



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0083424
(43) 공개일자 2022년06월20일

(51) 국제특허분류(Int. Cl.)
G06F 16/36 (2019.01) G06F 16/28 (2019.01)
G06F 40/279 (2020.01)
(52) CPC특허분류
G06F 16/36 (2019.01)
G06F 16/285 (2019.01)
(21) 출원번호 10-2020-0173653
(22) 출원일자 2020년12월11일
심사청구일자 2020년12월11일

(71) 출원인
경희대학교 산학협력단
경기도 용인시 기흥구 덕영대로 1732 (서천동, 경희대학교 국제캠퍼스내)
(72) 발명자
이영구
경기도 수원시 영통구 청명북로 33, 431동 1704호(영통동, 청명마을삼성아파트)
두영 티 투 반
경기도 용인시 기흥구 서그내로16번길 19, 303호(서천동, 화인빌라)
(74) 대리인
특허법인도담

전체 청구항 수 : 총 10 항

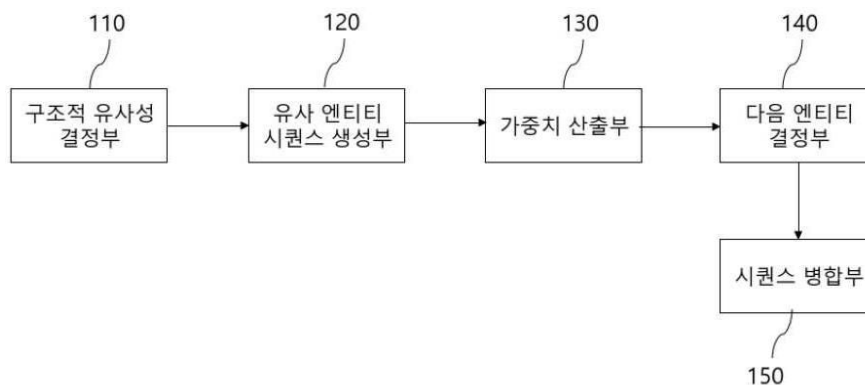
(54) 발명의 명칭 RDF 그래프 임베딩 방법 및 이를 실행하는 장치

(57) 요약

본 발명의 일 실시예에 따른 RDF 그래프 임베딩 장치에서 실행되는 RDF 그래프 임베딩 방법은 RDF 그래프 상의 현재 엔티티에 대해 현재 경로에 따라 상기 현재 엔티티의 속성과 상기 현재 엔티티와 연결된 다른 엔티티의 속성을 비교하는 단계, 상기 비교 결과에 따라 상기 다른 엔티티를 이용하여 유사 엔티티 시퀀스를 생성하는 단계, 상기 유사 엔티티 시퀀스 상의 각각의 후보 엔티티의 가중치를 산출하는 단계, 상기 후보 엔티티의 가중치를 기초로 상기 유사 엔티티 시퀀스 상의 후보 엔티티 중 특정 후보 엔티티를 추출하여 다음 엔티티로 결정하는 단계 및 상기 유사 엔티티를 포함하는 유사 엔티티 시퀀스를 병합하여 병합 엔티티 시퀀스를 생성하는 단계를 포함한다.

대표도

100



(52) CPC특허분류

G06F 40/279 (2020.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711111586
과제번호	2018R1A2A2A05023669
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	개인기초연구(과기정통부)(R&D)
연구과제명	대용량 RDF그래프 요약 및 압축 기반 고속 RDF질의 처리 기술
기 여 율	1/1
과제수행기관명	경희대학교(국제캠퍼스)
연구기간	2020.03.01 ~ 2021.02.28

명세서

청구범위

청구항 1

RDF 그래프 임베딩 장치에서 실행되는 RDF 그래프 임베딩 방법에 있어서,

RDF 그래프 상의 현재 엔티티에 대해 현재 경로에 따라 상기 현재 엔티티의 속성과 상기 현재 엔티티와 연결된 다른 엔티티의 속성을 비교하는 단계;

상기 비교 결과에 따라 상기 다른 엔티티를 이용하여 유사 엔티티 시퀀스를 생성하는 단계;

상기 유사 엔티티 시퀀스 상의 각각의 후보 엔티티의 가중치를 산출하는 단계;

상기 후보 엔티티의 가중치를 기초로 상기 유사 엔티티 시퀀스 상의 후보 엔티티 중 특정 후보 엔티티를 추출하여 다음 엔티티로 결정하는 단계; 및

상기 유사 엔티티를 포함하는 유사 엔티티 시퀀스를 병합하여 병합 엔티티 시퀀스를 생성하는 단계를 포함하는 것을 특징으로 하는

RDF 그래프 임베딩 방법.

청구항 2

제1항에 있어서,

상기 그래프 상의 현재 엔티티에 대해 현재 경로에 따라 상기 현재 엔티티의 속성과 상기 현재 엔티티와 연결된 다른 엔티티의 속성을 비교하는 단계는

상기 현재 엔티티 및 다른 엔티티의 에지 레이블, 아웃고잉 엔티티 및 인커밍 엔티티를 기초로 구조적 유사성을 결정하는 단계를 포함하는 것을 특징으로 하는

RDF 그래프 임베딩 방법.

청구항 3

제2항에 있어서,

상기 그래프 상의 현재 엔티티에 대해 현재 경로에 따라 상기 현재 엔티티의 속성과 상기 현재 엔티티와 연결된 다른 엔티티의 속성을 비교하는 단계는

상기 현재 엔티티의 이전 정점을 기초로 현재 엔티티 및 다른 엔티티 사이의 구조적 유사성 거리를 산출하는 단계를 포함하는 것을 특징으로 하는

RDF 그래프 임베딩 방법.

청구항 4

제1항에 있어서,

상기 유사 엔티티 시퀀스 상의 각각의 후보 엔티티의 가중치를 산출하는 단계는

상기 현재 엔티티 및 다른 엔티티의 에지 레이블, 인커밍의 에지의 수, 아웃고잉의 에지, 상기 후보 엔티티로부터 이동 가능한 엔티티의 수 및 상기 후보 엔티티와 구조적 유사성을 가지는 노드로부터 이동 가능한 엔티티의 수에 따라 상의 각각의 후보 엔티티의 가중치를 산출하는 단계를 포함하는 것을 특징으로 하는

RDF 그래프 임베딩 방법.

청구항 5

제1항에 있어서,

상기 유사 엔티티를 포함하는 유사 엔티티 시퀀스를 병합하여 병합 엔티티 시퀀스를 생성하는 단계는

유사 엔티티 시퀀스 상에 존재하는 엔티티의 수가 동일한 제1 유사 엔티티 시퀀스 및 제2 유사 엔티티 시퀀스 각각에 대해서 엔티티의 수 및 미리 결정된 특징에 속하는 엔티티의 수를 이용하여 제1 행렬 및 제2 행렬을 생성하는 단계;

상기 제1 행렬 및 상기 제2 행렬을 기초로 동일한 행 또는 열의 값에 따라 해당 값을 이용하여 병합 엔티티 시퀀스를 생성하는 단계를 포함하는 것을 특징으로 하는

RDF 그래프 임베딩 방법.

청구항 6

RDF 그래프 임베딩 장치에 있어서,

RDF 그래프 상의 현재 엔티티에 대해 현재 경로에 따라 상기 현재 엔티티의 속성과 상기 현재 엔티티와 연결된 다른 엔티티의 속성을 비교하는 구조적 유사성 결정부;

상기 비교 결과에 따라 상기 다른 엔티티를 이용하여 유사 엔티티 시퀀스를 생성하는 유사 엔티티 시퀀스 생성부;

상기 유사 엔티티 시퀀스 상의 각각의 후보 엔티티의 가중치를 산출하는 가중치 산출부;

상기 후보 엔티티의 가중치를 기초로 상기 유사 엔티티 시퀀스 상의 후보 엔티티 중 특정 후보 엔티티를 추출하고, 상기 특정 후보 엔티티를 다음 엔티티로 결정하는 다음 엔티티 결정부; 및

상기 유사 엔티티를 포함하는 유사 엔티티 시퀀스를 병합하여 병합 엔티티 시퀀스를 생성하는 시퀀스 병합부를 포함하는 것을 특징으로 하는

RDF 그래프 임베딩 장치.

청구항 7

제6항에 있어서,

상기 구조적 유사성 결정부는

상기 현재 엔티티 및 다른 엔티티의 에지 레이블, 아웃고잉 엔티티 및 인커밍 엔티티를 기초로 구조적 유사성을 결정하는 것을 특징으로 하는

RDF 그래프 임베딩 장치.

청구항 8

제7항에 있어서,

상기 구조적 유사성 결정부는

상기 현재 엔티티의 이전 정점을 기초로 현재 엔티티 및 다른 엔티티 사이의 구조적 유사성 거리를 산출하는 것을 특징으로 하는

RDF 그래프 임베딩 장치.

청구항 9

제6항에 있어서,

상기 가중치 산출부는

상기 현재 엔티티 및 다른 엔티티의 에지 레이블, 인커밍의 에지의 수, 아웃고잉의 에지, 상기 후보 엔티티로부터 이동 가능한 엔티티의 수 및 상기 후보 엔티티와 구조적 유사성을 가지는 노드로부터 이동 가능한 엔티티의 수에 따라 상기 각각의 후보 엔티티의 가중치를 산출하는 것을 특징으로 하는

RDF 그래프 임베딩 장치.

청구항 10

제6항에 있어서,

상기 시퀀스 병합부는

유사 엔티티 시퀀스 상에 존재하는 엔티티의 수가 동일한 제1 유사 엔티티 시퀀스 및 제2 유사 엔티티 시퀀스 각각에 대해서 엔티티의 수 및 미리 결정된 특징에 속하는 엔티티의 수를 이용하여 제1 행렬 및 제2 행렬을 생성하고, 상기 제1 행렬 및 상기 제2 행렬을 기초로 동일한 행 또는 열의 값에 따라 해당 값을 이용하여 병합 엔티티 시퀀스를 생성하는 것을 특징으로 하는

RDF 그래프 임베딩 장치.

발명의 설명

기술 분야

- [0001] 본 발명은 RDF 그래프 임베딩 방법 및 이를 실행하는 장치에 관한 것으로, 보다 상세하게는 크기가 N 인 그래프를 낮은 차원의 벡터에 매핑하기 위한 RDF 그래프 임베딩 방법 및 이를 실행하는 장치에 관한 것이다.

배경 기술

- [0003] RDF(Resource Description Framework)는 웹상의 자원의 정보를 표현하기 위한 규격이다. 상이한 메타데이터 간의 어의, 구문 및 구조에 대한 공통적인 규칙을 지원한다.
- [0004] 웹상에 존재하는 기계 해독형(machineunderstandable) 정보를 교환하기 위하여 월드와이드 웹 컨소시엄에서 제안한 것으로, 메타데이터 간의 효율적인 교환 및 상호호환을 목적으로 한다.
- [0005] 메타데이터 교환을 위해서 명확하고 구조화된 의미표현을 제공해 주는 공통의 기술언어로 XML(eXtensible Markup Language)을 사용하기도 한다. RDF는 데이터 모형, 데이터의 상호교환을 위한 구문, 스키마 모형, 기계 해독형 스키마를 위한 구문, 질문과 프로파일 프로토콜과 같은 요소로 구성된다.
- [0006] RDF는 주어, 술어, 목적어(subject, predicate, object)로 된 그래프 형태로 표현된다. RDF 그래프는 기존의 테이블 기반으로 한 데이터베이스와 구조가 다르지만, 구현의 편리성, 안정된 결과 등을 이유로 RDF 그래프를 저장하기 위하여 기존의 데이터베이스를 그대로 사용하고 있다.
- [0007] RDF(Resource Description Framework)는 시멘틱 웹을 표현하는 표준 모형이다. RDF는 RDF 트리플로 이루어져 있다. RDF 트리플은 주어(subject), 술어(predicate), 목적어(object)로 이루어지며, 공백으로 구분된다. 주어는 IRI(Internationalized Resource Identifier)나 블랭크 노드(Blank node), 술어는 IRI, 목적어는 IRI, 블랭크 노드나 리터럴(literal)이 될 수 있다. RDF 그래프는 RDF 트리플의 집합으로 데이터 셋을 이룬다.

- [0008] 이러한 트리플 형태의 RDF 데이터는 주어와 목적어를 정점(vertex)으로 표현하고, 술어를 정점 간의 관계인 간선(edge)으로 표현하여 그래프로 변환이 가능하다.
- [0009] RDF 데이터를 그래프 형태로 관리하면 질의 처리 시 정점의 연결 정보로 다른 정점을 접근할 수 있어서 조인 연산을 줄여 빠른 질의 응답을 제공할 수 있다.
- [0010] 이러한 RDF 그래프의 경우 엔티티를 RDF2vec, biaRDF2vec과 같은 RDF 그래프의 벡터로 변환하는 기존의 많은 작업도 제안되었다. 하지만, 일반 그래프와 마찬가지로 입력이 시퀀스인 모델링 기술로 word2vec를 사용한다.
- [0011] 그러나 RDF 그래프 임베딩에는 몇 가지 어려움이 있다.
- [0012] 첫째, graph walk는 가능한 모든 시퀀스를 생성 할 수 있다. 그러나 엄청난 양의 시퀀스가 생성되지만, 실제로, 훈련을 위해 엄청난 양의 시퀀스들을 모두 사용할 수는 없다.
- [0013] 둘째, 일반 그래프 문제와 달리 RDF 엔티티는 동일한 의미를 가질 때 유사하다는 문제점이 있다. 하지만, RDF 그래프의 유형 예지 때문에 graph walk로 일부 잠재 시퀀스를 생성 할 수 없는 문제가 발생한다.
- [0014] 셋째, 리터럴 엔티티를 가진 RDF 그래프에 그래프 임베딩을 적용할 때, RDF 데이터 세트의 리터럴 유형은 문자열, 숫자, 날짜 등과 같이 값이 매우 다양하기 때문에 이를 식별하는 것이 어렵다는 문제점이 있다.

발명의 내용

해결하려는 과제

- [0016] 본 발명은 크기가 N 인 그래프를 낮은 차원의 벡터에 매핑하기 위한 RDF 그래프 임베딩 방법 및 이를 실행하는 장치를 제공하는 것을 목적으로 한다.
- [0017] 또한, 본 발명은 그래프 임베딩을 사용하여 분류, 클러스터링 또는 회귀와 같은 모든 머신 러닝 기술을 적용 할 수 있도록 하는 RDF 그래프 임베딩 방법 및 이를 실행하는 장치를 제공하는 것을 목적으로 한다.
- [0018] 또한, 본 발명은 대략적인 쿼리 응답을 위해 유사한 엔티티를 사용하여 쿼리에 빠르게 응답할 수 있도록 하는 RDF 그래프 임베딩 방법 및 이를 실행하는 장치를 제공하는 것을 목적으로 한다.
- [0020] 본 발명의 목적들은 이상에서 언급한 목적으로 제한되지 않으며, 언급되지 않은 본 발명의 다른 목적 및 장점들은 하기의 설명에 의해서 이해될 수 있고, 본 발명의 실시예에 의해 보다 분명하게 이해될 것이다. 또한, 본 발명의 목적 및 장점들은 특허 청구 범위에 나타난 수단 및 그 조합에 의해 실현될 수 있음을 쉽게 알 수 있을 것이다.

과제의 해결 수단

- [0022] 이러한 목적을 달성하기 위한 RDF 그래프 임베딩 장치에서 실행되는 RDF 그래프 임베딩 방법은 RDF 그래프 상의 현재 엔티티에 대해 현재 경로에 따라 상기 현재 엔티티의 속성과 상기 현재 엔티티와 연결된 다른 엔티티의 속성을 비교하는 단계, 상기 비교 결과에 따라 상기 다른 엔티티를 이용하여 유사 엔티티 시퀀스를 생성하는 단계, 상기 유사 엔티티 시퀀스 상의 각각의 후보 엔티티의 가중치를 산출하는 단계, 상기 후보 엔티티의 가중치를 기초로 상기 유사 엔티티 시퀀스 상의 후보 엔티티 중 특정 후보 엔티티를 추출하여 다음 엔티티로 결정하는 단계 및 상기 유사 엔티티를 포함하는 유사 엔티티 시퀀스를 병합하여 병합 엔티티 시퀀스를 생성하는 단계를 포함한다.
- [0023] 또한 이러한 목적을 달성하기 위한 RDF 그래프 임베딩 장치는 RDF 그래프 상의 현재 엔티티에 대해 현재 경로에 따라 상기 현재 엔티티의 속성과 상기 현재 엔티티와 연결된 다른 엔티티의 속성을 비교하는 구조적 유사성 결정부, 상기 비교 결과에 따라 상기 다른 엔티티를 이용하여 유사 엔티티 시퀀스를 생성하는 유사 엔티티 시퀀스 생성부, 상기 유사 엔티티 시퀀스 상의 각각의 후보 엔티티의 가중치를 산출하는 가중치 산출부, 상기 후보 엔티티의 가중치를 기초로 상기 유사 엔티티 시퀀스 상의 후보 엔티티 중 특정 후보 엔티티를 추출하고, 상기 특정 후보 엔티티를 다음 엔티티로 결정하는 다음 엔티티 결정부 및 상기 유사 엔티티를 포함하는 유사 엔티티 시

퀵스를 병합하여 병합 엔티티 시퀀스를 생성하는 시퀀스 병합부를 포함한다.

발명의 효과

- [0025] 전술한 바와 같은 본 발명에 의하면, 크기가 N 인 그래프를 낮은 차원의 벡터에 매핑할 수 있다는 장점이 있다.
- [0026] 또한 본 발명에 의하면, 그래프 임베딩을 사용하여 분류, 클러스터링 또는 회귀와 같은 모든 머신 러닝 기술을 적용 할 수 있다는 장점이 있다.
- [0027] 또한 본 발명에 의하면, 대략적인 쿼리 응답을 위해 유사한 엔티티를 사용하여 쿼리에 빠르게 응답할 수 있다는 장점이 있다.

도면의 간단한 설명

- [0029] 도 1은 본 발명의 일 실시예에 따른 RDF 그래프 임베딩 장치를 설명하기 위한 블록도이다.
- 도 2는 본 발명에 따른 RDF 그래프 임베딩 방법의 일 실시예를 설명하기 위한 흐름도이다.
- 도 3 및 도 4는 본 발명에 따른 RDF 그래프 임베딩 과정을 설명하기 위한 예시도이다.

발명을 실시하기 위한 구체적인 내용

- [0030] 전술한 목적, 특징 및 장점은 첨부된 도면을 참조하여 상세하게 후술되며, 이에 따라 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 본 발명의 기술적 사상을 용이하게 실시할 수 있을 것이다. 본 발명을 설명함에 있어서 본 발명과 관련된 공지 기술에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 상세한 설명을 생략한다. 이하, 첨부된 도면을 참조하여 본 발명에 따른 바람직한 실시예를 상세히 설명하기로 한다. 도면에서 동일한 참조부호는 동일 또는 유사한 구성요소를 가리키는 것으로 사용된다.
- [0032] 본 명세서에서 사용된 용어 중 "엔티티"는 RDF 그래프를 구성하는 각각의 노드를 의미한다.
- [0033] 본 명세서에서 사용된 용어 중 "아웃고잉 엔티티"은 현재 엔티티를 기준으로 에지의 화살표가 가리키는 엔티티를 의미한다.
- [0034] 본 명세서에서 사용된 용어 중 "인커밍 엔티티"는 현재 엔티티를 기준으로 에지의 화살표가 가리키는 곳의 반대에 해당하는 엔티티를 의미한다.
- [0037] 도 1은 본 발명의 일 실시예에 따른 RDF 그래프 임베딩 장치를 설명하기 위한 블록도이다.
- [0038] 도 1을 참조하면, RDF 그래프 임베딩 장치(100)는 구조적 유사성 결정부(110), 유사 엔티티 시퀀스 생성부(120), 가중치 산출부(130), 다음 엔티티 결정부(140) 및 시퀀스 병합부(150)를 포함한다.
- [0039] 구조적 유사성 결정부(110)는 현재 시퀀스의 이전 엔티티를 기반으로 다음 엔티티를 선택한다.
- [0040] 이를 위해, 구조적 유사성 결정부(110)는 현재 시퀀스에서 이전 엔티티를 기반으로 다음 엔티티가 어떤 엔티티 인지 결정하기 위해 현재 엔티티를 사용하여 유사한 경로의 수를 나타내는 가중치를 산출한다.
- [0041] 본 발명은 유사한 경로의 수를 세기 위하여 그래프에서 유사한 엔티티를 정의하는 RDF 그래프 임베딩 방법을 제안한다. 다른 유사성 측정 방법과 달리, 본 발명은 연결 레이블과 이웃 사이의 균형을 유지하며, RDF 그래프에서 서로 연결된 연결된 엔티티 사이의 속성을 이용하여 유사성을 측정한다.
- [0042] 일 실시예에서, 구조적 유사성 결정부(110)는 그래프에서 [수학식 1]을 기초로 현재 엔티티 및 다른 엔티티 사이의 구조적 유사성을 결정할 수 있다. 이때, 현재 엔티티 및 다른 엔티티 사이의 구조적 유사성은 전체 연결에 대한 유사한 연결 비율과 전체 이웃에 대한 유사한 이웃 비율의 합으로 결정된다.

[0044] [수학식 1]

$$s(u,v) = w \frac{|predicate(u,v)|}{|totalPredicate(u,v)|} + (1-w) \frac{|neighbor(u,v)|}{|totalNeighbor(u,v)|}$$

[0045]

[0047] $s(u,v)$: 현재 엔티티 u 및 다른 엔티티 v 의 구조적 유사성,

[0048] $|predicate(u,v)|$: 현재 엔티티 u 및 다른 엔티티 v 의 유사한 에지 레이블의 수,

[0049] $|neighbor(u,v)|$: 현재 엔티티 u 및 다른 엔티티 v 의 유사한 에지(즉, 동일한 레이블 에지와 인커밍 또는 아웃고잉 정점)의 수,

[0050] $|totalPredicate(u,v)|$: 현재 엔티티 u 및 다른 엔티티 v 의 에지의 전체 인커밍 레이블 및 아웃고잉 레이블의 수,

[0051] $|totalNeighbor(u,v)|$: 현재 엔티티 u 및 다른 엔티티 v 의 인커밍 에지 및 아웃고잉 에지의 수,

[0052] w : 우선 순위를 조정하기 위한 가중치

[0054] 즉, 구조적 유사성 결정부(110)는 상기의 [수학식 1]를 기초로 현재 엔티티 u 및 다른 엔티티 v 의 구조적 유사성을 에지 레이블, 아웃고잉 엔티티 및 인커밍 엔티티를 기초로 결정된다.

[0056] [수학식 2]

$$f(u,v)\{a\} = \frac{\sum_i^n s(a,b_i)}{n} + s(u,v)$$

[0057]

[0059] $f(u,v)\{a\}$: 현재 엔티티 u 및 다른 엔티티 v 사이의 구조적 유사성 거리,

[0060] $s(u,v)$: [수학식 1]에서 산출된 현재 엔티티 u 및 다른 엔티티 v 의 구조적 유사성,

[0061] b_i : 다른 엔티티 v 의 인커밍 엔티티,

[0062] a : 현재 엔티티 u 의 이전 엔티티,

[0063] n : 각각의 엔티티의 이동 수

[0065] 그런 다음, 유사 엔티티 시퀀스 생성부(120)는 상기의 [수학식 2]를 기초로 그래프의 각 엔티티에 대해 구조적 유사성 거리를 산출하고, 구조적 유사성 거리가 특정 거리 이하인 다른 엔티티 v 를 이용하여 유사 엔티티 시퀀스를 생성한다.

[0066] 즉, 유사 엔티티 시퀀스 생성부(120)는 상기의 [수학식 2]를 기초로 이전 엔티티 a 에 대해서 현재 엔티티 u 와 구조적 유사성을 가지는 정점을 결정할 수 있다.

[0067] 이때, 유사 엔티티 시퀀스 생성부(120)가 유사 엔티티 시퀀스를 추출하는 이유는 유사 엔티티 시퀀스 상의 후보

엔티티 중 어느 하나의 엔티티를 다음 엔티티로 결정하기 위해서이다. 하지만, 유사 엔티티 시퀀스 상의 존재하는 후보 엔티티가 많기 때문에 유사 엔티티 시퀀스 상의 후보 엔티티의 가중치를 산출하여 가중치에 따라 어느 하나의 후보 엔티티를 다음 엔티티로 결정한다.

[0068] 이를 위해, 가중치 산출부(130)는 [수학식 3]을 기초로 상기 유사 엔티티 시퀀스 상의 후보 엔티티의 가중치를 산출한다.

[0070] [수학식 3]

$$w(p,q)=\frac{M(p|q)}{|totalNeighbor(p)|+|totalPredicate(p)|}$$

[0073] $w(p,q)$: 후보 엔티티 p가 다음 엔티티 q로 이동하는 가중치,

[0074] $|totalPredicate(u,v)|$: 현재 엔티티 u 및 다른 엔티티 v의 에지의 총 레이블의 식별자의 수,

[0075] $|totalNeighbor(u,v)|$: 현재 엔티티 u 및 다른 엔티티 v의 인커밍의 에지의 수 및 아웃고잉의 에지의 수의 합,

[0076] $|M(p|q)|$: 유사 엔티티 시퀀스 중 후보 엔티티 p 또는 후보 엔티티 p가 다음 엔티티 q로 결정될 모든 경우의 수를 지시하며, [수학식 4]에 의해 산출

[0078] [수학식 4]

$$M(p|q) = |edge(sc(p,a),q,any)| + |edge(sc(p,a),any,q)|$$

[0081] $|M(p|q)|$: 유사 엔티티 시퀀스 중 후보 엔티티 P 또는 후보 엔티티 P가 다음 엔티티 q로 결정될 경우의 수,

[0082] $|edge(sc(p,a),q,any)|$: sc(p, a)에서 후보 엔티티 P를 통과하는 엔티티 a로부터 다음 엔티티 q를 연결하는 에지

[0084] [수학식 3]에서 가중치가 높은 경우 엔티티 p가 다음 엔티티 q로 이동할 가능성이 높으며, 가중치가 낮은 경우 엔티티 p가 엔티티 q로 이동할 가능성이 낮은 것이다.

[0085] 그런 다음, 다음 엔티티 결정부(140)는 후보 엔티티의 가중치를 기초로 유사 엔티티 시퀀스 상의 후보 엔티티 중 특정 후보 엔티티를 추출하고, 특정 엔티티를 다음 엔티티로 결정한다.

[0086] 시퀀스 병합부(150)는 동일한 깊이를 가지는 두 개의 스퀀스를 병합하여 병합 시퀀스를 생성한다.

[0087] 이를 위해, 시퀀스 병합부(150)는 유사 엔티티 시퀀스 상에 존재하는 엔티티의 수가 동일한 제1 유사 엔티티 시퀀스 및 제2 유사 엔티티 시퀀스 각각에 대해서 엔티티의 수 및 미리 결정된 특징에 속하는 엔티티의 수를 이용하여 제1 행렬 및 제2 행렬을 생

- [0089] 도 2는 본 발명에 따른 RDF 그래프 임베딩 방법의 일 실시예를 설명하기 위한 흐름도이다.
- [0090] 도 2를 참조하면, RDF 그래프 임베딩 장치(100)는 RDF 그래프 상의 현재 엔티티에 대해 현재 경로에 따라 상기 현재 엔티티의 속성과 상기 현재 엔티티와 연결된 다른 엔티티의 속성을 비교한다(단계 S210).
- [0091] 단계 S210에 대한 일 실시예에서, RDF 그래프 임베딩 장치(100)는 현재 엔티티 및 다른 엔티티의 에지 레이블, 아웃고잉 엔티티 및 인커밍 엔티티를 기초로 구조적 유사성을 결정할 수 있다.
- [0092] RDF 그래프 임베딩 장치(100)는 비교 결과에 따라 상기 다른 엔티티를 이용하여 유사 엔티티 시퀀스를 생성한다(단계 S220).
- [0093] 단계 S220에 대한 일 실시예에서, RDF 그래프 임베딩 장치(100)는 현재 엔티티의 이전 정점을 기초로 현재 엔티티 및 다른 엔티티 사이의 구조적 유사성 거리를 산출한 후, 특정 거리 이하의 다른 엔티티를 이용하여 유사 엔티티 시퀀스를 생성한다.
- [0094] RDF 그래프 임베딩 장치(100)는 유사 엔티티 시퀀스 상의 각각의 후보 엔티티의 가중치를 산출한다(단계 S230).
- [0095] 단계 S230에 대한 일 실시예에서, RDF 그래프 임베딩 장치(100)는 현재 엔티티 및 다른 엔티티의 에지 레이블, 인커밍의 에지의 수, 아웃고잉의 에지, 상기 후보 엔티티로부터 이동 가능한 엔티티의 수 및 상기 후보 엔티티와 구조적 유사성을 가지는 노드로부터 이동 가능한 엔티티의 수에 따라 상의 각각의 후보 엔티티의 가중치를 산출할 수 있다.
- [0096] RDF 그래프 임베딩 장치(100)는 상기 후보 엔티티의 가중치를 기초로 상기 유사 엔티티 시퀀스 상의 후보 엔티티 중 특정 후보 엔티티를 추출하여 다음 엔티티로 결정한다(단계 S240).
- [0097] RDF 그래프 임베딩 장치(100)는 유사 엔티티를 포함하는 유사 엔티티 시퀀스를 병합하여 병합 엔티티 시퀀스를 생성한다(단계 S250).
- [0098] 단계 S250에 대한 일 실시예에서, RDF 그래프 임베딩 장치(100)는 유사시퀀스 시퀀스 상에 존재하는 엔티티의 수가 동일한 제1 유사 엔티티 시퀀스 및 제2 유사 엔티티 시퀀스 각각에 대해서 엔티티의 수 및 미리 결정된 특징에 속하는 엔티티의 수를 이용하여 제1 행렬 및 제2 행렬을 생성한다.
- [0099] 그런 다음, RDF 그래프 임베딩 장치(100)는 제1 행렬 및 상기 제2 행렬을 기초로 동일한 행 또는 열의 값에 따라 해당 값을 이용하여 병합 엔티티 시퀀스를 생성한다.
- [0101] 도 3 및 도 4는 본 발명에 따른 RDF 그래프 임베딩 과정을 설명하기 위한 예시도이다.
- [0102] 도 3을 참조하여 그래프 G4에서 엔티티 A1 및 엔티티 A2의 이웃 엔티티가 유사하기 때문에 엔티티 A1 및 엔티티 A2는 유사한 구조성을 가진다. 하지만, 모든 유사 엔티티 시퀀스가 생성되지 않은 경우, 특히 A1 → 1 → B1 → 3 → D1 → 6 → F1 경로 대신에 A1 → 1 → B1 → 3 → D1 → 8 → H1이 생성될 수 있다.
- [0103] 따라서, 본 발명은 현재 엔티티의 이전 엔티티를 기초로 다음 노드를 선택할 수 있는 유사 엔티티 시퀀스를 생성한다. 순차적으로, 다음 노드가 어떤 노드인지 결정하기 위해, 본 발명은 현재 노드에 대한 유사한 구조성을 가지는 유사 엔티티 시퀀스 상의 각각의 후보 엔티티의 가중치를 산출한다. 이하에서는 도 4를 참조하여 이러한 과정을 설명하기로 한다.
- [0104] 도 4를 참조하면, RDF 그래프 임베딩 장치(100)는 현재 시퀀스의 이전 엔티티를 기반으로 다음 엔티티를 선택한다.
- [0105] 이를 위해, RDF 그래프 임베딩 장치(100)는 현재 시퀀스에서 이전 엔티티를 기반으로 다음 엔티티가 어떤 엔티티인지 결정하기 위해 현재 엔티티 u 및 현재 엔티티 u와 연결된 다른 엔티티 v의 속성을 비교하는 유사한 구조성을 가지는지 확인한다.
- [0106] 이때, RDF 그래프 임베딩 장치(100)는 현재 엔티티 및 다른 엔티티의 에지 레이블, 아웃고잉 엔티티 및 인커밍 엔티티를 기초로 구조적 유사성을 결정할 수 있다.
- [0107] RDF 그래프 임베딩 장치(100)는 현재 엔티티 u 및 현재 엔티티 u와 연결된 다른 엔티티 v의 속성을 기초로 유사한 구조성을 가지는 다른 엔티티 v를 이용하여 유사 엔티티 시퀀스(v1, v2, ..., vn)을 생성한다.
- [0108] 이때, RDF 그래프 임베딩 장치(100)는 현재 엔티티의 이전 정점을 기초로 현재 엔티티 및 다른 엔티티 사이의

구조적 유사성 거리를 산출하고, 구조적 유사성 거리에 따라 다른 엔티티 v 를 이용하여 유사 엔티티 시퀀스($v_1, v_2, \dots, v_n$)을 생성한다.

- [0109] 즉, RDF 그래프 임베딩 장치(100)는 구조적 유사성 거리가 특정 거리 이하인 다른 엔티티 v 를 이용하여 유사 엔티티 시퀀스($v_1, v_2, \dots, v_n$)을 생성할 수 있다.
- [0110] 그런 다음, RDF 그래프 임베딩 장치(100)는 유사 엔티티 시퀀스($v_1, v_2, \dots, v_n$) 상의 각각의 후보 엔티티의 가중치를 산출한다.
- [0111] 예를 들어, RDF 그래프 임베딩 장치(100)는 유사 엔티티 시퀀스($v_1, v_2, \dots, v_n$) 상의 각각의 후보 엔티티 q_1 를 추출한 후, 상기의 [수학식 3]을 기초로 현재 엔티티 u 가 다음 엔티티 q_1 으로 이동 가능한 가중치 $w(u, q_1)$ 를 산출한다.
- [0112] 다른 예를 들어, RDF 그래프 임베딩 장치(100)는 유사 엔티티 시퀀스($v_1, v_2, \dots, v_n$) 상의 각각의 후보 엔티티 q_2 를 추출한 후, 상기의 [수학식 3]을 기초로 현재 엔티티 u 가 다음 엔티티 q_2 으로 이동 가능한 가중치 $w(u, q_2)$ 를 산출한다.
- [0113] 상기 후보 엔티티 $q_1 \sim q_n$ 의 가중치를 기초로 상기 유사 엔티티 시퀀스 상의 후보 엔티티 중 특정 후보 엔티티를 추출하고, 상기 특정 후보 엔티티를 다음 엔티티로 결정한다.
- [0114] 마지막으로, RDF 그래프 임베딩 장치(100)는 유사 엔티티를 포함하는 유사 엔티티 시퀀스를 병합하여 병합 엔티티 시퀀스를 생성한다.
- [0115] 예를 들어, RDF 그래프 임베딩 장치(100)는 동일한 깊이 d 를 갖는 제1 유사 엔티티 시퀀스 s 및 제2 유사 엔티티 시퀀스 s' 그리고 $s = \langle v_1 \rightarrow p_1 \rightarrow v_2 \rightarrow \dots \rightarrow v_d \rangle$ 및 $s' = \langle v'_1 \rightarrow p'_1 \rightarrow v \rangle$ 인 특징을 F 를 통해 병합 엔티티 시퀀스를 생성한다.
- [0116] 이를 위해, RDF 그래프 임베딩 장치(100)는 유사 엔티티 시퀀스 상에 존재하는 엔티티의 수가 동일한 제1 유사 엔티티 시퀀스 s 및 제2 유사 엔티티 시퀀스 s' 각각에 대해서 엔티티의 수 및 미리 결정된 특징에 속하는 엔티티의 수를 이용하여 제1 행렬 및 제2 행렬을 생성한다.
- [0117] 그런 다음, RDF 그래프 임베딩 장치(100)는 제1 행렬 및 상기 제2 행렬을 기초로 동일한 행 또는 열의 값에 따라 해당 값을 이용하여 병합 엔티티 시퀀스를 생성한다.
- [0119] 한정된 실시예와 도면에 의해 설명되었으나, 본 발명은 상기의 실시예에 한정되는 것은 아니며, 이는 본 발명이 속하는 분야에서 통상의 지식을 가진 자라면 이러한 기재로부터 다양한 수정 및 변형이 가능하다. 따라서, 본 발명 사상은 아래에 기재된 특허청구범위에 의해서만 파악되어야 하고, 이의 균등 또는 등가적 변형 모두는 본 발명 사상의 범주에 속한다고 할 것이다.

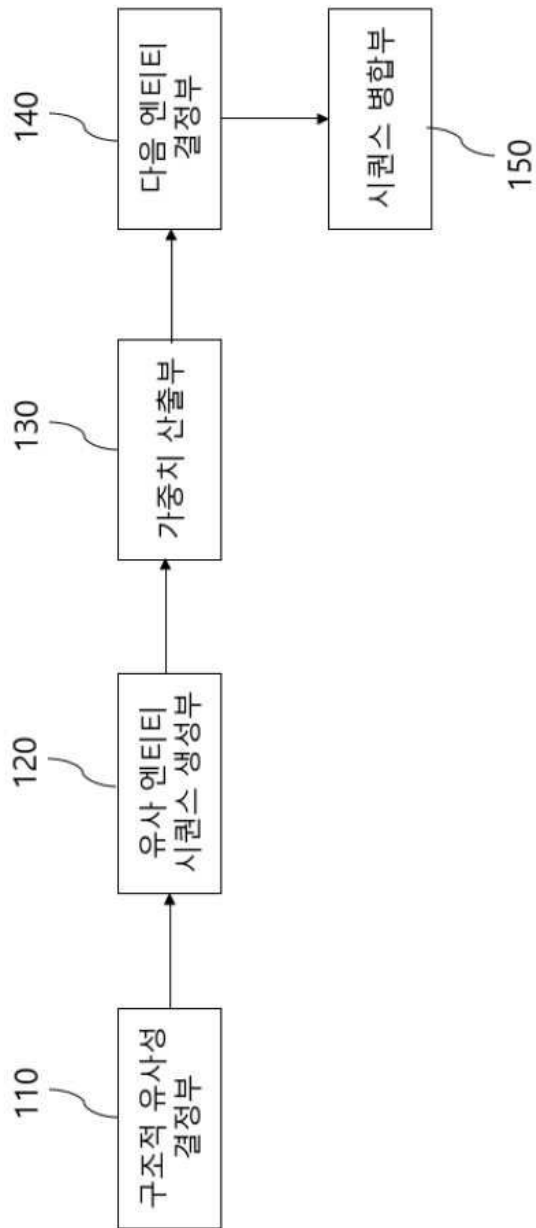
부호의 설명

- [0121] 100: RDF 그래프 임베딩 장치,
110: 구조적 유사성 결정부,
120: 유사 엔티티 시퀀스 생성부,
130: 가중치 산출부
140: 다음 엔티티 결정부,
150: 시퀀스 병합부

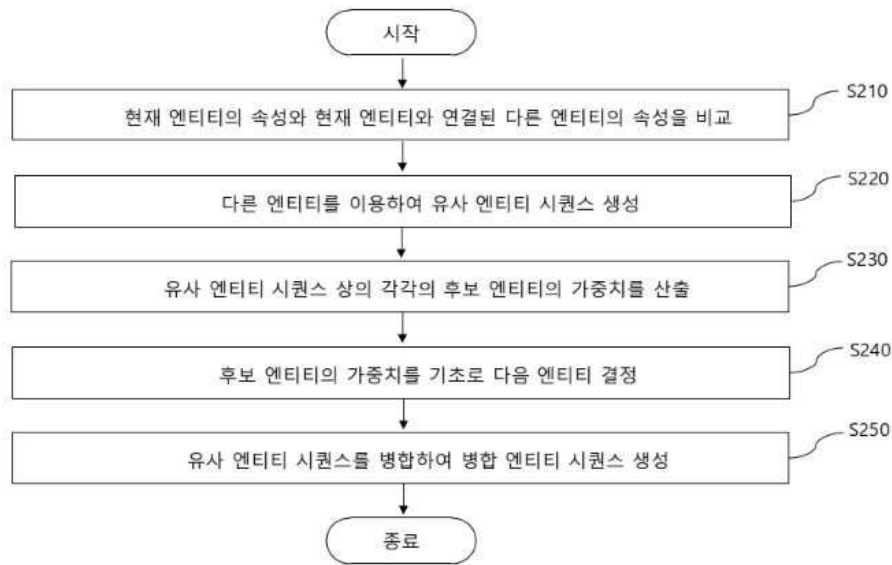
도면

도면1

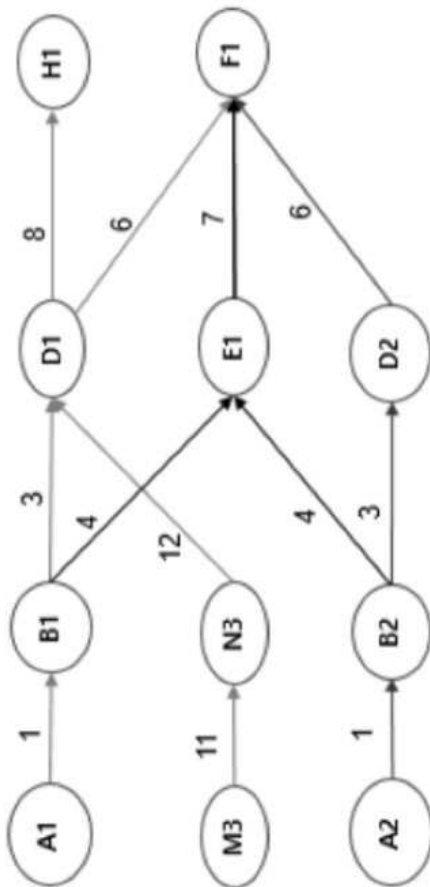
100



도면2



도면3



도면4

