SCIENCE CHINA
Information Sciences

• LETTER •
July 2020, Vol. 63 170209:1–170209:3
https://doi.org/10.1007/s11432-019-2751-4
Special Focus on Bio-Robotic Systems: Modeling, Design, Control and Perception

Data fusion using Bayesian theory and reinforcement learning method

Tongle ZHOU[1], Mou CHEN[1,3*], Chenguang YANG[2] & Zhiqiang NIE[3]

[1]*College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing* 211106*, China;*
[2]*School of Automation Science and Engineering, South China University of Technology, Guangzhou* 510641*, China;*
[3]*Science and Technology on Electro-optic Control Laboratory, Luoyang* 471000*, China*

Received 30 August 2019/Accepted 29 November 2019/Published online 30 April 2020

Citation Zhou T L, Chen M, Yang C G, et al. Data fusion using Bayesian theory and reinforcement learning method. Sci China Inf Sci, 2020, 63(7): 170209, https://doi.org/10.1007/s11432-019-2751-4

Dear editor,

In recent years, the theory and method of mission planning technology have been widely applied to bio-robotic systems. Increasing demands can be addressed by the application of innovative and advanced technology in mission planning and decision-making of bio-robotic systems, which are related to the complexity of mission environments and data fusion. The sizeable amount, geographical distribution, uncertainty, diversity and dynamics of target information make mission planning and decision-making of bio-robotic systems have a huge challenge. To solve this problem, multiple sensors have been used to fully support the information processing system [1]. However, because of differences in the performance and function of sensors, multi-source data fusion technology must be used to improve the accuracy of the detected information [2]. Therefore, the study of data fusion technology is of significance for bio-robotic systems.

Data fusion technology can use detected information to reflect actual situations and lay a foundation for decision-making in bio-robotic task planning systems. In general, the traditional data fusion algorithms face difficulties in processing online data, that are fused by an existing data set [2]. However, in a real scenario of task planning, sensor systems need to detect data in real time. Moreover, the large amount of mission information, complexity of the mission environment, and performance of the sensors may result in inaccurate data, which may lead to large errors in data fusion systems.

In such a case, a data fusion algorithm based on Bayesian theory and the reinforcement learning method is proposed to solve these problems. Reinforcement learning is suitable for solving learning and optimizing problems. Therefore, in this study, reinforcement learning and Bayesian theory are used to improve the fusion performance of data fusion systems.

Problem descriptions and preliminaries. This study aims to develop an active multi-sensor data fusion method. It is assumed that m sensors are used to simultaneously detect the same target and the time interval of each sensor may be different. Hence, the arrival time of the detected data is different.

The observed value of sensor i can be expressed as [3]

$$O_i = O_0 + \Delta O_i, \quad i = 1, 2, \ldots, m, \tag{1}$$

where O_0 is the actual value of the detected target attribute, ΔO_i is the uncertainty of sensor i, which is determined by the performance of sensor i.

The objective of this study is to design a data fusion algorithm based on Bayesian theory and reinforcement learning, so that the detected information $O_1, O_2, \ldots, O_m$ can be utilized to reflect the actual situation and to obtain the optimal fused value.

* Corresponding author (email: chenmou@nuaa.edu.cn)

To develop the data fusion algorithm, the following assumptions are required.

Assumption 1 ([2])**.** All sensors work independently and have no interference with each other.

Assumption 2 ([3])**.** The uncertainty ΔO_i $(i = 1, 2, \ldots, m)$ obeys the Gaussian distribution. To illustrate, $\Delta O_i \sim N(0, \sigma_i^2)$, where σ_i^2 is the variance of sensor i.

Data fusion algorithm. According to Assumption 2, we have

$$O_i \sim N(O_0, \sigma_i^2). \tag{2}$$

According to Bayesian theory, the traditional data fusion result can be expressed as [4]

$$\hat{O} = \sum_{i=1}^{m} \frac{O_i}{\sigma_i^2} \Big/ \sum_{i=1}^{m} \frac{1}{\sigma_i^2}. \tag{3}$$

The defect of the traditional fused result is that the influence of the faulty observations is ignored. In this study, this problem is formulated as a reinforcement learning task. The data fusion is denoted as a Markov decision process (MDP) $\{S, A, R, \gamma\}$, where S is the state set, A is the set of actions, R is the reward function, and $\gamma \in [0, 1)$ is the discount factor [5].

At time instant t, the data fusion system receives the state s_t and the reward r_t, and then it executes the action a_t while the sensor system receives the new detected data of the target. Then, the data fusion system produces the next state s_{t+1} and the reward r_{t+1}. The objective of reinforcement learning is to determine a policy $\pi : S \to A$ to maximize the expected reward across all episodes. In this study, the Q-learning algorithm is used to solve this problem. The Q-value function for the action $a \in A$ with the state $s \in S$ is defined as [5]

$$Q(s, a) = R(s, a) + \sum_{t=1}^{+\infty} \gamma^t R(s_t, a_t). \tag{4}$$

The optimal Q-function is expressed as the Bellman equation [5]:

$$Q^*(s, a) = R(s, a) + \gamma \max_{a \in A} Q^*(s', a'), \tag{5}$$

where s' is the state of the next time-step and a' is the selected action in state s'.

Then, the action at the time instant t can be obtained as

$$a_t = \max_{a \in A} Q^*(s_t, a). \tag{6}$$

Because of the difference in the time intervals of the sensors, the action set is designed as $A=$ {Retain, Delete}, where the next observation is retained or deleted based on the reward [6]. Additionally, we denote the current fused value as the state:

$$s_{t+1} = \begin{cases} \hat{O}_{t+1}, & a_{t+1} = \text{Retain}, \\ \hat{O}_t, & a_{t+1} = \text{Delete}. \end{cases} \tag{7}$$

Moreover, with the action set, the data set of each state is different. Therefore, the information entropy is used to evaluate the quality of data sets in different states. The information entropy $I(\Omega_t)$ in state t can be calculated as [7]

$$I(\Omega_t) = -\sum_{i=1}^{t} \left(O_i \Big/ \sum_{i=1}^{t} O_i \right) \ln \left(O_i \Big/ \sum_{i=1}^{t} O_i \right), \tag{8}$$

where Ω_t is the dataset of state s_t.

If the information $I(\Omega_{t+1})$ is smaller than $I(\Omega_t)$, then the newly state s_{t+1} is beneficial. Therefore, a positive reward should be provided. Otherwise, the reward should be negative. Hence, the reward is defined as

$$r_{s_t \to s_{t+1}} = \begin{cases} 1, & I(\Omega_{t+1}) \leqslant I(\Omega_t), \\ -1, & I(\Omega_{t+1}) > I(\Omega_t), \end{cases} \tag{9}$$

where Ω_{t+1} is the dataset of state s_{t+1}.

In this study, the data fusion system receives observations from multiple sensors. Because of the different sampling periods of the sensors, the reinforcement learning method is used to evaluate the quality of the new sampling data and to process the fault observations. Then, the data fusion system takes action (retain or delete the new sampling data) according to the reward (information entropy) until information detection has been completed. Finally, Bayesian theory is applied to the data fusion system based on the new data set without the inaccurate data. The whole algorithm is summarized in Algorithm 1.

Algorithm 1 Reinforcement learning based Bayesian data fusion algorithm

Require: The observations $O_1, O_2, \ldots, O_m$, the variances of sensors $\sigma_1, \sigma_2, \ldots, \sigma_m$.
Ensure: The fused data.
Initialize $Q = 0$, and set the discount factor γ;
for each episode **do**
  for $t = 1$ to m **do**
    while state s_t is not terminal **do**
      Initialize state s_t;
      $a' \leftarrow$ action in state s_t;
      Take action a', calculate reward r, and obtain the next available state s';
      Update Q according to (5);
      Calculate the fused data $\hat{O}_t$ based on the Bayesian theory (3);
      $s_t \leftarrow$ optimal new state s';
    end while
  end for
end for

Simulation results. To verify the effectiveness of the algorithm developed above, the simulating examples of the data fusion are provided.

In a situation where 10 sensors simultaneously detect the same target, the observations of each sensor O_i $(i = 1, 2, \ldots, 10)$ are 1.000, 0.990, 0.980, 0.970, 0.500, 0.650, 1.010, 1.020, 1.030 and 1.500, respectively. Furthermore, the sensor measurements σ_i $(i = 1, 2, \ldots, 10)$ are 0.05, 0.07, 0.10, 0.20, 0.30, 0.25, 0.10, 0.10, 0.20 and 0.30, respectively. The observations $O_5 = 0.500$, $O_6 = 0.650$ and $O_{10} = 1.500$ are obviously faulty.

The following methods are compared to demonstrate the efficiency and feasibility of the approach proposed in this study.

• RDFM (robust data fusion method). In this method, the optimal fused data of the sensors are obtained by a method based on statistical theory and the eigenvector theory [8].

• RDM (reliability degree-based method). The ellipse curve relation matrix is used to express the reliability degree and the fused data are obtained based on the reliability degree [9].

• TBM (traditional Bayesian method). In this method, the fused data are directly calculated by (3) [4].

• RLBM (reinforcement learning-based Bayesian method). The method that is developed used reinforcement learning for fault observations processing and Bayesian theory is used for data fusion.

The fused data of RDFM, RDM, TBM and RLBM are 0.9786, 0.9425, 0.9830 and 0.9989, respectively. Additionally, the errors of RDFM, RDM, TBM and RLBM are 0.0214, 0.0575, 0.0170 and 0.0011, respectively.

The error histogram is shown in Figure 1(a). Additionally, 10 sets of eligible samples are randomly generated, and the error curve is shown in Figure 1(b).

Obviously, the RLBM developed in this study has higher fusion accuracy. That is mainly because the influence of the fault observations is eliminated by the reinforcement learning process. Moreover, because the fault observations are gradually deleted, the proposed method can manage online data fusion situations. Furthermore, another benefit of this method is that time alignment is not a concern. The results show that the developed method is effective and feasible for multisensor data fusion.

Conclusion. To improve fusion accuracy, a data fusion method of bio-robotic systems using reinforcement learning was developed in this study. The proposed method could avoid the time alignment problem by fusing the observations step-by-step. Reinforcement learning was used to address inaccurate data and Bayesian theory is used for data fusion. The simulation results demonstrated the performance of the proposed algorithm. In future work, the decision-making problem will be considered based on the fused data for bio-robotic systems.

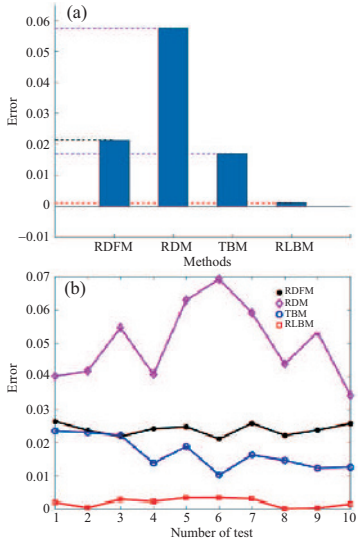

Figure 1 (Color online) (a) Error histogram; (b) the error curve of random samples.

Acknowledgements This work was supported by Major Projects for Science and Technology Innovation 2030 (Grant No. 2018AA0100800) and Equipment Pre-research Foundation of Laboratory (Grant No. 61425040104).

References

1 Du Y K, Jeon M. Data fusion of radar and image measurements for multi-object tracking via Kalman filtering. Inf Sci, 2014, 278: 641–652
2 Nguyen H, Cressie N, Braverman A. Spatial statistical data fusion for remote sensing applications. J Am Statist Assoc, 2012, 107: 1004–1018
3 Bass T. Intrusion detection systems and multisensor data fusion. Commun ACM, 2000, 43: 99–105
4 Wan S P. Method of fusion for multi-sensor data based on fisher information. Chin J Sensor Actuat, 2008, 21: 2035–2038
5 Li X X, Peng Z H, Liang L, et al. Policy iteration based Q-learning for linear nonzero-sum quadratic differential games. Sci China Inf Sci, 2019, 62: 052204
6 Zhang T, Huang M, Zhao L. Learning structured representation for text classification via reinforcement learning. In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence, 2018. 1–8
7 Yan X H, Zhu J H, Kuang M C, et al. Missile aerodynamic design using reinforcement learning and transfer learning. Sci China Inf Sci, 2018, 61: 119204
8 Wang H H, Wu Y, Fu Y, et al. Data fusion using empirical likelihood. Open J Statist, 2012, 2: 547–556
9 Wang W, Zhou J, Wang R. A method of the multisensor data fusion. J Transduc Technol, 2003, 22: 39–41