

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0152013

(43) 공개일자 2022년11월15일

(51) 국제특허분류(Int. Cl.)

G06Q 30/06 (2012.01) G06N 20/00 (2019.01)

G06Q 30/02 (2012.01)

(52) CPC특허분류

G06Q 30/0631 (2013.01)

G06N 20/00 (2021.08)

(21) 출원번호 10-2021-0059438

(22) 출원일자 2021년05월07일

심사청구일자 2021년05월07일

(71) 출원인

경희대학교 산학협력단

경기도 용인시 기흥구 덕영대로 1732 (서천동, 경희대학교 국제캠퍼스내)

(72) 발명자

홍민성

서울특별시 성동구 천호대로 416, 힐데스하임동 801호(용답동, 힐데스하임)

정남호

서울특별시 서초구 동광로 110, 201호(방배동, 현대힐스빌라)

구철모

서울특별시 서대문구 이화여대길 50-12, 102동 1103호(대현동, 럭키대현아파트)

(74) 대리인

김연권

전체 청구항 수 : 총 10 항

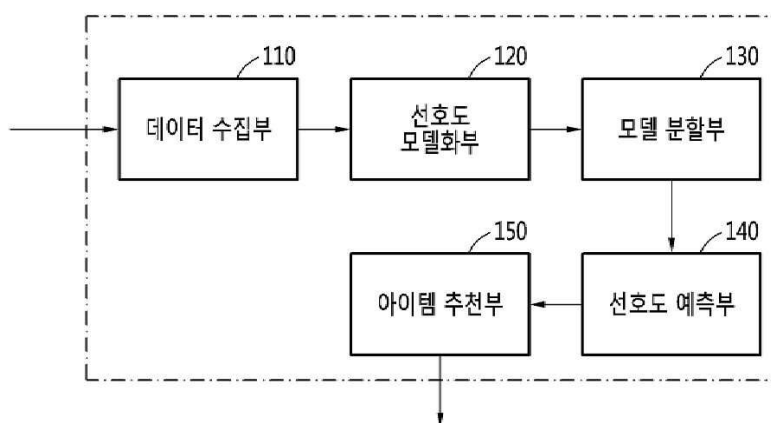
(54) 발명의 명칭 리뷰 빅데이터 기반의 아이템 추천 장치 및 그 방법

(57) 요약

본 발명은 리뷰 빅데이터 기반의 아이템 추천 장치 및 그 방법에 관한 것으로서, 일실시예에 따른 아이템 추천 장치는 복수의 아이템에 대한 복수의 사용자의 리뷰 데이터를 수집하는 데이터 수집부와, 수집된 리뷰 데이터에 대한 선호도 분석과 감성 분석을 수행하여 선호도 텐서 모델을 생성하는 선호도 모델화부와, 생성된 선호도 텐서 모델에 기초한 유사도 측정을 통해 생성된 선호도 텐서 모델을 복수의 하위 모델로 분할하는 모델 분할부와, 분할된 하위 모델에서 복수의 사용자 중 타겟 사용자의 리뷰 데이터가 포함된 적어도 하나의 하위 모델을 선정하고, 선정된 하위 모델 각각에 대한 타겟 사용자의 선호도를 예측하는 선호도 예측부 및 선호도의 예측 결과에 기초하여 복수의 아이템 중 적어도 하나의 아이템을 추천하는 아이템 추천부를 포함한다.

대표도 - 도1

100



(52) CPC특허분류

G06Q 30/0278 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1345327830
과제번호	2019S1A3A2098438
부처명	교육부
과제관리(전문)기관명	한국연구재단
연구사업명	학술·인문사회사업/집단연구/사회과학연구지원사업/한국사회과학(SSK)지원사업
연구과제명	스마트 관광도시의 지속가능발전을 위한 실천과 확산
기 여 율	1/1
과제수행기관명	경희대학교 산학협력단
연구기간	2019.09.01 ~ 2023.08.31

명세서

청구범위

청구항 1

복수의 아이템에 대한 복수의 사용자의 리뷰 데이터를 수집하는 데이터 수집부;

상기 수집된 리뷰 데이터에 대한 선호도 분석과 감성 분석을 수행하여 선호도 텐서 모델을 생성하는 선호도 모델화부;

상기 생성된 선호도 텐서 모델에 기초한 유사도 측정을 통해 상기 생성된 선호도 텐서 모델을 복수의 하위 모델로 분할하는 모델 분할부;

상기 분할된 하위 모델에서 상기 복수의 사용자 중 타겟 사용자의 리뷰 데이터가 포함된 적어도 하나의 하위 모델을 선정하고, 상기 선정된 하위 모델 각각에 대한 상기 타겟 사용자의 선호도를 예측하는 선호도 예측부 및

상기 선호도의 예측 결과에 기초하여 상기 복수의 아이템 중 적어도 하나의 아이템을 추천하는 아이템 추천부를 포함하는 아이템 추천 장치.

청구항 2

제1항에 있어서,

상기 리뷰 데이터는,

상기 복수의 아이템 각각에 대한 복수의 선호도 종류 정보 및 상기 복수의 선호도 종류 정보 각각에 대응되는 평점 정보를 포함하는

아이템 추천 장치.

청구항 3

제1항에 있어서,

상기 선호도 모델화부는,

상기 수집된 리뷰 데이터에 대한 상기 선호도 분석을 통해 다차원 텐서 모델을 생성하고, 상기 수집된 리뷰 데이터에 대한 상기 감성 분석의 수행 결과를 이용하여 상기 생성된 다차원 텐서 모델을 보완하여 상기 선호도 텐서 모델을 생성하는

아이템 추천 장치.

청구항 4

제3항에 있어서,

상기 선호도 모델화부는,

플래어엘모(FlairElmo), 텍스트블랍(Textblob), 베이더(Vader), 로지스틱 회귀(Logistic Regression), 패스트 텍스트(Fasttext) 및 서포트 벡터 머신(support vector machine) 중 적어도 하나에 기초한 상기 감성 분석의 수행 결과에 기초하여 상기 생성된 다차원 텐서 모델을 보완하는

아이템 추천 장치.

청구항 5

제2항에 있어서,

상기 모델 분할부는,

상기 생성된 선호도 텐서 모델에 구비되는 선호도 종류 정보와 평점 정보에 대한 유사도 거리 연산을 수행하고,

상기 유사도 거리 연산의 수행 결과에 기초하여 상기 생성된 선호도 텐서 모델을 복수의 제1 하위 모델로 분할하는

아이템 추천 장치.

청구항 6

제5항에 있어서,

상기 모델 분할부는,

상기 분할된 제1 하위 모델 중 기설정된 크기 비율을 초과하는 모델을 제2 하위 모델로 분할하는

아이템 추천 장치.

청구항 7

제5항에 있어서,

상기 모델 분할부는,

유클리드 거리 및 코사인 유사도 중 적어도 하나에 기초한 상기 유사도 거리 연산을 수행하는

아이템 추천 장치.

청구항 8

제1항에 있어서,

상기 선호도 예측부는,

상기 선정된 하위 모델 각각에 대하여 텐서 분해를 병렬로 수행하여, 상기 선정된 하위 모델 각각에 대한 선호도를 동시에 예측하는

아이템 추천 장치.

청구항 9

제8항에 있어서,

상기 아이템 추천부는,

상기 텐서 분해를 통해 도출되는 상기 선정된 하위 모델 각각의 선호도 값에 대하여 합산 연산 및 평균값 연산 중 적어도 하나의 연산을 수행하고, 상기 연산의 수행 결과에 기초하여 기설정된 개수의 아이템을 추천하는

아이템 추천 장치.

청구항 10

데이터 수집부에서, 복수의 아이템에 대한 복수의 사용자의 리뷰 데이터를 수집하는 단계;

선호도 모델화부에서, 상기 수집된 리뷰 데이터에 대한 선호도 분석과 감성 분석을 수행하여 선호도 텐서 모델을 생성하는 단계;

모델 분할부에서, 상기 생성된 선호도 텐서 모델에 구비된 리뷰 데이터에 기초한 유사도 측정을 통해 상기 생성된 선호도 텐서 모델을 복수의 하위 모델로 분할하는 단계;

선호도 예측부에서, 상기 분할된 하위 모델에서 상기 복수의 사용자 중 타겟 사용자의 리뷰 데이터가 포함된 적어도 하나의 하위 모델을 선정하고, 상기 선정된 하위 모델에 대한 상기 타겟 사용자의 선호도를 예측하는 단계 및

아이템 추천부에서, 상기 선호도의 예측 결과에 기초하여 상기 복수의 아이템 중 적어도 하나의 아이템을 추천하는 단계

를 포함하는 아이템 추천 방법.

발명의 설명

기술 분야

[0001] 본 발명은 아이템 추천 장치 및 그 방법에 관한 것으로, 보다 상세하게는 빅데이터 환경에서 분할 감소 기법에 기초하여 실시간으로 아이템을 추천하는 기술적 사상에 관한 것이다.

배경 기술

[0002] 현존하는 대표 추천 서비스(넷플릭스와 유튜브, 아마존 등)에서 가장 많이 활용되는 협업 기반 추천 알고리즘은 한가지 종류의 선호도를 기반으로 동작하기 때문에 사용자들의 아이템 항목의 세부 속성들에 대한 선호도를 반영할 수 없으며 이에 따라 추천의 정확도에 한계를 보이고 있다.

[0003] 이를 해결하고자 아이템 항목의 다중 속성 사이의 내재된 관계를 유지하기 위해 다차원 모델인 텐서들을 활용한 연구들이 진행되었으며, 이를 통해 높은 추천 정확도를 구현할 수 있었으나, 이러한 연구들은 다중 특성에 대한 선호도를 기반으로 모델링을 하기 때문에 모델 학습에 이용되는 관측 선호도의 수가 적어지는 모델 희박성 문제가 일반 협업 필터링 기법들보다 더 심각하게 발생하고 추천 정확도가 떨어지는 결과로 이어질 수 있다.

[0004] 또한, 텐서들을 활용한 기존 연구에서는 높은 추천 정확도에도 불구하고 전체 사용자에게 유사한 아이템 항목들만 추천하는 아이템 다양성이 부족해지는 문제를 야기할 수 있으며 이는 추천과 같은 개인화 서비스에 실패로 이어질 수 있다. 또한 기존 연구의 결과에 따르면 추천의 정확도와 아이템의 다양성은 상반 관계를 갖기 때문에, 두 성능 지표를 동시에 만족하는 것은 일반적으로 어려우며, 다차원 텐서 모델에 대한 연산은 일반적으로 많은 컴퓨팅 자원을 요구하기 때문에 빅데이터 환경에서의 실시간 추천을 달성하기에 제한이 있다.

선행기술문헌

특허문헌

[0005] (특허문헌 0001) 한국등록특허 제10-1977536호, "클라우드 서비스 추천/결합 장치 및 클라우드 서비스 추천/결합 방법"

(특허문헌 0002) 한국등록특허 제10-1840163호, "감정 정보를 이용하여 사용자 그룹에 대해 아이템을 추천하는 장치 및 방법"

(특허문헌 0003) 한국등록특허 제10-1821790호, "사용자 선호도를 고려한 군집화 기반 추천 장치 및 방법"

발명의 내용

해결하려는 과제

[0006] 본 발명은 빅데이터 환경에서 다중 기준을 이용하여 실시간으로 보다 정확하게 아이템을 추천할 수 있는 아이템 추천 장치 및 그 방법을 제공하고자 한다.

[0007] 또한, 본 발명은 다중 특성에 대한 선호도를 다차원 텐서를 기반으로 모델링하여 대응되는 특성들 사이의 관계를 반영한 보다 정확한 예측이 가능하며, 이를 통해 산출되는 결과물들을 활용하여 설명 가능한 추천 서비스를 제공함으로써, 사용자들의 신뢰성을 높일 수 있는 아이템 추천 장치 및 그 방법을 제공하고자 한다.

[0008] 또한, 본 발명은 스마트 관광 분야에 적용되어, 관광객들에게 그들의 상세 선호도에 맞는 관광 아이템 항목들을 추천할 수 있는 아이템 추천 장치 및 그 방법을 제공하고자 한다.

과제의 해결 수단

[0009] 본 발명의 일 실시예에 따른 아이템 추천 장치는 복수의 아이템에 대한 복수의 사용자의 리뷰 데이터를 수집하는 데이터 수집부와, 수집된 리뷰 데이터에 대한 선호도 분석과 감성 분석을 수행하여 선호도 텐서 모델을 생성하

는 선호도 모델화부와, 생성된 선호도 텐서 모델에 기초한 유사도 측정을 통해 생성된 선호도 텐서 모델을 복수의 하위 모델로 분할하는 모델 분할부와, 분할된 하위 모델에서 복수의 사용자 중 타겟 사용자의 리뷰 데이터가 포함된 적어도 하나의 하위 모델을 선정하고, 선정된 하위 모델 각각에 대한 타겟 사용자의 선호도를 예측하는 선호도 예측부 및 선호도의 예측 결과에 기초하여 복수의 아이템 중 적어도 하나의 아이템을 추천하는 아이템 추천부를 포함할 수 있다.

- [0010] 일측에 따르면, 리뷰 데이터는 복수의 아이템 각각에 대한 복수의 선호도 종류 정보 및 복수의 선호도 종류 정보 각각에 대응되는 평점 정보를 포함할 수 있다.
- [0011] 일측에 따르면, 선호도 모델화부는 수집된 리뷰 데이터에 대한 선호도 분석을 통해 다차원 텐서 모델을 생성하고, 수집된 리뷰 데이터에 대한 감성 분석의 수행 결과를 이용하여 생성된 다차원 텐서 모델을 보완하여 선호도 텐서 모델을 생성할 수 있다.
- [0012] 일측에 따르면, 선호도 모델화부는 플레어엘모(FlairElmo), 텍스트블랍(Textblob), 베이더(Vader), 로지스틱 회귀(Logistic Regression), 패스트텍스트(Fasttext) 및 서포트 벡터 머신(support vector machine) 중 적어도 하나에 기초한 감성 분석의 수행 결과에 기초하여 생성된 다차원 텐서 모델을 보완할 수 있다.
- [0013] 일측에 따르면, 모델 분할부는 생성된 선호도 텐서 모델에 구비되는 선호도 종류 정보와 평점 정보에 대한 유사도 거리 연산을 수행하고, 유사도 거리 연산의 수행 결과에 기초하여 생성된 선호도 텐서 모델을 복수의 제1 하위 모델로 분할할 수 있다.
- [0014] 일측에 따르면, 모델 분할부는 분할된 제1 하위 모델 중 기설정된 크기 비율을 초과하는 모델을 제2 하위 모델로 분할할 수 있다.
- [0015] 일측에 따르면, 모델 분할부는 유클리드 거리 및 코사인 유사도 중 적어도 하나에 기초한 유사도 거리 연산을 수행할 수 있다.
- [0016] 일측에 따르면, 선호도 예측부는 선정된 하위 모델 각각에 대하여 텐서 분해를 병렬로 수행하여, 선정된 하위 모델 각각에 대한 선호도를 동시에 예측할 수 있다.
- [0017] 일측에 따르면, 아이템 추천부는 텐서 분해를 통해 도출되는 선정된 하위 모델 각각의 선호도 값에 대하여 합산 연산 및 평균값 연산 중 적어도 하나의 연산을 수행하고, 연산의 수행 결과에 기초하여 기설정된 개수의 아이템을 추천할 수 있다.
- [0018] 본 발명의 일실시예에 따른 아이템 추천 방법은 데이터 수집부에서 복수의 아이템에 대한 복수의 사용자의 리뷰 데이터를 수집하는 단계와, 선호도 모델화부에서 수집된 리뷰 데이터에 대한 선호도 분석과 감성 분석을 수행하여 선호도 텐서 모델을 생성하는 단계와, 모델 분할부에서 생성된 선호도 텐서 모델에 구비된 리뷰 데이터에 기초한 유사도 측정을 통해 생성된 선호도 텐서 모델을 복수의 하위 모델로 분할하는 단계와, 선호도 예측부에서 분할된 하위 모델에서 복수의 사용자 중 타겟 사용자의 리뷰 데이터가 포함된 적어도 하나의 하위 모델을 선정하고, 선정된 하위 모델에 대한 타겟 사용자의 선호도를 예측하는 단계 및 아이템 추천부에서 선호도의 예측 결과에 기초하여 복수의 아이템 중 적어도 하나의 아이템을 추천하는 단계를 포함할 수 있다.

발명의 효과

- [0019] 일실시예에 따르면, 본 발명은 빅데이터 환경에서 다중 기준을 이용하여 실시간으로 보다 정확하게 아이템을 추천할 수 있다.
- [0020] 일실시예에 따르면, 본 발명은 다중 특성에 대한 선호도를 다차원 텐서를 기반으로 모델링하여 대응되는 특성들 사이의 관계를 반영한 보다 정확한 예측이 가능하며, 이를 통해 산출되는 결과물들을 활용하여 설명 가능한 추천 서비스를 제공함으로써, 사용자들의 신뢰성을 높일 수 있다.
- [0021] 일실시예에 따르면, 본 발명은 스마트 관광 분야에 적용되어, 관광객들에게 그들의 상세 선호도에 맞는 관광 아이템 항목들을 추천할 수 있다.

도면의 간단한 설명

- [0022] 도 1은 일실시예에 따른 아이템 추천 장치를 설명하기 위한 도면이다.
- 도 2는 일실시예에 따른 아이템 추천 장치에서 아이템을 추천하는 과정을 보다 구체적으로 설명하기 위한 도면

이다.

도 3은 일실시예에 따른 아이템 추천 장치에서 선호도 텐서 모델을 하위 모델로 분할하는 예시를 설명하기 위한 도면이다.

도 4는 일실시예에 따른 아이템 추천 장치의 분할 방법에 따른 성능 비교 결과를 설명하기 위한 도면이다.

도 5는 일실시예에 따른 아이템 추천 방법을 설명하기 위한 도면이다.

도 6은 일실시예에 따른 아이템 추천 방법을 보다 구체적으로 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0023] 본 명세서에 개시되어 있는 본 발명의 개념에 따른 실시예들에 대해서 특정한 구조적 또는 기능적 설명들은 단지 본 발명의 개념에 따른 실시예들을 설명하기 위한 목적으로 예시된 것으로서, 본 발명의 개념에 따른 실시예들은 다양한 형태로 실시될 수 있으며 본 명세서에 설명된 실시예들에 한정되지 않는다.
- [0024] 본 발명의 개념에 따른 실시예들은 다양한 변경들을 가할 수 있고 여러 가지 형태들을 가질 수 있으므로 실시예들을 도면에 예시하고 본 명세서에 상세하게 설명하고자 한다. 그러나, 이는 본 발명의 개념에 따른 실시예들을 특정한 개시형태들에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 변경, 균등물, 또는 대체물을 포함한다.
- [0025] 제1 또는 제2 등의 용어를 다양한 구성요소들을 설명하는데 사용될 수 있지만, 구성요소들은 용어들에 의해 한정되어서는 안 된다. 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만, 예를 들면 본 발명의 개념에 따른 권리 범위로부터 이탈되지 않은 채, 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소는 제1 구성요소로도 명명될 수 있다.
- [0026] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소가 다른 구성요소에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는, 중간에 다른 구성요소가 존재하지 않는 것으로 이해되어야 할 것이다. 구성요소들 간의 관계를 설명하는 표현들, 예를 들면 "~사이에"와 "바로~사이에" 또는 "~에 직접 이웃하는" 등도 마찬가지로 해석되어야 한다.
- [0027] 본 명세서에서 사용한 용어는 단지 특정한 실시예들을 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 실시된 특징, 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것이 존재함으로 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.
- [0028] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가진다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 의미를 갖는 것으로 해석되어야 하며, 본 명세서에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.
- [0030] 이하, 실시예들을 첨부된 도면을 참조하여 상세하게 설명한다. 그러나, 특허출원의 범위가 이러한 실시예들에 의해 제한되거나 한정되는 것은 아니다. 각 도면에 제시된 동일한 참조 부호는 동일한 부재를 나타낸다.
- [0032] 도 1은 일실시예에 따른 아이템 추천 장치를 설명하기 위한 도면이다.
- [0033] 도 1을 참조하면, 일실시예에 따른 아이템 추천 장치(100)는 빅데이터 환경에서 다중 기준을 이용하여 실시간으로 보다 정확하게 아이템을 추천할 수 있다.
- [0034] 또한, 아이템 추천 장치(100)는 다중 특성에 대한 선호도를 다차원 텐서를 기반으로 모델링하여 대응되는 특성들 사이의 관계를 반영한 보다 정확한 예측이 가능하며, 이를 통해 산출되는 결과물들을 활용하여 설명 가능한

추천 서비스를 제공함으로써, 사용자들의 신뢰성을 높일 수 있다.

- [0035] 또한, 아이템 추천 장치(100)는 스마트 관광 분야에 적용되어, 관광객들에게 그들의 상세 선호도에 맞는 관광 아이템 항목들을 추천할 수 있다.
- [0036] 구체적으로, 아이템 추천 장치(100)는 사용자의 리뷰 텍스트에 대한 감성 분석을 통해 생성된 다차원 텐서 모델의 희박성 문제를 일반적인 협업 필터링 기법의 2차원 행렬 모델의 희박성과 같은 수준으로 완화시키고, 사용자의 아이템 선호도 평점과 평점 행위를 기반으로 한 자동화된 2단계 군집 기법을 제안하여 전체 선호도 모델을 유사한 사용자 선호도들로 구성되는 하위 모델로 분하여 추천의 정확도를 향상시킬 수 있다.
- [0037] 또한, 아이템 추천 장치(100)는 아이템 다양성 보장하기 위해 시스템 관리자에 의해 조정 가능한 모델 세분화를 통해 분할된 하위 모델을 더 작은 모델로 세분화할 수 있으며, 추가적으로 빅데이터 환경에서 아이템을 실시간으로 추천하기 위해 세분화된 모델의 선호도 예측을 위해 텐서 인수분해를 병렬로 진행할 수 있다.
- [0038] 이를 위해 아이템 추천 장치(100)는 데이터 수집부(110), 선호도 모델화부(120), 모델 분할부(130), 선호도 예측부(140) 및 아이템 추천부(150)를 포함할 수 있다.
- [0039] 일실시예에 따른 데이터 수집부(110)는 복수의 아이템에 대한 복수의 사용자의 리뷰 데이터를 수집할 수 있다.
- [0040] 일측에 따르면, 리뷰 데이터는 복수의 아이템 각각에 대한 복수의 선호도 종류 정보 및 복수의 선호도 종류 정보 각각에 대응되는 평점 정보를 포함할 수 있다. 또한, 리뷰 데이터는 복수의 아이템 각각에 대한 리뷰 텍스트를 더 포함할 수 있다.
- [0041] 예를 들면, 복수의 아이템은 관광 아이템으로, 식당, 숙소, 교통수단 및 관광명소 중 적어도 하나의 아이템일 수 있으나, 이에 한정되지 않고 디지털 콘텐츠와 같이 사용자의 리뷰가 존재하는 모든 아이템 항목을 포함할 수 있다.
- [0042] 또한, 복수의 선호도 종류 정보는 아이템 각각에 대응되는 평가 항목을 포함할 수 있다. 보다 구체적인 예를 들면, 아이템이 식당인 경우에 선호도 종류 정보는 음식(food), 가격(price), 서비스(service) 및 분위기(mood)와 같은 평가 항목을 포함할 수 있다.
- [0043] 또한, 평점 정보는 사용자가 0 내지 5의 점수로 선호도 종류 정보 각각에 대응되는 항목을 평가한 정보를 의미할 수 있다.
- [0044] 일실시예에 따른 선호도 모델화부(120)는 수집된 리뷰 데이터에 대한 선호도 분석과 감성 분석을 수행하여 선호도 텐서 모델을 생성할 수 있다.
- [0045] 일측에 따르면, 선호도 모델화부(120)는 수집된 리뷰 데이터에 대한 선호도 분석을 통해 다차원 텐서 모델을 생성하고, 수집된 리뷰 데이터에 대한 감성 분석의 수행 결과를 이용하여 생성된 다차원 텐서 모델을 보완하여 선호도 텐서 모델을 생성할 수 있다.
- [0046] 구체적으로, 선호도 모델화부(120)는 데이터 수집부(110)를 통해 수집된 리뷰 데이터를 리뷰 데이터베이스에 저장할 수 있고, 리뷰 데이터베이스에 저장된 리뷰 데이터 중 리뷰 텍스트에 대한 전처리를 수행할 수 있으며, 이후 전처리된 리뷰 텍스트와, 리뷰 데이터 중 복수의 선호도 종류 정보 및 이에 대응되는 평점 정보를 입력으로 수신하는 감성분석 모델을 통해 학습하여 복수의 선호도 종류 정보에 대응되는 감성 정보(즉, 암시적 평점)를 예측할 수 있다.
- [0047] 이와 동시에, 선호도 모델화부(120)는 리뷰 데이터베이스에 저장된 리뷰 데이터를 이용하여 사용자의 선호도(즉, 명시적 평점)를 다차원 텐서를 활용하여 모델화하고, 최종적으로 모델의 희박성을 보완하기 위해 예측된 감성 정보를 이용하여 다차원 텐서 모델을 보완할 수 있다. 예를 들면, 다차원 텐서 모델은 사용자 정보를 x축, 선호도 종류 정보를 y축, 아이템 항목을 z축으로 하여 형성되는 모델일 수 있다.
- [0048] 일측에 따르면, 선호도 모델화부(120)는 플레어엘모(FlairElmo), 텍스트블랍(Textblob), 베이더(Vader), 로지스틱 회귀(Logistic Regression), 패스트텍스트(Fasttext) 및 서포트 벡터 머신(support vector machine) 중 적어도 하나에 기초한 감성 분석의 수행 결과에 기초하여 생성된 다차원 텐서 모델을 보완할 수 있다. 바람직하게는, 선호도 모델화부(120)는 플레어엘모에 기초한 감성 분석을 통해 다차원 텐서 모델을 보완할 수 있다.
- [0049] 일실시예에 따른 모델 분할부(130)는 생성된 선호도 텐서 모델에 기초한 유사도 측정을 통해 생성된 선호도 텐서 모델을 복수의 하위 모델로 분할할 수 있다.

- [0050] 일측에 따르면, 모델 분할부(130)는 생성된 선호도 텐서 모델에 구비되는 선호도 종류 정보와 평점 정보에 대한 유사도 거리 연산을 수행하고, 유사도 거리 연산의 수행 결과에 기초하여 생성된 선호도 텐서 모델을 복수의 제1 하위 모델로 분할할 수 있다.
- [0051] 또한, 모델 분할부(130)는 분할된 제1 하위 모델 중 기설정된 크기 비율을 초과하는 모델을 제2 하위 모델로 분할할 수 있다.
- [0052] 예를 들면, 모델 분할부(130)는 유클리드 거리 및 코사인 유사도 중 적어도 하나에 기초한 유사도 거리 연산을 수행할 수 있다.
- [0053] 구체적으로, 모델 분할부(130)는 선호도 텐서 모델에 포함된 리뷰 데이터에 대해서, 사용자가 평가한 선호도 종류 정보(즉, 평가 항목) 간의 선호도 평점의 차이를 반영하는 평점 행위와, 평점 행위에 따른 다차원 선호도(즉, 평점 정보)에 대한 두가지 거리 계산 연산 방법에 기초하여 2단계로 군집화를 수행하여 유사한 리뷰를 제1 하위 모델로 그룹화할 수 있다.
- [0054] 또한, 모델 분할부(130)는 추가적으로 추천 아이템의 다양성을 보장하기 위해 시스템 관리자에 의해 조정 가능한 상수 요소(전체 모델 대비 모델 그룹의 크기 비율) ϕ 를 기반으로 더 세분화된 하위 모델인 제2 하위 모델로 적어도 하나의 제1 하위 모델을 분할할 수 있다.
- [0055] 일측에 따르면, 2단계 군집화 과정 중 첫번째 군집화 과정에서는 평점 점수를 고려하여 군집화를 수행하고, 두번째 군집화 과정에서는 평점 행위를 고려하여 군집화를 수행할 수 있다.
- [0056] 일실시예에 따른 선호도 예측부(140)는 분할된 하위 모델에서 복수의 사용자 중 타겟 사용자의 리뷰 데이터가 포함된 적어도 하나의 하위 모델을 선정하고, 선정된 하위 모델 각각에 대한 타겟 사용자의 선호도를 예측할 수 있다.
- [0057] 일측에 따르면, 선호도 예측부(140)는 선정된 하위 모델 각각에 대하여 텐서 분해를 병렬로 수행하여, 선정된 하위 모델 각각에 대한 선호도를 동시에 예측할 수 있다.
- [0058] 구체적으로, 선호도 예측부(140)는 적어도 하나의 타겟 사용자로부터 아이템의 추천 요청이 있을 때마다 동작하는 온라인 처리부에 속하며, 우선 타겟 사용자의 리뷰가 포함된 적어도 하나의 하위 모델을 선정하고, 실시간 응답을 위해 하위 모델 각각에 대한 텐서 인수분해를 병렬로 처리하여, 최종적으로 다수의 예측(근사) 하위 모델을 타겟 사용자의 선호도 예측 결과로 출력할 수 있다.
- [0059] 이때, 하위 모델 각각에 대한 텐서 인수분해는 예측한 선호도와 기존에 관측된(사용자가 입력한) 선호도 들 간의 차이(에러)를 최소화하는 매개변수를 찾기 위해 반복하여 최적의 변수들로 모델을 최적화하는 과정을 의미한다.
- [0060] 예를 들면, 선호도 예측부(140)는 확률적 경사하강법(Stochastic Gradient Descent, SGD), 모멘텀(Momentum) 및 아담(Adam) 중 적어도 하나에 기초한 최적화 과정을 통해 텐서 분해를 수행할 수 있으며, 바람직하게는, 선호도 예측부(140)는 확률적 경사하강법을 이용하여 텐서 분해를 수행할 수 있다.
- [0061] 일실시예에 따른 아이템 추천부(150)는 선호도의 예측 결과에 기초하여 복수의 아이템 중 적어도 하나의 아이템을 추천할 수 있다.
- [0062] 일측에 따르면, 아이템 추천부(150)는 텐서 분해를 통해 선정된 하위 모델 각각의 선호도 값에 대하여 합산 연산 및 평균값 연산 중 적어도 하나의 연산을 수행하고, 연산의 수행 결과에 기초하여 기설정된 개수의 아이템을 추천할 수 있다.
- [0063] 구체적으로, 아이템 추천부(150)는 사용자들의 리뷰가 다수의 근사 하위 모델(즉, 사용자의 선호도 예측 결과)에 포함되어 있기 때문에 이에 대한 통합 과정을 수행하며, 이때 아이템 추천에 따른 응답 시간을 최소화하기 위해 간단한 합산 연산이나 평균 연산을 수행하는 집계함수를 활용하여 통합 과정을 수행할 수 있다.
- [0064] 또한, 아이템 추천부(150)는 최종적으로 집계된 아이템 항목에 대한 타겟 사용자의 선호도를 정렬하여 상위 N개(여기서, N은 기설정된 개수)의 아이템을 타겟 사용자에게 추천할 수 있다.
- [0066] 도 2는 일실시예에 따른 아이템 추천 장치에서 아이템을 추천하는 과정을 보다 구체적으로 설명하기 위한 도면이다.

- [0067] 도 2를 참조하면, 이하에서 설명하는 아이템 추천 장치(200)는 도 1의 아이템 추천 장치일 수 있다.
- [0068] 일실시예에 따른 선호도 모델화부(210)는 수집된 리뷰 데이터에 대한 선호도 분석과 감성 분석을 수행하여 선호도 텐서 모델을 생성할 수 있다.
- [0069] 구체적으로, 211 단계에서 선호도 모델화부(210)는 리뷰 데이터베이스(리뷰 DB)에 저장된 리뷰 데이터를 이용하여 사용자의 선호도(즉, 명시적 평점)를 다차원 텐서를 활용하여 모델화할 수 있다.
- [0070] 보다 구체적으로, 다차원 텐서 모델($\mathcal{T}$)은 3차원 텐서 모델(사용자 정보 x 선호도 종류 정보 x 아이템 항목)로서, J번째 아이터에 대한 K개의 다중 평가항목(즉, 선호도 종류) 각각에 대하여 I번째 사용자가 평가한 리뷰 데이터가 주어지면 $\mathcal{T} = \{t_{ijk}\} \in \mathbb{R}^{I \times J \times K}$ 로 정의될 수 있으며, 여기서 t_{ijk} 는 K번째 평가항목에 대해 I번째 사용자가 평가한 평점으로, 다시 말해 t_{ijk} 는 다차원 텐서 모델($\mathcal{T}$)을 구성하는 각각의 큐브(텐서 값)을 의미할 수 있다.
- [0071] 다음으로, 212 단계에서 선호도 모델화부(210)는 리뷰 데이터베이스에 저장된 리뷰 데이터 중 리뷰 텍스트에 대하여 감성 분석을 위한 전처리를 수행할 수 있다.
- [0072] 다음으로, 213 단계에서 선호도 모델화부(210)는 전처리된 리뷰 텍스트와, 리뷰 데이터 중 복수의 선호도 종류 정보 및 이에 대응되는 평점 정보를 입력으로 수신하는 감성분석 모델을 통해 학습할 수 있다.
- [0073] 일측에 따르면, 감성분석 모델은 플레어엘모(FlairElmo), 텍스트블랍(Textblob), 베이더(Vader), 로지스틱 회귀(Logistic Regression), 패스트텍스트(Fasttext) 및 서포트 벡터 머신(support vector machine) 중 적어도 하나에 기초하는 모델일 수 있다. 바람직하게는, 감성분석 모델은 플레어엘모에 기초하는 모델일 수 있다.
- [0074] 보다 구체적으로, 복수의 사용자의 음식점 리뷰 데이터 세트에 대하여 텍스트블랍(TD), 베이더(VD), 서포트 벡터 머신(SVM), 로지스틱 회귀(LR), 패스트텍스트(FT) 및 플레어엘모(FE)에 기초한 감성 분석의 수행 결과를 나타내는 하기 표1을 참조하면, 플레어엘모(FE)에 기초한 감성분석 모델이 가장 높은 성능을 보이는 것을 확인할 수 있다.

표 1

Type	Metric	TB	VD	SVM	LR	FT	FE
Overall	MAE	0.603	0.599	0.659	0.537	0.391	0.367
	Kapp	0.254	0.250	0.195	0.337	0.462	0.491
Food	MAE	0.615	0.625	0.667	0.603	0.448	0.422
	Kapp	0.218	0.209	0.138	0.261	0.394	0.415
Price	MAE	0.772	0.771	0.732	0.730	0.626	0.568
	Kapp	0.167	0.167	0.224	0.237	0.289	0.328
Service	MAE	0.662	0.648	0.668	0.631	0.493	0.468
	Kapp	0.223	0.222	0.137	0.290	0.364	0.375
Mood	MAE	0.703	0.695	0.716	0.924	0.617	0.580
	Kapp	0.174	0.174	0.163	0.155	0.256	0.277

- [0075]
- [0076] 표1에 따르면, 종합 평가(Overall), 음식(Food), 가격(Price), 서비스(Service) 및 분위기(Mood)의 평가 항목(즉, 선호도 종류 정보)에 대한 감성 분석을 수행하고, 수행 결과에 대한 평균 절대 오차(MAE) 및 Cohen의 Kappa 점수를 이용한 감성분석 성능의 비교결과, 플레어엘모(FE)에 기초한 감성분석 모델이 모든 평가 항목에서 감성분석 성능이 가장 우수한 것으로 나타났다.
- [0077] 즉, 일실시예에 따른 선호도 모델화부(210)는 플레어엘모(FE)에 기초한 감성 분석을 통해 감성 분석의 정확도를 향상시킬 수 있다.
- [0078] 다음으로, 214 단계에서 선호도 모델화부(210)는 감성분석 모델을 이용한 학습의 결과로 선호도 종류 정보에 대응되는 감성 정보(즉, 암시적 평점)를 예측할 수 있다.
- [0079] 다음으로, 215 단계에서 선호도 모델화부(210)는 모델의 회소성을 보완하기 위해 예측된 감성 정보를 이용하여 다차원 텐서 모델을 보완하여 선호도 텐서 모델을 생성할 수 있다.

[0080] 보다 구체적으로, 선호도 모델화부(210)에서 선호도 분석을 통해 도출된 다차원 텐서 모델은 일부 텐서 값이 누락된 영역이 존재할 수 있으며, 이에 텐서 값이 누락된 영역에 암시적 평점을 적용하여 다차원 텐서 모델을 보완할 수 있다.

[0081] 215 단계에서 선호도 모델화부(210)는 다차원 텐서 모델($\mathcal{T}$)과, 텐서($\mathcal{S} = \{s_{ijk}\} \in \mathbb{R}^{I \times J \times K}$)로 구성된 암

시적 평점 세트가 주어지면 모델 보완 함수 $\mathcal{MC}(\mathcal{T}, \mathcal{S}) = \begin{cases} s_{ijk}, & \text{if } (t_{ijk} = \emptyset) \text{ and } (s_{ijk} \neq \emptyset) \\ t_{ijk}, & \text{otherwise.} \end{cases}$ 를 도출할 수 있다.

[0082] 일실시예에 따른 모델 분할부(220)는 선호도 텐서 모델에 기초한 유사도 측정을 통해 생성된 선호도 텐서 모델을 복수의 하위 모델로 분할할 수 있다.

[0083] 구체적으로, 221 단계에서 모델 분할부(220)는 선호도 텐서 모델에 포함된 리뷰 데이터에 대해서 평점 행위와 이에 대응되는 다차원 선호도(즉, 평점 정보)에 대한 두가지 거리 계산 연산 방법(유클리드 거리 및 코사인 유사도에 기반한 연산)에 기초하여 2단계 군집화 과정을 수행하여 유사한 리뷰를 제1 하위 모델로 그룹화할 수 있다.

[0084] 보다 구체적으로, 모델 분할부(220)는 유클리드 거리를 연산하기 위해 평점 행위와, 평점 행위에 대응되는 다차원 선호도(즉, 평점 정보)를 고려할 수 있다.

[0085] 두 개의 리뷰 벡터 X와 Y가 주어지면 n 개의 다중 기준(즉, 복수의 선호도 종류 정보)에 대한 n 개의 평가(즉, 평점)으로 구성된 이들 사이의 유사도 거리 $\mathcal{MD}(X, Y)$ 는 하기 수학적식1과 같이 정의될 수 있다.

[0086] [수학적식1]

$$[0087] \mathcal{MD}(X, Y) = \|x - y\|^2 = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}.$$

[0088] 여기서, 수학적식1은 두 리뷰(X 및 Y) 사이의 거리가 다중 기준 평가 측면에서 서로 얼마나 떨어져 있는지를 나타낼 수 있다. 즉, $\mathcal{MD}$ 의 연산 결과는 n 차원 유클리드 공간에서 두 리뷰 사이의 선분의 길이를 나타낼 수 있다.

[0089] 한편, 사용자의 평점 행위 및 평점 행위에 따른 평점 정보는 사용자의 선호도에 영향을 받을 수 있다. 일례로, 어떤 사용자는 호텔에 대하여 청결도 항목이 중요하다고 생각할 수 있고, 다른 사용자는 직원의 친절도 항목 및 서비스 항목이 중요하다고 생각할 수 있다.

[0090] 즉, 동일한 호텔에 대한 전체 평점은 평점 행위에 따라 달라질 수 있으며, 이에 모델 분할부(220)는 이러한 현상을 반영하기 위해 두 리뷰 벡터간의 코사인 유사도를 고려할 수 있다.

[0091] n 개의 다중 기준에 대해 n 개의 양수 값으로 구성된 두 개의 리뷰 벡터 x 및 y가 주어지면, 코사인 유사도 $\cos(x, y) = \sum_i x_i y_i / \sqrt{\sum_i x_i^2} \sqrt{\sum_i y_i^2}$ 로 정의될 수 있으며, 이는 $1 - \cos(x, y)$ 의 연산을 통해 코사인 거리($\mathcal{CD}$)로 변환될 수 있다.

[0092] 결과적으로, 유클리드 거리를 기반으로 한 군집은 모든 리뷰를 하나의 그룹으로 분류하고 코사인 거리를 기반으로 한 군집은 두 개의 그룹으로 분류할 수 있다.

[0093] 이에, 모델 분할부(220)는 두 평가 동작 간의 거리를 유클리드 거리의 형식으로 고려하기 위해 l^2 정규화를 이용할 수 있으며, 두 개의 l^2 정규화된 리뷰 벡터 x 및 y (여기서, $\sum x_i^2 = \sum y_i^2 = 1$ 및 $\cos(x, y) = \sum x_i y_i$)간의 유클리드 및 코사인 거리는 하기 수학적식2와 같이 선형적으로 연결(즉, 비례 관계)될 수 있다.

[0094] [수학적식2]

$$[0095] \|x - y\|^2 = \sum (x_i - y_i)^2 = \sum (x_i^2 + y_i^2 - 2x_i y_i) = \sum x_i^2 + \sum y_i^2 - 2 \sum x_i y_i \\ = 1 + 1 - 2\cos(x, y) = 2(1 - \cos(x, y)) = 2\mathcal{CD}.$$

- [0096] 여기서, l^2 정규화된 리뷰 벡터는 평점 행위일 수 있다.
- [0097] 다음으로, 222 단계에서 모델 분할부(220)는 추천 아이템의 다양성을 보장하기 위해 시스템 관리자에 의해 조정 가능한 상수 요소(전체 모델 대비 모델 그룹의 크기 비율) ϕ 를 기반으로 더 세분화된 하위 모델인 제2 하위 모델로 적어도 하나의 제1 하위 모델을 분할할 수 있다.
- [0098] 보다 구체적으로, 상수 요소 ϕ 는 백분율 단위의 상수로, 하위 모델에 속하는 리뷰 데이터의 수, 즉 하위 모델의 크기를 제어할 수 있다. 예를 들면, $\phi = 10$ 은 전체 리뷰 데이터의 10 %를 하위 모델의 최대 크기로 구성함을 의미할 수 있다.
- [0099] 한편, 모델 분할부(220)는 K-평균 군집 알고리즘(K-means clustering algorithm)을 통해 서로 유사한 사용자의 리뷰 데이터를 군집화할 수도 있다.
- [0100] 일실시예에 따른 선호도 예측부(230)는 분할된 하위 모델에서 복수의 사용자 중 타겟 사용자의 리뷰 데이터가 포함된 적어도 하나의 하위 모델을 선정하고, 선정된 하위 모델 각각에 대한 타겟 사용자의 선호도를 예측할 수 있다.
- [0101] 구체적으로, 231 단계에서 선호도 예측부(230)는 타겟 사용자의 리뷰가 포함된 적어도 하나의 하위 모델을 선정할 수 있다.
- [0102] 다음으로, 232 단계에서 선호도 예측부(230)는 선정된 하위 모델 각각에 대한 텐서 인수분해를 병렬로 처리하여, 최종적으로 다수의 근사 하위 모델을 타겟 사용자의 선호도 예측 결과로 출력할 수 있다.
- [0103] 보다 구체적으로, 선호도 예측부(230)는 타겟 사용자의 리뷰가 포함된 하위 모델에서 사용자가 평가하지 않은 평가항목에 대한 누락된 평점을 예측하기 위해 텐서 분해를 통한 근사화를 수행하되, 고차 특이 값 분해(Higher-order singular value decomposition, HOSVD)를 이용하여 타겟 사용자의 리뷰가 포함된 하위 모델을 병렬 방식으로 근사화할 수 있다.
- [0104] 즉, 선호도 예측부(230)는 텐서 분해를 통해 누락된 평점이 예측된 다수의 근사 하위 모델을 획득할 수 있다.
- [0105] 일실시예에 따른 아이템 추천부(240)는 텐서 분해를 통해 선정된 하위 모델 각각의 선호도 값에 대하여 합산 연산 및 평균값 연산 중 적어도 하나의 연산을 수행하고, 연산의 수행 결과에 기초하여 기설정된 개수의 아이템을 추천할 수 있다.
- [0106] 구체적으로, 241 단계에서 아이템 추천부(240)는 합산 연산이나 평균 연산을 수행하는 집계함수를 활용하여 다수의 근사 하위 모델(즉, 사용자의 선호도 예측 결과)에 따른 평점 정보를 통합할 수 있다.
- [0107] 다음으로, 242 단계에서 아이템 추천부(240)는 통합된 평점 정보가 높은 순으로 아이템 항목을 정렬할 수 있다.
- [0108] 다음으로, 243 단계에서 아이템 추천부(240)는 정렬된 아이템 항목 중 상위 N개(여기서, N은 기설정된 개수)의 아이템을 타겟 사용자에게 추천할 수 있다.
- [0110] 도 3은 일실시예에 따른 아이템 추천 장치에서 선호도 텐서 모델을 하위 모델로 분할하는 예시를 설명하기 위한 도면이다.
- [0111] 도 3을 참조하면, 310 단계 내지 320 단계에서 일실시예에 따른 아이템 추천 장치는 선호도 텐서 모델에 포함된 리뷰 데이터(UR) 중 선호도 종류 정보와 평점 정보에 대한 유사도 거리 연산을 수행하고, 유사도 거리 연산의 수행 결과에 기초하는 2단계 군집화 과정을 통해 선호도 텐서 모델을 복수의 제1 하위 모델(RC 1-1 내지 RC k-3)로 분할할 수 있다.
- [0112] 구체적으로, 310 단계에서 일실시예에 따른 아이템 추천 장치는 평점 정보를 고려한 첫번째 군집화 과정을 수행하여 선호도 텐서 모델을 복수의 리뷰 그룹(리뷰그룹1 내지 리뷰그룹k)로 분할할 수 있다.
- [0113] 다음으로, 320 단계에서 일실시예에 따른 아이템 추천 장치는 복수의 선호도 종류 정보 중 사용자가 평가한 항목에 대한 정보를 포함하는 평점 행위를 고려한 두번째 군집화 과정을 수행하여 복수의 리뷰 그룹(리뷰그룹1 내지 리뷰그룹k)을 복수의 제1 하위 모델(RC 1-1 내지 RC k-3)로 분할할 수 있다.
- [0114] 다음으로, 330 단계에서 일실시예에 따른 아이템 추천 장치는 분할된 제1 하위 모델(RC 1-1 내지 RC k-3) 중 기

설정된 크기 비율을 초과하는 모델(RC k-1)을 제2 하위 모델(RC k-1-1, RCK-1-2)로 분할할 수 있다.

- [0115] 일측에 따르면, 330 단계에서 일실시예에 따른 아이템 추천 장치는 추천 아이템의 다양성을 보장하기 위해 시스템 관리자에 의해 조정 가능한 상수 요소 ϕ 를 기반으로 제1 하위 모델(RC k-1)을 보다 더 세분화된 하위 모델인 제2 하위 모델(RC k-1-1, RCK-1-2)로 분할할 수 있다.
- [0117] 도 4는 일실시예에 따른 아이템 추천 장치의 분할 방법에 따른 성능 비교 결과를 설명하기 위한 도면이다.
- [0118] 도 4를 참조하면, 도 4의 (a)는 2단계 군집화만을 이용하여 리뷰 데이터를 그룹화하는 DCTF(Double clustering tensor factorization) 분할 방법과, 상수 요소 ϕ 만을 이용하여 리뷰 데이터를 그룹화하는 DTF(Division-based tensor factorization) 분할 방법과, 2단계 군집화 및 상수 요소 ϕ 모두를 이용하여 리뷰 데이터를 그룹화하는 DnC(the decrease and conquer-based parallel tensor factorization) 방법에 따른 정확도(precision) 및 리콜(recall) 성능의 비교 결과를 도시한다.
- [0119] 또한, 도 4의 (b)는 DCTF 분할 방법, DTF 분할 방법 및 DnC 분할 방법에 따른 커버리지(coverage) 및 다양성(diversity) 성능의 비교 결과를 도시한다.
- [0120] 또한, 도 4의 (a) 내지 (b)에서 하첨자 RSB는 2단계 군집화 과정 중 첫번째 군집화 과정에서 평점 점수를 고려하고 두번째 군집화 과정에서 평점 행위를 고려함을 의미하고, 하첨자 RBS는 2단계 군집화 과정 중 첫번째 군집화 과정에서 평점 행위를 고려하고 두번째 군집화 과정에서 평점 점수를 고려함을 의미하며, 하첨자 *은 상수 계수 ϕ (여기서, $\phi \in 0.5, 1, 3, 5, 7, 10$)의 평균값을 의미한다.
- [0121] 또한, MP@10과 MR@10은 10개의 아이템을 추천 시 평균 정확도와 재현율을 의미하며, upCvrg는 커버리지 성능을 의미하고, Divs는 다양성 성능을 의미한다.
- [0122] 도 4의 (a)에 따르면, 사용자 선호 구조(즉, RSB 및 RBS)의 적용 순서 측면에서 $DCTF_{RSB}$ 는 $DCTF_{RBS}$ 보다 MP@10 및 MR@10 성능이 약간 더 우수한 것으로 나타났다. 이 결과는 두번째 군집화 과정에서 평점 행위를 고려하는 것이 더 적절하고, 사용자들의 평점 정보와 관련하여 다른 평점 행위를 나타낼 수 있음을 의미한다.
- [0123] 예를 들면, 음식점 리뷰의 경우 사용자에게 의한 나쁜 리뷰 데이터에는 주로 음식 항목에 대한 낮은 평가가 반영되고, 좋은 리뷰 데이터에는 주로 서비스 항목에 대한 높은 평가가 반영되며, 이를 통해 나쁜 리뷰와 좋은 리뷰 사이에 서로 다른 평점 행위가 반영된다는 것을 확인할 수 있다.
- [0124] 한편, DCTF는 DnC 대비 조금 더 높은 정밀도 성능을 보이고 있으나, DnC의 리콜 성능을 고려하였을 때, DnC를 이용하는 것이 바람직할 수 있다.
- [0125] 도 4의 (b)에 따르면, 이진 분할이 랜덤 분할보다 DnC 및 DTF에 더 긍정적인 영향을 미치는 것을 확인할 수 있으며, 이는 랜덤 분할보다 이진 분할을 사용하는 하위 모델의 크기가 더 균일하기 때문인 것으로 나타났다. 즉, 아이템을 하위 모델로 다양하게 분배할 가능성이 높아져 추천의 다양성을 향상시킬 수 있다.
- [0126] 한편, DnC는 DTF에 비해 커버리지 성능이 낮게 측정되었지만 다양성 성능은 높게 측정되었다. 특히, DnC_{RSBBD} 의 다양성 성능은 DTF에 비해 평균적으로 약 15% 향상된 것으로 나타났지만 커버리지 저하는 5%에 불과한 것으로 나타났으며, 이 밖에 다른 성능 평가에서도 최고의 성능을 보이는 것을 확인할 수 있었다.
- [0127] 이에, 일실시예에 따른 아이템 추천 장치는 DnC_{RSBBD} 분할 방법을 적용하여 선호도 텐서 모델을 복수의 하위 모델로 분할할 수 있다.
- [0129] 도 5는 일실시예에 따른 아이템 추천 방법을 설명하기 위한 도면이다.
- [0130] 다시 말해, 도 5는 도 1 내지 도 4를 통해 설명한 일실시예에 따른 아이템 추천 장치의 동작방법을 설명하기 위한 도면으로, 이후 도 5를 통해 설명하는 내용 중 도 1 내지 도 4를 통해 설명한 내용과 중복되는 설명은 생략하기로 한다.
- [0131] 도 5를 참조하면, 510 단계에서 일실시예에 따른 아이템 추천 방법은 데이터 수집부에서 복수의 아이템에 대한 복수의 사용자의 리뷰 데이터를 수집할 수 있다.

- [0132] 다음으로, 520 단계에서 일실시예에 따른 아이템 추천 방법은 선호도 모델화부에서 수집된 리뷰 데이터에 대한 선호도 분석과 감성 분석을 수행하여 선호도 텐서 모델을 생성할 수 있다.
- [0133] 다음으로, 530 단계에서 일실시예에 따른 아이템 추천방법은 모델 분할부에서 생성된 선호도 텐서 모델에 구비된 리뷰 데이터에 기초한 유사도 측정을 통해 생성된 선호도 텐서 모델을 복수의 하위 모델로 분할할 수 있다.
- [0134] 다음으로, 540 단계에서 일실시예에 따른 아이템 추천방법은 선호도 예측부에서 분할된 하위 모델에서 복수의 사용자 중 타겟 사용자의 리뷰 데이터가 포함된 적어도 하나의 하위 모델을 선정하고, 선정된 하위 모델에 대한 타겟 사용자의 선호도를 예측할 수 있다.
- [0135] 다음으로, 550 단계에서 일실시예에 따른 아이템 추천방법은 아이템 추천부에서 선호도의 예측 결과에 기초하여 복수의 아이템 중 적어도 하나의 아이템을 추천할 수 있다.
- [0137] 도 6은 일실시예에 따른 아이템 추천 방법을 보다 구체적으로 설명하기 위한 도면이다.
- [0138] 이하에서 설명하는 도 6의 610 단계 내지 670 단계는 도 5의 520 단계 내지 550 단계를 통해 수행될 수 있다.
- [0139] 도 6을 참조하면, 이하에서 설명하는 610 내지 630 단계는 오프라인 처리 단계에 해당하며, 시스템 관리자에 의해 일정한 간격을 두고 반복 수행될 수 있으며, 이를 통해 새로 수집된 리뷰 데이터가 반영된 선호도 텐서 모델을 생성 및 분할할 수 있다.
- [0140] 또한, 이하에서 설명하는 640 내지 670 단계는 온라인 처리 단계에 해당하며, 타겟 사용자의 아이템 추천 요청이 있을 때마다 수행될 수 있다.
- [0141] 구체적으로, 610 단계에서 일실시예에 따른 아이템 추천 방법은 선호도 모델화부에서 리뷰 데이터에 대한 선호도 분석을 통해 다차원 텐서 모델을 생성할 수 있으며, 여기서 생성된 다차원 텐서 모델은 일반적으로 높은 희박성 즉, 관측되지 않는 선호도로써 예측 대상이 되는 복수의 흰색 큐브를 포함할 있다.
- [0142] 다음으로, 620 단계에서 일실시예에 따른 아이템 추천 방법은 선호도 모델화부에서 리뷰 데이터에 대한 감성 분석의 수행 결과를 이용하여 생성된 다차원 텐서 모델을 보완하여 선호도 텐서 모델을 생성할 수 있다.
- [0143] 보다 구체적으로, 620 단계에서 일실시예에 따른 아이템 추천 방법은 감성 분석을 통해 사용자의 리뷰에 한 개 이상의 선호도 점수(즉, 평점 정보)가 관측되면, 관측된 선호도 점수를 통해 흰색 큐브가 보완되어 빗금이 쳐진 큐브가 될 수 있다.
- [0144] 다음으로, 630 단계에서 일실시예에 따른 아이템 추천 방법은 모델 분할부에서 생성된 선호도 텐서 모델에 기초한 유사도 측정을 통해 생성된 선호도 텐서 모델을 복수의 하위 모델로 분할할 수 있다.
- [0145] 보다 구체적으로, 630 단계에서 일실시예에 따른 아이템 추천 방법은 선호도 텐서 모델(즉, 보완된 다차원 텐서 모델)이 실시간 연산/생성되어 크기가 크기 때문에, 모델 분할을 통해 선호도 텐서 모델을 복수의 하위 모델로 분할할 수 있다.
- [0146] 다음으로, 640 단계에서 일실시예에 따른 아이템 추천 방법은 선호도 예측부에서 분할된 하위 모델에서 복수의 사용자 중 타겟 사용자의 리뷰 데이터가 포함된 적어도 하나의 하위 모델을 선정할 수 있다.
- [0147] 다시 말해, 640 단계에서 일실시예에 따른 아이템 추천 방법은 타겟 사용자로부터 아이템 추천 요청을 수신하면 타겟 사용자의 리뷰 데이터가 포함된 하위 모델만을 선정할 수 있으며, 이때 모델에 포함되는 아이템은 각 하위 모델마다 다르기 때문에, 기존 방법에 비해 훨씬 높은 아이템의 다양성을 달성할 수 있다.
- [0148] 다음으로, 650 단계에서 일실시예에 따른 아이템 추천 방법은 선호도 예측부에서 선정된 하위 모델 각각에 대하여 텐서 분해를 병렬로 수행하여, 선정된 하위 모델 각각에 대한 선호도를 동시에 예측할 수 있다.
- [0149] 다시 말해, 650 단계에서 일실시예에 따른 아이템 추천 방법은 전체 하위 모델보다 적은 수의 하위 모델이 병렬 인수분해를 통해 동시에 예측되기 때문에, 실시간으로 아이템 예측 및 추천을 수행할 수 있다.
- [0150] 보다 구체적으로, 650 단계에서 일실시예에 따른 아이템 추천 방법은 모든 선호도 값이 예측된 하위 텐서와 같고 이중 추천 요청 사용자의 아이템들에 대한 선호도 부분(즉, 선호도 값)(검은 큐브)만이 아이템 추천부로 전달될 수 있다.
- [0151] 다음으로, 660 단계에서 일실시예에 따른 아이템 추천 방법은 아이템 추천부에서 선정된 하위 모델 각각의 선호

도 값에 대하여 합산 연산 및 평균값 연산 중 적어도 하나의 연산을 수행하여 선호도 값을 통합할 수 있다.

[0152] 다음으로, 670 단계에서 일실시예에 따른 아이템 추천 방법은 아이템 추천부에서 빗금이 있는 아이템의 전체 특성에 대한 선호도들을 활용하여 아이템들을 정렬하고 상위 N개의 항목을 타겟 사용자에게 추천할 수 있다.

[0154] 결국, 본 발명을 이용하면, 빅데이터 환경에서 다중 기준을 이용하여 실시간으로 보다 정확하게 아이템을 추천할 수 있다.

[0155] 또한, 본 발명은 다중 특성에 대한 선호도를 다차원 텐서를 기반으로 모델링하여 대응되는 특성들 사이의 관계를 반영한 보다 정확한 예측이 가능하며, 이를 통해 산출되는 결과물들을 활용하여 설명 가능한 추천 서비스를 제공함으로써, 사용자들의 신뢰성을 높일 수 있다.

[0156] 또한, 본 발명은 스마트 관광 분야에 적용되어, 관광객들에게 그들의 상세 선호도에 맞는 관광 아이템 항목들을 추천할 수 있다.

[0158] 이상과 같이 실시예들이 비록 한정된 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들면, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 장치, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

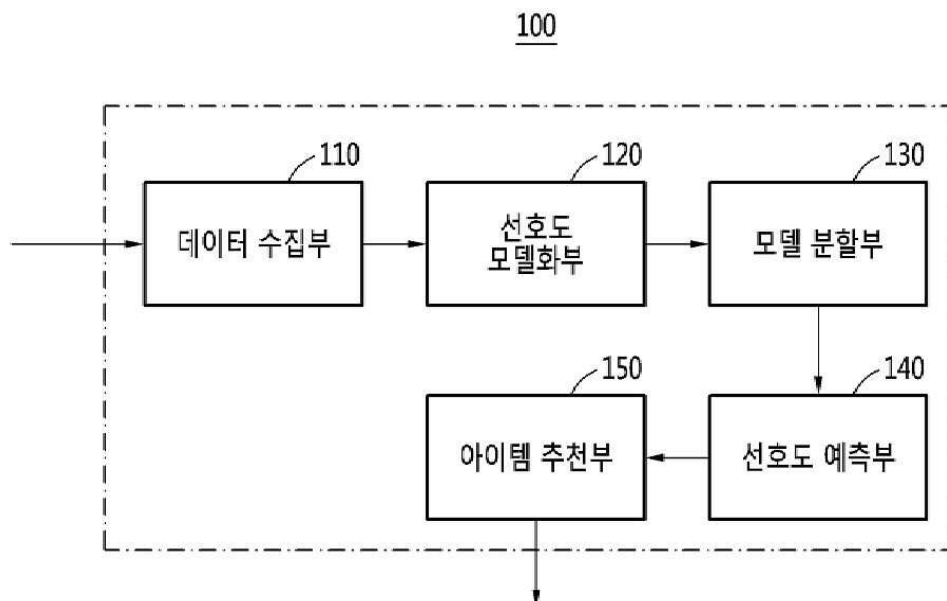
[0159] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

부호의 설명

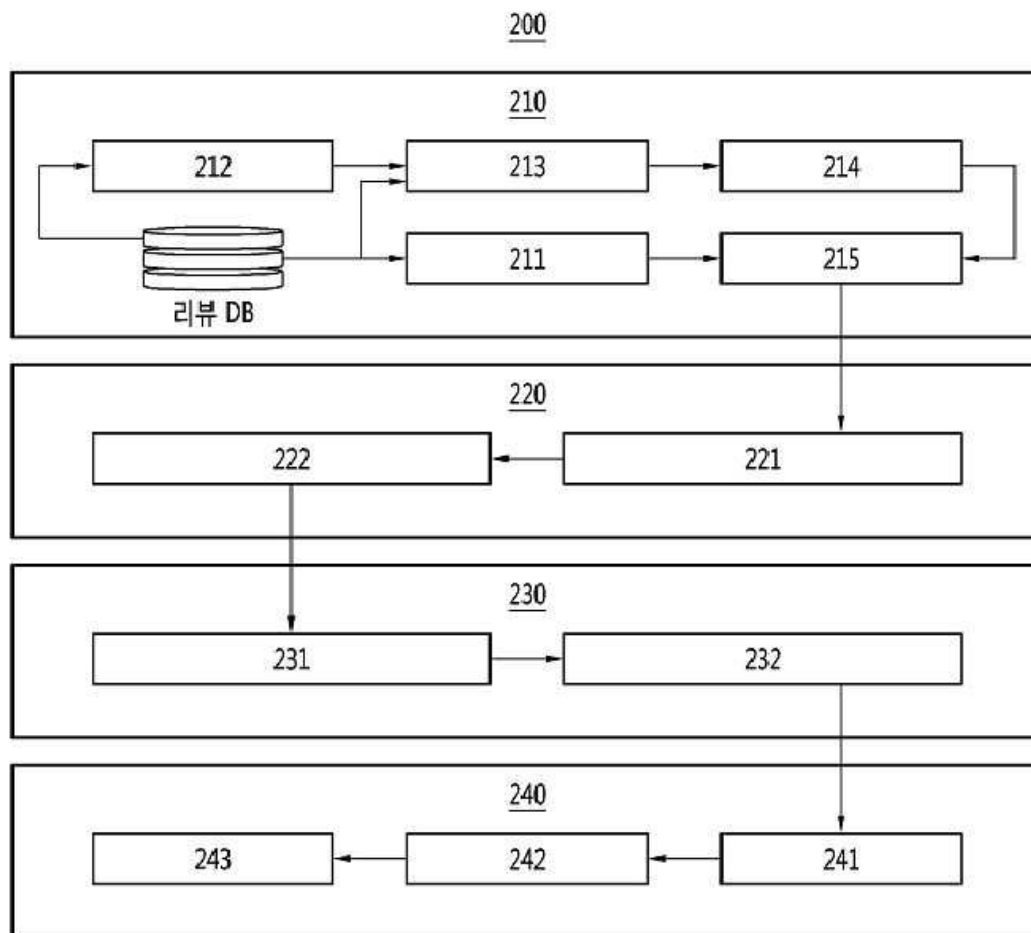
[0161] 100: 아이템 추천 장치 110: 데이터 수집부
120: 선호도 모델화부 130: 모델 분할부
140: 선호도 예측부 150: 아이템 추천부

도면

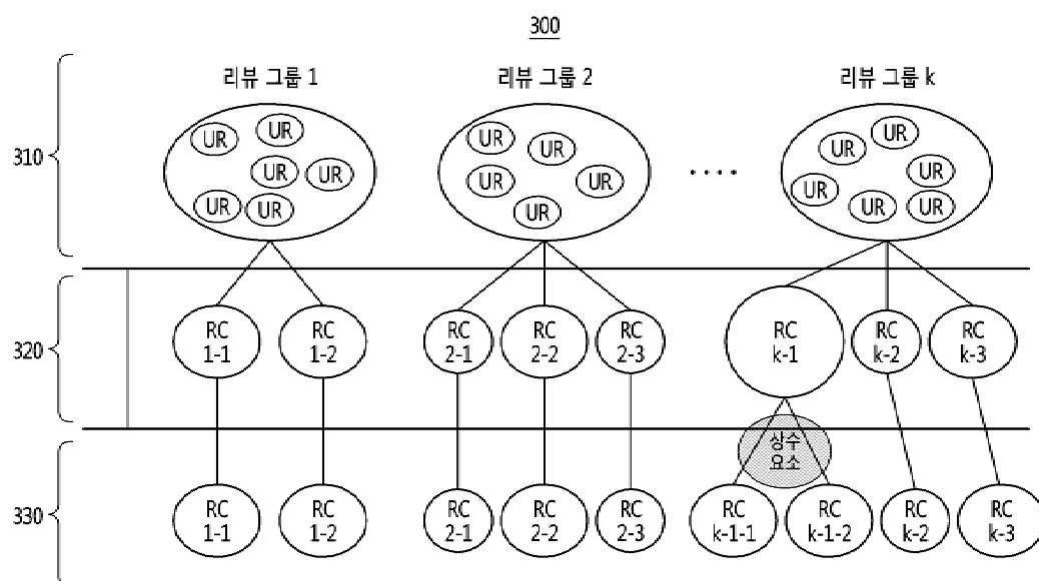
도면1



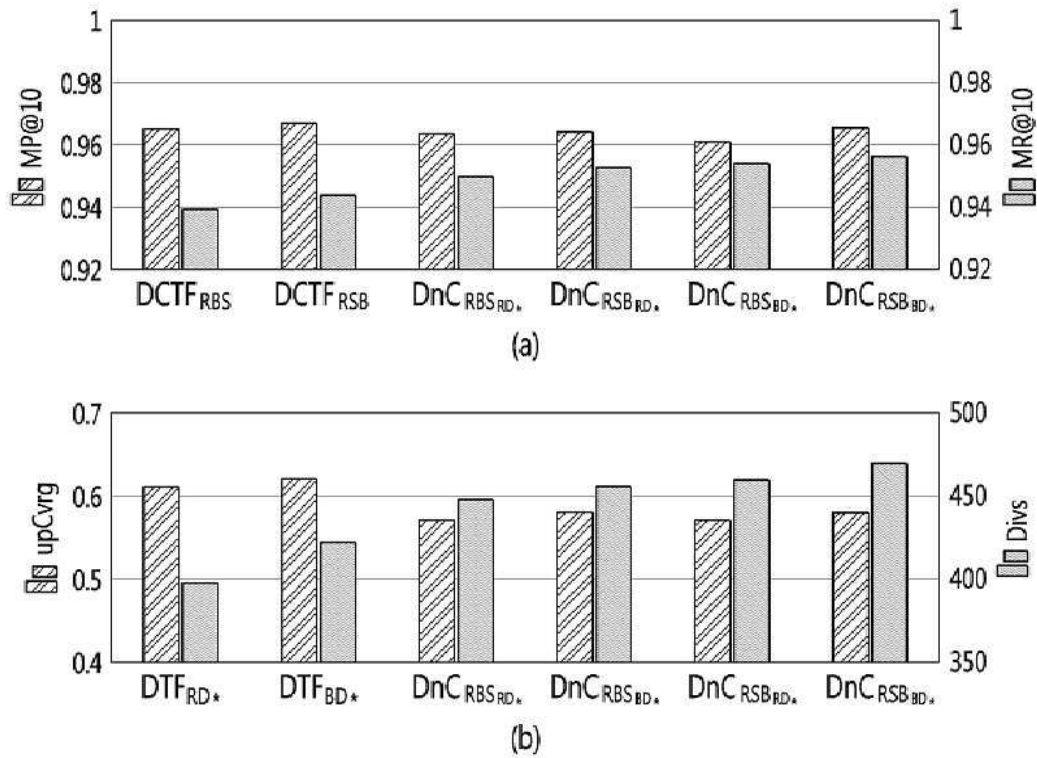
도면2



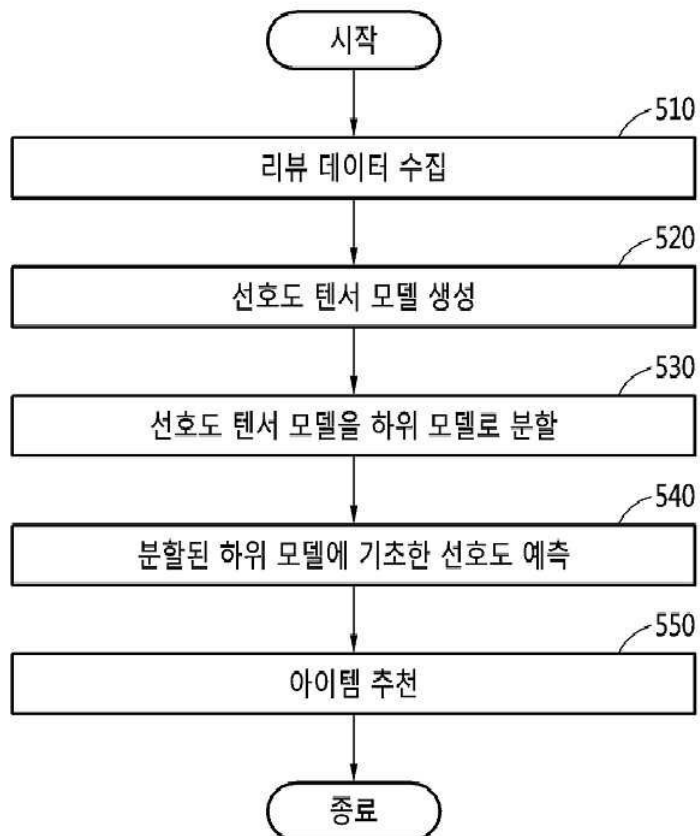
도면3



도면4



도면5



도면6

