

명세서

발명의 명칭: 라이프로그 데이터 간 인과 관계 분석 장치 및 그 방법 기술분야

- [1] 본 발명은 라이프로그 데이터 간 인과 관계 분석 장치 및 그 방법에 관한 것으로서, 더욱 상세하게는 라이프로그 데이터를 이용하여 사람과 환경 사이의 인과 관계 및 영향력을 분석하기 위한 라이프로그 데이터 간 인과 관계 분석 장치 및 그 방법에 관한 것이다.

배경기술

- [2] 체화된 인지 또는 연장된 마음 이론에 따르면 인간의 몸과 뇌 그리고 인간을 둘러싸고 있는 환경은 유기적인 상호 작용이 이루어짐으로써 하나의 통합적 단위를 이룬다. 환경은 인간의 신체 반응과 뇌 인지과정에 영향을 미치며 역으로 인간의 신체 반응과 뇌의 인지활동이 환경을 변화시킨다.
- [3] 따라서 인간과 환경을 정확히 파악하기 위해서는 인간의 신체 반응(즉, 내현적 반응)과 뇌의 인지적 반응인 행동(즉, 외현적 반응), 그리고 인간을 둘러싸고 있는 환경 사이의 상호 작용을 파악하는 것이 중요하다.
- [4] 이러한 내현적 반응, 외현적 반응, 환경 사이에 상호 작용이 있다는 것은 상관관계 분석을 통해 알 수 있다. 그러나 이들 사이에 단순히 상관성이 있다는 것만으로는 상호 작용의 원인과 결과에 대한 정보를 취득할 수 없다.
- [5] 그러므로 인간의 내현적 반응, 외현적 반응 그리고 환경을 반영하는 다양하고 이질적인 데이터에 대한 인과 관계가 분석될 수 있다면 인간과 환경의 상호 작용 사이에서 발생하는 현상에 대한 해석과 판단이 용이해 질 것이다.
- [6] 본 발명의 배경이 되는 기술은 한국등록특허 제10-1262922호(2013.05.09.공고)에 개시되어 있다.

발명의 상세한 설명

기술적 과제

- [7] 본 발명이 이루고자 하는 기술적 과제는 라이프로그 데이터를 이용하여 사람과 환경 사이의 인과 관계 및 영향력을 분석하기 위한 라이프로그 데이터 간 인과 관계 분석 장치 및 그 방법을 제공하기 위한 것이다.

과제 해결 수단

- [8] 이러한 기술적 과제를 이루기 위한 본 발명의 실시예에 따르면 라이프로그 데이터 간 인과 관계 분석 장치를 이용한 인과 관계 분석 방법에 있어서, 인과 관계 분석 방법은 생체 신호 정보, GPS 신호 정보 및 음성 신호 정보를 포함하는 라이프로그 데이터를 입력받는 단계, 상기 라이프로그 데이터의 변수 중 두 개 이상의 변수를 선택하여 변수 조합 및 상기 변수 조합에 대한 경로 모델(path model)을 생성하고, 상기 라이프로그 데이터 및 상기 경로 모델을 이용하여 상기 변수 조합 각각에 대한 변수간 인과 정보를 산출하는 단계, 기 설정된 임계값을

이용하여 상기 산출된 변수간 인과 정보가 통계적 유의성을 가지는지 여부를 판단하는 단계, 그리고 상기 통계적 유의성을 가진다고 판단된 변수간 인과 정보를 화면상에 출력하는 단계를 포함한다.

- [9] 상기 라이프로그킹 데이터의 변수는, 상기 생체 신호 정보에 대한 변수, 상기 GPS 신호 정보에 대한 변수 및 상기 음성 신호 정보에 대한 변수 중에서 적어도 하나를 포함할 수 있다.
- [10] 상기 생체 신호 정보에 대한 변수는 BPM(Beat Per Minutes), HF(High Frequency), LF(Low Frequency) 및 VLF(Very Low Frequency) 중에서 적어도 하나를 포함하고, 상기 GPS 신호 정보에 대한 변수는 이동 패턴의 불규칙성을 의미하는 이동 규칙성(circadian movement), 위치 클러스터에서의 시간 변동성을 의미하는 위치 엔트로피(entropy), 위도 및 경도의 변화량을 의미하는 위치 변화량(location variance), 위치 클러스터의 수(number of clusters), 총 이동 거리(total distance) 및 총 이동 시간(transition time) 중에서 적어도 하나를 포함하며, 상기 음성 신호 정보에 대한 변수는 소음 크기의 분당 평균값을 의미하는 주변 소음 레벨(noise amplitude)을 포함할 수 있다.
- [11] 상기 변수간 인과 정보는, 상기 변수간 인과 관계의 방향성 및 상기 경로 모델($\hat{y}=ax+b$)의 경로 계수(path coefficient) 중 적어도 하나를 포함할 수 있다.
- [12] 상기 변수간 인과 정보를 산출하는 단계는, 상기 변수 조합에 대한 변수 간 공분산 값을 산출하는 단계, 상기 변수 조합을 구성하는 변수 사이의 인과 경로에 따른 복수의 경로 모델을 생성하는 단계, 상기 산출된 공분산 값을 이용하여 상기 복수의 경로 모델에 대한 경로 계수 및 표준 오차를 산출하는 단계, 상기 복수의 경로 모델에 대한 표준 오차 중 가장 작은 값을 가지는 표준 오차에 대응하는 경로 모델을 선택하여 상기 변수간 인과 관계의 방향성을 판단하는 단계, 그리고 상기 표준 오차를 이용하여 상기 선택된 경로 모델에 대한 유의 지표를 산출하는 단계를 포함할 수 있다.
- [13] 상기 경로 계수(r) 및 표준 오차(SE)는 아래의 수학식을 이용하여 산출될 수 있다.

$$[14] \quad r = \frac{(n-1)cov(x,y)}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

$$[15] \quad SE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n-2)}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

[16] 여기서, $\hat{y}=ax+b$ 이고, $a=r \times \sqrt{\frac{(y-\bar{y})^2}{(x-\bar{x})^2}}$ 이고, $b = \frac{\sum_{i=1}^n y_i - a \times \sum_{i=1}^n x_i}{n}$ 이고, i 는

상기 라이프로그 데이터의 인덱스 값을 의미하고, x 는 상기 라이프로그 데이터의 제1 변수를 의미하고, y 는 상기 라이프로그 데이터의 제2 변수를 의미하고, \bar{x} 는 상기 제1 변수의 평균값을 의미하고, \bar{y} 는 상기 제2 변수의 평균값을 의미한다.

- [17] 상기 산출된 변수간 인과 정보의 유의성 여부를 판단하는 단계는, 상기 유의 지표($t = \frac{a}{SE}$)가 상기 기 설정된 임계값보다 크면 상기 선택된 경로 모델에

대응하는 변수간 인과 정보가 통계적 유의성을 가진다고 판단할 수 있다.

- [18] 상기 변수간 인과 정보를 화면상에 출력하는 단계는, 상기 변수는 노드(node) 형태로 형성되고, 인과 관계를 가지는 변수끼리는 화살표 형상의 엣지(edge)로 연결시킨 그래프로 표시하며, 상기 노드는 상기 라이프로그 데이터의 변수에 따라 서로 다른 색상으로 표시되고, 상기 엣지는 상기 경로 계수가 클수록 굵게 표시될 수 있다.

- [19] 본 발명의 다른 실시예에 따른 인과 관계 분석 장치는 생체 신호 정보, GPS 신호 정보 및 음성 신호 정보를 포함하는 라이프로그 데이터를 입력받는 입력부, 상기 라이프로그 데이터의 변수 중 두 개 이상의 변수를 선택하여 변수 조합 및 상기 변수 조합에 대한 경로 모델(path model)을 생성하고, 상기 라이프로그 데이터 및 상기 경로 모델을 이용하여 상기 변수 조합 각각에 대한 변수간 인과 정보를 산출하는 산출부, 기 설정된 임계값을 이용하여 상기 산출된 변수간 인과 정보가 통계적 유의성을 가지는지 여부를 판단하는 판단부, 그리고 상기 통계적 유의성을 가진다고 판단된 변수간 인과 정보를 화면상에 출력하는 제어부를 포함한다.

발명의 효과

- [20] 이와 같이 본 발명에 따르면, 다양한 라이프로그 데이터 간 인과 관계를 분석하여 제공하므로 사람과 환경 사이의 인과 관계를 용이하게 판단할 수 있다. 또한 사람이 환경에 미치는 영향력이나 환경이 사람에게 미치는 영향력을 정확히 분석할 수 있으므로, 환경 개선이나 인력 개발과 같은 다양한 업무 개선에 이용되는 유용한 정보를 제공할 수 있다.

도면의 간단한 설명

- [21] 도 1은 본 발명의 실시예에 따른 인과 관계 분석 장치의 구성도이다.
 [22] 도 2는 본 발명의 실시예에 따른 인과 관계 분석 방법의 순서도이다.
 [23] 도 3은 도 2의 S240 단계를 상세하게 나타낸 순서도이다.
 [24] 도 4는 본 발명의 실시예에 따른 인과 관계 정보를 화면에 출력하는 방법을 설명하기 위한 도면이다.

발명의 실시를 위한 최선의 형태

- [25] 아래에서는 첨부한 도면을 참고로 하여 본 발명의 실시예에 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 실시예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

- [26] 명세서 전체에서, 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있다는 것을 의미한다.
- [27] 그러면 첨부한 도면을 참고로 하여 본 발명의 실시예에 대하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다.
- [28] 우선 도 1을 통해 본 발명의 실시예에 따른 라이프로그 데이터 간 인과 관계 분석 장치에 대해 살펴보도록 한다. 도 1은 본 발명의 실시예에 따른 인과 관계 분석 장치의 구성도이다.
- [29] 도 1에 도시된 바와 같이, 본 발명의 실시예에 따른 인과 관계 분석 장치(100)는 입력부(110), 산출부(120), 판단부(130) 및 제어부(140)를 포함한다.
- [30] 먼저 입력부(110)는 생체 신호 정보, GPS 신호 정보 및 음성 신호 정보를 포함하는 라이프로그 데이터를 입력받는다.
- [31] 다음으로 산출부(120)는 라이프로그 데이터의 변수 중 두 개 이상의 변수를 선택하여 변수 조합 및 변수 조합에 대한 경로 모델(path model)을 생성한다. 그리고 산출부(120)는 라이프로그 데이터 및 경로 모델을 이용하여 변수 조합 각각에 대한 변수간 인과 정보를 산출한다.
- [32] 구체적으로 산출부(120)는 변수 조합에 대한 변수 간 공분산 값을 산출한다. 그러면 산출부(120)는 변수 조합을 구성하는 변수 사이의 인과 경로에 따른 복수의 경로 모델을 생성한다. 그리고 산출부(120)는 산출된 공분산 값을 이용하여 복수의 경로 모델에 대한 경로 계수 및 표준 오차를 산출한다.
- [33] 그러면 산출부(120)는 복수의 경로 모델에 대한 표준 오차 중 가장 작은 값을 가지는 표준 오차에 대응하는 경로 모델을 선택하여 변수간의 인과 관계의 방향성을 판단한다. 그리고 산출부(120)는 표준 오차를 이용하여 선택된 경로 모델에 대한 유의 지표를 산출한다.
- [34] 여기서 변수간 인과 정보는 변수간 인과 관계의 방향성 및 경로 모델의 경로 계수(path coefficient) 중 적어도 하나를 포함한다.
- [35] 다음으로 판단부(130)는 기 설정된 임계값을 이용하여 산출된 변수간 인과 정보가 통계적 유의성을 가지는지 여부를 판단한다.
- [36] 구체적으로 판단부(130)는 유의 지표가 기 설정된 임계값보다 크면 선택된 경로 모델에 대응하는 변수간 인과 정보가 통계적 유의성을 가진다고 판단한다.
- [37] 다음으로 제어부(140)는 통계적 유의성을 가진다고 판단된 변수간 인과 정보를 화면상에 출력한다.
- [38] 이때 변수는 노드(node) 형태로 형성되고, 인과 관계를 가지는 변수끼리는

화살표 형상의 엣지(edge)로 연결시킨 그래프로 표시된다. 그리고 노드는 라이프로그킹 데이터의 변수에 따라 서로 다른 색상으로 표시되고, 엣지는 경로 계수가 클수록 굵게 표시된다.

[39] 아래에서는 표 1을 통해 본 발명의 실시예에 따른 라이프로그킹 데이터 및 라이프로그킹 데이터의 변수를 설명한다.

[40] 표 1은 본 발명의 실시예에 따른 라이프로그킹 데이터 및 이에 따른 변수를 나타낸다.

[41] [표1]

라이프로그킹 데이터	변수
생체 신호	BPM(Beat Per Mintues)
	HF(High Frequency)
	LF(Low Frequency)
	VLF(Very Low Frequency)
GPS 신호	이동 규칙성(circadian movement)
	위치 엔트로피(entropy)
	위치 변화량(location variance)
	위치 클러스터의 수(number of clusters)
	총 이동 거리(total distance)
	총 이동 시간(transition time)
음성 신호	주변 소음 레벨(noise amplitude)

[42]

[43] 표 1에 나타난 바와 같이, 본 발명의 실시예에 따른 라이프로그킹 데이터의 변수는 생체 신호 정보에 대한 변수, GPS 신호 정보에 대한 변수, 음성 신호 정보에 대한 변수 중에서 적어도 하나를 포함한다.

[44] 우선 생체 신호 정보에 대한 변수는 사용자의 생체 신호의 변화와 같은 내현적 반응에 따른 변수를 의미한다. 여기서 생체 신호 정보는 사용자의 심혈관 반응에 따른 심전도, 맥파, 심탄도와 같은 생체 신호를 포함하며, 생체 신호 측정 센서가 구비된 웨어러블 장치(wearable device)나 스마트폰(smart phone)과 같은 사용자 단말을 통해 측정될 수 있다.

[45] 구체적으로 생체 신호 정보에 대한 변수는 BPM(Beat Per Minutes), HF(High Frequency), LF(Low Frequency) 및 VLF(Very Low Frequency) 중에서 적어도 하나를 포함한다. 여기서 피크 검출 알고리즘이나 심박 변동성 분석 알고리즘은 본 발명의 실시예에 따른 인과 관계 분석 장치(100)에 기 저장될 수 있으며, BPM은 생체 신호의 피크 검출 알고리즘을 통해 산출될 수 있다. 또한, HF, LF,

VLF는 심박 변동성(Heart Rate Variability, HRV) 분석 알고리즘을 통해 산출될 수 있다.

[46] 다음으로 GPS 신호 정보에 대한 변수는 사용자의 이동이나 움직임 등 생활 패턴과 같은 외현적 반응에 따른 변수를 의미한다. 여기서 GPS 신호는 GPS 센서가 구비된 웨어러블 장치나 스마트폰과 같은 사용자 단말을 통해 측정될 수 있다.

[47] 구체적으로 GPS 신호 정보에 대한 변수는 이동 규칙성(circadian movement), 위치 엔트로피(entropy), 위도 및 경도의 변화량을 의미하는 위치 변화량(location variance), 위치 클러스터의 수(number of clusters), 총 이동 거리(total distance) 및 총 이동 시간(transition time) 중에서 적어도 하나를 포함한다.

[48] 먼저 이동 규칙성은 이동 패턴의 규칙성을 의미하며, 이동 규칙성의 수치가 높을수록 일정한 패턴에 따라 이동함을 나타낸다.

[49] 이때 이동 규칙성(CM)은 아래의 수학적 식 1을 통해 연산될 수 있다.

[50] [수식1]

$$CM = \log(E_{latitude} + E_{longitude})$$

[51]

$$E = \sum_i \frac{psd(f_i)}{i_1 - i_2}$$

[52] 여기서 $E_{latitude}$ 는 위도에 따른 기 설정된 주파수 대역에서의 전력 값의 평균을 의미하고, $E_{longitude}$ 는 경도에 따른 기 설정된 주파수 대역에서의 전력 값의 평균을 의미하고, psd는 기 설정된 주파수 대역(f_i)에서 전력 값의 합을 의미하고, i_1 은 기 설정된 주파수 대역에서 최대 주파수를 의미하고, i_2 는 기 설정된 주파수 대역에서 최소 주파수를 의미한다.

[53] 그리고 위치 엔트로피는 위치 클러스터에서의 시간 변동성을 의미한다. 여기서 위치 클러스터는 GPS 신호의 군집을 의미하며 특정 장소를 기준으로 형성될 수 있다. 예를 들어 집, 학교, 쇼핑몰, 관광 명소 등이 각각 하나의 클러스터로 설정될 수 있다.

[54] 위치 엔트로피가 높다는 것은 위치 클러스터에서 보낸 시간들이 유사하다는 것을 의미하며, 사용자가 어느 특정 장소에서 주로 시간을 보내는지 불확실하다는 것을 뜻한다.

[55] 이때 위치 엔트로피는 아래의 수학적 식 2를 통해 연산될 수 있다.

[56] [수식2]

$$Entropy = - \sum_i p_i \log p_i$$

[57] 여기서 Entropy는 위치 엔트로피를 의미하고, p는 하루(24시간)동안 특정 장소(i)에서 사용자가 보낸 시간 비율을 의미한다.

[58] 다음으로 위치 변화량은 위도 및 경도의 변화량을 의미하며, 위치 변화량이 클수록 사용자의 움직임이 많음을 의미한다.

[59] 이때 위치 변화량은 아래의 수학적 식 3을 통해 연산될 수 있다.

[60] [수식3]

$$\text{Location variance} = \log(\sigma_{\text{latitude}}^2 + \sigma_{\text{longitude}}^2)$$

[61] 여기서 Location variance는 위치 변화량을 의미하고, σ_{latitude} 는 위도의 표준편차를 의미한다. $\sigma_{\text{longitude}}$ 는 경도의 표준편차를 의미한다.

[62] 다음으로 위치 클러스터의 수(number of clusters)는 분류된 위치 클러스터의 수를 의미한다. 구체적으로 K-means 알고리즘을 이용하여 GPS 신호들을 복수의 위치 클러스터로 분류할 수 있으며, 위치 클러스터는 분류된 위치 클러스터의 개수를 나타낸다.

[63] 그리고 총 이동거리는 하루동안 이동한 총 거리를 의미하며, km단위로 나타낼 수 있다.

[64] 그리고 총 이동시간은 하루동안 이동에 소비한 총 시간을 의미한다.

[65] 다음으로 음성 신호 정보에 대한 변수는 사용자 주변의 청각 요소와 같은 환경적 요인을 반영한 변수를 의미하며, 음성 신호 정보는 음성 녹음 기기가 구비된 웨어러블 장치나 스마트폰과 같은 사용자 단말을 통해 측정될 수 있다.

[66] 구체적으로 음성 신호 정보에 대한 변수는 주변 소음 레벨(noise amplitude, NA)을 포함한다.

[67] 이때 주변 소음 레벨(NA)은 수학적 식 4와 같이 소음 크기의 분당 평균값을 의미하며, 음성 신호를 이용하여 산출될 수 있다.

[68] [수식4]

$$NA = \frac{\sum_{t=1}^{60} \text{Amplitude}_t}{60}$$

[69] 여기서 Amplitude_t 는 시간 t에서 소음의 크기를 의미한다.

[70] 이하에서는 도 2 내지 도 4를 통해 본 발명의 실시예에 따른 라이프로그 데이터 간 인과 관계 분석 장치를 이용한 인과 관계 분석 방법에 대해 살펴보도록 한다.

[71] 도 2는 본 발명의 실시예에 따른 인과 관계 분석 방법의 순서도이다.

[72] 도 2에 나타난 바와 같이, 입력부(110)는 생체 신호 정보, GPS 신호 정보 및 음성 신호 정보를 포함하는 라이프로그 데이터를 입력받는다(S210).

[73] 그러면 산출부(120)는 라이프로그 데이터의 변수 중 두 개 이상의 변수를 선택하여 변수 조합을 생성하고(S220), 변수 조합에 대한 경로 모델(path model)을 생성한다(S230).

[74] 예를 들어 산출부(120)는 두 개의 변수를 포함하는 변수 조합을 생성할 수 있으며, 이 경우 일차 방정식의 형태로 변수 조합에 대한 경로 모델이 생성될 수 있다.

[75] 본 발명의 실시예에 따른 13개의 변수를 이용할 경우 최대 78개(${}_{13}C_2$)의 변수 조합이 생성될 수 있으며, 156개의 경로 모델이 생성될 수 있다. 경로 모델이

변수 조합의 2배인 이유는 두 변수의 인과 관계가 바뀔 수 있어 하나의 변수 조합 당 두 개의 경로 모델이 생성되기 때문이다.

[76] 한편 두 개 이상의 변수는 산출부(120)에 의해 무작위로 선택될 수 있으나, 사용자에게 의해 선택될 수도 있다.

[77] 그리고 산출부(120)는 라이프로그 데이터 및 경로 모델을 이용하여 변수 조합에 대한 변수간 인과 정보를 산출한다(S240). 여기서 인과 정보는 변수간 인과 관계의 방향성 및 경로 모델의 경로 계수(path coefficient) 중 적어도 하나를 포함한다.

[78] 아래에서는 도 3을 통해 S240 단계를 구체적으로 살펴보도록 한다. 도 3은 도 2의 S240 단계를 상세하게 나타낸 순서도이다.

[79] 우선 산출부(120)는 변수 조합에 대한 변수 간 공분산 값을 산출한다(S241).

[80] 예를 들어 변수 조합이 두 개의 변수를 포함하는 경우, 산출부(120)는 아래의 수학적식 5를 이용하여 공분산 값(cov(x,y))을 연산할 수 있다.

[81] [수식5]

$$cov(x,y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

[82] 여기서 x 는 제1 변수의 값을 의미하고, y 는 제2 변수의 값을 의미하고,

\bar{x}

는 제1 변수의 평균값을 의미하고,

\bar{y}

는 제2 변수의 평균값을 의미한다.

[83] 그러면 산출부(120)는 변수 조합을 구성하는 변수 사이의 인과 경로에 따른 복수의 경로 모델을 생성한다(S242).

[84] 구체적으로 산출부(120)는 제1 변수(x)가 원인이 되고 제2 변수(y)가 결과가 되는 $y=ax+b$ 와 제2 변수(y)가 원인이 되고 제1 변수(x)가 결과가 되는 $x=ay+b$ 의 경로 모델을 생성한다.

[85] 그리고 산출부(120)는 산출된 공분산 값을 이용하여 복수의 경로 모델에 대한 경로 계수 및 표준 오차를 산출한다(S243).

[86] 구체적으로 산출부(120)는 아래의 수학적식 6을 이용하여 경로 계수를 산출할 수 있다.

[87] [수식6]

$$r = \frac{(n-1)cov(x,y)}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

[88]

[89] 그리고 산출부(120)는 아래의 수학적식 7을 이용하여 경로 모델에 대한 표준

오차를 연산할 수 있다.

[90] [수식7]

$$SE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n-2)}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

[91] 여기서

$$\hat{y} = ax + b$$

이고,

$$a = r \times \sqrt{\frac{(y - \bar{y})^2}{(x - \bar{x})^2}}$$

이고,

$$b = \frac{\sum_{i=1}^n y_i - a \times \sum_{i=1}^n x_i}{n}$$

이고, i 는 라이프로그 데이터의 인덱스 값을 의미한다.

[92] 그러면 산출부(120)는 복수의 경로 모델에 대한 표준 오차 중 가장 작은 값을 가지는 표준 오차에 대응하는 경로 모델을 선택하여 변수간의 인과 관계의 방향성을 판단한다(S244).

[93] 즉, $y=ax+b$ 와 $x=ay+b$ 두 개의 경로 모델 중 표준 오차가 작은 경로 모델을 선택한다. 만약 $y=ax+b$ 의 경로 모델이 선택된 경우 변수 x 가 원인이 되고 변수 y 가 결과가 되므로, 산출부(120)는 변수 조합(x,y)에 대한 인과 관계의 방향성을 변수 x 가 y 에 영향을 미친다고 판단한다.

[94] 예를 들어 선택된 변수 조합이 (BPM, LF)인 경우, BPM이 원인이 되고 LF가 결과가 되는 제1 경로 모델과 LF가 원인이 되고 BPM이 결과가 되는 제2 경로 모델이 생성된다. 이때 제1 경로 모델의 표준 오차가 제2 경로 모델의 표준 오차보다 작으면, 산출부(120)는 제1 경로 모델을 선택하고 변수 조합 (BPM, LF)가 BPM이 원인이 되고 LF가 결과가 되는 인과 관계를 가진다고 판단한다.

[95] 그리고 산출부(120)는 표준 오차를 이용하여 선택된 경로 모델에 대한 유의 지표를 산출한다(S245). 여기서 유의 지표는 경로 모델의 경로 계수가 통계적으로 얼마나 유의미한지를 판단하는 지표를 의미한다.

[96] 이때 산출부(120)는 아래의 수학적 식 8을 이용하여 선택된 경로 모델에 대한 유의 지표(t)를 산출한다.

[97] [수식8]

$$t = \frac{a}{SE}$$

[98]

- [99] 다음으로 판단부(130)는 기 설정된 임계값을 이용하여 산출된 변수간 인과 정보가 통계적 유의성을 가지는지 여부를 판단한다(S250).
- [100] 구체적으로 판단부(130)는 유의 지표가 기 설정된 임계값보다 크면 선택된 경로 모델에 대응하는 변수간 인과 정보가 통계적 유의성을 가진다고 판단한다. 이때 기 설정된 임계값은 기 저장된 룩업 테이블(lookup table)과 기 설정된 확률 값(α), 그리고 경로 모델의 자유도를 이용하여 설정될 수 있다. 이때 경로 모델의 자유도는 경로 모델 생성에 이용된 데이터의 개수에서 특정 수를 뺀 값을 의미한다.
- [101] 표 2는 룩업 테이블의 일 실시예를 나타낸다.

[102] [표2]

$\alpha \backslash v$	0.4	0.25	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.320	318.310	636.620
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.850	2.306	2.898	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.168	3.581	4.144	4.587
11	0.260	0.697	1.363	1.795	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.751	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.705	2.056	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.482	2.756	3.038	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

[103]

[104]

예를 들어 선택된 기 설정된 확률 값(α)이 0.05이고, 이용된 데이터의 개수가 10개이며 자유도 계산에 이용되는 상기 특정 수가 2라고 가정한다. 그러면 판단부(130)는 확률 값(α)가 0.05이고 자유도(v)가 8에 해당하는 값을 표 2의 록업 테이블로부터 선택하여 임계값(1.860)으로 설정한다.

- [105] 그러면 제어부(140)는 통계적 유의성을 가진다고 판단된 변수간 인과 정보를 화면상에 출력한다(S260).
- [106] 도 4는 본 발명의 실시예에 따른 인과 관계 정보를 화면에 출력하는 방법을 설명하기 위한 도면이다.
- [107] 도 4에 도시된 바와 같이, 제어부(140)는 변수를 노드(node) 형태로 형성하고 인과 관계를 가지는 변수끼리는 화살표 형상의 엣지(edge)로 연결시킨 그래프로 표시한다.
- [108] 이때 노드는 라이프로그 데이터의 변수에 따라 서로 다른 색상으로 표시된다. 예를 들어 도 4에 도시된 바와 같이, 제어부(140)는 생체 신호 정보에 관한 변수의 노드는 빨강, GPS 신호 정보에 관한 변수의 노드는 초록, 환경 신호 정보에 관한 변수의 노드는 파랑으로 표시할 수 있다.
- [109] 그리고 엣지는 경로 계수가 클수록 굵게 표시된다. 예를 들어 BPM과 HF 사이의 경로 계수가 BPM과 VLF 사이의 경로 계수보다 크다고 가정한다. 그러면 도 4에 도시된 바와 같이, 제어부(140)는 BPM에서 HF로 향하는 엣지를 BPM에서 VLF로 향하는 엣지보다 굵게 표시한다.
- [110] 본 발명의 실시예에 따르면, 다양한 라이프로그 데이터 간 인과 관계를 분석하여 제공하므로 사람과 환경 사이의 인과 관계를 용이하게 판단할 수 있다. 또한 사람이 환경에 미치는 영향력이나 환경이 사람에게 미치는 영향력을 정확히 분석할 수 있으므로, 환경 개선이나 인력 개발과 같은 다양한 업무 개선에 이용되는 유용한 정보를 제공할 수 있다.
- [111] 본 발명은 도면에 도시된 실시예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 다른 실시예가 가능하다는 점을 이해할 것이다. 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 특허청구범위의 기술적 사상에 의하여 정해져야 할 것이다.

청구범위

- [청구항 1] 라이프로그 데이터 간 인과 관계 분석 장치를 이용한 인과 관계 분석 방법에 있어서,
 생체 신호 정보, GPS 신호 정보 및 음성 신호 정보를 포함하는 라이프로그 데이터를 입력받는 단계,
 상기 라이프로그 데이터의 변수 중 두 개 이상의 변수를 선택하여 변수 조합 및 상기 변수 조합에 대한 경로 모델(path model)을 생성하고, 상기 라이프로그 데이터 및 상기 경로 모델을 이용하여 상기 변수 조합 각각에 대한 변수간 인과 정보를 산출하는 단계,
 기 설정된 임계값을 이용하여 상기 산출된 변수간 인과 정보가 통계적 유의성을 가지는지 여부를 판단하는 단계, 그리고
 상기 통계적 유의성을 가진다고 판단된 변수간 인과 정보를 화면상에 출력하는 단계를 포함하는 인과 관계 분석 방법.
- [청구항 2] 제1항에 있어서,
 상기 라이프로그 데이터의 변수는,
 상기 생체 신호 정보에 대한 변수, 상기 GPS 신호 정보에 대한 변수 및 상기 음성 신호 정보에 대한 변수 중에서 적어도 하나를 포함하는 인과 관계 분석 방법.
- [청구항 3] 제2항에 있어서,
 상기 생체 신호 정보에 대한 변수는 BPM(Beat Per Minutes), HF(High Frequency), LF(Low Frequency) 및 VLF(Very Low Frequency) 중에서 적어도 하나를 포함하고,
 상기 GPS 신호 정보에 대한 변수는 이동 패턴의 불규칙성을 의미하는 이동 규칙성(circadian movement), 위치 클러스터에서의 시간 변동성을 의미하는 위치 엔트로피(entropy), 위도 및 경도의 변화량을 의미하는 위치 변화량(location variance), 위치 클러스터의 수(number of clusters), 총 이동 거리(total distance) 및 총 이동 시간(transition time) 중에서 적어도 하나를 포함하며,
 상기 음성 신호 정보에 대한 변수는 소음 크기의 분당 평균값을 의미하는 주변 소음 레벨(noise amplitude)을 포함하는 인과 관계 분석 방법.
- [청구항 4] 제1항에 있어서,
 상기 변수간 인과 정보는,
 상기 변수간 인과 관계의 방향성 및 상기 경로 모델($\hat{y}=ax+b$)의 경로 계수(path coefficient) 중 적어도 하나를 포함하는 인과 관계 분석 방법.
- [청구항 5] 제4항에 있어서,
 상기 변수간 인과 정보를 산출하는 단계는,
 상기 변수 조합에 대한 변수 간 공분산 값을 산출하는 단계,

상기 변수 조합을 구성하는 변수 사이의 인과 경로에 따른 복수의 경로 모델을 생성하는 단계,
 상기 산출된 공분산 값을 이용하여 상기 복수의 경로 모델에 대한 경로 계수 및 표준 오차를 산출하는 단계,
 상기 복수의 경로 모델에 대한 표준 오차 중 가장 작은 값을 가지는 표준 오차에 대응하는 경로 모델을 선택하여 상기 변수간 인과 관계의 방향성을 판단하는 단계, 그리고
 상기 표준 오차를 이용하여 상기 선택된 경로 모델에 대한 유의 지표를 산출하는 단계를 포함하는 인과 관계 분석 방법.

[청구항 6]

제5항에 있어서,
 상기 경로 계수(r) 및 표준 오차(SE)는 아래의 수학적식을 이용하여 산출되는 인과 관계 분석 방법:

$$r = \frac{(n-1)cov(x,y)}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

$$SE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n-2)}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

여기서, $\hat{y} = ax + b$ 이고, $a = r \times \sqrt{\frac{(y - \bar{y})^2}{(x - \bar{x})^2}}$ 이고, $b = \frac{\sum_{i=1}^n y_i - a \times \sum_{i=1}^n x_i}{n}$

이고, i 는 상기 라이프로그 데이터의 인덱스 값을 의미하고, x 는 상기 라이프로그 데이터의 제1 변수를 의미하고, y 는 상기 라이프로그 데이터의 제2 변수를 의미하고, \bar{x} 는 상기 제1 변수의 평균값을 의미하고, \bar{y} 는 상기 제2 변수의 평균값을 의미한다.

[청구항 7]

제5항에 있어서,
 상기 산출된 변수간 인과 정보의 유의성 여부를 판단하는 단계는,
 상기 유의 지표($t = \frac{a}{SE}$)가 상기 기 설정된 임계값보다 크면 상기 선택된

경로 모델에 대응하는 변수간 인과 정보가 통계적 유의성을 가진다고 판단하는 인과 관계 분석 방법.

[청구항 8]

제4항에 있어서,
 상기 변수간 인과 정보를 화면상에 출력하는 단계는,
 상기 변수는 노드(node) 형태로 형성되고, 인과 관계를 가지는 변수끼리는 화살표 형상의 엣지(edge)로 연결시킨 그래프로 표시하며,
 상기 노드는 상기 라이프로그 데이터의 변수에 따라 서로 다른 색상으로

표시되고, 상기 옛지는 상기 경로 계수가 클수록 굵게 표시되는 인과 관계 분석 방법.

- [청구항 9] 생체 신호 정보, GPS 신호 정보 및 음성 신호 정보를 포함하는 라이프로그 데이터를 입력받는 입력부,
상기 라이프로그 데이터의 변수 중 두 개 이상의 변수를 선택하여 변수 조합 및 상기 변수 조합에 대한 경로 모델(path model)을 생성하고, 상기 라이프로그 데이터 및 상기 경로 모델을 이용하여 상기 변수 조합 각각에 대한 변수간 인과 정보를 산출하는 산출부,
기 설정된 임계값을 이용하여 상기 산출된 변수간 인과 정보가 통계적 유의성을 가지는지 여부를 판단하는 판단부, 그리고
상기 통계적 유의성을 가진다고 판단된 변수간 인과 정보를 화면상에 출력하는 제어부를 포함하는 인과 관계 분석 장치.
- [청구항 10] 제9항에 있어서,
상기 라이프로그 데이터의 변수는,
상기 생체 신호 정보에 대한 변수, 상기 GPS 신호 정보에 대한 변수 및 상기 음성 신호 정보에 대한 변수 중에서 적어도 하나를 포함하는 인과 관계 분석 장치.
- [청구항 11] 제10항에 있어서,
상기 생체 신호 정보에 대한 변수는 BPM(Beat Per Minutes), HF(High Frequency), LF(Low Frequency) 및 VLF(Very Low Frequency) 중에서 적어도 하나를 포함하고,
상기 GPS 신호 정보에 대한 변수는 이동 패턴의 불규칙성을 의미하는 이동 규칙성(circadian movement), 위치 클러스터에서의 시간 변동성을 의미하는 위치 엔트로피(entropy), 위도 및 경도의 변화량을 의미하는 위치 변화량(location variance), 위치 클러스터의 수(number of clusters), 총 이동 거리(total distance) 및 총 이동 시간(transition time) 중에서 적어도 하나를 포함하며,
상기 음성 신호 정보에 대한 변수는 소음 크기의 분당 평균값을 의미하는 주변 소음 레벨(noise amplitude) 중에서 적어도 하나를 포함하는 인과 관계 분석 장치.
- [청구항 12] 제9항에 있어서,
상기 변수간 인과 정보는,
상기 변수간 인과 관계의 방향성 및 상기 경로 모델($\hat{y}=ax+b$)의 경로 계수(path coefficient) 중 적어도 하나를 포함하는 인과 관계 분석 장치.
- [청구항 13] 제12항에 있어서,
상기 산출부는,
상기 변수 조합에 대한 변수 간 공분산 값을 산출하고, 상기 변수 조합을

구성하는 변수 사이의 인과 경로에 따른 복수의 경로 모델을 생성하며, 상기 산출된 공분산 값을 이용하여 상기 복수의 경로 모델에 대한 경로 계수 및 표준 오차를 산출하고, 상기 복수의 경로 모델에 대한 표준 오차 중 가장 작은 값을 가지는 표준 오차에 대응하는 경로 모델을 선택하여 상기 변수간 인과 관계의 방향성을 판단하며, 상기 표준 오차를 이용하여 상기 선택된 경로 모델에 대한 유의 지표를 산출하는 인과 관계 분석 장치.

[청구항 14]

제13항에 있어서,

상기 경로 계수(r) 및 표준 오차(SE)는 아래의 수학적식을 이용하여 산출되는 인과 관계 분석 장치:

$$r = \frac{(n-1)cov(x,y)}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

$$SE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n-2)}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

여기서, $\hat{y} = ax + b$ 이고, $a = r \times \sqrt{\frac{(y - \bar{y})^2}{(x - \bar{x})^2}}$ 이고, $b = \frac{\sum_{i=1}^n y_i - a \times \sum_{i=1}^n x_i}{n}$

이고, i 는 상기 라이프로그 데이터의 인덱스 값을 의미하고, x 는 상기 라이프로그 데이터의 제1 변수를 의미하고, y 는 상기 라이프로그 데이터의 제2 변수를 의미하고, \bar{x} 는 상기 제1 변수의 평균값을 의미하고, \bar{y} 는 상기 제2 변수의 평균값을 의미한다.

[청구항 15]

제13항에 있어서,

상기 판단부는,

상기 유의 지표($t = \frac{a}{SE}$)가 상기 기 설정된 임계값보다 크면 상기 선택된

경로 모델에 대응하는 변수간 인과 정보가 통계적 유의성을 가진다고 판단하는 인과 관계 분석 장치.

[청구항 16]

제12항에 있어서,

상기 제어부는,

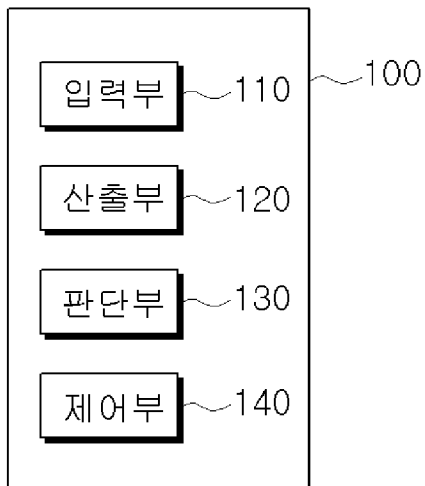
상기 변수는 노드(node) 형태로 형성되고, 인과 관계를 가지는 변수끼리는 화살표 형상의 엣지(edge)로 연결시킨 그래프로 표시하며, 상기 노드는 상기 라이프로그 데이터의 변수에 따라 서로 다른 색상으로 표시되고, 상기 엣지는 상기 경로 계수가 클수록 굵게 표시되는 인과 관계 분석 장치.

요약서

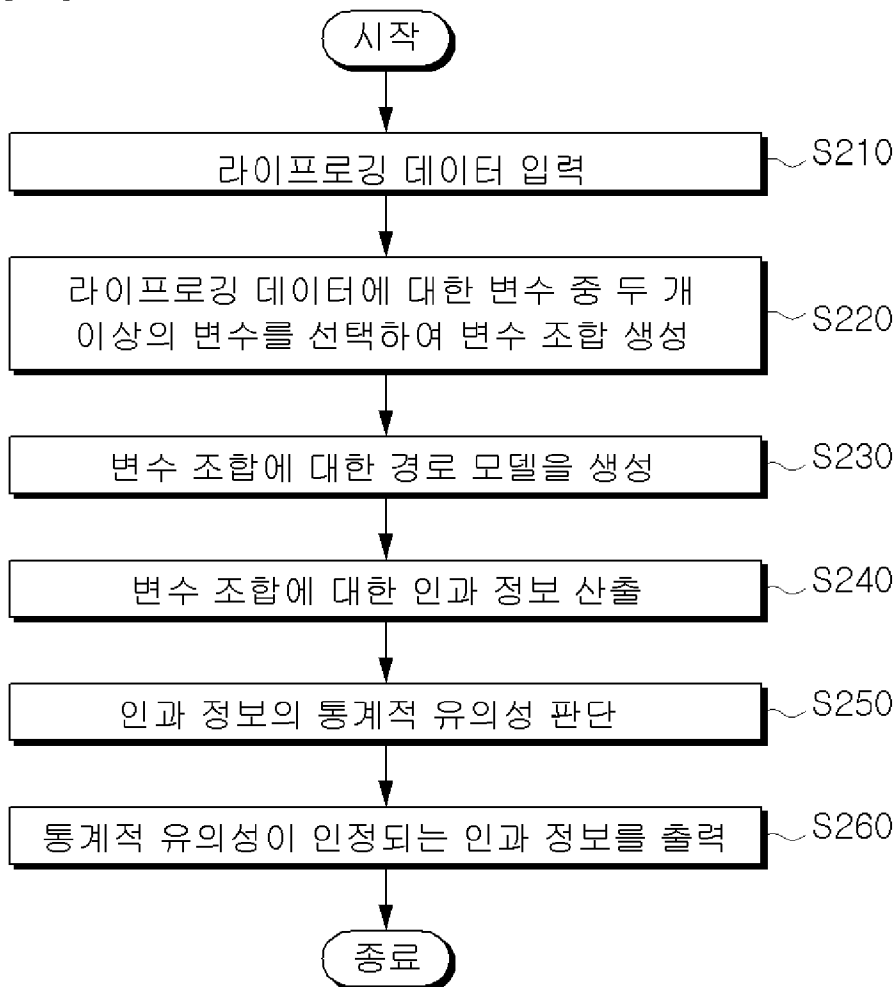
본 발명은 라이프로그 데이터 간 인과 관계 분석 장치 및 그 방법에 관한 것이다. 본 발명에 따르면, 라이프로그 데이터 간 인과 관계 분석 장치를 이용한 인과 관계 분석 방법에 있어서, 인과 관계 분석 방법은 생체 신호 정보, GPS 신호 정보 및 음성 신호 정보를 포함하는 라이프로그 데이터를 입력받는 단계, 상기 라이프로그 데이터의 변수 중 두 개 이상의 변수를 선택하여 변수 조합 및 상기 변수 조합에 대한 경로 모델(path model)을 생성하고, 상기 라이프로그 데이터 및 상기 경로 모델을 이용하여 상기 변수 조합 각각에 대한 변수간 인과 정보를 산출하는 단계, 기 설정된 임계값을 이용하여 상기 산출된 변수간 인과 정보가 통계적 유의성을 가지는지 여부를 판단하는 단계, 그리고 상기 통계적 유의성을 가진다고 판단된 변수간 인과 정보를 화면상에 출력하는 단계를 포함한다.

이와 같이 본 발명에 따르면, 다양한 라이프로그 데이터 간 인과 관계를 분석하여 제공하므로 사람과 환경 사이의 상관 관계를 용이하게 판단할 수 있다. 또한 사람이 환경에 미치는 영향력이나 환경이 사람에게 미치는 영향력을 정확히 분석할 수 있으므로, 환경 개선이나 인력 개발과 같은 다양한 업무 개선에 이용되는 유용한 정보를 제공할 수 있다.

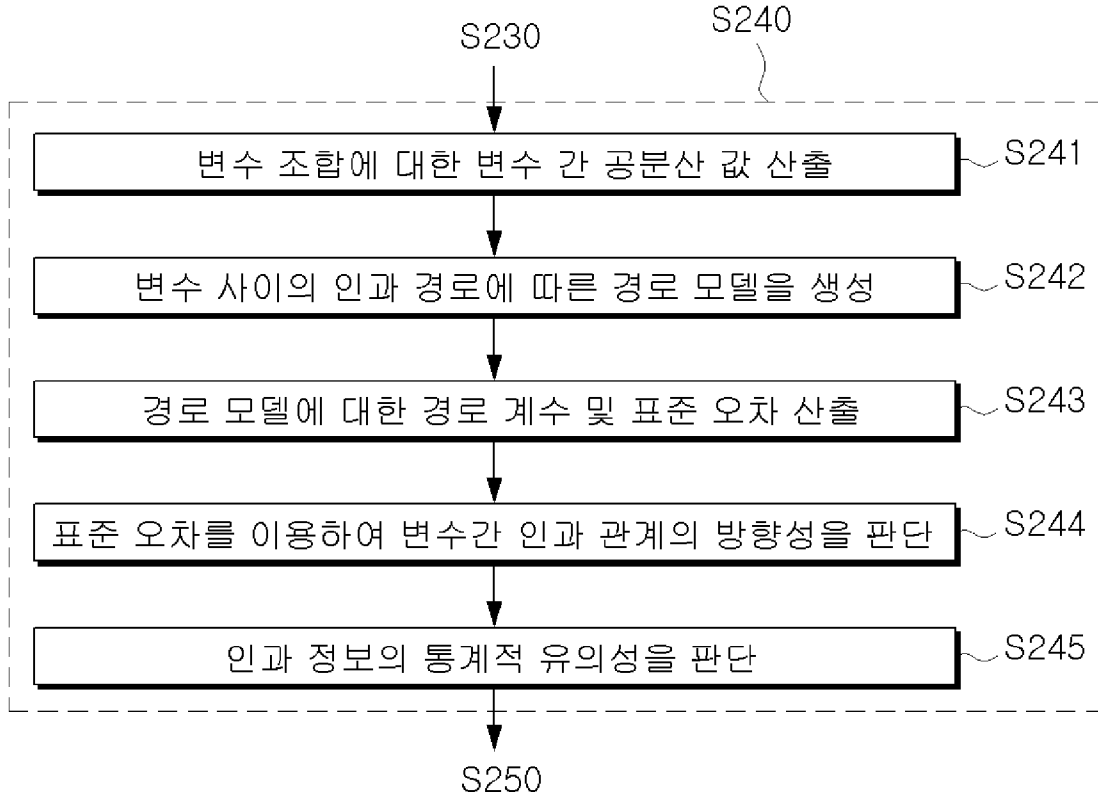
[도1]



[도2]



[도3]



[도4]

