

SVM classification model in depression recognition based on mutation PSO parameter optimization

Ming Zhang^{1,2,3,4}, Shengfu Lu^{1,2,3,4}, Mi Li^{1,2,3,4,a}, Qian Zhai^{5,6}, Jia Zhou⁵, Xiaofeng Lu^{1,2,3,4}, Jiying Xu^{1,2,3,4}, Jia Xue^{1,2,3,4}, and Ning Zhong^{1,2,3,4,7}

¹Beijing Advanced Innovation Center for Future Internet Technology, Beijing University of Technology, Beijing, China

²Laboratory of Intelligent Science & Technology, International WIC Institute, Beijing University of Technology, Beijing, China

³Beijing International Collaboration Base on Brain Informatics and Wisdom Services, Beijing, China

⁴Beijing Key Laboratory of MRI and Brain Informatics, Beijing, China

⁵Beijing Anding Hospital, Capital Medical University, Beijing, China

⁶Beijing Key Laboratory for Mental Disorders, Beijing, China

⁷Department of Life Science and Informatics, Maebashi Institute of Technology, Maebashi, Japan

Abstract. At present, the clinical diagnosis of depression is mainly through structured interviews by psychiatrists, which is lack of objective diagnostic methods, so it causes the higher rate of misdiagnosis. In this paper, a method of depression recognition based on SVM and particle swarm optimization algorithm mutation is proposed. To address on the problem that particle swarm optimization (PSO) algorithm easily trap in local optima, we propose a feedback mutation PSO algorithm (FBPSO) to balance the local search and global exploration ability, so that the parameters of the classification model is optimal. We compared different PSO mutation algorithms about classification accuracy for depression, and found the classification accuracy of support vector machine (SVM) classifier based on feedback mutation PSO algorithm is the highest. Our study promotes important reference value for establishing auxiliary diagnostic used in depression recognition of clinical diagnosis.

1 Introduction

With the rapid development of society, people's living standard is constantly raising, the pressure on living for people is bigger and bigger. The mental diseases are more common, such as depression. Depression is mainly characterized by psychological disorder, and the prevalence of depression is increased year by year. Depression has become a serious harm to people's physical and mental health. According to the report of the WHO in 2001, the harm caused by depression ranked first among all kinds of diseases [1]. Data shows that 70% of people are under a condition of unhealthy or subhealthy, and psychological stress related diseases accounted for about 5% -10% of the population, physical and mental disorders has become a common disease gradually. At present, the identification and treatment of depression are mainly based on the way of face-to-face interview with the doctor according to the diagnostic manual (DSM- IV)[2]. When patients are in severe depression, the misdiagnosis rate of the doctors will be low. However, the rate misdiagnosis of the doctors is likely higher when patients are in

^a Corresponding author: limi@bjut.edu.cn

mild depression. What's more, even some healthy people may be misdiagnosed and they get wrong medicine. The reason is that the experience of doctors has great impact on the identification of depression. Therefore, the auxiliary diagnosis of depression is very important.

At present, many medical devices have been used in auxiliary diagnosis depression, such as EEG (electroencephalo-graph) and fMRI (Functional Magnetic Resonance Imaging) [3]. The EEG can acquire the signal data of the cerebral cortex. Some sensors will be placed on the patient head while collecting data. After receiving the signal from the sensors, the signal will be amplified and converted into the waveform by amplifying system. Under normal circumstances, the waveform of the EEG follows certain regularity. The regularity will be destroyed and the waveform will be changed, when the cortical lesions. Accordingly, through analysing the waveform of the EEG, it can be used in auxiliary diagnosis depression. In addition, the fMRI is also a common auxiliary diagnosis method. The fMRI can observe the brain activity in during the cognitive task process, when combing with the blood oxygen level dependence and the echo planar imaging technology, which is different from the EEG. What's more, the fMRI is a non-invasive and no radiation exposing auxiliary diagnostic method. However, the fMRI can't achieve the purpose of diagnosis depression.

The eyes are the source of emotion. The reason why eyes can express emotion is because of the expansion and narrowing of the pupil, the rotation of the eyeball and the gaze of the eyes. People's thoughts and emotions are related to the changes of pupils closely, and an unpleasant stimulus can invoke pupillary constriction [4]. Contrarily, the pleasant stimulus will make the pupil expand. Therefore, the changing of pupil is a sign of the activity of central nervous system. Data shows that the patients of depression with the features of gazing the positive emotional information (such as happy expression) less and gazing negative emotional information (such as sad expression) frequently [5]. Accordingly, we can know people's emotional state by tracking the attention of a person's bias attention, the rate of pupil diameter, the saccadic slope and other visual information.

We will use an infrared eye tracking device Tobii T120 to obtain the visual information and the pupil size changes related to cognition, and then we use support vector machine (SVM) and the particle swarm algorithm to extract and classify the collected features to achieve the purpose of diagnosis of depression. The parameters in the SVM have great influence on the accuracy of data recognition, and setting the reasonable parameters of the SVM will improve the recognition accuracy effectively.

Based on above description and discussion, the main purpose of this paper is to optimize the parameters of the SVM by using an improved particle swarm optimization algorithm, so as to improve the recognition accuracy of the SVM. First, a feedback mutation particle swarm optimization algorithm is proposed in this paper. Then, we use the proposed particle swarm optimization algorithm to optimize the parameters based on radial basis function function (RBF) of the SVM. Last, we use the optimized SVM to classify the data of depression patients, and the result proves our proposed method is effective.

2 PSO algorithms

Particle swarm optimization (PSO) algorithm is a swarm intelligence algorithm which is proposed by Eberhart and Kennedy [6]. It is a kind of optimization algorithm based on iteration. At the beginning of the algorithm, the speed and position of the particles are required to be randomly initialized. Then the particles try to find the optimal solution by iterating in the solution space. In PSO, each particle is flying at a certain speed in D dimension of the search space. The position of the i th particle is expressed as $x_i=(x_{i1},x_{i2},\dots,x_{id})$, where $x_{id}\in [popmin,popmax]$, d is the dimension of the search space and $d\in [1,2,\dots,D]$. $popmin$ and $popmax$ are the minimum and maximum value of the search space. The velocity of the i th particle is expressed as: $v_i=(v_{i1},v_{i2},\dots,v_{id})$, where $v_{id}\in [v_{min},v_{max}]$, $d\in [1,2,\dots,D]$, and v_{min} and v_{max} are the minimum and maximum value of the velocity of the particle.

The position x_i and velocity v_i of the particles are updated by the following equations:

$$v_{ij}(t+1) = v_{ij}(t) + c_1r_1[pbest_{ij}(t) - x_{ij}(t)] + c_2r_2[gbest_j(t) - x_{ij}(t)] \quad (1)$$

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1) \quad (2)$$

where i denotes the i th particle, j denotes the j th dimension of particles, t denotes iteration, v_{ij} denotes j th dimension velocity of i th, x_{ij} denotes j th dimension position of i th, $pbest_{ij}$ is the j th dimension of the best previous position of the particle itself and $gbest_j$ denotes the j th dimension of the best previous position of all particles in the swarm. c_1 and c_2 are the acceleration parameters, both r_1 and r_2 are randomly numbers and they change in $[0,1]$.

2.1 Mutation PSO algorithm

The PSO algorithm need set only few parameters, and it convergences fast, but it falls into local optimum easily. In order to avoid the shortcomings, many new methods are proposed. Such as Natsuki and Higashi promoted a Gaussian mutation PSO algorithm [7]. The Gaussian mutation of the basic idea is that the position of the particle will be changed when the particle i fall into the local optimum. The change method for the position of particle is adding a distribution of Gaussian values to original position. Through the Gaussian mutation, particles jump in the vicinity of the original position, to a certain extent, improving the particles' ability to jump out of the local optimal position. In addition, there are many variants of Gaussian mutation, such as Zhan D et al. introduced Gaussian cloud learning strategy [8].

The probability density function image of Gaussian distribution is characteristic with both ends are narrow, high in the middle so that the mutation particles will not have large change when the Gaussian mutation is used. For this reason, Wang H et al. proposed a Cauchy mutation algorithm [9]. The Cauchy mutation also disturbs the position of particles, but the difference is the method of disturbance which adds a value of Cauchy distribution to the original position.

The probability density function of the Cauchy distribution is a relatively smooth strip, which has the characteristics of both ends of large and middle small. This leads to return a larger probability of Cauchy distribution, which makes the position of particles have a larger change on the original basis. Therefore, the Cauchy mutation has a strong ability of disturbance, which is more suitable for long-distance jump. In order to reduce the variation of long range hopping and to better control the jump distance of the particles, Zhang L et al. added a scaling factor on the basis of Cauchy PSO mutation [10].

Dirk Brockmann found Levy flight pattern [11]. From the distribution point of view, Levy flight is most of the value changes in a small range, but it also produces a long distance occasionally. The small probability time may bring great influence. Based on this feature, Hakl H and Uğuz H applied Levy flight pattern to particle swarm optimization [12]. Because most cases of the Levy flight pattern moves in a small range, and produce a long-distance movement occasionally. On the one hand, it is advantageous for the fine search in a small range; on the other hand, it can generate long range hopping for the particles which trapped in the local optimum. However, such a large range of jump phenomenon is accidental, and the stability of the algorithm needs to be improved.

The Gaussian mutation changes the range of particle position smaller, and the Cauchy mutation have great change to the particles position, Levy mutation can also bring a large range to change the location of the particle but with chance. Wang H et al. used an adaptive method combines the three mutation algorithm to balance the influence of this three mutation algorithm on the movement of the particle position [13]. Although this method combined with three mutation algorithm, in the process of, it also changed the particle population optimal location. As we all known, the particles move toward the optimal position of the population in the process of iteration, the convergence speed of the algorithm is decreased because of the change of the population optimal position. Similarly, Nishio T et al. used the individual best position on the mutation strategy [14].

Andrews, P. S. adapted a random mutation algorithm [15]. The mutation idea was that when the population of the global optimal particle continuous N generation is not improved, then use the particle mutation operation to generate a new "social part" learning sample. However, the position of the particles in the mutation operation is not in accordance with a certain distribution, but randomly

selected location in the search area. Through the mutation operation, particles distribution is more extensive and location diversity has been promoted. To some extent, avoiding the particles fall into local optimal solution. However, it's so random on the particles position mutation, the speed of particles convergence slow down.

Beyond all that, there are also many other improvement methods. Such as Dong w et al. [16] used adaptive mutation strategy, Ngo T T et al. [17] proposed extraordinary motion for particles in the PSO all got better results.

2.2 Feedback mutation PSO algorithm (FBPSO)

All the above mutation algorithms, to a certain extent, improve the ability of particles jump out of the local optimum solution [7, 9, 12, 13, 15]. However, these algorithms don't consider the information of the particles in the population when the particles are mutating. Therefore, this paper presents a feedback mutation particle swarm optimization algorithm. We use the fitness's feedback of particles to balance the global exploration and local search.

The fitness value particles can reflect the current state of particles, and fitness is also a kind of judgment whether particle swarm convergence property. Therefore, this paper uses particles' average fitness value as the condition to decide when particles need to mutate. The mutation thought is that if the current particle's fitness less than the average fitness, indicating where the particle position is better, but the better position may be a local optimization. So select a portion of the particles from these particles to mutate.

The formula of the particle's average fitness is defined as follows:

$$fitness(i) \leq fitness\ avg \tag{3}$$

$$fitnessavg = \frac{\sum_{i=1}^s fitness(i)}{s} \tag{4}$$

where s denotes population size, $fitness(i)$ denotes fitness of i th particle, $fitnessavg$ denotes average fitness of all particles in swarm.

Then, we select some of particles mutate, and the selection method is as follows:

$$rand() \leq P_u \tag{5}$$

$$P_u = 1 - t / T \tag{6}$$

where t denotes current iteration, T denotes total number of iterations, $rand()$ is a random value uniformly distributed in $[0, 1]$.

The condition (3) is used to judge the convergence of the population, The condition (5) is used to choice which particle to make mutate. If the two conditions are met, we will mutate the particle. The mutation formula is defined as follows:

$$x_i' = x_i + \beta \cdot x_i \tag{7}$$

$$\beta \sim N(0, \sigma^2) \tag{8}$$

$$\sigma = \sqrt{\frac{|fitness(i) - fitnesszbest|}{|fitnessavg - fitnesszbest|} + 0.1} \tag{9}$$

where x_i' denotes the after mutation position of i th particle. x_i is position of current particle. β obeys Gaussian distribution the expectations of 0, the standard deviation is σ . $fitness(i)$ denotes fitness of i th particle. $fitnesszbest$ denotes the best fitness in the whole swarm. If the $fitness(i)$ become larger, we can find σ is larger in formula (9), it means the larger standard deviation of Gauss distribution. At this time, the probability density image of Gauss distribution is relatively flat, the distribution is more dispersed, and the possibility of obtaining greater value is increased. The position of the particle will change in a large range, which can complete the long distance movement of the particle position to improve the ability to jump out of the local optimal solution. On the contrary, if the $fitness(i)$ become

smaller, we can find σ is smaller in formula (9), at this time, Gauss distribution probability density image is taller, distribution is relatively concentrated, the possibility of getting a larger value is declined. The position of the particle will change in a small range, and the particles can complete fine search near the optimal solution.

The mutation method that we propose controls the global exploration and local search ability by using the fitness feedback of current particle.

3 Depression Recognition Methods Based on PSO and SVM

3.1 Methods

In this study, 48 subjects, including 24 patients with depression and 24 healthy subjects whose gender, age and education level are similar, participate in the study. The experiment materials are obtained from the international standard expressions library, the NimStim set of facial expressions [18]. We selected 36 people's (18 males, 18 females) three types (happy, sad and neutral) expressions. During the experiment, we use infrared vision eye tracking systems (Tobii T120) to record eye related information. We identified a total of five features, including bias positive attention, bias negative attention, the rate of positive pupil diameter, the rate of negative pupil diameter and the positive saccadic slope. The total number of subjects is 1728, including 864 depressed and 864 healthy subjects.

We used SVM classifiers based on the RBF to classify depressed and healthy subjects, with the SVM penalty coefficient C and parameter g . For any two models, different (C , g) values will produce a different recognition promotion. During parameter selection and optimization, we chose the highest recognition promotion of (C , g) as the optimal parameter values.

3.2 Feature Selection

There are many methods for selecting features, including entropy, probability distribution, and reparability criterion based on statistical tests. In this paper, we used statistical methods to judge the classification indicators. In the features of the normal group and the patient group, if the feature is significant difference, it can be used as the feature of classification. Before selecting feature, we also cleared exception data, for example, we cleared beyond three times the standard deviation range data. What's more, in order to unify the dimension of data, we normalized the data.

Generally, training sets should account for 70%-85% of the total sample, and we use 75%. We divided the data into training set and test set. The training set is 1296, including 648 depressed and healthy subjects. The test set is 432, including 216 depressed and healthy subjects.

The experiment platform is lenovo G470 PC, 64-bit Windows 7 operating system, the CPU is Intel core i32350M, and it's frequency is 2.30 GHz, memory is 4GB. In experiment, we use LIBSVM toolkit to test the ability of the six kinds of PSO algorithm in optimizing the parameters of the SVM base on MATLAB R2010b.

3.3 Parameter Optimization based on Mutation Method

In order to analyze the effect of FPSO, in this paper, we use different mutation algorithm to compare with FBPSO. In the experiment, we compare PSO (particle swarm optimization)[7], GPSO(Gaussian mutation particle swarm optimization)[9], HPSO(Cauchy mutation particle swarm optimization)[12], LFPSO(Levy flight particle swarm optimization)[13], RPSO(Rand mutation particle swarm optimization)[15] and FBPSO(Feedback mutation particle swarm optimization) total six algorithms to optimize the parameters of SVM. The six mutation algorithms are shown in Table 1:

Table 1 different mutation algorithms

Algorithm	Mutation
PSO	—
GPSO	$x_i' = x_i(1 + \text{Gaussian}(\sigma))$
HPSO	$x_i' = x_i + \text{Cauchy}(\sigma)$
LFPSO	$x_i' = x_i + S$ $S = \alpha + \text{Levy}(\beta)$
RPSO	$x_i' = x_i + (\text{pop}_{\max} - \text{pop}_{\min}) \cdot r$ $x_i' = x_i + \text{Gaussian}(\sigma) \cdot x_i$
FBPSO	$\sigma = \sqrt{\left \frac{\text{fitness}(i) - \text{fitnesszbest}}{\text{fitnessavg} - \text{fitnesszbest}} \right + 0.1}$

In this paper, the FBPSO algorithm process is as follows:

Step 1: Initialize the position and velocity of the particles, set the size of the population and the number of iterations.

Step 2: Train the SVM model based on the parameters C and g , compute the classification accuracy.

Step 3: Compute the fitness values of all the particles.

Step 4: Update the personal best and global best values.

Step 5: Update the particle's speed and position.

Step 6: Calculate $\text{fitness}(i)$ and fitnessavg , if $\text{fitness}(i) \leq \text{fitnessavg}$ then continue, else goes to Step9.

Step 7: If $\text{rand}() \leq P_u$ then continue, else goes to Step9.

Step 8: Calculate σ according to the formula (9) and make the mutation according to the formula (7).

Step 9: If the maximum number of iterations is reached then continue, else jump to step 2.

Step 10: Output the parameters C and g .

4 Comparisons of the Results from Different Mutation Algorithms

The experiment results of different mutation algorithms are shown in Table 2.

4.1 Comparison of the classification accuracy

From the Table 1, we can see that the classification accuracy of the original PSO classifier without mutation is lowest, it's 85.65%. After adding the mutation strategy, the recognition accuracy of all the classifiers is improved, which shows that the mutation strategy improve the ability of jumping the local optimal solution. In all the mutation algorithms, the classification accuracy of RPSO is lowest. The main reason is that the RPSO adapt random mutation strategy, which leads to the position of particles disperse after the mutation, and affects the final classification accuracy. In this paper, the classification accuracy of FBPSO algorithm is the highest, and it is proved that the improved algorithm is beneficial to balance the global exploration and local search.

4.2 Comparison of the programs running time

By comparing the running time of PSO and other mutation algorithms, the without mutation PSO algorithm and other mutation algorithms except RPSO of the runing time is almost the same. The running time of RPSO algorithm is the longest, which shows that the convergence time of the algorithm is affected by the random mutation strategy. The running time of GPSO algorithm is the shortest, which shows that the *Gaussian* mutation, to a certain extent, accelerate the convergence of

algorithm. In this paper, the running time of improved FBPSO is the second. It shows that using fitness feedback of to adjust the position of the particles can improve the convergence of the algorithm.

Table 2 Results of different algorithms

Algorithm	C	g	Accuracy (%)	Running time (s)
PSO	2.2141	3.6930	85.65	559.66
GPSO	2.3327	3.2975	86.34	541.52
HPSO	2.3526	3.2948	86.13	585.77
LFPSO	2.6811	3.2562	86.36	578.53
RPSO	2.3448	3.3352	86.11	892.71
FBPSO	1.9884	3.5110	88.56	552.07

5 Discuss and Conclusion

In this paper, the feedback mutation particle swarm optimization (FBPSO) improves the classification accuracy after it's applied to the SVM classification model. Furthermore, the running time of the algorithm needs less time than other algorithms except the *Gaussian* mutation (including *Cauchy* mutation, *Levy* mutation and *Rand* mutation algorithms). On the one hand, this paper improves the PSO algorithm through the mutation of the position of particles. In addition to the mutation PSO algorithm, the inertia weight, the accelerating factor and other parameters can also influence the updating formula of particles directly. If these parameters are selected properly, they will also have a great influence on convergence accuracy and convergence time of particles [19]. Therefore, the particle velocity updating formula can also be improved [20]. On the other hand, this paper classifies the data of depression recognition based on the SVM classification model, if we use the other classification models that are popular in the last few years, such as deep learning, may improve the classification accuracy further. For the above problems still need further study.

Acknowledgment

This work is supported by the National Basic Research Programme of China (2014CB744600), the National Natural Science Foundation of China (61602017), the Beijing Natural Science Foundation (4164080), the Beijing Outstanding Talent Training Foundation (2014000020124G039), the National Natural Science Foundation of China (61420106005), the International Science & Technology Cooperation Program of China (2013DFA32180), the Special fund of Beijing Municipal Science and Technology Commission (Z151100003915117), Beijing Municipal Administration of Hospitals Clinical Medicine Development of Special Funding Support (ZYLX201607 and ZYLX201403), and Beijing Municipal Administration of Hospitals' Ascent Plan (DFL20151801).

References

1. Guilbert J J. The world health report 2002 - reducing risks, promoting healthy life.[J]. Education for Health, 2003, 16(16):230-230.
2. Segal, Daniel L. Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR)[M]// Diagnostic and statistical manual of mental disorders :. American Psychiatric Association, 2013:4189-4189.
3. Christmann C, Koeppel C, Braus D F, et al. A simultaneous EEG-fMRI study of painful electric stimulation[J]. Neuroimage, 2007, 34(4):1428-37.
4. Hess E H. ATTITUDE AND PUPIL SIZE.[J]. Scientific American, 1965, 212(4):46-54.

5. Duque A, Vázquez C. Double attention bias for positive and negative emotional faces in clinical depression: Evidence from an eye-tracking study[J]. *Journal of Behavior Therapy & Experimental Psychiatry*, 2015, 46:107-14.
6. Y.H. Shi, R.C. Eberhart, Parameter selection in particle swarm optimization, in: *Annual Conference on Evolutionary Programming*, San Diego, 1998
7. Higashi N, Iba H. Particle swarm optimization with mutation[C]//*Swarm Intelligence Symposium*, 2003. SIS'03. Proceedings of the 2003 IEEE. IEEE, 2003: 72-79.
8. Zhan D, Lu H, Hao W, et al. Improving particle swarm optimization: Using neighbor heuristic and Gaussian cloud learning[J]. *Intelligent Data Analysis*, 2016, 20(1): 167-182.
9. Wang, H., Li, C., Liu, Y., & Zeng, S. (2007). A Hybrid Particle Swarm Algorithm with Cauchy Mutation. *Swarm Intelligence Symposium*, 2007. *Sis* (pp.356-360).
10. Zhang L, Tang Y, Hua C, et al. A new particle swarm optimization algorithm with adaptive inertia weight based on Bayesian techniques[J]. *Applied Soft Computing*, 2015, 28: 138-149.
11. Brockmann, D., & Sokolov, I. M. (2002). Lévy flights in external force fields: from models to equations. *Chemical Physics*, 284(1), 409-421.
12. Hakl H, Uğuz H. A novel particle swarm optimization algorithm with Levy flight[J]. *Applied Soft Computing*, 2014, 23(5):333-345.
13. Wang H, Wang W, Wu Z. Particle swarm optimization with adaptive mutation for multimodal optimization[J]. *Applied Mathematics & Computation*, 2013, 221(9):296-305.
14. Nishio T, Kushida J, Hara A, et al. Adaptive particle swarm optimization with multi-dimensional mutation[C]//2015 IEEE 8th International Workshop on Computational Intelligence and Applications (IWCIA). IEEE, 2015: 131-136.
15. Andrews, P. S. (2006, July). An investigation into mutation operators for particle swarm optimization. In *2006 IEEE International Conference on Evolutionary Computation* (pp. 1044-1051). IEEE.
16. Dong W, Kang L, Zhang W. Opposition-based particle swarm optimization with adaptive mutation strategy[J]. *Soft Computing*, 2016:1-10.
17. Ngo T T, Sadollah A, Kim J H. A cooperative particle swarm optimizer with stochastic movements for computationally expensive numerical optimization problems[J]. *Journal of Computational Science*, 2016, 13:68-82.
18. Nim Tottenham, et al., The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research* 168 (2009) 242-249.
19. Lim W H, Isa N A M. Two-layer particle swarm optimization with intelligent division of labor[J]. *Engineering Applications of Artificial Intelligence*, 2013, 26(10): 2327-2348.
20. Mandal S, Ghoshal S P, Kar R, et al. Design of optimal linear phase FIR high pass filter using craziness based particle swarm optimization technique[J]. *Journal of King Saud University-Computer and Information Sciences*, 2012, 24(1): 83-92.