

Genetics Pattern Recognition by Using Intelligent Techniques

أ.د. باسل يونس الخياط د. عمر صابر قاسم
كلية علوم الحاسبات والرياضيات
جامعة الموصل

Abstract

In this research a light is shed on converting DNA series to amino acid being responsible for forming protein through intelligent techniques. Some comparisons have been made between particularly Artificial Neural Network, Fuzzy Logic and Genetic Algorithms for discovering the powerful and the week ones in particle way. The hybrid operation has been made between Artificial Neural Network and Fuzzy Logic t to get a hybrid techniques in a new formulae having robust results than the original ones.

الملخص

(DNA)

- :

،(DNA)

،(DNA) وذلك

- **:Patterns**

(Object)

(Pattern Class)

(Classifier) والتي

(Machine)

تقوم بمهام عملية التصنيف [Joshi, et al.; 1997].

- **: The Patterns Classification Stage**

- :

(DNA)

(Data Base)

. EMBL و GenBank و NCBI

(International Union of Pure and Applied Chemistry)

IUPAC

:(DNA)

،(A,C, G, T)

.(٦٤)

(٢٠)

(١)

(DNA)

(DNA)

:(١)

	T	C	A	G
T	TTT - Phe (F) TTC - Phe (F) TTA - Leu (L) TTG - Leu (L)	TCT - Ser (S) TCC - Ser (S) TCA - Ser (S) TCG - Ser (S)	TAT - Tyr (Y) TAC - Tyr (Y) TAA - STOP TAG - STOP	TGT - Cys (C) TGC - Cys (C) TGA - STOP TGG - Trp (W)
C	CTT - Leu (L) CTC - Leu (L) CTA - Leu (L) CTG - Leu (L)	CCT - Pro (P) CCC - Pro (P) CCA - Pro (P) CCG - Pro (P)	CAT - His (H) CAC - His (H) CAA - Gln (Q) CAG - Gln (Q)	CGT - Arg (R) CGC - Arg (R) CGA - Arg (R) CGG - Arg (R)
A	ATT - Ile (I) ATC - Ile (I) ATA - Ile (I) ATG - Met (M) START	ACT - Thr (T) ACC - Thr (T) ACA - Thr (T) ACG - Thr (T)	AAT - Asn (N) AAC - Asn (N) AAA - Lys (K) AAG - Lys (K)	AGT - Ser (S) AGC - Ser (S) AGA - Arg (R) AGG - Arg (R)
G	GTT - Val (V) GTC - Val (V) GTA - Val (V) GTG - Val (V)	GCT - Ala (A) GCC - Ala (A) GCA - Ala (A) GCG - Ala (A)	GAT - Asp (D) GAC - Asp (D) GAA - Glu (E) GAG - Glu (E)	GGT - Gly (G) GGC - Gly (G) GGA - Gly (G) GGG - Gly (G)

:

-

(DNA)

:



.(DNA)

: (١)

،(Neural Networks)

(Genetic Algorithm)

(Neuro Fuzzy)

(Fuzzy Logic)

.(Amino Acid)

(DNA)

(A,C,G,T)

: (64)

$4^3 = 64$.

:(DNA) -

(BP)

(DNA)

(DNA)

:(BP) -

(BP)

المخفية (Hidden Layers) -1

العقد (Nodes) -2

(Activation Functions) -3

.(Error Training) -4

: **٩- تحديد**

.(BP)

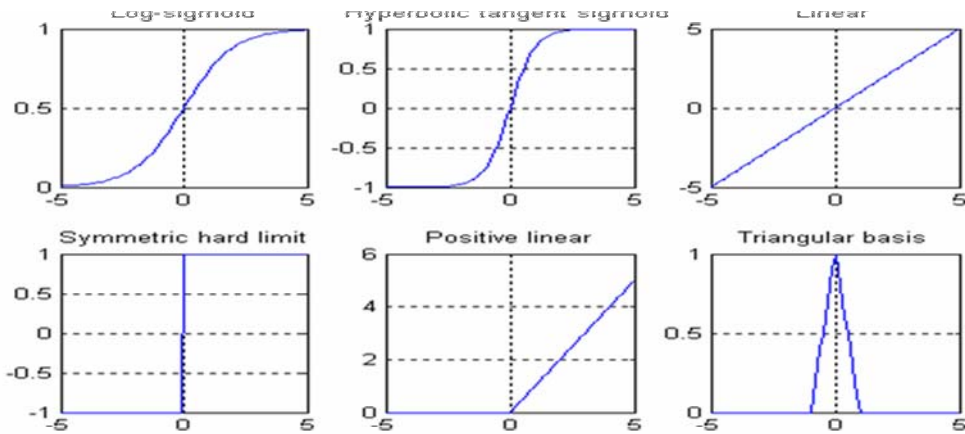
.(Efficiency)

$$EFF = 1/(MSE * Time) \quad (1)$$

(Positive linear)

،(DNA)

: (٢)



: (٢)

(Linear Function)

: ()

مقياس الكفاءة <i>EFF</i>	الزمن بالثانية <i>Time</i>	معدل مربع الخطأ <i>MSE</i>	عدد العقد في الطبقة المخفية	نوع دالة التنشيط <i>Type of Activation Function</i>
0.0767	0.686	19.016	38	Log-sigmoid
0.0530	0.739	25.530	4	Hyperbolic tangent sigmoid
0.0448	0.565	39.483	25	Linear
0.0444	0.532	42.343	45	Symmetric hard limit
0.1234	0.776	10.445	15	Positive linear
0.0762	0.738	17.792	5	Triangular basis

(EFF)

(٢)

(١٥)

،(Positive linear)

.(٢)

: (BP)

Conjugate)

،(Levenberg-Marquardt)

Polak Fletcher-Reeves Powell-Beale restarts)

(Gradient

(BFGS)

(Gradient descent) خوارزمية

(Ribiere

.(quasi-Newton)

.(MATLAB)

(DNA)

(BP)

:

(٣)

: (٣)

مقياس الكفاءة <i>EFF</i>	الزمن بالثانية <i>Time</i>	معدل مربع الخطأ <i>MSE</i>	خوارزمية التدريب <i>Training Algorithm</i>
0.0975	0.982	10.445	Levenberg-Marquardt
0.0396	0.775	32.613	Powell-Beale restarts
0.0519	1.220	15.779	Fletcher-Reeves
0.0334	0.719	41.664	Polak-Ribiere
0.0028	0.550	657.91	Gradient descent
0.0313	0.983	32.467	BFGS quasi-Newton

(٣)

(Levenberg-Marquardt)

(٢) و (٣)

النووي (DNA)

(BP)

:

(BP)

: (٤)

(BP)	
<i>Positive linear</i>	
<i>Linear</i>	
<i>Levenberg-Marquardt</i>	

:(DNA)

-

(3) و (64)

.(Amino Acid)

(A,C,G,T) (DNA)

(Simulation)

الاختبار (Test)

(٥)

(DNA)

(BP)

،(EFF)

.(GA)

(BP)

: ()

مقياس الكفاءة (EFF)	الزمن بالثانية (Time)	معدل مربع الخطأ (MSE)	نوع التقنية الذكية (Type of Technique)
1.539 e+3	8.661	7.501 e-5	الخوارزمية الجينية (GA)
0.0975	0.982	10.445	الشبكة العصبية الاصطناعية (BP)

(٥)

،(BP)

(Simulation)

(DNA)

(BP)

:(DNA)

١٢ - تطبيق

IF)

(THEN

،(DNA)

(3) (64) ،(P) وهي

$$P = [p1; p2; p3] , p1 = P(:,1) , p2 = P(:,2) , p3 = P(:,3)$$

$(p_i, i = 1,2,3)$

(n)

(n^3)

.(p1,p2,p3)

(,cσ)

(Gauss)

(a,b,c,d)

(Trapezoidal)

،(DNA)

.(Rules)

:

-

Genetic)

(Neural Networks)

،(Fuzzy Logic)

(Algorithm

،(DNA)

:

.(MSE)

-1

.(Time)

-2

.(Simulation)

-3

) (٦)
 .(
 : ()

(Time)	(MSE)	(Simulation)	(Type of Technique)
			(NN)
			(GA)
			(FL)

(٦)

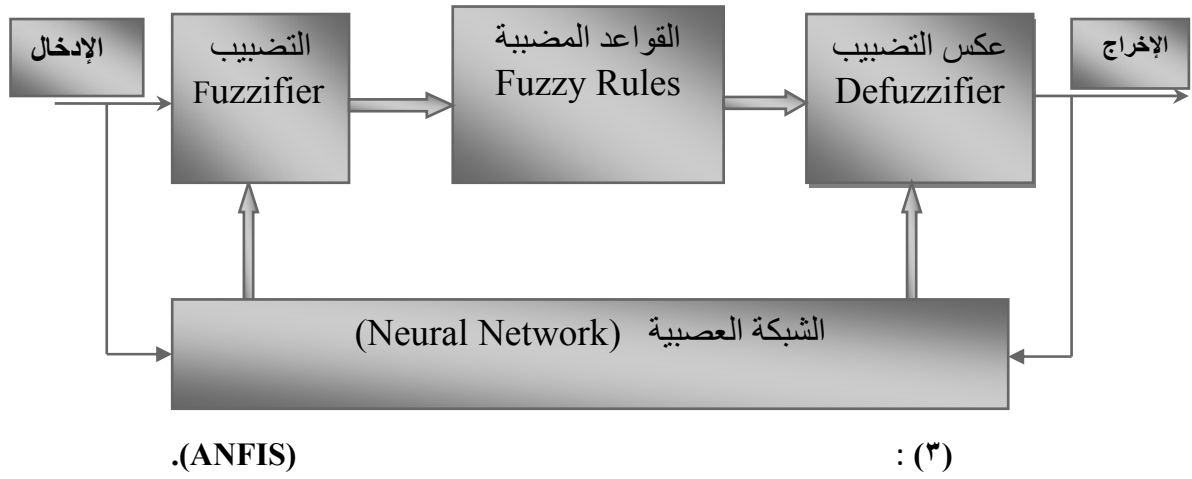
.(DNA)

:

.(BP)

(Gradient Descent)

(Sugeno)
 (BP) (LS)
 (BP)
 .(ANFIS)



(ANFIS)

(ANFIS)

: (ANFIS)

-

(DNA)

(ANFIS)

$(p_i, i = 1,2,3)$

$(P = [p1;p2; p3])$

(Triangular)

:

.(ANFIS)

: (V)

مقياس الكفاءة (EFF)	الزمن بالثانية (Time)	معدل مربع الخطأ (MSE)	الإدخال الثالث (p3)	الإدخال الثاني (p2)	الإدخال الأول (p1)
7.292e+11	2.179	6.293 e-13	Gaussian	Trapezoidal	Triangular
6.565e+11	2.388	6.378 e-13	Trapezoidal	Gaussian	Triangular
8.442e+11	1.905	6.218 e-13	Gaussian	Triangular	Trapezoidal
8.071e+11	1.956	6.334 e-13	Triangular	Gaussian	Trapezoidal
8.666e+11	1.818	6.347 e-13	Trapezoidal	Triangular	Gaussian
9.137e+11	1.751	6.250 e-13	Triangular	Trapezoidal	Gaussian
9.305e+11	1.807	5.947e-13	Triangular	Triangular	Triangular
9.175e+11	1.725	6.318 e-13	Trapezoidal	Trapezoidal	Trapezoidal
6.034e+11	2.091	7.925 e-13	Gaussian	Gaussian	Gaussian

:

-

(FL)

(ANFIS)

(BP)

(GA)

)

:

(^)

.(

.(DNA)

: ()

مقياس كفاءة التقنيات (EFF)	وقت التنفيذ بالثانية (Time)	معدل مربع الخطأ (MSE)	هل تتضمن عملية المحاكاة (Simulation)	نوع التقنية الذكية المستخدمة (Type of Technique)
9.305e+11	1.807	5.947e-013	تتضمن	التقنية المهجنة ANFIS
0.0975	0.982	10.445	تتضمن	الشبكة العصبية BP
0.4525	0.037	59.730	تتضمن	المنطق المضطرب FL
1.539e+3	8.661	7.5011e-5	لا تتضمن	الخوارزمية الجينية GA

(ANFIS)

(١٠)

١٧- الاستنتاجات والتوصيات :

(DNA)

المصادر:

1. Cartwright, H., (2008), “ **Using Artificial Intelligence in Chemistry and Biology: A Practical Guide**”, Taylor & Francis Group, LLC.
2. Jain, A.K., Duin, P.W. and Mao, J.,(2000) , “**Statistical Pattern Recognition: A Review**”, IEEE Transactions on Pattern Analysis and Machine Intelligence, VOL. 22, NO. 1.
3. Keedwell, E. and Narayanan, A., (2005), “**Intelligent Bioinformatics: The application of artificial intelligence techniques to bioinformatics problems**”, John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, England.
4. Steel, R.G.D. and Torrie, J.H., (1980), “**Principles and Procedures of Statistics a Biometrical Approach**”, Mc Graw-Hill, Inc.
5. Richards, J.E. and Hawley, R.S., (2005), “**The Human Genome: A users Guide**” , Second Edition, Elsevier Inc. All rights reserved.
6. Joshi, A. , Ramakrishman, N., Houstis, E.N. and Rice, J.R., (1997), “**On Neurobiological, Neuro-Fuzzy, Machine Learning, and Statistical Pattern Recognition Techniques**”, IEEE Transactions on Neural Networks, VOL. 8, NO.1.