Autoencoder classification algorithm based on swam intelligence optimization

Wen Feng ,Zhiguo Chen
School of IoT Engineering
Jiangnan University
Wuxi, China
427533@qq.com

Yi Fu
Wu Xi Research Center of Environmental Science and Engineering
Wuxi, China
miss_fuyi@yahoo.com.cn

***Abstract*—BP algorithm used by autoencoder classification algorithm. But the BP algorithm is not only complicated and inefficient, but sometimes falls into local optimum. This makes autoencoder classification algorithm are not very good .So in this paper we combie Quantum Particle Swarm Optimization (QPSO) and autoencoder classification algorithm. QPSO used to optimize the weight of autoencoder neural network and the parameter of softmax. This method has been tested on some database, and the experimental result shows that this method has got good results.**

Keywords-autoencoder neural network, classification, Quantum Particle Swarm Optimization (QPSO)

I. INTRODUCTION *(HEADING 1)*

In recent years, neural networks have become the hotspot for machine learning. As a simple and efficient network, autoencoder neural network can effectively extract effective information from data for classification. Autoencoder obtains the parameters that best represent the characteristics of the original data by encoder and decoder. Many scholars have improved and applied it, Vincent et al. proposed denoising autoencoder [3], Cambridge University O. Chen et al. also proposed the convolution self-encoding [4]. autoencoder network as an unsupervised learning model is widely used in computer science and related fields.

Particle Swarm Optimization [5] has been proposed by Kennedy and Eberhart in 1995, since then it has attracted the attention and research of many scholars in related fields at home and abroad[1,7,8,9], because of its simple calculation, easy implementation, and few control parameters. But PSO is easy to fall into local optimum and this algorithm's velocity and position evolution formula make the particle swarm's randomness and intelligence low; in addition, the dependence of the algorithm performance on the max speed limit makes it less robust. Aiming to these shortages Sun[6] has proposed Quantum-behaved Particle Swarm Optimization (QPSO).QPSO has global convergence, fewer control parameters, faster convergence, and better searchability it has been widely used in computers and related fields.

Dong[1] has combine autoencoder with PSO and tested on Email classification. XU [2] optimized autoencoder by Extreme Learning Machine (ELM) and tested on some database of UCI. In [1], we find that the recall of spam in the experimental results of Email classification is not good. In her method, the parameters of autoencoder network are optimized by PSO, and the parameters of softmax by BP. We know that PSO and BP are easy to fall into local optimum and inefficient. The reason for this phenomenon may be due to the shortages of PSO and BP. So we replace PSO and BP with QPSO.

II. AUTOENCODER AND QPSO

A. Autoencoder network

Autoencoder neural network is an unsupervised model, it makes the output and input as equal as possible. One Autoencoder neural network module has three layers including input layer, hidden layer (encoder layer) and output layer (decoder layer). The input data is mapped to the hidden layer by formula (1).

$$h = f\left(W_1 * x + b_1\right) \tag{1}$$

W_1 is the weight from the input layer to the output layer.

x is the input data. b_1 is the bias. f is the sigmoid function.

The input of decoder layer:

$$Y = f(W_2 * h + b2) \tag{2}$$

W_2 is the weight from the encoder layer to the decoder layer and $W_2 = W_1{}'$. For every sample the cost function of encode-decode is defined as follow:

$$C(X,Y) = \frac{1}{2} * \left\|X - Y\right\|^2 \tag{3}$$

So the cost function of the whole process：

$$SUM(X,Y) = \sum_{i=1}^{t} \frac{1}{2} * \left\|X_t - Y_t\right\|^2 \tag{4}$$

The t is the number of samples. Our target is to make the $SUM(X,Y)$ as minimum as possible. Figure 1 is the structure of autoencoder network .

In autoencoder network decoder layer is to reconstruct the input data. The smaller the $SUM(X,Y)$, the better the reconstruction. So in autoencoder network the focus is on encoder.

978-1-5386-7445-1/18/$31.00 ©2018 IEEE
DOI 10.1109/DCABES.2018.00068

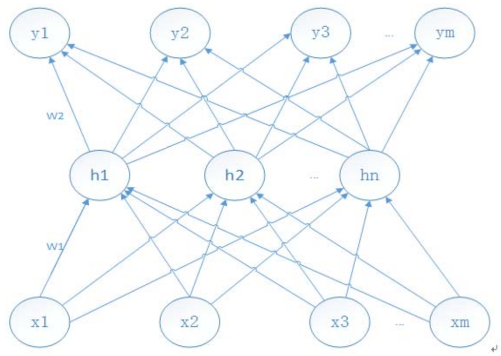

Figure 1. The structure Autoencoder neural network

B. Quantum Particle Swarm Optimization (QPSO)

QPSO is based on the classical particle swarm optimization algorithm. It mainly combines the idea of quantum physics to modify the "evolution" of PSO. The method focuses on the current local optimal position information and global optimal position information of each particle when updating the particle position, and the specific operation process is defined as follows.

$$X_t = P \pm 0.5 * L * \ln(\frac{1}{u})$$

(5)

u is a random number from 0 to 1. L is defined as follows

$$L_{t+1} = 2 * \beta \left| mbest - X_{\mathrm{t}} \right|$$

(6)

$$mbest = \sum_{i=1}^{M} \frac{P_{pi}}{M}$$

(7)

M is the number of particle. P_{pi} is the historical optimal solution of the ith particle

$$P_d = \varphi * p_{best} + (1 - \varphi) p_{gbest}$$

(8)

P_d is the search area of each particle

β is contraction expansion factor defined as

$$\beta = \frac{0.5 * (\max iter - iter)}{\max iter}$$

(9)

$\max iter$ is the maximum number of iteration. $iter$ is the current number of iteration.

The algorithm flow of QPSO :

一 Initialize the M particles $X_i(0)$ randomly, and let the current best position of each particle $P_{pi}(0) = X_i(0)$ and let the global best position

$$P_g(0) = \min\{X_1(0), X_2(0), ..., X_M(0)\}$$

(10)

二 Calculate the objective function value of the particle according to the objective function $f(x)$

三 According to the following formula updates the optimal position of each particle $P_{\mathrm{pi}}(t+1)$. Assume that the objective function is to make the $f(x)$ as small as possible:

$$P_{\mathrm{pi}}(t+1) = \begin{cases} P_{\mathrm{pi}}(t); if\ f(P_{\mathrm{pi}}(t)) \le f(X_{\mathrm{i}}(t+1)) \\ X_i(t+1); f\ f(P_{\mathrm{pi}}(t) \ge f(X_{\mathrm{i}}(t+1)) \end{cases}$$

四 Update the global optimal position $P_g(t+1)$

$$P_g(t+1) = \max\{P_1(t+1), P_2(t+1), ..., P_M(t+1)\}$$

(11)

五 Compute $mbest$ according (7)

六 Compute P_d for every particle according (8)

七 Update the new position of each particle according (5).

Repeat the step 二 to step 七 until the iteration reach the maximum value

III. AUTOENCODER CLASSIFICATION ALGORITHM BASED ON QPSO

Autoencoder classification algorithm is composed of autoencoder network and softmax. In this method, we optimized the weights of autoencoder network and the parameters of softmax with QPSO.

The cost function of softmax is :

$$J(\theta) = -\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{k} I[y^i = j] \log \frac{\exp(\theta_j^T x^{(i)})}{\sum_{j=1}^{k} \exp(\theta_j^T x^{(i)})} + \frac{\lambda}{2} \sum_{i=1}^{n} \sum_{j=1}^{k} \theta_{ij}^2 \quad (12)$$

$I[y^i = j]$ is an indicator function, when $y^i = j$ is true $I[y^i = j]$=1 otherwise $I[y^i = j]$=0. $\frac{\lambda}{2} \sum_{i=1}^{n} \sum_{j=1}^{k} \theta_{ij}^2$ is regularization term, λ is penalty factor.

When the parameters of softmax and weights of autoencoder network are optimized by QPSO, the target makes the objective function value as small as possible, the objective function is defined as follow:

$$C = \eta SUM(X, Y) + \mu J(\theta) \quad (13)$$

η and μ are the coefficients of $SUM(X, Y)$ and $J(\theta)$, respectively. η is very small, μ bigger than η ,because of $SUM(X, Y)$ much larger than $J(\theta)$.

Algorithm flow description:

(一) Perform step (一) initialize m particles, each particle is a candidate for the weights in the autoencoder network.

(二) Perform step (一) initialize n particles, each particle is a candidate for the parameters in the softmax

㈢ Input data, encoder-decoder, obtain the individual optimal solution for each particle and optimal solution for the global optimal solution based on the evaluation function (11). Select the global optimal solution as the weights and parameters, respectively

㈣ while iter<=itermax

① Keep the parameters of softmax unchanged, perform step 二 to step 七 and objection function replace with (13), the global optimal position is the weights of autoencoder network

② Keep the weights of autoencoder network unchanged,

perform step 二 to step 七 and objection function replace with (13), the global optimal position is the parameters of softmax.

㈤ Get weights and parameters optimized with QPSO

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Experimental datasets

In the experiment, we tested on the dataset 'Ling-Spam', 'iris' and 'glass'. 'iris' and 'glass' is randomly selected 2/3 for the training data set, and the remaining 1/3 for the test data set. 'Ling-Spam' is selected 320 for train, 240 for test. The basics information of each data set are described as below:.

TABLE 1 INTRODUCTION OF DATA SETS

name	Number of samples	attributes	categories
Ling-Spam	2893	5	2
glass	214	10	6
iris	150	4	3

B. Experimental results and analysis

Firstly, we have tested the number of neurons in hidden layer impact on accuracy for 'iris' and 'glass'. The result as show in figure 2. we can see that for 'iris' the accuracy is better than others when the number of neurons in hidden layer is 16. And glass is 20

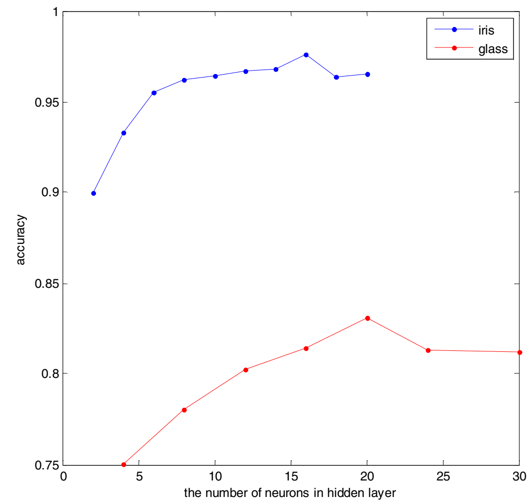

Figure 2 the number of neurons in hidden layer impact on accuracy

The experimental results of 'iris' ,'glass' and 'Lim-Spam' compare with other methods are showed at table 2 and table 3. For 'iris' data set, the depth for our method is 1 and the number of neurons in hidden layer is 16, in DELME[2] the depth is 2, and the number of neurons in hidden layer is 24. For 'glass' data set, the depth for our method is 1 and the number of neurons in hidden layer is 20, in DELME[2] the depth is 2, and the number of neurons in hidden layer is 30. In comparison, our parameters are less than DELME[2] and the results are better. For 'Ling-Spam' our method has greatly improved the recall of spam and other indicators also have improved.

Table 2 THE RESULTS IN IRIS AND GLASS(%)

name	DELME[2]	DBN[2]	RBF[2]	ELM[2]	Our method
iris	95.78	94	97.46	95.68	97.61
glass	80.87	81.27	74.51	68.00	83.1

Table 3 THE RESULT OF LING-SPAM(%)

method	index	Ling-Spam	
		norm	*spam*
PDNN[1]	*Precision*	94.98	92.01
	recall	97.85	74.25
	F1	96.39	82.18

RBFNN[1]	***Precision***	93.75	84.38
	recall	97.50	67.50
	F1	95.59	75.00
BPNN[1]	***Precision***	94.54	70.64
	recall	91.20	69.00
	F1	95.59	69.81
Our method	***Precision***	98.66	92.25
	recall	98.35	91.15
	F1	98.52	91.67

From the experimental results, we can see that our method has got good results in 'Ling-Spam', 'iris' and 'glass'. Compare with the methods of [1] and [2] our model is simpler and more effective.

V. CONCLUSION

In this paper, according to the merit of QPSO, we compare QPSO with autoencoder network. Replace BP with QPSO, we make the Autoencoder classification algorithm more effective and the parameters less .the experimental results compare with other methods show that our method convenient and efficient, and classification is more accurate

VI. ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China under Grant Nos. 61502203, the Natural Science Foundation of Jiangsu Province under Grant Nos. BK20150122, the Natural Science Research Project of Jiangsu Province University under Grant Nos. 17KJB520039

REFERENCES

[1] DONG Qing , SONG Wei. Deep neural network classification algorithm based on particle swarm optimization[J]. Transducer and Microsystem Technologies, 2017, 36(9):143-146.

[2] XU Yi, DONG Qing, DAI Xin, SONG Wei. ELM Optimized Deep Autoencoder Classification Algorithm. Journal of Frontiers of Computer Science and Technology, 2018, 12(5): 820-827..

[3] Vincent, Pascal, Larochelle, Hugo, Lajoie, Isabelle, et al. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion.[J]. Journal of Machine Learning Research, 2010, 11(12):3371-3408.

[4] Masci J, Meier U, Cireşan D, et al. Stacked convolutional auto-encoders for hierarchical feature extraction[C]// International Conference on Artificial Neural Networks. Springer-Verlag, 2011:52-59.

[5] Shi Y, Eberhart R C. Empirical study of particle swarm optimization[C]// Evolutionary Computation, 1999. CEC 99. Proceedings of the 1999 Congress on. IEEE, 2002:320-324 Vol.1..

[6] Sun J, Xu W, Feng B. A global search strategy of quantum-behaved particle swarm optimization[C]// Cybernetics and Intelligent Systems, 2004 IEEE Conference on. IEEE Xplore, 2004:111-116 vol.1..

[7] WUZhipan,ZHAOYueLong,LUOZhongLiang. License plate recognition technology based on PSO-BP neural network [J]. Acta Scientiarum Naturalium Universitatis Sunyatseni, 2017, 56(1):46-52..

[8] Liu B, Wang L, Jin Y H. An effective PSO-based memetic algorithm for flow shop scheduling[J]. IEEE Transactions on Systems Man & Cybernetics Part B Cybernetics A Publication of the IEEE Systems Man & Cybernetics Society, 2007, 37(1):18-27.

Gao hao. The Study of the Particle Swarm Optimization and Its Application in Image Segmentation [D]. Jiangnan University, 2009.