(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2018-0029509
(43) 공개일자 2018년03월21일

(51) 국제특허분류(Int. Cl.)
G06F 17/30 (2006.01)
(52) CPC특허분류
G06F 17/30601 (2013.01)
G06F 17/30713 (2013.01)
(21) 출원번호 10-2016-0117551
(22) 출원일자 2016년09월12일
심사청구일자 2016년09월12일

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
이성환
서울특별시 강남구 언주로30길 13, B동 2506호 (도곡동, 대림아크로빌)
원동욱
전라북도 남원시 금지면 독우물길 45-5
이홍복
서울특별시 서대문구 세검정로4다길 29-3, 2층 (홍제동)
(74) 대리인
특허법인엠에이피에스

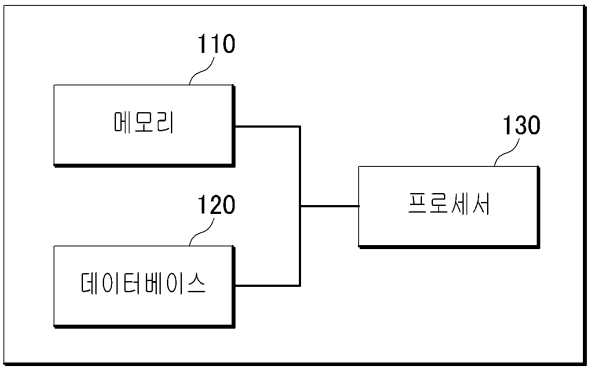
전체 청구항 수 : 총 14 항

(54) 발명의 명칭 **사용자를 위한 항목을 추천하는 장치 및 방법**

(57) 요 약

본 발명은 항목에 대한 추천 프로그램이 저장된 메모리 및 메모리에 저장된 프로그램을 실행하는 프로세서를 포함한다. 이때, 프로세서는 프로그램의 실행에 따라, 입력 데이터에 대한 클러스터링을 적용하여 입력 데이터에 포함된 복수의 항목 데이터에 대한 히스토그램을 생성하고, 입력 데이터 및 히스토그램에 기초하여 목표 사용자에 대한 신규 데이터를 생성하고, 생성된 신규 데이터 및 입력 데이터에 기초하여 목표 사용자에게 추천할 항목을 추출한다. 그리고 입력 데이터는 목표 사용자 및 복수의 타 사용자의 복수의 항목 각각에 대한 선호도를 포함한다.

대 표 도 - 도1



100

이 발명을 지원한 국가연구개발사업

과제고유번호	1711035286
부처명	미래창조과학부
연구관리전문기관	정보통신기술진흥센터
연구사업명	SW컴퓨팅산업원천기술개발사업
연구과제명	이동 환경에서 뇌-컴퓨터 인터페이스를 위한 지능형 패턴인식 소프트웨어 개발
기 여 율	1/1
주관기관	고려대학교 산학협력단
연구기간	2016.03.01 ~ 2017.02.28

명 세 서

청구범위

청구항 1

사용자를 위한 항목을 추천하는 장치에 있어서,

항목에 대한 추천 프로그램이 저장된 메모리 및

상기 메모리에 저장된 프로그램을 실행하는 프로세서를 포함하되,

상기 프로세서는 상기 프로그램의 실행에 따라, 입력 데이터에 대한 클러스터링을 적용하여 상기 입력 데이터에 포함된 복수의 항목 데이터에 대한 히스토그램을 생성하고, 상기 입력 데이터 및 상기 히스토그램에 기초하여 목표 사용자에 대한 신규 데이터를 생성하고, 상기 생성된 신규 데이터 및 상기 입력 데이터에 기초하여 상기 목표 사용자에게 추천할 항목을 추출하되,

상기 입력 데이터는 상기 목표 사용자 및 복수의 타 사용자의 상기 복수의 항목 각각에 대한 선호도를 포함하는, 항목 추천 장치.

청구항 2

제 1 항에 있어서,

상기 프로세서는 상기 입력 데이터에 포함된 복수의 사용자의 데이터와 상기 목표 사용자의 데이터 간의 유사도를 산출하고, 상기 산출된 유사도가 큰 순서대로 k(k는 자연수)명의 사용자를 추출하고, 상기 k 명의 사용자의 데이터에 기초하여 상기 히스토그램을 생성하는, 항목 추천 장치.

청구항 3

제 1 항에 있어서,

상기 프로세서는 상기 입력 데이터에 포함된 복수의 사용자의 데이터와 상기 목표 사용자의 데이터 간의 기계학습 기법을 적용하여 하나 이상의 클러스터를 생성하고,

상기 하나 이상의 클러스터 중 상기 목표 사용자에 대응하는 클러스터에 포함된 복수의 사용자의 데이터에 기초하여 상기 히스토그램을 생성하는, 항목 추천 장치.

청구항 4

제 1 항에 있어서,

상기 프로세서는 상기 히스토그램을 통하여 산출된 복수의 항목에 대한 선호도 및 상기 입력 데이터에 포함된 상기 목표 사용자에 대한 복수의 항목에 대한 선호도에 기초하여 상기 신규 데이터를 생성하는, 항목 추천 장치.

청구항 5

제 1 항에 있어서,

상기 프로세서는 상기 신규 데이터 및 상기 입력 데이터에 기초하여 복수의 추천 대상 항목에 대한 각각의 선호도를 예측하고, 상기 복수의 추천 대상 항목 중 예측된 선호도가 큰 n(n은 자연수) 개의 항목을 추출하되,

상기 추천 대상 항목은 상기 입력 데이터에 포함된 상기 목표 사용자의 데이터 중 선호도를 입력하지 않은 항목인, 항목 추천 장치.

청구항 6

제 5 항에 있어서,

상기 프로세서는 상기 신규 데이터 및 상기 입력 데이터에 기초하여 상기 목표 사용자의 데이터와 복수의 타 사

용자의 데이터 간의 유사도를 산출하고, 상기 산출된 유사도가 큰 순서대로 k명(k는 자연수)의 사용자를 선택하며,

상기 선택된 n명의 사용자의 데이터에 기초하여 상기 추천 대상 항목에 대한 선호도를 예측하는 것인, 항목 추천 장치.

청구항 7

제 5 항에 있어서,

상기 프로세서는 상기 입력 데이터에 포함된 복수의 항목의 데이터 간의 유사도를 산출하고,

상기 목표 사용자에 대한 신규 데이터를 생성된 이후, 상기 복수의 항목에 각각 대응되는 유사도, 상기 생성된 신규 데이터 및 상기 입력 데이터에 기초하여 상기 추천 대상 항목에 대한 선호도를 예측하는 것인, 항목 추천 장치.

청구항 8

제 1 항에 있어서,

상기 프로세서는 상기 입력 데이터와, 상기 입력 데이터에 포함된 각 사용자에 대한 정보 및 각 항목에 대한 정보 중 적어도 하나 이상을 이용하여 상기 클러스터링을 수행하되,

상기 사용자에 대한 정보는 상기 사용자의 인구통계학적 정보를 포함하며,

상기 항목에 대한 정보는 상기 항목에 대한 내용 정보를 포함하는, 항목 추천 장치.

청구항 9

추천 장치의 사용자를 위한 항목 추천 방법에 있어서,

입력 데이터에 대한 클러스터링을 수행하는 단계;

상기 클러스터링 결과에 기초하여, 상기 입력 데이터에 포함된 복수의 항목의 데이터에 대한 히스토그램을 생성하는 단계;

상기 입력 데이터 및 상기 히스토그램에 기초하여 목표 사용자에 대한 신규 데이터를 생성하는 단계; 및

상기 생성된 신규 데이터 및 상기 입력 데이터에 기초하여 상기 목표 사용자에게 추천할 항목을 추출하는 단계를 포함하되,

상기 입력 데이터는 상기 목표 사용자 및 복수의 타 사용자의 상기 복수의 항목 각각에 대한 선호도를 포함하는, 항목 추천 방법.

청구항 10

제 9 항에 있어서,

상기 신규 데이터를 생성하는 단계는,

상기 히스토그램을 통하여 산출된 복수의 항목에 대한 선호도 및 상기 입력 데이터에 포함된 상기 목표 사용자에 대한 복수의 항목에 대한 선호도에 기초하여 상기 신규 데이터를 생성하는, 항목 추천 방법.

청구항 11

제 9 항에 있어서,

상기 목표 사용자에게 추천할 항목을 추출하는 단계는,

상기 신규 데이터 및 상기 입력 데이터에 기초하여 복수의 추천 대상 항목에 대한 각각의 선호도를 예측하는 단계; 및

상기 복수의 추천 대상 항목 중 예측된 선호도가 큰 n(n은 자연수) 개의 항목을 추출하는 단계를 포함하되,

상기 추천 대상 항목은 상기 입력 데이터에 포함된 상기 목표 사용자의 데이터 중 선호도를 입력하지 않은 항목

인, 항목 추천 방법.

청구항 12

제 10 항에 있어서,

상기 신규 데이터 및 상기 입력 데이터에 기초하여 복수의 추천 대상 항목에 각각의 선호도를 예측하는 단계는,

상기 신규 데이터 및 상기 입력 데이터에 기초하여, 상기 목표 사용자의 데이터와 복수의 타 사용자의 데이터 간의 유사도를 산출하는 단계;

유사도가 큰 순서대로 k명(k는 자연수)의 사용자를 선택하는 단계; 및

상기 선택된 k명의 사용자의 데이터에 기초하여 상기 추천 대상 항목에 대한 선호도를 예측하는 단계를 포함하는, 항목 추천 방법.

청구항 13

제 10 항에 있어서,

상기 입력 데이터에 대한 클러스터링을 수행하는 단계 이전에, 상기 입력 데이터에 포함된 복수의 항목의 데이터 간의 유사도를 산출하는 단계를 더 포함하고,

상기 신규 데이터 및 상기 입력 데이터에 기초하여 복수의 추천 대상 항목에 각각의 선호도를 예측하는 단계는,

상기 복수의 항목에 각각 대응되는 유사도, 상기 생성된 신규 데이터 및 상기 입력 데이터에 기초하여 상기 추천 대상 항목에 대한 선호도를 예측하는 것인, 항목 추천 방법.

청구항 14

제 9 항 내지 제 13 항 중 어느 한 항에 기재된 방법을 컴퓨터 상에서 수행하기 위한 프로그램을 기록한 컴퓨터 판독 가능한 기록 매체.

발명의 설명

기 술 분 야

[0001] 본 발명은 사용자를 위한 항목을 추천하는 장치 및 방법에 관한 것이다.

배 경 기 술

[0002] 추천 시스템(recommendation system)은 다량의 항목(item) 중에서 사용자에게 적합한 항목을 필터링하여 사용자에게 제공하는 시스템이다. 추천 시스템에서 사용하는 추천 알고리즘은 협업 여과(collaborative filtering), 내용기반 여과(content-based filtering) 및 인구통계학적 정보 여과(demographical information filtering) 등이 있다. 협업 여과는 추천을 수행하는 목표 사용자와 다른 사용자 간의 유사성에 기초하여 항목의 추천을 수행한다. 내용기반 여과는 목표 사용자가 선호하는 내용 정보(content information)에 대한 유사성에 기초하여 항목의 추천을 수행한다. 또한, 인구통계학적 정보 여과는 목표 사용자의 인구통계학적 정보에 대한 유사성에 기초하여 항목에 대한 추천을 수행한다.

[0003] 추천 방법 중 대표적인 협업 여과는 사용자 기반 협업 여과(user-based collaborative filtering) 및 항목 기반 협업 여과(item-based collaborative filtering)가 있다. 이때, 사용자 기반 협업 여과는 사용자 간의 유사도(similarity)를 이용하여 목표 사용자에게 항목을 추천한다. 항목 기반 협업 여과는 항목 간의 유사도를 이용하여 목표 사용자에게 항목을 추천한다.

[0004] 이와 관련되어, 한국 공개특허공보 제10-2016-006447호(발명의 명칭: "협업 필터링의 예측 선호도를 이용한 처음 사용자에 대한 추천 방법 제공 방법")는 일부 선호 정보가 누락된 처음 사용자에 대한 예측 선호도를 이용한 추천 방법을 개시하고 있다.

발명의 내용

해결하려는 과제

[0005] 본 발명은 전술한 종래 기술의 문제점을 해결하기 위한 것으로서, 사용자에 대한 클러스터링 및 히스토그램에 기초하여 사용자를 위한 항목을 추천하는 장치 및 방법을 제공한다.

[0006] 다만, 본 실시예가 이루고자 하는 기술적 과제는 상기된 바와 같은 기술적 과제로 한정되지 않으며, 또 다른 기술적 과제들이 존재할 수 있다.

과제의 해결 수단

[0007] 상술한 기술적 과제를 달성하기 위한 기술적 수단으로서, 본 발명의 제 1 측면에 따른 사용자를 위한 항목을 추천하는 장치는 항목에 대한 추천 프로그램이 저장된 메모리 및 메모리에 저장된 프로그램을 실행하는 프로세서를 포함한다. 이때, 프로세서는 프로그램의 실행에 따라, 입력 데이터에 대한 클러스터링을 적용하여 입력 데이터에 포함된 복수의 항목 데이터에 대한 히스토그램을 생성하고, 입력 데이터 및 히스토그램에 기초하여 목표 사용자에 대한 신규 데이터를 생성하고, 생성된 신규 데이터 및 입력 데이터에 기초하여 목표 사용자에게 추천할 항목을 추출한다. 그리고 입력 데이터는 목표 사용자 및 복수의 타 사용자의 복수의 항목 각각에 대한 선호도를 포함한다.

[0008] 또한, 본 발명의 제 2 측면에 따른 추천 장치의 사용자를 위한 항목 추천 방법은 입력 데이터에 대한 클러스터링을 수행하는 단계; 클러스터링 결과에 기초하여, 입력 데이터에 포함된 복수의 항목의 데이터에 대한 히스토그램을 생성하는 단계; 입력 데이터 및 히스토그램에 기초하여 목표 사용자에 대한 신규 데이터를 생성하는 단계; 및 생성된 신규 데이터 및 입력 데이터에 기초하여 목표 사용자에게 추천할 항목을 추출하는 단계를 포함한다. 이때, 입력 데이터는 목표 사용자 및 복수의 타 사용자의 복수의 항목 각각에 대한 선호도를 포함한다.

발명의 효과

[0009] 본 발명은 입력 데이터에 대한 클러스터링 및 히스토그램에 기초하여, 사용자에게 항목을 추천할 수 있다. 그러므로 본 발명은 사용자 수가 항목의 수보다 많거나, 사용자가 선호도를 입력한 항목의 수가 적어 발생할 수 있는 희박성(sparsity) 문제를 해결할 수 있다. 또한, 본 발명은 선호도가 입력된 항목의 수가 비교적 적은 초기 사용자의 추천 정확도가 낮아지는 초기 사용자(cold start) 문제를 해결할 수 있다. 본 발명은 기계학습 기법에 기초하여 클러스터링을 수행하므로 항목의 수가 많은 경우에도 효율적으로 추천을 수행할 수 있다.

도면의 간단한 설명

[0010] 도 1 은 본 발명의 일 실시예 에 따른 항목 추천 장치의 블록도이다.

도 2는 본 발명의 일 실시예에 따른 입력 데이터의 예시도이다.

도 3은 본 발명의 일 실시예에 따른 k-근접 이웃 기반 신규 데이터 생성 과정의 예시도이다.

도 4는 본 발명의 일 실시예에 따른 클러스터링 알고리즘 기반 신규 데이터 생성 과정의 예시도이다.

도 5는 본 발명의 일 실시예에 따른 추천 장치의 항목 추천 방법에 대한 순서도이다.

발명을 실시하기 위한 구체적인 내용

[0011] 아래에서는 첨부한 도면을 참조하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 본 발명의 실시예를 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.

[0012] 명세서 전체에서, 어떤 부분이 다른 부분과 "연결"되어 있다고 할 때, 이는 "직접적으로 연결"되어 있는 경우뿐 아니라, 그 중간에 다른 소자를 사이에 두고 "전기적으로 연결"되어 있는 경우도 포함한다. 또한, 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.

[0013] 다음은 도 1 내지 도 4를 참조하여, 본 발명의 일 실시예에 따른 항목 추천 장치(100)를 설명한다.

[0014] 도 1 은 본 발명의 일 실시예에 따른 항목 추천 장치(100)의 블록도이다.

[0015] 항목 추천 장치(100)는 복수의 항목 중 사용자에게 사용자가 선호할 만한 항목을 추천한다. 이때, 항목 추천

장치(100)는 메모리(110) 및 프로세서(130)를 포함한다.

[0016] 메모리(110)는 항목에 대한 추천 프로그램이 저장된다. 이때, 메모리(110)는 전원이 공급되지 않아도 저장된 정보를 계속 유지하는 비휘발성 저장장치 및 저장된 정보를 유지하기 위하여 전력이 필요한 휘발성 저장장치를 통칭하는 것이다.

[0017] 또한, 항목 추천 장치(100)는 데이터베이스(120)를 더 포함할 수 있다. 이때, 데이터베이스(120)는 항목에 대한 사용자의 선호도를 포함하는 데이터 또는 항목에 대한 사용자의 구입 여부를 저장한 데이터를 저장할 수 있다. 또한, 데이터베이스(120)는 사용자의 인구통계학 데이터를 포함하는 사용자 정보 데이터 및 항목에 대한 내용 정보를 포함하는 항목 정보 데이터를 저장할 수 있다.

[0018] 이때, 사용자의 인구통계학 데이터는 사용자의 성별, 나이, 거주 지역 및 학력 등이 될 수 있다. 또한, 항목에 대한 내용 정보는 항목의 유형, 생산 년도 및 장르 등이 될 수 있다.

[0019] 항목은 사용자가 구매하는 각종 제품, 상품 및 서적 등이 될 수 있다. 또한, 항목은 영화, 음악, 신문 기사 및 웹 페이지 등이 될 수 있다. 그러므로 항목은 사용자가 구매하거나, 선호도를 입력할 수 있는 모든 종류의 재화 또는 서비스가 될 수 있으나, 이에 한정된 것은 아니다.

[0020] 이때, 선호도는 사용자가 항목에 대하여 입력한 선호 점수일 수 있다. 예를 들어, 선호도는 5점 척도에 따라 입력된 점수일 수 있다. 이때, 5점 척도에서 선호도는 1점에 가까울수록 불호를 의미하며, 5점에 가까울수록 선호에 가까울 수 있다.

[0021] 또한, 선호도는 사용자가 해당 항목을 선호하는지 여부 또는 구입했는지 여부가 될 수 있다. 예를 들어, 선호도는 해당 항목을 구입한 경우 1 또는 'true'가 될 수 있으며, 구입하지 않은 경우 0 또는 'false'가 될 수 있다.

[0022] 한편, 프로세서(130)는 데이터베이스(120)에 저장된 데이터를 이용하여 입력 데이터를 생성할 수 있다. 이때, 입력 데이터는 각 사용자 별 항목에 대한 선호도를 포함하는 레코드(recode)의 집합일 수 있다. 입력 데이터는 도 2를 참조하여 설명한다.

[0023] 도 2는 본 발명의 일 실시예에 따른 입력 데이터의 예시도이다.

[0024] 예를 들어, 입력 데이터는 n 명의 사용자 및 m 명의 항목에 대한 행렬 형태로 표현될 수 있다. 이때, 특정 사용자가 항목에 대한 선호도를 입력한 경우 입력 데이터는 특정 사용자의 해당 항목에 선호도를 포함할 수 있다.

[0025] 이와 반대로 특정 사용자의 항목에 대한 선호도가 존재하지 않은 경우, 입력 데이터는 특정 사용자의 해당 항목의 선호도가 없거나, 선호도가 없다는 의미의 '0' 또는 '-1'과 같은 값을 포함할 수 있다. 이때, 특정 사용자의 선호도가 없는 항목은 해당 사용자가 선호도를 아직 입력하지 않은 항목일 수 있다. 또한, 선호도가 없는 항목은 해당 사용자가 아직 구입하지 못했거나 서비스 받지 못한 항목일 수 있다.

[0026] 이하에서 항목 추천 장치(100)는 복수의 항목 중 선호도가 없는 항목을 사용자가 아직 구입하지 못했거나 서비스 받지 못한 항목으로 판단한다. 그러므로 항목 추천 장치(100)는 추천 목표 사용자의 선호도가 없는 항목을 목표 사용자에 대응하는 추천 대상 항목으로 선택할 수 있다.

[0027] 다시 도 2를 참조하면, 목표 사용자(200)인 제 2 사용자의 추천 대상 항목은 제 2 사용자가 아직 선호도를 입력하지 않은 제 1 항목 및 제 3 항목이 될 수 있다. 즉, 제 1 항목은 목표 사용자(200)의 제 1 목표 항목(210)이 될 수 있으며, 제 3 항목은 목표 사용자(200)의 제 2 목표 항목(220)이 될 수 있다.

[0028] 그러므로 프로세서(130)는 목표 사용자(200)와 유사한 다른 사용자의 항목에 대한 선호도에 기초하여, 제 1 목표 항목(210) 및 제 2 목표 항목(220)에 대한 선호도를 예측할 수 있다. 그리고 프로세서(130)는 제 1 목표 항목(210) 및 제 2 목표 항목(220) 중 선호도가 큰 순으로 목표 사용자(200)에게 추천할 수 있다.

[0029] 또는, 프로세서(130)는 제 1 목표 항목(210)과 유사한 다른 항목의 선호도에 기초하여 제 1 목표 항목(210)의 선호도를 예측할 수 있다. 그리고 프로세서(130)는 제 2 목표 항목(220)과 유사한 다른 항목의 선호도에 기초하여, 제 2 목표 항목(220)의 선호도를 예측할 수 있다. 그리고 프로세서(130)는 제 1 목표 항목(210) 및 제 2 목표 항목(220) 중 선호도가 큰 순으로 목표 사용자(200)에게 추천할 수 있다.

[0030] 한편, 프로세서(130)는 추출된 입력 데이터에 대한 클러스터링을 수행한다. 그리고 프로세서(130)는 클러스터

링 결과에 기초하여 입력 데이터에 포함된 복수의 항목에 대한 히스토그램을 생성한다. 프로세서(130)는 입력 데이터 및 히스토그램에 기초하여 목표 사용자에 대한 신규 데이터를 생성할 수 있다. 신규 데이터 생성과정은 도 3 및 도 4를 참조하여 구체적으로 설명한다.

[0031] 프로세서(130)는 입력 데이터를 이용하여 클러스터링을 수행할 수 있다. 이때, 클러스터링은 k-근접 이웃(k-nearest neighbor) 알고리즘 일 수 있다.

[0032] 도 3은 본 발명의 일 실시예에 따른 k-근접 이웃 기반 신규 데이터 생성 과정의 예시도이다.

[0033] 프로세서(130)는 k-근접 이웃 알고리즘에 기초하여 목표 사용자와 유사한 미리 정해진 사용자 수의 사용자의 입력 데이터를 선택할 수 있다. 예를 들어, 미리 정해진 사용자 수가 2인 경우, 프로세서(130)는 추천 대상이 되는 목표 사용자인 제 2 사용자와 유사한 2명의 사용자를 선택할 수 있다.

[0034] 이때, 유사한 사용자는 유사도 척도를 이용하여 선택할 수 있다. 유사도 척도는 피어슨 상관계수(Pearson's correlation coefficient), 자카드 계수(jaccard coefficient), 스피어만 상관계수(Spearman' s correlation coefficient), 유클리드 거리(Euclidean distance), 코사인(cosine) 유사도 등이 될 수 있으며, 이에 한정된 것은 아니다.

[0035] 도 3의 (a)를 참조하면, 프로세서(130)는 유사도에 기초하여, 제 1 사용자 및 제 4 사용자를 선택할 수 있다. 그리고 프로세서(130)는 선택된 제 1 사용자 및 제 4 사용자를 제 1 클러스터로 생성할 수 있다.

[0036] 도 3의 (b)를 참조하면, 프로세서(130)는 제 1 클러스터에 포함된 각 사용자의 각 항목에 대한 선호 정도에 기초하여 히스토그램을 생성할 수 있다. 이때, 히스토그램은 제 1 클러스터에 포함된 제 1 사용자의 각 항목에 대한 선호도 및 제 4 사용자의 각 항목에 대한 선호도의 평균값에 기초하여 산출될 수 있다. 그리고 프로세서(130)는 히스토그램에 포함된 각 항목의 값에 기초하여 제 2 사용자에 대응하는 신규 데이터를 생성할 수 있다.

[0037] 만약 제 3 사용자에 대한 신규 데이터를 추가로 생성하는 경우, 프로세서(130)는 제 3 사용자에 대하여 k-근접 이웃 알고리즘을 적용하여 제 3 사용자와 유사한 2명의 사용자를 선택하여, 새로운 클러스터를 생성할 수 있다. 그리고 프로세서(130)는 새로운 클러스터의 히스토그램에 기초하여 제 3 사용자에 대한 신규 데이터를 생성할 수 있다.

[0038] 이와 같이, 프로세서(130)는 추천을 수행하고자 하는 하나 이상의 사용자에 대하여, 각각 k-근접 이웃 알고리즘을 적용하여 히스토그램을 생성할 수 있다.

[0039] 도 4는 본 발명의 일 실시예에 따른 클러스터링 알고리즘 기반 신규 데이터 생성 과정의 예시도이다.

[0040] 도 4를 참조하면, 프로세서(130)는 k-근접 이웃 알고리즘 대신 기계학습(machine learning) 기반 클러스터링 알고리즘을 사용할 수 있다. 이때, 클러스터링 알고리즘은 k 평균(k-means) 알고리즘, EM(expectation-maximization) 알고리즘 및 DBSCAN 알고리즘 등이 될 수 있으나, 이에 한정된 것은 아니다.

[0041] 도 4의 (b)를 참조하면 프로세서(130)는 입력 데이터에 대한 클러스터링 알고리즘에 기초하여 복수의 클러스터를 생성할 수 있다. 이때, 프로세서(130)는 입력 데이터만을 이용하여 클러스터를 생성할 수 있다. 또는, 프로세서(130)는 입력 데이터 및 입력 데이터에 대응하는 사용자의 정보 또는 항목의 정보를 이용하여 클러스터를 생성할 수 있다.

[0042] 그리고 도 4의 (c)를 참조하면 프로세서(130)는 복수의 클러스터 각각에 대한 히스토그램을 생성할 수 있다. 이때, 히스토그램은 해당 클러스터에 포함된 복수의 사용자의 항목에 대한 항목의 선호도의 평균값에 기초하여 산출될 수 있다. 그리고 프로세서(130)는 복수의 클러스터 중 목표 사용자가 포함된 클러스터의 히스토그램에 기초하여 목표 사용자의 신규 데이터를 생성할 수 있다.

[0043] 도 4의 (d)를 참조하면, 프로세서(130)는 입력 데이터에 대한 클러스터링 알고리즘을 통하여, p개의 클러스터를 생성할 수 있다. 그리고 프로세서(130)는 p개의 클러스터 각각에 대하여, 히스토그램을 생성할 수 있다.

[0044] 그러므로 목표 사용자인 제 2 사용자가 제 1 클러스터에 포함되는 경우, 프로세서(130)는 제 1 클러스터를 통하여 생성된 히스토그램에 기초하여, 제 2 사용자의 신규 데이터를 생성할 수 있다.

[0045] 만약 제 3 사용자에 대한 신규 데이터를 추가로 생성하는 경우, 프로세서(130)는 생성된 p개의 클러스터 중 제 3 사용자가 포함된 클러스터를 선택하고, 선택된 클러스터의 히스토그램에 기초하여 제 3 사용자의 신규 데이터를 생성할 수 있다.

[0046] 이와 같이, 프로세서(130)는 클러스터링 알고리즘을 이용하는 경우, 입력 데이터 및 클러스터링 알고리즘에 기초하여 클러스터를 생성한 이후 생성된 클러스터를 이용하여 추천을 수행하고자 하는 복수의 사용자에 대하여 신규 데이터를 생성할 수 있다.

[0047] 한편, 특정 클러스터의 히스토그램은 가로축이 항목, 세로축이 항목에 대한 해당 클러스터의 선호도가 될 수 있다. 이때, 해당 클러스터의 선호도는 해당 클러스터에 포함된 사용자의 전체 항목에 대한 선호도에 기초하여 정규화된 항목에 대한 선호도일 수 있다.

[0048] 프로세서(130)는 입력 데이터에 포함된 목표 사용자의 복수의 항목에 대한 선호도 및 히스토그램을 통하여 목표 사용자에 대한 신규 데이터를 생성할 수 있다. 이때, 프로세서(130)는 입력 데이터에 포함된 목표 사용자의 복수의 항목에 대한 선호도 및 히스토그램을 통하여 산출된 복수의 항목에 대한 선호도에 대하여 각각 가중치를 부여할 수 있다. 그리고 프로세서(130)는 가중치에 기초하여 신규 데이터를 생성할 수 있다.

[0049] 각각의 가중치는 미리 정해진 값이 될 수 있다. 예를 들어, 각각의 가중치는 항목 추천 장치(100)의 관리자를 통하여 정해질 수 있다.

[0050] 또는, 각각의 가중치는 추천 정확도에 기초하여, 설정될 수 있다.

[0051] 예를 들어, 프로세서(130)는 입력 데이터에 포함된 목표 사용자의 일부 데이터를 테스트 데이터로 추출할 수 있다. 이때, 테스터 데이터는 목표 사용자가 선호도를 입력한 항목 중 하나 이상이 될 수 있다. 그리고 프로세서(130)는 입력 데이터 중 테스트 데이터를 제외한 나머지 데이터를 이용하여, 앞에서 설명한 바와 같이, 클러스터 및 히스토그램을 생성할 수 있다.

[0052] 프로세서(130)는 생성된 히스토그램 및 나머지 데이터에 기초하여, 신규 데이터를 생성할 수 있다. 이때, 프로세서(130)는 히스토그램 및 나머지 데이터의 가중치를 조절하여 복수의 신규 데이터를 생성할 수 있다. 그리고 프로세서(130)는 복수의 신규 데이터에 기초하여, 테스터 데이터에 포함된 항목에 대한 선호도를 예측할 수 있다.

[0053] 그리고 프로세서(130)는 예측된 선호도 및 실제 목표 사용자의 선호도에 기초하여 복수의 신규 데이터에 대한 정확도를 산출한다. 프로세서(130)는 정확도가 가장 큰 신규 데이터에 적용된 가중치를 선택할 수 있다. 이때, 정확도는 각 항목에 대한 예측 선호도 및 실제 선호도의 일치 여부 또는 유사성을 산출하는 것일 수 있다. 또는, 정확도는 추천된 항목 중 실제 목표 사용자가 선호도를 입력한 항목의 수를 산출하는 것일 수 있으나, 이에 한정되는 것은 아니다.

[0054] 가중치가 설정되면, 프로세서(130)는 입력 데이터에 포함된 복수의 항목에 대한 선호도 및 히스토그램을 통하여 산출된 복수의 항목에 대한 선호도의 가중합을 산출할 수 있다. 그리고 프로세서(130)는 산출된 가중합에 기초하여 신규 데이터를 생성할 수 있다.

[0055] 한편, 프로세서(130)는 신규 데이터를 이용하여 목표 사용자의 추천 대상 항목에 대한 선호도를 예측할 수 있다. 이때, 프로세서(130)는 사용자 기반 추천 방법 및 항목 기반 추천 방법 중 어느 하나를 사용할 수 있다.

[0056] 만약, 사용자 기반 추천 기법을 사용하는 경우, 프로세서(130)는 추천 알고리즘 및 신규 데이터를 이용하여 복수의 유사 사용자를 선택하여 추천을 수행할 수 있다. 이때, 추천 알고리즘은 앞에서 설명한 k-근접 이웃 외에도 연관규칙(association rule) 기반 추천 알고리즘, 신경망(neural network) 기반 추천 알고리즘 및 행렬 분해(matrix factorization) 기반 추천 알고리즘 등이 될 수 있으며, 이에 한정되는 것은 아니다.

[0057] 예를 들어, 추천 알고리즘이 k-근접 이웃인 경우, 프로세서(130)는 앞에서 설명한 유사도 척도를 이용하여, 신규 데이터와 유사한 사용자를 미리 정해진 유사 사용자 수만큼 선택할 수 있다. 또한, 프로세서(130)는 선택된 사용자의 추천 대상 항목에 대한 선호도에 기초하여, 목표 사용자의 추천 대상 항목에 대한 선호도를 예측할 수 있다. 그리고 프로세서(130)는 추출된 추천 대상 항목에 대한 선호도 중 상위 n개를 추출하여, 추천 할 수 있다. 이때, n은 추천을 수행하는 항목의 수에 기초하여 변경될 수 있다.

[0058] 한편, 항목 기반으로 목표 사용자에게 항목을 추천하기 위하여, 프로세서(130)는 클러스터링을 수행하기 전에 입력 데이터에 기초하여 복수의 항목 간의 유사도를 산출할 수 있다. 이때, 유사도는 앞에서 설명한 유사도 척도 중 어느 하나를 이용할 수 있다.

[0059] 그리고 프로세서(130)는 앞에서 설명한 바와 같이, 입력 데이터에 대한 클러스터링을 적용하여 입력 데이터에 포함된 복수의 항목에 대한 히스토그램을 생성할 수 있다. 또한, 프로세서(130)는 입력 데이터 및 히스토그램

에 기초하여 목표 사용자에 대한 신규 데이터를 생성할 수 있다.

[0060] 프로세서(130)는 목표 사용자에 대응하는 신규 데이터 및 산출된 복수의 항목 간의 유사도를 이용하여 추천 알고리즘을 적용하여 추천 대상 항목 각각에 대한 선호도를 예측할 수 있다. 그리고 프로세서(130)는 추천 대상 항목 각각에 예측된 선호도에 기초하여, 목표 사용자의 추천 대상 항목 중 목표 사용자에게 추천할 항목을 선택할 수 있다.

[0061] 예를 들어, 추천 알고리즘이 k-근접 이웃인 경우, 프로세서(130)는 미리 산출된 복수의 항목 간 유사도에 기초하여 추천 대상 항목 각각에 대하여 유사한 항목을 미리 정해진 유사 항목 수만큼 선택할 수 있다.

[0062] 그리고 프로세서(130)는 각각의 추천 대상 항목에 대하여, 선호도를 예측할 수 있다. 그리고 프로세서(130)는 예측된 추천 대상 항목에 대한 선호도 중 상위 n개를 추출하여 목표 사용자에게 추천 할 수 있다.

[0063] 다음은 도 5를 참조하여, 본 발명의 일 실시예에 따른 추천 장치(100)의 항목 추천 방법을 설명한다.

[0064] 도 5는 본 발명의 일 실시예에 따른 추천 장치(100)의 항목 추천 방법에 대한 순서도이다.

[0065] 추천 장치(100)는 입력 데이터에 대한 클러스터링을 수행한다(S500). 이때, 입력 데이터는 목표 사용자 및 복수의 타 사용자의 복수의 항목 각각에 대한 선호도를 포함한다.

[0066] 그리고 추천 장치(100)는 클러스터링 결과에 기초하여, 입력 데이터에 포함된 복수의 항목에 대한 히스토그램을 생성한다(S510).

[0067] 추천 장치(100)는 입력 데이터 및 히스토그램에 기초하여 목표 사용자에 대한 신규 데이터를 생성한다(S520). 이때, 추천 장치(100)는 히스토그램을 통하여 산출된 복수의 항목의 선호도 및 입력 데이터에 포함된 목표 사용자에 대한 복수의 항목의 선호도에 기초하여 신규 데이터를 생성할 수 있다.

[0068] 또한, 추천 장치(100)는 생성된 신규 데이터 및 입력 데이터에 기초하여 목표 사용자에게 추천할 항목을 추출한다(S530).

[0069] 이때, 추천 장치(100)는 신규 데이터 및 입력 데이터에 기초하여 복수의 추천 대상 항목에 대한 각각의 선호도를 예측할 수 있다. 그리고 추천 장치(100)는 복수의 추천 대상 항목 중 예측된 선호도가 큰 미리 정해진 항목 수의 항목을 추출할 수 있다. 이때, 추천 대상 항목은 입력 데이터 중 목표 사용자가 선호도를 입력하지 않은 항목일 수 있다.

[0070] 또한, 신규 데이터 및 입력 데이터에 기초하여 복수의 추천 대상 항목에 각각의 선호도를 예측하기 위하여, 추천 장치(100)는 신규 데이터 및 입력 데이터에 기초하여, 목표 사용자와 복수의 타 사용자 간의 유사도를 산출할 수 있다. 그리고 추천 장치(100)는 유사도가 큰 순서대로 미리 정해진 유사 사용자 수에 대응하는 사용자를 선택할 수 있다. 추천 장치(100)는 선택된 사용자에 기초하여, 추천 대상 항목에 대한 선호도를 예측할 수 있다.

[0071] 한편, 추천 장치(100)는 입력 데이터에 포함된 복수의 항목 간의 유사도를 산출할 수 있다. 그리고 추천 장치(100)는 신규 데이터 및 입력 데이터에 기초하여 복수의 추천 대상 항목에 대한 각각의 선호도를 예측하기 위하여, 복수의 항목 간의 유사도, 상기 생성된 신규 데이터 및 입력 데이터에 기초하여 추천 대상 항목에 대한 선호도를 예측할 수 있다.

[0072] 본 발명의 일 실시예에 따른 사용자를 위한 항목을 추천하는 장치 및 방법은 입력 데이터에 대한 클러스터링 및 히스토그램에 기초하여, 사용자에게 항목을 추천할 수 있다. 그러므로 사용자를 위한 항목을 추천하는 장치 및 방법은 사용자 수가 항목의 수보다 많거나, 사용자가 선호도를 입력한 항목의 수가 적어 발생할 수 있는 희박성(sparsity) 문제를 해결할 수 있다. 또한, 사용자를 위한 항목을 추천하는 장치 및 방법은 선호도가 입력된 항목의 수가 비교적 적은 초기 사용자의 추천 정확도가 낮아지는 초기 사용자(cold start) 문제를 해결할 수 있다. 사용자를 위한 항목을 추천하는 장치 및 방법은 기계학습 기법에 기초하여 클러스터링을 수행하므로 항목의 수가 많은 경우에도 효율적으로 추천을 수행할 수 있다.

[0073] 본 발명의 일 실시예는 컴퓨터에 의해 실행되는 프로그램 모듈과 같은 컴퓨터에 의해 실행가능한 명령어를 포함하는 기록 매체의 형태로도 구현될 수 있다. 컴퓨터 판독 가능 매체는 컴퓨터에 의해 액세스될 수 있는 임의의 가용 매체일 수 있고, 휘발성 및 비휘발성 매체, 분리형 및 비분리형 매체를 모두 포함한다. 또한, 컴퓨터 판독가능 매체는 컴퓨터 저장 매체를 포함할 수 있다. 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현된 휘발성 및

비휘발성, 분리형 및 비분리형 매체를 모두 포함한다.

[0074] 본 발명의 방법 및 시스템은 특정 실시예와 관련하여 설명되었지만, 그것들의 구성 요소 또는 동작의 일부 또는 전부는 범용 하드웨어 아키텍처를 갖는 컴퓨터 시스템을 사용하여 구현될 수 있다.

[0075] 전술한 본 발명의 설명은 예시를 위한 것이며, 본 발명이 속하는 기술분야의 통상의 지식을 가진 자는 본 발명의 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 쉽게 변형이 가능하다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다. 예를 들어, 단일형으로 설명되어 있는 각 구성 요소는 분산되어 실시될 수도 있으며, 마찬가지로 분산된 것으로 설명되어 있는 구성 요소들도 결합된 형태로 실시될 수 있다.

[0076] 본 발명의 범위는 상기 상세한 설명보다는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

부호의 설명

[0077] 100: 항목 추천 장치

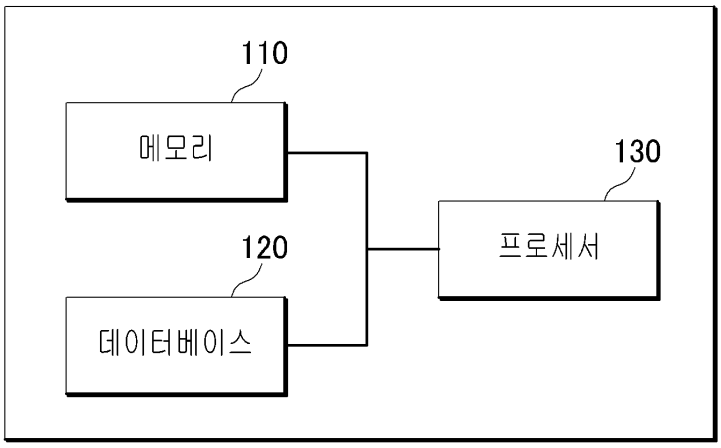
110: 메모리

120: 데이터베이스

130: 프로세서

도면

도면1



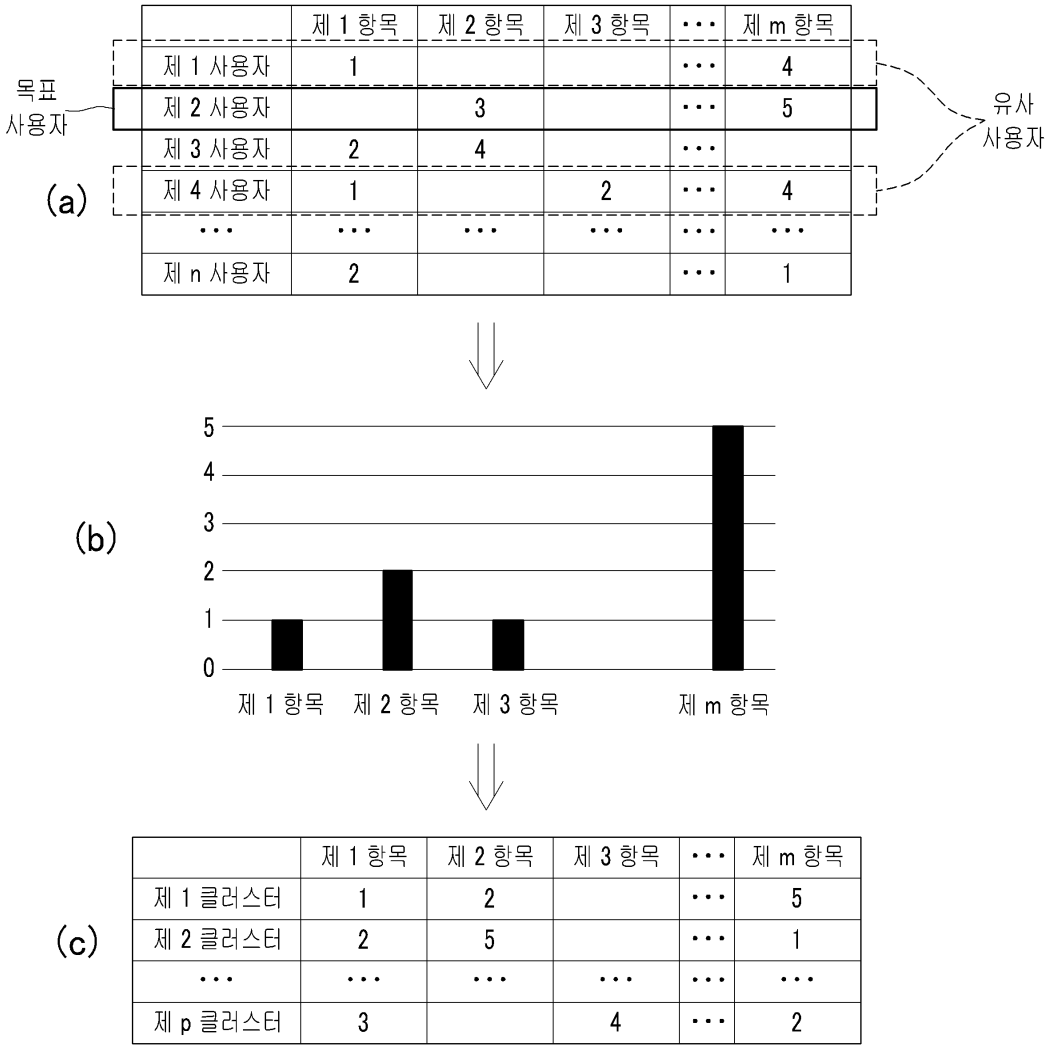
100

도면2



	제 1 항목	제 2 항목	제 3 항목	···	제 m 항목
제 1 사용자	1			···	4
제 2 사용자		3		···	5
제 3 사용자	2	4		···	
제 4 사용자	1		2	···	4
···	···	···	···	···	···
제 n 사용자	2			···	1

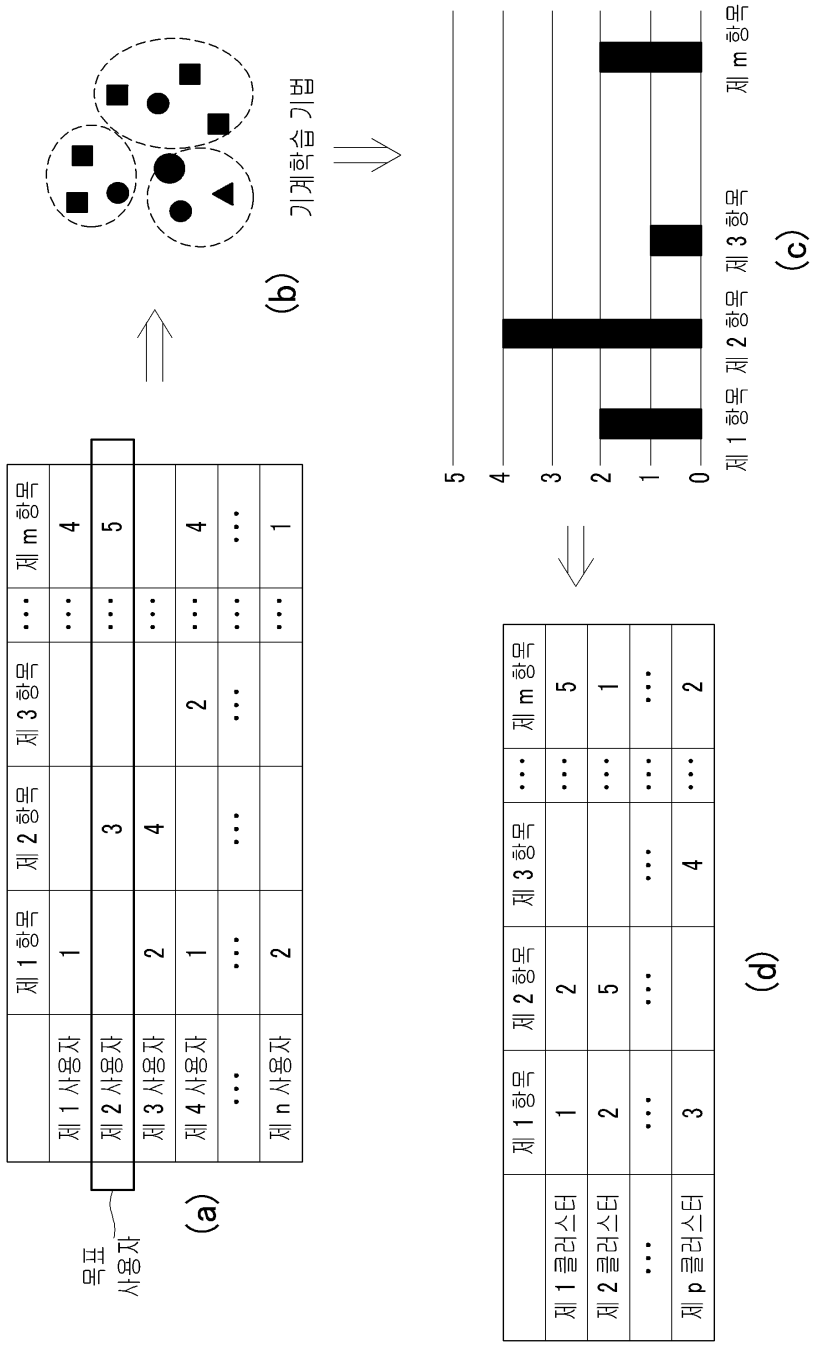
도면3



	제 1 항목	제 2 항목	제 3 항목	···	제 m 항목
제 1 사용자	1			···	4
제 2 사용자		3		···	5
제 3 사용자	2	4		···	
제 4 사용자	1		2	···	4
···	···	···	···	···	···
제 n 사용자	2			···	1

	제 1 항목	제 2 항목	제 3 항목	···	제 m 항목
제 1 클러스터	1	2		···	5
제 2 클러스터	2	5		···	1
···	···	···	···	···	···
제 p 클러스터	3		4	···	2

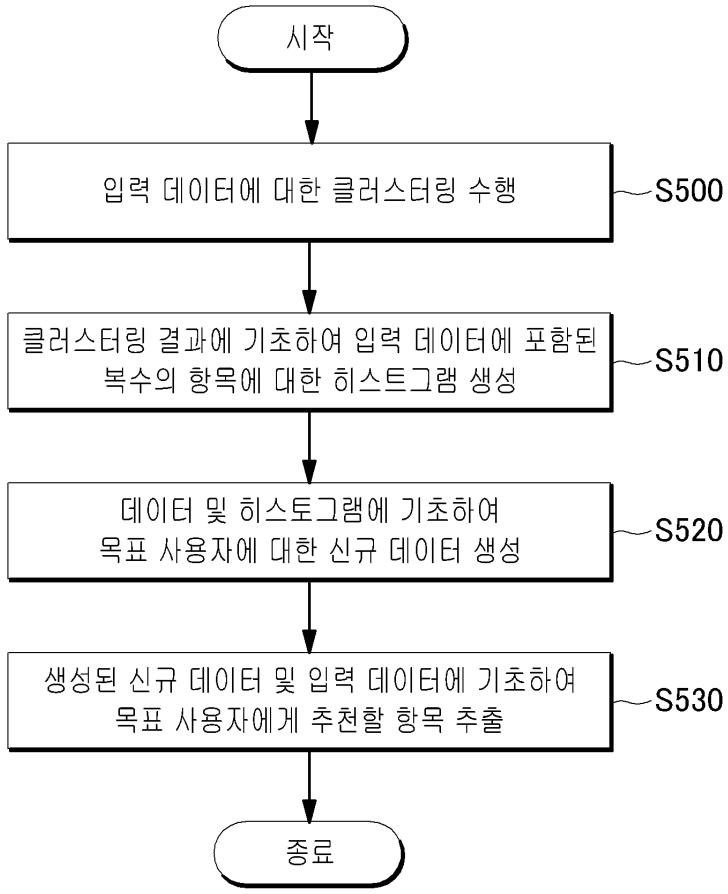
도면4



	제 1 항목	제 2 항목	제 3 항목	...	제 m 항목
제 1 사용자	1			...	4
제 2 사용자		3		...	5
제 3 사용자	2	4		...	
제 4 사용자	1		2	...	4
...	...	...	...	...	...
제 n 사용자	2			...	1

	제 1 항목	제 2 항목	제 3 항목	...	제 m 항목
제 1 클러스터	1	2		...	5
제 2 클러스터	2	5		...	1
...	...	...	...	...	...
제 p 클러스터	3		4	...	2

도면5


시작
입력 데이터에 대한 클러스터링 수행
S500
클러스터링 결과에 기초하여 입력 데이터에 포함된 복수의 항목에 대한 히스트그램 생성
S510
데이터 및 히스토그램에 기초하여 목표 사용자에 대한 신규 데이터 생성
S520
생성된 신규 데이터 및 입력 데이터에 기초하여 목표 사용자에게 추천할 항목 추출
S530
종료