4th International Conference on Agricultural and Biological Sciences IOP Publishing
IOP Conf. Series: Earth and Environmental Science **185** (2018) 012013 doi:10.1088/1755-1315/185/1/012013

# Price simulation and prediction of leaf vegetable based on gene expression algorithm

**Z H Su**[1]

Institute of Agricultural Economics and Rural Development, GAAS, Guangzhou 510640, China

E-mail: suzhuhua@163.com

**Abstract.** Based on the traditional gene expression algorithm, the paper proposed an improved gene expression algorithm, adding "inverted string" and "gene extraction" operator. An improved GEP algorithm is used in the leafy vegetables prices forecast, create the mathematical model through the analysis and evolution of training data, realize the simulation and forecast of the price of leafy vegetables. Through several experiments, the results proved that the new type of gene expression programming algorithm has faster convergence rate and higher precision in predicting the price of leafy vegetables.

## 1. Introduction

Guangdong is a large vegetable consumption province, and has a strong demand for vegetables. There are a large number of vegetables that enter the market through various channels, and the quality of vegetable circulation is directly related to the quality of life of the residents. Therefore, how to understand and monitor vegetable circulation organizations in terms of industrial circulation quantity, price, quality and safety, and carry out a basic and objective evaluation of operational risk is a key problem to mitigate the risk of vegetable circulation industry.

From the research trend at home and abroad, the price forecast has made great progress in the fields of social economy, energy, transportation and other fields [1-5]. In the market price prediction of agricultural products, Henry made a regression prediction for the US cotton, and the forecast results coincide with the actual results at that time [6]. Cui Guoli found that the chaotic neural network model was more suitable on the price of Chinese cabbage [7]. Zhu Xiaoxia used Markov chain analysis to simulate the predictability of vegetable price fluctuation cycle [8]. To sum up, there are few studies on vegetable price prediction, and the existing methods of prediction are mainly based on econometrics.

Gene expression absorbs the advantages of the two and overcomes the shortcomings of the two ways, and its remarkable feature is that it can be used to solve complex problems by simple coding. It has widely applied to solve spatial location, price prediction, and image recognition and method evaluation and so on, and achieves good results [9-11]. In this study, vegetables leafy vegetables were selected as representatives, and gene expression programming was applied to improve the price prediction of leafy vegetables, so as to provide references for monitoring and early warning decisions of government in the operation of vegetable circulation industry.

## 2. Application of new gene expression algorithm in price prediction of leaf vegetable

Since the design of the gene expression program was proposed in 2001, many scholars have improved it from different angles. In order to improve the efficiency of the price forecast, a new method of "gene extraction" was proposed in this paper. In the genome, the genome appears when the first gene as a terminator, not only to retain the gene, but also let the back of the hidden gene expression, the solution is to use the gene is extracted, it connected to the rest of the genome, because the head primordium group of a forward finally, the head of a gene complement.. The improved algorithm part process is as follows:

### *2.1. New new type of gene expression program design*

The new genetic operators of gene expression have such parts as single point recombination, two point recombination, and gene recombination.

*2.1.1. Single point recombination.* At single point recombination, the parent chromosomes match each other and randomly select a position to exchange all the genes behind the two parent chromosome. The following is an example of a single point recombination (a dark blue gene, Figure 1).

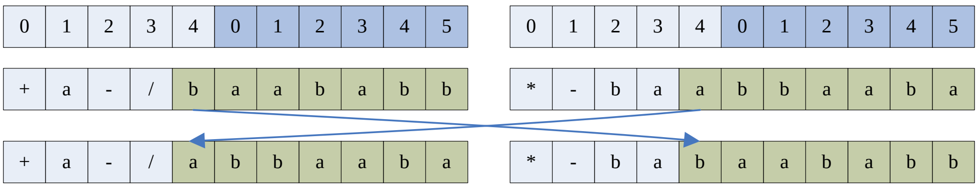


**Figure 1.** Single point recombination.

Two point recombination. Two point recombination is similar to single point recombination, only two points are randomly selected on the chromosome to exchange the genes between two points. Single point recombination and two point recombination are the most common and most effective genetic operators in traditional algorithms *(*dark blue as the exchange of genes, Figure 2*).*

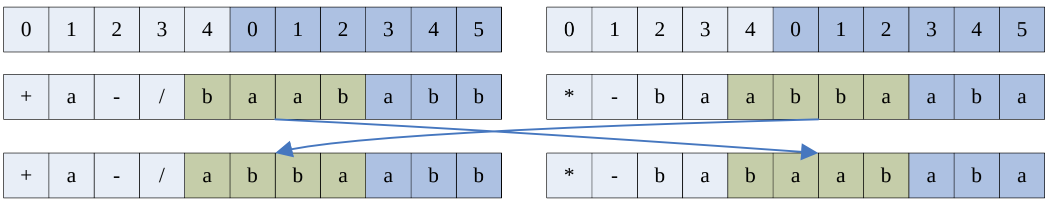


**Figure 2.** Two point recombination.

*2.1.2. Gene recombination.* Gene recombination acts on multi genome chromosomes, selects two parent chromosomes, selects a genome randomly, and exchanges two parent chromosomes (the deep blue is the exchanged genome, Figure 3).

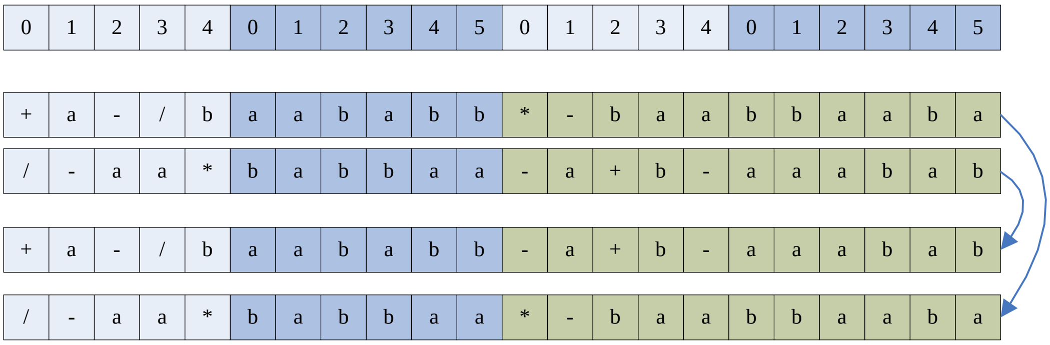


**Figure 3.** Gene recombination.

*2.1.3. Algorithm of gene extraction operator.* After input and output, Check the genome and mark the first genomes with the terminator, and turn to 2. If the genome is not found in all genomes, the algorithm ends. Search he first gene is non-terminating, the shortest effective gene and at least two bit shorter than the length of the gene head, if found, turn to 3, otherwise the algorithm is over. Move the genome head back to two bits, then fill in the connector in the first place, and fill in the terminator found in step 1 with the second bit and go to 1.

The important link is the terminator gene extracted in other genes are not at the end of the first and the effective length of the shortest genomic gene, gene behind so as to let the first end of the genome to be expressed, and the effective length of each chromosome genome tend to average.

### *2.2. Empirical analysis*

As leafy vegetables are seasonality, a leaf vegetable listing period is relatively short. Therefore, from the quarterly perspective, it is not suitable for prediction for a long time as accuracy consideration.

*2.2.1. Import data.* The experimental data selected the average daily average price of leafy vegetables in Guangdong province in July 31st to May 12 in 2015. The data are derived from the Shenzhen Middle Agriculture Data Network (¥/500g).

3.89 3.95 4.05 4.08 4.12 4.19 4.27 4.33 4.45 4.31 4.25 4.20 4.18 4.05 3.98 3.90 3.90 3.85 3.78 3.75

3.70 3.75 3.78 3.80 3.85 3.90 4.05 4.20 4.27 4.35 4.40 4.45 4.56 4.58 4.65 4.65 4.73 4.70 4.70 4.72

4.66 4.68 4.66 4.50 4.55 4.42 4.40 4.33 4.27 4.28 4.19 4.30 4.39 4.56 4.72 4.80 4.69 4.82 4.90 4.95

4.99 4.98 5.12 5.26 5.22 5.34 5.46 5.42 5.31 5.50 5.69 5.67 5.57 5.63 5.69 5.71 5.75 5.65 5.52 5.43

The above total of 80 data, the first 70 data as a training set, the last 10 data as the evaluation of the evaluation of the evaluation set. When the width of the sliding window is 5, 65 sets of training data can be obtained after the data processing.

*2.2.2. Parameter setting.* In this experiment, the length of the head is 10, the number of terminator is 5, the number of genome is 8, and the number of chromosomes is 10. According to the formula t=h*(n-1) +1, the length of the genomic tail of this chromosome is 11, so the length of the genome is 21. The length of a chromosome is 21 * 8, that is, 168. The operator set {+, -, *, /, Q, S, C, E, L}, terminal set for the {a, B, C, D, e} respectively represent the five consecutive days of price. After the data processing, the training set has 62 sets of data, because the selection range is 10, so the maximum fitness is 620.

*2.2.3. Traditional algorithm evolution.* When the data is imported and the parameters are set, the traditional algorithms are beginning to evolve. The groups of genes through variation, inverted string, insert string, string root genetic operators such as modification, continue to move toward the direction of maximum fitness evolution.

*2.2.4. End of evolution and get the prediction formula.* When the evolutionary algebra is set, the traditional algorithm stops evolving. The optimal output is selected as the optimal solution under the setting condition. In this experiment, the optimal formula for the prediction of chromosomes and leaf vegetables is as follows:

```
SL+**a//b*baedbadaaecS~SSb+-dbQeebcacadcbbSc/C~Ee/Seeedbdbbadcd~Q*cQS~
SSLbacccdddbcdS-L+-S/+-*decaccadeabCLca~cS*/Cddbdbddddac+c-EcLc~S*dbcda
aecdbd~-ee/cSeCScbbcedbbcba
```

The final formula for predicting the price of leaf vegetables can be obtained by connecting each expression with the symbol "+". As shown below:

$$f_i = \sin(\log(f_{i-5} * f_{i-1} + f_{i-5} * f_{i-4} / f_{i-1})) + \sin(-\sin(\sin f_{i-3})) + \sin f_{i-3} - \sqrt{f_{i-3} * \sqrt{\sin(-\sin(\sin(\log(f_{i-4}))))}} + \sin(\log(-(f_{i-3} + f_{i-5})) - (\sin(f_{i-5} * f_{i-2}) + f_{i-2} / f_{i-1})) + \cos(\log f_{i-3}) + (f_{i-3} + e^{\log f_{i-3}})$$

## *2.3. Comparison and analysis of prediction results*

In order to verify the feasibility of the application of the traditional algorithm in the prediction of the price of leaf vegetable and the superiority of the improved algorithm, a series of experimental analyses are carried out in this paper.

*2.3.1. Comparison of gene utilization ratio.* The advantage of the improved algorithm is to improve the efficiency of gene utilization. Figure 4 shows the 200 generation in one experiment, and the gene utilization ratio of each generation's traditional algorithm and improved algorithm. The average utilization rate of the improved algorithm is 18%, while the traditional algorithm has a gene utilization rate of 12%. It can be seen that the gene utilization rate of the improved algorithm is obviously higher than that of the traditional one.

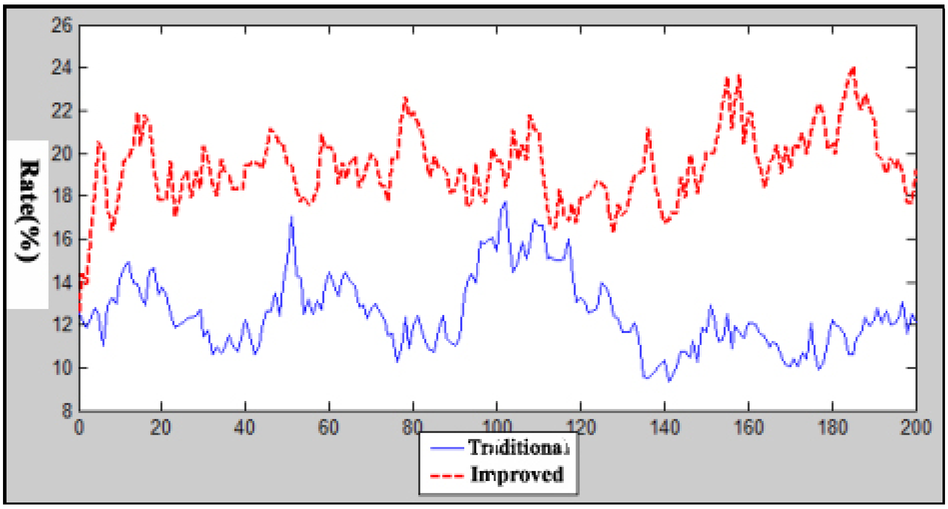


**Figure 4.** Comparison of gene utilization ratio between GEP and IGEP.

*2.3.2. Effect of the increase in gene utilization on evolution.* In order to find out the effect of the increase in gene utilization on evolution, a comparative experiment was designed in this paper. First, the conditions of evolution end are set, and the evolution is stopped when the given error range is reached. For example, when the set error is 10%, the maximum fitness 620 of 90% is 558, that is to say, the evolutionary algebra that the traditional algorithm and the improved algorithm achieve 558 respectively. The following are the average evolutionary algebra required for the different error ranges in the 100 experiment, the traditional algorithm and the improved algorithm.

**Table 1.** Algebraic contrast of evolution needs.

| Error (%) | Traditionalalgorithm | Improved algorithm |
|---|---|---|
| 10% | 72 | 98 |
| 8% | 96 | 136 |
| 6% | 166 | 192 |
| 4% | 288 | 368 |
| 2% | 512 | 384 |
| 0.8% | 702 | 466 |
| 0.6% | 783 | 683 |
| 0.4% | 843 | 742 |
| 0.2% | 1124 | 926 |
| 0.1% | 1440 | 1020 |

The results in Table 1 showed that the evolutionary speed of the improved algorithm is not faster than that of the traditional algorithm when the error range is large. The main reason is that the fitness evaluation function used in the experiment is absolute error, and the price value of the leaf vegetable is small, which leads to a very high value of adaptability very quickly. The improved algorithm is an increase in genetic operators on the basis of the traditional algorithm, the algorithm is equivalent to the basic role of chromosome, the improved algorithm for further modification, resulting in the error is low, but because of improved rather than basic algorithm takes the evolution algebra. Therefore, in the price prediction of the designated evolutionary algebra as the end condition, the improved algorithm's evolution solution, that is, the prediction function, will be better than the traditional algorithm.

In order to get more accurate conclusions, 10 prediction experiments were carried out in this paper, and the conditions of each experiment were consistent. Then the average value of the prediction is taken and the basic algorithm and the improved algorithm are evaluated according to the prediction error. The traditional algorithm parameters in the experiment are as follows (Table 2):

**Table 2.** Gene expression algorithm parameter.

| Parameter | numerical value |
|---|---|
| Maximum evolutionary algebra | 1080 |
| Population size | 30 |
| Function set | + - * / ~ Q S C L E |
| Terminator set | a b c d e f g h |
| Head length | 30 |
| Number of genes | 20 |
| Connector | + |
| Mutation probability | 0.005 |
| Single point recombination probability | 0.25 |
| Two point recombination probability | 0.20 |
| Gene recombination probability | 0.15 |
| IS insertion probability | 0.20 |
| RIS insertion probability | 0.20 |
| Gene string probability | 0.15 |
| Gene extraction probability | 0.55 |

After 10 experiments, 10 sets of predicted values were obtained and their average values were as follows (Table 3).

**Table 3.** Comparison of prediction results (¥/500g).

| Price | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Relative error (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Real price | 4. 43 | 4.48 | 4.58 | 4.52 | 4.66 | 4.72 | 4.64 | 4.51 | 4.62 | 4.44 | |
| Traditional | 4.29 | 4.34 | 4.42 | 4.61 | 4.76 | 4.89 | 4.83 | 4.39 | 4.78 | 4.52 | 14.03 |
| Improved | 4.48 | 4.52 | 4.54 | 4.55 | 4.68 | 4.82 | 4.75 | 4.43 | 4.56 | 4.40 | 7.25 |

The results showed that in leafy vegetable price prediction, the relative error between predicted price and real price based on gene expression operator and inverted string operator is 7.25%, but the traditional relative error is 14.03%. The improved is higher than that of GEP and closer to the real leafy vegetable price. It also shows that the new genetic expression programming algorithm is feasible in vegetable price prediction and is of practical reference value.

## 3. Conclusion

In view of the shortcomings of traditional gene expression programming in the process of gene utilization, a new improved method based on "gene extraction" operator and "inverted string" operator is proposed, and the algorithm is applied to leaf vegetable price prediction. The feasibility of gene expression programming in vegetable price prediction and the superiority of the new gene expression program design are proved by a number of comparison experiments.

The research shows that gene expression programming has certain practical value in vegetable price prediction, but this paper only uses time series prediction method, without considering other factors. There are many factors that influence the price of vegetables, such as temperature and weather. In the next step, we can introduce these factors into the prediction of gene expression program of vegetable price.

## Acknowledgments

We thank the supported by grants from the Guangdong Philosophy and Social Science Planning Project (GD16CGL11), Guangdong Soft Science Project (2016A070705037), Guangdong Science and Technology Program (2014B090908017, 2014B070706015), Guangdong Soft Science Project (2015A070704034), President Fund Project of Guangdong Academy of Agricultural Sciences (201532).

## References

[1] Nguyen D H and Widrow B, 1990, Neural networks for self-learning control system, *IEEE Con. Sys. Mag.*, **12**: 18-23.
[2] Dieng A, 2008, Alternative forecasting techniques for vegetable prices in senegal, *Intl. Conf.Power Sys. Tech. IEEE*, **1**: 5-10.
[3] Li G Q, Xu S W and Li Z M, 2010, Short-term price forecasting for agro-products using artificial neural network. *AASP*, **1**: 278-87.
[4] Ju J, Zhao L and Wang J, 2013, Forecasting the total power of China's agricultural machinery based on bp neural network combined forecast method, *CCTA*,**4**: 85-93.
[5] Li Z M, Xu S W and Cui L G, 2012, The short-term forecast model of pork price based on CNN-GA,*Adv. Mater. Res.*,**11**: 396-401.
[6] Moore H L, 1997, Forecasting the yield and the price of cotton, *Mac Comp.,* **15**:100-13.
[7] Li C G,2013, Short-term prediction of vegetable prices in China based on chaos neural network mode.Chin. *Academy Agri. Sci.*, **22**:101-15.
[8] Zhu X X, 2012, Markov chain analysis and prediction of vegetable price fluctuation cycle*, Productivity*, **8**: 143-6.

[9] Zhou A M, Cao H Q and Kang L S, 2003, Automatic modeling of complex functions based on genetic programming, *J. Sys. simu*, **6**: 44-6.
[10] Xie J Y, Wang M Z, Zhou Y, Gao H C and Xu S Q, 2018, Research on the differential expression gene selection algorithm of non-equilibrium gene data, *J. Comp. Sci.*, **5**:1-18.
[11] Wang Z L, Wang H K, Li X S, Gu F, Zhou Y and Ma W, 2016, Research on bi-directional clustering algorithm based on gene expression data, *J Natural Sci. Heilongjiang Uni.*, **33**: 444-9.