

Cepstrum방법과 신경회로망을 이용한 정상, 양성종양, 악성종양 상태의 식별에 관한 연구

조철우, 김대현

창원대학교 제어계측공학과

On the Classification of Normal, Benign, Malignant Speech Using Neural Network and Cepstral Method

Cheol-Woo Jo, Dae-Hyun Kim

Department of Control & Instrumentation Engineering

cwjo@sarim.changwon.ac.kr

요 약

본 논문에서는 환자의 음성을 정상, 양성종양, 악성종양으로 분류하는 실험을 cepstrum 파라미터를 통한 음원분리와 신경회로망을 이용하여 수행하고 그 결과를 보고한다. 기존의 상해음성 데이터베이스에는 정상음성과 양성종양의 경우만 수록되어 있었고 외국어 환자들을 대상으로한 경우만 있었기 때문에 국내의 환자들에게 직접 적용할 경우 어떠한 결과가 나올지 예측하기가 어려웠다. 최근 부산대학교 이마인후과팀에서 수집한 국내의 정상, 양성, 악성종양의 경우에 대한 데이터베이스를 분석하고 신경회로망에 의해 분류함으로써 사람의 음성신호만에 의한 후두질환의 식별이 가능하였다.

본 실험에서는 식별 파라미터로 음성신호의 선형 예측 코efficient에 관한 cepstrum으로부터 음원분리인 HNRR을 구하여 Jitter, Shimmer와 함께 사용하였다. 신경회로망은 일층의 층과 한 개의 은닉층을 갖는 다층신경망을 이용하였으며, 식별은 두 단계로 나누어 정상과 비정상을 분류한 후 다시 비정상을 양성과 악성으로 분류하였다.[1].

1. 서 론

최근 들어 인간의 건강에 대한 관심이 점점 증가하고 있다. 이마인후과 영역에서는 성대의 질환을 음성의 음향적 특성을 분석하여 질병의 징후를 발견하려는 시도가 여러 곳에서 이루어지고 있다. 이러한 시도는 질병의 정확한 진단의 목적

이러기보다는 질병의 가능성을 사전에 발견하여 질병의 상태가 깊어지기 않도록 예방한다는 차원에서 필요하다. 일반적으로 성대의 질환을 진단하기 위해서는 성대를 직접 검사할 수 있는 내시경과 같은 기구를 이용하여 직접 보는 것이 가장 효과적이라고 한다. 실제로 의사들도 이러한 방법을 많이 사용하고 있다. 그러나 이 방법은 정확한 진단이 가능한 반면 숙달된 전문의사와 전문적인 기구가 필요하여 일반의사나 환자 자신이 질병의 유무를 판단하기에 적합한 방법이 아니다. 이러한 목적으로는 환자의 음성만에 의하여 질병의 유부, 종류를 진단하려는 시도가 곳곳에서 이루어지고 있다. 성대의 질병은 많은 경우 성질(Voice Quality)의 변화를 수반하지만 그렇지 않은 경우도 있다. 목소리만으로 질병을 진단할 수 있는 경우는 질병으로 인해 음성의 음향적 특성이 변화된 경우에 한한다.

본 논문에서는 이와 같은 음향적 분석법에 의한 성대질환의 진단을 목적으로 하는 파라미터로 사용되는 jitter, shimmer 및 cepstrum방식의 음원분리에 의한 파라미터인 HNRR[1]의 조합으로 신경회로망을 이용하여 정상, 양성종양, 악성종양을 구분하는 실험을 행하고 그 결과를 고찰하였다.

2. 상해음성의 식별파라미터

장애음성의 음향적 특성을 나타내는 파라미터로는 여러 가지 종류가 있다. 그 중에서도 통계적으로 분석에 많이 사용되어지고 변별력이 있는 파라미터로 jitter, shimmer가 대표적인 파라미터이

다.[2]

그리고 기존의 파라미터의 신뢰성 문제로 인해 새로 제안된 캡스트럼 방식의 음원분석에 의한 파라미터인 HNRR(Harmonic to-Noise Ratio : Residual)를 장애음성 식별에 사용하려고 한다.

Jitter는 피치주기의 변화율을 나타내는데 사용되어지는 파라미터로 연속적인 피치주기사이의 평균 퍼센트 변화로 나타낼 수 있다.

$$Jitter = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |P(i) - P(i+1)|}{\frac{1}{N} \sum_{i=1}^N P(i)} \quad (2-1)$$

식2-1에서 $P(i)$ 는 i 번째 피치의 주기이고, N 은 측정된 피치의 개수이다.

Shimmer는 jitter와는 달리 peak사이의 크기에 대한 변화율을 나타낸다.

$$Shimmer = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |A(i) - A(i+1)|}{\frac{1}{N} \sum_{i=1}^N A(i)} \quad (2-2)$$

식2-2에서 $A(i)$ 는 i 번째 peak사이의 크기를 나타낸다.

마지막으로 HNRR은 음성에서 선형예측분석으로부터 구해진 예측오차신호로부터 캡스트럼을 구한 뒤 하모닉 성분과 잡음성분을 분리하여 그 비율을 나타낸 파라미터이다.[1][3] 선행실험에서 HNRR은 jitter나 shimmer에 비해 높은 변별력을 갖는 것이 확인되었다.

이러한 세 가지 파라미터를 이용하여 시료간의 통계적 처리결과를 바탕으로 장애음성(양성종양과 악성종양)과 정상음성사이의 관계 및 효과적인 식별방법을 찾고자 한다. 그 식별방법으로 신경회로망을 이용한다.

3. 패턴인식방법

근래 들어 인간의 두뇌가 대량의 복잡한 데이터를 병렬 처리할 수 있을 뿐만 아니라 학습능력이 있다는 사실에 근거하여 새로운 패턴인식의 방법으로 신경회로망이 자주 사용되고 있다.

신경회로망은 뉴런 즉 처리할 요소의 선이합수, 회로망의 구조 즉 layer의 수 및 처리할 요소들간의 연결상대 및 연결강도를 주어진 문제해결에 적절하게 조정하는 학습 규칙들에 따라 여러 가지의 형태가 있다. 대표적인 것을 들면 Hopfield network, perceptron, Boltzmann machine, error

propagation model, Grossberg model 및 neocognition 등이 있다. 이들에서 패턴인식 및 식별에 직접적으로 관련되는 모델들은 Hopfield network, Grossberg model, perceptron, Neocognitron등이 있으며, 그중 perceptron은 feed-forward 연결구조를 가지며 패턴 식별의 기능을 갖는 간단한 형태의 신경회로망이다. 초기의 단층 구조의 perceptron은 그 구조의 간결성과 특히 수학적으로 증명된 학습규칙으로 많은 관심의 대상이었으며, 그후 Rumelhart등에 의하여 back-propagation의 학습 알고리즘이 제안됨으로써 다층구조의 perceptron은 그 계산기능이 크게 확장되어 많은 문제에 응용될 수 있게 되었다.[4]

Back-propagation은 일종의 gradient 탐색방법으로서 기대되는 출력층의 node들에서 대하여 비용함수를 계산하여 이를 감소하는 방향으로 가중치를 조정하여 다음에 한층 낮은 층의 node에 대하여 에러를 계산하여 이를 감소하도록 가중치를 조절하는 식으로 반복한다.

본 논문에서는 그림1에서와 같은 3개의 계층구조를 갖는 신경회로망을 구성하고 back-propagation방법에 의해서 훈련시킨 다음 정상음성과 장애음성의 식별에 사용하였다.

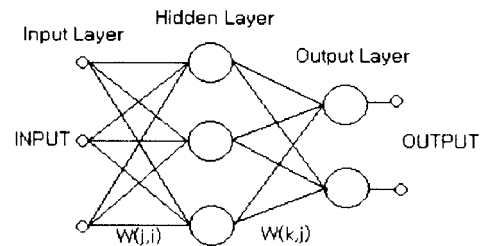


그림 2 다층 구조 신경회로망

4. 실험 및 검토

본 논문에서는 부산대학교병원에서 수집한 한국인 장애음성데이터를 대상으로 정상음성과 장애음성(양성, 악성종양)을 식별하는 실험을 수행하였다. 이 데이터베이스에서 정상음성 33개, 양성종양 17개, 악성종양 16개를 이용하여, 전체 66개의 데이터에 대해 신경망 훈련 및 식별 실험을 하였다. 실험에 사용한 각 음성은 모음 /아/를 일정한 시간동안 발음한 것으로 샘플링 주파수는 16kHz이며, 양자화 비트수는 16bit로 되어있다.

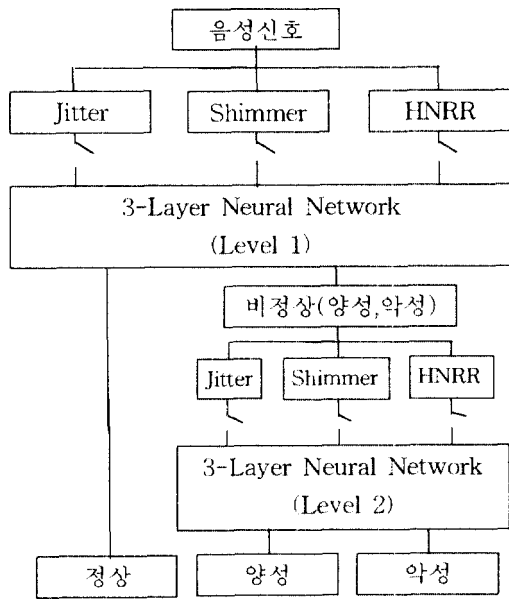


그림 2. 전체적인 실험방법

그림2는 전체적인 실험방법에 대해 간략하게 나타낸 것이며, 먼저 정상음성과 장애음성에 대한 파라미터로 jitter, shimmer, HNRR을 구해 각 파라미터들을 일정한 조합으로 단계 1의 신경회로망의 입력이 된다. 단계 1의 결과에서 식별된 장애음성을 다시 단계 2의 신경회로망을 통과시켜 양성과 악성으로 분류한다.

신경회로망은 그림1에 주어진 것과 같이 입력
표 1. 단계 1의 에러의 수와 식별율

		Jitt-Shim		Shim-HNRR		HNRR-Jitt		Jitt-Shim-HNRR	
		정상	비정상	정상	비정상	정상	비정상	정상	비정상
Train Data	정상	21	1	19	3	18	4	21	1
	비정상	6	16	2	20	3	19	4	18
식별율(에러/전체)		84.09%(7/44)		88.64%(5/44)		84.09%(7/44)		88.64%(5/44)	
Test Data	정상	11	.	9	2	9	2	10	1
	비정상	1	10	1	10	2	9	1	10
식별율((에러/전체))		95.45%(1/22)		86.36%(3/22)		81.81%(4/22)		90.9%(2/22)	

표 2. 단계 2에서의 에러의 수와 식별율

		Jitt-Shim		Shim-HNRR		HNRR-Jitt		Jitt-Shim-HNRR	
		양성	악성	양성	악성	양성	악성	양성	악성
Train Data	양성	11	.	9	2	9	2	11	.
	악성	.	11	0	11	.	11	.	11
식별율(에러/전체)		100%(0/22)		90.9%(2/22)		90.9%(2/22)		100%(0/22)	
Test Data	양성	6	.	6	.	5	1	6	.
	악성	3	3	3	3	.	6	6	.
식별율(에러/전체)		75%(3/12)		75%(3/12)		91.67%(1/12)		50%(6/12)	

층과 출력층 사이에 1개의 은닉층을 가지는 다층 구조 신경망을 사용하였다. 입력층은 세가지 파라미터의 조합이고, 은닉층에는 3개의 뉴런을 사용하였으며, 원하는 출력값은 단계 1.2에서 정상과 비정상(장애음성), 양성과 악성으로 구분하기 위해 1과 0으로 하였다. 신경망의 학습을 위해서는 역전파 학습알고리즘을 이용하였다.

신경회로망에 사용한 전체적인 입력데이터 중 훈련을 위해 2/3를 훈련데이터로 사용하고 나머지 1/3은 테스트 데이터로 사용하였다. 표 1은 단계 1에서의 훈련데이터와 테스트데이터에 대한 에러의 수와 식별율을 나타낸 것이고, 표 2는 단계 2에서의 에러의 수와 식별율을 나타내었다.

표1의 정상음성과 장애음성간의 식별에서 훈련데이터의 경우는 세가지 파라미터를 모두 사용한 경우가 88.64%로 가장 식별율이 높은 것으로 판정되었고 이렇게 훈련된 신경회로망을 이용한 나머지 데이터를 이용한 식별실험의 경우는 Jitter와 Shimmer를 이용한 신경회로망이 가장 높은 식별율을 나타내었다. 그러나 세가지 파라미터를 모두 사용한 경우도 90.9%로 높았기 때문에 양쪽의 경우 모두 높은 식별율을 보인 세가지의 조합이 유효한 것으로 판단된다.

표2의 경우 양성과 악성종양의 구분실험에서는 전반적으로 HNRR과 Jitter의 조합이 높은 식별율을 보였는데 악성과 양성종양의 환자데이터의 절

대수가 부족하기 때문에 신뢰성은 많이 떨어진다 고 볼 수 있다.

두 단계의 신경회로망을 이용한 실험으로부터 우리는 정상, 양성, 악성종양의 구분이 신경회로망과 파라미터의 적절한 조합에 의해 상당한 정확도를 가지고 수행될 수 있음을 확인하였다.

현 상태에서는 종양데이터의 수가 절대적으로 부족한 상태이지만 앞으로 자료가 충분히 확보된다면 본 실험에서 사용한 방법이 환자진단의 선행 단계로 사용이 가능함을 보여주었다.

5. 결론

본 논문에서는 정상음성과 장애음성을 구분하기 위해 3-layer의 신경회로망을 이용하였으며 신경회로망의 입력으로는 장애음성식별파라미터인 jitter, shimmer, HNRR을 이용하였다. 신경회로망의 학습을 위해서는 back-propagation을 사용하였다. 실험결과 세 가지 파라미터를 이용하고 단계별로 신경회로망을 적용함으로써 효과적으로 정상음성과 장애음성, 양성과 악성종양간의 구분이 가능하다는 것을 입증할 수 있었다.

차후의 연구에서는 보다 많은 종류의 질병에 대하여 정상음성과의 식별 및 각 질병간의 구분에 대해서도 실험을 계속 진행할 예정이며, 또한 악성종양 환자의 음성수집 및 국내의 정상인과 장애음성화자의 음성을 수집하여 동일한 실험을 계속할 예정이다.

감사의 글

본 연구는 한국학술진흥재단의 1997 학제간 연구 '음향신호의 분석에 의한 후두질환의 진단에 관한 연구'의 연구결과의 일부입니다. 지원에 감사드립니다.

참 고 문 헌

1. 조철우, 김대현, "음원분석을 통한 장애음성의 음향적 특성분석에 관한 연구", 한국음향학회 하계학술발표대회 논문집, pp163-166, 1998
2. Operations Manual, "Multi-Dimensional Voice Program(MDVP)", Model 4305, Kay Elemetrics Corp, 1993
3. B.Yegnanarayana, C.d'Alessandro, V.Darsinos, "An Iterative Algorithm for

Decomposition of Speech Signals into Periodic and Aperiodic Components", IEEE trans. on Speech and Audio Processing, Vol.6, No 1, Jan. 1998

4. R.P.Lippman, "An Introduction to computing with Neural Nets", IEEE ASSP Magazine, Vol.4, No 2, pp4-20, April, 1987

5. Operations Manual, "Disordered Voice Database", Model 4337, version 1.03, Kay Elemetrics Corp, 1994