Research and Application of Improved Probabilistic Neural Network Algorithm in Dynamics of Flexible Job-shop under the Situation of Arrival of New Workpiece

Xu Liang1,2
1.Computer School
Beijing Information
Science&technology University
Beijing, China
2656960183@qq.com
2.Software Technology Institute
Dalian Jiaotong University
Dalian, China
2656960183@qq.com

Yiming Huang
Software Technology Institute
Dalian Jiaotong University
Dalian, China
1164104505@qq.com

Ming Huang
(Corresponding author)
Software Technology Institute
Dalian Jiaotong University
Dalian, China
13604118232@163.com

Abstract—in this paper, an improved probabilistic neural network algorithm is designed for the problems such as failure to respond quickly to interference, resulted long waiting time and the extension of idle time of the machine existing in the dynamic dispatch of flexible job-shop under the situation of arrival of new workpiece. The algorithm uses the main component analysis method to reduce the dimension of high-dimensional feature data; uses the improved sparrow algorithm to optimize the smoothing factor, and completes the training and prediction of the model based on the simulation sample. Improved model can reduce the effect of noise and data redundancy in the sample, and improve the self-adaptiveness and certain immunity to disturbance. The experimental results show that the algorithm has significantly improved the classification ability of job-shop scheduling under the interference of the arrival of new workpieces.

Keywords—probabilistic neural network algorithm; arrival of new workpiece; flexible job-shop

I.Introduction (Heading 1)

Flexible Job-shop Scheduling Problem (FJSP) is widely used by enterprises because of its production flexibility, the sufficient utilization of resources and other characteristics that can effectively reduce production costs, bring high added value and competitiveness, and has also attracted the attention of scholars[1].

Zuo Le [2]uses the re-scheduling mixed drive mechanism based on shift coefficient to dig deep into the nature of multi-target dynamic scheduling, integrates the cost of handling and delay penalty into the optimization target, and efficiently selects an excellent rescheduling scheme. Adopting [3]algorithm and comparison between order cancellation and postpone insertion of order improved by NEH and exchange strategy, Pei Xiaobing[3] et al solve the problem of emergency order insertion under certain conditions. Zambrano et al [4] delivery delay in this complex situation, taking into account both machine failure and the immediate arrival of new workpieces. Li [5] et al have established a feasible mathematical model based on rescheduling strategy, and solved the FJSSP model under this case by a hybrid artificial bee colony algorithm (HABC) in conjunction with taboo search (TS).

The Probabilistic Neural Network algorithm (abbreviated as PNN) is known for its simple structure, few parameters and high stability, but the choice of unreasonable parameters can also have a significant impact on the performance of the algorithm. Porwik et al [7] also use PSO algorithm to solve optimal smoothing factors, and comparison with PNN and BP network shows that the stability of the improved PSO-PNN algorithm has been improved significantly. Liu et al [12] proposed a self-adaptive strategy to improve the smooth parameters of probabilistic neural network, and the simulation results show that these improvements are superior to traditional PNN network, but their efficiency decreases with the increase of self-adaptive range.

Scholars at home and abroad have carried out research of flexible workshop scheduling problem regarding arrival of new workpiece from different angles and obtained certain research results, but most of the current research results are based on a specific interference situation, there are few research for uncertainty of perturbation degree of the perturbation factors, it is impossible to respond quickly to interference, resulting in the machine waiting time, idle time extension, and waste of resources. In this paper, we study the rescheduling model in depth, combine the three dynamic scheduling performance indicators, comprehensively consider the efficiency and stability of scheduling, combine the actual job-shop data, get a large volume of labeled data through simulation, and provide a new method for flexible job-shop to respond quickly to interference caused by arrival of new workpiece through PNN learning.

II.Flexible job-shop dynamic scheduling method under condition of arrival of new workpiece on the basis of improved probabilistic neural network algorithm

A.Improved PNN accuracy

Enter feature dimension reduction. Unprocessed training samples will contain related noise interference, failure to perform standardized processing will have a great impact on the classification accuracy of probabilistic neural network[9], and the high-dimensional sample vector will make the probabilistic neural network structure complex, reduce the computational speed of the probabilistic neural network, and greatly increase the difficulty of hardware realization. This is because when training the probabilistic neural network, the probabilistic neural network sets the training sample as hidden layer neuron, and a large number of hidden layer neurons form the vector of the hidden center of the probabilistic neural network, and the training sample directly affects the structure and operation of the probabilistic neural network. Similar to training samples, unprocessed test samples also contain noise interference and high dimensional problems. Training and test sample vectors can also contain a lot of redundant information. These subjective unprocessed training and test samples need processing optimization to further improve the computational speed and accuracy of the probabilistic neural network.

Optimize smoothing factor values. The traditional smoothing factor setting process is subject to artificial subjective intervention, and in most cases, selecting a specific number of smoothing factor values according to experience can only be approximately reflected in the change of probabilistic neural network classification accuracy under different smoothing factors, and cannot fully reflect the overall change trend of probabilistic neural network classification accuracy when the smoothing factor changes. Selecting different smoothing factors will directly affect the probabilistic neural network classification accuracy regarding the same training sample and the test sample under the same target vector[12], and the smoothing factor can be selected by adopting optimization algorithm.

B.Data processing and dimension reduction analysis for optimization of PNN

Perform high-dimensional feature data dimension reduction using Principal Component Analysis (PCA). The PCA selects K-unit orthogonal base to map a set of N-dimensional vectors to this set of bases, reducing K-dimensional vectors (N>>K), in which it is necessary to ensure that the covariance between the variables is 0, and the variance between variables is as large as possible. The model is as follows: assume that the jth indicator of the ith sample in the mth sample set and in the nth evaluation indicators takes value of xij, and construct the original evaluation indicator data matrix:

$$X=(x_{ij})_{m\times n} \tag{1}$$

$$X=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \tag{2}$$

In which, the original variables: x1,x2,...,xn, dimension-reduced composite indicator (main component): y1,y2,...,yn, n indicator vectors of Xm*n are linearly combined, the formula is as follows:

$$\begin{cases} y_1=l_{11}x_1+l_{21}x_2+\cdots+l_{n1}x_n \\ y_2=l_{12}x_1+l_{22}x_2+\cdots+l_{n2}x_n \\ \vdots \\ y_n=l_{1n}x_1+l_{2n}x_2+\cdots+l_{nn}x_n \end{cases} \tag{3}$$

Formula (2) and (3) should meet:

(1) The quadratic sum of the coefficients of each equation is equal to 1, as shown in formula (4);

(2) Any two main components obtained are independent of each other and not related to each other;

(3) y1 is the one with the maximum variance in all linear combinations of x1,x2,...,xn, and yn is the one with the maximum variance in all linear combinations of x1,x2,...,xn not related to y1, y2, ···, yn-1

$$l^2_{1j}+l^2_{2j}+\cdots+l^2_{nj}=1 \quad (j=1,2,\cdots,n) \tag{4}$$

Here are the steps to solve the main component:

Step 1: standardize the raw data.

$$\dot{x}_{ij}=(x_{ij}-\overline{x_j})/s_j \quad (i=1,2,\cdots,m;j=1,2,\cdots,n) \tag{5}$$

$$\overline{x_j}=\sum_{i=1}^{m}x_{ij}/m \tag{6}$$

$$s_j=\sqrt{\sum_{i=1}^{m}(x_{ij}-\overline{x_j})^2/(m-1)} \tag{7}$$

In which, and sj are the mean and standard deviation of the jth indicator, respectively.

Step 2: calculate the Pearson correlation coefficient matrix between indicators, i.e.

$$R=(r_{kl})_{n\times n} \quad (k,l=1,2,\cdots,n) \tag{8}$$

In which rkl is the correlation coefficient between the kth indicator and the lth indicator, and rkl=rlk (i.e. symmetric matrix), the calculation formula is:

$$r_{kl}=\sum_{i=1}^{m}(x_{ik}-\overline{x_k})(x_{il}-\overline{x_l})\Big/\sqrt{\sum_{i=1}^{m}(x_{ik}-\overline{x_k})^2\sum_{i=1}^{m}(x_{il}-\overline{x_l})^2} \tag{9}$$

Step 3: calculate the characteristic value and characteristic vector of the relevant matrix R. The characteristic value is recorded as λ1, λ2, ..., λn and satisfies λi >= 0 (i=1⊠2⊠…⊠n), and the unitized characteristic vector corresponding to the characteristic value is recorded as p1,p2,..., pn.

Step 4: determine the number of main components. Calculate the cumulative contribution rate of the main component, generally take first kth main components of which the characteristic value is greater than 1 and the cumulative contribution rate reaches 85%~95%.

$$v_s=\lambda_s\Big/\sum_{s=1}^{n}\lambda_s \quad (s=1,2,\cdots,n) \tag{10}$$

In which, vs is the variance contribution rate of the sth main component.

$$v_{sumk}=\sum_{s=1}^{k}\lambda_s\Big/\sum_{s=1}^{n}\lambda_s \quad (k=1,2,\cdots,n) \tag{11}$$

In which, vsumk is the cumulative contribution rate of the first kth main component.

Step 5: calculate the corresponding score for extracting the main components. The main component coefficient matrix is : U-(p1,p2,...,pn), and if the first kth main components are extracted from the original indicator, then:

$$y_s=X^{*}p_s=(x_1^{*},x_2^{*},\cdots,x_n^{*})p_s \quad (s=1,2,\cdots,k) \tag{12}$$

In this paper, the PCA method is used to extract key predictive parameter variables. In the case of a large data set (5000), the matlab software provides the contribution rate and accumulative contribution rate as shown in Figure 1 below, contribution rate reflects the weighting information of the main components, and accumulative contribution rate determines the parameter variables that need to be extracted. From the figure, we can know that the cumulative contribution rate of the first three main components reaches 90%, this is sufficient to support the calculation. The two-dimensional figure after dimension reduction is Figure 2:

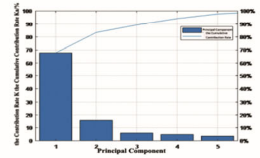

Fig. 1 Principal component contribution rate

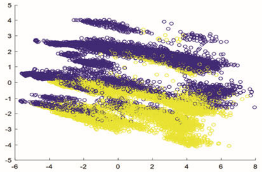

Fig. 2 Two dimensional sample distribution

C.Predictive analysis based on the rescheduling method of PCA-tSSA-PNN model

Sparrow search algorithm can be updated in two ways approximately: approach to the current optimal position and approach to the origin. A large number of simulation experiments regarding the data of this paper show that, when solving as per optimal rescheduling method, each convergence directly jumps to the adjacency of current optimal solution, it is easy to lose the global optimal method solution. In order to improve the global search ability and prevent being trapped in local optimal solution, this paper uses adaptive t distribution strategy to improve sparrow search algorithm.

The specific algorithm flow of this paper for improving algorithm PCA-tSSA-PNN is as follows:

Step 1: initialize population parameter, gen, initialize predator and intrant proportion.

Step 2: initialize the sparrow population as an alternative smoothing factor, build PNN network, calculate the number of correct classifications and accuracy, calculate the initial fitness value, and sort.

Step 3: sparrow algorithm updates the predator position.

Step 4: sparrow algorithm updates the intrant location.

Step 5: sparrow algorithm updates the alerter position.

Step 6: calculate the fitness value and update the sparrow position.

Step7: if rand < p, perform self-adaptive t distribution variation.

Step 8: calculate the current fitness value and update the sparrow position.

Step 9: whether the maximum number of iterations is reached or the error condition is met, exit and output the results if yes, otherwise repeat steps 2-8.

Step 10: the individual with the best output fitness value is taken into PNN network as a result of the smoothing factor to obtain the final identification model.

After running PCA-tSSA-PNN and PCA-SSA-PNN algorithms respectively for 10 times, the comparison between classification accuracy and running time is shown in Figures 3～4 below:

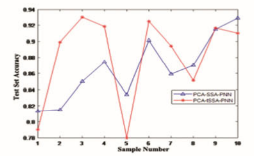

Fig. 3 Accuracy comparison chart

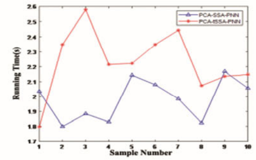

Fig. 4 Time comparison chart

To sum up, the following conclusions can be obtained from the analysis of the results in the figure:

The accuracy of the model is further improved by optimizing PNN network parameters using tSSA algorithm. This is because tSSA seeks optimal, avoiding the randomness of smoothing factors when they rely on empirical values, resulting in better recognition rate.

PCA-tSSA-PNN-based rescheduling decision-making model is better than PCA-SSA-PNN in terms of running speed and classification accuracy, so the method proposed in this paper is valid for the rescheduling problem in this paper.

References

[1]Gu Heping. Research of robust rescheduling problem of discrete manufacturing enterprises under machine failure [D]. Chongqing Polytechnic University, 2018.

[2]Zuo Le. Multi-objective dynamic scheduling research of flexible job-shop under uncertain environment [D]. Beijing Jiaotong University, 2015

[3]Pei Xiaobing and Yang Jingxia. Fruit fly optimization algorithm for solving emergency order insertion problem [J]. Systems Engineering, 2020, 38 (06): 139-146.

[4]Gabriel Zambrano Rey,Abdelghani Bekrar,Vittaldas Prabhu,et al. Coupling a genetic algorithm with the distributed arrival-time control for the JIT dynamic scheduling of flexible job-shops[J]. International Journal of Production Research,2014,52(12).

[5]Li X, Peng Z, Du B, et al. Hybrid artificial bee colony algorithm with a rescheduling strategy for solving flexible job shop scheduling problems[J]. Computers & Industrial Engineering, 2017, 113: 10-26.

[6]Gao K Z, Suganthan P N, Tasgetiren M F, et al. Effective ensembles of heuristics for scheduling flexible job shop problem with new job insertion[J]. Computers & Industrial Engineering, 2015, 90: 107-117.

[7]Porwik P, Doroz R, Orczyk T. Signatures verification based on PNN classifier optimized by PSO algorithm[J]. Pattern Recognition, 2016, 60: 998-1014.

[8]Bojun Liu,Yushun Fan,Yi Liu. A fast estimation of distribution algorithm for dynamic fuzzy flexible job-shop scheduling problem[J]. Computers & Industrial Engineering,2015,87.

[9]Li Weidong, Guo Rui and Zhang Lei et al. Multi-target large-scale cargo transport channel selection based on PCA improvement hierarchical analysis method [J]. China Safe Production Science and Technology, 2021, 17 (02): 135-139.

[10]Guo Xiaoyuan, Li Haiming and Xu Yunjie. Transformer fault diagnosis research based on improvement of fruit fly algorithm optimization PNN [J]. Journal of Shanghai Electric Power University, 2020, 36 (04): 395-400.

[11] Ban Dongdong. Data-driven mine ventilator bearing fault diagnosis research [D]. Xi'an University of Science and Technology, 2020.

[12]Bojun Liu,Yushun Fan,Yi Liu. A fast estimation of distribution algorithm for dynamic fuzzy flexible job-shop scheduling problem[J]. Computers & Industrial Engineering,2015,87.

[13] Liu Jinghao, Sun Xiaowei and Jin Jie. Intrusion detection model based on main component analysis and circulating neural network [J]. Journal of Chinese Information, 2020, 34 (10): 105-112

[14] Xue Jiankai. Research and application of a new type of cluster intelligence optimization technology [D]. Donghua University, 2020.