Research Article

A Two-Stage Exon Recognition Model Based on Synergetic Neural Network

Zhehuang Huang[1,2] and Yidong Chen[2,3]

[1] *School of Mathematics Sciences, Huaqiao University, Quanzhou 362021, China*
[2] *Cognitive Science Department, Xiamen University, Xiamen 361005, China*
[3] *Fujian Key Laboratory of the Brain-Like Intelligent Systems, Xiamen 361005, China*

Correspondence should be addressed to Yidong Chen; ydchen@xmu.edu.cn

Exon recognition is a fundamental task in bioinformatics to identify the exons of DNA sequence. Currently, exon recognition algorithms based on digital signal processing techniques have been widely used. Unfortunately, these methods require many calculations, resulting in low recognition efficiency. In order to overcome this limitation, a two-stage exon recognition model is proposed and implemented in this paper. There are three main works. Firstly, we use synergetic neural network to rapidly determine initial exon intervals. Secondly, adaptive sliding window is used to accurately discriminate the final exon intervals. Finally, parameter optimization based on artificial fish swarm algorithm is used to determine different species thresholds and corresponding adjustment parameters of adaptive windows. Experimental results show that the proposed model has better performance for exon recognition and provides a practical solution and a promising future for other recognition tasks.

1. Introduction

With the completion of human genome project, gene data increase exponentially. Identifying the genes encoding of DNA [1] has important theoretical and practical implications. How to quickly access accurate genetic information is an urgent problem to be solved.

Early exon recognition methods were based mainly on statistical models [2], which get their chromosomal order by statistical analysis of different genes. But with the increase of genomic number, statistical methods cannot meet the need for rapid recognition of exons. At present, exon recognition methods based on digital signal processing have also been widely used [3–5]. These techniques select a suitable mapping method and transformation method to get spectral values and identify exons according to fixed length window. Limitations of these methods include slow recognition speed and inability to accurately determine the threshold for different species.

Synergetic theory [6] is the science proposed by Haken to describe high dimension and nonlinear problem as a set of low-dimension nonlinear equations. One advantage of synergetic neural network is that the method is robust against noise and the method can better handle the fuzzy matching problem [7–9]. Exon recognition can also be considered a problem of pattern recognition, for which the proposed method can be used to solve.

Artificial fish swarm algorithm (AFSA) [10, 11] is a class of swarm intelligence optimization algorithms based on the behavior of animals proposed in 2002; the basic idea of AFSA is to imitate the fish behaviors such as praying, swarming, and following. AFSA is very suitable for solving a variety of numerical optimization problems, making the algorithm become a hot topic in the current optimization field quickly. Because of simplicity in principle and good robustness, AFSA has been applied successfully to all kinds of optimization problems such as image segmentation [12], color quantization [13], neural network [14], fuzzy logic controller [15], multi-robot task scheduling [16], fault diagnosis in mine hoist [17], data clustering [18], and other areas.

In this paper, we proposed a two-stage exon recognition model based on synergetic neural network and artificial fish
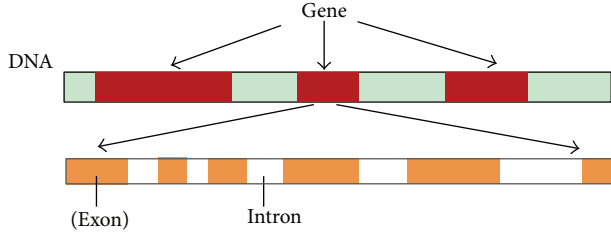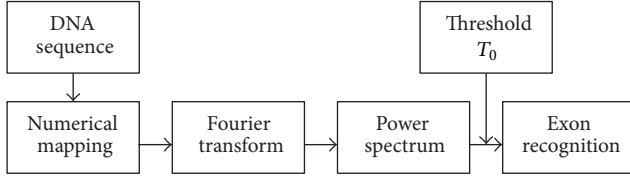
FIGURE 1: Structure of eukaryotic DNA sequence.

FIGURE 2: Exon recognition algorithm based on 3-Cycle spectrum.

TABLE 1

$\Delta x[n]$	1	−1	1	−1	−1	1	1
$\Delta y[n]$	1	1	−1	−1	−1	1	−1
$\Delta z[n]$	1	−1	−1	1	1	1	−1

swarm algorithm. This paper is organized as follows. Firstly, traditional exon recognition method based on digital signal processing and related work are presented. Secondly, an exon recognition model based on synergetic neural network and parameter optimization method based on artificial fish swarm algorithm are introduced. Finally some experimental tests, results, and conclusions are given on the systems.

2. Introduction to Exon Recognition Method Based on Digital Signal Processing

The gene is usually divided into many fragments. The coding sequence is called exons and noncoding part is called introns, as shown in Figure 1.

The objective of gene recognition is to identify the exons of DNA sequence. Gene recognition based on digital signal processing methods consists of several steps [19, 20]. First, gene sequences are transformed into digital symbol sequences using mapping methods [21–24]. This is followed by calculation of the corresponding frequency value by fast Fourier transform and the 3-Cycle properties of the spectrum are then used to identify exons [25, 26]. Finally, fixed sliding window method is used for automatic exon recognition, as shown in Figure 2.

2.1. Z-Curve Mapping.
In order to make digital processing, we must transform four nucleotide sequences A, T, G, and C into their corresponding numeric sequence based on certain rules.

Let the four instruction sequences be $\{u_b[n]\}$, $b \in I = \{A, C, G, T\}$, and cumulative sequence b_n $(n = 0, 1 \ldots, N - 1)$

is $b_n = \sum_{i=0}^{n-1} u_b[i]$; then we can define three sequences $x[n]$, $y[n]$, and $z[n]$:

$$x[n] = 2(A_n + G_n) - n,$$
$$y[n] = 2(A_n + C_n) - n, \tag{1}$$
$$z[n] = 2(A_n + T_n) - n.$$

Let

$$x[-1] = 0, \qquad y[-1] = 0, \qquad z[-1] = 0,$$
$$\Delta x[n] = x[n] - x[n-1], \qquad \Delta y[n] = y[n] - y[n-1],$$
$$\Delta z[n] = z[n] - z[n-1]. \tag{2}$$

Thus we can get the Z-curve mapping:

$$\begin{pmatrix} \Delta x[n] \\ \Delta y[n] \\ \Delta z[n] \end{pmatrix} = \begin{pmatrix} 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \begin{pmatrix} u_A[n] \\ u_C[n] \\ u_G[n] \\ u_T[n] \end{pmatrix}. \tag{3}$$

For example, the DNA sequence of $S(n)$ is ACGTTAG; then the corresponding Z-curve mapping sequence is shown in Table 1.

2.2. The Power Spectrum.
To study the characteristics of DNA coding sequences (exons), we can do the discrete Fourier transform (DFT), respectively, for the instruction sequences:

$$U_b[k] = \sum_{n=0}^{N-1} u_b[n] e^{-j(2\pi nk/N)}, \quad k = 0, 1, \ldots, N - 1. \tag{4}$$

Thus we can calculate the power spectrum:

$$P_z[k] = |\Delta X[k]|^2 + \|\Delta Y[k]\|^2 + |\Delta Z[k]|^2,$$
$$k = 0, 1, \ldots, N - 1, \tag{5}$$

where $\Delta X[k]$, $\Delta Y[k]$, and $\Delta Z[k]$ are the Fourier transform of $\Delta x[n]$, $\Delta y[n]$, and $\Delta z[n]$, respectively.

The spectral peaks of exon sequences are larger in $k = N/3$ and $k = 2N/3$ of the power spectrum curve, while they are not similar for intron. This statistical phenomenon is known as 3-Cycle. Suppose that the average power spectrum of DNA sequences is

$$\overline{E} = \frac{\sum_{k=0}^{N-1} P[k]}{N}. \tag{6}$$

The power spectrum ratio of the DNA sequence and the average spectrum of the entire sequence are known as SNR (signal-to-noise ratio):

$$R = \frac{P[N/3]}{\overline{E}}. \tag{7}$$

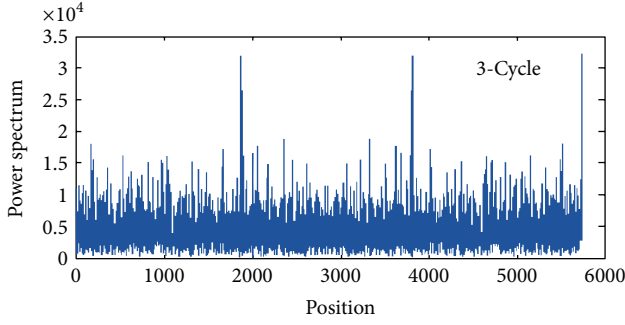Figure 3 shows the power spectrum of viral genes.

FIGURE 3: The power spectrum of viral gene sequence.

From Figure 3, we can see that the spectrum presents obvious 3-Cycle. The peaks appear roughly in 2000, 4000, and 6000. So the exon segment can be determined, enabling the recognition of genes.

The highest point of power spectrum may not appear in $k = N/3$ and $k = 2N/3$ but occur in the surrounding. So we can calculate average SNR R_1 and R_2 of intervals $[N/3 - \gamma, N/3 + \gamma]$ and $[2N/3 - \gamma, 2N/3 + \gamma]$, respectively:

$$R_1 = \frac{\sum_{k=N/3-\gamma}^{N/3+\gamma} P[E]}{(2\gamma + 1)\,\overline{E}}, \qquad R_2 = \frac{\sum_{k=2N/3-\gamma}^{2N/3+\gamma} P[E]}{(2\gamma + 1)\,\overline{E}}. \tag{8}$$

2.3. Automatic Recognition Algorithm Based on Fixed Sliding Windows. Supposed M is the length of fixed window; we can do four discrete Fourier transforms (DFT) for instruction sequences $\{u_b[n]\}$ ($0 \le n \le N - 1$),

$$U_b[k] = \sum_{i=n-(M-1)/2}^{i=n+(M-1)/2} u_b[i]\, e^{-j(2\pi ik/M)}, \quad k = 0, 1, \ldots, M - 1. \tag{9}$$

Then the total spectrum $p\,(n; M/3)$ at position $M/3$ is

$$P\left[\frac{M}{3}\right] = \left|U_A\left[\frac{M}{3}\right]\right|^2 + \left|U_T\left[\frac{M}{3}\right]\right|^2$$
$$+ \left|U_G\left[\frac{M}{3}\right]\right|^2 + \left|U_C\left[\frac{M}{3}\right]\right|^2 \tag{10}$$
$$\stackrel{\Delta}{=} p\left(n; \frac{M}{3}\right).$$

3. Related Work

The SNR of exon sequences reflects the distribution of spectrum peak. SNR greater than a given threshold is a characteristic of exons, while introns generally do not have this property.

Protein coding regions and noncoding regions can be distinguished using the value of SNR, but this method still has a large predictive error because the spectrum peak varies amongst different biological categories. A fixed threshold is unreasonable to use for different biological categories. Therefore, determining the SNR threshold has great significance for exon recognition. Note that it is difficult to find the proper prediction threshold for biological categories when relying only on prior biological knowledge.

Xu [27] proposed a method based on bootstrap algorithm to determine the best SNR threshold that can be obtained from marked exon sequences. The results of that study showed that the average prediction accuracy of the method was 81%, which is 19% higher than other methods that employ empirical thresholds. In paper [28], a novel model was proposed to determine the SNR threshold based on the means of biological categories and improved the recognition performance to some extent.

But all the methods mentioned above have problems, such as slow recognition speed, inaccurate determination of the threshold for different species, and the requirement to know the exon fragments of DNA sequences. In the following sections, we propose a novel two-stage exon recognition model based on synergetic neural network and artificial fish swarm algorithm to better deal with these problems.

4. A Novel Two-Stage Exon Recognition Model

In this section, a two-stage exon recognition model is presented. In the first stage, synergetic neural network is used to determine initial exon intervals. In the second stage, final accurate exon intervals determination based on adaptive sliding window and parameter optimization algorithm are introduced.

4.1. Initial Exon Intervals Determination Based on Synergetic Neural Network. The basic principle of synergetic neural network [29, 30] is that the pattern recognition procedure can be viewed as the competition progress of many order parameters. The strongest order parameter will win by competition and desired pattern will be recognized.

A pattern that remained to be recognized, q, is constructed by a dynamic process which translates q into one of prototype pattern vectors v_k through status $q(t)$; namely, this prototype pattern is closest to $q(0)$. The process is described as the following equation:

$$q \longrightarrow q(t) \longrightarrow v_k. \tag{11}$$

A dynamic equation can be given for an unrecognized pattern q:

$$\dot{q} = \sum_{k=1}^{M} \lambda_k v_k \left(v_k^+ q\right) - B \sum_{k' \neq k} \left(v_{k'}^+ q\right)^2 \left(v_k^+ q\right) v_k \tag{12}$$
$$- C\left(q^+ q\right) q + F(t),$$

where q is the status vector of input pattern with initial value q_0, λ_k is attention parameter, v_k is prototype pattern vector, and v_k^+ is the adjoint vector of v_k that satisfies

$$\left(v_k^+, v_{k'}^T\right) = v_k^+ \cdot v_{k'}^T = \delta_{kk'}. \tag{13}$$

TABLE 2: The signal-to-noise ratio of four different gene sequences.

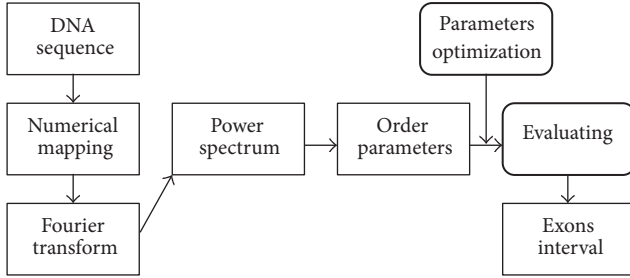| Gene categories	Exon			Intron		
	Number	R-mean	Variance	Number	R-mean	Variance
Human	35	3.02	3.071	26	0.82	0.533
Mus musculus	357	2.46	2.508	275	0.68	0.414
Sewer rat	45	3	5.233	35	0.83	0.624
Mammalian	827	2.72	6.243	626	0.67	0.394

FIGURE 4: Exon recognition based on synergetic neural network.

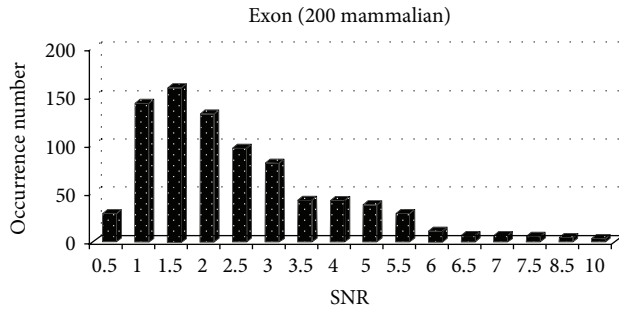FIGURE 6: The SNR distribution of 200 mammalian introns.

FIGURE 5: The SNR distribution of 200 mammalian exons.

Corresponding dynamic equation of order parameters is

$$\dot{\xi}_k = \lambda_k \xi_k - B \sum_{k' \neq k} \xi_{k'}^2 \xi_k - C \left| \sum_{k'=1}^{M} \xi_{k'}^2 \right| \xi_k. \tag{14}$$

Haken has proved that when $\lambda_k = c$ ($c > 0$), the largest initial order parameter will win and the network will then converge.

We firstly introduce the synergetic theory to exon recognition; an exon recognition algorithm based on synergetic neural network is shown in Figure 4.

We use synergetic neural network and N equal method to quickly determine the initial exon region, as shown in Algorithm 1.

4.2. Get Precise Exon Intervals Using Adaptive Smoothing Window. We can obtain several possible exon intervals by Algorithm 1. In this section, we propose an adaptive sliding window algorithm to get more accurate intervals, as shown in Algorithm 2.

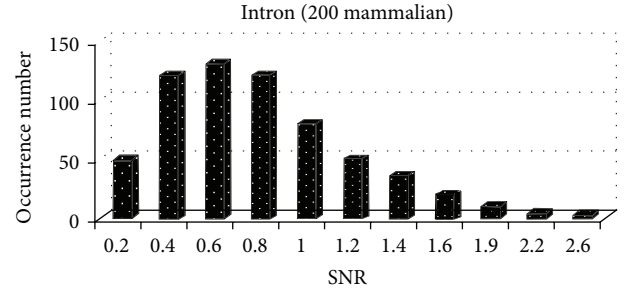*4.3. Parameter Optimization Based on Artificial Fish Swarm Algorithm.* The parameters T_0 and γ directly influence the performance of exon recognition. The adjustment of the parameters is a global behaviour and has no general research theory to control the parameters in the recognition process at present. In this section, artificial fish swarm algorithm is used to search the global optimum parameters (T_0, γ) in the corresponding parameter space.

The parameter optimization based on artificial fish swarm algorithm is shown as Algorithm 3.

5. Experiment

5.1. Data Description. In our experiments, we use some gene sequences provided by Chinese Graduate Mathematical Contest in Modeling. Chinese graduate Mathematical Contest in modeling is aimed at improving the students' comprehensive abilities of mathematical modeling and computer to solve practical problems. From different points of view, the integrated use of a variety of mathematical methods established the mathematical model of the characteristic.

We selected 100 human gene sequences, 100 rodent gene sequences (including *Mus musculus* and Sewer rat), and 200 mammalian gene sequences for testing. The signal-to-noise ratios of the sequences are gotten by SPSS statistical analysis software, as shown in Table 2.

From Table 2, we can find out that the difference between SNR standard deviation of exons is greater than SNR standard deviation of introns.

At the same time, we analyze the SNR distribution of exons and introns of 200 mammalian gene sequences, as shown in Figure 5 and Figure 6.

From Figure 5 and Figure 6, we can see that the mammalian introns are mostly less than 2, while exons are mostly distributed in the range of $[0, 2]$, which accounts for 55.38%. Therefore, it is unreasonable to set SNR threshold of different categories as fixed value. How to accurately determine SNR

(1) Let S is a given gene sequence, S_{start} and S_{end} are the beginning and end of the sequence respectively, T_0 is throdthod of spectral values;

(2) Using $Z\_$Curve mapping converted gene sequence to the corresponding numeric sequence;

(3) Using fast Fourier transform to get spectral values R_1 and R_2 according to the formula (8);

(4) Calculating gene sequence order parameter:
$$\xi_1 = \frac{R_1}{R_1 + R_2}, \qquad \xi_2 = \frac{R_2}{R_1 + R_2};$$

(5) Setting network parameter λ_k and B, C;

(6) Order parameter evolution according to formula (14);

(7) If $\xi_1 > T_0$ and $\xi_2 > T_0$, then $[S_{\text{start}}, S_{\text{end}}]$ is recorded as a possible interal, and S is divided equally into n interals $S_1, S_2, \ldots, S_n$, Repeat step 1 to step 7;

(8) End.

ALGORITHM 1: Determination of initial exon region based on synergetic neural network.

(1) Let W is a given gene sequence, W_{start} and W_{end} are the beginning and the end of the sequence respectively;

(2) Using $Z\_$Curve mapping converted gene sequence $[W_{\text{start}}, W_{\text{end}}]$ to the corresponding numeric sequence;

(3) Using fast Fourier transform to get spectral values;

(4) Calculating gene sequence order parameter:
$$\xi_1 = \frac{R_1}{R_1 + R_2}, \qquad \xi_2 = \frac{R_2}{R_1 + R_2};$$

(5) Order parameter evolution according to formula (14);

(6) If $\xi_1 > T_0$, $\xi_2 > T_0$ and $W_{\text{start}} + \gamma < W_{\text{end}} - \gamma$, Then $W_{\text{start}} = W_{\text{start}} + \gamma$, $W_{\text{end}} = W_{\text{end}} - \gamma$, Repeat step 2 to step 6;

(7) Output the final interval $[W_{\text{start}}, W_{\text{end}}]$.

ALGORITHM 2: Precise exon regions based on adaptive smoothing window.

threshold of each kind of biological gene has important significance.

5.2. Experiment Results. Suppose that sensitivity $S_N = T_P/(T_P + F_N)$ and specificity $S_P = T_N/(T_N + F_P)$, where T_P is the number of exons which are correctly identified, T_N is the number of introns which are correctly identified, F_P is the number of exons which are not correctly identified, and F_N is the number of introns which are not correctly identified. Then we can compute the accurate rate $A_c = (S_N + S_P)/2$.

For comparison, we use four strategies.

(1) Baseline: automatic recognition algorithm with threshold $R_0 = 2$.

(2) Bootstrap: the threshold selection algorithm based on bootstrap method.

(3) SNN: exon recognition based on synergetic neural network.

(4) SNN + AFSA: two-stage exon recognition model based on synergetic neural network and artificial fish swarm algorithm.

The testing performance of Baseline is shown as in Table 3.

The experiments showed that when the exon length is short, the recognition accuracy rate is low. In the short gene coding sequence, 3-base periodicity is not absolutely satisfied. In our experiments, we complete a two-stage exon recognition model based on synergetic neural network and artificial fish swarm algorithm. The parameter settings of artificial fish swarm algorithm are shown in Table 4.

In the experiment, we set the recognition accuracy rate as score function.

The testing performance of SNN + AFSA is shown as in Table 5.

Table 5 shows that the two-stage exon recognition algorithm improves precision compared to the Baseline system. Experiments also indicate that the improved model has a more powerful global exploration ability and a reasonable convergence speed.

The accurate rate A_c of different methods is shown in Table 6.

Detailed comparisons of results are given in Table 6. Experimental results show that the proposed model SNN and SNN + AFSA have good performance for exon recognition. The accurate rate we obtained for all four corpuses is comparable to the state-of-the-art systems, such as Baseline and bootstrap method. Through the evaluating of order parameter equation of SNN to obtain the best threshold, we can further improve the exon recognition performance.

At the same time, we can see that the performance of SNN + AFSA is better than SNN model. This is because

> (1) Initialize the parameters of artificial fish, such as *step*, *visual*, the number of exploratory, maximum number of iterations, and randomly generated *n* fishes;
> (2) Set bulletin board to record the current status of each fish, and select the optimal value;
> (3) Implementation of prey behavior, swarm behavior and follow behavior;
> (4) Optimal value in bulletin board is updated;
> (5) If termination condition is satisfied, output the result; otherwise return to step 2.

ALGORITHM 3: Parameter optimization based on artificial fish swarm algorithm.

TABLE 3: The testing performance of Baseline.

Gene categories	T_P	F_N	S_N	T_N	F_P	S_P	A_c
Human	17	18	0.485	24	2	0.923	0.71
Mus musculus	146	211	0.409	271	4	0.985	0.70
Sewer rat	17	28	0.378	31	4	0.886	0.63
Mammalian	369	458	0.446	621	5	0.992	0.72

TABLE 4: The parameter settings of artificial fish swarm algorithm.

Algorithm	Fish number	Visual	Delta	Step	Number of iterations
AFSA	100	2.85	9	1	60

TABLE 5: The test performance of SNN + AFSA.

Gene categories	T_P	F_N	S_N	T_N	F_P	S_P	A_c
Human	30	5	0.857	19	7	0.731	0.79
Mus musculus	295	62	0.826	220	55	0.80	0.81
Sewer rat	36	9	0.80	28	7	0.80	0.80
Mammalian	630	197	0.762	607	19	0.97	0.87

TABLE 6: The test performance comparison among different methods.

Gene categories	Baseline	Bootstrap	SNN	SNN + AFSA
Human	0.71	0.76	0.78	0.79
Mus musculus	0.70	0.78	0.80	0.81
Sewer rat	0.63	0.75	0.77	0.80
Mammalian	0.72	0.84	0.85	0.87

the attention parameters are very important for SNN and optimization algorithm is essential for better performance. Experimental results show that improved AFSA algorithm has better global and local parameter searching capabilities and thus a better recognition result.

It is worth noting that experimental results show that run times of our proposed model reduced with good speedup ratio compared with Baseline. Further studies show that the procedure exhibits data parallelism, so it can be effectively parallelized by running it concurrently. In the future work, we will utilize parallel processing techniques for rapid exon recognition based on SNN to further reduce the run time.

6. Conclusions

In the paper, we proposed a two-stage exon recognition model based on synergetic neural network and artificial fish swarm algorithm. Experiments show that the proposed model can improve the precision of exon recognition.

We got the following conclusions.

(1) The exon recognition procedure can be viewed as the competition progress of many order parameters. The proposed model based on synergetic neural network and *N* equal method can quickly determine the exon intervals.

(2) Artificial fish swarm algorithm has both global and local search ability and can effectively choose the parameters of our proposed model.

(3) Using *N* equal algorithm to obtain exon intervals may still miss some intervals which are in the middle; we will further improve the algorithm or use different pattern recognition algorithm in the future.

It must be noted that, although we have made some efforts to explore the intelligent exon recognition algorithm in this paper. But due to the special nature of life science itself, there are many problems such as how to accurately determine that the exon interval needs further study. But we believe that with the development of social progress and technology, gene identification technology will become increasingly perfect; we expect it can bring gospel to manking in the near future.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] C. B. Burge and S. Karlin, "Finding the genes in genomic DNA," *Current Opinion in Structural Biology*, vol. 8, no. 3, pp. 346–354, 1998.

[2] Z. Wang, Y. Chen, and Y. Li, "A brief review of computational gene prediction methods," *Genomics Proteomics Bioinformatics*, vol. 2, no. 4, pp. 216–221, 2004.

[3] S. D. Sharma, K. Shakya, and S. N. Sharma, "Evaluation of DNA mapping schemes for exon detection," in *Proceedings of the International Conference on Computer, Communication and Electrical Technology (ICCCET '11)*, pp. 71–74, March 2011.

[4] D. Kotlar and Y. Lavner, "Gene prediction by spectral rotation measure: a new method for identifying protein-coding regions," *Genome Research*, vol. 13, no. 8, pp. 1930–1937, 2003.

[5] C. Yin and S. S.-T. Yau, "Prediction of protein coding regions by the 3-base periodicity analysis of a DNA sequence," *Journal of Theoretical Biology*, vol. 247, no. 4, pp. 687–694, 2007.

[6] H. Haken, *Synergetic Computers and Cognition-A Top-Down Approach to Neural Nets*, Springer, Berlin, Germany, 1991.

[7] J. Shao, J. Gao, and X. Z. Yang, "Synergetic face recognition algorithm based on ICA," in *Proceedings of the International Conference on Neural Networks and Brain*, vol. 1, pp. 249–253, Beijing, China, October 2005.

[8] Z. Jiang and R. A. Dougal, "Synergetic control of power converters for pulse current charging of advanced batteries from a fuel cell power source," *IEEE Transactions on Power Electronics*, vol. 19, no. 4, pp. 1140–1150, 2004.

[9] X. L. Ma and L . C. Jiao, "Reconstruction of order parameters based on immunity clonal strategy for image classification," in *Image Analysis and Recognition*, A. Campilho and M. Kamel, Eds., vol. 3211 of *Lecture Notes in Computer Science*, pp. 455–462, Springer, Berlin, Germany, 2004.

[10] X. L. Li, S. H. Feng, J. X. Qian, and F. Lu, "Parameter tuning method of robust pID controller based on artificial fish school algorithm," *Chinese Journal of Information and Control*, vol. 33, no. 1, pp. 112–115, 2004.

[11] X. L. Li, F. Lu, G. H. Tian, and J. X. Qian, "Applications of artificial fish school algorithm in combinatorial optimization problems," *Chinese Journal of Shandong University: Engineering Science*, vol. 34, no. 5, pp. 64–67, 2004.

[12] W. Tian, Y. Geng, J. Liu, and L. Ai, "Optimal parameter algorithm for image segmentation," in *Proceedings of the 2nd International Conference on Future Information Technology and Management Engineering (FITME '09)*, pp. 179–182, December 2009.

[13] D. Yazdani, H. Nabizadeh, E. M. Kosari, and A. N. Toosi, "Color quantization using modified artificial fish swarm algorithm," in *Proceedings of the International conference Artificial Intelligence*, vol. 7106 of *Lecture Notes in Artificial Intelligence*, pp. 382–391, 2011.

[14] M. Zhang, C. Shao, F. Li, Y. Gan, and J. Sun, "Evolving neural network classifiers and feature subset using artificial fish swarm," in *Proceedings of the IEEE International Conference on Mechatronics and Automation (ICMA '06)*, pp. 1598–1602, Luoyang, China, June 2006.

[15] D. Chen, L. Shao, Z. Zhang, and X. Yu, "An image reconstruction algorithm based on artificial fish-swarm for electrical capacitance tomography system," in *Proceedings of the 6th International Forum on Strategic Technology (IFOST '11)*, pp. 1190–1194, August 2011.

[16] C.-J. Wang and S.-X. Xia, "Application of probabilistic causaleffect model based artificial fish-swarm algorithm for fault diagnosis in mine hoist," *Journal of Software*, vol. 5, no. 5, pp. 474–481, 2010.

[17] W. Tian, Y. Tian, L. Ai, and J. Liu, "A new optimization algorithm for fuzzy set design," in *Proceedings of the International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC '09)*, vol. 2, pp. 431–435, August 2009.

[18] Y. Cheng, M. Jiang, and D. Yuan, "Novel clustering algorithms based on improved artificial fish swarm algorithm," in *Proceedings of the 6th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD '09)*, vol. 3, pp. 141–145, August 2009.

[19] M. J. Berryman, A. Allison, C. R. Wilkinson, and D. Abbott, "Review of signal processing in genetics," *Fluctuation and Noise Letters*, vol. 5, no. 4, pp. 13–35, 2005.

[20] N. von Öhsen, I. Sommer, R. Zimmer, and T. Lengauer, "Arby: automatic protein structure prediction using profile-profile alignment and confidence measures," *Bioinformatics*, vol. 20, no. 14, pp. 2228–2235, 2004.

[21] D. Kotlar and Y. Lavner, "Gene prediction by spectral rotation measure: a new method for identifying protein-coding regions," *Genome Research*, vol. 13, no. 8, pp. 1930–1937, 2003.

[22] C. Yin and S. S.-T. Yau, "Prediction of protein coding regions by the 3-base periodicity analysis of a DNA sequence," *Journal of Theoretical Biology*, vol. 247, no. 4, pp. 687–694, 2007.

[23] A. Rushdi and J. Tuqan, "Gene identification using the stroke Z sign-curve representation," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '06)*, vol. 2, pp. 1024–1027, May 2006.

[24] S. D. Sharma, K. Shakya, and S. N. Sharma, "Evaluation of DNA mapping schemes for exon detection," in *Proceedings of the International Conference on Computer, Communication and Electrical Technology (ICCCET '11)*, pp. 71–74, March 2011.

[25] D. Kotlar and Y. Lavner, "Gene prediction by spectral rotation measure: a new method for identifying protein-coding regions," *Genome Research*, vol. 13, no. 8, pp. 1930–1937, 2003.

[26] L. Bo and K. Ding, "Graphical approach to analyzing DNA sequences," *Journal of Computational Chemistry*, vol. 26, no. 14, pp. 1519–1523, 2005.

[27] S. L. Xu, *Threshold selection of gene prediction Based on Bootstrap algorithm [M.S. thesis]*, University of Electronic Science and Technology, Sichuan, China, 2011.

[28] J. F. Shao, X. H. Yan, and S. Shao, "SNR of DNA sequences mapped by general affine transformations of the indicator sequences," *Journal of Mathematical Biology*, vol. 67, no. 2, pp. 433–451, 2013.

[29] Z. Jiang and R. A. Dougal, "Synergetic control of power converters for pulse current charging of advanced batteries from a fuel cell power source," *IEEE Transactions on Power Electronics*, vol. 19, no. 4, pp. 1140–1150, 2004.

[30] J. Gao, H. Dong, J. Shao, and J. Zhao, "Parameters optimization of synergetic recognition approach," *Chinese Journal of Electronics*, vol. 14, no. 2, pp. 192–197, 2005.