

데이터와 클러스터들의 대표값들 사이의 거리를 이용한 퍼지 학습법칙

Fuzzy Learning Rule Using the Distance between Datum and the Centroids of Clusters

김용수

대전시 동구 용운동 96-3 대전대학교 컴퓨터공학과
E-mail : kystj@dju.ac.kr

요 약

학습법칙은 신경회로망의 성능에 중요한 영향을 미친다. 본 논문은 데이터와 클러스터들의 대표값들 사이의 거리를 고려하여 학습률을 정하는 새로운 퍼지 학습법칙을 제안한다. 클러스터들의 대표값을 조정할 때, 이러한 고려는 outlier에 비하여 결정경계선 근처에 있는 데이터의 반영도를 높임으로써 outlier의 클러스터의 대표값에 미치는 영향도를 낮출 수 있다. 따라서 outlier들이 결정경계선을 악화시키는 것을 방지할 수 있다. 이 새로운 퍼지 학습법칙을 IAFC(Integrated Adaptive Fuzzy Clustering) 신경회로망에 적용하였다. 제안한 퍼지 신경회로망과 다른 감독 신경회로망들의 성능을 비교하기 위하여 iris 데이터를 사용하였다. iris 데이터를 사용하여 테스트한 결과 제안한 퍼지 신경회로망의 성능이 우수함을 보였다.

Key Words : Learning rates, Learning rule, IAFC neural network

1. 서 론

학습법칙은 신경회로망의 성능에 중요한 영향을 미친다. 학습법칙은 감독 학습법칙, 무감독학습법칙, 강화 학습법칙으로 분류된다[1]. 감독 학습법칙 중에 LVQ(Learning Vector Quantization)은 많이 사용되어왔다[1, 2]. LVQ는 학습률을 연결강도의 수렴을 위하여 0과 1사이의 작은 값을 초기값으로 사용하고 반복횟수가 증가함에 따라서 감소시키는 방법을 사용하였다. 그러나, 이러한 학습률을 선정하는 방법은 연결강도의 수렴을 보장하나 정교한 학습률의 조정을 통한 성능의 개선이라는 관점에서는 부족함이 있다. 이러한 한계점을 극복하고자 하는 연구가 꾸준히 이루어져왔다[3-8]. Chung과 Lee는 LVQ에 퍼지 소속도를 결합한 FLVQ(Fuzzy Learning Vector Quantization)를 제안하였다[4, 5]. 이 FLVQ는 퍼지 목적함수를 최적화하는 것에 근간을 두고 있으나, 목표 소속도와 실제 소속도의 차이값을 활용하기 때문에 목표 소속도를 알아야한다. 그러나 실제 상황에서 목표 소속도를 얻는 것이 쉽지 않다. Karayiannis도 목적함수를 최적화하는 것에 근간을 둔 FLVQ를 제안하였다[6, 7]. Bezdek등은 FCM(Fuzzy c-Means)을 활용한 FLVQ를 제안하였다[3, 8].

본 논문에서는 새로운 FLVQ를 제안한다. 이 새로운 FLVQ는 분류가 맞느냐 틀리느냐에 따라서 다른 학습률을 사용한다. 분류가 맞으면 데이터의 위치와 데이터와 존재하는 클러스터들의 대표값들 사이의 상대적인 위치를 동시에 고려한 학습률을 사용한다. 이는 결정경계선 근처에 있는 데이터는 최적의 결정경계선에 대한 정보를 가지고 있지만 outlier는 최적의 결정경계선에 대한 정보를 가지고 있지 않고 오히려 클러스터의 대표값을 outlier쪽으로 이동시켜 성능을 떨어뜨리기 때문이다. 따라서 클러스터의 대표값을 조정할 때 결정경계선 근처에 있는 데이터의 반영도를 높이고 outlier의 반영도는 낮추는 것이다. 분류가 틀리면 퍼지 소속도를 활용하여 클러스터의 대표값을 조정한다.

새로운 FLVQ를 학습법칙으로 IAFC(Integrated Adaptive Fuzzy Clustering) 퍼지 신경회로망에 적용하였다. IAFC 퍼지 신경회로망은 안정하면서도 유연하다. 즉 과거에 학습한 내용을 유지하면서도 새로운 데이터를 반영하는 것이다. 이는 IAFC 퍼지 신경회로망의 제어구조가 ART-1(Adaptive Resonance Theory) 신경회로망의 제어구조와 유사하기 때문이다[9]. 또한, 이러한 구조는 Kohonen의 SOM(Self-Organizing Feature Map)이 가지고 있는

underutilization 문제를 해결하였다. Underutilization 문제는 초기에 원하는 클러스터의 개수를 초기값으로 주었음에 불구하고 클러스터가 원하는 개수만큼 형성되지 않는 것이다. IAFC 퍼지 신경회로망은 클러스터의 개수와 크기를 vigilance parameter로 조정한다. Vigilance parameter가 크면 클러스터의 개수가 작고 클러스터의 크기가 크다. 반면에 vigilance parameter가 작으면 클러스터의 개수가 많고 클러스터의 크기가 작다.

제안한 퍼지 신경회로망의 성능과 LVQ 알고리즘과 오류 역전파 신경회로망의 성능을 비교하기 위하여 iris 데이터를 사용하였다. Iris 데이터는 클러스터링 알고리즘의 성능을 비교하는데 많이 사용되어 왔다.

2. Supervised IAFC 퍼지 신경회로망 5

입력 벡터가 supervised IAFC 퍼지 신경회로망 5에 가해진 후, 출력 뉴런들 간의 경쟁은 winner-take-all 식으로 일어난다. 입력벡터와의 유클리디안 거리가 제일 작은 연결강도를 가진 출력뉴런이 경쟁에서 승리한다. 즉 I번째 출력뉴런,

$$I = \min_i \| \mathbf{x} - \mathbf{v}_i(t) \|,$$

이 경쟁에서 승리한다. 여기서 \mathbf{x} 는 입력벡터이고, $\mathbf{v}_i(t)$ 는 i번째 출력뉴런의 연결강도이다.

출력 뉴런들 중에서 승자를 결정한 후에 supervised IAFC 퍼지 신경회로망 5는 vigilance 기준에 의해 vigilance test를 한다 :

$$e^{-u_i} \| \mathbf{x} - \mathbf{v}_I(t) \| \leq T,$$

여기서 T는 vigilance parameter이다. 퍼지 소속도 u_i 는 다음과 같이 정의된다.

$$u_i = \frac{\left[\frac{1}{\| \mathbf{x} - \mathbf{v}_I(t) \|^2} \right]^{\frac{1}{m-1}}}{\sum_{i=1}^n \left[\frac{1}{\| \mathbf{x} - \mathbf{v}_i(t) \|^2} \right]^{\frac{1}{m-1}}},$$

여기서 n은 존재하는 committed 출력 뉴런의 개수이고, $m \in [1, \infty]$ 는 weight exponent이며, 실험적으로 2로 놓았다.

만약 승자가 vigilance 테스트를 만족하면은 supervised IAFC 퍼지 신경회로망 5는 승자의 연결강도를 다음의 새로운 퍼지 학습법칙을 사용하여 조정한다.

$$\mathbf{v}_I(t+1) = \mathbf{v}_I(t) + \frac{r(d)}{t} \cdot (\mathbf{x} - \mathbf{v}_I(t)) \text{ if } \mathbf{x} \text{ is classified correctly}$$

$$\mathbf{v}_I(t+1) = \mathbf{v}_I(t) - \frac{u_i}{t} \cdot (\mathbf{x} - \mathbf{v}_I(t)) \text{ if } \mathbf{x} \text{ is classified incorrectly}$$

$$\mathbf{v}_i(t+1) = \mathbf{v}_i(t) \text{ for } i \neq I,$$

여기서 $r(d)$ 는 데이터의 위치와 데이터와 클러스터들의 대표값들 사이의 거리를 고려한 것이다. $r(d)$ 는

$$r(d) = \frac{\| \mathbf{x} - \mathbf{v}_I(t) \|}{\sum_{i=1}^n \| \mathbf{x} - \mathbf{v}_i(t) \| + \lambda \cdot d_{middle}}$$

이다. d_{middle} 는 선정된 출력뉴런의 연결강도와 다음으로 가까운 출력뉴런의 연결강도의 평균값이다. λ 는 선정된 출력뉴런의 연결강도를 조정할 때 결정경계선 근처에 있는 데이터가 미치는 영향도를 조절하는 상수이다.

위의 새로운 학습법칙은 클러스터의 대표값을 조정할 때, 결정경계선 근처에 있는 데이터를 반영하는 양이 outlier를 반영하는 양보다 크게 할 수 있다. 이는 결정경계선 근처에 있는 데이터는 최적의 결정경계선에 대한 정보를 가지고 있는데 비하여, 한쪽 클러스터에 치우쳐있는 outlier는 비록 퍼지 소속도는 크나 최적의 결정경계선에 대한 정보는 별로 가지고 있지 않다는 점을 고려한 것이다. Outlier는 결정경계선을 적절한 위치로부터 outlier쪽으로 끌어당기는 악영향을 미칠 수 있다. 이를 방지하기 위하여 새로운 퍼지 학습법칙을 제안하였다. d_{middle} 은 클러스터의 대표값을 조정할 때 결정경계선 근처에 있는 데이터의 영향도를 조절하기 위한 것으로, 그 영향도는 λ 로써 조절이 된다. 분류가 틀렸을 때는 데이터의 잘못 선정된 클러스터에 소속되어 있는 정도를 나타내는 퍼지 소속도를 사용하여 클러스터의 대표값을 조정한다.

알고리즘을 요약하면 아래와 같다.

- (1) 파라미터들(t, k)과 입력 뉴런들과 출력 뉴런들을 초기화 한다.
- (2) 입력 벡터 \mathbf{x} 를 가한다.

- (3) 입력 벡터가 첫 번째 입력 벡터이면 (4)로 가고; 그렇지 않으면 (5)로 간다.
- (4) 입력 벡터를 첫 번째 출력 뉴런의 연결 강도로 하고 (2)로 간다.
- (5) 출력 뉴런 중 승자를 찾는다.
- (6) 승자인 출력 뉴런이 vigilance 테스트를 만족하면 (7)로 가고; 그렇지 않으면 (8)로 간다.
- (7) 승자인 출력 뉴런의 연결강도를 조정하고 (2)로 간다.
- (8) 첫 번째 uncommitted 출력 뉴런을 활성화하고 입력벡터를 첫 번째 uncommitted 출력 뉴런의 연결강도로 한다. (2)로 간다.

3. 실험 및 결과

제안된 퍼지 학습법칙을 사용한 supervised IAFC 퍼지 신경회로망 5의 성능과 LVQ 알고리즘과 오류 역전파 신경회로망의 성능을 비교하기 위하여 클러스터링 알고리즘의 성능을 비교하는데 많이 쓰이는 iris 데이터를 사용하였다. iris 데이터는 150개의 4차원 데이터로 구성되어 있다. 이 데이터는 3개의 subspecies로부터 추출한 데이터이고, 각 subspecies는 50개의 데이터를 가지고 있다. 150개의 데이터 중 임의로 75개의 데이터를 선정해 훈련 데이터로 사용하였는데, 각 subspecies로부터 25개의 데이터를 사용하였다.

반복학습을 하는 중에 현재의 연결강도와 이전의 연결강도의 차이인 $\|v(t) - v(t-1)\|$ 이 0.01보다 작으면 연결강도들이 실험적으로 수렴한 것으로 보고 supervised IAFC 퍼지 신경 회로망 5의 훈련을 끝내도록 하였다. 훈련을 끝낸 후 supervised IAFC 퍼지 신경회로망 5를 테스트하였다. supervised IAFC 퍼지 신경회로망 5는 4개의 오류를 발생시켰으며 오류 역전파 신경회로망은 5개의 오류를 발생시켰고, LVQ 알고리즘은 7개의 오류를 발생시켰다. 그림1은 supervised IAFC 퍼지 신경회로망5와 오류 역전파 신경회로망과 LVQ 알고리즘의 성능을 비교한 것이다.

		Output		
		1	2	3
Desired	1	18	7	
	2		25	
	3			25

(a) LVQ 알고리즘

		Output		
		1	2	3
Desired	1	20	5	
	2		25	
	3			25

(b) 오류 역전파 신경회로망

		Output		
		1	2	3
Desired	1	23	2	
	2	2	23	
	3			25

(c) Supervised IAFC 퍼지 신경회로망 5

그림1. iris 데이터를 사용한 결과 비교

그림 2는 T가 2.5이고 λ 가 1.0일 때 supervised IAFC 퍼지 신경회로망 5를 훈련시키기 위하여 30번의 반복학습을 시켰으며 2개의 오류가 발생했음을 보여준다.

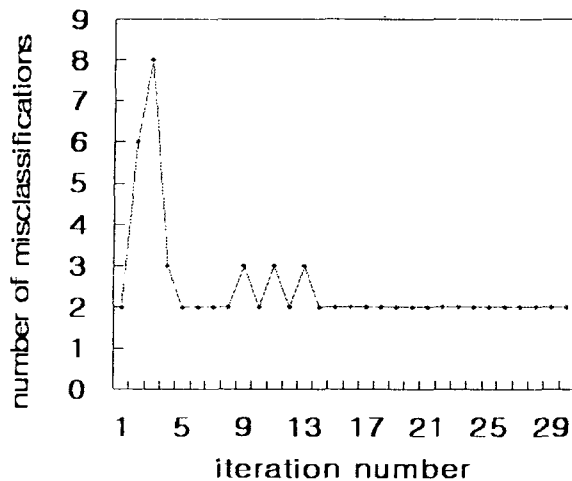


그림2. T가 2.5이고 λ 가 1.0일 때 supervised IAFC 퍼지 신경회로망 5의 훈련결과

4. 결론

본 논문에서 데이터의 위치와 데이터와 클러스터의 대표값 간의 위치를 고려한 새로운 퍼지 학습법칙을 제안하였고, 이 새로운 퍼지 학습법칙을 supervised IAFC 퍼지 신경회로망 5에 적용하였다. Iris 데이터를 사용하여 실험한 결과 supervised IAFC 퍼지 신경회로망 5의 성능이 오류 역전과 알고리즘과 LVQ 알고리즘보다 성능이 우수함을 보여주었다.

참 고 문 헌

- [1] S. Haykin, "Neural Networks-A Comprehensive Foundation 2nd Ed," Prentice Hall, Upper Saddle River, NJ, 1999.
- [2] T. Kohonen, "Learning Vector Quantization for Pattern Recognition," Technical Report TKK-F-A601, Helsinki University of Technology, Finland, 1986.
- [3] J. C. Bezdek, E. C. Tsao, and N. R. Pal, "Fuzzy Kohonen Clustering Networks," Proceedings of the First IEEE Conference on Fuzzy Systems, San Diego pp. 1035-1043, 1992.
- [4] F. -L. Chung and T. Lee, "A Fuzzy Learning Model for Membership Function Estimation and Pattern Classification," Proceedings of the IEEE Conference on Fuzzy Systems, Vol. 1, pp. 426-431, 1994.
- [5] F. -L. Chung and T. Lee, "Fuzzy Learning Vector Quantization," Proceedings of 1993 International Joint Conference on Neural Networks, Nagoya, Vol. 3, pp. 2739-2743, 1993.
- [6] N. B. Karayiannis, "Weighted Fuzzy Learning Vector Quantization and Weighted Fuzzy c-Means Algorithms," IEEE International Conference on Neural Networks, Vol. 2, pp. 1044-1049, 1996.
- [7] N. B. Karayiannis and J. C. Bezdek, "An Integrated Approach to Fuzzy Learning Vector Quantization and Fuzzy c-Means," IEEE Transactions on Fuzzy Systems, Vol. 5, pp. 662-629, 1997.
- [8] E. C. -K. Tsao, J. C. Bezdek, and N. R. Pal, "Fuzzy Kohonen Clustering Networks," Pattern Recognition, Vol. 27, No. 5, pp. 757-764, 1994.
- [9] G. A. Carpenter S. Grossberg, "A Massively Parallel Architecture for A Self-Organizing Neural Pattern Recognition Machine," Computer Vision, Graphics, and Image Processing, Vol. 37, pp. 54-115, 1987.