Knowledge-Based Systems 26 (2012) 69–74

Contents lists available at SciVerse ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

A new Fruit Fly Optimization Algorithm: Taking the financial distress model as an example

Wen-Tsao Pan *

Department of Information Management, Oriental Institute of Technology, Taiwan, ROC

ARTICLE INFO

Article history:
Received 29 May 2011
Received in revised form 29 June 2011
Accepted 1 July 2011
Available online 8 July 2011

Keywords:
Fruit Fly Optimization Algorithm
Financial distress
Optimization problem
General Regression Neural Network
Data mining

ABSTRACT

The treatment of an optimization problem is a problem that is commonly researched and discussed by scholars from all kinds of fields. If the problem cannot be optimized in dealing with things, usually lots of human power and capital will be wasted, and in the worst case, it could lead to failure and wasted efforts. Therefore, in this article, a much simpler and more robust optimization algorithm compared with the complicated optimization method proposed by past scholars is proposed; the Fruit Fly Optimization Algorithm. In this article, throughout the process of finding the maximal value and minimal value of a function, the function of this algorithm is tested repeatedly, in the mean time, the population size and characteristic is also investigated. Moreover, the financial distress data of Taiwan's enterprise is further collected, and the fruit fly algorithm optimized General Regression Neural Network, General Regression Neural Network and Multiple Regression are adopted to construct a financial distress model. It is found in this article that the RMSE value of the Fruit Fly Optimization Algorithm optimized General Regression Neural Network model has a very good convergence, and the model also has a very good classification and prediction capability.

1. Preface

In recent years, the treatment of optimization problems has caught everyone's attention, for example, the production schedule optimization problem [1,9], the shortest route problem in the logistics industry [13,12] or the shift arrangement problem in the traffic transport industry [7,22], etc. should all be treated by the optimization algorithm. Until now, algorithms that are commonly used to treat optimization problems and data mining include: regression and classification [18–21], genetic algorithm [14,2], Ant Colony Optimization Algorithm [6,4] and Particle Swarm Optimization Algorithm [8,25]. However, the common disadvantages of these algorithms are complicated computational processes, difficulty of understanding for beginners and there are very many parameters in GA-base methods.

Therefore, this article author Pan [16] proposes a new optimization algorithm, which is called the Fruit Fly Optimization Algorithm or Fly Optimization Algorithm (abbreviated as FOA). Such an optimization algorithm has advantages such as a simple computational process, ease of transformation of such concept into program code and ease of understanding, etc. In this article, the way of finding maximal value and minimal value of function is first used to test repeatedly, the function of this optimization algorithm. In the meantime, the correlation between the population size and the optimization capability of the fruit fly group is investigated. Furthermore, this article refer to Ravisankar [17], Li [15], Cho [5] and Hadavandi [10] adopts the financial distress data of enterprises with stocks listed in the regular stock market and over-the-counter stock market from the years 2003 to 2004 as the test data. Meanwhile, fruit fly algorithm optimized General Regression Neural Network (abbreviated as FOAGRNN), General Regression Neural Network (abbreviated as GRNN) and Multiple Regression (abbreviated as MR) are used to set up a financial distress model. Finally, the classification prediction capabilities of these three models are compared to verify the feasibility of applying the Fruit Fly Optimization Algorithm in real cases.

The main structure of this article is as follows: The first section introduces the research motivation and objective of this article. The second section introduces the Fruit Fly Optimization Algorithm proposed by this article and the test of applying it in finding minimal and maximal value. The third section introduces the sample data used by this article and the real case analysis. The fourth section proposes the research conclusion and suggestions.

2. A new Fruit Fly Optimization Algorithm

2.1. The basic concept of the Fruit Fly Optimization Algorithm

The Fruit Fly Optimization Algorithm (FOA) is a new method for finding global optimization based on the food finding behavior of the fruit fly. The fruit fly itself is superior to other species in sensing

* Address: 3F., No.12, Lane 271, Longjiang Rd., Jhongshan District, Taipei City 104, Taiwan, ROC.

E-mail address: teacherp0162@yahoo.com.tw

0950-7051/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.knosys.2011.07.001

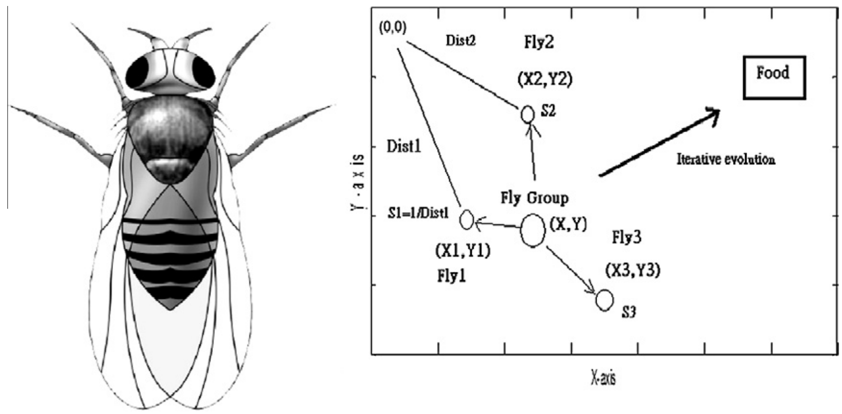

Fig. 1. Illustration of the body look of the fruit fly and group iterative food searching of fruit fly.

and perception, especially in osphresis and vision, which is as shown in Fig. 1. The osphresis organs of fruit flies can find all kinds of scents floating in the air; it can even smell food source from 40 km away. Then, after it gets close to the food location, it can also use its sensitive vision to find food and the company's flocking location, and fly towards that direction too.

In this article, based on the food finding characteristics of the fruit fly, it is divided into several necessary steps and a program example to be referred to by the readers. The steps are:

(1) Random initial fruit fly swarm location is shown in the figure to the right of Fig. 1.

$InitX\_axis; InitY\_axis$

(2) Give the random direction and distance for the search of food using osphresis by an individual fruit fly.

$X_i = X\_axis + RandomValue$

$Y_i = Y\_axis + RandomValue$

(3) Since the food location cannot be known, the distance to the origin is thus estimated first (Dist), then the smell concentration judgment value (S) is calculated, and this value is the reciprocal of distance.

$Dist_i = \sqrt{X_i^2 + Y_i^2}$

$S_i = 1/Dist_i$

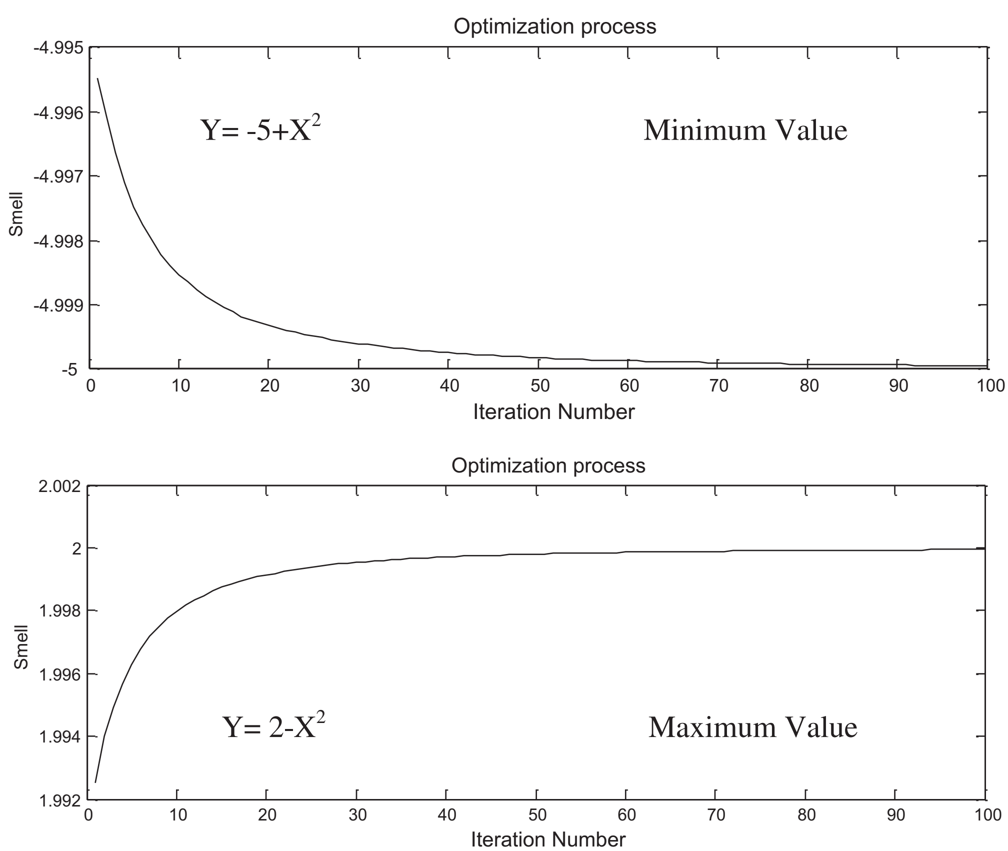

Fig. 2. Curve of extreme value solution of an iterative search solution.

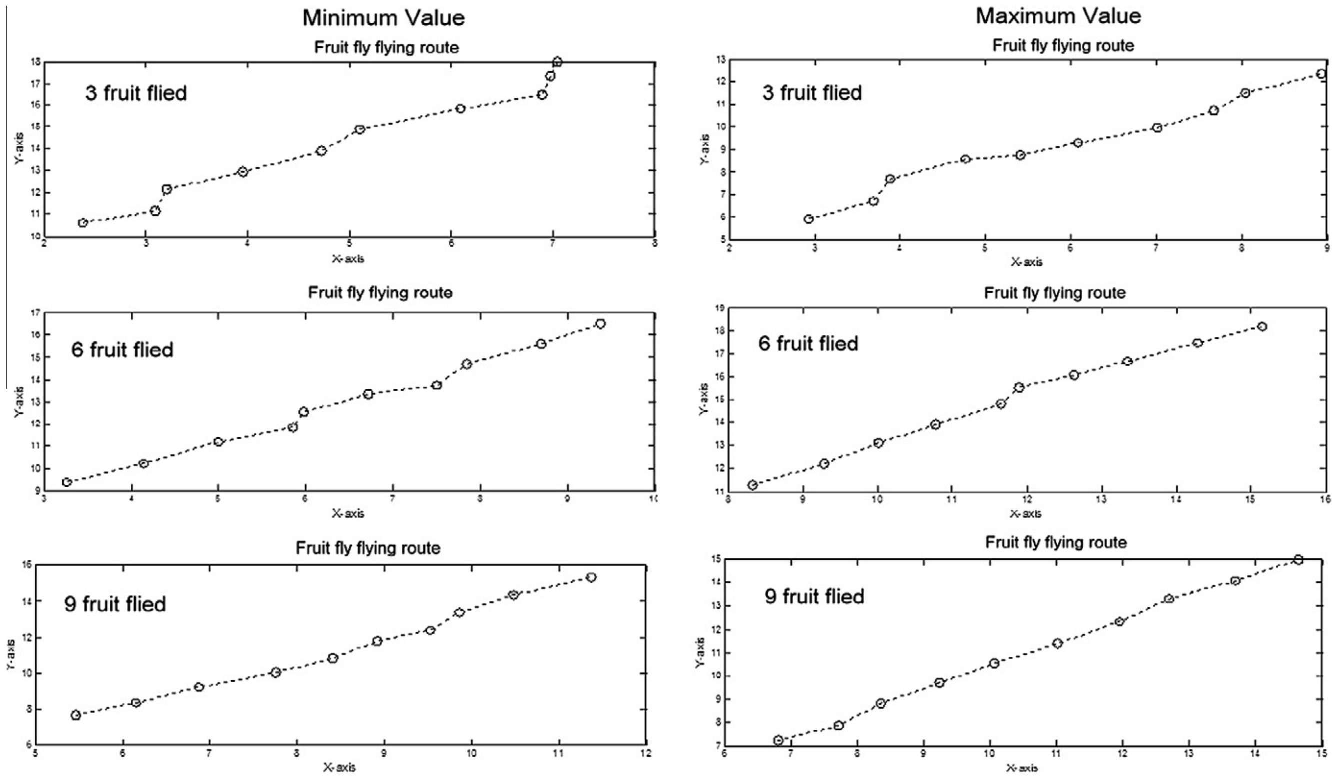

Fig. 3. The search route chart for different numbers of fruit flies.

(4) Substitute smell concentration judgment value (S) into smell concentration judgment function (or called Fitness function) so as to find the smell concentration (Smell_i) of the individual location of the fruit fly.

$$Smell_i = Function(S_i)$$

(5) Find out the fruit fly with maximal smell concentration (finding the maximal value) among the fruit fly swarm.

$$[bestSmellbestIndex] = max(Smell)$$

(6) Keep the best smell concentration value and x, y coordinate, and at this moment, the fruit fly swarm will use vision to fly towards that location.

$$Smellbest = bestSmell$$

$$X\_axis = X(bestIndex)$$

$$Y\_axis = Y(bestIndex)$$

(7) Enter iterative optimization to repeat the implementation of steps 2–5, then judge if the smell concentration is superior to the previous iterative smell concentration, if so, implement step 6.

2.2. Finding minimal value and maximal value example

In this article, FOA is used to get minimal value and maximal value with functions of respectively:

$$Y = -5 + X^2$$

$$Y = 2 - X^2$$

and in the former one, the solution of minimal value is -5, and in the latter, the solution of the maximal value is 2. The random initialization fruit fly swarm location zone is [0, 10], the random fly direction and distance zone of iterative fruit fly food searching is $[-1, 1]$. After 100 times of iterative search of minimal value and maximal value, the program implementation result will gradually approach the solution of the functional extreme value. Fig. 2 is the curve drawn by the solution of the extreme value of iterative search function. From the above figure, it can be seen that the curve gradually approaches a functional minimal value of -5, and the coordinate of the swarm of the fruit fly is (76.9049, 82.8723); from the following figure, it can be seen that the curve gradually approaches a functional maximal value of 2, and the coordinate of the swarm of the fruit fly is (80.4769, 75.2152). Further investigation will be performed in this section.

In this article, how the number of fruit flies affects the food searching efficiency will be investigated. From Fig. 3, it can be found that after 10 iterative searches, no matter in the search of minimal value or maximal value, the swarm with fewer fruit fly numbers will have disadvantages of an unstable search route and a slower convergence speed; but the advantage is the execution speed will be faster. On the contrary, swarms with a large number of fruit flies will have the advantage of a stable search route, and the convergence speed will be faster too; but the disadvantage is the program execution speed will be slower. Therefore, the researchers must consider the complexity of the optimization problem to select appropriately, the number of fruit flies to perform the treatment for the problem.

3. Case analysis

3.1. Sample data and variable

Financial distress historical data of 100 companies with stocks listed in the regular stock market and over-the-counter stock market from 2003 to 2004 are adopted as test data. Meanwhile, three financial distress models such as fruit fly algorithm optimized General Regression Neural Network (abbreviated as FOAGRNN),

Table 1
The descriptive statistical values of the financial distress test data from 2003 to 2004.

Factor	2003				2004			
	Max	Min	Avg	Std	Max	Min	Avg	Std
X1	3798.39	−99.66	53.455	390.3753	1645.88	−99.41	54.0975	209.1663
X2	296.67	−99.78	3.3369	46.02071	918.9	−85.64	6.9279	95.54782
X3	57.79	−4344.32	−67.8833	459.7565	55.9	−3671.59	−56.4762	379.4832
X4	197.05	−4641.7	−114.638	643.7703	2019.77	−23589.1	−267.093	2380.936

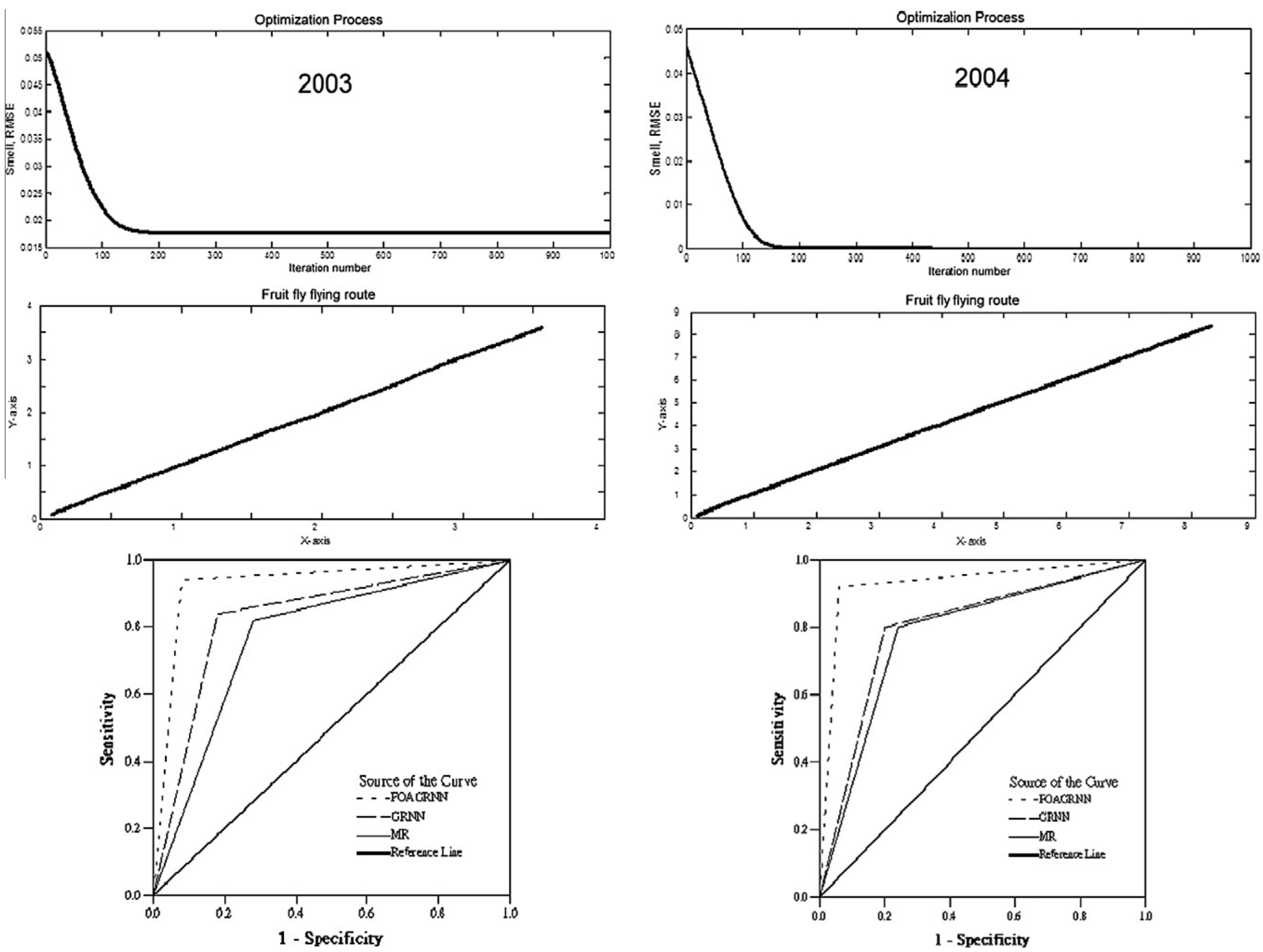

Fig. 4. The iterative RMSE trend of FOAGRNN search of optimization parameter; ROC curve of fruit fly swarm location coordinate route and classification prediction result.

General Regression Neural Network (abbreviated as GRNN) and Multiple Regression (abbreviated as MR) models are set up respectively to observe the efficacy of fruit fly algorithm optimization. The financial distress historical data includes the growth force and profitability force of the five financial forces. The four financial ratio index independent variables include revenue growth rate (X1), fixed-asset growth rate (X2), operating profit margin (X3) and profit margin (X4), etc. The dependent variable Y is 50 companies in risk (represented by 1) and the paired 50 normal companies (represented by 0), which is used to construct and test the sample data of fruit fly algorithm optimized General Regression Neural Network (abbreviated as FOAGRNN), General Regression Neural Network (abbreviated as GRNN) and Multiple Regression (abbreviated as MR). Table 1 shows the descriptive statistical values of four financial ratio indexes.

3.2. The construction of three financial distress classification prediction models

In this article, 100 sample data are normalized to make samples in the range from 0 to 1. It is then divided all sample data equally into five groups with each contains 20 data rows. Four groups of data used as training data to construct the model, and one group of data used as test data to test model stability. Thus, cross verification is accomplished. The output result of three models is defined as when it is smaller than or equal to 0.5, it is classified as 0 (that is, it is normal company); when it is larger than 0.5, it is classified as 1 (that is, company in risk). First, training data of four groups are used to construct three models of FOAGRNN, general GRNN and traditional Multiple Regression, respectively. For GRNN theories, the readers can refer to the related publications of professor [23,24]. In the FOAGRNN model, this article has adopted the GRNN matlab toolbox and self-written MATLAB program. This article has also designed the spread parameter value of the FOA dynamic search of optimized GRNN. And finally, the FOAGRNN model is also constructed in this article. The initial value of the spread parameter of GRNN is set up in the range of [0.01, 1], and the neuron number of the network input layer is 4, and the output neuron number is one.

In the initial parameter set up of the FOA, the random initialization fruit fly swarm location range is [0, 100], the random fly direction and distance zone of iterative fruit fly food searching is [−10, 10], fruit fly population size is 50, and iterative number is 1000. The way used in the FOA optimized GRNN is to first calculate the distance between the individual location of the fruit flies and the origin coordinate (0, 0), and then calculate the reciprocal. The smell concentration judgment value (S) is then calculated, and then substituted into the parameter spread of GRNN. After that, the training data is entered to get the network output value. Then with the target value, RMSE (or called Fitness) is calculated, the smaller the value, the better. Finally, the best smell concentration judgment value (S) is kept to be used as the spread value of GRNN, and the iterative search is made based on this method. Through the smell of the fruit fly's random food finding, and through the flocking at the location of the highest concentration of smell using vision, the spread value of GRNN can be adjusted to its optimal value, and the RMSE between network output value and target value can be adjusted to the minimal value.

The upper four figures in Fig. 4 represent the training data of 2003 and 2004, respectively. FOA is used to iteratively optimize the spread value of GRNN, the RMSE between generated output value and target value, and the trend of the fruit fly swarm flying route. After 1000 times of iterative evolution, in 2003, convergence can be seen in generation 183, with a coordinate of (3.5554, 3.5932), and the spread value and RMSE value is respectively [0.1978, 0.0177]; in 2004, convergence starts from generation 187 with a coordinate of (1.5845, 1.6680), spread value and RMSE value is respectively [0.4347, 8.5498e−086]. However, the down two figures in Fig. 4 show that for the sample data from 2003 to 2004, four groups of the data are used as training data to construct the model, and one group of data is used as test data to cross-verify the results generated by three models such as FOAGRNN, general GRNN and traditional Multiple Regression model, and the results are drawn into an ROC curve. Bradley [3] pointed out that the larger the area above the reference line and underneath the curve, the better the classification capability of the model. From the figure, it can be clearly seen that the FOAGRNN model has the best classification capability. Then, from an observation of the ROC curve analysis output result through Table 2, we see that sensitivity (Sen) means the percentage of the number of the prediction result of 1 (that is, the company in risk) to the number of the actual value of 1. Specificity (Spe) means the percentage of the number of the prediction result of 0 (that is, the normal company) to the number of the actual value of 0; and professor [11] pointed out that Gini Index is 2 $\times$ AUC – 1. Here, these index values are the larger the better. As shown in the table, for 2003 FOAGRNN model, the Sen is 0.940, Spe is 0.920, area under the curve (AUC) is 0.930, and Gini Index is 0.860; for 2004 FOAGRNN model, the Sen is 0.920, Spe is 0.940, the area under curve (AUC) is 0.930, and the Gini Index is 0.860, which are all higher than those of the general GRNN and traditional Multiple Regression model. Therefore, the FOAGRNN model has a very good classification prediction capability.

Table 2
ROC curve analytic index value of the test results of three models.

	Model	Sen	Spe	Auc	Gini
2003	FOAGRNN	0.940	0.920	0.930	0.860
	GRNN	0.840	0.820	0.830	0.660
	MR	0.820	0.720	0.770	0.540
2004	FOAGRNN	0.920	0.940	0.930	0.860
	GRNN	0.800	0.800	0.800	0.600
	MR	0.800	0.760	0.780	0.560

4. Conclusion

The major contribution of this article is to propose a new Fruit Fly Optimization Algorithm (FOA) with a real application in finding maximal value and minimal value. From the test result, it is found that the FOA can usually find solutions correctly, and the stability of the fruit fly swarm search route is obviously related to fruit fly quantity. In addition, this article further adopts financial distress historical data in the Taiwan area to test the optimization capability of the FOA. From the analysis result, it can be seen that through the FOA, the spread value of the GRNN network parameter can be optimized, and the classification prediction capability of the GRNN can obviously be enhanced. This algorithm not only has the feature of being easy to understand, but also is easy to be written into program code. Meanwhile, the program code, as compared to other algorithms, is not too long, and it is thus easy to be used to handle all kinds of optimization problems. However, the brand new algorithm as proposed by the FOA for this article might not be thorough, and we thus hope that more researchers can participate in the promotion and test.

References

[1] D.R. Anderson, C.L. Moodie, Optimal buffer storage capacity in production line, Int. J. Prod. Res. 7 (1969) 233–240.
[2] Y.D. Bertrand, D. Barba, Feature selection by a genetic algorithm application to seed discrimination by artificial vision, J. Science Food Agric. 76 (1998) 77–86.
[3] A.P. Bradley, The use of the area under the ROC curve in the evaluation of machine learning algorithms, Pattern Recognit. 30 (7) (1997) 1145–1159.
[4] B. Bullnheimer, R. Hartl, C. Strauss, An improved ant system algorithm for the vehicle routing problem, Ann. Oper. Res. 89 (1999) 319–328.
[5] V. Cho, MISMIS – A comprehensive decision support system for stock market investment, Knowl. Based Syst. 23 (6) (2010) 626–633.
[6] M. Dorigo, L.M. Gambardella, Ant colony system: a cooperative learning approach to the travelling salesman problem, IEEE Trans. Evol. Comput. 1 (1) (1997) 53–66.
[7] M. Friedman, A mathematical programming model for optimal scheduling for buses departure under deterministic condition, Transp. Res. 10 (2) (1976) 83–90.
[8] Y. Fukuyama, H. Yoshida, 2001. A particle swarm optimization for reactive power and voltage control in electric power systems. Congress Evolutionary Computation 1, 87–93.
[9] S.B. Gershwin, An efficient decomposition method for the approximate evaluation of tandem queues with finite storage space and blocking, Oper. Res. 35 (1987) 291–305.
[10] E. Hadavandi, H. Shavandi, A. Ghanbari, Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting, Knowl. Based Syst. 23 (8) (2010) 800–808.
[11] D.J. Hand, R.J. Till, A simple generalisation of the area under the ROC curve to multiple class classification problems, Mach. Learn. 45 (2) (2001) 171–186.
[12] P.E. Hart, N.J. Nilsson, B. Raphael, A formal basis for the heuristic determination of minimum cost paths, IEEE Trans. Syst. Sci. Cybern. 4 (2) (1968) 100–107.
[13] N. Koncz, J. Greenfield, K. Mouskos, A strategy for solving static multiple-optimal-path transit network problems, J. Transp. Eng. 122 (3) (1996) 218–225.
[14] Y.W. Leung, Y. Wang, An orthogonal genetic algorithm with quantization for global numerical optimization, IEEE Trans. Evolution. Comput. 5 (2001) 41–53.
[15] H. Li, J. Sun, Ranking-order case-based reasoning for financial distress prediction, Knowl.Based Syst. 21 (8) (2008) 868–878.
[16] W.T. Pan, 2011. A new evolutionary computation approach: Fruit Fly Optimization Algorithm, 2011 Conference of Digital Technology and Innovation Management, Taipei. Program code on the website <http://www.oitecshop.byethost16.com/FOA.html>.
[17] P. Ravisankar, V. Ravi, Financial distress prediction in banks using Group Method of Data Handling neural network, counter propagation neural network and fuzzy ARTMAP, Knowl. Based Syst. 23 (8) (2010) 823–831.
[18] Roman M. Balabina, Sergey V Smirnov, Variable selection in near-infrared spectroscopy: benchmarking of feature selection methods on biodiesel data, Anal. Chim. Acta 692 (1–2) (2011) 63–72.

[19] Roman M. Balabin, Ravilya Z. Safieva, Biodiesel classification by base stock type (vegetable oil) using near infrared spectroscopy data, Anal. Chim. Acta 689 (2) (2011) 190–197.
[20] Roman M. Balabin, Ekaterina I. Lomakina, Support vector machine regression (SVR/LS-SVM)—an alternative to neural networks (ANN) for analytical chemistry? Comparison of nonlinear methods on near infrared (NIR) spectroscopy data, Analyst 136 (2011) 1703–1712.
[21] Roman M. Balabin, Ravilya Z. Safieva, Ekaterina I. Lomakina, Near-infrared (NIR) spectroscopy for motor oil classification: from discriminant analysis to support vector machines, Microchem. J. 98 (1) (2011) 121–128.
[22] F.J.M. Salzborn, Optimum bus scheduling, Transp. Sci. 6 (2) (1972) 137–148.
[23] D.F. Specht, Probabilistic neural networks and the polynomial adaline as complementary techniques for classification, IEEE Trans. Neural Netw. 1 (1) (1990) 111–121.
[24] D.F. Specht, A general regression neural network, IEEE Trans. Neural Netw. 2 (6) (1991) 568–576.
[25] D. Srinivasan, W.H. Loo, R.L. Cheu, 2003. Traffic incident detection using particle swarm optimization. Swarm Intelligence Symposium, Proceedings of the 2003 IEEE, pp.144–151.