



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0030235
(43) 공개일자 2023년03월06일

(51) 국제특허분류(Int. Cl.)
G06N 3/08 (2023.01) G06N 3/04 (2023.01)
(52) CPC특허분류
G06N 3/082 (2023.01)
G06N 3/04 (2023.01)
(21) 출원번호 10-2021-0112209
(22) 출원일자 2021년08월25일
심사청구일자 2021년08월25일

(71) 출원인
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
(72) 발명자
신원용
서울특별시 서대문구 성산로 394, 503호(창천동, JK캠퍼스)
신용민
경기도 고양시 일산서구 대산로211번길 7-13, 201호(대화동)
(74) 대리인
민영준

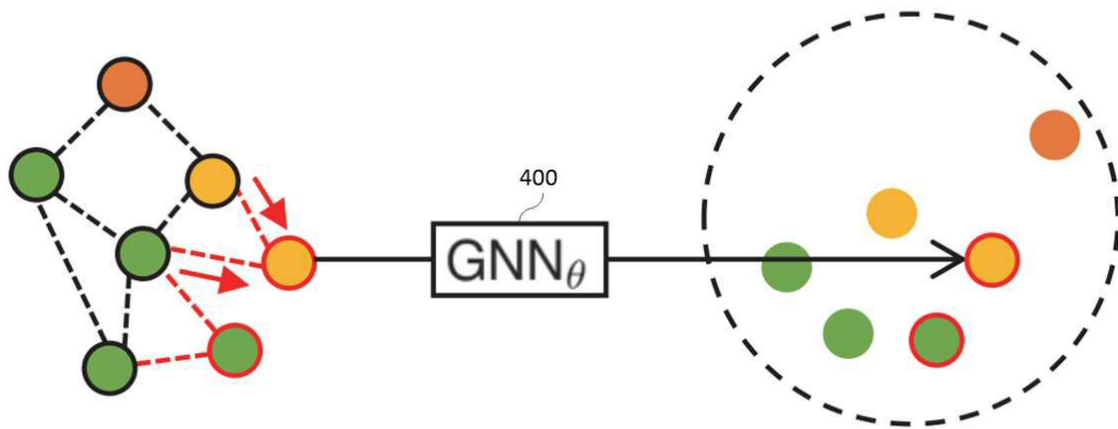
전체 청구항 수 : 총 16 항

(54) 발명의 명칭 그래프 인공 신경망 기반 엣지리스 네트워크 임베딩 장치 및 방법

(57) 요약

본 발명은 다수의 노드와 다수의 노드 사이의 연관 관계에 따라 대응하는 노드를 연결하는 다수의 엣지를 포함하는 네트워크 데이터에서 다수의 노드 각각에 대한 속성 정보를 획득하는 네트워크 획득부, 다수의 노드 각각에 미리 지정된 속성 사이의 유사도를 기지정된 방식으로 계산하는 노드 유사도 판별부, 판별된 유사도에 따라 다수 (뒷면에 계속)

대표도 - 도6



의 노드 중 대응하는 노드를 유사도 엣지로 서로 연결하여 유사도 네트워크 데이터를 생성하는 유사도 네트워크 구성부 및 미리 학습된 인공 신경망으로 구현되어, 학습된 신경망 연산 방식에 따라 유사도 네트워크 데이터로부터 네트워크 데이터의 노드 사이의 연관 관계를 추정하고, 추정된 연관 관계에 따라 다수의 노드를 가상의 임베딩 공간 상에 배치하는 네트워크 임베딩부를 포함하여, 기존의 노드들과의 연관 관계가 생성되지 않은 신규 노드와 같은 엣지리스 노드를 추가 학습 또는 재학습 없이 효과적으로 임베딩 공간에 임베딩할 수 있는 엣지리스 네트워크 임베딩 장치 및 방법을 제공한다.

이 발명을 지원한 국가연구개발사업

과제고유번호	1711137771
과제번호	2021R1A2C3004345
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	중견연구자지원사업
연구과제명	초불확실성 문제 해결을 위한 심층 네트워크 복구 방법 및 핵심 요소기술 개발
기 여 율	1/1
과제수행기관명	연세대학교 산학협력단
연구기간	2021.03.01 ~ 2022.02.28

명세서

청구범위

청구항 1

다수의 노드와 상기 다수의 노드 사이의 연관 관계에 따라 대응하는 노드를 연결하는 다수의 엣지를 포함하는 네트워크 데이터에서 다수의 노드 각각에 대한 속성 정보를 획득하는 네트워크 획득부;

상기 다수의 노드 각각에 미리 지정된 속성 사이의 유사도를 기지정된 방식으로 계산하는 노드 유사도 판별부;

판별된 유사도에 따라 상기 다수의 노드 중 대응하는 노드를 유사도 엣지로 서로 연결하여 유사도 네트워크 데이터를 생성하는 유사도 네트워크 구성부; 및

미리 학습된 인공 신경망으로 구현되어, 학습된 신경망 연산 방식에 따라 상기 유사도 네트워크 데이터로부터 상기 네트워크 데이터의 노드 사이의 연관 관계를 추정하고, 추정된 연관 관계에 따라 상기 다수의 노드를 가상의 임베딩 공간 상에 배치하는 네트워크 임베딩부를 포함하는 엣지리스 네트워크 임베딩 장치.

청구항 2

제1항에 있어서, 상기 노드 유사도 판별부는

상기 다수의 노드 사이를 잇는 엣지와 무관하게 상기 다수의 노드 각각에 대해 미리 설정된 속성 정보를 기반으로 유사도를 계산하는 엣지리스 네트워크 임베딩 장치.

청구항 3

제1항에 있어서, 상기 유사도 네트워크 구성부는

상기 다수의 노드 각각에 대해 가장 높은 유사도를 갖는 기지정된 개수의 노드를 상기 유사도 엣지로 서로 연결하여 유사도 네트워크 데이터를 생성하는 엣지리스 네트워크 임베딩 장치.

청구항 4

제1항에 있어서, 상기 네트워크 임베딩부는

임베딩 공간에 노드들 사이의 거리를 에너지 크기로 가정하는 에너지 기반 학습에 기반하여, 엣지로 연결되지 않은 엣지리스 노드가 포함되지 않은 네트워크 데이터를 학습 데이터에서 엣지에 따른 노드간 연관 관계가 클수록 임베딩 공간 상에서 노드 쌍 사이의 에너지가 줄어들도록 서로 가깝게 위치시키고, 연관 관계가 작을수록 노드 쌍 사이의 에너지가 증가되도록 멀리 위치하도록 학습시키면서, 공통으로 연결되는 노드의 수가 많은 경우에도 에너지가 줄어들어 가깝게 위치하도록 설정된 손실을 계산하고, 계산된 손실이 역전파되어 학습되는 엣지리스 네트워크 임베딩 장치.

청구항 5

제4항에 있어서, 상기 네트워크 임베딩부는

상기 학습 데이터에서 다수의 노드 중 하나를 대상 노드(v_i)로 선택하고, 선택된 대상 노드(v_i)와 엣지를 통해 직접 연결된 이웃 노드(v_j)와 이웃 노드를 통해 연결되는 2홉 노드(v_l) 및 2홉을 초과하는 원격 노드(v_n)를 획득하여 미니 배치(B')를 추출하고, 상기 미니 배치에 포함된 다수의 노드(v_i, v_j, v_n, v_l)가 엣지로 연결된 관계에 따른 에너지를 상기 손실로 계산하는 엣지리스 네트워크 임베딩 장치.

청구항 6

제5항에 있어서, 상기 손실(L)은

상기 미니 배치(B')의 상기 대상 노드(v_i)로부터 상기 이웃 노드(v_j)와 상기 원격 노드(v_n) 사이의 엣지 수에 따라 임베딩 공간에서 상기 대상 노드(v_i)와 상기 이웃 노드(v_j)가 서로 인접하고 상기 원격 노드(v_n)는 엣지 수만

를 이격되어 배치되도록 하는 연관 손실(L_0)과 상기 대상 노드(v_i)와 상기 2홉 노드(v_t)에서 공통의 이웃 노드 수에 따라 상기 2홉 노드(v_t)가 상기 대상 노드(v_i)에 인접하여 배치되도록 하는 공통 이웃 손실(L_{2nd})로부터 수학식

$$\mathcal{L} = \mathcal{L}_0 + \alpha \mathcal{L}_{2nd}$$

(여기서 α 는 손실 가중치)

에 따라 계산되는 엡지리스 네트워크 임베딩 장치.

청구항 7

제6항에 있어서, 상기 연관 손실(L_0)은

상기 미니 배치(B')의 다수의 노드(v_i, v_j, v_n, v_t)로부터 수학식

$$\mathcal{L}_0 = \frac{1}{|B'|} \sum_{(v_i, v_j, v_n, v_t) \in B'} (E_{ij}^+ + D_{in} E_{in}^-)$$

(여기서 $|B'|$ 는 미니 배치(B')를 구상하기 위해 구성된 대상 노드(v_i), 이웃 노드(v_j), 2홉 노드(v_t), 그리고 원격 노드(v_n) 모임의 총 갯수이고, E_{ij}^+ 는 상기 대상 노드(v_i)와 상기 이웃 노드(v_j)가 임베딩 공간에 배치된 대상 임베딩 노드(z_i)와 이웃 임베딩 노드(z_j) 사이의 양의 포괄 에너지 함수로서 $E_{ij}^+ = \phi(\text{sim}(\mathbf{z}_i, \mathbf{z}_j))$ 로 계산되고, E_{in}^- 는 상기 대상 노드(v_i)와 상기 원격 노드(v_n)가 임베딩 공간에 배치된 대상 임베딩 노드(z_i)와 이웃 임베딩 노드(z_t) 사이의 음의 포괄 에너지 함수로서 $E_{in}^- = \phi(-\text{sim}(\mathbf{z}_i, \mathbf{z}_n))$ 로 계산되며, $\text{sim}(\circ)$ 은 미리 지정된 유사도 함수이고, $\phi(\circ)$ 는 미리 지정된 양의 상수 파라미터(γ, b)에 따라 $\phi(x) = \gamma^{-1} \log(1 + \exp(-\gamma x + b))$ 로 계산되는 유사도 가중치이다. 그리고 D_{in} 은

$D_{in} = \exp\left(\frac{\beta}{d_{sp}(v_i, v_n)}\right)$ 로 계산되는 대상 노드(v_i)와 원격 노드(v_n) 사이의 최단 경로 $d_{sp}(v_i, v_n)$ 에 포함되는 엡지 개수에 따라 가변되는 경로 거리 가중치로서, β 는 대상 노드(v_i)와 원격 노드(v_n)의 노드 쌍에 의한 영향력을 조절하기 위한 제어 파라미터)

에 따라 계산되는 엡지리스 네트워크 임베딩 장치.

청구항 8

제7항에 있어서, 상기 공통 이웃 손실(L_{2nd})은

수학식

$$\mathcal{L}_{2nd} = \frac{1}{|B'|} \sum_{(v_i, v_j, v_n, v_t) \in B'} J_{it} E_{it}^+$$

(여기서 J_{it} 는 대상 노드(v_i)와 2홉 노드(v_t) 사이의 2차 근접 정도를 측정하는 자카드 유사도이고, E_{it}^+ 는 대상 노드(v_i)와 2홉 노드(v_t)가 임베딩 공간에 배치된 대상 임베딩 노드(z_i)와 2홉 임베딩 노드(z_t) 사이의 양의 포괄 에너지 함수)

에 따라 계산되는 엡지리스 네트워크 임베딩 장치.

청구항 9

다수의 노드와 상기 다수의 노드 사이의 연관 관계에 따라 대응하는 노드를 연결하는 다수의 엣지를 포함하는 네트워크 데이터를 가상의 임베딩 공간에 투영하는 네트워크 임베딩 장치의 엣지리스 네트워크 임베딩 방법에 있어서,

상기 네트워크 데이터에서 다수의 노드 각각 대한 속성 정보를 획득하는 단계;

기지정된 방식으로 상기 다수의 노드 각각에 미리 지정된 속성 사이의 유사도를 계산하는 단계;

판별된 유사도에 따라 상기 다수의 노드 중 대응하는 노드를 유사도 엣지로 서로 연결하여 유사도 네트워크 데이터를 생성하는 단계; 및

미리 학습된 인공 신경망을 이용하여, 학습된 신경망 연산 방식에 따라 상기 유사도 네트워크 데이터로부터 상기 네트워크 데이터의 노드 사이의 연관 관계를 추정하고, 추정된 연관 관계에 따라 상기 다수의 노드를 가상의 임베딩 공간 상에 배치하는 단계를 포함하는 엣지리스 네트워크 임베딩 방법.

청구항 10

제9항에 있어서, 상기 유사도를 계산하는 단계는

상기 다수의 노드 사이를 잇는 엣지와 무관하게 상기 다수의 노드 각각에 대해 미리 설정된 속성 정보를 기반으로 유사도를 계산하는 엣지리스 네트워크 임베딩 방법.

청구항 11

제9항에 있어서, 상기 유사도 네트워크 데이터를 생성하는 단계는

상기 다수의 노드 각각에 대해 가장 높은 유사도를 갖는 기지정된 개수의 노드를 상기 유사도 엣지로 서로 연결하여 유사도 네트워크 데이터를 생성하는 엣지리스 네트워크 임베딩 방법.

청구항 12

제9항에 있어서, 상기 엣지리스 네트워크 임베딩 방법은

상기 네트워크 임베딩 장치를 실제 이용하기 이전, 상기 인공 신경망을 학습시키기 위한 학습 단계를 더 구비하고,

상기 학습 단계는

임베딩 공간에 노드들 사이의 거리를 에너지 크기로 가정하는 에너지 기반 학습에 기반하여, 엣지로 연결되지 않은 엣지리스 노드가 포함되지 않은 네트워크 데이터를 학습 데이터에서 엣지에 따른 노드간 연관 관계가 클수록 임베딩 공간 상에서 노드 쌍 사이의 에너지가 줄어들도록 서로 가깝게 위치시키고, 연관 관계가 작을수록 노드 쌍 사이의 에너지가 증가되도록 멀리 위치하도록 학습시키면서, 공통으로 연결되는 노드의 수가 많은 경우에도 에너지가 줄어들어 가깝게 위치하도록 설정된 손실을 계산하고, 계산된 손실을 상기 인공 신경망으로 역전파하는 엣지리스 네트워크 임베딩 방법.

청구항 13

제12항에 있어서, 상기 학습 단계는

상기 학습 데이터에서 다수의 노드 중 하나를 대상 노드(v_i)로 선택하고, 선택된 대상 노드(v_i)와 엣지를 통해 직접 연결된 이웃 노드(v_j)와 이웃 노드를 통해 연결되는 2홉 노드(v_l) 및 2홉을 초과하는 원격 노드(v_n)를 획득하여 미니 배치(B')를 추출하고, 상기 미니 배치에 포함된 다수의 노드(v_i, v_j, v_n, v_l)가 엣지로 연결된 관계에 따른 에너지를 상기 손실로 계산하는 엣지리스 네트워크 임베딩 방법.

청구항 14

제13항에 있어서, 상기 학습 단계는

상기 미니 배치(B')의 상기 대상 노드(v_i)로부터 상기 이웃 노드(v_j)와 상기 원격 노드(v_n) 사이의 엣지 수에 따라 임베딩 공간에서 상기 대상 노드(v_i)와 상기 이웃 노드(v_j)가 서로 인접하고 상기 원격 노드(v_n)는 엣지 수만

큼 이격되어 배치되도록 하는 연관 손실(L_0)을 계산하는 단계;

상기 대상 노드(v_i)와 상기 2홉 노드(v_t)에서 공통의 이웃 노드 수에 따라 상기 2홉 노드(v_t)가 상기 대상 노드(v_i)에 인접하여 배치되도록 하는 공통 이웃 손실(L_{2nd})을 계산하는 단계; 및

상기 연관 손실(L_0)과 상기 공통 이웃 손실(L_{2nd})로부터 수학적식

$$\mathcal{L} = \mathcal{L}_0 + \alpha \mathcal{L}_{2nd}$$

(여기서 α 는 손실 가중치)

에 따라 상기 손실을 계산하여 역전파하는 단계를 포함하는 엡지리스 네트워크 임베딩 방법.

청구항 15

제14항에 있어서, 상기 연관 손실(L_0)을 계산하는 단계는

상기 연관 손실(L_0)을 상기 미니 배치(B')의 다수의 노드(v_i, v_j, v_n, v_t)로부터 수학적식

$$\mathcal{L}_0 = \frac{1}{|B'|} \sum_{(v_i, v_j, v_n, v_t) \in B'} (E_{ij}^+ + D_{in} E_{in}^-)$$

(여기서 $|B'|$ 는 미니 배치(B')를 구상하기 위해 구성된 대상 노드(v_i), 이웃 노드(v_j), 2홉 노드(v_t), 그리고 원격 노드(v_n) 모임의 총 갯수이고, E_{ij}^+ 는 상기 대상 노드(v_i)와 상기 이웃 노드(v_j)가 임베딩 공간에 배치된 대상 임베딩 노드(z_i)와 이웃 임베딩 노드(z_j) 사이의 양의 포괄 에너지 함수로서 $E_{ij}^+ = \phi(\text{sim}(\mathbf{z}_i, \mathbf{z}_j))$ 로 계산되고, E_{in}^- 는 상기 대상 노드(v_i)와 상기 원격 노드(v_n)가 임베딩 공간에 배치된 대상 임베딩 노드(z_i)와 이웃 임베딩 노드(z_t) 사이의 음의 포괄 에너지 함수로서 $E_{in}^- = \phi(-\text{sim}(\mathbf{z}_i, \mathbf{z}_n))$ 로 계산되며, $\text{sim}(\circ)$ 은 미리 지정된 유사도 함수이고, $\phi(\circ)$ 는 미리 지정된 양의 상수 파라미터(γ, b)에 따라 $\phi(x) = \gamma^{-1} \log(1 + \exp(-\gamma x + b))$ 로 계산되는 유사도 가중치이다. 그리고 D_{in} 은

$D_{in} = \exp\left(\frac{\beta}{d_{sp}(v_i, v_n)}\right)$ 로 계산되는 대상 노드(v_i)와 원격 노드(v_n) 사이의 최단 경로 $d_{sp}(v_i, v_n)$ 에 포함되는 엡지 개수에 따라 가변되는 경로 거리 가중치로서, β 는 대상 노드(v_i)와 원격 노드(v_n)의 노드 쌍에 의한 영향력을 조절하기 위한 제어 파라미터)

에 따라 계산하는 엡지리스 네트워크 임베딩 방법.

청구항 16

제15항에 있어서, 상기 공통 이웃 손실(L_{2nd})을 계산하는 단계는

상기 공통 이웃 손실(L_{2nd})을 수학적식

$$\mathcal{L}_{2nd} = \frac{1}{|B'|} \sum_{(v_i, v_j, v_n, v_t) \in B'} J_{it} E_{it}^+$$

(여기서 J_{it} 는 대상 노드(v_i)와 2홉 노드(v_t) 사이의 2차 근접 정도를 측정하는 자카드 유사도(Jaccard similarity)이고, E_{it}^+ 는 대상 노드(v_i)와 2홉 노드(v_t)가 임베딩 공간에 배치된 대상 임베딩 노드(z_i)와 2홉 임베딩 노드(z_t) 사이의 양의 포괄 에너지 함수)

에 따라 계산되는 엣지리스 네트워크 임베딩 방법.

발명의 설명

기술 분야

[0001] 본 발명은 엣지리스 네트워크 임베딩 장치 및 방법에 관한 것으로, 그래프 인공 신경망 기반 엣지리스 네트워크 임베딩 장치 및 방법에 관한 것이다.

배경 기술

[0002] 네트워크 데이터(또는 그래프 데이터)는 소셜 네트워크, 관계형 네트워크, 분자 구조 및 추천 시스템과 같은 다양한 객체와 객체들 간의 관계를 설명하기에 매우 유용한 데이터이다. 네트워크 데이터는 각 객체에 해당하는 노드와 노드들 사이의 관계에 따라 노드를 연결하는 엣지로 구성되어 여러 객체 간의 연관성을 분석할 수 있도록 한다. 일 예로 소셜 네트워크에서 노드는 사용자이고 엣지는 사용자 사이의 관계(예를 들면 친구)를 나타낼 수 있으며, 관계형 네트워크에서는 노드가 개별 논문이고 엣지는 인용 관계로 나타낼 수 있다. 또한 추천 시스템에서 노드는 사용자 또는 제품일 수 있으며, 엣지는 추천 관계로 나타낼 수 있다.

[0003] 최근 그래프 신경망(Graph Neural Network: 이하 GNN)과 같은 인공 신경망을 활용한 네트워크 임베딩(Network Embedding) 기술은 네트워크 데이터에서 다수의 노드 각각을 벡터화함으로써 각 노드들 사이의 유사성을 표현할 수 있도록 발전하였다. 네트워크 임베딩 방식을 이용하면, 다수의 노드 각각이 벡터화되어 저차원의 가상의 임베딩 공간 상에 배치되며, 임베딩 공간 상에 배치된 노드들 사이의 거리를 기반으로 각 노드들 사이의 유사성을 인지할 수 있도록 한다. 이러한 네트워크 임베딩 방식은 네트워크 데이터에 특화되어 설계된 모델로서, 각 노드 사이에 연결된 엣지를 통해 정보를 교환함으로써 노드들의 정보와 네트워크의 위상 정보를 결합하여 임베딩 공간에 표현되도록 함으로써 각 노드의 새로운 상태를 용이하게 업데이트할 수 있도록 한다.

[0004] 도 1은 네트워크 데이터의 노드 상태가 업데이트되는 개념을 설명하기 위한 도면이다.

[0005] 도 1의 (a)에 도시된 바와 같이, 5개의 노드(v)가 존재하고 5개의 노드(v) 중 서로 연관이 있는 노드들은 엣지(e)를 통해 서로 연결되어 있다. 그리고 각 노드는 엣지를 통해 연결된 노드들과 정보 교환을 수행하여 업데이트된다. 도 1에서는 제3 노드(3)가 연결된 제1, 제2 및 제5 노드(1, 2, 5)에서 엣지(e)를 통해 전달된 정보에 따라 제3' 노드(3')로 업데이트되는 과정을 도시하였다. 그리고 이러한 노드 업데이트 과정은 업데이트된 노드들(여기서는 제3' 노드(3'))를 기반으로 다시 엣지를 통해 다른 노드들로 정보가 전달되어 (b)와 같이 다른 노드들(여기서는 제1', 제2', 제4' 및 제5' 노드(1', 2', 4', 5')) 또한 업데이트되는 과정이 반복됨으로써 모든 노드들이 업데이트될 수 있다.

[0006] 이때 인공 신경망을 이용한 네트워크 임베딩 방식은 노드들 사이의 관계에 기반하여 각 노드를 벡터화한다. 즉 노드들 사이를 잇는 엣지를 분석함으로써 임베딩 공간 상에서 해당 노드의 위치를 벡터 형태로 결정한다. 따라서 노드를 임베딩 공간에 임베딩하기 위해서는 다른 노드와의 연관 관계, 즉 엣지가 우선 구성되어야 한다.

[0007] 그러나 일부 노드들의 경우, 다른 노드와의 연관 관계를 갖지 않는 경우가 있다. 대표적인 예로서 네트워크 데이터에서 신규로 생성된 노드(예를 들면 소셜 네트워크에서 친구 관계 비공개 설정 사용자와 신규 사용자 가입 등)가 발생하는 경우 등이 있다. 이와 같이 다른 노드와 연관 관계를 갖지 않아 다른 노드와 엣지로 연결되지 않는 노드를 엣지리스 노드(edgeless node)라 한다.

[0008] 도 2는 네트워크 데이터에서 엣지리스 노드의 일 예를 나타낸다.

[0009] 도 2에서 제6 및 제7 노드(6, 7)과 같이 다른 노드들과 엣지로 연결되지 않은 엣지리스 노드의 경우, 기존의 인공 신경망을 이용하는 네트워크 임베딩 방식에서는 해당 노드를 임베딩 공간에 투영할 수 없다는 한계가 있다. 즉 노드들 사이의 연관 관계가 불명확하므로, 엣지리스 노드를 임베딩할 수 없다는 문제가 있다.

선행기술문헌

특허문헌

[0010] (특허문헌 0001) 한국 공개 특허 제10-2021-0043881호 (2021.04.22 공개)

발명의 내용

해결하려는 과제

- [0011] 본 발명의 목적은 네트워크 데이터의 엣지리스 노드를 임베딩할 수 있는 엣지리스 네트워크 임베딩 장치 및 방법을 제공하는데 있다.
- [0012] 본 발명의 다른 목적은 기존의 노드들과의 연관 관계가 생성되지 않은 노드를 별도의 추가 학습 또는 재학습 없이 임베딩할 수 있는 엣지리스 네트워크 임베딩 장치 및 방법을 제공하는데 있다.

과제의 해결 수단

- [0013] 상기 목적을 달성하기 위한 본 발명의 일 실시예에 따른 엣지리스 네트워크 임베딩 장치는 다수의 노드와 상기 다수의 노드 사이의 연관 관계에 따라 대응하는 노드를 연결하는 다수의 엣지를 포함하는 네트워크 데이터에서 다수의 노드 각각에 대한 속성 정보를 획득하는 네트워크 획득부; 상기 다수의 노드 각각에 미리 지정된 속성 사이의 유사도를 기지정된 방식으로 계산하는 노드 유사도 판별부; 판별된 유사도에 따라 상기 다수의 노드 중 대응하는 노드를 유사도 엣지로 서로 연결하여 유사도 네트워크 데이터를 생성하는 유사도 네트워크 구성부; 및 미리 학습된 인공 신경망으로 구현되어, 학습된 신경망 연산 방식에 따라 상기 유사도 네트워크 데이터로부터 상기 네트워크 데이터의 노드 사이의 연관 관계를 추정하고, 추정된 연관 관계에 따라 상기 다수의 노드를 가상의 임베딩 공간 상에 배치하는 네트워크 임베딩부를 포함한다.
- [0014] 상기 노드 유사도 판별부는 상기 다수의 노드 사이를 잇는 엣지와 무관하게 상기 다수의 노드 각각에 대해 미리 설정된 속성 정보를 기반으로 유사도를 계산할 수 있다.
- [0015] 상기 유사도 네트워크 구성부는 상기 다수의 노드 각각에 대해 가장 높은 유사도를 갖는 기지정된 개수의 노드를 상기 유사도 엣지로 서로 연결하여 유사도 네트워크 데이터를 생성할 수 있다.
- [0016] 상기 네트워크 임베딩부는 임베딩 공간에 노드들 사이의 거리를 에너지 크기로 가정하는 에너지 기반 학습에 기반하여, 엣지로 연결되지 않은 엣지리스 노드가 포함되지 않은 네트워크 데이터를 학습 데이터에서 엣지에 따른 노드간 연관 관계가 클수록 임베딩 공간 상에서 노드 쌍 사이의 에너지가 줄어들도록 서로 가깝게 위치시키고, 연관 관계가 작을수록 노드 쌍 사이의 에너지가 증가되도록 멀리 위치하도록 학습시키면서, 공통으로 연결되는 노드의 수가 많은 경우에도 에너지가 줄어들어 가깝게 위치하도록 설정된 손실을 계산하고, 계산된 손실이 역전파되어 학습될 수 있다.
- [0017] 상기 네트워크 임베딩부는 상기 학습 데이터에서 다수의 노드 중 하나를 대상 노드(v_i)로 선택하고, 선택된 대상 노드(v_i)와 엣지를 통해 직접 연결된 이웃 노드(v_j)와 이웃 노드를 통해 연결되는 2홉 노드(v_l) 및 2홉을 초과하는 원격 노드(v_n)를 획득하여 미니 배치(B')를 추출하고, 상기 미니 배치에 포함된 다수의 노드(v_i, v_j, v_n, v_l)가 엣지로 연결된 관계에 따른 에너지를 상기 손실로 계산할 수 있다.
- [0018] 상기 손실(L)은 상기 미니 배치(B')의 상기 대상 노드(v_i)로부터 상기 이웃 노드(v_j)와 상기 원격 노드(v_n) 사이의 엣지 수에 따라 임베딩 공간에서 상기 대상 노드(v_i)와 상기 이웃 노드(v_j)가 서로 인접하고 상기 원격 노드(v_n)는 엣지 수만큼 이격되어 배치되도록 하는 연관 손실(L_0)과 상기 대상 노드(v_i)와 상기 2홉 노드(v_l)에서 공통의 이웃 노드 수에 따라 상기 2홉 노드(v_l)가 상기 대상 노드(v_i)에 인접하여 배치되도록 하는 공통 이웃 손실(L_{2nd})로부터 수학적

$$\mathcal{L} = \mathcal{L}_0 + \alpha \mathcal{L}_{2nd}$$

[0019]

[0020] (여기서 α 는 손실 가중치)에 따라 계산될 수 있다.

[0021]

상기 목적을 달성하기 위한 본 발명의 다른 실시예에 따른 다수의 노드와 상기 다수의 노드 사이의 연관 관계에 따라 대응하는 노드를 연결하는 다수의 엣지를 포함하는 네트워크 데이터를 가상의 임베딩 공간에 투영하는 네트워크 임베딩 장치의 엣지리스 네트워크 임베딩 방법은 상기 네트워크 데이터에서 다수의 노드 각각에 대한 속성 정보를 획득하는 단계; 기지정된 방식으로 상기 다수의 노드 각각에 미리 지정된 속성 사이의 유사도를 계산하는 단계; 판별된 유사도에 따라 상기 다수의 노드 중 대응하는 노드를 유사도 엣지로 서로 연결하여 유사도 네

트위크 데이터를 생성하는 단계; 및 미리 학습된 인공 신경망을 이용하여, 학습된 신경망 연산 방식에 따라 상기 유사도 네트워크 데이터로부터 상기 네트워크 데이터의 노드 사이의 연관 관계를 추정하고, 추정된 연관 관계에 따라 상기 다수의 노드를 가상의 임베딩 공간 상에 배치하는 단계를 포함한다.

발명의 효과

[0022] 따라서, 본 발명의 실시예에 따른 엡지리스 네트워크 임베딩 장치 및 방법은 네트워크 데이터에서 기존의 노드들과의 연관 관계가 생성되지 않은 신규 노드와 같은 엡지리스 노드를 추가 학습 또는 재학습 없이 효과적으로 임베딩 공간에 임베딩할 수 있다. 그러므로 엡지리스 노드들과 다른 노드들과의 유사도를 분석할 수 있도록 하여 소셜 네트워크나 추천 시스템과 같은 네트워크 데이터를 이용하는 다양한 시스템에서 다른 사용자나 제품과의 연관성이 불명확한 사용자에게도 친구 추천이나 상품 추천과 같은 적합한 서비스를 제공할 수 있도록 한다.

도면의 간단한 설명

[0023] 도 1은 네트워크 데이터의 노드 상태가 업데이트되는 개념을 설명하기 위한 도면이다.
 도 2는 네트워크 데이터에서 엡지리스 노드의 일 예를 나타낸다.
 도 3은 본 발명의 일 실시예에 따른 엡지리스 네트워크 임베딩 장치의 개략적 구조를 나타낸다.
 도 4는 본 실시예의 엡지리스 네트워크 임베딩 장치가 노드 속성 유사도에 기반하여 네트워크를 재구성하는 과정을 나타낸다.
 도 5는 도 1의 유사도 네트워크 구성부가 엡지 리스 노드를 포함하는 유사도 네트워크 데이터를 구성하는 과정을 설명하기 위한 도면이다.
 도 6은 도 1의 네트워크 임베딩부가 유사도 네트워크 데이터를 네트워크 임베딩하는 과정을 설명하기 위한 도면이다.
 도 7은 학습부가 본 실시예의 네트워크 임베딩부를 학습시키는 방법을 설명하기 위한 도면이다.
 도 8은 본 발명의 일 실시예에 따른 엡지리스 네트워크 임베딩 방법을 나타낸다.

발명을 실시하기 위한 구체적인 내용

[0024] 본 발명과 본 발명의 동작상의 이점 및 본 발명의 실시예에 의하여 달성되는 목적을 충분히 이해하기 위해서는 본 발명의 바람직한 실시예를 예시하는 첨부 도면 및 첨부 도면에 기재된 내용을 참조하여야만 한다.

[0025] 이하, 첨부한 도면을 참조하여 본 발명의 바람직한 실시예를 설명함으로써, 본 발명을 상세히 설명한다. 그러나, 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며, 설명하는 실시예에 한정되는 것이 아니다. 그리고, 본 발명을 명확하게 설명하기 위하여 설명과 관계없는 부분은 생략되며, 도면의 동일한 참조부호는 동일한 부재임을 나타낸다.

[0026] 명세서 전체에서, 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라, 다른 구성요소를 더 포함할 수 있는 것을 의미한다. 또한, 명세서에 기재된 "...부", "...기", "모듈", "블록" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.

[0027] 도 3은 본 발명의 일 실시예에 따른 엡지리스 네트워크 임베딩 장치의 개략적 구조를 나타내고, 도 4는 본 실시예의 엡지리스 네트워크 임베딩 장치가 속성 유사도에 기반하여 네트워크를 재구성하는 과정을 나타내며, 도 5는 네트워크 데이터에 신규 노드가 입력된 경우의 일 예를 나타낸다. 그리고 도 6은 도 1의 네트워크 임베딩부가 유사도 네트워크 데이터를 네트워크 임베딩하는 과정을 설명하기 위한 도면이다.

[0028] 도 3을 참조하면, 본 실시예의 엡지리스 네트워크 임베딩 장치는 네트워크 획득부(100), 노드 유사도 판별부(200), 유사도 네트워크 구성부(300) 및 네트워크 임베딩부(400)를 포함할 수 있다.

[0029] 네트워크 획득부(100)는 임베딩 공간 상에 노드간 유사도에 따라 각 노드가 투영될 네트워크 데이터를 획득하고, 획득된 네트워크 데이터에서 노드들(v) 사이의 연관 관계를 나타내는 엡지(e)를 제거하고 노드 속성 정보를 분석하여 획득한다. 본 실시예에서 네트워크 획득부(100)는 네트워크 데이터가 저장된 데이터베이스와 같은 저장 장치로 구현되거나 네트워크 데이터를 전송받는 통신 모듈 등으로 구현될 수 있다.

- [0030] 네트워크 획득부(100)는 노드 추출부(110)와 노드 속성 분석부(120)를 포함할 수 있다. 노드 추출부(110)는 도 4의 (a)에 도시된 바와 같이, 다수의 노드(v)와 다수의 노드(v) 사이의 연관 관계에 따라 서로 연관된 노드(v) 사이를 잇는 엣지(e)로 구성된 네트워크 데이터를 획득할 수 있다. 상기한 바와 같이 네트워크 데이터는 다수의 노드(v)와 다수의 엣지(e)로 구성되므로, 노드 추출부(110)는 이미 구성된 네트워크 데이터를 획득할 수 있다.
- [0031] 그리고 본 실시예에서 노드 추출부(110)는 노드(v)와 엣지(e)로 구성된 네트워크 데이터가 획득되면, 네트워크 데이터에서 노드(v) 사이를 잇는 엣지(e)를 모두 제거한다. 즉 도 4의 (b)에 도시된 바와 같이, 네트워크 데이터의 노드(v) 사이의 연관 관계를 나타내는 엣지(e)를 제거하여 다수의 노드(v)를 추출한다.
- [0032] 노드 속성 분석부(120)는 노드 추출부(110)에서 획득된 다수의 노드들(v) 각각에 대한 속성 정보를 획득한다. 네트워크 데이터에서 다수의 노드(v) 각각은 다른 노드들과의 연관 관계를 나타내지 않는 자체적인 속성 정보를 포함할 수 있다. 일 예로 소셜 네트워크에서 노드인 사용자는 다른 사용자와의 연관 관계를 나타내는 친구 설정 이외에도 회원 가입시에 다른 사용자와 전혀 무관한 생년 월일, 성별, 거주지와 같은 각종 정보를 제공하며, 이와 같이 노드 자체에 대한 정보를 여기서는 속성 정보라 한다. 이에 노드 속성 분석부(120)는 노드(v) 사이에 연관 관계와 무관한 노드 각각에 대한 속성 정보를 획득하여 분석한다.
- [0033] 한편, 네트워크 획득부(100)는 신규 노드 추가부(130)를 더 포함할 수 있다. 노드 추출부(110)가 획득하는 네트워크 데이터에는 다수의 노드(v)와 다수의 엣지(e)가 포함되고, 여기에는 신규로 생성된 노드(v)가 함께 포함될 수도 있다. 그러나 상기한 바와 같이, 신규 노드의 경우, 기존 네트워크 데이터를 구성하는 노드들과의 연관 관계가 불명확하여, 네트워크 데이터에 포함되는 것이 용이하지 않다. 따라서 본 실시예의 엣지리스 네트워크 임베딩 장치에서는 신규 노드를 추가해야 하는 경우, 신규 노드 추가부(130)가 속성 정보를 포함하는 노드를 추가로 생성하여 노드 속성 분석부(120)에 제공함으로써, 노드 속성 분석부(120)가 기존의 노드와 함께 신규 노드에 대한 속성을 분석하도록 할 수 있다.
- [0034] 노드 속성 분석부(120)는 각 노드(v)에 대한 속성을 분석하여, 각 노드(v)의 속성 정보를 미리 지정된 방식에 따라 벡터화하여 속성 벡터를 생성할 수도 있다.
- [0035] 노드 유사도 판별부(200)는 네트워크 획득부(100)의 노드 속성 분석부(120)에서 분석된 다수의 노드(v) 각각에 대한 속성 정보를 기반으로, 노드들(v) 사이의 유사도를 기지정된 방식으로 계산한다. 노드 유사도 판별부(200)는 일 예로 다수의 노드(v)들 사이의 속성 정보에 대한 코사인 유사도(cosine similarity)를 계산할 수 있다.
- [0036] 유사도 네트워크 구성부(300)는 노드 유사도 판별부(200)에서 판별된 각 노드별 속성 유사도에 기반하여 유사도 네트워크 데이터를 생성한다. 유사도 네트워크 구성부(300)는 노드 유사도 판별부(200)가 계산한 다수의 노드들(v) 사이의 유사도에 따라 다수의 노드(v) 각각에 대해 가장 높은 유사도를 갖는 k(일 예로 $k = 2$)개의 노드를 검출하고, 검출된 k개의 노드와 엣지로 연결하여 도 4의 (c)와 같이, 유사도 네트워크 데이터로 재구성한다.
- [0037] 도 4의 (c)의 유사도 네트워크 데이터는 도 4의 (a)에 도시된 입력 네트워크 데이터에서 엣지(e)로 나타나는 노드간 연관 관계가 아닌 노드의 속성 사이의 유사도를 연관 관계로 하는 새로운 네트워크 데이터를 재구성된 네트워크 데이터이다. 즉 도 4의 (a)에 나타난 엣지(e)와 (c)에 나타난 엣지(se)는 서로 무관하게 생성된 엣지이며, 이에 여기서는 (a)와 (c)에 도시된 엣지를 서로 구분하기 위해 (c)에 도시된 속성 사이의 유사도에 따라 생성된 엣지를 유사도 엣지(se)라고 한다.
- [0038] 네트워크 임베딩부(400)는 유사도 네트워크 데이터를 인가받아, 미리 학습된 신경망 연산 방식에 따라 유사도 네트워크 데이터의 다수의 노드를 가상의 임베딩 공간에 투영한다. 이때 네트워크 임베딩부(400)는 학습된 신경망 연산 방식에 따라 다수의 노드의 속성들 간의 유사도에 기반하여 생성된 유사도 네트워크 데이터로부터 입력된 네트워크 데이터에 나타난 노드간 연관 관계를 추정하고, 추정된 연관 관계에 따라 다수의 노드가 임베딩 공간에 위치하도록 학습된다. 즉 네트워크 임베딩부(400)는 입력된 네트워크 데이터의 노드간 연관 관계가 아닌 유사도 네트워크 데이터의 노드간 유사도를 기반으로 노드간 연관 관계를 추정하여 저차원의 임베딩 공간 상에 투영한다. 따라서 입력된 네트워크 데이터에서 엣지 정보가 없는 노드나 신규 생성 노드와 같은 엣지리스 노드에 대해서도 연관 관계를 추정하여 임베딩 공간에 배치할 수 있다.
- [0039] 이때 네트워크 임베딩부(400)는 추정된 연관 관계가 높은 노드들은 서로 인접하여 배치되고, 연관 관계가 낮은 노드들은 서로 이격되어 배치되도록 학습될 수 있다.

- [0040] 네트워크 임베딩부(400)는 미리 학습된 인공 신경망으로 구현될 수 있으며, 일 예로 GNN으로 구현될 수 있다.
- [0041] 상기한 바와 같이, 본 실시예에서는 네트워크 획득부(100)가 네트워크 데이터가 이미 확보하고 있는 노드 사이의 연관 관계에 대한 정보를 모두 제거하고, 다수의 노드 각각의 속성 정보만을 분석한다. 이는 이미 확인된 노드들 사이의 연관 관계 정보를 이용하지 않는다는 것을 의미한다.
- [0042] 따라서 도 5에 도시된 바와 같이, 노드들(1 ~ 5) 사이의 연관 관계로 구축된 네트워크 데이터에 다른 노드들과의 연관 관계가 존재하지 않는 신규 노드와 같은 엣지리스 노드(6, 7)가 추가되게 되면, 기존의 네트워크 임베딩 방식에서는 엣지리스 노드(6, 7)의 다른 노드와의 연관 관계가 존재하지 않아, 엣지리스 노드(6, 7)를 처리하기 어렵다는 문제가 있었다.
- [0043] 그러나 본 발명에서는 이와 같이 네트워크 획득부(100)가 연관 관계 정보를 미리 제거하고, 유사도 네트워크 구성부(300)가 도 5의 (b)에 도시된 바와 같이, 각 노드들의 속성에 기반하여 유사도 엣지로 노드들을 연결한 유사도 네트워크를 구축함으로써, 네트워크 임베딩부(400)가 기존 연관 관계 정보에 의한 편견없이, 노드들의 속성 유사도에 따라 네트워크 데이터의 노드들에 대한 연관 관계를 추정할 수 있도록 한다. 따라서 도 6에 도시된 바와 같이, 연관 관계가 존재하지 않는 엣지리스 노드들에 대한 연관 관계를 정확하게 추정하여 임베딩 공간에 배치할 수 있다.
- [0044] 한편, 본 실시예의 엣지리스 네트워크 임베딩 장치는 네트워크 임베딩부(400)의 학습 시에 학습부(600)를 더 포함할 수 있다. 상기한 바와 같이, 네트워크 임베딩부(400)는 인공 신경망으로 구현되므로, 엣지리스 네트워크 임베딩 장치를 실제로 사용하기 전에 미리 학습되어야 한다. 학습부(600)는 네트워크 임베딩부(400)를 학습시키기 위해 추가되는 구성으로, 네트워크 임베딩부(400)의 학습이 완료된 이후에는 생략될 수 있다.
- [0045] 도 7은 학습부가 본 실시예의 네트워크 임베딩부를 학습시키는 방법을 설명하기 위한 도면이다.
- [0046] 학습부(600)는 도 7에 도시된 바와 같이, 엣지리스 노드가 없는 네트워크 데이터(701)를 학습 데이터로서 네트워크 획득부(100)로 인가한다. 이때 상기한 바와 같이, 학습부(600)는 엣지가 제외된 노드(X)만을 네트워크 획득부(100)로 인가할 수 있다.
- [0047] 유사도 네트워크 구성부(300)는 네트워크 획득부(100)로 인가된 노드들의 속성 정보에 따라 유사도 네트워크 데이터(702)를 생성하여, 네트워크 임베딩부(400)로 전달한다. 네트워크 임베딩부(400)는 이전까지 학습된 신경망 연산 방식에 따라 유사도 네트워크 데이터(702)로부터 노드들 사이의 연관 관계를 추정하여 학습 데이터로서 입력된 네트워크 데이터의 다수의 노드들을 임베딩 공간에 배치한다.
- [0048] 이에 학습부(600)는 학습 데이터인 네트워크 데이터의 노드들 사이를 잇는 엣지에 따른 노드들 간의 관계와 네트워크 임베딩부(400)에 의해 임베딩 공간에 배치된 노드들 사이의 거리를 기반으로 손실(L)을 계산하여 네트워크 임베딩부(400)로 역전파함으로써, 네트워크 임베딩부(400)를 학습시킨다.
- [0049] 구체적으로 본 실시예에서 학습부(600)는 임베딩 공간에 노드들 사이의 거리를 에너지 크기로 가정하는 에너지 기반 학습에 기반하여, 학습 데이터에서 엣지에 따른 노드간 연관 관계가 클수록 임베딩 공간 상에서 노드 쌍 사이의 에너지가 줄어들도록 서로 가깝게 위치시키고, 연관 관계가 작을수록 노드 쌍 사이의 에너지가 증가되도록 멀리 위치하도록 학습시키면서, 공통으로 연결되는 노드의 수가 많은 경우에도 에너지가 줄어들어 가깝게 위치하도록 손실(L)을 수학적 1에 따라 설정하여 네트워크 임베딩부(400)를 학습시킨다.

수학적 1

$$\mathcal{L} = \mathcal{L}_0 + \alpha \mathcal{L}_{2nd}$$

- [0050]
- [0051] 여기서 \mathcal{L}_0 는 노드간 엣지를 통한 연관 관계에 따른 연관 손실을 나타내고, \mathcal{L}_{2nd} 는 공통으로 연결된 노드의 수에 따른 공통 이웃 손실을 나타내며, α 는 손실 가중치를 나타낸다.
- [0052] 이때 학습 데이터인 네트워크 데이터의 모든 노드들에 대한 손실(L)을 일괄하여 계산하는 것은 매우 많은 연산량을 요구할 뿐만 아니라, 동일한 연관 관계로 구성된 대량의 네트워크 데이터를 학습 데이터로 요구하므로 효율적인 학습을 수행하기 어렵게 한다. 이에 본 실시예에 따른 학습부(600)는 학습 데이터인 네트워크 데이터에서 특정 노드를 기반으로 미니 배치(B')를 추출하고, 추출된 미니 배치(B')를 기반으로 손실(L)을 계산하여 효

올적인 학습이 수행될 수 있도록 한다.

[0053] 학습부(600)는 네트워크 데이터의 다수의 노드들 중 임의의 노드를 대상 노드(v_i)로 선택하고, 선택된 대상 노드(v_i)와 엣지를 통해 직접 연결된 이웃 노드(v_j)와 이웃 노드를 통해 연결되는 2홉 노드(v_t) 및 2홉을 초과하는 원격 노드(v_n)를 획득하여 미니 배치(B')를 구성할 수 있다. 일 예로 도 7에 도시된 5개의 노드 중 제5 노드(5)가 대상 노드(v_i)로 선택된 경우, 이웃 노드(v_j)는 제3 및 제4 노드(3, 4)이고, 2홉 노드(v_t)는 제2 노드(2)이며, 원격 노드(v_n)는 제1 노드(1)이다.

[0054] 학습부(600)가 미니 배치(B')를 구성하여 학습을 수행하는 경우, 수학식 1에서 노드간 연관 관계에 따라 획득되는 연관 손실(L_0)은 2개의 노드가 엣지로 직접 연결된 이웃 노드이면 임베딩 공간 상에서 해당 노드들이 더 가깝게 위치하고, 2개의 노드가 3개 이상의 엣지를 통해 연결되는 원격 노드이면 2개의 노드 사이의 존재하는 엣지 수, 즉 홉(hop) 수에 따라 더 멀리 위치하도록 조절되는 손실로서 수학식 2로 정의될 수 있다.

수학식 2

$$\mathcal{L}_0 = \frac{1}{|B'|} \sum_{(v_i, v_j, v_n, v_t) \in B'} (E_{ij}^+ + D_{in} E_{in}^-)$$

[0055]

[0056] 수학식 2에서 $|B'|$ 는 미니 배치(B')를 구성하기 위해 구성된 대상 노드(v_i), 이웃 노드(v_j), 2홉 노드(v_t), 그리고 원격 노드(v_n) 모임의 총 갯수이고, E_{ij}^+ 는 대상 노드(v_i)와 이웃 노드(v_j)가 임베딩 공간에 배치된 대상 임베딩 노드(z_i)와 이웃 임베딩 노드(z_j) 사이의 양의 포괄 에너지 함수(generic energy function)로서 $E_{ij}^+ = \phi(\text{sim}(\mathbf{z}_i, \mathbf{z}_j))$ 로 계산되고, E_{in}^- 는 대상 노드(v_i)와 원격 노드(v_n)가 임베딩 공간에 배치된 대상 임베딩 노드(z_i)와 이웃 임베딩 노드(z_t) 사이의 음의 포괄 에너지 함수로서 $E_{in}^- = \phi(-\text{sim}(\mathbf{z}_i, \mathbf{z}_n))$ 로 계산된다. 여기서 $\text{sim}(\circ)$ 은 미리 지정된 유사도 함수(일 예로 코사인 유사도 함수)이고, $\phi(\circ)$ 는 미리 지정된 양의 상수 파라미터(γ, b)에 따라 $\phi(x) = \gamma^{-1} \log(1 + \exp(-\gamma x + b))$ 로 계산되는 유사도 가중치이다. 즉 $\phi(\circ)$ 는 유사도 함수($\text{sim}(\circ)$)에 의해 계산된 유사도에 따른 에너지(E)의 변화량을 조절한다.

[0057] 그리고 D_{in} 은 네트워크 데이터에서 대상 노드(v_i)와 원격 노드(v_n) 사이의 엣지를 통한 경로 거리에 따른 경로 거리 가중치로서, 대상 노드(v_i)와 원격 노드(v_n) 사이의 최단 경로 $d_{sp}(v_i, v_n)$ 에 포함되는 엣지 개수에 따라 가변되는 가중치이며, 수학식 3으로 계산될 수 있다.

수학식 3

$$D_{in} = \exp\left(\frac{\beta}{d_{sp}(v_i, v_n)}\right)$$

[0058]

[0059] 여기서 β 는 대상 노드(v_i)와 원격 노드(v_n)의 노드 쌍에 의한 영향력을 조절하기 위한 제어 파라미터로서 미리 지정된 상수값을 갖는다.

[0060] 한편, 수학식 1에서 공통 이웃 손실(L_{2nd})은 대상 노드(v_i)와 2홉 노드(v_t) 사이의 공통 이웃 노드(n_j)의 수에 따라 2홉 노드(v_t)가 임베딩 공간 상에서 대상 노드 노드(v_i)에 가깝게 배치되도록 손실을 적용하며, 수학식 4로 계산된다.

수학식 4

$$\mathcal{L}_{2nd} = \frac{1}{|\mathcal{B}'|} \sum_{(v_i, v_j, v_n, v_t) \in \mathcal{B}'} J_{it} E_{it}^+$$

[0061]

[0062]

여기서 J_{it} 는 대상 노드(v_i)와 2홉 노드(v_t) 사이의 2차 근접 정도를 측정하는 자카드 유사도(Jaccard similarity)이고, E_{it}^+ 는 대상 노드(v_i)와 2홉 노드(v_t)가 임베딩 공간에 배치된 대상 임베딩 노드(z_i)와 2홉 임베딩 노드(z_t) 사이의 양의 포괄 에너지 함수이다.

[0063]

수학식 4에 따르면, 2홉 노드(v_t)는 대상 노드(v_i)와 공통되는 이웃 노드(v_j)의 수가 많을수록 더 높은 가중치를 부여하는 방식으로 임베딩 공간에서 대상 노드(v_i)에 가깝게 배치된다.

[0064]

상기한 바와 같이 학습부(600)에 의해 에너지 기반 학습된 네트워크 임베딩부(400)는 도 7에 도시된 바와 같이, 이후 기존의 노드(X) 이외에 신규 노드(X')가 엣지 정보 없이 추가되더라도, 신규 노드(X')가 추가된 유사도 네트워크 데이터(703)를 인가받아 다수의 노드(X, X')를 임베딩 공간 상에 정확하게 배치할 수 있다. 즉 용이하게 네트워크 임베딩을 수행할 수 있다.

[0065]

도 8은 본 발명의 일 실시예에 따른 엣지리스 네트워크 임베딩 방법을 나타낸다.

[0066]

도 8을 참조하면, 본 실시예에 따른 엣지리스 네트워크 임베딩 방법은 크게 학습 단계(S10)와 임베딩 단계(S20)로 구분될 수 있다.

[0067]

학습 단계(S10)에서는 우선 엣지리스 노드가 없는 네트워크 데이터를 학습 데이터로서 획득한다(S11). 그리고 획득된 학습 데이터에서 엣지 정보를 제외하고, 다수의 노드(v)에 대한 속성 정보를 기반으로 다수의 노드 사이의 유사도를 판별한다(S12). 그리고 판별된 노드간 유사도에 기반하여 학습 유사도 네트워크 데이터를 생성한다(S13). 학습 유사도 네트워크 데이터가 생성되면, 인공 신경망을 이용하여 학습 유사도 네트워크 데이터를 이전까지 학습된 신경망 연산 방식에 따라 네트워크 임베딩하여 유사도 네트워크 데이터의 노드들을 임베딩 공간에 배치한다(S14).

[0068]

한편, 학습 데이터의 다수의 노드 중 하나를 대상 노드(v_i)로 선택하고, 선택된 대상 노드(v_i)와 엣지를 통해 직접 연결된 이웃 노드(v_j)와 이웃 노드를 통해 연결되는 2홉 노드(v_t) 및 2홉을 초과하는 원격 노드(v_n)를 획득하여 미니 배치(B')를 추출한다(S15).

[0069]

그리고 에너지 기반 학습 방법을 이용하여 미니 배치(B')의 대상 노드(v_i)로부터 이웃 노드(v_j)와 원격 노드(v_n) 사이의 엣지 수에 따라 임베딩 공간에서 대상 노드(v_i)와 이웃 노드(v_j)가 서로 인접하고 원격 노드(v_n)는 엣지 수만큼 이격되어 배치되도록 하는 연관 손실(L_0)을 수학식 2에 따라 계산하고, 대상 노드(v_i)와 2홉 노드(v_t)에서 공통의 이웃 노드 수에 따라 2홉 노드(v_t)가 대상 노드(v_i)에 인접하여 배치되도록 하는 공통 이웃 손실(L_{2nd})을 수학식 4에 따라 계산한다(S16).

[0070]

연관 손실(L_0)과 공통 이웃 손실(L_{2nd})이 계산되면 수학식 1에 따라 손실(L)을 계산하고, 계산된 손실(L)을 인공 신경망으로 역전파하여 학습을 수행한다(S17).

[0071]

그리고 학습 종료 여부를 판별한다(S18). 학습은 계산된 손실(L)이 기설정된 문턱값 미만이거나, 기설정된 횟수로 학습이 수행되면 종료될 수 있다. 만일 학습이 종료되지 않은 것으로 판별되면, 다시 학습 유사도 네트워크 데이터를 네트워크 임베딩하여 유사도 네트워크 데이터의 노드들을 임베딩 공간에 배치하여 학습을 수행할 수 있다(S14).

[0072]

한편 학습이 종료된 것으로 판별되면, 임베딩 단계를 수행할 수 있다(S20). 임베딩 단계에서는 네트워크 임베딩되어야 할 네트워크 데이터를 획득한다(S21). 네트워크 데이터에는 엣지리스 노드가 포함될 수 있다. 그리고 획득된 네트워크 데이터의 다수의 노드의 속성 정보를 분석하여, 분석된 다수의 노드(v)에 대한 속성 정보를 기반으로 다수의 노드 사이의 유사도를 판별한다(S22). 그리고 네트워크 데이터의 엣지(e)와 무관하게 판별된

노드간 유사도에 기반하여 다수의 노드를 유사도 엣지(se)로 연결하여 유사도 네트워크 데이터를 생성한다(S12). 유사도 네트워크 데이터가 생성되면, 학습 단계(S20)에서 학습된 인공 신경망을 이용하여 네트워크 데이터의 엣지(e)에 따른 다수의 노드(v) 사이의 연관 관계를 유사도 네트워크 데이터로부터 추정하고, 추정된 연관 관계에 따라 다수의 노드(v)를 네트워크 임베딩 하여 임베딩 공간에 배치한다(S24).

[0073] 본 발명에 따른 방법은 컴퓨터에서 실행시키기 위한 매체에 저장된 컴퓨터 프로그램으로 구현될 수 있다. 여기서 컴퓨터 판독가능 매체는 컴퓨터에 의해 액세스 될 수 있는 임의의 가용 매체일 수 있고, 또한 컴퓨터 저장 매체를 모두 포함할 수 있다. 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현된 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 모두 포함하며, ROM(판독 전용 메모리), RAM(랜덤 액세스 메모리), CD(컴팩트 디스크)-ROM, DVD(디지털 비디오 디스크)-ROM, 자기 테이프, 플로피 디스크, 광데이터 저장장치 등을 포함할 수 있다.

[0074] 본 발명은 도면에 도시된 실시예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다.

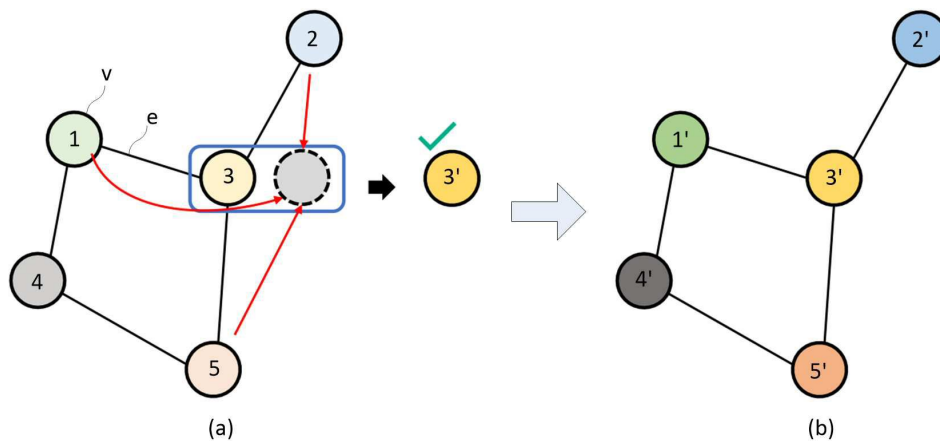
[0075] 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 청구범위의 기술적 사상에 의해 정해져야 할 것이다.

부호의 설명

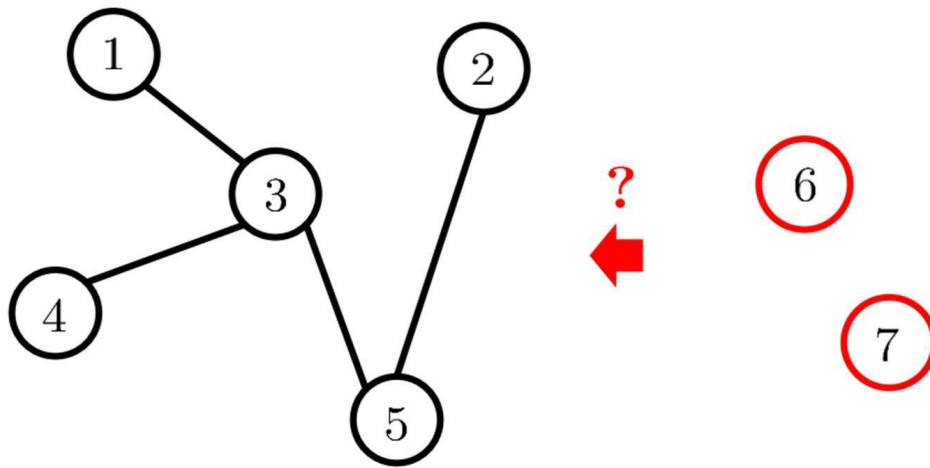
[0076] 100: 네트워크 획득부 110: 노드 추출부
120: 노드 속성 분석부 130: 신규 노드 추가부
200: 노드 유사도 판별부 300: 유사도 네트워크 구성부
400: 네트워크 임베딩부 500: 네트워크 재구축부
600: 학습부

도면

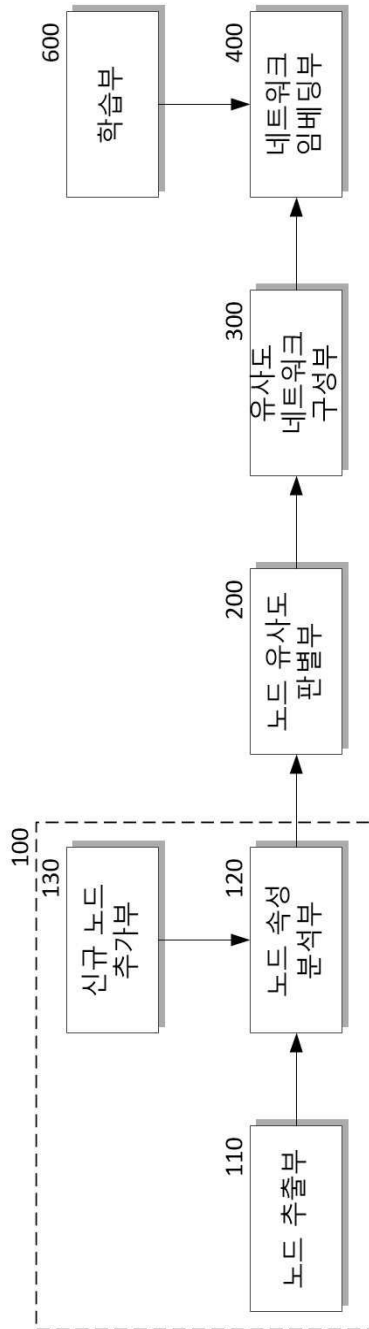
도면1



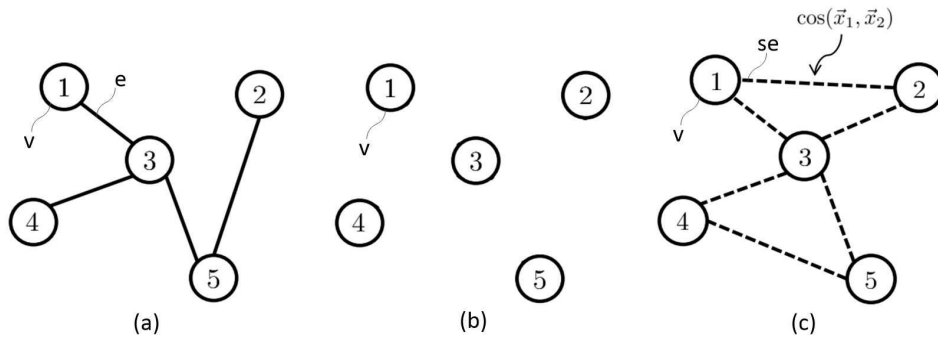
도면2



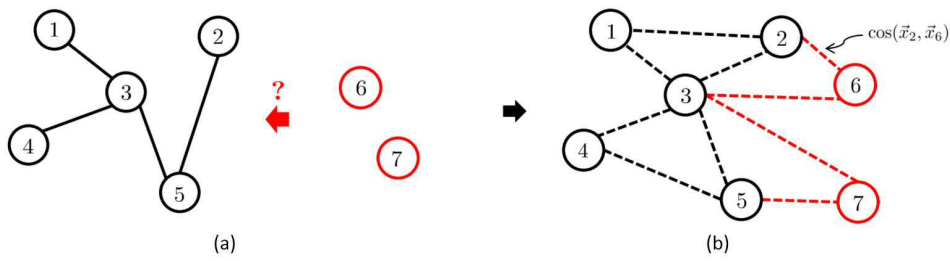
도면3



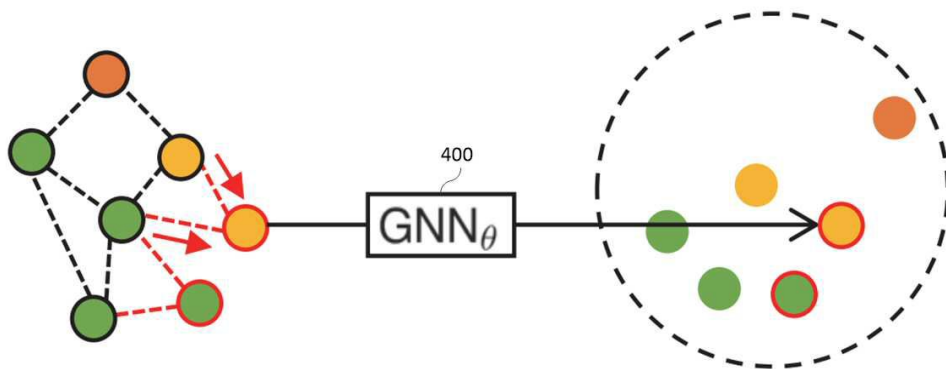
도면4



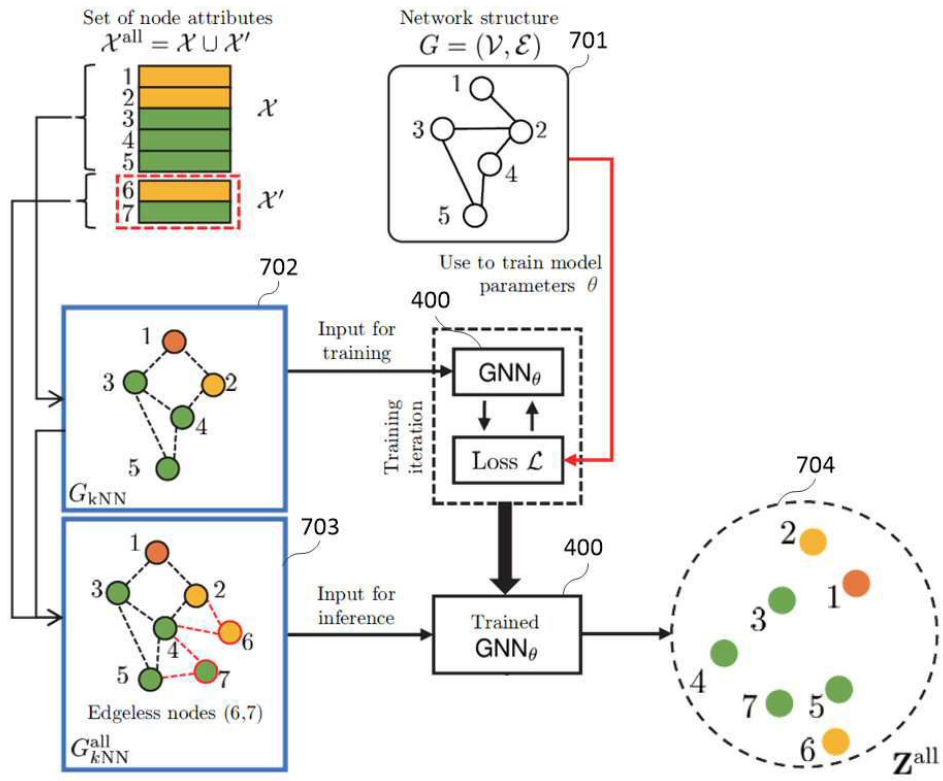
도면5



도면6



도면7



도면8

