



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2015년11월02일

(11) 등록번호 10-1561651

(24) 등록일자 2015년10월13일

(51) 국제특허분류(Int. Cl.)

G10L 25/63 (2013.01) G10L 25/09 (2013.01)

(21) 출원번호 10-2014-0062477

(22) 출원일자 2014년05월23일

심사청구일자 2014년05월23일

(56) 선행기술조사문헌

A. Mohamed et al., 'Deep belief networks for phone recognition', in Proc. NIPS Workshop Deep Learn. Speech Recogn. Rel. Applicat., 2009.*

Mohamed R. Amer et al., 'Emotion detection in speech using deep networks', 2014 IEEE International Conference on Acoustic, Speech and Sognal Processing(ICASSP), pp.3724~3728, 4 May 2014.*

*는 심사관에 의하여 인용된 문헌

(73) 특허권자

서강대학교산학협력단

서울특별시 마포구 백범로 35 (신수동, 서강대학교)

경북대학교 산학협력단

대구광역시 북구 대학로 80 (산격동, 경북대학교)

(72) 발명자

박형민

서울특별시 강남구 삼성로 212 3동 1002호 (대치동, 은마아파트)

조지원

서울 마포구 독막로32안길 15, 옥탑 (신수동)

(뒷면에 계속)

(74) 대리인

이지연

전체 청구항 수 : 총 7 항

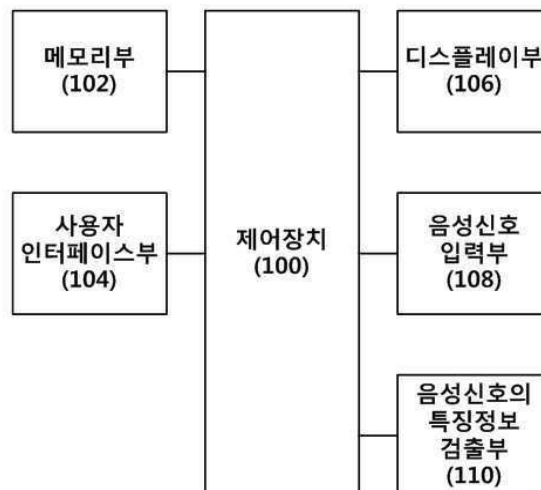
심사관 : 정성윤

(54) 발명의 명칭 딥 빌리프 네트워크를 이용한 음성기반 호불호 검출방법 및 장치, 그를 기록한 기록매체

(57) 요약

본 발명에 따르는 딥 빌리프 네트워크(Deep Belief Network; 이하 DBN라 칭함)를 이용한 음성기반 호불호 검출방법은, 음성신호들과 그에 대응되는 호불호 식별정보들로 구성되는 학습 데이터를 제공받아, 음성신호들에 대한 특징정보를 추출하고, 그 특징정보를 DBN을 통해 학습하여 DBN 데이터를 생성하고, 그 DBN 데이터에 호불호 식별정보로 구성되는 레이블 레이어를 추가하여 DBN 데이터의 생성을 완료하는 단계; 및 상기 테스트 음성신호가 입력되면, 그 테스트 음성신호에서 특징정보를 추출하고, 그 특징정보에 대응되는 DBN 데이터를 검출하고, 그 검출된 DBN 데이터의 레이블 레이어에 기록된 호불호 식별정보에 따라 호불호 검출결과를 출력하는 단계;를 구비함을 특징으로 한다.

대표도 - 도1



(72) 발명자

이호용

서울 마포구 백범로10길 27

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 ~ 2014.05.31

명세서

청구범위

청구항 1

딥 빌리프 네트워크(Deep Belief Network; 이하 DBN라 칭함)를 이용한 음성기반 호불호 검출방법에 있어서,

음성신호들과 그에 대응되는 호불호 식별정보들로 구성되는 학습 데이터를 제공받아, 음성신호들에 대한 특징정보를 추출하고, 그 특징정보를 DBN을 통해 학습하여 DBN 데이터를 생성하고, 그 DBN 데이터에 호불호 식별정보로 구성되는 레이블 레이어를 추가하여 DBN 데이터의 생성을 완료하는 단계;

테스트 음성신호가 입력되면, 그 테스트 음성신호에서 특징정보를 추출하고, 그 특징정보에 대응되는 DBN 데이터를 검출하고, 그 검출된 DBN 데이터의 레이블 레이어에 기록된 호불호 식별정보에 따라 호불호 검출결과를 출력하는 단계;를 구비하며,

상기 특징정보의 학습은,

특징정보에 대한 가중합으로 다음 층의 유닛을 정의하는 RBM(Restricted Boltzmann Machine) 을 다수회 수행하고, 각 RBM 수행결과를 적층하여 DBN 데이터를 구성하는 것이며,

상기 RBM은 특징정보에 포함된 정보들 각각의 상태에 대한 에너지를 최소화하여 평형상태에 이르도록 수학식 7에 따라 학습함을 특징으로 하는 DBN을 이용한 음성기반 호불호 검출방법.

수학식 7

$$\begin{aligned} h^{(t)} &\sim \text{sigm}(v^{(t)}w + c) \\ v^{(t+1)} &\sim \text{sigm}(h^{(t)}w' + b) \end{aligned}$$

상기 수학식 7에서 상기 $h^{(t)}$ 는 t-스텝의 힉스 샘플링을 통해 얻은 은닉 유닛의 상태이고, $\text{sigm}()$ 은 sigmoid function이고, $v^{(t)}$ 는 t-스텝의 힉스 샘플링을 통해 얻은 가시 유닛의 상태이고, w 는 가시 유닛과 은닉 유닛 사이의 가중치이고, c 는 은닉 유닛의 바이어스이고, b 는 가시 유닛의 바이어스이고, w' 는 가중치 매트릭스(weight matrix)의 치환(transpose)이며, 상기 가시 유닛이 음성신호의 특징정보임.

청구항 2

삭제

청구항 3

삭제

청구항 4

제1항에 있어서,

상기 가중치와 바이어스들은 수학식 8에 따라 근사화됨을 특징으로 하는 DBN을 이용한 음성기반 호불호 검출방법.

수학식 8

$$\begin{aligned} w_{ji} : CD_k(\theta, v^{(0)}) &= p(H_i = 1|v^{(0)})v_j^{(0)} - p(H_i = 1|v^{(k)})v_j^{(k)} \\ b_j : CD_k(\theta, v^{(0)}) &= v_j^{(0)} - v_j^{(k)} \\ c_i : CD_k(\theta, v^{(0)}) &= p(H_i = 1|v^{(0)}) - p(H_i = 1|v^{(k)}) \end{aligned}$$

상기 수학식 8에서, $CD_k()$ 는 k-step Contrastive divergence를 이용하여 구한 그레디언트(gradient)이고, w_{ji} 는 j번째 가시 유닛과 i번째 은닉 유닛 사이의 가중치(weight)이고, b_j 는 j번째 가시 유닛의 바이어스(bias)이고, c_i 는 i번째 은닉 유닛의 바이어스이고, $v^{(0)}, v^{(1)}, \dots, v^{(t)}$ 는 t-스텝의 깁스 샘플링(t-step gibbs sampling)을 통해 얻은 가시 유닛의 상태이고, Θ 는 가중치나 바이어스를 표현하기 위한 파라미터임.

청구항 5

제1항에 있어서,

상기 DBN 데이터를 재학습을 위해 파인 튜닝하는 단계;를 더 구비함을 특징으로 하는 DBN을 이용한 음성기반 호불호 검출방법.

청구항 6

제1항, 제4항 및 제5항 중 어느 한 항에 따르는 방법을 실행하기 위한 프로그램이 기록된 저장매체.

청구항 7

DBN을 이용한 음성기반 호불호 검출장치에 있어서,

학습 데이터와 DBN 데이터를 저장하는 메모리부;

음성신호에 대한 특징정보를 추출하는 음성신호의 특징정보 검출부;

음성신호들과 그에 대응되는 호불호 식별정보들로 구성되는 학습 데이터를 제공받아, 상기 음성신호의 특징정보 검출부를 통해 음성신호들에 대한 특징정보를 추출하고, 그 특징정보를 DBN을 통해 학습하여 DBN 데이터를 생성하고, 그 DBN 데이터에 호불호 식별정보로 구성되는 레이블 레이어를 추가하여 DBN 데이터의 생성을 완료하고,

테스트 음성신호가 입력되면, 상기 음성신호의 특징정보 검출부를 통해 상기 테스트 음성신호에서 특징정보를 추출하고, 그 특징정보에 대응되는 DBN 데이터를 검출하고, 그 검출된 DBN 데이터의 레이블 레이어에 기록된 호불호 식별정보에 따라 호불호 검출결과를 출력하는 제어장치;를 구비하며,

상기 특징정보의 학습은,

특징정보에 대한 가중합으로 다음 층의 유닛을 정의하는 RBM(Restricted Boltzmann Machine)을 다수회 수행하고, 각 RBM 수행결과를 적층하여 DBN 데이터를 구성하는 것이며,

상기 RBM은 특징정보에 포함된 정보들 각각의 상태에 대한 에너지를 최소화하여 평형상태에 이르도록 수학식 9에 따라 학습함을 특징으로 하는 DBN을 이용한 음성기반 호불호 검출장치.

수학식 9

$$\begin{aligned} h^{(t)} &\sim \text{sigm}(v^{(t)}w + c) \\ v^{(t+1)} &\sim \text{sigm}(h^{(t)}w' + b) \end{aligned}$$

상기 수학식 9에서 상기 $h^{(t)}$ 는 t-스텝의 깁스 샘플링을 통해 얻은 은닉 유닛의 상태이고, $\text{sigm}()$ 은 sigmoid function이고, $v^{(t)}$ 는 t-스텝의 깁스 샘플링을 통해 얻은 가시 유닛의 상태이고, w 는 가시 유닛과 은닉 유닛 사이의 가중치이고, c 는 은닉 유닛의 바이어스이고, b 는 가시 유닛의 바이어스이고, w' 는 가중치 매트릭스(weight

matrix)의 치환(transpose)이며, 상기 가시 유닛이 음성신호의 특징정보임.

청구항 8

삭제

청구항 9

삭제

청구항 10

제7항에 있어서,

상기 가중치와 바이어스들은 수학적식 10에 따라 근사화됨을 특징으로 하는 DBN을 이용한 음성기반 호불호 검출장치.

수학적식 10

$$\begin{aligned} w_{ji} &: CD_k(\theta, v^{(0)}) = p(H_i = 1 | v^{(0)})v_j^{(0)} - p(H_i = 1 | v^{(k)})v_j^{(k)} \\ b_j &: CD_k(\theta, v^{(0)}) = v_j^{(0)} - v_j^{(k)} \\ c_i &: CD_k(\theta, v^{(0)}) = p(H_i = 1 | v^{(0)}) - p(H_i = 1 | v^{(k)}) \end{aligned}$$

상기 수학적식 10에서, $CD_k()$ 는 k-step Contrastive divergence를 이용하여 구한 그래디언트(gradient)이고, w_{ji} 는 j번째 가시 유닛과 i번째 은닉 유닛 사이의 가중치(weight)이고, b_j 는 j번째 가시 유닛의 바이어스(bias)이고, c_i 는 i번째 은닉 유닛의 바이어스이고, $v^{(0)}, v^{(1)}, \dots, v^{(t)}$ 는 t-스텝의 깁스 샘플링(t-step gibbs sampling)을 통해 얻은 가시 유닛의 상태이고, Θ 는 가중치나 바이어스를 표현하기 위한 파라미터임.

청구항 11

제7항에 있어서,

상기 제어장치가 상기 DBN 데이터를 재학습을 위해 파인 튜닝함을 특징으로 하는 DBN을 이용한 음성기반 호불호 검출장치.

발명의 설명

기술 분야

[0001]

본 발명은 음성신호로부터 특징정보를 검출하는 기술에 관한 것으로, 더욱 상세하게는 딥 빌리프 네트워크(Deep Belief Network; 이하 DBN라 칭함)를 이용하여 음성신호들과 그에 대응되게 정해진 호불호 식별정보들로 구성되는 학습 데이터에 대해 음성신호의 특징이 향상되도록 학습하여 음성기반 호불호 검출결과와 신뢰도를 높일 수 있는 DBN을 이용한 음성기반 호불호 검출방법 및 장치, 그를 기록한 기록매체에 관한 것이다.

배경 기술

[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 DBN를 이용하여 음성신호들과 그에 대응되게 정해진 호불호 식별정보들로 구성되는 학습 데이터에 대해 음성신호의 특징이 향상되도록 학습하여 음성기반 호불호 검출결과와 신뢰도를 높일 수 있는 DBN을 이용한 음성기반 호불호 검출방법 및 장치, 그를 기록한 기록매체를 제공하는 것을 목적으로 한다.

과제의 해결 수단

[0010]

상기한 목적을 달성하기 위한 본 발명에 따르는 딥 빌리프 네트워크(Deep Belief Network; 이하 DBN라 칭함)를 이용한 음성기반 호불호 검출방법은, 음성신호들과 그에 대응되는 호불호 식별정보들로 구성되는 학습 데이터를 제공받아, 음성신호들에 대한 특징정보를 추출하고, 그 특징정보를 DBN을 통해 학습하여 DBN 데이터를 생성하고, 그 DBN 데이터에 호불호 식별정보로 구성되는 레이블 레이어를 추가하여 DBN 데이터의 생성을 완료하는 단계; 및 상기 테스트 음성신호가 입력되면, 그 테스트 음성신호에서 특징정보를 추출하고, 그 특징정보에 대응되는 DBN 데이터를 검출하고, 그 검출된 DBN 데이터의 레이블 레이어에 기록된 호불호 식별정보에 따라 호불호 검출결과를 출력하는 단계;를 구비함을 특징으로 한다.

발명의 효과

[0011]

상기한 본 발명은 DBN를 이용하여 음성신호들과 그에 대응되게 정해진 호불호 식별정보들로 구성되는 학습 데이터에 대해 음성신호의 특징이 향상되도록 학습하여 음성기반 호불호 검출결과와 신뢰도를 높일 수 있는 효과를 야기한다.

도면의 간단한 설명

- [0012] 도 1은 본 발명의 바람직한 실시예에 따른 음성신호의 특징정보를 이용한 호불호 검출 시스템의 구성도.
- 도 2는 본 발명의 바람직한 실시예에 따른 DBN를 이용한 학습방법의 절차를 도시한 도면.
- 도 3은 본 발명의 바람직한 실시예에 따른 음성기반 호불호 검출방법의 흐름도.
- 도 4는 본 발명의 바람직한 실시예에 따른 RBM 구조를 예시한 도면.
- 도 5는 본 발명의 바람직한 실시예에 따른 깁스 샘플링 구조를 예시한 도면.
- 도 6은 본 발명의 바람직한 실시예에 따른 DBN 구조를 예시한 도면.

발명을 실시하기 위한 구체적인 내용

- [0013] 본 발명은 DBN를 이용하여 음성신호들과 그에 대응되게 정해진 호불호 식별정보들로 구성되는 학습 데이터에 대해 음성신호의 특징이 향상되도록 학습하여 음성기반 호불호 검출결과와 신뢰도를 높일 수 있다.
- [0014] 이러한 본 발명의 바람직한 실시예를 도면을 참조하여 상세히 설명한다.
- [0015] 본 발명의 바람직한 실시예에 따른 음성신호의 특징정보를 이용한 호불호 검출장치의 구성을 도 1을 참조하여 설명한다.
- [0016] 상기 음성신호의 특징정보를 이용한 호불호 검출장치는 제어장치(100)와 메모리부(102)와 사용자 인터페이스부(104)와 디스플레이부(106)와 음성신호 입력부(108)와 음성신호의 특징정보 검출부(110)로 구성된다.
- [0017] 상기 제어장치(100)는 본 발명의 바람직한 실시예에 따라 미리 수집된 학습데이터를 제공받는다. 상기 학습 데이터는 음성신호들과 그에 대응되게 정해진 호불호 식별정보들로 구성된다.
- [0018] 상기 제어장치(100)는 상기 학습 데이터에 포함된 음성신호들 각각에 대해 특징정보를 추출하고, 그 특징정보를 DBN을 통해 학습하여 특징정보를 향상시킨다. 이후 상기 제어장치(100)는 상기 DBN을 통해 특징정보가 향상된 DBN 데이터에 호불호 식별정보를 레이블 레이어(label layer)에 부가하고, 다시 과인 튜닝(fine tuning)을 이행하여 학습을 완료한다. 즉, 상기 학습이 완료되면, 학습 데이터에 포함된 모든 음성신호들에 대해 레이블 레이어에 호불호 식별정보가 부가된 향상된 특징정보로 구성되는 DBN 데이터들을 획득하게 된다.
- [0019] 이후 상기 제어장치(100)는 테스트 음성신호를 입력받아 특징정보를 추출하고, 그 특징정보에 부합된 DBN 데이터를 검출하고, 그 DBN 데이터의 레이블 레이어를 통해 호불호를 검출한다.
- [0020] 상기 메모리부(102)는 상기 제어장치(100)의 처리 프로그램을 포함하는 다양한 정보를 저장한다. 특히, 상기 학습 데이터, DBN 데이터들, 테스트 음성신호에 대한 호불호 검출결과 등이 저장된다.
- [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 호불호 검출결과 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] 상기한 본 발명은 크게 학습과정과 호불호 검출과정으로 나눌 수 있다.
- [0026] 먼저 학습과정을 도 2를 참조하여 설명한다.
- [0027] 상기 음성신호의 특징정보를 이용한 호불호 검출장치의 제어장치(100)는 음성신호들 및 그에 대응되게 미리 설정된 호불호 식별정보들로 구성되는 학습 데이터를 제공받아 메모리부(102)에 저장한다(200단계).

- [0028] 이후 상기 제어장치(100)는 음성신호의 특징정보 검출부(110)를 통해 상기 학습 데이터의 음성신호들 각각에 대해 특징정보들을 검출하고(202단계), 그 음성신호들의 특징정보들을 DBN을 통해 향상시켜 DBN 데이터들을 구성하고, 그 DBN 데이터들에 미리 설정된 호불호 식별정보로 구성되는 레이블 레이어를 부가하고(206단계), 이후 전체 DBN 데이터에 대해 파인 튜닝을 수행하여 호불호 검출을 위한 DBN 데이터들의 생성을 완료한다(206단계).
- [0029]
- [0030] 그리고 상기 호불호 검출과정을 도 3을 참조하여 설명한다.
- [0031] 상기 음성신호의 특징정보를 이용한 호불호 검출장치의 제어장치(100)는 음성신호 입력부(108)를 통해 테스트 음성신호가 입력되면(300단계), 음성신호의 특징정보 검출부(110)를 통해 상기 음성신호로부터 특징정보를 추출한다(302단계).
- [0032] 이후 상기 제어장치(100)는 상기 음성신호의 특징정보에 대응되는 DBN 데이터를 검출하고, 그 DBN 데이터의 레이블 레이어에 기록된 호불호 식별정보를 독출하여, 해당 테스트 음성신호에 대응되는 호불호 식별정보를 검출한다(304단계).
- [0033] 이제 상기 학습과정을 좀더 상세히 설명한다.
- [0034] 본 발명의 학습을 위해 전문가들은 음성신호들을 수집하고, 각 음성신호들에 대한 호불호 식별정보를 부여하여 학습 데이터를 생성한다.
- [0035] 본 발명은 상기 학습 데이터에 포함된 음성신호들로부터 특징정보들을 추출하고, 그 특징정보들에 대해 RBM 수식에 의해 학습하고 그 RBM 학습결과를 적층하여 DBN 데이터를 생성하며, 그 DBN 데이터에 호불호 식별정보로 구성되는 레이블 레이어를 부가하고, 파인 튜닝을 이행하여 학습을 완료한다.
- [0036] 상기한 RBM 및 DBN은 음성신호가 입력되었을 때에 이로 인해 얻어지는 특징정보를 자율학습(unsupervised training)하는 과정이다.
- [0037] 좀더 설명하면, 입력되는 음성신호의 특징정보를 가시유닛에 넣고, RBM 수식에 의해 학습하면 상단에 위치하는 은닉유닛에 나오는 결과는 입력된 특징정보를 재분류하고 배열하여 최종 판단에 조금 더 도움이 되는 특징정보로 향상된다. 이 과정에서 특징정보의 차원(dimension)을 늘리거나 줄일 수 있다.
- [0038] 이러한 RBM을 여러번 거친 결과들을 적층하여 DBN 데이터를 생성하며, 이 DBN 데이터는 입력값이 적당히 수정되고 재배열된 결과를 가지게 된다.
- [0039] 상기 RBM의 학습은 입력 데이터에 의한 계산값과 복원 데이터에 의한 계산값의 차이가 적어지는 방향으로 진행되며, 여기서 복원 데이터란 네트워크의 연결 가중치를 랜덤으로 초기화한 상태에서 정방향 계산하고 다시 역방향 계산하여 얻어진 데이터를 말한다. 즉 입력 데이터에 의한 평균은 입력 데이터를 정방향으로 네트워크 통과시킨 값이며, 복원 데이터에 의한 평균은 복원 데이터를 다시 정방향으로 네트워크 통과시킨 값이다. 상기 복원 데이터를 구할 때 네트워크를 왔다갔다 하는 횟수를 무한대로 늘리면 보다 정확한 값을 얻어낼 수 있고 이것을 모델 데이터라 하며, 실제 계산량 측면에서 모델 데이터는 네트워크를 한번만 왔다 갔다한 복원 데이터를 사용한다.
- [0040] 이러한 학습과정을 도면과 수학적식을 토대로 설명한다.
- [0041] 도 4는 입력 유닛의 가중합으로 다음 층의 유닛을 정의하는 RBM 모델을 예시한 것이다. 도 4에서 h_0, h_1, h_2 는 은닉 유닛(hidden unit)이고, v_0, v_1, v_2, v_3 은 가시 유닛(visible unit)이다. 이 모델은 각 상태들에 대한 에너지를 최소화하여 평형상태에 이르도록 학습하며, 이때 정의되는 상태 에너지는 수학적식 1과 같다.

수학식 1

$$E(v, h) = - \sum_{i=1}^H \sum_{j=1}^V w_{ji} h_i v_j - \sum_{j=1}^V b_j v_j - \sum_{i=1}^H c_i h_i$$

[0042]

[0043]

상기 수학식 1에서, $E(v, h)$ 는 네트워크(network)의 에너지(energy)이고, v_j 는 j 번째 가시 유닛의 상태 바이너리(binary) 값이고, h_i 는 i 번째 은닉 유닛의 상태 바이너리 값이고, H 는 은닉 유닛의 개수이고, V 는 가시 유닛의 개수이고, i 는 은닉 유닛의 인덱스(index)이고, j 는 가시 유닛의 인덱스이고, w_{ji} 는 j 번째 가시 유닛과 i 번째 은닉 유닛 사이의 가중치(weight)이고, b_j 는 j 번째 가시 유닛의 바이어스(bias)이고, c_i 는 i 번째 은닉 유닛의 바이어스이다.

[0044]

상기한 수학식 1에 의해 정의된 상태 에너지로부터 깁스 분산(gibbs distribution)에 의해 수학식 2와 같은 상태 확률을 획득할 수 있다.

수학식 2

$$p(v, h) = \frac{1}{Z} e^{-E(v, h)}$$

$$Z = \sum_{v, h} e^{-E(v, h)}$$

[0045]

[0046]

상기 수학식 2에서, $p(v, h)$ 는 네트워크의 어떤 유닛의 상태 확률이고, Z 는 분배함수(partition function)이다.

[0047]

그리고 에너지를 최소화하는 파라미터를 추정하기 위해 상기 확률을 이용한 로그 우도 함수를 최대화한다. 그 로그우도 함수는 수학식 3과 같다.

수학식 3

$$\ln L(\theta|v) = \ln p(v|\theta) = \ln \frac{1}{Z} \sum_h e^{-E(v, h)}$$

[0048]

[0049]

상기 수학식 3에서, $\ln L(\theta|v)$ 는 우도의 대수값(log likelihood)이고, θ 는 가중치(weight)나 바이어스(bias) 등을 표현하는 파라미터(parameter)이다.

[0050]

상기 로그우도함수는 그레이디언트 어센트(gradient ascent) 방법을 이용하여 최대화하게 되는데, 이때 가중치와 가시 유닛, 은닉 유닛의 바이어스 각각에 대한 그레이디언트는 수학식 4와 같다.

수학식 4

$$\begin{aligned}\frac{\partial \ln L(\theta|v)}{\partial w_{ji}} &= p(H_i = 1|v)v_j - \sum_v p(v)p(H_i = 1|v)v_j \\ \frac{\partial \ln L(\theta|v)}{\partial b_j} &= v_j - \sum_v p(v)v_j \\ \frac{\partial \ln L(\theta|v)}{\partial c_i} &= p(H_i = 1|v) - \sum_v p(v)p(H_i = 1|v)\end{aligned}$$

[0051]

[0052]

상기 수학식 4를 간단하기 하기 위해 간단히 하기 위한 방법으로 CD(Contrastive divergence)를 사용한다.

[0053]

즉, 상기의 로그우도함수 그레이디언트(gradient)는 입력 데이터의 평균과 모델 데이터의 평균의 차이로 구성되어 있는데, 모델 데이터의 평균을 계산하는 것이 복잡하여 MCMC(Markov Chain Monte Carlo)에 기반한 깃스 샘플링(Gibbs sampling)을 이용하며, 깃스 샘플링을 예시한 것이 도 5이다.

[0054]

도 5에서, 상기 $h^{(0)}, h^{(1)}, \dots, h^{(t)}$ 는 t-스텝의 깃스 샘플링(t-step gibbs sampling)을 통해 얻은 은닉 유닛의 상태이고, $v^{(0)}, v^{(1)}, \dots, v^{(t)}$ 는 t-스텝의 깃스 샘플링(t-step gibbs sampling)을 통해 얻은 가시 유닛의 상태이다.

[0055]

상기한 깃스 샘플링은 처음 입력으로부터 다음 은닉 유닛의 확률을 계산하고, 이로부터 은닉 유닛 상태를 샘플링해낸다. 또한 이렇게 얻어낸 은닉 유닛값을 이용하여 다음 가시 유닛의 확률을 계산하여 가시 유닛 상태를 샘플링한다.

[0056]

이와 같은 과정을 반복하면, 모델 분포에 가까운 유닛의 상태를 얻어낼 수 있다. 이는 수학식 5와 같이 표현될 수 있다.

수학식 5

$$\begin{aligned}h^{(t)} &\sim \text{sigm}(v^{(t)}w + c) \\ v^{(t+1)} &\sim \text{sigm}(h^{(t)}w' + b)\end{aligned}$$

[0057]

[0058]

상기 수학식 5에서 상기 $h^{(t)}$ 는 t-스텝의 깃스 샘플링을 통해 얻은 은닉 유닛의 상태이고, $\text{sigm}()$ 은 sigmoid function이고, $v^{(t)}$ 는 t-스텝의 깃스 샘플링을 통해 얻은 가시 유닛의 상태이고, w 는 가시 유닛과 은닉 유닛 사이의 가중치이고, c 는 은닉 유닛의 바이어스이고, b 는 가시 유닛의 바이어스이고, w' 는 가중치 매트릭스(weight matrix)의 치환(transpose)이다.

[0059]

이와 같은 과정을 k번 반복하는 것을 k-step Constrative divergence라 하는데, 대개 k를 1로 하여 모델분포를 근사해낸다. 그 결과 얻어지는 수정된 그레이디언트(gradient) 식은 수학식 6과 같다.

수학식 6

$$\begin{aligned}w_{ji} : CD_k(\theta, v^{(0)}) &= p(H_i = 1|v^{(0)})v_j^{(0)} - p(H_i = 1|v^{(k)})v_j^{(k)} \\ b_j : CD_k(\theta, v^{(0)}) &= v_j^{(0)} - v_j^{(k)} \\ c_i : CD_k(\theta, v^{(0)}) &= p(H_i = 1|v^{(0)}) - p(H_i = 1|v^{(k)})\end{aligned}$$

[0060]

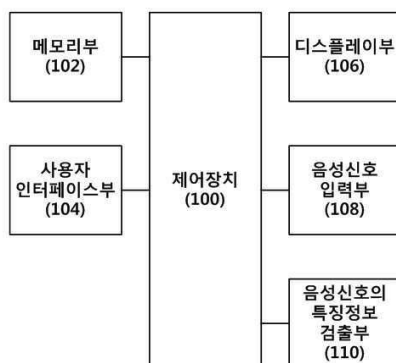
- [0061] 상기 수학적 식 8에서, $CD_k()$ 는 k-step Contrastive divergence를 이용하여 구한 그레디언트(gradient)이고, W_{ji} 는 j번째 가시 유닛과 i번째 은닉 유닛 사이의 가중치(weight)이고, b_j 는 j번째 가시 유닛의 바이어스(bias)이고, c_i 는 i번째 은닉 유닛의 바이어스이고, $v^{(0)}, v^{(1)}, \dots, v^{(t)}$ 는 t-스텝의 깁스 샘플링(t-step gibbs sampling)을 통해 얻은 가시 유닛의 상태이고, Θ 는 가중치나 바이어스를 표현하기 위한 파라미터이다.
- [0062] 이렇게 얻어낸 그레디언트를 이용하여 입력 데이터에 대한 평균과 복원 데이터에 의한 평균 차이가 줄어드는 방향으로 학습되어 보다 향상된 다음 단계의 특징정보를 얻어낼 수 있다.
- [0063] 이렇게 학습되는 RBM을 한층씩 쌓으면 다중층의 모델을 만들 수 있는 이를 DBN라 하며, 이를 예시한 것이 도 6이다. 상기한 RBM 학습(pre-training) 과정이 끝난 후, 최종 레이어(layer)에 레이블 레이어(label layer)를 추가로 연결한다.
- [0064] 또한 본 발명은 자율학습(Unsupervised learning)을 통해 학습된 DBN에 타겟(Target)을 추가하여, 통제된 학습(supervised training)을 진행하고, 원하는 호불호를 분류할 수 있게도 할 수 있다.
- [0065] 또한 전체 DBN을 재학습하는 파인 튜닝(fine-tuning) 과정에는 역전파(backpropagation) 학습 등 다양한 알고리즘이 사용될 수 있다.
- [0066] 상기한 본 발명의 바람직한 실시예에 따른 음성신호의 특징정보를 이용한 호불호 방법은 컴퓨팅 가능한 장치에 의해 독취될 수 있는 저장매체에 기록될 수 있으며, 이는 당업자에게 자명하다.

부호의 설명

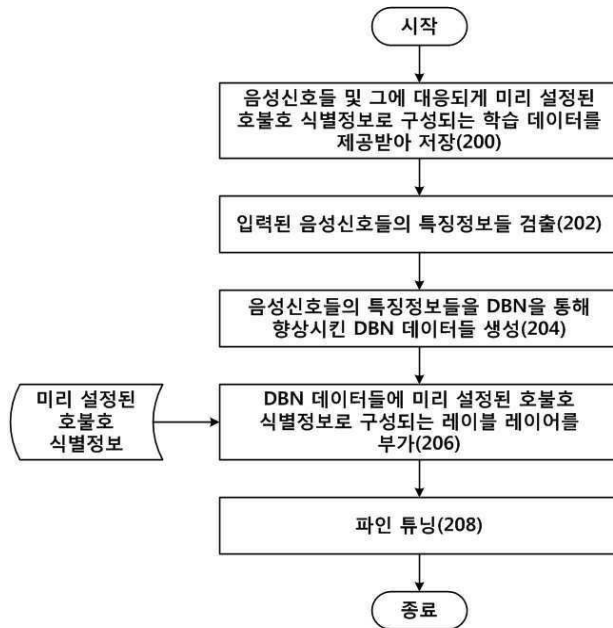
- [0067] 100 : 제어장치
102 : 메모리부
104 : 사용자 인터페이스부
106 : 디스플레이부
108 : 음성신호 입력부
110 : 음성신호의 특징정보 검출부

도면

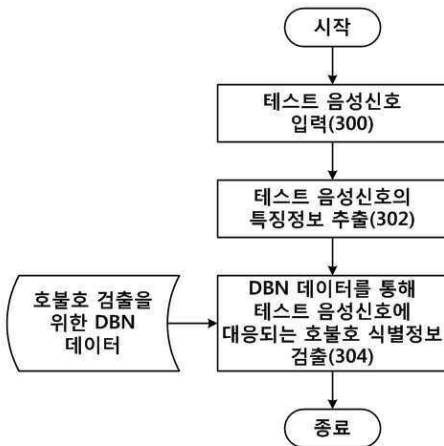
도면1



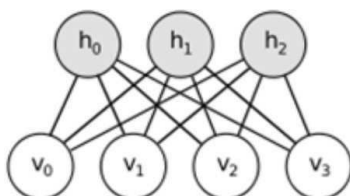
도면2



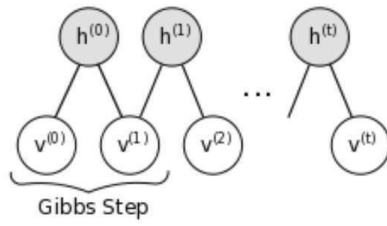
도면3



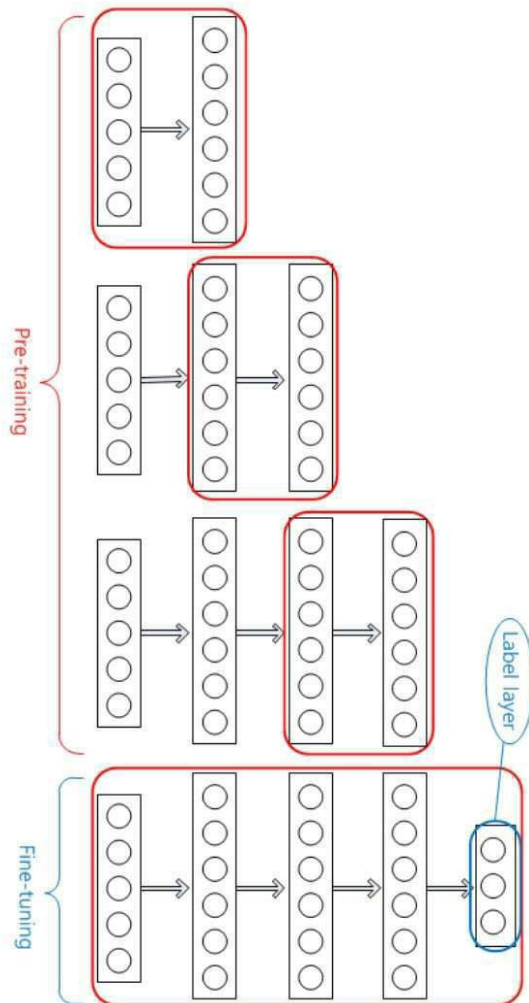
도면4



도면5



도면6



【심사관 직권보정사항】

【직권보정 1】

【보정항목】 청구범위

【보정세부항목】 청구항 7의 9번째 줄

【변경전】

상기 테스트 음성신호가 입력되면

【변경후】

테스트 음성신호가 입력되면

【직권보정 2】

【보정항목】 청구범위

【보정세부항목】 청구항 1의 7번째 줄

【변경전】

상기 테스트 음성신호가 입력되면

【변경후】

테스트 음성신호가 입력되면