선택적 패턴블럭 신경회로망을 이용한 불특정 화자 인식

°강 명광*, 서 민석*,이 형준**, 양 진우*, 김 순협*
*광운대학교 컴퓨터공학과 **한림전문대학교 전자통신과

A Study on Unspecified Speaker Recognition by Selective Pattern-Block Neural Netwotk

°MyungKwang Kang*,MinSuk Seo*,JinWoo Yang*,Hyungjun Lee*,SoonHyob Kim*
*Dept. of Computer Engineering, KwangWoon Univ.
*Dept. of Electronic Communication, Hallym Junior College.

요 약

본 연구는 특정 파라메터의 특성을 고려한 신경회로망에 관한 연구로서 패턴블럭 선택적 신경회로망을 제안하고, 제안한 신경회로망의 성능을 평가하기 위하여 한국어 단모음에 대한 불특정 화자 인식 실험을 하였다.

각 패턴에 따른 특정 파라메터의 변화를 고려하지 않은 기존의 패턴매칭(pattern matching) 알고리즘에 비하여 제안된 신경회로망은 인가된 패턴을 파라메터의 특성에 맞게 몇 개의 부패턴(sub-pattern)으로 분할한 후 가장 최적의 부패턴을 선택하여 학습하고 인지하는 것이 그 특징이다

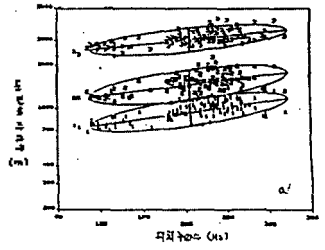
1. 서 론

McCulloch-Pitts이 뉴런의 동작을 논리적 표현으로 제시한 이래 많은 연구가 진행되어 오늘날 다양한 신경회로망 모델이 존재하고 계산량이 많은 패턴 인식 및 분류 기술이 필요한 다양한 분야에서 신경망을 이용하기 위하여 활발한 연구가 진행되고 있다.

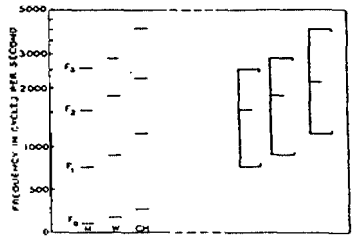
일반적으로 패턴 인식 또는 분류에서 좋은 결과를 얻기 위해서는 패턴 분류기의 성능뿐만 아니라 특정 파라메터의 선정또한 매우 중요하기 때문에 두 분야에서 많은 연구가 진행되어 왔다. 하지만, 기존의 접근 방법은 가장 효과적으로 패턴을 표현할 수 있는 특징 추출 방법에 대한 연구와 패턴에서 추출된 특정 파라메터의 확률 분포 공간을 가장 최적으로 분할할 수 있는 패턴 분류기에 대한 연구가 서로 독립적으로 진행되어 왔다. 그러므로 특정 파라메터가 본질적으로 가지고 있는 문제는 패턴분류시 그대로 가지고 있게 된다. 이러한 문제점은 분류되는 패턴에 대한 특정 메터 사이의 중첩된 분포로 나타나게 되어 아무리 뛰어난 패턴 분류기나 학습라도 좋은 결과를 기대하기는 힘들다. 본 연구에서는 이 두 분야를 상호 고려하여 설계된 패턴블럭 선택적 신경회로망을 제안하고 음성 인식 분야 중 불특정 화자 인식에 적용을 하여 본다.

2. 패턴블럭 선택적 신경회로망

패턴 인식 및 분류 분야에서 사용하는 특정 파라메터들은 동일한 클래스(class)에 속하는 다른 패턴들 사이에 서의 유사한 특정 파라메터만을 갖는 것이 아니다. 예를 들면, 한국어에 대한 단모음의 / ㅏ/와 영어 /a:/에 대하여 포만트 주파수가 화자별, 성인 남녀 그리고 어린이에 따라 달라지는 것을 그림 1에서[1][3] 알 수 있다. 같은 모음에 대하여 화자에 따라 포만트 주파수가 각 포만트 상대 위치를 어느 정도 유지하면서 주파수 축을 따라 이동하는 것을 알 수 있다. 그러므로 이러한 특정 파라메터의 특성은 파라메터 공간을 안정한 부공간(sub space)의 단위, 음성의 경우 포만트 주파수 영역이 이동하는 것을 알 수 있다. 이와 같은 현상을 고려하지 않을 경우 각 패



한국어 /a/



영어 /a:/

그림 1 한국어 /a/와 영어 /a:/의 포만트 주파수

턴의 중첩된 분포가 확대되어 패턴 분할이 어렵게 된다. 이와 같은 현상을 다른 관점에서는 하나의 특징 파라메터 속에 몇 개의 패턴블럭이 존재하고 그 패턴 블럭중 패턴의 특성을 나타내는 부 파라메터(sub-parameter)가 존재하면서 패턴에 따라 이동한다고 볼 수 있다. 그러므로 패턴 인식기는 이와 같은 특정 파라메터에 내재하고 있는 패턴블럭을 검출하고 학습할 수 있어야만 좋은 인식 결과를 기대할 수 있다.

본 연구에서는 이러한 문제점을 해결하기 위하여 1960년대에 시각 신경계에서 발견된 국부적 패턴에 선택적으로 반응하는 뉴런을 모델링하여 패턴블럭 선택적 신경회로망을 제안한다. 이러한 뉴런은 특정한 위치에서 선과 모서리 등과 같은 국부적인 특징들에 대하여 선택적으로 반응한다. 또한 시각 피질보다 상위 영역에서는 삼각형이나 사각형 등과 같은 패턴에 선택적으로 반응하는 셀의 존재를 1980년대 초에 발견되었다.

신경회로망에 인가되는 패턴에 대한 특징 벡터를 $X=(x_1, x_2, \cdots, x_N)$라하고 각 패턴을 나타내는 패턴블럭을 $X_b=(x_{b1}, x_{b2}, \cdots, x_{bM})$이라 하자. 제안된 신경회로망은 그림 2와 같은 구조로 되어 있다. N개의 뉴런으로 이루진 입력 층과 M개의 입력 뉴런을 가진 블럭매칭 신경회로망과 $L = N-M$개의 가상 뉴런으로 이루어진 블럭들로 이루어진 출력 층으로 이루어지 있다. 그리고 학습시 패턴블럭 정보를 입력 층에 전달해 주는 회귀 연결이 있는 구조이다. 입력 층에 N차원의 특징벡터 X가 인가되면 M개의 뉴런으로 입력층이 이루어진 블럭매칭 신경회로망은 M차원의 패턴블럭 단위로 인가된 특징벡터를 스캔(scan)하면서 매칭(matching) 결과를 패턴블럭에 대응하는 출력 층의 가상 뉴런들에게 전달한다 출력 층의 가상 뉴런들은 블럭내에서 경쟁하여 승자 뉴런만이 출력한다. 그리고 이들 출력중 최대 출력을 갖는 뉴런에 대응하는 패턴블럭이 선택되어 입력 층에 다음과 같은 회귀 정보를 전달한다.

$$S(X_b) = \arg\max O_j(b) \qquad (1)$$

$S()$은 블럭 선택 함수이고, $O_j(b)$은 패턴블럭 b에 대한 출력 층의 j번째 블럭의 출력을 나타낸다. 블럭매칭 신경회로망은 블럭선택함수 $S()$에 의해 허가된 패턴블럭에 대한 강화 학습과 나머지 패턴블럭에 대하여 반강화 학습 진행한다.

3. 불특정 화자 인식을 위한 패턴블럭 선택적 신경회로망

불특정 화자 인식에 제안된 신경회로망을 적용하기 위해서는 우선 패턴블럭의 크기와 특정 파라메터 내에서 패턴블럭의 변화에 대하여 고찰해 보아야 한다. 한국어 단모음에 대한 포만트 블럭의 크기는 표 1과 같기때문에 패턴블럭의 크기는 약 3KHZ 정도로 가정할 수 있다. 그리고 패턴블럭의 변화패턴은 그림 3에서 볼 수 있듯이 패턴블럭(포만트 주파수 영역)이 화자에 따라 주파수 축을 따라 이동[1][2]하는 것을 볼 수 있다. 그러므로 불특정 화자 인식을 위한 패턴블럭 선택적 신경회로망은 주파수 축의 왜곡이 없는 특징을 사용해야 하므로 필터 뱅크(filter bank)를 사용하는 것이 바람직하다. 필터 뱅크는 FFT를 이용하여 구현하였으며 인간의 청각 신경의 인지 방식을 적용한 멜 스케일(mel scale)을 사용[6]하므로써 그림 3에서 볼 수 있는 화자간의 각 포만트 주파수간의 상호 거리 차이를 어느 정도 흡수할 수 있다. 그리고 필터의 각 밴드(band)를 삼각창(triangular window)을 취하여 경계(boundary) 효과를 줄였다[7]. 또한 블럭매칭 신경회로망은 본 실험에서 사용된 데이터가 다소 미흡한 점을 보안하기 위하여 D. F. Specht가 제안한 확률 신경망[4][5]을 사용하였다. 확률 신경회로망은 Bayes decision rule을 사용하기 때문에 제안된 신경회로망과 같이 선택적 학습을 하는 경우에는 기존의 ML(Maximum Likelihood) 학습 방법을 그대로 사용할 수 있다. 그러므로 선택적 신경회로망의 학습은 다음과 같이 ML을 수정하여 수행할 수 있다.

표 1. 한국어 단모음에 대한 포만트 크기

Hz / vowel	formant size
/u/	about 2,120
/i/	about 2,890
/a/	about 1,760
/æ/	about 2,120
/ə/	about 2,020
/e/	about 2,200
/o/	about 2,120
/ʌ/	about 2,090

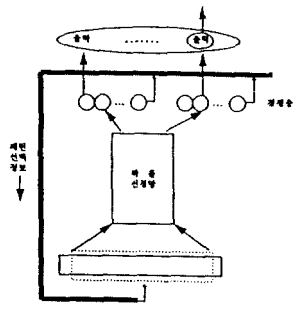


그림 2 PSNN의 구조

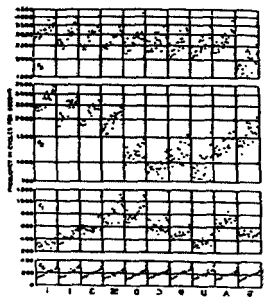


그림 3 영어의 모음에 대한 *F0, F1, F2, F3*

X_k가 신경회로망에 인가되었을 때 모델 φ의 출력확률은

$$f(X_k|\varphi) = f(X_k|\varphi, b=S(X_k)) + f(X_k|\varphi, b\neq S(X_k)) \quad (2)$$

이다. 식 (2)에서 우변의 두 항은 정규화 함수에 선택함수를 사용한 학습의 타당성을 얻을 수 있고 전체비용 함수를 식 (3)과 같이 정의 할 수 있다.

$$C(X_k|\varphi) = \prod_{k=1}^{N} f(X_k|\varphi, b=S(X_k)) \quad (3)$$

그리고 식 (3)의 비용함수를 최대화 하도록 학습함으로써 본 연구에서 설계한 신경회로망을 최적화 시킬 수 있다.

4. 실험 및 결과 고찰

실험에 사용된 특징 파라메터는 표 2와 같은 19차의 멜 스펙트럼을 사용하였고 데이터베이스는 한국어 단모음 7개(/ㅏ/ /ㅓ/ /ㅗ/ /ㅜ/ /ㅡ/ /ㅣ/ /ㅐ/)에 대하여 성인 남녀가 각각 발성한 1960개의 토큰에 대하여 학습을 수행하고 학습에 참여하지 않은 성인 남녀와 어린이가 발성한 1078개의 토큰에 대하여 인식실험이 수행되었다.

표3~4는 입력을 몇개의 부패턴으로 분할하지 않은 기존 신경회로망과 제안된 신경회로망의 비교분류 실험결과이다. 표 5는 성인 남성과 여성 그리고 어린이에 대한 전체 인식 실험결과 이다. 그리고 4는 인식된 패턴분석에 대한 결과이다.

5. 결 론

본 논문은 패턴에 대한 특징 파라메터의 특성을 고려하여 설계되는 패턴블럭 선택적 신경회로망에 관한 연구이다. 본 연구에서는 음성 인식 분야에 제안된 신경회로망을 적용하였지만 문자 인식이나 다른 분야에서 사용하는 특징의 특성에 알맞도록 패턴블럭을 설정하고 그 특성에 알맞게 패턴블럭 신경회로망을 구성하여 응용할 수 있을 것으로 생각된다.

참고문헌

[1] R. K. Potter, J. C. Steinberg, "Toward the Specification of Speech", *Bell Telephone Laboratories, Murray Hill, New Jersey*, August 4, 1950.

[2] R. K. Potter, G. A. Kopp, H. C Green, "Visiable Speech", *D. Van Nostrand Company, Inc.*, 1947.

[3] by Reported *ETRI*, Korea. 1988.

[4] X. D. Huang, Y. Ariki, M. A. Jack, Hidden Markov Models for Speech Recognition, 7th edit., *Edinburgh University Press*, 1990.

[5] D. F. Specht, "Probabilistic Neural Networks", *Neural Networks*, Vol. 3, pp. 109-118, 1990.

[6] L. Rabiner, B. H. Juang, "Fundamentals of Speech Recognition", *Prentice-Hall International, Inc.*, 1993

[7] D. O'Shaughnessy,"*Speech Communication Human and Machine*," Addison Wesley, 1990.

표 2. 멜-스펙트럼 구간(19 order)

Window[1]	0.000000000	(0.092933889) 0.185867779
Window[2]	0.092933889	(0.194504486) 0.296075084
Window[3]	0.185867779	(0.296075084) 0.406282388
Window[4]	0.296075084	(0.416524382) 0.536973680
Window[5]	0.406282388	(0.536973680) 0.667664972
Window[6]	0.536973680	(0.679810622) 0.822647564
Window[7]	0.667664972	(0.822647564) 0.977630156
Window[8]	0.822647564	(0.992033291) 1.161419018
Window[9]	0.977630156	(1.161419018) 1.345207880
Window[10]	1.161419018	(1.362288094) 1.563157169
Window[11]	1.345207880	(1.563157169) 1.781106459
Window[12]	1.563157169	(1.801361334) 2.039565499
Window[13]	1.781106459	(2.039565499) 2.298024539
Window[14]	2.039565499	(2.322044142) 2.604522786
Window[15]	2.298024539	(2.604522786) 2.911021033
Window[16]	2.604522786	(2.939505108) 3.274487429
Window[17]	2.911021033	(3.274487429) 3.637953825
Window[18]	3.274487429	(3.671732171) 4.068976913
Window[19]	3.637953825	(4.068976913) 4.500000000

표 3. 제안된 신경망의 분류 결과

	100						
	0.71	99.29					
			60.01	39.28	0.71		
				99.64	0.36		
					100		
						100	
							100

전체 인식율 = 94.13%(closed test)

표 4. 기존 방식의 분류 결과

	98.93				1.07		
	53.93	45.71		0.36			
			48.57	50.36			1.07
			30.71	66.08	3.21		
				1.43	98.57		
						98.93	1.07
					1.07	2.14	96.79

전체 인식율 = 79.08%(closed test)

표 5. 전체 인식률

남성	여성	어린이	전체
87.07%	86.28%	81.12%	84.78%

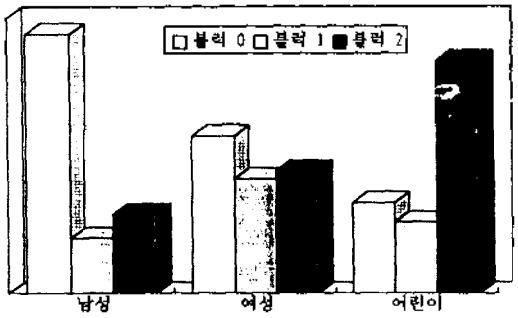


그림 4. 인식된 패턴블럭