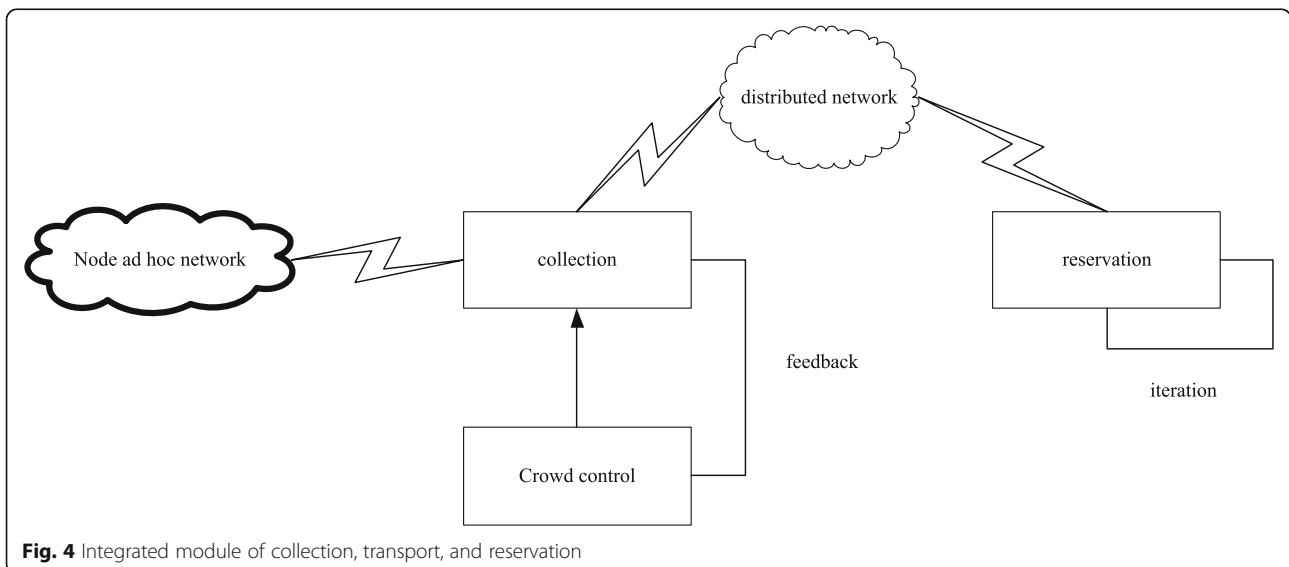

Fig. 4 Integrated module of collection, transport, and reservation

The crowd distribution transition probability P_{cd} (m, n) can be obtained from Eq. 5 from the m dataset to the n dataset.

$$P_{cd}(m,n) = \begin{cases} 1, CL(m) > CL(n) \\ \dfrac{\tau_{m \to n}\cos\alpha + \tau_{n \to m}\cos\beta}{\sum_{i \in m}[\tau_\beta(n)]^i + \sum_{i \in n}[\tau_\alpha(m)]^i}, CL(m) < CL(n) 0, CL(m) = CL(n) \end{cases} \tag{5}$$

Here, C_L represents the complexity of the data subset.

$$C_M = \begin{cases} 0, CL(m) \geq CL(n) \\ \dfrac{\sum_{i=1}^{m} \Delta\tau_i}{\sum_{i \in m}[\tau_\alpha(m)]^i [\tau_\beta(n)]^i}, CL(m) < CL(n) \end{cases} \tag{6}$$

Constructed filter input set for the above distributed crowd processing results $D_F = \{y : m \leq y \leq n\}$. The filter object of the collection is shown in Eq. 7.

$$\begin{cases} f(x) = C_M \|D_F(x)\|^2 \|A\| \\ A = \begin{bmatrix} -\alpha & 1 & 0 \\ 0 & -\beta & 1 \\ 1 & -1 & 0 \end{bmatrix} \end{cases} \tag{7}$$

The filtering algorithm is described as follows:

(1) Select the closed interval [mi, ni] in each data subset. According to Eq. 8, the dataset of filtered hierarchical FPi is obtained.

$$FP_i = \left\{ x_i : DF_i + \frac{\tau_{m \to n}(x_i - x_{i-1})}{\|A\|^2} \right\} \tag{8}$$

Here, the closed interval should be filtered based on the network delay, delay capture and the complexity of the distribution complex degree.

(2) Based on Eqs. 5 and 6, combined with the filtering requirement, the characteristic measure and reservation for each data packet are carried out. This will help improve the accuracy of the filter.

$$\begin{cases} ME_F = \sum_{i=1}^{m} A_{i,i}(\cos\beta x_i - DF_i) \\ F_R = \lambda \left(\sin\alpha - \sum_{i=1}^{n} DF_i x_i \right) \end{cases} \tag{9}$$

Here, ME_F represents the characteristic measure matrix. The matrix is obtained by the characteristic matrix, the crowd transfer starting point dataset, and the filter input set. F_R representation feature reservation matrix. The matrix is obtained by solving the target dataset and the feature set of distributed crowd data. λ represents the feature set aside weight.

(3) The updated rules of the filtering matrix of the distributed crowd excitation (Eq. 10) are shown in real time.

$$\begin{cases} x(t) = \cos\alpha DF(x) + \sin\beta A(x) \\ u(t) = \sqrt{|\cos\alpha - \sin\beta|} \sum_{i=1}^{n} \dfrac{FP_i}{\lambda} \end{cases} \tag{10}$$

(4) The conditions for the stability of the distributed crowd filtering are shown in Eq. 11. In order to meet the conditions, you can call the optimal polling strategy shown in Fig. 3.

$$\cos\beta DF(x)\left(\sin\alpha - \sum_{i=1}^{n} DF_i x_i \right) \leq \dfrac{\lambda \sum_{i=1}^{n} FP_i}{\sqrt{|\cos\alpha - \sin\beta|}} \tag{11}$$

4 Analysis of crowd-filter algorithm

Experimental filter data is the example dataset of machine learning data warehouse. Seventy-five percent of each dataset is used as a training set; 25 % of data is used as a test set. The training part and test part of the data size, network delay, and incentive weights are randomly generated. The transmission of the data elements in each dataset is subject to Poisson distribution. Data features and diversity are shown in Table 1.

We wanted the proposed Distributed Crowd Filtering mechanism based on Heterogeneous network delay and data packet loss Constraint (DCFHC) and constraints of the ant algorithm (ant filter algorithm based on constraint based on Ant Filter algorithm based on Constraint (AFC)) in data location, error filtering, classification accuracy, and the data of mean performance.

Figure 5 shows the relationship between the data position of the two filtering algorithms and the data transmission speed. Figure 6 shows the relationship between the filtering error and the number of samples of the two filtering algorithms. DCFHC can accurately locate the data at high transmission speed. The filtering error of DCFHC is obviously lower than that of the AFC algorithm. The DCFHC scheme can weaken the random nature of the error and the interference of the burst on the data transmission by accurately capturing the transmission delay of heterogeneous sensor networks, convergence delay, uplink delay and downlink delay. DCFHC through heterogeneous network delay controls logic to

Table 1 Data features and diversity

Dataset	Number of features	Sample	Training sample	Test sample
1	15	325	213	123
2	14	287	145	98
3	12	234	89	45

Fig. 5 Data position

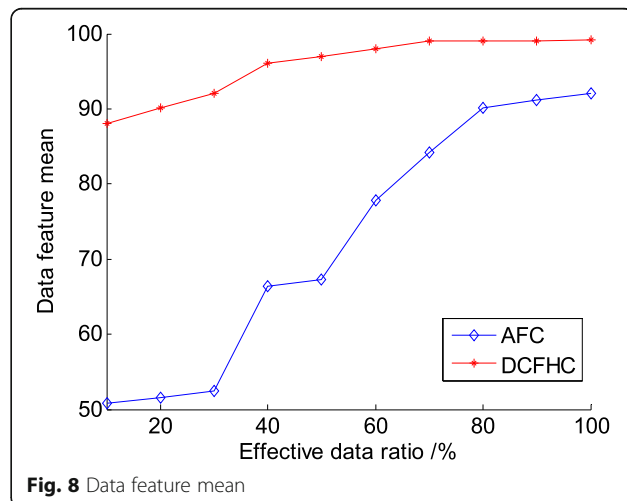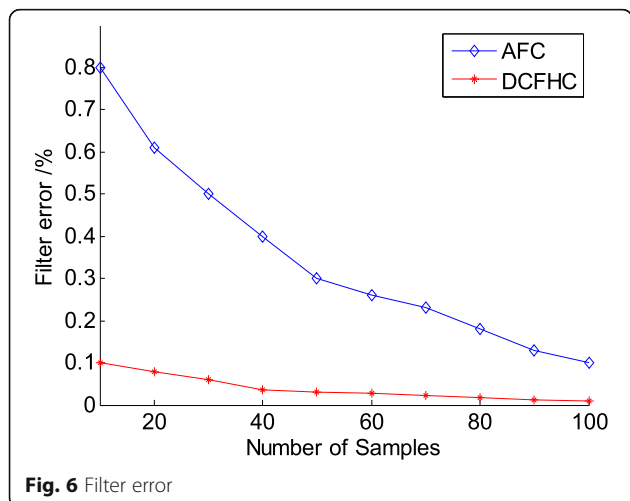
Fig. 7 Classification accuracy
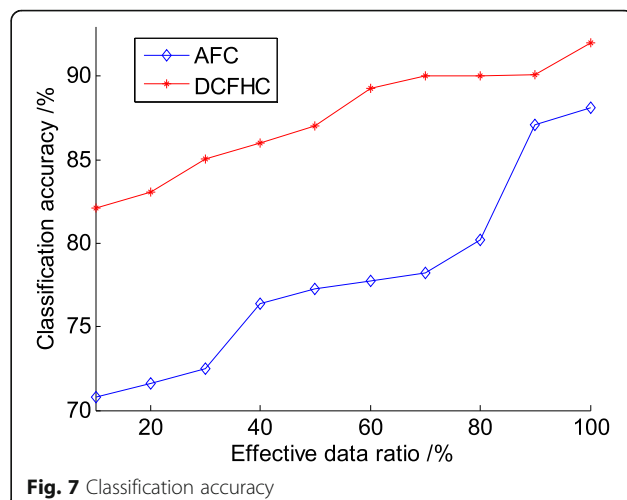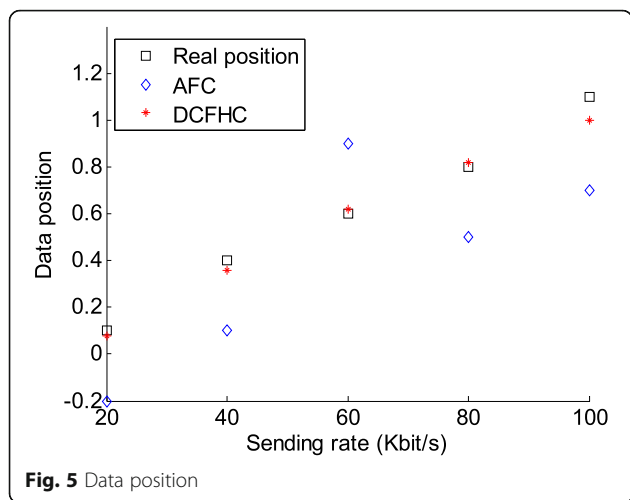
improve the network delay and data packet loss on the network reliability.

The relationship between the classification accuracy of the two algorithms and the effective data ratio is shown in Fig. 7. The relationship between the data characteristics of the two algorithms and the effective data ratio is shown in Fig. 8. Distributed swarm intelligence architecture is based on data features to capture the network delay. The network characteristics and delay characteristics are considered in the filter data stream. Using the crowd method to obtain the network delay feature set and the constraint feature set. The DCFHC can obtain the network delay feature set and the constraint feature set by using the crowd method. The DCFHC has the adaptive crowd efficiency excitation algorithm. Data set would be processed based on classification feature. DCFHC would refine the filtering result through coarse grain size and fine grains. Therefore, DCFHC has a

higher classification accuracy and more reasonable characteristics than AFC.

5 Conclusions

We study a distributed crowd-filtering mechanism to eliminate the heterogeneity of heterogeneous networks, network delay, and data packet loss on network performance. First, in the heterogeneous network, we set up the coordinate signal matrix, the aggregation signal matrix, and the control signal matrix. We updated the above matrix to effectively control the diversity of heterogeneous network delay. Second, based on the packet loss constraint, the transmission reliability of network packets is improved effectively. Then, considering the influence of heterogeneous network delay on the reliability of data mining results, the effect of packet loss on the efficiency of data filtering is studied. Finally, in the

Fig. 6 Filter error

Fig. 8 Data feature mean

aspect of the data location, the classification accuracy of filtering error, and the improvement of data characteristics, the distributed crowd-filtering algorithm has better performance.

Competing interests
The author declares that he has no competing interests.

References
1. A Oka, L Lampe, Distributed target tracking using signal strength measurements by a wireless sensor network [J]. IEEE J. Sel. Areas Commun. **28**(7), 1006–1015 (2010)
2. F Hashim, KS Munasinghe, A Jamalipour, On the negative selection and the danger theory inspired security for heterogeneous networks [J]. IEEE Wirel. Commun. **19**(3), 74–84 (2012)
3. BH Jung, NO Song, KS Dan, A network-assisted user-centric WiFi-offloading model for maximizing per-user throughput in a heterogeneous network [J]. IEEE Trans. Veh. Technol. **63**(4), 1940–1945 (2014)
4. M Gerasimenko, D Moltchanov, R Florea et al., Cooperative radio resource management in heterogeneous cloud radio access networks [J]. Access IEEE **3**, 1 (2015)
5. D Tian, J Zhou, Y Wang et al., A dynamic and self-adaptive network selection method for multimode communications in heterogeneous vehicular telematics [J]. IEEE Trans. Intell. Transp. Syst. **16**(6), 1–17 (2015)
6. K Zhang, X Liang, R Lu et al., PIF: A personalized fine-grained spam filtering scheme with privacy preservation in mobile social networks [J]. IEEE Trans. Comput. Soc. Syst. **2**(3), 41–52 (2015)
7. S Zhu, C Chen, W Li et al., Distributed optimal consensus filter for target tracking in heterogeneous sensor networks [J]. IEEE Trans. Cybern. **43**(6), 1963–1976 (2013)
8. JM Pak, CK Ahn, YS Shmaliy et al., Improving reliability of particle filter-based localization in wireless sensor networks via hybrid particle/FIR filtering [J]. IEEE Trans. Ind. Inf. **11**(5), 1089–1098 (2015)
9. W Tang, G Zhang, J Zeng et al., Information weighted consensus-based distributed particle filter for large-scale sparse wireless sensor networks [J]. IET Commun. **8**(17), 3113–3121 (2014)
10. J Yoo, W Kim, HJ Kim, Distributed estimation using online semi-supervised particle filter for mobile sensor networks [J]. IET Control Theory Appl. **9**(3), 418–427 (2015)
11. X Fu, S Li, Control of single-phase grid-connected converters with LCL, filters using recurrent neural network and conventional control methods [J]. IEEE Trans. Power Electron. **31**(7), 5354–5364 (2016)
12. A Eryildirim, MB Guldogan, A Bernoulli filter for extended target tracking using random matrices in a UWB sensor network [J]. IEEE Sensors J. **16**(11), 4362–4373 (2016)