(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0064630
(43) 공개일자 2021년06월03일

(51) 국제특허분류(Int. Cl.)
*G16H 50/70* (2018.01) *G16H 15/00* (2018.01)
*G16H 50/80* (2018.01) *G16H 70/60* (2018.01)
(52) CPC특허분류
*G16H 50/70* (2018.01)
*G16H 15/00* (2018.01)
(21) 출원번호 10-2019-0153057
(22) 출원일자 2019년11월26일
심사청구일자 2019년11월26일

(71) 출원인
**인천대학교 산학협력단**
인천광역시 연수구 아카데미로 119 (송도동)
(72) 발명자
**전광길**
인천광역시 연수구 경원대로119번길 21, 114동 904호(동춘동, 연수2차풍림아파트)
(74) 대리인
**특허법인충정**

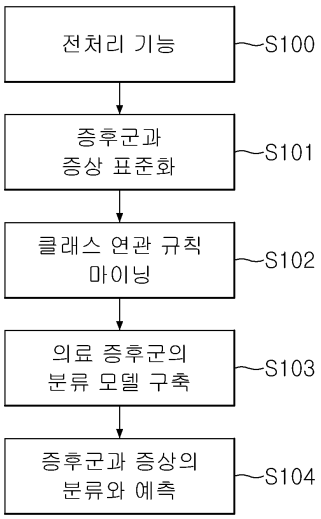
전체 청구항 수 : 총 7 항

(54) 발명의 명칭 **클래스 연관 규칙을 이용한 질병 분류 시스템 및 방법**

## (57) 요 약

클래스 연관 규칙을 이용한 질병 분류 시스템 및 방법은 증상과 증후군 사이의 강력한 연관 규칙을 적용하기 위해서 우선 순위가 높은 세트를 선택하여 티벳 의학 증후군의 분류 모델을 설정하여 티벳 의학 증후군을 진단 및 치료하고, 일반적인 임상 진단의 정확성을 향상시킬 수 있는 효과가 있다.

*대 표 도* - 도3



(52) CPC특허분류
*G16H 50/80* (2018.01)
*G16H 70/60* (2018.01)

# *명 세 서*

## *청구범위*

### 청구항 1

의료 정보 데이터베이스부로부터 티벳의 의료 기록 문서를 수신하는 입력부;

상기 의료 기록 문서에서 연관 규칙의 생성과 무관한 불용어를 제거하여 의료 기록 문서에 포함된 복수의 단어 중에서 연관 규칙 생성의 대상이 되는 복수의 의료 단어를 추출하고, 추출한 의료 단어를 이용하여 데이터 마이닝에 적합한 데이터 세트를 생성하는 전처리부;

상기 데이터 세트에서 병의 증상을 조건 속성으로 설정하고, 의료 증후군을 결정 속성으로 설정하고, 연관 규칙 마이닝(Mining) 알고리즘을 이용하여 상기 병의 증상과 상기 의료 증후군 간의 클래스 연관 규칙을 생성하는 데이터 마이닝부; 및

신뢰도와 지지도에 의한 클래스 연관 규칙을 정렬하는 방법 또는 Lift(상향법)에 의한 클래스 연관 규칙을 정렬을 이용하여 각각 우선 순위를 결정하고, 우선 순위가 가장 높은 규칙 세트를 분류 모델로 설정하는 의료 질병 분류부를 포함하는 것을 특징으로 하는 클래스 연관 규칙을 이용한 질병 분류 시스템.

### 청구항 2

제1항에 있어서,

상기 의료 질병 분류부는 신뢰도와 지지도에 의한 클래스 연관 규칙을 하기의 수학식 1에 의해 정렬하여 우선 순위를 결정하며, r1과 r2가 하기의 수학식 1의 규칙 중 어느 하나를 만족하면 r1에 우선권이 주어지고, r1 > r2로 표시하는 것을 특징으로 하는 클래스 연관 규칙을 이용한 질병 분류 시스템.

[수학식 1]

$$Conf(r_1) > Conf(r_2)$$

$$Conf(r_1) = Conf(r_2), Sup(r_1) > Sup(r_2)$$

$$Conf(r_1) = Conf(r_2), Sup(r_1) = Sup(r_2), r_1 \subset r_2$$

r1(규칙 1)의 신뢰도가 r2(규칙 2)의 신뢰도보다 클 때,

r1(규칙 1)의 신뢰도가 r2(규칙 2)의 신뢰도가 같고, r1(규칙 1)의 지지도가 r2(규칙 2)의 지지도보다 클 때,

r1(규칙 1)의 신뢰도가 r2(규칙 2)의 신뢰도가 같고, r1(규칙 1)의 지지도가 r2(규칙 2)의 지지도보다 같으며, r1의 범주가 r2의 보다 작을때(r2가 r1을 포함)

Sup는 규칙의 지지도를 Conf는 규칙의 신뢰도를 나타냄.

### 청구항 3

제1항에 있어서,

상기 의료 질병 분류부는 Lift(상향법)에 의한 클래스 연관 규칙을 하기의 수학식 2에 의해 정렬을 이용하여 우선 순위를 결정하고, r1 및 r2가 하기의 수학식 2의 조건을 만족하면, r1의 우선 순위는 r2보다 높고, r1 > r2로 표시하는 것을 특징으로 하는 클래스 연관 규칙을 이용한 질병 분류 시스템.

[수학식 2]

$$Lift(r_1) > Lift(r_2)$$

r1(규칙 1)의 리프트 > r2(규칙 2)의 리프트를 나타냄.

**청구항 4**

제1항에 있어서,

상기 데이터 마이닝부는 상기 데이터 세트에서 클래스 연관 규칙을 하기의 수학식 3과 같이 정의하고, 규칙 r의 지지도를 하기의 수학식 4와 같이 규칙의 증상과 증후군이 교집합(연결)이 되면 1이고, 교집한되는 것을 모두 카운트하고, 전체 집합 D로 나누며, 규칙 r의 신뢰도를 하기의 수학식 5와 같이 정의하는 것을 특징으로 하는 클래스 연관 규칙을 이용한 질병 분류 시스템.

[수학식 3]

$$\mathbf{x} => \mathbf{y}, \ |r.y|=1, \ x \in I, \ y \in C.$$

X는 규칙의 선행 조건이고, Y는 분류 속성이고, I는 데이터 세트(D)의 모든 항목(증상)들의 집합이고, C는 데이터 세트(D)에서 모든 클래스(증후군) 레벨들의 집합, D는 모든 경우의 전체 집합임.

[수학식 4]

$$Sup(r) = \frac{count(r.x \cap r.y)}{|D|}$$

[수학식 5]

$$Conf(r) = \frac{count(r.x \cap r.y)}{count(r.x)}$$

**청구항 5**

의료 정보 데이터베이스부로부터 티벳의 의료 기록 문서를 수신하는 단계;

상기 의료 기록 문서에서 연관 규칙의 생성과 무관한 불용어를 제거하여 의료 기록 문서에 포함된 복수의 단어 중에서 연관 규칙 생성의 대상이 되는 복수의 의료 단어를 추출하고, 추출한 의료 단어를 이용하여 데이터 마이닝에 적합한 데이터 세트를 생성하는 단계;

상기 데이터 세트에서 병의 증상을 조건 속성으로 설정하고, 의료 증후군을 결정 속성으로 설정하고, 연관 규칙 마이닝(Mining) 알고리즘을 이용하여 상기 병의 증상과 상기 의료 증후군 간의 클래스 연관 규칙을 생성하는 단계;

신뢰도와 지지도에 의한 클래스 연관 규칙을 정렬하는 방법 또는 Lift(상향법)에 의한 클래스 연관 규칙을 정렬을 이용하여 각각 우선 순위를 결정하고, 우선 순위가 가장 높은 규칙 세트를 분류 모델로 설정하는 단계를 포함하는 것을 특징으로 하는 클래스 연관 규칙을 이용한 질병 분류 방법.

**청구항 6**

제5항에 있어서,

상기 분류 모델로 설정하는 단계는,

상기 신뢰도와 지지도에 의한 클래스 연관 규칙을 하기의 수학식 1에 의해 정렬하여 우선 순위를 결정하며, r1과 r2가 하기의 수학식 1의 규칙 중 어느 하나를 만족하면 r1에 우선권이 주어지고, r1 > r2로 표시하는 것을 특징으로 하는 클래스 연관 규칙을 이용한 질병 분류 방법.

[수학식 1]

$$Conf(r_1) > Conf(r_2)$$

$$Conf(r_1) = Conf(r_2),\ Sup(r_1) > Sup(r_2)$$

$$Conf(r_1) = Conf(r_2),\ Sup(r_1) = Sup(r_2),\ r_1 \subset r_2$$

r1(규칙 1)의 신뢰도가 r2(규칙 2)의 신뢰도보다 클 때,

r1(규칙 1)의 신뢰도가 r2(규칙 2)의 신뢰도가 같고, r1(규칙 1)의 지지도가 r2(규칙 2)의 지지도보다 클 때,

r1(규칙 1)의 신뢰도가 r2(규칙 2)의 신뢰도가 같고, r1(규칙 1)의 지지도가 r2(규칙 2)의 지지도보다 같으며, r1의 범주가 r2의 보다 작을때(r2가 r1을 포함)

Sup는 규칙의 지지도를 Conf는 규칙의 신뢰도를 나타냄.

**청구항 7**

제5항에 있어서,

상기 분류 모델로 설정하는 단계는,

상기 Lift(상향법)에 의한 클래스 연관 규칙을 하기의 수학식 2에 의해 정렬을 이용하여 우선 순위를 결정하고, r1 및 r2가 하기의 수학식 2의 조건을 만족하면, r1의 우선 순위는 r2보다 높고, r1 > r2로 표시하는 것을 특징으로 하는 클래스 연관 규칙을 이용한 질병 분류 방법.

[수학식 2]

$$Lift(r_1) > Lift(r_2)$$

r1(규칙 1)의 리프트 > r2(규칙 2)의 리프트를 나타냄.

## *발명의 설명*

### *기 술 분 야*

[0001] 본 발명은 질병 분류 시스템에 관한 것으로서, 특히 증상과 증후군 사이의 강력한 연관 규칙을 적용하기 위해서 우선 순위가 높은 세트를 선택하여 티벳 의학 증후군의 분류 모델을 설정하여 티벳 의학 증후군을 진단 및 치료하는 클래스 연관 규칙을 이용한 질병 분류 시스템 및 방법에 관한 것이다.

### *배 경 기 술*

[0002] 티벳 의약은 오랜 역사를 가진 전통적인 국내 의약을 가지고 있다. 중국 의약의 중요한 부분은 독창적인 이론 시스템과 임상 효능이 있다.

[0003] 티벳 의약은 중국 북서부 지역의 농민을 위한 의약적 어드바스를 제공하고 있으나, 문화적 배경과 지역 환경의 복잡성으로 인해 임상 연구 기반이 취약하며, 임상 특성 기술이 표준화되지 못한 문제점이 있다.

[0004] 티벳은 고원 지대에 위치한 특성상 고산병의 발병률이 높고, 고산병의 반복으로 인하여 위암의 위험이 증가하고 있다.

[0005] 티벳은 의학 이론에 근거한 임상 원칙이 없기 때문에 질병 진단 및 치료 과정에서 의사의 주관적 요인이 크게 존재하는 문제점이 있다.

### *선행기술문헌*

#### *특허문헌*

[0006] (특허문헌 0001) 한국 등록특허번호 제10-1950112호

## 발명의 내용

### 해결하려는 과제

[0007] 이와 같은 문제점을 해결하기 위하여, 본 발명은 증상과 증후군 사이의 강력한 연관 규칙을 적용하기 위해서 우선 순위가 높은 세트를 선택하여 티벳 의학 증후군의 분류 모델을 설정하여 티벳 의학 증후군을 진단 및 치료하는 클래스 연관 규칙을 이용한 질병 분류 시스템 및 방법을 제공하는데 그 목적이 있다.

### 과제의 해결 수단

[0008] 상기 목적을 달성하기 위한 본 발명의 특징에 따른 클래스 연관 규칙을 이용한 질병 분류 시스템은,

[0009] 의료 정보 데이터베이스부로부터 티벳의 의료 기록 문서를 수신하는 입력부;

[0010] 상기 의료 기록 문서에서 연관 규칙의 생성과 무관한 불용어를 제거하여 의료 기록 문서에 포함된 복수의 단어 중에서 연관 규칙 생성의 대상이 되는 복수의 의료 단어를 추출하고, 추출한 의료 단어를 이용하여 데이터 마이닝에 적합한 데이터 세트를 생성하는 전처리부;

[0011] 상기 데이터 세트에서 병의 증상을 조건 속성으로 설정하고, 의료 증후군을 결정 속성으로 설정하고, 연관 규칙 마이닝(Mining) 알고리즘을 이용하여 상기 병의 증상과 상기 의료 증후군 간의 클래스 연관 규칙을 생성하는 데이터 마이닝부; 및

[0012] 신뢰도와 지지도에 의한 클래스 연관 규칙을 정렬하는 방법 또는 Lift(상향법)에 의한 클래스 연관 규칙을 정렬을 이용하여 각각 우선 순위를 결정하고, 우선 순위가 가장 높은 규칙 세트를 분류 모델로 설정하는 의료 질병 분류부를 포함하는 것을 특징으로 한다.

[0014] 본 발명의 특징에 따른 클래스 연관 규칙을 이용한 질병 분류 방법은,

[0015] 의료 정보 데이터베이스부로부터 티벳의 의료 기록 문서를 수신하는 단계;

[0016] 상기 의료 기록 문서에서 연관 규칙의 생성과 무관한 불용어를 제거하여 의료 기록 문서에 포함된 복수의 단어 중에서 연관 규칙 생성의 대상이 되는 복수의 의료 단어를 추출하고, 추출한 의료 단어를 이용하여 데이터 마이닝에 적합한 데이터 세트를 생성하는 단계;

[0017] 상기 데이터 세트에서 병의 증상을 조건 속성으로 설정하고, 의료 증후군을 결정 속성으로 설정하고, 연관 규칙 마이닝(Mining) 알고리즘을 이용하여 상기 병의 증상과 상기 의료 증후군 간의 클래스 연관 규칙을 생성하는 단계;

[0018] 신뢰도와 지지도에 의한 클래스 연관 규칙을 정렬하는 방법 또는 Lift(상향법)에 의한 클래스 연관 규칙을 정렬을 이용하여 각각 우선 순위를 결정하고, 우선 순위가 가장 높은 규칙 세트를 분류 모델로 설정하는 단계를 포함하는 것을 특징으로 한다.

### 발명의 효과

[0019] 전술한 구성에 의하여, 본 발명은 증상과 증후군 사이의 강력한 연관 규칙을 적용하기 위해서 우선 순위가 높은 세트를 선택하여 티벳 의학 증후군의 분류 모델을 설정하여 티벳 의학 증후군을 진단 및 치료하고 일반적인 임상 진단의 정확성을 향상시킬 수 있는 효과가 있다.

### 도면의 간단한 설명

[0020] 도 1은 본 발명의 실시예에 따른 클래스 연관 규칙을 이용한 질병 분류 시스템의 구성을 간략하게 나타낸 도면이다.

도 2는 본 발명의 실시예에 따른 증상과 증후군의 연결 관계의 개념을 나타낸 도면이다.

도 3은 본 발명의 실시예에 따른 클래스 연관 규칙을 이용한 질병 분류 방법을 나타낸 도면이다.

도 4는 본 발명의 실시예에 따른 Conf1Sup2와 Lift 정렬 방법을 기초로 한 분류 모델을 비교한 도면이다.

## *발명을 실시하기 위한 구체적인 내용*

[0021] 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 있는 것을 의미한다.

[0022] 도 1은 본 발명의 실시예에 따른 클래스 연관 규칙을 이용한 질병 분류 시스템의 구성을 간략하게 나타낸 도면이고, 도 2는 본 발명의 실시예에 따른 증상과 증후군의 연결 관계의 개념을 나타낸 도면이며, 도 3은 본 발명의 실시예에 따른 클래스 연관 규칙을 이용한 질병 분류 방법을 나타낸 도면이다.

[0023] 본 발명의 실시예에 따른 클래스 연관 규칙을 이용한 질병 분류 시스템(100)은 입력부(110), 전처리부(120), 데이터 마이닝부(130) 및 의료 질병 분류부(140)를 포함한다.

[0024] 입력부(110)는 의료 정보 데이터베이스부(101)로부터 티벳의 의료 기록 문서(임상 진료 문서, 처방 기록 문서, 진료 소견 문서, 의료 실험 데이터)를 수신한다.

[0025] 전처리부(120)는 의료 기록 문서에서 연관 규칙의 생성과 무관한 불용어를 제거하여 의료 기록 문서에 포함된 복수의 단어 중에서 연관 규칙 생성의 대상이 되는 복수의 의료 단어를 추출하고, 추출한 의료 단어를 이용하여 데이터 마이닝에 적합한 복수의 데이터 세트를 생성한다(S100, S101).

[0026] 데이터 마이닝부(130)는 단순히 문서에 나타나는 키워드의 빈출 정보를 이용하여 분류 카테고리를 지정하는 통계적인 방법과 다르게 연관규칙분석 기법인 아프리오리(Apriori) 알고리즘을 이용하여 추출한 의료 단어를 바탕으로 문서들 간에 연관성 있는 키워드들의 집합을 추출하여 각 카테고리 별로 의미적으로 대표성을 가진 키워드들로 분류 규칙을 생성할 수 있다.

[0027] 데이터 마이닝부(130)는 데이터 세트에서 고산병의 증상을 조건 속성으로 설정하고, 티벳 의료 증후군을 결정 속성으로 설정한다. 이로 인하여 Symptoms => Syndrome(증상이면 증후군이다)의 결과가 도출된다.

[0028] 데이터 마이닝부(130)는 아프리오리(Apriori) 연관 규칙 마이닝(Mining) 알고리즘을 이용하여 고산병의 증상과 티벳 의료 증후군 간의 클래스 연관 규칙을 생성한다(S102). 이때, 분류의 지지도(support)와 신뢰도(confidence)의 경계치(Threshold)가 미리 정해져 있다.

[0029] 연관 규칙(Asscoiation Rules)은 의료 단어와 같은 트랜잭션에서 추출한 "X => Y(support, confidence)" 형태의 조건과 결과(Condition-Conclusion) 식으로 표현되는 유용한 패턴을 의미한다.

[0030] 여기서, X는 트랜잭션을 구성하는 항목(Item)이고, "X => Y"는 조건에 해당하는 항목 X가 발생할 때 결과에 해당하는 항목 Y가 같이 발생함을 의미한다.

[0031] 연관 규칙의 목적은 미리 설정된 지지도 경계치와 신뢰도 경계치를 만족하는 규칙들을 데이터 마이닝하는 것이다.

[0032] 지지도(support)는 전체 트랜잭션에서 항목 X와 항목 Y가 동시에 발생하는 트랜잭션의 비율을 의미하고, 신뢰도(confidence)는 항목 X가 포함된 트랜잭션에서 항목 Y를 함께 포함하고 있는 트랜잭션의 비율을 의미한다.

[0033] 클래스 연관 규칙은 연관 규칙의 부분이고, 규칙들의 사후 항목은 분류 속성으로 고정되며, 이는 데이터 오브젝트의 카테고리를 구별하는데 사용될 수 있다.

[0034] 클래스 연관 규칙은 내부 관련성을 반영할 뿐만 아니라 규칙의 예측을 반영하고, 데이터 오브젝트를 예측하기 위해서 최소 지지도와 최소 신뢰도를 만족하는 규칙을 데이터 마이닝한다.

[0035] 데이터 마이닝부(130)는 트랜잭션 데이터 세트(D)라고 할 때, 데이터 세트에서 클래스 연관 규칙(Class Asscoiation Rules, CAR)을 다음의 수학식 1과 같이 정의한다.

*수학식 1*

[0036] $$x => y,\ |r.y|=1,\ x \in I,\ y \in C.$$

[0037] 규칙(Rule) r의 형태는 X => Y이고, I는 데이터 세트(D)의 모든 항목(증상)들의 집합이고, C는 데이터 세트(D)에서 모든 클래스(증후군) 레벨들의 집합이다. 예를 들면, I의 요소에는 기침, 가래, 콧물 등이 있고, C의 요소에는 폐렴, 감기 등이 있다. D는 모든 경우의 전체 집합이다.

[0038] 데이터 마이닝부(130)는 하기의 수학식 2를 이용하여 클래스 연관 규칙의 지지도를 계산한다.

[0039] 클래스 연관 규칙의 지지도(Sup(r))는 선행 조건(X)와 분류 속성(Y)의 규칙과 데이터 세트(D)의 모든 항목들의 총 개수를 매칭하는 항목들의 개수 비율을 나타낸다.

*수학식 2*

[0040] $$Sup(r) = \frac{count(r.x \cap r.y)}{|D|}$$

[0041] 규칙 r의 지지도는 수학식 2와 같이 정의하며, 규칙의 증상과 증후군이 교집합(연결)이 되면 1이다. 이러한 것들을 모두 카운트하고, 전체 집합 D로 나눈다.

[0042] 예를 들어, 기침 -> 감기, 콧물 -> 감기, 가래 -> 미정이면, 2/3이 Sup이다.

[0043] 데이터 마이닝부(130)는 하기의 수학식 3을 이용하여 클래스 연관 규칙의 신뢰도를 계산한다.

[0044] 클래스 연관 규칙의 신뢰도(Conf(r))는 선행 조건(X)와 분류 속성(Y)의 규칙과 데이터 세트(D)에서 선행 조건(X)의 항목들의 총 개수를 만족하는 항목들의 개수 비율을 나타낸다.

*수학식 3*

[0045] $$Conf(r) = \frac{count(r.x \cap r.y)}{count(r.x)}$$

[0046] 수학식 3은 전술한 수학식 2에서 D 대신에 count(r.x)로 교체한 것이다.

[0047] 예를 들어, 기침 -> 감기, 콧물 -> 감기, 가래 -> 미정이면, 2/3이 Conf이다.

[0048] 클래스 연관 규칙은 수학식 1, 수학식 2, 수학식 3의 조건을 만족하는 규칙이며, minSup와 minConf는 지지도와 신뢰도가 미리 설정된 최소 지지도 경계치와 최소 신뢰도 경계치이다(수학식 4).

*수학식 4*

[0049] $$a)\ |r.y|=1,\ x \in I,\ y \in C.$$
$$b)\ SupR(r) >= minSup$$
$$c)\ Conf(r) >= minConf$$

[0050] 데이터 마이닝부(130)는 각각의 클래스의 지지도를 계산하기 위해서 너비 우선 반복 검색 방법(Breadth First Iterative Search Method)이 사용되고, 최소 지지도 경계치와 비교한 후, 클래스 연결 규칙의 빈발(Frequent) 1-itemset(빈도가 가장 많이 나오는 항목 집합)을 검색한다.

[0051] 다음으로, 데이터 마이닝부(130)는 1-itemset(빈도가 가장 많이 나오는 항목 집합)의 클래스 연관 규칙의 결과를 이용하여 클래스 연관 규칙의 후보 2-itemset과 빈발(Frequent) 2-itemset이 결정된다. 이러한 과정은 클래스 연관 규칙의 k-itemset이 더 이상 발생할 수 없을 때까지 진행한다. 이로써 티벳 의료의 진단 및 치료에 기반하여 클래스 연관 규칙의 후보 집합은 "어떤 증상이면 어떤 증후군"이라는 규칙이 결정된다.

[0052] 다시 말해, 데이터 마이닝부(130)는 대용량의 데이터 세트의 단위 트랜잭션에서 빈번하게 발생하는 사건의 유형을 검색하고, 트랜잭션을 대상으로 최소 지지도 경계치 이상을 만족하는 빈발 항목 집합을 검색하고, 빈발 항목으로 후보 집합을 생성하고, 새로운 빈발 항목 집합이 생성되지 않을 때까지 반복 수행하며, 검색된 다량의 항목 집합 내에 포함된 항목들 중에서 최소 신뢰도 경계치 이상을 만족하는 항목들 간의 연관 규칙을 생성할 수 있다.

[0053] 후보 CAR 세트의 규칙은 중복적이고 우선 순위를 가지므로 분류 모델의 정확성을 향상시키기 위해 중복 규칙을 제거해야 하며, 우선 순위가 높은 CAR이 분류 모델로 선택된다.

[0054] 의료 질병 분류부(140)는 샘플을 올바르게 분류하는 기능을 기반으로 클래스 연관 규칙을 제거하는데, 후보 CAR 세트의 각 규칙 r에 대해서 훈련 세트에 규칙 r이 포함된 항목이 있는지 여부가 검출된다.

[0055] 의료 질병 분류부(140)는 규칙 r이 하나 이상의 샘플을 올바르게 분류하면 규칙 r은 잠재적으로 유용한 클래스 연관 규칙으로 표시된다.

[0056] 의료 질병 분류부(140)는 후보 CAR 세트의 규칙 또는 훈련 세트의 항목이 모두 순회되면, 항목을 올바르게 분류할 수 없는 규칙이 차단된다.

[0057] 의료 질병 분류부(140)는 CAR의 신뢰도를 첫 번째 키워드로 하고, 지지도를 두 번째 키워드로 하는 ConflSup2 및 리프트(Lift) 방법으로 각각 우선 순위를 두고, 우선 순위가 높은 규칙 세트를 선택하여 분류 모델을 설정한다(S103). 위의 두 가지 정렬 방법에 대한 설명은 다음과 같다.

[0058] (1) 신뢰도와 지지도에 의한 클래스 연관 규칙을 정렬하는 방법을 ConflSup2라 부른다.

*수학식 5*

[0059]
$$\mathrm{Conf}(r_1) > \mathrm{Conf}(r_2)$$

$$\mathrm{Conf}(r_1) = \mathrm{Conf}(r_2),\ \mathrm{Sup}(r_1) > \mathrm{Sup}(r_2)$$

$$\mathrm{Conf}(r_1) = \mathrm{Conf}(r_2),\ \mathrm{Sup}(r_1) = \mathrm{Sup}(r_2),\ r_1 \subset r_2$$

[0060] r1(규칙 1)의 신뢰도가 r2(규칙 2)의 신뢰도보다 클 때,

[0061] r1(규칙 1)의 신뢰도가 r2(규칙 2)의 신뢰도가 같고, r1(규칙 1)의 지지도가 r2(규칙 2)의 지지도보다 클 때,

[0062] r1(규칙 1)의 신뢰도가 r2(규칙 2)의 신뢰도가 같고, r1(규칙 1)의 지지도가 r2(규칙 2)의 지지도보다 같으며, r1의 범주가 r2의 보다 작을때(r2가 r1을 포함)

[0063] 여기서, Sup는 규칙의 지지도를 Conf는 규칙의 신뢰도를 나타낸다. 만일 r1과 r2가 상기 규칙 중 어느 하나를 만족하면 r1에 우선권이 주어지고, r1 > r2로 표시한다.

[0064] (2) Lift(상향법)에 의한 클래스 연관 규칙을 정렬한다.

*수학식 6*

[0065] $$Lift(r_1) > Lift(r_2)$$

[0066] r1(규칙 1)의 리프트 > r2(규칙 2)의 리프트를 나타낸다. r1 및 r2가 위의 조건을 만족하면, r1의 우선 순위는 r2보다 높고, r1 > r2로 표시한다.

[0067] 의료 질병 분류부(140)는 신뢰도와 지지도에 의한 클래스 연관 규칙을 정렬하는 방법(Conf1Sup2) 또는 Lift(상향법)에 의한 클래스 연관 규칙을 정렬하는 방법을 수행하고, 분류 결과의 정확성이 높은 정렬 방법을 분류 모델로 선택한다.

[0068] 하기의 표 1 및 도 4에 도시된 바와 같이, 클래스 연관 규칙이 Conf1Sup2를 기준으로 정렬 된 경우, 규칙의 개수가 40 내지 400의 범위에 있을 때 규칙의 개수가 증가함에 따라 정확도가 증가하고, 선택한 규칙의 개수가 80이면 최대 80.78%이다.

[0069] 그러나 규칙의 개수가 400에서 800 사이인 경우 규칙의 개수가 증가해도 정확도는 변경되지 않는다.

[0070] 리프트를 기준으로 클래스 연관 규칙을 정렬하는 경우 규칙의 개수가 40 내지 400 범위에 있을 때 규칙의 개수가 증가함에 따라 정확도가 증가하고 규칙의 개수가 400인 경우 정확도가 최대 82.72%이다. 그러나 규칙의 개수가 400에서 800사이일 때 규칙의 개수가 증가하면 정확도가 떨어진다.

[0071] 클래스 연관 규칙을 정렬하는 데 사용되는 방법에 관계없이 Lift를 기반으로 구성된 분류 모델의 정확도는 200을 갖는 규칙의 개수를 제외하고 Conf1Sup2를 기준으로 한 정확도보다 높다.

[0072] 의료 질병 분류부(140)는 두 가지 정렬 방법을 비교할 때 리프트 정렬 방법을 기준으로 클래스 연관 규칙을 정렬하고 규칙의 개수가 400인 경우 정확도가 82.72%로 가장 높기 때문에 규칙의 개수에 따라 Conf1Sup2 또는 Lift 중 정확도가 높은 정렬 방식을 선택한 분류 모델을 설정할 수 있다.

[0073] 즉, 의료 질병 분류부(140)는 도 4에 도시된 바와 같이, 규칙의 개수가 200인 경우, Conf1Sup2 정렬 방법으로 클래스 연관 규칙의 우선 순위를 정렬하고, 40, 80, 400, 800, 1000인 경우, Lift 정렬 방법으로 클래스 연관 규칙의 우선 순위를 정렬한다.

[0074] 또한, 의료 질병 분류부(140)는 Conf1Sup2 또는 Lift 정렬 방법을 기반으로 클래스 연관 규칙의 우선 순위를 정렬하고, 40, 80, 200, 400, 800, 1000개의 서로 다른 크기의 규칙을 선택하여 높은 순서에서 낮은 순서로 우선 순위 순서에 따라 분류 모델을 설정할 수 있다.

*표 1*

[0075]

| | Accuracy (%) | | | | | |
|---|---|---|---|---|---|---|
| *Number of CAR* / *Sorting methods* | 40 | 80 | 200 | 400 | 800 | 1000 |
| **Conf1Sup2** | 74.5 | 76.4 | 79.8 | 80.8 | 80.8 | 80.8 |
| **Lift** | 75.0 | 76.4 | 78.9 | 82.8 | 81.2 | 81.0 |

[0076] 의료 질병 분류부(140)는 클래스 연관 규칙를 기반으로 하는 분류 모델을이용하여 질병에 대한 증상을 티벳 의학 증후군의 연관 관계가 분류 모델의 규칙 우선 순위에 따라 항목이 올바르게 분류되는지 검증하는 프로세스를 실행한다(S104).

[0077] 이상에서 본 발명의 실시예는 장치 및/또는 방법을 통해서만 구현이 되는 것은 아니며, 본 발명의 실시예의 구성에 대응하는 기능을 실현하기 위한 프로그램, 그 프로그램이 기록된 기록 매체 등을 통해 구현될 수도 있으며, 이러한 구현은 앞서 설명한 실시예의 기재로부터 본 발명이 속하는 기술분야의 전문가라면 쉽게 구현할 수 있는 것이다.

[0078] 이상에서 본 발명의 실시예에 대하여 상세하게 설명하였지만 본 발명의 권리범위는 이에 한정되는 것은 아니고 다음의 청구범위에서 정의하고 있는 본 발명의 기본 개념을 이용한 당업자의 여러 변형 및 개량 형태 또한 본 발명의 권리범위에 속하는 것이다.

## *부호의 설명*

[0079] 100: 질병 분류 시스템

101: 의료 정보 데이터베이스부

110: 입력부

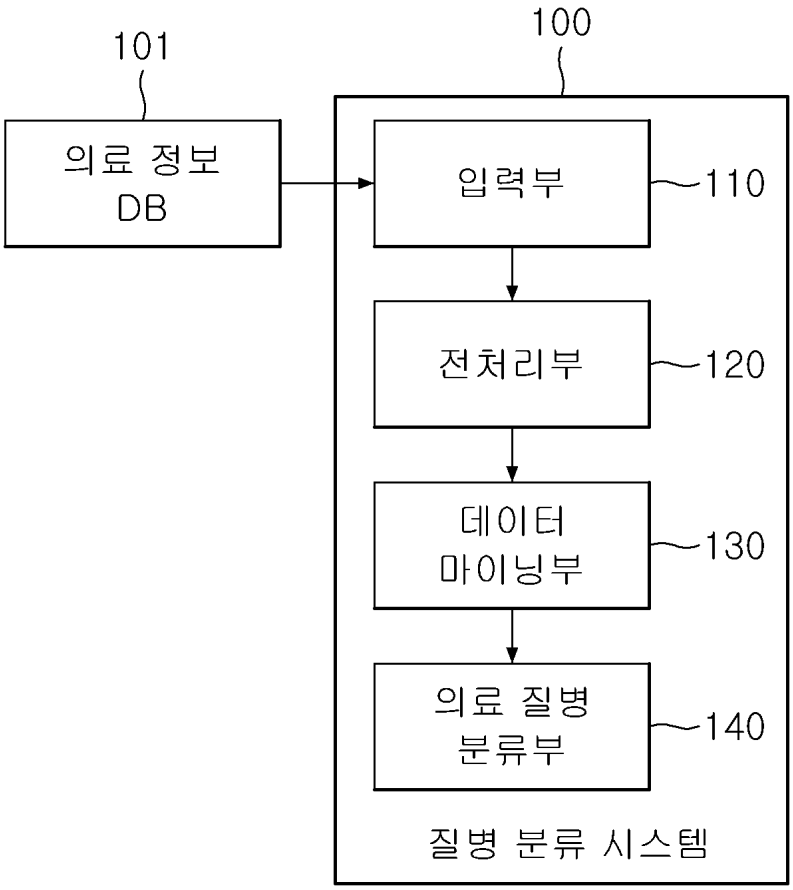
120: 전처리부

130: 데이터 마이닝부

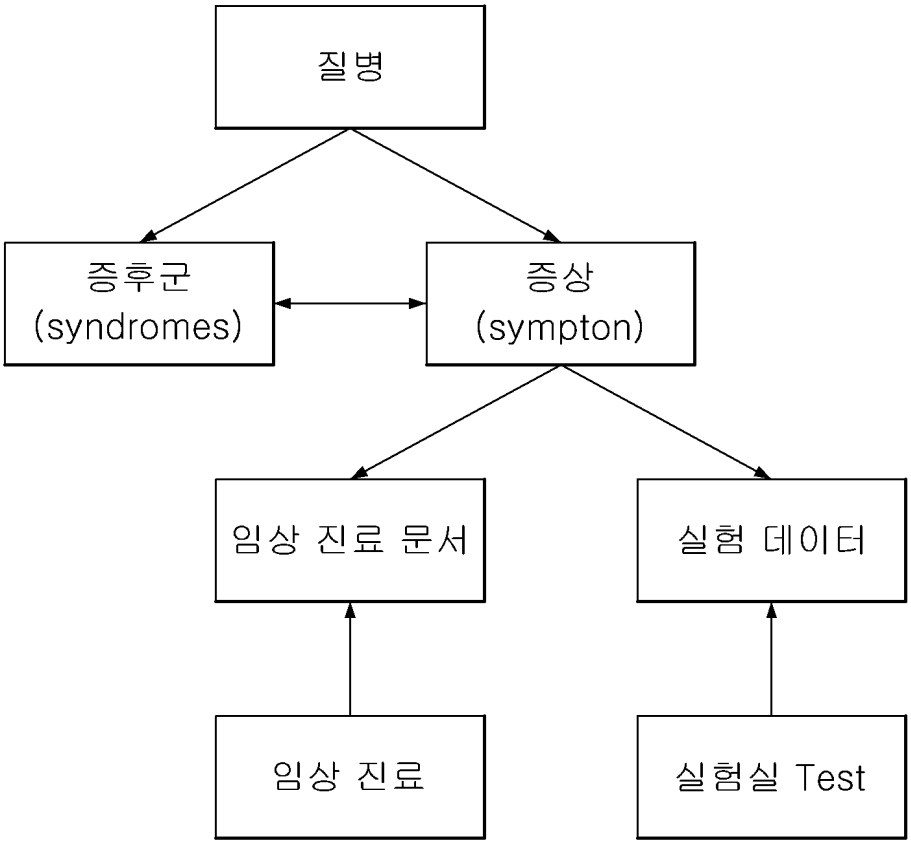
140: 의료 질병 분류부

## *도면*

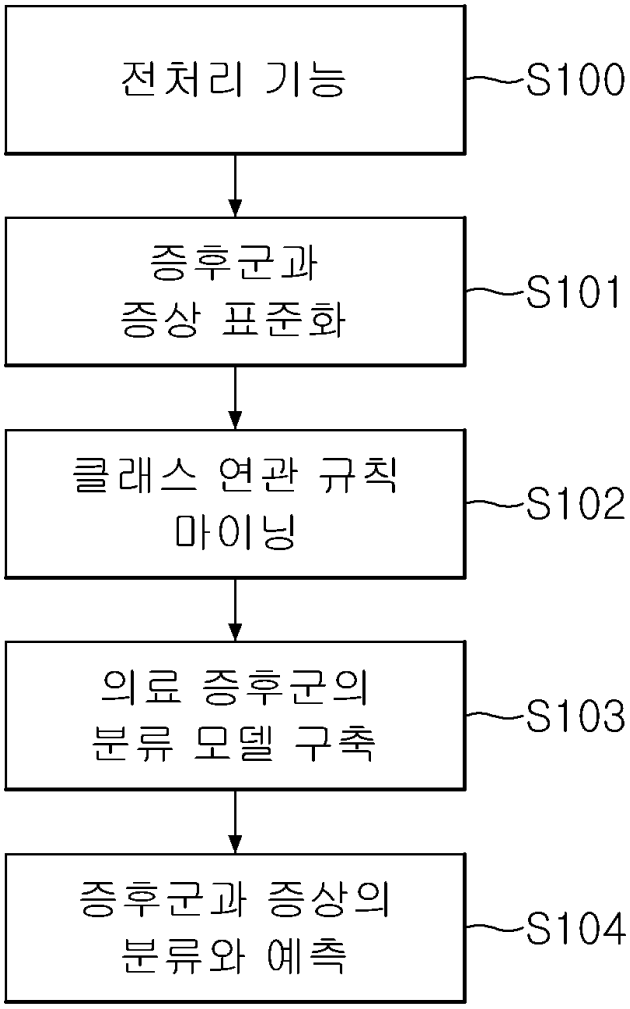
### *도면1*



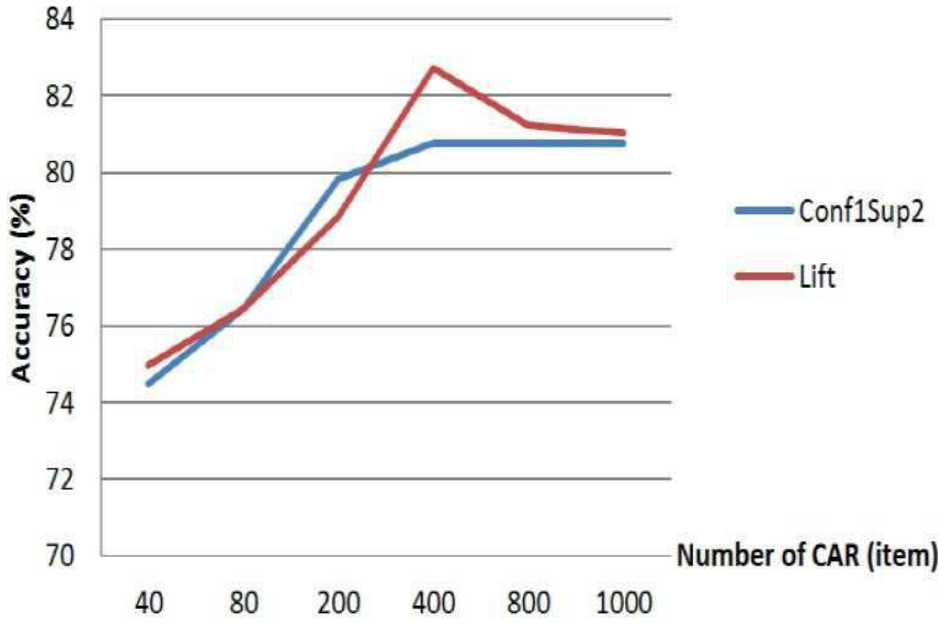
도면2



도면3


전처리 기능
S100
증후군과
증상 표준화
S101
클래스 연관 규칙
마이닝
S102
의료 증후군의
분류 모델 구축
S103
증후군과 증상의
분류와 예측
S104

도면4


84
82
80
78
76
74
72
70
Accuracy (%)
40
80
200
400
800
1000
Number of CAR (item)
Conf1Sup2
Lift