SVM을 이용한 화자의 감정상태 인식

이나라*, 최훈하*, 김현정**, 원일용*
*서울호서전문학교 사이버해킹보안과
**건국대학교 컴퓨터 공학과
e-mail : , lnr32@naver.com, p0si0n@nate.com, nygirl@konkuk.ac.kr, clccclcc@shoseo.ac.kr

Recognition of Emotional State of Speaker Using Machine learning

Na-Ra, Lee*, Hoon-Ha, Choi*, Hyun-jung, Kim**, Il-Young, Won*
*Cyber Hacking Security Seoul Hoseo Technical College
**Dept, of Computer Science and Engineering KonKuk University

요 약

음성을 통한 자동화된 감정 인식은 편리하고 다양한 서비스를 제공할 수 있어 중요한 연구분야라고 할 수 있다. 기계학습의 다양한 알고리즘을 사용하여 감정을 인식하는 연구가 진행되어 왔지만 그 성능은 아직 초보적 단계를 벗어나지 못하고 있는 실정이다. 앞선 연구에서 우리는 비감독 학습 방법으로 감성을 그룹화 하고 이것을 이용하여 다시 감독 학습을 하는 시스템을 소개 하였다. 본 연구에서 우리는 감독 학습 방법에서 사용했던 오류 역전파 알고리즘을 support vector machine(SVM) 으로 변경하고 몇 가지 구조를 변경하여 기능을 개선하였다. 실험을 통하여 성능을 측정하였으며 어느 정도 개선된 결과를 얻을 수 있었다.

1. 서론

삶의 질을 향상 시키기 위한 목적으로 끊임없이 진보해온 인공지능 기술은 점차적으로 우리 생활 속 가까이에서 접할 수 있게 되었다. 자동차나 가전제품은 물론이고 휴대폰과 같은 우리 생활에 밀접한 제품들은 손을 이용하던 인터페이스를 거쳐 음성으로 구동되는 형태의 인터페이스로 진화해가고 있다[1].

이러한 흐름으로 인해 최근 IT 연구의 전체적인 흐름도 PC 나 네트워크 중심이 아니라 사용자 중심의 서비스로 흘러가고 있으며, 사용자 중심의 서비스를 제공하기 위해서는 사용자의 행동은 물론 감정, 기호 등을 종합적으로 파악하여 맞춤형 서비스를 제공하는 데에 중점을 두고 있다[2].

사용자 중심의 서비스를 제공하는데 있어서 음성은 사용자의 감정을 인지하는데 가장 간단한 수단으로 많은 연구가 진행되어 왔다[3]. 여기서 감정이란 기쁨, 슬픔, 즐거움, 노여움, 평상심 등을 의미한다.

음성을 통해 감정을 인식하기 위한 기존 연구는 아직 초보적인 단계이며 주로 신경망을 이용한 기계학습 방법이 시도되고 있다[3]. 기계학습 분야에서 일반적으로 데이터에 대한 어떤 도메인적 특징을 발견하지 못하는 경우 신경망을 주로 사용하지만, 이 방법은 성능에 한계가 있다는 문제를 가지고 있다.

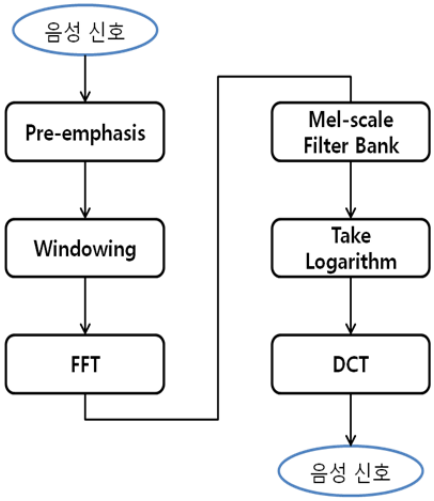
앞선 연구에서 우리는 비감독 학습을 이용하여 감정을 적당한 그룹으로 분리하고, 이 결과를 기반으로 감독학습을 실시하는 방법을 사용하였다[4]. 본 연구에서는 기존에 사용한 감독학습 알고리즘에 오류 역전파 알고리즘 대신 support vector machine(SVM)을 사용하였으며 결합 구조를 변경하여 계산의 양을 줄이고 성능을 향상시키고자 하였다.

본 논문은 다음과 같이 구성된다. 2 장에는 관련 연구를 설명하고, 3 장에서는 본 논문에서 제안하는 시스템을 설명하였다. 4 장에서는 성능 실험을 언급하였으며, 마지막 5 장에서는 결론 및 향후 과제에 대하여 언급하였다.

2. 에이전트 개발도구의 요구사항

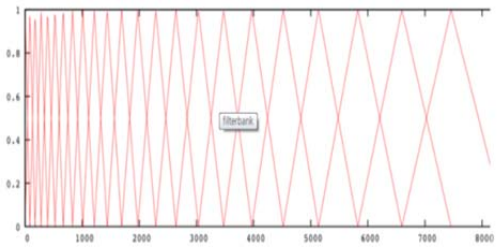
2.1 Mel-Frequency Cepstral Coefficient (MFCC)

MFCC 는 FFT 기반 켑스트럼인 멜 켑스트럼(MFCC)은 음성신호의 대표적 특징 파라미터이다. MFCC 는 인간의 귀가 저주파영역에서 민감하고, 고주파영역에서 둔감한 사실을 이용하여 filter bank 를 통과시킨 것이다. FFT 기반 켑스트럼 계수는 복수 켑스트럼이 Fourier 스펙트럼의 진폭만으로 계산될 수 있도록 최소 위상열로써 음성 신호를 가정함으로써 얻어진다. FFT 기반 파라미터의 장점은 잡음에 강하며 비균일(bark, mel) 스케일로 주파수를 묶을 수 있다는 점이다. MFCC 특징벡터를 얻기 위한 일련의 과정은 (그림 1)과 같다.



(그림 1) MFCC 동작과정

(그림 2)는 멜 스케일(Mel-Scale)필터 뱅크의 예를 보여준다.



(그림 2) Mel scale Filter bank

멜 스케일을 구하기 위해서는 (수식 1)과 같이 정의된다.

$$\mathbf{m = 1127.01048 \log_e(1 + \frac{f}{700})} \quad \text{(수식 1)}$$

2.2 Support Vector Machine (SVM)

최근에 패턴분류에 있어서 각광을 받고 있는 SVM 모델은 1998 년 V.N.Vapnik[6] 에 의해 개발된 통계적 학습 이론으로서 학습데이터와 범주 정보의 학습 진단을 대상으로 학습과정에서 얻어진 확률분포를 이용하여 의사결정함수를 추정한 후 이 함수에 따라 새로운 데이터를 이원 분류하는 것으로, VC(Vapnik-Chervonenkis) 이론이라고도 한다. 특히 SVM 은 분류 문제에 있어서 일반화 기능이 높기 때문에 많은 분야에서 응용되고 있다[7]. 외국의 경우 Tay[5]는 미국 금융지표의 예측을 위해 BP 와 SVM 을 적용한 결과 SVM 이 우수함을 입증했다.

Vapnik[9]이 제안한 기계학습 알고리즘, 즉 SVM 의 기본 원리는 훈련 데이터들을 고차원의 특징공간으로 사상(mapping) 시킨 후 두 분류 사이의 여백(margine)을 최대화 시키는 결정함수(hyperplane)를 찾는 것이다.

SVM 은 (수식 2)와 같이 비선형적 변환함수를 이용하여 입력된 샘플을 고차원의 특징 공간으로 투사하고 학습 데이터에 대한 인식 오류율을 최소화하는 최적 초평면을 찾는다.

$$\mathbf{X: x = (x_1, \ldots, x_n) \rightarrow F: \Phi(x) = (\Phi_1(x), \ldots, \Phi_n(x))} \quad \text{(수식 2)}$$

학습 데이터의 수가 n 일 때, 클래스 레이블 $c_i \epsilon \{1, -1\}$를 가지는 샘플 x_i에 대해 SVM 은 (수식 3)과 같이 샘플과 초평면 사이의 거리를 계산한다.

$$\mathbf{f(x) = \sum_{i=1}^{n} \alpha_i c_i K(x, x_i) + b,\ K(x, x_i) = \Phi(x) \cdot \Phi(x_i)} \quad \text{(수식 3)}$$

(수식 3)의 α_i는 x_i가 초평면을 구성하는 지지벡터일 때에는 0 이 아니며, 커널 함수 $K(x, x_i)$는 비선형 매핑 함수의 내적으로 쉽게 계산된다.

사상(mapping)에 대한 정보가 없더라도 SVM 은 특정 공간에서 커널(Kernel)이라는 내적 함수를 활용하여 원하는 최적의 결정함수를 찾는다. 최적의 결정함수는 지지벡터(support vector)라는 몇 개의 입력 벡터들의 결합으로 나타낸다.

SVM 은 입력벡터 x 를 고차원의 특징공간(high-dimensional feature space)으로 사상(mapping)시킨 후 두 클래스 사이의 마진(margin)을 최대화시키는 분리 경계면을 찾는 것을 목적으로 한다. 이러한 최대마진 분리 경계면(maximum margin hyperplane)은 두 클래스 사이를 최대로 분리한다. 최대마진 분리 경계면에 가장 근접한 훈련 데이터를 support vector 라고 부른다[10].

2.3 코호넨 자기조직화 신경회로망(SOM)

신경 회로망(Neural network)은 신경 세포들(Neurons)이 연결 강도들(weights)을 통하여 상호 연결된 네트워크이다. 연결강도들은 학습을 통하여 적응 및 조정되며 이를 위한 학습법칙은 신경회로망의 성능을 좌우하는 중요한 요소의 하나이다. 학습 법칙에는 감독 학습(Supervised learning)과 비감독 학습(Unsupervised learning)이 있는데 비감독 학습을 사용하는 신경회로망들을 자기조직화 신경망(Self-Organizing neural network)이라고 부른다[8].

코호넨 네트워크의 생성시 층내의 뉴런의 연결강도 벡터(연결 가중치)가 임의값을 가지면서 적합하게 초기화되어야 하고 연결강도 벡터와 입력벡터가 통상 0 에서 1 사이의 정규화된 값을 사용하여야 한다.

코호넨 네트워크의 학습은 경쟁 학습 모델과 같이 승자전취(winner take all)라는 방식을 따르며 이 방식은 승자만이 출력을 낼 수 있으며, 승자와 그의 이웃들만이 그들의 연결강도를 조정할 수 있다. 승자 뉴

런의 연결강도 벡터(연결 가중치)는 입력 벡터(활성화)와 가장 가까운 것이다. 이 규칙은 (수식 4)와 같이 표현된다.

$$\mathbf{w(new)ij = w(old) + (ai - w(old)ij)} \quad \text{(수식 4)}$$

연결강도를 초기화하고 새로운 입력벡터를 제시한다. 그리고 입력벡터와 모든 뉴런들 간의 거리를 계산하고 최소거리에 있는 출력 뉴런을 선택하며, 출력 뉴런의 이웃 뉴런들도 선택한다. 승자 뉴런과 이웃 뉴런들의 연결강도를 조정해준다. 모든 뉴런들이 변화가 없을 때 종료한다.

3. 에이전트 개발도구의 요구사항

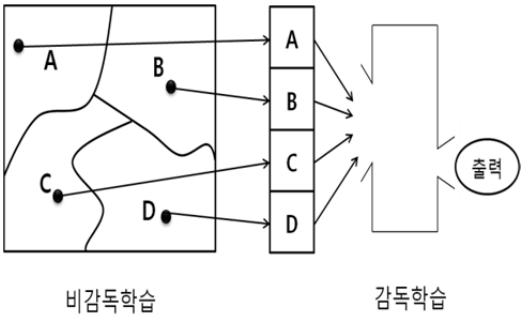
3.1 학습 및 인식

감정인식 시스템의 구성은 다음과 같다. 감정이 포함된 데이터를 전처리 하고 전처리 된 데이터를 비감독 학습으로 세분화 한다. 이렇게 세분화된 데이터를 감독 학습의 입력으로 사용한다. 제안된 시스템의 논리적 흐름은 (그림 3)과 같다. 전처리는 원본 음성데이터에서 학습 요소를 추출하는 단계로 제안한 시스템에서는 음성 분야에 널리 사용되는 MFCC 를 사용한다.

(그림 3) 음성 신호를 이용한 감정인식 순서도

3.2 비감독 학습과 감독 학습의 결합

비감독 학습에 의해 학습용 데이터를 다양한 클러스터로 분류 할 수 있다. 이렇게 분류된 클러스터는 각각의 기초 감정에 대한 다양한 표현으로 이해 될 수 있다. 학습이 끝나고 만들어진 자기조직화 맵에서 각 대표 클러스터의 승자 노드를 구하고, 이 노드와 입력 벡터와의 거리를 측정하여 감독학습의 입력 데이터로 사용한다. 각각의 입력 데이터에 대한 감정 구분은 초기 입력 데이터의 클래스를 따른다.



(그림 4) 비감독 학습과 감독학습의 결합

4. 에이전트 개발도구의 요구사항

4.1 실험 데이터 및 환경

실험에 사용된 데이터는 기존의 BP 와 성능 비교를 위해 앞선 연구에서[3] 사용한 데이터를 그대로 사용하였다. 이 데이터는 분노, 기쁨, 슬픔, 평소의 4 가지 감정 상태를 정의 하였다. 15 명의 화자를 지정하고, 감정당 10 개의 문장을 준비하였다. 우리가 사용한 문장은 논문[3]을 참고하여 지정 하였으며 <표 1>에 예시를 제시 하였다.

<표 1> 감정이 포함된 녹음 문장 예시

기본 감정	문장
평소	어리다는 사실은 제약이 아니라 무기입니다.
기쁨	우리팀이 좋은 성적을 거두었답니다.
슬픔	가슴에 피멍이 드는 그런 이별이였습니다.
분노	너 나를 어떻게 보고 그런 소리를 하는거야?
…	…

녹음된 파일은 44.1khz, 16 bit 단일 채널로 구성되어 있으며, 감정별로 화자가 감정에 잘 몰입할 수 있는 시간에 데이터를 수집하였다. 사용된 알고리즘은 VC++을 사용하여 윈도우즈 환경에서 구현되었다. 특히 SVM 은 오픈소스인 [12,13]를 사용하였으며, 비감독 학습으로는 SOM 을 사용하였다.

4.2 실험 결과 및 분석

실험은 총 10 회 실시하였다. 수집 데이터의 70%를 학습에 사용하였으며, 30%를 테스트에 사용하였다. 비교를 위하여 SOM+BP 결합과, SOM+SVM 결합을 사용하였으며, 10 회에 대한 실험 결과는 <표 2>와 같다.

SOM 과 SVM 의 결합은 SOM 과 BP 의 결합에 비하여 평균 5%정도의 성능 향상을 보이지만, 전체적인 입장에서 볼 때는 만족할 만한 수준이라고 할 수 없다. SOM 이 매 학습 때마다 서로 다른 맵을 만들어 내기 때문에 성능의 유동적인 면이 존재 하며, SOM 이 만든 클러스터의 개수에 따라서도 성능의 차이기 있었다. SOM 과 SVM 의 결합이 SOM 과 BP 의 결합에 비해 개선된 결과를 보이는 것은 BP 가 적용되는 분야에 SVM 이 적용하면 더 좋은 성능을 보인다는

일반적인 원칙이 통용되었기 때문인 것으로 추측된다.

<표 2> 실험 결과

회차	SOM+BP	SOM+SVM
1	57%	60%
2	60%	67%
…	…	..
평균	61.2%	66.3%

5. 결론 및 향후 과제

본 연구에서 우리는 비감학습인 SOM 과 감독학습인 SVM 의 결합을 통한 화자의 감정상태를 자동으로 인식하는 방법에 대하여 언급하였다. 앞선 연구의 성능 개선을 위해 감독 학습 부분을 SVM 을 적용하였다. 또 원 음성의 특징을 추출하기 위한 MFCC 를 사용하는 전처리 부분에서 데이터의 압축률을 낮추어 데이터 손실을 보정하고자 하였으며, 비감독 학습과 감독 학습의 결합 부분을 개선하였다.

실험으로 결과를 분석하였지만 앞선 연구에 비하여 성능의 향상은 만족할 만한 수준은 되지 못했다. 가장 큰 이유는 실험 데이터가 문장 단위를 대상으로 하기 때문에 사람마다 문장을 발음하는 속도도 다르며, 감정을 표현하는 핵심적 단어의 위치도 다르기 때문일 것으로 예상된다. 추후 문장이 아닌 단어 단위로 범위를 축소한 조건에서 실험이 필요하다. 또한 다양한 종류의 MFCC 적용이 필요하며[11], 음성에서 감정적 요소를 추출할 수 있는 음성 전처리 방법을 찾는 것이 필요하다.

참고문헌

[1] 말소리와 음성과학 제 2 권 제 1 호, 잡음 환경에서의 음성 감정 인식을 위한 특징 벡터 처리, 2010

[2] 신동일, “감정인식 기술 동향”, 주간기술동향, 통권 1283 호, 2007

[3] 배상호, 이장훈, 김현정, 원일용, 비감독 학습과 감독 학습의 결합을 통한 음성 감정 인식, 한국정보처리학회, 2011

[4] 한학용 저, 패턴인식 개론 MATLAB 실습을 통한 입체적 학습 개정판, 한빛미디어

[5] F.E.H.Tay, “Application of support vector machines in financial time series forecasting”, Omega 29, 309-317, 2002

[6] J.T.Jeng, “Support Vector Machines for the Fuzzy Neural Networks”, www.kernel-machines.org.

[7] 이수용, 이일병, “Fuzzy 이론과 SVM 을 이용한 KOSPI 200 지수 패턴분류기”, 2002

[8] 김용수, 함창현, 백용선, “가우시안 데이터에 대한 개선된 신경회로망과 코호넨 자기 조직화 특징 지도 성능 비교 연구”

[9] N.Cristianini, “An Introduction to Support Vector Machines”, Cambridge University Press, 2000

[10] 박정민, 김경재, 한인구, “Support Vector Machine 을 이용한 기업부도 예측”, Asia Pacific Journal of Information Systems 제 15 권 제 2 호 (2005. 6) pp.51-63 1229-0270 KCI

[11] T. Ganchev, N. Fakotakis, and G. Kokkinakis (2005), "Comparative evaluation of various MFCC implementations on the speaker verification task," in *10th International Conference on Speech and Computer (SPECOM 2005)*, Vol. 1, pp. 191–194.

[12] http://www.support-vector-machines.org/

[13] http://www.cs.cornell.edu/people/tj/svm_light/