변종 몬테 칼로 신경망을 이용한 패턴 분류

전 성 해[†]·최 성 용[††]·오 임 걸[†††]·이 상 호[††††]·전 홍 석[†††††]

요 약

일반적인 다층 신경망에서 가중치의 갱신 알고리즘으로 사용하는 오류 역전파 방식은 가중치 갱신 결과를 고정된(fixed) 한 개의 값으로 결정한다. 이는 여러 갱신의 가능성을 오직 한 개의 값으로 고정하기 때문에 다양한 가능성들을 모두 수용하지 못하는 면이 있다. 하지만 모든 가능성을 확률적 분포로 표현하는 갱신 알고리즘을 도입하면 이런 문제는 해결된다. 이러한 알고리즘을 사용한 베이지안 신경망 모형(Bayesian Neural Networks Models)은 주어진 입력값(Input)에 대해 블랙 박스(Black-Box)와 같은 신경망 구조의 각 층(Layer)을 거친 출력값(Output)을 계산한다. 이 때 주어진 입력 데이터에 대한 결과의 예측값은 사후분포(posterior distribution)의 기대값(mean)에 의해 계산할 수 있다. 주어진 사전분포(prior distribution)와 학습데이터에 의한 우도함수(likelihood functions)에 의해 계산한 사후확률의 함수는 매우 복잡한 구조를 가짐으로 기대값의 적분계산에 대한 어려움이 발생한다. 따라서 수치해석적인 방법보다는 확률적 추정에 의한 근사 방법인 몬테 칼로 시뮬레이션을 이용할 수 있다. 이러한 방법으로서 Hybrid Monte Carlo 알고리즘은 좋은 결과를 제공하여준다(Neal 1996). 본 논문에서는 Hybrid Monte Carlo 알고리즘을 적용한 신경망이 기존의 CHAID, CART 그리고 QUEST와 같은 여러 가지 분류 알고리즘에 비해서 우수한 결과를 제공하는 것을 나타내고 있다.

Pattern Classification Using Hybrid Monte Carlo Neural Networks

Sung Hae Jun[†]·Seong Yong Choi[††]·Im Geol Oh[†††]·Sangho Lee[††††]·Hong Suk Jorn[†††††]

ABSTRACT

There are several algorithms for classification in modeling relations, patterns, and rules which exist in data. We learn to classify objects on the basis of instances presented to us, not by being given a set of classification rules. The hybrid monte carlo neural networks uses the probability distribution to express our knowledge about unknown parameters and update our knowledge by the law of probability as the evidence gathered from data. Also, the neural network models are designed for predicting an unknown category or quantity on the basis of known attributes by training. In this paper, we compare the misclassification error rates of hybrid monte carlo neural networks with those of other classification algorithms, CHAID, CART, and QUEST using several data sets. In terms of error rate, the Bayesian method works better than others in all data sets. The only trouble at this point is computing time.

키워드 : Hybrid Monte Carlo Algorithms, Pattern Classification, Bayesian Neural Network, Bayesian Learning

1. 서 론

지문 인식(pattern recognition)등의 수많은 생체 인식의 시스템, 문서 분류(text classification), 웹 문서분류 등 정보 검색(information retrieval) 엔진, 데이터 마이닝(data mining)등 매우 다양한 분야에서 분류 알고리즘을 사용한다. 이러한 분류 알고리즘은 오래 전부터 매우 다양한 방법으로 연구, 적용되어 왔다. 본 논문에서는 이러한 기존의 여러 알고리즘 중에서 특히 인간의 뇌 구조를 모형화 한 신경망(neural network)에 베이지안 추론(bayesian inference)을 적용한 변종 몬테 칼로(hybrid monte carlo) 신경망에 대한 연구를 수행하였다. 오류 역전파 방식(back-propagation)의 기존의 신경망의 가중치(weights) 갱신 알고리즘에 사전(prior), 사후(posterior) 분포를 적용하였다. 이러한 사후 분포의 기대값(expectation)으로서 새로운 개체에 대한 분류를 수행하게 된다.

베이지안 신경망 모형은 네트워크 모수(network parameters)인 가중치(weight)와 바이어스(bias)를 결정하는 초기의 사전확률분포(prior distribution)와 데이터에 의한 우도함수(likelihood function)의 곱에 의해 네트워크 모수의 사

† 정 회 원 : NCR Korea 데이터 마이닝팀 컨설턴트
†† 정 회 원 : 인하대학교 대학원 전자계산공학과
††† 정 회 원 : 한서대학교 컴퓨터통신공학과 교수
†††† 정 회 원 : 강릉대학교 통계학과 교수
††††† 정 회 원 : 인하대학교 통계학과 교수
논문접수 : 2001년 1월 19일, 심사완료 : 2001년 6월 25일

후분포(posterior distribution)을 계산하여 새로운 입력값에 대한 출력값은 입력값에 대한 사후확률의 기대값으로 예측된다. 하지만 사후분포의 기대값은 매우 복잡하여 원시함수를 구하여 정적분을 하기가 어렵다. 또한 모수의 차원이 다차원(multi-dimension)일 경우는 수치해석적인 방법에 의해서도 구하기가 어렵다. 따라서 확률적 추정에 의한 적분값의 계산인 몬테 칼로 적분이 필요하게 된다. 이러한 몬테 칼로의 기법들 중에서 본 논문에서는 Hybrid Monte Carlo (HMC) 알고리즘을 적용하여 오분류율이 낮아지는 우수한 결과를 얻었다.

2. 베이지안 예측 모델

베이지안 추론은 신경망의 학습에 있어서, 특히 무한 신경망(infinite neural networks)의 학습에 있어서 효율적인 대안이 될 수 있다. 베이지안 추론이 적용되지 않는 일반 오류 역전파 방식(Back-Propagation)의 신경망과는 달리 데이터들이 갖고 있는 규칙들을 자동으로 결정해 주고 또한 예측에 대한 불확실성을 양적으로 나타내어 준다. 신경망 모형에 대한 베이지안 추론의 목적은 n개의 학습 데이터로부터 입력값(input value ; $x^{(i)}$)과 목적값(target value ; $y^{(i)}$)으로부터 다음 식과 같은 새로운 입력값에 대한 출력값의 예측분포(Predictive Distribution)를 찾아내는 것이다.

$$P(y^{(n+1)}|x^{(n+1)},(x^{(1)},y^{(1)}),\ldots,(x^{(n)},y^{(n)}))=\int P(y^{(n+1)}|x^{(n+1)},\theta)\,P(\theta|(x^{(1)},y^{(1)}),\ldots,(x^{(n)},y^{(n)}))\,d\theta$$

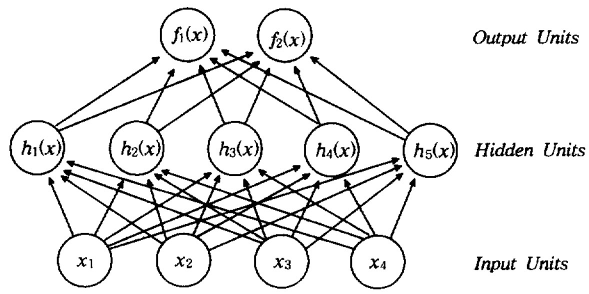
여기서 θ는 가중치와 편이로 구성된 네트워크의 모수(network parameters)이다. 위의 예측분포를 이용하여 새로운 입력값 $x^{(n+1)}$에 대한 목적값의 예측은 다음의 식과 같이 예측분포의 사후 평균(mean)으로 구할 수 있다.

$$\hat{y}_k^{(n+1)}=\int f_k(x^{(n+1)},\theta)P(\theta|(x^{(1)},y^{(1)}),\ldots,(x^{(n)},y^{(n)}))\,d\theta$$

위의 평균값을 구하는 것은 수치적으로 매우 어렵거나 불가능한 경우가 대부분이다. 따라서 몬테칼로 방법(Monte Carlo Methods)을 이용하여 근사값을 구한다. 특히 본 논문에서는 Hybrid Monte Carlo 방법이 기존의 다른 방법에 비해 우수함을 제시하고 최근의 방법들과도 비교한다.

3. 베이지안 신경망

(그림 1)과 같은 일반적인 신경망 구조에서 네트워크의 모수의 갱신과 새로운 입력값에 대한 예측값의 추정에 확률분포를 적용하게 된다.



(그림 1) 베이지안 신경망 구조

$$h_j(x)=A_1(a_j+\sum_{i=1}^{I}u_{ij}x_i)$$

$$f_k(x)=A_2(b_k+\sum_{j=1}^{H}v_{jk}h_j(x))$$

위의 모형에서 $h(x)$는 은닉층을 통해 계산된 결과값이고, $f(x)$는 최종적인 출력값이다. u_{ij}는 입력층에서 은닉층으로의 가중치이고 a_j는 은닉층의 편이(Bias)이다. 또한 v_{jk}는 은닉층에서 출력층으로의 가중치이고 b_k는 출력층의 편이이다. 각각의 네트워크 모수는 $u_{ij}\sim N(0,\sigma_u^2)$, $a_j\sim N(0,\sigma_a^2)$, $v_{jk}\sim N(0,\sigma_v^2)$, $b_k\sim N(0,\sigma_b^2)$의 사전 분포(Prior Distribution)를 갖는다.

4. HMC(Hybrid Monte Carlo)

Hybrid Monte Carlo 알고리즘은 Gibbs Sampling기법과 Metropolis 알고리즘을 결합한 구조를 갖는다.(Duane, Kennedy, Pendleton, Roweth, 1987). HMC 알고리즘의 결과는 Monte Carlo 방법에 의한 기대치 추정에 사용될 표본이 된다. 이러한 HMC 알고리즘은 Hamitonian dynamic에 의한 Metropolis 갱신알고리즘에 의해 수행된다. 즉 HMC 알고리즘은 leapfrog step과 dynamic transition을 번갈아 가면서 반복 수행한다.

4.1 Hamiltonian Dynamics

각각의 q와 p는 다음의 식에 의해 시간 (t)을 통하여 변화한다.

$$\frac{dq_i}{dt}=+\frac{\partial H}{\partial p_i}\qquad \frac{dp_i}{dt}=-\frac{\partial H}{\partial q_i}$$

위 식에서 H는 Hamiltonian 함수이다.

4.2 Leapfrog step

step 1) $\hat{p}_i(t+\frac{\epsilon}{2})=\hat{p}_i(t)-\frac{\epsilon}{2}\frac{\partial E}{\partial q_i}(\hat{q}(t))$

여기서 ε은 step size이다.

step 2) $\hat{q}_i(t+\varepsilon) = \hat{q}_i(t) + \varepsilon \dfrac{\hat{p}_i(t+\frac{\varepsilon}{2})}{m_i}$

step 3) $\hat{p}_i(t+\varepsilon) = \hat{p}_i(t+\frac{\varepsilon}{2}) - \frac{\varepsilon}{2}\frac{\partial E}{\partial q_i}(\hat{q}(t+\varepsilon))$

step 4) step1)부터 step3)을 L(leapfrog step의 수)번 반복한다.

4.3 Dynamical transition

step 1) Initial state $(q, p) = (\hat{q}(0), \hat{p}(0))$로 부터 ε의 step size로 leapfrog step을 L번 반복해서 $(\hat{q}(\varepsilon L), \hat{p}(\varepsilon L))$을 얻는다.

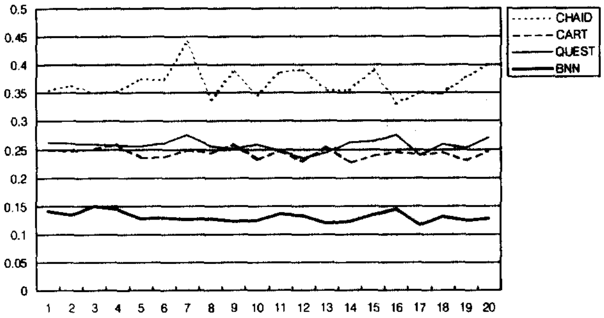
step 2) 운동량변수들(momentum variables)의 부호를 반대로 하여서 $(q^*, p^*) = (\hat{q}(\varepsilon L), -\hat{p}(\varepsilon L))$를 구한다. 이러한 (q^*, p^*)를 새로운 state에 대한 candidate state로 한다.

step 3) 이렇게 구해진 candidate state에 대해 Metropolis 알고리즘에서처럼 $\min(1, \exp(-(H(q^*, p^*) - H(q, p))))$의 확률로 받아들이고 그렇지 않으면 새로운 state는 이전의 state와 같게 한다.

5. 실험 및 결과

5.1 Wave Data 실험

브라이만 등의 시뮬레이션 자료로 제안한 파동형태의 자료는 Wave 형태에 관한 자료(Breiman et. Al.,1984)로서 21개의 연속적인 입력변수(x1 - x21)를 갖는다. 입력변수에 따라 0, 1, 2의 3개의 그룹으로 분류했다.



(그림 2) Wave Data에 대한 오분류율 비교

<표 1>과 (그림 2)의 결과를 보면 총 20회 실험의 모든 경우에서 변종 몬테 칼로 신경망(BNN)의 방법이 오분류율이 가장 낮은 것을 알 수 있다.

5.2 Digit Data 실험

고장난 전자계산기로부터 얻어진 자료로서 7개의 입력변수(x1 - x7)를 가지며 숫자 0 ~ 9의 10개의 집단으로 분류했다.

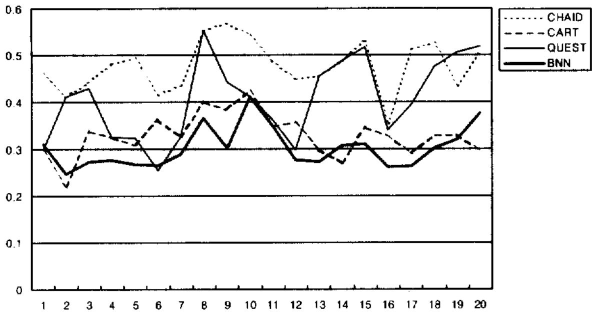
<표 2>와 (그림 3)의 결과를 보면 총 20회의 실험의 대부분의 경우에서 변종 몬테 칼로 신경망의 방법의 오분류율이 작게 나타나는 것을 알 수 있다.

〈표 1〉 20회의 실험결과에 대한 오분류율

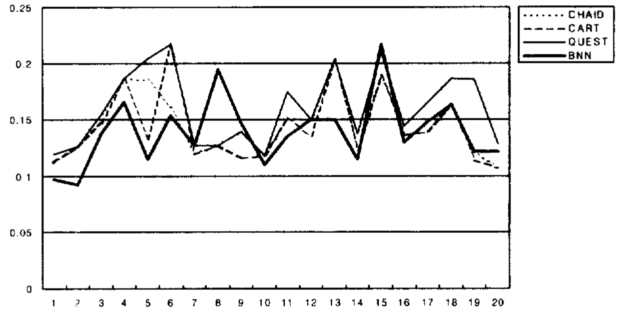
	CHAID	CART	QUEST	BNN
1	0.3561	0.25	0.2629	0.1414
2	0.3633	0.248	0.26	0.1342
3	0.3511	0.2517	0.2596	0.1489
4	0.3528	0.2594	0.2569	0.1439
5	0.3752	0.2359	0.2557	0.1286
6	0.3739	0.2382	0.2608	0.1291
7	0.4397	0.2493	0.2765	0.1266
8	0.3371	0.2438	0.2554	0.1282
9	0.3877	0.2593	0.2511	0.1228
10	0.3449	0.2318	0.2586	0.1247
11	0.3874	0.2472	0.2472	0.1359
12	0.3924	0.23	0.235	0.1326
13	0.3554	0.2544	0.2444	0.12
14	0.3563	0.2271	0.2627	0.1227
15	0.3908	0.2384	0.2649	0.1349
16	0.3307	0.2453	0.2757	0.1439
17	0.3499	0.2409	0.2388	0.1168
18	0.3483	0.2461	0.2587	0.1309
19	0.3777	0.2308	0.252	0.1248
20	0.4019	0.2469	0.2715	0.1284

〈표 2〉 20회의 실험결과에 대한 오분류율

	CHAID	CART	QUEST	BNN
1	0.4598	0.2989	0.2989	0.3103
2	0.411	0.2192	0.411	0.2466
3	0.4416	0.3377	0.4286	0.2727
4	0.4819	0.3253	0.3253	0.2771
5	0.493	0.3099	0.3239	0.2676
6	0.4149	0.3617	0.2553	0.266
7	0.4337	0.3253	0.3253	0.2892
8	0.5529	0.4	0.5529	0.3647
9	0.5698	0.3837	0.4419	0.3023
10	0.5412	0.4235	0.4118	0.4118
11	0.4875	0.35	0.3625	0.35
12	0.4483	0.3563	0.2989	0.2759
13	0.4545	0.2987	0.4545	0.2727
14	0.4872	0.2692	0.4872	0.3077
15	0.5287	0.3448	0.5172	0.3103
16	0.3523	0.3295	0.3409	0.2614
17	0.5132	0.2895	0.3947	0.2632
18	0.5263	0.3289	0.4737	0.3026
19	0.433	0.3299	0.5052	0.3196
20	0.5059	0.2941	0.5176	0.3765



(그림 3) Wave Data에 대한 오분류율 비교



(그림 4) Credit Data에 대한 오분류율 비교

5.3 Credit Data 실험

고객의 신용정도에 관한 자료로서 4개의 입력변수(x1 - x4)를 갖는다. 입력 변수가 되는 직위(x1 ; 1(경영자), 2(전문가), 3(일반사무직), 4(기술자), 5(무기술자)), 급여형태(x2 ; 1(주급), 2(월급)), 나이(x3 ; 1(젊음 : < 25), 2(보통 : 25 - 35), 3(많음 : 35 <)), 아멕스 카드 소지여부(x4 ; 1(비소지), 2(소지))에 따라 개인의 신용여부를 불량등급(분류집단 0), 우량등급(분류집단 1)의 두 집단으로 분류했다.

<표 3>과 (그림 4)의 결과를 보면 총 20회의 실험의 모든 경우에서 변종 몬테 칼로 신경망의 방법의 오분류율이 다른 것에 비해서 눈에 띄게 작게 나타나지는 않았지만, 전체 오분류율의 평균을 비교하면 0.1395의 오분류율로 근소한 차이지만 비교적 분류를 잘 했음을 알 수 있다.

5.4 Pen-based Recognition of Handwritten Digits 실험

손으로 직접 쓴 문자를 판독하는 자료로서 각 문자의 좌표를 나타내는 16개의 입력 변수를 갖는다. 총 학습 데이터의 개수는 7494개이다. 이 실험은 총 50회의 실험을 하였다. 실험결과에서 변종 몬테 칼로 신경망에 의한 오분류율이 가장 낮음을 확인할 수 있다. 또한 이번 실험에서는 학습 시간에 대한 비교를 하였다. <표 5>에 나타난 학습 시간에 대한 결과는 다른 분류 알고리즘들보다 변종 몬테 칼로 신경망이 오래 걸리는 것으로 나타났다. 이는 Hybrid Monte Carlo 알고리즘에서의 Leapfrog step에서 step size를 작게 하였기 때문이다. 따라서 변종 몬테 칼로 신경망의 학습 시간을 단축시키기 위해서는 적절한 step size에 대한 고려가 필요하다.

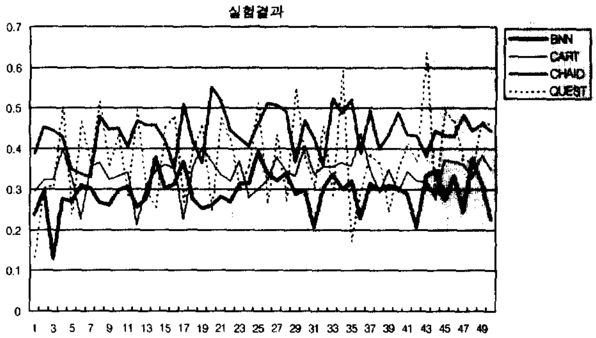
또한 <표 4>와 (그림 5)는 50회의 Pen-based Recog-

〈표 3〉 20회의 실험결과에 대한 오분류율

	CHAID	CART	QUEST	BNN
1	0.1119	0.1119	0.1194	0.097
2	0.1261	0.1261	0.1261	0.0924
3	0.1552	0.1466	0.1552	0.1379
4	0.1862	0.1862	0.1862	0.1655
5	0.1858	0.1327	0.2035	0.115
6	0.1608	0.2168	0.2168	0.1538
7	0.1194	0.1194	0.1269	0.1269
8	0.1269	0.1269	0.1269	0.194
9	0.1395	0.1163	0.1395	0.1473
10	0.1181	0.1181	0.1181	0.1102
11	0.1508	0.1508	0.1746	0.1349
12	0.1504	0.1353	0.1504	0.1504
13	0.203	0.203	0.203	0.1504
14	0.1221	0.1221	0.1374	0.1145
15	0.1888	0.1888	0.2098	0.2168
16	0.1367	0.1367	0.1439	0.1295
17	0.1391	0.1391	0.1652	0.1478
18	0.1628	0.1628	0.186	0.1628
19	0.1214	0.1143	0.1857	0.1214
20	0.1071	0.1071	0.1286	0.1214

〈표 4〉 50회의 실험결과에 대한 오분류율

	CHAID	CART	QUEST	BNN
1	0.2365	0.296	0.3889	0.1327
2	0.2967	0.3238	0.4538	0.3038
3	0.1286	0.3228	0.4438	0.3102
4	0.2756	0.3999	0.4302	0.5
5	0.2686	0.3271	0.3492	0.2381
6	0.3089	0.2238	0.3377	0.466
7	0.3028	0.3559	0.3302	0.3658
8	0.2673	0.3642	0.4801	0.5139
9	0.2587	0.3227	0.4464	0.3587
10	…	…	…	…
41	0.289	0.3453	0.4349	0.4125
42	0.2051	0.3212	0.4332	0.3695
43	0.3315	0.3195	0.3826	0.6353
44	0.3474	0.2816	0.4453	0.2754
45	0.2737	0.3716	0.4326	0.4905
46	0.3319	0.3666	0.4319	0.466
47	0.2432	0.3622	0.4856	0.3775
48	0.3767	0.3297	0.4451	0.3622
49	0.3105	0.3839	0.4619	0.4674
50	0.2244	0.3482	0.4458	0.4622



(그림 5) Pen Recognition Data에 대한 오분류율 비교

nition of Handwritten Digits 실험을 통한 각 분류 알고리즘의 학습 시간의 평균값이다.

〈표 5〉 각 분류 알고리즘의 평균 학습시간

분류 방법	평균학습시간
BNN	42분 11초
CART	37분 24초
CHAID	39분 16초
QUEST	31분 09초

6. 결 론

다른 분류 알고리즘들에 비해 HMC 알고리즘을 적용한 변종 몬테 칼로 신경망은 좋은 결과를 제공하여 준다. 적은 수의 학습 데이터를 이용하여 실험을 하였지만 많은 데이터를 이용하는 실제 적용에서도 같은 결과를 이루게 될 것이다. 다만 학습 시간이 기존의 분류 알고리즘보다 다소 오래 걸리는 면이 있다. 앞으로의 연구는 변종 몬테 칼로 알고리즘의 학습 시간을 단축하는 방향으로 연구되리라 기대한다. 이러한 연구를 이루기 위해서는 HMC 알고리즘 속의 Dynamical transition 시간에 대한 단축 연구가 선행되어야 할 것이다. 또한 본 논문의 신경망은 자율 학습(unsupervised learning) 모형으로서 코호넨(kohonen)이 제안한 신경망인 자기 조직화 형상 지도(SOM ; Self - Organizing Map)에도 수정 없이 적용되어 질 수 있다.

참 고 문 헌

[1] Bishop, C. M. , Neural Networks for Pattern Recognition, Clarendon Press, 1995.

[2] Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C, J, Classification and Regression Trees, Chapman & Hall, 1984.

[3] Duane, S., Kennedy, A. D., Pendleton, B. J., Roweth, D., Hybrid Monte Carlo, Phisics Letters B, 1987.

[4] Neal, R. M., Bayesian Learning for Neural Networks, Springer, 1996.

[5] Ripley, B. D., Pattern Recognition and Neural Networks, Cambridge University Press, 1996.

[6] Johnson, Richard A. and Wichern Dean W, Applied Multivariate Statistical Analysis, Prentice Hall, 1992.

[7] T. Kohonen, Self - Organizing Maps, 2nd Edition, Springer, 1997

전 성 해

e-mail : shjun@anova.inha.ac.kr
1993년 인하대학교 통계학과 졸업(이학사)
1996년 인하대학교 대학원 통계학과 (이학석사)
2001년 인하대학교 대학원 통계학과 (이학박사)
2001년 서강대학교 대학원 컴퓨터학과(박사과정)
1996년~1997년 효성그룹 전자통신연구소 연구원
2000년~2001년 한서대학교 컴퓨터통신과 겸임교수
2000년~현재 NCR Korea 데이터 마이닝팀 컨설턴트
관심분야 : 베이지안 학습, 신경망, 데이터마이닝

최 성 용

e-mail : sychoi@anova.inha.ac.kr
1993년 인하대학교 통계학과 졸업 (이학사)
2001년 인하대학교 대학원 통계학과 (이학석사)
2001년~현재 인하대학교 대학원 전자계산공학과(박사과정)
관심분야 : 베이지안 학습, 신경망, 데이터마이닝

오 임 걸

e-mail : oig@hanseo.ac.kr
1983년 인하대학교 수학과 졸업(이학사)
1986년 인하대학교 대학원 수학과 (이학석사) / 응용수학
1993년 인하대학교 대학원 수학과 (이학박사) / 수리통계학전공
2000년~현재 인하대학교 대학원 전자계산공학과(박사과정)
1995년~1997년 한서대학교 전산통계학과 전임강사
1997년~1999년 한서대학교 컴퓨터과학과 조교수
1999년~현재 한서대학교 컴퓨터통신공학과 조교수
관심분야 : 패턴인식, 신경망, 보안, 인공지능

이 상 호

e-mail : shlee@knusun.kangnung.ac.kr

1977년 인하대학교 수학과 졸업 (이학사)

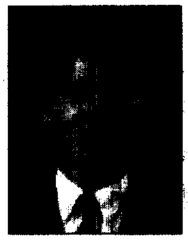
1980년 인하대학교 대학원 수학과 (이학석사)

1887년 인하대학교 대학원 수학과 (이학박사)

1997년~1998년 스탠포드 대학교 통계학과 객원교수

1985년~현재 강릉대학교 자연대학 통계학과 교수

관심분야 : 신경망, 데이터마이닝, 정보공학



전 홍 석

e-mail : hsjorn@anova.inha.ac.kr

1972년 서울대학교 수학과 졸업(이학사)

1977년 서울대학교 대학원(이학석사)/통계학

1986년 미국 Wisconsin대 (이학박사)/통계학

1979년~1983년 충북대학교 통계학과 교수

1983년~1988년 충북대학교 전산통계학과 교수

1988년~현재 인하대학교 통계학과 교수

관심분야 : 통계계산, 베이지안 학습, 신경망, 데이터마이닝