

ELM을 이용한 개선된 속성선택 기법

Effective Feature Selection Algorithm by Extreme Learning Machine

조재훈¹, 이대종², 전명근¹

¹충북 청주시, 충북대학교 전기전자컴퓨터공학부

²충북 청주시, 충북대학교 BK21 충북정보기술사업단

E-mail : mgchun@chungbuk.ac.kr

요 약

본 논문에서는 ELM(Extreme Learning Machine)을 이용하여 계산속도 뿐만 아니라 성능면에서도 우수한 입력 속성선택 기법을 제안한다. 일반적으로 입력 속성 선택문제는 다양한 속성들의 영향을 고려함으로써 모든 입력속성들을 평가하는데 많은 계산량이 요구되는 단점이 있다. 이러한 문제점을 개선하기 위하여 학습속도가 기존의 신경회로망에 비하여 월등히 우수한 ELM 알고리즘을 적용한다. 입력속성 선택은 ELM으로부터 산출된 출력값을 이용하여 출력 오차에 영향이 큰 속성들 순으로 순위를 결정한 후, 전방향 선택이나 후방향 선택기법을 이용하여 입력속성을 선택한다. 제안된 방법은 다양한 데이터에 적용하여 타당성을 검증한다.

Key Words : Feature selection, ELM, Forward selection, Backword selection

1. 서 론

신경회로망의 우수한 성능을 위해서 정확한 데이터와 학습기법은 중요한 요소들이다. 불필요한 데이터와 잡음이 섞인 부적절한 데이터들은 신경회로망의 정확도와 성능에 큰 영향을 주고, 많은 양의 데이터들은 오랜 학습시간과 실제 문제의 적용을 어렵게 만든다. 이러한 문제점들을 해결하기 위하여 주요 변수만을 추출하여 데이터의 수를 줄이는 방법들이 많이 연구되어지고 있다 [1-3].

주요 변수를 추출하기 위해 주로 사용하는 방법에는 변수생성, 변수 추출 및 부분집합 선정 기법이 있다 [4][5]. 이 중에서, 부분집합선정 방법은 적용방법에 따라 필터 기법과 래퍼(Wrapper) 기법으로 구분된다. 필터 기법은 데이터의 전처리 작업을 통해 주요 부분속성집합을 선택하고 선정된 집합들을 이용하여 기계학습 기법으로, 계산상의 효율성은 높으나 래퍼 기법에 비하여 성능이 저조한 단점이 있다[6].

래퍼기법은 주요 변수의 선택에서부터 해당 신경망과 같은 귀납적 기법을 결합하여 수행하여 선택된 부분속성집합에 대한 평가를 통하여 최적의 부분속성집합을 찾는 방식이다. Hsu et[7]는 신경회로망에서 가중치를 기반으로 하는 ANNIGMA(artificial neural net input gain

measurement approximation)-래퍼 기법을 제안하였다. 이 방법은 입력데이터에 대한 신경회로망의 가중치들을 계산하여 입력에 대한 출력을 근사화한 이득을 계산하여 신경회로망의 출력에 미치는 입력 속성들의 영향을 평가하고 정렬한 후, 순차적으로 입력속성을 제거하거나 첨가하여 부분집합을 탐색하는 방법이다. 이 방법은 기존 방법보다 빠른 탐색 시간을 갖지만, 입력 속성들이 많은 데이터에 대해서는 계산속도가 많이 요구되는 문제점을 갖는다.

따라서 본 논문에서는 신경회로망의 한 구조인 ELM을 이용하여 계산속도 뿐만 아니라 성능면에서도 우수한 입력 속성선택 기법을 제안한다. ELM은 기존의 신경회로망과 달리 후방향을 고려하는 것 없이 한 번의 전방향 단계에서 학습이 종료됨으로 학습시간이 다층신경망에 비해 월등히 우수한 장점을 갖는다. 제안된 입력속성 알고리즘은 ELM 알고리즘을 적용함으로써 출력값을 산출한 후, 전방향 선택이나 후방향 선택기법을 이용하여 속성선택을 수행하는 방법으로 빠른 속성선택을 할 수 있는 장점이 있다. 제안된 방법의 유용성과 성능을 평가하기 위하여 UCI Machine-Learning Repository[8]데이터에 적용하고 기존 기법들의 결과들과 비교하여 그 타당성을 보이고자 한다.

2. ELM의 구조 및 학습 알고리즘

ELM 알고리즘은 고전적 학습 방법들의 단점과 빠른 학습시간을 위하여 입력 가중치와 뉴런들의 바이어스들의 학습을 하지 않고 오직 출력 가중치들을 넘 최소자승해(norm least-squares solution)와 모어-페로스의 일반화된 역행렬을 이용하여 구한다 [9].

출력층에 사용되는 활성화 함수는 선형함수를 사용한다. ELM 알고리즘의 주요 장점을 열거하면 다음과 같다. ① 반복학습을 하지 않기 때문에 다른 학습 방법들에 비해 속도가 매우 빠르다. ② 경사 기반 학습 방법들은 오직 출력 오차를 최소화 하기 때문에 가중치들의 크기를 고려하지 않지만 ELM 알고리즘은 가장 작은 출력 오차와 노름최소자승 방법을 이용하여 가장 작은 가중치들까지 구하기 때문에 일반화 성능이 우수하다. ③ 경사 기반 학습 방법은 오직 미분 가능한 활성화 함수를 사용하지만 ELM 알고리즘은 미분가능하지 않은 활성화 함수를 가지는 한 개의 은닉층으로 구성된 신경 회로망을 학습할 수 있다. ④ 경사 기반 학습 방법이 국부최소, 부적절한 학습률의 효과 그리고 over-fitting과 같은 문제점이 있는 반면 ELM은 한 번의 학습으로 해에 도달하는 경향이 있다.

그림 1에서는 ELM의 구조 및 학습과정에 대한 흐름도를 나타냈다. ELM의 학습과정을 설명하기 위해서 N 개의 데이터 중에서 임의의 입출력 데이터 (x_i, t_i) 를 고려하자. 여기서, 입력 데이터 $x_i = [x_{i1}, \dots, x_{in}]^T \in \mathbb{R}^n$ 이고, 목표 출력값 $t_i = [t_{i1}, \dots, t_{im}]^T \in \mathbb{R}^m$ 이다. 은닉층의 뉴런의 개수가 \tilde{N} 개이고 활성화 함수가 $g(x)$ 인 단일 은닉층을 갖는 신경회로망의 실제 출력값 o_j 는 다음과 같다.

$$\sum_{i=1}^{\tilde{N}} \beta_i \{ g(w_i \cdot x_j + b_i) \} = o_j, \quad j = 1, \dots, N \quad (1)$$

여기서, $w_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T$ 은 입력 뉴런들과 i 번째 은닉 뉴런 사이에 연결된 가중치 벡터, $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$ 은 출력 뉴런들과 i 번째 은닉 뉴런 사이에 연결된 가중치 벡터이며, b_i 는 i 번째 은닉 뉴런의 바이어스 값을 나타낸다. 은닉 뉴런 \tilde{N} 을 가지는 단일 은닉층을 갖는 신경회로망은 학습을 통하여 N 개의 모든 학습 데이터에 대하여 오차가 제로인 다음 식 (2)를 만족할 경우 식 (3)를 만족하는 β_i, w_i, b_i 가 존재한다.

$$\sum_{j=1}^{\tilde{N}} \| o_j - t_j \| = 0 \quad (2)$$

$$\sum_{i=1}^{\tilde{N}} \beta_i g(w_i \cdot x_j + b_i) = t_j, \quad j = 1, \dots, N \quad (3)$$

N 개의 방정식으로 표현되는 식 (4)를 아래와 같이 간단히 표현하자.

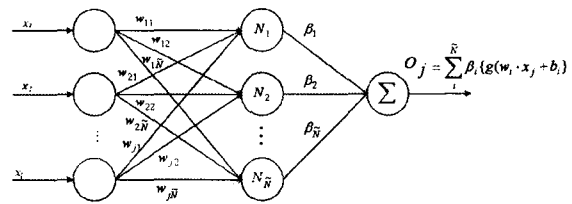
$$H\beta = T \quad (4)$$

H 는 신경망의 은닉층의 출력 행렬값을 나타낸다. H 의 i 번째 열은 입력 x_1, x_2, \dots, x_N 의 각각에 대한 i 번째 은닉 뉴런의 출력 벡터이다. 여기서 β 의 극소노름최소자승해 $\hat{\beta}$ 는 아래와 같이 구해질 수 있다.

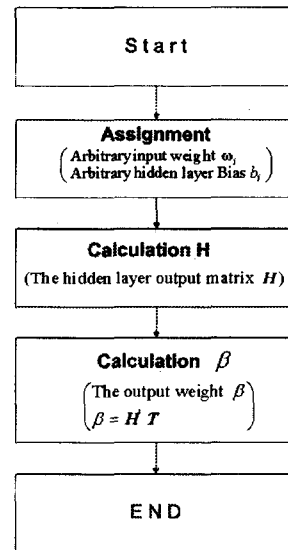
$$\hat{\beta} = H^\dagger T \quad (5)$$

여기서 H^\dagger 는 H 에 대한 모어-페로스의 일반화된 역행렬을 의미한다.

위의 계산 방법과 같이 ELM 기법은 매우 빠른 학습 속도를 가지고 있다.



(a) ELM의 구조



(b) ELM의 학습 흐름도

그림 1. ELM의 구조 및 학습 흐름도

3. ELM 기법을 이용한 속성선택

그림 2는 일반적인 신경회로망의 가중치를 이용한 Wrapper 기법의 프로그램과 데이터 흐름도를 보이고 있다[7]. 가중치 기반의 일반적인 Wrapper 기법은 속성집합들의 오차를 평가하는데 많은 시간이 걸리는 단점을 가지고 있다.

본 장에서는 ELM기법을 이용하여 속성선택을 수행하는 방법에 대해서 설명한다. 그림 2에서 (1)로 표시된 부분의 입력은 모든 입력 속성과 학습 데이터들의 집합이고, 출력은 속성부분 집합(feature subset)이다. 전체의 입력 데이터를 10-fold로 나누고 데이터의 90%를 학습데이터 나머지 10%를 검증데이터로 사용한다. 그림 2에서 (2)로 표시된 부분은 (1)로 표시된 단계에서 학습된 가중치를 기반으로 검증데이터를 사용하여 검증 오차값 E를 계산한다. 검증 오차값 E는 (3)으로 표시된 속성들의 이득에 가중치로 작용하여 속성순위를 결정할 때 사용한다. 각각의 속성부분집합들에 대한 가중치 이득은 아래의 식으로부터 계산한다.

$$LG_{ik} = \sum_j |V_{ij} \times W_{jk}| \quad (6)$$

여기서 i, j, k 는 신경회로망에서 각각 입력, 은닉층, 출력층 노드 번호들이다. V 는 입력층과 은닉층에 연결된 노드들의 가중치이고, W 는 은닉층과 출력층에 연결된 노드들의 가중치이다. 식 (6)에서 계산된 지역 이득 LG_{ik} 를 이용하여 ELMIGMA 점수를 아래와 같이 산출한 후 최대값이 100이 되도록 정규화 시켰다.

$$ELMScore_{ik} = \left[\frac{LG_{ik}}{\max(LG_k)} \times 100 \right] \times E \quad (7)$$

그림 2의 (8)로 표시된 부분에서는 (3)으로 표시된 부분에서 계산된 ELM 점수를 기반으로 속성들의 순위를 결정하고, (4)로 표시된 부분에서 후보 속성 부분 집합이 생성된다. 이러한 순환을 계속 반복하면서 속성들의 제거와 추가를 통하여 최적의 속성부분집합을 탐색하게 된다.

제안된 ELM 기반 속성선택 방법에서의 탐색은 BE, BEB, BSE 를 포함하는 후방 선택의 3 가지 탐색 방법으로 수행되어졌다. BE (Backward elimination)는 ELMscore로 결정된 순위와 모든 속성들을 가지고 오차 비율을 얻기 위하여 학습 반복을 실행 하는 후방 제거 기법이다. 이 기법은 반복적으로 오차 비율이 커질 때까지 속성선택의 반복을 통해 가장 나쁜 ELM점수가 좋지 못한 순서대로 속성을 제

거한다.

BEB(backward elimination with backtracking)는 backtracking을 허용하는 BE의 기법이다. 이 기법은 에러 비율이 올라갈 때 속성 선택을 종결시키는 대신 이전의 제거되었던 속성들이 다시 저장되고 더 나쁜 속성이 제거된다. 이 과정은 각각의 속성 부분집합에 대하여 제거를 통하여 성능이 향상될 때 까지 반복된다.

BSE(backward stepwise elimination) 기법은 BEB가 너무 느릴 때 많은 데이터들에 대한 속성 선택을 빠르게 수행하는 기법이다. 이 기법은 조기에 ELM 점수가 좋지 않은 속성들을 하나씩 제거하는 것이 아니라 여러 개씩 제거해 나가고 성능이 떨어지면 제거했던 속성들을 다시 저장하여 속성 부분집합들의 성능을 평가하는 방법이다 [9].

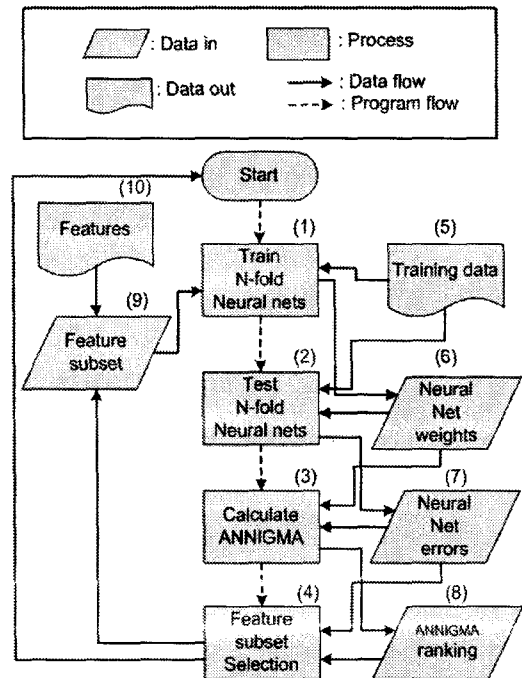


그림 2. 일반적인 Wrapper 기법에서의 데이터 및 프로그램 흐름도

4. 시뮬레이션 및 결과 고찰

제안한 기법의 성능을 평가하기 위하여 UCI(University of California at Irvine)의 기계 학습 연구에서 제공하는 연구용 데이터들 사용하였다. 모든 분석과 검증은 타당성의 확보를 위해서 데이터들을 무작위로 10번 구성하여 실험하는 10-fold 방식에 의해 수행하였다. 실험 환경은 Pentium4 1.6GHz, 1G 메모리에서 수행하였다. 본 연구에서 실험한 데이터는 표 1과 같다. 표 2는 제안된 알고리즘과 기존의 다른 속성부분집합 선택 기법들과의 성능을 비교하

였다. NN은 일반적인 신경회로망을 이용한 결과이며 속성 수는 반복 연산을 하여 평균적으로 사용되는 속성의 수를 나타낸 것이다. 표 2에서 알 수 있듯이 제안된 기법은 수렴속도 뿐만 아니라 인식성능 면에서도 다른 속성선택 기법에 비해 우수한 결과를 보임을 확인할 수 있다.

표 1. 실험에 사용된 데이터들의 특성 및 탐색전략

데이터명	학습 데이터	검증 데이터	반복 횟수	뉴런수	활성함수	탐색 전략
3P	100	40	8	6(60)	{logsig+purelin}	BEB
Monk3a	122	432	15	6(59)	{tansig+purelin}	BE
Monk3b	122	432	8	3(100)	{purelin+purelin}	BE
Cancer	399	300	10	12(9)	{logsig+logsig}	BEB
Credit	490	200	10	10(35)	{tansig+purelin}	BE
Heart(LB)	133	67	10	2(15)	{purelin+purelin}	BEB
Ionosphere	200	151	10	22(50)	{logsig+purelin}	BSE
Pima	576	192	10	6(20)	{tansig+purelin}	BEB
Vote	218	217	8	3(30)	{purelin+purelin}	BE

참고) 뉴런의 수중에서 ()는 제안된 방법에 대한 뉴런의 수를 의미함

5. 결론

신경회로망의 성능은 정확하고 적절한 입력 데이터들에 따라 차이가 현저하기 때문에 최근에는 데이터들의 주요 변수를 선택하는 연구들이 많이 진행되어져 왔다. 그러나 기존의 연구들은 전체의 데이터들에 반영하여 부분 집합을 산출하여 평가하기 때문에 시간이 많이 걸리는 단점들이 있다. 본 논문에서는 가중치들의 이득값들과 신경회로망의 빠른 학습방법인 ELM 기법을 융합하여 빠르게 속성부분집합을 선택하는 기법을 제안하였다. 제안된 방법은 기존의 다른 기법들보다 속도와 탐색성능에서 우수한 결과를 보임을 확인할 수 있었다.

2. 제안된 기법에 의한 속성선택과 다른 기법들의 성능비교

데이터	신경회로망		Conventional Wrapper			ANNIGMA-Wrapper			ELMGMA-Wrapper		
	속성수	오차(%)	속성수	오차(%)	시간(sec)	속성 수	오차(%)	시간(sec)	속성 수	오차(%)	시간(sec)
3P	13	9.3±14.7	4.4±3.1	0.5±1.4	1165	3.0±0.0	0.0±0.0	404	3.18±0.3	0.0±0.0	167
Monk3a	6	10.0±5.2	3.4±1.6	5.1±3.4	125	2.3±0.7	2.9±0.8	435	5.4±0.6	2.67±0.8	16.95
Monk3b	15	2.8±0.0	4.4±1.1	2.8±0.0	519	2.2±0.4	2.8±0.0	623	6.4±0.7	2.22±0.1	139
Cancer	9	4.1±4.7	7.2±1.2	3.6±1.1	451	5.8±1.3	3.5±1.2	513	6.1±1.3	1.4±0.1	28.3
Credit	9	14.1±1.7	13.4±1.0	14.4±0.8	1712	6.7±2.5	12.0±0.8	1616	9.2±3.6	11.71±0.3	202
Heart(LB)	13	20.2±0.2	7.3±1.9	25.7±1.8	352	2.7±1.2	22.3±2.0	795	8.4±1.7	18.6±0.6	27.3
Ionosphere	34	11.4±3.9	32	10.2	138065	9.0±2.5	9.8±1.3	11497	29.4±1.8	8.5±0.4	406.3
Pima	8	24.1±5.0	6.9±1.0	23.0±1.3	179	5.2±1.4	22.2±1.4	997	5.5±0.9	21.8±0.4	67.5
Vote	16	3.2±0.0	3.2±1.4	3.2±0.2	1052	3.3±1.9	3.1±0.2	522	12.4±2.56	2.9±0.3	35.2

참 고 문 헌

[1] Kohavi. R. and D. Sommerfield, "Feature subset selection using the wrapper method : Overfitting and dynamic search space topology," Proceedings of the First International Conference on Knowledge Discovery and Data Mining(KDD '95). Menlo Park, CA: AAAI Press, 1995

[2] M. Dash and H. Liu, "Hybrid search of feature subsets," Proc. PRICAI, Singapore, 1998

[3] Jiang Li, M.T. Manry, P.L. Narasimha, C.Yu, "Feature Selection Using a Piecewise Linear Network," IEEE Trans. on Neural Networks. V.17, No 5, 2006.

[4] Wnek, J. and R.S. Michalski, "Hypothesis Driven Constructive Induction in AQ17-HCI: a Method and Experiments," Machine Learning, V.14, 1994.

[5] Wyse, N., R.Dubes, and A.K. Jain, "a Critical Evaluation of intrinsic Dimensionality Algorithms," In Pattern Recognition in Practice, E.S. Gelsema and L.N. Kanal, (Eds.), Morgan Kaufmann Publishers, 1980.

[6] 강부식, 박상찬, "신경망의 민감도 분석을 이용한 귀납적 학습 기법의 변수 부분집합 선정." 한국지능 정보시스템학회논문지, 제7권, 2호 2001.

[7] Chun-Nan Hsu, Hung-Ju Huang, and Dietrich Schuschel, "The ANNIGMA-Wrapper Approach to Fast Feature Selection for Neural Nets," IEEE Trans. on Syst. man and Cybernetics-PART B: CYBERNETICS, V. 32, NO. 2, 2002

[8] C. J. Merz and P. M. Murphy. UCI repository of machine learning databases. Dept. Computer Science, Univ. California, Irvine. [Online]. Available: <http://www.ics.uci.edu/~mllearn/MLRepo-sitory.html>.

[9] G. B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: a new learning scheme of feedforward neural networks, in: Proceedings of the International Joint Conference on Neural Networks (IJCNN2004), 25 - 29 July 2004.