

A New Algorithm for Detection of Animal and Plant Ion Concentration Based on Gene Expression Programming

Kangshun Li, Dongguan City College, China*

Leqing Lin, South China Agricultural University, China

Jiaming Li, South China Agricultural University, China

Siwei Chen, South China Agricultural University, China

Hassan Jalil, South China Agricultural University, China

ABSTRACT

In order to accurately predict the concentration detection data of ion sensors for animal and plant, this paper proposes a gene expression programming (GEP) based concentration detection method. The method includes collecting ion concentration data as well as voltage timing data; preprocessing all the collected data to obtain an initial sample set; constructing a prediction model of ion concentration, which is an explicit functional relationship between voltage and the concentration of a specific ion. The Gene Expression Programming is used to train and evaluate the prediction model, and obtain a trained model. By comparing gene expression programming with other two modeling methods, it is found that the accuracy of the model established by gene expression programming has greater advantages than that established by polynomial fitting and neural network in processing animal and plant ion concentration data.

KEYWORDS

Gene Expression Programming, ion concentration, prediction model

INTRODUCTION

Inorganic ions are an important component of animals and plants and have the function of regulating physiological activities, which is a necessary condition for maintaining the normal growth of animals and plants. Obtaining the exchange and absorption information of inorganic ions in the environment during the growth of animals and plants is helpful to reveal the life activity mechanisms of animals and plants, assist agricultural scientific research, and promote agricultural production. In order to dynamically and accurately monitor the changes in ion content around the plant growth environment, the most suitable growth environment conditions for the plant can be obtained by comparing the

DOI: 10.4018/IJCNIN.318144

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

growth status of the plant in different concentration environments of specific ions. As a result, it can provide a method for large-scale vegetable and fruit farms to accurately measure the suitable plant growth environment, thereby improving the yield and quality of vegetables and fruits. The authors conducted a large number of detection experiments on standard solution samples and mixed solution samples of eight ions, including Mn^{2+} , Ca^{2+} , K^+ , Na^+ , Mg^{2+} , Cl^- , NH_4^+ , and NO_3^- , and obtained a large amount of data. The authors will use gene expression programming to model the ion-selective membrane properties of these data detection results. On this basis, pattern recognition technology is used to achieve the automatic identification of these ions as well as the automatic concentration reading.

GENE EXPRESSION PROGRAMMING

Gene expression programming was created by Portuguese scientist Ferreira. It is a development of GA (genetic algorithm) and GP (genetic programming design). GEP combines the advantages of both, including the simple and fixed-length coding characteristics of GA, and the indefinite length and indefinite characteristics of the tree structure in GP. Therefore, gene expression programming is much faster than GA or GP.

Gene expression programming processes chromosomes, which consist of genes connected by linking functions. A gene consists of a head and a tail, and the head contains function sets and terminals, while the tail contains only terminals. Thereinto:

$$t = h \times (n - 1) + 1 \tag{1}$$

where t represents the length of the gene tail, h represents the length of the gene head, and n denotes the maximum number of parameters in the function set.

K-Expressions

Chromosomes are made up of one or more fixed-length, linear, equal-length genes, so genes are also linear and fixed-length. Chromosomes can determine the size and shape of the expression tree. For example, in this simple algebraic equation:

$$\sqrt{(a - b) \times (c - d)} \tag{2}$$

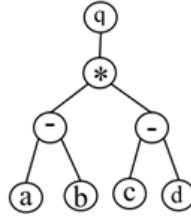
The expression tree of this equation is shown in Figure 1, where q represents the square root. Traversing the expression tree in Figure 1 from top to bottom and left to right yields the corresponding K-expression, as shown in equation (3). The genotype in gene expression programming is:

$$q * - a b c d 1 2 3 4 5 6 7 8 \tag{3}$$

Fitness Function

In order to obtain the best solution, the newly produced chromosomes need to be evaluated for environmental adaptability. In the application of functional modeling, the final result is to find an expression that fits the sample data well within the limits of the error. The evaluation of the pros and cons of the expression is based mainly on the degree of consistency between the results of the expression calculation and the training data. In gene expression programming, the coefficient of determination R^2 , root mean square error $RMSE$, mean absolute error MAE , etc. are usually used as evaluation criteria.

Figure 1.
The Expression Tree for Equation (2)



Evolutionary Operators

Replicate: Chromosomes are copied to the next generation based on the probability of selection for fitness.

Mutate: Variation can occur anywhere on a chromosome. However, the organizational structure of chromosomes must remain intact. At the head of a gene, any symbol can become another symbol (a function or endpoint); at the end of the gene, it can only become an endpoint.

Transpose: 1) Translocation of short fragments with functions and endpoints to the head of the gene, except for the root (the root intersection element or RIS element); 2) Short fragments with functions that translocate to the root of the gene (root insertion element or RIS element); 3) Translocation of the entire gene to the beginning of the chromosome.

Recombine: 1) Single-point recombination, that is two chromosomes exchange all codes after a randomly selected point, producing offspring chromosomes; 2) Two-point recombination, in which two paired chromosomes exchange all codes between randomly selected intersections to obtain two new chromosomes; 3) Gene recombination, in which two chromosomes randomly select a gene exchange, producing offspring chromosomes.

The Basic Steps of Gene Expressions Programming

The basic steps of standard gene expression programming can be summarized as follows:

Step1: Initialize parameter settings and randomly generate population individuals; the relative errors or absolute errors in the fitness function are used to solve the fitness of each individual in the population, and then it is judged whether there are individuals who have reached the set conditions of the adaptation value or whether the number of population iterations has reached the set evolutionary iteration value.

If the conditions are met, the final result will be output; otherwise, it will go to the next step.

Step2: According to the set probability, the mutated individual p is arbitrarily selected, and the original element of the position is transformed by the operation position of p .

Step3: Transpose IS, RIS, and genes; select individuals according to transposable probability, and perform gene fragment transfer operations.

Step4: Recombine one-point, two-point, and gene, and then randomly select individuals for corresponding recombination operations according to the recombination rate.

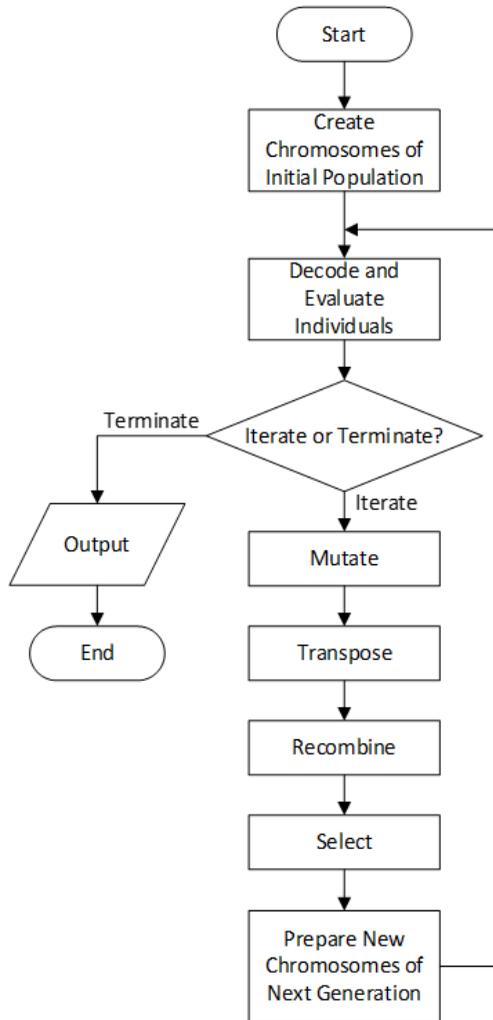
Step5: Select the individuals who enter the next iteration according to the selection operator.

The chromosome undergoes genetic manipulation and continuous iteration, until it finally obtains the most suitable individual, which is the optimal solution. The flowchart is shown in Figure 2.

Cross-Mutation Operator Parameters Settings

The parameters in the gene expression programming are: population size N , maximum evolutionary algebra ME , function set F , terminal set T , number of genes NG , head length HL , link function

Figure 2.
 Basic Steps of Gene Expression Programming Algorithms



LF , mutation rate MR , one-point recombination rate OR , two-point reorganization rate TR , gene recombination rate GRR , IS transposition rate ISR , IS element length ISL , RIS transposable rate R , RIS element length RL , gene transposition rate GTR , the strategy of choice SC , and the fitness function FIT .

Its initial parameter settings are as shown in Table 1.

METHOD AND PROCEDURE OF CONCENTRATION DETECTION OF ANIMAL AND PLANT ION SENSOR BASED ON GENE EXPRESSION PROGRAMMING

The concentration detection method of animal and plant ion sensors based on gene expression programming consists of the following steps:

Step1: Collect ion concentration data and ion voltage timing data.

Step2: Preprocess all the collected data to obtain the initial sample set.

Table 1.
Gene Expression Programming Algorithm Parameter Setting

Parameter name	Symbol	The parameter values
Population size	N	60
Maximum evolutionary algebra	ME	800
Function set	F	{ '+', '-', '*', '/', 'sin', 'cos', 'tan', 'log', 'Q' }
Terminal set	T	{ a }
Number of genes	NG	2
Head length	HL	15
Link function	LF	{ '+', '-', '*', '/' }
Mutation rate	MR	0.044
One-point recombination rate	OR	0.1
Two-point reorganization rate	TR	0.1
Gene recombination rate	GRR	0.1
IS transposition rate	ISR	0.3
IS element length	ISL	{ 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15 }
RIS transposable rate	R	0.3
RIS element length	RL	{ 1,2,3,4,5,6,7,8,9,10,11,12,13,14 }
Gene transposition rate	GTR	0.1
The strategy of choice	SC	championship
Fitness function	FIT	relative error

Step3: Build a prediction model of ion concentration that is an explicit function of voltage and the concentration of a specific ion.

Step4: Use gene expression programming to train the ion concentration prediction model and obtain the trained ion concentration prediction model.

Step5: Write the trained prediction model into the animal and plant ion sensors, and detect the concentration of specific ions in the solution through the animal and plant ion sensors. The above steps can be summarized in Figure 3.

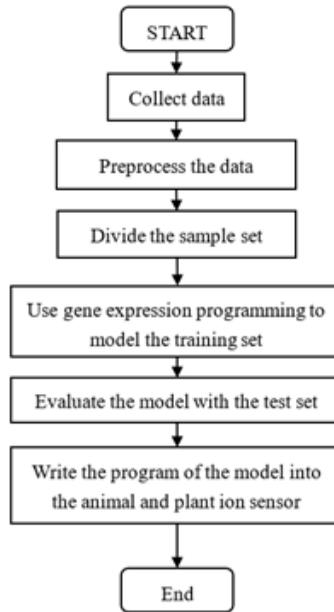
Collect Ion Concentration Data and Voltage Timing Data for Ions

The voltage timing data of ions is collected in liquid based on animal and plant ion sensors based on microelectrophoresis and capacitively coupled non-contact conductivity detection technology. Each set of timing data is measured with a sensor in a glass of solution of a known concentration, and the timing data for each set of ion voltages corresponds to a known concentration.

Preprocessing Preprocesses All the Collected Data to Obtain an Initial Sample Set

The pretreatment refers to calculating the average value of the voltage value in each group of voltage timing data after using outlier diagnosis to remove outliers, replacing the average values of the group of time series data, and forming a data pair of the group of time series data with a known concentration; the initial sample set consists of multiple pairs of data.

Figure 3. Detection Method and Procedure Flowchart



Evaluate The Trained Ion Concentration Model

The performance of the prediction is evaluated by using three indicators: the coefficient of determination R^2 , the root mean square error $RMSE$, and the mean absolute error MAE . The calculation formulas of R^2 , $RMSE$, and MAE are shown below:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2} \quad (4)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (5)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (6)$$

where N is the number of sample groups in the dataset, y_i represents the actual data, \hat{y}_i represents the model prediction, and \bar{y}_i represents the average of the actual data.

The value range of R^2 is $[0,1]$, and the closer its value is to 1, the better the fitting effect. The root mean square error measures the deviation between the true values and predicted values, and the mean absolute error is used to evaluate the closeness of the predicted results to the actual data. The smaller the values of both, the better the fitting effect.

DATA EXPERIMENTAL ANALYSIS

In order to show the superiority of gene expression programming in processing animal and plant ion concentration data, polynomial fitting, neural networks, and gene expression programming are used

in this experiment to establish mathematical models. Moreover, the predicted values, Coefficient of Determination, Root Mean Square Error, and Mean Absolute Error of animal and plant ion concentrations will be calculated through the fitting model of the three algorithms. Finally, we conducted a comparative analysis of the models.

Three Algorithms Predict Concentrations

The authors of this paper used a large amount of data on the concentration of ions in animals and plants (parameter setting reference Table 1) to establish an animal and plant ion concentration model that is based on gene expression programming:

$$Y = 2 \times a^2 \times \{\sqrt{a} \times [a \times (a + 1) \times \ln 10 - \ln a] + \ln a\}^5 \times (\log_{10} a)^5 \quad (7)$$

where Y represents concentration and a represents voltage.

This paper uses this model to curve fit and predict the sample data. The results are then visually compared with the results of polynomial fitting and neural network fitting to the data and graphs. When performing experiments on polynomial fitting, a variety of models were obtained. The polynomial-fitted model in this article is the best of them. The process of polynomial fitting will not be described in this article.

Table 2 shows the fitted prediction values of the sample data and the three algorithms. As can be seen from Table 2, the predicted values obtained using gene expression programming are more accurate than polynomial fitting and neural network fitting. Moreover, the distribution of predicted values is uniform, which is more stable than the traditional polynomial fitting and the neural network fitting. In general, gene expression programming is more advantageous. There are three plots below: a polynomial fitting plot (Figure 4), a neural network fitting plot (Figure 5), and a gene expression programming fitting plot (Figure 6). They more intuitively show that curves fitted with gene expression programming are significantly closer to the real curve than those of the other two methods.

In order to further support the above statement, we calculated the root mean square error, squared absolute error, and coefficient of determination of the three fitted models, as shown in Table 3. From Table 3, it can be clearly seen that the root mean square error and squared absolute error calculated

Figure 4.
 Fitting Plot for Polynomial

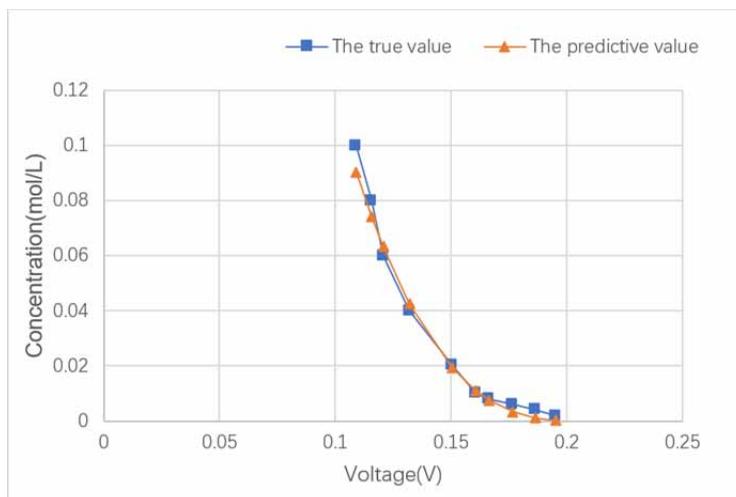


Figure 5.
Fitting Plot for Neural Network

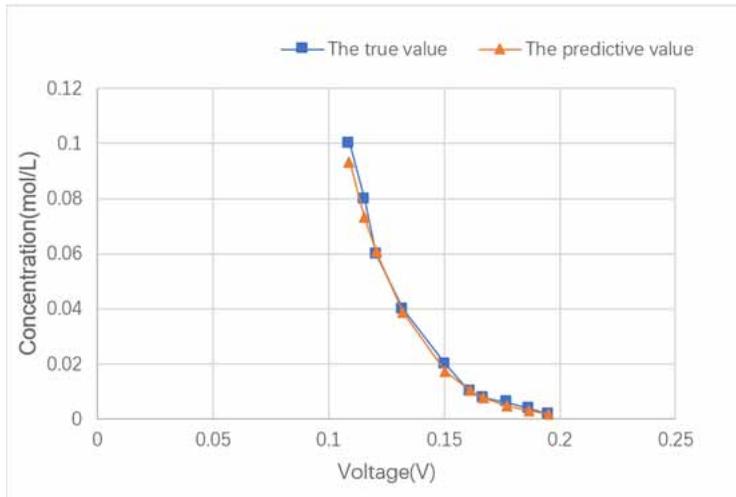
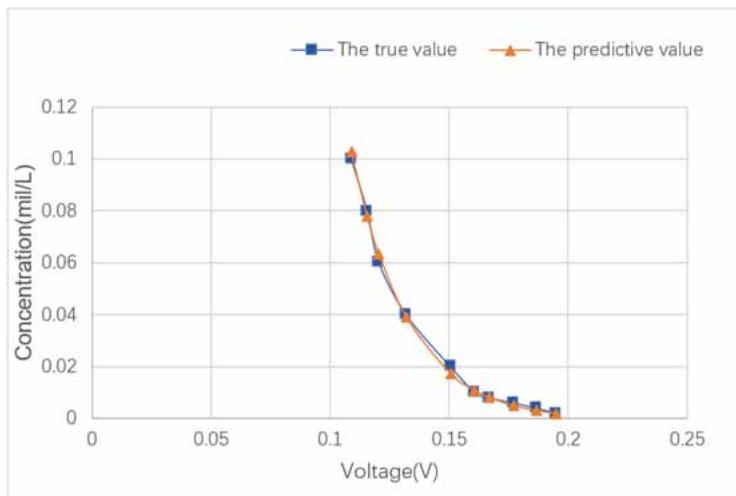


Figure 6.
Fitting Plot for Gene Expression Programming



by the gene expression programming model are much smaller than those calculated by the other two models. The R^2 calculated by the gene expression programming model is larger than the other two models. In summary, it can be shown that the model obtained by gene expression programming is much better than the other two models in terms of fit and prediction accuracy.

CONCLUSION

Based on the detection model of animal and plant ion concentrations, this paper researched the prediction of specific ions by introducing gene expression programming, polynomial fitting, and a neural network algorithm. After comparing and analyzing the three approaches, the following

Table 2. Summary of Fitted Data for the Three Algorithms

Voltage	The true values of concentration	The predicted values of the polynomial	The predicted values of the neural network	The predicted values of the GEP
0.194941	0.002	0.000297	0.002	0.001942
0.18656	0.004	0.001153	0.003	0.00305
0.176964	0.006	0.003353	0.0048	0.004987
0.166679	0.008	0.007593	0.008	0.008242
0.160958	0.01	0.010971	0.0106	0.010796
0.150597	0.02	0.019277	0.0173	0.017354
0.132045	0.04	0.042594	0.0387	0.039135
0.120632	0.06	0.063366	0.0608	0.063446
0.115698	0.08	0.074074	0.0732	0.077961
0.109029	0.1	0.09034	0.0932	0.102782

Table 3. Evaluation Index of the Fitted Data of the Three Algorithms

Algorithm	Evaluation criteria		
	<i>RMSE</i>	<i>MAE</i>	<i>R²</i>
Polynomial	0.00407591	0.00308449	0.98533716
Neural network	0.00323883	0.00212	0.9907414
Gene expression programming	0.00184794	0.00148378	0.99698597

conclusions were obtained. The model of animal and plant ion concentration data studied in this paper is more complex, and it is more advantageous to use gene expression programming to build the model than other approaches. It is shown that the ion concentration prediction model constructed by gene expression programming can more accurately predict the concentration of specific ions in solution. In turn, the detection of the ion concentration of animals and plants can be achieved with faster efficiency and better results.

ACKNOWLEDGMENT

This work was supported by Key Realm R&D Program of Guangdong Province with Grant No. 2019B020219003, Key Field Special Project of Guangdong Provincial Department of Education with No.2021ZDZX1029.

COMPETING INTEREST

Although the journal’s leadership is present within the authorship of this manuscript, the formal evaluation of this manuscript was performed by appointed members of the journal’s editorial review board and all subsequent revisions were made in consideration of the reviewer’s comments. The peer-review process was completed completely independently of the journal’s leadership.

REFERENCES

- Chen, W. J., Qiu, J., & Yang, J. (2018). Apparent temperature modeling based on improved GEP algorithm. *Kongzhi Gongcheng*, 25(1), 137–141.
- Cui, W., Wang, W., & Huang, Z. (2018). Forecasting models of high order ordinary differential equations based on GEP. *Computer Engineering and Applications*, 54(18), 256–262.
- Dai, W., & Zhang, Y. (2007). Application of GEP in the soft sensing model of translate percentage of carbon dioxide. *Computers and Applied Chemistry*, 24(10), 1371–1374.
- Ferreira, C. (2001a). Gene expression programming: A new adaptive algorithm for solving problems. *Complex Systems*, 13(2), 87–129. doi: doi:10.48550/arXiv.cs/0102027
- Ferreira, C. (2001b, September 10–24). *Gene expression programming in problem Solving* [Conference tutorial]. 6th Online World Conference on Soft Computing in Industrial Applications, Berlin, Germany.
- Gong, X. Y. (2020). Study on estimation of landslide motion information based on polynomial fitting. *Foundation and Basement*, 47(19), 155–157.
- Hu, X., Tu, Z., Luo, Y., Zhou, F., Li, Y., Liu, J., & Yi, P. (2022). Shale gas well productivity prediction model with fitted function-neural network cooperation. *Petroleum Science Bulletin*, 03, 394–405. doi:10.3969/j.issn.2096-1693.2022.03.034
- Hu, Z., & Yan, M. (2019). Optimal dispatching of new energy power grid with considering of uncertainty effect of wind power. *Henan Science and Technology*, 31, 130–132.
- Jiang, M., Lu, Q., & Zuo, Y. (2021). Target track maintenance based on polynomial fitting. *Ship Electronic Engineering*, 41(9), 83–85.
- Li, K. S., Huang, H. H., & Zhang, W. S. (2012). Prüfer code optimization algorithm for multi-layer logistic networks based on gene expression programming. *Xitong Fangzhen Xuebao*, 24(03), 594–602.
- Liang, D., Xie, L., & Sui, T. (2006). Scheduling optimization based on hybrid genetic-tabu search algorithm for dual-resource constrained job shop. [Natural Science]. *Journal of Northeastern University*, 27(8), 895–989.
- Liu, C. L., Liu, X. S., & Li, Y. Y. (2013). Deformation prediction model of metal mine tailings dam based on GEP. *Nonferrous Metals Science and Engineering*, 4(6), 63–68.
- Liu, J. X., Zhong, Y. Q., Qiu, L. Q., & Zhang, W. T. (2020). Optimization of pre-camber fitting curve for continuous rigid frame bridge construction. *Journal of Changjiang Institute of Technology*, 37(3), 4–8.
- Liu, M., Yuan, G., & Zhu, J. (2016). Gene expression programming algorithm for multi-objective spatial siting optimization. *Journal of Genomics*, 41(3), 82–85.
- Liu, Y. Q., Qie, Z. H., & Han, W. M. (2017). Fitting model optimization of the compaction test curve based on the AIC criterion. *Hebei Nongye Daxue Xuebao*, 40(5), 120–124.
- Syarif, A., & Gen, M. (2003). Solving exclusionary side constrained transportation problem by using a hybrid spanning tree-based genetic algorithm. *Journal of Intelligent Manufacturing*, 14(3), 389–399. doi:10.1023/A:1024610128238
- Tu, X. Q., & Lan, H. (2019). Gabor-based genetic algorithm for static gesture recognition. *Tongxin Jishu*, 52(10), 2395–2400.
- Wang, C., Jia, X., & Sun, J. (2021). Application of double system co-evolutionary gene expression programming in PM2.5 concentration prediction. *Jisuanji Celiang Yu Kongzhi*, 29(10), 71–75.
- Wang, C., Yang, Q., Li, J., Zhong, W., & Chen, Z. (2022). Fast demodulation method of optical fiber temperature and strain based on neural network. *Laser Technology*, 46(2), 254–259.
- Zhao, Y., Zhou, H., & Li, M. (2010, September 23–25). *Gene expression programming model for determining location in wireless LANs* [Conference paper]. International Conference on Computational Intelligence and Software Engineering, Chengdu, China. doi:10.1109/WICOM.2010.5601087

Kangshun Li received the B.Sc. degree in computational mathematics from Jiangxi University (currently Nanchang University), Nanchang, China, in 1983, and the Ph.D. degree in computer software and theory from Wuhan University, Wuhan, China, in 2006. He was a Post-Doctoral Research Fellow with the Institute of Automation, Chinese Academy of Sciences, Beijing, China, from 2007 to 2009. His current research interests include evolutionary computation and computer vision.

Lin Leqing, a major in Statistics of grade 2020, College of Mathematics and Information Science, South China Agricultural University.

Jiaming Li, a major in information management and information system of grade 2020, College of Mathematics and Information Science, South China Agricultural University.

Siwei Chen, a major in Information Management and Information System of grade 2020, College of Mathematics and Information Science, South China Agricultural University.

Hassan Jalil received the M.Sc degree in software engineering from the South China University of Technology, Guangzhou, China, in 2019. He is currently pursuing his Ph.D. degree in agricultural electrification and automation from the South China Agriculture University, Guangzhou, China. His current research interests include evolutionary algorithms, multi-objective optimization, computer vision, and image segmentation.