Multi-Sub-Swarm PSO Classifier Design and Rule Extraction

Yanwei Chang
School of Mechanical & Electrical Engineering
Jiangsu Normal University
Xuzhou, Jiangsu 221116, China
changyw@jsnu.edu.cn

Guofang Yu
School of Information and Electrical Engineering
China University of Mining and Technology
Xuzhou, Jiangsu 221116, China
gfyu@cumt.edu.cn

***Abstract*—In this paper, a new classifier, based on multi-sub-swarm PSO algorithm, is proposed. Natural number coding is used in the classifier to avoid the updating inconvenience of binary encoding that has different properties dimensions. Classification is done by parallel search of multi-sub-swarm PSO algorithm. According to the characteristics of the coal mine gas emission concentration data, an extraction model is constructed of classification rules of coal gas emission concentration. The results showed that this classifier have high prediction accuracy rate, and the gas emission concentration rules, extracted from its rule space using this classifier, run efficiency significantly with less redundancy.**

Keywords- multi-sub-swarm PSO; classifier; gas emission concentration; rule extraction

I. INTRODUCTION

Particle Swarm Optimization[1] is a swarm intelligence optimization algorithms[2,3] based on the theory of stochastic optimization. Being composed of simple individuals and through interaction between individuals, particle swarm produces unpredictable behavior on using the local information.

Purely from a biological point of view, the particles in the evolution of process should have four characteristic as follow: 1. Convergence. All the particles in motion move to the optimum particle. 2. Exclusion. With exclusion, swarm will be able to maintain the diversity. 3. Learnability. This feature includes the particle learn from the other neighbors and from the swarm. 4. Memory. Memory is the basis of all the features.

Inspired by these features, we propose the vertical parameter PSO in literature [4]. In the algorithm, with evolutionary process the particles dynamic cluster to form the small sub-swarm through the exchange of information between individuals, and the small sub-swarm dynamic cluster to merge into swarm through the mutual exchange information between sub-swarms. So the algorithm keeps the diversity of particle sub-swarm and improves learning ability of the particles.

Based on vertical parameters PSO Algorithm, we propose a multi-sub-swarm PSO classification algorithm in this paper, which search the best classification rules by clustering and classification in the rule space. As having parallel multi-point search feature, multi-sub-swarm PSO classification algorithm ensure that the best rules are extracted with the minimum redundancy rules, and that is applied to coal mine emission concentration classification rule extraction.

II. THE CLASSIFICATION RULE EXTRACTION BASED ON MSSPSO ALGORITHM TYPE STYLE AND FONTS

A. Particle Coding

Let D be the classification problem that will be solved, class C be the category attribute and N be the characteristic attribute. A particle represents a rule, and particle swarm does the rule set. Literature [5] uses real-coded, and particle dimension is $2*C*N$, as (1) below.

$$X=\left(\overline{X_1},\overline{X_2},\cdots,\overline{X_C},\overline{V_1},\overline{V_2},\cdots,\overline{V_C}\right) \quad (1)$$

where：

$$\overline{X_i}=\left\{x_{i,1},x_{i,2},\cdots,x_{i,N},\right\},\ \overline{V_i}=\left\{v_{i,1},v_{i,2},\cdots,v_{i,N},\right\}.$$

Rule pair real-coded is used in Literature [6-8], that is a (l_i,u_i) pair constituting a particle the ith dimension component. Here l_i and u_i are the left and right borders of the ith dimension component.

There are two models in particle binary coding, such as Holland's Michigan model[9] and De Jong's Pittsburgh model[10,11]. Literature [12,13] use the Michigan model binary coding. As particles characteristic attributes are a range of real numbers, and the binary codes have different length, the encoding also includes a categories attribute, which does not participate in the evolution. The binary coding form is as follow formula (2) and $N+1$ dimensional decimal transformed form as follow formula(3).

$$X=\left(\left(x_1\right)_B,\left(x_2\right)_B,\cdots,\left(x_N\right)_B,\left(C_k\right)_B\right) \quad (2)$$

$$X=\left(x_1,x_2,\cdots,x_N,C_k\right) \quad (3)$$

Here, $(x_i)_B$ represents the binary code of x_i, C_k represents the kth categories. Corresponding to the particle encoding, and each dimension component of speed coding is required to participate in the evolution, speed encoding thus have n dimensional component in which there is not class attribute. N dimensional form can be expressed as follow:

$$V=\left(\left(v_1\right)_B,\left(v_2\right)_B,\cdots,\left(v_N\right)_B\right) \quad (4)$$

$$V=\left(v_1,v_2,\cdots,v_N\right) \quad (5)$$

In summary, as the binary codes have different length and the rule pair real-coded have the left and right borders, using Holland or Michigan models binary coding can shaped like literature [12,13] in which encoded form could easily lead to confusion in the calculation and could not conducive to calculate vertical parameters such as the approximation degree and the position degree.

Thus, combining advantages of these two kinds of particles encoding, particle-coded use natural number coding in MSSPSO classification algorithm. That is, for continuous attributes, the data set is required preprocessing to a discrete attribute value. Particle-coded and speed-coded are as formula(3) and(5). Where, x_i is the ith characteristic attributes category after discretization and C_k is the kth category of the dataset which do not participate in evolution.

B. Fitness Function

The common measure used to evaluate the performance of classification rules is classification error rate that is the ratio of classification error in the dataset to all classification. However, we consider the corresponding classification success rate, the sum of which and the error rate is 1. So the following concepts[14] are introduced.

For a classification, the correct classification can be divided into the true positive*(TP)* and the true negative *(TN*), and misclassification error can be divided into the false positive*(FP)* and false negatives*(FN*). Therefore, the specificity and the sensitivity are defined as follows.

$$\frac{TP}{TP+FN} \tag{6}$$

$$\frac{TN}{FP+TN}=1-\frac{FP}{FP+TN} \tag{7}$$

The product of the sensitivity and the specificity is often used to measure overall classification, and can be used as fitness measuring performance of classification rules as follow.

Therefore, we use formula (8) as classification rule fitness[15].

$$f\left(x\right)=\frac{TP\cdot TN}{\left(TP+FN\right)\cdot\left(FP+TN\right)} \tag{8}$$

C. Classifier Design

The number of categories is fixed in classification problem, and the MSSPSO algorithm is a multi-sub-swarm, dynamic clustering and adaptive parallel algorithms, so the category has been unable to determine by the sub-swarm in the particles search space, i.e. rule space. From the particles encoding and fitness function representation, a sub-swarm search for an optimal rule and multiple sub-swarm do multiple rules. The particles with different categories will be belonging to different swarm as the categories attribute does not participate in the evolution. In particle evolution there are two evolutionary manners that are parallel search, which simultaneously search for different categories, and serial search, which sequentially for categories.

As particles encoding is natural number, the speed of particles is continuous quantity. The position of particle is rounded as formula (9) and (10), and the fitness function is as formula (8).

$$v_{i,k}^{\ t+1}=w_{i,k}^{\ t}v_{i,k}^{\ t}+c_1r_1(pb_{i,k}^{\ t}-x_{i,k}^{\ t})+c_2r_2(sb_i^t-x_{i,k}^{\ t}) \tag{9}$$

$$x_{i,k}^{\ t+1}=\mathrm{int}\left(x_{i,k}^{\ t}+v_{i,k}^{\ t}\right) \tag{10}$$

For dataset D in which the category attribute is C and the characteristic attribute is N, the rules search space is $N+1$ dimension, the ith particle of which is a $N+1$ dimension vector as follow.

$$x_i=\left(x_{i1},x_{i2},\cdots,x_{iN},C_k\right) \tag{11}$$

where, $i=1,2,3,\ldots,n$, $k=1,2,\ldots,C$, and x_{ij} are the natural numbers.

Classification algorithm based MSSPSO process is shown in Figure 1.

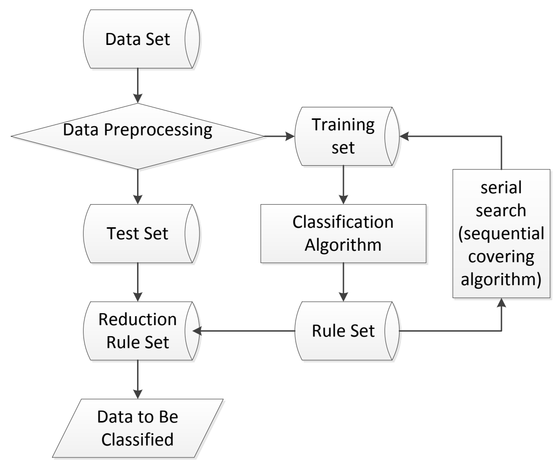

Figure 1. Classification algorithm process based on MSSPSO

Step 1 is data cleansing that is preprocessing the data set. The key is making the each attribute values discretization using dynamic clustering algorithm described previously.

Step 2 calculate the upper bound of each attribute value (lower bound is 1), and initialize the particles and its' speed.

Step 3 is Parallel search or serial search. Parallel search is to search for each category swarm until it reaches the stagnation conditions applying multi-sub-swarm PSO algorithm, and serial search is to search in turn for the categories swarm determined before.

Step 4 is rule set creating by collecting the optimal value of each sub-swarm. If use serial search, this step remove the sample data covered by the optimal value from sequential covering algorithm and then turn to step 3.

Step 5 is rules reduction for the containment relationship rules of the rule set. The rule of the optimal value of sub-swarm and the rule of the large sample data covering intersection of swarms have the containment relationship.

Step 6 is classification of the final rule set according to the classification accuracy. For data have the same precision, use the distance of the samples to be classified and conflicts rules to resolve.

Step 7 is ending.

III. CLASSIFICATION ALGORITHM TEST

To verify the effectiveness of the classification algorithm based MSSPSO, we select Iris data set for testing in UCI[15] data set.

Data preprocessing contains the standardizing attribute values and clustering characteristics attribute. We derive the number of attributes from the attribute discretization value by clustering each attribute of data set. By MSSPSO clustering algorithm[16], the number of processed data characteristic attributes shown in following table I.

TABLE I. CLUSTERING NUMBERS OF IRIS DATASET FEATURE ATTRIBUTE

Sepal length	Sepal width	Petal length	Petal width
4	4	3	3

Cluster centers successively is as following: sepal length: (2.7,5.8,6.7,7.55), sepal width: (1.2,2.65,3.4,4.15), petal length: (0.8,4.05,5.5), petal width: (0.3,1.25,2). Then the feature data are classified into each corresponding class as shown in following table II.

By MSSPSO classifier mining the processed data sets, rule sets are as follows:

Petal width = 1 => setosa;
Petal width = 3 => virginica;
Petal width = 2 ∧Petal length = 3 => virginica;
Petal width = 2 ∧Petal length = 2 => versicolor;

TABLE II. PREPROCESSING RESULT OF IRIS DATASET FEATURE ATTRIBUTE

Sepal length	Sepal width	Petal length	Petal width	Categories
1	3	1	1	setosa
1	3	1	1	setosa
1	3	1	1	setosa
1	3	1	1	setosa
2	4	1	1	setosa
3	3	2	2	versicolor
2	3	2	2	versicolor
3	3	2	2	versicolor
2	1	2	2	versicolor
3	2	2	2	versicolor
3	2	3	3	virginica
4	3	3	3	virginica
3	3	2	3	virginica
3	2	3	3	virginica
3	3	3	3	virginica

In order to understand MSSPSO classifier performance, we classify the Iris data set to comparative experiments results by using Naive Bayesian classifier and C4.5 classifier built-in Weka and MSSPSO classifier in this paper. The results are shown in following table III.

TABLE III. THE IRIS DATA COMPARISON OF THREE CLASSIFIER

Classification	The average accuracy rate
NaiveBayes	94.7
C4.5	95.333
MSSPSO	95.137

The experimental results show that, MSSPSO classifier has the same prediction accuracy and performance as C4.5 decision tree classifier and is slightly better than the naive Bayes classifier. Using MSSPSO classifier algorithm is feasible and effective.

IV. CLASSIFICATION RULE EXTRACTION GAS EMISSION CONCENTRATION

A. The Characteristics of Gas Emission Concentration Data

A lot of gas emission concentration data were stored in Information Decision Support System in Liaoning Xiaoming and Xiaoqing Coal Mine. We use the coal gas emission concentration data of Xiaoqing Coal Mine, whose period is from the July 16, 2003 - September 20, measuring point is 01A01, and the number of the recorded is 1840, of which 20 records are shown in table IV.

TABLE IV. DATA SET OF INSTANCE OF GAS EMISSION IN XIAOQING COALMINE

No.	Burial depth (m)	Coal seam thickness (m)	Gas content (m^3/t)	Daily advance (m/d)	Coal seam interval (m)	Daily output (t/d)	Gas concentration (m3/min)
1	408	2.0	1.92	4.42	20	1825	3.34
2	411	2.0	2.15	4.16	22	1527	2.97
3	420	1.8	2.14	4.13	19	1751	3.56
4	432	2.3	2.58	4.67	17	2078	3.62
5	456	2.2	2.40	4.51	20	2104	4.17
6	516	2.8	3.22	3.45	12	2242	4.60
7	527	2.5	2.80	3.28	11	1979	4.92
8	531	2.9	3.35	3.68	13	2288	4.78
9	550	2.9	3.61	4.02	14	2325	5.23
10	563	3.0	3.68	3.53	12	2410	5.56
11	590	5.9	4.21	2.85	18	3239	7.24
12	604	6.2	4.03	2.64	16	3354	7.80
13	607	6.1	4.34	2.77	17	3087	7.68
14	634	6.5	4.80	2.92	15	3620	8.51
15	640	6.3	4.67	2.75	15	3412	7.95
16	450	2.2	2.43	4.32	16	1996	4.06
17	544	2.7	3.16	3.81	13	2207	4.92
18	629	6.4	4.62	2.80	19	3456	8.04
19	517	2.6	3.10	3.51	18	2090	4.16
20	435	2.3	2.50	4.71	17	2083	3.60

Analysis of data in the table shows, the characteristic attributes, such as burial depth, coal seam thickness, gas content, daily advance, coal seam interval and daily output, have a causal relationship with gas emission concentrations, and there are certain rules which exist between them.

For the factors affect the concentration of gas emission, some of it have been mastered by engineering and technical personnel, and unknown or uncontrollable factors still exist, which caused the collected data have a certain loss. Characteristic properties appeared in table IV, already mastered for engineering and technical personnel, have a significant impact on the gas emission concentration. Gas emission concentration data in Coal Mine inherently relates to attributes and their characteristics properties.

B. Data Preprocessing

The number of categories of characteristic attributes is derived by data discretization, which is data preprocessing including data cleansing and characteristic attributes clustering operation. Using clustering algorithm based on MSSPSO clustering algorithm, the number of categories of each characteristic are shown in the following table V.

TABLE V. ATTRIBUTE CLASSIFICATION OF GAS EMISSION SAMPLES

No.	Burial depth	Coal seam thickness	Gas content	Daily advance	Coal seam interval	Daily output	Gas concentration
1	1	1	1	4	4	1	1
2	1	1	2	3	5	1	1
3	1	1	2	3	4	1	1
4	1	2	3	4	3	2	1
5	2	2	3	4	4	2	1

6	4	3	4	2	1	2	1
7	4	2	4	2	1	1	2
8	4	3	4	2	1	2	2
9	5	3	5	3	2	3	2
10	5	3	5	2	1	3	2
11	5	6	6	1	3	5	3
12	5	6	5	1	2	5	3
13	5	6	6	1	3	5	3
14	6	7	7	1	2	6	3
15	6	6	7	1	2	6	3
16	2	2	3	3	2	1	1
17	4	3	4	2	1	2	2
18	6	7	7	1	4	6	3
19	4	3	4	2	3	2	1
20	1	2	3	4	3	2	1

C. Classification Rules Minings

Using MSSPSO classifier, classification rules are extracted from coal gas emission concentration data, some of which rules are as follows:

Burial depth = 6 ∧ Coal seam thickness = 7 ∧ Gas content = 7 ∧ Daily output = 5 => Gas emission concentration = 3 (high);

Burial depth = 6 ∧ Coal seam thickness = 6 ∧ Daily output = 7 ∧ = 6 => Gas emission concentration = 3 (high);

Burial depth = 5 ∧ Coal seam thickness = 6 ∧ Gas content = 6 ∧ Daily output = 5 => Gas emission concentration = 3 (high);

Burial depth = 6 ∧ Coal seam thickness = 7 ∧ Daily output = 7 ∧ = 6 => Gas emission concentration = 3 (high);

Burial depth = 1 ∧ Coal seam thickness = 1 ∧ Gas content = 2 ∧ Daily advance = 3 ∧ Coal seam interval = 4 ∧ Daily output = 1 => Gas emission concentration = 1 (low);

Burial depth = 1 ∧ Coal seam thickness = 1 ∧ Gas content = 1 => Gas emission concentration = 1 (low);

Burial depth) = 4 ∧ Coal seam thickness = 3 ∧ Gas content = 4 ∧ Daily advance = 2 ∧ Coal seam interval = 1 ∧ Daily output = 2 => Gas emission concentration = 2 (middle);

Burial depth) = 4 ∧ Coal seam thickness = 2 ∧ Gas content = 4 ∧ Daily advance = 2 ∧ Coal seam interval = 1 ∧ Daily output) = Gas emission concentration = 2 (middle).

D. Reduction Rules

Reduction rules should follow the principles of the safe side. For the above rules, the emission concentration rules reduction should be retained important attributes taken a relatively safe value. For example, the 4 rules above, having equal importance to the reserved property, shall be taken as security value reduction as follows:

Burial depth = 5 ∧ Coal seam thickness = 6 ∧ Gas content = 6 ∧ Daily output = 5 => Gas Emission concentration = 3 (high).

Obtaining gas emission concentration rules, we can interpret and assessed the rules, of which the final rules will be stored in the knowledge database for assisting mine gas forecasts.

V. CONCLUSIONS

MSSPSO classification algorithm forms the best category on multi-sub-swarm particle swarm optimization clustering features. Classifier uses natural number coding, avoiding the use of binary encoding to avoid the inconvenience of particle velocity updates with the different dimensions of the particle properties. Data preprocessing using MSSPSO algorithm is in order to make the data attribute values for particle encoding. According to the characteristics of Mine gas emission concentration data, the multiple sub-swarm particle swarm classifier is used for gas emission concentrations classification rule extraction. Through training mine gas emission concentration data, emission rules of it are get for auxiliary mine gas forecasts with ensuring a minimum of redundant rules.

REFERENCES

[1] J. Kennedy and R. C. Eberhart, "Particle swarm optimization.," in In Proceedings of IEEE International Conference on Neural Networks, Piscataway, NJ., 1995, pp. 1942-1948.

[2] E. Bonabeau, M. Dorigo and G. Theraulaz, Swarm intelligence: from natural to artificial systems: Oxford University Press, Inc., 1999.

[3] R. Eberhart, Y. Shi and J. Kennedy, Swarm Intelligence. U.S.A: The Morgan Kaufmann Series in Artificial Intelligence, 2001.

[4] Y.W. Chang and Y.C. Wang, " Neural network ensemble based on vertical parameter PSO and its' application of mine inflow prediction," Safety in Coal Mines, pp. 45-48, 2008.

[5] I. DeFalco, A. Della Cioppa and E. Tarantino, "Facing classification problems with particle swarm optimization," Applied Soft Computing Journal, vol. 7, pp. 652-658, 2007.

[6] L.Gao, H.B. Gao, C. Zhou and D.Y. Yu, " Acquisition of pattern classification rule based on particle swarm optimization," Journal of Huazhong University of Science and Technology, pp. 24-26, 2004.

[7] T.Liu, F.Yin, J.Y. Chen and W.L. He, " Acquisition of classification rule based on quantum-behaved particle swarm optimization," Application research of computers, pp. 496-499, 2009.

[8] L.P.Yan and J.C.Zeng, " Design Classification Rules Using the PSO of Multiple Swarms," computer engineering & science, pp. 98-100+129, 2007.

[9] J. H. Holland and J. S. Reitman, "Cognitive systems based on adaptive algorithms," ACM SIGART Bulletin, p. 49-49, 1977.

[10] K. A. DeJong and W. M. Spears, "Learning concept classification rules using genetic algorithms,", 1991, pp. 651-656.

[11] K. A. DeJong, W. M. Spears and D. F. Gordon, "Using genetic algorithms for concept learning," Machine learning, vol. 13, pp. 161-188, 1993.

[12] J. R. Quinlan, "Learning efficient classification procedures and their application to chess end games," Machine learning: An artificial intelligence approach, vol. 1, pp. 463-482, 1983.

[13] X.D.Duan, Q.R.Wang, N.N.Wang, X.D.Liu and L.Shi, "Design of Classifier Based on Particle Swarm Algorithm," Computer Engineering, pp. 107-109, 2005.

[14] I. H. Witten and E. Frank, Data Mining: Practical machine learning tools and techniques: Morgan Kaufmann, 2005.

[15] X.H.Xu, A.Zhang, Y.Y.Jiang and Q.X.Hao, " One-dimensional image recognition of radar target rased on particle swarm classifier," Acta photonica sinica, pp. 2157-2162, 2007.

[16] A. Asuncion and D. J. Newman, "UCI Machine Learning Repository,", 2007.

[17] T.Liu, " Immune and Data Mining Applications," Doctor: China university of mining and technology, 2005, p. 130.