

난수발생기와 일반화된 회귀 신경망을 이용한 DNA 서열 분류

論 文

53D-7-8

DNA Sequence Classification Using a Generalized Regression Neural Network and Random Generator

金 性 模* · 金 根 鎬** · 金 秉 桓***
(Sungmo Kim · Kunho Kim · Byungwhan Kim)

Abstract - A classifier was constructed by using a generalized regression neural network (GRNN) and random generator (RG), which was applied to classify DNA sequences. Three data sets evaluated are eukaryotic and prokaryotic sequences (Data-I), eukaryotic sequences (Data-II), and prokaryotic sequences (Data-III). For each data set, the classifier performance was examined in terms of the total classification sensitivity (TCS), individual classification sensitivity (ICS), total prediction accuracy (TPA), and individual prediction accuracy (IPA). For a given spread, the RG played a role of generating a number of sets of spreads for gaussian functions in the pattern layer. Compared to the GRNN, the RG-GRNN significantly improved the TCS by more than 50%, 60%, and 40% for Data-I, Data-II, and Data-III, respectively. The RG-GRNN also demonstrated improved TPA for all data types. In conclusion, the proposed RG-GRNN can effectively be used to classify a large, multivariable promoter sequences.

Key Words : Promoter, Generalized Regression Neural Network, Random Generator, Classification

1. 서 론

Deoxyribonucleic acid(DNA) 칩 정보로부터 질병진단과 신약개발을 위한 유용한 생물학 정보를 추출하기 위한 연구가 활발히 진행이 되고 있다. 인공신경망은 무정형의 DNA 염기서열상의 전사초기점과 종점등과 같은 프로모터 확인을 위한 중요한 정보를 예측하고 분류하는데 이용되고 있다 [1-3]. 염기서열분석에는 다양한 종류의 신경망이 응용되고 있으며, 그 대표적인 구조가 역전파 신경망 (Backpropagation neural network-BPNN) [4]이다. BPNN의 최적화에는 다수의 학습인자가 관여하고 있으며 [5-6], 초기 웨이트의 불규칙성으로 인해 그 최적화가 어렵다. 특히, 다 변수, 대용량의 바이오 데이터 처리에 소요되는 계산이 상당하여, 그 응용이 제한을 받고 있다. 이에 따라, 구조가 간단하고, 성능최적화도 용이한 신경망이 요구되고 있다. 이러한 요구에 가장 적합한 신경망중의 하나는 일반화된 회귀신경망 (Generalized regression neural network-GRNN) [7]이다. GRNN은 학습이 용이하고, 학습제어인자도 spread 변수 하나밖에 없어, 분류기 내지 예측기의 설계가 용이하다. 바이오데이터 처리와 관련하여, 다른 신경망 구조와 비교사례가 보고 된 사례가 극히 드물어 그 성능 파악이 어려우나, 플라즈마 데이터에의 응용사례 [8]를 비추어 볼 때, GRNN의 예

측성능은 통계적 회귀모델과 비슷하였다. 한편 BPNN은 플라즈마 데이터에 대한 예측모델개발에서, 통계적인 회귀모델에 비해 월등한 예측성능을 보였으며 [6], BPNN에 비해 상대적으로 낮은 GRNN의 성능은 패턴층 뉴런의 가우시안 함수를 구성하는 spread가 동일한 값에서 최적화하는 것에 주로 기인한다. 패턴층 뉴런의 수는 학습데이터의 수와 일치하며, 따라서 학습패턴에 따라 서로 다른 spreads를 사용할 때 예측성능의 증진이 예상된다. 즉 각 spread를 변수화하고, 이러한 다변수 spreads를 최적화하는 기법이 요구된다.

본 연구에서는 난수발생기 (Random generator-RG)를 이용하여 GRNN의 성능이 다중 값을 가지는 spread에서 최적화가 되게 하는 기법을 제안하며, 본 기법을 세 종류의 프로모터 군에 각기 적용한다. 편의상, RG기초의 GRNN은 "RG-GRNN"이라 칭한다. 고정된 난수범위에서 다중 분류기를 발생하며, 이중 결정된 최적의 분류기의 성능을 분류민감도와 예측정확도 측면에서 평가한다. 분류민감도는 임계치 (Threshold)의 함수로 세분화 한다. 제안된 RG-GRNN은 종래의 GRNN와 비교 평가한다.

2. 프로모터 데이터

프로모터 데이터는 Oriza Sativa (OS), Arabidopsis Thaliana (AT), Escherichia Coli (EC), Zymomonas Mobils (ZM)의 서로 다른 4가지 프로모터 서열들로 구성된다. 이중에서 OS와 AT는 eukaryotic 프로모터, 그리고 EC와 ZM은 prokaryotic 프로모터에 속한다. 각각의 프로모터는 NCBI에 등재된 각 개체의 유전체 염기서열 및 유전자 정보를 통해 각 유전자 번역개시인자의 앞쪽 서열을 취하였고 이를 프로

* 學生會員 : 世宗大 電子工學科 碩士課程

** 學生會員 : 世宗大 電子工學科 博士課程

*** 正 會 員 : 世宗大 電子工學科 副教授

接受日字 : 2003年 7月 8日

最終完了 : 2004年 4月 29日

모터 데이터로 활용하였다. 각 유전체의 염기서열은 해당 데이터베이스에서 추출하였다. AT 프로모터 서열은 기존에 알려진 cDNA 염기서열[9]로 Arabidopsis genomic DNA 데이터 베이스[10]를 검색, 비교하여 유전자의 5' upstream 부위로부터 추출하였다. OS 프로모터 서열도 AT의 경우와 마찬가지로 rice 데이터 베이스 [11]를 이용하여 수집하였다. EC 프로모터 서열은 National Center for Biotechnology Information (NCBI)로 부터, 그리고 ZM 프로모터 서열은 두 과정, 즉 Glimmer V2.0 [12] 프로그램을 이용한 유전자 추출과 BlastX [13] 프로그램을 한 non-redundant NCBI 단백질 데이터 베이스와의 유사성 (homology) 비교과정을 통해 수집하였다. 수집된 데이터는 다시 세 종류의 프로모터 데이터, 즉 eukaryotic과 prokaryotic 프로모터 서열, eukaryotic 프로모터 서열, 그리고 prokaryotic 서열로 나누었다. Eukaryotic과 prokaryotic 서열의 경우, 학습에 이용된 서열 패턴의 수는 115개로 OS 20, AT 25, EC 35, ZM 35개로 구성되었으며, 테스트 패턴의 수는 58개로 OS 13, AT 15, EC 15, ZM 15개로 구성되었다. Eukaryotic 프로모터와 prokaryotic 프로모터의 경우, 학습과 테스트 패턴수는 각 45개와 28개 그리고 70개와 30개로 구성이 되었다. 세 종류의 프로모터 데이터의 각 서열 패턴은 146개의 base pair로 이루어졌다.

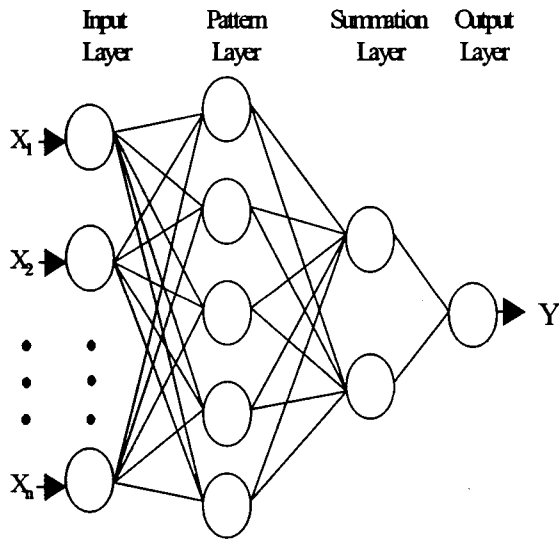


그림 1 일반화된 회귀 신경망 구조
Fig. 1 Architecture of generalized regression neural network

3. 일반화된 회귀 신경망

그림 1에서와 같이, GRNN은 총 4개의 층, 즉 입력층, 패턴층, 합층, 그리고 출력층으로 구성된다. 입력층의 뉴런 수는 독립 변수의 수와 일치하며, 패턴층의 뉴런 수는 학습패턴의 수와 일치한다. 합층은 두 개의 뉴런으로 구성된다. 입력층과 패턴층간의 하중치(W_p)는 입력패턴 (X)에 의해 결정되며, 이를 표현하면,

$$W_p = X^T \quad (1)$$

여기서, "T"는 이항 (Transposition)을 의미한다. 패턴층의 하나의 뉴런은 합층의 두 개의 뉴런에 연결되며, 패턴층의 i 번째의 뉴런과 합층 첫 뉴런간의 연결 하중치는 y_i 가 된다. 이 i번째의 뉴런과 합층의 다른 하나의 뉴런과의 연결 하중치는 1이 된다. 합층과 출력층간의 하중치 (W_o)는 y_i 와 1에 의해 다음과 같이 결정된다.

$$W_o = [Y \ 1] \quad (2)$$

출력층에서는, 단순히 합층의 두 뉴런의 출력을 나누어 예측치를 출력한다. 임의의 입력패턴 x에 대한 예측치는 (3)으로 구해진다.

$$\hat{y}_i(x) = \frac{\sum_{i=1}^n y_i \exp[-D(x, x_i)]}{\sum_{i=1}^n \exp[-D(x, x_i)]} \quad (3)$$

여기서 x_i 는 저장된 i번째의 학습패턴을 지칭하며, n은 전체 학습데이터의 수를 의미한다. (3)에서 함수 D는

$$D(x, x_i) = \sum_{j=1}^p \left(\frac{x_j - x_{ij}}{\zeta} \right)^2 \quad (4)$$

여기서 p는 각 입력패턴을 구성하는 전체 독립변수의 수를 지칭한다. x_j 와 x_{ij} 는 x와 x_i 의 j번째의 요소를 의미한다. 그리고 변수 ζ 는 spread라 불리며, GRNN의 성능을 결정하는 유일한 학습인자이다. 일반적으로 spread는 실험적으로 일정한 범위에서 결정하며, 결정된 spread는 그림 1의 패턴층을 구성하는 모든 뉴런에 대해서 동일하다. spread 값을 다중값으로 결정할 때, GRNN의 성능이 향상될 수 있으며, 이와 관련한 연구는 매우 드물다.

4. 분류기의 성능 분석

4.1 성능지표

테스트 입력 패턴이 특정 부류 (class)로 분류되는지의 여부는 다음과 같이 결정한다. 즉 테스트 입력 패턴에 대한 예측 패턴에서 가장 큰 값이 놓인 위치가, 해당 프로모터에 대한 타겟 패턴의 위치와 동일한 경우 그 입력 패턴은 분류가 된다고 본다. 이때, 분류민감도 (Classification sensitivity)는 정의된 부류 (class)로 정확히 분류되는 테스트 입력패턴의 수로 결정된다. 분류민감도는 전체분류민감도 전체 테스트 패턴에 대해서 계산되는 (Total classification sensitivity-TCS)와 개별적 프로모터에 대해서 계산되는 개별적 분류민감도 (Individual classification sensitivity-ICS)로 나누어 평가하였다. 각 TCS와 ICS는 (5)의 임계치를 기준으로 더욱 세분화하였다.

$$|d_{ij} - out_{ij}| < \text{Threshold} \quad (5)$$

여기서, d_{ij} 와 out_{ij} 는 j 번째의 테스트 입력에 대한 i 번째 출력 뉴런에 주어지는 실제치와 그 뉴런으로부터의 예측치를 의미한다. 임계점은 0.1-0.4로 변화시켰다. 한편, 분류민감도는 다수의 spread에서 동일할 수 있으며, 따라서 최적분류기의 결정을 위해서는 각 분류기의 예측성능을 평가하는 것이 요구되며, 그 성능은 (6)으로 정의되는 Root mean squared error (RMSE)를 이용하여 계산한다.

$$RMSE = \sqrt{\frac{\sum_{i=1}^q \sum_{j=1}^r (d_{ij} - out_{ij})^2}{qr}} \quad (6)$$

여기서 q 와 r 은 출력층 뉴런의 수와 테스트 패턴의 수를 의미한다. 분류민감도와 마찬가지로, 예측정확도는 전체예측정확도 (Total prediction accuracy-TPA)와 개별적 예측정확도 (Individual prediction accuracy-IPA)로 나누어 평가하였다. TPA는 전체 테스트 패턴에 대해서 계산되며, IPA는 각 프로모터에 해당하는 테스트 패턴에 대해서 계산된다. 예를 들어 Eukaryotic과 prokaryotic 프로모터 서열의 경우, 테스트 패턴이 58개이며, 4가지 프로모터 서열로 구성되어 있으므로, TPA를 계산할 때, 식 (6)의 q 와 r 은 4와 58이 된다. OS 프로모터의 경우 테스트 패턴의 수가 13이며, 따라서 IPA의 계산을 위한 q 와 r 은 4와 13이 된다.

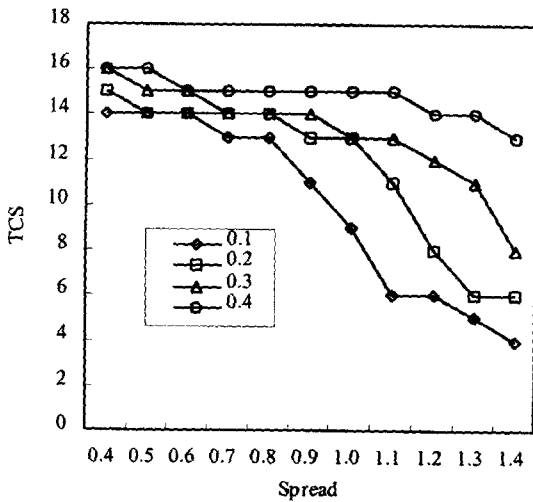


그림 2 TCS의 spread에 따른 변화
Fig. 2 TCS as a function of spread

4.2 Eukaryotic과 prokaryotic 프로모터 서열

비교를 위해, 분류기를 종래의 방식, 즉 spread를 일정한 실험범위에서 변화시키면서 분류기를 설계하였다. 본 연구에서 채택한 실험범위는 [0.4, 1.4]이며, spread 간격을 0.1씩 증가시키면서 분류기를 개발하였다. 각 spread에서 얻어진 분류기로부터 테스트 데이터에 대한 TCS를 먼저 구하고, (5)의 관계를 이용하여 TCS를 다시 임계치의 함수로 계산

하였으며, 그 결과가 그림 2에 도시되어 있다. 그림 2에서와 같이 TCS는 spread의 증가에 따라 일반적으로 감소하고 있다. 모든 임계치에 대해서, 분류기는 0.4에서 가장 우수한 TCS를 보이고 있다.

표 1 임계치에 따른 TCS와 ICS
Table 1 TCS and ICS as a function of threshold

임계치	ICS					Spread
	OS	AT	EC	ZM	TCS	
0.1	6	0	4	4	14	0.4*, 0.5, 0.6
0.2	6	0	4	5	15	0.4
0.3	6	0	5	5	16	0.4
0.4	6	0	5	5	16	0.4*, 0.5

표 2 임계치에 따른 TCS와 ICS
Table 2 TCS and ICS as a function of threshold

임계치	ICS					난수범위
	OS	AT	EC	ZM	TCS	
0.1	1	8	4	8	21	0.4
0.2	1	9	6	8	24	0.4
0.3	1	9	6	8	24	0.4, 0.6*
0.4	1	9	6	8	24	0.4, 0.6*

각 임계치에서 결정된 TCS와 최적화된 spread가 표 1에 나타나 있다. 표 1에서와 같이, 임계치 0.1과 0.4인 경우, TCS가 다중 spread에서 동일하였으며, 최적 분류기를 결정하기 위해 (6)을 이용하여 각 spread에 대한 RMSE를 계산하였다, 그 결과, 가장 작은 RMSE에 대한 spread는 *로 표기하였으며, 이 분류기에 대한 ICS가 표 1에 정리되어 있다. 다음은, RG를 이용하여 패턴층 뉴런의 가우시안 함수가 다중 spread 값을 갖게 한 다음, 그 성능을 평가하였다. 이를 위해 주어진 spread 값에서 RG를 이용하여 200개의 난수 sets을 발생하고, 각 set에 대해서 분류기의 성능을 평가하였다. 예컨대, 주어진 spread가 0.4일 경우, 가우시안 함수의 spreads는 0에서 0.4사이에서 발생한 임의의 값을 가지게 된다. 발생한 200개의 분류기 중, 최고의 TCS를 보이는 분류기를 최적 분류기로 결정하였다. 최적 분류기의 TCS와 서로 다른 프로모터에 대한 ICS를 임계치의 함수로 계산하였으며, 그 결과가 표 2에 나타나 있다. 표 2에서와 같이, TCS는 임계치 0.1에서 21개 (36.2%)이며, 이는 표 1의 동일 임계치에서의 TCS와 비교해 7개가 증진이 되어 50%의 향상된 분류정확도를 보이고 있다. 나머지 0.3과 0.4에 대해서는 8개가 향상이 되었다. 이는 RG-GRNN이 TCS를 모든 임계치에 대해서 50%이상 향상시켰음을 의미한다. 한편, ICS 측면에서, OS의 경우 종래의 GRNN에 비해, 그 성능이 저하되었다. 그러나, AT의 경우, 분류성능을 상당히 증진되었다. EC와 ZM에 대해서도 분류성능이 향상되었다. 다음에는 주어진 임계치에 대해서, 결정된 최적분류기의 예측성능을 비교 평가하였으며, 그 결과는 표 3에 정리되어 있다. 표 3의 상대적인 개선도는 GRNN에 대해서 계산된 양이다. 개

선도에서 알 수 있듯이, RG-GRNN은 TPA측면에서 향상되었다. IPA측면에서도, 임계치 0.1과 0.2에서 OS를 제외하고 나머지 세 프로모터에 대해서 향상되었으며, 임계점 0.3과 0.4에 대해서는 모든 종류의 프로모터에 대해서 향상되었다. 결국, GRNN에 비해 RG-GRNN은 모든 성능 지표에서 향상된 성능을 보이고 있으며, 특히 TCS에서의 개선도는 주목할 정도로 상당하였다.

표 3 임계치에 따른 TPA와 IPA
Table 3 TPA and IPA as a function of threshold

임계치	프로모터	RG-GRNN		GRNN		Improvement (%)	
		TPA	IPA	TPA	IPA	TPA	IPA
		OS		0.675		0.518	
0.1	AT	0.528	0.419	0.573	0.701	7.8	40.2
	EC		0.538		0.577		6.7
	ZM		0.48		0.574		16.3
	OS		0.675		0.518		-30.3
0.2	AT	0.528	0.419	0.573	0.701	7.8	40.2
	EC		0.538		0.577		6.7
	ZM		0.48		0.574		16.3
	OS		0.214		0.518		58.6
0.3	AT	0.470	0.673	0.573	0.701	17.9	3.9
	EC		0.442		0.577		23.3
	ZM		0.418		0.574		27.1
	OS		0.214		0.518		58.6
0.4	AT	0.470	0.673	0.573	0.701	17.9	3.9
	EC		0.442		0.577		23.3
	ZM		0.418		0.574		27.1

표 4 임계치에 따른 TCS와 ICS
Table 4 TCS and ICS as a function of threshold

임계치	ICS			Spread
	OS	AT	TCS	
0.1	9	2	11	0.4, 0.5*
0.2	9	2	11	0.4, 0.5*, 0.6
0.3	9	2	11	0.4, 0.5*, 0.6, 0.7
0.4	9	2	11	0.4, 0.5*, 0.6, 0.7, 0.8

4.3 Eukaryotic 프로모터 서열

종래의 GRNN을 주어진 서열에 적용하여, 분류민감도를 분석하였으며, 그 결과가 표 4에 정리되어 있다. 표 4에서와 같이, 모든 임계치에 대해서 TCS는 11 (39.3%)로 동일하며, ICS도 OS와 AT에 대해서 각 9개와 2개로 동일하다. RG-GRNN에 의해 분류하였을 때, 그 결과가 표 5에 정리되어 있다. TCS측면에서 GRNN에 비해 모든 임계치에 대해서 7개가 증진이 되었으며, 이는 63.6%정도 향상된 분류정확도를 의미한다. AT의 경우, ICS가 상당히 증진되었으나, OS의 경우, 저하되었다. 각 임계치에서 최적화된 분류기의 RMSE는 표 6에 나타나 있다. 표 6에서와 같이, TPA측

면에서 RG-GRNN은 GRNN에 비해 20%이상 향상된 성능을 보인다. IPA측면에서 AT의 경우 향상되었으나, OS의 경우에는 그 반대이다. 이는 앞서의 OS에 대한 저하된 ICS의 주 원인이 된다.

표 5 임계치에 따른 TCS와 ICS
Table 5 TCS and ICS as a function of threshold

임계치	ICS			난수범위
	OS	AT	TCS	
0.1	6	12	18	0.5
0.2	6	12	18	0.5
0.3	3	15	18	0.4, 0.5, 0.6*
0.4	7	11	18	0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0

표 6 임계치에 따른 TPA와 IPA
Table 6 TPA and IPA as a function of threshold

임계치	프로모터	RG-GRNN		GRNN		Improvement (%)	
		TPA	IPA	TPA	IPA	TPA	IPA
0.1	OS	0.582	0.7264	0.779	0.5536	25.2	-31.20
	AT		0.419		0.9312		54.90
0.2	OS	0.582	0.7264	0.779	0.5536	25.2	-31.20
	AT		0.419		0.9312		54.90
0.3	OS	0.548	0.8049	0.779	0.5536	23.1	-45.30
	AT		0.0174		0.9312		98.10
0.4	OS	0.520	0.5448	0.779	0.5536	25.9	-0.10
	AT		0.4983		0.9312		46.50

4.4 Prokaryotic 프로모터 서열

종래의 GRNN을 이용하여 주어진 서열을 분류하였으며, 그 분류성능이 표 7에 정리되어 있다. 표 7에서와 같이, 임계치 0.1에 대해서 TCS는 13 (43.3%)이며, ZM에 비해 EC에 대한 ICS가 더 크다. 이는 다른 임계치 0.2-0.3에 대해서도 동일하다. 한편, RG-GRNN을 이용하여 분류하였을 때의 결과가 표 8에 나타나 있다. 표 8에서와 같이 RG-GRNN은 임계치 0.1에서 TCS가 19 (63.6%)이며, 표 7의 동일 임계치의 TCS와 비교하여 46.1%의 향상된 분류정확도를 보이고 있다. Eukaryotic 프로모터와는 달리, ICS 측면에서도 각 임계치에 대해서 향상되었거나 동일하였다. 결국, TCS와 ICS측면에서 RG-GRNN이 더 우수한 성능을 보였다. RMSE 비교결과는 표 9에 정리되어 있다. 표 9에서와 같이 RG-GRNN은 TPA측면에서 20%이상 향상된 성능을 보이고 있

으며, IPA측면에서도 향상되었다. 결국, RG-GRNN은 모든 성능지표에서 GRNN보다 향상되었다.

표 7 임계치에 따른 TCS와 ICS
Table 7 TCS and ICS as a function of threshold

임계치	ICS			Spread
	EC	ZM	TCS	
0.1	8	5	13	0.4*, 0.5, 0.6
0.2	9	6	15	0.4*
0.3	9	6	15	0.4*, 0.5, 0.6
0.4	9	6	15	0.4*, 0.5, 0.6, 0.7, 0.8, 0.9

표 8 임계치에 따른 TCS와 ICS
Table 8 TCS and ICS as a function of threshold

임계치	ICS			난수범위
	EC	ZM	TCS	
0.1	8	11	19	0.4*
0.2	13	6	19	0.4, 0.5, 0.6*
0.3	14	6	20	0.4, 0.6*
0.4	14	8	22	0.6*

표 9 임계치에 따른 TPA와 IPA
Table 9 TPA and IPA as a function of threshold

임계치	프로모터	RG-GRNN		GRNN		Improvement (%)	
		TPA	IPA	TPA	IPA	TPA	IPA
		0.1	EC	0.509	0.6111	0.707	0.6332
	ZM		0.3806		0.7752		50.9
0.2	EC	0.498	0.2668	0.707	0.6332	20.9	57.9
	ZM		0.653		0.7752		15.7
0.3	EC	0.498	0.2668	0.707	0.6332	20.9	57.9
	ZM		0.653		0.7752		15.7
0.4	EC	0.498	0.2668	0.707	0.6332	20.9	57.9
	ZM		0.653		0.7752		15.7

5. 결 론

RG 기초의 GRNN 분류기 설계방식을 제안하였으며, 이를 세 종류의 프로모터 데이터에 적용하였다. 분류기의 성능 평가를 위해 4종류의 평가지표를 설정하였다. 비교를 위해 종래의 방식으로 GRNN분류기를 설계하였으며 이를 RG-GRNN과 비교하였다. 비교결과, RG-GRNN은 TCS와 TPA측면에서 우수한 성능을 보였으며, 특히 TCS측면에서 월등한 향상을 보였다. ICS와 IPA의 경우에도 eukaryotic promoter를 제외하고는 향상되었다. 결론적으로, 제안된 RG-GRNN은 패턴층 뉴런의 가우시안 함수를 다중 spread

에서 최적화가 되게 함으로써 DNA서열에 대한 분류성능을 획기적으로 증진하였다. 이 방식은 입력변수가 많은 다른 바이오 데이터의 예측과 분류에 향후 효과적으로 이용될 수 있다.

참 고 문 헌

- [1] M. V. Gils, H. Jansen, K. Nieminen, R. Summers, P. R. Weller, "Using artificial neural networks for classifying ICU patient states," *IEEE EMB Mag.*, pp. 41-47, 1997.
- [2] S. Knudsen, "Promoter 2.0: for the recognition of Pol II promoter sequences," *Bioinformatics*, vol. 15, pp. 356-361, 1999.
- [3] S. Matis, Y. Xu, M. Shah, X. Guan, J. R. Einstein, R. Mural, E. Uberhacher, "Detection of RNA polymerase II promoters and polyadenylation sites in human DNA sequence." *Comp. Chem.* pp. 135-140, 1996.
- [4] D. E. Rummelhart, J. L. McClelland, *Parallel Distributed Processing*, MIT Press, Cambridge, 1986.
- [5] B. Kim and Gary S. May, "An optimal neural network process model for plasma etching," *IEEE Trans. Semicond. Manufact.*, vol. 7, no. 1, pp. 12-21, 1994.
- [6] B. Kim and S. Park, "An optimal neural network plasma model: a case study," *Chemom. Intell. Lab. Syst.*, vol. 56, pp. 39-50, 2001.
- [7] Specht D F, "A generalized regression neural networks." *IEEE Trans. Neural Networks* vol. 2, pp. 568-576, 1991.
- [8] B. Kim and S. Park, "Modeling of process plasma using a radial basis function network: a case study," *Trans. Contr. Autom. Syst. Eng.*, vol.2, no. 4, pp. 268-273, 2000.
- [9] <http://signal.salk.edu/cgi-bin/tdnaexpress>.
- [10] <http://arabidopsis.org>.
- [11] <http://www.ncbi.nlm.nih.gov>.
- [12] F. R. Blattner, G. III Plunket, C. A. Bloc, N. T. Perna, V. Burland, M. Riley, J. Colladovides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao, "The complete genome sequence of Escherichia coli K-12," *Science* 277, vol. 5331, pp. 1453-1474, 1997.
- [13] A. L. Delcher, D. Harmon, S. Kasif, O. White, S. L. Salzberg, "Improved microbial gene identification with GLIMMER," *Nucleic Acids Res.*, vol. 27, pp. 4636-4641, 1999.

저 자 소 개



김성모 (金性模)

1976년 12월 14일생. 2002년 남서울대학교 전자공학과 졸업. 2002년~현재 세종대학교 일반대학원 전자공학과 석사과정

Tel : 02-3408-3729

Fax : 02-3408-3329

E-mail : ksm1214@hanmir.com



김근호 (金根鎬)

1975년 6월 16일생. 2002년 세종대학교 전자공학과 졸업. 2002년~현재 동 대학원 전자공학과 박사과정

Tel : 02-3408-3729

Fax : 02-3408-3329

E-mail : gno0616@hanmail.net



김병환 (金秉桓)

1962년 10월 30일생. 1985년 고려대 전기공학과 졸업. 1987년 고려대 전기공학과 졸업(석사). 1995년 Georgia Institute of Technology 졸업(공학). 1996-1998년 현대전자반도체 연구소 책임연구원. 1999-2001년 전남대 전기공학과 전임강사, 2001- 현재 세종대 전자공학과 부교수.

Tel : 02-3408-3729

Fax : 02-3408-3329

E-mail : kbwhan@sejong.ac.kr