

모호성을 포함하고 있는 시계열 패턴인식을 위한 새로운 모델 RFAM과 그 응용

김 원* · 이 중 재** · 김 계 영*** · 최 형 일****

요 약

본 논문에서는 모호성을 포함하고 있는 시계열 패턴인식을 위한 새로운 인식모델인 순환퍼지기억장치를 제안한다. 순환퍼지기억장치는 기존의 퍼지기억장치에 순차적인 입력패턴을 처리하고 시간적 관련성을 표현할 수 있는 순환층을 추가함으로써 확장된 모델이다. 본 논문에서 제안하는 순환퍼지기억장치는 입력과 출력사이의 관련정도를 설정하기 위해 헤비안 방식의 학습알고리즘을 사용한다. 그리고 순환퍼지기억장치의 순환층에 필요한 가중치를 학습하기 위해서 오류역전파 알고리즘을 이용한다. 본 논문에서는 제안하는 모델을 음성신호의 경계를 추출하는 문제에 적용하여 성능을 평가한다.

A Novel Model, Recurrent Fuzzy Associative Memory, for Recognizing Time-Series Patterns Contained Ambiguity and Its Application

Won Kim* · Joongjae Lee** · Gyeyoung Kim*** · Hyungil Choi****

ABSTRACT

This paper proposes a novel recognition model, a recurrent fuzzy associative memory(RFAM), for recognizing time-series patterns contained an ambiguity. RFAM is basically extended from FAM(Fuzzy Associative memory) by adding a recurrent layer which can be used to deal with sequential input patterns and to characterize their temporal relations. RFAM provides a Hebbian-style learning method which establishes the degree of association between input and output. The error back-propagation algorithm is also adopted to train the weights of the recurrent layer of RFAM. To evaluate the performance of the proposed model, we applied it to a word boundary detection problem of speech signal.

키워드 : 순환퍼지기억장치(RFAM ; Recurrent Fuzzy Associative Memory), 시계열 패턴인식(Time-Series Pattern Recognition), 헤비안 학습(Hebbian Learning), 음성경계추출(Word Boundary Detection)

1. 서 론

시계열 패턴인식(time-series pattern recognition)은 시스템 제어와 패턴인식 분야에서 중요한 관심을 모으고 있는 문제로서 동적 시스템 제어, 교통흐름 예측, 기상예측, 음성인식등과 같이 다양한 분야에 적용될 수 있다. 실생활에서 발생하는 방대한 자료처리에 응용하기 위해서 입출력에 대한 학습기능을 가지는 신경망 모델이 많이 사용되어 왔다 [1]. 신경망은 처리할 자료의 시간적 특성을 모델에 반영할 수 있는가 하는 여부에 따라 크게 두 종류로 구분된다. 첫째 자료의 시간적인 특성을 반영하지 않는 정적구조의 신경망이 있다. 둘째 신경망 내부에 시간적 특성을 처리할 수 있는

방법이 있는데 그 대표적인 방법으로 신경망 노드사이에 지연된 피드백(feedback) 연결을 가지는 순환 신경망(recurrent neural network)이 있다[2-4]. 정적구조의 신경망은 정적패턴을 가지는 입력에 대해서는 비교적은 좋은 인식률을 보이거나 시간에 따라 특성변화가 일어나는 시계열 패턴에 대해서는 인식률이 급격히 저하되는 문제점이 있다. 이에 반해서 순환 신경망은 순환부 연결 방법을 통해 입력자료의 시간적 관련성을 표현함으로써 순차적인 입력패턴의 처리가 가능하다. 하지만 순환 신경망 모델 역시 다음과 같은 문제점을 가지고 있다. 첫째, 모호성을 포함하고 있는 시계열 패턴을 처리하기에 부적합하다. 실생활에서 발생하는 대부분의 시계열패턴은 모호성을 포함하고 있는데, 기존의 순환 신경망 모델에서는 이산적인 학습과 추론방식을 사용하기 때문에 입력패턴의 모호성과 오류에 민감한 결과를 보인다. 둘째, 신경망이 가지고 있는 장점인 학습기능을 통해 동적인 특성변화에 적응적으로 대처할 수 있지만 어떤 규칙이 생성되었는

* 본 연구는 숭실대학교 교내연구비 지원으로 이루어졌음.

† 정 회 원 : 전주기전여자대학 실용예술학부 교수

** 준 회 원 : 숭실대학교 대학원 컴퓨터학과

*** 중신회원 : 숭실대학교 컴퓨터학부 교수

**** 중신회원 : 숭실대학교 미디어학부 교수

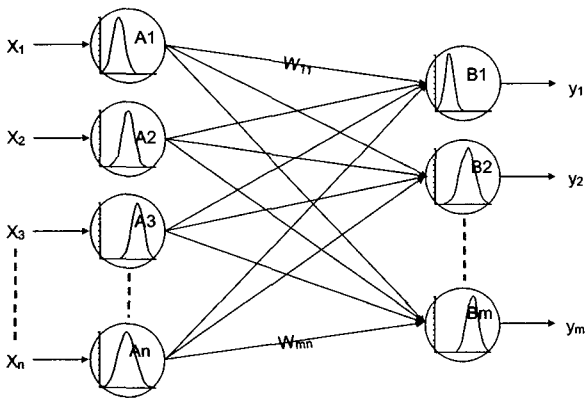
논문접수 : 2004년 3월 10일, 심사완료 : 2004년 5월 24일

지, 생성된 규칙에 어떤 특징이 어느 정도의 기여를 했는가에 대한 규칙생성 정보가 부족하다. 후자의 문제점을 해결하기 위한 기존의 방법으로는 비교적 간단한 형태의 인공 신경망 모델인 연상기억장치와 퍼지집합 이론을 결합시킨 퍼지연상기억장치가 있다[5]. 이것은 모호성이 포함된 자료에 대해 소속함수를 인간 전문가의 도움 없이 시스템이 자동으로 추출하고, 학습 기능을 통해 동적으로 변화하는 환경에 쉽게 적용할 수 있는 퍼지추론시스템이다. 그러나 이 방법 또한 입력자료의 모호성을 처리할 수 있고 학습기능을 통해 생성된 규칙 정보를 획득할 수 있는 반면에 시계열패턴을 처리하기에는 부족하다. 따라서 본 논문에서는 기존의 퍼지연상기억장치에 시간에 동적으로 변화하는 특성을 처리할 수 있는 순환 연결부를 추가한 순환 퍼지연상기억장치를 제안한다. 제안한 모델은 퍼지연상기억장치를 구성하는 조건부층의 출력값을 다시 퍼지화층의 입력으로 피드백 처리하는 순환부층을 삽입하여 현재 시점의 출력이 다음 시점의 결과에 영향을 미치도록 구성되어 있다. 본 논문에서는 제안한 퍼지기억장치를 음성과 비음성사이의 경계를 찾아내는 음성경계 추출문제에 적용하여 성능평가를 수행한다.

본 논문의 구성은 다음과 같다. 2장에서는 본 논문에서 제안한 순환 퍼지기억장치의 구조에 대해서 기술하고, 3장에서는 순환 퍼지기억장치의 학습방법에 대해서 설명한다. 4장에서는 실험결과를 보이며, 5장에서는 결론 및 향후연구에 대해서 기술한다.

2. 순환 퍼지연상기억장치

퍼지연상기억장치는 인간의 두뇌를 모방하여 학습을 통해 관련 있는 개념들을 서로 연결하여 연상해 내는 신경망 모델과, 모호한 개념을 표현하고 처리할 수 있는 퍼지집합 이론을 결합한 것이다.

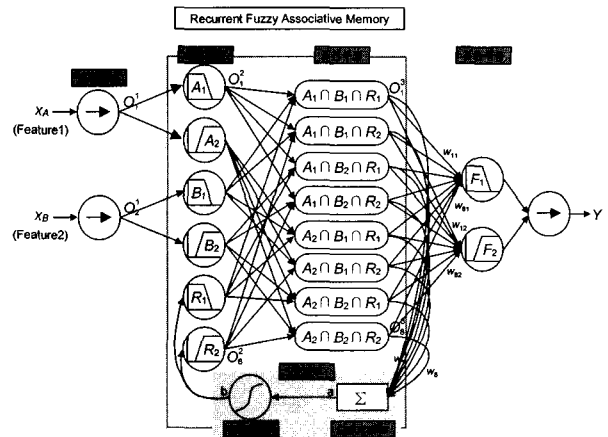


(그림 1) 퍼지연상기억장치

(그림 1)에서 볼 수 있듯이 퍼지연상기억장치는 연상기억 신경망 모델과 매우 유사한 형태를 가지고 있다. 그러나 퍼

지연상기억장치는 각 노드들에 대해서 퍼지집합을 구성하고 입력층과 출력층의 퍼지집합의 결합 정도를 나타내는 가중치의 학습과 입력퍼지집합의 만족도로부터 출력퍼지집합의 만족도를 유도하는 과정에 있어서 연상기억장치 신경망 모델과는 달리 퍼지 최대 연산자와 퍼지 최소 연산자를 이용한다는 차이점이 있다. 그리고 퍼지집합 이론을 사용하기 때문에 모호한 입력패턴을 인식하기에 적합한 장점이 있다. 그런데 앞서 언급 했듯이 퍼지연상기억장치에는 입력자료가 동적으로 변화하는 시간적 특성을 반영하기가 어려운 문제점이 있다. 이러한 문제점을 개선하기 위해서 본 논문에서는 순환 퍼지기억장치를 제안한다.

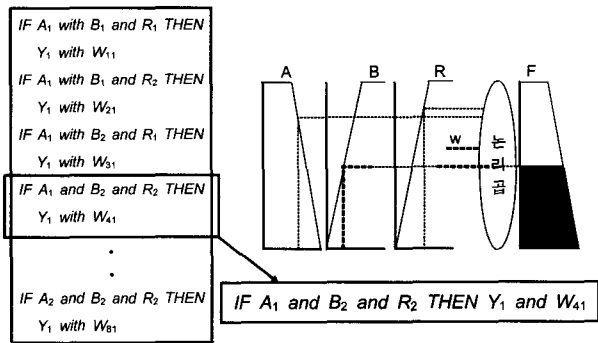
순환 퍼지기억장치는 모호성을 포함한 입력자료를 처리할 수 있는 기존의 퍼지기억장치에 시계열 특징을 동시에 반영할 수 있도록 하기 위하여 순환부층을 추가한 모델로서 (그림 2)는 순환 퍼지연상기억장치의 구조를 보여주고 있다.



(그림 2) 순환 퍼지연상기억장치 구조도

순환 퍼지기억장치는 크게 입력부층, 퍼지화층, 조건부층, 순환부층 그리고 결론부층의 다섯 부분으로 구성된다. 입력부층은 입력 자료에 대한 각각의 특징값을 별도의 처리 없이 해당하는 퍼지화층으로 전달하는 역할을 담당한다. 퍼지화층은 소속함수 수만큼의 노드로 구성된다. 조건부층에서는 퍼지화층으로부터 입력된 특징벡터의 퍼지값을 이용하여 퍼지곱의 형태로 퍼지규칙을 생성한다. 이때, 특히 주목할 점은 시간적 관련성을 표현하는데 필요한 순환부층 노드들이 추가되었다는 것이다. 순환부층은 역퍼지화 함수와 시그모이드 함수로 구성된다. 역퍼지화 함수에서는 조건부층의 출력과 가중치를 곱한 후 모두 더하여 하나의 출력을 획득한다. 시그모이드 함수에서는 역퍼지화 함수의 출력을 0과 1사이의 퍼지값으로 변환하여 퍼지화층의 입력으로 전달한다. 그리고 조건부층은 결론부층과 순환부층으로 동시에 연결된다. 마지막으로 결론부층에서는 조건부층의 퍼지곱 연산을 통해 산출된 결과값을 입력값으로 받아들여 소속함수와의 논리곱을 통해 결과값을 획득한다.

(그림 2)에서 볼 수 있듯이 2개의 특징 X_A, X_B 가 주어지고, 각각의 특징에 대해 n_1, n_2 개의 퍼지집합이 주어졌다고 가정했을 때 생성되는 퍼지규칙의 수는 기존의 퍼지연상기역장치에 n_3 개의 순환 퍼지집합이 추가됨으로써 $n_1 \times n_2 \times n_3$ 개를 가진다. 그리고 O_k^i 는 k 번째 특징에 대한 i 번째 소속함수의 유용도를 나타내며 1에 가까울수록 결론을 명확하게 도출해 낼 수 있는 퍼지집합이라는 의미를 가진다. 그리고 w_{ji} 는 i 번째 조건부와 j 번째 결론부로 이루어지는 퍼지규칙의 신뢰도를 나타낸다.



(그림 3) 퍼지규칙 (그림 4) 퍼지연상기역장치를 이용한 추론모델

(그림 3)은 생성된 퍼지규칙의 한 예를 보여 주고 있으며 (그림 4)에서는 조건부가 A, B, R의 세 퍼지집합의 최소 연산으로, 결론부 집합이 F로 구성된 퍼지규칙을 이용한 퍼지 추론의 예를 보여준다. (그림 4)에서 함수 F의 검정색 영역은 결론부층의 입력값과 소속함수와의 논리곱을 나타내는데 최종 결론값을 얻기 위해 결론부층을 이루는 소속함수의 결과를 가로축을 기준으로 누적시켜 그 무게중심을 구하는 방법을 사용한다.

식 (1)은 전체 결론에 대한 부분적인 결론을 얻는 추론방법을 보여준다. 각 입력 특징값 A, B, R의 소속함수와 각 소속함수의 유용도 O_A, O_B, O_R 와의 곱을 최소 연산에 의해 결합하고 다시 퍼지규칙의 신뢰도 w 와 최소 연산을 수행한다. 그리고 이 결과를 결론부 퍼지집합 F의 소속함수와 논리곱으로 결합하여 부분적인 결론을 획득한다. 식 (1)에서

X_A 는 퍼지집합 A에 대한 입력 특징값, X_B 는 퍼지집합 B에 대한 입력 특징값, X_R 는 퍼지집합 R에 대한 입력 특징값을 나타내며 μ_A 는 퍼지집합 A의 소속함수, μ_B 는 퍼지집합 B의 소속함수, y_k 는 k 번째 규칙의 부분적인 결론, w_k 는 k 번째 규칙의 신뢰도, 그리고 \wedge 기호는 조건부의 만족도와 결론부 퍼지 소속함수와의 논리곱을 의미한다.

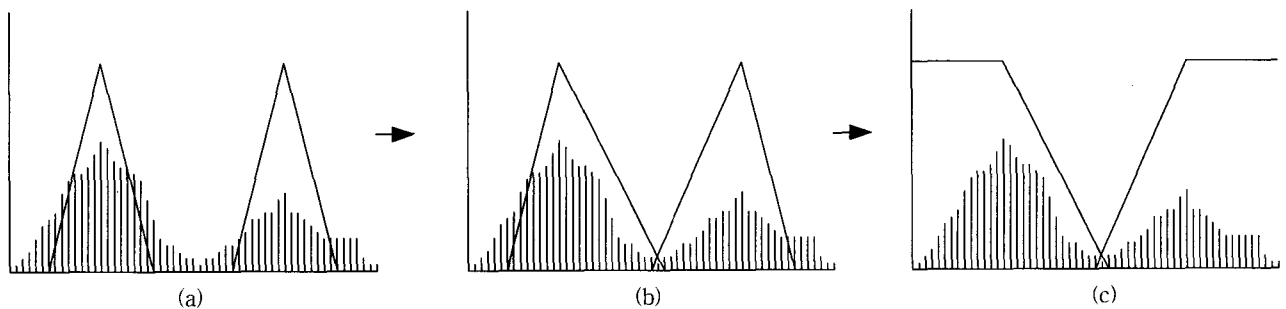
$$y_k = \min(\min(\mu_A(X_A) \times O_A, \mu_B(X_B) \times O_B, \mu_R(X_R) \times O_R), w_k) \wedge \mu_F \quad (1)$$

3. 순환 퍼지기역장치의 학습단계

본 장에서는 순환 퍼지기역장치의 학습단계에 대해서 설명한다. 입력패턴을 인식하기 위해서는 정확하게 퍼지규칙을 자동 생성하는 것이 핵심문제이다. 구체적으로 퍼지 소속 함수의 생성과 학습을 통한 퍼지규칙을 생성하는 것이 필요하다.

3.1 퍼지소속함수 결정

퍼지 소속함수는 퍼지집합의 원소들이 그 퍼지집합에 소속되는 정도를 나타내며 즉, 퍼지규칙을 이루는 각 퍼지집합을 특징짓는 함수이다. 따라서 퍼지규칙의 기본이 되는 가장 중요한 역할을 담당한다. 일반적인 경우에 퍼지 소속함수는 인간 전문가에 의해서 주관적으로 생성되어지지만 복잡한 문제를 해결해야 할 때는 그 문제에 적합한 소속함수를 만들기가 어렵다. 그러므로 본 논문에서는 소속함수를 자동으로 생성하기 위해 특징벡터의 히스토그램을 분석하는 방법을 이용한다. (그림 5)는 특징벡터의 히스토그램에서 두 개의 봉우리(Peak)가 나타나는 경우에 대해 소속함수를 생성하는 예를 보여준다. 소속함수를 생성하기 위한 단계는 다음과 같다. 첫째, 각 특징에 대한 히스토그램을 생성한다. 둘째, 잡음 또는 불규칙한 값들을 제거하기 위해 히스토그램 스무딩을 수행한다. 셋째, 스무딩된 히스토그램으로부터 지역적 최대값 즉 봉우리의 위치를 찾는다. 넷째, 선택된 최대값을 기준으로 삼각형 형태의 함수를 생성하고 (그림 5)(c)와 같이 좌우 함수의 하한과 상한이 1이 되도록 조정하면 2개의 소속함수가 생성된다.



(그림 5) 히스토그램 분석을 이용한 소속함수 생성

3.2 학습에 의한 퍼지규칙 생성

히스토그램 분석 방법을 통해 획득된 퍼지 소속함수를 기반으로 초기 가중치를 결정한 다음에는 학습을 통하여 조건부층과 순환부층, 조건부층과 결론부층을 연결하는 가중치를 결정하게 된다.

먼저 조건부과 결론부층을 연결하는 가중치에 대해서는 헤비안 학습방법을 사용한다[6]. 헤비안 학습은 연결된 두 개의 뉴런들이 동시에 활성화된다면 이들의 연결 강도가 증가한다는 개념을 이용한 학습방법으로 신경망의 생물학적 학습 방법을 모방한 것이다. 식 (2)는 헤비안 학습방법을 이용한 가중치 학습을 나타낸 것으로서 \oplus 기호는 퍼지합 연산을 나타낸다.

$$w_{ij}(new) = \Delta w_{ij} \oplus w_{ij}(old)$$

$$\Delta w_{ij} = \min(\min(A, B, R), y_i) \quad (2)$$

그리고 조건부층과 순환부층을 연결하는 가중치에 대한 학습을 위해서는 오류 역전파 알고리즘을 사용하는데 이것은 학습시에 입력값 X_A, X_B 를 알고 있고 원하는 출력값을 알고 있기 때문에 지도 학습이 가능하다. 식 (3), 식 (4)는 오류 역전파 알고리즘을 수식으로 나타낸 것으로서 목표 출력값 d 와 실제 출력값 b 사이의 오류값 δ 가 임계값 보다 작을 때까지 반복 수행된다. 그리고 식 (4)에서 c 는 학습률로서 값이 클수록 수렴하는 속도가 빨라지지만 정확도는 떨어지게 되는데 본 논문에서는 실험적으로 0.2를 사용하였다.

$$\delta = b(1-b)(b-d) \quad (3)$$

$$w_i = w_i - c\delta O_i \quad (4)$$

4. 실험 및 결과

본 장에서는 본 논문에서 제안한 순환 퍼지연상기억장치를 음성경계 추출에 적용한 실험결과를 기술한다. 실험에 사용한 음성자료는 KAIST에서 제공한 낭독체 형식으로서 음소와 출신지역이 적절하게 분포되어 있다. 또한 연령별로는

<표 1> 음성자료

파일명	성별	연령	샘플링시간(sec)
10Fm	여성	10대	6.575
10Mm	남성	10대	8.375
20FFm	여성	20대	4.375
20Fm	여성	20대	5.195
20MMm	남성	20대	7.0175
20Mm	남성	20대	5.675
30Fm	여성	30대	6.175
30Mm	남성	30대	6.375
40Fm	여성	40대	7.375
40Mm	남성	40대	6.975

10대, 20대, 30대, 40대로 구분되며 성별에 따라 남성과 여성으로 구분된다. <표 1>은 학습에 사용된 자료를 보여주는데 각 자료들은 8KHz로 샘플링되었으며 1회 학습 시 모든 자료가 학습에 한번씩 사용된다.

4.1 특징추출

음성경계 추출 실험을 위해 본 논문에서는 사용한 특징은 시간영역 정보에 대해서 정규화된 RMS(Root Mean Square)와 주파수영역 정보에 대해서는 정규화된 멜주파수 대역 최대에너지를 각각 사용한다. 정규화된 RMS는 음성경계추출에 대해 영교차율과 더불어 가장 많이 사용되는 시간영역 정보로서 영교차율에 비해 음성과 비음성을 잘 구분할 수 있고 계산이 쉽다는 장점이 있다. 그리고 정규화된 멜주파수 대역 에너지는 잡음에 강인하며 인간의 청각적 특성을 잘 반영할 수 있는 특징이다. 본 논문에서는 이렇게 시간영역과 주파수영역의 대표적인 특징의 장점을 동시에 사용함으로써 상호보완적인 효과를 얻을 수 있다. 식 (5)는 주어진 구간에서 시간에 따른 정보의 변화량을 나타내는 RMS를 구하는 식으로서 입력 신호가 0에 가까울수록 RMS는 감소하고 반대의 경우는 RMS가 증가하게 된다. 식 (5)에서 $x_{rms}(m)$ 는 m 번째 프레임의 RMS값을 나타내며 $x_{time}(m, n)$ 은 m 번째 프레임내에서 n 번째 신호의 샘플값을 의미한다. 그리고 L, m, n 은 각각 음성 프레임의 길이, 프레임 인덱스, 프레임 내에서의 신호에 대한 인덱스를 나타낸다.

$$x_{rms}(m) = \log \sqrt{\frac{\sum_{n=0}^{L-1} x_{time}^2(m, n)}{L}} \quad (5)$$

본 논문에서는 보다 정확한 퍼지소속함수를 얻기 위해서 이웃하는 프레임 사이에 스무딩을 수행하고 잡음에 덜 민감한 정규화된 RMS값을 통해 특징벡터의 히스토그램을 획득한다. 식 (6)은 이웃하는 프레임 사이의 정규화 방법을 보여준다.

$$\hat{x}_{rms}(m) = \frac{x_{rms}(m-1) + x_{rms}(m) + x_{rms}(m+1)}{3} \quad (6)$$

식 (7)은 정규화된 RMS를 나타내는데 이것은 잡음이 추가된 RMS값이 잡음이 없는 RMS에 값에 비해 상대적으로 큰 값을 가지기 때문에 정규화를 하지 않고 비교했을 때 발생하는 인식오류를 줄이기 위해서 사용된다. 식 (7)에서 rms_{max} 는 RMS의 최대값을 나타내고 \hat{x}_{rms}^{max} 는 현재 프레임의 \hat{x}_{rms} 값 중 최대값을 의미한다.

$$T(m) = \frac{\hat{x}_{rms}(m) \times rms_{max}}{\hat{x}_{rms}^{max}} \quad (7)$$

음성경계추출에 사용되는 주파수영역 정보는 고속 푸리에 변환된 주파수를 멜주파수로 변환한 후 포먼트 분석을 통하여 획득된다. 여기서, 멜주파수 변환을 사용하는 이유는 고주파 보다는 저주파에 민감한 인간의 청각적 특성을 잘 반영할 수 있기 때문이다[7].

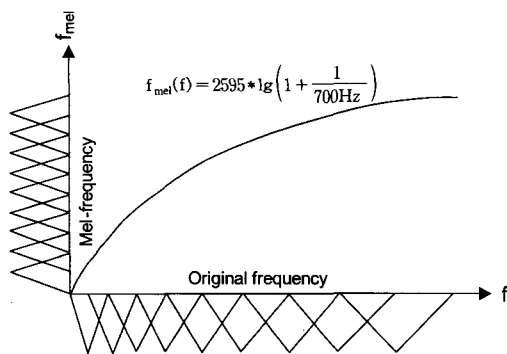
식 (8)은 고속 푸리에 변환을 통해 획득된 원시 주파수 f 를 멜주파수로 변환하는 방법을 보여준다. 식 (8)에서 알 수 있듯이 원시 주파수 f 에 대해 로그 스케일을 취함으로써 고주파에 비해 저주파 부분을 더 많이 샘플링할 수 있게 된다.

$$f_{mel} = 2595 \times \ln \left(1 + \frac{f}{700} \right) \quad (8)$$

이렇게 멜주파수로 변환된 입력 음성으로부터 주파수영역 특징을 추출하기 위해서 포먼트 분석을 사용한다. 포먼트 분석은 신호 전체에 걸쳐 백색잡음이 추가되더라도 음성 고유의 포먼트 정보가 유지되어서 잡음이 있는 경우에도 음성영역과 비음성영역을 잘 구분해 낼 수 있다는 장점이 있다[6]. 따라서, 본 논문에서 사용되는 주파수영역 특징인 멜주파수 대역최대에너지는 각 대역의 최대 주파수 강도를 후보로 하여 이 후보 중에서 최대값을 선택함으로써 특정한 시점을 대표하는 포먼트 정보를 계산한다.

(그림 6)은 고주파에 비해서 저주파영역을 더 강조하는 멜주파수 포먼트 분석에 필요한 멜스케일 필터뱅크를 보여주고 있다.

(그림 6)(a)는 원시 주파수와 멜주파수의 축소 비율에 대한 곡선의 그래프와 필터뱅크를 수행하기 위한 윈도우를 보여준다. 여기서 가변적 윈도우 크기를 갖는 가로축의 원시 주파수를 사용하는 대신 균일한 크기의 윈도우를 갖는 세로축의 멜주파수를 사용하는 것은 멜주파수의 필터뱅크를 수행할 때 구현하기 쉽기 때문이다[7]. 그리고 원시 주파수를 멜주파수로 확대 및 축소한 후에 홀(hole)이 발생한 자료는 이웃하는 정보들을 이용한 선형 보간으로 채워 넣는다.



(a) 멜스케일(mel-scale)

식 (9)는 시간영역의 입력 음성을 주파수영역으로 변환하는 식을 나타내고 식 (10)은 필터뱅크를 수행하는 식을 보여준다. 식 (9), 식 (10)에서 $x_{freq}(m, k)$ 는 m 번째 프레임에서 k 번째 스펙트럼의 크기를 나타내고 $x(m, i)$ 는 m 번째 프레임에서 i 번째 필터뱅크의 크기를, W_N^{kn} 는 고속 푸리에 변환, $f(i, k)$ 는 윈도우를 의미한다.

$$x_{freq}(m, k) = \sum_{n=0}^{N-1} x_{ime}(m, n) W_N^{kn} \quad (9)$$

$$x(m, i) = \sum_{k=0}^{N-1} |x_{freq}(m, k) f(i, k)| \quad (10)$$

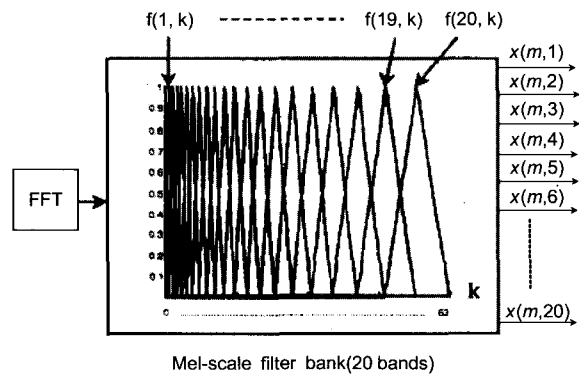
식 (9), 식 (10)을 통해 각 프레임마다 20개의 멜주파수 대역 에너지를 획득하면 식 (6)과 같은 방법으로 식(11)을 이용하여 스무딩을 수행한다.

$$\hat{x}(m, i) = \frac{x_{rms}(m-1, i) + x_{rms}(m, i) + x_{rms}(m+1, i)}{3} \quad (11)$$

지금까지의 방법으로 멜주파수 대역 에너지에 대한 히스토그램을 획득했을 때의 문제점은 음성영역과 비음성영역이 잘 분리되어 있으나 히스토그램이 넓은 범위에 걸쳐 분포되어 있다는 것이다. 따라서 더 구분력이 있는 정확한 퍼지소속함수의 생성을 위해서는 좀 더 좁은 범위를 갖는 히스토그램이 필요하기 때문에 본 논문에서는 대역 에너지의 누적값을 사용하는 대신 프레임 내에서 가장 큰 대역값을 취하는 멜주파수 대역최대에너지를 사용한다. 식 (12)는 멜주파수 대역최대에너지를 구하는 방법을 보여주며 여기서 i 는 필터뱅크의 인덱스를 나타낸다.

$$F_{max}(m) = \max [\hat{x}(m, i)]_{i=1,2,\dots,20} \quad (12)$$

본 논문에서는 최종적으로 주파수영역 정보로 식 (13)와 같은 정규화된 멜주파수 대역 에너지를 사용한다. 식 (13)에서 F_{max} 는 모든 프레임의 $F_{max}(m)$ 값 중 최대값이고 $mfbe_{max}$



(b) 멜스케일 필터뱅크(mel-scale filter bank)

(그림 6) 멜주파수 포먼트 분석

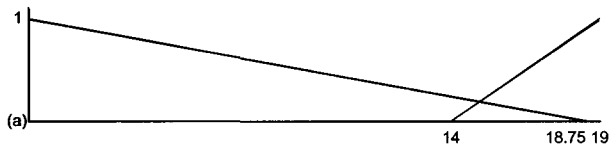
는 정규화를 위한 상수값으로 본 논문에서는 50을 사용한다.

$$F(m) = \frac{F_{\max}(m) \times mfbe_{\max}}{F_{\max}} \quad (13)$$

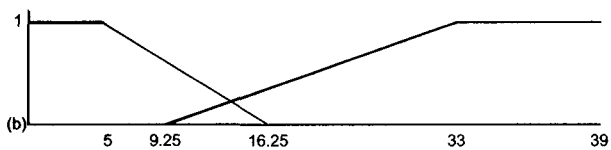
4.2 실험 및 평가

본 절에서는 퍼지연상기억장치를 이용한 음성경계추출에 대한 실험결과와 기존방법과의 성능평가를 보인다.

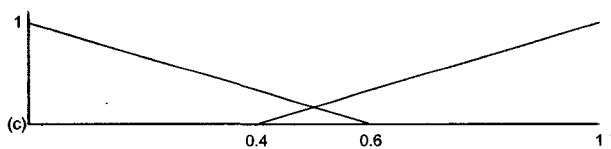
(그림 7)은 히스토그램 분석을 통해 생성된 퍼지 소속함수를 보여주는데 가로축은 각 특징벡터의 값을, 세로축은 소속함수를 통한 퍼지값을 의미한다. (그림 7)(a)는 정규화된 RMS에 대한 소속함수를 (그림 7)(b)는 정규화된 멜주파수 대역최대에너지에 대한 소속함수, 그리고 (그림 7)(c)는 순환노드에 대한 소속함수를 나타낸다.



(a) 정규화된 RMS

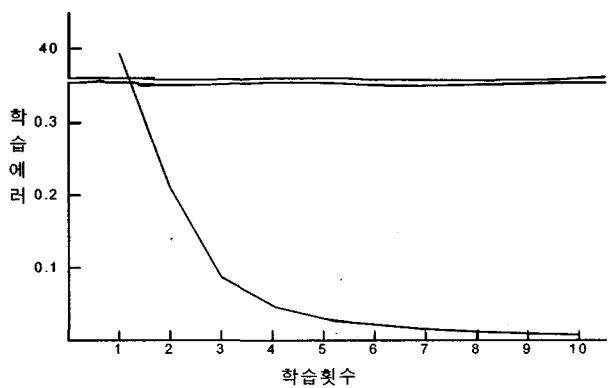


(b) 정규화된 멜주파수 대역최대에너지



(c) 순환부

(그림 7) 생성된 소속함수

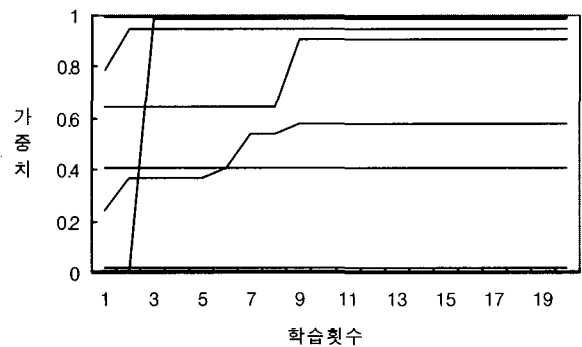


(그림 8) 순환부 가중치 학습 횟수에 따른 오류변화

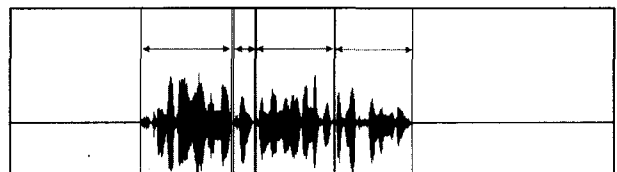
(그림 8)은 학습이 진행됨에 따라 순환부 가중치의 학습 오류율이 변화하는 과정을 그래프로 보여주며 총 50회의 학

습을 수행하였다. (그림 8)에서 볼 수 있듯이 2회 학습부터 오류율이 급격히 감소하여 10회 학습에서부터는 거의 0에 가까운 매우 작은 값을 나타내고 있다.

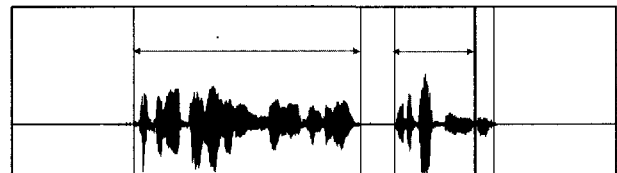
(그림 9)는 총 50회 동안 순환 퍼지연상기억장치의 조건부에서 학습이 진행됨에 따라 조건부 가중치가 변화하는 과정을 그래프로 보여주는데 대략 9회 이후에는 가중치의 변화가 거의 없음을 알 수 있다.



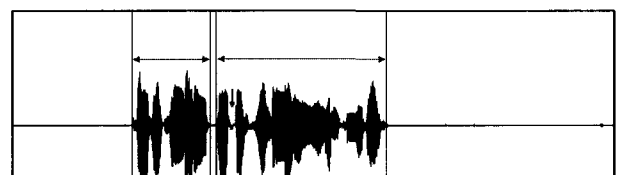
(그림 9) 조건부 학습 횟수에 따른 가중치 변화



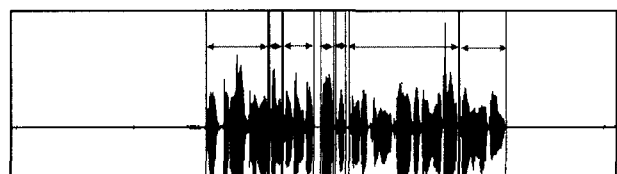
(a) 샘플음성 1



(b) 샘플음성 2



(c) 샘플음성 3



(d) 샘플음성 4

(그림 10) 음성과 비음성 경계추출 결과의 예

이와 같이 학습을 통해 구해진 가중치와 특징벡터의 히스토그램 분석을 이용해 획득한 소속함수를 이용하여 음성 경

계추출에 필요한 퍼지연상기억장치가 마련된 후에 실제 음성자료에 적용한 결과를 (그림 10)에서 볼 수 있다. (그림 10)(a)~(그림 10)(d)는 4개의 샘플음성에 대해 각각 음성영역과 비음성영역을 검출한 결과이며 음성영역을 양끝이 화살표로 된 실선으로, 비음성영역은 음성영역사이의 구간으로 구분하여 보여주고 있다. 특히 (그림 10)(c)에서 ↓기호 부분은 음성을 구성하는 자음부분으로서 RMS값이 작기 때문에 비음성영역으로 오검출할 수 있으나 벨주파수 대역 최대에너지 값이 크므로 음성영역으로 제대로 검출하고 있다.

본 논문에서는 식 (14), 식 (15)에서 정의한 오인식율과 오거부율을 기반으로 음성 경계추출에 대한 성능을 평가한다.

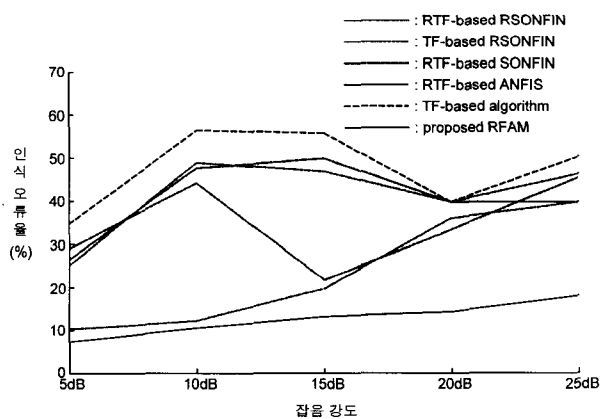
$$\text{오인식율} = \frac{\text{음성으로 인식된 샘플 중 비음성 샘플의 수}}{\text{음성으로 인식된 샘플의 수}} \quad (14)$$

$$\text{오거부율} = \frac{\text{비음성으로 인식된 샘플 중 음성 샘플의 수}}{\text{비음성으로 인식된 샘플의 수}} \quad (15)$$

<표 2>는 제안한 방법을 이용한 음성 경계추출에 대한 실험결과를 보여주고 있는데 약 98%의 정확도를 보인다.

<표 2> 음성경계추출 결과

샘플번호	오인식율(%)	오거부율(%)	정확도(%)
1	1.21	1.1	97.69
2	0.18	1.28	98.54
3	0.45	0.78	98.77
4	1.81	0.52	97.67
5	0.57	0.92	98.51
평균	0.844	0.92	98.27



(그림 11) 인식 오류율 비교

(그림 11)은 잡음에 따른 인식 오류율을 기존 방법과 비교 평가한 그래프를 보여주고 있다. 여기서 가로축은 음성자료에 추가된 잡음의 강도를 나타내며 세로축은 인식 오류율을 나타낸다. 그리고 기존 방법으로는 RTF(refined Time-Frequency)기반 RSONFIN(Recurrent Self-Organizing Neural Fuzzy Inference Network), TF(Time-Frequency)기반 RSONFIN, RTF기반 SONFIN, RTF기반 ANFIS(Adaptive Network based Fuzzy Inference System), TF기반 알고리

즘을 사용하였다[8-11]. (그림 11)에서 알 수 있듯이 실선으로 표시된 제안한 방법이 기존 방법에 비해서 잡음 강도가 증가함에 따라 인식 오류율이 낮고 안정적인 것을 볼 수 있다.

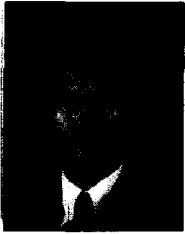
5. 결론 및 향후연구

본 논문에서는 기존의 퍼지연상기억장치에 동적인 시간 특성변화를 처리할 수 있는 순환 연결부를 추가한 순환 퍼지연상기억장치를 제안하였다. 제안한 방법은 퍼지기억장치에 입력자료의 시간적 관련성을 기술할 수 있는 순환부층을 추가함으로써 모호성을 포함하고 있는 시계열 패턴인식에 적합하다. 이것은 시간적 특성을 많이 가지고 있는 음성자료에 대한 경계검출 실험을 통해서 기존 방법에 비해 좋은 성능을 보인다는 것을 통해 나타났다. 본 논문에서 제안한 순환 퍼지 기억연상장치는 모호성을 포함하고 있는 시계열 패턴인식과 시스템 제어와 같은 응용분야에 효율적으로 적용될 수 있으리라 기대된다.

참고 문헌

- [1] Martin T. Hagan, Howard B. Demuth, "Neural Network Design," PWS Publishing Company, 1995.
- [2] J. T. Connor, R. D. Martin and L. E. Atlas, "Recurrent Neural Networks and Robust Time Series Prediction," IEEE Transaction on Neural Networks, Vol.5, No.2, 1994.
- [3] Barak Cohen, Daid Saad and Emanuel Marom, "Efficient Training of Recurrent Neural Network with Time Delays," Neural Networks, Vol.10, No.1, pp.51-59, 1997.
- [4] Alex Aussem, Fionn Murtagh and Marc Sarazin, "Dyanmical Recurrent Neural Networks - Towards Environmental Time Series Prediction," International Journal of Neural Systems, Vol.6, No.2, pp.145-170, 1995.
- [5] 장대식, "퍼지연상기억장치에 기반한 퍼지 추론 시스템," 숭실대학교 석사청구논문, 1995.
- [6] D. O. Hebb, "The Organization of Behavior," John Wiley & Sons, New York, 1949.
- [7] Gin-Der Wu, Chin-Teng Lin, "Word Boundary Detection with Mel-Scale Frequency Bank in Noisy Environment," IEEE Speech and Audio Processing, Vol.8, No.5, pp 541-554, 2000.
- [8] Gin-Der Wu, Chin-Teng Lin, "A Recurrent Neural Fuzzy Network for Word Boundary Detection in Variable Noise-Level Environments," IEEE Systems, Man and Cybernetics, Vol.31, No.1, pp.84-97, 2001.
- [9] J. C. Junqua, B. Mak and B. Reaves, "A robust algorithm for word boundary detection in the presence of noise," IEEE Trans. Speech Audio Processing, Vol.2, pp.406-412, 1994.
- [10] C. F. Juang and C. T. Lin, "An on-line self-constructing neural fuzzy inference network and its application," IEEE Trans. Fuzzy System, Vol.6, pp.12-32, 1998.

[11] J. S. R. Jang, "Self-learning fuzzy controllers based on temporal back propagation," IEEE Trans. Neural Networks, Vol.3, pp.714-723, 1992.



김 원

e-mail : wkim@kijeon.ac.kr
1988년 송실대학교 전자계산학과(공학사)
1993년 송실대학교대학원 컴퓨터학과
(공학석사)
1997년 송실대학교대학원 컴퓨터학과
(공학박사)

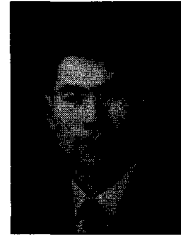
1993년~1995년 송실대학교 전자계산학과 시간강사
1995년~현재 전주기전여자대학 실용예술학부 조교수
관심분야 : 멀티미디어 통신, 멀티미디어 저작권 보호 등



이 중 재

e-mail : ljhop@vision.ssu.ac.kr
1999년 송실대학교 전자계산학과(공학사)
2001년 송실대학교 대학원 컴퓨터학과
(공학석사)
2001년~현재 송실대학교 대학원 컴퓨터
학과 박사과정

관심분야 : 컴퓨터 비전, 증강현실, 의료영상처리 등



김 계 영

e-mail : gykim@computing.soongsil.ac.kr
1990년 송실대학교 전자계산학과(공학사)
1992년 송실대학교대학원 컴퓨터학과
(공학석사)
1996년 송실대학교대학원 컴퓨터학과
(공학박사)

1996년~1997년 한국전자통신연구원(Post Doc.)
1997년~2001년 한국전력공사 전력연구원(선임연구원)
2001년~현재 송실대학교 컴퓨터학부(조교수)
관심분야 : 컴퓨터비전, 형태인식, 생체인식, 증강현실, 영상 및
신호처리 등



최 형 일

e-mail : hic@computing.ssu.ac.kr
1979년 연세대학교 전자공학과(공학사)
1982년 미시간대학교 전산공학과(공학석사)
1987년 미시간대학교 전산공학과(공학박사)
1987년~현재 송실대학교 미디어학부 교수
관심분야 : 컴퓨터비전, 패턴인식, 퍼지이론,
비디오검색, 인터페이스 에이전
트 등