

패턴인식에서 온라인 오류역전파 알고리즘의 학습속도 향상방법

이태승⁰, 황병원
한국항공대학교 항공전자공학과
thestaff@hitel.net⁰, bhwang@mail.hankong.ac.kr

An Improvement of the Online Mode Error Backpropagation Algorithm Learning Speed for Pattern Recognition

Tae-Seung Lee⁰, Byong Won Hwang
Department of Avionics, Hankuk Aviation University

요약

MLP(multilayer perceptron)는 다른 패턴인식 방법에 비해 몇 가지 이점이 있어 다양한 문제영역에서 사용되고 있다. 그러나 MLP의 학습에 일반적으로 사용되는 EBP(error backpropagation) 알고리즘은 학습시간이 비교적 오래 걸린다는 단점이 있으며, 이는 실시간 처리를 요구하는 문제나 대규모 데이터 및 MLP 구조로 인해 학습시간이 상당히 긴 문제에서 제약으로 작용한다. 패턴인식에 사용되는 학습데이터는 풍부한 중복특성을 내포하고 있으므로 패턴마다 MLP의 내부변수를 갱신하는 온라인 계열의 학습방식이 속도의 향상에 상당한 효과가 있다. 일반적인 온라인 EBP 알고리즘에서는 내부 가중치 갱신시 고정된 학습률을 적용한다. 고정 학습률을 적절히 선택함으로써 패턴인식 응용에서 상당한 속도개선을 얻을 수 있지만, 학습률을 고정함으로써 온라인 방식에서 패턴별 갱신의 특성을 완전히 활용하지 못하는 비효율성이 발생한다. 또한, 학습도중 패턴군이 학습된 패턴과 그렇지 못한 패턴으로 나뉘고 이 가운데 학습된 패턴은 학습을 위한 계산에 포함될 필요가 없음에도 불구하고, 기존의 온라인 EBP에서는 예폭에 할당된 모든 패턴을 일률적으로 계산에 포함시킨다. 이 문제에 대해 본 논문에서는 학습이 진행됨에 따라 패턴마다 적절한 학습률을 적용하고 필요한 패턴만을 학습에 반영하는 패턴별 가변학습률 및 학습생략(COIL) 방법을 제안한다. 제안한 COIL의 성능을 입증하기 위해 화자증명파 음성인식을 실험하고 그 결과를 제시한다.

1. 서론

패턴인식을 위한 여러 인식방법 가운데 MLP(multilayer perceptron)는 다음과 같은 이점 때문에 다양한 문제에 응용되고 있다[1][2].

- 논파라메트릭(nonparametric) 방식이기 때문에 문제에서 가정해야 하는 하부확률분포가 필요없다.
- 학습되는 각 모델 사이의 차이를 최대한 구별하는 거부학습능력이 있기 때문에 인식오류 가능성을 최소화한다.
- 학습모델별로 +1, 0(또는 -1)의 학습목표치를 사용할 때 LDA(linear discriminant analysis)와 유사한 특징공간 변환능력을 갖는다.

MLP는 0개 이상의 은닉 뉴런층과 1개의 출력 뉴런층으로 구성되며, 은닉층에서 MLP의 학습 복잡도를 결정하고 출력층에서 학습모델의 개수를 결정한다. MLP에서 학습을 위해 사용하는 오류역전파(error backpropagation, 이하 EBP) 알고리즘은 최대 기울기 감소 방법을 바탕으로 한 것으로, 신경망의 현재출력과 목표출력 사이의 오류를 출력층에서 은닉층으로 역방향으로 전파하면서 가중치 벡터를 조정하는 방법으로 최종적인 목표치를 달성한다[2]. 각 가중치의 현재 값에 대해 목표치에 가장 빠르게 접근할 수 있는 변위는 아래와 같이 계산된다.

$$\frac{\partial e_p}{\partial w_{ij}} = \frac{\partial e_p}{\partial s_i} \frac{\partial s_i}{\partial n_i} \frac{\partial n_i}{\partial w_{ij}} \quad (1)$$

여기서, e_p 는 현재 패턴 p 에 대한 출력뉴런층의 오류측정 함수이고, w_{ij} 는 j 번째 뉴런과 i 번째 뉴런 사이의 연결 가중치, s_i 는 i 번째 뉴런의 동작치, n_i 는 i 번째 뉴런에 대한 가중된 입력의 총합을 나타낸다. 이 식에서 계산된 변화량을 이전의 가중치 벡터에 적용하면 목표치에 더욱 가까운 값을 도출할 수 있으며, 이를 아래의 방법으로 실현한다.

$$w_{ij}(t+1) = w_{ij}(t) - \eta \frac{\partial e(t)}{\partial w_{ij}(t)} \quad (2)$$

여기서, t 는 가중치 벡터의 특정 상태 시각을 나타내며, η 는 적용할 변화량의 비율을 결정하는 학습률이다.

식(2)에서 알 수 있듯이 목표치를 향한 가중치 변화는 학습률 η 에 상당한 영향을 받는다. η 가 크면 학습이 빠르게 이루어지지만 출력치가 목표치 근방에서 진동하면서 수렴하지 못할 가능성이 증가한다. 반면에 η 가 작으면 목표치에 수렴할 가능성은 높지만 수렴하기까지 많은 갱신회수가 필요해지며, 경우에 따라서는 학습이 로컬 미니마(local minima)에 빠져 목표치에 도달하지 못하는 수도 있다. 따라서 가장 적절한 학습속도를 알아내기 위해 η 를 여러 값으로 변경해가며 반복 실험하여 수렴이 보장되는 최대의 수치를 선택하는 것이 일반적이다.

일반적으로 위와 같이 최대의 학습속도를 사용하더라도 MLP의 학습에는 상당한 시간이 소요된다. 이는 최대 기울기 감소 방법이 현재의 가중치에 대한 지역적인 정보만 사용하는 것에서 연유한다. MLP의 학습속도를 개선하려는 시도는 크게 두 방향으로 이루어졌다. 첫 번째 방향은 경험과 실험결과를 활용한 것으로, 출력치가 목표치에서 멀 경우에는 η 를 크게 하고 가까울 경우에는 작게 하는 것이다. 이것은 다시 가중치 벡터 전체에 일괄적으로 영향을 미치는 전역 η 를 변경하는 방법[3]과 각 가중치마다 식(1)의 변화에 따라서 지역 η 를 변경하는 방법[4]으로 나뉜다. 두 번째 방향은 최적화 이론을 활용한 것으로, 가중치에 대한 2차 미분정보를 사용한다. 이러한 부류로는 모멘텀(momentum)을 사용하여 이전의 학습추세를 현재 갱신에 반영하거나[3], Newton의 최적화 이론[5] 또는 이를 변형한 알고리즘[6][7]을 이용하여 목표치로 가장 빠르게 수렴할 수 있는 가중치 벡터 갱신치를 계산하는 방법이 있다.

식(2)의 가중치 갱신은 두 가지 방식으로 이루어진다. 하나는 모든 학습데이터를 제시한 후 그에 따른 변경치들의 평균을 적용하는 방법이고, 다른 하나는 학습데이터를 하나씩 제시할 때마다 변경치를 적용하는 방법이다. 전자를 오프라인(또는 일괄적) 방식이라고 하고, 후자를 온라인(또는 확률적) 방식이라고 부른다. 두 방식 모두 모든 학습데이터가 제시되는 한 주기를 에폭(epoch)이라 하고, 에폭마다 MLP 목표치와 출력치 사이의 차이를 검사하여 학습의 속행여부를 결정한다.

패턴인식에서는 문제에서 설정된 모델마다 다수의 패턴이 포함되고 학습은 각 모델을 구별할 수 있도록 진행된다. 일반적으로 모델 내 패턴 사이의 차이는 모델 간 패턴 사이의 차이보다 작다. 패턴인식에서 MLP를 사용할 경우 오프라인 학습보다 온라인 학습이 빠른 속도로 이루어지는데, 그 원인으로 아래와 같은 이유를 찾을 수 있다[8].

- 모델 내의 모든 패턴이 서로에 대해 상당한 중복성을 내포하므로 모든 패턴이 식(1)의 최대 기울기 계산에 기여한다. 이 때문에 모델에 포함된 패턴수가 많을수록 에폭 단위의 학습속도가 빨라진다.
- 식(1)에 의해 계산되는 모델 내 패턴의 최대 기울기가 90°이내일 경우 오류를 최소화하는 방향으로 학습이 진행된다.
- 로컬 미니마에 빠질 가능성을 크게 줄인다. 이러한 특성은 모델 내의 모든 패턴마다 가중치 벡터의 갱신이 이루어질 때 중심위치에서 상대적으로 멀리 떨어진 패턴에 의해 전체진행방향과 다른 임의적 진동이 발생하기 때문이다.

이처럼 온라인 학습방식이 오프라인 방식에 비해 상당한 이점을 내포하고 있지만, 그럼에도 불구하고 여전히 개선할 수 있는 여지가 남아있다. 기존의 온라인 방식에서는 고정 학습률을 사용했는데, 이것은 온라인 방식에서 모델 내의 패턴이 학습이 진행됨에 따라 학습에 기여하는 정도가 달라지는 점을 활용하지 못하는 것이다. EBP 알고리즘으로 학습되는 MLP는 각 모델의 중심위치를 시작으로 가장자리 방향으로 모델영역을 학습해 나간다. 이 때 가장자리 쪽으로 진행해 갈수록 학습에 기여하는 패턴의 부분이 줄어들고 이에 알맞은 학습률이 달라질 뿐 아니라, 학습에 기여하지 않는 패턴이 계속해서 계산과정에 포함되는 비효율성이 발생한다. 고정 학습률을 사용하는 기존의 온라인 학습방법은 이러한 변화를 수용하지 못하므로 최적의 속도로 학습목표를 달성하지 못한다.

이러한 사실을 기반으로 하여 본 논문에서는 온라인 학습방식에서 고정 학습률의 단점을 개선하여 패턴마다 개별적으로 적용되는 가변 학습률을 도입하고 학습의 진행단계에 따라 학습에 기여하지 못하는 패턴의 학습계산을 생략하는 방법을 제안한다.

II. 문제제기

패턴인식을 위한 EBP 학습 방법에서 각 패턴의 모델구별학습은 학습이 진행됨에 따라 변화를 일으킨다. 즉, 학습의 초기단계에서는 모든 패턴이 모델간 구별학습에 참여하지만, 학습이 진행됨에 따라 각 모델의 중심부근에 위치한 패턴들은 학습에 참여하지 않는 현상이 발생한다. 이에 따라 빠른 학습을 위해서는 (1) 전체 패턴이 학습에 참여하는 초기 단계에서는 패턴에 큰 학습률을 적용하고, (2) 모델의 정밀한 외형이 학습됨에 따라 학습의 진동을 막기 위해 작은 학습률을 적용하며, (3) 이미 학습된 패턴은 학습에 기여하는 바 없이 비효율적인 학습을 발생시키므로 학습계산에서 제외해야 한다.

기존의 온라인 EBP 학습에서는 학습전에 고정된 학습률을 설정하므로 이러한 학습률의 변동가능성에 대응하지 못할 뿐 아니라, 학습이 진행됨에 따라 학습에 기여하지 못하는 패턴이 불필요한 학습계산을 일으킨다. 이러한 점은 온라인 방식 EBP 알고리즘이 보다 빠른 학습을 달성하지 못하는 원인이 된다.

III. 제안방법

본 논문에서는 온라인 학습의 위와 같은 문제를 해결하기 위해 두 가지 방법을 제안한다.

(i) 학습패턴별 학습률 가변방법

학습률은 모델의 학습진행에 따라 큰 값에서 작은 값으로의 변화가 필요하다.

식(1)의 e_p 는 다음과 같이 정의되며,

$$e_p(t) = \frac{1}{2} \sum_{\sigma=1}^N e_{\sigma,p}^2(t) \quad (3)$$

여기서, $e_{\sigma,p}$ 는 현재 패턴에 대한 출력뉴런의 오류함수이고, N 은 출력뉴런의 개수이다.

식(3)에서 현재패턴의 모델과 대응하는 출력뉴런의 $e_{\sigma,p}^2(t)$ 는 현재패턴의 학습상태를 알 수 있는 수치적 측정수단을 제공한다. 즉, 패턴이 충분히 학습되지 않았을 경우 큰 값을 나타내고, 충분히 학습되었을 경우 작은 값을 나타낸다. 따라서, 초기학습에서는 큰 값이, 그 뒤 학습이 완결되어감에 따라 작은 값이 구해진다.

온라인 학습방식에서 보편적으로 선택되는 학습률의 범위는 1~0.0001이다[9]. 이 범위에서 상한을 넘어서면 내부변수의 값이 발산하기 쉽고, 하한을 넘어서면 학습이 불필요하게 길어진다. 그러나 문제에 따라 적절한 값의 범위가 다르므로 이 범위를 실험을 통해 결정해야 한다. 식(3)의 $e_{\sigma,p}^2(t)$ 를 학습률로 사용하면 하한에 대해서는 걱정할 필요가 없으므로 적절한 상한을 알아내지만 하면 된다. 그런 다음 0에서부터 이렇게 알아낸 상한까지 $e_{\sigma,p}^2(t)$ 의 범위를 제한한다.

$$y(x) = \frac{2 \cdot V_{UL}}{1 + e^{-2x}} - V_{UL} \quad (4)$$

여기서, V_{UL} 은 실험을 통해 알아내는 학습률의 상한이고, x 는 $e_{\sigma,p}^2(t)$ 의 출력치이다.

그러나 식(4)은 현재패턴에 한정된 오류정보만을 사용한다. 만약 학습모델 자체가 필연적인 오류를 크게 내포하는 경우, 즉 여러 모델 사이에 모델 분포영역이 겹치는 부분이 있을 경우에는 그 부분의 패턴에 의한 $e_{\sigma,p}^2(t)$ 가 현재 학습상황과 대체적으로 무관하게 큰 값을 나타냄으로써 전체 학습을 방해하는 결과를 초래할 수 있다. 이런 경우에 대처하기 위해 식(4)의 값을 한 단계 이전 에폭의 평균오류로 제한한다.

$$y'(x) = \begin{cases} V_{ASEE} \cdot V_{UL} & \text{if } y(x) \geq V_{ASEE} \cdot V_{LL} \\ y(x) & \text{otherwise} \end{cases} \quad (5)$$

여기서, V_{ASEE} 는 아래와 같이 정의되는 평균오류제곱에너지를 나타낸다.

$$V_{ASEE} = \frac{1}{NM} \sum_{p=1}^M e_p(t-1) \quad (6)$$

여기서, M 은 학습에폭에 소요되는 패턴의 개수이다.

(ii) 패턴학습 생략방법

온라인 EBP에서 이뤄지는 주요 계산은 패턴의 오류계산, 오류 역전파, 가중치 갱신이다. 현재 학습단계에서 현재 패턴의 기여도는 $e_{o,p}^2(t)$ 를 통해 알 수 있으므로, $e_{o,p}^2(t)$ 의 값이 학습 전에 설정하는 식(6)의 최종목표 오류치보다 작은 경우 현재 패턴이 학습에 기여하는 바가 적다고 판단하여 오류 역전파와 내부변수 갱신 계산과정을 생략할 수 있다. 만일 이후 다른 패턴의 학습으로 현재 패턴의 기여도가 높아진다면 이 상태는 $e_{o,p}^2(t)$ 에 의해 발견되기 때문에 다시 학습에 참여할 수 있게 된다.

본 논문에서는 이 두 방법을 패턴별 가변학습률 및 학습생략 방법(Changing rate and Omitting patterns in Instant Learning)이라고 부른다.

IV. 실험

COIL의 성능을 증명하기 위해 화자증명[9]과 음성인식[10]의 MLP 학습속도를 측정한다.

먼저, 음성데이터를 이용한 화자증명(speaker verification)에서 화자를 증명시스템에 등록하고 증명시험을 수행하는 인식기로서 MLP를 사용할 때 화자를 등록하는 MLP 학습에서 여러 방법의 속도차이를 비교하여 중소규모 데이터를 사용하는 패턴인식을 실험한다. 이 실험에 사용된 데이터는 한국인 남녀 40명의 4연숫자 발성을 녹음한 것으로, 각 화자가 총 35개의 서로 다른 숫자음 배열을 4회씩 발성한다. 4회 발성 중 3회를 각 화자의 등록음성으로 사용하고 1회를 증명시험 음성으로 사용한다. 각 등록음성에서 한국어 지속음(/a/, /e/, /ə/, /o/, /u/, /i/, /l/, /r/, 비음)을 추출하여 지속음마다 MLP를 학습시키며, 각 MLP에 주어지는 학습데이터의 수는 에폭 당 580개이다. 실험 중 수행된 MLP 학습회수는 대략 9,600회이며 그 평균을 기록한다. 사용된 MLP의 구조는 입력뉴런 50, 은닉뉴런 2, 출력뉴런 1의 2층 구조이다.

그 다음으로, 7개의 한국어 단모음(/a/, /e/, /ə/, /o/, /u/, /i/, /l/)을 화자독립(speaker independent)으로 인식하기 위한 인식기로서 MLP를 사용할 때 MLP 학습에서 여러 방법의 속도차이를 비교하여 대규모 데이터를 사용하는 패턴인식을 실험한다. 이 실험에 사용된 데이터는 한국인 남녀 70명의 PBW(phone balanced words) 발성에서 추출된 단모음이다. 이 실험에서는 1개의 MLP를 5회 학습시키며 그 평균을 기록한다. 학습데이터의 수는 에폭 당 20,748개이고, 사용된 MLP의 구조는 입력뉴런 16, 은닉뉴런 20, 출력뉴런 7의 2층 구조이다.

실험결과를 오프라인 EBP, RPROP[3], 온라인 EBP, CIL, COIL별로 기록한다. 여기서 오프라인 EBP를 OffEBP로, 온라인 EBP를 OnEBP로 표기하고, 제안방법 (i)만 적용한 것을 CIL, (i) 및 (ii)를 모두 적용한 것을 COIL이라고 표기한다. 기록수치는 에폭수, 학습시간, 인식률을 표시한다.

Table 1: 화자증명 실험결과

학습방법	OffEBP	RPROP	OnEBP	CIL	COIL
에폭수	2554.49	1254.74	489.40	266.60	248.61
학습시간 (초)	4.45	2.12	1.47	0.81	0.48
인식률 (%)	98.06	97.66	98.16	98.34	98.35

Table 2: 음성인식 실험결과

학습방법	OffEBP	RPROP	OnEBP	CIL	COIL
에폭수	157.60	79.10	14.00	8.30	8.50
학습시간 (초)	77.47	37.98	11.03	7.02	6.02
인식률 (%)	68.26	68.42	69.05	69.03	69.04

V. 결론

패턴인식에 이용되는 MLP의 온라인 방식 EBP 학습 알고리즘은 패턴인식 데이터 내의 풍부한 중복성으로 인해 오프라인 방식보다 빠른 학습을 달성한다. 그러나, 학습이 진행됨에 따라 각 모델에 속한 패턴들의 학습에 대한 기여도가 달라지는 특성을 활용하면 속도를 더욱 향상시킬 수 있다. 본 논문에서는 기존의 고정 학습률 대신 패턴마다 학습률을 다르게 적용하고 학습에 기여하지 못하는 패턴의 학습계산을 생략하는 COIL 방법을 제안하여 온라인 방식 EBP 알고리즘의 학습속도 개선을 시도하였다. 제안한 방법을 화자증명과 음성인식에 적용하여 실험한 결과 기존 방법보다 학습이 더 빠르게 달성됨을 확인하였다.

References

- [1] N. Morgan and H. Bourlard, "Hybrid Connectionist Models for Continuous Speech Recognition," Automatic Speech and Speaker Recognition, Kluwer Academic Publishers, 1996.
- [2] S. Haykin, Neural Networks, Prentice Hall, 1999.
- [3] M. Riedmiller and H. Braun, "A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm," CNN, Vol. 1, pp. 586-591, 1993.
- [4] H. Demuth and M. Beale, Neural Network Toolbox, The MathWorks, 2001.
- [5] R. Fletcher, Practical Methods of Optimization, Wiley, 1987.
- [6] M. Moller, "Supervised Learning on Large Redundant Training Sets," Proceedings of the 1992 IEEE-SP Workshop Neural Networks for Signal Processing, pp. 79-89, 1992.
- [7] S. Becker and Y. LeCun, "Improving the Convergence of Back-Propagation Learning with Second-Order Methods," Proceedings of the 1988 Connectionist Models Summer School, pp. 29-37, 1988.
- [8] D. R. Wilson and T. R. Martinez, "The Need for Small Learning Rates on Large Problems," IJCNN, Vol. 1, pp. 115-119, 2001.
- [9] J. M. Naik, "Speaker Verification: A Tutorial," IEEE Communications Magazine, Vol. 1, pp. 42-48, 1990.
- [10] C. Becchetti and L. P. Ricotti, Speech Recognition, John Wiley & Sons, 1999.