



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0019353
(43) 공개일자 2025년02월10일

- | | |
|--|--|
| <p>(51) 국제특허분류(Int. Cl.)
G06Q 30/06 (2023.01) G06F 16/9035 (2019.01)
G06F 16/9535 (2019.01) G06N 20/00 (2019.01)</p> <p>(52) CPC특허분류
G06Q 30/0631 (2013.01)
G06F 16/9035 (2019.01)</p> <p>(21) 출원번호 10-2023-0100387
(22) 출원일자 2023년08월01일
심사청구일자 2023년08월01일</p> | <p>(71) 출원인
인하대학교 산학협력단
인천광역시 미추홀구 인하로 100(용현동, 인하대학교)</p> <p>(72) 발명자
서영덕
인천광역시 미추홀구 인하로 100(용현동, 인하대학교)</p> <p>이승주
인천광역시 미추홀구 소성로 16, 인하 아리스타 1동 723호(용현동)</p> <p>(74) 대리인
양성보</p> |
|--|--|

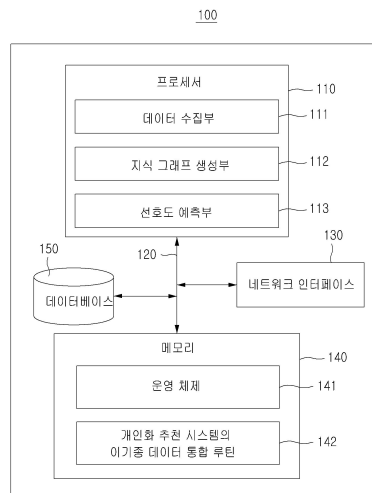
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 지식 그래프 상의 개체 간 관계 모델링을 통한 개인화 추천 시스템의 이중 데이터 통합 방법 및 시스템

(57) 요약

지식 그래프 상의 개체 간 관계 모델링을 통한 개인화 추천 시스템의 이중 데이터 통합 방법 및 시스템이 제시된다. 본 발명에서 제안하는 지식 그래프 상의 개체 간 관계 모델링을 통한 개인화 추천 시스템의 이중 데이터 통합 시스템은 사용자별 개인화 추천을 위한 사용자 및 아이템에 관한 보조 정보를 수집하는 데이터 수집부, 상기 수집된 보조 정보를 기반으로 지식 그래프 상의 개체 간 관계의 삼중항 형태로 지식 그래프를 생성하는 지식 그래프 생성부 및 상기 생성된 지식 그래프 상의 개체와 신규 개체 간의 잠재 관계를 모델링하고, 딥러닝을 사용하여 신규관계를 정의하고, 상기 정의된 신규 관계를 상기 지식 그래프 상에 추가하여 지식 그래프를 확장하며, 상기 확장된 지식 그래프를 학습하여 사용자 및 아이템에 관한 특성을 학습하고, 사용자의 선호도를 예측하고, 예측된 사용자의 선호도에 따라 사용자에게 제품을 추천하는 선호도 예측부를 포함한다.

대 표 도 - 도1



(52) CPC특허분류

G06F 16/9535 (2019.01)

G06N 20/00 (2021.08)

G06Q 30/0201 (2023.01)

G06Q 30/0202 (2023.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711181937
과제번호	2022R1C1C1012408
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	우수신진연구
연구과제명	[Ezbaro] 게임이론과 그래프 신경망 기반의 저비용/고효율그룹 추천 시스템
기 여 율	7/10
과제수행기관명	인하대학교
연구기간	2023.03.01 ~ 2024.02.29

이 발명을 지원한 국가연구개발사업

과제고유번호	1711193695
과제번호	2022-0-00448-002
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	방송통신산업기술개발사업
연구과제명	[Ezbaro] 인간처럼 회상이 가능한 인공 신경망 지속학습플랫폼 개발(2차년도)
기 여 율	2/10
과제수행기관명	인하대학교 산학협력단
연구기간	2023.01.01 ~ 2023.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711179347
과제번호	00155915
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신인재양성(교육운영수익)
연구과제명	[Ezbaro][정부+민간] 인공지능융합혁신인재양성(2차년도)
기 여 율	1/10
과제수행기관명	인하대학교 산학협력단
연구기간	2023.01.01 ~ 2023.12.31

명세서

청구범위

청구항 1

사용자별 개인화 추천을 위한 사용자 및 아이템에 관한 보조 정보를 수집하는 데이터 수집부;

상기 수집된 보조 정보를 기반으로 지식 그래프 상의 개체 간 관계의 삼중항 형태로 지식 그래프를 생성하는 지식 그래프 생성부; 및

상기 생성된 지식 그래프 상의 개체와 신규 개체 간의 잠재 관계를 모델링하고, 딥러닝을 사용하여 신규관계를 정의하고, 상기 정의된 신규 관계를 상기 지식 그래프 상에 추가하여 지식 그래프를 확장하며, 상기 확장된 지식 그래프를 학습하여 사용자 및 아이템에 관한 특성을 학습하고, 사용자의 선호도를 예측하고, 예측된 사용자의 선호도에 따라 사용자에게 제품을 추천하는 선호도 예측부

를 포함하는 개인화 추천을 위한 이중 데이터 통합 시스템.

청구항 2

제1항에 있어서,

상기 선호도 예측부는,

상기 생성된 지식 그래프 상의 개체와 신규 개체 간의 잠재 관계를 모델링하기 위해 관계를 정의하고자 하는 상기 두 개체 유형의 동일 여부를 확인하고, 동일 여부에 따라 이중 개체 관계 모델링 또는 동종 개체 관계 모델링을 수행하여 신규 관계를 정의하는

개인화 추천을 위한 이중 데이터 통합 시스템.

청구항 3

제1항에 있어서,

상기 선호도 예측부는,

단일 지식 베이스 기반으로 생성된 지식 그래프에 대하여, 개체 유형에 따른 협업 필터링을 통해 사용자 및 아이템 간의 예측 평점 기반 관계 모델링을 수행하고, 상기 예측 평점 및 사용자 및 아이템 간 신규 관계 여부에 따라 신규 관계를 정의 하고, 상기 정의된 신규 관계를 상기 지식 그래프 상에 추가하여 지식 그래프를 확장하는

개인화 추천을 위한 이중 데이터 통합 시스템.

청구항 4

제1항에 있어서,

상기 선호도 예측부는,

단일 지식 베이스 기반으로 생성된 지식 그래프에 대하여, 외부 지식 베이스로부터 추출한 신규 개체를 추가하고, 기존 개체들 간, 기존 개체-신규 개체 간, 신규 개체들 간에 따라 예측 평점 기반 관계 모델링을 수행하고, 상기 예측 평점에 따라 신규 관계를 정의 하고, 상기 정의된 신규 관계를 상기 지식 그래프 상에 추가하여 지식 그래프를 확장하는

개인화 추천을 위한 이중 데이터 통합 시스템.

청구항 5

제1항에 있어서,

상기 선호도 예측부는,

단일 지식 베이스 기반으로 생성된 지식 그래프 및 외부 지식 베이스 기반으로 생성된 지식 그래프 각각에 대하여, 상기 생성된 지식 그래프 상의 개체 간 관계 모델링을 수행하고, 상기 개체 간 관계 모델링을 통해 단일 지식 베이스 기반으로 생성된 지식 그래프 및 외부 지식 베이스 기반으로 생성된 지식 그래프를 통합하여 지식 그래프를 확장하는

개인화 추천을 위한 이중 데이터 통합 시스템.

청구항 6

데이터 수집부를 통해 사용자별 개인화 추천을 위한 사용자 및 아이템에 관한 보조 정보를 수집하는 단계;

지식 그래프 생성부를 통해 상기 수집된 보조 정보를 기반으로 지식 그래프 상의 개체 간 관계의 삼중항 형태로 지식 그래프를 생성하는 단계; 및

선호도 예측부를 통해 상기 생성된 지식 그래프 상의 개체와 신규 개체 간의 잠재 관계를 모델링하고, 딥러닝을 사용하여 신규관계를 정의하는 단계;

선호도 예측부를 통해 상기 정의된 신규 관계를 상기 지식 그래프 상에 추가하여 지식 그래프를 확장하는 단계; 및

선호도 예측부를 통해 상기 확장된 지식 그래프를 학습하여 사용자 및 아이템에 관한 특성을 학습하고, 사용자의 선호도를 예측하고, 예측된 사용자의 선호도에 따라 사용자에게 제품을 추천하는 단계

를 포함하는 개인화 추천을 위한 이중 데이터 통합 방법.

청구항 7

제6항에 있어서,

상기 선호도 예측부를 통해 상기 생성된 지식 그래프 상의 개체와 신규 개체 간의 잠재 관계를 모델링하고, 딥러닝을 사용하여 신규관계를 정의하는 단계는,

상기 생성된 지식 그래프 상의 개체와 신규 개체 간의 잠재 관계를 모델링하기 위해 관계를 정의하고자 하는 상기 두 개체 유형의 동일 여부를 확인하고, 동일 여부에 따라 이중 개체 관계 모델링 또는 동종 개체 관계 모델링을 수행하여 신규 관계를 정의하는

개인화 추천을 위한 이중 데이터 통합 방법.

청구항 8

제6항에 있어서,

상기 선호도 예측부를 통해 상기 정의된 신규 관계를 상기 지식 그래프 상에 추가하여 지식 그래프를 확장하는 단계는,

단일 지식 베이스 기반으로 생성된 지식 그래프에 대하여, 개체 유형에 따른 협업 필터링을 통해 사용자 및 아이템 간의 예측 평점 기반 관계 모델링을 수행하고, 상기 예측 평점 및 사용자 및 아이템 간 신규 관계 여부에 따라 신규 관계를 정의 하고, 상기 정의된 신규 관계를 상기 지식 그래프 상에 추가하여 지식 그래프를 확장하는

개인화 추천을 위한 이중 데이터 통합 방법.

청구항 9

제6항에 있어서,

상기 선호도 예측부를 통해 상기 정의된 신규 관계를 상기 지식 그래프 상에 추가하여 지식 그래프를 확장하는 단계는,

단일 지식 베이스 기반으로 생성된 지식 그래프에 대하여, 외부 지식 베이스로부터 추출한 신규 개체를 추가하고, 기존 개체들 간, 기존 개체-신규 개체 간, 신규 개체들 간에 따라 예측 평점 기반 관계 모델링을 수행하고, 상기 예측 평점에 따라 신규 관계를 정의 하고, 상기 정의된 신규 관계를 상기 지식 그래프 상에 추가하여 지식

그래프를 확장하는

개인화 추천을 위한 이중 데이터 통합 방법.

청구항 10

제6항에 있어서,

상기 선호도 예측부를 통해 상기 정의된 신규 관계를 상기 지식 그래프 상에 추가하여 지식 그래프를 확장하는 단계는,

단일 지식 베이스 기반으로 생성된 지식 그래프 및 외부 지식 베이스 기반으로 생성된 지식 그래프 각각에 대하여, 상기 생성된 지식 그래프 상의 개체 간 관계 모델링을 수행하고, 상기 개체 간 관계 모델링을 통해 단일 지식 베이스 기반으로 생성된 지식 그래프 및 외부 지식 베이스 기반으로 생성된 지식 그래프를 통합하여 지식 그래프를 확장하는

개인화 추천을 위한 이중 데이터 통합 방법.

발명의 설명

기술 분야

[0001] 본 발명은 지식 그래프 상의 개체 간 관계 모델링을 통한 개인화 추천 시스템의 이중 데이터 통합 방법 및 시스템에 관한 것이다.

배경 기술

[0002] 최근 개인화 서비스의 중요성이 매우 높아지고 있으며 다양한 분야에서 추천 시스템 연구가 활발하게 진행되고 있다. 그러나 대부분의 추천 시스템은 데이터 희소성(Data sparsity) 및 콜드 스타트(Cold start) 문제로 여전히 어려움을 겪고 있고, 이를 해결하기 위해 다양한 이중 데이터를 활용하는 방안을 제시하고 있다.

[0003] 추천 시스템 상에서 활용할 수 있는 데이터는 크게 사용자의 피드백 데이터와 아이템의 보조 정보 데이터가 존재한다. 사용자의 피드백 데이터에는 평점과 같이 사용자의 선호도를 직접적으로 표현하는 명시적 피드백과 구매력, 클릭 횟수와 같이 간접적으로 사용자의 선호도를 표현하는 암묵적 피드백이 존재한다. 이런 다양한 피드백 데이터는 사용자의 선호도를 표현하는데 이용된다. 반면, 아이템의 보조 정보 데이터에는 외부 지식 웹사이트 또는 지식 베이스에서 추출한 아이템을 보조적으로 설명하는 방대한 양의 지식 데이터가 존재한다. 최근 많은 추천 시스템 연구에서는 이러한 많은 양의 지식 데이터를 활용하여 데이터 부족 문제를 해결하고 추천 시스템의 성능을 높이는 연구가 등장하고 있다.

[0004] 지식 데이터를 활용하는 연구에는 크게 온톨로지 기반 연구와 지식 그래프 기반 추천 연구가 존재한다. 온톨로지 구축을 위해서는 개체(Entity)의 속성을 사람이 직접 분석하고 그들 간의 관계를 계층적으로 정의하는 추상화 작업이 필요하다. 하지만, 온톨로지는 "사용자 A가 상품 B를 구매했다"와 같은 개체 간의 명시적 관계(Explicit relation)만 정의할 수 있다. 추천 시스템 데이터의 특성상 명시적 관계의 수는 전체 개체 수와 비교하여 매우 적기 때문에 온톨로지만 사용할 경우 개체 간 관계의 대부분을 표현할 수 없다. 또한, 기존의 온톨로지 구축 방법은 정해진 지식 베이스를 기반으로 데이터 간 관계를 정의하기 때문에, 새로운 지식 베이스가 추가될 경우 새로운 온톨로지 구축 작업이 필수적이다. 이러한 이유로 대부분 온톨로지 기반 추천 시스템 연구는 새로운 온톨로지 구축 방법을 제시하기보단 기존의 온톨로지를 활용하여 추천 성능을 개선하는 데 집중하고 있다. 따라서 온톨로지 기반 추천 시스템 연구는 다중 지식 베이스를 기반으로 하는 이중 데이터 통합에 대한 궁극적인 해결책을 제공한다고 말할 수 없다.

[0005] 한편 지식 그래프 기반 추천 시스템 연구에서는 사전 정의된 온톨로지를 활용하거나 노동 집약적인 지식 베이스의 개체 명 일치 작업을 통해 지식 그래프를 생성한다. 이로 인해 지식 그래프 상에 존재하는 개체 간 관계도 여전히 명시적 관계에만 치우쳐져 있다는 한계가 존재한다. 또한 지식 그래프 기반 추천 시스템 연구에서는 대부분 단일 지식 베이스만 활용하여 지식 그래프를 구축한다. 따라서, 개체 간 관계가 풍부하지 못하고 누락된 지식이 존재할 가능성이 크며, 이는 정확한 추천을 제공하기에 지식이 불완전함을 의미한다.

[0006] 따라서, 최근 연구에서 다양한 지식 데이터를 활용하기 위해 이중 데이터 통합의 중요성을 강조하고 있지만 대부분 이중 데이터 간의 명시적 관계에만 집중한 낮은 데이터 품질 문제를 간과한다. 또한, 단일 지식 베이스 기

반 이중 데이터 활용이라는 점에서 여전히 지식 데이터 부족을 해결하지 못한다. 그러므로 데이터의 품질을 개선하면서 지식 데이터의 부족을 해결할 수 있는 지식 그래프 상에서의 이중 데이터 통합 방식이 필요하다.

선행기술문헌

특허문헌

- [0007] (특허문헌 0001) [1] Xiou Ge, Yun-Cheng Wang, Bin Wang and C.C. Jay Kuo, "CORE: A knowledge graph entity type prediction method via complex space regression and embedding," Pattern Recognition Letters, Vol. 157, pp. 97 103, May 2022.
- (특허문헌 0002) [2] R. Xie, Z. Liu and M. Sun, "Representation Learning of Knowledge Graphs with Hierarchical Types," Proc. of the 25th International Joint Conference on Artificial Intelligence, pp. 2965-2971, New York, NY, US, Jul. 2016.

발명의 내용

해결하려는 과제

- [0008] 본 발명이 이루고자 하는 기술적 과제는 지식 그래프 상에 새로운 개체 및 관계를 추가하여 지식 그래프를 확장하고 이를 통해 데이터 부족 문제를 해결하기 위한 지식 그래프 상의 개체 간 관계 모델링을 통한 개인화 추천 시스템의 이중 데이터 통합 방법 및 시스템을 제공하는데 있다. 또한, 본 발명은 딥러닝 기반 관계 모델링을 통해 개체 간 신규 관계를 정의함으로써, 기존의 명시적 관계뿐만 아니라 잠재 관계를 추가하여 지식 그래프의 데이터 품질을 개선하고자 한다.

과제의 해결 수단

- [0009] 일 측면에 있어서, 본 발명에서 제안하는 지식 그래프 상의 개체 간 관계 모델링을 통한 개인화 추천 시스템의 이중 데이터 통합 시스템은 사용자별 개인화 추천을 위한 사용자 및 아이템에 관한 보조 정보를 수집하는 데이터 수집부, 상기 수집된 보조 정보를 기반으로 지식 그래프 상의 개체 간 관계의 삼중항 형태로 지식 그래프를 생성하는 지식 그래프 생성부 및 상기 생성된 지식 그래프 상의 개체와 신규 개체 간의 잠재 관계를 모델링하고, 딥러닝을 사용하여 신규관계를 정의하고, 상기 정의된 신규 관계를 상기 지식 그래프 상에 추가하여 지식 그래프를 확장하며, 상기 확장된 지식 그래프를 학습하여 사용자 및 아이템에 관한 특성을 학습하고, 사용자의 선호도를 예측하고, 예측된 사용자의 선호도에 따라 사용자에게 제품을 추천하는 선호도 예측부를 포함한다.
- [0010] 상기 선호도 예측부는 상기 생성된 지식 그래프 상의 개체와 신규 개체 간의 잠재 관계를 모델링하기 위해 관계를 정의하고자 하는 상기 두 개체 유형의 동일 여부를 확인하고, 동일 여부에 따라 이중 개체 관계 모델링 또는 동종 개체 관계 모델링을 수행하여 신규 관계를 정의한다.
- [0011] 상기 선호도 예측부는 단일 지식 베이스 기반으로 생성된 지식 그래프에 대하여, 개체 유형에 따른 협업 필터링을 통해 사용자 및 아이템 간의 예측 평점 기반 관계 모델링을 수행하고, 상기 예측 평점 및 사용자 및 아이템 간 신규 관계 여부에 따라 신규 관계를 정의 하고, 상기 정의된 신규 관계를 상기 지식 그래프 상에 추가하여 지식 그래프를 확장한다.
- [0012] 상기 선호도 예측부는 단일 지식 베이스 기반으로 생성된 지식 그래프에 대하여, 외부 지식 베이스로부터 추출한 신규 개체를 추가하고, 기존 개체들 간, 기존 개체-신규 개체 간, 신규 개체들 간에 따라 예측 평점 기반 관계 모델링을 수행하고, 상기 예측 평점에 따라 신규 관계를 정의 하고, 상기 정의된 신규 관계를 상기 지식 그래프 상에 추가하여 지식 그래프를 확장한다.
- [0013] 상기 선호도 예측부는 단일 지식 베이스 기반으로 생성된 지식 그래프 및 외부 지식 베이스 기반으로 생성된 지식 그래프 각각에 대하여, 상기 생성된 지식 그래프 상의 개체 간 관계 모델링을 수행하고, 상기 개체 간 관계 모델링을 통해 단일 지식 베이스 기반으로 생성된 지식 그래프 및 외부 지식 베이스 기반으로 생성된 지식 그래프를 통합하여 지식 그래프를 확장한다.
- [0014] 또 다른 일 측면에 있어서, 본 발명에서 제안하는 지식 그래프 상의 개체 간 관계 모델링을 통한 개인화 추천

방법은 데이터 수집부를 통해 사용자별 개인화 추천을 위한 사용자 및 아이템에 관한 보조 정보를 수집하는 단계, 지식 그래프 생성부를 통해 상기 수집된 보조 정보를 기반으로 지식 그래프 상의 개체 간 관계의 삼중항 형태로 지식 그래프를 생성하는 단계 및 선호도 예측부를 통해 상기 생성된 지식 그래프 상의 개체와 신규 개체 간의 잠재 관계를 모델링하고, 딥러닝을 사용하여 신규관계를 정의하는 단계, 선호도 예측부를 통해 상기 정의된 신규 관계를 상기 지식 그래프 상에 추가하여 지식 그래프를 확장하는 단계 및 선호도 예측부를 통해 상기 확장된 지식 그래프를 학습하여 사용자 및 아이템에 관한 특성을 학습하고, 사용자의 선호도를 예측하고, 예측된 사용자의 선호도에 따라 사용자에게 제품을 추천하는 단계를 포함한다.

발명의 효과

[0015] 본 발명의 실시예들에 따르면 지식 그래프 상의 개체 간 관계 모델링을 통한 개인화 추천 시스템의 이중 데이터 통합 방법 및 시스템을 통해 지식 그래프 상에 새로운 개체 및 관계를 추가하여 지식 그래프를 확장하고 이를 통해 데이터 부족 문제를 해결할 수 있다. 또한, 본 발명은 딥러닝 기반 관계 모델링을 통해 개체 간 신규 관계를 정의함으로써, 기존의 명시적 관계뿐만 아니라 잠재 관계를 추가하여 지식 그래프의 데이터 품질을 개선할 수 있다. 따라서 본 발명은 최종적으로 다중 지식 베이스로부터 추출한 이중 데이터 간의 상호 호환적 통합이 지식 그래프 상에서 가능하도록 하여 추천 시스템의 상호운용성을 해결할 수 있다.

도면의 간단한 설명

[0016] 도 1은 본 발명의 일 실시예에 따른 개인화 추천을 위한 이중 데이터 통합 시스템의 구성을 나타내는 도면이다.
 도 2는 본 발명의 일 실시예에 따른 개인화 추천을 위한 이중 데이터 통합 방법을 설명하기 위한 흐름도이다.
 도 3은 본 발명의 일 실시예에 따른 지식 그래프 상의 개체와 신규 개체 간의 관계를 모델링하는 과정을 설명하기 위한 흐름도이다.
 도 4는 본 발명의 일 실시예에 따른 개체 유형에 기반한 관계 모델링을 설명하기 위한 도면이다.
 도 5는 본 발명의 일 실시예에 따른 협업 필터링에 기반한 관계 모델링을 통한 지식 그래프 확장 방법을 설명하기 위한 흐름도이다.
 도 6은 본 발명의 일 실시예에 따른 협업 필터링에 기반한 관계 모델링을 설명하기 위한 도면이다.
 도 7은 본 발명의 일 실시예에 따른 신규 개체의 관계 모델링을 통한 지식 그래프 확장 방법을 설명하기 위한 흐름도이다.
 도 8은 본 발명의 일 실시예에 따른 신규 개체의 관계 모델링을 설명하기 위한 도면이다.
 도 9는 본 발명의 일 실시예에 따른 다중 지식 베이스 기반 이중 데이터의 통합 관계 모델링을 통한 지식 그래프 확장 방법을 설명하기 위한 흐름도이다.
 도 10은 본 발명의 일 실시예에 따른 다중 지식 베이스 기반 이중 데이터의 통합 관계 모델링을 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0017] 본 발명은 추천 시스템 상의 다양한 보조 정보를 지식 그래프 상에 통합하여 지식 그래프를 확장하는 지식 그래프 상의 개체 간 관계 모델링을 통한 개인화 추천 시스템의 이중 데이터 통합 방법 및 시스템에 관한 것이다. 본 발명은 개체 유형 정보를 활용하여 지식 그래프 상의 개체 간 관계 모델링을 통해 신규 관계를 점진적으로 추가함으로써 확장된다. 더 나아가 다양한 외부 지식 베이스를 활용하여 신규 개체와 그들 간의 잠재 관계를 지식 그래프 상에 추가하여 부족한 지식 데이터를 보완하고 지식 그래프의 품질을 향상시켜 추천 시스템의 성능을 개선할 수 있도록 하는 이중 데이터 통합 방법에 관한 것이다. 이하, 본 발명의 실시 예를 첨부된 도면을 참조하여 상세하게 설명한다.

[0019] 도 1은 본 발명의 일 실시예에 따른 개인화 추천을 위한 이중 데이터 통합 시스템의 구성을 나타내는 도면이다.

[0020] 본 실시예에 따른 이중 데이터 통합 시스템(100)은 프로세서(110), 버스(120), 네트워크 인터페이스(130), 메모리(140) 및 데이터베이스(150)를 포함할 수 있다. 메모리(140)는 운영체제(141) 및 개인화 추천을 위한 이중

데이터 통합 루틴(142)을 포함할 수 있다. 프로세서(110)는 데이터 수집부(111), 지식 그래프 생성부(112) 및 선호도 예측부(113)를 포함할 수 있다. 다른 실시예들에서 이종 데이터 통합 시스템(100)은 도 1의 구성요소들보다 더 많은 구성요소들을 포함할 수도 있다. 그러나, 대부분의 종래기술적 구성요소들을 명확하게 도시할 필요성은 없다. 예를 들어, 이종 데이터 통합 시스템(100)은 디스플레이나 트랜시버(transceiver)와 같은 다른 구성요소들을 포함할 수도 있다.

[0021] 메모리(140)는 컴퓨터에서 판독 가능한 기록 매체로서, RAM(random access memory), ROM(read only memory) 및 디스크 드라이브와 같은 비소멸성 대용량 기록장치(permanent mass storage device)를 포함할 수 있다. 또한, 메모리(140)에는 운영체제(141)와 개인화 추천을 위한 이종 데이터 통합 루틴(142)을 위한 프로그램 코드가 저장될 수 있다. 이러한 소프트웨어 구성요소들은 드라이브 메커니즘(drive mechanism, 미도시)을 이용하여 메모리(140)와는 별도의 컴퓨터에서 판독 가능한 기록 매체로부터 로딩될 수 있다. 이러한 별도의 컴퓨터에서 판독 가능한 기록 매체는 플로피 드라이브, 디스크, 테이프, DVD/CD-ROM 드라이브, 메모리 카드 등의 컴퓨터에서 판독 가능한 기록 매체(미도시)를 포함할 수 있다. 다른 실시예에서 소프트웨어 구성요소들은 컴퓨터에서 판독 가능한 기록 매체가 아닌 네트워크 인터페이스(130)를 통해 메모리(140)에 로딩될 수도 있다.

[0022] 버스(120)는 이종 데이터 통합 시스템(100)의 구성요소들간의 통신 및 데이터 전송을 가능하게 할 수 있다. 버스(120)는 고속 시리얼 버스(high-speed serial bus), 병렬 버스(parallel bus), SAN(Storage Area Network) 및/또는 다른 적절한 통신 기술을 이용하여 구성될 수 있다.

[0023] 네트워크 인터페이스(130)는 이종 데이터 통합 시스템(100)을 컴퓨터 네트워크에 연결하기 위한 컴퓨터 하드웨어 구성요소일 수 있다. 네트워크 인터페이스(130)는 이종 데이터 통합 시스템(100)을 무선 또는 유선 커넥션을 통해 컴퓨터 네트워크에 연결시킬 수 있다.

[0024] 데이터베이스(150)는 개인화 추천을 위한 이종 데이터 통합을 위해 필요한 모든 정보를 저장 및 유지하는 역할을 할 수 있다. 도 1에서는 이종 데이터 통합 시스템(100)의 내부에 데이터베이스(150)를 구축하여 포함하는 것으로 도시하고 있으나, 이에 한정되는 것은 아니며 시스템 구현 방식이나 환경 등에 따라 생략될 수 있고 혹은 전체 또는 일부의 데이터베이스가 별개의 다른 시스템 상에 구축된 외부 데이터베이스로서 존재하는 것 또한 가능하다.

[0025] 프로세서(110)는 기본적인 산술, 로직 및 이종 데이터 통합 시스템(100)의 입출력 연산을 수행함으로써, 컴퓨터 프로그램의 명령을 처리하도록 구성될 수 있다. 명령은 메모리(140) 또는 네트워크 인터페이스(130)에 의해, 그리고 버스(120)를 통해 프로세서(110)로 제공될 수 있다. 프로세서(110)는 데이터 수집부(111), 지식 그래프 생성부(112) 및 선호도 예측부(113)를 위한 프로그램 코드를 실행하도록 구성될 수 있다. 이러한 프로그램 코드는 메모리(140)와 같은 기록 장치에 저장될 수 있다.

[0026] 데이터 수집부(111), 지식 그래프 생성부(112) 및 선호도 예측부(113)는 도 2의 단계들(210~250)을 수행하기 위해 구성될 수 있다.

[0027] 이종 데이터 통합 시스템(100)은 데이터 수집부(111), 지식 그래프 생성부(112) 및 선호도 예측부(113)를 포함할 수 있다.

[0028] 본 발명의 실시예에 따른 데이터 수집부(111)는 사용자별 개인화 추천을 위한 사용자 및 아이템에 관한 보조 정보를 수집한다.

[0029] 본 발명의 실시예에 따른 데이터 수집부(111)는 사용자 및 아이템 관련 보조 정보 데이터를 여러 지식 베이스 상에서 추출하여 수집한다. 사용자 관련 정보는 인구통계학적인 정보, 아이템 관련 정보는 아이템의 특성을 보조적으로 설명하는 지식 데이터를 포함할 수 있다.

[0030] 본 발명의 실시예에 따른 지식 그래프 생성부(112)는 상기 수집된 보조 정보를 기반으로 지식 그래프 상의 개체 간 관계의 삼중항 형태로 지식 그래프를 생성한다. 각 삼중항은 주어 개체(Head entity), 관계(Relation), 술어 개체(Tail entity)로 이루어져 있다.

[0031] 본 발명의 실시예에 따른 지식 그래프 생성부(112)는 데이터 수집부(111)에서 수집한 여러 종류의 데이터를 지식 그래프 상에 정점과 간선으로 표현하여 반영한다. 일반적으로, $\mathcal{E} = \{\epsilon_1, \epsilon_2, \dots, \epsilon_k\}$ 은 지식 그래프 상의 정점 즉, 개체의 집합을 표현한다. $R = \{r_1, r_2, \dots, r_l\}$ 은 지식 그래프 상의 간선, 즉 관계의 집합을 나타내며, k, l 은 개체와 관계 각각의 개수를 의미한다. 따라서, 지식 그래프 KG는 카데시안 곱인 $\mathcal{E} \times R \times \mathcal{E}$ 의 부분집합이며

지식 그래프는 $KG = \{(h, r, t) | h, t \in \varepsilon, r \in R\}$ 와 같이 정의한다.

- [0032] 본 발명의 실시예에 따른 선호도 예측부(113)는 상기 생성된 지식 그래프 상의 개체와 신규 개체 간의 잠재 관계를 모델링하고, 딥러닝을 사용하여 신규관계를 정의한다. 본 발명의 실시예에 따른 선호도 예측부(113)는 상기 정의된 신규 관계를 상기 지식 그래프 상에 추가하여 지식 그래프를 확장한다. 본 발명의 실시예에 따른 선호도 예측부(113)는 상기 확장된 지식 그래프를 학습하여 사용자 및 아이টে에 관한 특성을 학습하고, 사용자의 선호도를 예측하고, 예측된 사용자의 선호도에 따라 사용자에게 제품을 추천한다.
- [0033] 본 발명의 실시예에 따른 선호도 예측부(113)는 상기 생성된 지식 그래프 상의 개체와 신규 개체 간의 잠재 관계를 모델링하기 위해 관계를 정의하고자 하는 상기 두 개체 유형의 동일 여부를 확인하고, 동일 여부에 따라 이중 개체 관계 모델링 또는 동종 개체 관계 모델링을 수행하여 신규 관계를 정의한다.
- [0034] 본 발명의 일 실시예에 따른 선호도 예측부(113)는 단일 지식 베이스 기반으로 생성된 지식 그래프에 대하여, 개체 유형에 따른 협업 필터링을 통해 사용자 및 아이টে 간의 예측 평점 기반 관계 모델링을 수행하고, 상기 예측 평점 및 사용자 및 아이টে 간 신규 관계 여부에 따라 신규 관계를 정의 하고, 상기 정의된 신규 관계를 상기 지식 그래프 상에 추가하여 지식 그래프를 확장한다.
- [0035] 본 발명의 또 다른 실시예에 따른 선호도 예측부(113)는 단일 지식 베이스 기반으로 생성된 지식 그래프에 대하여, 외부 지식 베이스로부터 추출한 신규 개체를 추가하고, 기존 개체들 간, 기존 개체-신규 개체 간, 신규 개체들 간에 따라 예측 평점 기반 관계 모델링을 수행하고, 상기 예측 평점에 따라 신규 관계를 정의 하고, 상기 정의된 신규 관계를 상기 지식 그래프 상에 추가하여 지식 그래프를 확장한다.
- [0036] 본 발명의 또 다른 실시예에 따른 선호도 예측부(113)는 단일 지식 베이스 기반으로 생성된 지식 그래프 및 외부 지식 베이스 기반으로 생성된 지식 그래프 각각에 대하여, 상기 생성된 지식 그래프 상의 개체 간 관계 모델링을 수행하고, 상기 개체 간 관계 모델링을 통해 단일 지식 베이스 기반으로 생성된 지식 그래프 및 외부 지식 베이스 기반으로 생성된 지식 그래프를 통합하여 지식 그래프를 확장한다.
- [0038] 도 2는 본 발명의 일 실시예에 따른 개인화 추천을 위한 이중 데이터 통합 방법을 설명하기 위한 흐름도이다.
- [0039] 본 발명의 실시예에 따른 개인화 추천을 위한 이중 데이터 통합 방법은 데이터 수집부를 통해 사용자별 개인화 추천을 위한 사용자 및 아이টে에 관한 보조 정보를 수집하는 단계(210), 지식 그래프 생성부를 통해 상기 수집된 보조 정보를 기반으로 지식 그래프 상의 개체 간 관계의 삼중항 형태로 지식 그래프를 생성하는 단계(220), 선호도 예측부를 통해 상기 생성된 지식 그래프 상의 개체와 신규 개체 간의 잠재 관계를 모델링하고, 딥러닝을 사용하여 신규관계를 정의하는 단계(230), 선호도 예측부를 통해 상기 정의된 신규 관계를 상기 지식 그래프 상에 추가하여 지식 그래프를 확장하는 단계(240) 및 선호도 예측부를 통해 상기 확장된 지식 그래프를 학습하여 사용자 및 아이টে에 관한 특성을 학습하고, 사용자의 선호도를 예측하고, 예측된 사용자의 선호도에 따라 사용자에게 제품을 추천하는 단계(250)를 포함한다.
- [0040] 단계(210)에서, 데이터 수집부를 통해 사용자별 개인화 추천을 위한 사용자 및 아이টে에 관한 보조 정보를 수집한다.
- [0041] 본 발명의 실시예에 따르면, 데이터 수집부는 사용자 및 아이টে 관련 보조 정보 데이터를 여러 지식 베이스 상에서 추출하여 수집한다. 사용자 관련 정보는 인구통계학적인 정보, 아이টে 관련 정보는 아이테의 특성을 보조적으로 설명하는 지식 데이터를 포함할 수 있다.
- [0042] 단계(220)에서, 지식 그래프 생성부를 통해 상기 수집된 보조 정보를 기반으로 지식 그래프 상의 개체 간 관계의 삼중항 형태로 지식 그래프를 생성한다.
- [0043] 본 발명의 실시예에 따르면, 지식 그래프 생성부는 데이터 수집부에서 수집한 여러 종류의 데이터를 지식 그래프 상에 정점과 간선으로 표현하여 반영한다. 일반적으로, $\varepsilon = \{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_k\}$ 은 지식 그래프 상의 정점 즉, 개체의 집합을 표현한다. $R = \{r_1, r_2, \dots, r_l\}$ 은 지식 그래프 상의 간선, 즉 관계의 집합을 나타내며, k, l 은 개체와 관계 각각의 개수를 의미한다. 따라서, 지식 그래프 KG는 카테시안 곱인 $\varepsilon \times R \times \emptyset$ 의 부분집합이며 지식 그래프는 $KG = \{(h, r, t) | h, t \in \varepsilon, r \in R\}$ 와 같이 정의한다.

- [0044] 단계(230)에서, 선호도 예측부를 통해 상기 생성된 지식 그래프 상의 개체와 신규 개체 간의 잠재 관계를 모델링하고, 딥러닝을 사용하여 신규관계를 정의한다.
- [0045] 단계(240)에서, 선호도 예측부를 통해 상기 정의된 신규 관계를 상기 지식 그래프 상에 추가하여 지식 그래프를 확장한다.
- [0046] 본 발명의 실시예에 따르면, 선호도 예측부는 상기 정의된 신규 관계를 상기 지식 그래프 상에 추가하여 지식 그래프를 확장한다.
- [0047] 본 발명의 실시예에 따르면, 선호도 예측부는 상기 생성된 지식 그래프 상의 개체와 신규 개체 간의 잠재 관계를 모델링하기 위해 관계를 정의하고자 하는 상기 두 개체 유형의 동일 여부를 확인하고, 동일 여부에 따라 이중 개체 관계 모델링 또는 동종 개체 관계 모델링을 수행하여 신규 관계를 정의한다.
- [0048] 본 발명의 일 실시예에 따르면, 선호도 예측부는 단일 지식 베이스 기반으로 생성된 지식 그래프에 대하여, 개체 유형에 따른 협업 필터링을 통해 사용자 및 아이템 간의 예측 평점 기반 관계 모델링을 수행하고, 상기 예측 평점 및 사용자 및 아이템 간 신규 관계 여부에 따라 신규 관계를 정의 하고, 상기 정의된 신규 관계를 상기 지식 그래프 상에 추가하여 지식 그래프를 확장한다.
- [0049] 본 발명의 또 다른 실시예에 따르면, 선호도 예측부는 단일 지식 베이스 기반으로 생성된 지식 그래프에 대하여, 외부 지식 베이스로부터 추출한 신규 개체를 추가하고, 기존 개체들 간, 기존 개체-신규 개체 간, 신규 개체들 간에 따라 예측 평점 기반 관계 모델링을 수행하고, 상기 예측 평점에 따라 신규 관계를 정의 하고, 상기 정의된 신규 관계를 상기 지식 그래프 상에 추가하여 지식 그래프를 확장한다.
- [0050] 본 발명의 또 다른 실시예에 따르면, 선호도 예측부는 단일 지식 베이스 기반으로 생성된 지식 그래프 및 외부 지식 베이스 기반으로 생성된 지식 그래프 각각에 대하여, 상기 생성된 지식 그래프 상의 개체 간 관계 모델링을 수행하고, 상기 개체 간 관계 모델링을 통해 단일 지식 베이스 기반으로 생성된 지식 그래프 및 외부 지식 베이스 기반으로 생성된 지식 그래프를 통합하여 지식 그래프를 확장한다.
- [0051] 단계(250)에서, 선호도 예측부를 통해 상기 확장된 지식 그래프를 학습하여 사용자 및 아이템에 관한 특성을 학습하고, 사용자의 선호도를 예측하고, 예측된 사용자의 선호도에 따라 사용자에게 제품을 추천한다.
- [0053] 도 3은 본 발명의 일 실시예에 따른 지식 그래프 상의 개체와 신규 개체 간의 관계를 모델링하는 과정을 설명하기 위한 흐름도이다.
- [0054] 도 3에 도시된 바와 같이, 본 발명의 실시예에 따른 지식 그래프 상의 개체와 신규 개체 간의 관계를 모델링 방법은 관계를 정의하고자 하는 두 개체 유형의 동일 여부 확인(310)과, 동일 여부에 따른 이중 관계 모델링 수행(320) 및 동종 관계 모델링 수행(330), 신규 관계 정의(340), 그리고 신규 관계 추가에 따른 지식 그래프 확장(350)을 포함한다.
- [0055] 개체 간 관계 모델링을 하기에 앞서, 본 발명의 실시예에 따르면 개체의 유형이라는 개념을 활용하여 모델링을 수행한다. 따라서, 두 개체 유형 동일 여부 확인(310)은 비교하고자 하는 두 개의 개체 유형을 확인하고 이를 기반으로 이중 개체 관계 모델링(320)과 동종 개체 관계 모델링(330)으로 나뉜다. 적절한 관계를 정의하기 위해 개체 유형이 $a \in A$ 인 모든 개체 집합을 고려하고, 각 개체는 $\varepsilon_a = \{e | e \in \varepsilon, \emptyset(e) = a\}$ 로 표현한다.
- [0056] 이중 개체 관계 모델링(320)에서는 서로 다른 두 개체 유형 집합인 ε_a 및 ε_b 가 관계 r_{ab} 를 기반으로 모델링 될 수 있는 경우, 일부 개체인 $e \in \varepsilon_a$ 및 $f \in \varepsilon_b$ 에 대해 (e, r_{ab}, f) 의 삼중항을 생성할 수 있다. 예를 들어, $\varepsilon_a = \text{Movie}$ 및 $\varepsilon_b = \text{Genre}$ 로 설정하면 두 개체 집합 간의 관계는 일부 영화 개체 $m \in \text{Movie}$ 와 일부 장르 개체 $j \in \text{Genre}$ 의 관련성을 기반으로 관계를 정의할 수 있고 다음과 같은 2개의 삼중항인 $(j, r_{\text{IsGenreToMovie}}, m), (m, r_{\text{GenreOfMovie}}, j)$ 를 생성할 수 있다. 만약 개체 간 관련성 여부가 기존에 존재하지 않는 경우, 두 개체 간의 새로운 관계를 정의(340)하여 지식 그래프에 추가하고 확장할 수 있다.
- [0057] 동종 개체 관계 모델링(330)에서는 두 개체가 속한 하나의 개체 유형 집합인 ε_a 가 관계 r_a 에 기반을 두어 모델

링 될 수 있다면, 일부 개체인 $e, f \in \varepsilon_a$ 에 대해 (e, r_a, f) 형식의 삼중항을 생성할 수 있다. 예를 들어, $\varepsilon_a = \text{Item}$ 로 설정하면 두 아이템 $i_1, i_2 \in \text{Item}$ 간의 관계는 i_1, i_2 의 아이템 유사도를 기반으로 정의할 수 있고, 다음과 같은 2개의 삼중항인 $(i_1, r_{\text{IsSimilarItem}}, i_2), (i_2, r_{\text{IsSimilarItem}}, i_1)$ 를 생성할 수 있다. 또한, 다중 지식 베이스의 이중 데이터 통합 과정에서 기존 지식 그래프에 존재하지 않았던 새로운 관계와 신규 개체를 추가할 수 있다.

[0058] 위와 같이 개체 유형 동일 여부에 따른 관계 모델링 방식(310-330)을 통해 개체 간 유의미한 신규 관계를 도출하고 이를 정의(340)하여 지식 그래프에 추가하여 확장하도록 한다.

[0060] 도 4는 본 발명의 일 실시예에 따른 개체 유형에 기반한 관계 모델링을 설명하기 위한 도면이다.

[0061] 본 발명의 실시예에 따르면, 지식 그래프 상의 개체 간 관계 모델링을 통해 새로운 관계 지식을 추가하기 위해 개체 유형 정보를 활용한다. 관계를 정의하고자 하는 두 개체 유형의 동일 여부에 따라 두 가지 방법으로 관계 모델링을 수행하며, 도 4에 도시된 바와 같이 이중 개체 관계 모델링(Inter-entity relation)(도 4(a))과 동종 개체 관계 모델링(Intra-entity relation)(도 4(b))으로 나뉜다. 적절한 관계를 정의하기 위해 개체 유형이 $a \in A$ 인 모든 개체 집합을 고려하고 각 개체는 $\varepsilon_a = \{e | e \in \varepsilon, \emptyset(e) = a\}$ 로 표현한다. 본 발명에서는 딥러닝 기반의 잠재 관계 모델링을 통해 지식 그래프를 확장하여 개체 간의 관계성을 강화하고, 다중 지식 베이스의 이중 데이터 통합을 달성한다.

[0063] 도 5는 본 발명의 일 실시예에 따른 협업 필터링에 기반한 관계 모델링을 통한 지식 그래프 확장 방법을 설명하기 위한 흐름도이다.

[0064] 도 5를 참조하면, 도 3의 과정을 통해 관계 모델링을 수행하고, 이를 기반으로 본 발명의 일 실시예에 따른 단일 지식 베이스 상에서 지식 그래프를 확장하는 과정을 보여준다.

[0065] 지식 그래프에 사용자와 아이템의 피드백 정보를 추가한 기존 사례가 존재하지만, 이는 추가된 피드백 정보의 양이 데이터 희소성 문제를 해결하기에 충분하지 않다. 따라서 도 5에서 설명하는 과정은 이러한 부족함을 보충하기 위해 추천 시스템 알고리즘을 기반으로 예측 평점과 아이템 간의 유사도 정보를 지식 그래프에 추가 관계로 반영한다.

[0066] 도 5에 도시된 바와 같이 단일 지식 베이스 기반으로 지식 그래프를 생성하고(510), 도 3의 개체 유형을 고려한 관계 모델링 방식을 기반으로 추천 시스템에서 대표적인 알고리즘인 협업 필터링을 통해 사용자와 아이템 간의 선호도 또는 아이템 간의 유사 관계를 유추한다(520).

[0067] 먼저, 예측 평점 기반 관계 모델링을 예시로 도 5를 설명하도록 한다. 이미 매겨진 평점에 따라 특정 사용자 u 가 아이템 i 에 미리 정해진 임계치 이상의 평점을 주었다면(530), 사용자와 아이템 개체 사이의 관계를 추가한다. 미리 정해진 임계치 미만일 경우 해당 관계는 누락한다(540).

[0068] 정의하고자 하는 관계의 신규 여부를 확인(550)하고 신규 관계라면 이중 개체 관계인 $r_{\text{Like}}, r_{\text{PositivelyRatedBy}}$ 를 새롭게 정의(570)하여 지식 그래프에 추가한다(580). 신규 관계가 아닌 기존에 정의된 관계라면, 해당 관계를 지식 그래프 상의 두 개체 사이에 추가한다(560). 따라서, 지식 그래프에 2개의 삼중항 (u, r_{Like}, i) 및 $(i, r_{\text{PositivelyRatedBy}}, u)$ 를 추가할 수 있다.

[0069] 만약 결측 평점이라면, 협업 필터링을 통해 예측한 평점(520)이 특정 임계치 이상(530)일 경우, 해당 관계 역시 신규 관계 여부를 확인한 후(550) 그에 맞게 관계를 정의(570)하고 지식 그래프에 추가한다(580). 예측 평점은 실제 평점과 비교하면 정확도가 다소 떨어지므로 예측 평점에 대한 임계치는 기존 평점에 대한 임계치보다 높게 설정해야 한다.

[0070] 위와 같이 협업 필터링을 통해 신규 관계를 추가함(580)으로써, 내부 지식 베이스의 지식 그래프를 확장하여 이를 기반으로 선호도를 예측할 수 있다.

[0072] 도 6은 본 발명의 일 실시예에 따른 협업 필터링에 기반한 관계 모델링을 설명하기 위한 도면이다.

[0073] 지식 그래프에 사용자와 아이템의 피드백 정보를 추가한 기존 사례가 존재하지만, 이는 추가된 피드백 정보의 양이 데이터 희소성 문제를 해결하기에 충분하지 않다. 첫 번째 시나리오는 이러한 부족함을 보충하기 위해 추천 시스템 알고리즘을 기반으로 예측 평점과 아이템 간의 유사도 정보를 지식 그래프에 추가 관계로 반영한다. 오 6과 같이 기존의 아이템 지식 그래프(610)를 기반으로, 협업 필터링(620)을 통해 사용자와 아이템 간의 선호도와 아이템 간의 유사 관계를 유추한다. 예를 들어, 사용자 u_2 개체와 아이템 i_1 개체 사이의 예측 평점이 특정한 임계치 이상으로 유추되면 두 개체 간의 선호 관계를 추가할 수 있다. 또한, 아이템 i_3 개체와 아이템 i_5 개체 사이의 유사도가 특정한 임계치 이상으로 유추되면 두 아이템 개체 간의 유사 관계를 추가할 수 있다. 이와 같이 신규 관계를 추가(630)함으로써, 내부 지식 베이스의 지식 그래프를 확장(640)할 수 있다. 도 6의 시나리오 상에서 고려할 수 있는 신규 관계와 구체적인 예시는 표 1과 같다.

[0074] <표 1>

Relation Type	Relation Name	Triples
Inter - entity relation	r_{Like}	(u, r_{Like}, i)
	$r_{Positively Rated By}$	$(i, r_{Positively Rated By}, u)$
Intra - entity relation	$r_{Is Similar User}$	$(u, r_{Is Similar User}, v)$
	$r_{Is Similar Item}$	$(i, r_{Is Similar Item}, j)$

[0075]

[0076] 본 발명의 실시예에 따른 예측 평점 기반 관계 모델링은 사용자와 아이템 개체 사이의 관계를 추가하기 위해 사전에 정의한 이중 개체 관계인 r_{Like} 와 $r_{Positively Rated By}$ 를 고려한다. 먼저, 이미 매겨진 평점에 따라 특정 사용자 u 가 아이템 i 에 특정한 임계치 이상의 평점을 주었다면, 해당 관계를 지식 그래프에 추가한다. 따라서, 지식 그래프에 2개의 삼중항 (u, r_{Like}, i) 및 $(i, r_{Positively Rated By}, u)$ 를 추가할 수 있다.

[0077] 만약 결측 평점이라면, 협업 필터링을 통해 예측한 평점이 특정한 임계치 이상일 경우, 해당 관계 역시 지식 그래프에 추가한다. 예측 평점은 실제 평점과 비교하면 정확도가 다소 떨어지므로 예측 평점에 대한 임계치는 기존 평점에 대한 임계치보다 높게 설정해야 한다.

[0078] 본 발명의 실시예에 따른 유사도 기반 관계 모델링은 사용자와 사용자, 그리고 아이템과 아이템 사이의 관계를 추가하기 위해 사전에 정의한 동종 개체 관계인 $r_{Is Similar User}$ 와 $r_{Is Similar Item}$ 을 고려한다. 만약 협업 필터링에 의해 특정한 두 사용자 u, v 의 유사도가 특정한 임계치 이상으로 예상 되면, $r_{Is Similar User}$ 관계를 지식 그래프에 추가한다. 마찬가지로, 특정한 두 아이템 i, j 의 유사도가 특정한 임계치 이상으로 예상되면, $r_{Is Similar Item}$ 관계를 지식 그래프에 추가한다. 따라서, 지식 그래프에 2개의 삼중항 $(u, r_{Is Similar User}, v)$ 및 $(i, r_{Is Similar Item}, j)$ 를 추가할 수 있다.

[0080] 도 7은 본 발명의 일 실시예에 따른 신규 개체의 관계 모델링을 통한 지식 그래프 확장 방법을 설명하기 위한 흐름도이다.

[0081] 도 7을 참조하면, 도 3의 과정을 통해 개체 유형을 고려한 관계 모델링을 수행하고, 도 5의 신규 관계 여부에 따른 관계 정의(550-570)를 활용하여 외부 지식 베이스 기반의 신규 개체 및 관계 추가에 따른 지식 그래프 확장 과정을 보여준다.

[0082] 본 발명의 실시예에 따른 관계 모델링은 사용자 또는 아이템의 개체 뿐만 아니라 지식 기반의 개체도 다룬다. 따라서, 온라인 데이터베이스(예를 들어, IMDB, DBpedia, Wikipedia)를 기반으로 지식 기반의 신규 개체와 새로운 관계를 정의할 수 있다. 따라서, 도 7에 도시된 본 발명의 일 실시예는 단일 지식 베이스 기반으로 지식 그래프를 생성하고(710), 외부 지식 베이스로부터 추출한 신규 개체가 추가된 경우(720)에도, 도 3에서 도시된 신

규 관계 모델링 방식으로 기존 개체 간, 신규 개체와 기존 개체간, 신규 개체 간의 관계를 정의한다(730). 미리 정해진 임계치를 기준으로 해당 관계를 추가할지 여부를 판단하고(740), 임계치 이상일 경우, 지식 그래프에 추가한다(760). 미리 정해진 임계치 미만일 경우 해당 관계는 누락한다(750). 이렇게 확장된 지식 그래프를 기반으로 학습하여 사용자의 아이템 선호도를 예측하도록 한다. 새로운 개체 유형이 추가되는 지식 그래프 확장 시나리오를 개체 유형 동일 여부에 따라 구성할 수 있으며, 해당 예시는 아래와 같다.

[0083] 본 발명의 일 실시예에 따라 영화 지식 베이스에 새로운 개체 유형인 카테고리 개체가 추가(720)되었고, 이 신규 개체들과 기존 지식 그래프(710)상의 개체 간 관계 모델링을 한다(730)고 가정한다. 이 경우, 신규 개체 유형인 카테고리에 대해 다음과 같은 두 개의 관계(즉, 동종 개체 관계, 이종 개체 관계)를 생각해볼 수 있다. 먼저, 이종 개체 간 관계의 관점에서, $\mathcal{E}_{category}$ 와 I 사이의 두 가지 관계, 즉 $r_{BelongsTo}$ 와 $r_{IsRelatedTo}$ 을 정의할 수 있다. 아이템의 특성 벡터를 기반으로 클러스터링(730)하여 임계치를 기준으로 가장 유사한 아이템이 속한 카테고리에 따라 해당 아이템 개체와 카테고리 개체의 신규 관계를 정의(740-760)한다. 다시 말해, 아이템 가 카테고리 $e \in \mathcal{E}_{category}$ 에 해당하는 경우, 지식 그래프에 2개의 삼중항 $(i, r_{BelongsTo}, e)$, $(e, r_{IsRelatedTo}, i)$ 를 추가할 수 있다는 뜻이다(760).

[0084] 본 발명의 일 실시예에 따라 동종 개체 간 관계의 관점에서 유사 카테고리 개체 간 관계인 $r_{IsSimilarCategory}$ 를 생각해볼 수 있다. 아이템 특성 벡터의 클러스터링 결과(730), 두 아이템에 해당하는 두 카테고리가 서로 유사한지 아닌지를 판단할 수 있고 만약 두 카테고리 사이의 유사도가 특정한 임계치를 넘는다면(740), 두 카테고리 개체 사이에 관계 $r_{IsSimilarCategory}$ (760)를 추가한다. 따라서 지식 그래프에 2개의 삼중항인 $(e_1, r_{IsSimilarCategory}, e_2)$ 와 $(e_2, r_{IsSimilarCategory}, e_1)$ 를 추가할 수 있다(760). 이러한 관계는 IMDB나 Wikipedia와 같은 온라인 데이터베이스에 정의되어 있지 않으므로 결과적으로 지식 그래프 상에 더 다양한 관계를 제공할 수 있다는 장점이 있다.

[0085] 위와 같이 확장한 지식 그래프를 학습하여 보다 풍부한 사용자 및 아이템의 특성을 학습하고 이를 통해 사용자의 선호도를 예측하여 사용자에게 제품을 추천하도록 한다.

[0087] 도 8은 본 발명의 일 실시예에 따른 신규 개체의 관계 모델링을 설명하기 위한 도면이다.

[0088] 본 발명의 실시예에 따른 관계 모델링은 사용자 또는 아이템의 개체뿐만 아니라 지식 기반의 개체도 다룬다. 따라서, 온라인 데이터베이스(예를 들어, IMDB, DBpedia, Wikipedia)를 기반으로 지식 기반의 신규 개체와 새로운 관계를 정의할 수 있다. 도 8의 시나리오는 기존 지식 그래프(810)에 신규 개체 e_4 와 e_5 가 추가된 경우에도, 신규 관계 모델링(830)으로 e_4 와 i_3 간의 관계 또는 e_5 와 i_2 간의 관계를 추가하여, 지식 그래프를 확장(840)하는 방법이다. 새로운 개체 유형이 추가되는 경우와 이미 존재하는 개체 유형 집합 내에 새로운 개체가 추가되는 경우로 나누어 시나리오를 구성할 수 있으며, 도 8의 시나리오 상에서 고려하는 신규 관계는 표 2와 같다.

[0089] <표 2>

Relation Type	Relation Name	Triples
Inter - entity relation	$r_{BelongsTo}$	$(i, r_{BelongsTo}, e)$
	$r_{IsRelatedTo}$	$(e, r_{IsRelatedTo}, i)$
Intra - entity relation	$r_{IsSimilarCategory}$	$(c_1, r_{IsSimilarCategory}, c_2)$

[0090]

[0091] 본 발명의 실시예에 따른 새로운 개체 유형의 관계 모델링은 영화 지식 베이스에 새로운 개체 유형인 카테고리 개체가 추가되었고, 이 신규 개체들과 기존 지식 그래프 상의 개체 간 관계 모델링을 한다고 가정한다. 이 경우, 신규 개체 유형인 카테고리에 대해 다음과 같은 두 개의 관계(즉, 동종 개체 관계, 이종 개체 관계)를 생각해볼 수 있다. 먼저, 이종 개체 간 관계의 관점에서, $\mathcal{E}_{Category}$ 와 I 사이의 두 가지 관계, 즉 $r_{BelongsTo}$ 와 $r_{IsRelatedTo}$ 을 정의할 수 있다. 아이템의 특성 벡터를 기반으로 클러스터링하여 가장 유사한 아이템이 속한 카테고리

고리에 따라 해당 아이템 개체와 카테고리 개체의 신규 관계를 정의한다. 다시 말해, 아이템 i 가 카테고리 $c \in \mathcal{E}_{Category}$ 에 해당하는 경우, 지식 그래프에 2 개의 삼중항 $(i, r_{Belongs To}, c)$ 와 $(c, r_{Is Related To}, i)$ 를 추가할 수 있다는 뜻이다. 두 번째, 동종 개체 간 관계의 관점에서 유사 카테고리 개체 간 관계인 $r_{Is Similar Category}$ 를 생각해볼 수 있다. 아이템 특성 벡터의 클러스터링 결과, 두 아이템에 해당하는 두 카테고리가 서로 유사한지 아닌지를 판단할 수 있고 만약 두 카테고리 사이의 유사도가 특정한 임계치를 넘는다면, 두 카테고리 개체 사이에 관계 $r_{Is Similar Category}$ 를 추가한다. 따라서 지식 그래프에 2개의 삼 중항인 $(c_1, r_{Is Similar Category}, c_2)$ 와 $(c_2, r_{Is Similar Category}, c_1)$ 를 추가할 수 있다. 이러한 관계는 IMDB나 Wikipedia와 같은 온라인 데이터베이스에 정의되어 있지 않으므로 결과적으로 지식 그래프 상에 더 다양한 관계를 제공할 수 있다는 장점이 있다.

[0092] 본 발명의 실시예에 따른 기존 개체 집합 내의 새로운 개체에 대한 관계 모델링은 새로운 개체 유형뿐만 아니라, 기존에 이미 존재하는 개체 유형을 가지는 신규 개체가 추가되었을 때에도 관계 모델링을 수행할 수 있다. 시간 $T-1$ 까지 존재하지 않았던 신규 개체 e_T 가 시간 T 에 추가되었다고 가정한다. 개체 e_T 의 개체 유형이 $a \in A$ 인 경우, 집합 $\mathcal{E}_a$ 에서 $\overline{\mathcal{E}_a} := \mathcal{E}_a \cup \{e_T\}$ 로 업데이트할 수 있다. 그런 다음, 집합 $\overline{\mathcal{E}_a}$ 과 함께 정의된 두 관계(즉, 동종 개체 관계, 이종 개체 관계)를 고려하여 개체 간의 새로운 관계를 연결한다. 다시 말해, 해당 개체는 유형이 이미 정해져 있고, 개체 유형이 해당하는 이종 개체 관계와 동종 개체 관계가 사전에 정의되어 있으므로, 구성된 관계를 기준으로 신규 개체와 다른 개체 사이의 관계를 모델링 할 수 있다. 이는 본 발명의 실시예에 따른 방식이 실시간으로 유입되는 동적 데이터를 고려하여 지식 그래프를 확장하는 것 또한 가능성을 보여준다.

[0094] 도 9는 본 발명의 일 실시예에 따른 다중 지식 베이스 기반 이종 데이터의 통합 관계 모델링을 통한 지식 그래프 확장 방법을 설명하기 위한 흐름도이다.

[0095] 도 9를 참조하면, 도 3의 과정을 통해 개체 유형을 고려한 관계 모델링을 수행하고, 도 5와 도 7의 흐름을 기반으로 여러 개의 외부 지식 베이스를 통합하는 지식 그래프 확장 과정을 설명한다.

[0096] 도 9에 도시된 바와 같이, A 데이터 셋(910)과 IMDB, DBpedia, Wikipedia와 같은 외부 지식 베이스(920) 각각을 기반으로 생성된 두 개의 지식 그래프가 존재한다고 가정한다. 이때, 도 5와 도 7을 참조하여, 두 지식 그래프 상의 개체 간 관계 모델링을 통해 각 지식 그래프에 신규 개체 및 관계를 추가(930, 940)하고 두 개의 지식 그래프를 개별적으로 확장할 수 있다.

[0097] 다음으로, 확장된 두 지식 그래프는 도 7을 참조하여 개체 간 관계 모델링을 활용하여 통합된 지식 그래프를 생성한다(950). 이를 통해 A 데이터 셋은 기존의 지식 그래프에 존재하지 않았던 새로운 개체와 신규 관계 정보를 갖는 확장된 지식 그래프를 얻을 수 있다(960).

[0098] 최종적으로 A 데이터 셋은 외부 지식 베이스의 이종 데이터 통합으로 다양한 보조 정보를 반영한 지식 그래프를 얻을 수 있으므로 이로 인해 아이템 특성 벡터의 품질이 개선되고 추천 정확도가 향상된다. 이처럼 개체 간 관계 모델링을 통한 지식 그래프 확장 방식은 다중 지식 베이스 상의 이종 데이터 통합, 즉 상호 운용성을 달성한다.

[0100] 도 10은 본 발명의 일 실시예에 따른 다중 지식 베이스 기반 이종 데이터의 통합 관계 모델링을 설명하기 위한 도면이다.

[0101] 본 발명의 실시예에 따른 여러 개의 외부 지식 베이스를 통합하여 이종 데이터 간의 상호 호환이 가능한 지식 그래프 확장 예시를 설명한다. 도 10의 시나리오는 A 데이터 셋과 IMDB, DBpedia, Wikipedia와 같은 외부 지식 베이스를 기반으로 구축한 두 개의 지식 그래프(10101, 1012)가 존재한다고 가정한다. 이때, 두 지식 그래프를 각각을 활용한 개체 간 관계 모델링(1021, 1022)을 통해 e_3 와 i_5 간의 관계 또는 e_1 와 e_2 간의 관계와 같은 신규 관계를 추가하고 지식 그래프를 확장할 수 있다. 두 지식 그래프는 앞서 소개한 두 시나리오상의 관계 모델링을 활용하여 통합된 지식 그래프를 생성한다(1030). 이를 통해 A 데이터 셋은 기존의 지식 그래프에 존재하지

않았던 새로운 개체와 신규 관계 정보를 갖는 확장된 지식 그래프를 얻을 수 있다(1040). 이중 데이터의 통합을 위해 일반적으로 두 개 이상의 외부 데이터 집합을 고려해야 하지만 도 10의 시나리오에서는 간단하게 하나의 데이터 셋과 해당 데이터 셋과 관련된 하나의 외부 지식 베이스만 있는 상황을 고려한다. A 데이터 셋과 외부 지식 베이스는 도 6의 시나리오와 도 8의 시나리오를 통해 각각의 지식 그래프에 신규 관계를 추가한다. 이렇게 확장된 두 지식 그래프를 도 8의 시나리오의 방법으로 합치면 단독 데이터 셋에서는 유도할 수 없는 새로운 개체와 신규 관계를 추가할 수 있다. 최종적으로 A 데이터 셋은 외부 지식 베이스의 이중 데이터 통합으로 다양한 보조 정보를 반영한 지식 그래프를 얻을 수 있으므로 이로 인해 아이템 특성 벡터의 품질이 개선되고 추천 정확도가 향상된다. 이처럼 개체 간 관계 모델링을 통한 지식 그래프 확장 방식은 다중 지식 베이스 상의 이중 데이터 통합을 달성한다.

[0102] 이와 같이, 본 발명의 일실시예에 따르면, 다양한 이중 데이터를 지식 그래프상에 통합하여 확장함으로써 지식 그래프의 불완전성 문제를 해결하고 추천 시스템의 데이터 부족 문제를 해결할 수 있다.

[0103] 또한, 본 발명으로 명시적 관계뿐만 아니라 신뢰성 높은 잠재 관계를 지식 그래프에 반영하여 풍부한 관계 표현이 가능하고 사용자 및 아이템 특성의 품질을 개선시킬 수 있다.

[0104] 본 발명은 확장된 지식 그래프의 품질 향상으로 인해 추천 시스템의 성능 또한 개선시킴을 확인할 수 있다.

[0105] 본 발명의 일실시예에 따라 확장된 지식 그래프를 활용한 임베딩 기반 학습 방법은 사용자와 아이템의 특성 벡터를 더욱 풍부하게 추출할 수 있으며, 이를 통해 추천 시스템의 선호도 예측 성능을 향상시킬 수 있다. 또한, 관계 모델링을 통해 추가된 정점(개체)과 간선(관계)도 고정된 임베딩 공간에서 임베딩을 진행하므로, 확장된 지식 그래프는 KGCN, KGAT, MKGAT 등 기존 임베딩 기반 학습 방법에 모두 적용할 수 있다.

[0107] 이상에서 설명된 장치는 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성요소는, 예를 들어, 프로세서, 콘트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPGA(field programmable gate array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

[0108] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치에 구체화(embodiment)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

[0109] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를

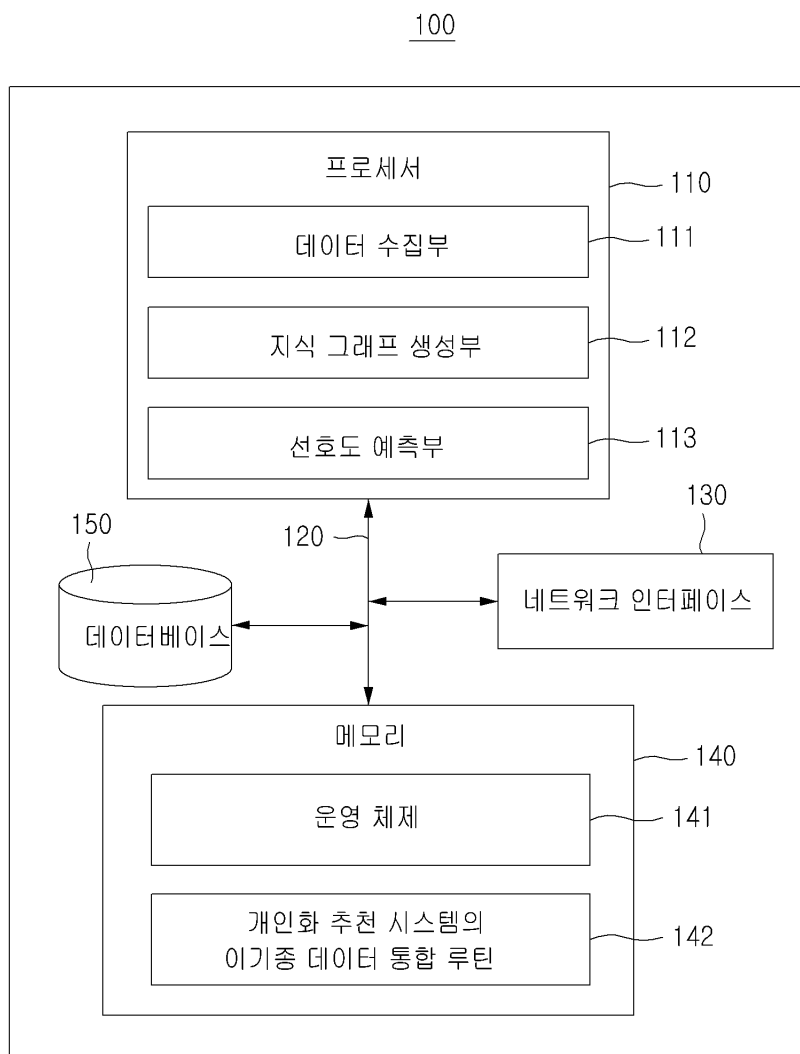
포함한다.

[0110] 이상과 같이 실시예들이 비록 한정된 실시예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

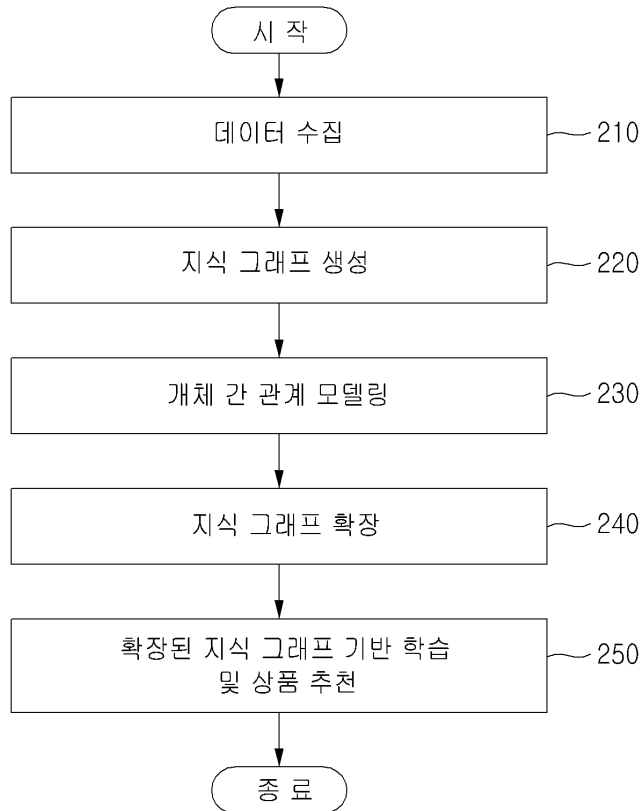
[0111] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

도면

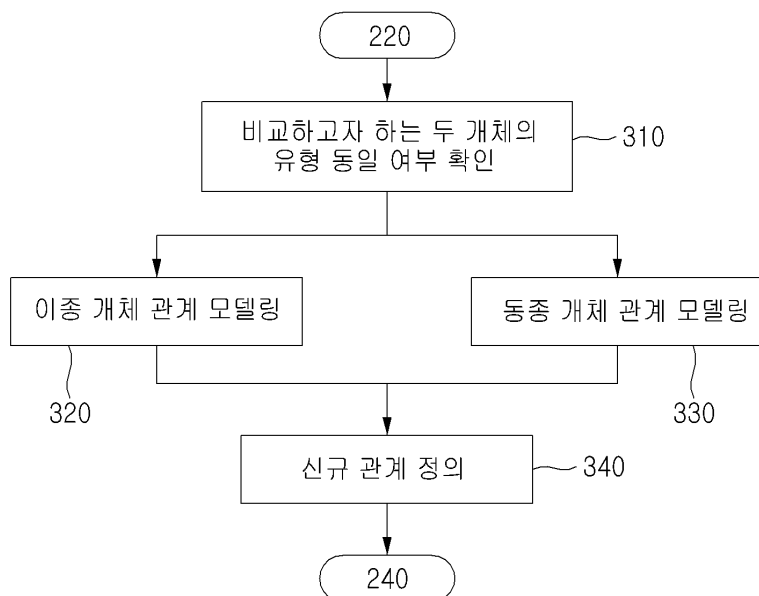
도면1



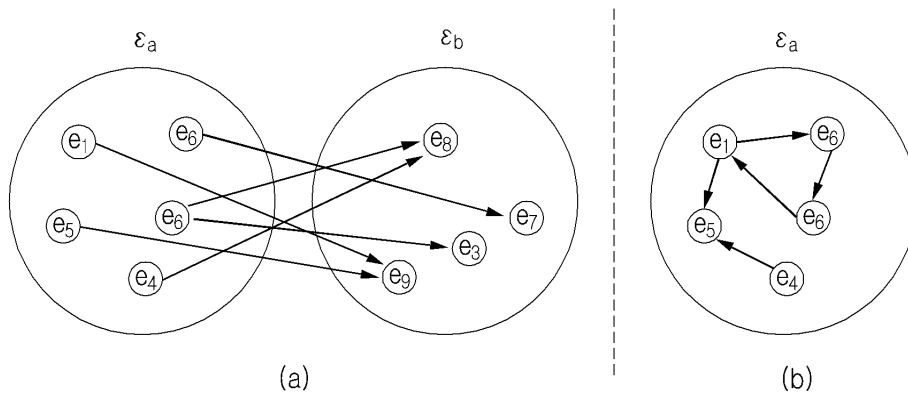
도면2



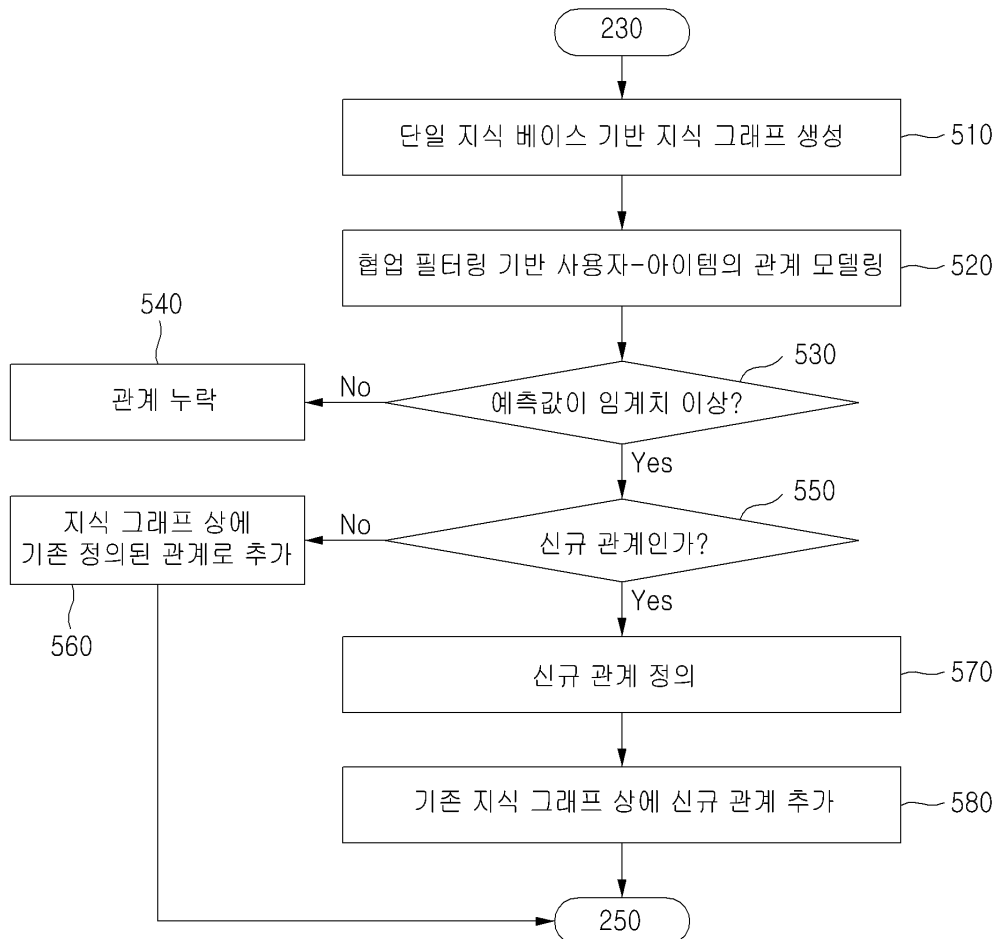
도면3



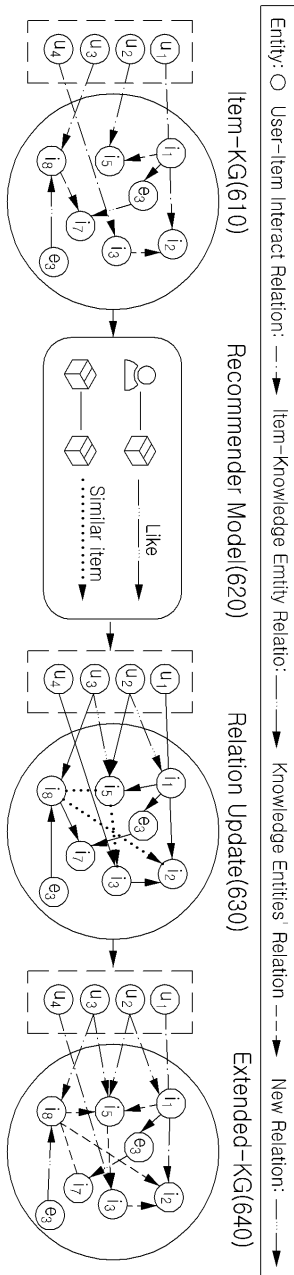
도면4



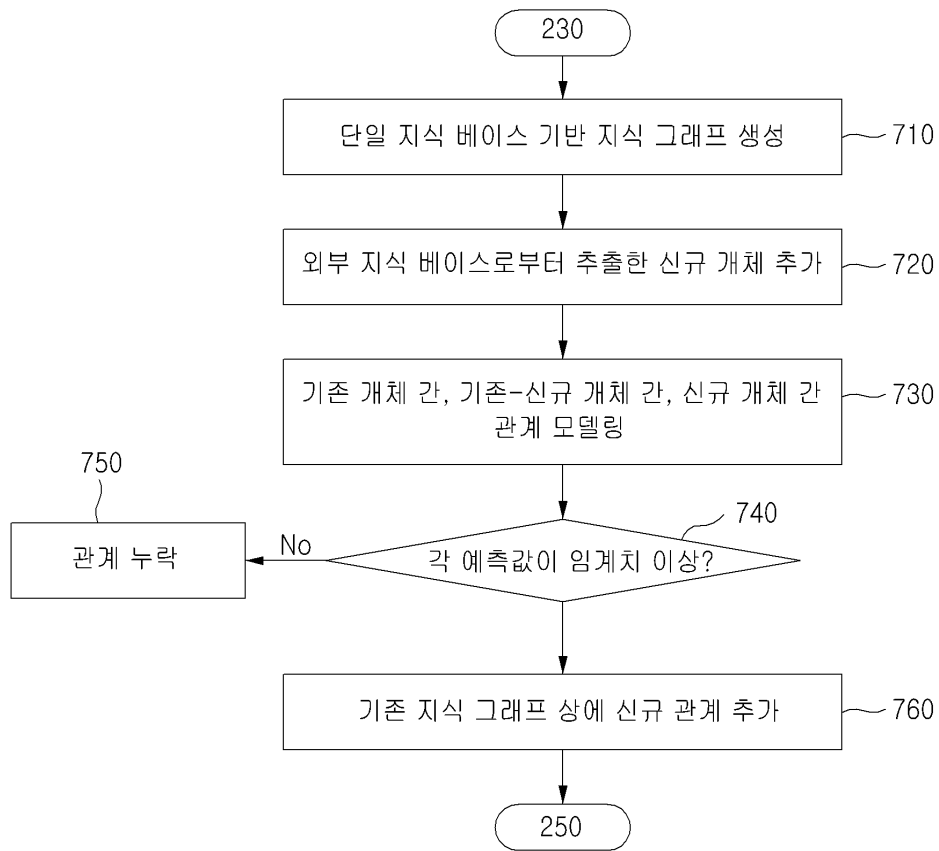
도면5



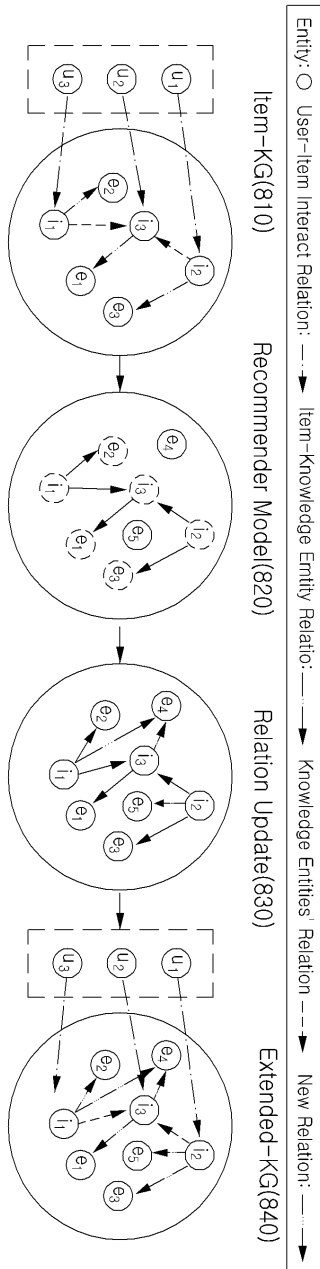
도면6



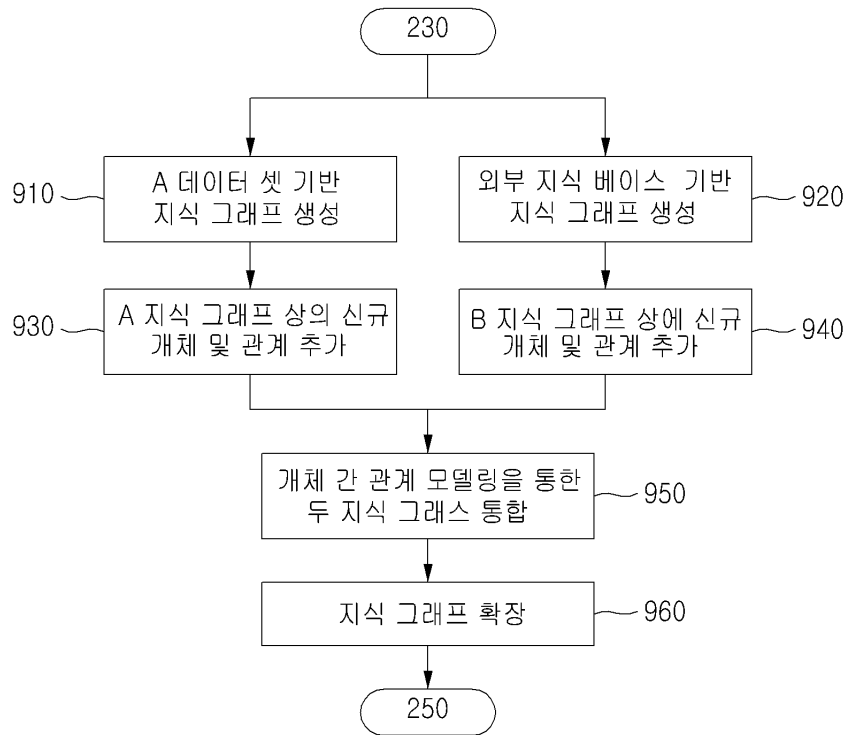
도면7



도면8



도면9



도면10

