러프집합을 이용한 신경망 학습알고리즘

Learning Algorithm of Neural Networks Using Rough Set

손 현숙*, 피 수 영*, 정 환 묵*
*대구효성가톨릭대학교 전자정보 공학부

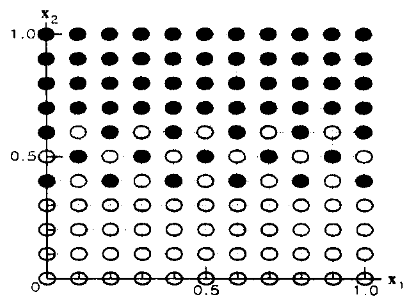
Son Hyun Sook*, Pi Su Young*, Chung Hwan Mook*
*Faculty of Electronics & info. Engineering,
Catholic Univ. of Taeguhyosung

요 약

패턴인식중에서 가장 기본적인 문제인 판별문제를 대상으로 러프집합을 이용한 판별분석을 행하는 신경망의 학습알고리즘을 제안한다. 어떤군에 속할 것인가의 경계영역을 명확히 하는 것을 목적으로 한다. 2군 판별의 문제를 각 데이터가 각 군에 속한 정도를 표현하는 소속함수(membership function)을 이용하며, 경계영역에 대한 문제는 소속함수를 구간치 함수로 확장하여 가능성과 필연성을 동시에 표현할 수 있는 학습 알고리즘을 제안한다.

1. 서론

일반적으로 2군판별을 행하는 가장 간단한 방법은 선형판별함수를 이용하는 것이다. 한편 계층형 신경망을 이용하면 비선형함수에 의해 판별이 가능하게 된다. 이러한 비선형함수는 백프로파게이션알고리즘을 이용하여 학습에 의해 구할수 있다. 그러나 실제학습에서 그러한 신경망이 얻어진다는 보증은 없다. 또 잘못된 학습의 문제도 있고 주어진 데이터점을 완전히 판별할 수 있도록 신경망이 미학습 데이터에 대해서도 유효하다고는 제한하지는 않는다. 따라서 본 연구에서는 우리들의 직관적이해와 일치한 판별분석을 행하기 위해 2군 판별의 문제에 대해 러프집합을 이용한 가능성 및 필연성을 고려한 판별분석을 행한다. 구체적으로는 어떤군에도 속할 가능성이 있다라고 생각되어지는 경계영역을 명확히 하는 것을 목적으로 신경망의 학습알고리즘을 제안한다. 본논문에서는 [그림 1]의 도트로 표시된 영역을 어떤군에도 속할 수 있는 경계영역이라고 인정한다고 하는 입장을 취하고 있다.



[그림 1] 2군의 경계영역

2. 러프 집합

2.1 가능성 판별분석

러프집합이론[3]은 불명확한 또는 부정확한 정보를 다루기 위한 수학적 도구로써 제안되었다 $X \subseteq U$ 이고 R이 동치관계일때 X가 R-기본집합의 합집합이면 X는 R-definable이라고 하며,그렇지않으면 X는 R-undefinable이라고 한다. 즉 R-definable 집합은 지식베이스 K에서 정확하게 정의되는 부분집합이며, R-undefinable 집합은 이 지식베이스에서 잘 정의될 수 없는 집합이다. 러프집합은 U가 전체집합일 때 R은 $U \times U$상의 동치관계로써 $K=(U,R)$를 근사공간(Approximation Space)라 한다. 부분집합 $U1 \subseteq U$, $U2 \subseteq U$, 동치관계 $R \in IND(K)$일때 통상의 판별분석의 문제는 주어진 데이터점을 이용하여 전체집합 U를 다음과 같이 2분할하는 문제를 생각할 수 있다.

$$U = U_1 \cup U_2$$
$$U_1 \cap U_2 = \varnothing$$

다시말하면 전체집합 U 가 제1군을 표시하는 부분집합 $U1$ 과 제2군을 표시하는 부분집합 $U2$에 2분할 되어지고 있다. 이때 다음의 관계가 성립하면 $U1$, $U2$ 에 의한 전체집합 U의 분할은 주어진 데이터점을 올바르게 판별하는 것으로 된다.

$$x_p \in U1 \quad \text{for } all\ p\ in\ P_{G1}$$
$$x_p \in U2 \quad \text{for } all\ p\ in\ P_{G2}$$

그러나 경계영역이 존재하도록한 문제에서는 위의 식들은 적합하지가 않다. 본논문에서는 전체집합의 2분할이라고 하는방법을 이용하지 않고 러프집합을 이용한 각군에 속하는 가능성 및 필연성을 고려한 판별분석을 제안한다.

우선 가능성을 생각한 경우에는 다음과 같은 2개의 부분집합 R^*U1 , R^*U2를 이용한다.

$$U1\text{의 } R-\text{가능성}: R^*U1 = \bigcup\{Y \in U/R : Y \cap U1 \neq \varnothing\}$$
$$U = R^*U1 \cup R^*U2$$

여기서 R^*U1 은 제1군에 속하는 가능성이 있는 영역이고, R^*U2 은 제2군에 속하는 가능성이 있는 영역이다. 한편 필연성을 생각할 경우에는 다음과 같이 2개의 R_*U1 , R_*U2을 이용한다.

$$U1\text{의 } R \text{ 필연성}: R_*U1 = \bigcup\{Y \in U/R : Y \subseteq U1\}$$

$$U \supseteq (R_*U1 \cup R_*U2)$$
$$R_*U1 \cap R_*U2 = \varnothing$$

여기서 R_*U1은 필연적으로 제 1군에 속하는 영역이고 R_*U2은 필연적으로 제1군에 속하는 영역이다.

$$R^*U2 = U - R_*U1$$

$$R_*U2 = U - R^*U1$$

위 첫 번째 식은 필연적으로 제1군에 속하는 영역을 전체집합에서 제외되어 있는것이지만 제2군에 속할 가능성있는 영역으로 있다는 것을 나타내고 있다. 역으로 두 번째 식은 제1군에 속하는 가능성이 있는 영역을 전체집합에서부터 제외된것이지만 필연적으로 제2군에 속하는 영역으로 있는 것을 표시하고 있다.

따라서 구체적인 계산으로서는 제1군에 관계하는 2개의 영역 R^*U1 , R_*U1을 구하는것에 의해 제2군에 관계하는 2개의 영역 R^*U2 , R_*U2을 얻을수 있다. 어떤군에도 속할 가능성이 있는 영역 $R^*U1 \cap R^*U2$ 은 다음과 같이 계산을 할 수 있다.

$$\begin{aligned} R^*U1 \cap R^*U2 &= R^*U1 \cap (U - R_*U1) \\ &= (R^*U1 \cap U) - (R^*U1 \cap R_*U1) \\ &= R^*U1 - (R^*U1 \cap R_*U1) \end{aligned}$$

더욱이 가능성이론으로부터 $R_*U1 \subseteq R^*U1$가 성립된다고 생각할수 있으므로 어떤군에도 속할 가능성이 있는 영역은 다음과 같이 계산되어진다.

$$R^*U1 \cap R^*U2 = R^*U1 - R_*U1 \cdots \ (1)$$

3. 학습알고리즘

3.1 BP알고리즘

BP알고리즘에 대해 간단히 알아보면 하나의 출력 유닛과 n_2 의 중간층 유닛수를 가진 3층구조로 표시할수 있다. 입력유닛로부터의 입력벡터는 P번째의 데이터점 위치를 표시하는 n차원 벡터이다. 각 유닛으로부터 출력치는 다음과 같이 계산되어 진다.

$$\text{입력층} \quad O_x = x_{pi}, \qquad i = 1, 2, \cdots, n \cdots \ (2)$$

중간층

$$O_{pj} = f(net_{pj}),\ j=1,2,\cdots,n_2$$

$$net_{pj} = \sum_{i=1}^{n} W_{ij} O_{pi} + \theta_j \quad \cdots (3)$$

출력층

$$O_p = f(net_p)$$

$$net_p = \sum_{i=1}^{n_2} W_j O_{pj} + \theta \quad \cdots (4)$$

단 입.출력함수는 다음과 같은 시그모이드함수를 이용한다.

$$f(net) = 1 / \{1+\exp(-net)\} \quad \cdots (5)$$

여기서 벡터 x_p가 입력되어질때 신경망으로부터 기대하는 출력치(교사신호)를 t_p 로 한다. 이때 p번째의 데이터점에 대한 신경망학습은 다음과 같은 평가함수가 최소가 되도록 행하여 진다.

$$E_p = (t_p - o_p)^2 / 2 \quad \cdots (6)$$

따라서 결합강도 및 임계치의 수정량은 다음과 같다.

$$\Delta_p W_j = \eta(-\partial E_p / \partial W_j) = \eta \delta_p o_{pj} \quad \cdots (7)$$

$$\Delta_p W_{ji} = \eta(-\partial E_p / \partial W_{ji}) = \eta \delta_{pj} o_{pi} \quad \cdots (8)$$

$$\Delta_p \theta = \eta(-\partial E_p / \partial \theta) = \eta \delta_p \quad \cdots (9)$$

$$\Delta_p \theta_j = \eta(-\partial E_p / \partial \theta_j) = \eta \delta p_j \quad \cdots (10)$$

단 δp , δp_j은 다음과 같이 주어진다.

$$\delta p = (t_p - O_p) O_p (1 - O_p) \quad \cdots (11)$$

$$\delta p_j = O_{pj}(1 - O_{pj}) \delta p W_j \quad \cdots (12)$$

3.2 신경망에 의한 판별분석

우선 입력벡터 x_p 에 대응한 교사신호 t_p를 다음과 같이 정할수 있다.

$$t_p = \begin{cases} 1 & \text{for } all\ P\ in\ P_{G1} \\ 0 & \text{for } all\ P\ in\ P_{G2} \end{cases}$$

3.1에서 표시한 BP알고리즘에 의해 학습을 행한다. 학습후 신경망에 두어 입력벡터 x에 대응한 출력치를 $A1(x)$로 표시한다면 다음과 같이 전체집합 U 의 2분할을 행할수 있다.

$$U1 = \{ x / A1(x) \ge 0.5 ,\ x \in U \} \quad \cdots (13)$$

$$U2 = \{ x / A2(x) < 0.5 ,\ x \in U \} \quad \cdots (14)$$

이때 출력치 $A1(x)$가 다음관계를 만족하면 신경망에 의해 주어진 모든 데이터점이 올바르게 판별되어진다.

$$A1(x_p) \ge 0.5 \quad \text{for } all\ P\ in\ P_{G1}$$

$$A1(x_p) < 0.5 \quad \text{for } all\ P\ in\ P_{G2}$$

3.3 가능성판별분석을 위한 학습알고리즘

제1군에 속할 가능성이 있는 영역 R^*U1을 구하기위해 학습알고리즘을 표시한다. 3.1과 동일상태의 신경망으로서 출력층에 1개의 유닛을 갖는 계층형신경망을 생각하고 각 유닛 입.출력함수가 식 (2)-(5)에서 주어진다고 한다.

제안하는 학습알고리즘에는 학습에 있어 최소화되어지는 평가함수로써 데이터점에 가중치를 가한 2승오차의 총계를 이용한다. 구체적으로는 제1군에 속하는 데이터점에 대한 가중치를 1로 하고 제2군에 속하는 데이터점에 대한 가중치를 w(u)로 한다. 여기서 u 는 학습횟수이고 w(u) ≤ 1이다. 제1군에 속할 가능성이 있는 영역 R^*U1을 구하기위해 다음과 같은 학습알고리즘을 제안한다.

step 1 u:=1

step 2 p=1,2,...,m 에 대해 다음의 cost함수를 이용하고 신경망의 결합강도 및 임계치의 학습을 행한다.

$$p \in p_{G1} \ :\ E_p = (t_p - O_p)^2 / 2 \quad \cdots (15)$$

$$p \in p_{G2} \ :\ E_p = w(u) \cdot (t_p - O_p)^2 / 2 \cdots (16)$$

step 3 u:=u+1 로하고 step 2로 되돌아간다.

또 무게가중치 w(u)는 다음과 같은 조건을 만족하도록 설정되어진다.

$$w(u) \fallingdotseq 1 \quad \cdots (17)$$

$$w(u) \le w(u-1) \le 1 \quad \cdots (18)$$

$$\lim_{u\to\infty} w(u) \to 0 \quad \cdots (19)$$

가중치 $w(u)$로써는 다음과 같은 감소함수를 이용한다.

$$w(u) = a^u ,\quad 0 < a < 1$$

$$w(u) = \frac{1}{1+(cu)^b} ,\quad 0 < b ,\ c$$

이와같이 학습알고리즘을 이용하여 학습을 행한 신경망의 출력치를 $R^*A1(x)$로 표시하는 것으로 하면 충분히 학습을 행한후에는 다음관계가 성립한다고 고찰할 수 있다.

$$R^*A1(x_p) \fallingdotseq 1 \quad \text{for } all\ P\ in\ P_{G1} \cdots (20)$$

$$R^*U1 = \{x / R^*A1(x) \ge 0.5 ,\ x \in U\} \cdots (21)$$

3.4 필연성판별분석을 위한 학습알고리즘

제1군에 속하는 필연성이 있는 영역 R_*U1을 구하기 위해서는 step 2의 (15)식과 (16)식을 역으로 하면 좋다. 즉 신경망의 학습알고리즘은 다음과 같다.

step 1 u:=1

step 2 p=1,2,...,m 에 대해 다음의 cost함수를 이용하고 신경망의 결합강도 및 임계치의 학습을 행한다.

$$p \in p_{G1} \ : \ E_{p} = w(u).(t_p - O_p)^2 \ / \ 2 \quad \cdots (22)$$

$$p \in p_{G2} \ : \ E_{p} = (t_p \ - \ O_p)^2 \ / \ 2 \quad \cdots (23)$$

step 3 u:=u+1 로하고 step 2로 되돌아간다.

여기서 가중치 $w(u)$은 3.3과 같은 상태에 식 (17)-(19)을 만족하도록 주어지는 것으로 한다. 신경망의 출력치 $R_*A1(x)$로 표시한다면 충분히 학습을 행한후에는 다음 관계가 성립한다고 생각할 수 있다.

$$R_*A1(x_p) \fallingdotseq 0 \quad \text{for} \quad all \ \ P \ \ in \ \ P_{G2} \quad \cdots (24)$$

여기서 3.3과 같은 방식으로 생각하면 제2군에 속하는 가능성있는 영역 R^*U2가 다음과 같이 설정되어 진다.

$$R^*U2 = \{x \ / \ 1 - R_*A1(x) \geq 0.5 \ , x \in U\} \quad \cdots (25)$$

$$R_*U1 = \{x \ / \ 1 - R_*A1(x) < 0.5 \ , x \in U\} \quad \cdots (26)$$

즉

$$R_*U1 = \{x \ / \ R_*A1(x) > 0.5 \ , x \in U\} \quad \cdots (27)$$

로 한다.

4. Membership 함수

본 논문에서 제안한 2군 판별의 문제는 데이터점이 각군에 속한 정도를 표현한 멤버쉽함수를 구하는 문제라고 생각 할 수 있다. 그러나 경계영역이 존재하도록 한 경우에는 통상의 BP알고리즘에·의해 얻어진 결과를 그대로 소속함수라고 생각하는 것은 우리들의 직관과는 일치하지 않는다. 이러한 문제에 대해 본 논문에서는 소속함수를 구간치함수에 확장할수 있는 경우를 생각해 본다. 이경우에는 제1군을 표현한 퍼지집합의 소속함수를 구간치함수 $[R_*A1(x) \ , \ R^*A1(x)]$로 표현하는것에 의해 데이터 점 x 가 제1군에 속하는 필연성과 가능성을 동시에 표현할 수 있다. 다음에는 소속함수를 실수치구간에만 한정하는 경우에 대해 생각해보자. 이러한 경우에는 제안한 2개의 알고리즘에 의해 얻어진 신경망우로부터의 출력치의 평균을 취한것에 의해 소속함수를 얻을수 있다.

$$\widehat{U1(x)} = \{R^*A1(x) \ + \ R_*A1(x)\} \ / \ 2$$

5. 결론

본 논문에서는 인간의 직관적 이해와 일치하는 판별분석을 행하기 위해 러프집합을 이용한 가능성과 필연성의 개념을 기초로한 2군 판별분석을 행하는 것을 목적으로 신경망의 학습알고리즘을 제안 했다. 학습알고리즘에관한 종래의 연구들은 주어진 데이터에 가능한 일치하는 신경망을 가능한한 짧은 계산시간으로 얻을수 있는 것을 목적으로 하는 것이다. 이것에 대해 제안한 것은 다차원속성공간내에 어떤 군에도 속할 가능성이 있는 경계영역이 존재하고 있다는 것을 전제로하고 있다. 따라서 본논문에서 제안한 것을 이용하여 미지의 데이터의 판별을 행할경우에는 2개의 군이 강제적으로 분할하는 것이 아니고 어떤 군에도 속할 가능성이 있다라고 판단을 행하는 것이 가능하다.

참고문헌

1. K.Funahashi : On the Approximate Realization of Continuous Mapping by Neural Networks,Neural Networks, VOL. 2, pp 183-193, 1989
2. D. E. Rumelhart, G. E. Hinton & R. J. Williams: Learing Representations by Back-Propagating Errors, Nature, Vol. 323,pp. 533-536, 1986.
3. D. Dubois & H. Prade : Possibility Theory, Plenumpress, 1988
4. Z, Pawlak, Rough sets, Theoretical Aspects of Reasoning about Data, Kluwer Academic Publisher,1991.
5. Z, Pawlak, "Rough Sets", International Journal of computer and Information Science, 1982
6. D. E. Rumelhart and J. L. McCelland, Parallel processing, Explorations in the Microstructure of cognition, Vol.1, pp318-362, 1986
7. M. R. Anderberg, Cluster Analysis for Applications, New York San Francisco London, 1973