Scholar Publishing Group
Distributed Processing System
https://doi.org/10.38007/DPS.2021.020302
ISSN 2790-0916 Vol. 2, Issue 3: 12-21

Distributed System Optimization Analysis Based on Data Fusion and Data Transmission Method

Ransa Lisen*

University of Rochester, America

**corresponding author*

Keywords: Data Fusion, Data Transmission, Distributed System, Optimization Analysis

Abstract: With the continuous development of network technology and the release of various software products, people's lives are increasingly connected with the Internet. With the rapid increase in the number of users of software products, the processing requirements of servers are also increasing rapidly, and distributed systems begin to play an important role. The purpose of this paper is to analyze the optimization of distributed systems based on data fusion and transmission methods. First, it reveals the research and development background of the subject and its importance in scientific research, and studies the data acquisition technology based on RabbitMQ. Some new RabbitMQ implementation functions are proposed in the log collection service, and the RabbitMQ message processing method inside the model is analyzed and optimized. After comparative experiments in this paper, it is confirmed that the optimized RabbitMQ can support more log generators and thus have more efficient processing performance. Through experimental comparison and data analysis, it can be found that in this use case, when there is only one producer, the maximum delivery speed of the server before optimization can be reached faster, with a speed of 3800msg/s, while the optimized server with the maximum delivery speed is 3950msg/s, 2.3% higher than the highest delivery speed.

1. Introduction

Today, various new Internet services and applications are becoming more and more complex, such as large shopping platforms such as Taobao and Pinduoduo, mobile games, social networks, mobile video, etc. The common feature of these applications and services is that the number of users is huge and active, which makes network requests. The number has increased dramatically [1]. However, the development of distributed systems also brings up many difficulties and problems, such as operating systems, programming languages, database systems and other fields. A new phenomenon in communication between applications in distributed systems. Because message middleware can bridge the differences between computer networks and application platforms and realize fast and efficient transmission of distributed application information, message middleware has been widely used in distributed systems [2].

With the rapid development of modern information technology and communication technology, the control mode of distributed system is gradually transformed into distributed coordinated control under network communication. Samala R K proposed an efficient method to obtain better distributed generation (DG) location and capacity in radial distribution scheme (RDS). Here an efficient technique is obtained by hybrid Antlion Optimization (ALO) and the performance is improved with the help of the PSO function. This ALO technique typically mimics the antlion's catching behavior. The novelty of the proposed technique is to find the position of maximum power loss of RDS and place the best DG from a given PV (photovoltaic), WT (wind turbine) and diesel generator. First, the power flow survey of the RDS is to measure the power loss of the RDS. Due to load variation, the voltage profile, actual power loss and voltage stability index (VSI) are studied. Afterwards, confirm weak busbars based on VSI and power loss factors and fixed DG. After placing the DG, the VSI and power losses are reduced and the voltage distribution is maximized. The scheme of this plan was implemented with MATLAB software and validated on an IEEE 33 radial system. The effectiveness of the predicted method is determined and evaluated using independent techniques such as traditional ALO, the Gravity Search Algorithm (GSA), and the Bat algorithm [3]. Bashar A believes that when a large amount of data is transmitted, the number of collisions that occur is high. This is especially true when edge wireless sensor networks (WSNs) are densely distributed. Some of the major conflicts that affect the overall operation of the system include severe transmission delays and massive data loss. In this proposed work, multi-path reliable transmission is used for wireless sensor networks. To determine the reliability of the system, WSN implements redundancy methods. As a first step, the data is subdivided into packets and data redundancy. These packets are then multipathed to their corresponding destination nodes. Experimental observations show that the proposed work demonstrates a significant increase in network lifetime and a reduction in transmission delay and packet loss rate [4]. How to optimize the analysis of distributed systems is still a big problem.

The theoretical research of distributed coordinated optimal control has been widely valued by industry and academia, and has become a hot topic in current control theory research. Considering the widespread existence and practical application of distributed control system in industrial network system, this paper discusses how to design network communication, information update mechanism and coordinated optimal control strategy for distributed control system, so as to realize the control of each subsystem in complex distributed control system. parameter coordination and global performance optimization. This not only promotes the development of intelligent control theory, but also enriches the application of distributed optimization, intelligent computing and other theories, and promotes the interdisciplinarity, which has important engineering value.

2. Research on Distributed System Optimization Based on Data Fusion and Data Transmission

2.1. Distributed Multi-sensor Data Fusion

(1) Classical Kalman filter

Kalman filter is based on a series of data observed over time in a period of time, using a recursive data processing algorithm, by fusing the observed variables at the current moment with the predictors predicted by the optimal estimation at the previous moment. The iterative process is used to obtain the optimal estimation of the system state; compared with using a single state observation variable, a more accurate system state estimation in the sense of minimum mean square error can be obtained [5-6].

(2) Distributed Kalman Filter

In order to realize the distributed Kalman filter, researchers have made a lot of theoretical and practical attempts. The main idea is to use the data communication and information exchange between node devices to make multiple node devices work together, and each node device shares a small amount of computation [7-8].

(3) Information filter

When M sensor node devices are installed in the measurement system, and the observed variables obtained by each sensor node are not correlated with each other, the information filter that transmits the system information form matrix is obtained, and the information form state estimation and covariance matrix are obtained. The information that is independent of other nodes is extracted from each observation of a node, and all the information required for the one-step measurement update process can be obtained by accumulating the observation data of each sensor, eliminating the influence of cross-correlation [9-10].

2.2. Features of Distributed Systems

(1) Distribution

Multiple computers in a distributed system will be randomly distributed in space, and the distribution of machines will change at any time [11-12].

(2) Equivalence

The distributed system is not divided into master and slave, all parts of the distributed system are equal. Replication is one of the most common concepts of distributed systems and refers to the unique way a distributed system provides data and services. In distributed systems, we often change data and services in order to provide the most available services to the outside world. Data replication refers to maintaining the same data on different nodes. When data stored on a node is lost, data can be read from the replica. Another feature is service redundancy, which means that several nodes provide the same service, and each node has the ability to receive external requests and process them accordingly [13-14].

(3) Concurrency

Concurrency during program execution is a very common behavior in computer networks. For example, multiple nodes in the same distributed system can simultaneously work on some shared resource, such as a database or shared storage [15-16].

2.3. Problems Existing in Distributed Coordinated Optimal Control

(1) Feasibility and stability

In the coordinated optimization and MPC research of distributed nonlinear control system, feasibility and stability problems are still a major difficulty. Many existing distributed MPC research results are aimed at linear distributed control systems, while for nonlinear subsystems, it is generally necessary to consider not only the interconnection between subsystems, but also the control effect of distributed coordination. More complicated [17-18].

(2) Nonlinear physical coupling

Many existing achievements cannot easily deal with the physical coupling of the controlled objects, especially the nonlinear physical coupling objects. The common distributed predictive control method can coordinate the dynamic information of different individuals through the exchange of individual information in the system, so as to achieve the purpose of cooperative optimal control, but for the objects of physical coupling, it is necessary to estimate the change of the coupling state in order to carry out large-scale control. Scale system distributed MPC design.

(3) Update and control cycle

According to the existing research results of distributed coordination optimization and predictive control, most of them assume that the information update and control cycles of all subsystems are completely synchronized, which is difficult to achieve in practical engineering. Although the existing researches on distributed coordinated optimization and predictive control have adopted different optimization problems and information exchange strategies according to different problems, most of the researched distributed control systems are based on similar assumptions: each subsystem has a similar time constant; Although the dynamic characteristics of each subsystem are different, they all belong to the same type of system.

3. Experiment and Research on Distributed System Optimization Based on Data Fusion and Data Transmission

3.1. System Model

Description Consider a multi-agent system consisting of N linear subsystems, each agent Ai can be represented as a subsystem Ai, regarded as a tracker, and is described by the following dynamic equations.

$$\dot{x}_i(t) = A_i x_i(t) + B_i u_i(t) + \Delta_i(t), x_i(0) = x_{i0}, t \geq 0 \quad (1)$$

$$y_i(t) = C_i x_i(t) + D_i u_i(t), i = 1, \ldots, N \quad (2)$$

Among them, x∈Rni, ui∈Rmi, and yi(t)∈RPi represent the state, input and output of the distributed subsystem Ai, respectively.

3.2. Experimental Parameters

We conduct experimental evaluations through 5 metrics: resource usage (CPU and RAM), execution time (the time it takes for various algorithms to process a certain number of requests), latency (the time a task waits in a waiting queue before being processed), Stability and query performance. Finally, the algorithm proposed in this paper compares iterative time-consuming with the current popular deep learning methods. Different server nodes are simulated by a separate thread, and the communication between servers is inter-thread communication. Since the game model proposed in this paper has nothing to do with network parameters, the simulation method of the network does not affect the experimental results.

3.3. Neighbor Node Screening Based on Remaining Available Delay

The end-to-end delay set by the user means that the total data transmission delay from the source node to the destination node should be less than or equal to the threshold. According to this threshold, we can update a corresponding remaining available delay at each hop in the transmission process, which indicates the maximum allowable delay for transmitting data from the current sending node to the destination node. The definition of the remaining available delay is shown in formula (3):

$$t_{RAD} = t_{TD} - t_{UD} \tag{3}$$

tTD is the end-to-end delay set by the user, and tUD indicates the used delay of the data packet, that is, how long it has been since the source node started sending the data packet.

Using these shortest paths as a reference, we can estimate the minimum delay required for the rest of the transmission process, as shown in equations (4) and (5):

$$t_{MD}^{i} = (t_{DIFS} + t_{backoff} + \frac{S}{B} + t_{SIFS} + \frac{S_{ack}}{B} + t_{r\_ack}) + \frac{d}{v} \tag{4}$$

$$t_{MD} = \sum_{i} t_{MD}^{i} \tag{5}$$

Among them, t_{MD}^{i} represents the minimum delay required by the i-th link in the case of successful transmission, and tMD represents that the current sending node uses a neighbor node as a relay to deliver the data packet to the destination node through multiple hops according to its shortest path. Minimum delay required.

4. Analysis and Research of Distributed System Optimization Based on Data Fusion and Data Transmission Method

4.1. Processing Performance was Tested

This paper firstly tests the processing performance of RabbitMQ before optimization. Table 1 is the experimental data, and there are eight attributes in the table. The first item is the serial number of the test, confirm represents whether the confirm mechanism is used, ack represents whether the acknowledged mechanism is used, publisher represents the number of producers, consumer represents the number of consumers, messagesize represents the size of the message used in the test, and publishrates represents the producer The rate of messages sent to the RabbitMQ server, in msg/s (the number of messages per second), and delivers represent the number of messages delivered by RabbitMQ to the consumer, in msg/s (the number of messages per second).

Table 1. Experimental data before RabbitMQ optimization

Serial number	Publisher	Consumer	Publish rates	Deliver rates
1	1	1	6345	6345
2	2	1	6412	3245
3	3	1	7035	2365
4	4	1	7564	1789
5	5	1	7984	1246
6	1	1	4567	4568
7	3	2	5468	2647
8	2	1	6354	6358
9	1	1	7589	4879
10	2	2	4987	7968

From the test data in Table 1, it can be seen that with the increase of the number of producers, the delivery rate of the server decreases, resulting in message accumulation. It is proved by test 3 and test 7 that increasing the number of queue consumers does not improve the processing performance of the server. The comparison of test 2 and test 5 found that the size of the message also affects the processing rate of the RabbitMQ server, and the processing speed decreases when the message is large. After analyzing the test data in Table 1, this section conducts experiments on the performance of the optimized RabbitMQ server. The experimental data is shown in Table 2.

Table 2. Experimental data after RabbitMQ optimization

Serial number	Publisher	Consumer	Publish rates	Deliver rates
1	1	1	6230	6345
2	2	1	6340	6302
3	3	1	6387	4689
4	4	1	6468	3324
5	5	1	6789	2896
6	1	1	4984	4800
7	3	2	5236	5136
8	2	1	5046	5046
9	1	1	4879	4879
10	2	2	4968	3645

The data in Table 2 proves that when the number of producers increases, it will lead to the accumulation of messages and reduce the processing performance of the server. The use of the confirm and acknowledged mechanisms will also reduce the processing performance of the server. When the confirm mechanism is used, the delivery rate of the server is reduced by 21.3%, the delivery rate of the server using the acknowledged mechanism is reduced by 36.5%, and the server performance is reduced by 41.5% when the confirm and acknowledged mechanisms are used at the same time. The test results of Table 2 are consistent with the test results of Table 1.

4.2. Comparative Analysis

At the same time, this paper conducts a comparative analysis of the data, as well as the comparative analysis of the server before and after optimization.

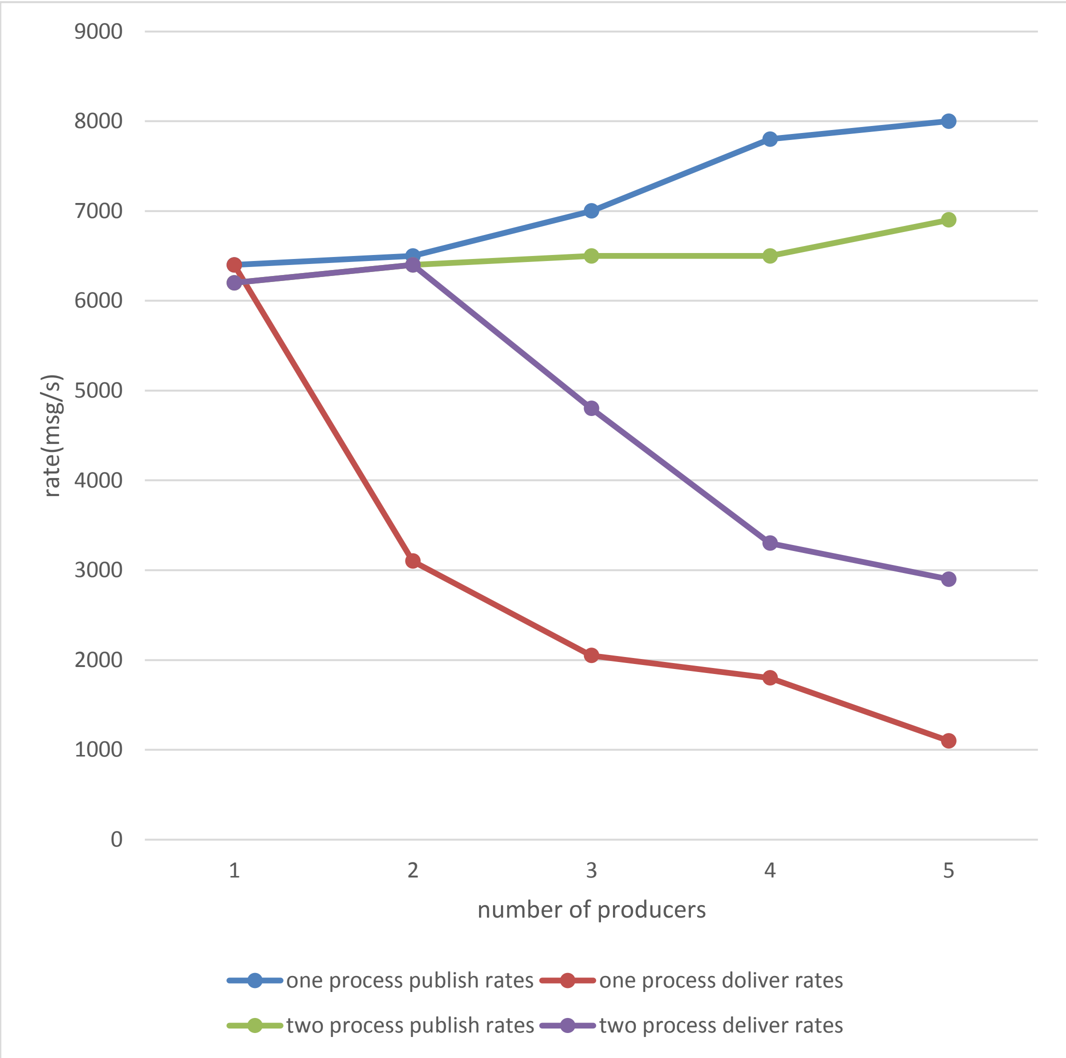

Figure 1. Comparative analysis before and after optimization1

The data in Figure 1 was collected without using either the confirm mechanism nor the acknowledged mechanism. Through the comparative analysis of the experimental data, it can be seen that in this usage scenario, the delivery rate of the server before optimization reaches the highest when there is only one producer, which is 6400msg/s. With the increase of the number of producers, the delivery rate of the server decreases very much. Obviously, the production rate continues to increase, the accumulation of messages increases, and the performance of the server processing drops very quickly. After optimization, the maximum delivery rate of the server is 6300msg/s, and the maximum delivery rate is reduced by 0.2%. However, when there are two producers, messages can still be processed normally without message accumulation. As the number of producers increases, the delivery rate decreases, but Both are higher than the delivery rate before optimization, and the message accumulation rate is slower.

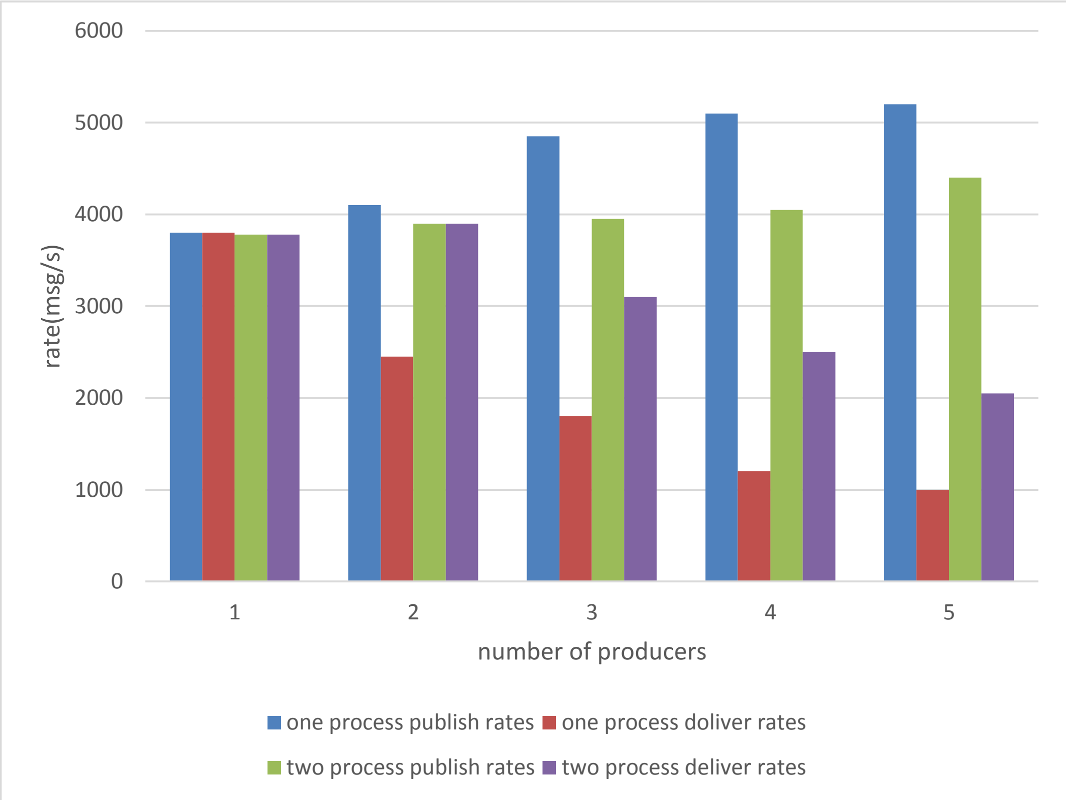

Figure 2. Comparative analysis before and after optimization2

The data in Figure 2 is collected using the confirm mechanism and using the acknowledged mechanism. Through the comparative analysis of experimental data, it can be seen that under this usage scenario, the delivery rate of the server before optimization reaches the highest when there is only one producer, which is 3800msg/s. With the increase of the number of producers, the delivery rate of the server decreases very much. Obviously, the production rate continues to increase, the accumulation of messages increases, and the performance of the server processing drops very quickly. After optimization, the maximum delivery rate of the server is 3950msg/s, and the maximum delivery rate is increased by 2.3%. With the increase of producers, the delivery rate decreases, but it is higher than the delivery rate before optimization, and the accumulation rate of messages is slow.

5.Conclusion

At present, the operation optimization of distributed multi-energy systems is mainly carried out through numerical analysis or limited experimental research. If the research is carried out through pure mathematical simulation, some simulation models are not accurate enough or difficult to establish. If experimental research is carried out through pure physical models, the research and development cycle is long and the research and development cost is high. Therefore, the hardware-in-the-loop simulation technology, which combines organic mathematical models and physical models (or objects), has become an effective means of conducting research on distributed multi-energy systems. On the basis of the large-scale utilization of distributed renewable energy, the physical interconnection and energy interaction of the regional energy network can be realized through ER or EH, and further research on the system planning and operation optimization of the energy Internet is conducive to the creation of resource-saving and sustainable development. The topic selection of the thesis not only has the practical application value of the research content, but also has the innovation of the method and technology system.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Stepin Y P, Leonov D G, Papilina T M, et al. System modeling, risks evaluation and optimization of a distributed computer system. Computer Research and Modeling, 2020, 12(6):1349-1359.

[2] Caballero-Aguila R, Hermoso-Carazo A, Linares-Perez J. Networked distributed fusion estimation under uncertain outputs with random transmission delays, packet losses and multi-packet processing. Signal processing, 2019, 156(MAR.):71-83. https://doi.org/10.1016/j.sigpro.2018.10.012

[3] Samala R K, Kotapuri M R. Distributed Generation Allocation in Distribution System using Particle Swarm Optimization based Ant-Lion Optimization. International Journal of Control and Automation, 2020, 13(1):414-426.

[4] Bashar A, Kumar D. Reliable Data Transmission with Sub-Packet Generation and Fusion in Wireless Sensor Networks. IRO Journal on Sustainable Wireless Systems, 2021, 2(4):160-164. https://doi.org/10.36548/jsws.2020.4.004

[5] Kumar R, Kaushik B K, Balasubramanian R. Multispectral Transmission Map Fusion Method and Architecture for Image Dehazing. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2019, PP(99):1-5.

[6] Nawaz M A, Raheem A, Shakoor R, et al. feasibility and optimization of standalone pv-biogas hybrid distributed renewable system for rural electrification: a case study of a cholistan community. Mehran University Research Journal of Engineering and Technology, 2019, 38(2):453-462. https://doi.org/10.22581/muet1982.1902.19

[7] Tabasi M, Asgharian P. Optimal operation of energy storage units in distributed system using social spider optimization algorithm. AIMS Electronics and Electrical Engineering, 2019, 3(4):309-327. https://doi.org/10.3934/ElectrEng.2019.4.309

[8] Sameti M, Haghighat F. Optimization of 4th generation distributed district heating system: Design and planning of combined heat and power. Renewable Energy, 2018, 130(JAN.):371-387. https://doi.org/10.1016/j.renene.2018.06.068

[9] Hsu C C, Lin C H, Kao C H, et al. DCSN: Deep Compressed Sensing Network for Efficient Hyperspectral Data Transmission of Miniaturized Satellite. IEEE Transactions on Geoscience and Remote Sensing, 2020, PP(99):1-17.
[10] Yahav I, Sheffi N, Biofcic Y, et al. Multi-Gigabit Spatial-Division Multiplexing Transmission Over Multicore Plastic Optical Fiber. Journal of Lightwave Technology, 2021, 39(8):2296-2304.
[11] Alsafasfeh M, Arida Z A, Saraereh O A, et al. An Optimized Data Fusion Paradigm for WSN Based on Neural Networks. Computers, Materials and Continua, 2021, 69(1):1097-1108. https://doi.org/10.32604/cmc.2021.018187
[12] Iqbal N, Zerguine A, Khan S. OFDMA-TDMA based Seismic Data Transmission Over TV White Space. IEEE Communications Letters, 2021, PP(99):1-1.
[13] Timofeev G A, Potrakhov N N, Gryaznov A Y. Data Transmission in the X-Ray Emission Frequency Range of Electromagnetic Radiation. Journal of the Russian Universities Radioelectronics, 2021, 24(2):6-17.
[14] Yamanaka T, Iwai T, Kubo R. Quality of Performance Aware Data Transmission for Energy-Efficient Networked Control. IEEE Access, 2021, PP(99):1-1. https://doi.org/10.1109/ACCESS.2020.3048796
[15] Selvaraj S, Dhamodharavadhani S, Rathipriya R. Reduce Data Transmission Energy in Wireless Body Area Network using LRNN Prediction Model. International Journal of Future Generation Communication and Networking, 2021, 14(1):1039-1053.
[16] Abdolkarimi E S, Mosavi M R, Rafatnia S, et al. A Hybrid Data Fusion Approach to AI-assisted Indirect Centralized Integrated SINS/GNSS Navigation System during GNSS Outage. IEEE Access, 2021, PP(99):1-1. https://doi.org/10.1109/ACCESS.2021.3096422
[17] Gomathi B, Sujatha R. Prediction of Breast Cancer using Data Fusion of SVM with Optimization Technique. Journal of Information and Computational Science, 2021, 9(12):504-511.
[18] Habbouche H, Benkedjouh T, Amirat Y, et al. Gearbox Failure Diagnosis Using a Multisensor Data-Fusion Machine-Learning-Based Approach. Entropy, 2021, 23(697):1-20. https://doi.org/10.3390/e23060697