



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0161496
(43) 공개일자 2024년11월12일

(51) 국제특허분류(Int. Cl.)
G06N 3/08 (2023.01) G06F 17/16 (2006.01)
G06N 3/04 (2023.01)
(52) CPC특허분류
G06N 3/082 (2023.01)
G06F 17/16 (2013.01)
(21) 출원번호 10-2023-0058697
(22) 출원일자 2023년05월04일
심사청구일자 2023년05월04일

(71) 출원인
경희대학교 산학협력단
경기도 용인시 기흥구 덕영대로 1732 (서천동, 경희대학교 국제캠퍼스내)
(72) 발명자
이영구
경기도 수원시 영통구 청명북로 33, 431동 1704호 (영통동, 청명마을삼성아파트)
앰디 콜람 모세드
경기도 수원시 영통구 청명로59번길 53, B03호 (영통동)
(74) 대리인
두호특허법인

전체 청구항 수 : 총 18 항

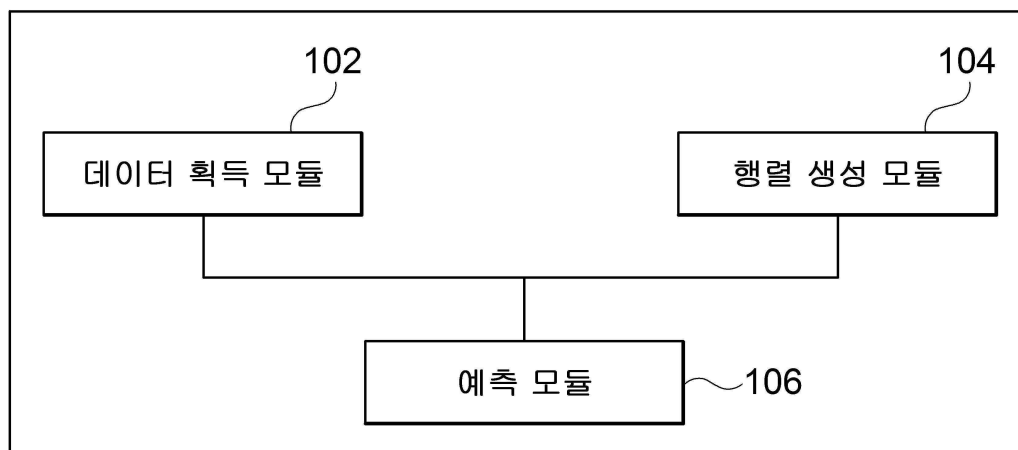
(54) 발명의 명칭 그래프의 링크 예측 장치 및 방법

(57) 요약

그래프의 링크 예측 장치 및 방법이 개시된다. 개시되는 일 실시예에 따른 그래프의 링크 예측 장치는, 그래프 데이터를 수집하는 데이터 획득 모듈, 그래프 데이터로부터 노드 특징 행렬, 인접 행렬, 및 에지 특징 행렬을 각각 생성하는 행렬 생성 모듈, 및 노드 특징 행렬, 인접 행렬, 및 에지 특징 행렬을 그래프 신경망 모델에 입력하여 해당 그래프에 대한 링크 예측을 하도록 그래프 신경망 모델을 학습하는 예측 모듈을 포함한다.

대표도 - 도1

100



(52) CPC특허분류
G06N 3/04 (2023.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711193818
과제번호	2021-0-00859-003
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	SW컴퓨팅산업원천기술개발(R&D, 정보화)
연구과제명	초거대 그래프의 지능적 고속 처리를 위한 그래프 DBMS 기술 개발
기 여 율	1/1
과제수행기관명	포항공과대학교 산학협력단
연구기간	2023.01.01 ~ 2023.12.31

명세서

청구범위

청구항 1

하나 이상의 프로세서들, 및

상기 하나 이상의 프로세서들에 의해 실행되는 하나 이상의 프로그램들을 저장하는 메모리를 구비한 그래프의 링크 예측 장치로서,

그래프 데이터를 수집하는 데이터 획득 모듈;

상기 그래프 데이터로부터 노드 특징 행렬, 인접 행렬, 및 에지 특징 행렬을 각각 생성하는 행렬 생성 모듈; 및

상기 노드 특징 행렬, 상기 인접 행렬, 및 상기 에지 특징 행렬을 그래프 신경망 모델에 입력하여 해당 그래프에 대한 링크 예측을 하도록 상기 그래프 신경망 모델을 학습하는 예측 모듈을 포함하는, 링크 예측 장치.

청구항 2

청구항 1에 있어서,

상기 그래프 데이터는, 복수 개의 노드 및 노드들 간의 연결 관계인 에지를 포함하며,

상기 행렬 생성 모듈은,

상기 그래프 데이터의 각 노드들로부터 특징을 추출하여 노드 특징 벡터로 표현하고 상기 노드 특징 벡터들의 집합을 통해 상기 노드 특징 행렬을 구성하며,

상기 그래프 데이터의 각 노드들의 연결 관계를 인접 특징 벡터로 표현하고 상기 인접 특징 벡터들의 집합을 통해 상기 인접 행렬을 구성하며,

상기 그래프 데이터의 각 에지들의 특징을 에지 특징 벡터로 표현하고 상기 에지 특징 벡터들의 집합을 통해 상기 에지 특징 행렬을 구성하는, 링크 예측 장치.

청구항 3

청구항 2에 있어서,

상기 행렬 생성 모듈은,

상기 그래프 데이터에서 노드 간 공통 이웃에 대한 정보를 기초로 에지 특징 벡터를 추출하여 제1 에지 특징 행렬을 구성하고,

상기 그래프 데이터에서 노드 간 최단 경로에 대한 정보를 기초로 에지 특징 벡터를 추출하여 제2 에지 특징 행렬을 구성하며,

상기 그래프 데이터에서 노드 간 연결 강도를 기초로 에지 특징 벡터를 추출하여 제3 에지 특징 행렬을 구성하는, 링크 예측 장치.

청구항 4

청구항 3에 있어서,

상기 에지 특징 행렬은,

상기 제1 에지 특징 행렬, 상기 제2 에지 특징 행렬, 및 상기 제3 에지 특징 행렬 중 둘 이상을 포함하여 다차

원 행렬로 구성되며, 상기 다차원 행렬의 에지 특징 행렬이 상기 노드 특징 행렬 및 상기 인접 행렬과 함께 상기 그래프 신경망 모델로 입력되는, 링크 예측 장치.

청구항 5

청구항 4에 있어서,

상기 그래프 신경망 모델은, 복수 개의 그래프 신경망 층 및 분류기를 포함하고,

상기 그래프 신경망 모델에서, $l+1$ 번째 그래프 신경망 층의 출력($h^{(l+1)}$)은 하기 수학식으로 산출되는, 링크 예측 장치.

(수학식)

$$h^{(l+1)} = \sigma(\hat{A}h^l W^l)$$

σ : 활성화 함수

$\hat{A}$: 정규화된 인접 행렬

h^l : 1번째 그래프 신경망 층의 출력으로 $h^l = (X_l, E_l)$

W^l : 1번째 그래프 신경망 층의 학습 가능한 가중치

X_l : 1번째 그래프 신경망 층에서 출력되는 노드 특징 행렬

E_l : 1번째 그래프 신경망 층에서 출력되는 에지 특징 행렬

청구항 6

청구항 5에 있어서,

1번째 그래프 신경망 층에서 출력되는 노드 특징 행렬(X_l)은 아래 수학식을 통해 산출되는, 링크 예측 장치.

(수학식)

$$X_l = \sigma(\text{CONCAT}(\alpha_l(X_{l-1}, E_{l-1}), g_l(X_{l-1})))$$

σ : 활성화 함수

CONCAT: 괄호 안의 두 개의 행렬을 연결(concatenated)하기 위한 함수

α_l : 1번째 그래프 신경망 층에서 어텐션 계수(attention coefficient)를 구하는 함수

X_{l-1} : $l-1$ 번째 그래프 신경망 층에서 출력되는 노드 특징 행렬

E_{l-1} : $l-1$ 번째 그래프 신경망 층에서 출력되는 에지 특징 행렬

g_l : 1번째 그래프 신경망 층에서 행렬의 차원을 매핑하기 위한 함수

청구항 7

청구항 6에 있어서,

1번째 그래프 신경망 층에서 출력되는 노드 특징 행렬(X_l)이 1+1번째 그래프 신경망 층의 노드 특징 행렬의 입력 값으로 사용되고, 1번째 그래프 신경망 층의 어텐션 계수(α_l)가 1+1번째 그래프 신경망 층의 에지 특징 행렬의 입력 값으로 사용되는, 링크 예측 장치.

청구항 8

하나 이상의 프로세서들, 및

상기 하나 이상의 프로세서들에 의해 실행되는 하나 이상의 프로그램들을 저장하는 메모리를 구비한 컴퓨팅 장치에서 수행되는 방법으로서,

그래프 데이터를 수집하는 단계;

상기 그래프 데이터로부터 노드 특징 행렬, 인접 행렬, 및 에지 특징 행렬을 각각 생성하는 단계; 및

상기 노드 특징 행렬, 상기 인접 행렬, 및 상기 에지 특징 행렬을 그래프 신경망 모델에 입력하여 해당 그래프에 대한 링크 예측을 하도록 상기 그래프 신경망 모델을 학습하는 단계를 포함하는, 그래프의 링크 예측 방법.

청구항 9

청구항 8에 있어서,

상기 그래프 데이터는, 복수 개의 노드 및 노드들 간의 연결 관계인 에지를 포함하며,

상기 생성하는 단계는,

상기 그래프 데이터의 각 노드들로부터 특징을 추출하여 노드 특징 벡터로 표현하고 상기 노드 특징 벡터들의 집합을 통해 상기 노드 특징 행렬을 구성하는 단계;

상기 그래프 데이터의 각 노드들의 연결 관계를 인접 특징 벡터로 표현하고 상기 인접 특징 벡터들의 집합을 통해 상기 인접 행렬을 구성하는 단계; 및

상기 그래프 데이터의 각 에지들의 특징을 에지 특징 벡터로 표현하고 상기 에지 특징 벡터들의 집합을 통해 상기 에지 특징 행렬을 구성하는 단계를 포함하는, 그래프의 링크 예측 방법.

청구항 10

청구항 9에 있어서,

상기 에지 특징 행렬을 구성하는 단계는,

상기 그래프 데이터에서 노드 간 공통 이웃에 대한 정보를 기초로 에지 특징 벡터를 추출하여 에지 특징 행렬을 구성하는, 그래프의 링크 예측 방법.

청구항 11

청구항 9에 있어서,

상기 에지 특징 행렬을 구성하는 단계는,

상기 그래프 데이터에서 노드 간 최단 경로에 대한 정보를 기초로 에지 특징 벡터를 추출하여 에지 특징 행렬을

구성하는, 그래프의 링크 예측 방법.

청구항 12

청구항 9에 있어서,

상기 에지 특징 행렬을 구성하는 단계는,

상기 그래프 데이터에서 노드 간 연결 강도를 기초로 에지 특징 벡터를 추출하여 에지 특징 행렬을 구성하는, 그래프의 링크 예측 방법.

청구항 13

청구항 9에 있어서,

상기 에지 특징 행렬을 구성하는 단계는,

상기 그래프 데이터에서 노드 간 공통 이웃에 대한 정보를 기초로 에지 특징 벡터를 추출하여 제1 에지 특징 행렬을 구성하는 단계;

상기 그래프 데이터에서 노드 간 최단 경로에 대한 정보를 기초로 에지 특징 벡터를 추출하여 제2 에지 특징 행렬을 구성하는 단계; 및

상기 그래프 데이터에서 노드 간 연결 강도를 기초로 에지 특징 벡터를 추출하여 제3 에지 특징 행렬을 구성하는 단계를 포함하는, 그래프의 링크 예측 방법.

청구항 14

청구항 13에 있어서,

상기 에지 특징 행렬은, 상기 제1 에지 특징 행렬, 상기 제2 에지 특징 행렬, 및 상기 제3 에지 특징 행렬 중 둘 이상을 포함하여 다차원 행렬로 구성되며,

상기 학습하는 단계는,

상기 다차원 행렬의 에지 특징 행렬을 상기 노드 특징 행렬 및 상기 인접 행렬과 함께 상기 그래프 신경망 모델로 입력하는 단계를 포함하는, 그래프의 링크 예측 방법.

청구항 15

청구항 14에 있어서,

상기 그래프 신경망 모델은, 복수 개의 그래프 신경망 층 및 분류기를 포함하고,

상기 그래프 신경망 모델에서, $l+1$ 번째 그래프 신경망 층의 출력($h^{(l+1)}$)은 하기 수학식으로 산출되는, 그래프의 링크 예측 방법.

(수학식)

$$h^{(l+1)} = \sigma(\hat{A}h^l W^l)$$

σ : 활성화 함수

$\hat{A}$: 정규화된 인접 행렬

h^l : 1번째 그래프 신경망 층의 출력으로 $h^l = (X_l, E_l)$

W^l : 1번째 그래프 신경망 층의 학습 가능한 가중치

X_l : 1번째 그래프 신경망 층에서 출력되는 노드 특징 행렬

E_l : 1번째 그래프 신경망 층에서 출력되는 에지 특징 행렬

청구항 16

청구항 15에 있어서,

1번째 그래프 신경망 층에서 출력되는 노드 특징 행렬(X_l)은 아래 수학적식을 통해 산출되는, 그래프의 링크 예측 방법.

(수학적식)

$$X_l = \sigma(\text{CONCAT}(\alpha_l(X_{l-1}, E_{l-1}), g_l(X_{l-1})))$$

σ : 활성화 함수

CONCAT: 괄호 안의 두 개의 행렬을 연결(concatenated)하기 위한 함수

α_l : 1번째 그래프 신경망 층에서 어텐션 계수(attention coefficient)를 구하는 함수

X_{l-1} : 1-1번째 그래프 신경망 층에서 출력되는 노드 특징 행렬

E_{l-1} : 1-1번째 그래프 신경망 층에서 출력되는 에지 특징 행렬

g_l : 1번째 그래프 신경망 층에서 행렬의 차원을 매핑하기 위한 함수

청구항 17

청구항 16에 있어서,

1번째 그래프 신경망 층에서 출력되는 노드 특징 행렬(X_l)이 1+1번째 그래프 신경망 층의 노드 특징 행렬의 입력 값으로 사용되고, 1번째 그래프 신경망 층의 어텐션 계수(α_l)가 1+1번째 그래프 신경망 층의 에지 특징 행렬의 입력 값으로 사용되는, 그래프의 링크 예측 방법.

청구항 18

비일시적 컴퓨터 판독 가능한 저장 매체(non-transitory computer readable storage medium)에 저장된 컴퓨터 프로그램으로서,

상기 컴퓨터 프로그램은 하나 이상의 명령어들을 포함하고, 상기 명령어들은 하나 이상의 프로세서들을 갖는 컴퓨팅 장치에 의해 실행될 때, 상기 컴퓨팅 장치로 하여금,

그래프 데이터를 수집하는 단계;

상기 그래프 데이터로부터 노드 특징 행렬, 인접 행렬, 및 에지 특징 행렬을 각각 생성하는 단계; 및

상기 노드 특징 행렬, 상기 인접 행렬, 및 상기 에지 특징 행렬을 그래프 신경망 모델에 입력하여 해당 그래프에 대한 링크 예측을 하도록 상기 그래프 신경망 모델을 학습하는 단계를 수행하도록 하는, 컴퓨터 프로그램.

발명의 설명

기술 분야

[0001] 본 발명의 실시예는 그래프의 링크 예측 기술과 관련된다.

배경 기술

[0003] 그래프는 표, 그리드, 또는 시계열 등의 데이터 보다 실제 세계의 데이터를 보다 직관적으로 표현하는 방법 중 하나이다. 실제 세계의 사물들은 그래프의 노드로 표현되고, 그 사물들 사이에 존재하는 상호 작용과 관계는 에지로 표현된다. 예를 들어, 소셜 네트워크는 사용자들을 노드로 표현하고, 사용자들 사이의 관계를 에지로 표현할 수 있다.

[0004] 최근, 그래프 학습을 위해 신경망 방법이 제안되고 있으나, 링크 예측과 관련하여서는 오래된 휴리스틱(heuristic) 접근 방식이 여전히 좋은 성능을 보이고 있다. 그 이유는 그래프 신경망은 그래프 구조보다는 노드 속성에 크게 의존하는 반면, 링크 예측은 그래프 구조 정보가 매우 중요하기 때문이다. 따라서, 링크 예측을 보다 정확하게 수행할 수 있는 새로운 그래프 신경망 및 그 학습 기법이 요구된다.

선행기술문헌

특허문헌

[0006] (특허문헌 0001) 한국공개특허공보 제10-2022-0076028호(2022.06.08)

발명의 내용

해결하려는 과제

[0007] 본 발명의 실시예는 링크 예측을 정확하게 수행할 수 있는 그래프의 링크 예측 장치 및 방법을 제공하기 위한 것이다.

과제의 해결 수단

[0009] 개시되는 일 실시예에 따른 링크 예측 장치는, 하나 이상의 프로세서들, 및 상기 하나 이상의 프로세서들에 의해 실행되는 하나 이상의 프로그램들을 저장하는 메모리를 구비한 그래프의 링크 예측 장치로서, 그래프 데이터를 수집하는 데이터 획득 모듈; 상기 그래프 데이터로부터 노드 특징 행렬, 인접 행렬, 및 에지 특징 행렬을 각각 생성하는 행렬 생성 모듈; 및 상기 노드 특징 행렬, 상기 인접 행렬, 및 상기 에지 특징 행렬을 그래프 신경망 모델에 입력하여 해당 그래프에 대한 링크 예측을 하도록 상기 그래프 신경망 모델을 학습하는 예측 모듈을 포함한다.

[0010] 상기 그래프 데이터는, 복수 개의 노드 및 노드들 간의 연결 관계인 에지를 포함하며, 상기 행렬 생성 모듈은, 상기 그래프 데이터의 각 노드들로부터 특징을 추출하여 노드 특징 벡터로 표현하고 상기 노드 특징 벡터들의 집합을 통해 상기 노드 특징 행렬을 구성하며, 상기 그래프 데이터의 각 노드들의 연결 관계를 인접 특징 벡터로 표현하고 상기 인접 특징 벡터들의 집합을 통해 상기 인접 행렬을 구성하며, 상기 그래프 데이터의 각 에지들의 특징을 에지 특징 벡터로 표현하고 상기 에지 특징 벡터들의 집합을 통해 상기 에지 특징 행렬을 구성할 수 있다.

- [0011] 상기 행렬 생성 모듈은, 상기 그래프 데이터에서 노드 간 공통 이웃에 대한 정보를 기초로 에지 특징 벡터를 추출하여 제1 에지 특징 행렬을 구성하고, 상기 그래프 데이터에서 노드 간 최단 경로에 대한 정보를 기초로 에지 특징 벡터를 추출하여 제2 에지 특징 행렬을 구성하며, 상기 그래프 데이터에서 노드 간 연결 강도를 기초로 에지 특징 벡터를 추출하여 제3 에지 특징 행렬을 구성할 수 있다.
- [0012] 상기 에지 특징 행렬은, 상기 제1 에지 특징 행렬, 상기 제2 에지 특징 행렬, 및 상기 제3 에지 특징 행렬 중 둘 이상을 포함하여 다차원 행렬로 구성되며, 상기 다차원 행렬의 에지 특징 행렬이 상기 노드 특징 행렬 및 상기 인접 행렬과 함께 상기 그래프 신경망 모델로 입력될 수 있다.
- [0013] 상기 그래프 신경망 모델은, 복수 개의 그래프 신경망 층 및 분류기를 포함하고, 상기 그래프 신경망 모델에서, 1+1 번째 그래프 신경망 층의 출력($h^{(l+1)}$)은 하기 수학식으로 산출될 수 있다.
- [0014] (수학식)
- $$h^{(l+1)} = \sigma(\hat{A}h^l W^l)$$
- [0015]
- [0016] σ : 활성화 함수
- [0017] $\hat{A}$: 정규화된 인접 행렬
- [0018] h^l : 1번째 그래프 신경망 층의 출력으로 $h^l = (X_l, E_l)$
- [0019] W^l : 1번째 그래프 신경망 층의 학습 가능한 가중치
- [0020] X_l : 1번째 그래프 신경망 층에서 출력되는 노드 특징 행렬
- [0021] E_l : 1번째 그래프 신경망 층에서 출력되는 에지 특징 행렬
- [0022] 1번째 그래프 신경망 층에서 출력되는 노드 특징 행렬(X_l)은 아래 수학식을 통해 산출될 수 있다.
- [0023] (수학식)
- $$X_l = \sigma(\text{CONCAT}(\alpha_l(X_{l-1}, E_{l-1}), g_l(X_{l-1})))$$
- [0024]
- [0025] σ : 활성화 함수
- [0026] CONCAT: 괄호 안의 두 개의 행렬을 연결(concatenated)하기 위한 함수
- [0027] α_l : 1번째 그래프 신경망 층에서 어텐션 계수(attention coefficient)를 구하는 함수
- [0028] X_{l-1} : 1-1번째 그래프 신경망 층에서 출력되는 노드 특징 행렬
- [0029] E_{l-1} : 1-1번째 그래프 신경망 층에서 출력되는 에지 특징 행렬
- [0030] g_l : 1번째 그래프 신경망 층에서 행렬의 차원을 매핑하기 위한 함수
- [0031] 1번째 그래프 신경망 층에서 출력되는 노드 특징 행렬(X_l)이 1+1번째 그래프 신경망 층의 노드 특징 행렬의 입력 값으로 사용되고, 1번째 그래프 신경망 층의 어텐션 계수(α_l)가 1+1번째 그래프 신경망 층의 에지 특징 행렬의 입력 값으로 사용될 수 있다.
- [0032] 개시되는 일 실시예에 따른 그래프의 링크 예측 방법은, 하나 이상의 프로세서들, 및 상기 하나 이상의 프로세서들에 의해 실행되는 하나 이상의 프로그램들을 저장하는 메모리를 구비한 컴퓨팅 장치에서 수행되는 방법으로

서, 그래프 데이터를 수집하는 단계; 상기 그래프 데이터로부터 노드 특징 행렬, 인접 행렬, 및 에지 특징 행렬을 각각 생성하는 단계; 및 상기 노드 특징 행렬, 상기 인접 행렬, 및 상기 에지 특징 행렬을 그래프 신경망 모델에 입력하여 해당 그래프에 대한 링크 예측을 하도록 상기 그래프 신경망 모델을 학습하는 단계를 포함할 수 있다.

발명의 효과

[0034] 개시되는 실시예에 의하면, 그래프 데이터의 인접 행렬 및 노드 특징 행렬뿐만 아니라 에지 특징 행렬도 그래프 신경망 모델의 입력으로 사용되어 링크 예측을 수행함으로써, 그래프 데이터에서 링크 예측의 정확도를 높일 수 있게 된다. 그로 인해, 커뮤니티 탐지, 질병 발생 예측, 스팸 전자 메일 탐지, 클라우드 데이터 분석, 트래픽 패턴에 기반한 탐색 경로 등 다양한 분야에서 링크 예측을 효율적으로 수행할 수 있게 된다.

도면의 간단한 설명

[0036] 도 1은 본 발명의 일 실시예에 따른 그래프의 링크 예측 장치를 나타낸 도면

도 2는 개시되는 일 실시예에 따른 그래프 데이터로부터 노드 특징 행렬 및 인접 행렬을 생성하는 상태를 개략적으로 나타낸 도면

도 3은 본 발명의 일 실시예에서 노드 간 공통 이웃에 대한 정보를 기초로 에지 특징 벡터를 생성하는 상태를 나타낸 도면

도 4는 본 발명의 일 실시예에서 노드 간 최단 경로에 대한 정보를 기초로 에지 특징 벡터를 생성하는 상태를 나타낸 도면

도 5는 본 발명의 일 실시예에 따른 예측 모듈에서 링크 예측을 수행하는 상태를 개략적으로 나타낸 도면

도 6은 본 발명의 일 실시예에 따른 그래프의 링크 예측 방법을 나타낸 흐름도

도 7은 예시적인 실시예들에서 사용되기에 적합한 컴퓨팅 장치를 포함하는 컴퓨팅 환경을 예시하여 설명하기 위한 블록도

발명을 실시하기 위한 구체적인 내용

[0037] 이하, 도면을 참조하여 본 발명의 구체적인 실시형태를 설명하기로 한다. 이하의 상세한 설명은 본 명세서에서 기술된 방법, 장치 및/또는 시스템에 대한 포괄적인 이해를 돕기 위해 제공된다. 그러나 이는 예시에 불과하며 본 발명은 이에 제한되지 않는다.

[0038] 본 발명의 실시예들을 설명함에 있어서, 본 발명과 관련된 공지기술에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략하기로 한다. 그리고, 후술되는 용어들은 본 발명에서의 기능을 고려하여 정의된 용어들로서 이는 사용자, 운용자의 의도 또는 관례 등에 따라 달라질 수 있다. 그러므로 그 정의는 본 명세서 전반에 걸친 내용을 토대로 내려져야 할 것이다. 상세한 설명에서 사용되는 용어는 단지 본 발명의 실시예들을 기술하기 위한 것이며, 결코 제한적이지는 안 된다. 명확하게 달리 사용되지 않는 한, 단수 형태의 표현은 복수 형태의 의미를 포함한다. 본 설명에서, "포함" 또는 "구비"와 같은 표현은 어떤 특성들, 숫자들, 단계들, 동작들, 요소들, 이들의 일부 또는 조합을 가리키기 위한 것이며, 기술된 것 이외에 하나 또는 그 이상의 다른 특성, 숫자, 단계, 동작, 요소, 이들의 일부 또는 조합의 존재 또는 가능성을 배제하도록 해석되어서는 안 된다.

[0039] 또한, 제1, 제2 등의 용어는 다양한 구성 요소들을 설명하는데 사용될 수 있지만, 상기 구성 요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성 요소를 다른 구성 요소로부터 구별하는 목적으로 사용될 수 있다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제1 구성 요소는 제2 구성 요소로 명명될 수 있고, 유사하게 제2 구성 요소도 제1 구성 요소로 명명될 수 있다.

[0041] 도 1은 본 발명의 일 실시예에 따른 그래프의 링크 예측 장치를 나타낸 도면이다.

- [0042] 도 1을 참조하면, 그래프의 링크 예측 장치(100)는 데이터 획득 모듈(102), 행렬 생성 모듈(104), 및 예측 모듈(106)을 포함할 수 있다. 링크 예측 장치(100)는 그래프에서 노드와 노드 간 링크(link)(연결)가 존재하는지를 예측하기 위한 장치일 수 있다. 링크 예측 장치(100)는 인공 신경망 기술에 기반하여 그래프에서 노드와 노드 간 링크 존재 여부를 예측할 수 있다. 개시되는 실시예에서 링크는 에지(edge)와 동일한 용어로 사용될 수 있다.
- [0043] 데이터 획득 모듈(102)은 그래프 데이터를 수집할 수 있다. 데이터 획득 모듈(102)은 예측 모듈(106)에 포함되는 인공 신경망의 학습을 위한 학습 데이터로서 그래프 데이터를 수집할 수 있다. 일 실시예에서, 그래프 데이터는 소셜 네트워크(social network) 데이터, 전자 메시지를 교환하는 네트워크 데이터, 생태계 네트워크 데이터, 단백질 네트워크 데이터 등을 포함할 수 있다. 또한, 그래프 데이터는 무방향 그래프(undirected graph)일 수도 있고, 유방향 그래프(directed graph)일 수도 있다.
- [0044] 행렬 생성 모듈(104)은 수집된 그래프 데이터로부터 노드 특징 행렬, 인접 행렬, 및 에지 특징 행렬을 각각 생성할 수 있다. 그래프 데이터는 복수 개의 노드(node) 및 노드들 간의 연결 관계인 에지(edge)로 구성된다. 즉, 그래프 데이터 $G = (V, E)$ 로 표현할 수 있고, $V = (v_1, v_2, \dots, v_n)$ 으로 노드들의 집합이며, $E = (e_{ij} \mid v_i, v_j \in V)$ 으로 에지의 집합이다. 여기서, n 은 전체 노드의 개수이고, e_{ij} 는 노드 v_i 와 노드 v_j 를 연결하는 에지를 의미한다.
- [0045] 행렬 생성 모듈(104)은 그래프 데이터의 각 노드들의 특징을 추출하여 노드 특징 행렬(node feature matrix)(X_0)을 생성할 수 있다. 또한, 행렬 생성 모듈(104)은 그래프 데이터의 각 노드 및 에지에 기초하여 인접 행렬(adjacency matrix)(A)을 생성할 수 있다.
- [0046] 도 2는 개시되는 일 실시예에 따른 그래프 데이터로부터 노드 특징 행렬 및 인접 행렬을 생성하는 상태를 개략적으로 나타낸 도면이다. 도 2를 참조하면, 노드가 5개인 그래프 데이터(50)가 주어진 경우, 행렬 생성 모듈(104)은 각 노드(node 1, node 2, node 3, node 4, node 5)로부터 특징을 추출하여 노드 특징 벡터로 표현할 수 있다.
- [0047] 예를 들어, 노드 1(node 1)은 그 특징에 따라 [11100]이라는 노드 특징 벡터로 표현하고, 노드 2(node 2)는 그 특징에 따라 [11111]이라는 노드 특징 벡터로 표현하며, 노드 3(node 3)은 그 특징에 따라 [11111]이라는 노드 특징 벡터로 표현하고, 노드 4(node 4)는 그 특징에 따라 [01111]이라는 노드 특징 벡터로 표현하고, 노드 5(node 5)는 그 특징에 따라 [01111]이라는 노드 특징 벡터로 표현하며, 이러한 노드 특징 벡터들을 통해 노드 특징 행렬(node feature matrix)(X_0)을 생성할 수 있다. 즉, 노드 특징 행렬은 각 노드의 노드 특징 벡터로 구성된 행렬일 수 있다.
- [0048] 또한, 행렬 생성 모듈(104)은 각 노드들의 연결 관계를 인접 특징 벡터로 표현할 수 있다. 즉, 노드 1(node 1)은 자기 자신(즉, 노드 1)과 연결되고, 노드 2 및 노드 3과 연결되며, 노드 4 및 노드 5와는 연결되지 않으므로 [01100]이라는 인접 특징 벡터로 표현할 수 있다. 마찬가지로, 노드 2(node 2)는 [10111]이라는 인접 특징 벡터로 표현하고, 노드 3(node 3)은 [11011]이라는 인접 특징 벡터로 표현하며, 노드 4(node 4)는 [01101]이라는 인접 특징 벡터로 표현하고, 노드 5(node 5)는 [01110]이라는 인접 특징 벡터로 표현할 수 있다. 그리고, 각 노드의 인접 특징 벡터를 집합하여 인접 행렬(adjacency matrix)(A)을 생성할 수 있다.
- [0049] 여기서, 인접 행렬(A)은 그래프 데이터(50)의 구조를 나타내는 것으로, 그래프 데이터(50)에서 두 노드가 연결되어 있으면 1로 표현하고, 연결되어 있지 않으면 0으로 표현하는 이진 행렬을 의미할 수 있다. 즉, 인접 행렬(A)은 각 노드들 간의 연결 관계를 0과 1로 표현한 행렬일 수 있다.
- [0050] 또한, 행렬 생성 모듈(104)은 노드 및 에지로 구성되는 그래프 데이터로부터 에지 특징 행렬(edge feature matrix)(E_0)을 생성할 수 있다. 에지 특징 행렬은 그래프 데이터의 에지들의 특징인 에지 특징 벡터로 구성된 행렬일 수 있다. 여기서, 에지 특징 벡터는 단순히 이웃하는 노드 간의 연결 관계만을 1차원적으로 나타내는 인접 특징 벡터와는 달리 해당 에지의 특징을 고차원적으로 나타낸 것일 수 있다. 일 실시예에서, 에지 특징 벡터는 그래프 데이터의 인접 행렬을 분석하여 획득할 수 있다.
- [0051] 예를 들어, 에지 특징 벡터는 제1 노드와 제2 노드의 공통 이웃에 대한 정보를 포함할 수 있다. 즉, 도 3에 도시된 바와 같이, 그래프 데이터에서 A 노드와 B 노드에 공통적인 이웃은 C 노드, D 노드, 및 E 노드의 3개로서 이는 **Common neighbors $C_{A,B} = 3$** 으로 표현할 수 있다.

[0052] 또한, 에지 특징 벡터는 제1 노드와 제2 노드 간의 최단 경로에 대한 정보를 포함할 수 있다. 즉, 도 4에 도시된 바와 같이, 노드 a에서 노드 f까지의 최단 경로는 노드 a - 노드 c - 노드 b - 노드 d - 노드 e - 노드 f로 그 최단 경로 길이는 노드 a부터 노드 c까지 거리인 2, 노드 c부터 노드 b까지의 거리인 1, 노드 b부터 노드 d까지의 거리인 5, 노드 d부터 노드 e까지의 거리인 2, 노드 e부터 노드 f까지의 거리인 5를 합산한 15가 되며 이는 **Shortest path $S_{a,f} = 15$** 로 표현할 수 있다.

[0053] 그 이외에도, 에지 특징 벡터는 제1 노드와 제2 노드 간의 연결 강도를 수치로 표현할 것일 수 있다. 예를 들어, 그래프 데이터가 SNS 데이터인 경우, 에지 특징 벡터는 제1 사용자(제1 노드)와 제2 사용자(제2 노드) 간의 친밀도를 수치로 표현한 것일 수 있다. 이때, 제1 사용자와 제2 사용자 간의 친밀도는 팔로워 기간, 댓글 수, 좋아요 수 등을 통해 산출할 수 있다. 행렬 생성 모듈(104)은 이러한 에지 특징 벡터들로부터 에지 특징 행렬을 생성할 수 있다.

[0054] 일 실시예에서, 에지 특징 행렬은 다차원의 행렬일 수 있다. 예를 들어, 에지 특징 행렬은 노드 간 공통 이웃에 대한 정보를 기초로 하는 에지 특징 벡터들로 구성되는 제1 에지 특징 행렬, 노드 간 최단 경로에 대한 정보를 기초로 하는 에지 특징 벡터들로 구성되는 제2 에지 특징 행렬, 및 노드 간 연결 강도를 기초로 하는 에지 특징 벡터들로 구성되는 제3 에지 특징 행렬을 포함하여 다차원의 행렬로 구성될 수 있다.

[0055] 예측 모듈(106)은 그래프 데이터의 인접 행렬(A), 노드 특징 행렬(X_0), 및 에지 특징 행렬(E_0)을 입력 받고, 인접 행렬(A), 노드 특징 행렬(X_0), 및 에지 특징 행렬(E_0)에 기초하여 지정된 노드가 연결될 가능성이 있는지 여부(즉, 링크의 존재 여부)를 예측할 수 있다. 이를 위해, 예측 모듈(106)은 그래프 신경망(Graph Neural Network) 모델을 포함할 수 있다.

[0056] 일반적인 그래프 신경망 모델은 그래프 데이터의 인접 행렬과 노드 특징 행렬을 입력으로 하여 링크 예측을 수행하나, 개시되는 실시예에서는 그래프 데이터의 인접 행렬 및 노드 특징 행렬뿐만 아니라 에지 특징 행렬도 그래프 신경망 모델의 입력으로 사용되어 링크 예측을 수행하게 된다.

[0057] 도 5는 본 발명의 일 실시예에 따른 예측 모듈(106)에서 링크 예측을 수행하는 상태를 개략적으로 나타낸 도면이다. 도 5에서는 그래프 데이터의 노드들 중 노드 i와 노드 j에 링크가 존재하는지 여부를 예측하는 상태를 도시하였다.

[0058] 도 5를 참조하면, 예측 모듈(106)은 순차적으로 연결된 복수 개의 그래프 신경망 층(111-1, 111-2, ... 111-L) 및 분류기(113)를 포함할 수 있다. 여기서, l+1 번째 그래프 신경망 층의 출력($h^{(l+1)}$)은 아래 수학식 1로 산출할 수 있다.

[0059] (수학식 1)

[0060]
$$h^{(l+1)} = \sigma(\hat{A}h^l W^l)$$

[0061] σ : 활성화 함수

[0062] $\hat{A}$: 정규화된 인접 행렬

[0063] h^l : l번째 그래프 신경망 층의 출력

[0064] W^l : l번째 그래프 신경망 층의 학습 가능한 가중치

[0065] 여기서, $h^l = (X_l, E_l)$ 이다. X_l 은 l번째 그래프 신경망 층에서 출력되는 노드 특징 행렬이고, E_l 은 l번째 그래프 신경망 층에서 출력되는 에지 특징 행렬이다. 즉, l번째 그래프 신경망 층의 출력(h^l)은 l번째 그래프 신경망 층을 통해 임베딩 된 노드 특징 행렬 및 에지 특징 행렬을 포함할 수 있다.

[0066] 일 실시예에서, 인접 행렬(A)이 제1 에지 특징 행렬, 제2 에지 특징 행렬, 및 제3 에지 특징 행렬을 포함하는 다차원의 행렬인 경우, 제1 에지 특징 행렬, 제2 에지 특징 행렬, 및 제3 에지 특징 행렬의 범위를 맞추기 위해

인접 행렬(A)을 정규화 할 수 있다.

그리고, 1번째 그래프 신경망 층에서 출력되는 노드 특징 행렬(X_l)은 아래 수학적 2를 통해 산출할 수 있다.

(수학적 2)

$$X_l = \sigma(\text{CONCAT}(\alpha_l(X_{l-1}, E_{l-1}), g_l(X_{l-1})))$$

σ : 활성화 함수

CONCAT: 괄호 안의 두 개의 행렬을 연결(concatenated)하기 위한 함수

α_l : 1번째 그래프 신경망 층에서 어텐션 계수(attention coefficient)를 구하는 함수

X_{l-1} : 1-1번째 그래프 신경망 층에서 출력되는 노드 특징 행렬

E_{l-1} : 1-1번째 그래프 신경망 층에서 출력되는 에지 특징 행렬

g_l : 1번째 그래프 신경망 층에서 행렬의 차원을 매핑하기 위한 함수

구체적으로, 1번째 그래프 신경망 층에서 1-1번째 그래프 신경망 층에서 출력되는 노드 특징 행렬(X_{l-1})과 에지 특징 행렬(E_{l-1})은 어텐션 함수(α)의 입력 값으로 사용되어 1번째 그래프 신경망 층의 어텐션 계수(α_l)를 산출하게 된다. 여기서, 어텐션 계수는 어텐션 메커니즘(attention mechanism)을 통해 산출할 수 있으며, 이는 공지된 기술이므로 이에 대한 자세한 설명은 생략하기로 한다.

그리고, CONCAT 함수를 통해 1번째 그래프 신경망 층의 어텐션 계수(α_l)와 1-1번째 그래프 신경망 층에서 출력되는 노드 특징 행렬(X_{l-1})이 연결(concatenated)된다. 이때, 1-1번째 그래프 신경망 층에서 출력되는 노드 특징 행렬(X_{l-1})은 매핑 함수(g_l)로 입력되어 어텐션 계수(α_l)와 동일 차원으로 디멘전(dimension)을 매핑시킬 수 있다. 다음으로, CONCAT 함수를 통해 연결된 값은 활성화 함수(σ)를 거쳐서 1번째 그래프 신경망 층의 노드 특징 행렬(X_l)로 출력되게 된다.

여기서, 1번째 그래프 신경망 층에서 출력되는 노드 특징 행렬(X_l)이 1+1번째 그래프 신경망 층의 노드 특징 행렬의 입력 값으로 사용되게 된다. 한편, 1+1번째 그래프 신경망 층의 에지 특징 행렬의 입력 값은 1번째 그래프 신경망 층의 어텐션 계수(α_l)가 사용될 수 있다.

분류기(113)는 마지막 그래프 신경망 층(111-L)에서 출력되는 노드 특징 행렬 및 에지 특징 행렬에 기초하여 링크의 존재 여부를 예측(link prediction)할 수 있다.

개시되는 실시예에 의하면, 그래프 데이터의 인접 행렬 및 노드 특징 행렬뿐만 아니라 에지 특징 행렬도 그래프 신경망 모델의 입력으로 사용되어 링크 예측을 수행함으로써, 그래프 데이터에서 링크 예측의 정확도를 높일 수 있게 된다. 그로 인해, 커뮤니티 탐지, 질병 발생 예측, 스팸 전자 메일 탐지, 클라우드 데이터 분석, 트래픽 패턴에 기반한 탐색 경로 등 다양한 분야에서 링크 예측을 효율적으로 수행할 수 있게 된다.

본 명세서에서 모듈이라 함은, 본 발명의 기술적 사상을 수행하기 위한 하드웨어 및 상기 하드웨어를 구동하기 위한 소프트웨어의 기능적, 구조적 결합을 의미할 수 있다. 예컨대, 상기 "모듈"은 소정의 코드와 상기 소정의 코드가 수행되기 위한 하드웨어 리소스의 논리적인 단위를 의미할 수 있으며, 반드시 물리적으로 연결된 코드를 의미하거나, 한 종류의 하드웨어를 의미하는 것은 아니다.

- [0083] 도 6은 본 발명의 일 실시예에 따른 그래프의 링크 예측 방법을 나타낸 흐름도이다. 도시된 흐름도에서는 상기 방법을 복수 개의 단계로 나누어 기재하였으나, 적어도 일부의 단계들은 순서를 바꾸어 수행되거나, 다른 단계와 결합되어 함께 수행되거나, 생략되거나, 세부 단계들로 나뉘어 수행되거나, 또는 도시되지 않은 하나 이상의 단계가 부가되어 수행될 수 있다.
- [0084] 도 6을 참조하면, 링크 예측 장치(100)는 그래프 데이터를 수집할 수 있다(S 101). 여기서, 그래프 데이터는 소셜 네트워크(social network) 데이터, 전자 메시지를 교환하는 네트워크 데이터, 생태계 네트워크 데이터, 단백질 네트워크 데이터 등 노드와 에지로 표현할 수 있는 다양한 형태의 그래프 데이터를 포함할 수 있다.
- [0085] 다음으로, 링크 예측 장치(100)는 수집된 그래프 데이터로부터 노드 특징 행렬, 인접 행렬, 및 에지 특징 행렬을 각각 생성할 수 있다(S 103).
- [0086] 구체적으로, 링크 예측 장치(100)는 그래프 데이터의 각 노드들의 특징을 추출하여 노드 특징 행렬(node feature matrix)을 생성할 수 있다. 링크 예측 장치(100)는 그래프 데이터의 각 노드 및 에지에 기초하여 인접 행렬(adjacency matrix)을 생성할 수 있다. 링크 예측 장치(100)는 그래프 데이터의 인접 행렬을 분석하여 에지 특징 벡터를 추출하고, 에지 특징 벡터들로 구성된 에지 특징 행렬을 생성할 수 있다.
- [0087] 다음으로, 링크 예측 장치(100)는 그래프 데이터의 인접 행렬, 노드 특징 행렬, 및 에지 특징 행렬을 그래프 신경망 모델로 입력하여 그래프 데이터에 대한 링크 예측을 수행하도록 그래프 신경망 모델을 학습할 수 있다(S 105).
- [0089] 도 7은 예시적인 실시예들에서 사용되기에 적합한 컴퓨팅 장치를 포함하는 컴퓨팅 환경(10)을 예시하여 설명하기 위한 블록도이다. 도시된 실시예에서, 각 컴포넌트들은 이하에 기술된 것 이외에 상이한 기능 및 능력을 가질 수 있고, 이하에 기술된 것 이외에도 추가적인 컴포넌트를 포함할 수 있다.
- [0090] 도시된 컴퓨팅 환경(10)은 컴퓨팅 장치(12)를 포함한다. 일 실시예에서, 컴퓨팅 장치(12)는 그래프의 링크 예측 장치(100)일 수 있다.
- [0091] 컴퓨팅 장치(12)는 적어도 하나의 프로세서(14), 컴퓨터 판독 가능 저장 매체(16) 및 통신 버스(18)를 포함한다. 프로세서(14)는 컴퓨팅 장치(12)로 하여금 앞서 언급된 예시적인 실시예에 따라 동작하도록 할 수 있다. 예컨대, 프로세서(14)는 컴퓨터 판독 가능 저장 매체(16)에 저장된 하나 이상의 프로그램들을 실행할 수 있다. 상기 하나 이상의 프로그램들은 하나 이상의 컴퓨터 실행 가능 명령어를 포함할 수 있으며, 상기 컴퓨터 실행 가능 명령어는 프로세서(14)에 의해 실행되는 경우 컴퓨팅 장치(12)로 하여금 예시적인 실시예에 따른 동작들을 수행하도록 구성될 수 있다.
- [0092] 컴퓨터 판독 가능 저장 매체(16)는 컴퓨터 실행 가능 명령어 내지 프로그램 코드, 프로그램 데이터 및/또는 다른 적합한 형태의 정보를 저장하도록 구성된다. 컴퓨터 판독 가능 저장 매체(16)에 저장된 프로그램(20)은 프로세서(14)에 의해 실행 가능한 명령어의 집합을 포함한다. 일 실시예에서, 컴퓨터 판독 가능 저장 매체(16)는 메모리(랜덤 액세스 메모리와 같은 휘발성 메모리, 비휘발성 메모리, 또는 이들의 적절한 조합), 하나 이상의 자기 디스크 저장 디바이스들, 광학 디스크 저장 디바이스들, 플래시 메모리 디바이스들, 그 밖에 컴퓨팅 장치(12)에 의해 액세스되고 원하는 정보를 저장할 수 있는 다른 형태의 저장 매체, 또는 이들의 적합한 조합일 수 있다.
- [0093] 통신 버스(18)는 프로세서(14), 컴퓨터 판독 가능 저장 매체(16)를 포함하여 컴퓨팅 장치(12)의 다른 다양한 컴포넌트들을 상호 연결한다.
- [0094] 컴퓨팅 장치(12)는 또한 하나 이상의 입출력 장치(24)를 위한 인터페이스를 제공하는 하나 이상의 입출력 인터페이스(22) 및 하나 이상의 네트워크 통신 인터페이스(26)를 포함할 수 있다. 입출력 인터페이스(22) 및 네트워크 통신 인터페이스(26)는 통신 버스(18)에 연결된다. 입출력 장치(24)는 입출력 인터페이스(22)를 통해 컴퓨팅 장치(12)의 다른 컴포넌트들에 연결될 수 있다. 예시적인 입출력 장치(24)는 포인팅 장치(마우스 또는 트랙패드 등), 키보드, 터치 입력 장치(터치패드 또는 터치스크린 등), 음성 또는 소리 입력 장치, 다양한 종류의 센서 장치 및/또는 촬영 장치와 같은 입력 장치, 및/또는 디스플레이 장치, 프린터, 스피커 및/또는 네트워크 카드와 같은 출력 장치를 포함할 수 있다. 예시적인 입출력 장치(24)는 컴퓨팅 장치(12)를 구성하는 일 컴포넌트로서 컴퓨팅 장치(12)의 내부에 포함될 수도 있고, 컴퓨팅 장치(12)와는 구별되는 별개의 장치로 컴퓨팅 장치(12)와 연결될 수도 있다.

[0096] 이상에서 본 발명의 대표적인 실시예들을 상세하게 설명하였으나, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자는 상술한 실시예에 대하여 본 발명의 범주에서 벗어나지 않는 한도 내에서 다양한 변형이 가능함을 이해할 것이다. 그러므로 본 발명의 권리범위는 설명된 실시예에 국한되어 정해져서는 안 되며, 후술하는 특허 청구범위뿐만 아니라 이 특허청구범위와 균등한 것들에 의해 정해져야 한다.

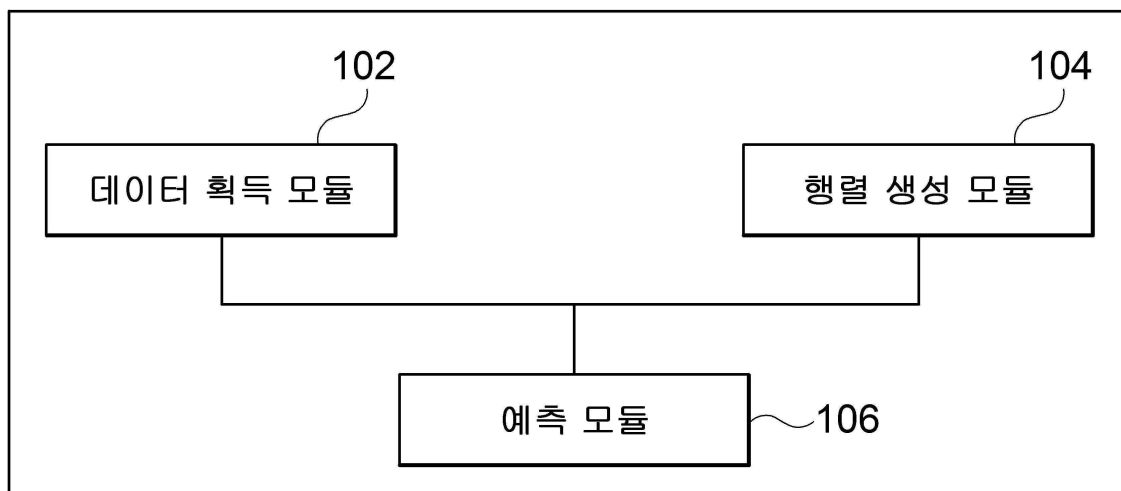
부호의 설명

[0098] 10 : 컴퓨팅 환경
12 : 컴퓨팅 장치
14 : 프로세서
16 : 컴퓨터 관독 가능 저장 매체
18 : 통신 버스
20 : 프로그램
22 : 입출력 인터페이스
24 : 입출력 장치
26 : 네트워크 통신 인터페이스
100 : 그래프의 링크 예측 장치
102 : 데이터 획득 모듈
104 : 행렬 생성 모듈
106 : 예측 모듈
111 : 그래프 신경망 층
113 : 분류기

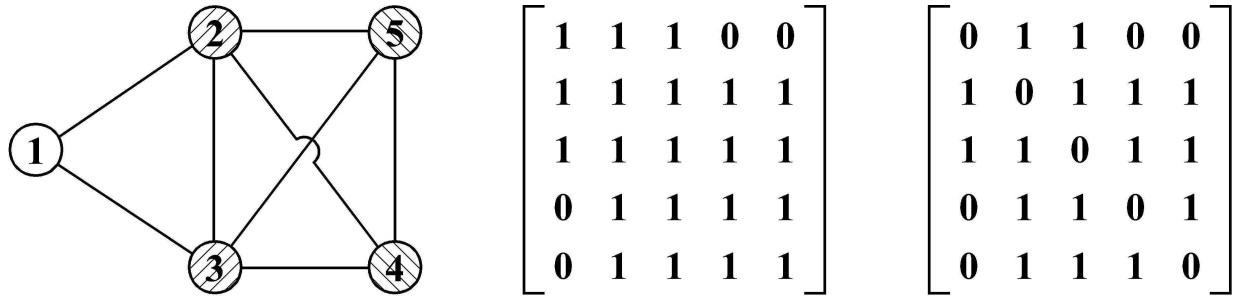
도면

도면1

100



도면2

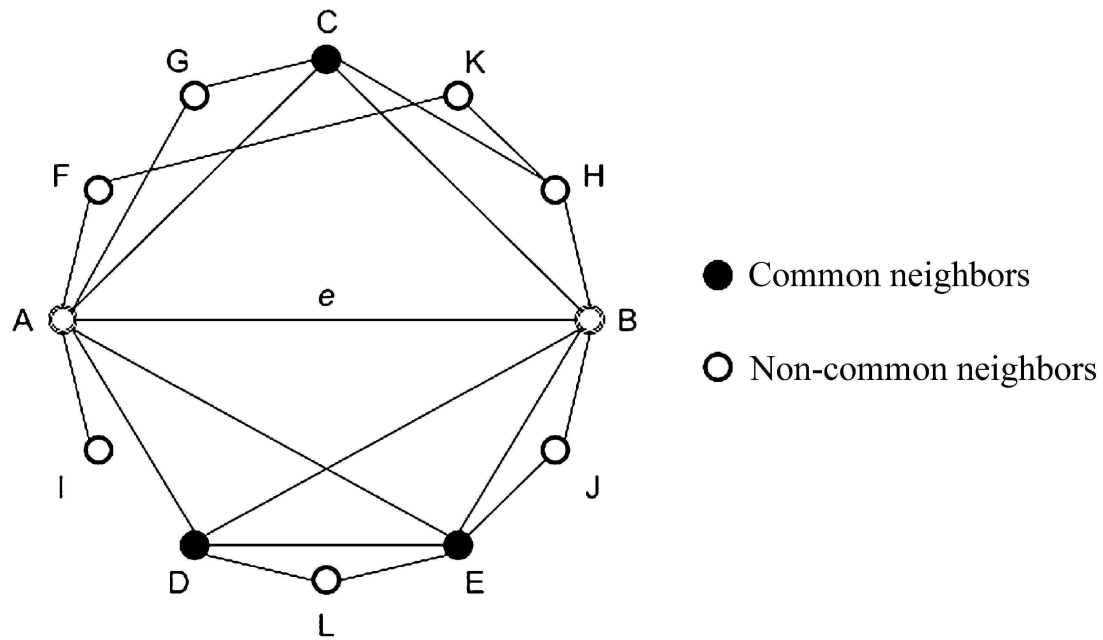


(a) 그래프 데이터

(b) 노드 특징 행렬(X_0)

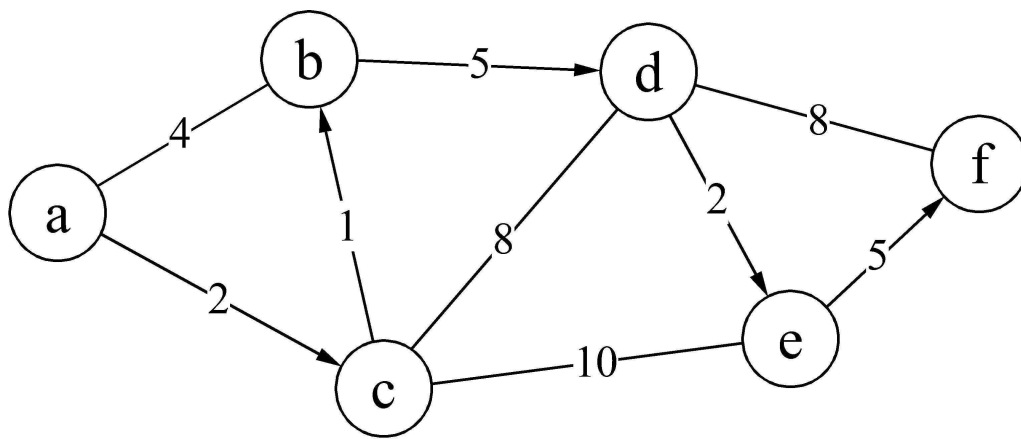
(c) 인접 행렬(A)

도면3



Common neighbors $C_{A,B} = 3$

도면4

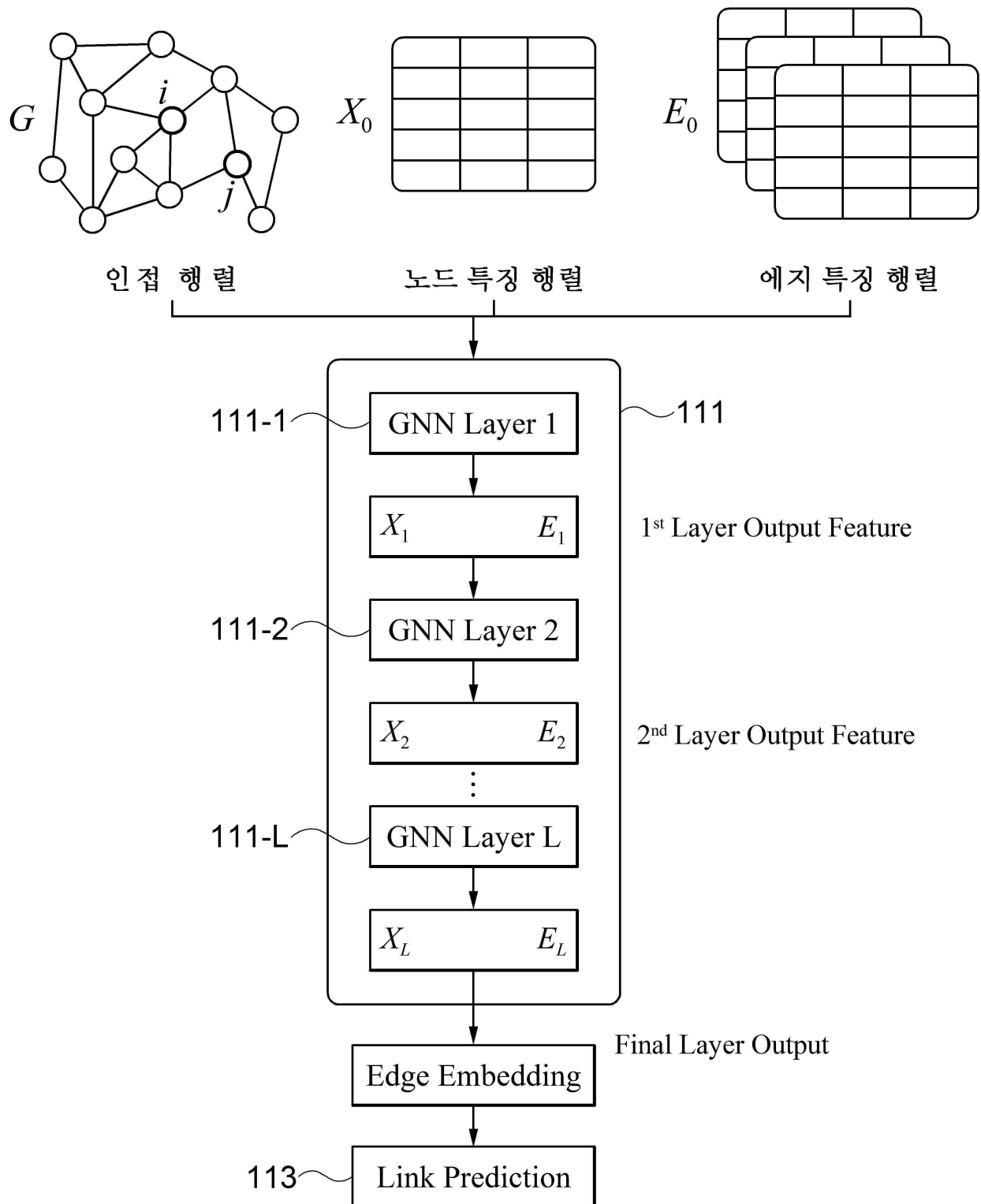


$$2 + 1 + 5 + 2 + 5 = 15$$

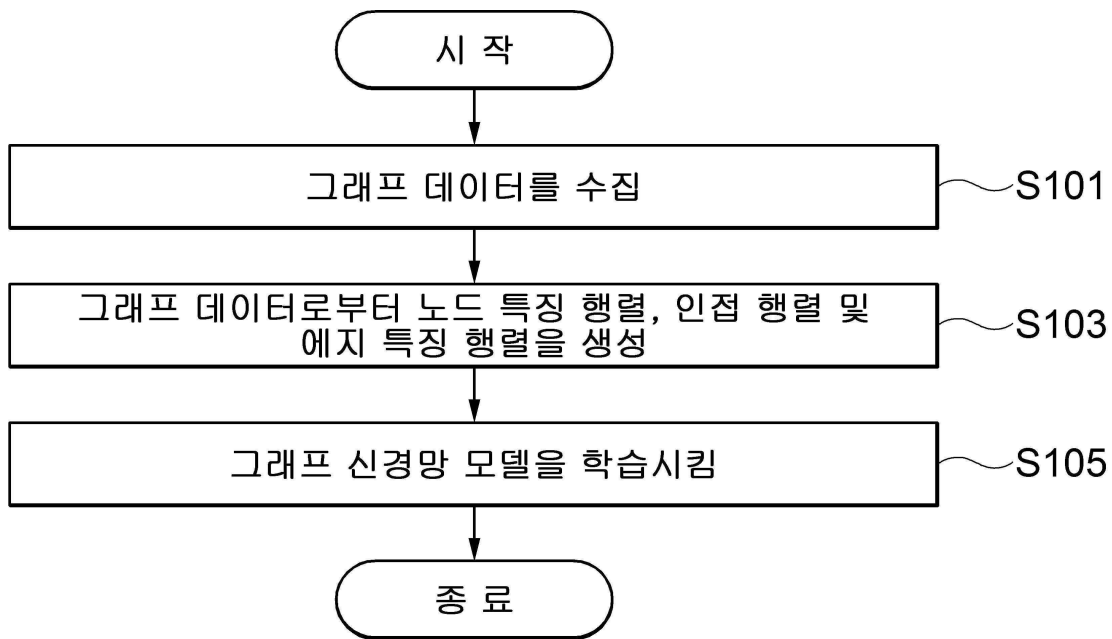
Shortest path: $S_{a,f} = 15$

도면5

106



도면6



도면7

10

