DE GRUYTER
OPEN

BULGARIAN ACADEMY OF SCIENCES

CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **16**, No 5
Special Issue on Application of Advanced Computing and Simulation in Information Systems

Sofia • 2016 Print ISSN: 1311-9702; Online ISSN: 1314-4081
DOI: 10.1515/cait-2016-0052

Using Fitness Value for Monitoring Kiwifruit's Variant Seedling in Tissue Culture

Shouguo Tang, Yong Li, Zhikun Zhang

Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, China
Emails: tandycool@qq.com Dr.Leon@foxmail.com redmars141@126.com

Abstract: *Based on Genetic Algorithm, a pattern recognition approach using fitness to dynamically monitor the sub cultured seeding of kiwifruit is proposed in order to decrease the loss of variant seedlings in tissue culture. By coding, selection, mutation and cross-overing the selected primer pairs of the sub cultured seeding, we simulate the process of optimizing the kiwifruit's genomic DNA polymorphism. The corresponding fitness values of the primer pairs are evaluated with fitness function for monitor the variation of kiwi's DNA. The result shows that kiwi's plantlets can better maintain their genes' genetic stability for the first to the ninth generation. But from the tenth generation, the fitness values become variation. The results are based on experimentation, which uses optimized AFLP system for analyzing genetic diversity of 75 samples of seventh to eleventh 5 generations of kiwi.*

Keywords: *Kiwifruit, Variant Seedling, monitoring, Genetic Algorithm, Fitness Function.*

1. Introduction

In accordance with the basic theory of genetics, the physical characteristics of the organism are determined by chromosomes in the nucleus of the gene. DeoxyriboNucleic Acid or DNA is the genetic material containing the whole information of an organism to be copied into the next generation of the species. Normally, the survival of the fittest individuals (chromosomes) are selected for the reproduction. But due to the influence of many factors such as the environment, genetic variation appears in DNA replication process. Variation can negatively affect the quality of the plant. The effect on the variation of the plant increases rapidly with the age of the subculture. If necessary measures are not taken in time, the number of variation seedlings will grow exponentially.

Cuttings, grafting, seeding and tissue culture are four kinds of cultivation methods of Kiwifruit. 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, culture medium, hormone composition and content, and explants regeneration mode case seedling variation. Some genetic variation is not easy to be found in the process of tissue culture, and if it is found in the later of planting in the field, the economic loss is even greater. In order to avoid the economic loss caused by variant seedlings, the variation in the process of tissue culture is needed to be monitored.

Genetic Algorithm (GA) is an adaptive probabilistic optimization technique based on biological genetic evolutionary mechanisms for searching the survival of the fittest individual. It simulates the natural process of genetic recombination and evolution, performing operations similar to natural selection, crossover and mutation to get the final optimization result after repeated iterations (i.e., generations genetic) [1, 2]. It is obviously that tissue culture of kiwifruit is an optimal process that the kiwifruit seed adapt to environment. There are some similarities between tissue culture of kiwifruit and GA operations. DNA, the genetic material that encodes for living organisms, is stable and predictable in its reactions and can be used to encode information for monitor the tissue culture of kiwifruit with GA.

Fitness function is proposed to evaluate the quality of chromosomes in GA. The fittest individuals are chosen by ranking them according to a pre-defined fitness function, which is evaluated for each member of this population. The individuals with high fitness values therefore represent better solution to the problem than individuals with lower fitness values. Following this initial process, the crossover and mutation operations are used where the individuals in the current population produce the children (offspring). These children are assigned fitness scores to be selected. After selection, crossover and mutation have been applied to the initial population, a new population will have been formed and the generational counter is increased by one. This process of selection, crossover and mutation is continued until a termination condition is reached [3]. Normally, the condition is the fitness values reach a highest value and convergence in a finite number of iterations. Thus, if the fitness values appear obviously various after convergence, it could indicate gene mutation in tissue culture. In the study, using fitness and genetic algorithm, we proposed a method to determine the variation of the subculture of kiwifruit in process of tissue culture.

2. Literature review

Variant seedling is the collective of seedlings that are mutated in the production process of the industry, which are beneath values. There are many general methods for monitoring variant seedlings in biological science, such as SSR, ISSR, RFLP, SNP, AFLP and PAPD [4-6], etc. For example, AFLP (Amplified Fragment Length Polymorphism) is a molecular marker technology invented by Holland scientist Vos Pieter in 1995. These techniques are used to monitor the variation of tissue culture seedlings with common experiment mode.

Computer science is generally used in the field of biological science. Pattern recognition refers to process and analysis information (numerical, text and logical relation) characterizing various forms of things or phenomena for description, identification, classification and interpretation of things or phenomenon. Genetic algorithm is a commonly used pattern recognition method. Voss method, *Z*-curve method, Tetrahedron method and so on are the typical genetic algorithms.

A genetic algorithm is an optimization technique inspired by Darwin's theory of evolution, which is first popularized by Holland and extensively studied by Goldberg [7]. Generally, it starts with a set of individuals that forms a population with crossover, mutation, and selection to generate new search points in a discontinuous search state space. It is one of the promising methods as the optimization algorithm that is based on an idea about evolution of life, and thus, receiving remarkable attention all over the world today [8-10]. Genetic Algorithms (GA) are search algorithms based on the mechanics of natural selection and natural genetics. The use of this optimization technique has given important results in other areas [11-13]. In the above works, the fundamental criterion to regulate the optimization is to demand the survival of the fittest individual as is usual in GA studies. The forecasting problem is one extensively studied [14-16]. It consists in the estimation of future values of securities and trends in data. Investment would be straightforward with perfect predictions of the future [17-19]. Although this is not possible, forecasting involves also an estimation of prediction error, allowing to making better decisions under uncertainty. All computational methods developed so far to predict replication origin rely on the fact that the leading and lagging strands are subject to different mutational pressures during the replication process, which leads to differences in the statistical properties of the DNA sequence on two sides of the replication site [20-22].

The Rosenbrock method is a widely used for testing new optimization algorithms, in its basic form a gradient free minimization algorithm [23]. Its' corresponding function is a continuous and nonlinear function, and there is only one peak in a steep parabolic valley with a flat bottom. Thus, in tissue culture of kiwi, if fitness values changed steeply after keeping long time steady, it can be taken as there is gene mutation and monitor of Variant Seedling.

3. Using fitness to dynamically monitor of kiwifruit's variant seedling

From what mention above, monitoring of Kiwifruit's variant seedling in tissue culture is a typical optimization problem. Normally, using genetic algorithm to solve an optimization problem considered is

(1) $$J = \max_{X \in K} F(X),\ X = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}.$$

Here, $F(X)$ is the evaluation function, X is the solution vector, and K is the definition field of the solution vector.

Supposing there are two variables x_1, x_2 in $F(X)$ as an example, the corresponding encoding length l of x'_i is usually given by the following:

(2) $$l \geq \ln\left[\frac{b-a}{X} + 1\right],\ x_i \in [a, b],\ x_i \leq X.$$

Here X is a given positive integer, and denotes the solution of chromosome with length l+1, which can be composed of two binary strings x'_1, x'_2. The decoding of real number x_i can be determined by the following formula:

(3) $$x_i = a + (b-a)\frac{X_{i(D)}}{2^l - 1},$$

where $X_{i(D)}$ is a decimal number corresponding to binary chromosome code x'_i.

Thus, using fitness to dynamically monitor of Kiwifruit's variant seedling should be done as in following steps.

3.1. Encoding and decoding

For different types of optimization problems, the process of genetic algorithms' solution is basically the same. However, due to the different nature of the variable, it will lead to differences in encoding and decoding of the chromosome (individual). There are typical four different encoding types such as binary, value, quaternary, and octal encodings. It can offers flexibility for effective utilization of genetic operators.

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, 6-aminopurine; *G*, guanine, 2-amino-6-hydroxy-purine; *C*, cytosine, 6-amino-2-hydroxy-pyrimidine; *T*, thymine, 2, 6-dihydroxy-5-methyl-pyrimidine) includes the genetic information necessary for the preservation of the base sequences. In mathematics, it means using a character set containing four characters Σ = (*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. For example, let $A = 1$, $G = 2$, $C = 0$, $T = 3$, according to the principle of the pairing of a Purine for a Pyrimidine, a quaternary combination is encoded. Based on the numerical sequence, a high quality gene can be obtained by using a kind of distribution function.

3.2. Calculating values of fitness function

The range of fitness function $f(x_1, x_2, \ldots, x_n)$ is always non negative, and the optimization goal is obtaining $x_1, x_2, \ldots, x_n$, when $f(x_1, x_2, \ldots, x_n)$ is at maximum. So the individual's fitness is directly corresponding to value of objective function, i.e.

(4) $$\text{fitness}(x) = f(x_1, x_2, \ldots, x_n).$$

In some cases, fitness function is morbid with the optimum located in a steep parabolic valley with a flat bottom where $x_1 = x_2 = \ldots = x_n$, such as Rosenbrock function, so it is easy to find the local maximum. What mentioned above discusses the situation of solving the maximum value of the objective function. When the minimum value of the objective function was need to be solved, the test function was considered as

(5) $$F(x_1, x_2, ..., x_n) = 1/(1 + c + f(x_1, x_2, ..., x_n)),$$

where c is constant and $c + f(x_1, x_2, \ldots, x_n) \geq 0$.

For example, the minimum of Rosenbrock function is 0 in the range of [−2.048, 2.048] when $x_1 = x_2$. To transform Rosenbrock function into a maximization problem, the test function was considered as

(6) $$F(x_1, x_2) = 1/(1 + f(x_1, x_2)),$$

where $-2.048 \leq x_1,\ x_2 \leq 2.048$. Then the global optimal solution of $F(x_1, x_2)$ is 1 when $F(x_1, x_2)$ locates in the minimum.

3.3. Selection operator

Selection strategy usually includes: Roulette, proportional selection, tournament competition and keeping the best individual. Roulette wheel selection method is first suggested alternative of genetic algorithm, that is, from the current group choose the high adaptive value of individuals to produce in the mating process. Roulette is usually implemented by accumulating probability, fitness of the selected probability is higher, and the fitness of low parent individuals reproduces less or not.

First step of roulette wheel selection method is calculating the function of the parent individual adaptive value, then putting the value of individual to sort ascending and calculate the probability of roulette, and finally, generating from 0 up to 1 of the real number. Selection of the corresponding probability of roulette is performed, and then the individual parent is put as a new parent.

3.4. Crossover and mutation

At a certain probability, two individuals (chromosomes) 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 the crossover is to produce a new gene, that is, the new chromosome. The mutation operation of the gene in the chromosome is executed according to the probability of mutation. At present, the main method of mutation is NOT to perform an operation, which aims to find new genes and overcome the premature convergence. Due to the amount of quality gene, the selection strategy needs to be performed based on the principle of base labeling. In calculating the sequence, initial populations are constructed with n-size arbitrary DNA chains, and coded with the four characters. In the genetic population of these generations, *m*-size DNA individuals were selected as the source of the offspring by a certain probability for the larger adaptive genes having more chance to reproduce.

3.5. Stop criteria

There are two stopping conditions: the maximum number of iterations and the satisfactory solution. Once one pre-specified stopping condition is satisfied, the algorithm ends. Otherwise, it returns to calculate values of fitness function.

4. Implementation of the proposed model

4.1. Individual encoding mode

The quaternary DNA coding, M gene and the calculation of 2 bands interchange are adopted. The numbers 0, 1, 2, 3 denote G, A, T, C respectively, i.e.,

$$G=0(00),\ A=1(01),\ T=2(10),\ C=3(11).$$

The encoding of the individual includes genotype coding and phenotype coding. According to the definition of genotype and phenotype, the expression of the phenotype is directly related to the fitness of the individual, so it is used to express the position of the individual in the search plane (coordinate plane). The experimenter selects eight primer pairs:

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 numbers of bands are the richest in experiments. With the selected eight primer pairs we are able to evaluate the polymorphism of Kiwifruit, so we choose the eight pairs of primers for polymorphism analysis of subculture plantlets of kiwifruit tissue culture.

The corresponding encoding of kiwifruit genomic DNA is as follows:

010000110110, 011110110100, 010000110100, 011100111000,
010100111000, 011100110101, 010100110101, and 011100110100.

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

(7) $$x_i = 4.096\frac{y_i}{2^6 - 1} - 2.048, \quad i = 1, 2.$$

4.2. Individual evaluation method

The Rosenbrock function $F(X)$ is taken in the study for evaluating the optimization problems. The Rosenbrock method is a gradient free minimization algorithm in its basic form. It was introduced by H. H. Rosenbrock and avoids the use of line searches. It is based on orthogonal search directions with alternating minimizations between these and using pattern search at the end of each orthogonal direction search cycle. It is

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

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

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

4.3. Selection operator: using the roulette selection operator

The probability of being chosen of each individual is proportional to its fitness. Here we use the roulette selection strategy to dynamically monitor the kiwifruit's variant seedling in tissue culture and to choose for cultivating the kiwifruit individuals with big fitness.

4.4. Genetic operators

Two kinds of genetic operators are used, crossover and mutation. Among them, the cross operation uses a single point crossover operator; the type of mutation operation uses the basic bit mutation operator. The initial operating parameters of the algorithm are as follows: Population size M=8, crossover probability p_c = 0.6, mutation probability p_m = 0.1, termination generation T = 15.

The genetic data of the individual is read 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 is done again and again until condition is reached.

4.5. Experimental results

Debugging with Java programming, the fitness of E-ACT+M-CAG, E-AAG+M-CTG and E-AAG+M-CAA are taken as examples. The corresponding conclusions are shown in Tables 1-3, and Figs 1-3.

Table 1. The evolution results of E-ACT+M-CAG

Generations	Binary coding	x_1	x_2	Fitness
R_1	011100110101	–0.228	1.33	176.3
R_2	011100110101	–0.228	1.40	183.0
R_3	011100111000	–0.228	1.40	183.0
R_4	011100111001	–0.228	1.60	239.0
R_5	101100111001	–0.228	1.60	239.0
R_6	101100111001	–0.228	1.60	239.05
R_7	101100111001	–0.228	1.60	239.0
R_8	101100111001	–0.228	1.60	239.0
R_9	101100111001	–0.228	1.60	239.0
R_{10}	101100111001	0.813	1.60	239.0
R_{11}	101100001001	0.813	1.66	99.5
R_{12}	101100001001	0.813	1.66	99.5

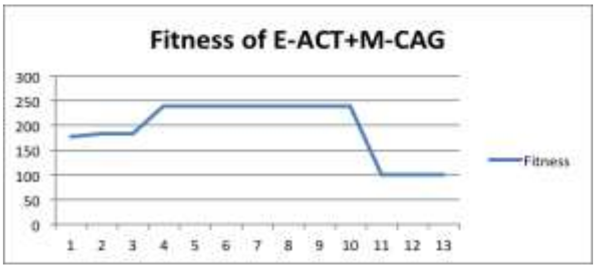

Fig. 1. Fitness of E-ACT+M-CAG

Table 2. The evolution results of E-AAG+M-CTG

Generations	Binary coding	x_1	x_2	Fitness
R_1	010100111000	–0.75	1.60	109.9
R_2	011100110100	–0.23	1.33	165.6
R_3	011100110101	–0.23	1.40	182.7
R_4	011100111000	–0.23	1.60	239.0
R_5	011100111001	–0.23	1.60	239.0
R_6	101100111001	–0.23	1.60	239.0
R_7	101100111001	–0.23	1.60	239.0
R_8	101100111001	–0.23	1.60	239.0
R_9	101100111001	–0.23	1.60	239.0
R_{10}	101100111001	–0.23	1.60	239.0
R_{11}	101101111101	0.813	1.66	99.5
R_{12}	101100001001	0.813	1.66	99.5

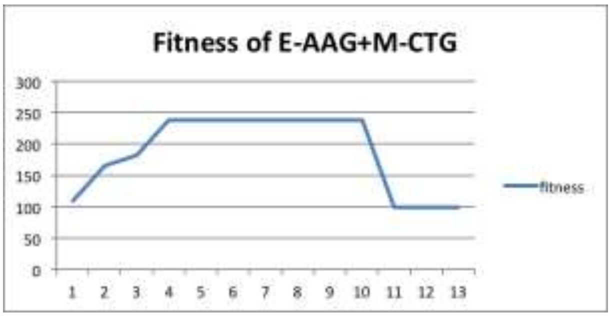

Fig. 2. Fitness of E-AAG+M-CTG

Table 3. The evolution results of E-AAG+M-CAA

Generations	Binary coding	x_1	x_2	Fitness
R_1	010100110101	–0.75	1.40	73.4
R_2	010000110110	–1.00	1.46	24.0
R_3	011100110101	–0.23	1.40	182.7
R_4	011100111000	–0.23	1.60	239.0
R_5	011100111001	–0.23	1.60	239.0
R_6	101100111001	–0.23	1.60	239.0
R_7	101100111001	–0.23	1.60	239.0
R_8	101100111001	–0.23	1.60	239.0
R_9	101100111001	–0.23	1.60	239.0
R_{10}	101100111001	–0.23	1.60	239.0
R_{11}	101100111001	0.813	1.66	99.5
R_{12}	101100001001	0.813	1.66	99.5

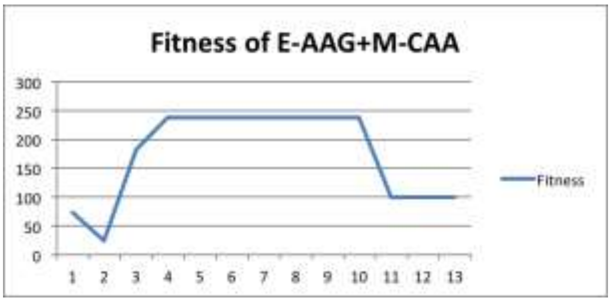

Fig. 3. Fitness of E-AAG+M-CAA

5. Conclusion and further work

From Figs 1-3, the variations of the R_7, R_8, and R_9 three generations are relatively stable. The variation rate of the R_{10} generation changes rapidly. In conclusion, the 9th generation of kiwifruit subculture can better maintain the genetic stability, but variation occurred from the 10th generation on. The result is accompanied with the tissue culture “Hort 16A” of kiwifruit seedling subculture R_7, R_8, R_9, R_{10}, R_{11} (the earliest sample is R_0, and R_1 as subculture) as experimental material. The 15 plant seedlings are randomly chosen from the every subculture seedlings, 1-2 fresh leaves of per plant are picked, put in zip lock bag, marked respectively, and preserved under –80 ℃. 300 mg genomic DNA marked on AFLP are used for endonuclease digestion with 4 h 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. It shows that the proposed method is reasonable and proposes a way to monitor Kiwifruit's variant seedling in tissue culture with information technology. But the samples size and kinds of tissue culture used in the study are relatively small; the proposed method needs to be validated with more samples and kinds of tissue culture. The probabilities of mutation and crossover are fixed, which is not according to practical cases. Our future work is to adopt the probabilities changing according to the length and numbers of chromosomes and other conditions.

References

1. Holland, J. H. Adaptation in Natural and Artificial Systems. Ann Arbor, Michigan, University of Michigan Press, 1975.
2. Aditi, K., Y. T. Oscar, J. Wenyi et al. Iterative Reliability-Based Decoding of Linear Block Codes with Adaptive Belief Propagation. – IEEE Communications Letters, Vol. **9**, 2005, No 1, pp. l067-1069.
3. Michalewicz, Z. Genetic Algorithms + Data Structures = Evolution Programs. Berlin, Springer-Verlag, 1996.
4. Simmons, M. P., L. B. Zhang, C. T. Webb, K. Müller. A Penalty of Using Anonymous Dominant Markers (AFLPs, ISSRs, and RAPDs) for Phylogenetic Inference. – Molecular Phylogenetics and Evolution, Vol. **42**, 2007, No 2, pp. 528-542.
5. Venkat, S. K., P. Bommisetty, M. S. Patil, L. Reddy, A. Chennareddy. The Genetic Linkage Maps of Anthurium Species Based on RAPD, ISSR and SRAP Markers. – Scientia Horticulturae, Vol. **178**, 2014, pp. 132-137.
6. Kayis, S. A. Evaluation of Confidence Limit Estimates of Cluster Analysis on Molecular Marker Data. – Journal of the Science of Food and Agriculture, Vol. **92**, 2012, No 4, pp. 776-780.
7. Goldberg, D. E. Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley, Reading: MA, 1989.
8. Davis, L. Handbook of Genetic Algorithms. New York, Van Nostrand Reinhold, 1991.
9. Back, T. Evolutionary Algorithms in Theory and Practice. Oxford, Oxford University Press, 1996.
10. Jones, D. R., M. Schonlau, W. J. Welch. Efficient Global Optimization of Expensive Black-Box Functions. – Journal of Global Optimization, Vol. **13**, 1998, pp. 445-492.
11. Beghi, A., L. Cecchinato, M. Rampazo. A Multi-Phase Genetic Algorithm for the Efficient Management of Multi-Chiller Systems. – Energy Conversion and Management, Vol. **52**, 2011, pp. 1650-1661.
12. Vakkas, U. S., M. Demirtas. Modeling and Control of V/f Controlled Induction Motor Using Genetic-ANFIS Algorithm. – Energy Conversion and Management, Vol. **50**, 2009, 786-791.
13. Baghernejad, A., M. Yaghoubi. Exergoeconomic Analysis and Optimization of an Integrated Solar Combined Cycle System (ISCCS) Using Genetic Algorithm. – Energy Conversion and Management, Vol. **52**, 2011, pp. 2193-2203.
14. Schippmann, B., H. Burchard. Rosenbrock Methods in Biogeochemical Modelling – A Comparison to Runge-Kutta Methods and Modified Patankar Schemes. – Ocean Model, Vol. **37**, 2011, pp. 112-121.
15. Berardi, M. Rosenbrock-Type Methods Applied to Discontinuous Differential Systems. – Mathematics and Computers in Simulation, Vol. **95**, 2013, pp. 229-243.
16. Lipowski, A., D. Lipowska. Roulette-Wheel Selection via Stochastic Acceptance. – Physica A: Statistical Mechanics and Its Applications, Vol. **391**, 2012, No 6, pp. 2193-2196.
17. Yang, X. F., J. Li, H. Pei et al. Pattern Recognition Analysis of Proteins Using DNA-Decorated Catalytic Gold Nanoparticles. – Small, Vol. **9**, 2013, No 17, pp. 2844-2849.
18. Li, L. B., S. H. He, S. Li et al. A Closer Look at the Russian Roulette Problem: A Re-Examination of the Nonlinearity of the Prospect Theory’s Decision Weight pi. – International Journal of Approximate Reasoning, Vol. **50**, 2009, No 3, pp. 515-520.
19. Li, Z., N. Wang. A Modified DNA Genetic Algorithm for Parameter Estimation of the 2-Chlorophenol Oxidation in Supercritical Water. – Applied Mathematical Modeling, Vol. **01.37**, 2013, No 3, pp. 1137-1146.
20. Lobry, J. R., N. Sueoka. Asymmetric Directional Mutation Pressures in Bacteria. – Genome Biology, Vol. **3**, 2002, No 10, pp. 00-14.
21. Shah, K., A. Krishnamachari. Nucleotide Correlation Based Measure for Identifying Origin of Replication in Genomic Sequences. – BioSystems, Vol. **107**, 2012, No 1, pp. 52-55.
22. Schneider, T. D. A Brief Review of Molecular Information Theory. – Nano Communication Networks, Vol. **1**, 2010, No 3, pp. 173-180.
23. Rosenbrock, H. H. An Automatic Method for Finding the Greatest or Least Value of a Function. – Computer Journal, Vol. **3**, 1960, pp. 175-184.