Construction and Application of Modifying Weights with Multi-Strategies Fruit Fly Optimization Algorithm Support Vector Regression

Yunlong Zhong[a], Qing Wu[b,*], Jianxin Chen[a,c], Yongli Niu[b], Shuchun Pang[b]

[a]School of Chemical Engineering, Hebei University of Technology, Tianjin 300130, China
[b]School of Computer Science and Engineering, Hebei University of Technology, Tianjin 300401, China
[c]Engineering Research Center of Seawater Utilization Technology, Ministry of Education, Hebei University of Technology
 Tianjin 300130, China
 qingwu@hebut.edu.cn

The present work constructs the fruit fly optimization algorithm integrated with multi-strategies modified weights support vector regression model (FOAAMWP-SVR). It is based on fruit fly optimization algorithm (FOA) by introducing multi-strategies modified searching weights and linear modified optimal fruit fly weight, and then combining it with support vector regression (SVR). To construct FOAAMWP-SVR model, the support vector regression model based on fruit fly optimization algorithm (FOA-SVR) is firstly formed to obtain the optimal parameter of support vector regression, and proceeds to integrate with multi-strategies to modify searching weight and the optimal fruit fly weight. In addition, a simulation experiment of activity product regression of solid solution in KCl-NH$_4$Cl-H$_2$O ternary system is applied to test FOA and FOAAMWP-SVR algorithms, and the calculation shows that the regression error of two models is small, while FOAAMWP-SVR is more accurate than FOA model.

1. Introduction

Fruit fly optimization algorithm (FOA) is an intelligent revolution algorithm which simulates fruit flies' foraging behaviour (Pan, 2012). FOA is usually used to search for the optimal solution and the principle is simple and understandable (Hazim and Mesut, 2014). Compared with other algorithms, it has the advantages of less adjusting parameters, smaller computational amount, stronger ability of global optimization and more accuracy of optimization, etc. (Hou, 2015). Recently, optimization algorithms based on FOA have been widely put forward to solve problems in different fields, such as operation optimization in chemical engineering (Niu, et al., 2015), cloud computing in mathematical analysis (Wu et al., 2015) and other applications in scientific disciplines (Wang, et al., 2016). Schematic diagram and flowchart of FOA is shown in Figure 1.

Despite FOA's strong global searching ability and high rate of convergence, while in the process of iterative optimization, its aggregation behaviour towards the optimal individual may easily lead to the loss of population diversity. If the individual is not the optimal, the algorithm is likely to fall into local optimum, causing premature convergence problem (Li and Chen, 2016). Meanwhile, in the next course of iterative optimization, the searching scope weights of the optimal fruit fly will be the same as other fruit flies, failing to represent the advantage over the former. It means that the algorithm may run for a long time, but fails to find the optimal solution.

To solve this problem, it has to determine whether the fruit fly is the optimal one firstly. As for those non optimal fruit flies, not only location of the optimal fruit fly but also the impact of location on food hunting should be taken into consideration. Besides, the locations of fruit flies should be modified and reset through linear increasing, and the searching scope should be narrowed down to improve the efficiency of obtaining the globally optimal solution (Jiang et al., 2016). In this case, support vector regression (SVR), a dot product algorithm, is proposed to combine with FOA to obtain the optimal fruit fly accurately (Zhao et al., 2013). As for nonlinear SVR, the input vector is mapped to high-dimensional space by nonlinear mapping $\varphi(x)$, then it

constructs the optimal decision function based on structural risk minimization and finally replaces high-dimensional feature space by kernel function $K(x_i, y_i) = \varphi(x_i) \cdot \varphi(y_i)$ in the original space (Zhang, 2000). By combining the Lagrange function and duality principle with kernel function, the decision function

$$y = f(x) = \sum_{i=1}^{i} (a_i - a_i^*) K(x_i, y_i) + b$$ can be then obtained, where a and b represents Lagrange multiplier and

the offset (Wang and Feng, 2016). From the construction of FOA-SVR model, the optimal parameter of SVR can be obtained (Niu et al., 2017). The minimum value of the fitness function corresponds to the particular flavour concentration judgment value, which represents the optimal solution parameter population size, maximum number of iterations and kernel function parameter of SVR.

However, in practical application, it depends on individual experience to choose parameter values of SVR, which directly affect the accuracy and the generalization ability of algorithm. Unsatisfactory result will be obtained if parameter value is chosen blindly and randomly. Therefore, the aim of this work is to realize the adaptive selection of parameter values based on FOA and SVR. The modifying weights with multi-strategies fruit fly optimization algorithm support vector regression model (MWMSFOA-SVR) is constructed, in which multi-strategies modified searching weights and linear modified optimal fruit fly weight are introduced. The optimal parameter of SVR can be obtained adaptively and the result of MWMSFOA-SVR model is more accurate and reliable.
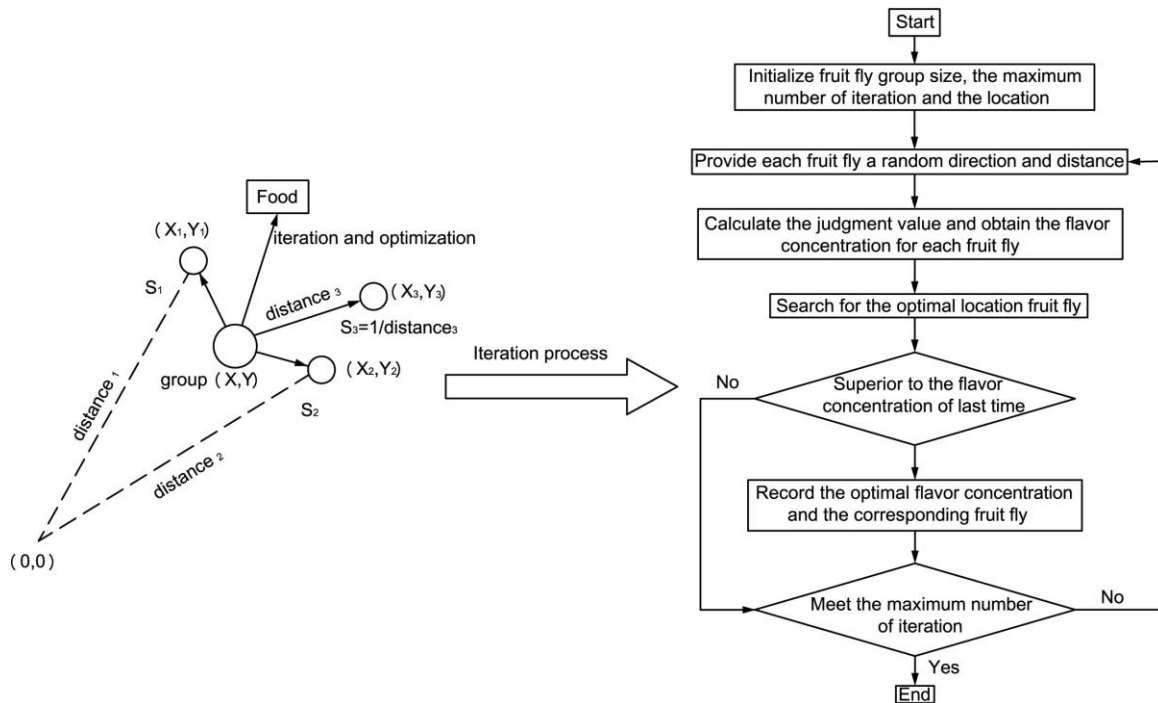

Figure 1: Schematic figure of the foraging and flowchart of FOA

2. Construction process and simulation experiment

2.1 Construction of MWMSFOA-SVR

As for the construction of modifying weights with multi-strategies fruit fly optimization algorithm support vector regression model (MWMSFOA-SVR), FOA and SVR are firstly combined to obtain the optimal parameter of

SVR. In this work, the radial basis function $K(x, y) = e^{-\|x-y\|^2/\delta^2}$ acts as the kernel function of SVR, where e represents exponential function and δ is kernel function parameter. It can map the sample set from input space to high-dimensional feature space, and it has the ability to deal with the complex nonlinear relation between the input and the output of the sample set.

The next step is to integrate modifying weights with support vector regression. Modified searching weight and linear modified optimal fruit fly weight are introduced to obtain the optimal parameter adaptively. With the multi-strategies, iteration of each time can effectively modify the weight of the optimal parameter, which expands the search range of individual fruit fly and improves the search speed of fruit fly group.

Thus, the framework of MWMSFOA-SVR model construction is firstly using FOA-SVR to obtain the optimal parameter, and then integrating with multi-strategies to introduce modified searching weights and linear modified optimal fruit fly weight.

The basic steps of MWMSFOA-SVR model are shown as follows:

Firstly, initialization parameter: population size, maximum number of iterations and random initialization location of the fruit fly group X_{axis}, Y_{axis}.

Secondly, provide random direction and distance of food hunting process which depends on fruit fly group's olfaction as Eq(1) and Eq(2).

$$X_i = X_{axis} + r \tag{1}$$

$$Y_i = Y_{axis} + r \tag{2}$$

where X, Y represent the location of fruit fly, a random value r is food hunting scope of fruit fly.

Thirdly, calculate D_i, distance between each fruit fly and original point; then calculate S_i and flavour concentration.

$$D_i = \sqrt{X_i^2 + Y_i^2} \tag{3}$$

$$S_i = 1/D_i \tag{4}$$

in which S_i is the reciprocal of D_i.

Fourthly, substitute S_i into flavour concentration function and calculate flavour concentration of each fruit fly C_i.

$$C_i = F(S_i) \tag{5}$$

where F represents the fitness function.

Fifthly, search, record and retain the best flavour concentration B_{smell} and the corresponding coordinate (X, Y).

$$[B_{smell}, B_{index}] = \min(C_i) \tag{6}$$

$$B_{temp} = B_{smell} \tag{7}$$

$$X_{best} = X(B_{index}) \tag{8}$$

$$Y_{best} = Y(B_{index}) \tag{9}$$

Sixthly, enter iterative optimization, and calculate weight coefficient of linear increasing w according to Eq(10).

$$w = (w_1 - w_2)/M \times g \tag{10}$$

in which w_1, w_2 are fixed parameters and g is the current number of iterations.

Seventhly, if the fruit fly is the optimal one, then relocate the fruit fly according to Eq(11) and Eq(12).

$$X_i = w \times X_{best} + r' \tag{11}$$

$$Y_i = w \times Y_{best} + r' \tag{12}$$

where r' represents the smaller food hunting scope of fruit fly.

Eighthly, if the fruit fly is not the optimal one, then relocate it according to Eq(13) and Eq(14).

$$X_i = w \times (X_{best} + X_i) + r \tag{13}$$

$$Y_i = w \times (Y_{best} + Y_i) + r \tag{14}$$

in which r is food hunting scope of fruit fly.

Ninthly, repeat the thirdly to eighthly step until the maximum number of iteration is met.

Then the algorithm comes to an end and gives the result.

Figure 2 shows the optimal parameter obtaining flow diagram and the construction of MWMSFOA-SVR model. In the left, FOA-SVR model is constructed, from which the optimal parameter of SVR can be obtained. And in the right, multi-strategy is introduced and it can realize the adaptive modification of the weight of parameters. By combing these two processes, MWMSFOA-SVR model is constructed.
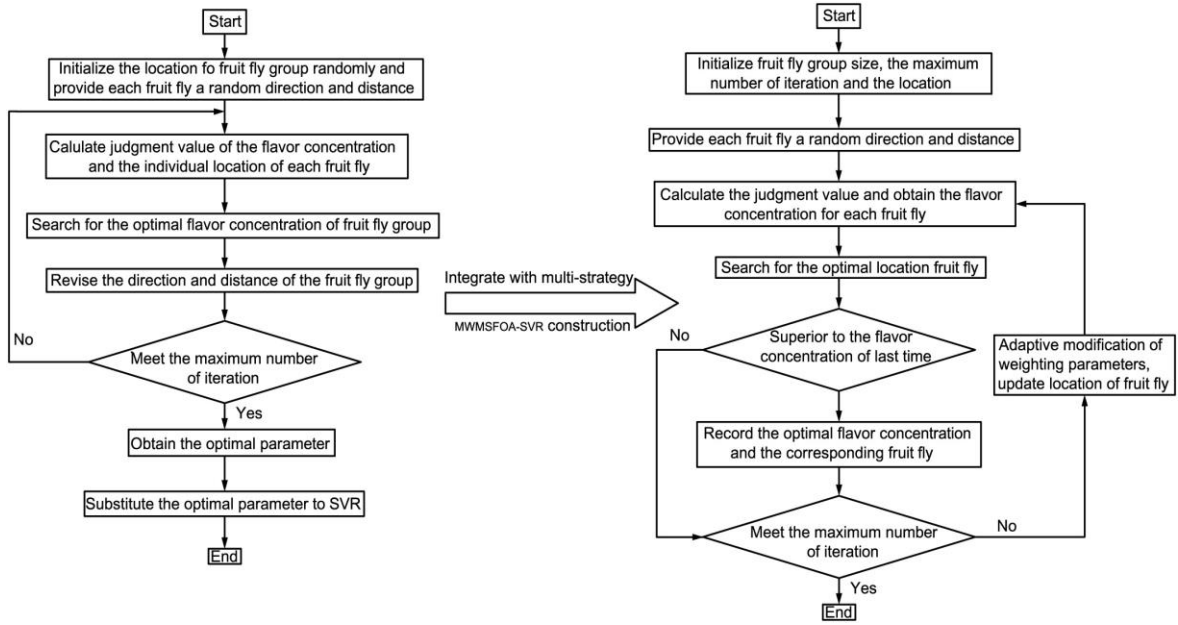
Figure 2: Optimal parameter obtaining flow diagram and the construction of MWMSFOA-SVR model

2.2 Set-up of simulation experiment

To test the accuracy and feasibility of MWMSFOA-SVR model, a simulation experiment of activity product regression of solid solution in KCl-NH$_4$Cl-H$_2$O ternary system is conducted. As for the study of electrolyte solution, Pitzer theory is often used to calculate activity coefficient and predict solid-liquid equilibrium (Pitzer, 1973). However, for some mixed salt-water system, it will form solid solution in the crystallization, which makes it impossible to calculate the activity of each component and predict the phase equilibrium. Taking KCl-NH$_4$Cl-H$_2$O ternary system for example, solid solutions of (K,NH$_4$)Cl and (NH$_4$,K)Cl will be formed in crystallization, the composition of which is variable and activity coefficient for each component is not a constant. Therefore, investigation of relation between activity and solid solution composition is necessary.

In this work, solid-liquid equilibrium data of KCl-NH$_4$Cl-H$_2$O ternary system at 30°C are taken as sample, which is measured according to literature (Zhong et al., 2017). To obtain the activity of solid solution, we assume the molar fraction of NH$_4$Cl in (K,NH$_4$)Cl and (NH$_4$,K)Cl solid solutions are P and Q, and the dissolution equilibrium of (K,NH$_4$)Cl and (NH$_4$,K)Cl solid solutions can be expressed as Eq(15) and Eq(16).

$$\left(K_{1-P}^+ NH_{4P}^+\right)Cl^- \Leftrightarrow (1-P)K^+ + PNH_4^+ + Cl^- \tag{15}$$

$$\left(K_{1-Q}^+ NH_{4Q}^+\right)Cl^- \Leftrightarrow (1-Q)K^+ + QNH_4^+ + Cl^- \tag{16}$$

Then, activity product of solid solutions can be calculated if the molarity of K$^+$ and NH$_4^+$ in experiment is determined, and the formulas to calculate the activity of solid solutions are given as Eq(17) and Eq(18).

$$K_{a1} = m_{K^+}^{(1-P)}\gamma_{K^+}^{(1-P)}m_{NH_4^+}^P\gamma_{NH_4^+}^P m_{Cl^-}\gamma_{Cl^-} = f_1(P) \tag{17}$$

$$K_{a2} = m_{K^+}^{(1-Q)}\gamma_{K^+}^{(1-Q)}m_{NH_4^+}^Q\gamma_{NH_4^+}^Q m_{Cl^-}\gamma_{Cl^-} = f_2(Q) \tag{18}$$

where m is molality of solute; K_{a1} and K_{a2} are activity of (K,NH$_4$)Cl and (NH$_4$,K)Cl solid solutions, which can be thought of as a function of the molar fraction of NH$_4$Cl. Moreover, ion activity coefficient γ is obtained by Pitzer's equation and literature has offered a detailed course of calculation (Xue et al., 2016), which could be regarded as a function of molarity at a constant temperature. Eq(19) and Eq(20) are relations between activity coefficient and molarity and the charge balance equation.

$$\gamma_i = f(m_i) \tag{19}$$

$$m_{K^+} + m_{NH_4^+} = m_{Cl^-} \tag{20}$$

With Eq(17) to Eq(20), the relation between activity product and molar fraction of NH_4Cl can be regressed.

3. Result and discussion

3.1 Experimental data analysis

Solid-liquid equilibrium data of $KCl-NH_4Cl-H_2O$ ternary system at 30°C is listed in Table 1. It can be seen that the equilibrium solid phase contains pure NH_4Cl, pure KCl, $(K,NH_4)Cl$ solid solution, $(NH_4,K)Cl$ solid solution and the mixed crystal of two solid solutions. As for two solid solutions, the composition is variable and the relation between activity and solid phase cannot be calculated using Pitzer theory. FOA and MWMSFOA-SVR models are used to obtain the regression relation.

Table 1: solid-liquid equilibrium data of $KCl(1)-NH_4Cl(2)-H_2O(3)$ ternary system at 30°C

Composition of liquid phase (wt%)			Composition of solid phase (g/100g dry salt)		Equilibrium solid phase
$100w_1$	$100w_2$	$100w_3$	$100w_1$	$100w_2$	
0.00	29.29	70.71	0.00	100.00	NH_4Cl
1.87	28.18	69.95	0.00	100.00	NH_4Cl
3.91	27.11	68.98	0.00	100.00	NH_4Cl
5.87	25.96	68.17	0.00	100.00	NH_4Cl
6.96	25.26	67.78	3.45	96.55	$(K,NH_4)Cl$
8.02	24.52	67.46	6.93	93.07	$(K,NH_4)Cl$
9.10	23.79	67.11	10.41	89.59	$(K,NH_4)Cl$
10.18	23.05	66.77	13.89	86.11	$(K,NH_4)Cl$
11.20	22.30	66.50	17.30	82.70	$(K,NH_4)Cl$
11.23	22.37	66.40	20.78	79.22	$(K,NH_4)Cl$
11.23	22.37	66.40	49.70	50.30	$(K,NH_4)Cl+(NH_4,K)Cl$
11.23	22.37	66.40	80.95	19.05	$(NH_4,K)Cl$
11.84	21.55	66.61	82.74	17.26	$(NH_4,K)Cl$
12.91	19.63	67.46	84.86	15.14	$(NH_4,K)Cl$
14.05	18.09	67.86	86.96	13.04	$(NH_4,K)Cl$
15.73	15.30	68.97	90.99	9.01	$(NH_4,K)Cl$
17.40	13.11	69.49	92.87	7.13	$(NH_4,K)Cl$
19.90	9.94	70.16	95.06	4.94	$(NH_4,K)Cl$
22.06	6.85	71.09	96.95	3.05	$(NH_4,K)Cl$
22.96	5.24	71.80	98.16	1.84	$(NH_4,K)Cl$
24.03	3.63	72.34	100.00	0.00	KCl
25.68	1.82	72.50	100.00	0.00	KCl
27.34	0.00	72.66	100.00	0.00	KCl

3.2 Regression equation of activity product

By converting the mass fraction (wt%) of each component into molarity and regressing with Eq(17) to Eq(20), the relation between activity product of solid solution and molar fraction of NH_4Cl (P and Q) can be obtained. Eq (21) and Eq(22) are the regression results of FOA.

$$K_{a1} = 87.83P^3 + 11.83P^2 + 1.88P + 1.35 \tag{21}$$

$$K_{a2} = 310.84Q - 236.16 \tag{22}$$

in which the correlation coefficient of Eq(21) is 0.998 and the correlation coefficient of Eq(22) is 0.953. Regression results with MWMSFOA-SVR model are shown in Eq(23) and Eq(24).

$$K_{a1} = 23.33P^3 + 35.33P^2 - 0.39P + 1.38 \tag{23}$$

$$K_{a2} = 3479.14Q^2 - 5924.3Q + 2550.31 \tag{24}$$

where the correlation coefficient of Eq(23) is 0.999 and the correlation coefficient of Eq(24) is 0.985. The difference of correlation coefficients, the maximum relative error and the average relative error of two models are listed in Table 2.

Table 2: Correlation coefficients, maximum relative error and average relative error of two models

Algorithm	Correlation coefficient of K_{a1}	Correlation coefficient of K_{a2}	Maximum relative error, %	Average relative error, %
FOA	0.998	0.953	1.623	1.544
MWMSFOA-SVR	0.999	0.985	0.796	0.771

4. Conclusions

MWMSFOA-SVR model was constructed on the basis of FOA and SVR. Parameters of SVR such as the penalty factor, kernel function parameter and loss function were initialized as a fruit fly group, and then the group proceeded to iterate and forage based on FOA fitness optimum to search for the optimal parameter of SVR. From this process, FOA-SVR model was established. MWMSFOA-SVR was then constructed by introducing the modifying weights multi-strategies with FOA-SVR model. Moreover, FOA and MWMSFOA-SVR models were compared in the regression of activity products of $(K,NH_4)Cl$ and $(NH_4,K)Cl$ solid solutions. Result showed that the error of both models is small, while MWMSFOA-SVR regression is more accurate than that of FOA model.

Acknowledgments

References

Pan W.T., 2012, A New Fruit Fly Optimization Algorithm: Taking the Financial Distress Model as An Example, Knowledge-Based Systems, 26, 69-74.

Hazim I., Mesut G., 2014, Parameter Analysis on Fruit Fly Optimization Algorithm, Journal of Computer and Communication, 2, 137-141.

Huo H.H., 2015, Research on Fruit Fly Optimization Algorithm and its Applications, Master's thesis, Taiyuan University of Technology, China.

Niu J.W., Zhong W.M., Liang Y., Luo N., Qian F., 2015, Fruit Fly Optimization Algorithm Based on Differential Evolution and Its Application on Gasification Process Operation Optimization, Knowledge-Based System, 88, 253-263.

Wu L.H., Zuo C.L., Zhang H.Q., 2015, A Cloud Model Based Fruit Fly Optimization Algorithm, Knowledge-Based Systems, 89, 603-617.

Wang L., Liu R., Liu S., 2016, An Effective and Efficient Fruit Fly Optimization Algorithm with Level Probability Policy and its Applications, Knowledge-Based Systems, 97, 158-174.

Li M.S., Chen H.L., 2016, Evolving Support Vector Machines using Fruit Fly Optimization for Medical Data Classification, Knowledge-Based System, 96, 61-75.

Jiang W.J., Shi Y.B., Zhao W.J., Wang X.X., 2016, The Parameters Identification of Magnetic Core Using Fruit Fly Optimization Algorithm, Chemical Engineering Transactions, 51, 169-174.

Zhao W., Tao, T., Zio E., 2013, Parameters Tuning in Support Vector Regression for Reliability Forecasting, Chemical Engineering Transactions, 33, 523-528.

Zhang X.G., 2000, Introduction to Statistical Learning Theory and Support Vector Machines, Acta Automatica Sinica, 26, 32-42.

Wang L., Feng Q., 2016, Vector Machines Regression Applied in Penicillin Fermentation Process Control, Chemical Engineering Transactions, 51, 1339-1344.

Niu D.X., Ma T.N., Liu B.Y., 2017, Power Load Forecasting by Wavelet Least Squares Support Vector Machine with Improved Fruit Fly Optimization Algorithm, Journal of Combinatorial Optimization, 33, 1122-1143.

Pitzer K.S., 1973, Thermodynamics of Electrolytes-I, Journal of Physical Chemistry, 77, 268-277.

Zhong Y.L., Chen J.X., Su M., Han J., Wu Q., 2017, Solid-Liquid Equilibrium, Crystal Type, Solid Solubility and Thermal Stability Studies of Potassium Ammonium Chloride Solid Solution, Fluid Phase Equilibria, 439, 24-30.

Xue C.Y., Zhao B., Guo H.F., Dou S.Y., Cao J.L., 2016, Phase Equilibrium of the Quaternary System K_2SO_4-$MgSO_4$-$(NH_4)_2SO_4$-H_2O at 25°C, Fluid Phase Equilibria, 408, 115-122.