한국음향학회지 제27권 제5호 pp. 252~255 (2008)

변별적 가중치 학습을 적용한 성별인식 알고리즘

Discriminative Weight Training for Gender Identification

강 상 익*, 장 준 혁*
(Sang-Ick Kang*, Joon-Hyuk Chang*)

*인하대학교 전자공학부

(접수일자: 2008년 4월 1일; 수정일자: 2008년 5월 21일; 채택일자: 2008년 6월 27일)

본 논문에서는 성별 인식 시스템의 성능향상을 위해 변별적 가중치 학습 (discriminative weight training) 기반의 최적화된 SVM (support vector machine)을 제안한다. MCE (minimum classification error)방법을 도입하여, 각각의 MFCC (mel-frequency cepstral coefficients) 특징벡터 차수별로 다른 가중치를 가지는 SVM을 제안한다. 제안된 알고리즘은 기존의 동일 가중치를 가지는 SVM 기반의 성별인식 시스템과 비교하였으며, 우수한 성능을 보인다.

핵심용어: 변별적 가중치 학습, 성별인식
투고분야: 음성처리 (2)

In this paper, we apply a discriminative weight training to a support vector machine (SVM) based gender identification. In our approach, the gender decision rule is expressed as the SVM of optimally weighted mel-frequency cepstral coefficients (MFCC) based on a minimum classification error (MCE) method which is different from the previous works in that different weights are assigned to each MFCC filter bank which is considered more realistic. According to the experimental results, the proposed approach is found to be effective for gender identification using SVM.

Keywords: Discriminative Weight Training, Gender Identification
ASK subject classification: Speech Signal Processing (2)

I. 서 론

음성신호를 기반으로 한 성별인식은 음성인식의 성능 향상을 위해 음성인식의 여러 분야에 적용되고 있다. 특히, 음성 부호화나 멀티미디어 신호분석 시스템에서 음성인식 전처리기로서의 성별인식은 성능 향상을 위한 중요한 요소로 작용한다 [1][2]. 실제로, 성별인식은 현재까지 다양한 접근방식으로 다루어지는데 HMM (hidden Markov model), GMM (Gaussian mixture model) 그리고 SVM (support vector machine) 등과 같이 모델 기반의 연구방법이 널리 사용되고 있다 [3][4][6].

본 논문에서는 고차원 공간으로의 확장을 통한 선형 패턴 분류에 있어서 변별적 가중치 학습 (discriminative weight training) 을 이용한 SVM 기반의 효과적인 성별인식 시스템을 제안한다. 단순히 특징벡터로 MFCC (mel-freqyency cepstral coefficients)를 이용하는 것이 아니라 변별적 가중치 학습을 위한 MCE (minimum classification error) 방법을 이용하여 도출된 최적화된 가중치를 MFCC 각각의 차수별로 적용하여 SVM으로 성별을 구분하는 새로운 방식을 제안하며, 기존의 SVM 기반의 방법과 성별인식 성능을 비교하였다.

본 논문의 II장에서는 MFCC 특징 벡터를 이용한 SVM, III 장에서는 변별적 가중치 학습에 대해 살펴본다. 그리고 V장에서는 실험결과를 종합적으로 검토하고 VI장에서 결론을 맺는다.

책임저자: 장 준 혁 (changjh@inha.ac.kr)
인천시 남구 용현동 253
인하대학교 전자공학부
(전화: 032-860-7423; 팩스: 032-868-3654)

II. MFCC 특징 벡터를 이용한 SVM 성별인식

인간이 음성을 인지할 때 각 주파수 성분을 선형적으로 인지하지 않고 비선형적인 Mel 스케일로 음성을 인지한다. Mel 스케일은 사람이 인지하는 톤의 변화를 측정하는 단위로, 사람의 청각 특성을 반영하고 있기 때문에 Mel 스케일과 캡스트럼을 적용한 MFCC 특징 파라미터를 음성 신호 기반의 인식 시스템에서 많이 사용되고 있다. 그림 1은 MFCC 특징벡터를 추출하는 과정을 보여주고 있으

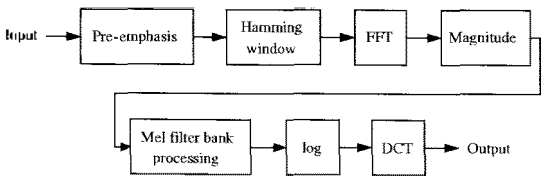

그림 1. MFCC 특징 벡터 추출 과정
Fig. 1. Block diagram of the MFCC extraction algorithm.

며 이 과정을 통해 도출한 최종 결과식은 아래와 같다 [5].

$$c_k = \sum_{m=1}^{M} X'(m)\cos\left(k\frac{\pi}{m}(m \quad \frac{1}{2})\right) \text{ for } k = 1,2,\ldots,M \qquad (1)$$

여기서 c_k와 M은 각각 k차 MFCC 그리고 필터뱅크의 수이다. 본 논문에서는 13차의 MFCC를 특징벡터로 사용하였고 $C = \{c_1, c_2, \ldots, c_M\}$와 같이 정의한다.

SVM은 통계학적 학습 이론에 기반을 둔 SRM (Structural Risk Minimization) 이론으로부터 발전한 이진 패턴 분류기이다 [6]. 선형 SVM에 있어서 positive, negative 클래스를 구분할 수 있는 초평면 (hyperplane)은 무수히 많으나 두 클래스 간 가장 가까운 점들의 거리 ρ (margin)을 최대화하도록 하면 OSH (Optimal Separating Hyperplane)는 유일한 해로 존재한다.

일반적으로 ρ를 최대화하는 초평면의 방정식은 최적 가중벡터 a_*와 바이어스 b_*로 아래와 같이 표현된다.

$$a_*^T \cdot c_k + b_* = 0 \qquad (2)$$

$$\rho = 2/\| a \| . \qquad (3)$$

이때 거리 ρ를 최대화하기 위해서는 Lagrangian의 안장점을 찾는 문제와 KKT (Karush-Kuhn-Tucker) 조건을 이용하여 Lagrange Multiplier를 찾는 Wolfe dual problem으로 바꿀 수 있다. 바뀐 식을 최대화하는 값을 가지고 최적 가중벡터 a_*와 바이어스 b_*를 구할 수 있다 [6]. 구해진 최적가중벡터와 바이어스에 따라 임의의 입력패턴 c는 아래와 같이 분류된다.

$$f(C) = sign(a_*^T C + b_*). \qquad (4)$$

보통의 입력 패턴의 경우 명확하게 선형분리가 되지 않는 경우가 대부분이며 입력 패턴의 선형 분리가 불가능한 경우 비선형 특성을 가진 SVM을 사용한다. 비선형 SVM은 커널 (kernel) 함수를 사용하여 선형 분류가 가능한 고차원 공간으로 확장된 특징 공간을 가지고 ρ을 최대화 하는 값을 찾는다.

고차원의 공간으로 확장시킬 경우 어느 정도는 원 공간에서의 거리 관계를 보존시킬 필요가 있기 때문에 커널 함수는 고차원 공간으로의 사상 함수 $\phi(x)$를 사용해 아래와 같이 정의한다.

$$K(C \cdot C') = \phi(C)^T \phi(C') \qquad (5)$$

여기서 연산 $\cdot$은 두 벡터의 내적을 의미한다.

이때 중요한 점은 커널 트릭 (kernel trick)을 사용하여 사상 함수에 대한 구체적인 설정 없이도 분류함수를 구현할 수 있다는 것이다. 그리고 커널 함수를 사용해서 선형 SVM과 마찬가지의 방법으로 a_*, b_*를 구할 수 있으며, 결론적으로 비선형 SVM은 다음과 같이 분류된다.

$$f(C) = \sum_{k=1}^{M} sign(a_*^T K(c_k \cdot C) + b_*). \qquad (6)$$

III. Discriminative Weight Training

기존의 SVM기반의 성별인식 시스템은 각 MFCC 특징벡터 성분을 동일한 가중치를 이용한 점을 살펴볼 수 있다 [6]. 그러나, 각 MFCC 특징벡터의 차수가 성별인식 성능에 균일한 기여를 한다는 것은 음성신호의 주파수특성의 분포 등을 고려하면 실제적이지 않다. 따라서, 본 논문에서는 각 MFCC 특징벡터 차수별 최적화된 가중치를 인가함으로서 보다 효과적인 성별인식 시스템을 제안하고 새로운 결정식을 다음과 같이 정의한다.

$$f(WC^T) = \sum_{k=1}^{M} sign(a_*^T K(w_k c_k \cdot WC^T) + b_*). \qquad (7)$$

여기서 $W = \{w_1, w_2, \cdots, w_M\}$으로 입력 신호로부터 구한 각 주파수 채널별 우도비에 각각 다른 가중치 w_k를 적용하여 새로운 우도비를 구하며 각 가중치는 다음의 조건을 만족한다.

$$\sum_{k=1}^{M} w_k = 1, \quad w_k \geq 0 \text{ for } k = 1, \cdots, M. \qquad (8)$$

훈련할 데이터의 각각의 프레임에서 남성 $g_m(\cdot)$와 여성 $g_f(\cdot)$을 구분하는 두 개의 함수를 다음과 같이 정

의한다.

$$g_m = f(\boldsymbol{WC}^T) - \theta \tag{9}$$

$$g_f = \theta - f(\boldsymbol{WC}^T) \tag{10}$$

여기서 θ는 남성과 여성을 구분하는 문턱값이며 T는 전치행렬이다. 실제로 식 (9)와 (10)을 $f(\boldsymbol{WC}^T)$와 θ를 비교하는 하나의 수식으로 나타낼 수 있지만 MCE 훈련에서 남성과 여성의 각각의 경우에서 변별적 함수가 필요하므로 두 개의 함수가 필요하다 [7]. 제안된 연구에서는 최적화 알고리즘에 기반한 가중치를 구하기 위해 generalized probabilistic descent (GPD) 기반의 MCE 훈련을 적용하며 [8], 훈련 데이터 프레임의 분류 오류 D를 다음과 같이 정의한다.

$$D(t) = \begin{cases} -g_m(t) + g_f(t) & \text{if } g_m \text{ is true class} \\ -g_f(t) + g_m(t) & \text{if } g_f \text{ is true class} \end{cases} \tag{11}$$

여기서 t는 시간이며, 식 (11)이 음수인 값을 가질 때 올바른 분류가 되며 이를 기반으로 하는 손실함수 (loss function) L은 다음과 같이 sigmoid 함수 형태로 정의된다.

$$L = \frac{1}{1+\exp(-\beta D)}, \quad \beta > 0 \tag{12}$$

여기서 β는 sigmoid 함수의 기울기를 나타낸다. 최적화된 가중치를 구하기 위해선 손실함수가 최소가 되어야한다. MCE 훈련과정을 통해 가중치를 조정하는 과정에서 식 (8)과 같은 제약조건 때문에 가중치 w를 $\tilde{w}$로 변환한다.

$$\tilde{\boldsymbol{W}} = \{\tilde{w}_1, \tilde{w}_2, \cdots, \tilde{w}_k\} \tag{13}$$

$$\tilde{w}_k = \log w_k . \tag{14}$$

가중치 $\tilde{w}_k$는 매 프레임마다 연속적으로 존재하는데, 각 주파수 가중치는 다음과 같은 식으로 갱신된다 [9].

$$\tilde{w}_k(n+1) = \tilde{w}_k(n) - \epsilon \frac{\partial L}{\partial \tilde{w}_k}\bigg|_{\tilde{w}_k = \tilde{w}_k(n)} \tag{15}$$

여기서 ϵ (> 0)는 단조롭게 감소하는 구간의 크기이다. $\tilde{w}_k$를 갱신한 후에 식 (16)과 같이 w_k로 복원된다.

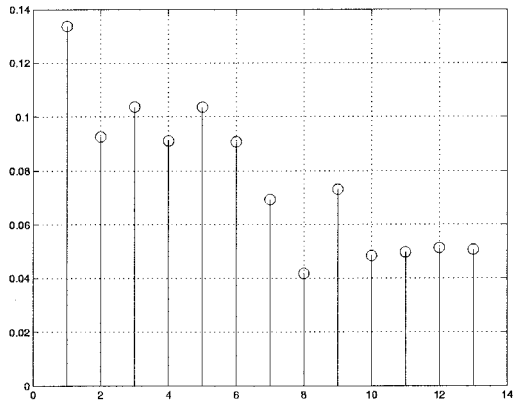

그림 2. 변별적 가중치 학습을 이용하여 도출한 가중치 분포
Fig 2. Weights distribution through discriminative weight training.

$$w_k = \frac{\exp(\tilde{w}_k)}{\sum_{i=1}^{M} \exp(\tilde{w}_i)} \tag{16}$$

식 (16)에서 정규화 된 가중치를 사용했을 때 식 (8)을 만족한다.

기존의 SVM의 결정식과 비교하여, 본 논문에서는 위에서 제시된 MCE 훈련방법을 이용해 구한 식 (16)의 가중치를 MFCC 각 차수에 곱한 후 SVM을 적용하여 식 (7)과 같이 최종적으로 성별을 구분하며 실제 구해낸 13차 가중치 벡터는 그림 2에 도시되어 있다.

IV. 성별인식실험 및 결과고찰

본 논문에서 제시한 알고리즘의 성능 평가를 위해 성별 인식 실험을 실시하였는데, 남, 여의 음성 파일은 OGI database를 사용하였다 [10]. 각각의 파일은 약 5 sec의 음성신호를 담고 있으며, 한사람이 여러 가지 문장과 단어를 영어로 읽는 정보를 담고 있다. SVM 패턴 인식기를 이용한 훈련에 남, 여 각각 500개의 파일을 사용했고, 테스트에는 훈련에 사용하지 않은 새로운 남, 여 각각 500개의 파일을 사용하였으며 인식의 단위는 한 파일이다. 일반적으로 특징벡터들이 선형적이지 않으므로 커널 함수를 이용하여 비선형 특성을 가지게 되는 SVM을 사용하는데 본 논문에서는 Radial-Basis Function (RBF) 커널 함수를 사용하였다 [6]. 또한 최적화된 가중치를 도출하기 위해 손실 함수 L에서 정의된 기울기 파라미터 $\beta = 1$으로 결정하였다.

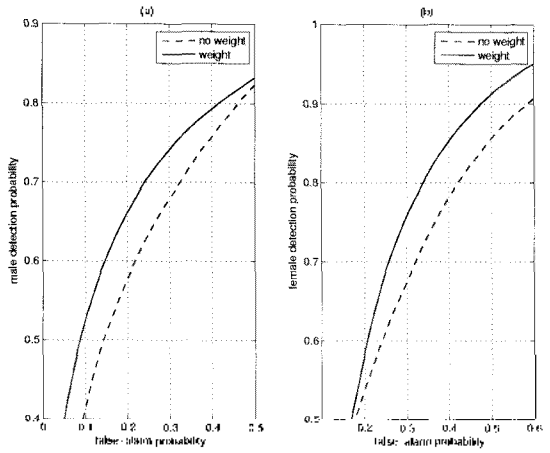

그림 3. ROC에 기반한 SVM과 제안한 방법의 성별인식 성능 비교 (a) 남성 (b) 여성

Fig. 3. Gender identification performance of SVM and proposed method based on ROC. (a) male (b) female.

그림 3은 가중치가 적용된 방법과 적용되지 않은 특징 벡터를 SVM 패턴인식기를 이용하여 도출한 성별 검출확률 (P_d)을 ROC (receiver operating characteristic) 곡선을 이용하여 보여주고 있다. 이때 식 (6)과 (7)의 바이어스 값 b를 변화시키면서 인식 성능을 비교하였으며, 실험 결과를 분석해 보면 변별적 가중치 학습을 이용하여 도출한 가중치를 적용한 실험 결과가 가중치를 적용하지 않은 실험 보다 우수한 성별인식 성능을 보여줌을 확인할 수 있다.

V. 결 론

본 논문에서는 성별인식 성능의 향상을 위해 변별적 가중치 학습을 이용하여 MFCC 특징벡터 각 차수에 서로 다른 가중치를 적용하는 방법을 제시하였고, SVM을 이용한 성별인식 시스템을 구현하여 기존의 성별인식 시스템과 성능을 비교하였다. 객관적인 실험 결과로부터 제안된 방법이 성별인식에서 성능이 우수함을 알 수 있었다.

감사의 글

본 연구는 지식경제부 및 정보통신연구진흥원의 IT핵심기술개발사업 [2008-F-045-01]과 지식경제부 및 정보통신연구진흥원의 대학 IT연구센터 지원사업의 연구결과로 수행되었음 (IITA-2008-C1090-0804-0007).

참 고 문 헌

1. C. Neti and S. Roukos, "Phone-context specific gender-dependent acoustic-models for continuous speech recognition," Proceedings of IEEE Automatic Speech Recognition Understanding Workshop, Santa Barbara, CA, 192-198, Dec. 1997.
2. D. F. Marston, "Gender adapted speech coding," Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 1, 12-15, 357-360, May 1998.
3. H. Harb and L. Chen, "Voice-based gender identification in multimedia applications," Journal of Intelligent Information Systems, 24, 179-198, May 2005.
4. C. Zheng, and B. Z. Yuan, "Text-dependent speaker identification using circular hidden Markov models," Proceeding of IEEE International Conference on Acoustic, Speech, Signal Processing, S 13.3, 580-582, Mar. 1988.
5. S. Sigurdsson, K. B. Petersen, and T. Lehn-Schiofer, "Mel Frequency Cepstral Coefficients: An evaluation of Robustness of MP3 Encoded Music," Proceeding of Int. Conf. Music Inf. Retrieval, 286-289, 2005.
6. 이계환, 강상익, 김덕환, 장준혁, "음성신호 기반의 성별인식을 위한 Support Vector Machines의 적용", 한국음향학회지, 26(2), 75-79, 2월, 2007.
7. Y. Kida, T. Kawahara, "Voice activity detection based on optimally weighted combination of multiple feature," Interspeech, 2621-2624, Sep. 2005.
8. B.-H. Juang, W. Chou, and C.-H. Lee, "Minimum classification error rate methods for speech recognition," IEEE Trans. Speech Audio Processing, 5(3), 257-265, May 1997.
9. S.-I. Kang, Q.-H. Jo, J.-H. Chang, "Discriminative Weight Training for A Statistical Model-Based Voice Activity Detection," IEEE Signal Processing Letters, 15, 170-173, Feb. 2008.
10. Y. K Muthusamy, R. A. Cole and B. T. Oshika, "The OGI multi-language telephone speech corpus," Proceedings of the 1992 International Conference on Spoken Language Processing, 2, 895-898, Oct. 1992.

저자 약력

•강 상 익 (Sang-Ick Kang)

2007년 2월: 인하대학교 전자공학과 학사
2007년 3월 ~ 현재: 인하대학교 전자공학부 석사과정

•장 준 혁 (Joon-Hyuk Chang)

1998년 2월: 경북대학교 전자공학과 학사
2000년 2월: 서울대학교 전기공학부 석사
2004년 2월: 서울대학교 전기컴퓨터공학부 박사
2000년 3월 ~ 2005년 4월: 디지넷시스 연구소장
2004년 5월 ~ 2005년 4월: 캘리포니아 주립대학, 산타바바라 (UCSB) 박사후연구원
2005년 5월 ~ 2005년 8월: 한국과학기술연구 (KIST) 전임연구원
2005년 9월 ~ 현재: 인하대학교 전자공학부 조교수