(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0075645
(43) 공개일자 2022년06월08일

(51) 국제특허분류(Int. Cl.)
G06N 20/00 (2019.01) *G06N 7/00* (2022.01)
(52) CPC특허분류
G06N 20/00 (2021.08)
G06N 7/005 (2013.01)
(21) 출원번호 10-2020-0163948
(22) 출원일자 2020년11월30일
심사청구일자 2020년11월30일

(71) 출원인
울산과학기술원
울산광역시 울주군 언양읍 유니스트길 50
(72) 발명자
이정혜
울산광역시 울주군 언양읍 유니스트길 50
김수현
울산광역시 울주군 언양읍 유니스트길 50
(뒷면에 계속)
(74) 대리인
특허법인태백

전체 청구항 수 : 총 15 항

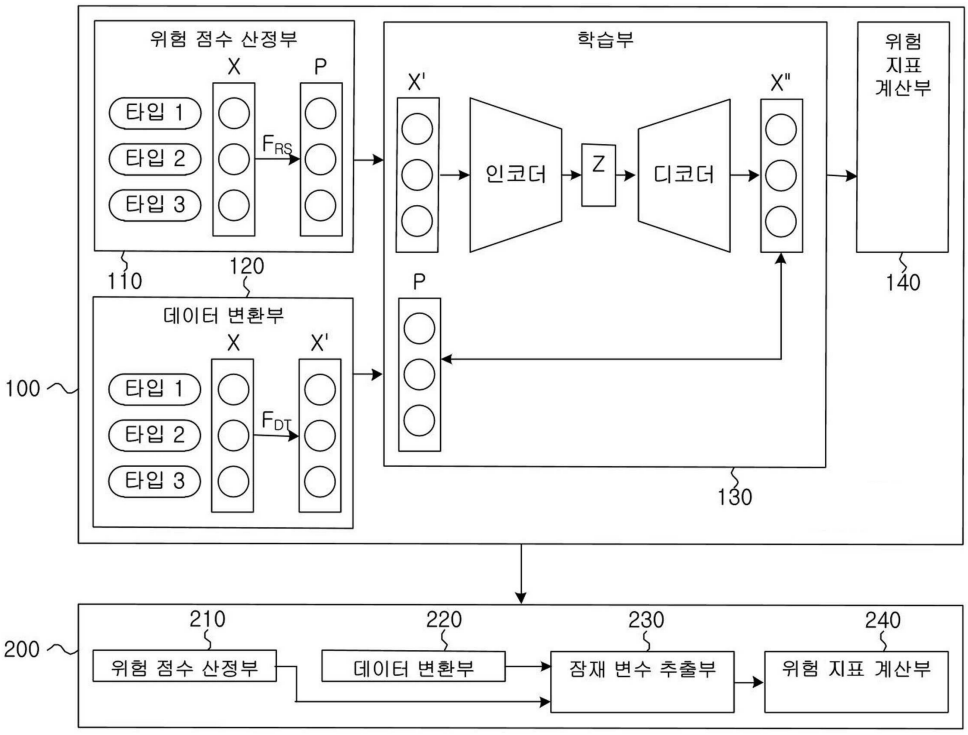
(54) 발명의 명칭 **위험 점수를 반영한 학습 기반의 위험 지표 산출 장치 및 방법**

(57) 요 약

본 발명은 위험 점수를 반영한 학습 기반의 위험 지표 산출 장치 및 방법에 관한 것이다. 본 발명에 따르면, 훈련부와 평가부를 포함하는 위험 지표 산출 장치가 개시된다. 훈련부는 평가 대상의 위험 정도에 영향을 미치는 각 변수별 훈련 데이터를 확률 분포에 기반하여 위험 점수 데이터로 변환하고, 훈련 데이터를 일정한 방향성을

(뒷면에 계속)

대 표 도



가지는 데이터로 변환하여 변환 데이터를 생성하며, 상기 위험 점수 데이터와 상기 변환 데이터를 이용하여 학습 모델의 파라미터를 결정한다. 평가부는 평가 대상의 현재 수집된 각 변수별 상태 데이터를 학습 모델의 결정된 파라미터를 이용하여 부호화하여 평가 대상의 위험 지표를 산출한다.

본 발명에 따르면, 평가 대상에 대한 실제 상태 데이터뿐만 아니라 위험 정도를 추가로 고려하여 평가 대상의 위험 지표를 계산하므로, 평가 대상의 상태를 보다 잘 반영하고 평가 대상의 전반적인 상태와 위험도를 쉽고 종합적으로 파악할 수 있는 신뢰성 있는 위험 지표가 얻어질 수 있다.

(72) 발명자

김한결

울산광역시 울주군 언양읍 유니스트길 50

임동철

울산광역시 울주군 언양읍 유니스트길 50

이 발명을 지원한 국가연구개발사업

과제고유번호	1711115544
과제번호	2020R1C1C1011063
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	개인기초연구(과기정통부)(R&D)
연구과제명	딥러닝 기반의 개인정보보호가 보장되는 인공지능 연합학습 플랫폼 개발
기 여 율	1/2
과제수행기관명	울산과학기술원
연구기간	2020.03.01 ~ 2021.02.28

이 발명을 지원한 국가연구개발사업

과제고유번호	1711112075
과제번호	2018R1C1B5086611
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	개인기초연구(과기정통부)(R&D)
연구과제명	분산형 데이터를 위한 개인 프라이버시 보호 및 시스템 보안이 보장되는 머신러닝 기반의 연합 예측 모형 개발
기 여 율	1/2
과제수행기관명	울산과학기술원
연구기간	2020.03.01 ~ 2020.08.31

명 세 서

청구범위

청구항 1

위험 지표 산출 장치를 이용한 위험 지표 산출 방법에 있어서,

평가 대상의 위험 정도에 영향을 미치는 각 변수별 훈련 데이터를 확률 분포에 기반하여 위험 점수 데이터로 변환하는 단계;

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

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

평가 대상의 현재 수집된 각 변수별 상태 데이터를 학습 모델의 결정된 파라미터를 이용하여 부호화하여 평가 대상의 위험 지표를 산출하는 단계를 포함하는 위험 지표 산출 방법.

청구항 2

제1항에 있어서,

상기 학습 모델은,

오토인코더, 주성분 분석, 비음수 행렬 분해 기반의 학습 모델 중에서 선택되는 위험 지표 산출 방법.

청구항 3

제1항에 있어서,

상기 학습 모델은 오토인코더이며,

상기 학습 모델의 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 오토인코더 파라미터를 결정하는 단계를 포함하는 위험 지표 산출 방법.

청구항 4

제3항에 있어서,

상기 오토인코더 파라미터는 상기 복호화 데이터 및 상기 변환 데이터 사이의 손실을 추가로 이용하여 결정되는 위험 지표 산출 방법.

청구항 5

제4항에 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 위험 지표 산출 방법.

청구항 6

제1항에 있어서,

상기 훈련 데이터를 상기 위험 점수 데이터로 변환하는 단계는,

상기 훈련 데이터가 따르는 가우시안 분포의 누적 확률 값을 상기 위험 점수 데이터로 결정하는 단계를 포함하는 위험 지표 산출 방법.

청구항 7

제1항에 있어서,

상기 변환 데이터를 생성하는 단계는,

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

상기 복수의 변환된 상태 변수의 범위를 일정 범위로 조절하여 상기 변환 데이터를 생성하는 단계를 포함하는 위험 지표 산출 방법.

청구항 8

제1항에 있어서,

상기 학습 모델은 오더인코더이고,

상기 위험 지표를 산출하는 단계는,

결정된 오토인코더 파라미터를 이용하여 평가 대상의 현재 수집된 상태 데이터의 복수의 상태 변수를 부호화하여 하나 이상의 잠재 변수를 생성하는 단계와,

상기 하나 이상의 잠재 변수를 이용하여 상기 위험 지표를 산출하는 단계를 포함하는 위험 지표 산출 방법.

청구항 9

제8항에 있어서,

상기 하나 이상의 잠재 변수의 개수는 상기 복수의 상태 변수의 개수보다 적은 위험 지표 산출 방법.

청구항 10

제8항에 있어서,

상기 하나 이상의 잠재 변수의 평균 값을 이용하여 상기 위험 지표가 산출되는 위험 지표 산출 방법.

청구항 11

제1항에 있어서,

상기 산출된 위험 지표를 적어도 하나의 임계값과 비교하여 평가 대상의 위험도 레벨을 판단하는 단계를 더 포함하는 위험 지표 산출 방법.

청구항 12

평가 대상의 위험 정도에 영향을 미치는 각 변수별 훈련 데이터를 확률 분포에 기반하여 위험 점수 데이터로 변환하고, 훈련 데이터를 일정한 방향성을 가지는 데이터로 변환하여 변환 데이터를 생성하며, 상기 위험 점수 데이터와 상기 변환 데이터를 이용하여 학습 모델의 파라미터를 결정하는 훈련부; 및

평가 대상의 현재 수집된 각 변수별 상태 데이터를 학습 모델의 결정된 파라미터를 이용하여 부호화하여 평가 대상의 위험 지표를 산출하는 평가부를 포함하는 위험 지표 산출 장치.

청구항 13

제12항에 있어서,

상기 학습 모델은 오토인코더이며,

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 오토인코더 파라미터를 결정하는 위험 지표 산출 장치.

청구항 14

제13항에 있어서,

상기 훈련부는 상기 복호화 데이터 및 상기 변환 데이터 사이의 손실을 추가로 이용하여 상기 오토인코더 파라미터를 결정하는 위험 지표 산출 장치.

청구항 15

제13항에 있어서,

상기 훈련부는,

상기 훈련 데이터가 따르는 가우시안 분포의 누적 확률 값을 상기 위험 점수 데이터로 결정하는 위험 지표 산출 장치.

발명의 설명

기 술 분 야

[0001] 본 발명은 위험 점수를 반영한 학습 기반의 위험 지표 산출 장치 및 방법에 관한 것이며, 구체적으로 계산된 위험 점수를 학습 모델에 반영하여 표현형 학습을 통해 평가 대상의 위험 지표를 산출하는 장치 및 방법에 관한 것이다.

배 경 기 술

[0002] "Index"라는 개념은 데이터를 기반으로 정해진 목적에 맞게 대상의 특정 수치나 level을 산정하여 대상의 상태를 표현하는 것을 의미한다. 일반적으로 특정 데이터 수치들의 계산식을 통해 인덱스가 생성된다. 예를 들어, 대표적인 경제지표인 무역수지와 상품수지는 상품의 수출입 거래에서 생기는 국제수지로서 상품의 수출액과 수입액의 차액계산식을 통해 새로운 지표로 계산되며, 수출액과 수입액을 같이 반영한 새로운 수치로 표현된다.

[0003] 그러나, 이러한 방식으로 계산된 인덱스는 많은 변수들의 특징을 한번에 반영하기는 굉장히 어렵다. 이러한 한계점을 극복하기 위해, 다양한 변수들의 상호간의 영향 및 특징을 반영할 수 있는 머신러닝/딥러닝 기반의 인덱스 개발의 필요성이 대두되고 있다.

[0004] 특히, 머신러닝 방법론 중 비지도적 학습 방법을 통해 각 변수들의 정보를 압축한 잠재 변수를 도출하여 계산한 인덱스들이 개발되고 있다. 예를 들어, 대표적인 헬스 지표(health index) 중 하나인 생체나이(biological age; BA)의 경우 주성분 분석 등의 머신러닝 모델을 활용하고 있는 추세이다.

[0005] 기존의 인덱스를 만드는 주요 목적은 대상의 상태가 좋고 나쁨을 표현하는 수치를 개발하는 데에 있다. 즉, 인덱스의 값에 따라 상태의 위험도를 판별하게 되는 경우가 많다.

[0006] 그러나, 실제로 해당 인덱스가 상태의 위험도를 판별할 수 있는 적합한 수치인지에 대한 기준이 모호하다. 즉, 위험 정도에 대한 판별 기준에 모호함이 있다.

[0007] 특히, 머신러닝/딥러닝 기반의 인덱스는 다양하고 많은 변수들을 활용하여 생성되지만, 해당 변수의 좋고 나쁨의 정도가 잘 반영되지 못할 수 있다.

[0008] 인덱스는 목표값으로 삼을 수 있는 실제 타겟(label)이 존재하지 않으며, 이로 인해 해당 인덱스를 검증함에 있어 명확한 방법이 존재하지 않는다.

[0009] 하나의 인덱스를 실제 상황에 적용하기 위해서는 다양한 측면에서의 검증이 필요하지만, 주로 해당 분야의 전문가들의 의견이나 설문조사 등을 통해 검증이 이루어지므로 많은 시간과 비용이 발생한다.

[0010] 또한, 생성된 인덱스가 대상의 특정 상태를 표현할 때 부적합하게 사용되는 경우도 많다. 예를 들어, 사람의 키와 몸무게 값의 계산식을 산정하여 만든 BMI(body mass index)는 비만의 주요 지표로 사용되지만, 실제로 비만은 키와 몸무게 외에 체내지방 및 근육량 등을 종합적으로 고려해야 한다.

[0011] 이러한 검증의 어려움 때문에, 인덱스가 개발되어도 실제로 활용되지 못하는 경우가 생길 수 있다.

[0012] 본 발명의 배경이 되는 기술은 한국공개특허 제10-2016-0036954호(2016.04.05 공개)에 개시되어 있다.

발명의 내용

해결하려는 과제

[0013] 본 발명이 이루고자 하는 기술적 과제는 새로운 머신 러닝 기법을 적용하여 평가 대상에 관한 위험 상태를 보다 잘 반영할 수 있는 위험 요인 별 확률 기반 위험 점수를 산정하고, 새로운 머신 러닝 기법을 적용하여 평가 대상의 위험 점수를 잘 반영할 수 있는 인덱스 산출 모델을 제시하고, 딥러닝 기반의 표현형 학습을 통해 평가 대상의 위험 인덱스를 효율적으로 산출하는 것을 제시하는 것에 있다.

과제의 해결 수단

[0014] 본 발명의 일 실시예에 따른 위험 지표 산출 장치를 이용한 위험 지표 산출 방법은, 평가 대상의 위험 정도에 영향을 미치는 각 변수별 훈련 데이터를 확률 분포에 기반하여 위험 점수 데이터로 변환하는 단계; 훈련 데이터를 일정한 방향성을 가지는 데이터로 변환하여 변환 데이터를 생성하는 단계; 상기 위험 점수 데이터와 상기 변환 데이터를 이용하여 학습 모델의 파라미터를 결정하는 단계; 및 평가 대상의 현재 수집된 각 변수별 상태 데이터를 학습 모델의 결정된 파라미터를 이용하여 부호화하여 평가 대상의 위험 지표를 산출하는 단계를 포함할 수 있다.

[0015] 상기 학습 모델은, 오토인코더, 주성분 분석, 비음수 행렬 분해 기반의 학습 모델 중에서 선택될 수 있다.

[0016] 상기 학습 모델은 오토인코더이며, 상기 학습 모델의 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 오토인코더 파라미터를 결정하는 단계를 포함할 수 있다.

[0017] 상기 오토인코더 파라미터는 상기 복호화 데이터 및 상기 변환 데이터 사이의 손실을 추가로 이용하여 결정될 수 있다.

[0018] 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.

[0019] 상기 훈련 데이터를 상기 위험 점수 데이터로 변환하는 단계는, 상기 훈련 데이터가 따르는 가우시안 분포의 누적 확률 값을 상기 위험 점수 데이터로 결정하는 단계를 포함할 수 있다.

[0020] 상기 변환 데이터를 생성하는 단계는, 상기 훈련 데이터의 복수의 상태 변수 각각이 복수의 방향성 중 하나를 가지는 경우에, 상기 복수의 상태 변수 각각이 상기 복수의 방향성 중 대표 방향성을 가지도록 상기 복수의 상태 변수 각각을 변환하여 복수의 변환된 상태 변수를 생성하는 단계와, 상기 복수의 변환된 상태 변수의 범위를 일정 범위로 조절하여 상기 변환 데이터를 생성하는 단계를 포함할 수 있다.

[0021] 상기 학습 모델은 오토인코더이고, 상기 위험 지표를 산출하는 단계는, 결정된 오토인코더 파라미터를 이용하여 평가 대상의 현재 수집된 상태 데이터의 복수의 상태 변수를 부호화하여 하나 이상의 잠재 변수를 생성하는 단계와, 상기 하나 이상의 잠재 변수를 이용하여 상기 위험 지표를 산출하는 단계를 포함할 수 있다.

[0022] 상기 하나 이상의 잠재 변수의 개수는 상기 복수의 상태 변수의 개수보다 적을 수 있다.

[0023] 상기 하나 이상의 잠재 변수의 평균 값을 이용하여 상기 위험 지표가 산출될 수 있다.

[0024] 상기 위험 지표 산출 방법은, 상기 산출된 위험 지표를 적어도 하나의 임계값과 비교하여 평가 대상의 위험도 레벨을 판단하는 단계를 더 포함할 수 있다.

[0025] 본 발명의 일 실시예에 따른 위험 지표 산출 장치는, 평가 대상의 위험 정도에 영향을 미치는 각 변수별 훈련 데이터를 확률 분포에 기반하여 위험 점수 데이터로 변환하고, 훈련 데이터를 일정한 방향성을 가지는 데이터로 변환하여 변환 데이터를 생성하며, 상기 위험 점수 데이터와 상기 변환 데이터를 이용하여 학습 모델의 파라미터를 결정하는 훈련부; 및 평가 대상의 현재 수집된 각 변수별 상태 데이터를 학습 모델의 결정된 파라미터를 이용하여 부호화하여 평가 대상의 위험 지표를 산출하는 평가부를 포함할 수 있다.

[0026] 상기 학습 모델은 오토인코더이며, 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 오토인코더 파라미터를 결정할 수 있다.

[0027] 상기 훈련부는 상기 복호화 데이터 및 상기 변환 데이터 사이의 손실을 추가로 이용하여 상기 오토인코더 파라미터를 결정할 수 있다.

[0028] 상기 훈련부는 상기 훈련 데이터가 따르는 가우시안 분포의 누적 확률 값을 상기 위험 점수 데이터로 결정할 수 있다.

발명의 효과

[0029] 본 발명에 따르면, 평가 대상에 대한 실제 상태 데이터뿐만 아니라 위험 정도를 추가로 고려하여 평가 대상의 위험 지표를 계산하므로, 평가 대상의 상태를 보다 잘 반영하고 평가 대상의 전반적인 상태와 위험도를 보다 쉽고 종합적으로 파악할 수 있는 신뢰성 있는 위험 지표가 얻어질 수 있다.

도면의 간단한 설명

[0030] 도 1은 본 발명의 실시예에 따른 위험 지표 산출 장치를 보여준다.

도 2는 본 발명의 실시예에 따른 위험 지표 산출 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 건강 검진 데이터의 데이터 변환의 예를 보여준다.

도 7a는 도 6의 과정을 요약한 결과를 보여준다.

도 7b는 데이터 변환부에 의해 수행된 재무 데이터의 데이터 변환 결과를 보여준다.

도 8은 본 발명의 실시예에 따른 학습부를 보여주는 블록도이다.

도 9은 본 발명의 실시예에 따른 학습부의 동작을 보여주는 알고리즘이다.

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.

발명을 실시하기 위한 구체적인 내용

[0031] 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.

[0032] 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.

[0033] 이하에서는 도 1 내지 도 13을 참고하여 본 발명의 실시예에 따른 위험 지표 산출 장치 또는 위험 지표 산출 시스템을 설명한다.

[0034] 도 1은 본 발명의 실시예에 따른 위험 지표 산출 장치를 보여준다.

[0035] 도 1에 도시된 바와 같이, 본 발명의 실시예에 따른 위험 지표 산출 장치(1)는 훈련부(100), 평가부(200), 위험 지표 표시부(300)를 포함한다.

[0036] 훈련부(100)는 평가 대상에 대응하여 기 수집된 훈련 데이터에 기반하여 위험 지표를 산출할 수 있는 파라미터를 생성한다. 훈련 데이터는 평가 대상의 위험 정도에 영향을 미치는 변수(인자)들에 관한 데이터를 포함한다.

[0037] 훈련부(100)는 훈련 데이터를 기반으로 학습 모델의 파라미터를 결정하며, 훈련 데이터로부터 확률 기반의 위험 점수를 계산하고 위험 점수를 학습 모델에 반영한다.

[0038] 훈련부(100)는 목표로 하는 특정 타겟이나 라벨 없이 데이터 그 자체의 정보를 함축적으로 표현하는 비지도적 학습 모델을 활용하여 훈련 데이터를 학습한다. 예컨대, 오토인코더(Autoencoder), 주성분 분석(Principal component analysis; PCA), 비음수 행렬 분해(Non-negative matrix factorization; NMF) 등에 기반한 학습 모델이 사용될 수 있다. 이들은 모두 잠재 변수를 도출하는 목적식을 가지는 모델이다.

[0039] 본 발명은 이러한 목적식에 위험 점수를 반영하는 목적식을 추가하여 잠재 변수 학습에 위험 점수를 반영하도록 하고, 이를 통해 위험 점수가 반영된 머신러닝/딥러닝 기반의 새로운 위험 지표 산출 기법을 제시한다.

[0040] 이하에서는 설명의 편의상 주로 오토인코더를 예시하지만, 학습 모델이 반드시 오토인코더로 한정되지 않으며, 대부분의 표현형 머신러닝/딥러닝 학습 모델(예: 주성분 분석, 비음수 행렬분해 기법 등)로 확장 가능하다.

[0041] 평가부(200)는 훈련부(100)에 의해 생성된 학습 모델의 파라미터 및 평가 대상의 현재 수집된 각 변수별 상태 정보를 이용하여 평가 대상의 위험 지표를 연산하며, 연산된 위험 지표로부터 그에 대응되는 위험도 레벨을 판별한다.

[0042] 위험 지표 표시부(300)는 평가부(200)가 평가 대상의 추정된 위험 지표 및 위험도 레벨을 표시한다.

[0043] 본 발명의 실시예에서 위험 평가 대상은 개인의 건강 상태, 기업 또는 개인의 금융 및 재무 상태, 각종 설비의 운영 상태 등 다양할 수 있다. 여기서 물론 평가 대상의 종류에 따라 해당 평가 대상의 위험 정도에 관여하는 변수(인자)들의 종류도 달라지게 된다. 이와 같은 본 발명의 실시예에 따르면, 개인의 건강 상태 데이터를 기반으로 건강 위험 지표를 산출하도록 하고 기업의 금융 및 재무 상태 데이터를 기반으로 재무 위험 지표를 산출하도록 한다.

[0044] 도 2는 본 발명의 실시예에 따른 위험 지표 산출 시스템을 보여준다.

[0045] 도 2에 도시된 바와 같이, 본 발명의 실시예에 따른 위험 지표 산출 시스템(2)은 하나 이상의 단말 장치(10), 위험 지표 산출 서버(20) 및 네트워크(30)를 포함한다.

[0046] 각 단말 장치(10)는 통신부(11) 및 위험 지표 표시부(300)를 포함한다. 통신부(11)는 네트워크(30)를 통해 위험 지표 산출 서버(20)에 접속하며, 위험 지표 산출 서버(20)에 평가 대상의 현재 수집된 상태 정보를 위험 지표 산출 서버(20)에 제공하며, 위험 지표 산출 서버(20)로부터 산출된 위험 지표 및 그에 대응되는 위험도 레벨을 수신한다. 단말 장치(10)는 수신한 위험 지표 및 위험도 레벨을 표시한다.

[0047] 위험 지표 산출 서버(20)는 통신부(21), 훈련부(100) 및 평가부(200)를 포함한다.

[0048] 통신부(21)는 네트워크(30)를 통해 단말 장치(10)와 통신하며, 단말 장치(10)로부터 평가 대상의 상태 정보를 수신하며, 평가된 위험 지표를 단말 장치(10)에 제공한다.

[0049] 훈련부(100)는 훈련 데이터에 기반하여 위험 지표를 산출할 수 있는 파라미터를 생성한다.

[0050] 평가부(200)는 훈련부(100)에 의해 생성된 파라미터 및 평가 대상의 상태 정보를 이용하여 평가 대상의 위험 지표를 추정한다.

[0051] 도 3은 본 발명의 실시예에 따른 훈련부 및 평가부를 보여주는 블록도이다.

[0052] 도 3에 도시된 바와 같이, 훈련부(100)는 위험 점수 산정부(110), 데이터 변환부(120), 학습부(130), 및 위험 지표 계산부(140)를 포함하고, 평가부(200)는 위험 점수 산정부(210), 데이터 변환부(220), 잠재 변수 추출부(230), 및 위험 지표 계산부(240)를 포함한다.

[0053] 훈련부(100)의 각 구성요소 및 평가부(200)의 각 구성요소에 대하여는 도 4 내지 도 13을 참고하여 상세히 설명한다.

[0054] 도 4는 본 발명의 실시예에 따른 훈련부의 동작 방법을 보여주는 흐름도이다.

[0055] 도 4에 도시된 바와 같이, 위험 점수 산정부(110)는 훈련에 사용될 훈련 데이터(X)의 각 변수의 값을 위험 점수

(P)로 변환한다(S401).

[0056] 여기서, 훈련 데이터는 N개의 훈련 데이터 제공자 및 복수의 변수에 대한 상태 정보를 포함할 수 있다. N이 클수록 훈련의 효과는 좋아진다.

[0057] 예컨대, 위험 평가 대상이 개인의 건강 상태인 경우, 훈련 데이터는 N명의 훈련 데이터 제공자 및 복수의 변수에 대한 건강 상태 정보를 포함할 수 있으며, 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.

[0058] 또한, 기업의 재무 상태가 위험 평가 대상인 경우, 훈련 데이터는 N개의 훈련 데이터 제공자(기업) 및 복수의 변수에 대한 상태 정보를 포함할 수 있다. 이때, 복수의 변수는 유동성 장기부재, 현금 및 현금성 자산, 재고자산, 부채비율, 자기자본 비율, 유동비율, 당좌비율, 차입금 의존도 등을 포함할 수 있으며 이에 한정되지 않는다.

[0059] 구체적으로, 위험 점수 산정부(110)는 훈련 데이터의 각 상태 변수의 타입을 결정한다. 각 상태 변수는 3가지 타입 중 하나로 결정될 수 있다. 예컨대, 변수의 값이 커질수록 상태(예: 건강 상태, 재무 상태)가 나쁘다는 것을 나타내는 경우, 해당 변수는 타입 1으로 결정되고, 변수의 값이 작아질수록 상태가 나쁘다는 것을 나타내는 경우, 해당 변수는 타입 2로 결정되고, 변수의 값이 평균보다 커지거나 작아질수록 상태가 나쁘다는 것을 나타내는 경우, 해당 변수는 타입 3으로 결정될 수 있다.

[0060] 위험 점수 산정부(110)는 각 상태 변수의 타입에 기초하여 훈련 데이터의 각 상태 변수의 평균과 분산을 결정한다.

[0061] 예컨대, 건강 상태를 평가하는 상황을 가정하면, 건강 상태 변수의 타입이 타입 1 또는 타입 2일 경우, 해당 건강 상태 변수가 반 가우시안 분포(half-Gaussian distribution)를 따른다는 가정하에, 평균과 분산이 결정된다.

[0062] 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. 예컨대, BMI 변수의 경우, 정상 수치 범위가 18.5에서 25에 해당하므로, BMI의 평균은 이들 두 값의 평균인 21.75로 계산될 수 있다.

[0063] 위험 점수 산정부(110)는 각 상태 변수가 위에서 결정된 평균과 분산을 가지는 가우시안 분포를 따른다는 가정하에, 이 가우시안 분포의 누적 확률 값으로 상태 변수의 값을 위험 점수로 변환한다.

[0064] 건강 상태를 평가하는 상황을 예시하면, 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 상태 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.

[0065] 다음, 데이터 변환부(120)는 훈련 데이터(X)를 일정한 방향성을 가지는 데이터로 변환하여 변환 데이터(X')를 생성한다(S403).

[0066] 이하에서는 도 5와 도 6을 참고하여 데이터 변환부(120)의 동작을 상세히 설명한다.

[0067] 도 5는 데이터 변환부(120)의 동작을 보여주는 알고리즘을 나타낸다.

[0068] 각 상태 변수에 대하여 데이터 변환부(120)는 훈련 데이터의 평균과 표준 편차를 구하고, 평균과 표준 편차를 이용하여 훈련 데이터를 정규화한다(S501).

[0069] 각 상태 변수에 대하여, 데이터 변환부(120)는 정규화된 데이터의 방향성을 통일시킨다. 이를 위하여, 상태 변수의 타입이 타입 2인 경우, 데이터 변환부(120)는 데이터에 -1을 곱할 수 있다(S503). 상태 변수의 타입이 타입 3인 경우, 데이터 변환부(120)는 데이터에 절대값을 취할 수 있다(S505). 상태 변수의 타입이 타입 1인 경우, 데이터 변환부(120)는 데이터를 바이패스할 수 있다(S507). 이처럼, 훈련 데이터의 복수의 상태 변수가 복수의 방향성 중 하나를 가지는 경우에, 복수의 상태 변수의 각각이 복수의 방향성 중 하나인 대표 방향성을 가지도록 복수의 상태 변수의 각각이 변환될 수 있다.

[0070] 각 상태 변수에 대하여, 데이터 변환부(120)는 방향성이 통일된 데이터가 일정한 범위 이내에 속하도록 데이터를 변환하여 변환 데이터(X')를 생성할 수 있다(S509). 예컨대, 데이터 변환부(120)는 시그모이드 (sigmoid) 함수를 통해 데이터가 0과 1사이의 범위에 속하도록 데이터를 변환할 수 있다.

[0071] 도 6은 데이터 변환부(120)에 의해 수행되는 건강 검진 데이터의 데이터 변환의 예를 보여준다.

[0072] 도 6에 도시된 바와 같이, 훈련 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. 이들은 건강 상태 변수에 해당한다.

[0073] 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.

[0074] 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.

[0075] 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.

[0076] 도 7a는 도 6의 과정을 요약한 결과를 보여준다. 이러한 도 7a에 나타낸 것과 같이, 건강 검진 데이터의 각 변수는 데이터 변환 과정을 통하여 모두 타입 1의 방향성을 가지는 데이터로 변환이 되고 시그모이드 함수를 통하여 0과 1 사이의 값으로 조정된다.

[0077] 도 7b는 데이터 변환부에 의해 수행된 재무 데이터의 데이터 변환 결과를 보여준다.

[0078] 도 7b에 도시된 바와 같이, 훈련 재무 데이터는 N개의 훈련 데이터 제공자 및 복수의 변수에 대한 재무 상태 정보를 포함하며, 복수의 변수는 유동성장기부채, 현금 및 현금성자산, 재고자산을 포함한다. 이들은 재무 상태 변수에 해당한다.

[0079] 유동성장기부채는 타입 1의 변수로 간주되며, 현금 및 현금성 자산은 타입2의 변수로 간주되며, 재고자산은 타입 3의 변수로 간주될 수 있다. 재무 데이터를 구성한 각 변수는 타입 1의 방향성을 가지는 데이터로 변환이 되고 시그모이드 함수를 통하여 0과 1 사이의 값으로 조정된다.

[0080] 다시 도 4를 설명한다.

[0081] 이하에서는 발명의 이해를 돕기 위하여 오토인코더 기반의 학습 모델을 활용하는 것을 대표 예시로 하여 설명한다.

[0082] 다음에, 학습부(130)는 위험 점수(P)와 변환 데이터(X')를 이용하여 학습 모델의 파라미터를 결정한다. 구체적으로, 학습부(130)는 위험 점수(P)와 변환 데이터(X')를 이용하여 오토인코더 파라미터 셋($\boldsymbol{\theta}$)을 결정한다(S405).

[0083] 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.

[0084] 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.

[0085] 다음은 도 8 및 도 9를 이용하여 오토인코더를 이용한 학습부(130)를 구체적으로 설명한다.

[0086] 도 8은 본 발명의 실시예에 따른 학습부(130)를 보여주는 블록도이다.

[0087] 도 8에 도시된 바와 같이, 본 발명의 실시예에 따른 학습부(130)는 인코더(131), 디코더(133), 오토인코더 파라미터 결정부(135)를 포함한다.

[0088] 도 9는 본 발명의 실시예에 따른 학습부의 동작을 보여주는 알고리즘이다.

[0089] 도 9에서 보여지는 바와 같이, 인코더(131)는 변환 데이터(X')를 선형 함수(linear function)을 통해 데이터($\widetilde{X'}$)를 생성한다(S901). 다음, 인코더(131)는 생성된 데이터($\widetilde{X'}$)를 인코딩 파라미터(W, b)를 이용하여 활성 함수(activation function)인 fa를 통해 인코딩하여 잠재 변수 벡터(h)를 생성한다(S903). 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.

[0090] 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(S905).

[0091] 오토인코더 파라미터 결정부(135)는 손실 값($\mathcal{L}$)을 계산한다(S907). 도 8에서 보여지는 바와 같이, 손실 값($\mathcal{L}$)은 입력 데이터(X')와 복원 데이터(X") 사이의 손실($\mathcal{L}(X', X'')$)과 위험 점수(P)와 복원 데이터(X") 사이의 손실($\mathcal{L}(X'', P)$)의 합으로 구해질 수 있다. 손실 함수 $\mathcal{L}$(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.

[0092] 오토인코더 파라미터 결정부(135)는 (iter+1)번째 손실 값($\mathcal{L}^{(iter+1)}$)과 iter 번째 손실 값($\mathcal{L}^{(iter)}$)의 차이가 소정의 값(1e-5)보다 작을 때의 오토인코더 파라미터 셋($\boldsymbol{\theta}$)을 결정한다(S909).

[0093] 다시 도 4를 설명한다.

[0094] 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.

[0095] 잠재 변수의 차원은 잠재 변수의 개수를 의미하고, 변환 데이터(X') 상의 상태 변수의 차원은 상태 변수의 개수를 의미할 수 있다. 예를 들어, 평가 대상이 건강 상태라 가정하면, 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.

[0096] 위험 지표 계산부(140)는 잠재 변수 셋(Z)을 이용하여 위험 지표를 산출한다(S409).

[0097] 위험 지표 계산부(140)는 각 차원의 잠재 변수의 평균 값을 구하고 평균 값을 위험 지표(Index)로 산출한다. 각 차원의 잠재 변수 l_k의 평균 값은 아래의 수학식 1을 통해 구해진다.

수학식 1

[0098]
$$\text{Index} = \sum_{i=1}^{K} \frac{l_k}{K}$$

[0099] 위험 지표 계산부(140)는 이렇게 구한 인덱스를 적어도 하나의 임계값과 비교하여 평가 대상의 위험도 레벨을

판단할 수 있다. 예를 들어, 위험 지표가 임계값 미만이면 안전 레벨, 임계값 이상이면 위험 레벨로 판단할 수 있으며, 다단의 임계값을 기반으로 위험 지표가 제1 임계값 미만이면 안전(저위험군), 제1 및 제2 임계값 사이인 경우 주의(중위험군), 제2 임계값 이상이면 위험(고위험군) 레벨로 판단할 수 있다.

[0100] 임계값은 훈련 데이터로부터 생성한 인덱스 값을 기준으로 인덱스 값이 큰 상위 n% 표본과 인덱스 값이 작은 하위 n% 표본으로부터 설정될 수 있다. 하위 n% 표본은 저위험군, 상위 n%의 표본은 고위험 군이라 할 때 제1 임계값은 하위 n% 표본에 대한 인덱스 값 중 최대값을 통해 결정되고 제2 임계값은 상위 n% 표본에 대한 인덱스 값 중 최소값을 통해 결정될 수 있다.

[0101] 도 10은 본 발명의 실시예에 따른 평가부(200)의 동작 방법을 보여주는 흐름도이다.

[0102] 위험 점수 산정부(210)는 평가 대상의 상태 데이터의 각 변수의 값을 위험 점수(P)로 변환한다(S1001).

[0103] 데이터 변환부(220)는 평가 대상의 상태 데이터를 일정한 방향성을 가지는 데이터로 변환하여 변환 데이터(X')를 생성한다(S1003).

[0104] 잠재 변수 추출부(230)는 학습부(130)가 결정한 오토인코더 파라미터 셋을 이용하여 변환 데이터(X')를 인코딩하여 잠재 변수 셋을 추출한다(S1007).

[0105] 위험 지표 계산부(240)는 잠재 변수 추출부(230)가 추출한 잠재 변수 셋을 이용하여 위험 지표를 계산한다(S1009).

[0106] 도 11은 본 발명의 실시예에 따른 평가부(200)의 위험 점수 산정부(210)의 동작 방법을 보여주는 흐름도이다.

[0107] 이하에서는 발명의 이해를 돕기 위하여 평가 대상 개인의 건강 상태 데이터로부터 건강 위험 지표를 산출하는 것을 예시한다.

[0108] 위험 점수 산정부(210)는 평가 대상개인의 건강 상태 데이터의 각 건강 상태 변수의 타입을 결정한다(S1101). 각 건강 상태 변수는 위에서 언급한 3가지 타입 중 하나로 결정될 수 있다.

[0109] 위험 점수 산정부(210)는 각 건강 상태 변수의 평균과 분산을 위험 점수 산정부(110)으로부터 수신한다(S1103).

[0110] 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 상태 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.

[0111] 도 12는 본 발명의 실시예에 따른 평가부(200)의 데이터 변환부(220)의 동작 방법을 보여주는 흐름도이다.

[0112] 각 건강 상태 변수에 대하여 데이터 변환부(220)는 훈련 건강 검진 데이터의 평균과 표준 편차를 데이터 변환부(120)로부터 수신한다(S1201).

[0113] 이후, 각 건강 상태 변수에 대하여 데이터 변환부(220)는 평균과 표준 편차를 이용하여 평가 대상 개인의 건강 상태 데이터를 정규화한다(S1203).

[0114] 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.

[0115] 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.

[0116] 도 13은 본 발명의 실시예에 따른 평가부(200)의 잠재 변수 추출부(230)의 동작 방법을 보여주는 흐름도이다.

[0117] 잠재 변수 추출부(230)는 학습부(130)로부터 오토인코더 파라미터 셋($\boldsymbol{\theta}$)을 수신한다(S1301).

[0118] 잠재 변수 추출부(230)는 변환 데이터(X')에 선형 함수(linear function)를 적용하여 데이터($\widetilde{\boldsymbol{X'}}$)를 생성한다(S1303).

[0119] 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).

[0120] 이후, 위험 지표 계산부(240)는 훈련 건강 검진 데이터의 잠재 변수의 평균 값을 이용하여 평가 대상 개인의 건강 위험 지표를 결정한다. 이때, 건강 위험 지표는 수학식 1을 이용하여 구해질 수 있으며, 위험 지표 계산부(240)는 건강 위험 지표를 임계값과 비교하여 평가 대상 개인의 건강 위험도 레벨을 판단하여 제공한다.

[0121] 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

[0122] 1: 위험 지표 산출 장치 2: 위험 지표 산출 시스템

10: 단말 장치 11: 통신부

20: 위험 지표 산출 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: 위험 지표 계산부 200: 평가부

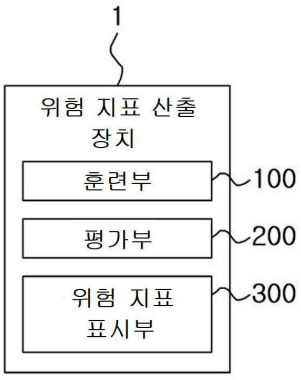
210: 위험 점수 산정부 220: 데이터 변환부

230: 잠재 변수 추출부 240: 위험 지표 계산부

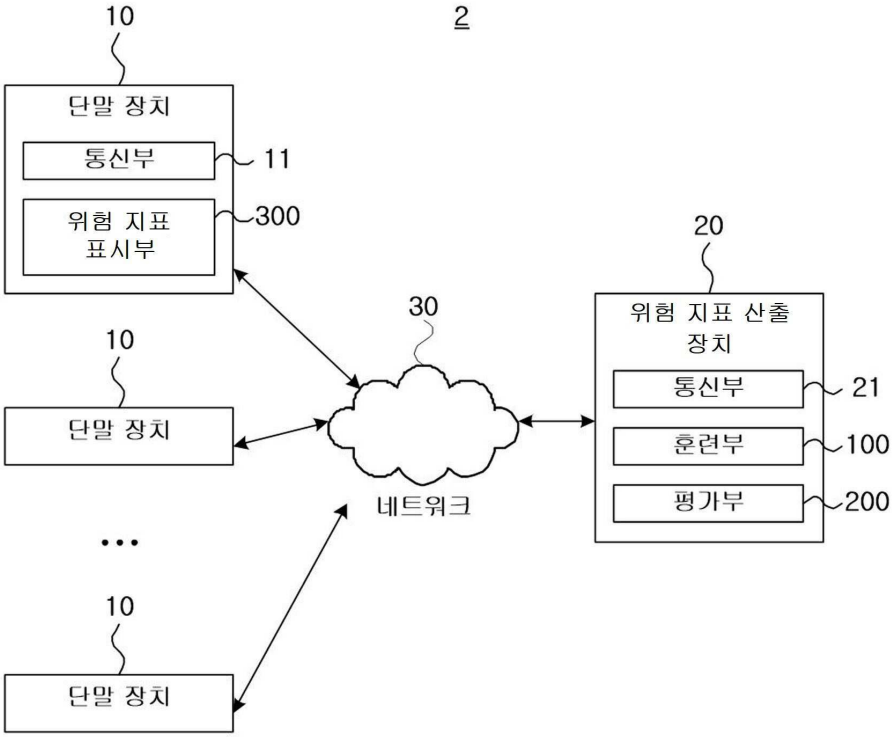
300: 위험 지표 표시부

도면

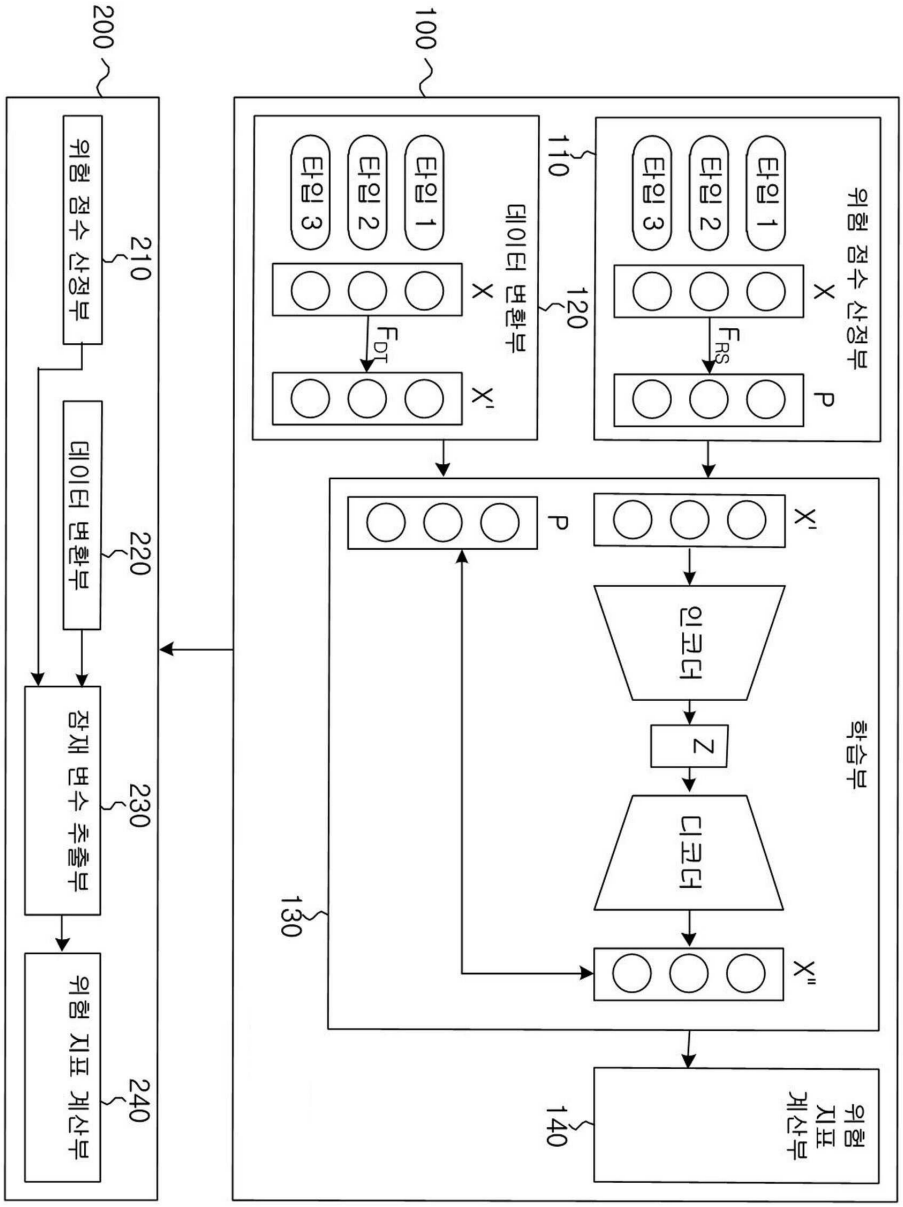
도면1



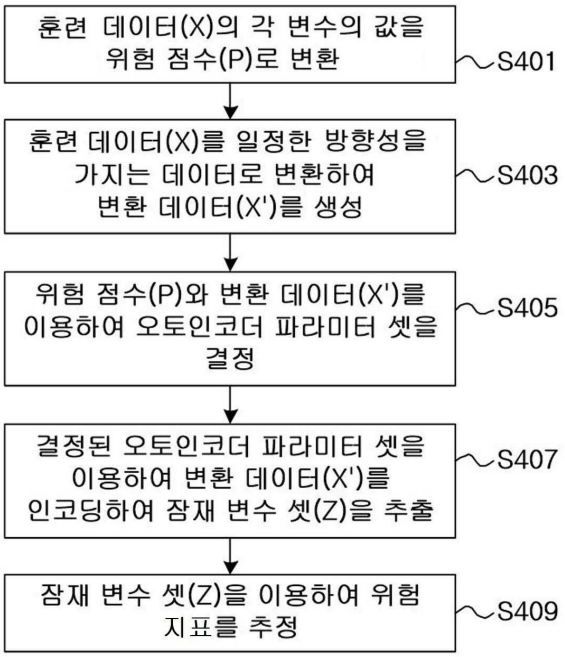
도면2


10
2
단말 장치
통신부
11
위험 지표
표시부
300
20
위험 지표 산출
장치
30
10
단말 장치
통신부
21
훈련부
100
평가부
200
네트워크
10
단말 장치

도면3


200
100
위험 점수 산정부
210
데이터 변환부
220
잠재 변수 추출부
230
위험 지표 계산부
240
110
데이터 변환부
120
위험 점수 산정부
타입 1
타입 2
타입 3
X
X'
P
F_{DT}
F_{RS}
130
학습부
인코더
Z
디코더
X''
140
위험 지표 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

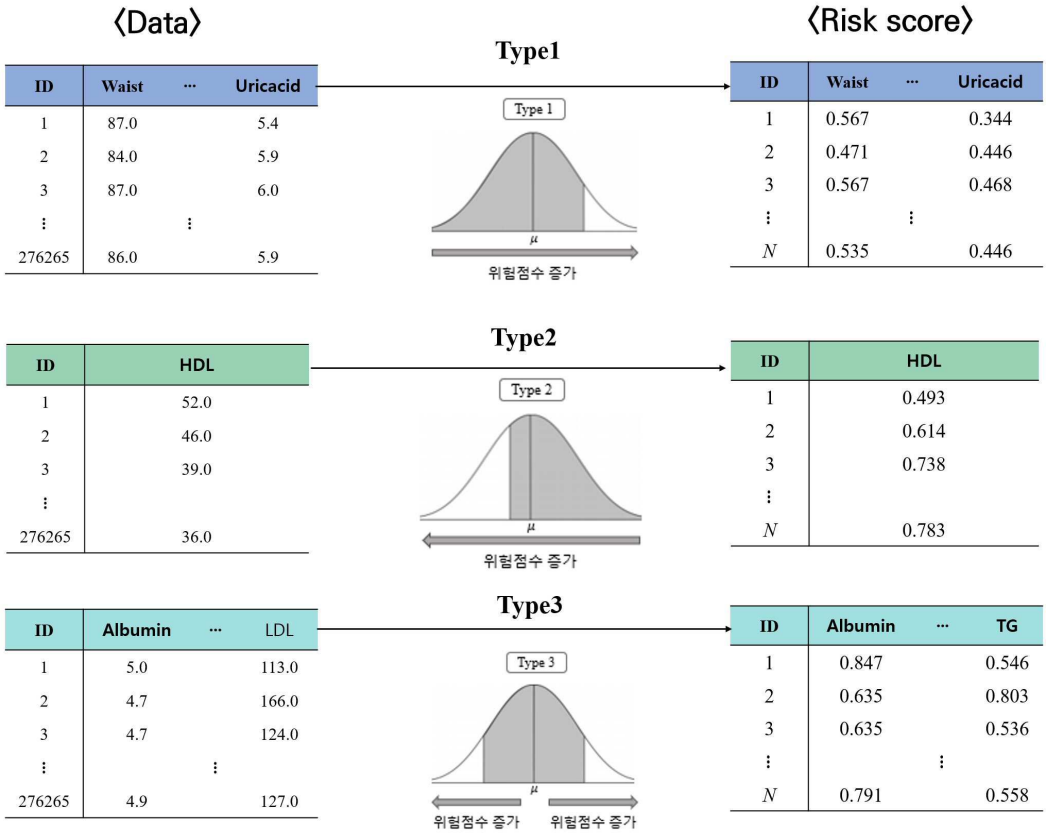
end function

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

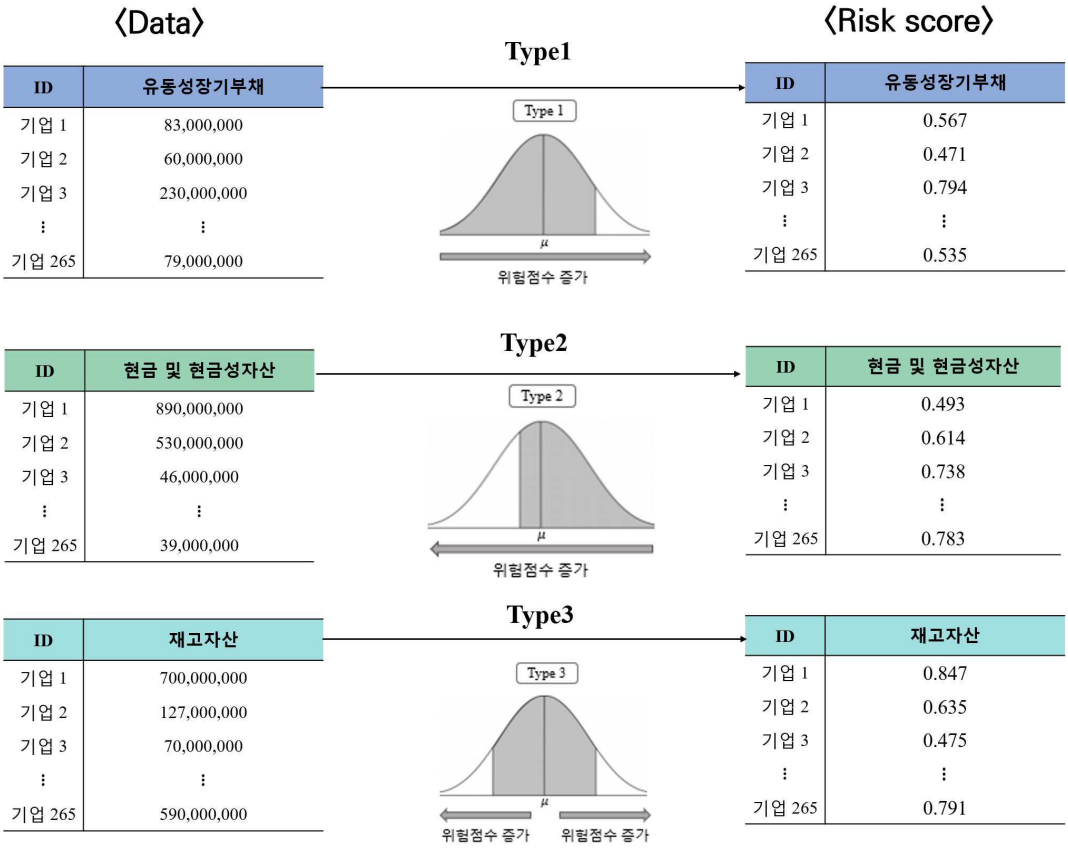
도면6

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

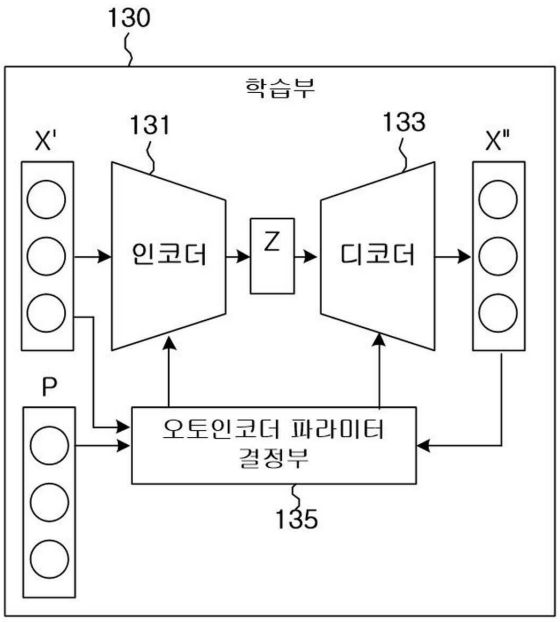
도면7a


〈Data〉
〈Risk score〉
Type1
ID
Waist
···
Uricacid
1
87.0
5.4
2
84.0
5.9
3
87.0
6.0
276265
86.0
5.9
0.567
0.344
0.471
0.446
0.567
0.468
N
0.535
0.446
Type 1
μ
위험점수 증가
Type2
HDL
52.0
46.0
39.0
36.0
0.493
0.614
0.738
0.783
Type 2
Type3
Albumin
LDL
TG
5.0
113.0
4.7
166.0
124.0
4.9
127.0
0.847
0.546
0.635
0.803
0.536
0.791
0.558
Type 3
위험점수 증가

도면7b


〈Data〉
〈Risk score〉
Type1
ID
유동성장기부채
기업 1
83,000,000
기업 2
60,000,000
기업 3
230,000,000
기업 265
79,000,000
0.567
0.471
0.794
0.535
Type 1
μ
위험점수 증가
Type2
현금 및 현금성자산
890,000,000
530,000,000
46,000,000
39,000,000
0.493
0.614
0.738
0.783
Type 2
Type3
재고자산
700,000,000
127,000,000
70,000,000
590,000,000
0.847
0.635
0.475
0.791
Type 3
위험점수 증가

도면8



도면9

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**

$\widetilde{\boldsymbol{X'}} = \boldsymbol{Linear}(\boldsymbol{X'})$ ~S901

$\boldsymbol{h} = f_a(\widetilde{\boldsymbol{X'}} * \boldsymbol{W} + \boldsymbol{b})$ ~S903

$\boldsymbol{X''} = f_a(\boldsymbol{h} * \boldsymbol{W^T} + \boldsymbol{b_h})$, f_a is an activation function ~S905

$\mathcal{L} = \mathcal{L}(\boldsymbol{X'}, \boldsymbol{X''}) + \mathcal{L}(\boldsymbol{X''}, \boldsymbol{P})$ ~S907

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: ~S909

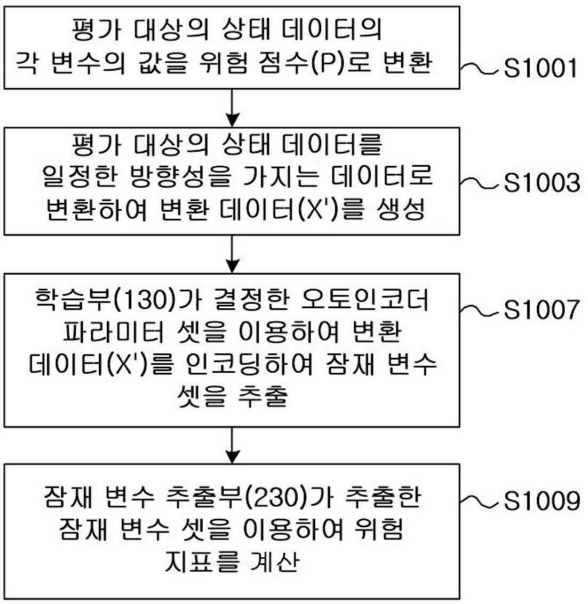
return $\boldsymbol{\theta}$

$iter \leftarrow iter + 1$

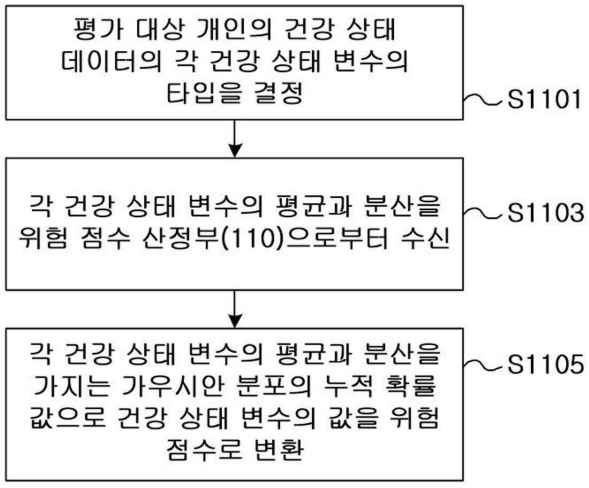
end for

end procedure

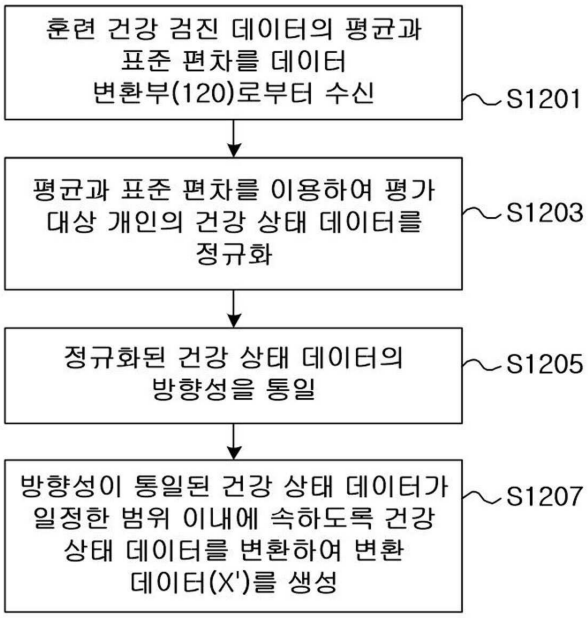
도면10


평가 대상의 상태 데이터의 각 변수의 값을 위험 점수(P)로 변환
S1001
평가 대상의 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 위험 지표를 계산
S1009

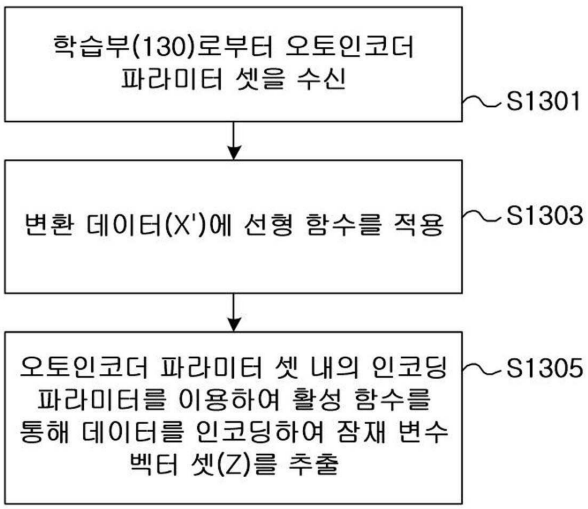
도면11


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