(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2014-0133257
(43) 공개일자 2014년11월19일

(51) 국제특허분류(Int. Cl.)
G10L 25/63 (2013.01) *G10L 25/09* (2013.01)
(21) 출원번호 10-2013-0053176
(22) 출원일자 2013년05월10일
심사청구일자 2013년05월10일

(71) 출원인
서강대학교산학협력단
서울특별시 마포구 백범로 35 (신수동, 서강대학교)
경북대학교 산학협력단
대구광역시 북구 대학로 80 (산격동, 경북대학교)
(72) 발명자
박형민
서울 서초구 잠원로 150, 5동 903호 (잠원동, 잠원한신아파트)
조지원
서울 마포구 독막로32안길 15, 옥탑 (신수동)
(뒷면에 계속)
(74) 대리인
이지연

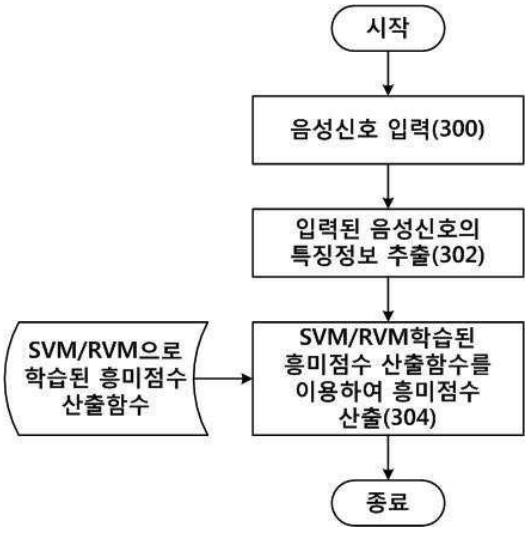
전체 청구항 수 : 총 9 항

(54) 발명의 명칭 **음성신호의 특징정보를 이용한 흥미점수 산출 시스템 및 방법, 그를 기록한 기록매체**

(57) 요 약

본 발명에 따르는 음성신호의 특징정보를 이용한 흥미점수 산출 시스템에 있어서, 질의를 위한 음성신호를 입력받아 특징정보를 검출하고, 상기 질의를 위한 음성신호의 특징정보를 수집된 음성신호들과 그에 대해 미리 설정된 점수정보들을 회귀분석 학습하여 생성한 흥미점수 산출함수에 따라 연산하여 흥미점수를 산출하는 제어장치; 상기 흥미점수 및 상기 흥미점수 산출함수를 저장하며, 상기 제어장치의 프로그램 수행을 위해 요구되는 저장영역을 제공하는 메모리부; 및 상기 제어장치의 제어에 따라 음성신호를 제공받아 특징정보를 생성하는 음성신호의 특징정보 검출부;를 구비함을 특징으로 한다.

대 표 도 - 도6



(72) 발명자

김비호

전남 여수시 도원로 140-1, (안산동)

이민호

대구 수성구 청호로 370, 102동 1203호 (범어동, 장원아파트)

이 발명을 지원한 국가연구개발사업

과제고유번호	10041826
부처명	지식경제부
연구관리전문기관	한국산업기술평가원
연구사업명	산업융합기술산업원천기술개발사업(IT융합)
연구과제명	잠재 역량 진단을 위한 감정특이점 기반 맞춤형 인지센싱 및 플랫폼 기술 개발
기 여 율	1/1
주관기관	한국전자통신연구원
연구기간	2012.06.01 ~ 2013.05.31

특허청구의 범위

청구항 1

음성신호의 특징정보를 이용한 흥미점수 산출 시스템에 있어서,

질의를 위한 음성신호를 입력받아 특징정보를 검출하고,

상기 질의를 위한 음성신호의 특징정보를 수집된 음성신호들과 그에 대해 미리 설정된 점수정보들을 회귀분석 학습하여 생성한 흥미점수 산출함수에 따라 연산하여 흥미점수를 산출하는 제어장치;

상기 흥미점수 및 상기 흥미점수 산출함수를 저장하며, 상기 제어장치의 프로그램 수행을 위해 요구되는 저장영역을 제공하는 메모리부; 및

상기 제어장치의 제어에 따라 음성신호를 제공받아 특징정보를 생성하는 음성신호의 특징정보 검출부;

를 구비함을 특징으로 하는 음성신호의 특징정보를 이용한 흥미점수 산출 시스템.

청구항 2

제1항에 있어서,

상기 제어장치가,

수집된 음성신호들 및 그 수집된 음성신호들 각각에 대해 미리 설정된 점수정보들을 입력받아 저장하고,

상기 수집된 음성신호들의 특징정보들을 검출하고,

상기 특징정보들과 상기 미리 설정된 점수정보들을 회귀분석 학습하여, 그 학습을 통해 흥미점수 산출함수를 생성함을 특징으로 하는 음성신호의 특징정보를 이용한 흥미점수 산출 시스템.

청구항 3

제1항에 있어서,

상기 음성신호의 특징정보는 음성신호의 MFCC이며,

상기 음성신호의 특징정보 생성부는,

음성신호에 대해 윈도우를 설정하는 허밍 윈도우 설정부;

입력되는 음성신호를 시간-주파수 평면의 음성신호로 변환하는 DFT 처리부;

상기 시간-주파수 평면의 신호로부터 시간-주파수 평면에서의 에너지를 구하는 에너지 검출부;

상기 시간-주파수 평면에서의 에너지에 대해 멜 주파수 필터링을 이행하는 멜 필터링부;

상기 멜 주파수 필터링을 통과한 주파수의 로그 연산을 이행하는 로그 연산부; 및

상기 로그 연산부를 통해 산출된 값으로 이산여현변환(discrete cosine transform)을 통해 특징정보를 산출하는 DCT(discrete cosine transform) 처리부;

로 구성됨을 특징으로 하는 음성신호의 특징정보를 이용한 흥미점수 산출 시스템.

청구항 4

제1항에 있어서,

상기 흥미점수 산출함수는 SVM 및 RVM에 따르는 수학식 21 내지 23임을 특징으로 하는 음성신호의 특징정보를 이용한 흥미점수 산출 시스템.

수학식 21

$$y(x) = \sum_{i=1}^{N} w_i k(\mathbf{x}, \mathbf{x_i}) + \mathrm{b}$$

상기 수학식 21에서, y(x)는 흥미점수 산출함수이고, w는 웨이트, k(x, X_i)는 논리니어 커널 평션이고, b는 바이어스이고, x는 질의를 위한 음성신호의 특징정보이고, X_i는 학습을 위한 음성신호들의 특징정보들임.

수학식 22

$$b = t_n - \epsilon - \sum_{m=1}^{N} (a_m - \widehat{a_m}) k(\mathbf{x_n}, \mathbf{x_m})$$

상기 수학식 22는 수학식 21의 b의 산출식으로, b는 바이어스이고, t_n은 입력정보에 대응되는 출력정보, 즉 미리 설정된 점수정보이고, a_m은 라그랑쥬 상수값이고, $\hat{a}_m$은 a_m과 베타적인 라그랑쥬 상수값(둘 중 하나는 0)이고, ϵ은 허용할 수 있는 에러 범위이고, $k(X_n, X_m)$는 비선형 커널 평션이고, X_n,X_m는 입력정보 중 임의의 두 정보임.

수학식 23

$$p(\mathbf{w}|\mathbf{t}, \mathbf{X}, \boldsymbol{\alpha}, \beta) = N(\mathbf{w}|\mathbf{m}, \Sigma)$$

여기서, $\boldsymbol{m} = \beta \Sigma \Phi^T \boldsymbol{t}$ 이고,
$\Sigma = (\boldsymbol{A} + \beta \Phi^T \Phi)^{-1}$

여기서, $\Phi_{ni} = \phi_i(\mathbf{x_n})$임.

상기 수학식 23은 수학식 21의 w의 산출식으로, w는 웨이트이고, X는 입력정보이고, t는 입력정보에 대응되게 설정된 출력정보이고, $\beta = \sigma^{-2}$는 분산(variance)을 의미하고, α, α_n^{-1}은 변수이며, $\Phi_{ni} = \phi_i(\mathbf{x_n})$는 비선형 베이시스 함수이고, A는 α를 대각선(daigonal) 텀으로 가지는 정방행렬임.

청구항 5

음성신호의 특징정보를 이용한 흥미점수 산출 방법에 있어서,

질의를 위한 음성신호를 입력받아 특징정보를 검출하는 단계;

상기 질의를 위한 음성신호의 특징정보를 수집된 음성신호들과 그에 대해 미리 설정된 점수정보들을 회귀분석 학습하여 생성한 흥미점수 산출함수에 따라 연산하여 흥미점수를 산출하는 단계;

를 포함하는 것을 특징으로 하는 음성신호의 특징정보를 이용한 흥미점수 산출방법.

청구항 6

제5항에 있어서,

수집된 음성신호들과 그 수집된 음성신호들 각각에 대해 미리 설정된 점수정보를 제공받는 단계;

상기 수집된 음성신호들의 특징정보들을 검출하는 단계;

상기 특징정보들과 상기 미리 설정된 점수정보들을 회귀분석 학습하고, 그 학습과정에서 흥미점수 산출함수를 생성하는 단계;를 더 구비함을 특징으로 하는 음성신호의 특징정보를 이용한 흥미점수 산출방법.

청구항 7

제5항에 있어서,

상기 음성신호의 특징정보는 음성신호의 MFCC임을 특징으로 하는 음성신호의 특징정보를 이용한 흥미점수 산출 방법.

청구항 8

제5항에 있어서,

상기 흥미점수 산출함수는 SVM 및 RVM에 따르는 수학식 24 내지 26임을 특징으로 하는 음성신호의 특징정보를 이용한 흥미점수 산출 시스템.

수학식 24

$$y(x)=\sum_{i=1}^{N} w_i k(\mathbf{x},\mathbf{x_i})+b$$

상기 수학식 24에서, y(x)는 흥미점수 산출함수이고, w는 웨이트, k(x, x_i)는 논리니어 커널 평션이고, b는 바이어스이고, x는 질의를 위한 음성신호의 특징정보이고, x_i는 학습을 위한 음성신호들의 특징정보들임.

수학식 25

$$b=t_n-\epsilon-\sum_{m=1}^{N}(a_m-\widehat{a_m})k(\mathbf{x_n},\mathbf{x_m})$$

상기 수학식 25는 수학식 24의 b의 산출식으로, b는 바이어스이고, t_n은 입력정보에 대응되는 출력정보, 즉 미리 설정된 점수정보이고, a_m은 라그랑쥬 상수값이고, $\hat{a}_m$은 a_m과 베타적인 라그랑쥬 상수값(둘 중 하나는 0)이고, ϵ은 허용할 수 있는 에러 범위이고, $k(x_n,x_m)$는 비선형 커널 평션이고, X_n,X_m는 입력정보 중 임의의 두 정보임.

수학식 26

$$p(\mathbf{w}|\mathbf{t},\mathbf{X},\boldsymbol{\alpha},\beta)=N(\mathbf{w}|\mathbf{m},\Sigma)$$

여기서, $\boldsymbol{m}=\beta\Sigma\Phi^T\boldsymbol{t}$ 이고,
$\Sigma=(\boldsymbol{A}+\beta\Phi^T\Phi)^{-1}$

여기서, $\Phi_{ni}=\phi_i(\mathbf{x_n})$임.

상기 수학식 26은 수학식 24의 w의 산출식으로, w는 웨이트이고, X는 입력정보이고, t는 입력정보에 대응되게 설정된 출력정보이고, $\beta=\sigma^{-2}$는 분산(variance)을 의미하고, α,α_n^{-1}은 변수이며, $\Phi_{ni}=\phi_i(\mathbf{x_n})$는 비선형 베이시스 함수이고, A는 α를 대각선(daigonal) 텀으로 가지는 정방행렬임.

청구항 9

제5항 내지 제8항 중 어느 한 항의 방법을 실행하기 위한 프로그램을 기록한 컴퓨팅 장치에 의해 판독 가능한 기록매체.

명 세 서

기 술 분 야

[0001] 본 발명은 음성신호로부터 특징정보를 검출하는 기술에 관한 것으로, 더욱 상세하게는 수집된 음성신호들로부터 검출한 특징정보들을 상기 음성신호들 각각에 미리 설정된 T 점수에 대응되게 SVM(Support Vector Regression) 및 RVM(Relevance Vector Machines) 학습하고, 그 학습과정에서 추출한 흥미점수 산출함수를 기반으로 화자의 음성신호에 대해 흥미점수를 산출하는 음성신호의 특징정보를 이용한 흥미점수 산출 시스템 및 방법, 그를 기록한 기록매체에 관한 것이다.

배 경 기 술

[0002] 음성은 인간의 가장 자연스러운 의사소통 수단이면서 정보전달 수단이자, 언어를 구현하는 수단으로서 인간이 내는 의미있는 소리이다. 이에 음성을 통해 화자의 감정 상태의 추정이 가능하다.

[0003] 좀 더 설명하면, 사람은 다른 사람의 음성만을 듣고도 그 감정 상태를 어느 정도 추정하는 것이 가능하다. 이는 감정이 격앙되고 분노한 화자의 발화는 평상시의 상태인 경우에 비해 음 높이가 높아지거나 에너지가 커지는 등의 특징이 있기 때문이다.

[0004] 이에 상기한 음성신호의 특징을 검출하여 화자의 감정을 인식하는 기술이 제안되고 있다.

[0005] 예를들어 대한민국 특허청에 감정 인식 장치 및 방법, 이를 이용한 콜센터 시스템을 명칭으로 하여 특허 출원된 제1020080078708호는 [음성 신호를 입력받는 입력부; 입력된 상기 음성 신호로부터 특징 벡터(feature vector)를 추출하는 특징 벡터 추출부; 상기 학습 음성 신호에 대하여 하나 이상의 감정 상태에 대한 특징 파라미터가 저장된 데이터베이스; 상기 음성 신호가 상기 대상 음성 신호인 경우 활성화되며, 상기 추출된 특징 벡터를 기초로 하여 상기 특징 파라미터를 HMM(Hidden Markov Models)에 적용하여 상기 감정 상태별 출력 확률값을 획득하는 HMM 인식부; 및 상기 출력 확률값을 기초로 하여 상기 음성 신호의 감정 상태를 판단하는 감정 판단부를 포함하는 감정 인식 장치]를 개시한다.

[0006] 그리고 대한민국 특허청에 감정 인식 장치 및 방법을 명칭으로 하여 특허출원된 제1020090047608호는 [입력되는 음성신호를 스펙트로그램으로 변환하는 스펙트로그램 변환부와, 상기 변환된 스펙트로그램에서 모음 성분을 추출하는 제로 크로싱 검출부와, 상기 추출된 모음 성분을 시간 축 정보와 주파수 축 정보로 분리한 후 상기 주파수 축 정보의 벡터 성분을 트레이닝 데이터베이스에 저장하는 다변수 데이터 분리부와, 저장되는 주파수 축 정보의 벡터 성분에 대해 매칭 검사를 실시하고, 상기 매칭 검사의 실시 결과에 따른 감정 인식 결과 데이터를 출력하는 매칭부를 포함하는 감정 인식 장치]를 개시한다.

[0007] 상술한 바와 같이 음성신호에는 언어 이외에도 감정 등 다양한 정보를 내포하고 있으므로, 상기 음성신호로부터 더욱 다양한 정보를 획득해내기 위한 노력이 계속되고 있다.

선행기술문헌

특허문헌

[0008] (특허문헌 0001) 대한민국 특허출원번호 제1020090047608호

(특허문헌 0002) 대한민국 특허출원번호 제1020080078708호

발명의 내용

해결하려는 과제

[0009] 본 발명은 수집된 음성신호들로부터 검출한 특징정보들을 상기 음성신호들 각각에 미리 설정된 T 점수에 대응되게 SVM(Support Vector Regression) 및 RVM(Relevance Vector Machines) 학습하고, 그 학습과정에서 추출한 흥미점수 산출함수를 기반으로 화자의 음성신호에 대해 흥미점수를 산출하는 음성신호의 특징정보를 이용한 흥미점수 산출 시스템 및 방법, 그를 기록한 기록매체를 제공하는 것을 목적으로 한다.

과제의 해결 수단

[0010] 상기한 목적을 달성하기 위한 본 발명에 따르는 음성신호의 특징정보를 이용한 흥미점수 산출 시스템에 있어서, 질의를 위한 음성신호를 입력받아 특징정보를 검출하고, 상기 질의를 위한 음성신호의 특징정보를 수집된 음성신호들과 그에 대해 미리 설정된 점수정보들을 회귀분석 학습하여 생성한 흥미점수 산출함수에 따라 연산하여

흥미점수를 산출하는 제어장치; 상기 흥미점수 및 상기 흥미점수 산출함수를 저장하며, 상기 제어장치의 프로그램 수행을 위해 요구되는 저장영역을 제공하는 메모리부; 및 상기 제어장치의 제어에 따라 음성신호를 제공받아 특징정보를 생성하는 음성신호의 특징정보 검출부;를 구비함을 특징으로 한다.

발명의 효과

[0011] 상기한 본 발명은 유아나 아이 등과 같이 흥미 정도를 파악하기 힘든 화자의 흥미 정도를 용이하게 확인할 수 있게 하는 효과를 야기한다.

[0012] 특히 본 발명은 상기 흥미 정도를 점수로 산출함으로써, 화자의 흥미 정도를 용이하게 이해할 수 있게 하는 효과를 야기한다.

도면의 간단한 설명

[0013] 도 1은 본 발명의 바람직한 실시예에 따른 음성신호의 특징정보를 이용한 흥미정도 산출 시스템의 구성도.

도 2는 도 1의 음성신호의 특징정보 검출부의 상세구성도.

도 3은 본 발명에 따른 에러 함수의 형태를 도시한 도면.

도 4는 본 발명에 따른 e-무감각 손실 함수의 형태를 도시한 도면.

도 5는 본 발명의 바람직한 실시예에 따른 흥미점수 산출함수 생성방법의 흐름도.

도 6은 본 발명의 바람직한 실시예에 따른 흥미점수 산출방법의 흐름도.

발명을 실시하기 위한 구체적인 내용

[0014] 본 발명은 유아나 아이 등과 같이 흥미 정도를 파악하기 힘든 화자의 흥미 정도를 용이하게 확인할 수 있다.

[0015] 또한 본 발명은 상기 흥미 정도를 점수로 환산함으로써, 화자의 흥미 정도를 용이하게 이해할 수 있게 한다.

[0016] <음성신호의 특징정보를 이용한 흥미정도 검출 시스템의 구성>

[0017] 본 발명의 바람직한 실시예에 따른 음성신호의 특징정보를 이용한 흥미정도 검출 시스템의 구성을 도 1을 참조하여 상세히 설명한다.

[0018] 상기 음성신호의 특징정보를 이용한 흥미정도 검출 시스템은 제어장치(100)와 메모리부(102)와 사용자 인터페이스부(104)와 디스플레이부(106)와 음성신호 입력부(108)와 음성신호의 특징정보 검출부(110)로 구성된다.

[0019] 상기 제어장치(100)는 본 발명의 바람직한 실시예에 따라 미리 수집된 음성신호들 각각에 대해 설정된 T 점수들을 입력받아 저장하고, 상기 수집된 음성신호들의 특징정보들을 검출하고, 상기 특징정보들을 상기 T 점수에 대응되게 SVM 및 RVM 학습하고, 그 학습과정에서 흥미점수 산출함수를 생성하고, 질의를 위한 음성신호를 입력받아 특징정보를 검출하고, 상기 질의를 위한 음성신호의 특징정보를 상기 흥미점수 산출함수에 따라 연산하여 흥미점수를 산출한다.

[0020] 상기 메모리부(102)는 상기 제어장치(100)의 처리 프로그램을 포함하는 다양한 정보를 저장한다. 특히, 상기 수집된 음성신호들에 대한 T 점수 설정정보, 흥미점수 산출함수 등을 저장한다.

[0021] 상기 사용자 인터페이스부(104)는 사용자로부터 제공되는 각종 정보를 상기 제어장치(100)에 제공한다.

[0022] 상기 디스플레이부(106)는 상기 제어장치(100)의 제어에 따라 각종 정보를 표시하여 사용자에게 안내하며, 예를 들어 흥미점수 산출결과 등이 표시될 수 있다.

[0023] 상기 음성신호 입력부(108)는 음성신호를 입력받아 상기 제어장치(100)에 제공한다. 상기 음성신호 입력부(108)는 직접적으로 음성을 입력받는 마이크 장치이거나, 외부저장매체로부터 음성신호를 제공받기 위한 외부기기 인터페이스 장치가 될 수도 있다.

[0024] 상기 음성신호의 특징정보 검출부(110)는 상기 제어장치(100)의 제어에 따라 입력된 음성신호의 특징정보를 검출하여 출력한다.

[0025] <음성신호의 특징정보 검출부(110)의 상세>

[0026] 여기서, 상기 음성신호의 특징정보 검출부(110)의 상세 구성을 도 2를 참조하여 설명한다.

[0027] 상기 음성신호의 특징정보 검출부(110)는 입력되는 음성신호에 대해 윈도우를 설정하는 허밍 윈도우 설정부(200)와, 입력되는 음성신호를 시간-주파수 평면의 음성신호 $Y(l,k)$로 변환하는 DFT 처리부(202)와, 상기 시간-주파수 평면의 음성신호 $Y(l,k)$로부터 시간-주파수 평면에서의 에너지를 구하는 에너지 검출부(204)와, 상기 시간-주파수 평면에서의 에너지에 대해 멜 주파수 필터링을 이행하는 멜 필터링부(206)와, 상기 멜 주파수 필터링을 통과한 주파수의 로그 연산을 이행하는 로그 연산부(208)와, 상기 로그 연산부(208)를 통해 산출된 값으로 이산여현변환(discrete cosine transform)을 통해 특징정보를 산출하는 DCT(discrete cosine transform) 처리부(210)로 구성된다.

[0028] 상기 멜 필터링부(206)의 출력은 수학식 1과 같다.

수학식 1

[0029]
$$y_m(l,l_m) = \sum_{k=0}^{K/2} C((k,l_m)|Y(l,k)|^2$$

[0030] 상기 수학식 1에서, $y_m(l,l_m)$는 멜 필터링부(306)가 출력하는 음성신호이고, $Y(l,k)$는 시간-주파수 평면의 음성신호이고, l_m는 멜 필터 윈도우 인덱스이고, l은 프레임 인덱스이고, C는 멜 필터 계수이고, k 주파수 인덱스이다.

[0031] <음성신호의 특징정보를 이용하여 흥미정도 검출을 위한 수식 정의 관련>

[0032] SVM(Support Vector Machine)을 에러 함수(Error Function)를 정의해서 회귀분석(Regression)에 응용한 것을 SVR(Support Vector Regression)이라 한다. 이 SVR를 위해 먼저 비선형 커널 함수 $\phi(x_n)$를 정의하며, 일반적인 경우에 RBF(Radial Basis Function) 커널을 사용한다. 그리고 상기 에러 함수(error function)는 도 3과 같이 정의된다.

[0033] 상기 도 3의 에러 함수는 ε-무감각 손실 함수(ε-insensitive loss function)이며, 이를 이용해서 회귀 분석(regression)을 수행하면 다음과 같이 식이 전개될 수 있다.

[0034] 먼저 ε-무감각 손실 함수(ε-insensitive loss function) $L_\epsilon(y)$는 수학식 2와 같이 표현될 수 있다.

수학식 2

[0035]
$$L_\epsilon(y) = \begin{cases} 0 & \text{for } |f(x)-y| < \epsilon \\ |f(x)-y|-\epsilon & otherwise \end{cases}.$$

[0036] 상기 수학식 2에서 $L_\epsilon(y)$는 ε-무감각 손실 함수, x는 입력 벡터, y는 x에 대응하는 출력 벡터, ϵ은 허용할

수 있는 에러 범위, f(x)는 입력 벡터를 논리니어 평션을 통과한 값을 의미한다.

[0037] 상기 수학식 2에 따른 모델을 사용한 SVR은 도 5에 도시한 바와 같이 표현된다. 상기 도 5에서 엡시론(ε) 범위 밖을 여유 변수(slack variable) ξ를 이용하여 표현하면 다음과 같다.

수학식 3

[0038]
$$t_n \le y(x_n) + \epsilon + \xi_n$$
$$t_n \ge y(x_n) - \epsilon - \xi_n.$$

[0039] 상기 수학식 3은 도 5의 그림을 수식화한 것으로, 엡실론(ε)의 범위 안에서는 여유변수(slack variable) ξ가 0이고 그 밖에서는 넌-제로(non-zero) 값을 가진다.

[0040] 상기 수학식 3에서 x_n은 N개의 입력정보 중 임의의 하나의 입력정보이고, t_n은 해당하는 출력정보이고, $y(x_n)$은 x_n이 입력일 때 SVM을 통해서 나온 결과이고, ξ_n은 에러의 양을 의미하는 여유 변수(slack variable)를 의미한다. 이 이후의 아래첨자 n 또는 m은 N개의 데이터 중 임의의 데이터 하나를 의미한다.

[0041] 이를 이용해 SVR을 하기 위한 비용 함수(cost function)는 수학식 4 및 수학식 5와 같다.

수학식 4

[0042]
$$C\sum_{n=1}^{N}(\xi_n + \hat{\xi_n}) + \frac{1}{2}\|\mathbf{w}\|^2$$

[0043] 상기 수학식 4에서, C는 $\frac{1}{2}\|\mathbf{w}\|^2$ 와 ξ의 비중을 조절하는 상수이고, $\frac{1}{2}\|\mathbf{w}\|^2$ 텀은 "가 발산하지 않도록 고정시켜주는 역할을 하고, ξ는 여유변수이고, w는 웨이트를 의미한다.

수학식 5

[0044]
$$L = C\sum_{n=1}^{N}(\xi_n + \hat{\xi_n}) + \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{n=1}^{N}(\mu_n\xi_n + \hat{\mu_n}\hat{\xi_n}) - \sum_{n=1}^{N}a_n(\epsilon + \xi_n + y_n - t_n) - \sum_{n=1}^{N}\hat{a_n}(\epsilon + \hat{\xi_n} - y_n + t_n)$$

[0045] 상기 수학식 5에서, $C\sum_{n=1}^{N}(\xi_n + \hat{\xi_n}) + \frac{1}{2}\|\mathbf{w}\|^2$ 는 비용 함수이고, 나머지는 ξ가 0보다 크거나 같다라는 제한(constraint) 및 수학식 3에 해당하는 제한(constraint) 텀을 라그랑쥬 상수인 $a_n, \mu_n, \hat{a_n}, \hat{\mu_n}$를 이용해서 적용시킨 텀이다.

[0046] 이를 각 값($w, b, \xi, \hat{\xi}$) 에 대하여 극대점들을 찾으면 다음과 같다.

수학식 6

$$\frac{\partial L}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{n=1}^{N} (a_n - \hat{a}_n)\phi(\mathbf{x}_n)$$
$$\frac{\partial L}{\partial b} = 0 \Rightarrow b = \sum_{n=1}^{N} (a_n - \hat{a}_n) = 0$$
$$\frac{\partial L}{\partial \xi_n} = 0 \Rightarrow a_n + \mu_n = C$$
$$\frac{\partial L}{\partial \hat{\xi}_n} = 0 \Rightarrow \hat{a}_n + \hat{\mu}_n = C$$

[0047]

[0048] 이를 다시 수학식 5에 대입하면 수학식 7과 같다. 상기 수학식 6에서 b는 yn안에 포함되어있는 바이어스 텀이다.

수학식 7

$$\tilde{L}(a,\hat{a}) = -\frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}(a_n - \hat{a}_n)(a_m - \hat{a}_m)k(\mathbf{x}_n,\mathbf{x}_m) - \epsilon\sum_{n=1}^{N}(a_n + \hat{a}_n) + \sum_{n=1}^{N}(a_n - \hat{a}_n)t_n$$

[0049]

[0050] 상기 수학식 7에서, 상기 $\tilde{L}(a,\hat{a})$ 는 a 및 $\hat{a}$ 에 해당하는 코스트 함수이고, 나머지는 위에서 언급한 값들을 계속해서 전개한 식이다.

[0051] 여기서, 나머지 변수들은 위와 동일하고 $k(\mathbf{x}_n,\mathbf{x}_m) = \Phi(\mathbf{x}_n)^T\Phi(\mathbf{x}_m)$ 을 의미한다. 상기 $\Phi(\mathbf{x}_n)^T$, $\Phi(\mathbf{x}_m)$ 는 논리니어 커널 펑션을 통과한 입력변수를 의미한다

[0052] 그리고 상기 a_n 은 라그랑쥬 멀티플라이어이기 때문에 수학식 8에 따른 조건을 만족시키고, 또한 μ_n 도 라그랑쥬 멀티플라이어이기 때문에 0보다 크거나 같다는 조건을 가진다.

수학식 8

$$0 \le a_n \le C$$
$$0 \le \hat{a}_n \le C$$
$$\sum_{n=1}^{N}(a_n - \hat{a}_n) = 0$$

[0053]

[0054] 상기 수학식 7에 상기 수학식 8을 적용하여 이차 프로그래밍 문제(quadratic programming problem)를 풀어 a, $\hat{a}$ 을 구하면, 이 값이 서포트 벡터의 웨이트가 된다.

[0055] 그리고 바이어스 b는 다음과 같이 구해진다.

수학식 9

[0056] $$b = t_n - \epsilon - \sum_{m=1}^{N} (a_m - \widehat{a_m}) k(\mathbf{x_n}, \mathbf{x_m})$$

[0057] 상기 수학식 9에서 t_n은 입력정보에 대응되는 출력정보, 즉 T 점수이고, a_m은 라그랑쥬 상수값이고, $\hat{a}_m$은 a_m과 베타적인 라그랑쥬 상수값(둘 중 하나는 0)이고, b는 바이어스이고, ε은 허용할 수 있는 에러 범위이고, $k(x_n,x_m)$는 비선형 커널 평션이고, x_n,x_m는 입력정보 중 임의의 두 정보이다.

[0058] 회귀분석(regression)은 수학식 10을 통해 얻어진다.

수학식 10

[0059] $$y(x) = \sum_{n=1}^{N} (a_n - \widehat{a_n}) k(\mathbf{x}, \mathbf{x_n}) + b$$

[0060] 상기 수학식 10에서, $\widehat{a}$ 및 b는 위에서 구해진 값이며 커널 평션은 $k(\mathbf{x_n}, \mathbf{x_m}) = \Phi(\mathbf{x_n})^T \Phi(\mathbf{x_m})$ 를 의미한다.

[0061] 상기 수학식 10에서 a, $\hat{a}_n$은 라그랑쥬 멀티플라이어이고, 상기 b는 바이어스이고, 상기 $k(\mathbf{x_n}, \mathbf{x_m}) = \Phi(\mathbf{x_n})^T \Phi(\mathbf{x_m})$ 을 의미하며, 상기 $\Phi(\mathbf{x_n})^T$, $\Phi(\mathbf{x_m})$, 는 비선형 커널 평션을 통과한 입력변수를 의미한다

[0062] 상기 SVM의 타겟값이 정규분포를 가진다고 가정하고 확장한 알고리즘이 RVM(Relevance Vector Machines)이다.

[0063] 상기 RVM(relevance Vector Machines)을 이용한 회귀분석(Regression) 함수 $p(t|\mathbf{x}, \mathbf{w}, \beta)$는 수학식 11과 같다.

수학식 11

[0064] $$p(t|\mathbf{x}, \mathbf{w}, \beta) = N(t|y(\mathbf{x}), \beta^{-1})$$

$$\text{where } \beta = \sigma^{-2}$$

[0065] 상기 수학식 11에서, x는 입력벡터, t는 x에 대응되는 출력벡터, $\beta = \sigma^{-2}$ 는 정규분포의 분산(variance)을 의미한다. 즉, RVM은 SVM에 비해서 타겟값이 정규 분포를 가진다고 정의되는 것이 다르다.

[0066] 상기 수학식 11에서, y(x)는 비선형 평션을 통과한 입력벡터, N은 정규 분포를 의미한다.

[0067] 앞에서 전개한 바처럼 비선형 베이시스(basis) 평션 파이를 이용해 상기 y(x)를 정의하면 수학식 12와 같다.

수학식 12

$$y(x)=\sum_{i=1}^{N} w_i\phi(\mathbf{x})=\mathbf{w}^{\mathrm{T}}\boldsymbol{\Phi}(\mathbf{x})$$

[0068]

[0069] 상기 수학식 12에서, w는 웨이트를 의미하고, $\Phi(x)$는 논 리니어 베이시스 평션이고, N은 학습 데이터 개수를 의미한다.

[0070] 이를 바이어스 b를 분리해낸 SVM 유사형태, 즉 수학식 10과 같은 형태로 다시 정의하면 수학식 13과 같다.

수학식 13

$$y(x)=\sum_{i=1}^{N} w_i k(\mathbf{x},\mathbf{x_i})+\mathrm{b}$$

[0071]

[0072] 이하, 본 발명에서는 상기 수학식 13을 기본 형태로 사용한다.

[0073] 이를 구하기 위해 다시 N개의 입력정보 x가 있다고 가정하면, 수학식 14와 같이 정의할 수 있다.

수학식 14

$$p(\mathrm{t}|\mathbf{X},\mathbf{w},\beta)=\prod_{\mathrm{n}=1}^{N} p(\mathrm{t_n}|\mathbf{x_n},\mathbf{w},\beta^{-1})$$

[0074]

[0075] 상기 수학식 14에서, n은 입력정보의 순서, N은 입력정보의 최대 순서값, w는 웨이트, X_n은 입력정보, t_n은 X_n에 대응되게 미리 설정된 출력정보인 T 점수, $\beta=\sigma^{-2}$는 분산(variance)을 의미한다.

[0076] 또한 이때 상기 w를 제로-민 가우시안 프라이어(zero-mean Gaussian prior)로 정의하면 수학식 15와 같다.

수학식 15

$$p(\mathbf{w}|\boldsymbol{\alpha})=\prod_{\mathrm{n}=1}^{N} N(\mathrm{w_n}|0,\alpha_{\mathrm{n}}^{-1})$$

[0077]

[0078] 상기 수학식 15에서 α,α_n^{-1}은 변수, N은 정규분포를 의미한다.

[0079] 상기 수학식 15를 베이지안 선형 회귀분석(bayesian linear regression) 모델에 적용시키면 수학식 16과 같다.

수학식 16

$$p(\mathbf{w}|\mathbf{t},\mathbf{X},\boldsymbol{\alpha},\beta) = N(\mathbf{w}|\mathbf{m},\Sigma)$$

여기서, $\boldsymbol{m} = \beta\Sigma\Phi^T\boldsymbol{t}$ 이고,
$\Sigma = (\boldsymbol{A} + \beta\Phi^T\Phi)^{-1}$

[0080] 여기서, $\Phi_{ni} = \phi_i(\mathbf{x_n})$임.

[0081] 상기 수학식 16에서, w는 웨이트이고, X는 입력정보, t는 입력정보에 대응되게 설정된 출력정보, $\beta = \sigma^{-2}$는 분산(variance)을 의미하고, α, α_n^{-1} 은 변수이며, $\Phi_{ni} = \phi_i(\mathbf{x_n})$ 는 비선형 베이시스 함수이고, A는 α를 대각선(daigonal) 텀으로 가지는 정방행렬이다.

[0082]

[0083] 이제 α와 β를 산출하는 과정을 설명한다.

수학식 17

$$p(\mathrm{t}|\mathrm{X},\boldsymbol{\alpha},\beta) = \int \mathrm{p}(\mathrm{t}|\mathrm{X},\mathrm{w},\beta)\mathrm{p}(\mathrm{w}|\boldsymbol{\alpha})\mathrm{dw}$$

[0084]

[0085] 상기 수학식 17을 다시 풀어쓰면, 수학식 18과 같다.

수학식 18

$$\ln p(\mathrm{t}|\mathrm{X},\boldsymbol{\alpha},\beta) = \ln N(\boldsymbol{t}|\mathbf{0},\boldsymbol{C}) = -\frac{1}{2}\{N\ln(2\pi) + \ln|\boldsymbol{C}| + \boldsymbol{t}^T\boldsymbol{C}^{-1}\boldsymbol{t}\}$$

[0086] 여기서, $\boldsymbol{C} = \beta^{-1}\boldsymbol{I} + \Phi\boldsymbol{A}^{-1}\Phi^T$이다.

[0087] 이 식을 최소화(maximize)하는 α와 β를 찾으면 모든 식을 풀 수 있다.

[0088] 이 값을 얻기 위한 방법이 여러 가지 있는데 베이지안 회귀분석(bayesian regression) 모델의 에비던스 근사화(Evidence Approximation)를 하면 수학식 19에 따라 구할 수 있다.

수학식 19

$$\alpha_i^{new} = \frac{\gamma_i}{m_i^2}$$

$$(\beta^{new})^{-1} = \frac{\|\boldsymbol{t} - \boldsymbol{\Phi m}\|^2}{N - \sum_i \gamma_i}$$

where

$$\gamma_i = 1 - \alpha_i \Sigma_{ii}$$

[0089]

[0090] 위 식에 해당하는 변수들은 모두 앞에서 언급된 변수들이다.

[0091] 이를 통해 음성신호의 특징정보 x를 입력받아 그 음성신호에 대한 흥미점수를 수집된 음성신호들과 그 음성신호들 각각에 대해 미리 설정된 T 점수에 따라 회귀분석 학습된 수학식 20을 획득할 수 있다.

수학식 20

$$y(x) = \sum_{i=1}^{N} w_i k(\mathbf{x}, \mathbf{x_i}) + \mathrm{b}$$

[0092]

[0093] 상기 수학식 20에서, w는 웨이트, $k(x,x_i)$는 논리니어 커널 평션을 의미하고, b는 바이어스를 의미하고, x는 음성신호의 특징정보, x_i는 학습을 위한 음성신호들의 특징정보들이다.

[0094] <음성신호의 특징정보를 이용한 흥미정도 검출 처리 과정>

[0095] 이제 상술한 함수들을 이용하여 본 발명에 따라 음성신호의 특징정보를 이용하여 흥미정도를 검출하는 과정을 상세히 설명한다.

[0096] 먼저 음성신호의 특징정보를 이용하여 흥미점수를 산출하는 흥미점수 산출함수를 생성하는 과정을 도 5의 흐름도를 참조하여 설명한다.

[0097] 상기 음성신호의 특징을 이용한 흥미정도 검출 시스템의 제어장치(100)는 수집된 음성신호들을 입력받아(200단계), 음성신호의 특징정보 검출부(110)를 통해 음성신호의 특징정보들을 검출한다(202단계).

[0098] 상기 음성신호의 특징정보들이 검출된 후, 상기 제어장치(100)는 비선형 커널 함수 및 오차 함수를 정의한다(204단계). 상기 비선형 커널 함수는 RBF 커널을 사용하고, 오차 함수는 ε-무감각 손실함수를 사용한다.

[0099] 이후 상기 제어장치(100)는 비선형 커널 함수와 오차 함수에 따라 음성신호들의 특징정보들과 그에 대응되게 미리 설정된 T 점수정보를 SVM 및 RVM 학습하여, 흥미점수 산출함수를 생성한다(206단계).

[0100] 그리고 SVM 및 RVM 학습으로 생성된 흥미점수 산출함수를 이용하여 흥미점수를 산출하는 과정을 도 6의 흐름도를 참조하여 설명한다.

[0101] 상기 음성신호의 특징정보를 이용한 흥미정도 검출 시스템의 제어장치(100)는 흥미검사를 위한 화자의 음성신

호가 입력되면(300단계), 음성신호의 특징정보 검출부(110)를 통해 상기 음성신호의 특징정보를 검출한다(302단계).

[0102] 이후 상기 제어장치(100)는 SVM 및 RVM 학습에 의해 생성된 흥미점수 산출함수를 이용하여 상기 음성신호의 특징정보에 대한 흥미점수를 산출한다(304단계).

[0103] 상기한 본 발명의 바람직한 실시예에 따른 음성신호의 특징정보를 이용한 흥미점수 산출 방법은 컴퓨팅 가능한 장치에 의해 독취될 수 있는 저장매체에 기록될 수 있으며, 이는 당업자에게 자명하다.

부호의 설명

[0104] 100 : 제어장치

102 : 메모리부

104 : 사용자 인터페이스부

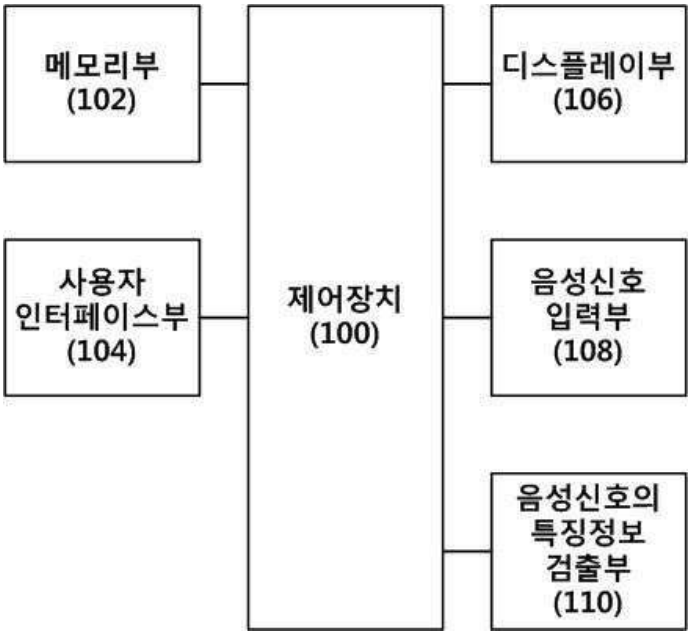
106 : 디스플레이부

108 : 음성신호 입력부

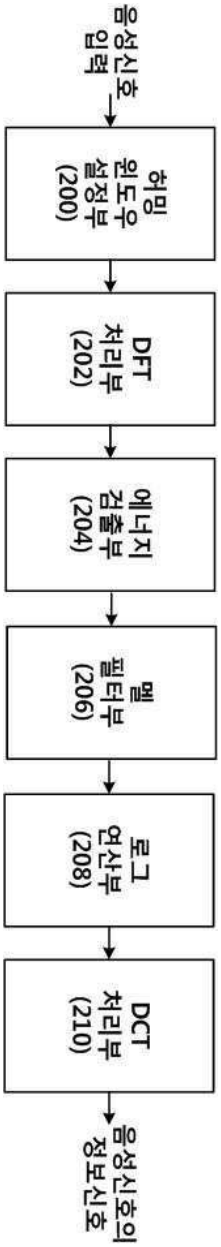
110 : 음성신호의 특징정보 검출부

도면

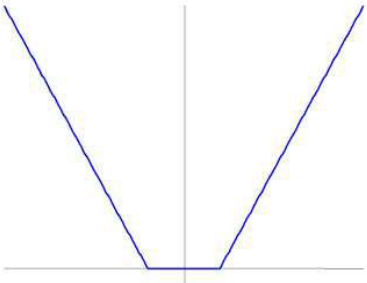
도면1



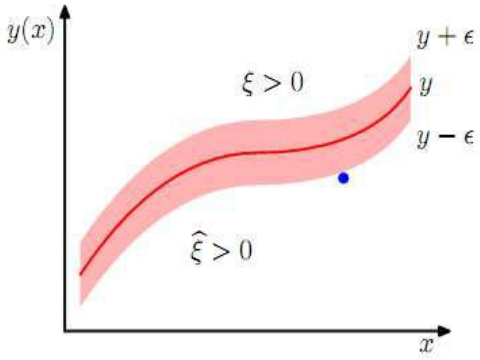
도면2


음성신호 입력
해밍 윈도우 설정부 (200)
DFT 처리부 (202)
에너지 검출부 (204)
멜 필터부 (206)
로그 연산부 (208)
DCT 처리부 (210)
음성신호의 정보신호

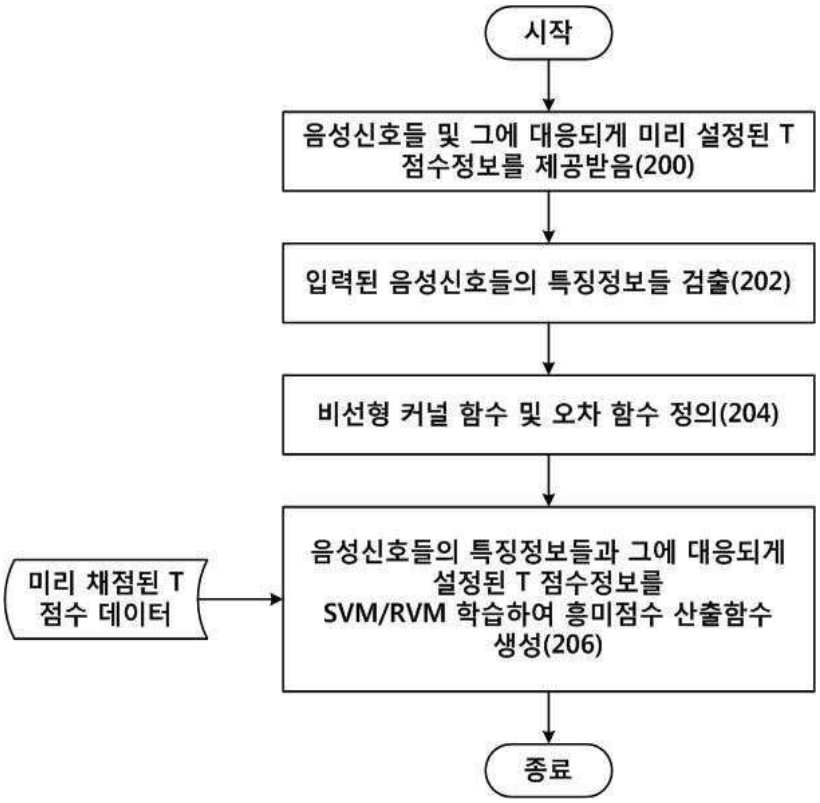
도면3



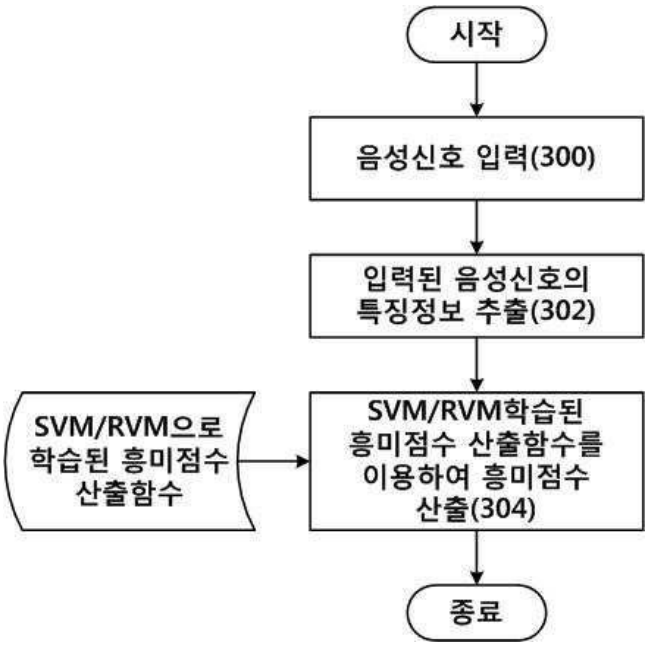
도면4


$y(x)$
$y+\epsilon$
$\xi > 0$
y
$y-\epsilon$
$\hat{\xi} > 0$
x

도면5


시작
음성신호들 및 그에 대응되게 미리 설정된 T 점수정보를 제공받음(200)
입력된 음성신호들의 특징정보들 검출(202)
비선형 커널 함수 및 오차 함수 정의(204)
미리 채점된 T 점수 데이터
음성신호들의 특징정보들과 그에 대응되게 설정된 T 점수정보를 SVM/RVM 학습하여 흥미점수 산출함수 생성(206)
종료

도면6


시작
음성신호 입력(300)
입력된 음성신호의 특징정보 추출(302)
SVM/RVM으로 학습된 흥미점수 산출함수
SVM/RVM학습된 흥미점수 산출함수를 이용하여 흥미점수 산출(304)
종료