

음성처리에서 온라인 오류역전파 알고리즘의 학습속도 향상방법

A Method on the Learning Speed Improvement of the Online Error Backpropagation Algorithm in Speech Processing

이 태 승*, 이 백 영*, 황 병 원*

(Lee Tae-Seung*, Rhee Baeg-Yeong*, Hwang Byong-Won*)

* 한국항공대학교 항공전자공학과

(접수일자: 2002년 2월 22일; 수정일자: 2002년 5월 22일; 채택일자: 2002년 6월 29일)

다층신경망 (MLP: multilayer perceptron)은 다른 패턴인식 방법에 비해 여러 가지 훌륭한 특성을 가지고 있어 음성인식 및 화자인식 영역에서 폭넓게 사용되고 있다. 그러나 다층신경망의 학습에 일반적으로 사용되는 오류역전파 (EBP: error backpropagation) 알고리즘은 학습시간이 비교적 오래 걸린다는 단점이 있으며, 이는 화자인식이나 화자적응과 같이 실시간 처리를 요구하는 응용에서 상당한 제약으로 작용한다. 패턴인식에 사용되는 학습데이터는 풍부한 중복특성을 내포하고 있으므로 패턴마다 다층신경망의 내부변수를 갱신하는 온라인 계열의 학습방식이 속도의 향상에 상당한 효과가 있다. 일반적인 온라인 오류역전파 알고리즘에서는 가중치 갱신시 고정된 학습률을 적용한다. 고정 학습률을 적절히 선택함으로써 패턴인식 응용에서 상당한 속도개선을 얻을 수 있지만, 학습률이 고정된 상태에서는 학습이 진행됨에 따라 학습에 기여하는 패턴영역이 달라지는 현상에 효과적으로 대응하지 못하는 문제가 있다. 이 문제에 대해 본 논문에서는 패턴의 기여도에 따라 가변하는 학습률과 학습에 기여하는 패턴만을 학습에 반영하는 패턴별 가변학습률 및 학습생략 (COIL: Changing rate and Omitting patterns in Instant Learning) 방법을 제안한다. 제안한 COIL의 성능을 입증하기 위해 화자증명과 음성인식을 실험하고 그 결과를 제시한다.

핵심용어: 다층신경망, 오류역전파 알고리즘, 고속 학습 알고리즘, 화자증명, 음성인식, 패턴인식

투고분야: 음향 신호처리 분야 (1.1, 1.7), 음성처리 분야 (2.5)

Having a variety of good characteristics against other pattern recognition techniques, the multilayer perceptron (MLP) has been widely used in speech recognition and speaker recognition. But, it is known that the error backpropagation (EBP) algorithm that MLP uses in learning has the defect that requires relatively long learning time, and it restricts severely the applications like speaker recognition and speaker adaptation requiring real time processing. Because the learning data for pattern recognition contain high redundancy, in order to increase the learning speed it is very effective to use the online-based learning methods, which update the weight vector of the MLP by the pattern. A typical online EBP algorithm applies the fixed learning rate for each update of the weight vector. Though a large amount of speedup with the online EBP can be obtained by choosing the appropriate fixed rate, fixing the rate leads to the problem that the algorithm cannot respond effectively to different learning phases as the phases change and the number of patterns contributing to learning decreases. To solve this problem, this paper proposes a Changing rate and Omitting patterns in Instant Learning (COIL) method to apply the variable rate and

the only patterns necessary to the learning phase when the phases come to change. In this paper, experimentations are conducted for speaker verification and speech recognition, and results are presented to verify the performance of the COIL.

Keywords: Multilayer perceptron, Error backpropagation, Fast learning algorithm, Speaker verification, Speech recognition, Pattern recognition

ASK subject classification: Acoustic signal processing (1.1, 1.7), Speech signal processing (2.5)

I. 서론

다층신경망 (MLP: multilayer perceptron)은 Rumelhart 등[1]이 학습 알고리즘으로서 오류역전파 (EBP: error backpropagation)를 제안한 이후 되살아난 관심에 힘입어 90년대 들어 음성인식 분야에서도 괄목할만한 성장을 이루었다. 이러한 다층신경망을 음성처리에 활용하려는 시도로서, 패턴 간의 시간적 차이 정합방법으로 널리 사용되었던 DTW (dynamic time warping)[2]와 결합하여 고립단어음을 인식하려 한 시도[3,4]와 함께, HMM (hidden Markov model)[5]의 발생확률 예측기로서 다층신경망을 사용하거나[6], 패턴의 시간지연을 수용할 수 있는 구조의 다층신경망을 통해 연속음을 인식하려는 시도[7]가 있었다. 또한 화자의 특성을 음성을 통해 구분 짓는 화자증명 및 식별 응용에도 다층신경망이 사용되어 좋은 성과를 거둔 바 있다[8,9].

음성처리를 위한 여러 인식방법 가운데 다층신경망은 다음과 같은 이점 때문에 다양한 문제에 응용되고 있다[10,11].

- 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.

다층신경망은 0개 이상의 은닉 뉴런층과 1개의 출력 뉴런층으로 구성되며, 은닉층에서 다층신경망의 학습 복잡도를 결정하고 출력층에서 학습모델의 개수를 결정한다. 오류역전파 알고리즘은 최대 기울기 감소 방법을 바탕으로 한 것으로, 다층신경망의 현재출력과 목표출력 사이의 오류를 출력층에서 은닉층으로 역방향으로 전파하면서 가중치를 조정하는 방법으로 최종적인 목표치를 달성한다. 각 가중치의 현재 값에 대해 목표치에 가장 빠

르게 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.

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

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 보편적이다.

일반적으로 위와 같이 최대의 학습속도를 사용하더라도 다층신경망의 학습에는 상당한 시간이 소요된다. 이는 최대 기울기 감소 방법이 현재의 가중치에 대한 지역적인 정보만 사용하는 것에서 연유한다. 음성처리 분야 중에서도 화자증명 및 식별 분야는 화자의 등록을 실시간으로 처리해야 하는 특성을 가지고 있으므로 다층신경망의 느린 학습속도는 충분히 문제가 될 수 있다. 또한 화자 독립방식 음성인식의 성능향상을 위해 특정 화자에 인식

시스템을 적용시키는 화자적응 응용[12]에서도 실시간 처리가 요구되므로 음성인식 시스템이 다층신경망을 채택하고 있다면 이 경우에서도 다층신경망의 학습속도는 해결해야 할 과제이다.

다층신경망의 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[13]과 각 가중치마다 식 (1)의 변화에 따라서 지역 η 를 변경하는 방법[14]으로 나뉜다. 두 번째 방향은 최적화 이론을 활용한 것으로, 가중치에 대한 2차 미분 정보를 사용한다. 이러한 부류로는 모멘텀 (momentum)을 사용하여 이전의 학습추세를 현재 갱신에 반영하거나 [13], Newton의 최적화 이론[15] 또는 이를 변형한 알고리즘[6,17]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 후에 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[11]. 두 방식 모두 고든 학습데이터가 제시되는 한 주기를 에폭 (epoch) 이라 하고, 에폭마다 다층신경망 목표치와 출력치 사이의 차이를 검사하여 학습의 속행여부를 결정한다.

음성인식 및 화자인식을 비롯한 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 다층신경망을 사용할 경우 오프라인 학습보다 온라인 학습이 빠른 속도로 이루어지는데 그 원인으로 아래와 같은 이유를 찾을 수 있다[18].

- 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 전체 진행방향과 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 오류역전파에서는 고정 학습률을 사용했는데, 이것은 온라인 방식에서 학습이 진행됨에 따라 모델 내에서 학습에 기여하는 패턴영역이 변화하는 점을 활용하지 못하는 것이다. 패턴인식에서 오류역전파 알고리즘으로 학습되는 다층신경망은 3단계의 학습단계를 밟는다. 즉, (1)모델의 중심위치를 학습하고, (2)모델의 분산을 학습한 다음, (3)모델분포의 윤곽을 학습한다. 후자 쪽으로 진행해 갈수록 학습에 기여하는 패턴영역이 외곽으로 한정되고 패턴마다 효율적인 학습률이 달라질 뿐 아니라, 학습에 기여하지 않는 패턴이 계속해서 계산과정에 포함되는 비효율성이 발생한다. 고정 학습률을 사용하는 기존의 온라인 오류역전파 학습방법은 이러한 변화를 수용하지 못하므로 최적의 속도로 목표를 달성하지 못한다.

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

II. 척력개념에 의한 오류역전파 학습해석

패턴인식을 위한 오류역전파 학습에서 각 모델은 식 (2)에 의해 반복적인 방식으로 학습되므로 다음과 같은 과정을 거친다.

- (1) 중심위치학습
- (2) 영역분산학습
- (3) 영역윤곽학습

2모델 분류문제에서 이 과정을 그림 1에서 보여준다. 이 그림의 (b)~(d)에서 검은색 띠는 두 모델의 결정 경계선을 나타내며 색이 짙을수록 경계가 뚜렷함을 의미한다. 여기서 경계선이 (b)에서는 두 모델의 중심을 가로지르는 모습을 보여주고, (c)에서는 두 모델의 영역분산정도를 구분하고 있으며, (d)에서는 세부적인 윤곽을 표현하고 있음을 알 수 있다.

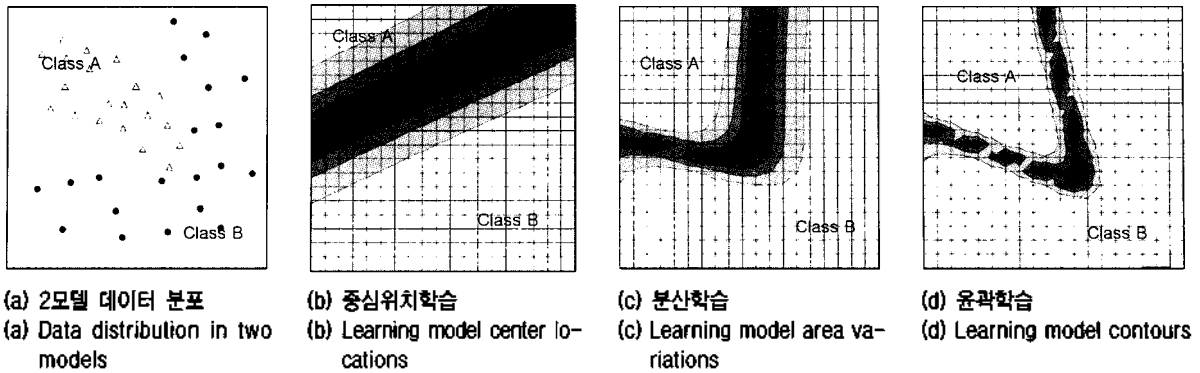


그림 1. 오류역전파 학습의 3단계
Fig. 1. The three phases in EBP learning.

위와 같은 학습영역의 변화를 척력개념을 통해 해석해 볼 수 있다.

식 (1)의 e_p 는 아래와 같이 표현된다.

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

여기서 $e_{o,p}$ 는 현재 패턴 p 에 대한 개별 출력뉴런 o 의 오류치이고, N 은 출력뉴런의 개수이다. $e_{o,p}$ 는 다시 다음과 같이 표현된다.

$$e_{o,p}(t) = d_{o,p}(t) - y_{o,p}(t) \quad (4)$$

여기서 $d_{o,p}$ 는 학습 목표치이고, $y_{o,p}$ 는 현재 출력치이다.

그림 1의 각 단계에서 식 (4)에 의해 각 패턴이 나타내는 오류치의 절댓치는 결정 경계선을 움직이는 척력으로 생각할 수 있으며, 이는 곧 각 단계에서 개별 패턴이 학습에 기여하는 정도로 간주할 수 있다. 그림 2에서 그림 1을 패턴별 척력 개념으로 묘사한 모습을 보여준다. 이 그림에서 점선은 결정 경계선을 나타내고, 화살표는 패턴별 척력을 나타낸다. (a)에 해당하는 미학습 상태에서는 모

든 패턴이 강한 척력을 나타내어 중심위치를 학습할 수 있게 하며, (b) 이후부터는 결정 경계선 인근과 경계선을 넘어 상대방 모델 영역에 위치한 패턴이 강한 척력을 발생시켜 경계선의 형상과 위치를 변화시킨다.

패턴인식에서 온라인 오류역전파가 오프라인 오류역전파에 비해 빠른 학습을 달성하는 원인을 패턴별 척력의 적용방식에서 찾을 수 있다. 오프라인 오류역전파에서는 모든 패턴별 척력벡터의 합벡터를 적용하지만, 온라인 오류역전파에서는 이들 척력벡터를 개별적으로 적용한다. 따라서 온라인 오류역전파가 결정 경계선의 모양을 변화시킬 기회를 더 많이 갖게 되며, 모델의 영역이 복잡한 형상을 띠는 경우 그 효과는 더욱 크다.

III. 온라인 오류역전파의 학습속도개선 가능성

온라인 오류역전파에서 학습단계에 따른 패턴의 척력 분포를 보다 적극적으로 활용하면 학습시간을 더욱 단축할 수 있다.

그림 2에서 모델의 결정 경계선이 아직 학습되지 않

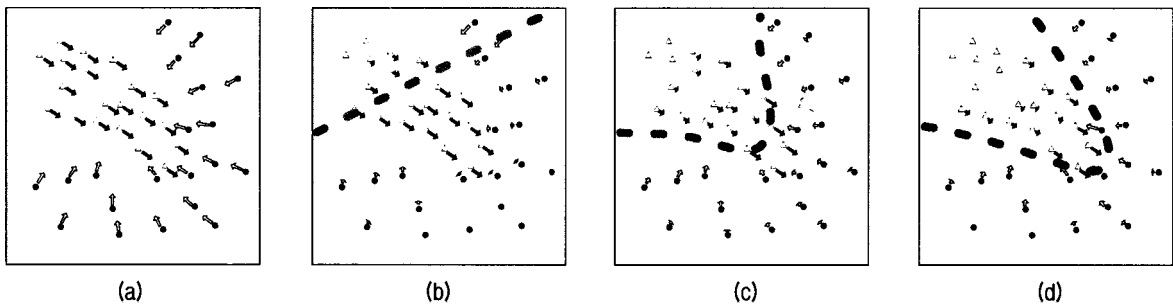


그림 2. 오류역전파 학습의 패턴별 척력
Fig. 2. The repulsive force of each pattern in EBP learning.

은 패턴의 큰 척력에 의해 변형되는 것을 볼 수 있다. 이 척력이 결정 경계선에 미치는 영향력은 식 (2)의 학습률 η 에 의해 조정된다. 즉, 학습이 덜된 패턴에 대해 η 가 클수록 그 패턴의 인근 결정 경계선은 급격히 이동하게 된다. 그러나 학습이 진행되면서 결정 경계선이 학습목표 주변에 도달하면, 그 부근의 패턴에 대한 η 가 점차 작아져야만 결정 경계선이 학습목표에 수렴할 가능성이 높아진다. 또한 여러 모델 사이의 중첩된 영역은 완벽한 학습이 불가능하므로, 결정 경계선의 불필요한 진동을 막기 위해 이 영역의 패턴에 대해서는 η 가 작아야 한다.

이를 위해 기존의 온라인 오류역전파 알고리즘에서와 같이 η 를 학습기간 동안 고정하는 대신, η 가 상황에 맞춰 변경될 수 있어야 한다. 이와 같은 가변 η 는 이미 오프라인 오류역전파에서 시도된 바 있지만[13,19], 온라인 오류역전파에서 이를 적용하면 온라인 방식의 데이터 중복성 활용 이점을 더욱 향상시킬 수 있다.

한편, 기존의 온라인 오류역전파 알고리즘에서는 패턴이 이미 학습되었다라든가 그 패턴이 계속해서 학습계산에 참여한다. 일정한 학습성과의 기준을 설정하고 패턴이 그 기준을 만족하면 학습계산에서 그 패턴을 제외하더라도 전체 학습진행에는 거의 영향을 주지 않으면서 학습시간을 단축할 수 있다.

IV. 제안방법

3절에서 제기한 온라인 방식 오류역전파 알고리즘의 학습속도 향상을 실현하기 위해 본 논문에서는 두 가지 방법을 제안한다.

첫 번째 방법은 학습패턴별로 학습률을 가변하는 방법이다. 학습률은 모델 내 패턴의 학습진행에 따라 큰 값에서 작은 값으로의 변화가 필요하다. 식 (3)에서 현재 패턴의 모델과 대응하는 출력뉴런의 $e_{o,p}^2(t)$ 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 학습 초기에는 큰 값이, 그 뒤 학습이 완결되어감에 따라 작은 값이 구해진다.

온라인 학습방식에서 보편적으로 선택되는 학습률의 범위는 1~0.0001이다[19]. 이 범위에서 상한을 넘어서면 내부변수의 값이 발산하기 쉽고, 하한을 넘어서면 학습이 불필요하게 길어진다. 그러나 문제에 따라 적절한 값

의 범위가 다르므로 이 범위는 실험을 통해 결정해야 한다. 식 (3)의 $e_{o,p}^2(t)$ 를 학습률로 사용하면 하한에 대해서는 걱정할 필요가 없으므로 적절한 상한을 알아내기만 하면 된다. 그런 다음 0에서부터 이렇게 알아낸 상한까지 $e_{o,p}^2(t)$ 의 범위를 아래의 시그모이드 (sigmoid) 형태를 갖는 식으로 제한한다.

$$y'(x) = \frac{2 \cdot V_{UL}}{1 + e^{-2x}} - V_{UL} \quad (5)$$

여기서 V_{UL} 은 실험을 통해 알아내는 학습률의 상한이고, x 는 $e_{o,p}^2(t)$ 의 출력치이다. 식 (5)가 시그모이드 형태를 띠는 이유는 상한까지 학습률을 제한하되 상한과 0 부근에서 급격한 값의 변화를 피하기 위한 것이다.

그러나 식 (5)는 현재 패턴에 한정된 오류정보만을 사용한다. 만약 학습모델 자체가 필연적인 오류를 크게 내포하는 경우, 즉 여러 모델 사이에 분포영역이 중첩되는 부분이 있을 경우에는 그 부분의 패턴에 의한 $e_{o,p}^2(t)$ 가 현재 학습상황과 무관하게 큰 값을 나타냄으로써 전체적인 학습진행을 방해하는 결과를 초래할 수 있다. 이런 경우에 대처하기 위해 식 (5)의 값을 한 단계 이전 에폭의 평균오류로 제한한다.

$$y(x) = \begin{cases} V_{ASEE} \cdot V_{UL} & \text{if } y(x) \geq V_{ASEE} \cdot V_{UL} \\ y'(x) & \text{otherwise} \end{cases} \quad (6)$$

여기서 V_{ASEE} 는 아래와 같이 정의되는 평균오류제곱에 너지[11]를 나타낸다.

$$V_{ASEE} = \frac{1}{NM} \sum_{p=1}^M e_p(t-1) \quad (7)$$

여기서 M 은 1에폭 동안 사용된 학습패턴의 총합이다.

두 번째 방법은 패턴의 학습을 생략하는 방법이다.

온라인 방식 오류역전파에서 이뤄지는 주요 계산은 패턴의 오류계산, 오류 역전파, 가중치 갱신이다. 현재 학습 단계에서 현재 패턴의 기여도는 $e_{o,p}^2(t)$ 를 통해 알 수 있으므로, $e_{o,p}^2(t)$ 의 값이 학습 전에 설정하는 식 (7)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. 이를 수식으로 표현하면 다음과 같다.

$$P(e_{o,p}^2(t)) = \begin{cases} \text{The learning of the current pattern is canceled,} & \text{if } e_{o,p}^2(t) < V_{ASEE, Objective} \\ \text{The learning of the current pattern is continued,} & \text{if } e_{o,p}^2(t) \geq V_{ASEE, Objective} \end{cases} \quad (8)$$

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

여기서 P 는 현재 패턴의 학습처리를 나타내고, ASEE, Objective는 최종목표 오류치를 나타낸다.

V. 실험 및 결과분석

COIL의 성능을 증명하기 위해 화자증명[20]과 음성인

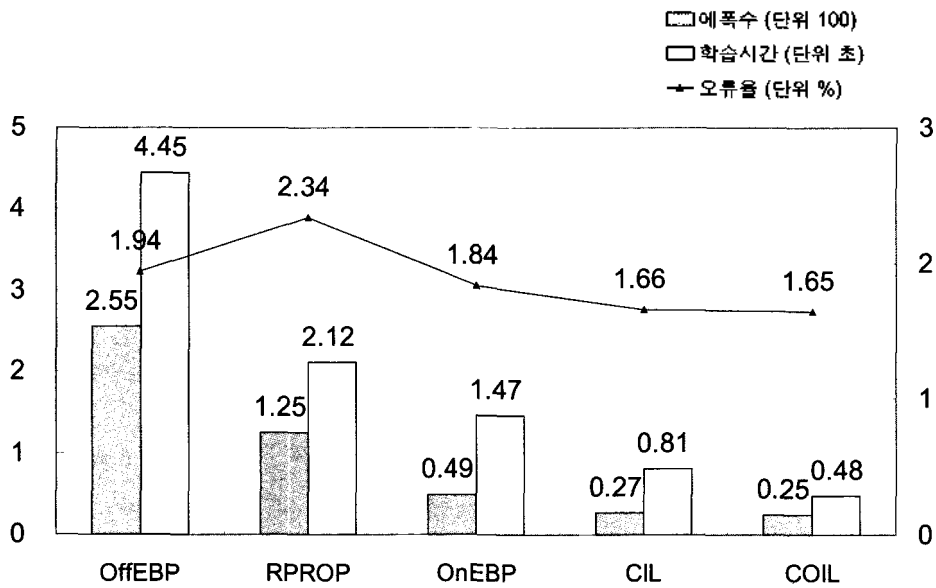


그림 3. 화자증명 실험의 학습방법별 에폭수, 학습시간, 오류율 비교결과

Fig. 3. The comparison of the numbers of epochs, learning duration times, and error rates with each learning method for the speaker verification experiment.

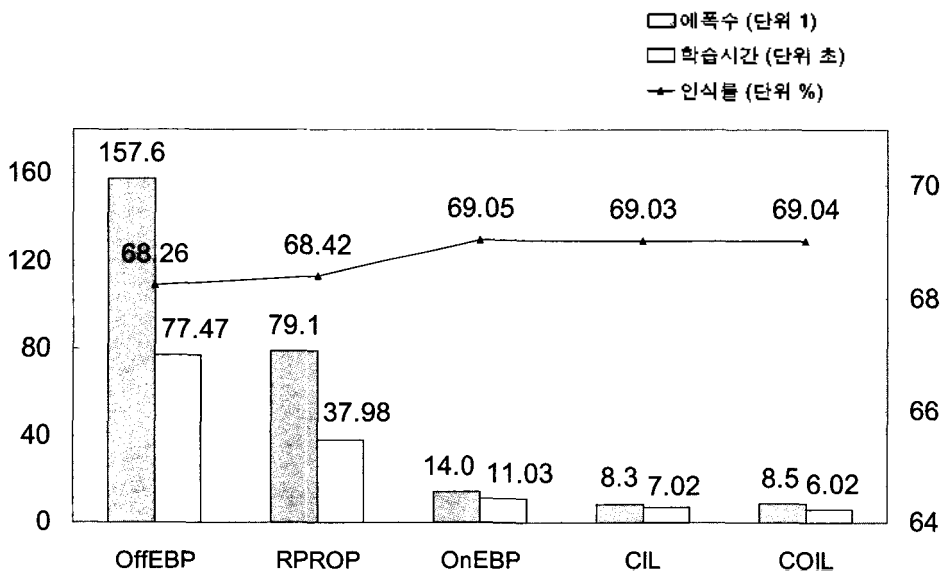


그림 4. 음성인식 실험의 학습방법별 에폭수, 학습시간, 인식률 비교결과

Fig. 4. The comparison of the numbers of epochs, learning duration times, and recognition rates with each learning method for the speech recognition experiment.

식[21]에서의 다층 신경망 학습속도를 측정한다.

먼저, 음성데이터를 이용한 화자증명에서 화자를 증명시스템에 등록하고 증명시험을 수행하는 인식기로서 다층신경망을 사용할 때 화자를 등록하는 다층신경망 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/, /i/, /l/, 비음)을 추출하고 지속음마다 다층신경망을 학습시키며, 이 때 각 다층신경망에 주어지는 학습데이터의 수는 에폭 당 580개이다. 실험 중 수행된 다층신경망 학습 회수는 대략 9,600회이며 그 평균을 기록한다. 사용된 다층신경망의 구조는 입력뉴런 50, 은닉뉴런 2, 출력뉴런의 2층 구조이다.

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

실험결과는 오프라인 오류역전파, RPROP[14], 온라인 오류역전파, CIL, COIL별로 기록한다. 여기서 오프라인 오류역전파를 OffEBP로, 온라인 오류역전파를 OnEBP로 표기하고, 제안방법 중 첫 번째 방법만 적용한 것을 CIL, 두 방법을 모두 적용한 것을 COIL이라고 표기한다. 기록수치는 에폭수, 학습시간, 오류율 또는 인식률을 기록한다.

먼저 화자증명 실험의 결과를 보여주는 그림 3에서 에폭수와 학습시간이 OffEBP, RPROP, OnEBP, CIL, COIL의 순서로 향상됨을 알 수 있다. RPROP는 가장 느린 학습 방법으로 나타난 OffEBP를 향상시키는 오프라인 계열의 학습방법으로서 OffEBP에 비해 2배 가까운 속도향상을 기록한다. OnEBP는 이러한 RPROP보다도 2배 이상의 속도향상을 보여주는데, 이 사실은 패턴인식에서 데이터의 중복성 이용이 얼마나 중요한지 잘 보여주고 있다. CIL은 다시 OnEBP보다 2배에 약간 못 미치는 수준의 향상을

보여줌으로써 데이터의 중복성을 이용한 가변학습률 적용의 실효를 증명한다. COIL은 에폭수 면에서는 CIL에서 별다른 향상을 보이지 않지만 학습시간 면에서는 역시 2배 가까운 향상을 보이고 있다. 이는 패턴학습을 생략하는 추가된 방법의 특성상 학습추이의 양상은 변하지 않지만 각 학습단계마다 학습에 거의 영향을 주지 못하는 패턴이 있음을 증명한다.

이에 비해 음성인식 실험의 결과를 보여주는 그림 4에서는 향상의 추이는 화자인식 실험과 같지만 다른 학습방법들에 비해 CIL과 COIL의 향상폭이 상대적으로 작음을 알 수 있다. 이러한 현상은 화자증명 실험에서 사용된 에폭당 데이터 수에 비해 음성인식에 사용된 데이터 수가 35배 가량 많기 때문인 것으로 추측된다. 데이터 수가 많을수록 중복성을 이용하는 온라인 계열의 속도향상 효과가 커진다. CIL과 COIL은 OnEBP 알고리즘을 개량한 방법이고 이미 OnEBP에서 중복성의 혜택을 크게 활용했으므로 CIL과 COIL에 의해 개선되는 폭이 상대적으로 줄어드는 것으로 추측된다.

마지막으로 그림 3과 4에서 두드러지는 현상 중 각 학습방법마다 인식을 면에서 차이를 보인다는 점을 발견할 수 있다. 전반적으로 오프라인에 비해 온라인 계열의 학습방법이 향상된 인식률을 기록하고 있다. 이러한 결과의 원인은 오프라인 방법이 모든 패턴의 오류값을 에폭마다 평균하여 그 추세를 적용하고, 온라인 방법은 각 패턴의 차이를 모두 적용하기 때문인 것으로 판단된다. 학습 모델 데이터의 분포가 복잡한 형상을 띠면 평균연산이 이를 뭉툭한 형상으로 만드는 역할을 함으로써 인식오류를 증가시킬 수 있다.

VI. 결론

패턴인식에 이용되는 다층신경망의 온라인 방식 오류역전파 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 오류역전파 알고리즘의 학습속도 개선을 시도하였다. 제안한 방법을 화자증명과 음성인식에 적용하여 실험한 결과 기존의 온라인 오류역전파

알고리즘보다 학습이 2~3배 가량 더 빠르게 달성됨을 확인하였다.

참고 문헌

1. D. E. Rumelhart, G. E. Hinton and R. J. Williams, "Learning Internal representations by error propagation," *Parallel Distributed Processing*, MIT Press, Cambridge, 1, chap. 8, 318-362, 1986.
2. H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26 (1), 43-49, Feb. 1978.
3. H. Sakoe, R. Isotani, K. Yoshida, K. I. Iso and T. Watanabe, "Speaker-independent word recognition using dynamic programming neural networks," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1, 29-32, 1989.
4. P. Haffner, M. Franzini and A. Waibel, "Integrating time alignment and connectionist networks for high performance continuous speech recognition," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1, 105-108, 1991.
5. L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, 77, 257-285, Feb. 1989.
6. H. Bourlard and C. Wellekens, "Links between hidden markov models and multilayer perceptrons," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12, 1167-1178, Dec. 1990.
7. A. Waibel, T. Hanazawa, G. Hinton, K. Shikano and K. J. Lang, "Phoneme recognition using time-delay neural networks," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37, 328-339, Mar. 1989.
8. N. Fakotakis and J. Sirigos, "A high performance text independent speaker recognition system based on vowel spotting and neural nets," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2, 661-664, 1996.
9. Y. Bannani and P. Gallinari, "A modular connectionist architecture for text-independent talker identification," *Proceedings of the International Joint Conference on Neural Networks*, 2, 857-860, 1991.
10. N. Morgan and H. Bourlard, "Hybrid connectionist models for continuous speech recognition," *Automatic Speech and Speaker Recognition*, Kluwer Academic Publishers, 259-283, Massachusetts, 1996.
11. S. Haykin, *Neural Networks*, Prentice Hall, New Jersey, 1999.
12. K. Fukuzawa, Y. Komori, H. Sawai and M. Sugiyama, "A segment-based speaker adaptation neural network applied to continuous speech recognition," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1, 433-436, 1992.
13. H. Demuth and M. Beale, *Neural Network Toolbox*, The MathWorks, 2001.

14. M. Riedmiller and H. Braun, "A direct adaptive method for faster backpropagation learning: the rprop algorithm," *Proceedings of the IEEE International Conference on Neural Networks*, 1, 586-591, 1993.
15. R. Fletcher, *Practical Methods of Optimization*, Wiley, New York, 1987.
16. M. Moller, "Supervised learning on large redundant training sets," *Proceedings of the 1992 IEEE-SP Workshop Neural Networks for Signal Processing*, 79-89, 1992.
17. S. Becker and Y. LeCun, "Improving the convergence of back-propagation learning with second-order methods," *Proceedings of the 1988 Connectionist Models Summer School*, 29-37, 1988.
18. Y. Bengio, *Neural Networks for Speech and Sequence Recognition*, International Thomson Computer Press, London, 1995.
19. D. R. Wilson and T. R. Martinez, "The need for small learning rates on large problems," *Proceedings of the International Joint Conference on Neural Networks*, 1, 115-119, 2001.
20. J. M. Naik, "Speaker verification: a tutorial," *IEEE Communications Magazine*, 1, 42-48, Jan. 1990.
21. C. Becchetti and L. P. Ricotti, *Speech Recognition*, John Wiley & Sons, New York, 1999.

저자 약력

● 이 태 승 (Lee Tae-Seung)



1997년 2월: 한국항공대학교 항공전자공학과 학사 (공학사)
 2000년 2월: 한국항공대학교 항공전자공학과 석사 (공학석사)
 2000년 3월 ~ 현재: 한국항공대학교 항공전자공학과 박사과정 재학
 ※ 주관심분야: 음성인식, 패턴인식, 인공지능

● 이 백 영 (Rhee Baeg-Yeong)



1969년 2월: 한국항공대학교 항공전자공학과 학사 (공학사)
 1982년 2월: 연세대학교 전자공학과 석사 (공학석사)
 1999년 9월 ~ 현재: 한국항공대학교 항공전자공학과 박사과정 재학
 1974년 3월 ~ 1990년 9월: 롯데파이오니아(주) 개발 부장
 1990년 10월 ~ 1998년 12월: 현대전자산업(주) 전장 연구소장 (상무)
 ※ 주관심분야: 음향신호처리, 음성인식

● 황 병 원 (Hwang Byong-Won)



1972년 2월: 한국항공대학교 항공전자공학과 학사 (공학사)
 1981년 2월: 도쿄대학교 전자공학과 석사 (공학석사)
 1984년 2월: 도쿄대학교 전자공학과 박사 (공학박사)
 1973년 ~ 1984년: 여수대학 교수
 1984년 ~ 1985년: 국방과학연구소 연구원
 1985년 ~ 현재: 한국항공대학교 전자정보통신컴퓨터 공학부 교수
 ※ 주관심분야: 이미지처리, 음성인식, 패턴인식