(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0104828
(43) 공개일자 2020년09월04일

(51) 국제특허분류(Int. Cl.)
A61B 5/00 (2006.01) *G16H 50/20* (2018.01)
(52) CPC특허분류
A61B 5/7275 (2013.01)
A61B 5/0033 (2018.08)
(21) 출원번호 10-2020-0024553
(22) 출원일자 2020년02월27일
심사청구일자 2020년02월27일
(30) 우선권주장
1020190023295 2019년02월27일 대한민국(KR)

(71) 출원인
울산과학기술원
울산광역시 울주군 언양읍 유니스트길 50
(72) 발명자
이정혜
울산광역시 울주군 언양읍 유니스트길 50
임치현
울산광역시 울주군 언양읍 유니스트길 50 울산과학기술원 교원숙소 404동 703호
(74) 대리인
리앤목특허법인

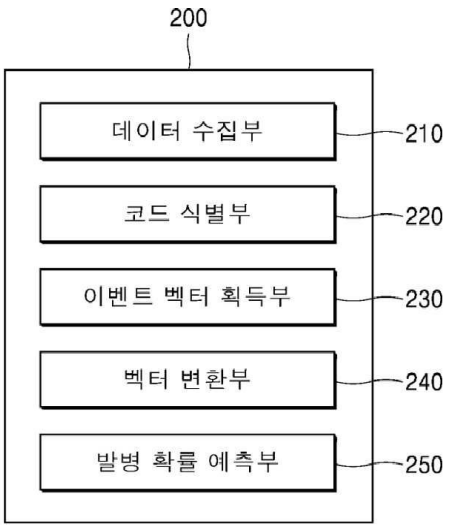
전체 청구항 수 : 총 8 항

(54) 발명의 명칭 **데이터 연합 기반의 질병 진단 시스템 및 그 운용 방법**

(57) 요 약

본 발명은 시스템을 포함하여 물리적으로 서로 다른 병원에서 수집된 데이터를 통합하여 향후 폐암 발병을 예측하고, 질병의 조기 진단 서비스를 제공하는 진단 시스템 및 그 운용 방법에 관한 것이다. 본 발명의 데이터 연합 기반의 질병 진단 방법은 제1 데이터 소스로부터 적어도 하나의 진단 코드를 포함하는 제1 의료 코드를 획득하고, 제1 데이터 소스와 상이한 제2 데이터 소스로부터 적어도 하나의 진단 코드를 포함하는 제2 의료 코드를 획득하는 단계, 제1 의료 코드 및 제2 의료 코드 중 기준 진단 코드를 식별하는 단계, 제1 의료 코드 및 제2 의료 코드로부터 각각 제1 이벤트 벡터 및 제2 이벤트 벡터를 획득하는 단계, 진단 코드를 기초로 제1 이벤트 벡터 및 제2 이벤트 벡터를 통합 벡터 공간으로 변환하는 단계 및 통합 벡터 공간 및 기준 진단 코드를 기초로 제1 의료 코드 및 제2 의료 코드와 관련된 환자의 발병 확률을 예측하는 단계를 포함한다.

대 표 도 - 도2



(52) CPC특허분류
G16H 50/20 (2018.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 1711077315
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 1/1
과제수행기관명 울산과학기술원
연구기간 2018.09.01 ~ 2019.02.28

명 세 서

청구범위

청구항 1

데이터 연합 기반의 질병 진단 방법에 있어서,

제1 데이터 소스로부터 적어도 하나의 진단 코드를 포함하는 제1 의료 코드를 획득하고, 상기 제1 데이터 소스와 상이한 제2 데이터 소스로부터 적어도 하나의 진단 코드를 포함하는 제2 의료 코드를 획득하는 단계;

상기 제1 의료 코드 및 상기 제2 의료 코드 중 기준 진단 코드를 식별하는 단계;

상기 제1 의료 코드 및 상기 제2 의료 코드로부터 각각 제1 이벤트 벡터 및 제2 이벤트 벡터를 획득하는 단계;

상기 진단 코드를 기초로 상기 제1 이벤트 벡터 및 제2 이벤트 벡터를 통합 벡터 공간으로 변환하는 단계; 및

상기 통합 벡터 공간 및 상기 기준 진단 코드를 기초로 상기 제1 의료 코드 및 상기 제2 의료 코드와 관련된 환자의 발병 확률을 예측하는 단계; 를 포함하는 데이터 연합 기반의 질병 진단 방법.

청구항 2

제1항에 있어서,

상기 기준 진단 코드를 식별하는 단계는 상기 제1 의료 코드 및 제2 의료 코드에 포함된 상기 진단 코드 중 중복되어 포함된 진단 코드를 상기 기준 진단 코드로 식별하는 데이터 연합 기반의 질병 진단 방법.

청구항 3

제2항에 있어서,

상기 제1 이벤트 벡터 및 제2 이벤트 벡터를 획득하는 단계는,

상기 제1 의료 코드 및 상기 제2 의료 코드를 시계열적으로 정렬하는 단계; 및

상기 시계열적으로 정렬된 제1 의료 코드 및 제2 의료 코드를 기초로 각각 제1 이벤트 시퀀스 코드 및 제2 이벤트 시퀀스 코드를 획득하는 단계;를 더 포함하고,

상기 제1 이벤트 시퀀스 코드 및 상기 제2 이벤트 시퀀스 코드를 변환 알고리즘에 입력하여 각각 상기 제1 이벤트 벡터 및 제2 이벤트 벡터를 획득하는 것인 데이터 연합 기반의 질병 진단 방법.

청구항 4

제1항에 있어서,

상기 통합 벡터 공간으로 변환하는 단계는,

상기 제1 이벤트 벡터 중 상기 진단 코드에 대응하는 제1 진단 코드 벡터 및 상기 제2 이벤트 벡터 중 상기 진단 코드에 대응하는 제2 진단 코드 벡터를 추출하는 단계;

상기 제1 진단 코드 벡터를 기초로 상기 제1 이벤트 벡터를 제1 벡터 공간으로 변환하고, 상기 제2 진단 코드 벡터를 기초로 상기 제2 이벤트 벡터를 제2 벡터 공간으로 변환하는 단계; 및

상기 제1 벡터 공간 및 상기 제2 벡터 공간을 상기 통합 벡터 공간으로 병합하는 단계;를 더 포함하는 데이터 연합 기반의 질병 진단 방법.

청구항 5

제1항에 있어서,

상기 질병을 판단하는 단계는,

상기 통합 벡터 공간 상 상기 기준 진단 코드에 대한 벡터를 식별하는 단계;

상기 제1 이벤트 벡터 및 제2 이벤트 벡터를 기초로 적어도 하나의 환자 벡터를 획득하는 단계; 및

상기 식별된 기준 진단 코드에 대한 벡터와 상기 적어도 하나의 환자 벡터를 비교하는 단계;를 더 포함하고,

상기 비교 결과를 기초로 질병 여부를 판단하는 데이터 연합 기반의 질병 진단 방법.

청구항 6

데이터 연합 기반의 질병 진단 장치에 있어서,

제1 데이터 소스로부터 적어도 하나의 진단 코드를 포함하는 제1 의료 코드를 획득하고, 상기 제1 데이터 소스와 상이한 제2 데이터 소스로부터 적어도 하나의 진단 코드를 포함하는 제2 의료 코드를 획득하는 데이터 수집부;

상기 제1 의료 코드 및 상기 제2 의료 코드 중 기준 진단 코드를 식별하는 코드 식별부;

상기 제1 의료 코드 및 상기 제2 의료 코드로부터 각각 제1 이벤트 벡터 및 제2 이벤트 벡터를 획득하는 이벤트 벡터 획득부;

상기 진단 코드를 기초로 상기 제1 이벤트 벡터 및 제2 이벤트 벡터를 통합 벡터 공간으로 변환하는 벡터 변환부; 및

상기 통합 벡터 공간 및 상기 기준 진단 코드를 기초로 상기 제1 의료 코드 및 상기 제2 의료 코드와 관련된 환자의 질병 여부를 판단하는 발병 확률 예측부;를 포함하는 데이터 연합 기반의 질병 진단 장치.

청구항 7

제1항 내지 제5항 중 어느 한 항의 상기 데이터 연합 기반의 질병 진단 방법을 실행 시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체.

청구항 8

컴퓨터를 이용하여 제1항 내지 제5항의 데이터 연합 기반의 질병 진단 방법을 실행하기 위하여 매체에 저장된 컴퓨터 프로그램.

발명의 설명

기 술 분 야

[0001] 본 발명은 시스템을 포함하여 물리적으로 서로 다른 병원에서 수집된 데이터를 통합하여 향후 폐암 발병을 예측하고, 질병의 조기 진단 서비스를 제공하는 진단 시스템 및 그 운용 방법에 관한 것이다.

배 경 기 술

[0002] 근래의 병원은 환자가 병원에 방문하여 겪는 모든 종류의 이벤트에 대하여 코드를 부여한다. 이때, 이벤트는 상태, 증상, 검진, 진단, 처방 등이 있다. 병원은 각각의 이벤트에 대하여 이벤트 코드를 부여한다. 한편, 서로 다른 병원의 경우, 병원이 환자에 대한 이벤트에 부여하는 이벤트 코드가 서로 상이하다.

[0003] 종래에는 병원마다 이벤트 코드가 상이하기 때문에, 서로 다른 병원에 방문한 동일한 환자에 대한 이벤트를 통합하는데 문제다. 동일한 환자가 다른 병원에서 수행한 이벤트를 즉각적으로 활용할 수 없고, 병원은 동일한 이벤트를 재수행하는 등 환자에 대한 효과적이고 신속한 질병 예측이 어렵다는 문제가 있어왔다.

발명의 내용

해결하려는 과제

[0004] 본 발명은 상술한 문제를 해결하기 위한 필요성에 따라, 서로 다른 병원의 이종 시스템에 저장된 환자 이벤트 데이터를 효과적으로 통합하고 관리할 수 있는 시스템을 제공하고자 한다.

[0005] 또한, 본 발명은 통합된 환자의 정보를 이용하여 보다 신속하고, 정확도가 향상된 질병 진단 시스템을 제공하고

자 한다.

과제의 해결 수단

[0006] 본 발명의 데이터 연합 기반의 질병 진단 방법은 제1 데이터 소스로부터 적어도 하나의 진단 코드를 포함하는 제1 의료 코드를 획득하고, 상기 제1 데이터 소스와 상이한 제2 데이터 소스로부터 적어도 하나의 진단 코드를 포함하는 제2 의료 코드를 획득하는 단계; 상기 제1 의료 코드 및 상기 제2 의료 코드 중 기준 진단 코드를 식별하는 단계; 상기 제1 의료 코드 및 상기 제2 의료 코드로부터 각각 제1 이벤트 벡터 및 제2 이벤트 벡터를 획득하는 단계; 상기 진단 코드를 기초로 상기 제1 이벤트 벡터 및 제2 이벤트 벡터를 통합 벡터 공간으로 변환하는 단계; 및 상기 통합 벡터 공간 및 상기 기준 진단 코드를 기초로 상기 제1 의료 코드 및 상기 제2 의료 코드와 관련된 환자의 발병 확률을 예측하는 단계; 를 포함한다.

[0007] 또한, 상기 기준 진단 코드를 식별하는 단계는 상기 제1 의료 코드 및 제2 의료 코드에 포함된 상기 진단 코드 중 중복되어 포함된 진단 코드를 상기 기준 진단 코드로 식별한다.

[0008] 또한, 상기 제1 이벤트 벡터 및 제2 이벤트 벡터를 획득하는 단계는, 상기 제1 의료 코드 및 상기 제2 의료 코드를 시계열적으로 정렬하는 단계; 및 상기 시계열적으로 정렬된 제1 의료 코드 및 제2 의료 코드를 기초로 각각 제1 이벤트 시퀀스 코드 및 제2 이벤트 시퀀스 코드를 획득하는 단계;를 더 포함하고, 상기 제1 이벤트 시퀀스 코드 및 상기 제2 이벤트 시퀀스 코드를 변환 알고리즘에 입력하여 각각 상기 제1 이벤트 벡터 및 제2 이벤트 벡터를 획득한다.

[0009] 또한, 상기 통합 벡터 공간으로 변환하는 단계는, 상기 제1 이벤트 벡터 중 상기 진단 코드에 대응하는 제1 진단 코드 벡터 및 상기 제2 이벤트 벡터 중 상기 진단 코드에 대응하는 제2 진단 코드 벡터를 추출하는 단계; 상기 제1 진단 코드 벡터를 기초로 상기 제1 이벤트 벡터를 제1 벡터 공간으로 변환하고, 상기 제2 진단 코드 벡터를 기초로 상기 제2 이벤트 벡터를 제2 벡터 공간으로 변환하는 단계; 및 상기 제1 벡터 공간 및 상기 제2 벡터 공간을 상기 통합 벡터 공간으로 병합하는 단계;를 더 포함한다.

[0010] 또한, 상기 질병을 판단하는 단계는, 상기 통합 벡터 공간 상 상기 기준 진단 코드에 대한 벡터를 식별하는 단계; 및 상기 제1 이벤트 벡터 및 제2 이벤트 벡터를 기초로 적어도 하나의 환자 벡터를 획득하는 단계; 및 상기 식별된 기준 진단 코드에 대한 벡터와 상기 적어도 하나의 환자 벡터를 비교하는 단계;를 더 포함하고, 상기 비교 결과를 기초로 질병 여부를 판단한다.

[0011] 본 발명의 데이터 연합 기반의 질병 진단 장치는 제1 데이터 소스로부터 적어도 하나의 진단 코드를 포함하는 제1 의료 코드를 획득하고, 상기 제1 데이터 소스와 상이한 제2 데이터 소스로부터 적어도 하나의 진단 코드를 포함하는 제2 의료 코드를 획득하는 데이터 수집부; 상기 제1 의료 코드 및 상기 제2 의료 코드 중 기준 진단 코드를 식별하는 코드 식별부; 상기 제1 의료 코드 및 상기 제2 의료 코드로부터 각각 제1 이벤트 벡터 및 제2 이벤트 벡터를 획득하는 이벤트 벡터 획득부; 상기 진단 코드를 기초로 상기 제1 이벤트 벡터 및 제2 이벤트 벡터를 통합 벡터 공간으로 변환하는 벡터 변환부; 및 상기 통합 벡터 공간 및 상기 기준 진단 코드를 기초로 상기 제1 의료 코드 및 상기 제2 의료 코드와 관련된 환자의 질병 여부를 판단하는 발병 확률 예측부;를 포함한다.

[0012] 본 발명의 일 실시예에 따른 기록매체는 상술한 데이터 연합 기반의 질병 진단 방법을 실행 시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체일 수 있다.

[0013] 또한, 본 발명의 일 실시예에 따른 프로그램은 컴퓨터를 이용하여 상술한데이터 연합 기반의 질병 진단 방법을 실행하기 위하여 매체에 저장된 컴퓨터 프로그램일 수 있다.

발명의 효과

[0014] 본 발명의 시스템에 의하면, 물리적으로 서로 다른 병원의 이종 데이터를 프라이버시 누출 없이 통합하여 질병 발병 예측의 정확도를 향상시킬 수 있다. 즉, 본 발명은 병원간에 공유할 수 있는 검진 데이터 (예: CT 영상)를 통합하여 사용함으로써 질병 예측의 정확도를 향상시킬 수 있다.

[0015] 또한, 본 발명은 하나의 병원 데이터에만 적용하더라도, 환자의 시간의 흐름에 따라 수집된 시퀀스 데이터를 활용함으로써, 종래의 연구 방식인 특정 시간대에 수집된 정적인 데이터를 통해 구성한 예측 모형에 비해 우수한 성능을 제공할 수 있다.

[0016] 나아가, 본 발명은 폐암을 포함하는 다양한 암 및 희귀암 진단 및 예측, 고혈압/합병증 예방, 건강관리 서비스

제공, 신약후보물질 생성 모델 개발 및 공공건강 DB 구축에 활용될 수 있다.

도면의 간단한 설명

[0017] 도 1은 본 발명의 일 실시예에 따른 진단 시스템을 설명하기 위한 시스템도이다.

도 2는 본 발명의 일 실시예에 따른 서버의 구성요소를 설명하기 위한 블록도이다.

도 3은 본 발명의 일 실시예에 따른 이벤트 시퀀스 코드를 예시적으로 도시한다.

도 4는 본 발명의 일 실시예에 따른 서버가 이벤트 벡터를 획득하는 것을 도시한 도면이다.

도 5a는 제1 병원의 의료 코드에 포함된 진단 코드를 기초로 획득한 이벤트 벡터의 벡터 공간을 도시하고, 도 5b는 제2 병원의 의료 코드에 포함된 진단 코드를 기초로 획득한 이벤트 벡터의 벡터 공간을 도시한다.

도 6은 본 발명의 일 실시예에 따른 통합 벡터 공간에 진단 코드에 대응하는 이벤트 벡터가 사영된 것을 도시한다.

도 7은 본 발명의 일 실시예에 따른 서버가 통합 벡터 공간을 기초로 환자의 발병 확률을 예측하는 것을 도시한다.

발명을 실시하기 위한 구체적인 내용

[0018] 본 발명은 다양한 변환을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세한 설명에 상세하게 설명하고자 한다. 본 발명의 효과 및 특징, 그리고 그것들을 달성하는 방법은 도면과 함께 상세하게 후술되어 있는 실시예들을 참조하면 명확해질 것이다. 그러나 본 발명은 이하에서 개시되는 실시예들에 한정되는 것이 아니라 다양한 형태로 구현될 수 있다.

[0019] 이하, 첨부된 도면을 참조하여 본 발명의 실시예들을 상세히 설명하기로 하며, 도면을 참조하여 설명할 때 동일하거나 대응하는 구성 요소는 동일한 도면부호를 부여하고 이에 대한 중복되는 설명은 생략하기로 한다.

[0020] 이하의 실시예에서, 제1, 제2 등의 용어는 한정적인 의미가 아니라 하나의 구성 요소를 다른 구성 요소와 구별하는 목적으로 사용되었다. 이하의 실시예에서, 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 이하의 실시예에서, 포함하다 또는 가지다 등의 용어는 명세서상에 기재된 특징, 또는 구성요소가 존재함을 의미하는 것이고, 하나 이상의 다른 특징들 또는 구성요소가 부가될 가능성을 미리 배제하는 것은 아니다. 도면에서는 설명의 편의를 위하여 구성 요소들이 그 크기가 과장 또는 축소될 수 있다. 예컨대, 도면에서 나타난 각 구성의 크기 및 형태는 설명의 편의를 위해 임의로 나타내었으므로, 본 발명이 반드시 도시된 바에 한정되지 않는다.

[0021] 도 1은 본 발명의 일 실시예에 따른 진단 시스템을 설명하기 위한 시스템도이다.

[0022] 도 1을 참조하면, 본 발명의 진단 시스템(1)은 적어도 하나의 병원(100) 및 서버(200)를 포함할 수 있다.

[0023] 적어도 하나의 병원(100)은 환자에 대한 의료 코드와 같은 의료 데이터를 저장 및 관리하는 데이터 소스일 수 있다. 이때, 환자에 대한 의료 데이터는 EMR(Electronice Medical Records) 데이터, 이미지 데이터, 바이오센서(Biosensor) 데이터, 오믹스(Omics) 데이터 및 행동(Behavioral) 데이터 등 다양한 데이터를 포함할 수 있다.

[0024] 이때, EMR 데이터는 명목형 또는 숫자형 데이터, 정형 또는 비정형 데이터, 정적 또는 동적 데이터 등으로 각각 상이하게 분류될 수 있고, 기입의 차이에 따라 일관성 없는 데이터로 분류될 수 있다. 이미지 데이터는 행렬 형식 및 크기에 따라 상이하게 분류될 수 있다. 바이오 센서 데이터는 시계열적인 데이터로써 간헐적으로 수집되기 때문에 노이즈가 크도록 측정 및 분류될 수 있다. 오믹스 데이터는 명목형 또는 숫자형 데이터로 상이하게 분류될 수 있고, 행동 데이터는 표준화되어 있지 않아 결측치가 높게 분류될 수 있다.

[0025] 즉, 제1 병원(110) 및 제2 병원(120)은 각각의 의료 데이터에 대한 분류가 상이할 수 있다. 특히, 제1 병원(110) 및 제2 병원(120)은 동일한 환자가 내방하여 동일한 이벤트가 발생하더라도 이에 대하여 서로 상이한 이벤트 코드로 분류할 수 있다. 이때, 이벤트의 종류로는 상태, 증상, 검진, 진단, 처방 등이 있으나, 이에 한정하지 않는다.

[0026] 본 발명의 일 실시예에 따른 서버(200)는 적어도 하나의 병원(100)으로부터 의료 데이터를 수신할 수 있다. 서버(200)는 수신한 의료 데이터로부터 환자별 이벤트 시퀀스 코드를 수집할 수 있다. 이때, 이벤트 시퀀스 코드

란 특정 환자가 적어도 하나의 병원(100) 각각에 내방하여 발생한 모든 의료 코드를 시계열로 정렬하여 획득한 데이터일 수 있다.

[0027] 한편, 의료 코드 중 진단 코드는 ICD 9 또는 10으로 전세계적으로 통일되어있지만, 대부분의 의료 코드의 경우 병원마다 다르게 지정할 수 있다. 이에 따라, 서버(200)는 적어도 하나의 병원(100)으로부터 수집하여 획득한 이벤트 시퀀스 코드 중 진단 코드를 기초로 환자의 질병 예측 진단을 수행할 수 있다. 이에 대하여 추후에 추가적으로 설명하기로 한다.

[0028] 도 2는 본 발명의 일 실시예에 따른 서버의 구성요소를 설명하기 위한 블록도이다.

[0029] 도 2를 참조하면, 본 발명의 서버(200)는 데이터 수집부(210), 코드 식별부(220), 이벤트 벡터 획득부(230), 벡터 변환부(240) 및 발병 확률 예측부(250)를 포함할 수 있다.

[0030] 데이터 수집부(210)는 적어도 하나의 병원(100)으로부터 적어도 하나의 환자에 대한 의료 데이터를 수집하기 위한 구성이다. 이때, 의료 데이터는 적어도 하나의 진단 코드를 포함할 수 있다.

[0031] 데이터 수집부(210)는 적어도 하나의 병원(100)과 통신하기 위한 통신 수단을 포함하여 구현된 것일 수 있다. 이때, 통신 수단(미도시)는 다양한 유형의 통신방식에 따라 다양한 유형의 외부 장치와 통신을 수행하는 구성일 수 있다. 통신 수단은 와이파이칩, 블루투스 칩, 무선 통신 칩 중 적어도 하나를 포함할 수 있다. 특히, 무선통신 칩은 IEEE, Zigbee, 3G(3rd Generation), 3GPP(3rd Generation Partnership Project), LTE(Long Term Evolution), 5G(Fifth Generation)등과 같은 다양한 통신 규격에 따라 통신을 수행하는 칩을 의미한다. 5G 통신 규격에 따른 무선 통신 칩은 3.5GHz 등 주파수(Below 6GHz) 대역뿐만 아니라, 26, 28, 38, 39, 60 GHz 등의 밀리미터파(mmWave), 즉 초고주파 대역(Above 6GHz)까지 사용할 수 있다.

[0032] 코드 식별부(220)는 데이터 수집부(210)에서 수집한 의료 코드 중 기준 진단 코드를 식별하기 위한 구성일 수 있다. 이때, 기준 진단 코드는 복수의 병원(또는 데이터 소스)로부터 수신한 의료 코드에 포함된 진단 코드에 중복되어 포함된 진단 코드일 수 있다.

[0033] 이벤트 벡터 획득부(230)는 데이터 수집부(210)에서 수신한 의료 코드를 기초로 이벤트 벡터를 획득하기 위한 구성이다. 구체적으로, 이벤트 벡터 획득부(220)는 이벤트 데이터를 시계열적으로 정렬하여 이벤트 시퀀스 코드를 획득할 수 있다. 이벤트 데이터 획득부(220)는 획득한 이벤트 시퀀스 코드에 변환 알고리즘을 적용하여 이벤트 벡터를 획득할 수 있다.

[0034] 본 발명의 변환 알고리즘은 Word2Vec을 이용할 수 있으나, 이에 한정하지 않는다. 이때, 이벤트 벡터는 이벤트 코드에 대한 숫자형 벡터일 수 있다.

[0035] 본 발명의 일 실시예에 따른 벡터 변환부(240)는 진단 코드를 기초로 이벤트 벡터를 통합 벡터 공간으로 변환하기 위한 구성일 수 있다. 벡터 변환부(240)는 이벤트 벡터 중 진단 코드에 대응하는 진단 코드 벡터를 추출할 수 있다.

[0036] 벡터 변환부(240)는 진단 코드 벡터를 기초로 이벤트 벡터를 벡터 공간으로 변환할 수 있다. 본 발명의 일 실시예에 따르면, 데이터 소스가 2개인 경우, 벡터 변환부(240)는 제1 이벤트 벡터 중 진단 코드에 대응하는 제1 진단 코드 벡터 및 제2 이벤트 벡터 중 진단 코드에 대응하는 제2 진단 코드 벡터를 추출할 수 있다. 이후, 벡터 변환부(240)는 제1 진단 코드 벡터를 기초로 제1 이벤트 벡터를 제1 벡터 공간으로 변환하고, 제2 진단 코드 벡터를 기초로 제2 이벤트 벡터를 제2 벡터 공간으로 변환할 수 있다. 벡터 변환부(240)는 제1 벡터 공간 및 제2 벡터 공간을 통합 벡터 공간으로 병합할 수 있다.

[0037] 구체적으로, 벡터 변환부(240)는 이벤트 벡터 획득부(230)가 획득한 이벤트 벡터를 기초로 매핑(mapping) 함수 또는 사상 함수를 획득할 수 있다. 이때, 매핑 함수는 적어도 2개의 이벤트 시퀀스 코드로부터 획득한 이벤트 벡터의 행렬을 기초로 획득된 것일 수 있다.

[0038] 본 발명의 일 실시예에 따른 매핑 함수는 상이한 이벤트 시퀀스 코드에대응하는 이벤트 벡터를 하나의 벡터 공간에 사영함으로써, 상이한 의료 코드를 하나로 통합하기 위한 목적 함수일 수 있다. 즉, 이벤트 벡터(또는 이벤트 벡터 행렬)가 있을 때, 그 이벤트 벡터가 속하는 벡터 공간에서 이벤트 벡터를 매개변수로 갖는 함수의 선형계획문제의 목적 함수가 상기 매핑 함수일 수 있다.

[0039] 본 발명의 일 실시예에 따른 발병 확률 예측부(250)는 통합 벡터 공간 및 기준 진단 코드를 기초로 의료 코드와 관련된 환자의 발병 확률을 예측하기 위한 구성이다.

[0040] 본 발명의 발병 확률 예측부(250)는 통합 벡터 공간 상 기준 진단 코드에 대한 벡터를 식별하고, 식별된 기준 진단 코드에 대한 벡터와 이벤트 벡터를 기초로 획득한 환자 벡터를 비교할 수 있다. 이후, 발병 확률 예측부(250)는 비교 결과를 기초로 질병 여부를 판단할 수 있다. 이에 대하여 추후에 자세히 설명하기로 한다.

[0041] 도 3은 본 발명의 일 실시예에 따른 이벤트 시퀀스 코드를 예시적으로 도시한다.

[0042] 본 발명의 일 실시예에 따른 서버(200)는 병원(100)으로부터 수신한 의료데이터에 포함된 의료 코드를 기초로 이벤트 시퀀스 코드를 획득할 수 있다.

[0043] 도 3에 도시된 이벤트 시퀀스 코드는 병원에서 수행된 모든 이벤트에 대한 의료 코드(예로, l_509701, l_50971, l_50820 등)가 시계열적으로 정렬된 것일 수 있다.

[0044] 본 발명의 일 실시예에 따르면, 서버(200)는 의료 코드 중 진단 코드가 별도로 구분된 이벤트 시퀀스 코드를 획득할 수 있다. 도 3의 이벤트 시퀀스 코드는 진단 코드(l_51249 등)가 Prediction window로 분류된 것일 수 있다.

[0045] 도 4는 본 발명의 일 실시예에 따른 서버가 이벤트 벡터를 획득하는 것을 도시한 도면이다.

[0046] 제1 병원에 n_1명의 환자가 있다면, 서버(200)는 n_1개의 이벤트 시퀀스 코드를 획득할 수 있다. 마찬가지로, 제2 병원에 명의 n_2환자가 있다면, 서버(200)는 n_2개의 이벤트 시퀀스 코드를 획득할 수 있다.

[0047] 본 발명의 서버(200)는 제1 병원에서 획득한 적어도 하나의 제1 의료 시퀀스 코드(410)를 변환 알고리즘을 통해 제1 이벤트 벡터 행렬(411)로 변환할 수 있다. 도 4의 예에서는, 제1 이벤트 벡터 행렬(411)는 n_A행의 행렬일 수 있다.

[0048] 마찬가지로, 서버(200)는 제2 병원에서 획득한 적어도 하나의 제2 의료 시퀀스 코드(420)를 변환 알고리즘을 통해 제2 이벤트 벡터 행렬(421)로 변환할 수 있다. 제2 이벤트 벡터 행렬(421)는 n_B행의 행렬일 수 있다.

[0049] 본 발명의 변환 알고리즘은 각각의 이벤트 시퀀스 코드에 포함된 의료 코드의 전후 순서를 기초로 이벤트 벡터가 출력한다. 이때, 본 발명의 변환 알고리즘은 Word2Vec일 수 있으나, 이에 한정하지 않는다.

[0050] 본 발명의 일 실시예에 따른 변환 알고리즘은 환자의 이벤트 시퀀스 코드에 대해 각 이벤트에 대응하는 시간을 반영하여, 이를 중첩하여 환자에 대한 이벤트 벡터를 도출할 수 있다.

[0051] 본 발명의 변환 알고리즘은 변환 시 매번 새로운 이벤트 벡터를 출력하도록 구현된 것일 수 있다. 따라서, 병원별로 도출된 의료 시퀀스 코드에 대한 이벤트 벡터 행렬은 각기 다른 벡터 공간에 위치하게 되고 이를 함께 활용할 수 없다.

[0052] 이에 따라, 본 발명의 서버(200)는 두 이벤트 벡터 행렬을 하나의 공간으로 사상시킬 수 있다. 전세계적으로 진단 코드는 통일되어 있기 때문에, 서버(200)는 제1 병원의 이벤트 시퀀스 코드에 포함된 진단 코드와 제2 병원의 이벤트 시퀀스 코드에 포함된 진단 코드를 활용할 수 있다.

[0053] 서버(200)는 제1 병원의 진단 코드들에 대한 이벤트 벡터들을 기초로 제2 쪽 병원의 진단 코드들에 대한 이벤트 벡터로의 사영을 통해 사상 함수를 학습할 수 있다. 본 발명의 서버(200)가 학습할 사상 함수의 목적 함수는 아래의 수학식 1과 같다.

수학식 1

[0054] $$\min_{Q,k} \|(A - \mu_A 1_n^T) - kQ(B - \mu_B 1_n^T)\|_F$$

[0055] 이때, A, B는 각각 제1 병원의 진단 코드 및 제2 병원의 진단 코드들에 대한 진단 코드 벡터 행렬을 의미하고, μ_A, μ_B는 각 데이터의 변수에 대한 평균 벡터를 의미한다. 수학식 1의 목적 함수를 최적화하여 구해진 Q, k 값은 아래의 수학식 2와 같다.

수학식 2

$$(A - \mu_A 1_n^T)(B - \mu_B 1_n^T) = U\Sigma V^T$$
$$Q = UV^T, k = trace(\Sigma)/trace(BB^T)$$

[0056]

[0057] 서버(200)는 의료 코드 중 진단 코드가 아닌 코드의 이벤트 벡터에 대하여 수학식 2를 통해 구해진 Q와 k를 기초로 사영을 시킬 수 있다. 이 경우, 제1 병원 및 제2 병원의 모든 이벤트 시퀀스 코드들에 대한 이벤트 벡터들이 하나의 공간에 위치할 수 있다.

[0058] 도 5a는 제1 병원의 의료 코드에 포함된 진단 코드를 기초로 획득한 이벤트 벡터의 벡터 공간을 도시하고, 도 5b는 제2 병원의 의료 코드에 포함된 진단 코드를 기초로 획득한 이벤트 벡터의 벡터 공간을 도시한다.

[0059] 특히, 도 5a 및 도 5b는 제1 병원 및 제2 병원에 대응하는 이벤트 시퀀스 코드 중 진단 코드에 대한 이벤트 벡터를 도시한다.

[0060] 상술한 바와 같이, 본 발명의 서버(200)는 수학식 1 및 수학식 2를 통해 제1 병원 및 제2 병원에 대응하는 이벤트 시퀀스 코드 중 단 코드가 아닌 코드의 이벤트 벡터에 대하여 사영할 수 있고, 통합 벡터 공간으로 병합할 수 있다.

[0061] 도 6은 본 발명의 일 실시예에 따른 통합 벡터 공간에 진단 코드에 대응하는 이벤트 벡터가 사영된 것을 도시한다.

[0062] 도 6을 참조하면, 제1 병원의 진단 코드 벡터 및 제2 병원 진단 코드 벡터 중 중복하여 포함된 진단 코드 벡터를 확인할 수 있다. 예를 들어, cirrhosis, hepatitis-c, Renal-failure, myocardial, lungs 등의 진단 코드는 제1 병원의 이벤트 시퀀스 코드 및 제2 병원의 이벤트 시퀀스 코드에 중복하여 포함된 진단 코드로 기준 진단 코드일 수 있다.

[0063] 본 발명의 일 실시예에 따른 서버(200)는 상술한 기준 진단 코드를 기초로 환자의 발병 확률을 예측할 수 있다.

[0064] 도 7은 본 발명의 일 실시예에 따른 서버가 통합 벡터 공간을 기초로 환자의 발병 확률을 예측하는 것을 도시한다.

[0065] 도 7을 참조하면, 서버(200)는 환자에 대한 이벤트 벡터와 기준 진단 코드에 대한 이벤트 벡터 사이의 코사인 유사도를 구할 수 있고, 이를 각 환자가 폐암에 발병할 확률로 판단할 수 있다.

[0066] 예를 들어, 서버(200)는 적어도 하나의 병원(100)을 통해 수신한 의료 코드를 통해 획득한 이벤트 벡터를 기초로 환자 벡터를 획득할 수 있다. 이때, 환자 벡터는 제1 이벤트 벡터 및 제2 이벤트 벡터를 환자 별로 대응시켜 재구성하여 획득한 벡터일 수 있다.

[0067] 환자 벡터가 $\vec{S}$ 이고, 기준 진단 질병 폐암(Lung Caner)에 대한 기준 진단 벡터가 $\vec{LC}$ 인 경우, 서버(200)는 수학식 3을 통해 획득한 코사인 유사도 값을 기초로 폐암(Lung Cancer)의 발병 확률을 예측할 수 있다.

수학식 3

$$\frac{\vec{S} \cdot \vec{LC}}{\|\vec{S}\| \cdot \|\vec{LC}\|}$$

[0068]

[0069] 이는 일예에 불과하고, 본 발명의 서버(200)는 폐암 진단명 코드가 아닌 다른 진단명 코드에 대해서도 코사인 거리를 구하여 발병 확률을 예측할 수 있다.

[0070] 본 개시에서는 데이터 소스로써, 병원(100)이 제1 병원 및 제2 병원으로 2개인 경우를 상정하여 설명하였으나, 병원(100)은 2개 초과하는 임의의 수일 수 있음은 물론이다.

[0072] 이상 설명된 본 발명에 따른 실시예는 컴퓨터 상에서 다양한 구성요소를 통하여 실행될 수 있는 컴퓨터 프로그램의 형태로 구현될 수 있으며, 이와 같은 컴퓨터 프로그램은 컴퓨터로 판독 가능한 매체에 기록될 수 있다. 이때, 매체는 컴퓨터로 실행 가능한 프로그램을 저장하는 것일 수 있다. 매체의 예시로는, 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체, CD-ROM 및 DVD와 같은 광기록 매체, 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical medium), 및 ROM, RAM, 플래시 메모리 등을 포함하여 프로그램 명령어가 저장되도록 구성된 것이 있을 수 있다.

[0073] 한편, 상기 컴퓨터 프로그램은 본 발명을 위하여 특별히 설계되고 구성된 것이거나 컴퓨터 소프트웨어 분야의 당업자에게 공지되어 사용 가능한 것일 수 있다. 컴퓨터 프로그램의 예에는, 컴파일러에 의하여 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용하여 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드도 포함될 수 있다.

[0074] 본 발명에서 설명하는 특정 실행들은 일 실시 예들로서, 어떠한 방법으로도 본 발명의 범위를 한정하는 것은 아니다. 명세서의 간결함을 위하여, 종래 전자적인 구성들, 제어 시스템들, 소프트웨어, 상기 시스템들의 다른 기능적인 측면들의 기재는 생략될 수 있다. 또한, 도면에 도시된 구성 요소들 간의 선들의 연결 또는 연결 부재들은 기능적인 연결 및/또는 물리적 또는 회로적 연결들을 예시적으로 나타낸 것으로서, 실제 장치에서는 대체 가능하거나 추가의 다양한 기능적인 연결, 물리적인 연결, 또는 회로 연결들로서 나타내어질 수 있다. 또한, "필수적인", "중요하게" 등과 같이 구체적인 언급이 없다면 본 발명의 적용을 위하여 반드시 필요한 구성 요소가 아닐 수 있다.

[0075] 따라서, 본 발명의 사상은 상기 설명된 실시예에 국한되어 정해져서는 아니 되며, 후술하는 특허청구범위뿐만 아니라 이 특허청구범위와 균등한 또는 이로부터 등가적으로 변경된 모든 범위는 본 발명의 사상의 범주에 속한다고 할 것이다.

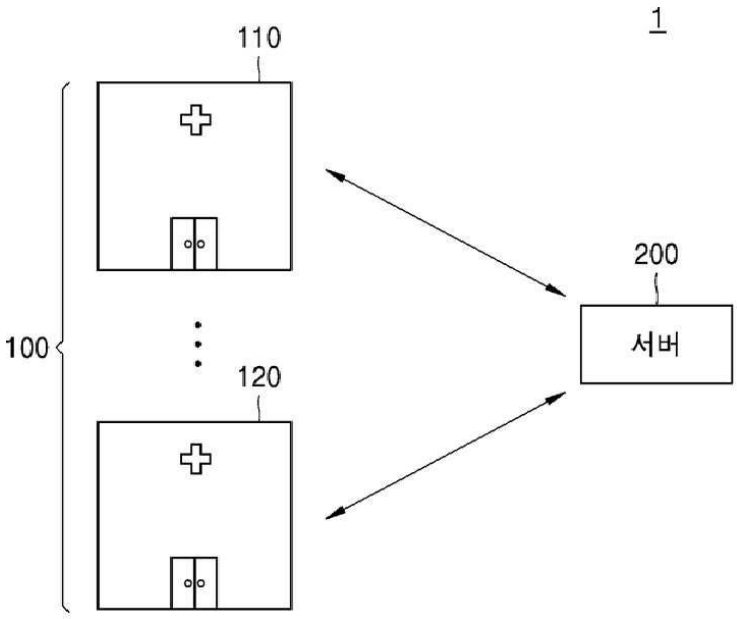
부호의 설명

[0076] 100: 병원

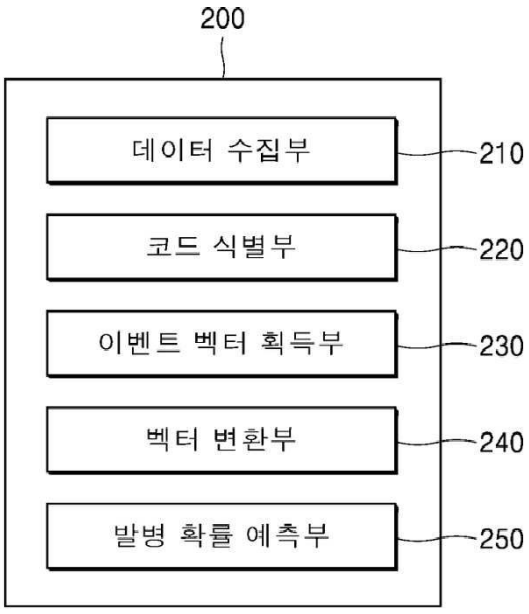
200: 서버

도면

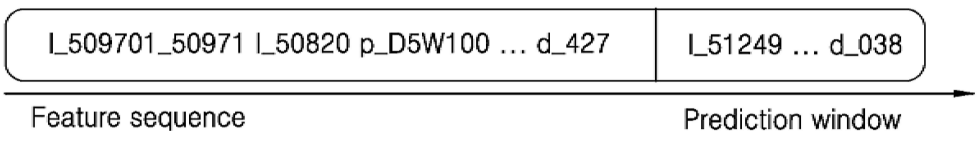
도면1



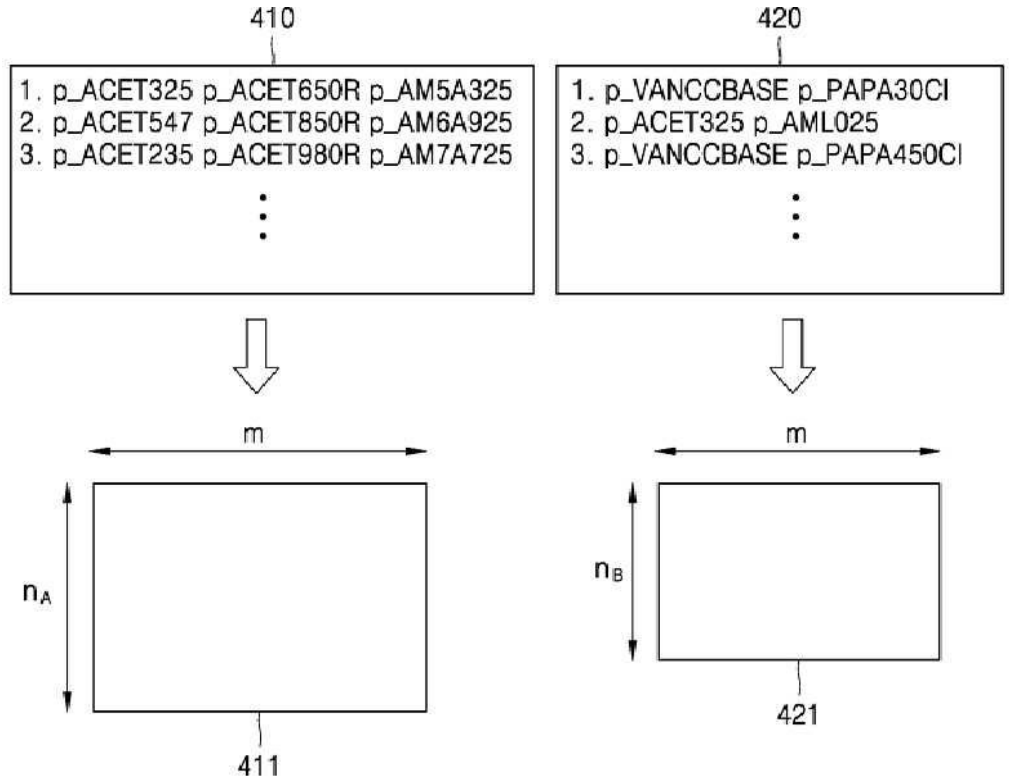
도면2



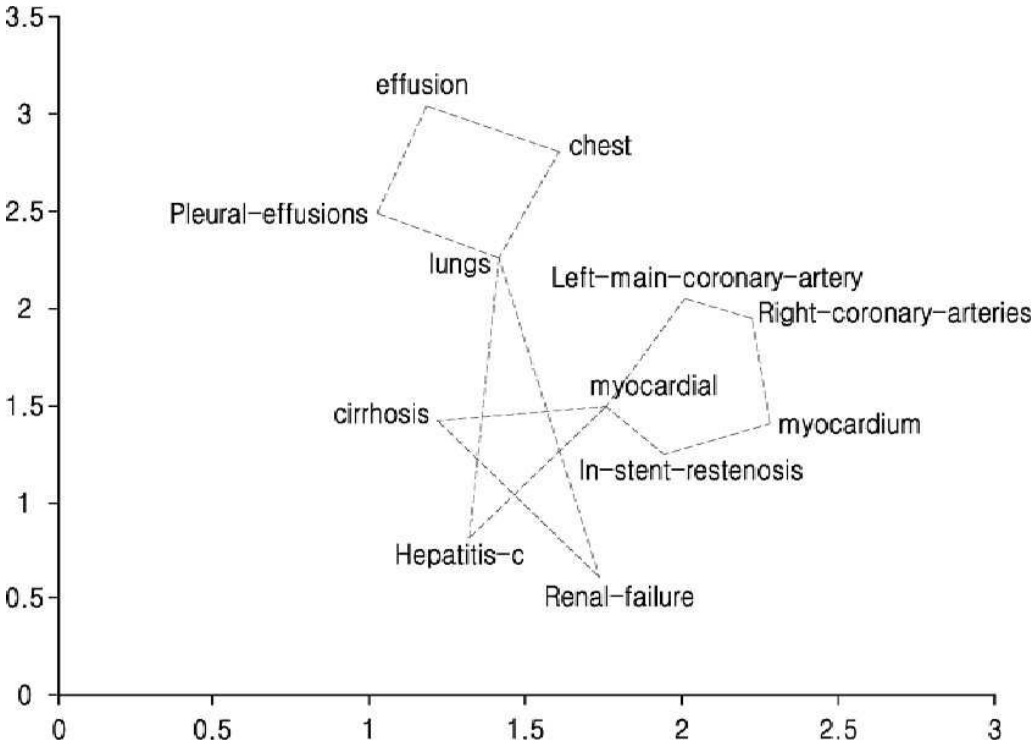
도면3



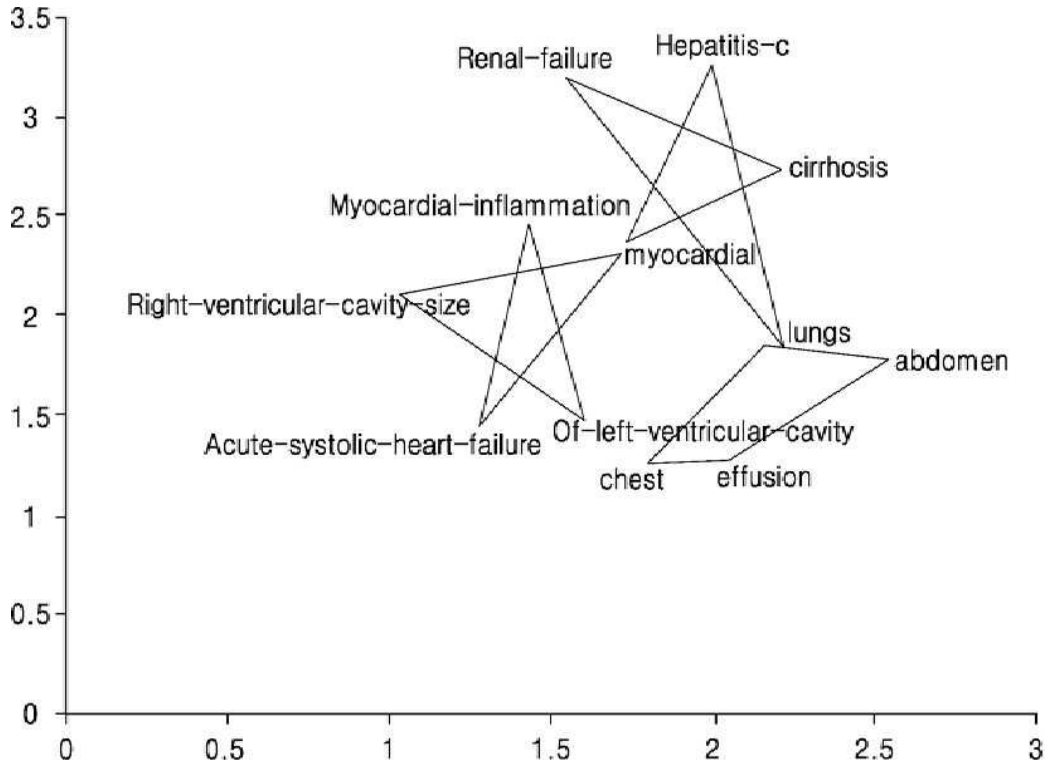
도면4


410
1. p_ACET325 p_ACET650R p_AM5A325
2. p_ACET547 p_ACET850R p_AM6A925
3. p_ACET235 p_ACET980R p_AM7A725
420
1. p_VANCCBASE p_PAPA30CI
2. p_ACET325 p_AML025
3. p_VANCCBASE p_PAPA450CI
m
m
n_A
n_B
411
421

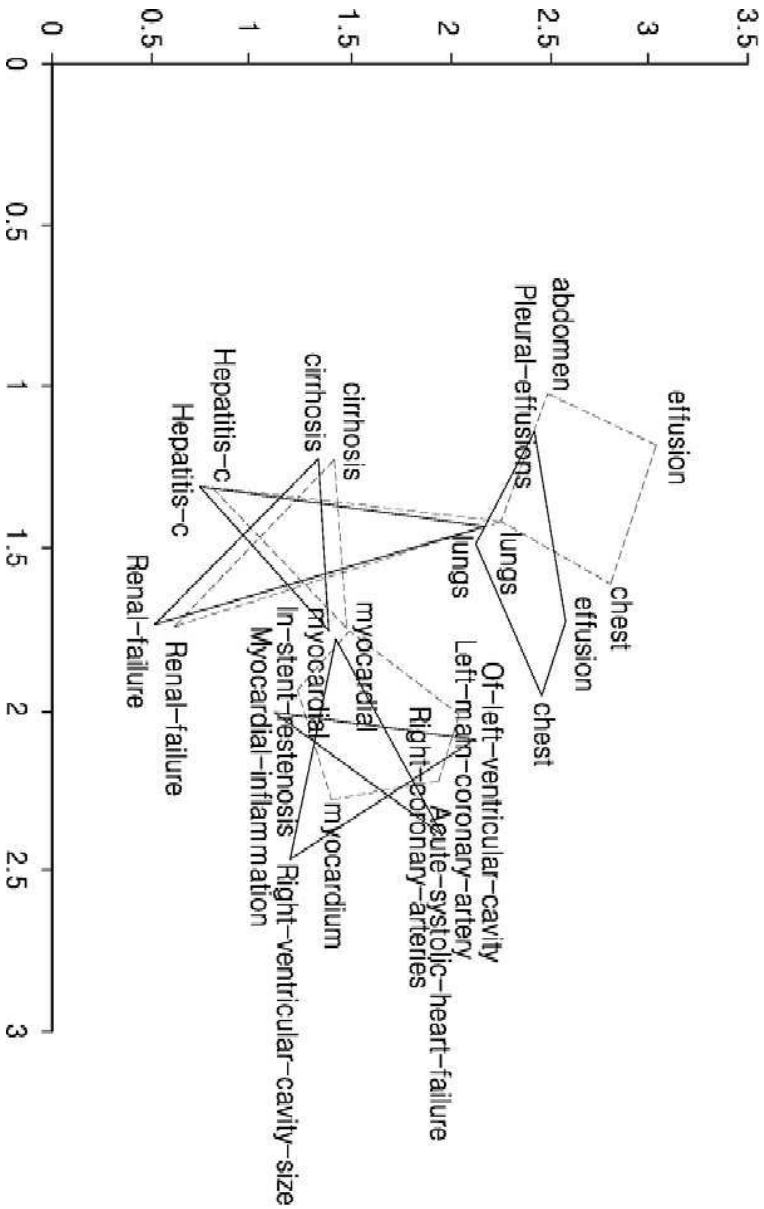
도면5a


3.5
3
2.5
2
1.5
1
0.5
0
0
0.5
1
1.5
2
2.5
3
effusion
chest
Pleural-effusions
lungs
Left-main-coronary-artery
Right-coronary-arteries
myocardial
cirrhosis
myocardium
In-stent-restenosis
Hepatitis-c
Renal-failure

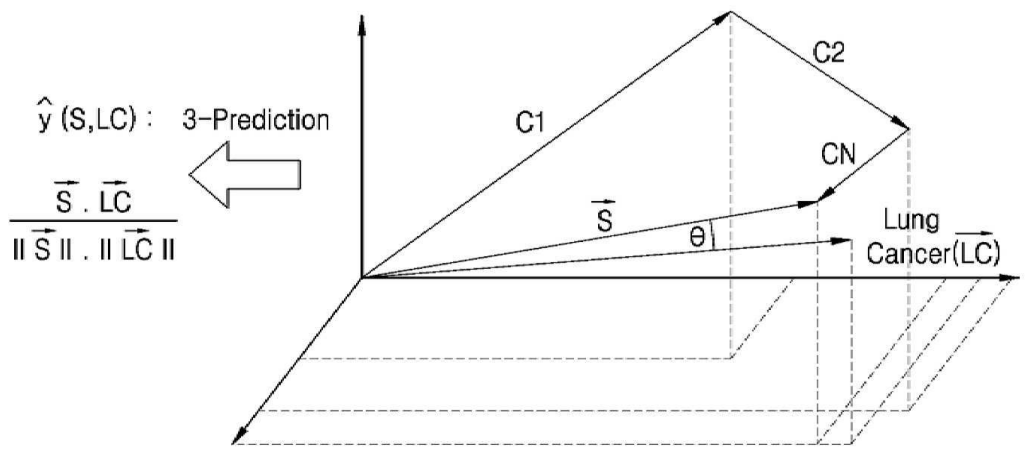
도면5b


Renal-failure
Hepatitis-c
cirrhosis
Myocardial-inflammation
myocardial
Right-ventricular-cavity-size
lungs
abdomen
Acute-systolic-heart-failure
Of-left-ventricular-cavity
chest
effusion
3.5
3
2.5
2
1.5
1
0.5
0
0
0.5
1
1.5
2
2.5
3

도면6


0
0.5
1
1.5
2
2.5
3
3.5
abdomen
Pleural-effusions
effusion
chest
effusion
chest
Hepatitis-c
Hepatitis-c
cirrhosis
cirrhosis
lungs
lungs
Renal-failure
Renal-failure
myocardial
myocardial
In-stent-restenosis
Myocardial-inflammation
myocardium
Right-ventricular-cavity-size
Of-left-ventricular-cavity
Left-main-coronary-artery
Right-coronary-arteries
Acute-systolic-heart-failure

도면7


ŷ (S,LC) : 3-Prediction
S . LC
‖S‖ . ‖LC‖
C1
C2
CN
S
θ
Lung
Cancer(LC)