(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2022년03월17일
(11) 등록번호 10-2375755
(24) 등록일자 2022년03월14일

(51) 국제특허분류(Int. Cl.)
G06Q 50/20 (2012.01) *G06F 16/2458* (2019.01)
G06F 16/9035 (2019.01) *G06F 40/279* (2020.01)
(52) CPC특허분류
G06Q 50/20 (2013.01)
G06F 16/2462 (2019.01)
(21) 출원번호 10-2021-0042245
(22) 출원일자 2021년03월31일
심사청구일자 2021년03월31일
(56) 선행기술조사문헌
KR101590976 B1*
KR101754723 B1*
KR101923312 B1*
KR1020070091843 A
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
김규태
경기도 안양시 동안구 시민대로327번길 55,108동 2105호 (관양동, 평촌더샵센트럴시티)
문기범
서울특별시 도봉구 덕릉로60카길 38, 2층(창동)
(뒷면에 계속)
(74) 대리인
윤귀상

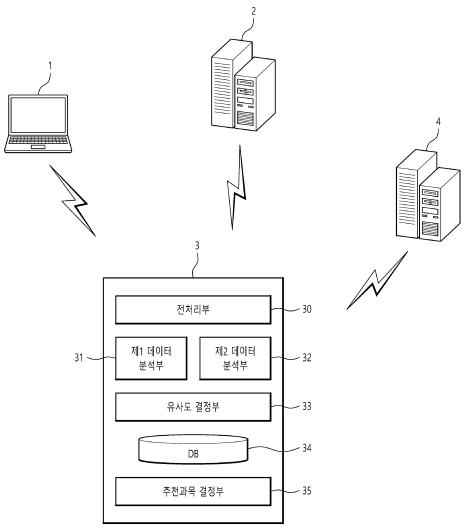
전체 청구항 수 : 총 11 항 심사관 : 김종태

(54) 발명의 명칭 **과목 유사도에 기반한 과목 추천 시스템과 방법 및 이를 위한 컴퓨터 프로그램**

(57) 요 약

과목 유사도에 기반한 과목 추천 시스템은, 하나 이상의 과목에 연관된 텍스트 데이터를 분석함으로서 상기 하나 이상의 과목 각각에 대한 강의내용 정보를 생성하도록 구성된 제1 데이터 분석부; 과목별 수강생의 수강이력 데이터를 분석함으로써 상기 하나 이상의 과목 각각에 대한 수강생 특성 정보를 생성하도록 구성된 제2 데이터 분석부; 상기 강의내용 정보 및 상기 수강생 특성 정보를 이용하여 과목별 유사도 정보를 생성하도록 구성된 유사도 결정부; 및 사용자로부터 수신된 관심 과목 정보 및 상기 유사도 정보에 기초하여 상기 사용자에 대한 과목 추천 정보를 생성하도록 구성된 추천과목 결정부를 포함할 수 있다. 상기 과목 추천 시스템을 이용하면, 강의내용 차원과 수강생 특성을 동시에 고려하여 과목 간의 유사도를 판별함으로써 종래에 비해 높은 정확도로 과목을 추천할 수 있으며, 학생별 관심사에 부합하는 폭 넓고 다양한 과목을 추천할 수 있는 이점이 있다.

대 표 도 - 도1



(52) CPC특허분류
G06F 16/2465 (2019.01)
G06F 16/9035 (2019.01)
G06F 40/279 (2020.01)

(72) 발명자
한재호
서울특별시 노원구 공릉로58길 161,103동 104호(하계동, 코오롱마들마을아파트)
권혜정
서울특별시 강남구 선릉로69길 19,102동 203호(역삼동, 역삼래미안아파트)
이수강
서울특별시 마포구 월드컵북로 501,914동302호 (상암동, 상암월드컵파크9단지)
이진숙
서울특별시 종로구 난계로29가길 24,710호(숭인동, 종로 동광모닝스카이)
한수연
서울특별시 용산구 백범로81길 10,202호(원효로1가)

공지예외적용 : 있음

명 세 서

청구범위

청구항 1

과목 유사도에 기반한 과목 추천 시스템으로서,

하나 이상의 과목에 연관된 텍스트 데이터를 분석함으로서 상기 하나 이상의 과목 각각에 대한 강의내용 정보를 생성하도록 구성된 제1 데이터 분석부;

과목별 수강생의 수강이력 데이터를 분석함으로써 상기 하나 이상의 과목 각각에 대한 수강생 특성 정보를 생성하도록 구성된 제2 데이터 분석부;

상기 강의내용 정보 및 상기 수강생 특성 정보를 이용하여 과목별 유사도 정보를 생성하도록 구성된 유사도 결정부; 및

사용자로부터 수신된 관심 과목 정보 및 상기 유사도 정보에 기초하여 상기 사용자에 대한 과목 추천 정보를 생성하도록 구성된 추천과목 결정부를 포함하고,

상기 과목의 교수요목, 상기 과목의 학습목표, 학과별 커리큘럼 및 학문분야별 연구논문을 포함하는 과목 정보로부터 학문영역별 키워드를 추출하여 상기 텍스트 데이터 형식으로 전처리하는 전처리부를 더 포함하고,

상기 관심 과목 정보는 상기 사용자의 수강이력 및 수강희망과목 중 하나 이상의 정보를 포함하며,

상기 추천과목 결정부는, 상기 사용자의 수강이력 및 수강희망과목 중 하나 이상의 정보를 이용하여 상기 사용자의 선호과목을 결정하고, 상기 선호과목에 대한 유사도에 기초하여 추천과목을 결정함으로써 상기 과목 추천 정보를 생성하도록 더 구성되고,

상기 추천과목 결정부는, 상기 사용자의 수강이력만을 기초하여 상기 사용자의 선호과목을 결정하는 경우, 상기 사용자의 수강이력에서 상기 사용자가 의무적으로 수강한 과목을 배제하여 선호과목을 선별하도록 더 구성되고,

상기 추천과목 결정부는, 상기 관심 과목 정보에 상기 사용자의 수강이력 및 수강희망과목이 입력되지 않은 경우, 상기 사용자와 동일한 특성을 가지는 서로 다른 수강생의 선호과목에 기초하여 추천과목을 결정하고,

상기 사용자와 동일한 특성을 가지는 서로 다른 수강생의 선호과목은,

서로 다른 수강생이 입력한 수강희망과목 및 서로 다른 수강생들의 기존 수강이력에 기초하여 결정된 과목 중 하나 이상의 과목을 포함하는, 과목 유사도에 기반한 과목 추천 시스템.

청구항 2

삭제

청구항 3

제1항에 있어서,

상기 제1 데이터 분석부는, 상기 텍스트 데이터에 대한 단어 빈도-역 문서 빈도 기반의 분석에 의하여 상기 과목의 강의내용을 학습하고, 상기 강의내용에 대한 학습 결과에 기초하여 상기 강의내용 정보를 생성하도록 더 구성된 과목 유사도에 기반한 과목 추천 시스템.

청구항 4

제1항에 있어서,

상기 제2 데이터 분석부는, 복수의 수강생의 상기 수강이력 데이터에 기초하여 학과별 수강비율 데이터를 생성

하고, 과목별 수강생의 상기 학과별 수강비율 데이터를 이용하여 상기 수강생 특성 정보를 생성하도록 더 구성된 과목 유사도에 기반한 과목 추천 시스템.

청구항 5

제1항에 있어서,

상기 유사도 결정부는, 상기 강의내용 정보 및 상기 수강생 특성 정보에 각각 상응하는 매트릭스 간의 코사인 유사도를 이용하여 상기 과목별 유사도 정보를 생성하도록 더 구성된 과목 유사도에 기반한 과목 추천 시스템.

청구항 6

삭제

청구항 7

제1항에 있어서,

상기 추천과목 결정부는, 과목별 특성에 상응하는 복수 개의 차원에 의해 정의되며 과목 간의 유사도가 높을수록 과목들이 서로 가까이 위치하는 과목지도를 이용하여 상기 과목 추천 정보를 상기 사용자에게 제공하도록 더 구성된 과목 유사도에 기반한 과목 추천 시스템.

청구항 8

과목 추천 시스템이 하나 이상의 과목에 연관된 텍스트 데이터를 분석함으로써 상기 하나 이상의 과목 각각에 대한 강의내용 정보를 생성하는 단계;

상기 과목 추천 시스템이 과목별 수강생의 수강이력 데이터를 분석함으로써 상기 하나 이상의 과목 각각에 대한 수강생 특성 정보를 생성하는 단계;

상기 과목 추천 시스템이 상기 강의내용 정보 및 상기 수강생 특성 정보를 이용하여 과목별 유사도 정보를 생성하는 단계; 및

상기 과목 추천 시스템이, 사용자로부터 수신된 관심 과목 정보 및 상기 유사도 정보에 기초하여, 상기 사용자에 대한 과목 추천 정보를 생성하는 단계를 포함하고,

상기 과목 추천 시스템이, 상기 과목의 교수요목, 상기 과목의 학습목표, 학과별 커리큘럼 및 학문분야별 연구논문을 포함하는 과목 정보로부터 학문영역별 키워드를 추출하여 상기 텍스트 데이터 형식으로 전처리하는 단계를 더 포함하고,

상기 관심 과목 정보는 상기 사용자의 수강이력 및 수강희망과목 중 하나 이상의 정보를 포함하며,

상기 과목 추천 정보를 생성하는 단계는,

상기 과목 추천 시스템이, 상기 수강이력 및 수강희망과목 중 하나 이상의 정보를 이용하여 상기 사용자의 선호과목을 결정하는 단계; 및

상기 과목 추천 시스템이, 상기 선호과목에 대한 유사도에 기초하여 추천과목을 결정하는 단계를 포함하고,

상기 사용자의 선호과목을 결정하는 단계는,

상기 사용자의 수강이력만을 기초하여 상기 사용자의 선호과목을 결정하는 경우, 상기 사용자의 수강이력에서 상기 사용자가 의무적으로 수강한 과목을 배제하여 선호과목을 결정하고,

상기 추천과목을 결정하는 단계는,

상기 관심 과목 정보에 상기 사용자의 수강이력 및 수강희망과목이 입력되지 않은 경우, 상기 사용자와 동일한 특성을 가지는 서로 다른 수강생의 선호과목에 기초하여 추천과목을 결정하고,

상기 사용자와 동일한 특성을 가지는 서로 다른 수강생의 선호과목은,

서로 다른 수강생이 입력한 수강희망과목 및 서로 다른 수강생들의 기존 수강이력에 기초하여 결정된 과목 중 하나 이상의 과목을 포함하는, 과목 유사도에 기반한 과목 추천 방법.

청구항 9

삭제

청구항 10

제8항에 있어서,

상기 강의내용 정보를 생성하는 단계는,

상기 과목 추천 시스템이, 상기 텍스트 데이터에 대한 단어 빈도-역 문서 빈도 기반의 분석에 의하여 상기 과목의 강의내용을 학습하는 단계; 및

상기 과목 추천 시스템이, 상기 강의내용에 대한 학습 결과에 기초하여 상기 강의내용 정보를 생성하는 단계를 포함하는 과목 유사도에 기반한 과목 추천 방법.

청구항 11

제8항에 있어서,

상기 수강생 특성 정보를 생성하는 단계는,

상기 과목 추천 시스템이, 복수의 수강생의 상기 수강이력 데이터에 기초하여 학과별 수강비율 데이터를 생성하는 단계; 및

상기 과목 추천 시스템이, 과목별 수강생의 상기 학과별 수강비율 데이터를 이용하여 상기 수강생 특성 정보를 생성하는 단계를 포함하는 과목 유사도에 기반한 과목 추천 방법.

청구항 12

제8항에 있어서,

상기 유사도 정보를 생성하는 단계는, 상기 과목 추천 시스템이, 상기 강의내용 정보 및 상기 수강생 특성 정보에 각각 상응하는 매트릭스들 간의 코사인 유사도를 산출하는 단계를 포함하는 과목 유사도에 기반한 과목 추천 방법.

청구항 13

삭제

청구항 14

제8항에 있어서,

상기 과목 추천 시스템이, 과목별 특성에 상응하는 복수 개의 차원에 의해 정의되며 과목 간의 유사도가 높을수록 과목들이 서로 가까이 위치하는 과목지도를 이용하여 상기 과목 추천 정보를 상기 사용자에게 제공하는 단계를 더 포함하는 과목 유사도에 기반한 과목 추천 방법.

청구항 15

하드웨어와 결합되어 제8항, 제10항 내지 제12항 및 제14항 중 어느 한 항에 따른 과목 유사도에 기반한 과목 추천 방법을 실행하도록 컴퓨터로 판독 가능한 기록매체에 저장된 컴퓨터 프로그램.

발명의 설명

기 술 분 야

[0001] 실시예들은 과목 유사도에 기반한 과목 추천 시스템과 방법 및 이를 위한 컴퓨터 프로그램에 대한 것이다. 보다 자세하게는, 실시예들은 강의내용과 수강생 특성을 함께 고려하여 산출되는 과목 유사도에 기반하여 추천과목을 결정할 수 있는 과목 추천 모델과 이를 서비스하는 기술에 대한 것이다.

배 경 기 술

[0002] 최근 많은 전문가들은 인공지능(Artificial Intelligence; AI) 기술이 우리 사회에 혁명적인 변화를 일으킬 것으로 전망하고 있다. 특히, 코로나바이러스 감염증-19(COVID-19)로 인해 사회 각 분야의 디지털화가 빠르게 진행되면서 많은 양의 데이터가 축적되고 있으며, AI 기술의 도입과 확산이 가속화되고 있다.

[0003] 교육 영역에서도 AI 도입에 관한 논의가 활발하게 이루어지고 있으며, 대한민국 정부에서는 AI기술을 활용한 개인 맞춤형 교육을 교육의 발전 방향으로 상정 및 추진하고 있다. 또한, 이와 관련된 종래 기술로, 공개특허공보 제10-2021-0026234호는 AI 기반의 사용자 맞춤형 수학 교육 서비스를 제공하는 장치 및 방법을 개시한다.

[0004] 그러나, 공개특허공보 제10-2021-0026234호에 개시된 것을 비롯한 종래의 교육 관련 AI 기술에서는, 연구 논문의 키워드, 교양강의의 교수요목 및 학습목표와 같은 비정형 데이터를 AI를 통해 분석하는 데에는 통찰이 미치지 못하고 있으며, 또한 이러한 비정형 데이터를 이용하여 과목간의 유사도를 도출하거나 이를 시각적으로 제공하는 방법은 전혀 개시하고 있지 못하다.

선행기술문헌

특허문헌

[0005] (특허문헌 0001) 공개특허공보 제10-2021-0026234호

발명의 내용

해결하려는 과제

[0006] 본 발명의 일 측면에 따르면, 비정형 데이터를 유사 과목 도출에 이용할 수 있으며, 강의내용 차원과 수강생 특성을 고려하여 과목 간의 유사도를 판단할 수 있고, 유사 과목을 과목지도 형태로 시각화하여 제공할 수 있는 과목 유사도에 기반한 과목 추천 시스템과 방법 및 이를 위한 컴퓨터 프로그램을 제공할 수 있다.

[0007] 본 발명의 기술적 과제들은 이상에서 언급한 기술적 과제들로 제한되지 않으며, 언급되지 않은 또 다른 기술적 과제들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

[0008] 본 발명의 일 측면에 따른 과목 유사도에 기반한 과목 추천 시스템은, 하나 이상의 과목에 연관된 텍스트 데이터를 분석함으로써 상기 하나 이상의 과목 각각에 대한 강의내용 정보를 생성하도록 구성된 제1 데이터 분석부; 과목별 수강생의 수강이력 데이터를 분석함으로써 상기 하나 이상의 과목 각각에 대한 수강생 특성 정보를 생성하도록 구성된 제2 데이터 분석부; 상기 강의내용 정보 및 상기 수강생 특성 정보를 이용하여 과목별 유사도 정보를 생성하도록 구성된 유사도 결정부; 및 사용자로부터 수신된 관심 과목 정보 및 상기 유사도 정보에 기초하여 상기 사용자에 대한 과목 추천 정보를 생성하도록 구성된 추천과목 결정부를 포함한다.

[0009] 일 실시예에 따른 과목 추천 시스템은, 상기 과목의 교수요목, 상기 과목의 학습목표, 학과별 커리큘럼 및 학문분야별 출간논문 키워드 중 하나 이상으로부터 상기 텍스트 데이터를 추출하도록 구성된 전처리부를 더 포함한다.

[0010] 일 실시예에서, 상기 제1 데이터 분석부는, 상기 텍스트 데이터에 대한 단어 빈도-역 문서 빈도 기반의 분석에

의하여 상기 과목의 강의내용을 학습하고, 상기 강의내용에 대한 학습 결과에 기초하여 상기 강의내용 정보를 생성하도록 더 구성된다.

[0011] 일 실시예에서, 상기 제2 데이터 분석부는, 복수의 수강생의 상기 수강이력 데이터에 기초하여 학과별 수강비율 데이터를 생성하고, 과목별 수강생의 상기 학과별 수강비율 데이터를 이용하여 상기 수강생 특성 정보를 생성하도록 더 구성된다.

[0012] 일 실시예에서, 상기 유사도 결정부는, 상기 강의내용 정보 및 상기 수강생 특성 정보에 각각 상응하는 매트릭스 간의 코사인 유사도를 이용하여 상기 과목별 유사도 정보를 생성하도록 더 구성된다.

[0013] 일 실시예에서, 상기 관심 과목 정보는 상기 사용자의 수강이력 및 수강희망과목 중 하나 이상의 정보를 포함한다. 이때, 상기 추천과목 결정부는, 상기 수강이력 및 수강희망과목 중 하나 이상의 정보를 이용하여 상기 사용자의 선호과목을 결정하고, 상기 선호과목에 대한 유사도에 기초하여 추천과목을 결정함으로써 상기 과목 추천 정보를 생성하도록 더 구성된다.

[0014] 일 실시예에서, 상기 추천과목 결정부는, 과목별 특성에 상응하는 복수 개의 차원에 의해 정의되며 과목 간의 유사도가 높을수록 과목들이 서로 가까이 위치하는 과목지도를 이용하여 상기 과목 추천 정보를 상기 사용자에게 제공하도록 더 구성된다.

[0015] 본 발명의 일 측면에 따른 과목 유사도에 기반한 과목 추천 방법은, 과목 추천 시스템이 하나 이상의 과목에 연관된 텍스트 데이터를 분석함으로써 상기 하나 이상의 과목 각각에 대한 강의내용 정보를 생성하는 단계; 상기 과목 추천 시스템이 과목별 수강생의 수강이력 데이터를 분석함으로써 상기 하나 이상의 과목 각각에 대한 수강생 특성 정보를 생성하는 단계; 상기 과목 추천 시스템이 상기 강의내용 정보 및 상기 수강생 특성 정보를 이용하여 과목별 유사도 정보를 생성하는 단계; 및 상기 과목 추천 시스템이, 사용자로부터 수신된 관심 과목 정보 및 상기 유사도 정보에 기초하여, 상기 사용자에 대한 과목 추천 정보를 생성하는 단계를 포함한다.

[0016] 일 실시예에 따른 과목 추천 방법은, 상기 과목 추천 시스템이, 상기 과목의 교수요목, 상기 과목의 학습목표, 학과별 커리큘럼 및 학문분야별 출간논문 키워드 중 하나 이상으로부터 상기 텍스트 데이터를 추출하는 단계를 더 포함한다.

[0017] 일 실시예에서, 상기 강의내용 정보를 생성하는 단계는, 상기 과목 추천 시스템이, 상기 텍스트 데이터에 대한 단어 빈도-역 문서 빈도 기반의 분석에 의하여 상기 과목의 강의내용을 학습하는 단계; 및 상기 과목 추천 시스템이, 상기 강의내용에 대한 학습 결과에 기초하여 상기 강의내용 정보를 생성하는 단계를 포함한다.

[0018] 일 실시예에서, 상기 수강생 특성 정보를 생성하는 단계는, 상기 과목 추천 시스템이, 복수의 수강생의 상기 수강이력 데이터에 기초하여 학과별 수강비율 데이터를 생성하는 단계; 및 상기 과목 추천 시스템이, 과목별 수강생의 상기 학과별 수강비율 데이터를 이용하여 상기 수강생 특성 정보를 생성하는 단계를 포함한다.

[0019] 일 실시예에서, 상기 유사도 정보를 생성하는 단계는, 상기 과목 추천 시스템이, 상기 강의내용 정보 및 상기 수강생 특성 정보에 각각 상응하는 매트릭스들 간의 코사인 유사도를 산출하는 단계를 포함한다.

[0020] 일 실시예에서, 상기 관심 과목 정보는 상기 사용자의 수강이력 및 수강희망과목 중 하나 이상의 정보를 포함한다. 이때, 상기 과목 추천 정보를 생성하는 단계는, 상기 과목 추천 시스템이, 상기 수강이력 및 수강희망과목 중 하나 이상의 정보를 이용하여 상기 사용자의 선호과목을 결정하는 단계; 및 상기 과목 추천 시스템이, 상기 선호과목에 대한 유사도에 기초하여 추천과목을 결정하는 단계를 포함한다.

[0021] 일 실시예에 따른 과목 추천 방법은, 상기 과목 추천 시스템이, 과목별 특성에 상응하는 복수 개의 차원에 의해 정의되며 과목 간의 유사도가 높을수록 과목들이 서로 가까이 위치하는 과목지도를 이용하여 상기 과목 추천 정보를 상기 사용자에게 제공하는 단계를 더 포함한다.

[0022] 본 발명의 일 측면에 따른 컴퓨터 프로그램은 하드웨어와 결합되어 전술한 실시예들에 따른 과목 유사도에 기반한 과목 추천 방법을 실행하기 위한 것으로서 컴퓨터로 판독 가능한 기록매체에 저장된 것일 수 있다.

발명의 효과

[0023] 실시예들에 따른 과목 유사도에 기반한 과목 추천 시스템과 방법에 의하면, 강의내용 차원과 수강생 특성을 동시에 고려하여 과목 간의 유사도를 판별함으로써 종래 기술, 예컨대, 강의내용 또는 수강생 특성 중 어느 하나만을 사용하는 모델에 비해 훨씬 높은 정확도로 과목을 추천할 수 있으며, 학생별 관심사에 부합하는 폭 넓고 다양한 과목을 추천함으로써 종래 기술에 비해 추천의 범위와 다양성 측면에서도 더 양호한 결과를 도출할 수

있는 이점이 있다.

[0024] 본 발명의 효과들은 이상에서 언급한 효과들로 제한되지 않으며, 언급되지 않은 또 다른 효과들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

[0025] 도 1은 일 실시예에 따른 과목 유사도에 기반한 과목 추천 시스템의 개략적인 블록도이다.

도 2는 일 실시예에 따른 과목 유사도에 기반한 과목 추천 방법의 각 단계를 나타내는 순서도이다.

도 3은 일 실시예에 따른 과목 유사도에 기반한 과목 추천 방법에 의해 강의내용 정보를 생성하는 과정의 개념도이다.

도 4는 일 실시예에 따른 과목 유사도에 기반한 과목 추천 방법에 의해 수강생 특성 정보를 생성하는 과정의 개념도이다.

도 5 내지 도 7는 일 실시예에 따른 과목 유사도에 기반한 과목 추천 방법의 효과를 종래 기술과 비교하여 나타낸 그래프이다.

도 8a 및 8b는 일 실시예에 따른 과목 유사도에 기반한 과목 추천 방법에 의하여 제공되는 과목지도의 예시를 나타내는 개념도이다.

발명을 실시하기 위한 구체적인 내용

[0026] 이하에서, 도면을 참조하여 본 발명의 실시예들에 대하여 상세히 살펴본다.

[0027] 도 1은 일 실시예에 따른 과목 유사도에 기반한 과목 추천 시스템의 개략적인 블록도이다.

[0028] 도 1을 참조하면, 과목 유사도에 기반한 과목 추천 시스템(3)은 유선 및/또는 무선 네트워크를 통하여 수강생의 사용자 장치(1), 교육기관 서버(2) 및/또는 학술정보 제공 서버(4) 등과 통신하면서 동작하도록 구성될 수 있다. 유선 및/또는 무선 네트워크를 통한 통신 방법은 객체와 객체가 네트워킹 할 수 있는 모든 통신 방법을 포함할 수 있으며, 유선 통신, 무선 통신, 3G, 4G, 혹은 그 이외의 방법으로 제한되지 않는다.

[0029] 예를 들어, 유선 및/또는 무선 네트워크는 LAN(Local Area Network), MAN(Metropolitan Area Network), GSM(Global System for Mobile Network), EDGE(Enhanced Data GSM Environment), HSDPA(High Speed Downlink Packet Access), W-CDMA(Wideband Code Division Multiple Access), CDMA(Code Division Multiple Access), TDMA(Time Division Multiple Access), 블루투스(Bluetooth), 지그비(Zigbee), 와이-파이(Wi-Fi), VoIP(Voice over Internet Protocol), LTE Advanced, IEEE802.16m, WirelessMAN-Advanced, HSPA+, 3GPP Long Term Evolution (LTE), Mobile WiMAX (IEEE 802.16e), UMB (formerly EV-DO Rev. C), Flash-OFDM, iBurst and MBWA (IEEE 802.20) systems, HIPERMAN, Beam-Division Multiple Access (BDMA), Wi-MAX(World Interoperability for Microwave Access) 및 초음파 활용 통신으로 이루어진 군으로부터 선택되는 하나 이상의 통신 방법에 의한 통신 네트워크를 지칭할 수 있으나, 이에 한정되는 것은 아니다.

[0030] 일 실시예에서, 과목 추천 시스템(3)은 제1 데이터 분석부(31), 제2 데이터 분석부(32), 유사도 결정부(33) 및 추천과목 결정부(35)를 포함한다. 일 실시예에서, 과목 추천 시스템(3)은 전처리부(30)를 더 포함할 수 있다. 또한 일 실시예에서, 과목 추천 시스템(3)은 데이터베이스(database; DB)(34)를 더 포함할 수 있다.

[0031] 본 명세서에 기재된 시스템, 장치 및 서버는 전적으로 하드웨어이거나, 또는 부분적으로 하드웨어이고 부분적으로 소프트웨어인 측면을 가질 수 있다. 예컨대, 본 명세서의 시스템, 장치 및 서버와, 이들에 포함된 각 부(unit)는 특정 형식 및 내용의 데이터를 처리하거나 또는/또한 전자통신 방식으로 주고받기 위한 하드웨어 및 이에 관련된 소프트웨어를 통칭할 수 있다. 본 명세서에서 "부", "모듈", "장치", "단말기", "서버" 또는 "시스템" 등의 용어는 하드웨어 및 해당 하드웨어에 의해 구동되는 소프트웨어의 조합을 지칭하는 것으로 의도된다. 예를 들어, 하드웨어는 CPU 또는 다른 프로세서(processor)를 포함하는 데이터 처리 기기일 수 있다. 또한, 하드웨어에 의해 구동되는 소프트웨어는 실행중인 프로세스, 객체(object), 실행파일(executable), 실행 스레드(thread of execution), 프로그램(program) 등을 지칭할 수 있다.

[0032] 또한, 본 실시예에 따른 과목 추천 시스템(3)을 구성하는 각각의 요소는 반드시 서로 물리적으로 구분되는 별개의 장치를 지칭하는 것으로 의도되지 않는다. 즉, 도 1의 전처리부(30), 제1 데이터 분석부(31), 제2 데이터 분

석부(32), 유사도 결정부(33), DB(34) 및 추천과목 결정부(35) 등은 과목 추천 시스템(3)을 구성하는 하드웨어를 해당 하드웨어에 의해 수행되는 동작에 따라 기능적으로 구분한 것일 뿐, 반드시 각각의 부가 서로 독립적으로 구비되어야 하는 것이 아니다. 물론, 실시예에 따라서는 과목 추천 시스템(3)의 각 부 중 하나 이상이 서로 물리적으로 구분되는 별개의 장치로 구현되는 것도 가능하다.

[0033] 과목 추천 시스템(3)은 추천의 대상에 해당하는 하나 이상의 과목에 연관된 텍스트 데이터와 복수의 수강생들의 수강이력 데이터 등을 분석함으로써 강의내용 정보와 수강생 특성 정보를 생성하고, 이를 토대로 과목 간의 유사도를 결정하도록 구성된다. 또한, 과목 추천 시스템(3)은 사용자 장치(1)로부터 과목을 추천받고자 하는 사용자의 관심 과목 정보를 수신하고, 관심 과목 정보와 전술한 유사도 정보를 기초로 과목 추천 정보를 생성하여 사용자 장치(1)에 제공할 수 있다.

[0034] 사용자가 관심 과목 정보를 전송하고 과목 추천 정보를 수신하는 것이 가능하도록, 과목 추천 시스템(3)은 사용자 장치(1) 상에서 실행되는 애플리케이션(또는, 앱(app))과 통신함으로써 애플리케이션의 기능 수행을 가능하게 하는 애플리케이션 서비스 서버의 기능을 수행할 수 있다. 또는, 다른 실시예에서 과목 추천 시스템(3)은 사용자 장치(1) 상에서 실행되는 웹 브라우저에 의하여 접속 가능한 웹 페이지를 제공하는 웹 서버 등의 형태로 구현될 수도 있다.

[0035] 도 1에 도시된 실시예에서 사용자 장치(1)는 노트북 컴퓨터의 형태로 도시되었다. 그러나 이는 예시적인 것으로서, 사용자 장치(1)의 종류는 도면에 도시된 것으로 한정되는 것은 아니다. 예를 들어, 과목을 추천받고자 하는 사용자는 스마트폰(smartphone)과 같은 이동 통신 단말기, 개인용 컴퓨터(personal computer), 노트북 컴퓨터, PDA(personal digital assistant), 태블릿(tablet), IPTV(Internet Protocol Television) 등을 위한 셋톱박스(set-top box) 등 임의의 컴퓨팅 장치를 이용하여 과목 추천 시스템(3)이 제공하는 기능을 사용할 수 있다.

[0036] 전처리부(30)는 과목 간의 유사도를 계산하기 위한 기초가 되는 텍스트 데이터를 추출하기 위한 부분이다. 전처리부(30)는 과목의 교수요목 및/또는 학습목표, 학과별 커리큘럼 등으로부터 텍스트 데이터를 추출할 수 있다. 이상의 동작을 위하여, 전처리부(30)는 각 과목의 강의를 제공하는 대학 등 교육기관 서버(2)로부터 정보를 수집하도록 구성될 수 있다. 또한, 전처리부(30)는 강의내용 정보를 생성함에 있어서 과목에 해당하는 학문분야의 키워드를 결정하기 위하여 학문분야별 출간논문 데이터로부터 텍스트 데이터를 추출할 수 있다. 이를 위하여, 전처리부(30)는 연구 논문 데이터를 제공하는 학술정보 제공 서버(4)와 통신하도록 구성될 수 있다.

[0037] 또한, 전처리부(30)는 과목별 수강생들의 수강이력 데이터에 대한 전처리를 통하여 과목 수강생의 특성을 추출하도록 구성될 수 있다. 나아가, 전처리부(30)는 과목을 추천받고자 하는 사용자가 사용자 장치(1)로부터 전송한 관심 과목 정보에 대한 전처리를 통하여 선호과목 정보를 추출하도록 더 구성될 수 있다.

[0038] 제1 데이터 분석부(31)는, 하나 이상의 과목에 연관된 텍스트 데이터를 분석함으로서 각 과목에 대한 강의내용 정보를 생성할 수 있다. 일 실시예에서, 제1 데이터 분석부(31)는 텍스트 데이터에 대한 단어 빈도-역 문서 빈도(Term Frequency-Inverse Document Frequency; TF-IDF) 기반의 분석에 의하여 과목의 강의내용을 학습하고, 강의내용에 대한 학습 결과에 기초하여 강의내용 정보를 생성할 수 있다. 제1 데이터 분석부(31)의 구체적인 동작에 대해서는 도 3을 참조하여 상세히 후술한다.

[0039] 제2 데이터 분석부(32)는, 과목별 수강생의 수강이력 데이터를 분석함으로써 각 과목에 대한 수강생 특성 정보를 생성할 수 있다. 예를 들어, 제2 데이터 분석부(32)는 복수의 수강생의 수강이력 데이터에 기초하여 학과별 수강비율 데이터를 생성하고, 각 과목별 수강생들의 학과별 수강비율 데이터를 이용하여 수강생 특성 정보를 생성할 수 있다. 제2 데이터 분석부(32)의 구체적인 동작에 대해서는 도 4를 참조하여 상세히 후술한다.

[0040] 유사도 결정부(33)는, 제1 데이터 분석부(31)에 의하여 생성된 강의내용 정보 및 제2 데이터 분석부(32)에 의하여 생성된 수강생 특성 정보를 이용하여 과목별 유사도 정보를 생성할 수 있다. 예를 들어, 유사도 결정부(33)는 과목별 학문영역 데이터가 임베딩(embedding)된 강의내용 매트릭스와, 과목을 수강한 학생들의 학과별 수강비율이 임베딩된 수강생 특성 매트릭스 사이의 코사인 유사도를 이용하여 과목별 유사도를 결정할 수 있다. 과목별 유사도를 결정하는 구체적인 과정에 대해서는 도 2를 참조하여 상세히 후술한다.

[0041] 일 실시예에서, 제1 데이터 분석부(31)에 의하여 생성된 강의내용 정보, 제2 데이터 분석부(32)에 의하여 생성된 수강생 특성 정보, 및/또는 유사도 결정부(33)에 의하여 생성된 유사도 정보는 과목 추천 시스템(3)의 DB(34)에 저장될 수 있다.

[0042] 추천과목 결정부(35)는, 사용자 장치(1)로부터 수신된 관심 과목 정보와 유사도 결정부(33)에 의하여 생성된 과목별 유사도 정보를 이용하여, 대상 사용자에 대한 과목 추천 정보를 생성할 수 있다. 예를 들어, 추천과목 결

정부(35)는 추천 대상 사용자의 수강이력 및/또는 수강희망과목을 토대로 사용자의 선호과목을 결정하고, 선호과목과의 유사도에 기초하여 추천과목을 결정할 수 있다.

[0043] 추천과목 결정부(35)는 생성된 과목 추천 정보를 사용자 장치(1)에 제공할 수 있다. 이때, 과목 추천 정보는 과목 간의 유사도를 나타내는 과목지도의 형태로 시각화될 수도 있다. 과목지도는 과목별 특성에 상응하는 복수개의 차원에 의해 정의되며 과목 간의 유사도가 높을수록 과목들이 서로 가까이 위치하는 평면 또는 입체 지도의 형태를 가질 수 있다. 이때, 과목별 학문영역 및 학과특성을 한정된 수의 차원(예컨대, 2차원)으로 축소하기 위하여 UMAP(Uniform Manifold Approximation and Projection) 알고리즘이 이용될 수도 있다. 과목지도의 형태에 대해서는 도 8a 및 8b를 참조하여 상세히 후술한다.

[0044] 도 2는 일 실시예에 따른 과목 유사도에 기반한 과목 추천 방법의 각 단계를 나타내는 순서도이다. 실시예들에 따른 과목 추천 방법은, 실시예들에 따른 과목 추천 시스템을 이용하여 수행될 수 있다. 이하에서는, 설명의 편의를 위하여 도 1 및 도 2를 참조하여 본 실시예에 따른 과목 추천 방법에 대하여 설명한다.

[0045] 본 발명의 실시예들은 과목을 추천받고자 하는 대상 사용자가 관심있는 강의와 유사한 강의를 추천하는 과목 기반 추천 모델을 이용하여 구현된다. 더 구체적으로는, 실시예들은 각 과목별 데이터를 강의내용 차원과 수강생 특성 차원에서 특성 매트릭스에 임베딩하는 방식으로 과목별 강의내용 정보와 수강생 특성 정보를 산출하고, 이를 활용하여 과목간 유사도를 산출할 수 있다. 이후, 사용자의 선호과목과 유사도가 높은 과목을 추천하는 방식으로 과목 추천 정보를 사용자에게 제공할 수 있다.

[0046] 구체적으로, 과목 추천 시스템(3)의 제1 데이터 분석부(31)는 연구 논문 데이터를 대상으로 키워드 분석을 수행함으로써, 각 과목의 학문분야 키워드 매트릭스를 생성할 수 있다(S11). 또한, 제1 데이터 분석부(31)는 과목별 교수요목과 같은 과목정보 데이터에 대한 형태소 분석을 통해 과목정보 키워드 매트릭스를 생성할 수 있다(S12). 다음으로, 제1 데이터 분석부(31)는 학문분야 키워드 매트릭스와 과목정보 키워드 매트릭스를 활용하여 각 과목을 강의내용 매트릭스에 임베딩함으로써 강의내용 정보를 생성할 수 있다(S13).

[0047] 제1 데이터 분석부(31)에 의한 이상의 동작을 위하여, 과목 추천 시스템(3)의 전처리부(30)가 연구논문 데이터로부터 학문영역별 키워드를 텍스트 데이터 형태로 추출할 수 있다. 또한, 전처리부(30)는 각 과목의 교수요목이나 학습목표, 학과별 커리큘럼 데이터 등 과목정보로부터 형태소 분석을 통하여 키워드에 해당하는 텍스트 데이터를 추출할 수 있다. 또한, 제1 데이터 분석부(31)에 의해 생성되는 각 매트릭스에서 각 키워드의 중요도를 반영하기 위하여 TF-IDF 값이 이용될 수도 있다.

[0048] 도 3은 일 실시예에 따른 과목 유사도에 기반한 과목 추천 방법에 의해 강의내용 정보를 생성하는 과정의 개념도이다.

[0049] 도 3을 참조하면, 추천 대상 과목들에 해당하는 각 과목의 학문영역 특성을 임베딩하기 위하여 연구논문 데이터를 이용할 수 있다. 예컨대, 연구논문 데이터의 각 논문은 통상 5 내지 7개 정도의 키워드를 포함하고 있으며, 투고된 저널의 특성에 따라 통상 1 내지 3개 정도의 학문 분야에 할당되어 있다. 일 실시예에서, 연구논문 데이터는 학술정보 제공 서버(4; 도 1)로부터 획득될 수도 있다.

[0050] 본 실시예에서는, 논문이 포함된 학문분야에 따라 논문에 포함된 키워드의 출현빈도를 집계하고, 집계된 데이터를 바탕으로 학문분야 키워드의 문서-단어 매트릭스(301)를 구축하였다. 문서-단어 매트릭스(301)의 각 행은 키워드에 대응되며, 각 열은 학문영역에 대응된다. 일 실시예에서는, 기록된 키워드가 일정 횟수 미만인 학문영역, 및/또는 모든 학문영역을 총합하여 일정 횟수 미만 등장한 키워드는 배제하고 전술한 문서-단어 매트릭스(301)를 구성할 수도 있다. 그러나 이는 예시적인 것으로서, 다른 실시예에서는 모든 학문영역과 모든 키워드를 포함하도록 문서-단어 매트릭스(301)를 구성할 수도 있다.

[0051] 이때, 문서-단어 매트릭스(301)의 각 성분은 연구논문의 키워드와 학문분야의 조합별 TF-IDF 값을 포함할 수 있다. TF-IDF 값은 키워드의 전체 등장 빈도와 특정 학문분야 내 등장 빈도를 고려해 하나의 키워드가 특정 학문분야 내에서 얼마나 중요한 키워드인지 나타내는 수치이며, 학문분야 전반에 걸쳐 높은 빈도로 등장하는 일반적인 키워드의 개별 분야 내 중요성을 적절하게 조절하기 위하여 사용된다. 일 실시예에서는, 전체 키워드-학문분야 조합별 TF-IDF 값이 0 내지 1 사이가 되도록 TF-IDF 값을 정규화할 수도 있다.

[0052] 한편, 추천 대상 과목들에서 각 과목의 콘텐츠에 해당하는 특성을 임베딩하기 위하여 과목정보가 이용될 수 있다. 이때 과목정보란 각 과목의 강의에서 학습하는 내용을 나타내는 것으로서, 예를 들어 두 강의에서 다루는 내용이 학문적으로 유사할수록 두 과목은 서로 유사한 것으로 간주할 수 있다. 본 실시예에서는, 과목의 학습내용을 분석하기 위하여 교수자가 입력한 교수요목 및/또는 학습목표, 교육기관 서버에서 제공하는 커리큘럼 정보

등을 과목정보로 이용할 수 있다. 일 실시예에서, 교수요목, 학습목표, 커리큘럼 정보 등은 교육기관 서버(2; 도 1)로부터 확득될 수도 있다.

[0053] 본 실시예에서는, 먼저 비정형 데이터인 과목정보 데이터에 대한 형태소 분석을 통하여 과목정보 키워드를 추출할 수 있다. 이때, 일 실시예에서는 예컨대 [기말, 페이퍼, 제출, ... 공부, 교시, 필요] 등 강의별 교육 내용과 밀접한 관련이 없는 단어들은 미리 설정된 불용어로 분류하여 키워드로부터 배제할 수도 있다. 과목정보 데이터로부터 추출된 키워드들의 매트릭스(302)는 전치(transpose)를 통하여 키워드-과목 매트릭스(303)로 변환될 수 있으며, 키워드-과목 매트릭스(303)의 각 행은 각 과목에 상응하고 각 열은 각 키워드에 상응한다.

[0054] 일 실시예에서는, 학문분야 키워드 매트릭스(301)와 마찬가지로 특정 키워드가 한 과목 내에서 얼마나 많이 등장했는지(TF)와, 얼마나 많은 과목에서 등장했는지(DF)를 고려해 키워드-과목 쌍별 TF-IDF 값을 산출할 수도 있으며, TF-IDF값이 높을수록 특정 키워드가 해당 과목을 대표하는 중요 키워드일 수 있다는 것을 의미한다. 또한 일 실시예에서, 각 과목정보 키워드의 과목별 TF-IDF 값은 0 내지 1 사이의 값으로 정규화될 수도 있다.

[0055] 다음으로, 학문분야 키워드 매트릭스(301)와 과목정보 키워드 매트릭스(303)를 이용하여 강의내용 매트릭스(310)를 생성할 수 있다. 이때 강의내용 매트릭스(310)의 각 행은 각각의 과목에 해당하며, 각 열은 각각의 학문영역에 해당된다. 또한, 강의내용 매트릭스(310)의 각 성분은 과목정보 키워드의 학문영역 특성을 TF-IDF 가중치를 적용하여 평균을 구한 값으로 산출될 수 있다. 즉, 과목별로 중요한 키워드의 학문영역 특성이 그 과목의 학문영역 특성을 산출할 때 더 크게 반영된다.

[0056] 예를 들어, 학문분야 m에서 키워드 i의 TF-IDF 값을 하기 수학식 1의 F로 나타내고, 과목정보 j에서 키워드 i의 TF-IDF 값을 하기 수학식 2의 C로 나타내며, 과목정보 j의 키워드 수를 N_j로 나타낼 경우, 학문분야 m에 속하는 과목정보 j의 가중치가 반영된 TF-IDF 값 F'은 하기 수학식 3과 같이 산출될 수 있다.

[0057] [수학식 1]

[0058] $$F = K = [k_{i,m}]$$

[0059] [수학식 2]

[0060] $$C = [c_{i,j}]$$

[0061] [수학식 3]

[0062] $$F' = [f'_{j,m}] = \frac{(C^T \times K)}{N_j}$$

[0063] 다시 도 1 및 도 2를 참조하면, 과목 추천 시스템(3)의 제2 데이터 분석부(32)는 각 과목별로 해당 과목의 강의를 들은 수강생의 특성을 반영하기 위하여 과목별 수강생 특성 정보를 생성할 수 있다. 이를 위하여, 제2 데이터 분석부(32)는 복수의 수강생의 수강이력 데이터를 이용할 수 있다. 먼저, 제2 데이터 분석부(32)는 수강이력 데이터의 각 수강생에 대하여 수강생이 이수한 전체 학점 중 특정 학과에서 이수한 학점의 비율을 산출함으로써, 수강생별 수강이력 특성 매트릭스를 산출할 수 있다(S14). 이를 위하여, 전처리부(30)는 수강이력 데이터에서 학번, 과목, 개설학과 등의 정보를 텍스트 데이터 형태로 추출할 수 있다.

[0064] 다음으로, 제2 데이터 분석부(32)는 추천의 대상이 되는 하나 이상의 과목을 대상으로, 수강이력 데이터에서 해당 과목을 수강한 수강생들의 수강이력 특성 매트릭스를 이용하여 과목별 수강생 특성을 산출할 수 있다(S15). 예를 들어, 동일한 과목을 수강한 수강생들의 수강이력 특성 매트릭스의 평균이 해당 과목의 수강생 특성 정보로 이용될 수 있다.

[0065] 도 4는 일 실시예에 따른 과목 유사도에 기반한 과목 추천 방법에 의해 수강생 특성 정보를 생성하는 과정의 개념도이다.

[0066] 도 4를 참조하면, 본 실시예에서 수강이력 데이터는 수강생들의 수강이력을 기반으로 과목별 학과 TF-IDF 값을 산출하기 위하여 이용된다. 각 수강생의 식별정보(예컨대, 학번)와 각 수강생이 수강한 과목 정보(예컨대, 과목 및 개설학과)를 포함하는 수강이력 데이터(400)로부터, 수강과목 비율 매트릭스(401)를 산출할 수 있다. 수강과

목 비율 매트릭스(401)의 각 행은 수강생에 해당하며, 각 열은 과목이 개설된 학과에 해당된다. 즉, 수강과목 비율 매트릭스(401)의 각 성분의 값은 해당 행에 해당하는 수강생의 전체 취득 학점 중 해당 열에 해당하는 학과에서 취득한 과목의 학점 총합의 비율을 나타낸다.

[0067] 또한, 수강이력 데이터로부터 과목별 수강생 매트릭스(402)를 추출할 수 있다. 과목별 수강생 매트릭스(402)의 각 행은 각각의 과목에 해당하며, 수강생 매트릭스(402)의 각 열은 각 수강생에 해당한다. 이와 같이 산출된 수강과목 비율 매트릭스(401)와 과목별 수강생 매트릭스(402)를 이용하여, 각각의 과목을 수강한 수강생들의 학과별 수강비율 특성이 임베딩된 수강생 특성 매트릭스(410)를 얻을 수 있다. 즉, 수강생 특성 매트릭스(401)의 각 행은 각각의 과목을 나타내며, 각 열은 해당 과목을 수강한 수강생들의 평균적인 개설학과별 수강비율을 나타낸다.

[0068] 다시 도 1 및 도 2를 참조하면, 과목 추천 시스템(3)의 유사도 결정부(33)는, 전술한 것과 같이 생성된 강의내용 매트릭스(310; 도 3)와 수강생 특성 매트릭스(410; 도 4)를 이용하여 과목간 유사도를 산출할 수 있다(S16). 일 실시예에서, 과목간 유사도 정보는 강의내용 매트릭스와 수강생 특성 매트릭스 각각을 기반으로 한 코사인 유사도에 의하여 산출될 수도 있다. 예를 들어, 서로 상이한 두 과목의 강의내용 매트릭스(또는 수강생 특성 매트릭스)를 각각 A, B라고 할 경우, 두 매트릭스 사이의 코사인 유사도는 하기 수학식 4와 같이 산출될 수 있다.

[0069] [수학식 4]

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}}$$

[0070]

[0071] 상기 수학식 4와 같이 산출되는 코사인 유사도에 기초하여, 강의내용 정보를 기초로 한 과목간 유사도 정보와 수강생 특성 정보를 기초로 한 과목간 유사도 정보를 각각 생성할 수 있다.

[0072] 다음으로, 과목 추천 시스템(3)의 추천과목 결정부(35)는, 과목을 추천받고자 하는 대상 사용자에게 적합한 추천과목을 결정하기 위하여 대상 사용자의 관심 과목 정보를 수신할 수 있다(S17). 이때 관심 과목 정보란, 사용자가 직접 지정한 수강희망과목을 포함할 수도 있고, 또는/또한 사용자의 기존 수강이력을 포함할 수도 있다.

[0073] 다음으로, 추천과목 결정부(35)는 관심 과목 정보를 토대로 사용자의 선호과목을 결정하고, 유사도 정보에서 선호과목과의 유사도가 높은 과목을 미리 설정된 개수의 과목을 추천과목으로 결정할 수 있다(S18). 관심 과목 정보가 사용자의 수강희망과목을 포함하는 경우에는 수강희망과목 자체가 선호과목에 해당하며, 관심 과목 정보가 기존 수강이력을 포함하는 경우에는 기존 수강이력 중 선별된 과목이 선호과목에 해당될 수 있다.

[0074] 일 실시예에서는, 기존 수강이력에서 선호과목을 선별하기 위하여, 해당 사용자가 필수적으로 이수한 과목(전공과목 또는 필수교양과목 등)을 제외한 나머지 과목을 선호과목으로 결정할 수도 있다. 이는 사용자가 졸업을 위하여 의무적으로 수강한 과목 등을 배제하고 오로지 사용자의 선호에 의하여 수강한 과목을 선별하기 위한 것이다.

[0075] 추천과목 결정부(35)는 이상과 같이 결정된 추천과목을 포함하는 과목 추천 정보를 사용자 장치(1)에 제공할 수 있다(S19). 예를 들어, 추천과목 결정부(35)는 사용자의 선호과목에 대한 유사도에 기초하여 유사도가 상위인 n개의 과목을 추천과목으로 결정할 수 있다(n은 임의의 자연수). n 값이 클수록 넓은 범위의 추천이 이루어지고 n값이 작을수록 세부적인 범위의 추천이 이루어지며, 따라서 상기 n 값은 본 명세서에서 추천 범위로도 지칭된다.

[0076] 일 실시예에서, 추천과목 결정부(35)는 (i) 선정 횟수, (ii) 수강희망 순위, 및 (iii) 유사도 백분위를 고려한 가중치를 반영하여 최종 추천과목을 결정할 수도 있다. 이때 선정 횟수란 얼마나 많은 수의 선호과목에 대한 유사도를 기준으로 상위 n개의 추천과목에 해당하는지를 가중치로 반영하기 위한 것이며, 다수의 선호과목과 유사한 과목일수록 선정 횟수 가중치가 증가하도록 구성할 수 있다. 또한, 수강희망 순위는 추천과목과 유사한 선호과목에 대하여 사용자가 기재한 수강희망 순위를 가중치로 반영하기 위한 것이다. 또한, 유사도 백분위는 선호과목과의 유사도 수치를 백분위로 변환한 값을 가중치로 반영하기 위한 것이다.

[0077] 일 실시예에서, 추천과목 결정부(35)는 추천 대상 사용자가 수강희망 과목을 등록하지 않았으며 해당 사용자의 기존 수강이력도 없는 경우, 해당 사용자와 동일한 특성을 가지는 다른 수강생들의 선호과목에 기초하여 추천과

목을 결정할 수도 있다. 이때 다른 수강생들이란, 추천 대상 사용자와 동일한 전공을 갖는 수강생들을 지칭할 수 있으나, 이에 한정되는 것은 아니다. 또한 다른 수강생들의 선호과목은, 다른 수강생들이 입력한 수강희망 과목이거나, 또는 다른 수강생들의 기존 수강이력에 기초하여 결정되는 과목을 지칭할 수 있다.

[0078] 본 발명자들은, 이상에서 설명한 본 발명의 실시예들에 따른 과목 추천 모델의 성능을 검증하기 위해 추천 모델이 추천한 과목과 학생들의 실제 수강신청 결과를 비교하였다. 이를 위해 2020년 1학기까지 3과목 이상 과목을 신청한 학생 중 2020년 2학기에 수강신청을 한 9,275명의 데이터를 활용했다. 2020년 1학기까지 데이터를 기반으로 학생별 추천과목 목록을 산출하고, 이를 2020년 2학기에 학생이 선택한 과목과 비교했다. 본 명세서에서는, 표현의 간결성을 위하여 학생들이 수강 신청한 과목과 수강희망과목에 등록한 과목을 선택 과목으로 표현한다. 학생들이 선택한 과목은 평균 2.94개였으며, 표준편차는 1.86이었다.

[0079] 성능의 검증에는 추천 모델 성능 검증에 널리 활용되는 정밀도(precision rate)와 재현율(recall rate)을 사용하였다. 정밀도는 추천 횟수 대비 추천 성공 횟수를 통해 [정밀도 = (과목추천 성공 횟수) / (추천과목 수)]로 산출하였으며, 이때 추천 성공은 추천 모델이 추천한 과목이 실제 학생이 선택한 과목일 경우로 정의하였다. 또한 재현율은 선택 과목 수에서 과목추천이 성공한 수의 비율을 통해 [재현율 = (과목추천 성공 횟수) / (선택 과목 수)]로 산출하였다.

[0080] 도 5 내지 도 7는 일 실시예에 따른 과목 유사도에 기반한 과목 추천 방법의 효과를 종래 기술과 비교하여 나타낸 그래프이다.

[0081] 도 5는 일 실시예에 따른 과목 추천 방법에 의한 추천 정확도를 재현율과 정밀도 곡선을 통하여 종래 기술과 비교한 것을 나타내며, 도 5의 (a) 내지 (f)는 각각 추천 범위 값으로 3, 5, 10, 30, 50, 100을 사용한 경우를 나타낸다. 본 실시예와 대비되는 가상의 종래 기술은 활용한 데이터에 따라 "수강생 특성" 만을 사용하는 모델 및 "강의내용" 만을 사용하는 모델로 구분하였다. 추천 범위가 커질수록 학생의 전반적인 관심사가 반영된 추천 결과가 산출되며, 추천 범위가 좁을수록 학생의 특정 관심 과목에 기반한 추천 결과가 산출되었다.

[0082] 도 5에 도시되는 바와 같이, 본 실시예에 따른 과목 추천 방법이 "수강생 특성" 또는 "강의내용" 중 어느 하나만을 사용하는 모델에 비해 성능이 우수하였다. 특히 추천 범위가 3, 5 및 10인 경우 본 실시예에 따른 과목 추천 방법의 정밀도와 재현율이 우수하여 분석과 서비스 적용에 바람직한 것으로 확인되었다. 또한, 추천과목 수에 따라 정밀도와 재현율 사이의 트레이드 오프(trade-off)가 확인되었다. 즉, 많은 과목을 추천해줄수록 학생이 선택한 과목 중에서 모델이 추천한 과목이 있을 확률이 높아지지만, 추천된 과목을 학생이 실제로 선택할 확률은 낮았다. 이에 기초하여, 학생들에게 많은 과목을 추천하기 위하여 21개 과목을 추천하는 것을 기본값으로 하여 본 발명을 구현할 수 있으나, 이는 예시적인 것으로서, 추천 범위와 추천과목 수는 실시예들에 따라 상이할 수 있다.

[0083] 도 6은 일 실시예에 따른 과목 추천 방법에 의한 추천 정확도를 커버리지(coverage) 및 섀넌 엔트로피(Shannon entropy) 측면에서 종래 기술과 비교한 것을 나타내는 것으로서, 도 6의 (a)는 모델별 커버리지를 나타내고, 도 6의 (b)는 모델별 추천 다양성을 나타낸다.

[0084] 도 6에서 커버리지는 전체 개설 과목 281개 중 추천 모델이 추천한 과목의 비율로 산출하였으며, 모델별 추천 다양성은 학생별로 얼마나 다양한 과목을 추천하는지를 의미하는 것으로서 섀넌 엔트로피 공식을 통해 산출하였다. 도시되는 바와 같이, 본 발명의 실시예에 따른 과목 추천 방법이 "수강생 특성" 또는 "강의내용" 중 어느 하나만을 사용하는 모델에 비해 더 많은 과목을 추천하였으며, 가장 넓은 범위의 과목 커버리지를 가지며 종래의 모델에 비해 더 다양한 추천 결과를 산출했다는 것을 알 수 있다.

[0085] 도 7은 일 실시예에 따른 과목 추천 방법에 의한 추천 결과와 학과별 인기과목에 기반하여 과목을 추천한 결과를 비교한 것으로서, 도 7의 (a)는 정밀도와 재현율 곡선을 나타내며, 도 7의 (b)는 커버리지와 섀넌 엔트로피를 나타낸다.

[0086] 도 7에 도시된 결과에서 본 발명의 실시예에 따른 과목 추천 방법은 추천범위 3을 적용을 적용한 결과를 나타낸다. 또한, 학과별 인기 과목은 2020년 2020년 2학기 개설과목 중에서 2020년 1학기까지 해당 학과에서 가장 많이 수강한 과목을 추천한 결과를 통해 산출하였다. 도시되는 바와 같이, 정밀도와 재현율에서는 인기과목 기반 추천이 더 우수한 추천 성능을 보였지만, 추천과목 커버리지와 추천의 다양성 측면에서는 본 발명의 실시예에 따른 과목 추천 방법이 인기과목 기반의 추천 모델보다 폭 넓고 다양한 추천결과를 산출하는 것으로 나타났다.

[0087] 도 8a 및 8b는 일 실시예에 따른 과목 유사도에 기반한 과목 추천 방법에 의하여 제공되는 과목지도의 예시를

나타내는 개념도이다.

[0088] 도 8a 및 8b에 도시된 과목지도는, 과목별 강의내용 매트릭스 및 과목정보 매트릭스를 McInnes, L., Healy, J. 및 Melville, J. 공저의 논문 "Umap: Uniform manifold approximation and projection for dimension reduction" (arXiv preprint arXiv: 1802.03426, 2018)에 개시된 UMAP 알고리즘을 이용하여 2차원으로 축소한 후 시각화한 결과물을 나타내는 것으로서, 과목지도 상에서 거리가 가까울수록 과목 간의 특성이 유사하며, 멀리 떨어져 있을수록 과목 간의 특성이 상이하다는 것을 의미한다.

[0089] 도 8a는 과목지도에서 각 과목의 개설학과에 따라 노드(node)가 서로 상이한 해칭을 갖도록 구분하여 나타낸 것으로서, 동일한 학과에서 개설한 과목이 서로 인접하여 위치하는 것을 통하여 본 발명의 실시예에 따른 과목 추천 방법에 의한 과목간 유사도 계산이 정확하다는 것을 알 수 있다. 도 8a에서는 흑백 지도상에서 해칭을 통하여 각 노드를 구분하였으나, 과목지도를 컬러로 제공하는 경우 각 노드의 색상을 통하여 개설학과를 구분할 수도 있다.

[0090] 또한, 도 8b는 학문영역 특성 데이터를 활용하여 특정 학문 영역의 특성이 강한 과목을 밝게 나타낸 것으로, 투자경제분석, 기술경영및전략, 공학도를위한기업가정신, 기업법 등 회계학(accounting) 특성이 높은 과목을 밝게 나타내었다. 그러나, 그 외 다른 학문영역에 대해서도 동일한 방식의 시각화를 적용할 수 있다. 또한, 다른 실시예에서는 학문영역이 아닌 과목정보 매트릭스에 기초하여 노드의 밝기를 구분함으로서, 예를 들어 특정 학과의 수업을 듣는 학생들의 비율이 높은 과목을 밝게 나타내도록 과목지도를 구성하는 것도 가능하다. 이와 같이 UMAP 알고리즘을 사용하여 차원이 축소된 과목지도 형태로 추천과목 정보를 제공하는 경우 전역 구조(global structure)와 위상적 구조(topological structure)를 저차원에서도 효과적으로 포착할 수 있도록 정보의 제공이 가능한 이점이 있다.

[0091] 이상에서 살펴본 바와 같이, 본 발명의 실시예들에 의하면 과목별 교수요목 및 학습목표 등에 대한 형태소 분석을 실시해 키워드를 추출한 후 TF-IDF 값을 통해 과목-키워드별 가중치를 계산할 수 있고, 이렇게 산출된 과목-키워드별 가중치와 연구 논문 키워드 기반 용어 사전을 활용하여 각 과목내용의 강의내용 특성 벡터를 생성할 수 있다. 또한, 각 과목의 수강생들의 학과별 수강비율을 통해 수강생 특성 벡터를 생성할 수 있다. 과목간 유사도는 이와 같은 과목별 강의내용 특성 벡터 및 수강생 특성 벡터에 기반한 코사인 유사도를 통해 계산하고, 학생별 관심과목과 유사도가 높은 과목들을 추천과목으로 추천할 수 있다. 나아가, 과목별 학문영역 특성 벡터에 UMAP 알고리즘을 적용하여 생성된 과목지도를 이용하여 과목 추천 정보를 시각적으로 제공할 수 있다.

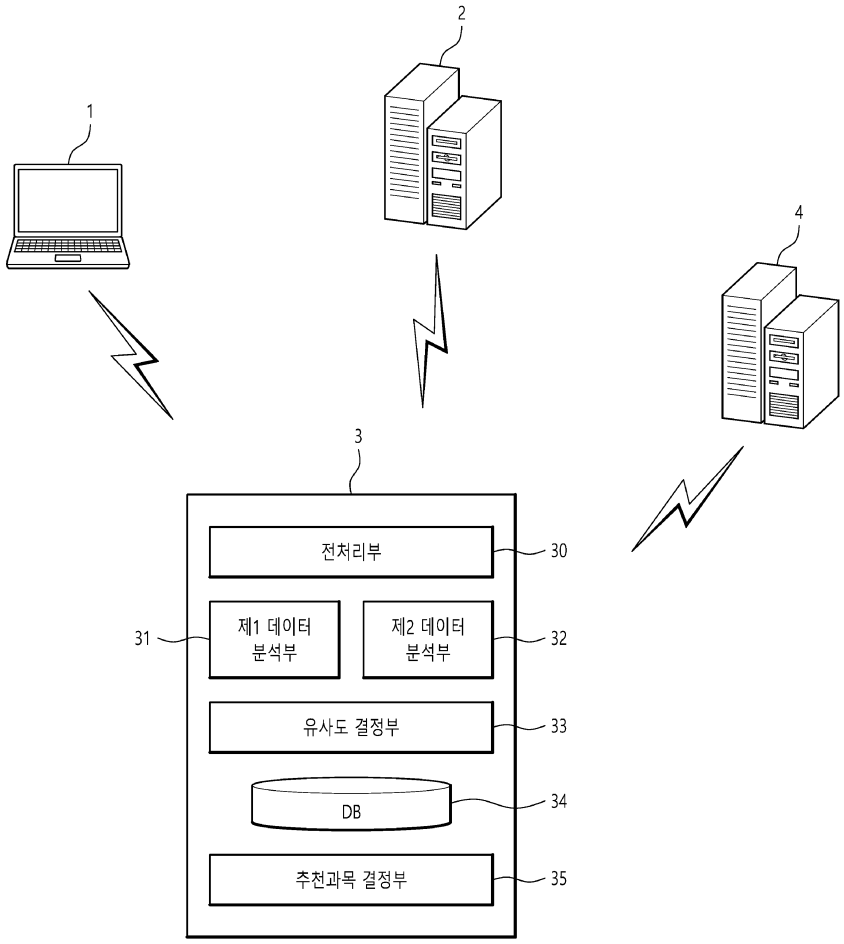
[0092] 본 발명의 실시예들에 의하면, 강의내용이나 수강생 특성 중 어느 하나의 정보만을 고려하여 과목을 추천하는 것에 비하여 실제 학생의 과목 선택과 일치하는 추천 결과를 산출할 수 있으며, 추천 모델의 정확도를 유지하면서도 추천의 범위와 다양성 측면에서 더 양호한 결과를 얻을 수 있는 이점이 있다. 이를 통하여, 본 발명의 실시예들을 이용할 경우 학생별 관심사에 부합하는 폭넓고 다양한 과목을 추천함으로써 창의·융합형 미래 인재 육성을 위한 맞춤형 서비스를 구현할 수 있고, 이는 제2 전공이나 진로의 추천, 마이크로 또는 나노 학위 관련 서비스 등으로 확장될 수 있는 이점이 있다.

[0093] 한편, 이상에서 설명한 실시예들에 따른 과목 유사도에 기반한 과목 추천 방법에 의한 동작은 적어도 부분적으로 컴퓨터 프로그램으로 구현되고 컴퓨터로 읽을 수 있는 기록매체에 기록될 수 있다. 실시예들에 따른 과목 유사도에 기반한 과목 추천 방법에 의한 동작을 구현하기 위한 프로그램이 기록되고 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치를 포함한다. 컴퓨터가 읽을 수 있는 기록매체의 예로는 ROM, RAM, CD-ROM, 자기 테이프, 플로피디스크, 광 데이터 저장장치 등이 있다. 또한 컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어, 분산 방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수도 있다. 또한, 본 실시예를 구현하기 위한 기능적인 프로그램, 코드 및 코드 세그먼트(segment)들은 본 실시예가 속하는 기술 분야의 통상의 기술자에 의해 용이하게 이해될 수 있을 것이다.

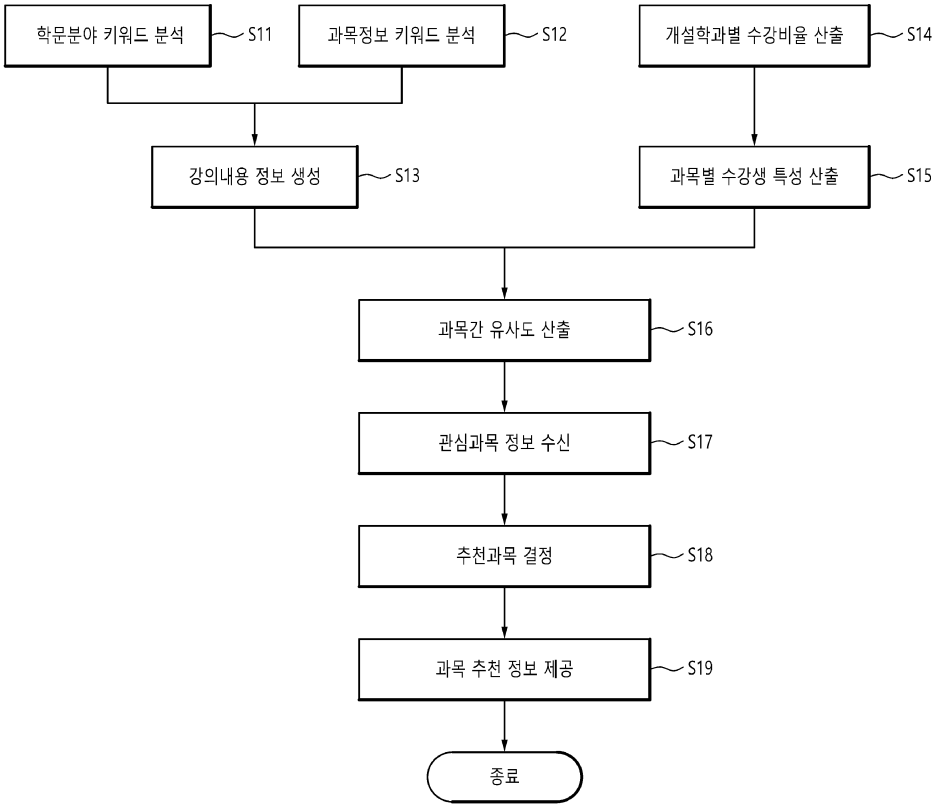
[0094] 이상에서 살펴본 본 발명은 도면에 도시된 실시예들을 참고로 하여 설명하였으나 이는 예시적인 것에 불과하며 당해 분야에서 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 실시예의 변형이 가능하다는 점을 이해할 것이다. 그러나, 이와 같은 변형은 본 발명의 기술적 보호범위 내에 있다고 보아야 한다. 따라서, 본 발명의 진정한 기술적 보호범위는 첨부된 청구범위의 기술적 사상에 의해서 정해져야 할 것이다.

도면

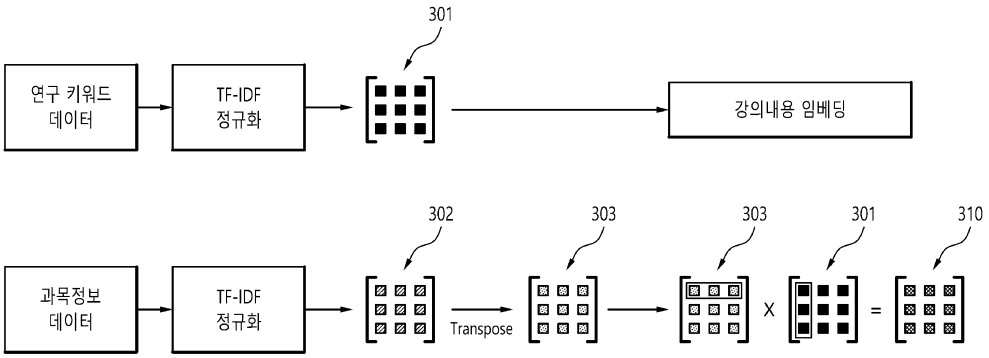
도면1


1
2
3
4
전처리부
30
제1 데이터 분석부
31
제2 데이터 분석부
32
유사도 결정부
33
DB
34
추천과목 결정부
35

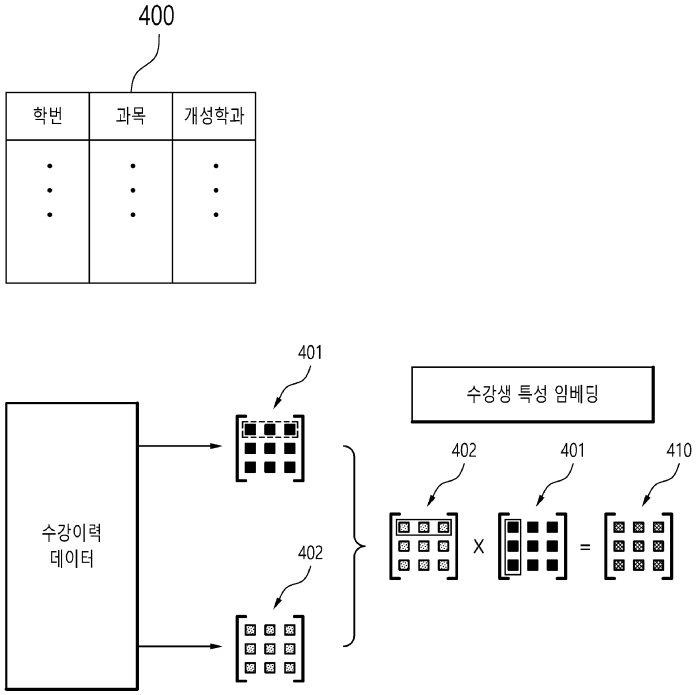
도면2


학문분야 키워드 분석
S11
과목정보 키워드 분석
S12
개설학과별 수강비율 산출
S14
강의내용 정보 생성
S13
과목별 수강생 특성 산출
S15
과목간 유사도 산출
S16
관심과목 정보 수신
S17
추천과목 결정
S18
과목 추천 정보 제공
S19
종료

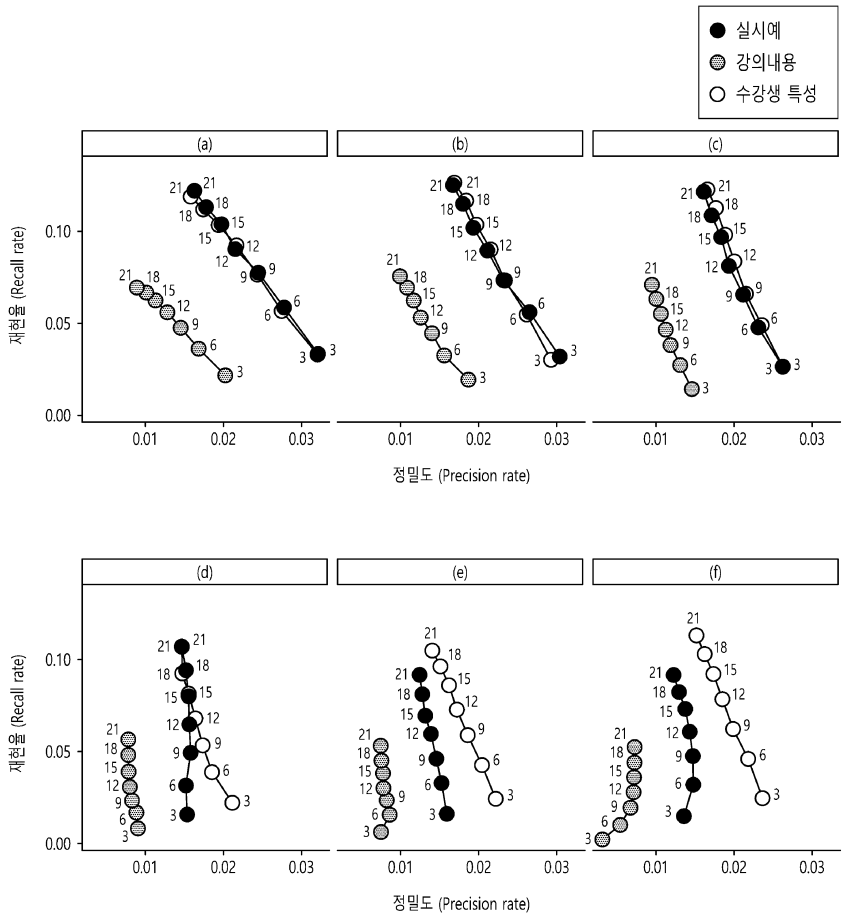
도면3


301
연구 키워드 데이터
TF-IDF 정규화
강의내용 임베딩
302
303
303
301
310
과목정보 데이터
TF-IDF 정규화
Transpose
X
=

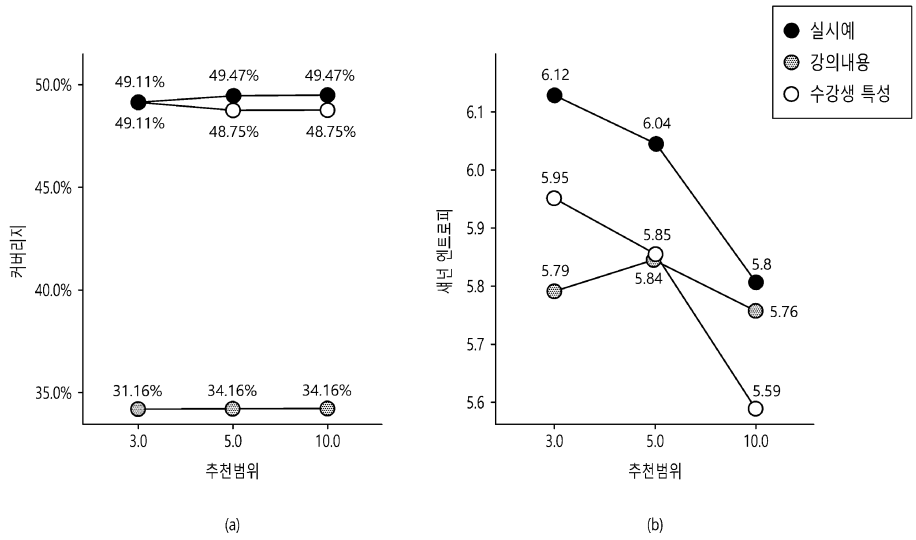
도면4


400
학번
과목
개성학과
401
수강생 특성 임베딩
수강이력 데이터
402
402
401
410
X
=

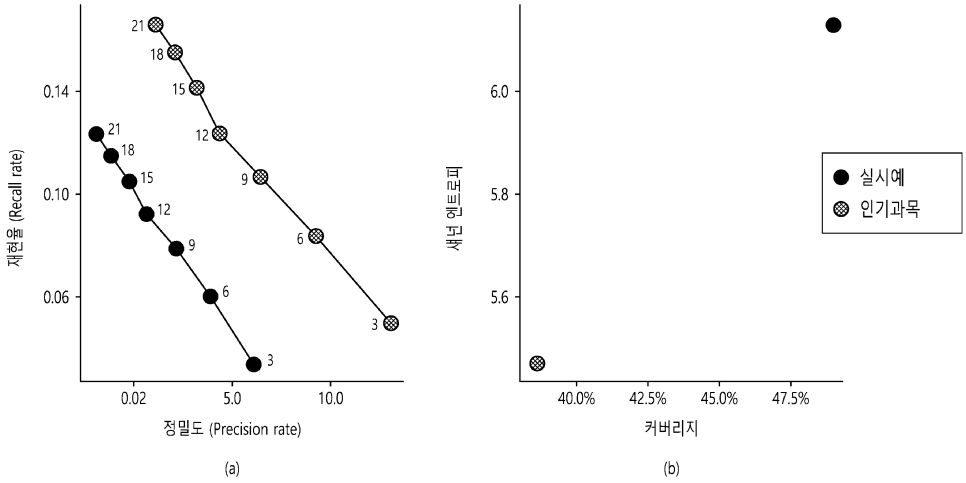
도면5


실시예
강의내용
수강생 특성
(a)
(b)
(c)
(d)
(e)
(f)
재현율 (Recall rate)
정밀도 (Precision rate)
0.00
0.05
0.10
0.01
0.02
0.03

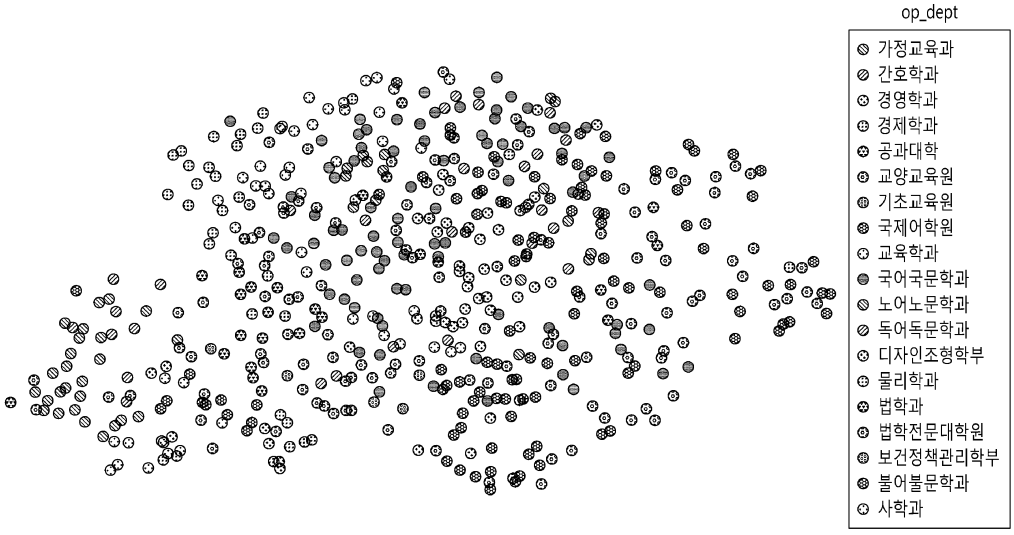
도면6


실시예
강의내용
수강생 특성
커버리지
50.0%
45.0%
40.0%
35.0%
49.11%
49.47%
49.47%
49.11%
48.75%
48.75%
31.16%
34.16%
34.16%
3.0
5.0
10.0
추천범위
(a)
섀넌 엔트로피
6.1
6.0
5.9
5.8
5.7
5.6
6.12
6.04
5.95
5.85
5.79
5.84
5.8
5.76
5.59
3.0
5.0
10.0
추천범위
(b)

도면7


재현율 (Recall rate)
0.14
0.10
0.06
21
18
15
12
9
6
3
0.02
5.0
10.0
정밀도 (Precision rate)
(a)
섀넌 엔트로피
6.0
5.8
5.6
실시예
인기과목
40.0%
42.5%
45.0%
47.5%
커버리지
(b)

도면8a


op_dept
가정교육과
간호학과
경영학과
경제학과
공과대학
교양교육원
기초교육원
국제어학원
교육학과
국어국문학과
노어노문학과
독어독문학과
디자인조형학부
물리학과
법학과
법학전문대학원
보건정책관리학부
불어불문학과
사학과

도면8b

