INT. J. BIOAUTOMATION, 2017, **21**(4), 349-356

The Use of Adaptive Genetic Algorithm for Detecting Kiwifruit's Variant Subculture Seedling

Jun Zeng, Yong Li*

College of Information Engineering and Automation
Kunming University of Science and Technology
Kunming, China
E-mails: 596683835@qq.com, Dr.Leon@foxmail.com

**Corresponding author*

Received: November 02, 2016 **Accepted: July 19, 2017**

Published: December 31, 2017

Abstract: *In order to reduce the possible economic loss brought by variant seedlings in tissue culture, we propose a pattern recognition approach using fitness to dynamically monitor subculture seedlings of kiwifruit based on adaptive Genetic Algorithm. By coding, selection, mutation and crossover the selected primer pairs of the subculture seedlings, we simulate the process of optimizing the kiwifruit's genomic DNA polymorphism. The result shows that fitness values of kiwifruit's subculture seedlings can better maintain their genetic stability from the first to the ninth generation in the simulation. But from the tenth generation, the rapid change of the fitness values of subculture seedlings happen. It is in accord with the experimentation, which uses optimized AFLP system for analyzing genetic diversity of 75 samples of seventh to eleventh 5 generations of kiwifruit subculture seedlings.*

Keywords: *Kiwifruit, Variant seedling, Tissue culture, Adaptive genetic algorithm, Fitness function.*

Introduction

According to the basic theory of genetics, chromosomes in the nucleus of gene determine the physical characteristics of the organism, and the survival of the fittest chromosomes (individuals) are selected for reproduction in [11] and [13]. DNA is the genetic material containing the whole information of an organism to be copied into the next generation of the species, and the reproduction is actual a DNA replication process. Genetic variation could appear in the process. Variation can negatively affect the quality of individuals. If necessary measures are not taken in time, the number of variation individuals will grow exponentially. In [15] and [6], researchers point out that DNA is stable and predictable in its reactions and can be used to encode information for detecting variation. Based on a genetic algorithm (GA), dynamic monitor of kiwifruit's variant seedlings in tissue culture is discussed as an example in this paper.

In [9], some researchers presented that cuttings, grafting, seeding and tissue culture are four kinds of kiwifruit cultivation methods. The tissue culture is mainly used to breed elite varieties of kiwifruit. In the process of tissue culture, different vitro conditions, plant growth regulators, medium osmotic pressure, culture temperature and time, and so on could cause genetic variation in seedlings. Some variations are not easy to be found in the tissue culture process, and if it is found in the later of planting in the field, the economic loss is even greater. Some researchers indicated that for reducing the economic loss caused by variant seedlings, the variation is needed to be monitored [3].

It is obviously that tissue culture of kiwifruit subculture seedlings is an optimal process that the kiwifruit seedlings adapt to environment. Researchers' works show GA is one of the promising methods as an adaptive probabilistic optimization algorithm that is based on biological genetic evolutionary mechanisms for searching the survival of the fittest individual [1, 4, 7, 12]. It simulates the natural process of genetic recombination and evolution, performing operations similar with natural selection, crossover and mutation to get the final optimization result after repeated iterations. In GA, fitness function is proposed to evaluate the quality of individuals. The fittest individual is chosen by ranking individuals according to a pre-defined fitness function, which evaluated each member of a population. The individuals with high fitness values therefore represent better solution to the problem than lower one. Following the initial process, the crossover and mutation operations are used, while the individuals in the current population produce offspring. And then, a new population will have been formed and the generational counter is increased by one. In [8], research work showed the process of selection, crossover and mutation is continued until a termination condition is met. In conventional mutation operator, mutation probability is fixed and the operation is performed by not paying any attention to whether it is necessary or not. As in conventional crossover operator, the probability of crossover, which determines the number of exchanged segments, is commonly given by user, depending on experience. The adaptive mutation and crossover operator are used to instead of the fixed mutation and crossover probabilities. The adaptive mutation maintains GA to have random nature and the adaptive crossover leads to keep an individual with high fitness in the next population, which is discussed in [14] and [16]. Thus, the adaptive genetic algorithm can be more enable to simulate evolution of tissue culture process. Normally, the termination condition of GA is the fitness values reach a highest value and (or) convergence in a finite number of repeated iterations. Thus, if the fitness values appear obviously various after convergence, it could indicate there are gene variations in tissue culture, and imply detecting kiwifruit's variant seedlings.

The model of detecting variant subculture seedlings

Individuals' encoding mode

The nuclear DNA is located within the nucleus of eukaryotic cells. The biochemical structure of the DNA containing the polynucleotide base alphabet {*A*, *G*, *C*, *T*} (*A*, adenine; *G*, guanine; *C*, cytosine; *T*, thymine) includes the genetic information necessary for the preservation of the base sequences. In mathematics, it means using a character set containing four characters $\sum = (A, G, C, T)$ to encode information. A numerical sequence could be obtained by mapping the base of DNA to the frequency of the base in the sequence. Here, the numbers 0, 1, 2, 3 are used to denote *G*, *A*, *T*, *C* respectively, i.e., G = 0(00), A = 1(01), T = 2(10), C = 3(11).

Individual evaluation method

Tissue culture is an optimal process. In information technology, many fitness functions are always used to simulate optimization problems. The Rosenbrock function is a single peak function, and always taken for evaluating the optimization problems. We use the function for mimic the tissue culture as example in the study. The Rosenbrock function is given as below:

$$f_r(x) = \sum_{r=1}^{r} 100(x_{i+1} - x_i^2)^2 + (1 - x_i^2)\,. \tag{1}$$

In this paper, $r = 2$ is implemented. The evaluation model is:

$$f_2(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1^2),\ \ -2.048 \le x_1, x_2 \le 2.048\,. \tag{2}$$

For this problem, variables x_1, x_2 can be represented with 6 bits of binary encoding string. The definition of x_1, x_2 is discretized into 1023 equal areas, forming 1024 discrete points, which includes the end of point. Let discrete points from –2.048 to 2.048 correspond to binary encoding from 000000 to 111111. In decoding process, the 12 bits long binary string of encoded kiwifruit genomic DNA is cut off for two 6 bits binary encoding strings, and then they are converted to the corresponding decimal integer code y_1 and y_2 respectively. The code y_i and x_i decoding formula is:

$$x_i = 4.096 \times \frac{y_i}{2^6 - 1} - 2.048, \quad i = 1, 2. \tag{3}$$

Selection, mutation and crossover operator

The probability of being chosen individual is proportional to its fitness. Here we use the roulette selection strategy way to dynamically monitor of kiwifruit's variant seedlings in tissue culture to choose the individual which the fitness is enough big to cultivate in [2].

The crossover operator is the operator to be in needs of a population and increases variety of solution. The cross operation uses a single point crossover operator, the type of mutation operation uses the basic bit mutation operator. At a certain probability, two individuals were randomly selected from the population, and some of them were partially exchanged. Crossover methods usually adopt one point, two points, more points and uniform crossover. The function of crossover is to produce a new gene, that is, the new chromosome. The adaptive crossover is operated by fitness of individuals as presented in Eq. (4):

$$p_m = \begin{cases} k_3 * \dfrac{f_{\max} - f}{f_{\max} - f_{avg}}, & f \geq f_{avg} \\ k_4 \quad , & f < f_{avg} \end{cases}, \tag{4}$$

where f_{max} is the maximum fitness in population, f_{avr} is the average; f is the fitness values of the variant individual. k_3 and k_4 are in interval (0, 1). In the scope of adaptive approach, the probability of crossover p_c is determined according to formulation of which characteristic form has been introduced in [14], where f is the lowest fitness between parents. p_c decreases if its' parent fitness increases. Increment of fitness of individual decreases the probability of crossover, so this leads to keep an individual with high fitness in the next population. This manner assists elitist strategy, which means that the best individual sustains its preference in the next generation.

The role of mutation operator in GA is to prevent GA to suboptimal solutions by restoring lost or unexplored genetic material into population. The suboptimal solutions result from premature convergence. For enabling the operator can play the role attractively, a mutation operator should be applied by controlling the degree of goodness (i.e. fitness) in the individuals. In [16] Yoshimoto use the method developed by Srinivas and Patnaik to determine the probability of the adaptive mutation operator, given as Eq. (5):

$$p_c = \begin{cases} k_1 * \dfrac{f_{\max} - f'}{f_{\max} - f_{avg}}, & f' \geq f_{avg} \\ k_2 \quad , & f' < f_{avg} \end{cases}. \tag{5}$$

Here, f_{max} is the maximum fitness value of individuals in population, f_{avr} is the average, f' is the larger of two individual fitness values to cross, k_1 and k_2 are in interval (0, 1). Increment of fitness of individual decreases the probability of mutation. In conventional mutation operator, however, mutation probability is fixed and the operation is performed by not paying any attention to whether it is necessary or not. In other words, adaptive mutation operator used in this study gives opportunity of survival to an individual having lower fitness than the others in the population by eliminating its bad design variables.

Normally, the use of the adaptive crossover and the adaptive mutation operators carries out that the individual having the greatest fitness is placed in the next generation, which can be viewed as elitist strategy. Hence, the adaptive approach covering the adaptive crossover and the adaptive mutation operators is preferred in the study.

The genetic data of the individual is initialized from an input file at the beginning of the algorithm. Based on their different genetic search strategy, individuals search the plane at different directions. At the same time, gene is evolved with crossover and mutation by using individual fitness information, and the search strategy is constantly adjusted. Finally, the best individual is maintained. Iteration continues until termination condition is reached.

Experiment

The monitor of kiwifruit subculture seedlings is taken as example to verify the proposed method. Fig. 1 is electrophoretogram of kiwifruit subculture seedlings, which is provided by the key laboratory of biodiversity conservation of China state forestry administration.

Fig. 1 Electropherogram of genomic DNA of kiwifruit subculture seedlings

Fig. 2 is the experimental results obtained by using 8 genomic prime pairs to detect the variation of tissue subculture seedlings of the kiwifruit in [10].

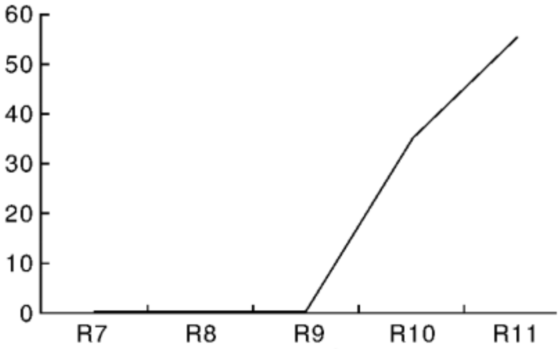

Fig. 2 The relationship between the number of generations
and the rate of variation of tissue cultures seedlings of kiwifruit with 8 prime pairs

Based on electrophoresis as Fig. 1, 8 prime pairs of kiwifruit genomic DNA are filtered from 64 primers: E-AGG+M-CAT, E-ACT+M-CAG, E-AGG+M-CAG, E-ACG+M-CTG, E-AAG+M-CTG, E-ACG+M-CAA, E-AAG+M-CAA, and E-ACG+M-CAG, which the numbers of bands are the richest in 64 primer pairs in experiments. The experimental subculture seedlings are randomly chosen from every subculture generation, 1-2 fresh leaves of per plant are picked, put in zip lock bag and marked respectively, preserved under –80 °C, which use 300 mg genomic DNA marked on AFLP for endonuclease digestion with 4h and adding 1U T4-DNA ligase, 1.5 uL Adapter and 2 uL ATP to link. In this teat, dosage pre-amplification primer is 0.8 uL and the selective amplification of Tap polymerase is 0.25 uL. The corresponding 12 bits encoding genomic DNA of kiwifruit primer pairs are as following: 010000110110, 011110110100, 010000110100, 011100111000, 010100111000, 011100110101, 010100110101, and 011100110100. The selected 8 prime pairs have been used to evaluate the polymorphism of kiwifruit genes in other research work in [9] and [5], so it can be used to identify variations of subculture seedlings in our study.

The initial operating parameters of the proposed algorithm are as follows: population scale $M = 8$, termination generation $T = 12$. Debugging with Java programming, the corresponding simulated results of the first three evaluations are listed in Table 1 to Table 4. In order to obtain better result to evaluate variations of the 8 primer pairs, fitness is instead by relative fitness, which is calculated with each fitness value divided by total fitness values.

Table 1. The coding result of initial population

No.	Binary coding of initial 8 prime pairs	X1	X2	Relative fitness
1	010000110110	-1.008	1.463	0.024
2	011110110100	-0.098	1.333	0.179
3	010000110100	-1.008	1.333	0.014
4	011100111000	-0.228	1.593	0.243
5	010100111000	-0.748	1.593	0.112
6	011100110101	-0.228	1.398	0.185
7	010100110101	-0.748	1.398	0.075
8	011100110100	-0.228	1.333	0.168

Table 2. The simulated result of 1st evolution of the 8 primer pairs

No.	Evaluation results of primer pairs	X1	X2	Relative fitness
1	001110111001	-0.228	1.333	0.125
2	001110111001	-0.228	1.333	0.125
3	010101110101	-0.748	1.398	0.055
4	001110111001	-0.228	1.333	0.125
5	101110001001	-0.098	1.333	0.133
6	101110001001	-0.098	1.333	0.133
7	010100110101	-0.228	1.593	0.180
8	011100110100	-0.228	1.333	0.125

Table 3. The simulated result of 2nd evolution of the 8 primer pairs

No.	**Evaluation results of primer pairs**	**X1**	**X2**	**Relative fitness**
1	101010111001	-0.228	1.333	0.123
2	100011011110	-0.098	1.333	0.131
3	101010111001	-0.228	1.333	0.123
4	101010111001	-0.228	1.333	0.123
5	101010111001	-0.228	1.333	0.123
6	101010111001	-0.228	1.333	0.123
7	101010111001	-0.228	1.333	0.123
8	100011011110	-0.098	1.333	0.131

Table 4. The simulated result of 3rd evolution of the 8 primer pairs

No.	**Evaluation results of prime pairs**	**X1**	**X2**	**Relative fitness**
1	110110011011	-0.228	1.398	0.129
2	010001010111	-0.748	1.593	0.078
3	011100111000	-0.228	1.593	0.169
4	010001010111	-0.748	1.593	0.078
5	011100111000	-0.228	1.593	0.169
6	011100111000	-0.228	1.593	0.169
7	110110011011	-0.228	1.398	0.129
8	010001010111	-0.748	1.593	0.078

The best individuals of 12 repeated iterations are as following:

Table 5. The simulated results of the 8 primer pairs after 12 iterations

No.	**Binary coding of each best individual**	**X1**	**X2**	**Relative fitness**
R1	011100111000	-0.228	1.593	0.199
R2	011100111000	-0.228	1.593	0.153
R3	011100111000	-0.228	1.593	0.148
R4	011100111000	-0.228	1.593	0.130
R5	011100111000	-0.228	1.593	0.130
R6	011100111000	-0.228	1.593	0.125
R7	111010010100	-0.228	1.593	0.125
R8	110100111110	-0.228	1.593	0.125
R9	111011100000	-0.228	1.593	0.125
R10	001100110101	1.788	0.033	0.375
R11	001100110101	1.788	0.033	0.175
R12	001100110101	1.788	0.033	0.138

From Fig. 3, the variations of the R4 to R9 generations are relative stable. But the variation rate of the R10 generation changes rapidly. In conclusion, the 1st to 9th generations of kiwifruit subculture seedlings can better maintain the genetic stability, but variation occurred from the 10th generation.

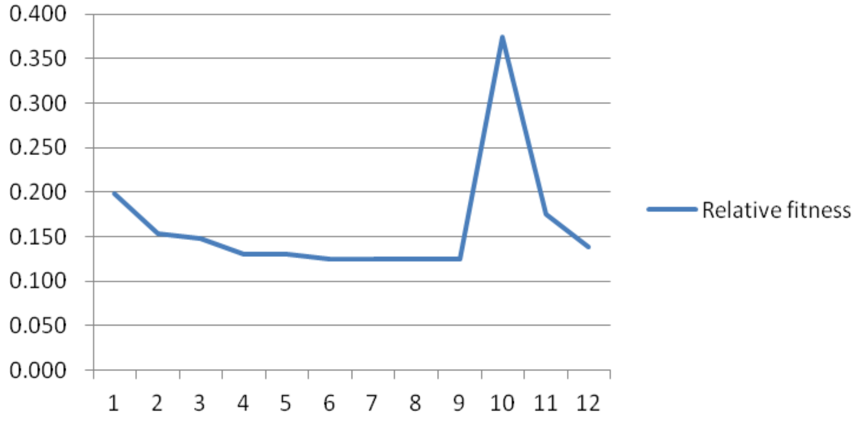

Fig. 3 The fitness variation of 12 generations' kiwifruit subculture seedlings in simulated tissue culture process

Conclusion

The result is accordance with subculture seedlings R7, R8, R9, R10, R11 (the earliest sample is R0, and R1 to R11 as subcultures) of kiwifruit 'Hort 16A' as experimental materials illustrated in Fig. 1 and Fig. 2. The study is also based on our pre-work that using normal GA, non-adaptive, which probability of crossover and mutation are fixed, and the similar result is also obtained. It shows the proposed method is a reasonable way to monitor of kiwifruit's variant seedlings in tissue culture with information technology.

References

1. Aditi K., Y. T. Oscar, J. Wenyi (2005). Iterative Reliability-based Decoding of Linear Block Codes with Adaptive Belief Propagation, IEEE Communications Letters, 9(1), l067-1069.
2. Althoff D. M., M. A. Gitzendanner, K. A. Segraves (2007). The Utility of Amplified Fragment Length Polymorphism in Phylogenetics: A Comparison of Homology within and between Genomes, Systems Biology, 156, 477-484.
3. Bairu M. W., A. O. Aremu, J. V. Staden (2011). Somaclonal Variation in Plants: Causes and Detection Methods, Plant Growth Regul, 63(2), 147-173.
4. Holland J. H. (1975). Adaptation in Natural and Artificial Systems, University of Michigan Press, Michigan.
5. Hu J., L. Pan, H. Liu, S. Wang (2012). Comparative Analysis of Genetic Diversity in Sacred Lotus (Nelumbo Nucifera Gaertn.) Using AFLP and SSR Markers, Molecular Biology Reports, 39(4), 3637-3647.
6. Kayis S. A. (2012). Evaluation of Confidence Limit Estimates of Cluster Analysis on Molecular Marker Data, Journal of the Science of Food and Agriculture, 92(4), 776-780.
7. Li Z. H., X. C. Wang, K. Q. Li, J. F Gu, L. Kang (2014). Molecular Docking Improvement: Coefficient Adaptive Genetic Algorithms for Multiple Scoring Functions, International Journal Bioautomation, 18(1), 5-14.
8. Michalewicz Z. (1996). Genetic Algorithms + Data Structures = Evolving Programs, Springer-Verlag, Berlin.

9. Predieri S. (2001). Mutation Induction and Tissue Culture in Improving Fruits, Plant Cell Tissue Organ Culture, 64(2), 185-210.
10. Qin X. B., J. H. Gao (2013). Studies on Genetic Diversity for Specialty Kiwifruits from Southwestern Regions of China Using AFLP System, Journal of Tropical and Subtropical Botany, 21(4), 315-322.
11. Schneider T. D. (2001). A Brief Review of Molecular Information Theory, Nano Communication Networks, 1(3), 173-180.
12. Semcheddine S., A. Aitouche (2015). Monitoring Biological Modes in a Bioreactor Process by Computer Simulation, International Journal Bioautomation, 19(4), 507-520.
13. Shah K., A. Krishnamachari (2012). Nucleotide Correlation Based Measure for Identifying Origin of Replication in Genomic Sequences, BioSystems, 107(1), 52-55.
14. Srinivas M., L. M. Patnaik (1994). Adaptive Probabilities of Crossover and Mutation in Genetic Algorithms, IEEE Transactions on Systems, Man and Cybernetics, 24, 656-667.
15. Sudarshini K. V., P. Bommisettya, M. S. Patilb, L. Reddya, A. Chennareddy (2014). The Genetic Linkage Maps of Anthurium Species Based on RAPD, ISSR and SRAP Markers, Scientia Horticulturae, 178, 132-137.
16. Yoshimoto F., T. Harada, Y. Yoshimoto (2003). Data Fitting with a Spline Using a Real Coded Genetic Algorithm, Computer-Aided Design, 35, 751-760.

Jun Zeng, B.Eng.
E-mail: 596683835@qq.com

Jun Zeng is a Master degree candidate in the College of Information Engineering and Automation, Kunming University of Science and Technology. She majors in Automation and her interests of research are intelligent information processing and data mining.

Assoc. Prof. Yong Li, Ph.D.
E-mail: Dr.Leon@foxmail.com

Yong Li is an Associate Professor of Information Engineering and Automation College, Kunming University of Science and Technology. He graduated from Yunnan University in 2010 and got Ph.D. degree with specialty of Communication and Information System. His research interests are in pattern recognition for biometrics and computer simulation.

© 2017 by the authors. Licensee Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).