(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0067724
(43) 공개일자 2022년05월25일

(51) 국제특허분류(Int. Cl.)
G16H 50/30 (2018.01) *G06N 20/00* (2019.01)
G06N 3/08 (2006.01) *G16H 50/20* (2018.01)
(52) CPC특허분류
G16H 50/30 (2018.01)
G06N 3/08 (2013.01)
(21) 출원번호 10-2020-0154262
(22) 출원일자 2020년11월18일
심사청구일자 2020년11월18일

(71) 출원인
울산과학기술원
울산광역시 울주군 언양읍 유니스트길 50
주식회사 바이오에이지
서울특별시 강남구 봉은사로 640(삼성동)
(72) 발명자
이정혜
울산광역시 울주군 언양읍 유니스트길 50
임치현
울산광역시 울주군 언양읍 유니스트길 50
(뒷면에 계속)
(74) 대리인
특허법인태백

전체 청구항 수 : 총 13 항

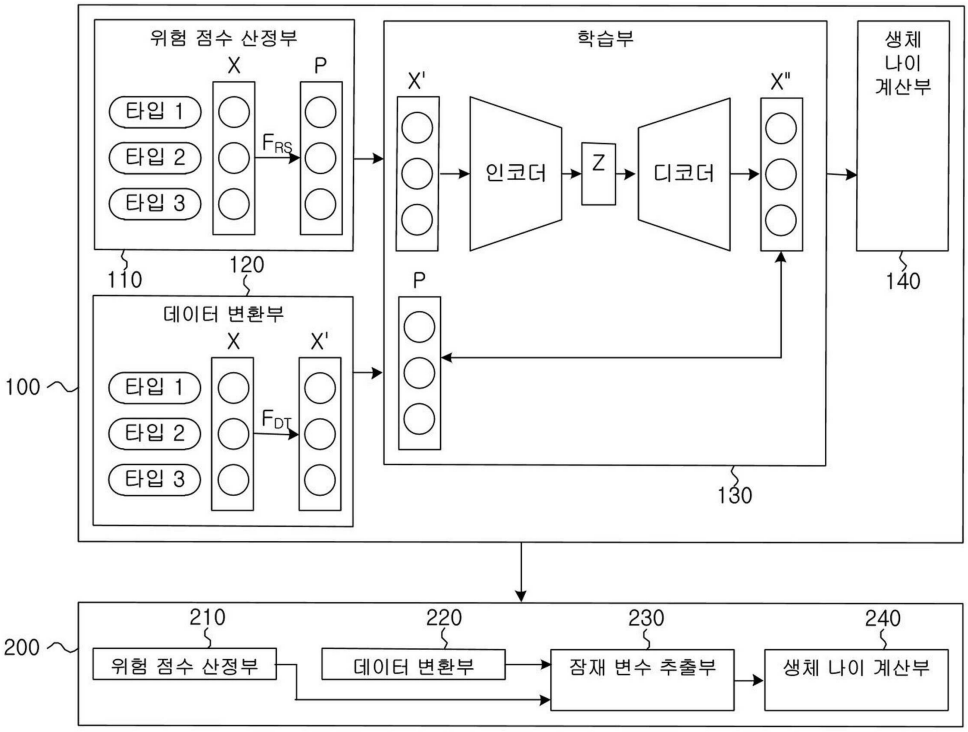
(54) 발명의 명칭 **딥러닝을 기반으로 하는 생체 나이 추정 장치 및 방법**

(57) 요 약

본 발명은 딥러닝을 기반으로 하는 생체 나이 추정 장치 및 방법에 관한 것이다. 본 발명에 따르면, 훈련부와 평가부를 포함하는 생체 나이 추정 장치가 개시된다. 훈련부는 훈련 건강 검진 데이터를 확률 분포에 기반하여 위험 점수 데이터로 변환하고, 훈련 건강 검진 데이터를 일정한 방향성을 가지는 데이터로 변환하여 변환 데이터를

(뒷면에 계속)

대 표 도



생성하며, 위험 점수 데이터와 변환 데이터를 이용하여 오토인코더 파라미터를 결정한다. 평가부는 오토인코더 파라미터를 이용하여 평가 대상 개인의 건강 상태 데이터를 부호화하여 생체 나이를 추정하는 평가부를 포함한다.

본 발명에 따르면, 실제의 건강 상태 데이터뿐만 아니라 위험 정도를 추가로 고려하여 생체 나이를 계산하므로, 개인의 건강 상태를 보다 잘 반영하고 개인의 전반적인 건강 상태와 위험도를 쉽고 종합적으로 파악할 수 있는 생체 나이가 얻어질 수 있다.

(52) CPC특허분류
G16H 10/60 (2021.08)
G16H 50/20 (2018.01)
G16H 50/50 (2018.01)
G16H 50/70 (2018.01)

(72) 발명자
김수현
울산광역시 울주군 언양읍 유니스트길 50
김한결
울산광역시 울주군 언양읍 유니스트길 50
임동철
울산광역시 울주군 언양읍 유니스트길 50
서현우
울산광역시 울주군 언양읍 유니스트길 50
이은솔
서울특별시 강남구 봉은사로 640(삼성동)

명 세 서

청구범위

청구항 1

훈련 건강 검진 데이터를 확률 분포에 기반하여 위험 점수 데이터로 변환하는 단계;

훈련 건강 검진 데이터를 일정한 방향성을 가지는 데이터로 변환하여 변환 데이터를 생성하는 단계;

상기 위험 점수 데이터와 상기 변환 데이터를 이용하여 오토인코더 파라미터를 결정하는 단계; 및

상기 오토인코더 파라미터를 이용하여 평가 대상 개인의 건강 상태 데이터를 부호화하여 생체 나이를 추정하는 단계를 포함하는 생체 나이 추정 방법.

청구항 2

제1항에 있어서,

상기 오토인코더 파라미터를 결정하는 단계는,

상기 변환 데이터를 부호화하여 부호화 데이터를 생성하는 단계와,

상기 부호화 데이터를 복호화 하여 복호화 데이터를 생성하는 단계와,

상기 복호화 데이터 및 상기 위험 점수 데이터 사이의 손실을 이용하여 상기 오토인코더 파라미터를 결정하는 단계를 포함하는 생체 나이 추정 방법.

청구항 3

제2항에 있어서,

상기 오토인코더 파라미터는 상기 복호화 데이터 및 상기 변환 데이터 사이의 손실을 추가로 이용하여 결정되는 생체 나이 추정 방법.

청구항 4

제3항에 있어서,

상기 오토인코더 파라미터는 상기 복호화 데이터 및 상기 위험 점수 데이터 사이의 손실 및 상기 복호화 데이터 및 상기 변환 데이터 사이의 손실의 합에 기초하여 결정되는 생체 나이 추정 방법.

청구항 5

제1항에 있어서,

상기 훈련 건강 검진 데이터를 상기 위험 점수 데이터로 변환하는 단계는,

상기 훈련 건강 검진 데이터가 따르는 가우시안 분포의 누적 확률 값을 상기 위험 점수 데이터로 결정하는 단계를 포함하는 생체 나이 추정 방법.

청구항 6

제1항에 있어서,

상기 변환 데이터를 생성하는 단계는,

상기 훈련 건강 검진 데이터의 복수의 건강 상태 변수의 각각이 복수의 방향성 중 하나를 가지는 경우에, 상기 복수의 건강 상태 변수의 각각이 상기 복수의 방향성 중 대표 방향성을 가지도록 상기 복수의 건강 상태 변수의 각각을 변환하여 복수의 변환된 건강 상태 변수를 생성하는 단계와,

상기 복수의 변환된 건강 상태 변수의 범위를 일정 범위로 조절하여 상기 변환 데이터를 생성하는 단계를 포함

하는 생체 나이 추정 방법.

청구항 7

제1항에 있어서,

상기 생체 나이를 추정하는 단계는,

상기 오토인코더 파라미터를 이용하여 평가 대상 개인의 건강 상태 데이터의 복수의 건강 상태 변수를 부호화하여 하나 이상의 잠재 변수를 생성하는 단계와,

상기 하나 이상의 잠재 변수를 이용하여 상기 생체 나이를 추정하는 단계를 포함하는 생체 나이 추정 방법.

청구항 8

제7항에 있어서,

상기 하나 이상의 잠재 변수의 개수는 상기 복수의 건강 상태 변수의 개수보다 적은 생체 나이 추정 방법.

청구항 9

제7항에 있어서,

상기 훈련 건강 검진 데이터에 속하는 개인들의 실제 나이의 분포를 추가로 이용하여 상기 생체 나이가 추정되는 생체 나이 추정 방법.

청구항 10

제9항에 있어서,

상기 평가 대상 개인의 실제 나이를 추가로 이용하여 상기 생체 나이가 추정되는 생체 나이 추정 방법.

청구항 11

훈련 건강 검진 데이터를 확률 분포에 기반하여 위험 점수 데이터로 변환하고, 훈련 건강 검진 데이터를 일정한 방향성을 가지는 데이터로 변환하여 변환 데이터를 생성하며, 상기 위험 점수 데이터와 상기 변환 데이터를 이용하여 오토인코더 파라미터를 결정하는 훈련부; 및

상기 오토인코더 파라미터를 이용하여 평가 대상 개인의 건강 상태 데이터를 부호화하여 생체 나이를 추정하는 평가부를 포함하는 생체 나이 추정 장치.

청구항 12

제11항에 있어서,

상기 훈련부는 상기 변환 데이터를 부호화하여 부호화 데이터를 생성하고, 상기 부호화 데이터를 복호화하여 복호화 데이터를 생성하며, 상기 복호화 데이터 및 상기 위험 점수 데이터 사이의 손실을 이용하여 상기 오토인코더 파라미터를 결정하는 생체 나이 추정 장치.

청구항 13

제12항에 있어서,

상기 훈련부는 상기 복호화 데이터 및 상기 변환 데이터 사이의 손실을 추가로 이용하여 상기 오토인코더 파라미터를 결정하는 생체 나이 추정 장치.

발명의 설명

기 술 분 야

[0001] 본 발명은 딥러닝을 기반으로 하는 생체 나이 추정 장치 및 방법에 관한 것이며, 구체적으로 딥러닝 기반의 표현형 학습을 통해 생체 나이를 추정하는 장치 및 방법에 관한 것이다.

배 경 기 술

[0002] 생체 나이(Biological Age, BA)는 실제 나이와 생체 나이는 다를 수 있다는 가정으로부터 출발하여 실제 나이와는 별개로 신체의 건강 및 노화 정도를 수치로 나타낸 것이다. 예컨대, 어떤 인물의 실제 나이가 40세이더라도, 생체 나이가 60세라고 한다면 건강이 20세만큼 좋지 않음을 의미하고, 반면 생체 나이가 20세라고 한다면 건강이 20세만큼 좋음을 의미할 수 있다.

[0003] 종래의 생체 나이 관련한 연구로서 다중 선형 회귀(Multiple Linear Regression, MLR) 분석, 주성분 분석(Principal Component Analysis, PCA), Klemera-Doubal Method (KDM), 코호트 데이터의 사망률/질병률을 이용한 검증 분석 등이 있다.

[0004] 그러나, 대부분의 종래 연구에서는 실제 데이터 수치들이 가지는 위험 정도에 생체 나이가 반영되지 않으며, 데이터 그 자체의 수치만을 활용하여 생체 나이가 계산되는 한계가 있고, 데이터의 분산을 최대화하는 방향으로 차원을 축소하는 PCA 기법이 생체 나이의 생성에 적합한가에 대한 의문점이 존재하고 있다. 실제 나이를 예측하는 방식의 회귀 분석으로 산출되는 생체 나이(BA)는 실제 나이에 편향되는 특성을 가지는 한계가 있다. 한편, KDM의 경우 다양한 가정들이 만족되지 못할 경우, 그 결과의 신뢰성이 매우 떨어지는 문제가 있다. 또한, 실제로 생체 나이를 대표하는 지표가 없기 때문에 생체 나이의 검증이 어려운 한계가 있다.

[0005] 본 발명의 배경이 되는 기술은 한국공개특허 제10-2016-0036954호(2016.04.05 공개)에 개시되어 있다.

발명의 내용

해결하려는 과제

[0006] 본 발명이 이루고자 하는 기술적 과제는 새로운 머신 러닝 기법을 적용하여 개인의 건강상태를 보다 잘 반영할 수 있는 생체 나이 모델을 제시하고, 딥러닝 기반의 표현형 학습을 통해 생체 나이를 산출하는 것을 제시하는 것에 있다.

과제의 해결 수단

[0007] 본 발명의 일 실시예에 따른 생체 나이 추정 방법은, 훈련 건강 검진 데이터를 확률 분포에 기반하여 위험 점수 데이터로 변환하는 단계; 훈련 건강 검진 데이터를 일정한 방향성을 가지는 데이터로 변환하여 변환 데이터를 생성하는 단계; 상기 위험 점수 데이터와 상기 변환 데이터를 이용하여 오토인코더 파라미터를 결정하는 단계; 및 상기 오토인코더 파라미터를 이용하여 평가 대상 개인의 건강 상태 데이터를 부호화하여 생체 나이를 추정하는 단계를 포함할 수 있다.

[0008] 상기 오토인코더 파라미터를 결정하는 단계는, 상기 변환 데이터를 부호화하여 부호화 데이터를 생성하는 단계와, 상기 부호화 데이터를 복호화 하여 복호화 데이터를 생성하는 단계와, 상기 복호화 데이터 및 상기 위험 점수 데이터 사이의 손실을 이용하여 상기 오토인코더 파라미터를 결정하는 단계를 포함할 수 있다.

[0009] 상기 오토인코더 파라미터는, 상기 복호화 데이터 및 상기 변환 데이터 사이의 손실을 추가로 이용하여 결정될 수 있다.

[0010] 상기 오토인코더 파라미터는, 상기 복호화 데이터 및 상기 위험 점수 데이터 사이의 손실 및 상기 복호화 데이터 및 상기 변환 데이터 사이의 손실의 합에 기초하여 결정될 수 있다.

[0011] 상기 훈련 건강 검진 데이터를 상기 위험 점수 데이터로 변환하는 단계는, 상기 훈련 건강 검진 데이터가 따르는 가우시안 분포의 누적 확률 값을 상기 위험 점수 데이터로 결정하는 단계를 포함할 수 있다.

[0012] 상기 변환 데이터를 생성하는 단계는, 상기 훈련 건강 검진 데이터의 복수의 건강 상태 변수의 각각이 복수의 방향성 중 하나를 가지는 경우에, 상기 복수의 건강 상태 변수의 각각이 상기 복수의 방향성 중 대표 방향성을 가지도록 상기 복수의 건강 상태 변수의 각각을 변환하여 복수의 변환된 건강 상태 변수를 생성하는 단계와, 상기 복수의 변환된 건강 상태 변수의 범위를 일정 범위로 조절하여 상기 변환 데이터를 생성하는 단계를 포함할 수 있다.

[0013] 상기 생체 나이를 추정하는 단계는, 상기 오토인코더 파라미터를 이용하여 평가 대상 개인의 건강 상태 데이터의 복수의 건강 상태 변수를 부호화하여 하나 이상의 잠재 변수를 생성하는 단계와, 상기 하나 이상의 잠재 변수를 이용하여 상기 생체 나이를 추정하는 단계를 포함할 수 있다.

[0014] 상기 하나 이상의 잠재 변수의 개수는 상기 복수의 건강 상태 변수의 개수보다 적을 수 있다.

[0015] 상기 훈련 건강 검진 데이터에 속하는 개인들의 실제 나이의 분포를 추가로 이용하여 상기 생체 나이가 추정될 수 있다.

[0016] 상기 평가 대상 개인의 실제 나이를 추가로 이용하여 상기 생체 나이가 추정될 수 있다.

[0017] 본 발명의 일 실시예에 따른 생체 나이 추정 장치는, 훈련 건강 검진 데이터를 확률 분포에 기반하여 위험 점수 데이터로 변환하고, 훈련 건강 검진 데이터를 일정한 방향성을 가지는 데이터로 변환하여 변환 데이터를 생성하며, 상기 위험 점수 데이터와 상기 변환 데이터를 이용하여 오토인코더 파라미터를 결정하는 훈련부; 및 상기 오토인코더 파라미터를 이용하여 평가 대상 개인의 건강 상태 데이터를 부호화하여 생체 나이를 추정하는 평가부를 포함할 수 있다.

[0018] 상기 훈련부는 상기 변환 데이터를 부호화하여 부호화 데이터를 생성하고, 상기 부호화 데이터를 복호화하여 복호화 데이터를 생성하며, 상기 복호화 데이터 및 상기 위험 점수 데이터 사이의 손실을 이용하여 상기 오토인코더 파라미터를 결정할 수 있다.

[0019] 상기 훈련부는 상기 복호화 데이터 및 상기 변환 데이터 사이의 손실을 추가로 이용하여 상기 오토인코더 파라미터를 결정할 수 있다.

발명의 효과

[0020] 본 발명에 따르면, 실제의 건강 상태 데이터뿐만 아니라 위험 정도를 추가로 고려하여 생체 나이를 계산하므로, 개인의 건강 상태를 보다 잘 반영하고 개인의 전반적인 건강 상태와 위험도를 쉽고 종합적으로 파악할 수 있는 생체 나이가 얻어질 수 있다.

도면의 간단한 설명

[0021] 도 1은 본 발명의 실시예에 따른 생체 나이 추정 장치를 보여준다.

도 2는 본 발명의 실시예에 따른 생체 나이 추정 시스템을 보여준다.

도 3은 본 발명의 실시예에 따른 훈련부 및 평가부를 보여주는 블록도이다.

도 4는 본 발명의 실시예에 따른 훈련부의 동작 방법을 보여주는 흐름도이다.

도 5는 데이터 변환부의 동작을 보여주는 알고리즘을 나타낸다.

도 6은 데이터 변환부에 의해 수행되는 데이터 변환의 예를 보여준다.

도 7은 본 발명의 실시예에 따른 학습부를 보여주는 블록도이다.

도 8은 본 발명의 실시예에 따른 학습부의 동작을 보여주는 알고리즘이다.

도 9는 본 발명의 실시예에 따른 생체 나이 계산부의 동작을 보여주는 흐름도이다.

도 10은 본 발명의 실시예에 따른 평가부의 동작 방법을 보여주는 흐름도이다.

도 11은 본 발명의 실시예에 따른 평가부의 위험 점수 산정부의 동작 방법을 보여주는 흐름도이다.

도 12는 본 발명의 실시예에 따른 평가부의 데이터 변환부의 동작 방법을 보여주는 흐름도이다.

도 13은 본 발명의 실시예에 따른 평가부의 잠재 변수 추출부의 동작 방법을 보여주는 흐름도이다.

도 14는 본 발명의 실시예에 따른 평가부의 생체 나이 계산부의 동작 방법을 보여주는 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0022] 그러면 첨부한 도면을 참고로 하여 본 발명의 실시 예에 대하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시 예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

[0023] 이하에서는 도 1 내지 도 14를 참고하여 본 발명의 실시예에 따른 생체 나이 추정 장치 또는 생체 나이 추정 시스템을 설명한다.

[0024] 도 1은 본 발명의 실시예에 따른 생체 나이 추정 장치를 보여준다.

[0025] 도 1에 도시된 바와 같이, 본 발명의 실시예에 따른 생체 나이 추정 장치(1)는 훈련부(100), 평가부(200), 생체 나이 표시부(300)를 포함한다.

[0026] 훈련부(100)는 훈련 건강 검진 데이터에 기반하여 생체 나이를 추정할 수 있는 파라미터를 생성한다.

[0027] 평가부(200)는 훈련부(100)에 의해 생성된 파라미터 및 평가 대상 개인의 건강 상태 정보를 이용하여 평가 대상 개인의 생체 나이를 추정한다.

[0028] 생체 나이 표시부(300)는 평가부(200)가 평가 대상 개인의 추정된 생체 나이를 표시한다.

[0029] 도 2는 본 발명의 실시예에 따른 생체 나이 추정 시스템을 보여준다.

[0030] 도 2에 도시된 바와 같이, 본 발명의 실시예에 따른 생체 나이 추정 시스템(2)은 하나 이상의 단말 장치(10), 생체 나이 추정 서버(20) 및 네트워크(30)를 포함한다.

[0031] 각 단말 장치(10)는 통신부(11) 및 생체 나이 표시부(300)를 포함한다. 통신부(11)는 네트워크(30)를 통해 생체 나이 추정 서버(20)에 접속하며, 생체 나이 추정 서버(20)에 평가 대상 개인의 건강 상태 정보를 생체 나이 추정 서버(20)에 제공하며, 생체 나이 추정 서버(20)로부터 추정된 생체 나이를 수신한다. 생체 나이 추정 서버(20)는 수신한 생체 나이를 표시한다.

[0032] 생체 나이 추정 서버(20)는 통신부(21), 훈련부(100) 및 평가부(200)를 포함한다.

[0033] 통신부(21)는 네트워크(30)를 통해 단말 장치(10)와 통신하며, 단말 장치(10)로부터 평가 대상 개인의 건강 상태 정보를 수신하며, 평가된 생체 나이를 단말 장치(10)에 제공한다.

[0034] 훈련부(100)는 훈련 건강 검진 데이터에 기반하여 생체 나이를 추정할 수 있는 파라미터를 생성한다.

[0035] 평가부(200)는 훈련부(100)에 의해 생성된 파라미터 및 평가 대상 개인의 건강 상태 정보를 이용하여 평가 대상 개인의 생체 나이를 추정한다.

[0036] 도 3은 본 발명의 실시예에 따른 훈련부 및 평가부를 보여주는 블록도이다.

[0037] 도 3에 도시된 바와 같이, 훈련부(100)는 위험 점수 산정부(110), 데이터 변환부(120), 학습부(130), 및 생체 나이 계산부(140)을 포함하고, 평가부(200)는 위험 점수 산정부(210), 데이터 변환부(220), 잠재 변수 추출부(230), 및 생체 나이 계산부(240)를 포함한다.

[0038] 훈련부(100)의 각 구성요소 및 평가부(200)의 각 구성요소에 대하여는 도 4 내지 도 14를 참고하여 상세히 설명한다.

[0039] 도 4는 본 발명의 실시예에 따른 훈련부의 동작 방법을 보여주는 흐름도이다.

[0040] 도 4에 도시된 바와 같이, 위험 점수 산정부(110)는 훈련에 사용될 훈련 건강 검진 데이터(X)의 각 변수의 값을 위험 점수(P)로 변환한다(S401).

[0041] 여기서, 훈련 건강 검진 데이터는 N명의 훈련 데이터 제공자 및 복수의 변수에 대한 건강 상태 정보를 포함할 수 있다. N이 클수록 훈련의 효과는 좋아지며, 복수의 변수는 체질량 지수(Body Mass Index, BMI), 허리 둘레(Waist), 요산 (Uric acid) 수치, 고밀도 지단백 (high-density lipoprotein, HDL) 콜레스테롤 수치, 알부민(Albumin) 수치, 중성 지방 (triglyceride, TG) 수치 등을 포함할 수 있으며 이에 한정되지 않는다.

[0042] 구체적으로, 위험 점수 산정부(110)는 훈련 건강 검진 데이터의 각 건강 상태 변수의 타입을 결정한다. 각 건강 상태 변수는 3가지 타입 중 하나로 결정될 수 있다. 예컨대, 변수의 값이 커질수록 건강 상태가 나쁘다는 것을 나타내는 경우, 해당 변수는 타입 1으로 결정되고, 변수의 값이 작아질수록 건강 상태가 나쁘다는 것을 나타내는 경우, 해당 변수는 타입 2로 결정되고, 변수의 값이 평균보다 커지거나 작아질수록 건강 상태가 나쁘다는 것을 나타내는 경우, 해당 변수는 타입 3으로 결정될 수 있다.

[0043] 위험 점수 산정부(110)는 각 건강 상태 변수의 타입에 기초하여 훈련 건강 검진 데이터의 각 건강 상태 변수의 평균과 분산을 결정한다. 건강 상태 변수의 타입이 타입 1 또는 타입 2일 경우, 해당 건강 상태 변수가 반 가우

시안 분포(half-Gaussian distribution)를 따른다는 가정하에, 평균과 분산이 결정된다.

[0044] 반 가우시안 분포는 정상 가우시안 분포의 평균에서 접힌 형태를 가지는 분포를 의미한다. 건강 상태 변수의 타입이 타입 3일 경우, 해당 건강 상태 변수가 보통 가우시안 분포(ordinary Gaussian distribution)를 따른다는 가정하에, 평균과 분산이 결정된다. 한편, 위험 점수 산정부(110)는 각 변수의 정상 수치 범위를 확인하고, 정상 수치 범위를 이용하여 평균을 구할 수도 있다.

[0045] 예컨대, BMI 변수의 경우, 정상 수치 범위가 18.5에서 25에 해당하므로, BMI의 평균은 아래의 수학식 1에 따라 계산될 수도 있다.

수학식 1

[0046] $$\mu = \frac{(18.5+25)}{2} = 21.75$$

[0047] 위험 점수 산정부(110)는 각 건강 상태 변수가 위에서 결정된 평균과 분산을 가지는 가우시안 분포를 따른다는 가정하에, 이 가우시안 분포의 누적 확률 값으로 건강 상태 변수의 값을 위험 점수로 변환한다.

[0048] 건강 상태 변수의 타입이 타입 1 또는 타입 2일 경우, 위험 점수 산정부(110)는 해당 건강 상태 변수가 위에서 결정된 평균과 분산을 가지는 반 가우시안 분포를 따른다는 가정하에, 이 반 가우시안 분포의 누적 확률 값으로 건강 상태 변수의 값을 위험 점수로 변환할 수 있다. 건강 상태 변수의 타입이 타입 3일 경우, 위험 점수 산정부(110)는 해당 건강 태 변수가 위에서 결정된 평균과 분산을 가지는 보통 가우시안 분포를 따른다는 가정하에, 이 가우시안 분포의 누적 확률 값으로 건강 상태 변수의 값을 위험 점수로 변환할 수 있다.

[0049] 다음, 데이터 변환부(120)는 훈련 건강 검진 데이터(X)를 일정한 방향성을 가지는 데이터로 변환하여 변환 데이터(X')를 생성한다(S403).

[0050] 이하에서는 도 5와 도 6을 참고하여 데이터 변환부(120)의 동작을 상세히 설명한다.

[0051] 도 5는 데이터 변환부(120)의 동작을 보여주는 알고리즘을 나타낸다.

[0052] 각 건강 상태 변수에 대하여 데이터 변환부(120)는 훈련 건강 검진 데이터의 평균과 표준 편차를 구하고, 평균과 표준 편차를 이용하여 훈련 건강 검진 데이터를 정규화한다(S501).

[0053] 각 건강 상태 변수에 대하여, 데이터 변환부(120)는 정규화된 건강 검진 데이터의 방향성을 통일시킨다. 이를 위하여, 건강 상태 변수의 타입이 타입 2인 경우, 데이터 변환부(120)는 건강 검진 데이터에 -1을 곱할 수 있다(S503). 건강 상태 변수의 타입이 타입 3인 경우, 데이터 변환부(120)는 건강 검진 데이터에 절대값을 취할 수 있다(S505). 건강 상태 변수의 타입이 타입 1인 경우, 데이터 변환부(120)는 검진 데이터를 바이패스할 수 있다(S507). 이처럼, 훈련 건강 검진 데이터의 복수의 건강 상태 변수가 복수의 방향성 중 하나를 가지는 경우에, 복수의 건강 상태 변수의 각각이 복수의 방향성 중 하나인 대표 방향성을 가지도록 복수의 건강 상태 변수의 각각이 변환될 수 있다.

[0054] 각 건강 상태 변수에 대하여, 데이터 변환부(120)는 방향성이 통일된 건강 검진 데이터가 일정한 범위 이내에 속하도록 건강 검진 데이터를 변환하여 변환 데이터(X')를 생성할 수 있다(S509). 예컨대, 데이터 변환부(120)는 시그모이드 (sigmoid) 함수를 통해 건강 검진 데이터가 0과 1사이의 범위에 속하도록 건강 검진 데이터를 변환할 수 있다.

[0055] 도 6은 데이터 변환부(120)에 의해 수행되는 데이터 변환의 예를 보여준다.

[0056] 도 6에 도시된 바와 같이, 건강 검진 데이터는 N명의 훈련 데이터 제공자 및 복수의 변수에 대한 건강 상태 정보를 포함하며, 복수의 변수는 Waist, Uric acid 수치, HDL 수치, Albumin 수치, TG 수치를 포함한다.

[0057] Waist와 Uric acid 수치는 타입 1의 변수로 간주될 수 있다. Waist를 살펴보면, Waist의 초기 입력 데이터는 87.0, 84.0, 87.0, ..., 86.0이며, 1.2, 1.0, 1.2, ..., 1.1로 정규화된다. Waist는 타입 1에 해당하므로, 데이터의 방향성 통일을 위하여 바이패스되고, 이전과 동일한 데이터인 1.2, 1.0, 1.2, ..., 1.1가 산출된다. 이후, Waist 데이터는 시그모이드 함수를 통해 0 내지 1 사이의 범위에 속하도록 변환되어, 0.567, 0.471, 0.567,

..., 0.535가 산출된다.

[0058] HDL 수치는 타입 2의 변수로 간주될 수 있다. HDL을 살펴보면, HDL의 초기 입력 데이터는 52.0, 46.0, 39.0, ..., 36.0이며, 1.3, 0.7, 0.1, ..., -0.1로 정규화된다. HDL는 타입 2에 해당하므로, 데이터의 방향성 통일을 위하여 -1이 곱해진 데이터인 -1.3, -0.7, -0.1, ..., 0.1이 산출된다. 이로써, HDL은 타입 1의 방향성을 가질 수 있다. 이후, HDL 데이터는 시그모이드 함수를 통해 0 내지 1 사이의 범위에 속하도록 변환되어, 0.493, 0.614, 0.738, ..., 0.783이 산출된다.

[0059] Albumin 수치와 TG 수치는 타입 3의 변수로 간주될 수 있다. Albumin를 살펴보면, Albumin의 초기 입력 데이터는 5.0, 4.7, 4.7, ..., 4.9이며, 0.1, -0.1, -0.1, ..., 0.0으로 정규화된다. Albumin는 타입 3에 해당하므로, 데이터의 방향성 통일을 위하여 절대값이 취해진 데이터인 0.1, 0.1, 0.1, ..., 0.0가 산출된다. 이로써, Albumin은 타입 1의 방향성을 가질 수 있다. 이후, Albumin 데이터는 시그모이드 함수를 통해 0 내지 1 사이의 범위에 속하도록 변환되어, 0.847, 0.635, 0.635, ..., 0.791이 산출된다.

[0060] 다시 도 4를 설명한다.

[0061] 다음에, 학습부(130)는 위험 점수(P)와 변환 데이터(X')를 이용하여 오토인코더 파라미터 셋($\boldsymbol{\theta}$)을 결정한다(S405).

[0062] 오토인코더(Autoencoder)는 인코더 네트워크와 디코더 네트워크로 이루어진 신경망 모델이며, 입력 데이터를 낮은 차원으로 투영한 뒤 다시 복원하는 과정에서 복원 오차를 낮추는 방향으로 신경망을 훈련하여 정보 손실을 최소화하면서도 저차원의 효율적인 특징 표현을 가능하게 한다.

[0063] 오토인코더 파라미터 셋($\boldsymbol{\theta}$)은 적어도 하나의 인코더 파라미터와 적어도 하나의 디코더 파라미터를 포함할 수 있다. 적어도 하나의 인코더 파라미터는 인코더 가중치 행렬(W)과 바이어스 벡터(b)를 포함할 수 있다. 적어도 하나의 디코더 파라미터는 디코더 가중치 행렬과 바이어스 벡터(b_h)를 포함할 수 있다. 디코더 가중치 행렬은 행렬(W^T)일 수 있다.

[0064] 다음은 도 7 및 도 8을 이용하여 오토인코더를 이용한 학습부(130)를 구체적으로 설명한다.

[0065] 도 7은 본 발명의 실시예에 따른 학습부(130)를 보여주는 블록도이다.

[0066] 도 7에 도시된 바와 같이, 본 발명의 실시예에 따른 학습부(130)는 인코더(131), 디코더(133), 오토인코더 파라미터 결정부(135)를 포함한다.

[0067] 도 8은 본 발명의 실시예에 따른 학습부의 동작을 보여주는 알고리즘이다.

[0068] 도 8에서 보여지는 바와 같이, 인코더(131)는 변환 데이터(X')를 선형 함수(linear function)을 통해 데이터($\widetilde{\boldsymbol{X'}}$)를 생성한다(S801). 다음, 인코더(131)는 생성된 데이터($\widetilde{\boldsymbol{X'}}$)를 인코딩 파라미터(W, b)를 이용하여 활성 함수(activation function)인 fa를 통해 인코딩하여 잠재 변수 벡터(h)를 생성한다(S803). 여기서 인코딩 파라미터인 W는 가중치 행렬(weight matrix)이고, b는 바이어스 벡터(bias vector)이다. 활성 함수인 fa는 시그모이드(sigmoid) 또는 Rectified Linear Unit (ReLU) 일 수 있으나, 이에 한정될 필요는 없다.

[0069] 디코더(133)는 잠재 변수 벡터(h)를 디코딩 파라미터(W^T, b_h)를 이용하여 활성 함수인 fa를 통해 디코딩하여 복원 데이터(X")를 출력한다(S805).

[0070] 오토인코더 파라미터 결정부(135)는 손실 값($\boldsymbol{\mathcal{L}}$)을 계산한다(S807). 도 8에서 보여지는 바와 같이, 손실 값($\boldsymbol{\mathcal{L}}$)은 입력 데이터(X')와 복원 데이터(X") 사이의 손실($\boldsymbol{\mathcal{L}(X', X'')}$)과 위험 점수(P)와 복원 데이터(X") 사이의 손실($\boldsymbol{\mathcal{L}(X'', P)}$)의 합으로 구해질 수 있다. 손실 함수 $\boldsymbol{\mathcal{L}}$(x,y)는 평균 제곱 오차 손실 함수(mean square error loss function)나 쿨백-라이블러 발산 손실 함수(Kullback-Leibler divergence loss function)일 수 있으나, 이에 한정될 필요는 없다.

[0071] 오토인코더 파라미터 결정부(135)는 (iter+1)번째 손실 값($\boldsymbol{\mathcal{L}^{(iter+1)}}$)과 iter 번째 손실 값($\boldsymbol{\mathcal{L}^{(iter)}}$)의 차이가

소정의 값(1e-5)보다 작을 때의 오토인코더 파라미터 셋($\boldsymbol{\theta}$)을 결정한다(S809).

[0072] 다시 도 4를 설명한다.

[0073] 이후, 학습부(130)는 결정된 오토인코더 파라미터 셋을 이용하여 변환 데이터(X')를 인코딩하여 잠재 변수 셋(Z)을 추출한다(S407). 잠재 변수 셋(Z)은 하나 이상의 잠재 변수 l_k (k = 1, 2, ..., K, 여기서 k는 잠재 변수의 차원을 의미함)를 포함할 수 있다.

[0074] 잠재 변수의 차원은 잠재 변수의 개수를 의미하고, 변환 데이터(X') 상의 건강 상태 변수의 차원은 건강 상태 변수의 개수를 의미할 수 있다. 학습부(130)의 학습의 결과 추출되는 잠재 변수의 차원은 건강 상태 변수의 차원보다 낮을 수 있다. 다시 말하면, 학습부(130)의 학습의 결과 추출되는 잠재 변수의 개수는 건강 상태 변수의 개수보다 작을 수 있다.

[0075] 생체 나이 계산부(140)는 잠재 변수 셋(Z)을 이용하여 생체 나이를 추정한다(S409).

[0076] 다음은 도 9를 이용하여 생체 나이 계산부(140)의 동작을 구체적으로 설명한다.

[0077] 도 9는 본 발명의 실시예에 따른 생체 나이 계산부의 동작을 보여주는 흐름도이다.

[0078] 도 9에 도시된 바와 같이, 생체 나이 계산부(140)는 각 차원의 잠재 변수의 평균(mean(l_k))과 표준 편차(sd(l_k))를 구한다(S901).

[0079] 생체 나이 계산부(140)는 구해진 잠재 변수 평균(mean(l_k))과 잠재 변수 표준 편차(sd(l_k))를 이용하여 생체 나이 점수(BA score, BAS)를 결정한다(S903). 이때, i번째 개인의 생체 나이 점수(BAS_i)는 아래의 수학식 2을 이용하여 구해질 수 있다.

수학식 2

[0080]
$$BAS_i = \frac{1}{K}\sum_{k=1}^{K}\frac{\{l_k - \mathrm{mean}(l_k)\}}{\mathrm{sd}(l_k)}$$

[0081] 단계(S901)과 단계(S903)의 예시를 표로 나타내면 아래와 같다. 표 1은 5명에 대한 3개의 건강 상태 변수에 대한 훈련 건강 검진 데이터를 보여주고, 표 2는 잠재 변수들의 값과 각 잠재 변수의 평균 및 표준 편차, 그리고 i번째 개인의 생체 나이 점수(BAS_i)를 보여준다.

표 1

[0082]

ID	Uric acid	HDL	TG
1	5.4	52.0	113.0
2	5.9	46.0	166.0
3	6.0	39.0	124.0
4	5.9	36.0	127.0
5	5.9	36.0	127.0

표 2

[0083]

ID (i)	l_1	l_2	$\mathrm{mean}(l_1)$	$\mathrm{sd}(l_1)$	$\mathrm{mean}(l_2)$	$\mathrm{sd}(l_2)$	BAS_i
1	0.1	3.8	0.14	0.05	1.4	1.22	1.17
2	0.2	1.2					1.04
3	0.2	0.8					0.71
4	0.1	0.6					-1.46
5	0.1	0.6					-1.46

[0084] 생체 나이 계산부(140)는 훈련 데이터 제공 그룹에 속한 개인들의 실제 나이의 평균($\overline{CA}$)과 표준 편차(sd(CA))를 구한다(S905).

[0085] 생체 나이 계산부(140)는 훈련 데이터 제공 그룹에 속한 개인들의 실제 나이의 평균($\overline{CA}$)과 표준 편차(sd(CA))를 이용하여 각 훈련 데이터 제공 개인의 생체 나이 점수(BAS)를 보정하여 보정된 생체 나이 점수(Corrected Biological Age Score, CBAS)를 구한다(S907). i번째 개인의 보정된 생체 나이 점수($CBAS_i$)는 아래의 수학식 3에 따라 결정될 수 있다.

수학식 3

[0086] $$CBAS_i = BAS_i \times \mathrm{sd}(CA) + \overline{CA}$$

[0087] 단계(S905)과 단계(S907)의 예시를 표로 나타내면 아래와 같다.

표 3

[0088]

ID	BAS_i	CA_i	$\overline{CA}$	$\mathrm{sd}(CA)$	$CBAS_i$
1	1.17	30	38.6	7.14	1.17x7.14+38.6 = 46.9
2	1.04	40			1.04x7.14+38.6 = 46
3	0.71	32			0.71x7.14+38.6 = 43.7
4	-1.46	50			-1.46x7.14+38.6 = 28.2
5	-1.46	41			-1.46x7.14+38.6 = 28.2

[0089] 생체 나이 계산부(140)는 훈련 데이터 제공 개인의 실제 나이(CA_i)와 보정된 생체 나이 점수(CBAS) 사이의 일차 방정식의 절편 a와 계수 b를 구한다(S909). 절편 a와 계수 b는 아래의 수학식 4에 따라 결정될 수 있다.

수학식 4

[0090] $$CBAS_i = a + bCA_i$$

[0091] 생체 나이 계산부(140)는 훈련 데이터 제공 그룹에 속한 i번째 개인의 보정된 생체 나이 점수($CBAS_i$), 훈련 데이터 제공 그룹에 속한 i번째 개인의 실제 나이(CA_i), 훈련 데이터 제공 그룹에 속한 개인들의 실제 나이의 평균($\overline{CA}$), 및 CA_i와 $CBAS_i$ 사이의 일차 방정식의 계수 b를 이용하여 i번째 개인의 생체 나이를 결정한다(S911). i번째 개인의 생체 나이(BA_i)는 아래의 수학식 5에 따라 결정될 수 있다.

수학식 5

[0092] $$BA_i = CBAS_i + (CA_i - \overline{CA})(1 - b)$$

[0093] 단계(S909)과 단계(S911)의 예시를 표로 나타내면 아래와 같다.

표 4

[0094]

CA	CBAS	a	b	BA
30	46.9	74.6	-0.93	30.3
40	46			48.7
32	43.7			31
50	28.2			50.2
41	28.2			32.8

[0095] 도 10은 본 발명의 실시예에 따른 평가부(200)의 동작 방법을 보여주는 흐름도이다.

[0096] 위험 점수 산정부(210)는 평가 대상 개인의 건강 상태 데이터의 각 변수의 값을 위험 점수(P)로 변환한다(S1001).

[0097] 데이터 변환부(220)는 평가 대상 개인의 건강 상태 데이터를 일정한 방향성을 가지는 데이터로 변환하여 변환 데이터(X')를 생성한다(S1003).

[0098] 잠재 변수 추출부(230)는 학습부(130)가 결정한 오토인코더 파라미터 셋을 이용하여 변환 데이터(X')를 인코딩하여 잠재 변수 셋을 추출한다(S1007).

[0099] 생체 나이 계산부(240)는 잠재 변수 추출부(230)가 추출한 잠재 변수 셋을 이용하여 생체 나이를 추정한다(S1009).

[0100] 도 11은 본 발명의 실시예에 따른 평가부(200)의 위험 점수 산정부(210)의 동작 방법을 보여주는 흐름도이다.

[0101] 위험 점수 산정부(210)는 평가 대상 개인의 건강 상태 데이터의 각 건강 상태 변수의 타입을 결정한다(S1101). 각 건강 상태 변수는 위에서 언급한 3가지 타입 중 하나로 결정될 수 있다.

[0102] 위험 점수 산정부(210)는 각 건강 상태 변수의 평균과 분산을 위험 점수 산정부(110)으로부터 수신한다(S1103).

[0103] 위험 점수 산정부(210)는 각 건강 상태 변수가 해당 평균과 분산을 가지는 가우시안 분포를 따른다는 가정하에, 이 가우시안 분포의 누적 확률 값으로 건강 상태 변수의 값을 위험 점수로 변환한다(S1105). 건강 상태 변수의 타입이 타입 1 또는 타입 2일 경우, 위험 점수 산정부(210)는 해당 건강 상태 변수가 위에서 결정된 평균과 분산을 가지는 반 가우시안 분포를 따른다는 가정하에, 이 반 가우시안 분포의 누적 확률 값으로 건강 상태 변수의 값을 위험 점수로 변환할 수 있다. 건강 상태 변수의 타입이 타입 3일 경우, 위험 점수 산정부(210)는 해당 건강 태 변수가 위에서 결정된 평균과 분산을 가지는 보통 가우시안 분포를 따른다는 가정하에, 이 가우시안 분포의 누적 확률 값으로 건강 상태 변수의 값을 위험 점수로 변환할 수 있다.

[0104] 도 12는 본 발명의 실시예에 따른 평가부(200)의 데이터 변환부(220)의 동작 방법을 보여주는 흐름도이다.

[0105] 각 건강 상태 변수에 대하여 데이터 변환부(220)는 훈련 건강 검진 데이터의 평균과 표준 편차를 데이터 변환부(120)로부터 수신한다(S1201).

[0106] 이후, 각 건강 상태 변수에 대하여 데이터 변환부(220)는 평균과 표준 편차를 이용하여 평가 대상 개인의 건강 상태 데이터를 정규화한다(S1203).

[0107] 각 건강 상태 변수에 대하여, 데이터 변환부(220)는 정규화된 건강 상태 데이터의 방향성을 통일시킨다(S1205). 이를 위하여, 건강 상태 변수의 타입이 타입 1인 경우, 데이터 변환부(220)는 건강 상태 데이터를 바이패스할 수 있다. 건강 상태 변수의 타입이 타입 2인 경우, 데이터 변환부(220)는 건강 상태 데이터에 -1을 곱할 수 있다. 건강 상태 변수의 타입이 타입 2인 경우, 데이터 변환부(220)는 건강 상태 데이터에 절대값을 취할 수 있다.

[0108] 각 건강 상태 변수에 대하여, 데이터 변환부(220)는 방향성이 통일된 건강 상태 데이터가 일정한 범위 이내에 속하도록 건강 상태 데이터를 변환하여 변환 데이터(X')를 생성할 수 있다(S1207). 예컨대, 데이터 변환부(220)는 시그모이드 (sigmoid) 함수를 통해 건강 상태 데이터가 0과 1사이의 범위에 속하도록 건강 상태 데이터를 변환할 수 있다.

[0109] 도 13은 본 발명의 실시예에 따른 평가부(200)의 잠재 변수 추출부(230)의 동작 방법을 보여주는 흐름도이다.

[0110] 잠재 변수 추출부(230)는 학습부(130)로부터 오토인코더 파라미터 셋($\boldsymbol{\theta}$)을 수신한다(S1301).

[0111] 잠재 변수 추출부(230)는 변환 데이터(X')에 선형 함수(linear function)를 적용하여 데이터($\widetilde{\boldsymbol{X'}}$)를 생성한다(S1303).

[0112] 다음, 잠재 변수 추출부(230)는 오토인코더 파라미터 셋($\boldsymbol{\theta}$) 내의 인코딩 파라미터(W, b)를 이용하여 활성 함수(activation function)인 fa를 통해 데이터($\widetilde{\boldsymbol{X'}}$)를 인코딩하여 잠재 변수 벡터 셋(Z)를 추출한다(S1305).

[0113] 도 14는 본 발명의 실시예에 따른 평가부(200)의 생체 나이 계산부(240)의 동작 방법을 보여주는 흐름도이다.

[0114] 생체 나이 계산부(240)는 훈련 건강 검진 데이터의 잠재 변수의 평균(mean(l_k))과 표준 편차(sd(l_k))를 생체 나이 계산부(140)으로부터 수신한다(S1401).

[0115] 생체 나이 계산부(240)는 훈련 건강 검진 데이터의 잠재 변수 평균(mean(l_k))과 잠재 변수 표준 편차(sd(l_k))를 이용하여 훈련 대상 개인의 생체 나이 점수(BA score, BAS)를 결정한다(S1403). 이때, 생체 나이 점수는 수학식 2을 이용하여 구해질 수 있다.

[0116] 생체 나이 계산부(240)는 훈련 데이터 제공 그룹에 속한 개인들의 실제 나이의 평균($\overline{CA}$)과 표준 편차(sd(CA))를 생체 나이 계산부(140)으로부터 수신한다(S1405).

[0117] 생체 나이 계산부(240)는 훈련 데이터 제공 그룹에 속한 개인들의 실제 나이의 평균($\overline{CA}$)과 표준 편차(sd(CA))를 이용하여 평가 대상 개인의 생체 나이 점수(BAS)를 보정하여 보정된 생체 나이 점수(Corrected Biological Age Score, CBAS)를 구한다(S1407). 보정된 생체 나이 점수는 수학식 3에 따라 결정될 수 있다.

[0118] 생체 나이 계산부(240)는 훈련 데이터 제공 개인의 실제 나이(CA_i)와 보정된 생체 나이 점수(CBAS) 사이의 일차 방정식의 절편 a와 계수 b를 생체 나이 계산부(140)으로부터 수신한다(S1409).

[0119] 생체 나이 계산부(240)는 평가 대상 개인의 보정된 생체 나이 점수($CBAS_i$), 평가 대상 개인의 실제 나이(CA_i), 훈련 데이터 제공 그룹에 속한 개인들의 실제 나이의 평균($\overline{CA}$), 및 계수 b를 이용하여 평가 대상 개인의 생체 나이를 결정한다(S1411). 평가 대상 개인의 생체 나이(BA_i)는 수학식 5에 따라 결정될 수 있다.

[0120] 본 발명은 도면에 도시된 실시 예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 다른 실시 예가 가능하다는 점을 이해할 것이다. 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 특허청구범위의 기술적 사상에 의하여 정해져야 할 것이다.

부호의 설명

[0121] 1: 생체 나이 추정 장치 2: 생체 나이 추정 시스템

10: 단말 장치 11: 통신부

20: 생체 나이 추정 서버 21: 통신부

30: 네트워크 100: 훈련부

110: 위험 점수 산정부 120: 데이터 변환부

130: 학습부 131: 인코더

133: 디코더 135: 오토인코더 파라미터 결정부

140: 생체 나이 계산부 200: 평가부

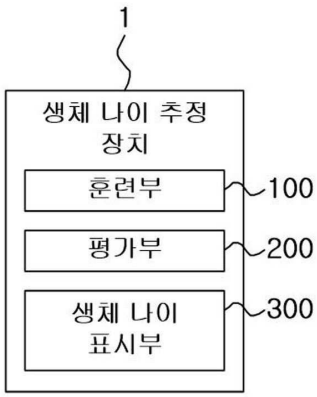
210: 위험 점수 산정부 220: 데이터 변환부

230: 잠재 변수 추출부 240: 생체 나이 계산부

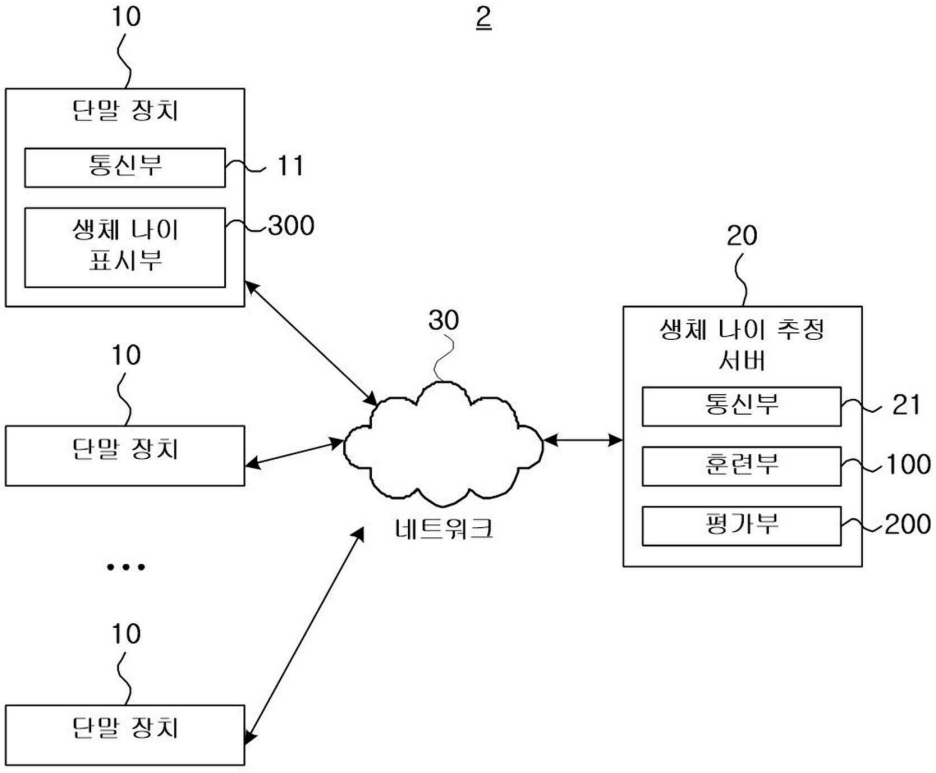
300: 생체 나이 표시부

도면

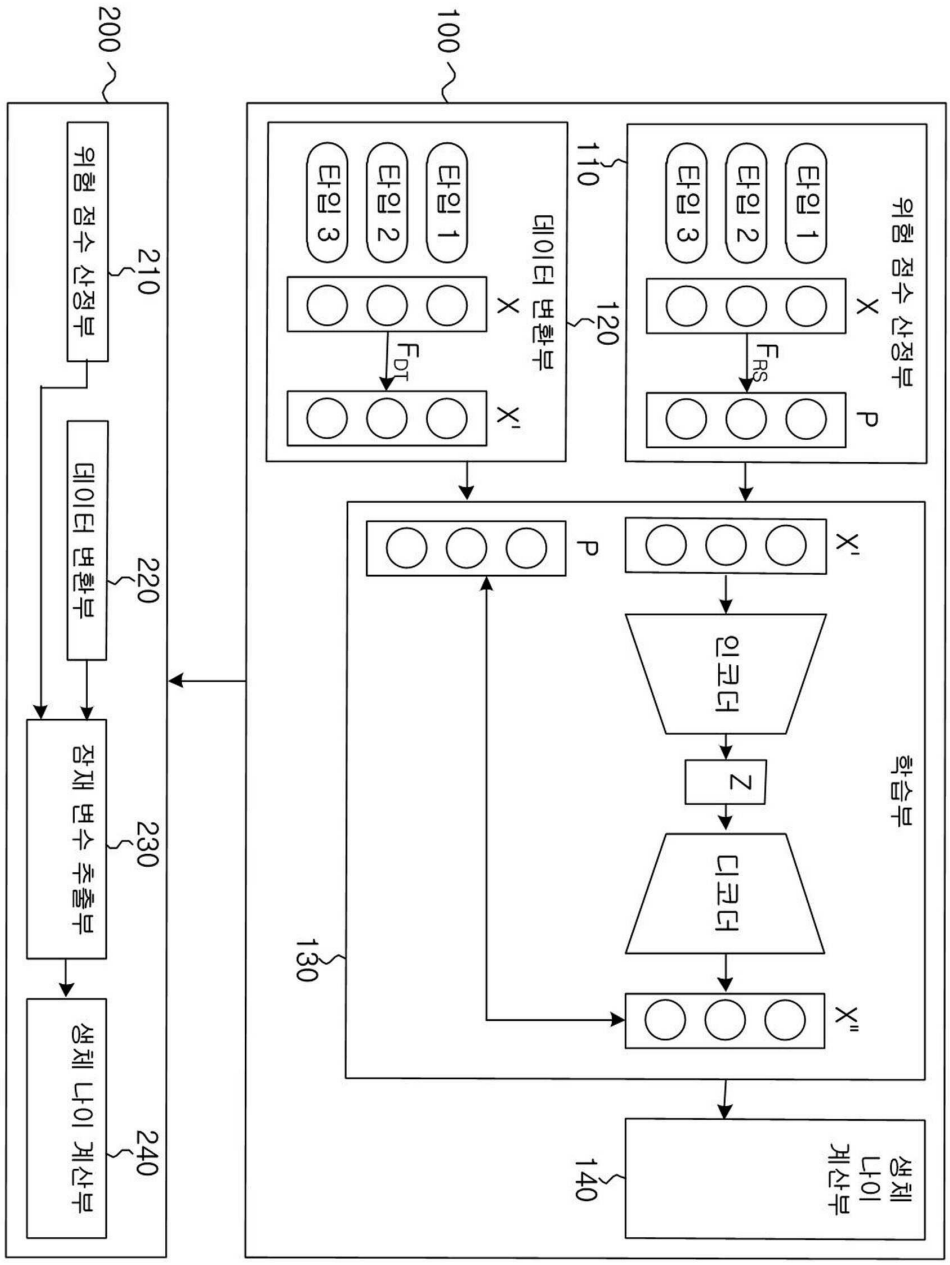
도면1


1
생체 나이 추정 장치
훈련부
100
평가부
200
생체 나이 표시부
300

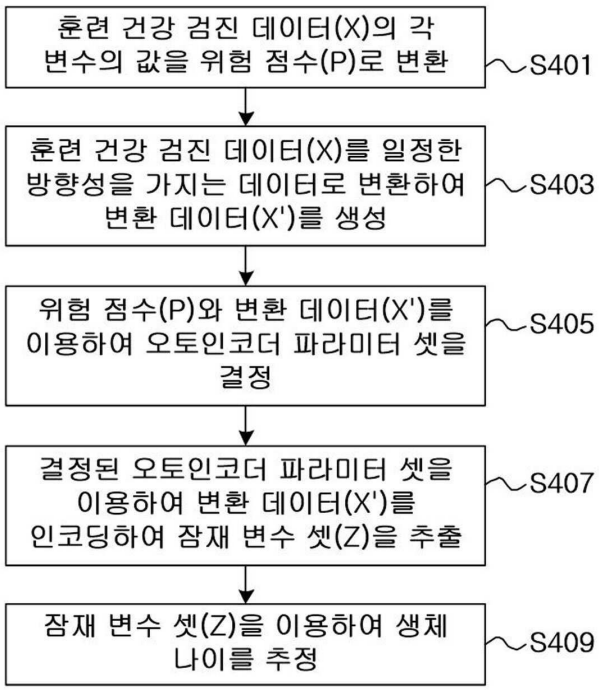
도면2


10
2
단말 장치
통신부
11
생체 나이 표시부
300
20
30
생체 나이 추정 서버
통신부
21
훈련부
100
평가부
200
10
단말 장치
네트워크
...
10
단말 장치

도면3


200
100
위험 점수 산정부
210
데이터 변환부
220
잠재 변수 추출부
230
생체 나이 계산부
240
110
120
130
140
타입 1
타입 2
타입 3
X
P
X'
인코더
Z
디코더
X''
학습부
생체 나이 계산부

도면4



도면5

Data transformation algorithm

Input : Raw data $\boldsymbol{X}$

Output : Transformed input data $\boldsymbol{X}'$

function TRANS(V_j):

V_j, $j = 1,2,\ldots,J$ is the features of $\boldsymbol{X}$

for 1 to p **do**

$V_j = \{V_j - \text{mean}(V_j)\}/\text{sd}(V_j)$ S501

if V_j in Type2 **then** S503

$V_j = (-1) \times V_j$

else if V_j in Type3 **then** S505

$V_j = abs(V_j)$

else S507

$V_j = V_j$

$V_j = sigmoid(V_j)$ S509

end for

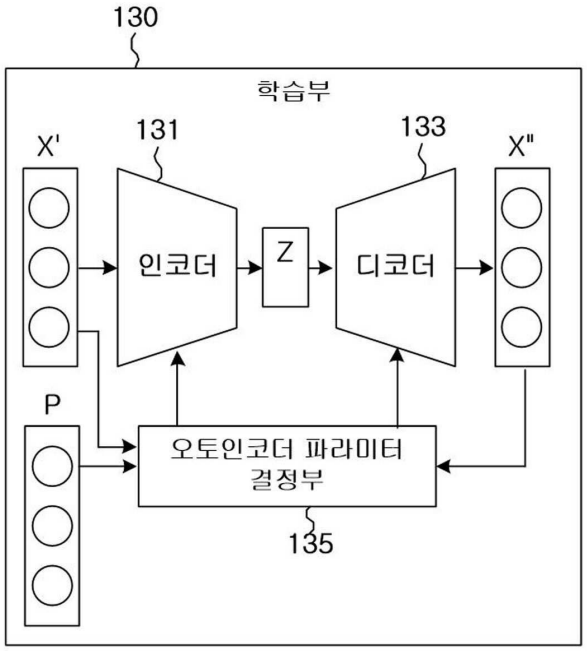
end function

$\boldsymbol{X}' = \text{TRANS}(V_j)$

도면6

정규화
데이터 방향성 통일
0-1 사이의 범위로 변환
Type1
바이패스
Type2
-1 을 곱함
Type3
절대값을 취함
ID
Waist
...
Uricacid
HDL
Albumin
TG
N

도면7



도면8

Risk embedded autoencoder learning algorithm

Input : Transformed input data $\boldsymbol{X'}$, risk score P

Output : Latent feature $l_{k'}$ $k = 1, \ldots, K$

procedure AM TRAINING($\boldsymbol{X'}$, $\boldsymbol{P}$, K, lr)

$\boldsymbol{\theta} = \{\boldsymbol{W}, \boldsymbol{b}, \boldsymbol{b_h}\}$, where $\boldsymbol{W}$, $\boldsymbol{b}, \boldsymbol{b_h}$ are the parameters of autoencoder

lr is the learning rate

$\boldsymbol{h}=[\boldsymbol{h_1}, \boldsymbol{h_2}, \ldots, \boldsymbol{h_c}] \in C^K$, where $\boldsymbol{h_c}$ is the number of hidden units in layer c and K is the number of hidden layers

Initialize : $\boldsymbol{\theta}^{(init)}$

$iter \leftarrow \mathbf{0}$

while $\mathcal{L}^{(iter+1)} < \mathcal{L}^{(iter)}$ **do**

$\widetilde{\boldsymbol{X'}} = \boldsymbol{Linear}(\boldsymbol{X'})$ ~S801

$\boldsymbol{h} = f_a(\widetilde{\boldsymbol{X'}} * \boldsymbol{W} + \boldsymbol{b})$ ~S803

$\boldsymbol{X''} = f_a(\boldsymbol{h} * \boldsymbol{W^T} + \boldsymbol{b_h})$, f_a is an activation function ~S805

$\mathcal{L} = \mathcal{L}(\boldsymbol{X'}, \boldsymbol{X''}) + \mathcal{L}(\boldsymbol{X''}, \boldsymbol{P})$ ~S807

g = compute the gradients of the loss with respect to $\boldsymbol{\theta}$

for $\boldsymbol{\theta}_i$, $\mathbf{g}_i$ in ($\boldsymbol{\theta}$, g) **do**

$\boldsymbol{\theta}_i = \boldsymbol{\theta}_i - l^* \mathbf{g}_i$

if $|\mathcal{L}^{(iter+1)} - \mathcal{L}^{(iter)}|$<1e-5: ~S809

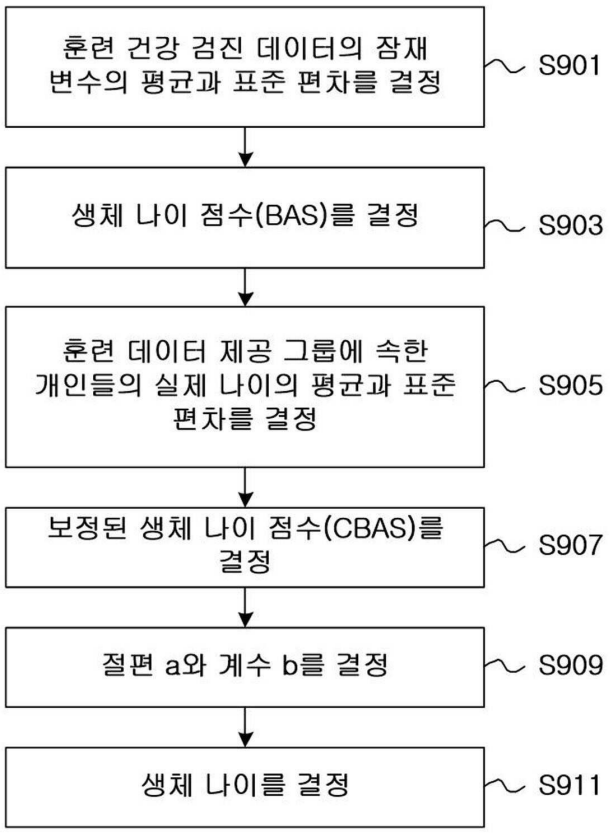
return $\boldsymbol{\theta}$

$iter \leftarrow iter + 1$

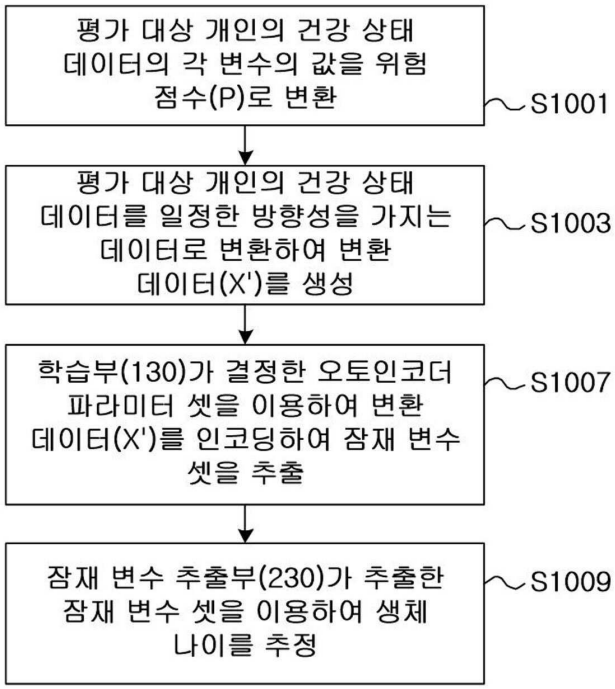
end for

end procedure

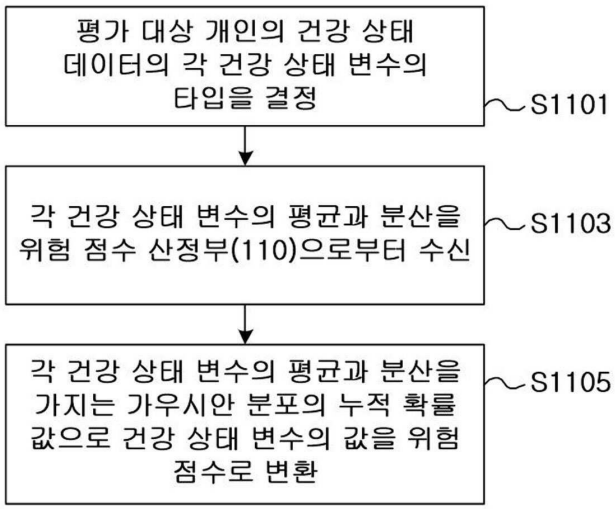
도면9


훈련 건강 검진 데이터의 잠재 변수의 평균과 표준 편차를 결정
S901
생체 나이 점수(BAS)를 결정
S903
훈련 데이터 제공 그룹에 속한 개인들의 실제 나이의 평균과 표준 편차를 결정
S905
보정된 생체 나이 점수(CBAS)를 결정
S907
절편 a와 계수 b를 결정
S909
생체 나이를 결정
S911

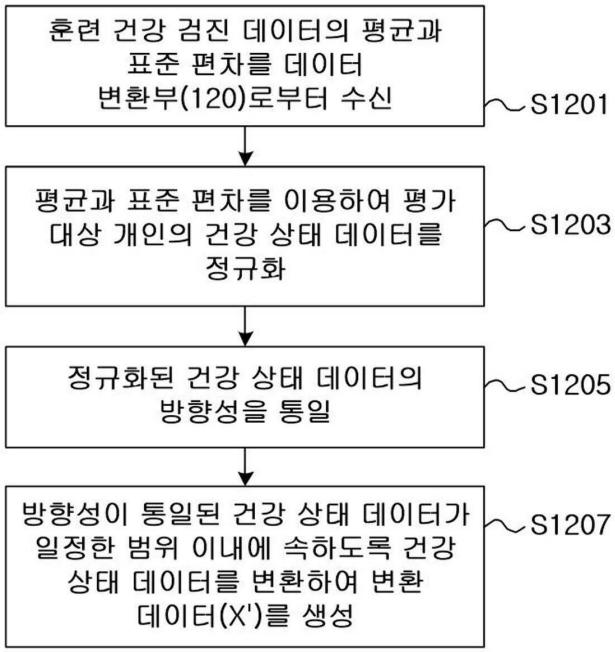
도면10


평가 대상 개인의 건강 상태 데이터의 각 변수의 값을 위험 점수(P)로 변환
S1001
평가 대상 개인의 건강 상태 데이터를 일정한 방향성을 가지는 데이터로 변환하여 변환 데이터(X')를 생성
S1003
학습부(130)가 결정한 오토인코더 파라미터 셋을 이용하여 변환 데이터(X')를 인코딩하여 잠재 변수 셋을 추출
S1007
잠재 변수 추출부(230)가 추출한 잠재 변수 셋을 이용하여 생체 나이를 추정
S1009

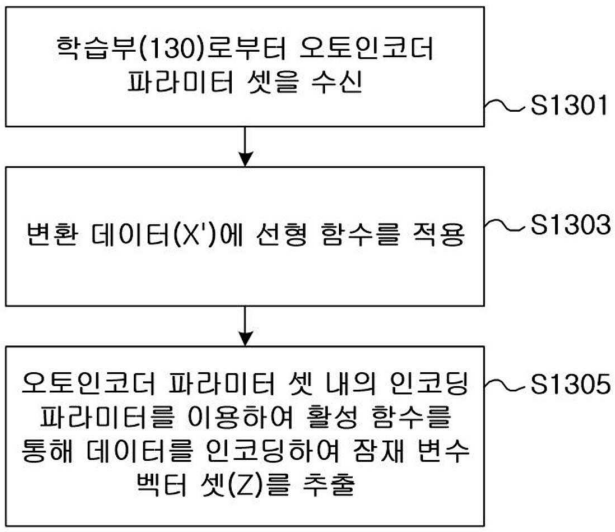
도면11


평가 대상 개인의 건강 상태 데이터의 각 건강 상태 변수의 타입을 결정
S1101
각 건강 상태 변수의 평균과 분산을 위험 점수 산정부(110)으로부터 수신
S1103
각 건강 상태 변수의 평균과 분산을 가지는 가우시안 분포의 누적 확률 값으로 건강 상태 변수의 값을 위험 점수로 변환
S1105

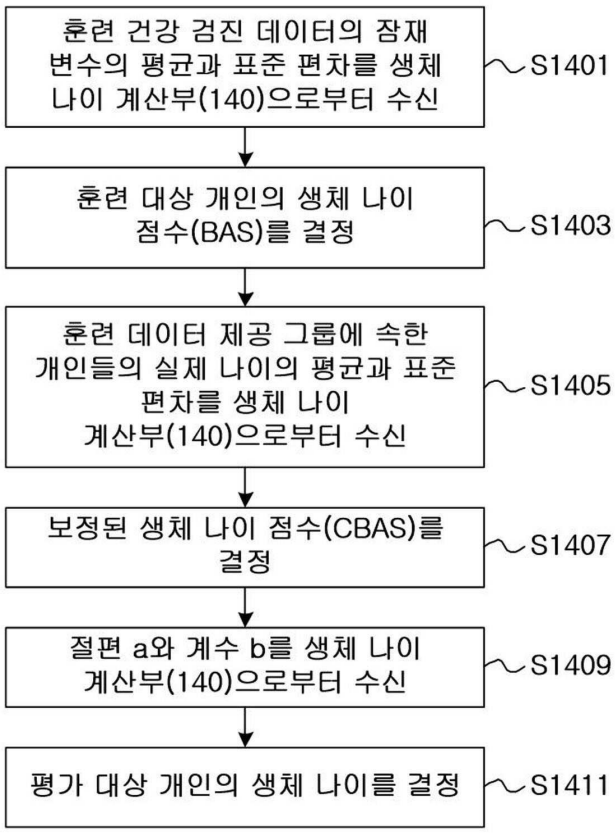
도면12


훈련 건강 검진 데이터의 평균과 표준 편차를 데이터 변환부(120)로부터 수신
S1201
평균과 표준 편차를 이용하여 평가 대상 개인의 건강 상태 데이터를 정규화
S1203
정규화된 건강 상태 데이터의 방향성을 통일
S1205
방향성이 통일된 건강 상태 데이터가 일정한 범위 이내에 속하도록 건강 상태 데이터를 변환하여 변환 데이터(X')를 생성
S1207

도면13


학습부(130)로부터 오토인코더 파라미터 셋을 수신
S1301
변환 데이터(X')에 선형 함수를 적용
S1303
오토인코더 파라미터 셋 내의 인코딩 파라미터를 이용하여 활성 함수를 통해 데이터를 인코딩하여 잠재 변수 벡터 셋(Z)를 추출
S1305

도면14


훈련 건강 검진 데이터의 잠재 변수의 평균과 표준 편차를 생체 나이 계산부(140)으로부터 수신
S1401
훈련 대상 개인의 생체 나이 점수(BAS)를 결정
S1403
훈련 데이터 제공 그룹에 속한 개인들의 실제 나이의 평균과 표준 편차를 생체 나이 계산부(140)으로부터 수신
S1405
보정된 생체 나이 점수(CBAS)를 결정
S1407
절편 a와 계수 b를 생체 나이 계산부(140)으로부터 수신
S1409
평가 대상 개인의 생체 나이를 결정
S1411