(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2008-0077719
(43) 공개일자 2008년08월26일

(51) Int. Cl.
G10L 15/08 (2006.01) *G10L 15/28* (2006.01)
G10L 15/00 (2006.01)
(21) 출원번호 10-2007-0017239
(22) 출원일자 2007년02월21일
심사청구일자 2007년02월21일

(71) 출원인
인하대학교 산학협력단
인천 남구 용현동 253 인하대학교
(72) 발명자
장준혁
서울 서초구 서초동 1660-31 B동 301호
이계환
경기 광명시 광명4동 한진아파트 110동902호
강상익
울산 남구 신정1동 1151-20
(74) 대리인
김건우

전체 청구항 수 : 총 1 항

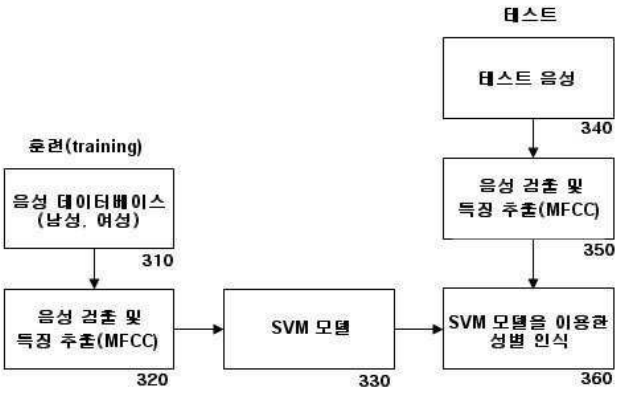
(54) 서포트 벡터머신을 이용한 음성신호에 기초한 성별 인식방법

(57) 요 약

본 발명은 서포트 벡터머신(Support Vector Machine; SVM)을 이용한 음성신호에 기초한 성별 인식 방법에 관한 것으로서, 보다 구체적으로는 (1) 남성 및 여성의 음성을 포함하고 있는 음성 데이터베이스로부터 음성을 검출하는 단계와, (2) 음성 검출 단계에서 검출된 음성으로부터 훈련(training)용 MFCC(Mel Frequency Cepstral Coefficient) 특징 벡터를 추출하는 단계와, (3) 추출한 훈련용 MFCC 특징 벡터를 이용하여 SVM 모델을 생성하는 단계와, (4) 테스트하고자 하는 음성신호로부터 음성을 검출하는 단계와, (5) 음성 검출 단계에서 검출된 음성으로부터 테스트 MFCC 특징 벡터를 추출하는 단계와, (6) 추출한 테스트 MFCC 특징 벡터를 SVM 모델 생성 단계에서 생성한 SVM 모델에 대입하여 성별을 인식하는 단계를 포함하는 음성신호에 기초한 성별 인식 방법에 관한 것이다.

음성신호에 기초한 본 발명의 성별 인식 방법에 따르면, 음성신호에 기초한 성별 인식에 있어서 SVM을 적용함으로써, 기존의 GMM을 이용한 성별 인식 방법과 비교하여 성별 인식의 성능을 대폭 향상시킬 수 있다.

대표도 - 도3



특허청구의 범위

청구항 1

서포트 벡터머신(Support Vector Machine; SVM)을 이용한 음성신호에 기초한 성별 인식 방법으로서,

남성 및 여성의 음성을 포함하고 있는 음성 데이터베이스로부터 음성을 검출하는 단계;

상기 음성 검출 단계에서 검출된 음성으로부터 훈련(training)용 MFCC(Mel Frequency Cepstral Coefficient) 특징 벡터 $X=\{x_1, x_2, \cdots, x_N\}$을 추출하는 단계;

추출한 상기 훈련용 MFCC 특징 벡터를 다음 수학식 1에 적용하여 최적의 가중벡터 w_*와 최적의 바이어스 b_*을 구하여 SVM 모델을 생성하는 단계;

<수학식 1>

$$f(x)=sign(w_*^T K(x_i,x)+b_*)$$

(여기서, T는 전치행렬을 의미하며, 함수 $K(x,x_i)$는 다음 수학식 2와 같이 정의되는 RBF(Radial-Basis Function) 커널 함수임.

<수학식 2>

$$K(x,x_i)=\exp(-\frac{1}{2\sigma^2}\|x-x_i\|^2)$$

여기서, σ는 상기 특징 벡터 집합의 표준편차임.)

테스트하고자 하는 음성신호로부터 음성을 검출하는 단계;

상기 음성 검출 단계에서 검출된 음성으로부터 테스트 MFCC 특징 벡터를 추출하는 단계;

추출한 상기 테스트 MFCC 특징 벡터를 상기 SVM 모델 생성 단계에서 생성한 상기 SVM 모델에 대입하여 성별을 인식하는 단계

를 포함하는 방법.

명 세 서

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

<19> 본 발명은 음성신호에 기초한 성별 인식 방법에 관한 것으로서, 특히 서포트 벡터머신(Support Vector Machine; SVM)을 이용한 음성신호에 기초한 성별 인식 방법에 관한 것이다.

<20> 음성신호에 기초한 성별 인식은 자동음성인식, 멀티미디어 및 인간과 컴퓨터와의 상호작용(Human Computer Interaction; HCI) 등의 성능을 좌우하는 중요한 문제로 다루어져 왔다. HCI를 위해서는 컴퓨터가 사람의 행동, 특히 음성신호를 제대로 인식하고 반응 하는 것이 필수적이며, 효과적인 HCI를 위한 성별 인식에 관한 연구가 활발히 진행되어져 왔다. 현재까지 제안된 성별 인식에 관한 연구는 대부분 HMM(Hidden Markov Model)이나 GMM(Gaussian Mixture Model)과 같은 경험적 위험을 최소화하는 방법에 기초하고 있는 것들이 대부분이다. 그러나 이들 방법은 성별 인식의 성능에 어느 정도 한계를 가지고 있는바, 성별 인식의 성능을 향상시킬 수 있는 새로운 접근법이 필요하다.

<21> 한편, 최근에 패턴 분류에 있어서 각광을 받고 있는 SVM 모델(Support Vector Machine Model)은 1995년 Vladimir Naumovich Vapnik에 의해 개발된 통계적 학습 이론으로서, 학습 데이터와 범주 정보의 학습 진단을 대상으로 학습 과정에서 얻어진 확률 분포를 이용하여 의사결정함수를 추정한 후, 이 함수에 따라 새로운 데이터

를 이원 분류하는 것으로 VC 이론 (Vapnik-Chervonenkis theory)이라고도 한다. 특히, SVM은 분류 문제에 있어서 일반화 기능이 높기 때문에 많은 분야에서 응용되고 있다.

<22> 앞서 소개한 HMM이나 GMM과 같은 기존의 학습 알고리즘은 학습 집단을 이용하여 학습오류(empirical error)를 최소화하는 경험적 위험 최소화 원칙(Empirical Risk Minimization; EMR)을 구현하는 것인데 비하여, SVM은 구조적 위험 최소화 원칙(Structural Risk Minimization; SRM)을 구현한 것이라 할 수 있는데, 이는 전체 집단을 하위 집단으로 세분화한 뒤 이 집단에 대한 경험적 위험도를 최소화하는 의사결정함수를 선택하는 것이다.

<23> SVM은 분류 문제를 해결하기 위한 최적의 분리 경계면(hyperplane; "초평면"이라고도 함)을 제공한다. SVM이 주목받는 이유는 (1) 명백한 이론적 근거에 기초하므로 결과 해석이 용이하고, (2) 실제 응용에 있어서 인공신경망 수준의 높은 성과를 나타내며, (3) 적은 학습 자료만으로 신속하게 분별 학습을 수행할 수 있기 때문이다.

<24> SVM은 처음에 이진 분류(binary classification)를 위하여 개발되었으며, 현재에는 생물정보학(bioinformatics), 문자인식, 필기인식, 얼굴 및 물체인식 등 다양한 분야에서 성공적으로 적용되고 있다.

<25> 이와 같이, SVM은 성능이 우수한 일종의 이진 분류 기법이라고 할 수 있는바, 일종의 이진 분류 방법이라고 볼 수 있는 성별 인식 방법에 SVM을 접목시켜볼 필요성이 있다.

발명이 이루고자 하는 기술적 과제

<26> 본 발명은, 상기와 같은 문제점 및 필요성에 대한 인식에서 비롯된 것으로서, 음성신호에 기초한 성별 인식에 있어서 SVM을 적용함으로써, 성별 인식의 성능을 향상시킬 수 있는 성별 인식 방법을 제안하는 것을 그 목적으로 한다.

발명의 구성 및 작용

<27> 상기한 목적을 달성하기 위한 본 발명의 특징에 따른, SVM을 이용한 음성신호에 기초한 성별 인식 방법은,

<28> (1) 남성 및 여성의 음성을 포함하고 있는 음성 데이터베이스로부터 음성을 검출하는 단계;

<29> (2) 상기 음성 검출 단계에서 검출된 음성으로부터 훈련(training)용 MFCC 특징 벡터 $X=\{x_1, x_2, \cdots, x_N\}$을 추출하는 단계;

<30> (3) 추출한 상기 훈련용 MFCC 특징 벡터를 다음 수학식 1에 적용하여 최적의 가중벡터 w_*와 최적의 바이어스 b_*을 구하여 SVM 모델을 생성하는 단계;

수학식 1

<31> $$f(x)=sign(w_*^T K(x_i,x)+b_*)$$

<32> 여기서, T는 전치행렬을 의미하며, 함수 $K(x,x_i)$는 다음 수학식 2와 같이 정의되는 RBF(Radial-Basis Function) 커널 함수이다.

수학식 2

<33> $$K(x,x_i)=\exp(-\frac{1}{2\sigma^2}\|x-x_i\|^2)$$

<34> 여기서, σ는 상기 특징 벡터 집합의 표준편차이다.

<35> (4) 테스트하고자 하는 음성신호로부터 음성을 검출하는 단계;

<36> (5) 상기 음성 검출 단계에서 검출된 음성으로부터 테스트 MFCC 특징 벡터를 추출하는 단계;

<37> (6) 추출한 상기 테스트 MFCC 특징 벡터를 상기 SVM 모델 생성 단계에서 생성한 상기 SVM 모델에 대입하여 성별을 인식하는 단계를 포함하는 것을 그 특징으로 한다.

<38> 이하에서는 본 발명의 실시예를 도면을 참조하여 상세히 설명한다.

<39> 본 발명의 상세한 설명에서는, 제안된 SVM을 이용한 성별 인식 방법을 기존의 GMM을 이용한 성별 인식 방법과 비교함으로써, 제안된 본 발명의 우수성을 보이려고 한다. 이를 위하여, 먼저 GMM을 이용한 성별 인식 방법에 대하여 간단히 설명한 후, 본 발명에서 제안된 SVM을 이용한 성별 인식 방법을 상세히 설명하기로 한다.

<40> 1. GMM(Gaussian Mixture Model)의 이해

<41> GMM은 EM 알고리즘(Expectation Maximization Algorithm)에 기반을 둔 패턴 분류기이다. 상태 열 N개의 특징 벡터를 $X=\{x_1, x_2, \cdots, x_N\}$, $x^t \subseteq R^d$라 하면, 우도(likelihood)는 다음 수학식 3과 같이 주어진다.

수학식 3

$$p(x_t|\lambda)=\sum_{i=1}^{M} p_i b_i(x_t)$$

<42>

<43> 여기서, $b_i(x_t)$는 다음 수학식 4와 같이 주어진다.

수학식 4

$$b_i(x_t)=\frac{1}{(2\pi)^{\frac{D}{2}}\left|\sum_i\right|^{\frac{1}{2}}} exp\{-\frac{1}{2}(x_t-\mu_i)^T(\sum_i)^{-1}(x_t-\mu_i)\}$$

<44>

<45> 상기 수학식 3 및 수학식 4에서 GMM을 위한 파라미터는 가우시안 혼합 성분 밀도의 가중치(mixture weight: p_i), 평균 벡터(mean vector: $\overrightarrow{\mu_i}$), 공분산 행렬(covariance matrix: $\sum_i$)에 의해 다음 수학식 5와 같이 구성된다.

수학식 5

$$\lambda=\{p_i, \overrightarrow{\mu_i}, \sum_i\} \quad i=1, \cdots, M$$

<46>

<47> 본 발명에서는 파라미터 λ의 추정을 위해, $P(X|\lambda') \geq P(X|\lambda)$가 되는 새로운 모델 λ'을 정해진 임계값(threshold)에 도달할 때까지 EM 알고리즘을 사용하여 i번째 사후 확률을 구한다. 이와 같은 방식으로 구한 사후 확률 중 가장 큰 확률을 가진 성별 인식 모델은 다음 수학식 6과 같이 주어진다.

수학식 6

$$\hat{G}=\arg \max_{1 \leq g \leq S} \sum_{t=1}^{T} \log p(\overrightarrow{x_t}|\lambda_g), \quad S=2(1:\text{남성}, 2:\text{여성})$$

<48>

<49> 이후 단락의 실험 결과에서는, 상기 수학식 6으로 표시되는 GMM을 이용한 성별 인식 결과를 본 발명에 따른 SVM을 이용한 성별 인식 결과와 비교한다.

<50> 2. SVM(Support Vector Machines)의 이해

<51> 앞서 언급한 바와 같이, SVM은 SRM(Structural Risk Minimization) 이론으로부터 발전한 이진 패턴 분류기이다.

<52> 도 1은 선형 분리(linear separation)를 이용한 SVM을 나타내는 도면이다. 도 1에 도시된 바와 같이, 선형 SVM에 있어서 2개의 클래스를 구분할 수 있는 초평면(Hyperplane)은 무수히 많으나, 2개의 클래스(110, 120) 사이에서 가장 가까운 점들의 거리 ρ(마진; margin)(150)을 최대화하도록 하면 유일한 해로 초평면을 구할 수 있다.

<53> 일반적으로 ρ을 최대화하는 초평면의 방정식은 최적의 가중벡터 w_*와 바이어스 b_*로 다음 수학식 7 및 8과 같이 표현된다.

수학식 7

<54> $$w_*^T \cdot x_i + b_* = 0$$

수학식 8

<55> $$\rho = 2/\| w \|$$

<56> 이때 거리 ρ을 최대화하기 위해서는, 다음 수학식 10을 만족하면서 다음 수학식 9를 최소화하여야 한다.

수학식 9

<57> $$\Phi(w) = 1/2 w^T w$$

수학식 10

<58> $$d_i(w^T x_i + b) \geq 1 \quad \text{for} \quad i = 1, \cdots, N$$

<59> 상기 수학식 9를 최소화하기 위해 Lagrangian의 안장점(saddle point)을 찾는 문제와 KKT(Karush-Kuhn-Tucker) 조건을 이용하여 Lagrange Multiplier를 찾는 Wolfe dual problem으로 변화시킬 수 있다. 변화된 식을 최대화하는 값을 가지고 최적 가중벡터 w_*와 바이어스 b_*를 구할 수 있다. 구해진 최적 가중벡터와 바이어스에 따라 임의의 입력패턴 x는 다음 수학식 11에 의해 분류된다.

수학식 11

<60> $$f(x) = sign(w_*^T x + b_*)$$

<61> 그러나 일반적인 입력 패턴의 경우 명확하게 선형 분리가 되지 않는 경우가 대부분이므로, 상기 수학식 11을 그대로 적용할 수는 없다. 즉, 선형 SVM을 비선형 SVM으로 확장할 필요가 있는바, 이에 대해서는 도 2를 참조하여 이하에서 상세하게 설명한다.

<62> 도 2는 선형 분리가 되지 않는 경우에 있어서, 커널 함수를 이용하여 고차원 공간(high dimensional space)으로 SVM을 확장하는 것을 나타내는 도면이다. 도 2에 도시된 바와 같이, 입력 패턴의 선형 분리가 불가능한 경우 비선형 특성을 가진 SVM을 사용하게 되는데, 비선형 SVM은 커널(kernel) 함수를 사용하여 선형 분류가 가능한 고차원 공간으로 확장된 특징 공간(feature space)을 가지고 ρ을 최대화 하는 값을 찾는다.

<63> 고차원의 공간으로 확장시킬 경우 어느 정도는 원 공간에서의 거리 관계를 보존 시킬 필요가 있기 때문에 커널 함수는 고차원 공간으로의 사상 함수 $\phi(x)$을 사용하여 다음 수학식 12와 같이 정의한다.

수학식 12

<64> $$K(x, x_i) = \phi(x)^T \phi(x')$$

<65> 여기서, 중요한 점은 커널 트릭(kernel trick)을 사용함으로써 사상 함수에 대한 구체적인 설정 없이도 분류 함수를 구현할 수 있다는 것이다. 본 발명에서는 다음 수학식 13과 같은 RBF(Radial-Basis Function) 커널 함수를 사용하였다.

수학식 13

<66> $$K(x, x_i) = \exp(-\frac{1}{2\sigma^2} \| x - x_i \|^2)$$

<67> 커널 함수를 사용할 경우, 선형 SVM과 동일한 방법으로 최적 가중벡터 w_*와 바이어스 b_*을 구할 수 있으며, 결론적으로 비선형 SVM은 다음 수학식 14로 정리될 수 있다.

수학식 14

<68> $$f(x) = sign(w_*^T K(x_i, x) + b_*)$$

<69> 3. 성별 인식을 위한 특징 벡터 추출

<70> 음성신호로부터 효과적인 정보를 얻기 위해서는 효율적인 특징 벡터를 추출하는 것이 중요하다. 특히, 성별 인식 성능 향상을 위해서는 효과적인 특징 벡터를 제시된 SVM과 기존의 GMM에 이용하는 것이 바람직하다. 본 발명에서는, 전체적으로 음성 파일에 8 kHz의 샘플링 주파수를 적용하였고, 음성신호의 정보를 얻기 위해서 AURORA2를 이용하여 MFCC 13개와 ΔMFCC 13개를 추출한 후, 음성 검출기(Voice Activity Detector; VAD)를 사용하여 음성 구간의 정보를 추출하였다. 추출한 MFCC는 25ms의 Hamming 윈도우를 15ms씩 이동하면서 계수를 추출하였다.

<71> 4. 본 발명에 따른 SVM을 이용한 성별 인식 방법

<72> 다음으로, 도 3을 참조하여 본 발명에 따른 SVM을 이용한 성별 인식 방법의 구체적인 구현 방법을 설명하기로 한다.

<73> 도 3은 본 발명에 따른 SVM을 이용한 성별 인식 방법의 구체적인 구현 단계들을 나타내는 블록도이다. 도 3에 도시된 바와 같이, 본 발명에 따른 성별 인식 방법은 크게 훈련(training) 단계와 테스트(test) 단계로 나누어진다. 훈련 단계는, 남성 및 여성의 음성을 포함하고 있는 음성 데이터베이스(310)로부터 음성을 검출한 후, 검출된 음성으로부터 훈련용 MFCC 특징 벡터를 추출하는 단계(320)와, 추출한 상기 훈련용 MFCC 특징 벡터를 앞서 언급한 수학식 1에 적용하여 최적의 가중벡터와 최적의 바이어스를 구하여 SVM 모델을 생성하는 단계(330)를 포함한다. 이렇게 SVM 모델이 구해지면, 테스트 단계로 들어간다. 테스트 단계는, 테스트하고자 하는 음성신호(340)로부터 음성을 검출한 후, 검출된 음성으로부터 테스트 MFCC 특징 벡터를 추출하는 단계(350)와 추출한 상기 테스트 MFCC 특징 벡터를 SVM 모델 생성 단계(330)에서 생성한 SVM 모델에 대입하여 성별을 인식하는 단계(360)를 포함한다.

<74> 5. 실험 결과

<75> 실험에 쓰인 남성과 여성의 음성 파일은 OGI database를 사용하였다. 각각의 파일은 약 5초 정도의 전화 음성신호이며, 한 사람이 여러 가지 문장과 단어를 영어로 읽는 정보를 담고 있다. 훈련(Training)은 한 명당 각각 10개의 파일을 선택하도록 하여 남녀 각 10명씩으로 구성하였으며, 테스트는 남성과 여성 각각에 대해 1000개의 파일을 사용하였다. SVM의 경우 수학식 14의 바이어스 b_*을 변화시키면서 인식 성능을 비교하였으며, GMM의 경우 수학식 6에 의해 구해진 사후 확률과 비교되는 임계값을 변화시킴으로써 인식 성능을 비교하였다. 사용된 GMM은 16개의 Mixture를 사용하였으며, SVM에 사용된 특징 벡터들은 모두 평균과 분산으로 정규화하였다.

<76> 도 4는 본 발명에 따른 SVM을 이용한 성별 인식 방법과 기존의 GMM을 이용한 성별 인식 방법의 성별 인식 성능을 비교하는 도면으로서, 남성과 여성의 성별 검출 확률(P_d)을 ROC(Receiver Operating Characteristic) 곡선을 이용하여 비교하는 도면이다. 도 4에 도시된 바와 같이, 기존의 성별 인식 방법에서 널리 쓰이고 있는 MFCC 특징 벡터를 가지고 본 발명에서 제안하고 있는 SVM을 이용하는 성별 인식 방법과 기존의 GMM을 이용하는 성별 인식 방법을 비교한 결과, 대부분의 구간에서 새롭게 제안된 SVM을 이용하는 성별 인식 방법이 GMM을 이용하는 성별 인식 방법보다 좋은 성능을 보였다.

<77> 이상 설명한 본 발명은 본 발명이 속한 기술분야에서 통상의 지식을 가진 자에 의하여 다양한 변형이나 응용이 가능하며, 본 발명에 따른 기술적 사상의 범위는 아래의 특허청구범위에 의하여 정해져야 할 것이다.

발명의 효과

<78> 본 발명의 음성신호에 기초한 성별 인식 방법에 따르면, 음성신호에 기초한 성별 인식에 있어서 SVM을 적용함으로써, 기존의 GMM을 이용한 성별 인식 방법과 비교하여 성별 인식의 성능을 대폭 향상시킬 수 있다.

도면의 간단한 설명

<1> 도 1은 선형 분리(linear separation)를 이용한 SVM을 나타내는 도면.

<2> 도 2는 선형 분리가 되지 않는 경우에 있어서, 커널 함수를 이용하여 고차원 공간으로 SVM을 확장하는 것을 나

타내는 도면.

<3> 도 3은 본 발명에 따른 SVM을 이용한 성별 인식 방법의 구체적인 구현 단계들을 나타내는 블록도.

<4> 도 4는 본 발명에 따른 SVM을 이용한 성별 인식 방법과 기존의 GMM을 이용한 성별 인식 방법의 성별 인식 성능을 비교하는 도면으로서, 남성과 여성의 성별 검출 확률(P_d)을 ROC(Receiver Operating Characteristic) 곡선을 이용하여 비교하는 도면.

<5> <도면 중 주요 부분에 대한 부호의 설명>

<6> 110, 210 : 클래스 1

<7> 120, 220 : 클래스 2

<8> 130, 230 : Support Vectors

<9> 140 : 선형 초평면

<10> 150 : 마진

<11> 240 : 비선형 초평면

<12> 250 : 커널 함수

<13> 310 : (훈련용) 음성 데이터베이스 (남성, 여성)

<14> 320 : (훈련용) 음성인식 및 특징 추출(MFCC)

<15> 330 : SVM 모델

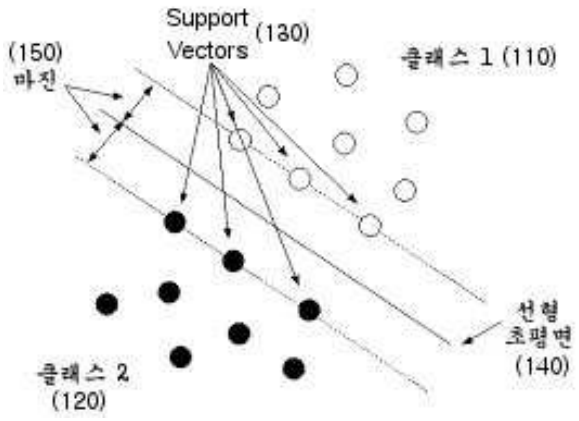
<16> 340 : 테스트 음성

<17> 350 : (테스트) 음성인식 및 특징 추출(MFCC)

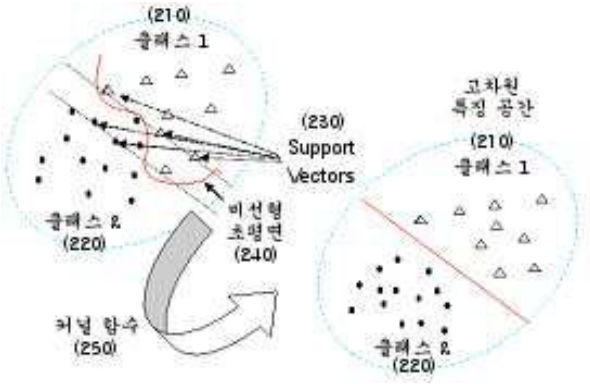
<18> 360 : SVM 모델을 이용한 성별 인식

도면

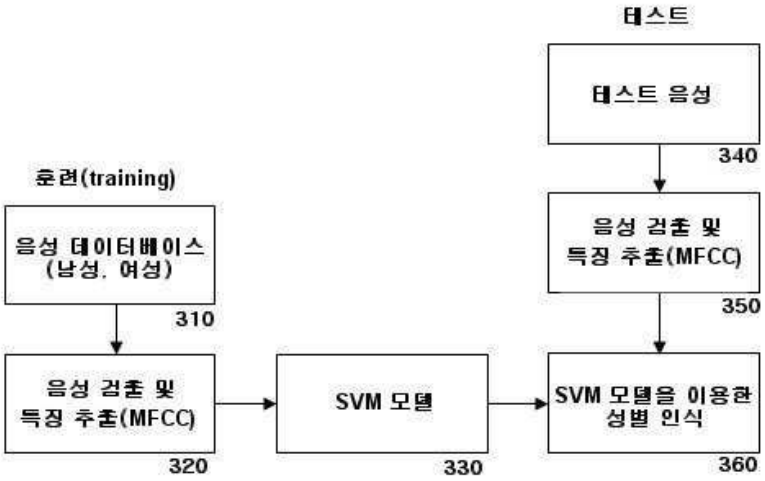
도면1



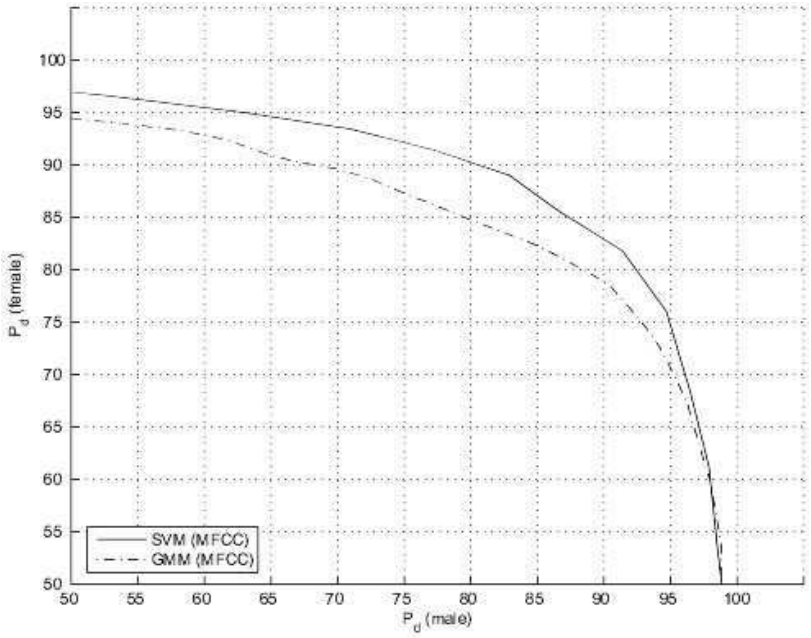
도면2


(210)
클래스 1
(230)
Support
Vectors
고차원
특징 공간
(210)
클래스 1
클래스 2
(220)
비선형
초평면
(240)
커널 함수
(250)
클래스 2
(220)

도면3


테스트
테스트 음성
340
훈련(training)
음성 데이터베이스
(남성, 여성)
310
음성 검출 및
특징 추출(MFCC)
350
음성 검출 및
특징 추출(MFCC)
320
SVM 모델
330
SVM 모델을 이용한
성별 인식
360

도면4


100
95
90
85
80
75
70
65
60
55
50
P_d (female)
50
55
60
65
70
75
80
85
90
95
100
P_d (male)
SVM (MFCC)
GMM (MFCC)