Proceedings of KFIS Autumn Conference 2006 Vol. 16, No. 2.

자기 조직화 신경망을 이용한 음성 신호의 감정 특징 패턴 분류 알고리즘

Emotion Feature Pattern Classification Algorithm of Speech Signal using Self Organizing Map

주종태[1], 박창현[2], 심귀보[1]

[1]서울시 동작구 흑석동 221, 중앙대학교 전자전기공학부
E-mail: kbsim@cau.ac.kr

[2]대전광역시 유성구 가정동 161, 한국전자통신연구원 전파방송연구단 전파기술연구그룹
E-mail: 3rr0r@etri.re.kr

요 약

현재 감정을 인식할 수 있는 방법으로는 음성, 뇌파, 심박, 표정 등 많은 방법들이 존재한다. 본 논문은 이러한 방법 중 음성 신호를 이용한 방법으로써 특징들은 크게 피치, 에너지, 포만트 3가지 특징 점을 고려하였으며 이렇게 다양한 특징들을 사용하는 이유는 아직 획기적인 특징점이 정립되지 않았기 때문이며 이러한 선택의 문제를 해결하기 위해 본 논문에서는 특징 선택 방법 중 Multi Feature Selection(MFS) 방법을 사용하였으며 학습 알고리즘은 Self Organizing Map 알고리즘을 이용하여 음성 신호의 감정 특징 패턴을 분류하는 방법을 제안한다.

Key Words : 감정 인식, Multi Feature Selection, SOM Algorithm, Pattern Classification

1. 서 론

음성은 영상에 비해 잡음에 강하다는 장점 때문에 현재 감정 인식 연구에서 주요 매체로 많이 사용되어지고 있다. 음성 신호를 이용하여 감정을 인식하는데 가장 중요시 되는 것은 특징점를 정립하는 것이며, 특징점를 정립하는 방법에는 크게 특징 선택(Feature Selection)과 특징 추출(Feature Extraction)방법이 있다.

전자의 경우는 어떠한 변환과정도 거치지 않고 보유하고 있는 특징 집합들 중 목적함수에 대해 좋은 성능을 보이는 하위 특징 집합을 선택하는 것이고, 후자의 경우는 보유한 특징 집합을 더 낮은 차원으로 변환하는 방법을 말한다. 특징 추출의 대표적인 예는 Principal Component Analysis(PCA)이고, 특징 선택은 Exhaustive Search, GA, Floating search방법 등이 있다[1]. 특히 Floating search 방법에서 Chul Min과 Shrinkanth는 Forward selection (FS)방법을 사용하였고[2], Dimitrios와 Constantine은 Sequential floating forward selection algorithm을 사용하여 87개의 특징 집합으로부터 10개의 최고 성능을 나타내는 특징들을 선택하여 실험 하였다. Yi-Lin과 Gang 또한 Sequential forward selection을 사용하여 39개의 후보 특징 집합에서 최적의 하위 특징 집합을 선택하였다[3]. 이러한 특징 선택 방법들은 '차원의 저주'에 대한 좋은 해결책이 되었고, 본 논문에서는 SFS방법에 LBG 알고리즘을 접목시킨 Multi Feature Selection(MFS)[4]을 적용시킨다.

이와 같은 방법으로 추출된 특징점들을 통해 패턴을 인식하는 방법에는 통계적인 방법과 기계학습을 이용한 방법이 있으며 통계적 방법은 확률 밀도 함수의 파라미터를 추정하는 모수 밀도 추정 방법과 특정 밀도 함수를 가정하지 않고 클러스터링을 수행하는 비모수 밀도 추정법, Bayes criterion, MAP criterion, ML criterion과 같이 LRT decision rule에서 변형된 방법, HMM등을 이용한다. 기계 학습을 이용한 방법은 뇌의 학습 방법이나 생명체의 진화, 발생 등의 메커니즘을 모방한 것으로써 통계적 방법으로 풀기 어려운 비선형 문제 등을 푸는데 좋은 성능을 보여주며 Artificial Neural Network(ANN), Self Organizing Map(SOM),

Genetic Algorithm(GA), Reinforcement Learning(RL) 등의 방법[5-8]이 있으며, 본 논문에서는 다른 알고리즘에 비해 인식 수행 속도가 빠르며 연속적인 학습이 가능한 SOM 알고리즘을 사용하여 감정 인식을 수행 하였다.

본 논문의 구성은 다음과 같다. 먼저 2장에서 언어적 특성을 배재한 음성 특징 벡터 추출에 대해서 언급하였으며, 3장에서는 본 논문에서 연구한 MFS 특징 선택 방법에 대한 내용, 4장에서는 본 논문에서 사용된 SOM알고리즘의 적용에 대해서 상세하게 기술하였다. 5장에서 Emotion Speech Database 구축에 대한 부분으로써 실제로 감정적 음성 샘플을 수집하기 위한 실험 환경에 대한 내용을 언급하였고, 6장에서 본 논문에서 제안한 알고리즘에 의한 감정 인식을 수행한 결과를 보이고, 마지막으로 7장에서 결론을 맺는다.

2. 음성 특징 벡터 추출

감성인식기는 두 부분으로 구성되어 있다. 첫 번째 부분은 음성으로 특징을 추출하는 부분이고 두 번째 부분은 그 특징들을 이용하는 패턴 인식 부분이다.

음성 신호에서 특징을 추출하기 위해 음성신호를 10ms씩 이동하면서 20ms 길이의 프레임으로 분할한다. 각 프레임의 음성신호에 해밍 윈도우(Hamming Window)를 씌워 음성 신호의 특징을 추출하는데 사용되어 진다. 이를 기반으로 각 프레임별로 고속 푸리에 변환(FFT) 분석을 거쳐 12차의 MFCC(Mel Frequency Cepstral Cofficients) 를 추출하여, 전후 2개씩의 프레임을 참조하여 12차의 차분 MFCC, 에너지, 차분에너지 등 총 26차의 특징 파라미터를 추출했다[7].

3. MFS 특징 선택

음성 신호에서 추출된 26차 특징 파라미터를 모두 감정별 패턴 분류하는데 사용할 경우 연산량이 증가하여 비효적인 시스템이 되므로 이를 해결하기 위해 특징 들을 줄여주는 작업이 필요하다.

본 논문에서는 이러한 문제점을 해결하기 위해서 MFS 알고리즘[4]을 적용하였는데, MFS 알고리즘의 기본 동작은 다음과 같다. 먼저 추출된 26차 파라미터들을 비어있는 집합에 순차적으로 Feature를 추가해본 뒤 최적의 적합도를 보여주는 Feature를 선택하고, 반복하면서 한 개씩 Feature를 추가하게 된다. 이렇게 하여 추출된 결과로 얻어진 벡터 열들을 k-means와 이진 분리를 결합한 LBG(Linde Buzo Gray) 알고리즘에 적용시켜 2개, 3개, 4개, 5개의 특징 벡터열로 분류한다. 다음 그림1은 MFS 알고리즘의 전체 개요를 나타내고 있다.

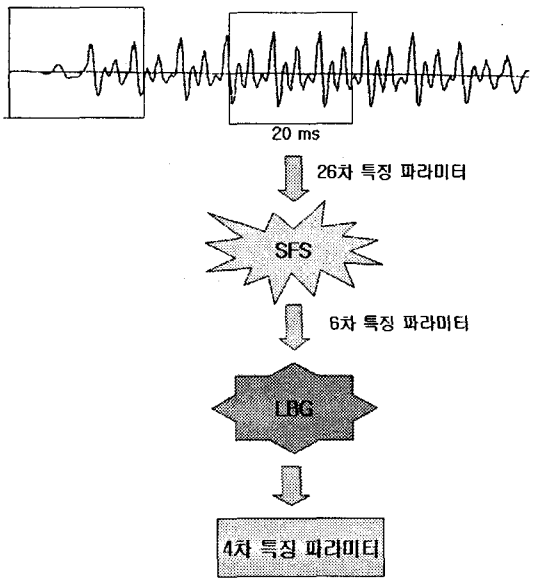

그림 1. Multi-Feature Selection 방법

4. SOM 학습 알고리즘

인간 뇌의 신경 생물학적 원리에 근거한 뇌반응의 메커니즘을 모델링하여 고안된 Self Organizing Map(SOM) 알고리즘은 학습 단계에서 유사한 패턴끼리 2차원의 특징 지도를 조직화하여 영역 지도를 형성 한 후 인식 단계에서 이미 학습 단계에서 훈련된 연결 가중치 하에서 미지의 특징 벡터에 대하여 경쟁층에서 반응이 일어나는 위치를 통하여 해당 클래스를 인식하는 알고리즘이다. SOM 알고리즘은 음성 인식, 이미지 패턴 분류에서 많이 사용되고 있으며 본 논문에서 사용된 SOM의 파라미터들은 다음과 같다.

표 1. SOM 학습 알고리즘의 파라미터

Parameter	Values
Input Units	24 × 24 (576)
Output Units	2
Learning Time	1000
Learning rate	0.04
Weight Units	24 × 24 × 2(1152)
Initialize weights	random

그림 2는 본 논문에서 사용된 SOM 알고리즘의 학습 과정(Flow chart)을 나타낸 것이다.

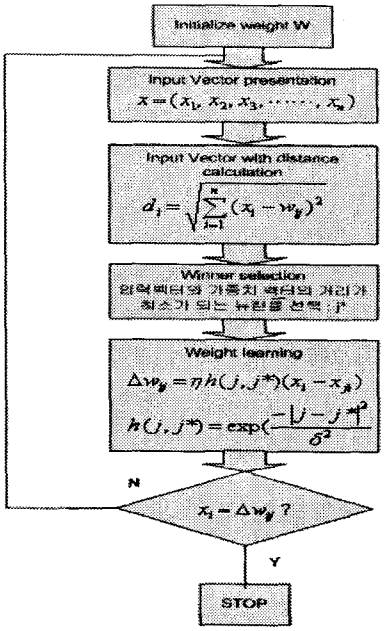

그림 2. SOM을 이용한 학습 과정

5. Emotion Speech Database 구축

15명의 남성과 여성 대학원생들(나이:23~30)에게 6가지 감정으로 총 300개의 음성 샘플을 얻었다. 그들은 보통 한국 남성이며 여러 지역의 출신으로 이루어져 있다. 녹음된 형태는 11KHz, 16bit, mono이고 마이크와 피험자와의 거리를 10Cm로 고정 하였다. 녹음된 문장들은 30개의 일상적이고 단순한 것들이었고 문장의 길이는 2~20음절로 제한해 놓았다. 30개의 미리 준비된 문장들은 그것들을 감정 데이터로 채택해도 될지 확인을 받아야 하기 때문에 녹음한 사람들 이외의 다른 30명에게 "녹음된 소리가 어떤 감정을 포함하고 있는 것 같은가?"라는 질문을 해서 90%의 동의를 얻은 10개의 문장에 대해서 녹음을 하였다.

6. 시뮬레이션 및 결과

본 논문에서는 2개의 여자 음성과 8개의 남성음성을 가지고 음성별 감정을 분류하였다. 각각 10개의 음성에 대해 앞 2장에서 설명한 방법으로 음성 특징 벡터 열을 추출하게 된다. 이렇게 추출된 특징 벡터 열을 SFS 방법을 이용하여 특징 벡터 열을 줄이게 되는데, 다음의 그림 3은 SFS 상태에서 최적의 특징 개수를 나타내고 있으며, 6차의 특징 벡터 열로 줄였을 경우 최고의 성능을 보임을 알 수 있다.

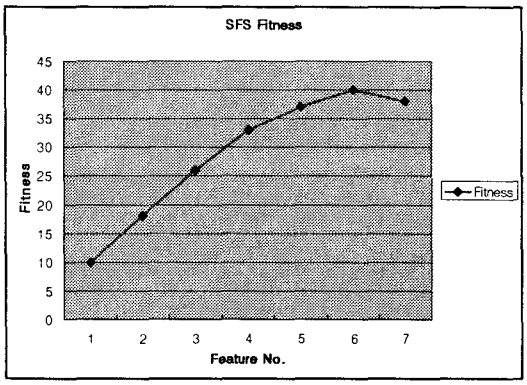

그림 3. SFS 알고리즘 성능 추이 그래프

다음으로는 이 결과로 얻어진 특징 벡터 열들을 k-means의 초기값을 이진 분리로 구한 중심을 사용하여 나타내는 LBG 알고리즘에 적용시켜 최종 벡터 열을 획득하게 되며 그 실험 결과는 그림 4와 같다.

실험 결과를 보면 4개의 특징 벡터 열로 줄였을 경우에 가장 최적의 특징이었음을 알 수 있으며, 4차 이상의 특징 벡터 열을 추출할 경우 더 비효율적임을 알 수 있다.

이와 같이 두 개의 알고리즘을 통해서 특징을 선택함으로써 1개의 특징 선택 방법을 사용하는 경우보다는 좀 더 우수한 특징 점을 추출할 수 있다는 것을 알 수 있었다.

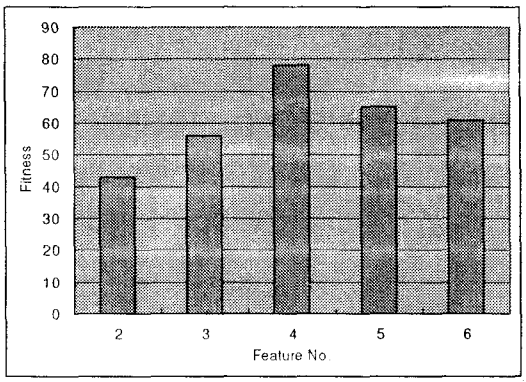

그림 4. LBG 알고리즘 성능 추이 그래프

이렇게 추출되어진 최적의 특징 계수들을 신경망의 한 종류인 SOM에 적용시켜 감정별 패턴을 분류하였으며, 그 결과는 그림 5와 같다.

실험 결과 평상시, 감탄, 화, 슬픔, 놀람, 울분 순으로 감정이 분류됨을 알 수 있었으며, 감정을 분류하는데 문장의 길이는 크게 좌우되지 않는다는 것도 알 수 있었다. 하지만 감정별 남성과 여성을 구별하지는 못했다.

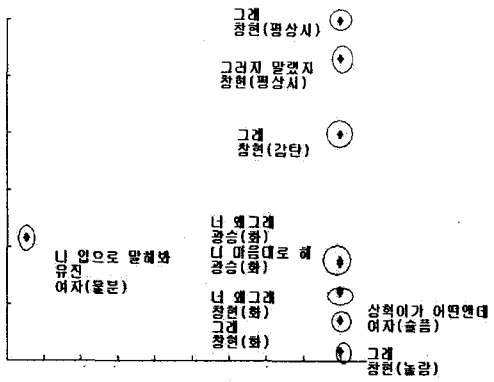

그림 5. SOM학습 알고리즘을 이용한 감정 분류 결과

7. 결 론

본 논문에서 특징점이 많은 패턴 인식의 경우 차원의 저주 문제의 해결책으로 제시될 수 있고, 성능향상에 도움을 줄 수 있는 MFS 알고리즘을 제안하였으며, 본 연구에서 구현한 MFS을 통해 Best Features를 찾아내어 SOM으로 감정 인식을 한 결과 감정별 인식이 잘 되었지만 남성과 여성의 차이는 구별하지 못하였다. 차후에는 더욱 다양한 경우에 대한 결과를 보여 알고리즘의 우수성을 확인하도록 할 것이다.

감사의 글 : 본 논문은 서울시 산학연 협력사업(과제번호 : 106876)에 의해 수행되었습니다. 연구비 지원에 감사드립니다.

참 고 문 헌

[1] P. Pudil and J. Novovicova, "Novel Methods for Subset Selection with Respect to Problem knowledge," *IEEE Intelligent System*, pp. 66-74, March, 1998.

[2] C. M. Lee and S. S. Narayanan, "Toward detecting emotions in spoken dialogs," *IEEE Transactions on Speech and Audio Processing*, vol. 13, pp. 293-303, March, 2005.

[3] Y. L. Lin and W. Gang, "Speech Emotion Recognition based on HMM and SVM," *Proc. of Machine Learning and Cybernetics*, vol. 8, pp. 4898-4901, Aug, 2005.

[4] 윤원중, 이강규, 박규식 "Multi-Feature Clustering을 이용한 강인한 내용 기반 음악 장르 분류 시스템에 관한 연구", *대한전자공학회 논문지*, 제42권, SP 제3호, pp. 393-398. 2005.

[5] 주종태, 김대욱, 심귀보, "신경망을 이용한 DNA칩 영상 패턴 분류 알고리즘", *한국퍼지 및 지능시스템학회 논문지*, 제16권, 제5호, pp. 556-561, 2006.

[6] T. Moriyama and S. Ozawa, "Emotion Recognition and Synthesis System on Speech," *IEEE International Conference on Multimedia Computing and Systems*, vol. 1, 1999.

[7] 김상덕, 이극, "연속분포 HMM을 이용한 음성인식 시스템에 관한 연구", *멀티미디어학회 추계학술대회 논문집*, pp. 221-225, 1998.

[8] C. H. Park, K. S. Byun, and K. B. Sim, "The Implementation of the Emotion Recognition from Speech and Facial Expression System," *Lecture Notes in Computer Science(LNCS) published in Springer*, vol. 3611, pp. 85-88, 2005. 7.