

	(19) 대한민국특허청(KR) (12) 공개특허공보(A)	(11) 공개번호 10-2014-0090483 (43) 공개일자 2014년07월17일
(51) 국제특허분류(Int. Cl.) G06Q 50/24 (2012.01)	(71) 출원인 경희대학교 산학협력단 경기도 용인시 기흥구 덕영대로 1732, 국제캠퍼스 내 (서천동, 경희대학교)	
(21) 출원번호 10-2013-0002601		
(22) 출원일자 2013년01월09일	(72) 발명자	
심사청구일자 2013년01월09일	이영구 경기 용인시 기흥구 덕영대로 1732, 전자정보관 314호 (서천동, 경희대학교)	
	팜더안 경기 용인시 기흥구 덕영대로 1732, 전자정보대학 351호 (서천동, 경희대학교)	
	홍지혜 경기 용인시 기흥구 덕영대로 1732, 전자정보관 351호 (서천동, 경희대학교)	
	(74) 대리인 서재승	

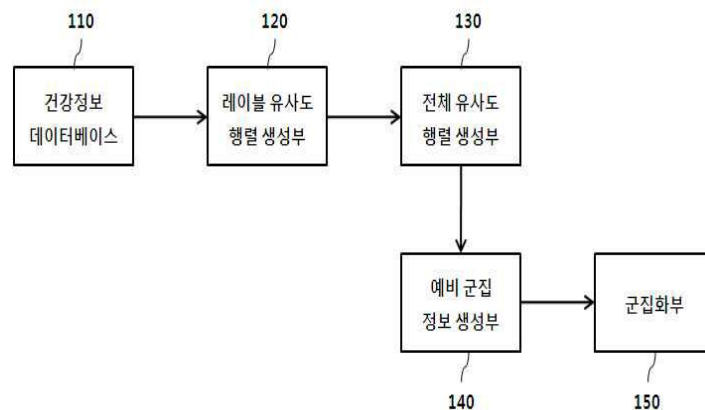
전체 청구항 수 : 총 16 항

(54) 발명의 명칭 건강정보의 군집화 방법

(57) 요약

본 발명은 정보 군집화 방법에 관한 것으로, 보다 구체적으로 건강정보에 특화되어 다수 환자의 환자별 치료 레이블이 있는 건강정보와 치료 레이블이 없는 건강정보로 이루어진 건강정보 데이터베이스를 활용하여 서로 유사한 건강정보를 군집화함으로써, 치료 레이블이 거의 존재하지 않는 통상의 건강정보 데이터베이스로부터 훈련 데이터를 생성할 수 있는 건강정보 군집화 방법에 관한 것이다.

대 표 도 - 도1



특허청구의 범위

청구항 1

- (a) 다수 환자의 환자별 치료 레이블이 있는 건강정보와 치료 레이블이 없는 건강정보로 이루어진 건강정보 데이터베이스에서 치료 레이블 정보를 포함하여 레이블 정보 종류에 따라 환자 사이의 각 레이블 정보별 유사도를 나타내는 레이블 유사도 행렬을 생성하는 단계;
- (b) 상기 각 레이블 정보별 유사도를 나타내는 레이블 유사도 행렬의 합으로부터 환자 사이의 전체 유사도 행렬을 생성하는 단계;
- (c) 상기 전체 유사도 행렬에서 서로 유사한 속성을 가지는 환자 사이의 군집 밀집도가 임계값보다 작도록 밀집시켜 상기 다수 환자의 예비 군집 정보를 생성하는 단계; 및
- (d) 상기 예비 군집 정보의 분포 특성에 기초하여 설정된 수의 군집으로 다수 환자를 군집화하는 단계를 포함하는 건강정보의 군집화 방법.

청구항 2

제 1 항에 있어서, 상기 환자 사이의 각 레이블 정보별 유사도는 환자별 각 레이블 정보의 유클리드 거리(euclidean distance)로 계산되는 것을 특징으로 하는 건강정보의 군집화 방법.

청구항 3

제 2 항에 있어서, 상기 치료 레이블 정보의 유사도는 유클리드 거리가 0을 초과하는 경우 1로 설정되는 것을 특징으로 하는 건강정보의 군집화 방법.

청구항 4

제 1 항에 있어서, 상기 레이블 정보의 종류는

감정 레이블 정보, 수치 레이블 정보, 징후 레이블 정보, 치료 레이블 정보 중 적어도 어느 하나를 포함하는 것을 특징으로 하는 건강정보의 군집화 방법.

청구항 5

제 4 항에 있어서,

상기 감정 레이블 정보, 수치 레이블 정보, 징후 레이블 정보, 치료 레이블 정보에 대한 레이블 유사도 행렬을 구성하는 각 원소의 값은 서로 동일한 제1 기준값으로 정규화되는 것을 특징으로 하는 건강정보의 군집화 방법.

청구항 6

제 5 항에 있어서, 상기 감정 레이블 정보, 수치 레이블 정보, 징후 레이블 정보, 치료 레이블 정보는 각 레이블 정보를 구성하는 항목별 식별자에 매칭되는 식별값으로 저장되어 있는 것을 특징으로 하는 건강정보의 군집화 방법.

청구항 7

제 1 항 내지 제 6 항 중 어느 한 항에 있어서, 상기 예비 군집 정보를 생성하는 단계는

- (c1) 제2 기준값으로 정규화된 초기 벡터를 생성하는 단계;
- (c2) 상기 초기 벡터를 이전 벡터로 설정하고 상기 전체 유사도 행렬과 상기 이전 벡터의 곱으로부터 다음 벡터를 생성하는 단계;
- (c3) 상기 이전 벡터와 상기 다음 벡터의 차가 임계값보다 작은지 판단하는 단계; 및
- (c4) 상기 이전 벡터와 상기 다음 벡터의 차가 임계값보다 작은 경우 상기 다음 벡터를 상기 다수 환자의 예비

군집 정보로 생성하는 단계를 포함하며,

상기 이전 벡터와 상기 다음 벡터의 차가 임계값보다 큰 경우 상기 다음 벡터를 이전 벡터로 설정하여 상기 (c2) 내지 (c3)를 반복 수행하는 것을 특징으로 하는 건강정보의 군집화 방법.

청구항 8

제 7 항에 있어서, 상기 예비 군집 정보는 K-means 군집화 알고리즘을 통해 상기 예비 군집 정보의 분포 밀집도 특성에 따라 설정된 수의 군집으로 군집화되는 것을 특징으로 하는 건강정보의 군집화 방법.

청구항 9

(a) 다수 환자의 환자별 치료 레이블이 있는 건강정보와 치료 레이블이 없는 건강정보로 이루어진 건강정보 데이터베이스에서 치료 레이블 정보를 포함하여 레이블 정보 종류에 따라 환자 사이의 각 레이블 정보별 유사도를 나타내는 레이블 유사도 행렬을 생성하는 단계;

(b) 상기 각 레이블 정보별 유사도를 나타내는 레이블 유사도 행렬의 합으로부터 환자 사이의 전체 유사도 행렬을 생성하는 단계;

(c) 상기 전체 유사도 행렬에서 서로 유사한 속성을 가지는 환자 사이의 건강정보를 밀집시켜 상기 다수의 환자의 예비 군집 정보를 생성하고 상기 예비 군집 정보의 분포 밀집도 특성에 기초하여 설정된 수로 상기 다수 환자를 군집화하는 단계;

(d) 상기 군집화된 환자의 레이블 정보 종류별 건강정보의 중심값으로 각 군집별 군집 정보를 생성하는 단계; 및

(e) 신규 환자의 건강정보와 상기 군집 정보의 유사도를 판단하여 상기 신규 환자와 가장 유사도가 높은 군집 정보의 치료 레이블 정보를 상기 신규 환자에 대한 치료 레이블 정보로 제공하는 단계를 포함하는 치료 조언 방법.

청구항 10

제 9 항에 있어서, 상기 예비 군집 정보를 생성하는 단계는

(c1) 기준값으로 정규화된 초기 벡터를 생성하는 단계;

(c2) 상기 초기 벡터를 이전 벡터로 설정하고 상기 전체 유사도 행렬과 상기 이전 벡터의 곱으로부터 다음 벡터를 생성하는 단계;

(c3) 상기 이전 벡터와 상기 다음 벡터의 차가 임계값보다 작은지 판단하는 단계; 및

(c4) 상기 이전 벡터와 상기 다음 벡터의 차가 임계값보다 작은 경우 상기 다음 벡터에 기초하여 상기 다수 환자의 예비 군집 데이터를 생성하는 단계를 포함하며,

상기 이전 벡터와 상기 다음 벡터의 차가 임계값보다 큰 경우 상기 다음 벡터를 이전 벡터로 설정하여 상기 (c2) 내지 (c3)를 반복 수행하는 것을 특징으로 하는 건강정보의 군집화 방법.

청구항 11

제 10 항에 있어서, 상기 레이블 정보 종류는

감정 레이블 정보, 수치 레이블 정보, 징후 레이블 정보, 치료 레이블 정보 중 적어도 어느 하나를 포함하는 것을 특징으로 하는 치료 조언 방법.

청구항 12

다수 환자의 환자별 치료 레이블이 있는 건강정보와 치료 레이블이 없는 건강정보로 이루어진 건강정보 데이터베이스에서 환자 사이의 각 레이블 정보별 유사도를 나타내는 레이블 유사도 행렬을 생성하는 레이블 유사도 행렬 생성부;

상기 각 레이블 정보별 유사도를 나타내는 레이블 유사도 행렬의 합으로부터 환자 사이의 전체 유사도 행렬을 생성하는 전체 유사도 행렬 생성부;

상기 전체 유사도 행렬에서 서로 유사한 속성을 가지는 환자 사이의 군집 밀집도가 임계값보다 작도록 밀집시켜 상기 다수 환자의 예비 군집 정보를 생성하는 예비 군집 정보 생성부; 및

상기 예비 군집 정보의 분포 밀집도 특성에 기초하여 설정된 수의 군집으로 다수 환자를 군집화하는 군집화부를 포함하는 것을 특징으로 하는 건강정보의 군집화 장치.

청구항 13

제 12 항에 있어서, 상기 레이블 유사도 행렬 생성부는

상기 환자 사이의 감정 레이블 정보의 유사도를 계산하고 상기 환자 사이의 감정 레이블 유사도 행렬을 생성하는 감정 행렬 생성부;

상기 환자 사이의 수치 레이블 정보의 유사도를 계산하고, 상기 환자 사이의 수치 레이블 유사도 행렬을 생성하는 수치 행렬 생성부;

상기 환자 사이의 징후 레이블 정보의 유사도를 계산하고 상기 환자 사이의 징후 레이블 유사도 행렬을 생성하는 징후 행렬 생성부; 및

상기 환자 사이의 치료 레이블 정보의 유사도를 계산하고 상기 환자 사이의 치료 레이블 유사도 행렬을 생성하는 치료 행렬 생성부를 포함하는 것을 특징으로 하는 건강정보 군집화 장치.

청구항 14

제 13 항에 있어서, 상기 레이블 유사도 행렬 생성부는

상기 감정 레이블 행렬, 수치 레이블 행렬, 징후 레이블 행렬, 치료 레이블 행렬을 동일한 제1 기준값으로 정규화하는 행렬 정규화부를 더 포함하는 것을 특징으로 하는 건강정보 군집화 장치.

청구항 15

제 12 항에 있어서, 상기 예비 군집 정보 생성부는

제2 기준값으로 정규화된 초기 벡터를 생성하는 초기 벡터 생성부;

상기 초기 벡터를 이전 벡터로 설정하고 상기 이전 벡터와 상기 전체 유사도 행렬을 곱하여 다음 벡터를 생성하는 다음 벡터 생성부;

상기 이전 벡터와 상기 다음 벡터의 차가 임계값보다 작은지 판단하며, 판단 결과에 기초하여 이전 벡터와 상기 다음 벡터의 차가 임계값보다 큰 경우 상기 다음 벡터를 이전벡터로 설정하여 다음 벡터를 반복 생성하도록 제어하는 밀집부; 및

상기 밀집부의 판단 결과에 기초하여 상기 이전 벡터와 상기 다음 벡터의 차가 임계값보다 작은 경우, 상기 다음 벡터를 상기 다수 환자의 예비 군집 정보로 생성하는 예비 군집 결정부를 포함하는 것을 특징으로 하는 건강정보 군집화 장치.

청구항 16

제 15 항에 있어서, 상기 군집화부는

상기 예비 군집 정보를 K-means 군집화 알고리즘을 통해 상기 예비 군집 정보의 분포 밀집도 특성에 따라 설정된 수로 군집화하는 것을 특징으로 하는 건강정보 군집화 장치.

명세서

기술분야

[0001]

본 발명은 정보 군집화 방법에 관한 것으로, 보다 구체적으로 건강정보에 특화되어 다수 환자의 환자별 치료 레이블이 있는 건강정보와 치료 레이블이 없는 건강정보로 이루어진 건강정보 데이터베이스를 활용하여 서로 유사한 건강정보를 군집화함으로써, 치료 레이블이 거의 존재하지 않는 통상의 건강정보 데이터베이스로부터 훈련 데이터를 생성할 수 있는 건강정보 군집화 방법에 관한 것이다.

배경 기술

- [0002] 유비쿼터스 환경에서 사용자의 건강정보는 다양한 종류의 디바이스를 통해 수집되며 수집된 정보는 통신 네트워크를 통해 건강관리 서버의 데이터베이스에 저장된다. 사용자는 요청에 의해 또는 설정된 주기에 따라 사용자 건강정보에 대한 치료 정보를 의사, 간호사 등의 전문 자문사로부터 획득하여 사용자의 건강을 관리하게 된다.
- [0003] 이렇게 통신 네트워크를 통해 수집된 사용자 건강 정보는 건강관리 서버의 데이터베이스에 저장되는데, 다수의 건강관리 서버가 운영되는 경우 다수의 사용자 건강정보는 다수의 건강관리 서버에 무작위로 분포되어 저장되어 있으며 다수의 건강관리 서버의 건강정보 데이터베이스에 저장되어 있는 건강 정보에는 사용자의 상태를 나타내는 감정 레이블 정보, 수치 레이블 정보, 징후 레이블 정보만이 저장될 뿐 대부분의 사용자 정보에는 사용자 건강 상태에 대한 치료 레이블 정보는 저장되어 있지 않다. 즉, 전문 자문사와 사용자의 오프라인 상담 또는 전화 상담을 통해 직접 사용자에게 치료 정보를 제공하는 것이 일반적이며, 더욱이 전문 자문사는 자신의 개인적인 치료 정보를 타인에게 공개하지 않는 것이 일반적이다.
- [0004] 따라서 건강 관련 소셜 네트워크에서 사용자 건강 상태에 따른 치료 레이블 정보를 제공하기 위하여 훈련 데이터를 생성하고자 하는 경우, 수집한 사용자의 건강정보는 치료 레이블이 없는 건강정보가 거의 대부분을 차지하며 치료 레이블이 있는 건강정보가 거의 존재하지 않는다. 그러므로 감독기반 학습 기법에 사용될 수 있는 훈련 데이터를 생성하기 곤란하다는 문제점을 가진다.

발명의 내용

해결하려는 과제

- [0005] 본 발명은 위에서 언급한 문제점을 해결하기 위한 것으로, 본 발명이 이루고자 하는 목적은 치료 레이블이 거의 존재하지 않는 건강정보 데이터베이스를 활용하여 건강정보 데이터베이스를 군집화할 수 있는 방법을 제공하는 것이다.
- [0006] 본 발명이 이루고자 하는 다른 목적은 치료 레이블이 거의 존재하지 않는 건강정보 데이터베이스를 활용하여 건강정보 데이터를 군집하고, 환자의 건강 정보와 가장 유사한 군집의 건강정보 중 치료 레이블 정보를 환자에게 제공할 수 있는 치료 조언 방법을 제공하는 것이다.
- [0007] 본 발명이 이루고자 하는 또 다른 목적은 선형 분포를 보장하지 않는 건강정보에 특화되어 건강정보의 밀집도에 기초하여 간단한 프로세스로 건강 정보를 군집화할 수 있는 방법을 제공하는 것이다.

과제의 해결 수단

- [0008] 본 발명의 목적을 달성하기 위하여 본 발명에 따른 건강정보의 군집화 방법은 다수 환자의 환자별 치료 레이블이 있는 건강정보와 치료 레이블이 없는 건강정보로 이루어진 건강정보 데이터베이스에서 치료 레이블 정보를 포함하여 레이블 정보 종류에 따라 환자 사이의 각 레이블 정보별 유사도를 나타내는 레이블 유사도 행렬을 생성하는 단계와, 각 레이블 정보별 유사도를 나타내는 레이블 유사도 행렬의 합으로부터 환자 사이의 전체 유사도 행렬을 생성하는 단계와, 전체 유사도 행렬에서 서로 유사한 속성을 가지는 환자 사이의 군집 밀집도가 임계값보다 작도록 밀집시켜 다수 환자의 예비 군집 정보를 생성하는 단계와, 예비 군집 정보의 분포 특성에 기초하여 설정된 수의 군집으로 다수 환자를 군집화하는 단계를 포함한다.
- [0009] 여기서 환자 사이의 각 레이블 정보별 유사도는 환자별 각 레이블 정보의 유클리드 거리(euclidean distance)로 계산된다. 치료 레이블 정보의 유사도는 유클리드 거리가 0을 초과하는 경우 1로 설정된다.
- [0010] 한편, 레이블 정보의 종류는 감정 레이블 정보, 수치 레이블 정보, 징후 레이블 정보, 치료 레이블 정보 중 적어도 어느 하나를 포함하며, 감정 레이블 정보, 수치 레이블 정보, 징후 레이블 정보, 치료 레이블 정보는 각 레이블을 구성을 항목 식별자에 매칭되는 식별값으로 저장되어 있다.
- [0011] 바람직하게, 감정 레이블 정보, 수치 레이블 정보, 징후 레이블 정보, 치료 레이블 정보에 대한 레이블 유사도 행렬을 구성하는 각 원소의 값은 서로 동일한 제1 기준값으로 정규화된다.
- [0012] 바람직하게, 예비 군집 정보를 생성하는 단계는 제2 기준값으로 정규화된 초기 벡터를 생성하는 단계(c1)와, 초기 벡터를 이전 벡터로 설정하고 전체 유사도 행렬과 이전 벡터의 곱으로부터 다음 벡터를 생성하는 단계(c2)와, 이전 벡터와 다음 벡터의 차가 임계값보다 작은지 판단하는 단계(c3)와, 이전 벡터와 다음 벡터의

차가 임계값보다 작은 경우 다음 벡터에 기초하여 다수 환자의 예비 군집 정보를 생성하는 단계(c4)를 포함하며, 이전 벡터와 다음 벡터의 차가 임계값보다 큰 경우 다음 벡터를 이전 벡터로 설정하여 (c2) 내지 (c3)를 반복 수행하는 것을 특징으로 한다.

[0013] 한편, 예비 군집 정보는 K-means 군집화 알고리즘을 통해 예비 군집 정보의 분포 밀집도 특성에 따라 설정된 수의 군집으로 군집화된다.

[0014] 한편, 본 발명에 따른 치료 조언 방법은 다수 환자의 환자별 치료 레이블이 있는 건강정보와 치료 레이블이 없는 건강정보로 이루어진 건강정보 데이터베이스에서 치료 레이블 정보를 포함하여 레이블 정보 종류에 따라 환자 사이의 각 레이블 정보별 유사도를 나타내는 레이블 유사도 행렬을 생성하는 단계와, 각 레이블 정보별 유사도를 나타내는 레이블 유사도 행렬의 합으로부터 환자 사이의 전체 유사도 행렬을 생성하는 단계와, 전체 유사도 행렬에서 서로 유사한 속성을 가지는 환자 사이의 건강정보를 밀집시켜 다수의 환자의 예비 군집 정보를 생성하고 예비 군집 정보의 분포 밀집도 특성에 기초하여 설정된 수로 다수 환자를 군집화하는 단계와, 군집화된 환자의 레이블 정보 종류별 건강정보의 중심값으로 각 군집별 군집 정보를 생성하는 단계와, 신규 환자의 건강정보와 군집 정보의 유사도를 판단하여 신규 환자와 가장 유사도가 높은 군집 정보의 치료 레이블 정보를 신규 환자에 대한 치료 레이블 정보로 제공하는 단계를 포함한다.

[0015] 한편, 본 발명의 목적을 달성하기 위하여 본 발명에 따른 건강정보의 군집화 장치는 다수 환자의 환자별 치료 레이블이 있는 건강정보와 치료 레이블이 없는 건강정보로 이루어진 건강정보 데이터베이스에서 환자 사이의 각 레이블 정보별 유사도를 나타내는 레이블 유사도 행렬을 생성하는 레이블 유사도 행렬 생성부와, 각 레이블 정보별 유사도를 나타내는 레이블 유사도 행렬의 합으로부터 환자 사이의 전체 유사도 행렬을 생성하는 전체 유사도 행렬 생성부와, 전체 유사도 행렬에서 서로 유사한 속성을 가지는 환자 사이의 군집 밀집도가 임계값보다 작도록 밀집시켜 다수 환자의 예비 군집 정보를 생성하는 예비 군집 정보 생성부와, 예비 군집 정보의 분포 밀집도 특성에 기초하여 설정된 수의 군집으로 다수 환자를 군집화하는 군집화부를 포함한다.

[0016] 여기서 레이블 유사도 행렬 생성부는 환자 사이의 감정 레이블 정보의 유사도를 계산하고 환자 사이의 감정 레이블 유사도 행렬을 생성하는 감정 행렬 생성부와, 환자 사이의 수치 레이블 정보의 유사도를 계산하고 환자 사이의 수치 레이블 유사도 행렬을 생성하는 수치 행렬 생성부와, 환자 사이의 징후 레이블 정보의 유사도를 계산하고 환자 사이의 징후 레이블 유사도 행렬을 생성하는 징후 행렬 생성부와, 환자 사이의 치료 레이블 정보의 유사도를 계산하고 환자 사이의 치료 레이블 유사도 행렬을 생성하는 치료 행렬 생성부를 포함한다.

[0017] 여기서 레이블 유사도 행렬 생성부는 감정 레이블 행렬, 수치 레이블 행렬, 징후 레이블 행렬, 치료 레이블 행렬을 동일한 제1 기준값으로 정규화하는 행렬 정규화부를 더 포함한다.

[0018] 여기서 예비 군집 정보 생성부는 제2 기준값으로 정규화된 초기 벡터를 생성하는 초기 벡터 생성부와, 초기 벡터를 이전 벡터로 설정하고 이전 벡터와 전체 유사도 행렬을 곱하여 다음 벡터를 생성하는 다음 벡터 생성부와, 이전 벡터와 다음 벡터의 차가 임계값보다 작은지 판단하며 판단 결과에 기초하여 이전 벡터와 다음 벡터의 차가 임계값보다 큰 경우 다음 벡터를 이전벡터로 설정하여 다음 벡터를 반복 생성하도록 제어하는 밀집부와, 밀집부의 판단 결과에 기초하여 이전 벡터와 다음 벡터의 차가 임계값보다 작은 경우 다음 벡터에 기초하여 설정된 수로 다수 환자의 예비 군집 정보를 생성하는 예비 군집 생성부를 포함한다.

발명의 효과

[0019] 본 발명에 따른 건강정보의 군집화 방법은 종래 군집화 방법과 비교하여 다음과 같은 다양한 효과를 가진다.

[0020] 첫째, 본 발명에 따른 건강정보의 군집화 방법은 치료 레이블 정보와 치료 레이블 정보 이외의 다른 건강 레이블 정보 각각에 대한 환자 사이의 유사도를 나타내는 레이블 유사도 행렬로부터 전체 유사도 행렬을 생성하고, 전체 유사도 행렬을 이용하여 예비 군집 정보를 생성함으로써, 치료 레이블이 거의 존재하지 않는 건강정보 데이터베이스를 이용하더라도 건강정보를 군집화할 수 있다.

[0021] 둘째, 본 발명에 따른 건강정보의 군집화 방법은 치료 레이블이 거의 존재하지 않는 건강정보 데이터베이스를 활용하여 건강정보 데이터를 군집화함으로써, 환자의 건강 정보와 가장 유사한 군집의 건강정보의 치료 레이블 정보를 환자에게 제공할 수 있다.

[0022] 셋째, 본 발명에 따른 건강정보의 군집화 방법은 선형 분포를 보장하지 않는 건강정보에 특화되어 건강정보의 밀집도에 기초하여 치료 레이블 정보가 거의 존재하지 않는 건강 정보를 간단한 프로세스로 군집화할 수 있다.

도면의 간단한 설명

- [0023] 도 1은 본 발명에 따른 건강정보 군집화 장치를 설명하기 위한 기능 블록도이다.
- 도 2는 본 발명에 따른 레이블 유사도 행렬 생성부를 보다 구체적으로 설명하기 위한 기능 블록도이다.
- 도 3은 본 발명에 따른 예비 군집 정보 생성부를 보다 구체적으로 설명하기 위한 기능 블록도이다.
- 도 4는 본 발명에 따른 건강정보 군집화 장치를 이용하여 생성한 건강정보의 훈련 데이터를 이용하여 새로운 환자에 치료 레이블 정보를 제공하는 치료 조언 시스템을 설명하기 위한 기능 블록도이다.
- 도 5는 본 발명에 따른 건강정보의 군집화 방법을 설명하기 위한 흐름도이다.
- 도 6은 건강정보 데이터베이스에 저장되어 있는 건강 레이블 정보의 일 예를 설명하기 위한 도면이다.
- 도 7은 본 발명에 따른 건강정보의 군집화 방법에서 레이블 유사도 행렬을 생성하는 단계를 보다 구체적으로 설명하기 위한 흐름도이다.
- 도 8은 본 발명에 따른 건강정보 군집화 방법에서 예비 군집 정보를 생성하는 단계를 보다 구체적으로 설명하기 위한 흐름도이다.
- 도 9는 본 발명에 따른 치료 조언 방법을 설명하기 위한 흐름도이다.
- 도 10은 K-means 군집화 알고리즘의 일 예를 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0024] 이하 첨부한 도면을 참고로 본 발명에 따른 건강정보 군집화 방법 및 그 장치에 대해 보다 구체적으로 설명한다.
- [0025] 도 1은 본 발명에 따른 건강정보 군집화 장치를 설명하기 위한 기능 블록도이다.
- [0026] 도 1을 참고로 살펴보면, 레이블 유사도 행렬 생성부(120)는 건강정보 데이터베이스(110)에 저장되어 있는 다수 환자의 건강정보로부터 각 건강 레이블 정보별 다수 환자 사이의 유사도를 나타내는 레이블 유사도 행렬을 생성한다. 건강정보 데이터베이스(110)에 저장되어 있는 건강 정보는 환자의 치료 레이블이 존재하는 건강정보와 환자의 치료 레이블이 존재하지 않는 건강정보가 혼합되어 저장되어 있다. 여기서 레이블 정보의 종류는 환자의 감정 상태를 나타내는 감정 레이블 정보, 환자의 건강 수치 상태를 나타내는 수치 레이블 정보, 환자의 징후 상태를 나타내는 징후 레이블 정보 및 환자의 건강 상태에 따라 처방된 치료 레이블 정보를 포함한다. 통상적으로 다수 환자 중 일부 환자는 모든 종류의 레이블 정보를 구비하는 반면 다른 일부 환자는 치료 레이블 정보를 구비하지 않거나 감정 레이블 정보, 수치 레이블 정보, 징후 레이블 정보 중 어느 하나를 구비하지 않고 저장될 수 있다.
- [0027] 본 발명이 적용되는 분야에 따라 감정 레이블 정보, 수치 레이블 정보, 징후 레이블 정보 및 치료 레이블 정보 이외의 추가적인 건강 레이블 정보가 건강정보 데이터베이스(110)에 저장되어 있거나 감정 레이블 정보, 수치 레이블 정보, 징후 레이블 정보 및 치료 레이블 정보와 다른 건강 레이블 정보가 건강정보 데이터베이스(110)에 저장될 수 있으며, 레이블 유사도 행렬 생성부(120)는 건강정보 데이터베이스(110)에 저장되어 있는 건강 레이블 정보의 종류에 따라 레이블 유사도 행렬을 생성한다.
- [0028] 전체 유사도 행렬 생성부(130)는 다수 환자 사이의 각 건강 레이블 정보별 유사도를 나타내는 레이블 유사도 행렬을 서로 합하여 환자 사이의 전체 유사도 행렬을 생성한다. 본 발명이 적용되는 분야에 따라 각 레이블 유사도 행렬은 서로 동일한 가중치로 합해지거나 서로 다른 가중치로 합해질 수 있으며 이는 본 발명의 범위에 속한다.
- [0029] 예비 군집 정보 생성부(140)는 생성한 전체 유사도 행렬을 이용하여 서로 유사한 속성을 가지는 환자 사이의 군집 밀집도를 반복적으로 증가시키며 이전 단계에서의 군집 밀집도와 현재 단계에서의 군집 밀집도의 차이가 임계값 이하가 되는 경우 즉, 환자 사이의 군집 밀집도가 임계값보다 작은 경우 다수 환자의 예비 군집 정보를 생성한다. 군집화부(150)는 생성한 예비 군집 정보의 분포 밀집도 특성에 기초하여 설정된 수의 군집으로 다수 환자를 군집화한다.

- [0030] 도 2는 본 발명에 따른 레이블 유사도 행렬 생성부를 보다 구체적으로 설명하기 위한 기능 블록도이다.
- [0031] 도 2를 참고로 살펴보면, 감정 행렬 생성부(111)는 건강정보 데이터베이스(110)에 저장되어 있는 다수 환자의 감정 레이블 정보 사이의 유사도를 계산하고, 계산한 다수 환자의 감정 정보 사이의 유사도를 원소로 하는 감정 레이블 유사도 행렬을 생성한다. 수치 행렬 생성부(113)는 건강정보 데이터베이스(110)에 저장되어 있는 다수 환자의 수치 레이블 정보 사이의 유사도를 계산하고, 계산한 다수 환자의 수치 레이블 정보 사이의 유사도를 원소로 하는 수치 레이블 유사도 행렬을 생성한다. 징후 행렬 생성부(115)는 건강정보 데이터베이스(110)에 저장되어 있는 다수 환자의 징후 레이블 정보 사이의 유사도를 계산하고, 계산한 다수 환자의 징후 레이블 정보 사이의 유사도를 원소로 하는 징후 레이블 유사도 행렬을 생성한다. 치료 행렬 생성부(117)는 건강정보 데이터베이스에 저장되어 있는 다수 환자의 치료 레이블 정보 사이의 유사도를 계산하고, 계산한 다수 환자의 치료 레이블 정보 사이의 유사도를 원소로 하는 치료 레이블 유사도 행렬을 생성한다.
- [0032] 바람직하게, 감정 정보 사이의 유사도, 수치 정보 사이의 유사도, 징후 정보 사이의 유사도, 치료 정보 사이의 유사도는 유클리드 거리(euclidean distance)를 이용하여 계산되는데, 각 환자의 감정 정보를 구성하는 각 항목별 식별값의 차이로 계산된다.
- [0033] 행렬 정규화부(119)는 감정 레이블 유사도 행렬을 구성하는 원소값, 수치 레이블 유사도 행렬을 구성하는 원소값, 징후 레이블 유사도 행렬을 구성하는 원소값을 동일한 기준값으로 정규화한다. 예를 들어, 감정 레이블 유사도 행렬을 구성하는 원소값, 수치 레이블 유사도 행렬을 구성하는 원소값, 징후 레이블 유사도 행렬을 구성하는 원소값은 0에서 1 사이의 값을 가지도록 정규화된다.
- [0034] 도 3은 본 발명에 따른 예비 군집 정보 생성부를 보다 구체적으로 설명하기 위한 기능 블록도이다.
- [0035] 도 3을 참고로 살펴보면, 초기 벡터 생성부(141)는 전체 유사도 행렬이 $n \times n$ 크기인 경우 $1 \times n$ 크기의 초기 벡터(v_0)를 생성한다. 초기 벡터 생성부(141)는 1에서 9의 값을 가지는 임의의 값으로 설정된 후 다시 0에서 1의 값을 가지는 값으로 정규화되거나, 전체 유사도 행렬의 각 행의 합을 전체 행렬의 합으로 나누어 정규화하여 생성된다.
- [0036] 다음 벡터 생성부(143)는 초기 벡터를 이전 벡터(v)로 설정하고 이전 벡터와 전체 유사도 행렬을 곱하여 다음 벡터(v')를 생성한다. 밀집부(145)는 차 계산부(145-1)와 밀집 제어부(145-3)를 구비하는데, 차 계산부(145-1)는 다음 벡터(v')와 이전 벡터(v)를 차감하여 차를 계산하며 밀집 제어부(145-3)는 차 계산부(145-1)에서 계산한 이전 벡터와 다음 벡터의 차가 임계값보다 작은지 판단하며, 판단 결과에 기초하여 이전 벡터와 다음 벡터의 차가 임계값보다 큰 경우 다음 벡터 생성부(143)를 통해 다음 벡터를 이전 벡터로 설정하여 다음 벡터를 반복 생성하도록 제어한다. 한편, 예비 군집 결정부(147)는 밀집 제어부(145-3)의 판단 결과에 기초하여 이전 벡터와 다음 벡터의 차가 임계값보다 작은 경우, 다음 벡터를 다수 환자의 예비 군집 정보로 결정한다.
- [0037] 도 4는 본 발명에 따른 건강정보 군집화 장치를 이용하여 생성한 건강정보의 훈련 데이터를 이용하여 새로운 환자에 치료 레이블 정보를 제공하는 치료 조언 시스템을 설명하기 위한 기능 블록도이다.
- [0038] 도 4를 참고로 보다 구체적으로 살펴보면, 치료 조언 시스템은 유선/무선 네트워크(300)에 접속되어 있는 건강정보 군집화 장치(100), 치료 조언 장치(200) 및 사용자 단말기(400)를 구비하고 있다.
- [0039] 건강 정보 군집화 장치(100)는 앞서 도 1 내지 도 3을 참고로 설명한 바와 같이 치료 레이블 정보를 일부만 구비하는 건강정보 데이터베이스를 활용하여 유사한 환자들을 군집화하여 유사한 건강상태를 가지는 환자를 군집화한 건강정보 훈련 데이터를 생성한다. 치료 조언 장치(200)는 새로운 환자가 소지하는 사용자 단말기(400)를 통해 새로운 환자의 건강정보, 즉 감정 레이블 정보, 수치 레이블 정보 및 징후 레이블 정보를 네트워크(300)를 통해 수신하는 경우, 수신한 새로운 환자의 건강정보와 가장 유사한 건강정보를 가지는 군집을 판단하고, 가장 유사한 건강정보를 가지는 군집의 치료 레이블 정보를 네트워크(300)를 통해 사용자 단말기(400)로 제공한다.
- [0040] 본 발명이 적용되는 분야에 따라 치료 조언 장치(200)는 별도의 사용자 단말기(400)를 이용하여 새로운 환자의 건강 정보를 수신하는 대신, 사용자 인터페이스부(미도시)를 더 구비하며 사용자 인터페이스부를 통해 새로운 환자의 건강 정보를 입력받을 수 있다.

- [0041] 도 5는 본 발명에 따른 건강정보의 군집화 방법을 설명하기 위한 흐름도이다.
- [0042] 도 5를 참고로 살펴보면, 다수 환자의 환자별 치료 레이블이 있는 건강정보와 치료 레이블이 없는 건강정보로 이루어진 건강정보 데이터베이스에서 치료 레이블 정보를 포함하여 건강 레이블 정보 종류에 따라 환자 사이의 각 건강 레이블 정보별 유사도를 나타내는 레이블 유사도 행렬을 생성한다(S100). 도 6은 건강정보 데이터베이스에 저장되어 있는 건강 레이블 정보의 일 예를 설명하기 위한 도면이다. 도 6(a)는 환자별 감정 레이블 정보의 일 예를 도시하고 있으며, 도 6(b)는 환자별 수치 레이블 정보의 일 예를 도시하고 있으며, 도 6(c)는 환자별 징후 레이블 정보의 일 예를 도시하고 있으며, 도 6(d)는 환자별 치료 레이블 정보의 일 예를 도시하고 있다. 바람직하게, 각 건강 레이블 정보에는 각 건강 레이블 정보를 구성하는 다수의 항목이 존재하며, 각 항목의 수치는 해당하는 레벨의 식별값으로 매칭되어 있다.
- [0043] 다시 도 5를 참고로 살펴보면, 각 레이블 정보별 유사도를 나타내는 레이블 유사도 행렬에 가중치를 곱하고 가중치가 곱해진 각 레이블 유사도 행렬을 서로 합하여 환자 사이의 전체 유사도 행렬을 생성한다(S200).
- [0044] 전체 유사도 행렬에서 서로 유사한 속성을 가지는 환자 사이의 군집 밀집도가 임계값보다 작도록 밀집시켜 다수 환자의 예비 군집 정보를 생성하며(S300), 예비 군집 정보의 분포 밀집도 특성에 기초하여 설정된 수의 군집으로 다수 환자를 군집화한다(S400). 바람직하게, 예비 군집 정보는 K-means 군집화 알고리즘을 통해 예비 군집 정보의 분포 밀집도 특성에 따라 설정된 수의 군집으로 군집화된다.
- [0045] 도 10은 K-means 군집화 알고리즘의 일 예를 설명하기 위한 도면으로, 도 10(a)와 같이 다수의 데이터가 정보가 존재하는 경우, 다수의 데이터를 2개의 군집으로 군집화하는 경우 군집화하고자 하는 수만큼 임의의 데이터를 선택한다. 도 10(b)와 같이 데이터 1과 2를 선택하는 경우, 나머지 데이터3, 4를 데이터1과 데이터 2 중 군집하는 데이터와 예비 군집을 형성한다. 도 10(c)와 같이 형성한 예비 군집을 구성하는 데이터의 평균으로 중심점을 계산하고 데이터 1 내지 4를 계산한 중심점(c1, c2)에 중 군집하는 데이터로 군집을 새로 형성한다. 도 10(d)와 같이 새로 형성한 군집을 구성하는 데이터의 평균으로 새로운 중심점(c3, c4)을 계산하고 데이터 1 내지 4를 새로운 중심점(c3, c4) 중 군집하는 데이터로 군집을 새로 형성한다. 새로 형성한 군집이 이전 형성한 군집과 동일한 경우 군집화가 완료된다.
- [0046] 도 7은 본 발명에 따른 건강정보의 군집화 방법에서 레이블 유사도 행렬을 생성하는 단계를 보다 구체적으로 설명하기 위한 흐름도이다.
- [0047] 도 7을 참고로 보다 구체적으로 살펴보면, 다수 환자의 다양한 건강 레이블 정보를 구비하는 건강정보 데이터베이스에서 동일한 종류의 건강 레이블 정보를 추출하여 환자 사이의 건강 레이블 정보의 유클리드 거리를 계산하고, 계산한 유클리드 거리로부터 환자 사이의 건강 레이블 정보의 유사도를 계산한다(S110).
- [0048] 예를 들어 설명하면, 건강정보 데이터베이스에 환자1 내지 환자5의 건강 레이블 정보가 저장되어 있는 경우, 환자1과 환자2 사이, 환자1과 환자3 사이, 환자1과 환자4 사이, 환자1과 환자5사이의 감정 레이블 정보 중 동일한 항목에 대한 차로부터 유클리드 거리를 계산한다. 건강 레이블 정보(S_k)에 대한 환자(i)와 환자(j) 사이의 유클리드 거리는 아래의 수학적식(1)에 의해 계산되는데,
- [0049] [수학적식 1]
- $$S_k(i,j)=\sqrt{(a_i-a_j)^2+(b_i-b_j)^2+....+(m_i-m_j)^2}$$
- [0050]
- [0051] 여기서 a, b, ..., j는 건강 레이블 정보(S_k)를 구성하는 환자(i)와 환자(j)의 항목별 식별값이다.
- [0052] 유클리드 거리로부터 계산한 환자 사이의 건강 레이블 정보의 유사도로부터 각 건강 레이블 정보의 유사도를 나타내는 레이블 유사도 행렬을 생성한다(S120). 여기서 치료 레이블 유사도 행렬을 구성하는 원소들의 값은 0, 1의 값을 가지는데, 환자 사이의 치료 정보 유사도가 일치하는 경우 유클리드 거리가 0을 가지며 원소값은 동일하게 0의 값을 가진다. 한편, 환자 사이의 치료 정보 유사도가 일치하지 않는 경우 유클리드 거리는 0을 초과하며 원소값은 1로 설정된다.
- [0053] 예를 들어, 환자1 내지 환자 5의 감정 레이블 유사도 행렬, 수치 레이블 유사도 행렬, 징후 레이블 유사도 행렬

및 치료 레이블 유사도 행렬의 일 예는 아래의 수학식(2) 내지 수학식(5)와 같다.

[0054] [수학식 2]

$$\text{감정레이블유사도행렬} = \begin{bmatrix} 0 & 2 & 3 & 2 & 4 \\ 2 & 0 & 3 & 1 & 2 \\ 3 & 3 & 0 & 1 & 2 \\ 2 & 1 & 1 & 0 & 1 \\ 4 & 2 & 2 & 1 & 0 \end{bmatrix}$$

[0056] [수학식 3]

$$\text{수치레이블유사도행렬} = \begin{bmatrix} 0 & 1 & 1 & 3 & 1 \\ 1 & 0 & 2 & 2 & 2 \\ 1 & 2 & 0 & 3 & 2 \\ 3 & 2 & 3 & 0 & 1 \\ 1 & 2 & 2 & 1 & 0 \end{bmatrix}$$

[0058] [수학식 4]

$$\text{징후레이블유사도행렬} = \begin{bmatrix} 0 & 1 & 1 & 2 & 3 \\ 1 & 0 & 1 & 1 & 2 \\ 1 & 1 & 0 & 1 & 2 \\ 2 & 1 & 1 & 0 & 1 \\ 3 & 2 & 2 & 1 & 0 \end{bmatrix}$$

[0060] [수학식 5]

$$\text{치료레이블유사도행렬} = \begin{bmatrix} 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

생성한 각 건강 레이블 정보에 대한 레이블 유사도 행렬을 동일한 기준값으로 정규화한다(S130). 바람직하게, 여기서 기준값은 0과 1인데 각 건강 레이블 정보에 대한 레이블 유사도 행렬의 원소는 0에서 1 사이의 값을 가지도록 정규화된다.

도 8은 본 발명에 따른 건강정보 군집화 방법에서 예비 군집 정보를 생성하는 단계를 보다 구체적으로 설명하기 위한 흐름도이다.

도 8을 참고로 보다 구체적으로 살펴보면, 전체 유사도 행렬의 크기가 $n \times n$ 인 경우 $1 \times n$ 의 크기를 가지는 초기 벡터를 생성하고(S310), 생성한 초기 벡터를 기준값으로 정규화한다(S320). 바람직하게, 초기 벡터는 임의값으

로 설정된 후 기준값, 예를 들어 0에서 1 사이의 값을 가지도록 정규화된다.

- [0065] 초기 벡터를 이전 벡터로 설정하고 전체 유사도 행렬과 이전 벡터를 곱하여 다음 벡터를 생성하며(S330), 생성한 다음 벡터와 이전 벡터를 차감하여 다음 벡터와 이전 벡터의 차가 설정한 임계값보다 작은지 판단한다(S340).
- [0066] 이전 벡터와 다음 벡터의 차가 임계값보다 작은 경우 다음 벡터에 기초하여 다음 벡터를 다수 환자의 예비 군집 데이터로 생성한다(S350). 그러나 이전 벡터와 다음 벡터의 차가 임계값보다 큰 경우 다음 벡터를 이전 벡터로 재설정하고, 재설정된 이전 벡터로부터 다음 벡터를 재생성하여 재설정된 다음 벡터와 재설정된 이전 벡터의 차가 설정한 임계값보다 작을 때까지 반복하여 다음 벡터를 재생성한다.
- [0067] 이러한 다음 벡터의 재생성을 통해 다수 환자 중 서로 밀접한, 즉 유사한 환자는 더욱 밀접하게 된다.
- [0068] 도 9는 본 발명에 따른 치료 조언 방법을 설명하기 위한 흐름도이다.
- [0069] 도 9를 참고로 보다 구체적으로 살펴보면, 다수 환자의 환자별 치료 레이블이 있는 건강정보와 치료 레이블이 없는 건강정보로 이루어진 건강정보 데이터베이스에서 치료 레이블 정보를 포함하여 레이블 정보 종류에 따라 환자 사이의 각 레이블 정보별 유사도를 나타내는 레이블 유사도 행렬을 생성한다(S500).
- [0070] 각 레이블 정보별 유사도를 나타내는 레이블 유사도 행렬의 합으로부터 환자 사이의 전체 유사도 행렬을 생성하고(S600), 전체 유사도 행렬에서 서로 유사한 속성을 가지는 환자 사이의 건강정보를 밀집시켜 다수의 환자의 예비 군집 정보를 생성하고 예비 군집 정보의 분포 밀집도 특성에 기초하여 설정된 수로 상기 다수 환자를 군집화한다(S700).
- [0071] 군집화된 환자의 레이블 정보 종류별 건강정보의 중심값으로 각 군집별 군집 정보를 생성한다(S800). 각 군집별 군집 정보는 각 군집을 구성하는 환자의 각 레이블 정보의 평균값을 각 군집의 각 레이블 정보로 계산한다.
- [0072] 신규 환자의 건강정보와 군집 정보의 유사도를 판단하여 신규 환자와 가장 유사도가 높은 군집 정보의 치료 레이블 정보를 신규 환자에 대한 치료 레이블 정보로 제공한다(S900).
- [0073] 한편, 상술한 본 발명의 실시 예들은 컴퓨터에서 실행될 수 있는 프로그램으로 작성 가능하고, 컴퓨터로 읽을 수 있는 기록 매체를 이용하여 상기 프로그램을 동작시키는 범용 디지털 컴퓨터에서 구현될 수 있다.
- [0074] 상기 컴퓨터로 읽을 수 있는 기록 매체는 마그네틱 저장 매체(예를 들어, 롬, 플로피 디스크, 하드디스크 등), 광학적 판독 매체(예를 들면, 시디롬, 디브이디 등) 및 캐리어 웨이브(예를 들면, 인터넷을 통한 전송)와 같은 저장 매체를 포함한다.
- [0075] 본 발명은 도면에 도시된 실시예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다. 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 등록청구범위의 기술적 사상에 의해 정해져야 할 것이다.

부호의 설명

- | | | |
|--------|--------------------|---------------------|
| [0076] | 110: 건강정보 데이터베이스 | 120: 레이블 유사도 행렬 생성부 |
| | 130: 전체 유사도 행렬 생성부 | 140: 예비 군집 정보 생성부 |
| | 150: 군집화부 | 111: 감정 행렬 생성부 |
| | 113: 수치 행렬 생성부 | 115: 징후 행렬 생성부 |
| | 117: 치료 행렬 생성부 | 119: 행렬 정규화부 |
| | 141: 초기 벡터 생성부 | 143: 다음 벡터 생성부 |
| | 145: 밀집부 | |

100: 건강정보 군집화 장치

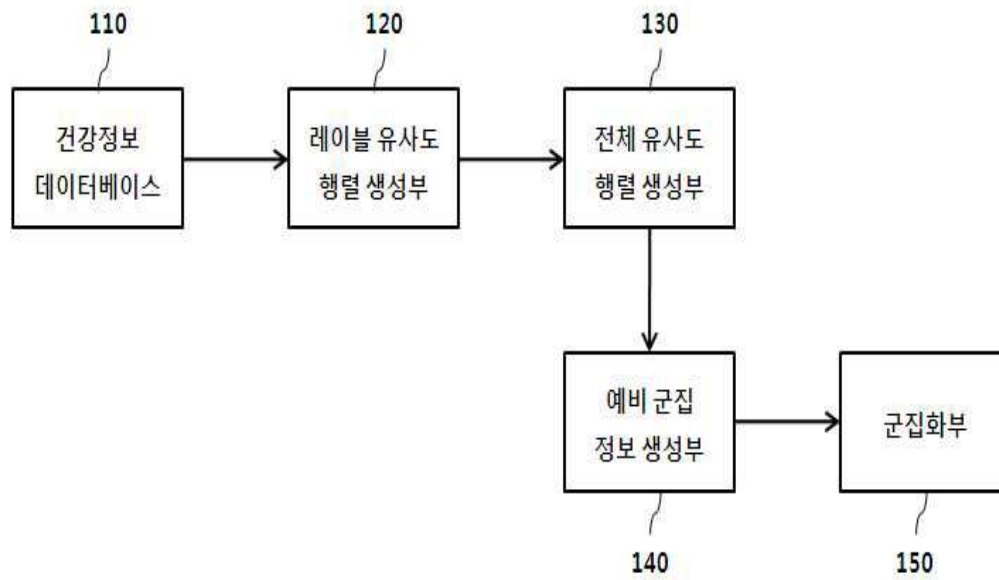
200: 치료 조언 장치

300: 네트워크

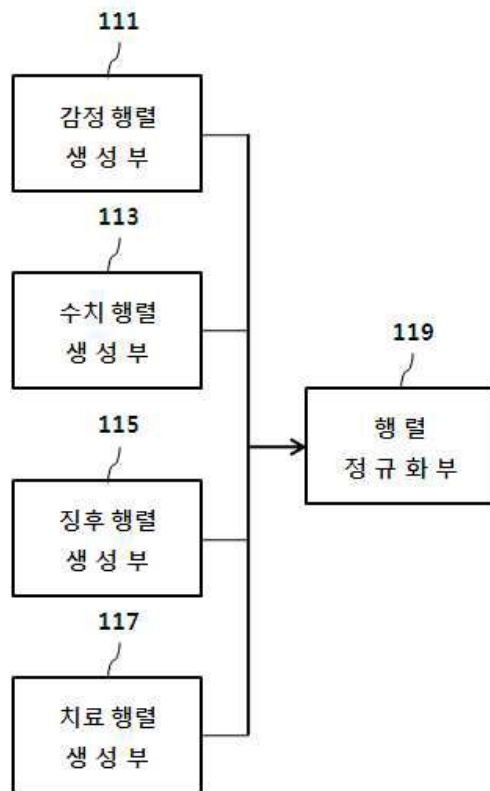
400: 사용자 단말기

도면

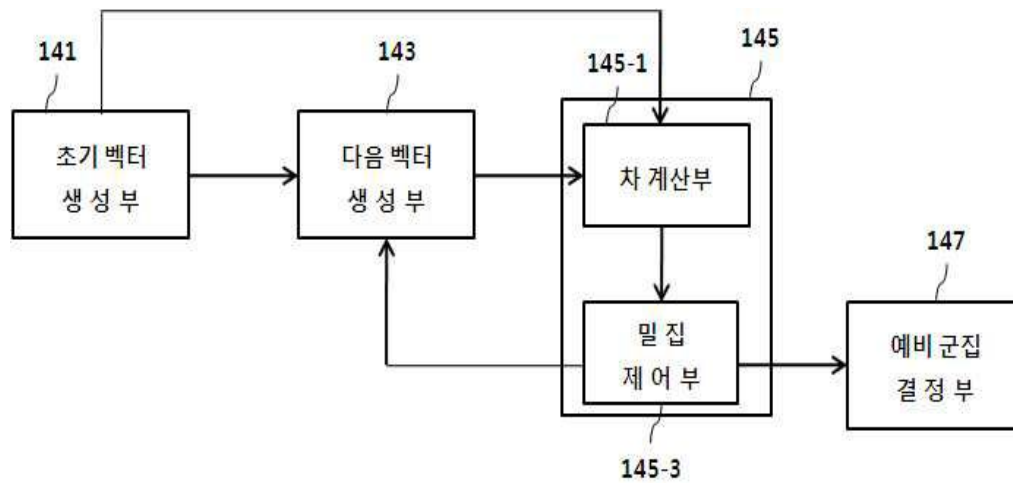
도면1



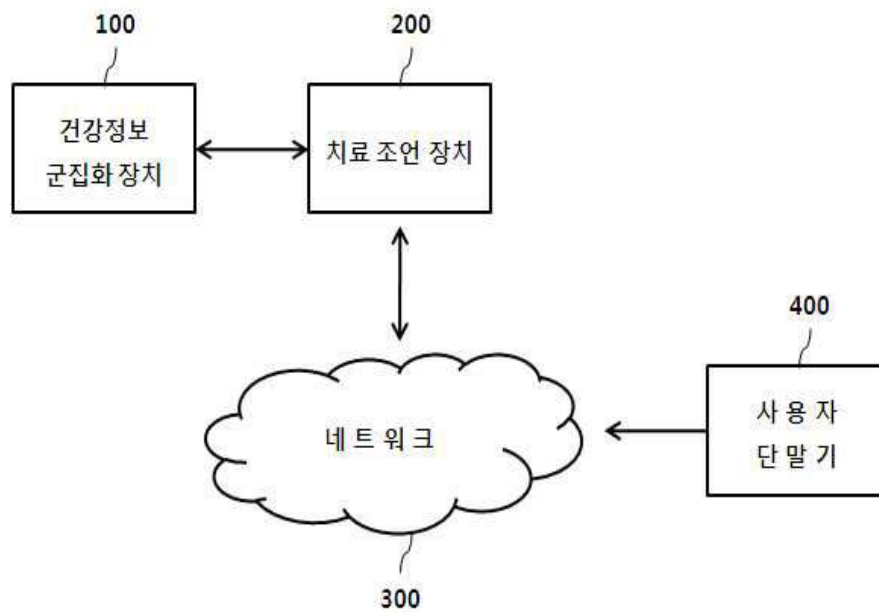
도면2



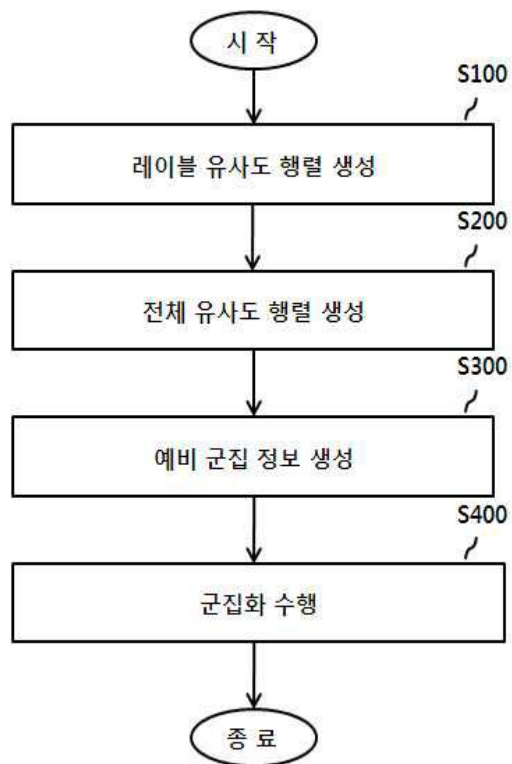
도면3



도면4



도면5



도면6

항 목	레 벨
사회 친밀도	3/(1~10)
내적 감성도	3/(1~10)
육체 피로도	4/(1~10)
정신 피로도	7/(1~10)

(a)

항 목	수 치	레 벨
혈 압	138.5	7/(1~10)
당수치	26.1	8/(1~10)
콜레스테롤	47.8	4/(1~10)
몸무게/키	0.53	7/(1~10)

(b)

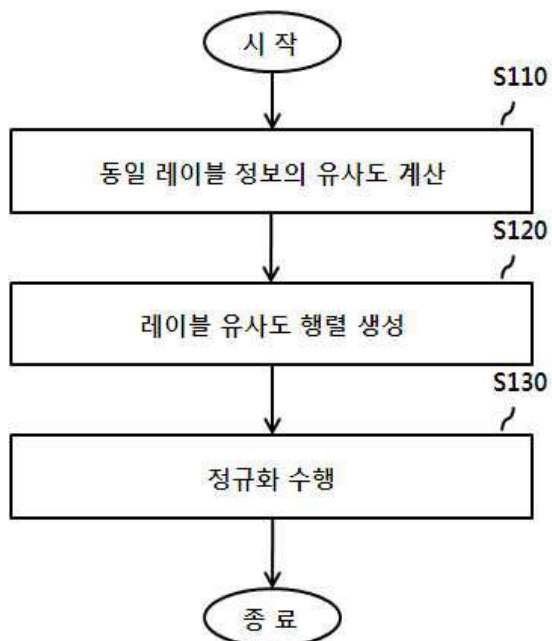
항 목	레 벨
흥분도	2/(1~10)
긴장도	3/(1~10)
고통도	5/(1~10)
우울도	7/(1~10)

(c)

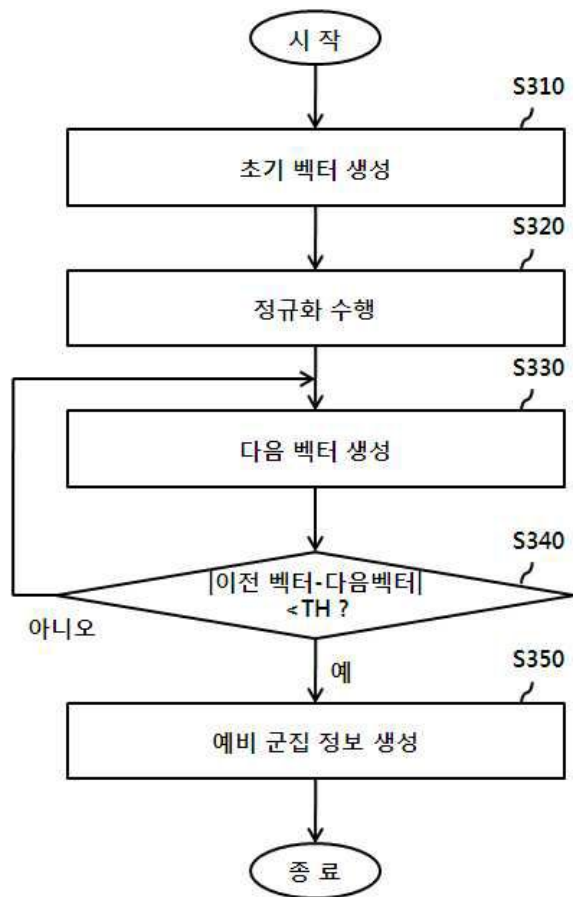
항 목	수 치	레 벨
약품 A	75.8	3/(1~10)
약품 B	21.8	2/(1~10)
약품 C	135.9	3/(1~10)
수면시간	7	5/(1~10)
섭취 열량	2450	4(1~10)

(d)

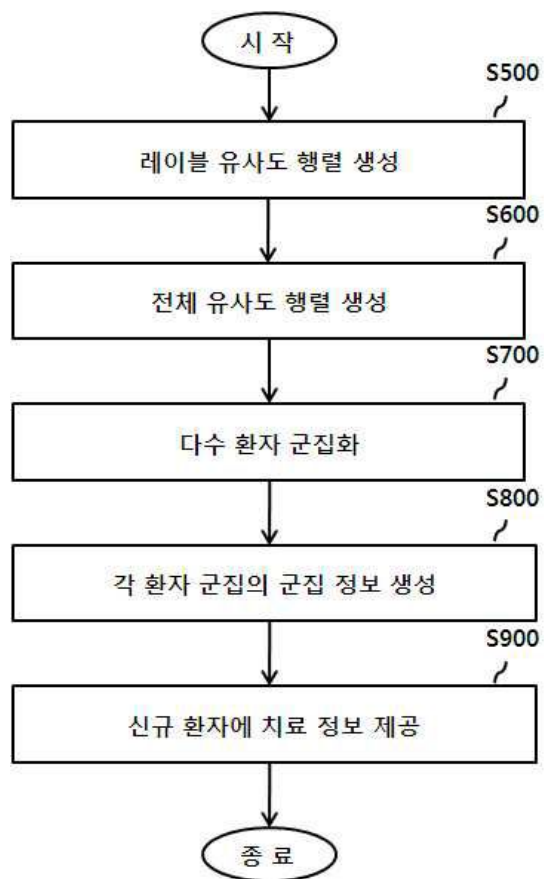
도면7



도면8



도면9



도면10

