동적 확장 가능한 다중 계층 신경망에 기반한 음성 질의의 onset 검출 기법*

한병준[1O] 노승민[2] 황인준[1]
[1]고려대학교 전기전자전파공학부
{hbj1147, ehwang04}@korea.ac.kr
[2]아주대학교 정보통신전문대학원
anycall@ajou.ac.kr

An Onset Detection Scheme for Vocal Queries Based on Dynamic Expansible MLP*

Byeong-jun Han[1O] Seungmin Rho[2] Eenjun Hwang[1]
[1]School of Electrical Engineering, Korea University
[2]Graduate School of Information and Communications, Ajou University

요 약

음성 질의에서 효율적으로 onset을 검출하기 위한 연구는 다양하게 이루어져 왔다. 특히 대부분의 연구는 확률론적 모델에서 큰 성과를 나타내고 있다. 그러나 이러한 모델들은 변화나 확장이 쉽지 않다는 단점을 가지고 있다. 본 논문에서는 동적 확장 가능한 다중 계층 신경망(Dynamic Expansible MLP)을 제안하여, 기존 방법론의 확장 가능성을 모색한다. 또한, 음성 질의의 onset을 검출하기 위해 MLP를 활용하기 위한 모델을 제시한다.

1. 서 론

음성 및 신호에서 onset을 검출하기 위한 연구는 오랜 기간 진행되어 왔다. 과거의 연구는 대부분 신호 특성을 특성 벡터로 분석하여 onset을 검출하는 것이었지만, 최근의 연구는 통계적, 확률론적 모델을 많이 활용하고 있다. 특히 음성의 발화를 검출하기 위해서 HMM (Hidden Markov Model)이 널리 쓰이고 있다. 그러나 이러한 모델링 방법은 사전에 음성 신호의 특성을 면밀히 분석해야 모델링 가능하거나, 동적 확장이 어렵다는 단점을 가지고 있다.

본 논문에서는 음성 신호에서 onset을 검출하기 위해 다중 계층 신경망(MLP)을 사용한다. MLP는 여러 가지 한계점을 지니고 있다. 신경망 훈련을 위해 보편적으로 사용되는 역 전파(BP) 알고리즘은 훈련 속도가 느리고, 지역 해에 빠질 수 있다는 단점을 가지고 있다. 또한, MLP은 사용 도중 확장되기가 어렵다는 단점이 있다.

따라서 논문에서는 필요에 따라 MLP의 은닉 계층을 확장시킬 수 있는 동적 확장 가능한 다중 계층 신경망(Dynamic Expansible MLP)의 모델링 방법을 제안한다. 또한, 이 과정에서 MLP의 출력 및 가중 벡터의 에너지 변화를 최소화하기 위한 방법을 제시한다.

HMM, 베이즈 모델과 같은 확률론적 방법이 음성 인식에 다양하게 활용되고 있음에도 불구, 본 논문에서 사용하고자 하는 다양한 특성들은 확실하지 않은 경계선에서 onset임을 판단할 수 있는 근거들이다. 따라서 MLP를 음성 신호에 적용할 수 있는 가능성을 모색해본다. 이 과정에서 특성 벡터의 높은 차원을 허물기 위해 통계적 방법론을 사용할 것을 제안한다. 또한, 지역 해에 수렴하는 것을 방지하기 위해, 제안하는 동적 확장 가능한 다중 계층 신경망을 활용함과 함께, weight scaling과 같은 방법을 사용할 것을 제안한다.

본 논문의 나머지는 다음과 같이 구성되어 있다. 2절에서는 신경망의 구조 및 학습 알고리즘, 음성 신호 분석에 있어서 차원 저주 문제, 그리고 onset 검출을 위한 멀티미디어 데이터 처리에 관한 연구를 다룬다. 3절에서는 제안하는 동적 확장 가능한 신경망 구조를 수식과 함께 다루고 있다. 4절에서는 onset 검출을 위해 연속된 프레임 데이터를 신경망에 활용하기 위한 모델을 제시한다. 마지막으로 5절에서는 제안한 모델에 대한 결론을 도출하고, 향후 실험 계획을 논한다.

2. 관련 연구

이 절에서는 onset 검출을 위해 사용하는 신경망 구조 및 학습 알고리즘과 통계적 문제 해결방법, 그리고 audio 신호를 분석하기 위한 여러 특성을 기술한다.

2.1. 신경망 및 학습 알고리즘

MLP (Multi-Layer Perceptions)에 적용되는 BP (Back-Propagation) 알고리즘은 입력 신호의 조합에서 최적의

* 본 연구는 정보통신부 및 정보통신연구진흥원의 대학 IT연구센터 지원사업의 연구 결과로 수행되었음 (IITA-2006-(C1090-0603-0002))

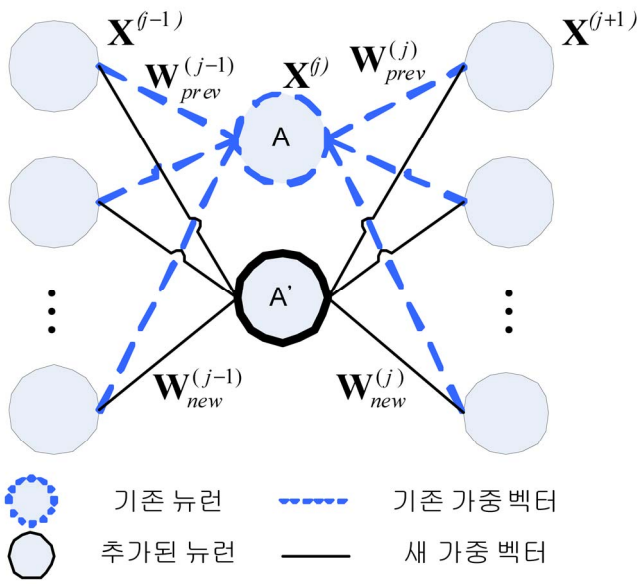

그림 1 은닉계층에서 새롭게 추가된 뉴런

분류 해를 찾아내기 위해 각각의 입력에 대한 출력의 오차를 에너지 함수로 만들고, 이를 Quadratic Form으로 간주한다. 그러나 BP 알고리즘의 문제점으로 지적되어온 것은 수렴 속도와 최적 해의 보장 여부이다[1]. 이 문제들을 해결하기 위해 weight scaling[1][2]과 mean field annealing[3][4], dynamic tunneling[5]과 같은 방법들을 사용해왔다. 또한 처음부터 weight vector set을 여러 개 만들어서 다양한 지역 해를 얻어 택일하거나, 입력 데이터에 비해 많은 수의 은닉 뉴런 구조에서 훈련시킨 다음 이들의 수를 감쇄시키는 방법(pruning) 등이 제안되었다.

시계열 데이터를 분석하기 용이하지 않은 MLP의 구조적 한계를 극복하기 위해 다양한 구조가 제안되었다. HRNN(Hopfield Recurrent Neural Network)[6]은 입력에 의한 출력을 다시 입력으로 넣는 회귀 구조를 가지고 있다. 이러한 신경망 구조는 입력이 변화함의 따라 내부의 구조도 동적으로 변화한다는 특성을 가지고 있다. 그러나, 동일한 입력에 대해 지속적인 변화가 이루어지기 때문에 출력의 동일성을 보장받을 수 없으며, 보다 넓은 범위의 시계열을 확실하게 고려하는 경우, 지연 처리를 위해 회귀 회로가 복잡해진다는 단점을 가지고 있다.

2.2. 차원의 저주

일반적으로 멀티미디어 데이터를 표현하기 위한 특성은 높은 차원을 가진다. 높은 차원의 데이터를 정제하지 않고 그대로 사용하면 복잡도와 처리시간이 늘어난다. 이러한 차원의 저주(the curse of dimensionality) 문제는 오래 전부터 논의되어온 문제이며 이를 해결하고자 하는 방법 또한 많이 제안되어 있다. 특히 데이터 본연의 특성을 보존하면서 데이터의 차원을 줄이기 위한 방법으로 PCA, ICA, LDA 등의 다양한 통계학적 방법론이 활용되고 있다. 특히 InMAF[7] 방법은 PCA 및 MLP가 결합된 방법론으로, 음악의 특성 차원을 감쇄시키는데 좋은 효과를 보이고 있다.

2.3. 멀티미디어 신호 처리

멀티미디어 데이터를 특성화하여 응용하기 위한 연구는 멀티미디어 신호처리 및 패턴인식 등의 분야에서 활발하게 이루어져 왔다.

본 논문에서 인식에 사용하고 있는 QBH (Query-By-Humming)[8][7] 인터페이스는 사람의 허밍 혹은 노래 음성을 입력 받아 이를 멀티미디어 검색에 활용한다. 특히 사람의 음성을 CMN (Common Music Notation)[9]으로 옮기기 위해서, 음성의 시작 부분과 끝 부분을 분석해야 한다. FFT와 같은 주파수 영역에서의 해석 방법과 달리, DTC[10], ADF[11]와 같은 특성은 시간 영역에서 이러한 특성을 쉽게 추출할 수 있게 해준다. 반면, MFCC (Mel-Frequency Constant Cepstrum)[12]는 FFT와 DCT 등에 근거한 방법으로, 다양한 음성 인식 분야에서 기본 특성으로 활용되고 있다.

3. 동적 확장 가능한 다중 계층 신경망

이 절에서는 본 논문에서 제안하고 있는 동적 확장 가능한 다중 계층 신경망(Dynamic Expansible MLP)에 대해 설명한다.

3.1. 개요

기존의 신경망은 은닉 뉴런의 개수가 고정된 모델 하에서 훈련하므로, 입력 데이터의 차원 복잡도가 늘어나는 경우 충분한 수의 은닉 뉴런을 필요로 한다. 그러나 입력 데이터에 비해 많은 은닉 뉴런은 편중된 데이터로 인해 필요 없는 지역 해가 계산될 수 있다. 따라서 보다 많은 은닉 뉴런이 필요해질 때 기존의 신경망과 동일한 결과를 산출하며 확장 가능한 신경망이 있다면, 좀더 효율적인 신경망의 구축이 가능할 것이다.

제안하는 다중 계층 신경망은 필요에 따라 은닉 뉴런을 추가하여 신경망을 재구축함과 동시에, 기존의 신경망과 동일한 결과를 산출하기 위한 모델링 수정 방법이다. 그림 1은 이러한 일반적인 추가 과정을 보이기 위해 축약된 모델을 보이고 있다.

3.2. 기본 설정

A는 j 번째 은닉계층을 구성하고 있는 기존의 뉴런이며, A'는 이 계층에 새롭게 추가된 뉴런이다. $\mathbf{X}^{(j)}$ 는 j 번째 은닉계층으로부터 $j+1$ 번째 은닉계층으로의 입력을 뜻하며, 이는 즉, $j-1$ 번째 은닉계층으로부터의 입력에 각각의 가중치와 활성화 함수를 적용한 결과이다. 또한 $\mathbf{X}_i^{(j)}$ 는 j 번째 은닉계층의 상위 i 번째 뉴런으로부터의 입력을 뜻한다. $\mathbf{W}^{(j)}$는 $\mathbf{X}^{(j)}$ 의 가중 벡터이며, $\mathbf{W}_i^{(j)}$는 이 중 $\mathbf{X}_i^{(j)}$ 에 해당하는 가중 벡터이다.

각각의 뉴런이 활성화 함수 φ 를 가지고, 활성화 함수는 역함수를 가지는 1:1 대응함수라 할 때, $\mathbf{X}^{(j-1)}$과 $\mathbf{W}_i^{(j-1)}$로부터 $\mathbf{X}_i^{(j)}$은 다음과 같이 계산된다.

$$\mathbf{X}_i^{(j)} = \varphi\left(\mathbf{X}^{(j-1)} \cdot \mathbf{W}_i^{(j-1)}\right) \quad (1)$$

이를 그림 1의 축소 모델에 적용한다. 노드가 추가되기 전에 j 번째 은닉계층에서는 $\mathbf{X}^{(j)}$ 이 출력되었으며, j 번째 은닉계층 전후로 $\mathbf{W}^{(j-1)}$, $\mathbf{W}^{(j)}$ 가 존재하였다고 하면, 노드가 추가됨으로써 새롭게 결정해야 하는 변수들은 다음과 같다.

$$\begin{cases} \mathbf{X}^{(j)} \rightarrow \left[\mathbf{X}_A^{(j)} \quad \mathbf{X}_{A'}^{(j)}\right] \\ \mathbf{W}^{(j-1)} \rightarrow \left[\mathbf{W}_{prev}^{(j-1)} \quad \mathbf{W}_{new}^{(j-1)}\right] \\ \mathbf{W}^{(j)} \rightarrow \left[\mathbf{W}_{prev}^{(j)} \quad \mathbf{W}_{new}^{(j)}\right] \end{cases} \quad (2)$$

기존 노드에서는 다음과 같은 관계가 성립한다.

$$\begin{cases} \mathbf{X}^{(j)} = \varphi\left(\mathbf{X}^{(j-1)} \cdot \mathbf{W}^{(j-1)}\right) \\ \mathbf{X}^{(j+1)} = \varphi\left(\mathbf{X}^{(j)} \cdot \mathbf{W}^{(j)}\right) \end{cases} \quad (3)$$

노드가 추가된 이후, 기존 노드와 새로운 노드에 대해 다음과 같이 표현할 수 있다.

$$\begin{cases} \mathbf{X}_A^{(j)} = \varphi\left(\mathbf{X}^{(j-1)} \cdot \mathbf{W}_{prev}^{(j-1)}\right) \\ \mathbf{X}_{A'}^{(j)} = \varphi\left(\mathbf{X}^{(j-1)} \cdot \mathbf{W}_{new}^{(j-1)}\right) \end{cases} \quad (4)$$

따라서 $j+1$번째 은닉계층으로부터의 출력은,

$$\mathbf{X}^{(j+1)} = \varphi\left(\mathbf{X}_A^{(j)} \cdot \mathbf{W}_{prev}^{(j)} + \mathbf{X}_{A'}^{(j)} \cdot \mathbf{W}_{new}^{(j)}\right) \quad (5)$$

이며, j 번째 은닉계층에 노드가 추가됨으로써 $j+1$번째 은닉계층에 영향을 끼치지 않기 위해서 (3)과 (5)의 $\mathbf{X}^{(j+1)}$이 다음과 같이 일치해야 한다.

$$\mathbf{X}^{(j)} \cdot \mathbf{W}^{(j)} = \mathbf{X}_A^{(j)} \cdot \mathbf{W}_{prev}^{(j)} + \mathbf{X}_{A'}^{(j)} \cdot \mathbf{W}_{new}^{(j)} \quad (6)$$

한편, 그림 1의 축약 모델에서는 하나의 뉴런에 또 하나의 뉴런을 추가하고 있다. $j-1$번째 및 $j+1$번째 은닉계층의 뉴런 개수를 각각 n 개, m 개라 할 경우, $\mathbf{W}^{(j-1)}$ 와 $\mathbf{W}_{prev}^{(j-1)}$, $\mathbf{W}_{new}^{(j-1)}$ 는 $n\times 1$ 행렬, $\mathbf{W}^{(j)}$ 와 $\mathbf{W}_{prev}^{(j)}$, $\mathbf{W}_{new}^{(j)}$ 는 $1\times m$ 행렬이다. 따라서 (3)과 (5)의 $\mathbf{X}^{(j)}$ 와 $\mathbf{X}_A^{(j)}$, $\mathbf{X}_{A'}^{(j)}$ 는 상수로 간주할 수 있으며, 이를 임의로 결정하면 새로이 생성되는 가중치 벡터 간 비율로 활용할 수 있다.

3.3. j+1번째 은닉계층의 파라메터 결정

새로운 뉴런이 추가됨으로써 은닉계층 출력값의 총 에너지가 변화하면, 전체 시스템의 에너지 총량이 변화한다. 따라서 전체 에너지를 일정하게 유지하게 하기 위한 $\mathbf{X}^{(j)}$ 와 $\mathbf{X}_A^{(j)}$, $\mathbf{X}_{A'}^{(j)}$ 관계를 다음의 관계식으로 표현할 수 있다.

$$\left(\mathbf{X}^{(j)}\right)^2 = \left(\mathbf{X}_A^{(j)}\right)^2 + \left(\mathbf{X}_{A'}^{(j)}\right)^2 \quad (7)$$

한편, $\mathbf{W}_{prev}^{(j)}$ 와 $\mathbf{W}_{new}^{(j)}$ 의 경우, 일반적으로 $\mathbf{X}_A^{(j)}$ 와 $\mathbf{X}_{A'}^{(j)}$ 를 (6)에 적용하여 좌변과 우변을 만족시키는 어떠한 형태의 벡터도 가능하다. 그러나 BP 알고리즘 등에 의해 가중 벡터 조정이 발생할 경우, 처음 노드와 연결된 가중 벡터와 전혀 다른 결과를 예상할 수 있다. 이와 같은 결과의 변화를 최소화하기 위해, 가중 벡터의 에너지 총합을 일정하게 한다.

$$\left(\mathbf{W}_i^{(j)}\right)^2 = \left(\mathbf{W}_{prev}^{(j)}\right)_i^2 + \left(\mathbf{W}_{new}^{(j)}\right)_i^2 \quad (8)$$

다음 명제에 의해 (6), (7), (8)의 선형방정식의 조건을 만족하는 $\mathbf{W}_{prev}^{(j)}$ 와 $\mathbf{W}_{new}^{(j)}$ 벡터의 요소들이 $\mathbf{W}^{(j)}$ 벡터의 요소와 항상 선형관계를 이룬다. 또한, $\mathbf{X}_A^{(j)}$ 나 $\mathbf{X}_{A'}^{(j)}$ 가 결정되어 있을 경우, $\mathbf{W}_{prev}^{(j)}$ 와 $\mathbf{W}_{new}^{(j)}$ 는 항상 유일하며, 이는 $\mathbf{X}^{(j)}$ 와 $\mathbf{X}_A^{(j)}$, $\mathbf{X}_{A'}^{(j)}$ 의 비례식을 통해 계산할 수 있다.

(명제) 다음의 세 등식을 만족하는 0 아닌 실수 a, b, c, x, y, z 가 있다.

$$ax+by=cz,\quad a^2+b^2=c^2,\quad x^2+y^2=z^2$$

이때, x, y, z는 주어진 a, b, c에 대해 항상 선형 비례한다.

(증명) $x^2+y^2=z^2$ 을 주어진 첫번째 식을 이용하여 x와 z에 관한 식으로 나타내어 정리하면 다음과 같다.

$$\left(a^2+b^2\right)x^2 - 2acxz + \left(c^2-b^2\right)z^2 = 0 \quad (9)$$

이때 $a^2+b^2=c^2$ 이고, $c^2-b^2=a^2$ 이므로, 이를 (9)에 대입하여 간단하게 표현하면 다음과 같다.

$$\left(cx-az\right)^2 = 0,\quad x=\frac{az}{c} \quad (10)$$

따라서 (10)에 의해 x, z는 항상 선형비례하며, (10)을 $x^2+y^2=z^2$ 에 대입하여 정리하면 다음의 결론이 도출된다.

$$a^2z^2 + c^2y^2 = c^2z^2,\quad y=\pm\frac{bz}{c} \quad (11)$$

따라서 (10)과 (11)에 의해 0 아닌 실수 a, b, c, x, y, z 에 대해 가정을 만족하는 경우, x, y, z 는 항상 선형 비례한다. (증명 끝)

3.4. j-1번째 은닉계층의 파라메터 결정

$j-1$ 번째 은닉계층의 가중 벡터에 대해서도 다음의 에너지 총합 조건을 만족하면 잠재적 변화가 크지 않다.

$$\left(\mathbf{W}_i^{(j-1)}\right)^2 = \left(\mathbf{W}_{prev}^{(j-1)}\right)_i^2 + \left(\mathbf{W}_{new}^{(j-1)}\right)_i^2 \quad (12)$$

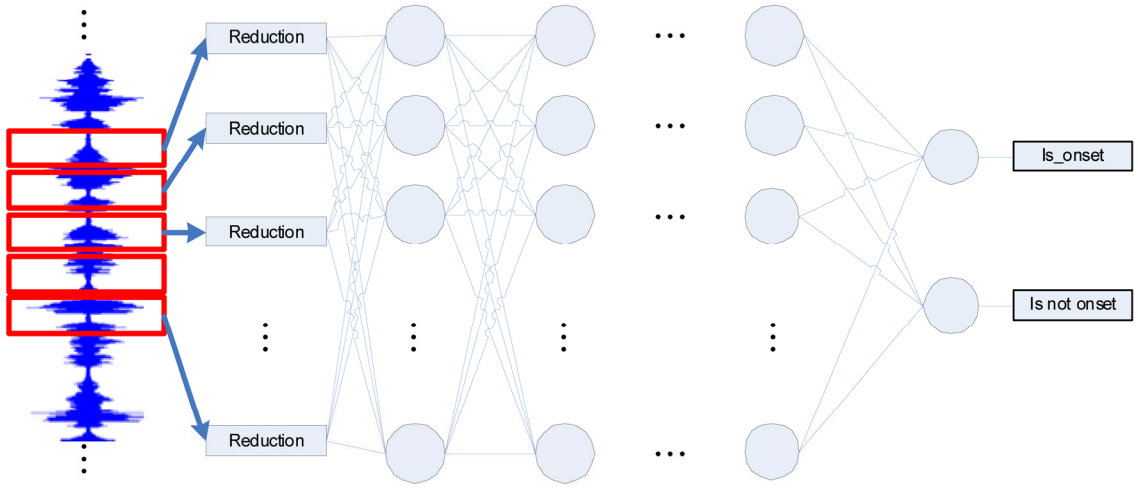

그림 2 음성 신호에서 onset 검출을 위한 다중 계층 신경망의 적용 모델

그러나 $\mathbf{X}^{(j)}$ 와 $\mathbf{X}_{A}^{(j)}$, $\mathbf{X}_{A'}^{(j)}$ 는 (7)의 관계를 가지며, 활성화 함수 φ 에 의존적이므로, 본 논문에서는 $\mathbf{W}_{prev}^{(j-1)}$ 과 $\mathbf{W}_{new}^{(j-1)}$ 를 $\mathbf{W}^{(j-1)}$ 에 선형 비례하며 (4) 및 (7)의 조건을 만족하는 벡터를 역산한다.

3.5. 신경망의 성장

제안하는 신경망 구조는 모든 상황 시점, 그리고 모든 은닉계층에 뉴런의 추가가 가능하다. 그러나 뉴런 추가의 필요성이 정의되지 않으면, 결국 최후에는 지나치게 많은 수의 뉴런으로 구성된 신경망을 최적화(pruning)해야 할 것이다.

따라서 훈련 데이터에 의해 충분히 학습된 신경망이 검증 데이터에 의해 적절한 성능을 발휘하지 못하는 경우, 신경망을 성장시키도록 한다. 또한, 기존의 뉴런을 대체함에 있어서, 입력 혹은 출력되는 가중 벡터의 에너지가 다른 벡터의 에너지보다 큰 뉴런을 가장 우선적으로 대체하도록 하여, 신경망 내부의 에너지가 특정 뉴런 혹은 특정 가중 벡터에 집중되지 않도록 한다.

4. 다중 계층 신경망의 적용

이 절에서는 사람의 음성의 발성 여부를 파악하기 위해 제안한 다중 계층 신경망을 학습시키기 위한 모델(그림 2)을 설명한다.

4.1. 연속된 음성 신호 처리

기본적으로 MLP는 시계열 데이터의 처리에 한계가 있다. 하지만 본 연구에서는 음성 신호를 프레임 단위로 분리하며, 각 프레임의 특성벡터와 현재 프레임으로부터 과거 혹은 미래의 일정 길이의 프레임과의 연관성을 고려하기에, 그림 2와 같은 모델의 형태로 MLP의 적용이 가능하다.

한편, 각 프레임의 특성은 음성의 onset 검출을 위한 대표적인 특성을 위주로 추출한다. 기존 onset 검출에 사용되어 온 AE (Average Energy), ZCR (Zero-Crossing Rate), MFCC 계수, DTC[10], ADF[11] 등을 추출한다. 특히 MFCC의 경우 사람 음성의 대역을 나타낼 수 있는 13개의 계수만을 고려하였으며, 이중 가장 저주파 영역을 고려하는 첫번째 계수(C0)는 onset 검출에 큰 영향을 미치지 않기에 고려하지 않는다. 따라서 각 프레임 당 특성 데이터는 16차원을 가지게 된다.

n개의 프레임이 입력으로 활용되는 경우, 전체 입력 데이터 벡터는 16n차원을 가지게 된다. 이러한 입력 벡터는 각 프레임 데이터의 주요한 특성이 추출되지 효과적으로 활용되지 않으므로, 주요 특성을 효과적으로 분리해야 할 필요성이 있다. 본 논문에서는 사전에 프레임으로부터 추출된 16차원의 데이터를 PCA와 ICA와 같은 통계적 방법을 사용하여 적절한 차원으로 사영하는 것을 제안한다.

4.2. 지역 해 수렴 문제

제안하는 신경망 구조는 뉴런의 동적 증가로 잠재적인 성장 가능성이 있으나, 지역 해에 수렴된 경우에는 기존 신경망과 동일한 출력을 보이므로 학습을 통해 벗어나기 힘들다. 따라서 실험에서는 뉴런을 동적 증가시킴과 함께, Weight Scaling 방법[1]을 사용하여 가중 벡터의 전체 에너지를 증가시키는 것을 제안한다.

5. 결론 및 실험계획

본 논문은 동적 확장 가능한 신경망 구조를 제안함으로써, 이미 구성된 신경망 구조와 동일한 결과를 보이면서, 동적으로 확장시킬 수 있는 방법을 제시하였다. 특히, 뉴런이 추가됨과 동시에 신경망 내부의 에너지를 정적으로 유지하게 하기 위해, 뉴런의 출력 및 가중 벡터의 에너지 합을 일정하게 유지하도록 하였다.

또한, 시계열 데이터를 분석하여 이산 결과를 도출하기 위해, 일정 수의 프레임으로부터 추출된 특성 벡터를 통계적 방법을 통해 사영하고, 이를 입력 데이터로 활용하는 것을 제안하였다. 이 과정에서 보다

효율적인 훈련을 위해, 뉴런을 동적 증가시키며, weight scaling 방법을 사용할 것을 제안하였다.

실험은 매우 다양하게 이루어져야 할 것이다. 우선 동일한 초기 가중 벡터 set에 대해 기존의 MLP와 제안된 동적 확장 가능한 신경망 구조의 훈련 효율성 등을 비교하여야 할 것이며, onset 검출에서는 프레임의 길이, 사영 차원 등을 고려하여 실험하여야 할 것이다.

6. 참고 문헌

[1] Y. Fukuoka, et al, "A modified back-propagation method to avoid false local minima," *Neural Network*, Vol.11, No.6, pp. 1059-1072, Jun. 1998.

[2] A.K. Rigler, et al, "Rescaling of variables in back propagation learning," *Neural Network*, Vol.4, No.2, pp. 225-229, 1991.

[3] G.L. Bilbro, et al, "Mean Field Annealing: A Formalism for Constructing GNC-like Algorithms," *IEEE Trans. on Neural Networks*, Vol.3, pp. 131-138, 1992.

[4] C. Perterson and J.R. Anderson, "A mean field theory learning algorithm for neural networks," *Complex Systems*, Vol.1, pp. 995-1019, 1987.

[5] P. RoyChowdhury, et al, "Dynamic Tunneling Technique for Efficient Training of Multilayer Perceptions," *IEEE Trans. on Neural Networks*, Vol.10, pp. 48-55, Jan. 1999.

[6] Hopfield, J.J., "Neural Networks and Physical Systems with Emergent Collective Computational Abilities," *Proc. Natl. Acad. Sci. USA*, Vol. 79, pp. 2554-2558, 1982.

[7] J. Shen, et al, "Towards Effective Content-Based Music Retrieval With Multiple Acoustic Feature Combination," *IEEE Trans. on Multimedia*, Vol.8, No.6, Dec. 2006.

[8] Ghias, et al, "Query By Humming: Musical Information Retrieval in an Audio Database," *ACM Multimedia 1995*, pp. 231-236, 1995.

[9] S. Rho and E. Hwang, "FMF: Query adaptive melody retrieval system," Journal of Systems and Software (JSS), Vol.79, No.1, pp.43-56, 2006.

[10] B. Han, S. Rho, E. Hwang, "An Efficient Voice Transcription Scheme for Music Retrieval," to appear in *IEEE Conf. Proc. on MUE 2007*, Apr. 2007.

[11] S. Park, S. Kim, K. Byeon, E. Hwang, "Automatic Voice Query Transformation for Query-by-Humming Systems," *In Proc. of IMSA 2005*, pp. 197-202, Aug. 2005.

[12] Davis, S.B., Mermelstein, P., "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences," *IEEE Trans. on Acoustic, Speech and Signal Processing*, Vol.28, No.4, pp. 357-366, 1980.

[13] L. Lu, D. Liu, H.-J. Zhang, "Automatic Mood Detection and Tracking of Music Audio Signals," *IEEE Trans. on Audio, Speech, and Language Processing*, Vol.14, No.1, Jan. 2006.