(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0126299
(43) 공개일자 2024년08월20일

(51) 국제특허분류(Int. Cl.)
G06N 3/0895 (2023.01) *G06N 3/08* (2023.01)
G06Q 30/06 (2023.01)
(52) CPC특허분류
G06N 3/0895 (2023.01)
G06N 3/082 (2023.01)
(21) 출원번호 10-2023-0018908
(22) 출원일자 2023년02월13일
심사청구일자 2023년02월13일

(71) 출원인
네이버 주식회사
경기도 성남시 분당구 정자일로 95(정자동, 네이버 1784)
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
이정태
경기도 성남시 분당구 정자일로 95(정자동)
손대능
경기도 성남시 분당구 정자일로 95(정자동)
(뒷면에 계속)
(74) 대리인
양성보

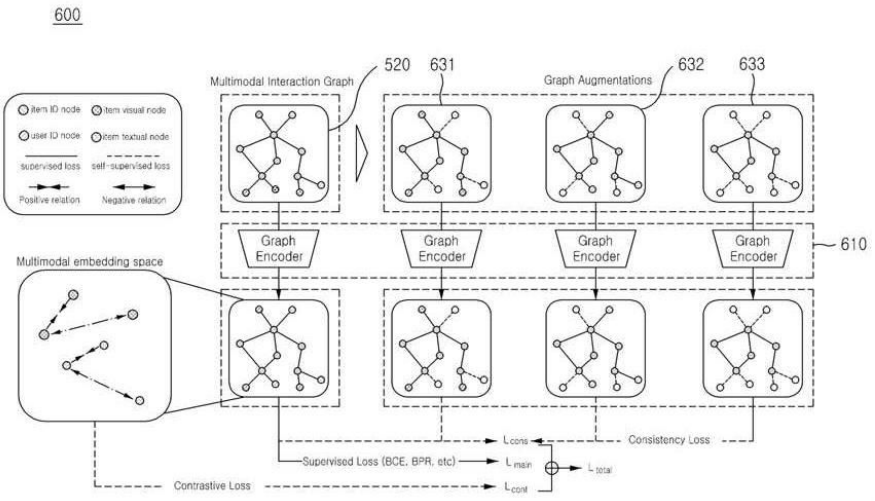
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 **그래프 기반 아이템 인기도 예측을 위한 자가지도 학습 방법, 컴퓨터 장치, 및 컴퓨터 프로그램**

(57) 요 약

그래프 기반 아이템 인기도 예측을 위한 자가지도 학습 방법, 컴퓨터 장치, 및 컴퓨터 프로그램이 개시된다. 일실시예에서는 사용자와 아이템의 관계를 나타내는 그래프에 각 아이템의 멀티모달 정보를 추가하여 멀티모달 그래프를 생성하고, 상기 멀티모달 그래프를 이용하여 각 노드 유형에 대한 임베딩을 수행할 수 있다.

대 표 도



(52) CPC특허분류
G06Q 30/0631 (2013.01)
(72) 발명자
김상필
서울특별시 성북구 안암로 145, 우정정보관 606호 (안암동5가)
김성준
서울특별시 동대문구 왕산로 86, 102동 1517호(용두동, 동대문 푸르지오시티)
윤성준
서울특별시 성북구 개운사길 83-2(안암동5가)
박소윤
경기도 성남시 분당구 정자일로 95(정자동)

명 세 서

청구범위

청구항 1

컴퓨터 장치에서 실행되는 방법에 있어서,

상기 컴퓨터 장치는 메모리에 포함된 컴퓨터 판독가능한 명령들을 실행하도록 구성된 적어도 하나의 프로세서를 포함하고,

상기 방법은,

상기 적어도 하나의 프로세서에 의해, 사용자와 아이템의 관계를 나타내는 그래프에 각 아이템의 멀티모달 정보를 추가하여 멀티모달 그래프를 생성하는 단계; 및

상기 적어도 하나의 프로세서에 의해, 상기 멀티모달 그래프를 이용하여 각 노드 유형에 대한 임베딩을 수행하는 단계

를 포함하는 방법.

청구항 2

제1항에 있어서,

상기 생성하는 단계는,

상기 아이템의 이미지, 텍스트, 오디오 중 적어도 하나의 멀티모달 정보를 별개의 노드 유형으로 추가하여 해당 아이템 노드와 연결하는 단계

를 포함하는 방법.

청구항 3

제1항에 있어서,

상기 생성하는 단계는,

상기 멀티모달 정보에 대해 사전 학습 모델을 이용하여 멀티모달 피처를 추출한 후 상기 멀티모달 피처를 상기 아이템과 같은 임베딩 공간에 매핑함으로써 학습 가능한 파라미터로 설정하는 단계

를 포함하는 방법.

청구항 4

제1항에 있어서,

상기 수행하는 단계는,

상기 멀티모달 그래프에서 이웃 노드 간의 메시지 전파(message passing)를 기초로 각 노드의 피처 표현(feature representation)을 생성하는 단계

를 포함하는 방법.

청구항 5

제2항에 있어서,

상기 수행하는 단계는,

아이템 노드에서 사용자 노드로의 메시지 전파에 따른 협업 신호를 인코딩하여 사용자 임베딩을 생성하는 단계;

멀티모달 노드와 사용자 노드에서 아이템 노드로의 메시지 전파에 따른 협업 신호를 인코딩하여 아이템 임베딩

을 생성하는 단계; 및

아이템 노드에서 멀티모달 노드로의 메시지 전파에 따른 협업 신호를 인코딩하여 멀티모달 임베딩을 생성하는 단계

를 포함하는 방법.

청구항 6

제5항에 있어서,

상기 사용자 임베딩과 상기 아이템 임베딩은 상기 아이템에 대한 상기 사용자의 인기도를 나타내는 친밀도 점수(affinity score)를 정량화하기 위해 사용되는 것

을 특징으로 하는 방법.

청구항 7

제1항에 있어서,

상기 수행하는 단계는,

상기 멀티모달 정보를 이용하여 대조 학습(contrastive learning)을 수행하는 단계

를 포함하고,

상기 대조 학습을 수행하는 단계는,

상기 멀티모달 그래프 상의 사용자와 아이템의 관계를 기초로 대조 학습 쌍을 구성하는 단계

를 포함하는 방법.

청구항 8

제7항에 있어서,

상기 대조 학습 쌍을 구성하는 단계는,

사용자와 인터랙션 이력이 있는 아이템 집합에서 샘플링된 아이템의 멀티모달 정보를 포지티브 샘플로 선정하는 단계; 및

상기 아이템 집합에 대한 여집합에서 샘플링된 아이템의 멀티모달 정보를 네거티브 샘플로 선정하는 단계

를 포함하는 방법.

청구항 9

제1항에 있어서,

상기 수행하는 단계는,

상기 멀티모달 그래프와 상기 멀티모달 그래프로부터 증강된 그래프를 이용하여 상기 멀티모달 그래프에 대한 일관성 정규화(consistency regularization)를 수행하는 단계

를 포함하는 방법.

청구항 10

제9항에 있어서,

상기 일관성 정규화를 수행하는 단계는,

모달리티 드롭아웃(modality dropout)을 통해 상기 멀티모달 그래프에서 적어도 하나의 노드 유형을 제거하여 증강 그래프를 생성하는 단계; 및

상기 멀티모달 그래프와 상기 증강 그래프를 이용한 모델 예측 결과를 정규화하는 단계

를 포함하는 방법.

청구항 11

제1항 내지 제10항 중 어느 한 항의 방법을 컴퓨터에 실행시키기 위해 컴퓨터 판독가능한 기록 매체에 저장되는 컴퓨터 프로그램.

청구항 12

컴퓨터 장치에 있어서,

메모리에 포함된 컴퓨터 판독가능한 명령들을 실행하도록 구성된 적어도 하나의 프로세서

를 포함하고,

상기 적어도 하나의 프로세서는,

사용자와 아이템의 관계를 나타내는 그래프에 각 아이템의 멀티모달 정보를 추가하여 멀티모달 그래프를 생성하는 과정; 및

상기 멀티모달 그래프를 이용하여 각 노드 유형에 대한 임베딩을 수행하는 과정

을 처리하는 컴퓨터 장치.

청구항 13

제12항에 있어서,

상기 적어도 하나의 프로세서는,

상기 아이템의 이미지, 텍스트, 오디오 중 적어도 하나의 멀티모달 정보를 별개의 노드 유형으로 추가하여 해당 아이템 노드와 연결하는 것

을 특징으로 하는 컴퓨터 장치.

청구항 14

제12항에 있어서,

상기 적어도 하나의 프로세서는,

상기 멀티모달 정보에 대해 사전 학습 모델을 이용하여 멀티모달 피처를 추출한 후 상기 멀티모달 피처를 상기 아이템과 같은 임베딩 공간에 매핑함으로써 학습 가능한 파라미터로 설정하는 것

을 특징으로 하는 컴퓨터 장치.

청구항 15

제12항에 있어서,

상기 적어도 하나의 프로세서는,

상기 멀티모달 그래프에서 이웃 노드 간의 메시지 전파를 기초로 각 노드의 피처 표현을 생성하는 것

을 특징으로 하는 컴퓨터 장치.

청구항 16

제13항에 있어서,

상기 적어도 하나의 프로세서는,

아이템 노드에서 사용자 노드로의 메시지 전파에 따른 협업 신호를 인코딩하여 사용자 임베딩을 생성하고,

멀티모달 노드와 사용자 노드에서 아이템 노드로의 메시지 전파에 따른 협업 신호를 인코딩하여 아이템 임베딩을 생성하고,

아이템 노드에서 멀티모달 노드로의 메시지 전파에 따른 협업 신호를 인코딩하여 멀티모달 임베딩을 생성하는 것

을 특징으로 하는 컴퓨터 장치.

청구항 17

제12항에 있어서,

상기 적어도 하나의 프로세서는,

상기 멀티모달 정보를 이용하여 대조 학습을 수행하되,

상기 멀티모달 그래프 상의 사용자와 아이템의 관계를 기초로 대조 학습 쌍을 구성하는 것

을 특징으로 하는 컴퓨터 장치.

청구항 18

제17항에 있어서,

상기 적어도 하나의 프로세서는,

사용자와 인터랙션 이력이 있는 아이템 집합에서 샘플링된 아이템의 멀티모달 정보를 포지티브 샘플로 선정하고,

상기 아이템 집합에 대한 여집합에서 샘플링된 아이템의 멀티모달 정보를 네거티브 샘플로 선정하는 것

을 특징으로 하는 컴퓨터 장치.

청구항 19

제12항에 있어서,

상기 적어도 하나의 프로세서는,

상기 멀티모달 그래프와 상기 멀티모달 그래프로부터 증강된 그래프를 이용하여 상기 멀티모달 그래프에 대한 일관성 정규화를 수행하는 것

을 특징으로 하는 컴퓨터 장치.

청구항 20

제19항에 있어서,

상기 적어도 하나의 프로세서는,

모달리티 드롭아웃을 통해 상기 멀티모달 그래프에서 적어도 하나의 노드 유형을 제거하여 증강 그래프를 생성하고,

상기 멀티모달 그래프와 상기 증강 그래프를 이용한 모델 예측 결과를 정규화하는 것

을 특징으로 하는 컴퓨터 장치.

발명의 설명

기 술 분 야

[0001] 아래의 설명은 아이템에 대한 멀티모달 표현(multimodal representation)을 학습하는 기술에 관한 것이다.

배 경 기 술

[0002] 추천 시스템은 전자 상거래, 광고, 소셜 미디어 등 각종 온라인 서비스에서 사용자 경험 향상에 중요한 역할을 한다.

[0003] 일반적으로 사용자가 인터넷 상의 아이템을 선택하면 선택된 아이템과 유사한 아이템을 추천할 수 있다.

[0004] 아이템 추천을 위해 주로 사용되는 방법은 크게 사용자들 간의 유사도를 활용한 협업 필터링(collaborative filtering) 방식, 그리고 아이템 및 사용자의 특성 값을 활용하는 내용 기반 필터링(content-based filtering) 방식 등으로 구분될 수 있다.

[0005] 예컨대, 한국 공개특허공보 제10-2021-0131198호(공개일 2021년 11월 02일)에는 인공 신경망을 통해 추출된 이미지를 기반으로 유사 상품을 추천하는 기술이 개시되어 있다.

발명의 내용

해결하려는 과제

[0006] 아이템에 대한 인기도 예측을 위한 멀티모달 융합 인공지능 모델을 생성할 수 있다.

[0007] 사용자와 아이템의 관계뿐만 아니라 각 아이템의 멀티모달 정보를 포함한 피처 표현(feature representation)을 통해 융합 멀티모달 그래프를 구성할 수 있다.

[0008] 멀티모달 정보에 기반하여 사용자와 아이템의 관계성을 강화할 수 있는 자가지도 학습을 제공할 수 있다.

[0009] 자가지도 학습의 일환으로 멀티모달 정보를 이용한 대조 학습(contrastive learning)과 일관성 정규화(consistency regularization) 학습을 수행할 수 있다.

과제의 해결 수단

[0010] 컴퓨터 장치에서 실행되는 방법에 있어서, 상기 컴퓨터 장치는 메모리에 포함된 컴퓨터 판독가능한 명령들을 실행하도록 구성된 적어도 하나의 프로세서를 포함하고, 상기 방법은, 상기 적어도 하나의 프로세서에 의해, 사용자와 아이템의 관계를 나타내는 그래프에 각 아이템의 멀티모달 정보를 추가하여 멀티모달 그래프를 생성하는 단계; 및 상기 적어도 하나의 프로세서에 의해, 상기 멀티모달 그래프를 이용하여 각 노드 유형에 대한 임베딩을 수행하는 단계를 포함하는 방법을 제공한다.

[0011] 일 측면에 따르면, 상기 생성하는 단계는, 상기 아이템의 이미지, 텍스트, 오디오 중 적어도 하나의 멀티모달 정보를 별개의 노드 유형으로 추가하여 해당 아이템 노드와 연결하는 단계를 포함할 수 있다.

[0012] 다른 측면에 따르면, 상기 생성하는 단계는, 상기 멀티모달 정보에 대해 사전 학습 모델을 이용하여 멀티모달 피처를 추출한 후 상기 멀티모달 피처를 상기 아이템과 같은 임베딩 공간에 매핑함으로써 학습 가능한 파라미터로 설정하는 단계를 포함할 수 있다.

[0013] 또 다른 측면에 따르면, 상기 수행하는 단계는, 상기 멀티모달 그래프에서 이웃 노드 간의 메시지 전파(message passing)를 기초로 각 노드의 피처 표현(feature representation)을 생성하는 단계를 포함할 수 있다.

[0014] 또 다른 측면에 따르면, 상기 수행하는 단계는, 아이템 노드에서 사용자 노드로의 메시지 전파에 따른 협업 신호를 인코딩하여 사용자 임베딩을 생성하는 단계; 멀티모달 노드와 사용자 노드에서 아이템 노드로의 메시지 전파에 따른 협업 신호를 인코딩하여 아이템 임베딩을 생성하는 단계; 및 아이템 노드에서 멀티모달 노드로의 메시지 전파에 따른 협업 신호를 인코딩하여 멀티모달 임베딩을 생성하는 단계를 포함할 수 있다.

[0015] 또 다른 측면에 따르면, 상기 사용자 임베딩과 상기 아이템 임베딩은 상기 아이템에 대한 상기 사용자의 인기도를 나타내는 친밀도 점수(affinity score)를 정량화하기 위해 사용될 수 있다.

[0016] 또 다른 측면에 따르면, 상기 수행하는 단계는, 상기 멀티모달 정보를 이용하여 대조 학습(contrastive learning)을 수행하는 단계를 포함하고, 상기 대조 학습을 수행하는 단계는, 상기 멀티모달 그래프 상의 사용자와 아이템의 관계를 기초로 대조 학습 쌍을 구성하는 단계를 포함할 수 있다.

[0017] 또 다른 측면에 따르면, 상기 대조 학습 쌍을 구성하는 단계는, 사용자와 인터랙션 이력이 있는 아이템 집합에서 샘플링된 아이템의 멀티모달 정보를 포지티브 샘플로 선정하는 단계; 및 상기 아이템 집합에 대한 여집합에서 샘플링된 아이템의 멀티모달 정보를 네거티브 샘플로 선정하는 단계를 포함할 수 있다.

[0018] 또 다른 측면에 따르면, 상기 수행하는 단계는, 상기 멀티모달 그래프와 상기 멀티모달 그래프로부터 증강된 그래프를 이용하여 상기 멀티모달 그래프에 대한 일관성 정규화(consistency regularization)를 수행하는 단계를 포함할 수 있다.

[0019] 또 다른 측면에 따르면, 상기 일관성 정규화를 수행하는 단계는, 모달리티 드롭아웃(modality dropout)을 통해 상기 멀티모달 그래프에서 적어도 하나의 노드 유형을 제거하여 증강 그래프를 생성하는 단계; 및 상기 멀티모달 그래프와 상기 증강 그래프를 이용한 모델 예측 결과를 정규화하는 단계를 포함할 수 있다.

[0020] 상기 방법을 컴퓨터에 실행시키기 위해 컴퓨터 판독가능한 기록 매체에 저장되는 컴퓨터 프로그램을 제공한다.

[0021] 컴퓨터 장치에 있어서, 메모리에 포함된 컴퓨터 판독가능한 명령들을 실행하도록 구성된 적어도 하나의 프로세서를 포함하고, 상기 적어도 하나의 프로세서는, 사용자와 아이템의 관계를 나타내는 그래프에 각 아이템의 멀티모달 정보를 추가하여 멀티모달 그래프를 생성하는 과정; 및 상기 멀티모달 그래프를 이용하여 각 노드 유형에 대한 임베딩을 수행하는 과정을 처리하는 컴퓨터 장치를 제공한다.

발명의 효과

[0022] 본 발명의 실시예들에 따르면, 사용자와 아이템의 관계뿐만 아니라 각 아이템의 멀티모달 정보를 포함한 피처 표현을 통해 모달리티 별 협업 신호를 보다 효과적으로 인코딩할 수 있다.

[0023] 본 발명의 실시예들에 따르면, 사용자와 아이템의 관계성을 기반으로 멀티모달 정보에 대한 대조 학습 쌍을 구성함으로써 더욱 정확하고 풍부한 메시지 전파를 통해 사용자와 아이템의 관계성을 강화할 수 있다.

[0024] 본 발명의 실시예들에 따르면, 융합 멀티모달 그래프에 대한 일관성 정규화를 통해 특정 피처나 모달리티에 대한 민감도나 의존도를 완화할 수 있고 피처 임베딩 간의 불필요한 편차(variance) 문제를 해결할 수 있다.

도면의 간단한 설명

[0025] 도 1은 본 발명의 일실시예에 따른 네트워크 환경의 예를 도시한 도면이다.

도 2는 본 발명의 일실시예에 따른 컴퓨터 장치의 예를 도시한 블록도이다.

도 3은 본 발명의 일실시예에 따른 컴퓨터 장치가 수행할 수 있는 방법의 예를 도시한 순서도이다.

도 4 내지 도 5는 본 발명의 일실시예에 있어서 그래프 임베딩 초기화를 설명하기 위한 예시 도면이다.

도 6은 본 발명의 일실시예에 있어서 SIMGNN 모델의 전반적인 아키텍처를 도시한 것이다.

도 7은 본 발명의 일실시예에 있어서 융합 멀티모달 그래프에 대한 임베딩을 설명하기 위한 예시 도면이다.

도 8은 본 발명의 일실시예에 있어서 아이템에 대한 사용자 친밀도를 산출하는 과정을 설명하기 위한 예시 도면이다.

도 9는 본 발명의 일실시예에 있어서 멀티모달 정보를 이용한 대조 학습 과정을 설명하기 위한 예시 도면이다.

도 10은 본 발명의 일실시예에 있어서 멀티모달 정보를 이용한 일관성 정규화 학습 과정을 설명하기 위한 예시 도면이다.

도 11은 본 발명의 일실시예에 있어서 SIMGNN 모델의 학습 알고리즘 예시를 도시한 것이다.

발명을 실시하기 위한 구체적인 내용

[0026] 이하, 본 발명의 실시예를 첨부된 도면을 참조하여 상세하게 설명한다.

[0027] 본 발명의 실시예들은 아이템에 대한 멀티모달 표현을 학습하는 기술에 관한 것이다.

[0028] 본 명세서에서 구체적으로 개시되는 것들을 포함하는 실시예들은 아이템 인기도 예측을 위한 멀티모달 융합 인공지능 모델을 제공하는 것으로, 이미지, 텍스트, 오디오 등의 멀티모달 정보를 사용자와 아이템 간의 관계성과 융합한 피처 표현을 학습할 수 있다.

[0029] 본 발명의 실시예들에 따른 멀티모달 임베딩 시스템은 적어도 하나의 컴퓨터 장치에 의해 구현될 수 있으며, 본 발명의 실시예들에 따른 멀티모달 임베딩 방법은 멀티모달 임베딩 시스템에 포함되는 적어도 하나의 컴퓨터 장치를 통해 수행될 수 있다. 이때, 컴퓨터 장치에는 본 발명의 일실시예에 따른 컴퓨터 프로그램이 설치 및 구동될 수 있고, 컴퓨터 장치는 구동된 컴퓨터 프로그램의 제어에 따라 본 발명의 실시예들에 따른 멀티모달 임베딩 방법을 수행할 수 있다. 상술한 컴퓨터 프로그램은 컴퓨터 장치와 결합되어 멀티모달 임베딩 방법을 컴퓨터에 실행시키기 위해 컴퓨터 판독 가능한 기록매체에 저장될 수 있다.

[0030] 도 1은 본 발명의 일실시예에 따른 네트워크 환경의 예를 도시한 도면이다. 도 1의 네트워크 환경은 복수의 전자 기기들(110, 120, 130, 140), 복수의 서버들(150, 160) 및 네트워크(170)를 포함하는 예를 나타내고 있다. 이러한 도 1은 발명의 설명을 위한 일례로 전자 기기의 수나 서버의 수가 도 1과 같이 한정되는 것은 아니다. 또한, 도 1의 네트워크 환경은 본 실시예들에 적용 가능한 환경들 중 하나의 예를 설명하는 것일 뿐, 본 실시예들에 적용 가능한 환경이 도 1의 네트워크 환경으로 한정되는 것은 아니다.

[0031] 복수의 전자 기기들(110, 120, 130, 140)은 컴퓨터 장치로 구현되는 고정형 단말이거나 이동형 단말일 수 있다. 복수의 전자 기기들(110, 120, 130, 140)의 예를 들면, 스마트폰(smart phone), 휴대폰, 내비게이션, 컴퓨터, 노트북, 디지털방송용 단말, PDA(Personal Digital Assistants), PMP(Portable Multimedia Player), 태블릿 PC 등이 있다. 일례로 도 1에서는 전자 기기(110)의 예로 스마트폰의 형상을 나타내고 있으나, 본 발명의 실시예들에서 전자 기기(110)는 실질적으로 무선 또는 유선 통신 방식을 이용하여 네트워크(170)를 통해 다른 전자 기기들(120, 130, 140) 및/또는 서버(150, 160)와 통신할 수 있는 다양한 물리적인 컴퓨터 장치들 중 하나를 의미할 수 있다.

[0032] 통신 방식은 제한되지 않으며, 네트워크(170)가 포함할 수 있는 통신망(일례로, 이동통신망, 유선 인터넷, 무선 인터넷, 방송망)을 활용하는 통신 방식뿐만 아니라 기기들간의 근거리 무선 통신 역시 포함될 수 있다. 예를 들어, 네트워크(170)는, PAN(personal area network), LAN(local area network), CAN(campus area network), MAN(metropolitan area network), WAN(wide area network), BBN(broadband network), 인터넷 등의 네트워크 중 하나 이상의 임의의 네트워크를 포함할 수 있다. 또한, 네트워크(170)는 버스 네트워크, 스타 네트워크, 링 네트워크, 메쉬 네트워크, 스타-버스 네트워크, 트리 또는 계층적(hierarchical) 네트워크 등을 포함하는 네트워크 토폴로지 중 임의의 하나 이상을 포함할 수 있으나, 이에 제한되지 않는다.

[0033] 서버(150, 160) 각각은 복수의 전자 기기들(110, 120, 130, 140)과 네트워크(170)를 통해 통신하여 명령, 코드, 파일, 컨텐츠, 서비스 등을 제공하는 컴퓨터 장치 또는 복수의 컴퓨터 장치들로 구현될 수 있다. 예를 들어, 서버(150)는 네트워크(170)를 통해 접속한 복수의 전자 기기들(110, 120, 130, 140)로 서비스(일례로, 아이템 추천 서비스, 아이템 검색 서비스, 아이템 쇼핑 서비스 등)를 제공하는 시스템일 수 있다.

[0034] 도 2는 본 발명의 일실시예에 따른 컴퓨터 장치의 예를 도시한 블록도이다. 앞서 설명한 복수의 전자 기기들(110, 120, 130, 140) 각각이나 서버들(150, 160) 각각은 도 2를 통해 도시된 컴퓨터 장치(200)에 의해 구현될 수 있다.

[0035] 이러한 컴퓨터 장치(200)는 도 2에 도시된 바와 같이, 메모리(210), 프로세서(220), 통신 인터페이스(230) 그리고 입출력 인터페이스(240)를 포함할 수 있다. 메모리(210)는 컴퓨터에서 판독 가능한 기록매체로서, RAM(random access memory), ROM(read only memory) 및 디스크 드라이브와 같은 비소멸성 대용량 기록장치(permanent mass storage device)를 포함할 수 있다. 여기서 ROM과 디스크 드라이브와 같은 비소멸성 대용량 기록장치는 메모리(210)와는 구분되는 별도의 영구 저장 장치로서 컴퓨터 장치(200)에 포함될 수도 있다. 또한, 메모리(210)에는 운영체제와 적어도 하나의 프로그램 코드가 저장될 수 있다. 이러한 소프트웨어 구성요소들은 메모리(210)와는 별도의 컴퓨터에서 판독 가능한 기록매체로부터 메모리(210)로 로딩될 수 있다. 이러한 별도의 컴퓨터에서 판독 가능한 기록매체는 플로피 드라이브, 디스크, 테이프, DVD/CD-ROM 드라이브, 메모리 카드 등의 컴퓨터에서 판독 가능한 기록매체를 포함할 수 있다. 다른 실시예에서 소프트웨어 구성요소들은 컴퓨터에서 판독 가능한 기록매체가 아닌 통신 인터페이스(230)를 통해 메모리(210)에 로딩될 수도 있다. 예를 들어, 소프트웨어 구성요소들은 네트워크(170)를 통해 수신되는 파일들에 의해 설치되는 컴퓨터 프로그램에 기반하여 컴퓨터 장치(200)의 메모리(210)에 로딩될 수 있다.

[0036] 프로세서(220)는 기본적인 산술, 로직 및 입출력 연산을 수행함으로써, 컴퓨터 프로그램의 명령을 처리하도록 구성될 수 있다. 명령은 메모리(210) 또는 통신 인터페이스(230)에 의해 프로세서(220)로 제공될 수 있다. 예를 들어 프로세서(220)는 메모리(210)와 같은 기록 장치에 저장된 프로그램 코드에 따라 수신되는 명령을 실행하도록 구성될 수 있다.

[0037] 통신 인터페이스(230)는 네트워크(170)를 통해 컴퓨터 장치(200)가 다른 장치(일례로, 앞서 설명한 저장 장치들)와 서로 통신하기 위한 기능을 제공할 수 있다. 일례로, 컴퓨터 장치(200)의 프로세서(220)가 메모리(210)와 같은 기록 장치에 저장된 프로그램 코드에 따라 생성한 요청이나 명령, 데이터, 파일 등이 통신 인터페이스(230)의 제어에 따라 네트워크(170)를 통해 다른 장치들로 전달될 수 있다. 역으로, 다른 장치로부터의 신호나 명령, 데이터, 파일 등이 네트워크(170)를 거쳐 컴퓨터 장치(200)의 통신 인터페이스(230)를 통해 컴퓨터 장치(200)로 수신될 수 있다. 통신 인터페이스(230)를 통해 수신된 신호나 명령, 데이터 등은 프로세서(220)나

메모리(210)로 전달될 수 있고, 파일 등은 컴퓨터 장치(200)가 더 포함할 수 있는 저장 매체(상술한 영구 저장 장치)로 저장될 수 있다.

[0038] 입출력 인터페이스(240)는 입출력 장치(250)와의 인터페이스를 위한 수단일 수 있다. 예를 들어, 입력 장치는 마이크, 키보드 또는 마우스 등의 장치를, 그리고 출력 장치는 디스플레이, 스피커와 같은 장치를 포함할 수 있다. 다른 예로 입출력 인터페이스(240)는 터치스크린과 같이 입력과 출력을 위한 기능이 하나로 통합된 장치와의 인터페이스를 위한 수단일 수도 있다. 입출력 장치(250)는 컴퓨터 장치(200)와 하나의 장치로 구성될 수도 있다.

[0039] 또한, 다른 실시예들에서 컴퓨터 장치(200)는 도 2의 구성요소들보다 더 적은 혹은 더 많은 구성요소들을 포함할 수도 있다. 그러나, 대부분의 종래기술적 구성요소들을 명확하게 도시할 필요성은 없다. 예를 들어, 컴퓨터 장치(200)는 상술한 입출력 장치(250) 중 적어도 일부를 포함하도록 구현되거나 또는 트랜시버(transceiver), 데이터베이스 등과 같은 다른 구성요소들을 더 포함할 수도 있다.

[0040] 이하에서는 그래프 기반 아이템 인기도 예측 모델을 위한 그래프 구성 및 자가지도 학습의 구체적인 실시예를 설명하기로 한다.

[0041] 아이템 추천을 위해 사용되는 방법 중 협업 필터링은 단순하고 직관적인 개념으로 인해 추천 시스템에서 널리 사용되고 있으며, 이는 사용자와 아이템 간의 인터랙션을 기반으로 사용자 행동을 예측하는 것이다.

[0042] GNN(graph neural network)이 여러 분야에서 만족할 만한 성과를 거두면서 GNN을 활용한 협업 필터링에서는 사용자와 아이템 간의 인터랙션을 그래프로 재구성하고 인접 노드의 협업 메시지를 집계하는 방식을 활용하고 있다.

[0043] 또한, 아이템에 대한 구체적인 정보로서 이미지, 텍스트, 오디오 등과 같은 멀티모달 정보는 아이템에 대한 사용자 행동을 이해하는데 풍부한 단서를 제공할 수 있다.

[0044] 본 실시예에서는 멀티모달 정보를 아이템에 대한 보조 정보(auxiliary information)로 활용하여 사용자와 아이템 간의 관계성과 융합한 피처 표현을 학습할 수 있다.

[0045] 최근 인공지능 분야에서는 이미지, 텍스트, 오디오 등의 멀티모달 정보를 융합하여 특정 모달리티의 정보를 보완하거나 새로운 의미를 도출함으로써 컴퓨터 비전, 자연어 처리 등의 다양한 분야에서 좋은 성능을 보여주고 있다.

[0046] 본 실시예에서는 멀티모달 정보의 학습 장점을 아이템 추천, 아이템 검색, 아이템 쇼핑 등 온라인 플랫폼에서의 사용자와 아이템 간의 관계성과 융합하여 아이템의 인기도를 예측할 수 있다.

[0047] 도 3은 본 발명의 일실시예에 따른 컴퓨터 장치가 수행할 수 있는 방법의 예를 도시한 흐름도이다.

[0048] 본 실시예에 따른 컴퓨터 장치(200)는 클라이언트를 대상으로 클라이언트 상에 설치된 전용 어플리케이션이나 컴퓨터 장치(200)와 관련된 웹/모바일 사이트 접속을 통해 아이템 추천 서비스를 제공할 수 있다. 컴퓨터 장치(200)에는 컴퓨터로 구현된 멀티모달 임베딩 시스템이 구성될 수 있다. 일례로, 멀티모달 임베딩 시스템은 독립적으로 동작하는 프로그램 형태로 구현되거나, 혹은 특정 어플리케이션의 인-앱(in-app) 형태로 구성되어 상기 특정 어플리케이션 상에서 동작이 가능하도록 구현될 수 있다.

[0049] 컴퓨터 장치(200)의 프로세서(220)는 도 3에 따른 멀티모달 임베딩 방법을 수행하기 위한 구성요소로서 적어도 하나 이상의 구성요소를 포함할 수 있다. 실시예에 따라 프로세서(220)의 구성요소들은 선택적으로 프로세서(220)에 포함되거나 제외될 수도 있다. 또한, 실시예에 따라 프로세서(220)의 구성요소들은 프로세서(220)의 기능의 표현을 위해 분리 또는 병합될 수도 있다.

[0050] 이러한 프로세서(220) 및 프로세서(220)의 구성요소들은 도 3의 멀티모달 임베딩 방법이 포함하는 단계들(S310 내지 S320)을 수행하도록 컴퓨터 장치(200)를 제어할 수 있다. 예를 들어, 프로세서(220) 및 프로세서(220)의 구성요소들은 메모리(210)가 포함하는 운영체제의 코드와 적어도 하나의 프로그램의 코드에 따른 명령(instruction)을 실행하도록 구현될 수 있다.

[0051] 여기서, 프로세서(220)의 구성요소들은 컴퓨터 장치(200)에 저장된 프로그램 코드가 제공하는 명령에 따라 프로세서(220)에 의해 수행되는 서로 다른 기능들(different functions)의 표현들일 수 있다.

[0052] 프로세서(220)는 컴퓨터 장치(200)의 제어와 관련된 명령이 로딩된 메모리(210)로부터 필요한 명령을 읽어들일

수 있다. 이 경우, 상기 읽어들인 명령은 프로세서(220)가 이후 설명될 단계들(S310 내지 S320)을 실행하도록 제어하기 위한 명령을 포함할 수 있다.

[0053] 이후 설명될 단계들(S310 내지 S320)은 도 3에 도시된 순서와 다른 순서로 수행될 수 있으며, 단계들(S310 내지 S320) 중 일부가 생략되거나 추가의 과정이 더 포함될 수 있다.

[0054] 도 3을 참조하면, 단계(S310)에서 프로세서(220)는 사용자와 아이템의 관계를 나타내는 그래프에 각 아이템의 멀티모달 정보를 추가하여 사용자와 아이템 간의 관계성과 멀티모달 정보가 융합된 멀티모달 그래프를 생성할 수 있다. 일례로, 프로세서(220)는 사용자, 아이템, 그리고 각 아이템의 이미지(예를 들어, Thumbnail image 등)와 텍스트(예를 들어, name, title 등)를 노드로 하는 융합 멀티모달 그래프를 구성할 수 있다. 이때, 사용자와 아이템 간의 인터랙션(예를 들어, 구매 이력, 클릭 이력 등)은 사용자 노드와 아이템 노드 간의 엣지로 표현될 수 있다. 그리고, 아이템의 이미지와 텍스트는 아이템에 대한 보조 정보로 활용되는 멀티모달 정보에 해당되는 것으로, 해당 아이템 노드와 엣지로 연결될 수 있다. 다시 말해, 이미지 노드와 텍스트 노드는 사용자와 아이템의 관계를 나타내는 이분 그래프에서 해당 아이템 노드에 연결되는 별개의 노드로 취급될 수 있다.

[0055] 단계(S320)에서 프로세서(220)는 사용자와 아이템 간의 관계성과 멀티모달 정보가 융합된 멀티모달 그래프에 대한 임베딩을 수행할 수 있다. 프로세서(220)는 사용자 노드, 아이템 노드, 이미지 노드(비주얼 노드), 및 텍스트 노드로 구성된 멀티모달 그래프를 노드 간의 연결 관계가 포함된 임베딩 벡터로 인코딩할 수 있다. 프로세서(220)는 융합 멀티모달 그래프를 통해 사용자와 아이템의 관계를 학습하고, 이때 여러 정보가 통합된 하나의 그래프에서 레이어 별 메시지 전파(layer-wise message passing)를 통해 인접한 이웃 노드로 정보를 전달할 수 있다. 다시 말해, 프로세서(220)는 그래프 인코더를 통해 멀티모달 정보(이미지와 텍스트)뿐만 아니라 협업 신호에 대해 레이어 별 집계(aggregation)를 수행할 수 있다. 여러 계층을 스태킹(stacking)함으로써 정보를 재귀적으로 집계하고 각 노드의 풍부한 피처 표현을 생성할 수 있다.

[0056] 사용자와 아이템의 관계성을 강화하기 위해 멀티모달 정보에 기반한 자가지도 학습을 적용할 수 있다. 프로세서(220)는 자기 지도 학습의 일환으로 멀티모달 정보를 활용하여 대조 학습을 수행할 수 있다. 본 실시예에서는 융합 멀티모달 그래프로 표현된 사용자와 아이템의 관계성을 기반하여 대조 학습 쌍(pair)을 구성할 수 있다. 예를 들어, 학습을 위해 입력된 사용자-아이템 쌍을 기준으로 보았을 때, 해당 아이템의 멀티모달 정보를 기준(anchor)으로 보고 사용자가 구매한 다른 아이템의 멀티모달 정보를 포지티브 샘플(positive sample)로 구성하고 소비자가 구매하지 않은 아이템의 멀티모달 정보를 네거티브 샘플(negative sample)로 구성하여 대조 학습을 수행할 수 있다. 이를 통해 아이템의 멀티모달 정보들은 사용자-아이템 간의 관계성 정보를 학습하여 구성한 멀티모달 그래프의 메시지 전파 과정에서 더욱 정확하고 풍부한 정보를 전달할 수 있다.

[0057] 더 나아가, 프로세서(220)는 멀티모달 정보를 활용하여 그래프의 일관성 정규화를 수행할 수 있다. 아이템의 멀티모달 정보들은 사전 학습(pre-train)된 이미지 또는 텍스트 인코더에서 추출될 수 있다. 그렇기 때문에 서로 다른 모달리티의 정보를 그래프를 통해 융합하는 과정에서 불필요한 편차 문제가 발생할 수 있다. 이를 방지하기 위해 본 실시예에서는 모달리티 드롭아웃(modality dropout) 기법을 활용할 수 있다. 기존의 그래프 증강(graph augmentation) 방법과는 다르게, 모달리티 드롭아웃은 각 증강에 대해 하나의 모달리티 노드 전체를 제거하는 방법을 사용한다. 이렇게 구성된 증강 그래프를 이용하여 메시지 전파에도 결과값이 일관될 수 있는 정규화 학습을 통해 특정 모달리티에 대한 민감도와 의존도를 낮추어 더욱 융통성있는 학습이 가능하다.

[0058] 멀티모달 정보와 융합된 구조의 그래프를 설계하여 학습한 멀티모달 융합 인공지능 모델을 SIMGNN(Self-supervised Integrated Multimodal Graph Neural Network)이라 칭하기로 한다.

[0059] 이하에서는 SIMGNN에 대한 방법론을 구체적으로 설명하기로 한다.

[0060] 도 4 내지 도 5는 본 발명의 일실시예에 있어서 그래프 임베딩 초기화를 설명하기 위한 예시 도면이다.

[0061] 도 4를 참조하면, 프로세서(220)는 사용자와 아이템 간의 인터랙션 데이터(41)를 이용하여 사용자 노드와 아이템 노드로 구성된 사용자-아이템 이분 그래프(410) $G_B=\{(u, i)|u\in U,i\in I\}$를 구성할 수 있다. 여기서, U는 사용자 집합, I는 아이템 집합을 나타낸다. 사용자와 아이템에 대한 초기 임베딩 벡터를 $e_u^{(0)} \in \mathbb{R}^d$, $e_i^{(0)} \in \mathbb{R}^d$ (여기서, d는 임베딩 차원)로 랜덤으로 초기화할 수 있다.

[0062] 그리고, 프로세서(220)는 도 5에 도시한 바와 같이 아이템의 멀티모달 정보로서 아이템 이미지(51)를 이미지 인코더(511)를 통해 인코딩하여 비주얼 피처(501)를 추출할 수 있다. 또한, 아이템 텍스트(52)를 텍스트 인코더(512)를 통해 인코딩하여 텍스트 피처(502)를 추출할 수 있다. 그런 다음, 프로세서(220)는 각 아이템의 비주

얼 피처(501)와 텍스트 피처(502)를 별개의 노드로 사용자-아이템 이분 그래프(410)에 통합하여 융합 멀티모달 그래프(520) $G_M=G_B \cup \{(i, m)|i \in I, m \in M\}$를 구성할 수 있다. 여기서, M은 아이템의 이미지 및 텍스트 엔티티 집합을 나타낸다.

[0063] 온라인 서비스에서 아이템의 이미지와 텍스트가 사용자 결정에 영향을 미친다는 점을 고려하여 이미지와 텍스트 모달리티를 해당 아이템의 멀티모달 피처로 사용한다. 모달리티 유형은 데이터셋에 따라 추가하거나 대체할 수 있으며, 예를 들어 이미지나 텍스트 이외에 오디오와 같은 모달리티를 활용하는 것 또한 가능하다.

[0064] 각 아이템 $i \in I$에 대해 해당 아이템의 대표 이미지와 아이템명을 사전 학습 모델(511, 512)(예를 들어, BERT 등)의 입력으로 제공하여 비주얼 피처(501) $f_{i,v} \in \mathbb{R}^{d_v}$와 텍스트 피처(502) $f_{i,t} \in \mathbb{R}^{d_t}$를 각각 추출할 수 있다. 융합 멀티모달 그래프(520) G_M에서 아이템 i의 멀티모달 노드의 초기 입력 임베딩은 수학식 1과 같이 공식화될 수 있다.

[0065] [수학식 1]

[0066] $$e_{i,m}^{(0)} = T_m \cdot f_{i,m}$$

[0067] 여기서, $T_m^{d \times d_m}$은 모달리티 m에 대한 훈련 가능한 변환 매트릭스(transformation matrix)를 의미한다. d와 d_m은 각각 변환된 피처와 초기 추출된 피처의 임베딩 차원을 의미한다. 다시 말해, 본 실시예에서는 인코더(511, 512)에서 추출된 멀티모달 피처에 대해 학습 가능한 변환 매트릭스를 곱한다. 이는 같은 임베딩 공간에 멀티모달 피처를 매핑하여 학습 가능한 파라미터로 설정하기 위한 것이다.

[0068] 추천 지도 태스크(supervised recommendation task)은 융합 멀티모달 그래프(520) G_M이 주어지면 사용자 u와 아이템 I 간에 인터랙션 할 수 있는 잠재적 확률을 나타내는 예측 함수 $\hat{y}_{ui} = F(G_M, E, (u, i); \Theta)$를 학습하는 것이다. 여기서, E는 모든 임베딩 벡터를 포함하는 행렬이고, Θ는 함수 F의 모델 파라미터를 의미한다. 이러한 예측 함수를 사용하여 모델 파라미터를 지도 방식으로 최적화할 수 있다. 상위 K개(top-K)의 추천 태스크의 경우 인터랙션 예측이 관찰되지 않은 샘플보다 더 높은 순위를 차지하도록 장려하기 위해 쌍별 베이지안(Bayesian) 개인화 랭킹(BPR) 손실을 주요 손실(main loss)로 사용할 수 있다.

[0069] [수학식 2]

[0070] $$L_{main} = \sum_{(u,i,k) \in O} -\log \sigma (\hat{y}_{ui} - \hat{y}_{uk})$$

[0071] 여기서, $O=\{(u,i,k)|(u,i) \in O^+, (u,k) \in O^-\}$는 관찰된 인터랙션으로 O^+를, 관찰되지 않은 인터랙션으로 O^-를 갖는 학습 데이터를 의미한다.

[0072] 사용자가 특정 아이템을 클릭(또는 구매, 뷰 등)할지 여부를 예측하는 이진 분류 태스크인 CRT(click-through rate) 예측의 경우 이진 교차 엔트로피(BCE) 손실이 주요 손실 함수로 사용될 수 있다.

[0073] [수학식 3]

[0074] $$L_{main} = \sum_{(u,i) \in O} -y_{ui} \cdot \log \sigma (\hat{y}_{ui}) - (1 - y_{ui}) \cdot \log (1 - \sigma(\hat{y}_{ui}))$$

[0075] 여기서, O는 학습 데이터이고, $y_{ui} \in \{0,1\}$는 실측 값(ground truth) 레이블을 의미하고, $\sigma(\cdot)$는 시그모이드 함수를 의미한다.

[0076] 도 6은 본 발명의 일실시예에 있어서 SIMGNN 모델의 전반적인 아키텍처를 도시한 것이다.

[0077] 도 6을 참조하면, 본 발명에 따른 SIMGNN 모델(600)은 융합 멀티모달 그래프(520)에서 멀티모달 표현을 집계할 수 있으며, 그런 다음 자가지도 대조 학습을 통해 멀티모달 표현을 안내하여 추천 성능을 향상시키기 위한 협업 신호를 보존할 수 있다. 또한, SIMGNN 모델(600)은 여러 그래프 증강(631, 632, 633)을 구성하여 서로 다른 모달리티에서 다중 시점(multiple view)을 생성함으로써 일관성을 최대화하도록 학습할 수 있다.

[0078] SIMGNN 모델(600)은 융합 멀티모달 그래프(520)에서 레이어 별 메시지 전달을 수행하는 그래프 인코더(610)를

포함할 수 있다. 융합 멀티모달 그래프(520)는 사용자 노드, 아이템 노드, 비주얼 노드, 및 텍스트 노드를 포함하는 다양한 노드 유형으로 구성되며, 멀티모달 정보를 전파하기 위해 추가적인 연산이 필요하다. 여기서, 단일 레이어에 대한 임베딩 방정식을 공식화하고 이를 여러 레이어로 일반화할 수 있다.

[0079] 각 아이템 노드에 비주얼 노드와 텍스트 노드를 연결함으로써 사용자와 아이템의 관계성과 각 아이템의 멀티모달 정보가 통합된 융합 멀티모달 그래프(520)를 구성할 수 있다. 멀티모달 정보를 각각 다루는 것이 아닌, 통합된 하나의 그래프로 구성함으로써 하나의 그래프 연산 레이어만으로 사용자의 전체적인 정보에 대한 선호도를 반영할 수 있다. 사용자가 실제 의사 결정을 할 때에 각각의 모달리티를 따로 고려하기보다 복합적인 정보를 통해 의사를 결정한다는 점을 반영한 모델을 만들 수 있다.

[0080] 도 7은 본 발명의 일실시예에 있어서 융합 멀티모달 그래프에 대한 임베딩을 설명하기 위한 예시 도면이다.

[0081] 먼저, 도 7의 (a) 도면을 참조하면, 프로세서(220)는 이웃하는 아이템 노드로부터 정보를 전달받는 사용자 노드를 임베딩 벡터로 인코딩함으로써 사용자 임베딩을 생성할 수 있다. 사용자 노드는 1-홉(hop) 인접 아이템 노드로부터 메시지를 수집한다. 또한, 이전 계층에서 자기 임베딩을 추가하여 노드의 자기 연결을 취한다. 이를 통해 사용자 노드는 연속된 레이어 전체에서 자신의 아이덴티티를 보존할 수 있다. 일례로, 레이어 l에서 l+1까지의 사용자 u에 대해 설명된 임베딩 함수를 수학식 4와 같이 정의할 수 있다.

[0082] [수학식 4]

[0083] $$e_u^{(l+1)} = g\left(\sum_{i \in N_u} \frac{1}{\sqrt{|N_u||N_i|}} e_i^{(l)}\right) + e_u^{(l)}$$

[0084] 여기서, N_u는 사용자 u와 연결된 아이템의 집합을 의미하고, N_i는 아이템 i와 연결된 사용자의 집합을 의미한다. 그리고, g(·)는 비선형 활성화 함수로, 예를 들어 LeakyReLU를 사용하여 메시지를 포지티브 값과 작은 네거티브 값으로 인코딩할 수 있다.

[0085] 다음으로, 도 7의 (b) 도면을 참조하면, 프로세서(220)는 이웃하는 멀티모달 노드(비주얼 노드와 텍스트 노드)와 사용자 노드로부터 정보를 전달받는 아이템 노드를 임베딩 벡터로 인코딩함으로써 아이템 임베딩을 생성할 수 있다. 아이템 노드는 먼저 사용자의 집계 함수를 따라 협업 신호를 캡처하고, 사용자 노드 이외에도 멀티모달 노드에서 메시지를 수집할 수 있다. 아이템 임베딩은 수학식 5와 같이 정의될 수 있다.

[0086] [수학식 5]

[0087] $$e_i^{(l+1)} = g\left(\sum_{i \in N_i} \frac{1}{\sqrt{|N_u||N_i|}} e_u^{(l)}\right) + e_i^{(l)} + \sum_{m \in \{v,t\}} W_m^{(l)} e_{i,m}^{(l)}$$

[0088] 여기서, $W_m^{(l)}$ 은 l번째 레이어에 있는 모달리티의 훈련 가능한 가중치 매트릭스를 나타낸다. 아이템 임베딩은 사용자와 아이템의 초기화된 ID 임베딩과 달리 구체적인 의미론을 포함하기 때문에 멀티모달 노드의 메시지에 대해서만 훈련 가능한 가중치 매트릭스를 곱한다.

[0089] 다음으로, 도 7의 (c) 도면을 참조하면, 프로세서(220)는 이웃하는 아이템 노드로부터 정보를 전달받는 멀티모달 노드(비주얼 노드와 텍스트 노드)를 임베딩 벡터로 인코딩함으로써 멀티모달 임베딩을 생성할 수 있다. 본 실시예에서는 아이템의 멀티모달 피처(비주얼 피처와 텍스트 피처)를 융합 멀티모달 그래프(520)에서 별개의 노드로 취급한다. 멀티모달 노드는 임베딩 레이어의 아이템 노드로부터 메시지를 수신하여 협업 신호를 인코딩할 수 있다. 멀티모달 임베딩은 수학식 6과 같이 정의될 수 있다.

[0090] [수학식 6]

[0091] $$e_{i,m}^{(l+1)} = g\left(e_i^{(l)}\right) + e_{i,m}^{(l)}$$

[0092] 아이템 노드의 메시지는 두 모달리티에 모두 동일하며, 추가된 자기 연결 $e_{i,m}^{(l)}$은 모달리티 간의 메시지를 구분한다.

[0093] 도 8에 도시한 바와 같이, 프로세서(220)는 융합 멀티모달 그래프(520)에 대한 L개의 연속적인 전파 레이어 후

에 사용자 $e_u^{(L)}$ 와 아이템 $e_i^{(L)}$ 의 최종 피처 표현을 얻을 수 있다. 피처 표현은 L-홉 인접 노드의 집계된 메시지를 암시하며, 이를 사용하여 아이템에 대한 사용자의 인기도를 나타내는 친밀도 점수(affinity score)를 정량화할 수 있다. 특히, 사용자 u와 아이템 i 간의 친밀도는 두 최종 표현 벡터의 내적(inner product)에 의해 계산될 수 있다.

[0094] [수학식 7]

[0095] $$\hat{y}_{ui} = e_u^{(L)\top} e_i^{(L)}$$

[0096] SIMGNN 모델(600)은 융합 멀티모달 그래프(520)가 입력으로 주어지는 경우 사용자 u와 아이템 i 간의 친밀도 점수를 출력으로 제공한다. 친밀도 점수는 각 사용자와 아이템의 피처 표현 벡터 간의 내적으로 하나의 스칼라 값을 구하고 그 값에 시그모이드 함수를 적용하여 산출될 수 있다.

[0097] 본 실시예에서는 융합 멀티모달 그래프(520)의 관계 구조를 기반으로 하는 새로운 대조 학습 방법을 제안한다. 구체적으로, 동일한 인스턴스의 증강을 통해 포지티브 쌍을 구성하는 기존의 대조 학습 방법과 달리, 본 발명에서는 사용자와 아이템 간의 관계를 기반으로 포지티브 쌍과 네거티브 쌍을 구성하고 이들 쌍에 대해 대조 학습을 수행한다. 이를 통해 각 모달리티의 표현이 협업 신호를 추가로 인코딩할 수 있으며, 더 많은 일반 사용자가 있는 표현이 임베딩 공간에 더 가깝게 배치될 수 있다.

[0098] 도 9는 본 발명의 일실시예에 있어서 멀티모달 정보를 이용한 대조 학습 과정을 도시한 것이다.

[0099] 본 실시예에서는 일반적인 대조 손실 개념에 사용자-아이템의 관계 구조를 반영하는 방식으로 학습을 진행할 수 있다. 도 9를 참조하면, 학습 샘플 (u,i)∈O가 주어지면, 해당 아이템 i의 멀티모달 피처를 각 모달리티에 대한 기준(anchor)으로 간주한다. 그런 다음, 사용자 u와 인터랙션 이력이 있는 아이템 집합 N_u에서 한 아이템을 샘플링하고 해당 아이템의 멀티모달 피처를 포지티브 인스턴스로 선정한다. 다음으로, 아이템 집합 N_u에 대한 여집합인 N_u^c, 즉 사용자 u와 인터랙션 이력이 없는 아이템 집합 N_u^c에서 랜덤으로 여러 개의 네거티브 샘플을 추출하여 집합 K를 선정할 수 있다.

[0100] 학습 샘플 (u,i)에 대한 비주얼 모달리티의 대조 손실은 수학식 8과 같다.

[0101] [수학식 8]

[0102] $$cl_{(u,i,j,K)}^{(v)} = -log \frac{\exp\left(\frac{< e_{i,v}, e_{j,v} >}{\tau}\right)}{\exp\left(\frac{< e_{i,v}, e_{j,v} >}{\tau}\right) + \sum_{k \in K} \exp\left(\frac{< e_{i,v}, e_{k,v} >}{\tau}\right)}$$

[0103] 여기서, j는 아이템 집합 N_u에서 랜덤으로 추출한 포지티브 아이템을 의미하고, k는 아이템 여집합 N_u^c에서 랜덤으로 추출한 네거티브 아이템을 의미한다. $<\cdot,\cdot>$은 두 요소 사이의 코사인 유사성을 나타내고, τ는 확률조건 파라미터(temperature parameter)를 의미한다.

[0104] 마찬가지로, 학습 샘플 (u,i)에 대한 텍스트 모달리티에 대한 대조 손실은 다음과 같이 정의할 수 있다.

[0105] [수학식 9]

[0106] $$cl_{(u,i,j,K)}^{(t)} = -log \frac{\exp\left(\frac{< e_{i,t}, e_{j,t} >}{\tau}\right)}{\exp\left(\frac{< e_{i,t}, e_{j,t} >}{\tau}\right) + \sum_{k \in K} \exp\left(\frac{< e_{i,t}, e_{k,t} >}{\tau}\right)}$$

[0107] 본 실시예에서는 비주얼 대조 손실과 텍스트 대조 손실이 결합된 대조 손실을 최소화하는 방향으로 SIMGNN 모델(600)을 학습할 수 있다.

[0108] [수학식 10]

[0109] $$L_{cont} = \sum_{(u,i) \in O} \left(cl_{(u,i,j,K)}^{(v)} + cl_{(u,i,j,K)}^{(t)}\right)$$

[0110] 도 10은 본 발명의 일실시예에 있어서 멀티모달 정보를 이용한 일관성 정규화 학습 과정을 도시한 것이다.

[0111] 각 모달리티의 피처는 서로 다른 사전 훈련 인코더에서 비롯되기 때문에 이러한 멀티모달 피처의 다양성은 종종 불필요한 분산 문제를 초래할 수 있다. 따라서, 특정 피처나 특정 모달리티에 대한 민감도 및 의존성을 완화하기 위해 본 실시예에서는 SIMGNN 모델(600)이 그래프 증강을 사용하여 다른 모달리티에서 생성된 뷰와 일치하도록 자기 지도 정규화 항을 추가할 수 있다.

[0112] 도 10을 참조하면, 본 실시예에서는 융합 멀티모달 그래프(520)는 물론이고, 다른 이기종 그래프에 적용할 수 있는 그래프 증강 방법으로 모달리티 드롭아웃 기법을 사용한다. 모달리티 드롭아웃은 각 증강에 대해 특정 모달리티의 전체 노드를 삭제하여 특정 모달리티에 대한 의존성을 줄이는 것이다. 본 실시예에서는 3가지 증강 그래프로서 융합 멀티모달 그래프(520)에서 비주얼 노드를 모두 삭제한 증강 그래프(631), 멀티모달 그래프(520)에서 텍스트 노드를 모두 삭제한 증강 그래프(632), 멀티모달 그래프(520)에서 비주얼 노드와 텍스트 노드를 모두 삭제한 증강 그래프(633)를 사용할 수 있다. 각 증강 그래프(631, 632, 633)는 앞서 설명한 메시지 전파 아키텍처와 예측 레이어를 따른다.

[0113] 융합 멀티모달 그래프(520)와 증강 그래프(631, 632, 633)는 각각 사용자와 아이템의 다른 표현을 생성하여 다른 예측을 초래한다. 학습 샘플 (u,i)에 대해 서로 다른 예측 점수의 평균을 계산할 수 있다. 즉, $\tilde{y}_{ui} = \frac{1}{S+1}(\hat{y}_{ui} + \sum_{s=1}^{S} \hat{y}_{ui(s)})$와 같으며, 여기서 S는 총 증강 횟수를 의미한다. 모든 학습 샘플 (u,i)∈O에 대해 평균과 각 예측 값 사이의 L2 거리를 최소화하여 일관성을 향상시킬 수 있다.

[0114] [수학식 11]

[0115] $$L_{cons} = \sum_{(u,i)\in O} \frac{1}{S+1}(\| \tilde{y}_{ui} - \hat{y}_{ui} \|_2^2 + \sum_{s=0}^{S} \| \tilde{y}_{ui} - \hat{y}_{ui(s)} \|_2^2$$

[0116] CTR 예측을 위해 예측 값 $\hat{z}_{ui(s)} = \sigma(\hat{y}_{ui(s)}$에 시그모이드 함수를 적용하여 평균 $\tilde{z}_{ui} = \frac{1}{S+1}(\hat{z}_{ui} + \sum_{s=1}^{S} \hat{z}_{ui(s)})$를 계산한다. 마찬가지로, 분산을 최소화하기 위해 일관성 학습을 적용할 수 있다.

[0117] [수학식 12]

[0118] $$L_{cons} = \sum_{(u,i)\in O} \frac{1}{S+1}(\| \tilde{z}_{ui} - \hat{z}_{ui} \|_2^2 + \sum_{s=0}^{S} \| \tilde{z}_{ui} - \hat{z}_{ui(s)} \|_2^2$$

[0119] SIMGNN 모델(600)은 세 가지 손실 함수인 (1) 추천 태스크에 대한 지도 손실 함수 L_{main}, (2) 사용자-아이템 그래프의 인접성을 보존하기 위해 멀티모달 피처를 강화하는 대조 손실 함수 L_{cont}, 멀티모달 피처의 분산 문제를 완화하기 위한 일관성 손실 함수 L_{cons}를 최소화할 수 있다.

[0120] SIMGNN 모델(600)의 최종 목적 함수는 수학식 13과 같이 정의될 수 있다. 각 손실 함수의 가중치를 결정하는 계수인 λ_1, λ_2, λ_3은 각 손실 함수와 선형적으로 결합된다.

[0121] [수학식 13]

[0122] $$L_{total} = L_{main} + \lambda_1 L_{cons} + \lambda_2 L_{cont} + \lambda_3 \| \Theta \|_2^2$$

[0123] 여기서, 지도 손실 함수 L_{main}는 BCE 또는 BPR 손실 함수를 사용할 수 있고, Θ는 모델 파라미터를 의미한다.

[0124] SIMGNN 모델(600)의 전반적인 학습 알고리즘은 도 11에 도시한 바와 같으며, 아이템의 멀티모달 정보가 융합된 그래프를 통해 사용자-아이템의 관계를 학습할 수 있고, 이때 멀티모달 정보를 이용한 대조 학습과 일관성 정규화 학습을 통해 모달리티 별 협업 신호를 같은 임베딩 공간에 효과적으로 인코딩할 수 있다.

[0125] SIMGNN 모델(600)의 복잡도는 다음과 같다. |V|와 |E|를 사용자-아이템 이분 그래프(410) G_B의 노드 수와 엣지 수라 하자. 모든 아이템 노드에 대해 두 개의 멀티모달 노드(비주얼 노드와 텍스트 노드)를 연결하여 융합 멀티모달 그래프(520) G_M을 구성한다. 융합 멀티모달 그래프(520) G_M은 추가적인 인접을 필요로 하지 않으므로 융합 멀티모달 그래프(520) G_M에 대한 복잡도는 O(2|E|)이다. 일관성 정규화를 위한 3개 그래프 증강에 따라 총

공간 복잡도는 O(8|E|)이다. 따라서, 그래프 컨볼루션에 대한 총 시간 복잡성은 $O(8(|E|d+|I|)L_S\frac{|E|}{B})$가 된다. 여기서, |I|는 아이템 수, d는 임베딩 차원, L은 컨볼루션 레이어 총 수, S는 에포크(epoch) 수, B는 학습 배치(batch) 크기를 의미한다. 대조 학습을 위해 포지티브 샘플에 대해 하나의 매트릭스 곱을 수행하고 네거티브에 대해 K개 매트릭스 곱을 수행할 수 있다. 결과적으로, $O((K+1)B^2d)$의 복잡성을 초래한다. 일관성 정규화는 4개의 그래프 결과와 평균 사이의 L2 거리를 계산하며, 복잡도는 O(4d)이다.

[0126] 이처럼 본 발명의 실시예들에 따르면, 사용자와 아이템의 관계뿐만 아니라 각 아이템의 멀티모달 정보를 포함한 피처 표현을 통해 모달리티 별 협업 신호를 보다 효과적으로 인코딩할 수 있다. 더 나아가, 본 발명의 실시예들에 따르면, 사용자와 아이템의 관계성을 기반으로 멀티모달 정보에 대한 대조 학습 쌍을 구성함으로써 더욱 정확하고 풍부한 메시지 전파를 통해 사용자와 아이템의 관계성을 강화할 수 있다. 또한, 본 발명의 실시예들에 따르면, 융합 멀티모달 그래프에 대한 일관성 정규화를 통해 특정 피처나 모달리티에 대한 민감도나 의존도를 완화할 수 있고 피처 임베딩 간의 불필요한 편차 문제를 해결할 수 있다.

[0127] 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, 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 어플리케이션을 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.

[0128] 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, 컴퓨터 저장 매체 또는 장치에 구체화(embody)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.

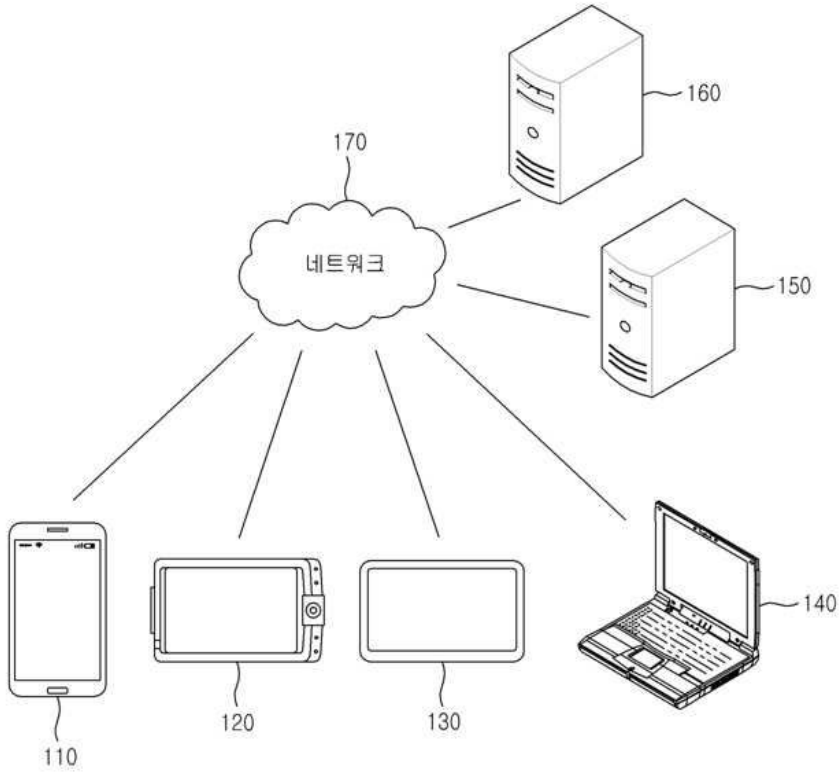
[0129] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 이때, 매체는 컴퓨터로 실행 가능한 프로그램을 계속 저장하거나, 실행 또는 다운로드를 위해 임시 저장하는 것일 수도 있다. 또한, 매체는 단일 또는 수 개의 하드웨어가 결합된 형태의 다양한 기록수단 또는 저장수단일 수 있는데, 어떤 컴퓨터 시스템에 직접 접속되는 매체에 한정되지 않고, 네트워크 상에 분산 존재하는 것일 수도 있다. 매체의 예시로는, 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체, CD-ROM 및 DVD와 같은 광기록 매체, 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical medium), 및 ROM, RAM, 플래시 메모리 등을 포함하여 프로그램 명령어가 저장되도록 구성된 것이 있을 수 있다. 또한, 다른 매체의 예시로, 어플리케이션을 유통하는 앱 스토어나 기타 다양한 소프트웨어를 공급 내지 유통하는 사이트, 서버 등에서 관리하는 기록매체 내지 저장매체도 들 수 있다.

[0130] 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.

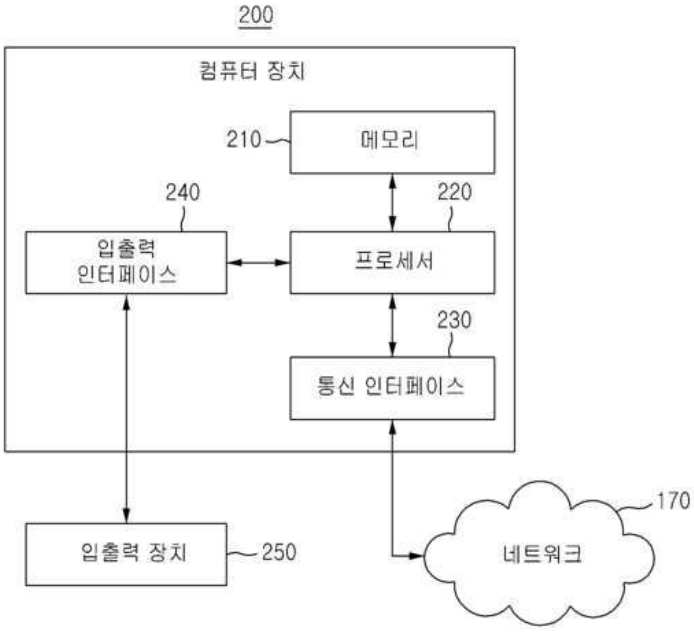
[0131] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

도면

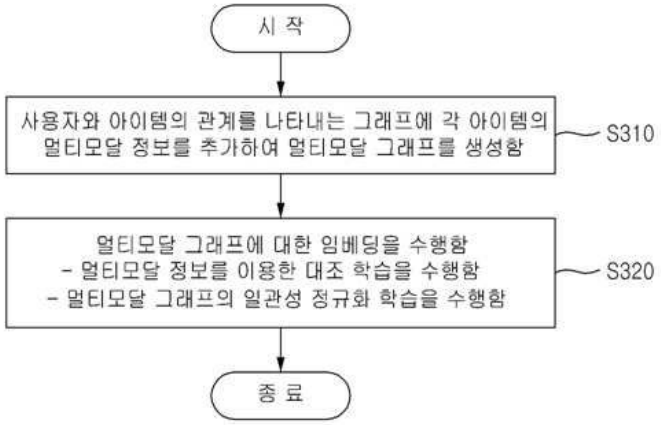
도면1


160
170
네트워크
150
140
110
120
130

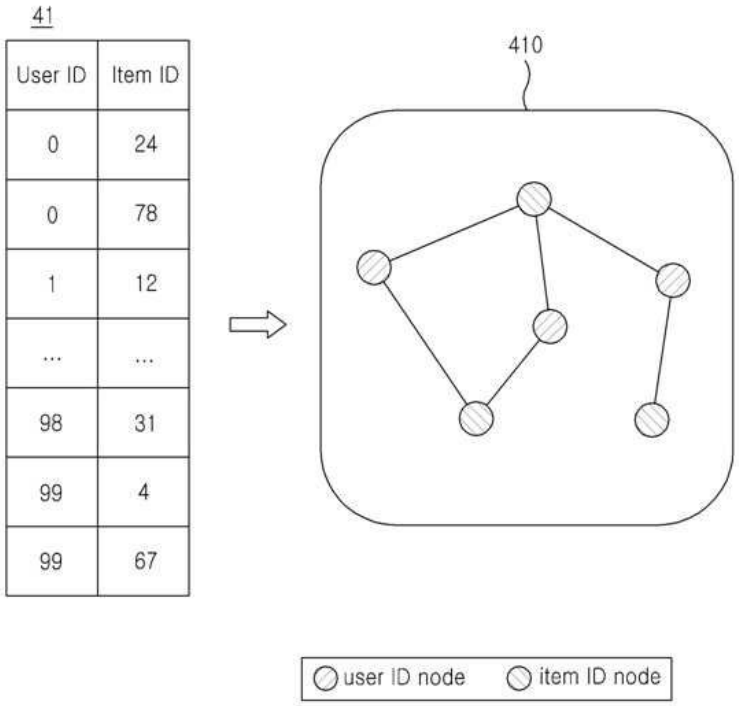
도면2


200
컴퓨터 장치
210
메모리
240
220
입출력
인터페이스
프로세서
230
통신 인터페이스
170
입출력 장치
250
네트워크

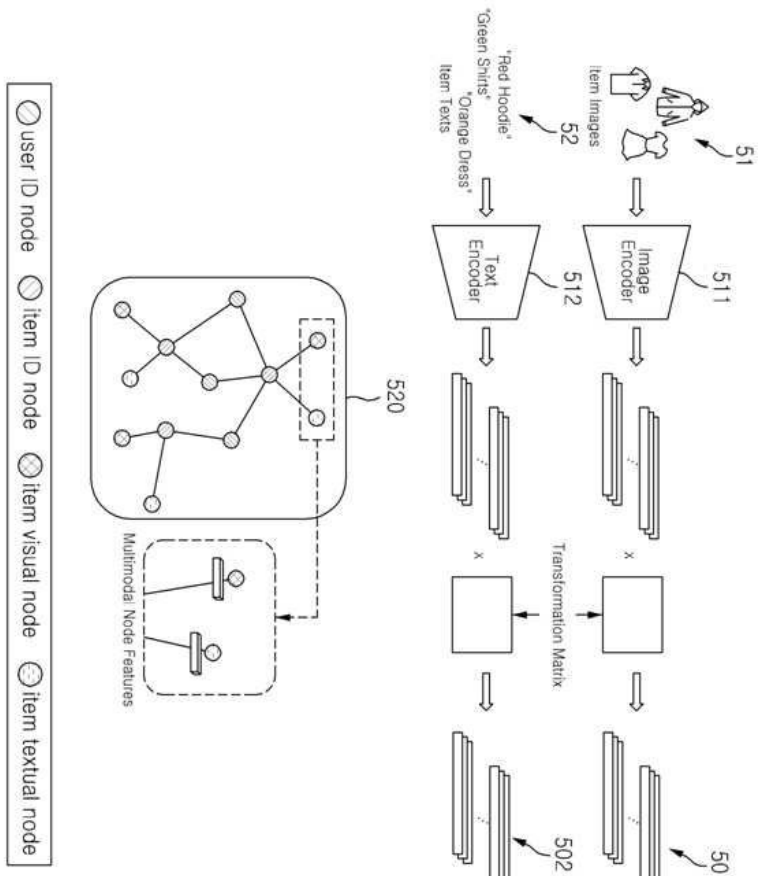
도면3


시 작
사용자와 아이템의 관계를 나타내는 그래프에 각 아이템의 멀티모달 정보를 추가하여 멀티모달 그래프를 생성함
S310
멀티모달 그래프에 대한 임베딩을 수행함
- 멀티모달 정보를 이용한 대조 학습을 수행함
- 멀티모달 그래프의 일관성 정규화 학습을 수행함
S320
종 료

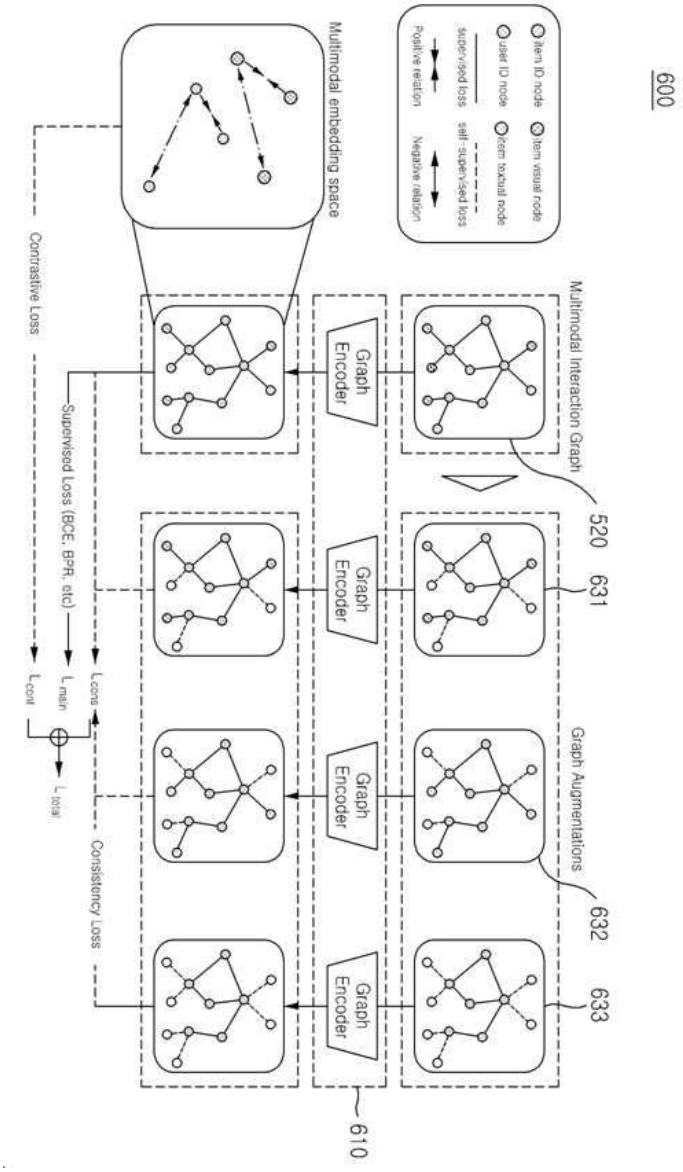
도면4


41
User ID
Item ID
0
24
0
78
1
12
...
...
98
31
99
4
99
67
410
user ID node
item ID node

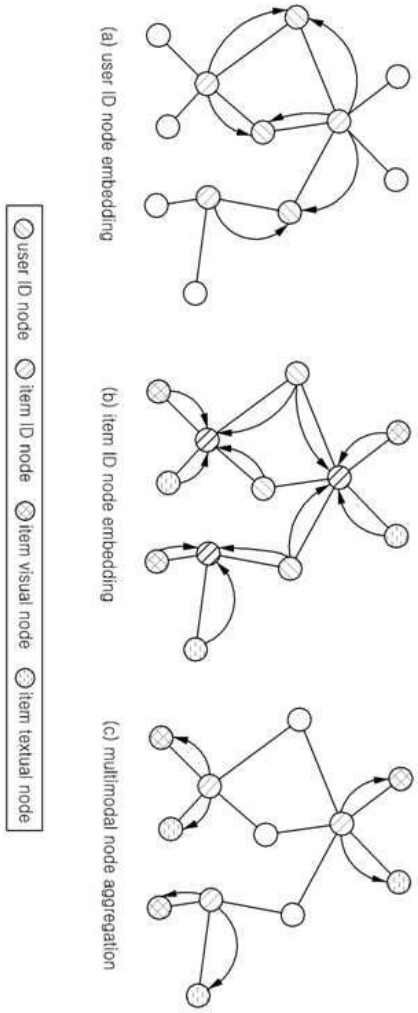
도면5


user ID node
item ID node
item visual node
item textual node
Item Images
"Red Hoodie"
"Green Shirts"
"Orange Dress"
Item Texts
Image Encoder
Text Encoder
Transformation Matrix
Multimodal Node Features
51
52
511
512
520
501
502

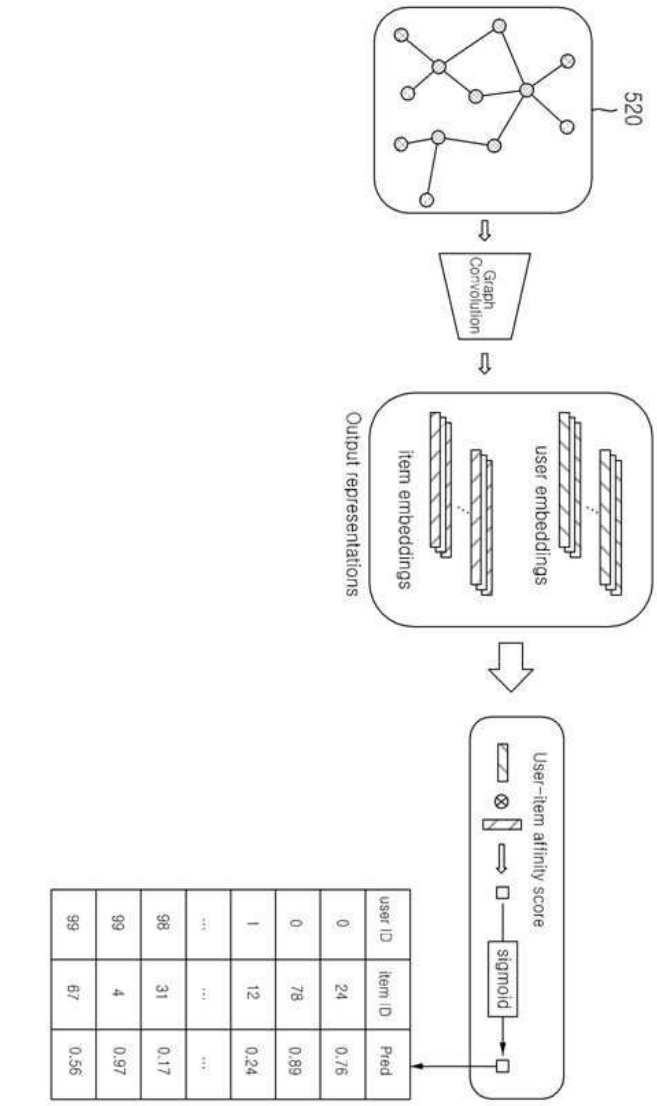
도면6


600
Item ID node
Item visual node
user ID node
Item textual node
supervised loss
self-supervised loss
Positive relation
Negative relation
Multimodal embedding space
Multimodal Interaction Graph
Graph Augmentations
Graph Encoder
520
631
632
633
610
Contrastive Loss
Supervised Loss (BCE, BPR, etc)
Consistency Loss

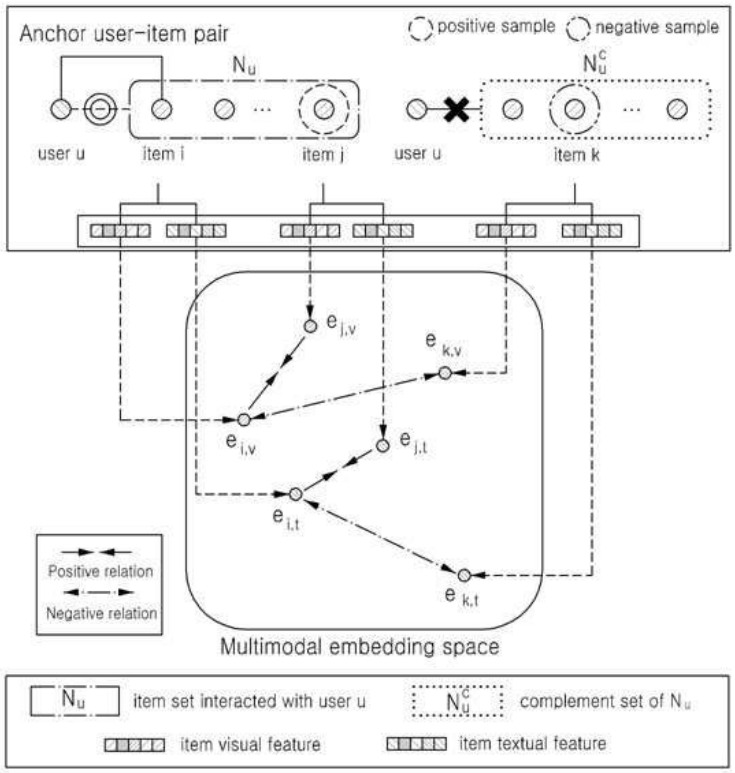
도면7


(a) user ID node embedding
(b) item ID node embedding
(c) multimodal node aggregation
user ID node
item ID node
item visual node
item textual node

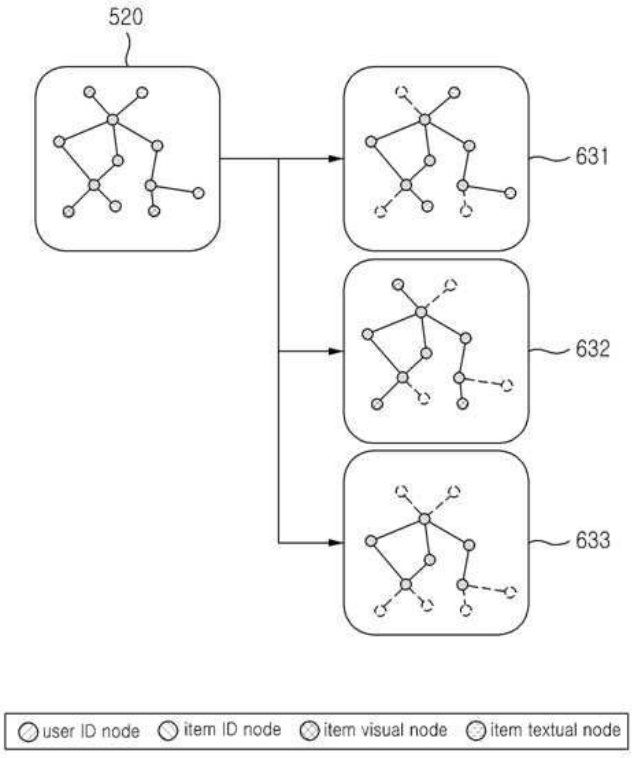
도면8


520
Graph Convolution
Output representations
item embeddings
user embeddings
User-item affinity score
sigmoid
user ID
Item ID
Pred
0
24
0.76
0
78
0.89
1
12
0.24
...
...
...
98
31
0.17
99
4
0.97
99
67
0.56

도면9


Anchor user-item pair
positive sample
negative sample
N_u
N_u^c
user u
item i
item j
user u
item k
$e_{j,v}$
$e_{k,v}$
$e_{i,v}$
$e_{j,t}$
$e_{i,t}$
$e_{k,t}$
Positive relation
Negative relation
Multimodal embedding space
N_u item set interacted with user u
N_u^c complement set of N_u
item visual feature
item textual feature

도면10


520
631
632
633
user ID node
item ID node
item visual node
item textual node

도면11

Algorithm 1 Learning algorithm for SMGNN

Input: Multimodal interaction graph G_M, embedding matrix E prediction function F, a set of training data O.

Initialize all parameters Θ
while until converge do
  res = []
  s = 0 and cont = 0
  for each $(u,i) \in O$ do
    $\hat{y}_{ui} \leftarrow F(G_M, E, (u,i); \Theta)$
    res . append $(\hat{y}_{ui})$
    j ← one randomly sampled item from N_u
    K ← multiple randomly sampled items from N_u^c
    $cont += \left(cl^{(v)}_{(u,i,j,K)} + cl^{(t)}_{(u,i,j,K)}\right)$ (Eq.8, 9)
    for each $n \in \{\{v\}, \{t\}, \{v,t\}\}$ do
      //Perform ModDrop
      $G'_M \leftarrow G_M - \{(i,m) \mid i \in I, m = n\}$
      $\hat{y}_{ui(s)} \leftarrow F(G'_M, E, (u,i); \Theta)$
      res . append($\hat{y}_{ui(s)}$)
      s += 1
  end for
end for
Compute L_{main} using res[0] (Eq.2, 3)
$\tilde{y}_{ui} \leftarrow$ average value of res
Compute L_{cons} using res and $\tilde{y}_{ui}$ (Eq.11, 12)
Compute L_{cont} using cont (Eq.10)
$L_{total} = L_{main} + \lambda_1 L_{cons} + \lambda_2 L_{cont} + \lambda_3 \|\Theta\|_2^2$ (Eq.13)
Updte Θ using L_{total}
end while