



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2021년01월05일
(11) 등록번호 10-2198302
(24) 등록일자 2020년12월28일

(51) 국제특허분류(Int. Cl.)

A61B 5/00 (2006.01) A61B 5/103 (2006.01)
A61B 5/107 (2006.01) G16H 50/20 (2018.01)
G16H 50/30 (2018.01)

(52) CPC특허분류

A61B 5/7275 (2013.01)
A61B 5/103 (2013.01)

(21) 출원번호 10-2020-0060627

(22) 출원일자 2020년05월20일

심사청구일자 2020년05월20일

(56) 선행기술조사문헌

KR1020190033972 A*

KR1020190072292 A*

KR1020200005404 A*

*는 심사관에 의하여 인용된 문헌

(73) 특허권자

주식회사 지피바이오

경기도 안산시 상록구 광덕1로 355, 209 (이동,
데코스포텔 월드비즈니스센터 35-4호)

세종대학교산학협력단

서울특별시 광진구 능동로 209 (군자동, 세종대학교)

(72) 발명자

성재혁

경기도 안산시 상록구 광덕1로 355, 209 35-4호(이동,
데코스포텔 월드비즈니스센터)

장윤

서울특별시 동대문구 정릉천동로 16, 103동 401호(용두동, 용두 두산위브)

손혜숙

경기도 남양주시 화도읍 먹갓로 54-3, 가동 202호(보아쉐르빌)

(74) 대리인

리엔목특허법인

전체 청구항 수 : 총 9 항

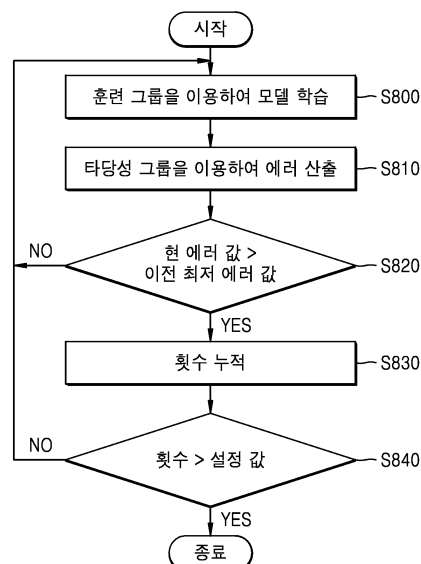
심사관 : 유창용

(54) 발명의 명칭 성장 예측 방법 및 그 장치

(57) 요약

성장 예측 방법 및 그 장치가 개시된다. 성장예측장치는 복수의 표본대상자에 대한 시계열적 신체 데이터를 기초로 학습데이터를 생성하고, 신체 데이터를 입력받아 성장 예측값을 출력하는 성장예측모델을 학습데이터를 이용하여 학습시킨 후 평가대상자의 신체 데이터를 상기 성장예측모델에 입력하여 평가대상자의 성장을 예측한다.

대표도 - 도8



(52) CPC특허분류

A61B 5/107 (2013.01)

G16H 50/20 (2018.01)

G16H 50/30 (2018.01)

명세서

청구범위

청구항 1

성장예측장치가 수행하는 성장 예측 방법에 있어서,

복수의 표본대상자에 대한 시계열적 신체 데이터를 기초로 학습데이터를 생성하는 단계;

신체 데이터를 입력받아 성장 예측값을 출력하는 성장예측모델을 상기 학습데이터를 이용하여 학습시키는 단계; 및

평가대상자의 신체 데이터를 상기 성장예측모델에 입력하여 상기 평가대상자의 성장을 예측하는 단계;를 포함하고,

상기 학습데이터를 생성하는 단계는,

단위기간별 시계열적 신체 데이터 중 백분위수를 기준으로 양단의 일정 범위에 속하는 신체 데이터를 이상치로 간주하여 제거하는 단계; 및

상기 시계열적 신체 데이터에서 단위기간별 연속하지 않는 표본대상자의 신체 데이터를 제거하는 단계;를 포함하는 것을 특징으로 하는 성장 예측 방법.

청구항 2

삭제

청구항 3

제 1항에 있어서, 상기 학습데이터를 생성하는 단계는,

각 표본대상자의 시계열적 신체 데이터를 단위기간별로 구분하는 단계;

각 표본대상자의 단위기간별 시계열적 신체 데이터에 대하여, 기 정의된 선별 방법을 이용하여 단위기간별 하나의 시계열적 신체 데이터를 선정하는 단계; 및

각 표본대상자의 단위기간별 시계열적 신체 데이터에 대하여, 단위기간별 연속하는 시계열적 신체 데이터를 일정 개수 단위로 분할하여 학습데이터를 생성하는 단계;를 포함하는 것을 특징으로 하는 성장 예측 방법.

청구항 4

제 1항에 있어서, 상기 학습시키는 단계는

신체 데이터를 훈련그룹 및 타당성그룹으로 분류하는 단계;

상기 성장예측모델을 상기 훈련그룹을 이용하여 1차 학습시키는 단계;

상기 1차 학습 후 상기 타당성그룹을 이용하여 상기 성장예측모델의 에러를 산출하는 단계; 및

n 차(n 은 2 이상의 자연수) 학습에서 산출한 에러가 이전의 학습과정들에서 산출된 $(n-1)$ 개의 에러 중 가장 작은 값보다 큰 경우가 기 정의된 횟수에 도달할 때까지 상기 성장예측모델을 상기 훈련그룹을 이용하여 학습시키고 상기 타당성그룹을 이용하여 에러를 산출하는 과정을 반복하는 단계;를 포함하는 것을 특징으로 하는 성장 예측 방법.

청구항 5

성장예측장치가 수행하는 성장 예측 방법에 있어서,

복수의 표본대상자에 대한 시계열적 신체 데이터를 기초로 학습데이터를 생성하는 단계;

신체 데이터를 입력받아 성장 예측값을 출력하는 성장예측모델을 상기 학습데이터를 이용하여 학습시키는 단계;

및

평가대상자의 신체 데이터를 상기 성장예측모델에 입력하여 상기 평가대상자의 성장을 예측하는 단계;를 포함하고,

상기 예측하는 단계는,

상기 평가대상자의 단위기간별 시계열적 신체 데이터를 상기 성장예측모델에 입력하여 제1 성장 예측값을 파악하는 단계; 및

상기 평가대상자의 시계열적 신체 데이터 중 가장 앞쪽의 신체 데이터를 제외한 신체 데이터와 상기 제1 성장 예측값을 새로운 입력데이터로 상기 성장예측모델에 입력하여 제2 성장 예측값을 파악하는 단계;를 포함하는 것을 특징으로 하는 성장 예측 방법.

청구항 6

제 1항에 있어서,

상기 성장예측모델은 LSTM(Long Short Term Memory) 아키텍처로 구현되는 것을 특징으로 하는 성장 예측 방법.

청구항 7

제 1항에 있어서,

상기 신체 데이터는 키와 함께 체중, 골격근량, 체질량지수, 기초 대사량, 목 둘레, 가슴 둘레, 복부 둘레, 허벅지 둘레, 엉덩이 둘레, 팔 둘레 중 적어도 하나 이상을 포함하는 것을 특징으로 하는 성장 예측 방법.

청구항 8

복수의 표본대상자에 대한 시계열적 신체 데이터를 기초로 학습데이터를 생성하는 데이터생성부;

신체 데이터를 입력받아 성장 예측값을 출력하는 성장예측모델을 상기 학습데이터를 이용하여 학습시키는 학습부; 및

평가대상자의 신체 데이터를 상기 성장예측모델에 입력하여 상기 평가대상자의 성장을 예측하는 예측부;를 포함하고,

상기 데이터생성부는,

단위기간별 시계열적 신체 데이터 중 백분위수를 기준으로 양단의 일정 범위에 속하는 신체 데이터를 이상치로 간주하여 제거하고, 상기 시계열적 신체 데이터에서 단위기간별 연속하지 않는 표본대상자의 신체 데이터를 제거하는 것을 특징으로 하는 성장 예측 장치.

청구항 9

제 8항에 있어서, 상기 학습부는,

상기 학습데이터를 훈련그룹 및 타당성그룹으로 분류하는 그룹분류부;

상기 성장예측모델을 상기 훈련그룹을 이용하여 1차 학습시키는 훈련부; 및

상기 1차 학습 후 상기 타당성그룹을 이용하여 상기 성장예측모델의 에러를 산출하는 에러산출부;를 포함하고,

상기 훈련부는 n차(n은 2 이상의 자연수) 학습에서 산출한 에러가 이전의 학습과정들에서 산출된 (n-1)개의 에러 중 가장 작은 값보다 큰 경우가 기 정의된 횟수에 도달할 때까지 상기 성장예측모델을 상기 훈련그룹을 이용하여 학습시키는 과정을 반복하는 것을 특징으로 하는 성장 예측 장치.

청구항 10

복수의 표본대상자에 대한 시계열적 신체 데이터를 기초로 학습데이터를 생성하는 데이터생성부;

신체 데이터를 입력받아 성장 예측값을 출력하는 성장예측모델을 상기 학습데이터를 이용하여 학습시키는 학습부; 및

평가대상자의 신체 데이터를 상기 성장예측모델에 입력하여 상기 평가대상자의 성장을 예측하는 예측부;를 포함하고,

상기 예측부는,

상기 평가대상자의 단위기간별 시계열적 신체 데이터를 상기 성장예측모델에 입력하여 제1 성장 예측값을 파악하고, 상기 평가대상자의 시계열적 신체 데이터 중 가장 앞쪽의 신체 데이터를 제외한 신체 데이터와 상기 제1 성장 예측값을 새로운 입력데이터로 상기 성장예측모델에 입력하여 제2 성장 예측값을 파악하는 것을 특징으로 하는 성장 예측 장치.

발명의 설명

기술 분야

[0001] 본 발명의 실시 예는 성장 예측 방법 및 그 장치에 관한 것으로, 보다 상세하게는 인공지능모델을 이용하여 성장기 청소년의 키 등의 성장을 예측하는 방법 및 그 장치에 관한 것이다.

배경 기술

[0002] 예측 분석은 데이터에서 정보를 추출하고 이를 사용하여 추세 및 행동 패턴 등을 예측하는 통계 및 데이터마이닝 영역의 기술이다. 이러한 예측 분석은 데이터에서 얻은 정보를 바탕으로 의사 결정에 필요한 모든 영역에 적용할 수 있다. 예측 분석의 핵심은 변수 간의 관계를 이해한 후에 알려지지 않은 변수를 예측해내는 부분이다. 이를 위해 데이터 특성 및 예측 대상에 따라 다양한 접근 방식들이 사용되고 있다.

[0003] 예측 분석이 필요한 다양한 분야 중에서, 청소년의 신체 성장 분야가 있다. 키 성장이 언제 나타나고 얼마나 성장할 것인지에 대하여 부모나 청소년 등의 많은 관심이 존재한다. 종래의 키 성장의 예측은 성장판을 X-ray 촬영하여 예측하거나 유전적/환경적 요인과의 관계를 분석하는 방법으로 이루어진다.

선행기술문헌

특허문헌

(특허문헌 0001) 등록특허공보 제10-2075743호

(특허문헌 0002) 등록특허공보 제10-1866208호

발명의 내용

해결하려는 과제

[0004] 본 발명의 실시 예가 이루고자 하는 기술적 과제는, 인공지능 학습을 통해 생성한 성장예측모델을 이용하여 키 등의 성장을 예측하는 방법 및 그 장치를 제공하는 데 있다.

과제의 해결 수단

[0005] 상기의 기술적 과제를 달성하기 위한, 본 발명의 실시 예에 따른 성장 예측 방법의 일 예는, 복수의 표본대상자에 대한 시계열적 신체 데이터를 기초로 학습데이터를 생성하는 단계; 신체 데이터를 입력받아 성장 예측값을 출력하는 성장예측모델을 상기 학습데이터를 이용하여 학습시키는 단계; 및 평가대상자의 신체 데이터를 상기 성장예측모델에 입력하여 상기 평가대상자의 성장을 예측하는 단계;를 포함한다.

[0006] 상기의 기술적 과제를 달성하기 위한, 본 발명의 실시 예에 따른 성장 예측 장치의 일 예는, 복수의 표본대상자에 대한 시계열적 신체 데이터를 기초로 학습데이터를 생성하는 데이터생성부; 신체 데이터를 입력받아 성장 예측값을 출력하는 성장예측모델을 상기 학습데이터를 이용하여 학습시키는 학습부; 및 평가대상자의 신체 데이터를 상기 성장예측모델에 입력하여 상기 평가대상자의 성장을 예측하는 예측부;를 포함한다.

발명의 효과

[0007] 본 발명의 실시 예에 따르면, 성장예측모델을 이용하여 정확한 키 등의 성장을 예측할 수 있다. 측정시기 또는

측정횟수가 각각 다른 표본대상자의 신체 데이터를 성장예측모델의 학습에 적합한 형태로 만들 수 있다. 또한, 학습과정에서 발생할 수 있는 과적합(overfitting) 문제를 해결할 수 있다. 키를 포함한 다양한 종류의 신체 정보를 함께 이용하여 정확한 성장 예측이 가능하다.

도면의 간단한 설명

[0008]

도 1은 본 발명의 실시 예에 따른 성장예측모델의 일 예를 도시한 도면,
 도 2는 본 발명의 실시 예에 따른 신체 데이터의 일 예를 도시한 도면,
 도 3은 본 발명의 실시 예에 따른 시계열적 신체 데이터의 전처리 방법의 일 예를 도시한 도면,
 도 4는 본 발명의 실시 예에 따른 시계열적 신체 데이터에서 이상치를 제거하는 방법의 일 예를 도시한 도면,
 도 5는 본 발명의 실시 예에 따른 시계열적 신체 데이터의 전처리 방법의 다른 예를 도시한 도면,
 도 6은 본 발명의 실시 예에 따른 학습 데이터의 생성 방법의 일 예를 도시한 도면,
 도 7은 본 발명의 실시 예에 따른 학습 데이터를 복수의 그룹으로 분류한 예를 도시한 도면,
 도 8은 본 발명의 실시 예에 따른 성장예측모델의 학습 방법의 일 예를 도시한 흐름도,
 도 9는 본 발명의 실시 예에 따른 성장예측모델을 이용하여 성장을 예측하는 방법의 일 예를 도시한 도면, 그리고,
 도 10은 본 발명의 실시 예에 따른 성장예측장치의 구성의 일 예를 도시한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0009]

이하에서, 첨부된 도면들을 참조하여 본 발명의 실시 예에 따른 성장 예측 방법 및 그 장치에 대해 상세히 설명한다.

[0010]

도 1은 본 발명의 실시 예에 따른 성장예측모델의 일 예를 도시한 도면이다.

[0011]

도 1을 참조하면, 성장예측모델(100)은 과거의 신체 데이터를 이용하여 미래의 성장을 예측하는 인공지능모델이다. 성장예측모델(100)은 시계열적 신체 데이터를 입력받으면 성장 예측값을 출력한다. 예를 들어, 평가대상자의 과거 3년치의 시계열적 신체 데이터를 입력받으면, 성장예측모델(100)은 평가대상자의 다음해 성장을 예측한 값(예를 들어, 예상 키 등)을 출력한다.

[0012]

성장예측모델(100)은 미리 정의된 학습 데이터를 이용한 학습과정이 필요하다. 본 실시 예에서는 학습 데이터로 복수의 표본대상자의 시계열적 신체 데이터를 이용한다. 신체 데이터의 일 예가 도 2에 도시되어 있다. 성장예측모델(100)이 출력하는 성장 예측값은 적어도 하나 이상의 신체 정보로 구성될 수 있다. 예를 들어, 성장예측모델은 도 2의 신체 데이터를 입력받으면, 도 2의 신체 데이터를 구성하는 정보와 동일한 종류의 정보를 포함하는 성장 예측값을 출력할 수 있다. 또는, 성장예측모델은 도 2의 신체 데이터를 입력받으면, 키 정보만을 포함하는 성장 예측값을 출력할 수 있다.

[0013]

표본대상자의 범위는 실시 예에 따라 다양하게 설정 가능하다. 예를 들어, 성장기 청소년을 대상으로 키 등의 성장 예측을 위한 성장예측모델(100)을 만든다면, 표본대상자의 범위를 8~19세 사이의 청소년으로 정할 수 있다.

[0014]

표본대상자의 키를 포함한 신체 데이터는 수집시기와 수집횟수는 다양하다. 예를 들어, 제1 표본대상자의 경우 8~19세 중 그 일부인 8~12세까지 측정된 신체 데이터가 존재하고, 제2 표본대상자의 경우 8, 10~12, 15세 등과 같이 불규칙적으로 측정된 신체 데이터가 존재할 수 있다. 또한 제3 표본대상자의 경우에 일정 기간 내 여러 번 측정된 신체 데이터가 존재하는 반면, 제4 표본대상자는 일정 기간 내 한 번만 측정된 신체 데이터가 존재할 수 있다.

[0015]

이와 같이, 수집시기와 수집횟수에 따라 일부 표본대상자의 신체 데이터가 다른 표본대상자보다 더 많거나 또는 일부 표본대상자의 신체 데이터가 불연속적으로 존재하므로, 이들 신체 데이터를 그대로 적용하여 성장예측모델을 학습시킬 경우 부정확한 학습 결과를 얻을 수 있다. 이를 해결하기 위하여 표본대상자의 신체 데이터를 가공하여 학습 데이터를 생성하는 방법에 대하여 도 3 내지 도 6에서 살펴본다.

[0016]

또한, 본 실시 예는 성장을 예측하는 것이므로 특정한 한 시점의 신체 데이터만으로 미래를 예측하는데 한계가

있을 수 있다. 예를 들어, 2019년에 측정한 키 정보만으로 2020년에 키를 예측하는 것보다 2017년, 2018년, 2019년에 순차적으로 측정한 키 정보를 모두 함께 이용하여 2020년에 키가 얼마가 될지 예측하는 것이 더 정확할 수 있다.

- [0017] 따라서 성장예측모델(100)은 현재 값뿐만 아니라 과거의 값을 함께 이용할 수 있도록 순환신경망(RNN, Recursive Neural Network) 구조의 인공지능으로 구현될 수 있다. 예를 들어, 성장예측모델(100)은 순환신경망인 LSTM(Long Short Term Memory) 또는 GRU(Gated Recurrent Units) 등의 아키텍처로 구현될 수 있다. 물론 이외에도 종래의 다양한 인공지능 아키텍처가 본 실시 예의 성장예측모델(100)에 적용될 수 있다.
- [0018] 도 2는 본 발명의 실시 예에 따른 신체 데이터의 일 예를 도시한 도면이다.
- [0019] 도 2를 참조하면, 신체 데이터(200)는 학년(또는 나이), 성별, 키 등의 기본정보뿐만 아니라 체중, 단백질이나 무기질 함량, 체지방, 체수분, 근육량(soft lean mass), 체지방량(fat free mass), 뼈 조직, 골격근량, 체질량지수(BMI, Body mass index), 기초대사량, 목 둘레, 가슴 둘레, 복부 둘레, 허벅지 둘레, 팔 둘레, 엉덩이 둘레 등의 추가정보를 더 포함할 수 있다.
- [0020] 도 2의 신체 데이터(100)는 본 발명의 이해를 돕기 위한 하나의 예일 뿐 본 실시 예가 이에 한정되는 것은 아니며, 실시 예에 따라 신체 데이터(100)를 구성하는 정보의 종류는 다양하게 변경 가능하다.
- [0021] 도 3은 본 발명의 실시 예에 따른 시계열적 신체 데이터의 전처리 방법의 일 예를 도시한 도면이다.
- [0022] 도 3을 참조하면, 신체 데이터의 측정시기에 따라 표본대상자의 신체 데이터는 T_0 부터 T_n 까지 시계열적으로 존재한다. 본 실시 예는 편의상 한 명의 표본대상자의 신체 데이터를 기준으로 설명한다.
- [0023] 시계열적으로 분포된 신체 데이터는 일정 시간 간격(이하, 단위기간)으로 구분된다. 단위기간은 실시 예에 따라 다양하게 설정 가능하다. 예를 들어, 1월1일부터 12월 31일까지를 단위기간으로 설정할 수 있다. 또는 단위기간을 1개월로 설정하거나 6개월로 설정하는 등 다양하게 설정할 수 있다. 이하에서는 설명의 편의를 위하여 단위기간이 1년 단위인 경우를 가정하여 설명한다.
- [0024] 본 실시 예에는 n 개의 단위기간이 존재한다. $T_0 \sim T_1$ 의 기간에서 신체 데이터(a,b)가 두 개 존재하고, $T_1 \sim T_2$ 기간에서는 신체 데이터(c)가 한 개 존재하고, $T_2 \sim T_3$ 기간에서는 신체 데이터가(d) 한 개 존재하고, $T_3 \sim T_4$ 기간에서는 신체 데이터(e,f)가 두 개 존재한다.
- [0025] 각 단위기간에 대하여 서로 다른 개수의 신체 데이터가 존재하므로, 성장예측장치는 각 단위기간에서 기 정의된 개수(예를 들어, 1개)의 신체 데이터를 선별한다. 예를 들어, 각 단위기간에 대하여 1개의 신체 데이터가 존재하도록 만드는 경우에, 성장예측장치는 $T_0 \sim T_1$ 기간에 존재하는 두 개의 신체 데이터(a,b) 및 $T_3 \sim T_4$ 기간에 존재하는 두 개의 신체 데이터(e,f)에서 각각 하나(a,f)씩 선택할 수 있다.
- [0026] 각 단위기간에서 어느 하나의 신체 데이터를 선택하는 다양한 기준이 미리 설정될 수 있다. 예를 들어, 키 성장의 예측이 주 목적인 경우에, 성장예측장치는 단위기간에 포함된 복수의 신체 데이터 중 키 값이 가장 큰 신체 데이터를 선택할 수 있다. 다른 예로, 성장예측장치는 단위기간에 존재하는 복수의 신체 데이터들 중 다양한 신체 정보(예를 들어, 체중이나 체질량 지수 등)가 미리 정의된 기준(예를 들어, 가장 큰 값 또는 가장 작은 값 또는 중간 값 등)에 해당하는 신체 데이터를 선택할 수 있다. 또는, 성장예측장치는 신체 데이터의 각 정보(키, 체중 등)에 대하여 기 정의된 가중치(예를 들어, 각 정보별 동일 또는 상이한 가중치)를 부여한 후 합산한 값이 미리 정의된 기준(예를 들어, 가장 큰 값 또는 가장 작은 값 또는 중간 값 등)에 해당하는 신체 데이터를 선택할 수 있다.
- [0027] 각 표본대상자에 대하여 단위기간별 기 정의된 개수의 신체 데이터를 선택하므로, 신체 데이터의 측정횟수와 관계없이 모든 표본대상자의 단위기간별 신체 데이터의 개수가 동일해진다.
- [0028] 도 4는 본 발명의 실시 예에 따른 시계열적 신체 데이터에서 이상치를 제거하는 방법의 일 예를 도시한 도면이다.
- [0029] 도 4를 참조하면, 복수의 표본대상자의 신체 데이터 중에서 키, 체중 등의 측정값이 평균에 비해 지나치게 크거나 작다면 이는 성장예측모델의 학습에서 제외하는 것이 바람직하다.
- [0030] 복수의 표본대상자 전체의 신체 데이터를 대상으로 이상치(outlier)를 제거할 경우에 가장 어린 나이 또는 가장 많은 나이의 표본대상자의 신체 데이터가 이상치로 분류되는 문제점이 발생하므로, 본 실시 예는 단위기간별 이

상치 제거를 수행한다.

- [0031] 예를 들어, 도 3의 예에서, 성장예측장치는 T0~T1 기간에 속하는 복수의 표본대상자의 신체 데이터를 대상으로 이상치 제거 과정을 수행하고, 또한 T1~T2, T2~T3, T3~T4의 각 단위기간에 대하여 이상치 제거 과정을 각각 수행한다.
- [0032] 이상치를 제거하는 다양한 기준은 미리 설정될 수 있다. 예를 들어, 성장예측장치는 어느 한 단위기간에 포함된 복수의 표본대상자의 신체 데이터 중 백분위수(percentile)를 기준으로 양단의 일정 범위에 속하는 신체 데이터를 이상치로 간주하여 제거할 수 있다. 일 예로, 성장예측장치는 75번째 백분위수(Q1)와 25번째 백분위수(Q3) 차이인 IQR(interquartile range)를 구하고, 신체 데이터의 값이 $(Q1 - 1.5 * IQR)$ 이하이거나, $(Q3 + 1.5 * IQR)$ 이상인 경우에 이상치로 간주하여 제거할 수 있다. 백분위수를 기준으로 이상치를 제거하는 것은 하나의 예일 뿐 이상치 제거에 평균이나 표준편차, 분산 등 다양한 통계적인 방법이 적용될 수 있다.
- [0033] 신체 데이터가 도 2와 같이 다양한 정보로 구성되는 경우에, 정보별로 이상치를 제거할 수 있다. 예를 들어, 성장예측장치는 키 값을 기준으로 이상치를 제거한다. 그리고 성장예측장치는 이상치를 제거하고 남은 신체 데이터들에 대하여 다시 체중을 기준으로 이상치를 제거한다. 이와 같은 방법으로 각 정보에 대한 이상치를 가진 신체 데이터를 제거할 수 있다.
- [0034] 예를 들어, 1~100번의 신체 데이터가 존재할 때 50번의 신체 데이터가 키에 대한 이상치로 파악된 경우에, 50번 신체 데이터는 제거된다. 제거 후 남은 1~49, 51~100번의 신체 데이터에 대하여 체중을 기준으로 검사한 결과 10, 20번의 신체 데이터가 이상치로 간주되는 경우, 10, 20번의 신체 데이터를 제거한다. 도 2와 같은 신체 데이터에서 이상치 제거에 사용되는 정보의 종류(예를들어, 키, 체중 등) 및 개수는 미리 설정될 수 있다. 즉, 키와 체중에 대해서 이상치를 제거하거나, 키, 체중, 체질량, 골격근량에 대해서 이상치를 제거하는 등 실시 예에 따라 이상치 제거에 사용되는 정보의 종류와 개수는 다양하게 설정 가능하다.
- [0035] 다른 실시 예로, 신체 데이터의 각 정보에 기 정의된 가중치(각 정보별 동일 또는 상이한 가중치)를 부여하여 합산한 값을 기준으로 백분위수를 구하고, 백분위수를 이용하여 이상치를 제거할 수도 있다.
- [0036] 도 5는 본 발명의 실시 예에 따른 시계열적 신체 데이터의 전처리 방법의 다른 예를 도시한 도면이다.
- [0037] 도 5를 참조하면, 표본대상자의 신체 데이터가 존재하지 않는 구간이 존재할 수 있다. 본 실시 예는 설명의 편의를 위하여 세 명의 표본대상자의 신체 데이터를 T0~T8 구간까지 도시하고 있다.
- [0038] 제1 표본대상자의 신체 데이터는 두 번째 단위기간(T1~T2), 7번째 단위기간(T6~T7)에 존재하지 않고, 제2 표본대상자의 신체 데이터는 4번째 단위기간(T3~T4) 및 8번째 단위기간(T7~T8)에 존재하지 않는다.
- [0039] 표본대상자의 불연속적인 신체 데이터가 성장예측모델의 학습 결과에 부정적인 영향을 미칠 수 있으므로, 성장예측장치는 불연속적인 신체 데이터를 제거한다.
- [0040] 신체 데이터의 최소 연속 구간은 미리 설정될 수 있다. 본 실시 예는 최소 연속 구간을 4로 설정한 경우의 예이다. 이 경우, 제1 표본대상자에 대하여 신체 데이터가 4개의 단위기간에 연속하여 존재하는 구간은 T2~T6 구간(500)이다. 제3 표본대상자에 대하여 신체 데이터가 4개의 단위기간에 연속하여 존재하는 구간은 T3~T7 구간(510)이다. 반면에, 제2 표본대상자의 신체 데이터는 4개의 단위기간에 연속하여 존재하는 구간이 존재하지 않는다.
- [0041] 이 경우에, 성장예측장치는 제1 표본대상자의 T2~T6 구간(500) 사이의 신체 데이터만을 선택하고 제1 표본대상자의 나머지 신체 데이터를 제거한다. 또한, 성장예측장치는 제3 표본대상자의 T3~T7 구간(510) 사이의 신체 데이터를 선택하고 제3 표본대상자의 나머지 신체 데이터를 제거한다. 제2 표본대상자의 경우 4개 연속하는 신체 데이터가 존재하지 않으므로, 성장예측장치는 제2 표본대상자의 신체 데이터 모두를 제거한다.
- [0042] 도 3 내지 도 5는 표본대상자의 신체 데이터에 대한 전처리 과정을 도시하고 있다. 실시 예에 따라 도 3 내지 도 5에 개시된 전처리 과정 중 어느 하나가 사용되거나 도 3 내지 도 5의 전처리 과정 전부가 사용될 수 있다. 또한, 도 3 내지 도 5의 각 전처리 과정의 수행순서는 실시 예에 따라 다양하게 변형 가능하다. 예를 들어, 도 5의 전처리 과정을 수행한 후 도 3의 전처리 과정을 수행하고, 그 다음에 도 4의 이상치 제거의 전처리 과정을 수행할 수 있다.
- [0043] 도 6은 본 발명의 실시 예에 따른 학습 데이터의 생성 방법의 일 예를 도시한 도면이다.
- [0044] 도 6을 참조하면, 표본대상자의 신체 데이터의 연속 구간의 개수는 서로 다를 수 있다. 예를 들어, 제1 표본대

상자의 신체 데이터는 T0~T7까지 7개의 신체 데이터가 연속하고, 제2 표본대상자의 신체 데이터는 T0~T4까지 4개의 신체 데이터가 연속할 수 있다. 각 표본대상자의 서로 다른 길이의 신체 데이터를 그대로 성장예측모델의 학습에 이용할 경우 성장예측모델이 불안정하므로 이를 그대로 학습에 사용하는데 어려움이 있다. 따라서 성장예측모델에 맞도록 미리 정의된 일정 개수의 신체 데이터를 포함하는 학습데이터의 생성이 필요하다.

- [0045] 본 실시 예는 설명의 편의를 위하여 어느 한 표본대상자의 신체 데이터 a,b,c,d,e,f의 6개가 각 단위기간에 존재하는 경우를 가정하여 설명한다. 또한, 성장예측장치의 학습을 위한 학습 데이터의 길이가 4인 경우를 가정하여 설명한다.
- [0046] 성장예측장치는 먼저 a,b,c,d로 구성된 제1 학습데이터(600)를 생성한다. 또한, 성장예측장치는 b,c,d,e로 구성된 제2 학습데이터(610) 및 c,d,e,f로 구성된 제3 학습데이터(620)를 생성한다. 이와 같은 방법으로 성장예측장치는 각 표본대상자의 신체 데이터에 대하여 일정 길이의 학습데이터를 생성할 수 있다. 즉, 성장예측장치는 4개의 단위기간으로 구성된 윈도우를 한칸씩 옮기면서 해당하는 신체 데이터를 추출하여 학습 데이터를 생성할 수 있다.
- [0047] 도 7은 본 발명의 실시 예에 따른 학습 데이터를 복수의 그룹으로 분류한 예를 도시한 도면이다.
- [0048] 도 7을 참조하면, 도 6에서 살핀 방법으로 생성한 학습 데이터(700)는 무작위로 훈련그룹(710), 타당성그룹(720) 및 테스트그룹(730)으로 분류될 수 있다. 각 그룹(710,720,730)에 속하는 학습 데이터의 개수는 실시 예에 따라 다양하게 변형 가능하며, 일 예로, 성장예측장치는 전체 학습 데이터를 (훈련그룹:타당성그룹:테스트그룹 = 8:1:1)로 분류할 수 있다.
- [0049] 훈련그룹(710)에 속한 학습 데이터는 성장예측모델의 학습에 사용된다. 학습 데이터가 도 6과 같이 4개의 연속하는 신체 데이터(예를 들어, a,b,c,d)로 구성된 경우에, 성장예측모델은 세 개의 연속하는 신체 데이터(a,b,c)를 입력받아 산출한 성장 예측값(d')과 학습 데이터의 네 번째 신체 데이터(d)를 비교하여 내부 파라미터 등을 조정하는 학습 과정을 수행할 수 있다.
- [0050] 실시 예에 따라 성장예측모델의 정확도를 높이기 위하여 동일한 훈련그룹(710)을 이용하여 성장예측모델의 학습 과정을 반복 수행할 수 있다. 그러나 동일 훈련그룹을 이용하여 성장예측모델의 학습과정을 반복 수행할 경우 과적합(overfitting)의 문제점이 발생할 수 있다. 과적합이 나타나지 않도록 반복 수행의 횟수를 조정하기 위하여 타당성그룹(720)이 이용된다.
- [0051] 타당성그룹(720)은 성장예측모델의 반복 학습의 종료 여부를 결정하기 위하여 사용되는 데이터이다. 타당성그룹(720)을 이용하여 성장예측모델의 반복 횟수를 제어하는 방법에 대하여 도 8에서 살펴본다.
- [0052] 테스트그룹(730)은 학습이 종료된 성장예측모델의 정확도 등을 파악하기 위하여 사용되는 데이터이다. 예를 들어, 성장예측장치는 테스트그룹(730)의 데이터를 성장예측모델에 입력하여 얻은 성장 예측값과 테스트그룹의 데이터를 비교하여 정확도 등을 산출할 수 있다. 실시 예에 따라 테스트그룹(730)은 생략 가능하다.
- [0053] 도 8은 본 발명의 실시 예에 따른 성장예측모델의 학습 방법의 일 예를 도시한 흐름도이다.
- [0054] 도 8을 참조하면, 성장예측장치는 훈련그룹을 이용하여 성장예측모델을 학습시킨다(S800). 훈련그룹을 이용한 1차 학습이 종료되면, 성장예측장치는 타당성그룹을 이용하여 성장예측모델의 에러를 산출한다(S810).
- [0055] 예를 들어, 성장예측장치는 타당성그룹에 속한 데이터를 성장예측모델에 입력하여 얻은 성장 예측값과 실제값을 비교하여 에러를 산출할 수 있다. MAPE(Mean Absolute Percentage Error), RMSE(Root Mean Square Error), MAE(Mean Absolute Error) 등이 에러 산출에 이용될 수 있다.
- [0056] 에러 산출 후 성장예측장치는 1차 학습시 사용했던 동일한 훈련그룹을 이용하여 성장예측모델을 다시 학습시킨다(S800). 2차 학습이 종료되면, 성장예측장치는 다시 타당성그룹을 이용하여 성장예측모델의 에러를 산출한다(S810).
- [0057] 성장예측장치는 1차 학습 후 에러와 2차 학습 후 에러를 비교한다(S820). 성장예측장치는 2차 학습 후 에러가 1차 학습 후 에러보다 더 큰 경우 횟수를 누적 저장한다(S830).
- [0058] 성장예측장치는 다시 훈련그룹을 이용하여 성장예측모델을 학습시키고, 타당성그룹을 이용하여 성장예측모델의 에러를 산출하는 과정을 반복한다. 이와 같이 n차 학습이 완료되면, 성장예측장치는 n차 학습 후의 에러가 1~(n-1)차까지 학습 후 측정된 에러값들 중 최저값보다 큰 값인지 파악한다(S820). n차 학습 후의 에러가 최저 에러값보다 크면, 성장예측장치는 해당 횟수를 누적저장하고(S840), 누적 저장된 횟수가 기 정의된 설정값(예를

들어, 20번) 보다 작으면(S840), (n+1)차 학습을 다시 수행하고 에러를 산출하는 과정을 반복한다.

- [0059] 성장예측장치는 n차 학습 후 에러가 최저 에러값보다 크고(S820), 이러한 경우를 누적 저장한 횟수가 기 정의된 설정값보다 크면(S840), 과적합을 방지하기 위하여 반복 학습을 종료한다.
- [0060] 다른 실시 예로, 성장예측장치는 반복 학습 횟수를 미리 정의할 수 있다. 예를 들어, 반복 학습 횟수를 10회로 설정한 경우에, 성장예측장치는 10회까지 연속하여 에러가 최저값에 도달하는 경우에도 10번째 학습 과정 수행 후 반복 학습을 종료할 수 있다.
- [0061] 또 다른 실시 예로, 성장예측장치는 현 학습 과정 후 에러가 이전의 에러 최저값보다 큰 경우의 횟수를 파악하는 방법(S840)과 반복 학습 횟수를 미리 설정하는 방법을 동시에 적용하여 반복 학습을 수행할 수 있다. 예를 들어, 현 에러가 에러 최저값보다 큰 경우의 횟수를 5로 설정하고, 반복 학습 횟수를 20으로 설정한 경우에, 성장예측장치는 반복 학습 횟수가 20회 될 때까지 에러가 에러 최저값보다 큰 경우의 횟수가 5 미만이어도 20회까지만 반복 학습을 수행할 수 있다.
- [0062] 도 9는 본 발명의 실시 예에 따른 성장예측모델을 이용하여 성장을 예측하는 방법의 일 예를 도시한 도면이다.
- [0063] 도 9를 참조하면, 학습이 완료된 성장예측모델을 이용하여 평가대상자의 장래 성장을 예측할 수 있다. 본 실시 예는 3개의 연속하는 신체 데이터가 성장예측모델의 입력데이터인 경우를 가정한다. 성장예측모델의 입력데이터의 길이는 실시 예에 따라 2개 이상으로 다양할 수 있다.
- [0064] 성장예측장치는 평가대상자의 연속하는 3개의 신체 데이터를 성장예측모델에 입력한다. 성장예측모델은 3개의 연속하는 신체 데이터를 이용하여 다음 단위기간에 대한 성장 예측값을 출력한다. 예를 들어, a,b,c의 3개의 신체 데이터가 2017~2019년의 측정 데이터인 경우에, 성장예측모델은 2020년에 대한 평가대상자의 성장 예측값을 출력한다.
- [0065] 장래의 여러 단위기간에 대한 성장 예측값을 구하는 경우에, 성장예측장치는 성장예측모델의 성장 예측값을 입력 데이터로 이용할 수 있다. 예를 들어, 성장예측모델이 a,b,c의 신체 데이터를 이용하여 그 다음 단위기간에 대한 성장 예측값 d를 구했다면, 성장예측장치는 평가대상자의 신체 데이터 중 가장 앞쪽의 신체 데이터 a를 제외하고, 성장예측값 d를 포함하는 새로운 입력 데이터를 성장예측모델에 입력하여, 그 다음 단위기간에 대한 성장 예측값 e를 산출한다.
- [0066] 도 10은 본 발명의 실시 예에 따른 성장예측장치의 구성의 일 예를 도시한 도면이다.
- [0067] 도 10을 참조하면, 성장예측장치(1000)는 데이터생성부(1010), 학습부(1020) 및 예측부(1030)를 포함한다. 성장예측장치(1000)는 메모리, 프로세서, 입출력장치 등을 포함한 컴퓨터일 수 있으며, 각 구성(1010,1020,1030)은 소프트웨어로 구현되어 메모리에 탑재되어 프로세서에 의해 실행될 수 있다.
- [0068] 데이터생성부(1010)는 복수의 표본대상자의 시계열적 신체 데이터를 이용하여 학습데이터를 생성한다. 데이터생성부(1010)는 도 3 내지 도 5에 개시된 전처리 과정 중 일부 또는 전부를 수행한 후 도 6과 같이 일정 개수의 신체 데이터로 구성된 학습 데이터를 생성할 수 있다. 또한, 데이터생성부(1010)는 도 7과 같이 학습 데이터를 훈련그룹, 타당성그룹 및 테스트그룹으로 분류한다. 실시 예에 따라 테스트그룹은 생략될 수 있다.
- [0069] 학습부(1020)는 학습 데이터를 이용하여 성장예측모델을 학습시킨다. 일 예로, 학습부(1020)는 훈련부(1024) 및 에러산출부(1026)를 포함할 수 있다. 훈련부(1024)는 훈련그룹에 속한 학습 데이터를 이용하여 성장예측모델을 학습시킨다. 에러산출부(1026)는 각 학습 회차의 성장예측모델의 학습이 완료되면 타당성그룹에 속한 데이터를 이용하여 성장예측모델의 에러를 산출한다. 훈련부(1024)는 동일한 훈련그룹을 이용하여 n차까지 성장예측모델을 반복 훈련시킬 수 있다. 훈련부(1024)는 에러산출부(1026)를 통해 구한 현 학습 회차의 에러가 이전까지 각 학습회차에서 구한 에러의 최저값보다 큰 경우가 일정 횟수 이상 나올 때까지 성장예측모델을 반복 훈련시킬 수 있다.
- [0070] 예측부(1030)는 학습이 완료된 성장예측모델을 이용하여 평가대상자의 성장 예측값을 구한다. 예를 들어, 평가대상자의 과거 일정 기간 동안의 신체 데이터를 입력받으면, 예측부(1030)는 성장예측모델을 통해 향후 성장을 예측한 값을 구하여 출력한다. 예측부(1030)는 장래의 일정 기간 동안의 성장을 예측할 때 성장예측모델이 구한 성장 예측값을 입력으로 사용할 수 있으며, 그 일 예가 도 9에 도시되어 있다.
- [0071] 본 발명은 또한 컴퓨터로 읽을 수 있는 기록매체에 컴퓨터가 읽을 수 있는 프로그램 코드로서 구현하는 것이 가능하다. 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 중

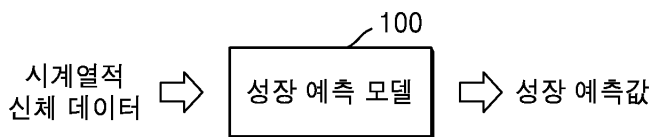
류의 기록장치를 포함한다. 컴퓨터가 읽을 수 있는 기록매체의 예로는 ROM, RAM, CD-ROM, 자기 테이프, 플로피 디스크, 광데이터 저장장치 등이 있다. 또한 컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어 분산방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수 있다.

[0072]

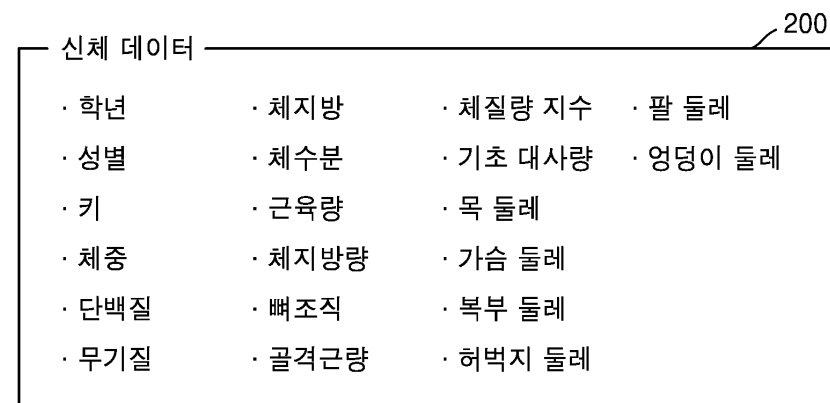
이제까지 본 발명에 대하여 그 바람직한 실시 예들을 중심으로 살펴보았다. 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 본 발명이 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 변형된 형태로 구현될 수 있음을 이해할 수 있을 것이다. 그러므로 개시된 실시 예들은 한정적인 관점이 아니라 설명적인 관점에서 고려되어야 한다. 본 발명의 범위는 전술한 설명이 아니라 특허청구범위에 나타나 있으며, 그와 동등한 범위 내에 있는 모든 차이점은 본 발명에 포함된 것으로 해석되어야 할 것이다.

도면

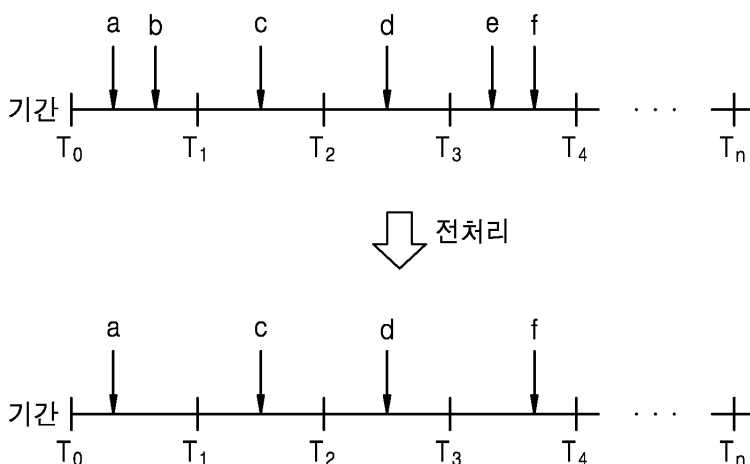
도면1



도면2



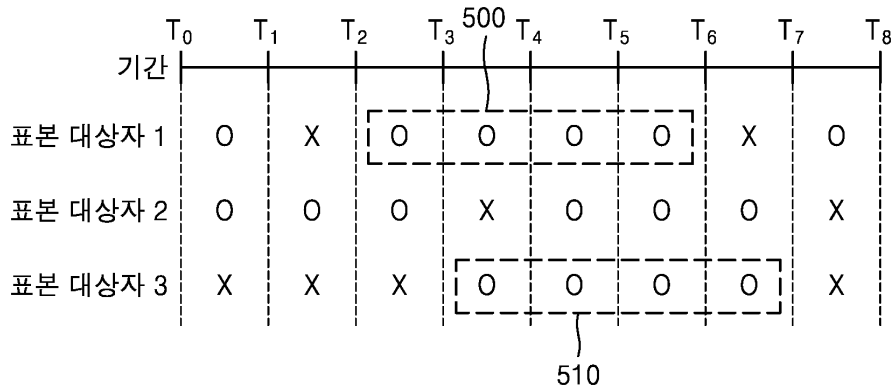
도면3



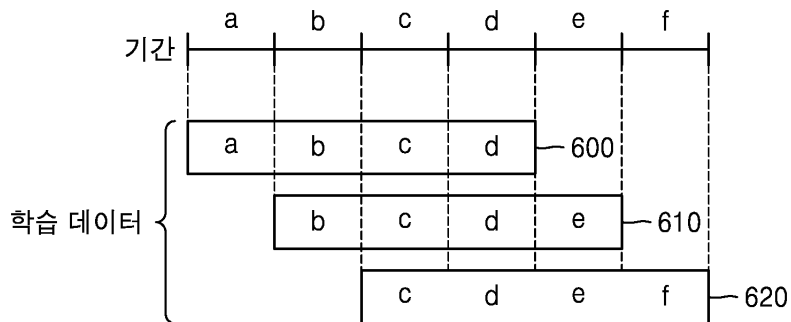
도면4

$IQR = Q_3 - Q_1$
 이상치 : $(Q_1 - 1.5 * IQR)$ 이하
 $(Q_3 + 1.5 * IQR)$ 이상

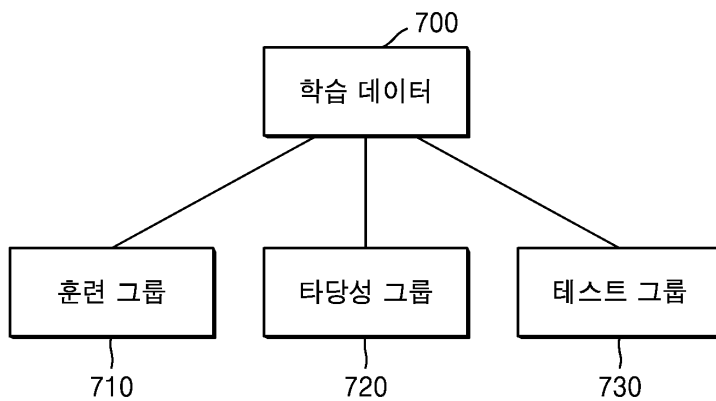
도면5



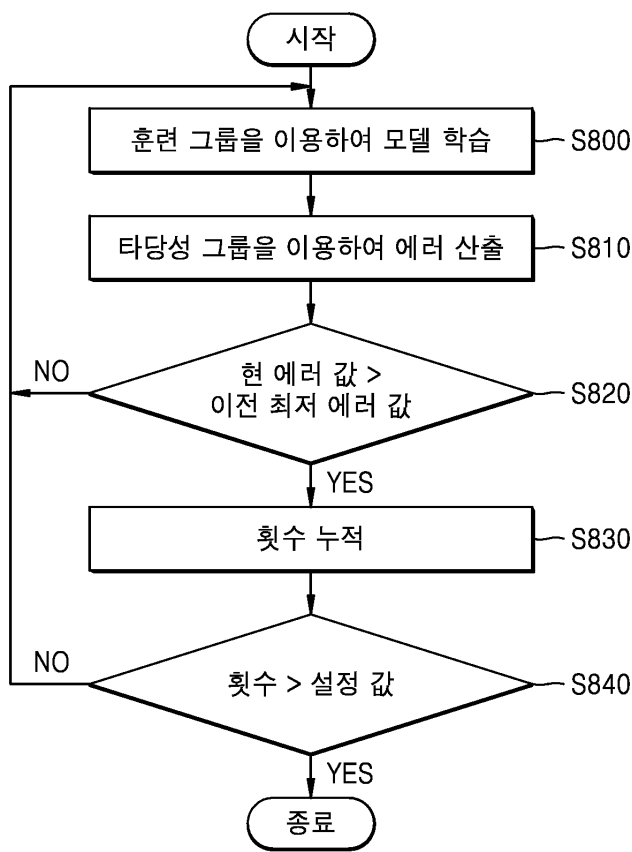
도면6



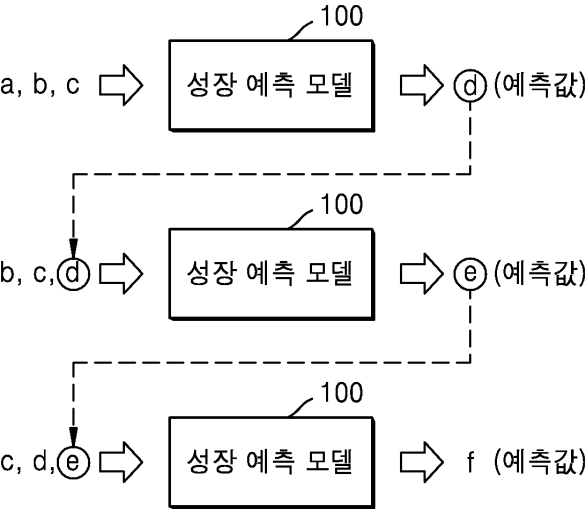
도면7



도면8



도면9



도면10

