

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)(11) 공개번호 10-2021-0043881
(43) 공개일자 2021년04월22일

(51) 국제특허분류(Int. Cl.)
G06N 3/08 (2006.01) G06N 3/04 (2006.01)
G06Q 50/30 (2012.01)
(52) CPC특허분류
G06N 3/08 (2013.01)
G06N 3/04 (2013.01)
(21) 출원번호 10-2019-0126749
(22) 출원일자 2019년10월14일
심사청구일자 2019년10월14일

(71) 출원인
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
(72) 발명자
신원용
서울특별시 서대문구 성산로 371, 102동 102호(연희동)
트란콩
경기도 용인시 수지구 죽전로168번길 3-1, 816호
(74) 대리인
민영준

전체 청구항 수 : 총 13 항

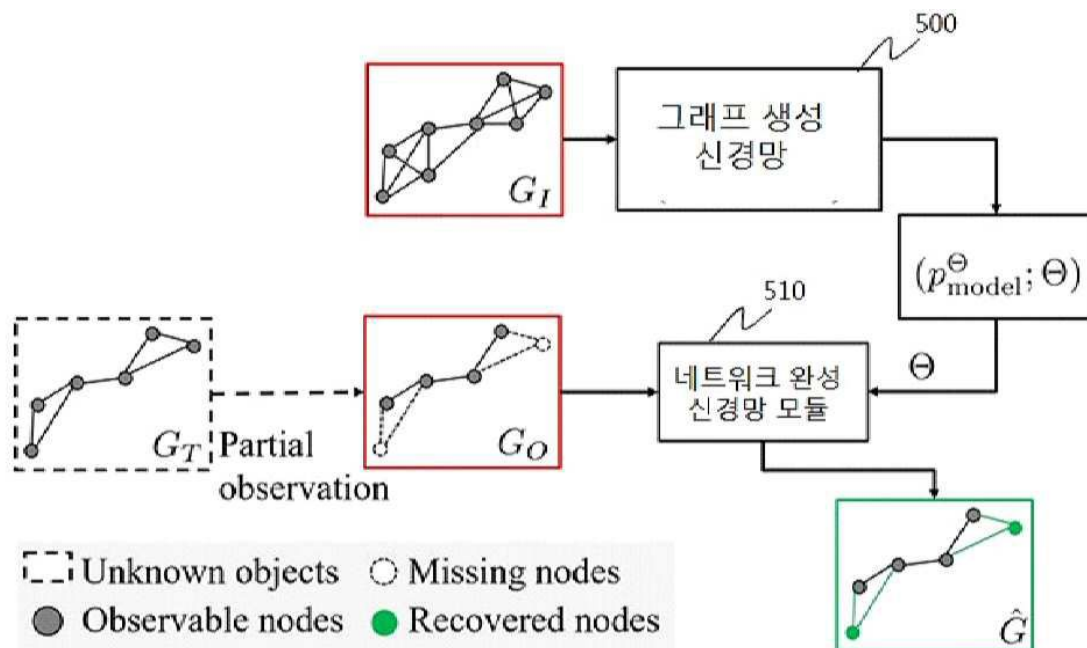
(54) 발명의 명칭 인공 신경망을 이용한 소셜 네트워크 완성 방법 및 장치

(57) 요약

인공 신경망을 이용한 소셜 네트워크 완성 방법 및 장치가 개시된다. 개시된 장치는, 미싱 노드가 노출되지 않은 대상 네트워크를 입력받아 미싱 노드의 연결 관계를 신경망을 통해 추론하며, 다양한 노드 순서에 따라 다수의 후보 완성 네트워크를 출력하는 신경망부; 및 상기 신경망부에서 출력하는 다수의 후보 완성 네트워크 중 어느

(뒷면에 계속)

대표도 - 도5



하나를 선정하는 선정하는 선정부를 포함하되, 상기 신경망부는 상기 대상 네트워크와 유사한 속성을 가지는 준거 네트워크들의 그래프 구조가 학습된 그래프 생성 신경망의 가중치를 이용하여 다수의 후보 완성 네트워크를 출력하며, 상기 선정부는 상기 학습된 그래프 생성 신경망으로부터 획득되는 연결 관계 확률 벡터를 이용하여 상기 연결 관계 확률 벡터에 기초할 때 상기 대상 네트워크와 확률적으로 가장 유사한 구조를 가지는 후보 완성 네트워크를 선정한다. 개시된 장치 및 방법에 의하면, 트루 데이터를 확보할 수 없는 소셜 네트워크의 한계를 극복하고 비교적 정확하게 미성 노드를 추론할 수 있는 장점이 있다.

(52) CPC특허분류

G06Q 50/01 (2013.01)

G06Q 50/30 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	2017RID1A1A09000835
부처명	교육부
과제관리(전문)기관명	한국연구재단
연구사업명	이공학개인지초연구지원사업
연구과제명	소셜 빅 데이터 마이닝을 통한 거대 스케일 포그 네트워크 설계 및 핵심 요소기술
개발(3/4)	
기 여 율	1/1
과제수행기관명	연세대학교 산학협력단
연구기간	2019.03.01 ~ 2020.02.29

명세서

청구범위

청구항 1

미싱 노드가 노출되지 않은 대상 네트워크를 입력받아 미싱 노드의 연결 관계를 신경망을 통해 추론하며, 다양한 노드 순서에 따라 다수의 후보 완성 네트워크를 출력하는 신경망부; 및

상기 신경망부에서 출력하는 다수의 후보 완성 네트워크 중 어느 하나를 선정하는 선정부를 포함하되, 상기 신경망부는 상기 대상 네트워크와 유사한 속성을 가지는 준거 네트워크들의 그래프 구조가 학습된 그래프 생성 신경망의 가중치를 이용하여 다수의 후보 완성 네트워크를 출력하며,

상기 선정부는 상기 학습된 그래프 생성 신경망으로부터 획득되는 연결 관계 확률 벡터를 이용하여 상기 연결 관계 확률 벡터에 기초할 때 상기 대상 네트워크와 확률적으로 가장 유사한 구조를 가지는 후보 완성 네트워크를 선정하는 것을 특징으로 하는 네트워크 완성 장치.

청구항 2

제1 항에 있어서,

상기 신경망부로는 상기 대상 네트워크의 가시 노드 및 추론하려는 미싱 노드에 대해 임의의 순서를 설정한 순서 정보가 입력되고, 상기 신경망부는 상기 순서 정보별로 후보 완성 네트워크를 출력하는 것을 특징으로 하는 네트워크 완성 장치.

청구항 3

제1 항에 있어서,

상기 학습되는 그래프 생성 신경망은 상기 준거 네트워크의 토폴로지에 관한 함수(f_{trans}) 및 연결 관계 확률(f_{ou})에 대한 함수를 학습하여 가중치를 설정하는 것을 특징으로 하는 네트워크 완성 장치.

청구항 4

제3 항에 있어서,

상기 학습되는 그래프 생성 신경망은 상기 토폴로지에 관한 함수 및 연결 관계 확률에 대한 함수를 이용하여 순차적으로 노드를 페인팅하여 그래프를 생성하는 것을 특징으로 하는 네트워크 완성 장치.

청구항 5

제1항에 있어서,

상기 연결 관계 확률 벡터는 상기 그래프 생성 신경망을 통해 생성되는 그래프 토폴로지의 연결 관계들에 대한 확률 정보인 것을 특징으로 하는 네트워크 완성 장치.

청구항 6

제1항에 있어서,

상기 선정부는 다음의 수학식에 기초하여 후보 완성 네트워크 중 하나를 선정하는 것을 특징으로 하는 네트워크

완성 장치.

$$\hat{G} = \operatorname{argmax}_G P(G|G_0, \Theta)$$

위 수학적식에서 G 는 후보 완성 네트워크들을 의미하고, G_0 는 입력되는 대상 네트워크를 의미하며, Θ 는 연결 관계 확률 벡터를 의미하고, \hat{G} 는 선정되는 후보 완성 네트워크를 의미함.

청구항 7

미싱 노드가 노출되지 않은 대상 네트워크를 입력받아 미싱 노드의 연결 관계를 신경망을 통해 추론하며, 다양한 노드 순서에 따라 다수의 후보 완성 네트워크를 출력하는 단계(a); 및

상기 단계(a)에서 출력하는 다수의 후보 완성 네트워크 중 어느 하나를 선정하는 선정하는 단계(b)를 포함하되,

상기 단계(a)는 상기 대상 네트워크와 유사한 속성을 가지는 준거 네트워크들의 그래프 구조가 학습된 그래프 생성 신경망의 가중치를 이용하여 다수의 후보 완성 네트워크를 출력하며,

상기 단계(b)는 상기 학습된 그래프 생성 신경망으로부터 획득되는 연결 관계 확률 벡터를 이용하여 상기 연결 관계 확률 벡터에 기초할 때 상기 대상 네트워크와 확률적으로 가장 유사한 구조를 가지는 후보 완성 네트워크를 선정하는 것을 특징으로 하는 네트워크 완성 방법.

청구항 8

제7 항에 있어서,

상기 단계(a)에서 상기 대상 네트워크의 가시 노드 및 추론하려는 미싱 노드에 대해 임의의 순서를 설정한 순서 정보가 입력되고, 상기 순서 정보별로 후보 완성 네트워크를 출력하는 것을 특징으로 하는 네트워크 완성 방법.

청구항 9

제7 항에 있어서,

상기 학습되는 그래프 생성 신경망은 상기 준거 네트워크의 토폴로지에 관한 함수(f_{trans}) 및 연결 관계 확률(f_{ou})에 대한 함수를 학습하여 가중치를 설정하는 것을 특징으로 하는 네트워크 완성 방법.

청구항 10

제9 항에 있어서,

상기 학습되는 그래프 생성 신경망은 상기 토폴로지에 관한 함수 및 연결 관계 확률에 대한 함수를 이용하여 순차적으로 노드를 페인팅하여 그래프를 생성하는 것을 특징으로 하는 네트워크 완성 방법.

청구항 11

제7항에 있어서,

상기 연결 관계 확률 벡터는 상기 그래프 생성 신경망을 통해 생성되는 그래프 토폴로지의 연결 관계들에 대한 확률 정보인 것을 특징으로 하는 네트워크 완성 방법.

청구항 12

제7항에 있어서,

상기 단계(b)는 다음의 수학식에 기초하여 후보 완성 네트워크 중 하나를 선정하는 것을 특징으로 하는 네트워크 완성 방법.

$$\hat{G} = \operatorname{argmax}_G P(G|G_0, \Theta)$$

위 수학식에서 G는 후보 완성 네트워크들을 의미하고, G₀는 입력되는 대상 네트워크를 의미하며, Θ는 연결 관계 확률 벡터를 의미하고, \hat{G} 는 선정되는 후보 완성 네트워크를 의미함.

청구항 13

제7항 내지 제12항 중 어느 한 항의 방법을 실행하기 위한 프로그램이 기록되어 있으며 디지털 데이터 처리 장치에 의해 판독 가능한 프로그램이 기록된 기록매체.

발명의 설명

기술 분야

[0001] 본 발명은 소셜 네트워크 완성 방법에 관한 것으로서, 더욱 상세하게는 인공 신경망을 이용한 소셜 네트워크 완성 방법 및 장치에 관한 것이다.

배경 기술

[0003] 근래에 들어 다양한 종류의 SNS 서비스가 제공되면서 사용자들은 SNS를 통해 다양한 친구 관계로 연결되어 있다. 이러한 소셜 네트워크에서의 사용자들과의 연결 관계는 다양한 용도로 유용하게 활용된다.

[0004] 예를 들어, 사용자들 간의 친구 관계를 이용하여 사용자 커뮤니티를 파악할 수 있으며, 이러한 커뮤니티는 사용자들에 대한 타겟 마케팅 및 사용자들의 성향 분석, 다운스트림 마이닝 등의 용도로 활용될 수 있다.

[0005] 그러나, 일부 사용자들은 친구 관계를 노출하지 않으며, 노출되지 않은 친구 관계로 인해 완벽한 소셜 네트워크 구조를 파악하는 것은 용이하지 않다. 노출되지 않은 친구 관계는 특정 소셜 네트워크가 가지는 구조적 특성을 명확히 파악하는데 장애 요인으로 작용하며, 이를 위해 노출되지 않은 친구 관계를 추론하려는 다양한 시도가 이루어졌다.

[0006] 이와 같이 소셜 네트워크에서 노출되지 않은 친구 관계를 추론하여 트루(True) 네트워크를 예측하는 것을 네트워크 완성이라고 한다. 노출되지 않은 친구 관계는 참값을 알 수 없는 정보이기에 정확한 네트워크 완성에는 본질적인 한계가 존재한다.

[0007] 네트워크 완성을 위한 다양한 알고리즘이 제시되었고, 대표적인 방법으로 크로네커 기반 네트워크 완성 방법이 있다. 그러나, 크로네커 기반 네트워크 완성 방법은 네트워크가 순수 멱 법칙(Power Law)을 따른다는 가정하에서 이루어지기에 네트워크가 이러한 멱 법칙 구조를 따르지 않는 속성을 가질 경우 매우 부정확한 네트워크 완성이 이루어지는 문제점이 있었다.

발명의 내용

해결하려는 과제

[0009] 본 발명은 트루 데이터를 확보할 수 없는 소셜 네트워크의 한계를 극복하고 비교적 정확하게 미싱 노드를 추론할 수 있는 인공 신경망을 이용한 네트워크 복원 방법 및 장치를 제안한다.

과제의 해결 수단

- [0011] 상기와 같은 목적을 달성하기 위해, 본 발명의 일 측면에 따르면, 미싱 노드가 노출되지 않은 대상 네트워크를 입력받아 미싱 노드의 연결 관계를 신경망을 통해 추론하며, 다양한 노드 순서에 따라 다수의 후보 완성 네트워크를 출력하는 신경망부; 및 상기 신경망부에서 출력하는 다수의 후보 완성 네트워크 중 어느 하나를 선정하는 선정부를 포함하되, 상기 신경망부는 상기 대상 네트워크와 유사한 속성을 가지는 준거 네트워크들의 그래프 구조가 학습된 그래프 생성 신경망의 가중치를 이용하여 다수의 후보 완성 네트워크를 출력하며, 상기 선정부는 상기 학습된 그래프 생성 신경망으로부터 획득되는 연결 관계 확률 벡터를 이용하여 상기 연결 관계 확률 벡터에 기초할 때 상기 대상 네트워크와 확률적으로 가장 유사한 구조를 가지는 후보 완성 네트워크를 선정하는 네트워크 완성 장치가 제공된다.
- [0012] 상기 신경망부로는 상기 대상 네트워크의 가시 노드 및 추론하려는 미싱 노드에 대해 임의의 순서를 설정한 순서 정보가 입력되고, 상기 신경망부는 상기 순서 정보별로 후보 완성 네트워크를 출력한다.
- [0013] 상기 학습되는 그래프 생성 신경망은 상기 준거 네트워크의 토폴로지에 관한 함수(f_{trans}) 및 연결 관계 확률(f_{ou})에 대한 함수를 학습하여 가중치를 설정한다.
- [0014] 상기 학습되는 그래프 생성 신경망은 상기 토폴로지에 관한 함수 및 연결 관계 확률에 대한 함수를 이용하여 순차적으로 노드를 페인팅하여 그래프를 생성한다.
- [0015] 상기 연결 관계 확률 벡터는 상기 그래프 생성 신경망을 통해 생성되는 그래프 토폴로지의 연결 관계들에 대한 확률 정보이다.
- [0016] 상기 선정부는 다음의 수학적식에 기초하여 후보 완성 네트워크 중 하나를 선정한다.

$$\hat{G} = \underset{G}{\operatorname{argmax}} P(G|G_o, \Theta)$$

- [0017]
- [0018] 위 수학적식에서 G 는 후보 완성 네트워크들을 의미하고, G_o 는 입력되는 대상 네트워크를 의미하며, Θ 는 연결 관계 확률 벡터를 의미하고, \hat{G} 는 선정되는 후보 완성 네트워크를 의미함.
- [0019] 본 발명의 다른 측면에 따르면, 미싱 노드가 노출되지 않은 대상 네트워크를 입력받아 미싱 노드의 연결 관계를 신경망을 통해 추론하며, 다양한 노드 순서에 따라 다수의 후보 완성 네트워크를 출력하는 단계(a); 및 상기 단계(a)에서 출력하는 다수의 후보 완성 네트워크 중 어느 하나를 선정하는 선정하는 단계(b)를 포함하되, 상기 단계(a)는 상기 대상 네트워크와 유사한 속성을 가지는 준거 네트워크들의 그래프 구조가 학습된 그래프 생성 신경망의 가중치를 이용하여 다수의 후보 완성 네트워크를 출력하며, 상기 단계(b)는 상기 학습된 그래프 생성 신경망으로부터 획득되는 연결 관계 확률 벡터를 이용하여 상기 연결 관계 확률 벡터에 기초할 때 상기 대상 네트워크와 확률적으로 가장 유사한 구조를 가지는 후보 완성 네트워크를 선정하는 네트워크 완성 방법이 제공된다.

발명의 효과

- [0021] 본 발명의 일 실시예에 따르면, 트루 데이터를 확보할 수 없는 소셜 네트워크의 한계를 극복하고 비교적 정확하게 미싱 노드를 추론할 수 있는 장점이 있다.

도면의 간단한 설명

- [0023] 도 1은 일반적인 소셜 네트워크의 구조를 나타낸 도면.
 도 2는 도 1에 도시된 소셜 네트워크를 구성하는 노드들의 연결 관계를 나타낸 매트릭스.
 도 3은 노출되지 않은 친구 관계를 포함하는 소셜 네트워크의 일례를 나타낸 도면.

도 4는 본 발명의 일 실시예에 따른 네트워크 완성을 설명하기 위한 도면.

도 5는 본 발명의 일 실시예에 따른 네트워크 완성 장치를 학습하기 위한 구조를 나타낸 도면.

도 6은 본 발명의 일 실시예에 따른 그래프 생성 신경망(500)의 추론 과정을 나타낸 개념도.

도 7은 본 발명의 일 실시예에 따른 네트워크 완성 신경망 모듈(510)의 개념적 구조를 나타낸 도면.

도 8은 본 발명의 일 실시예에 따른 신경망부(700)에서의 그래프 완성 과정의 일례를 나타낸 도면.

발명을 실시하기 위한 구체적인 내용

- [0024] 이하에서는 첨부한 도면을 참조하여 본 발명을 설명하기로 한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며, 따라서 여기에서 설명하는 실시예로 한정되는 것은 아니다.
- [0025] 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.
- [0026] 명세서 전체에서, 어떤 부분이 다른 부분과 "연결"되어 있다고 할 때, 이는 "직접적으로 연결"되어 있는 경우뿐 아니라, 그 중간에 다른 부재를 사이에 두고 "간접적으로 연결"되어 있는 경우도 포함한다.
- [0027] 또한 어떤 부분이 어떤 구성 요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성 요소를 제외하는 것이 아니라 다른 구성 요소를 더 구비할 수 있다는 것을 의미한다.
- [0028] 이하 첨부된 도면을 참고하여 본 발명의 실시예를 상세히 설명하기로 한다.
- [0029] 도 1은 일반적인 소셜 네트워크의 구조를 나타낸 도면이고, 도 2는 도 1에 도시된 소셜 네트워크를 구성하는 노드들의 연결 관계를 나타낸 매트릭스이다.
- [0030] 도 1을 참조하면, 도 1에 표시된 각 노드들은 소셜 네트워크를 구성하는 사용자이다. 도 1에서, 특정 노드는 다른 노드와 연결되어 있다. 예를 들어, 1번 노드는 2번 노드, 3번 노드 및 4번 노드와 연결되어 있으며, 노드의 연결은 1번 노드의 사용자가 2번 노드 사용자, 3번 노드 사용자 및 4번 노드 사용자와 친구 관계라는 것을 의미한다.
- [0031] 페이스북이나 인스타그램과 같은 소셜 네트워크는 사용자들이 다양한 관계로 연결되어 있으며 이러한 사용자들 간의 연결 관계를 사용자 커뮤니티를 탐지하는 유용한 정보로 작용한다. 사용자들 간의 연결 관계를 통해 탐지되는 커뮤니티는 다양한 마케팅 및 정보 제공 대상으로 사용될 수 있다.
- [0032] 도 2에 도시된 사용자들 간의 연결 관계를 네트워크는 매트릭스는 컬럼 및 로우 각각이 노드 번호를 나타내며, 1로 표시된 부분은 컬럼 및 로우에 해당되는 사용자들 간의 관계가 친구 관계라는 것을 의미하고 0으로 표시된 부분은 컬럼 및 로우에 해당되는 사용자들 간의 관계가 친구 관계가 아니라는 것을 의미한다.
- [0033] 도 1에 도시된 것과 같은 네트워크 연결 토폴로지는 도 2와 같은 연결 관계 매트릭스를 통해 벡터의 형태로 표현될 수도 있는 것이다.
- [0034] 그러나, 많은 소셜 네트워크의 사용자들은 다른 사용자와의 친구 관계를 노출시키지 않기도 한다. 이와 같이, 친구 관계를 노출시키지 않는 사용자들로 인해 정확한 소셜 네트워크를 파악하기 어려우며, 이는 부정확하게 커뮤니티를 탐지하는 주요한 요인이 된다. 부정확한 커뮤니티 탐지로 인해 부적절한 마케팅 및 정보 제공이 이루어질 수 있으며, 노출되지 않은 친구 관계를 파악하는 것은 다양한 이유로 필요할 수 있다.
- [0035] 도 3은 노출되지 않은 친구 관계를 포함하는 소셜 네트워크의 일례를 나타낸 도면이다.
- [0036] 도 3에서 실선으로 표시된 노드는 관찰 가능한 노드(사용자)들이고, 점선으로 표시된 노드는 친구 관계를 노출시키지 않아 숨겨진 노드이다. 본 명세서에서는 관찰 가능한 노드를 '가시 노드(Visible Node)'로 정의하고 숨겨진 노드를 미싱 노드(Missing Node)로 정의한다.
- [0037] 통상적으로 네트워크를 관찰할 때, 도 3에 도시된 노드들 중 가시 노드와 가시 노드들 간의 관계만 관찰되며, 미싱 노드의 존재 및 미싱 노드들의 연결 관계는 관찰되지 않는다.
- [0038] 관찰되지 않는 미싱 노드 및 미싱 노드들의 연결 관계를 복원하는 것을 네트워크 완성이라고 하며, 본 발명은 이러한 네트워크 완성에 관한 것이다. 결국, 본 발명은 가시 노드만이 관찰되는 특정 소셜 네트워크로부터 해당 소셜 네트워크에 숨겨진 미싱 노드 및 미싱 노드들의 연결 관계를 복원하는 것이다.

- [0039] 도 4는 본 발명의 일 실시예에 따른 네트워크 완성을 설명하기 위한 도면이다.
- [0040] 도 4의 좌측에 표시된 네트워크는 트루 네트워크(G_T)이며, 가시 노드 및 미싱 노드와 모든 노드들의 연결 관계가 표시된 네트워크이다. 중앙에 표시된 네트워크(G_0)는 실제로 관찰되는 네트워크이다. 물음표로 표시된 미싱 노드는 실제로는 관찰할 수 없으며, 도 4에는 2개의 노드가 미싱 노드인 경우가 도시되어 있다.
- [0041] 본 발명이 의도하는 네트워크 완성은 미싱 노드가 관찰되지 않는 네트워크(G_0)로부터 미싱 노드 및 미싱 노드들의 연결 관계를 추론하는 것이며, 완성된 네트워크는 도 4의 우측에 도시되어 있다.
- [0042] 본 발명은 인공 신경망을 통해 특정 소셜 네트워크의 미싱 노드 및 미싱 노드들의 연결 관계를 복원한다. 통상적인 인공 신경망의 학습은 ground truth로 명명되는 라벨 정보를 이용하여 이루어진다. 그러나, 미싱 노드 및 미싱 노드들의 연결 관계에 대한 라벨 정보는 현실적으로 획득하기 어려운 문제가 있다. 따라서, 본 발명은 ground truth를 사용하지 않으면서 숨겨진 미싱 노드들 및 미싱 노드들의 연결 관계를 복원하는 네트워크 완성 방법을 제안한다.
- [0043] 도 5는 본 발명의 일 실시예에 따른 네트워크 완성 장치를 학습하기 위한 구조를 나타낸 도면이다.
- [0044] 도 5를 참조하면, 본 발명의 네트워크 완성 장치는 그래프 생성 신경망(500)을 이용하여 학습을 수행한다. 그래프 생성 신경망(500)은 다양한 형태의 그래프들의 생성을 위해 기존에 이용되었던 신경망이다. 본 발명은 소셜 네트워크를 일종의 그래프로 간주하고, 그래프 생성 신경망(500)을 학습하여 네트워크를 완성하도록 한다.
- [0045] 또한, 그래프 생성 신경망(500)의 학습을 위한 학습 자료로 네트워크 복원 대상인 소셜 네트워크와 유사한 속성을 가지는 준거 소셜 네트워크를 활용한다. 예를 들어, 네트워크 복원 대상인 소셜 네트워크가 신촌 소재 제1대학 남학생의 소셜 네트워크라고 가정하면, 이와 유사한 속성을 가지는 다른 소셜 네트워크를 학습 자료로 활용하는 것이다. 예를 들어, 신촌 소재 제2대학 남학생의 소셜 네트워크를 준거 네트워크로 활용하여 그래프 생성 신경망(500)의 학습 자료로 활용할 수 있을 것이다.
- [0046] 그래프 생성 신경망(500)은 학습 자료로 입력되는 다양한 준거 소셜 네트워크를 이용하여 유사한 속성을 가지는 소셜 네트워크들의 구조를 학습한다. 구체적으로 그래프 생성 신경망(500)은 학습을 통해 연결 관계 확률 벡터(Θ)를 출력한다. 연결 관계 확률 벡터는 네트워크를 구성하는 노드들의 연결 관계들의 확률 정보를 나타낸 벡터이다.
- [0047] 그래프 생성 신경망(500)의 학습을 통해 출력되는 연결 관계 확률 벡터의 구체적인 구조 및 상세한 학습 방법은 별도의 도면을 참조하여 설명한다.
- [0048] 그래프 생성 신경망(500)에서의 학습이 완료된 후, 미싱 노드를 포함하는 소셜 네트워크의 완성은 네트워크 완성 신경망 모듈(510)에 의해 이루어진다. 네트워크 완성 신경망 모듈(510)로는 완성 대상 소셜 네트워크(G_0)가 입력된다. 입력되는 소셜 네트워크는 가시 노드 및 가시 노드들 간의 연결 관계 정보만을 포함하며, 네트워크 완성 신경망 모듈(510)은 입력된 소셜 네트워크로부터 미싱 노드 및 미싱 노드들의 연결 관계를 추론하여 네트워크 완성을 수행한다.
- [0049] 네트워크 완성 신경망 모듈(510)로는 그래프 생성 신경망(500)에서 학습을 통해 획득되는 연결 관계 확률 벡터가 입력되며, 네트워크 완성 신경망 모듈(510)은 연결 관계 확률 벡터에 기초하여 미싱 노드 및 미싱 노드들의 연결 관계를 추론한다.
- [0050] 앞서 설명한 바와 같이, 노출되지 않은 친구 관계에 대한 트루 데이터(True Data)는 획득하기 어려우며, 본 발명은 이러한 특성을 가지는 소셜 네트워크로부터 네트워크 완성을 수행하기 위해 그래프 생성 신경망(500)을 이용한다. 이를 위해, 소셜 네트워크를 그래프 생성을 위한 벡터 정보로 변환한 후 그래프 생성 신경망(500)에서 학습을 수행하도록 한다. 또한, 트루 데이터(True Data)가 획득되지 아니하므로 유사한 속성을 가지는 준거 집단의 소셜 네트워크를 그래프 생성 신경망(500)에서 그래프의 형태로 학습한다.
- [0051] 도 5를 참조하여 본 발명의 개괄적인 개요에 대해 설명하였으며, 이하에서는 그래프 생성 신경망(500) 및 네트워크 완성 신경망 모듈(510)의 구체적인 구조 및 동작에 대해 설명한다.
- [0052] 도 6은 본 발명의 일 실시예에 따른 그래프 생성 신경망(500)의 추론 과정을 나타낸 개념도이다.
- [0053] 그래프 생성 신경망(500)은 네트워크 구조를 학습하기 위한 신경망은 아니고, 그래프를 페인팅하기 위한 신경망이다. 그래프 생성 신경망(500)은 학습되는 그래프들의 특징 정보를 추론하고, 추론된 특징 정보에 기초하여 그

래프 페인팅을 학습하는 신경망이다. 본 발명은 네트워크 구조와 그래프의 구조가 유사하다는 점에 착안하여 그래프 완성을 위해 그래프 생성 신경망(500)을 이용하며, 이와 같은 그래프 생성 신경망(500)의 이용은 본 발명의 주요한 특징 중 하나이다.

- [0054] 그래프 생성 신경망(500)은 입력되는 네트워크 구조에 기초하여 두 개의 정보를 추론하며, 추론하는 두 개의 정보는 그래프 토폴로지 벡터와 연결 관계 확률 벡터이다.
- [0055] 그래프 토폴로지 벡터는 그래프의 토폴로지 구조를 나타낸 벡터이다. 예를 들어, 도 6을 참조하면, 생성되는 최종 그래프가 1-2-4-3-1로 연결된 그래프인데 이와 같은 그래프 구조를 벡터의 형태로 나타낸 것이 그래프 토폴로지 벡터이다.
- [0056] 연결 관계 확률 벡터는 그래프를 구성하는 노드들의 연결 관계에 대한 확률 정보에 대한 벡터이다. 예를 들어, 도 6을 참조하면, 노드1과 노드2가 연결되어 있는데, 노드1과 노드2가 연결될 확률 정보가 연결 관계 확률 벡터에 포함되며, 연결 관계 확률 벡터는 그래프의 모든 연결 관계에 대한 확률 정보를 포함한다.
- [0057] 그래프 생성 신경망(500)은 그래프를 구성하는 노드들을 순차적으로 페인팅하는 방식으로 그래프 생성을 학습하며, 페인팅하는 노드의 순서는 미리 주어진다.
- [0058] 네트워크 완성 관점에서, 생성해야 하는 그래프의 총 노드 수는 가시 노드의 수 ($|V_O|$)와 미시 노드의 수 ($|V_M|$)를 합친 값이다. 이때, 가능한 노드들의 순서에 대한 경우의 수는 $(|V_O|+|V_M|)!$ 이다.
- [0059] 주어진 노드 순서에 대한 그래프 토폴로지 벡터는 다음의 수학적 식 1과 같이 표현된다.

수학적 식 1

$$\mathbf{S}^\pi \triangleq (\mathbf{S}_1^\pi, \dots, \mathbf{S}_{|V_O|+|V_M|}^\pi),$$

- [0060]
- [0061] 위 수학적 식 1에서, S_i^π 의 i 는 순차적으로 페인팅(생성)하는 노드의 인덱스를 나타내며, S_i^π 는 i 번째 순서에서 노드들간의 연결 관계를 나타내는 벡터이다.
- [0062] 도 6을 참조하면, S_1^π 는 공집합에서 시작한다. 첫번째 노드인 1번 노드가 h1에서 생성되고, S_2^π 에는 다음에 생성되는 노드의 연결 관계가 기록된다. 도 6에서, S_2^π 는 {1}로 정의되어 있으며, 이는 다음에 생성되는 노드가 1번 노드와 연결된다는 것을 의미한다. 또한, Θ_2 에는 S_2^π 의 연결 관계에 대한 확률 정보가 기록된다. 도 6에는 Θ_2 가 0.9인 경우가 도시되어 있으며, 이는 1번 노드와 다음에 생성되는 노드가 연결될 확률이 0.9라는 것을 의미한다.
- [0063] h2에는 S_2^π 에 기초하여 다음 노드인 2번 노드가 생성된다. S_2^π 가 {1}이므로 다음에 생성되는 2번 노드는 1번 노드와 연결되어 생성되는 것을 h2로부터 확인할 수 있다.
- [0064] S_3^π 는 {1,0}으로 정의되어 있으며, 이는 다음에 생성되는 3번 노드와 기 생성된 노드들간의 연결 관계를 순차적으로 나타내며, {1,0}을 통해 다음에 생성되는 3번 노드가 1번 노드와는 연결되고 2번 노드와는 연결되지 않는 것을 의미한다. 한편, Θ_3 는 {0.8, 0.1}로 설정되어 있으며, 이는 1번 노드와 3번 노드가 연결될 확률이 0.8이고 2번 노드와 3번 노드가 연결될 확률이 0.1이라는 것을 의미한다.
- [0065] h3에는 S_3^π 에 기초하여 1번 노드와는 연결되고 2번 노드와는 연결되지 않는 3번 노드가 도시되어 있다.
- [0066] S_4^π 는 {0, 1, 1}로 설정되어 있으며, 이는 다음에 생성되는 4번 노드와 기 생성된 노드들의 연결 관계를 나타낸 것이고, 4번 노드는 1번 노드와는 연결되지 않고 2번 노드 및 3번 노드와는 연결된다는 것을 의미한다. Θ_4 는

{0.2, 0.7, 0.8}로 설정되어 있으며, 이는 S_4^π 의 연결 관계에 대한 확률 정보이다.

[0067] h4에는 S_4^π 에 기초하여 2번 및 3번 노드와 연결되며 1번 노드와 연결되지 않는 4번 노드가 도시되어 있다.

[0068] 그래프 토폴로지 벡터를 구성하는 S_i^π 는 다음의 수학적 식 2와 같이 표현될 수 있다.

수학적 식 2

$$\mathbf{S}_i^\pi = (a_{1,i}^\pi, \dots, a_{i-1,i}^\pi), \quad \forall i \in \{2, \dots, |V_O| + |V_M|\},$$

[0069]

위 수학적 식 2에서, $a_{u,v}^\pi$ 는 (u,v)번째 노드와의 연결 관계를 나타낸다.

[0070]

또한, 그래프 토폴로지 벡터에 대한 확률 분포는 다음의 수학적 식 3과 같이 표현될 수 있다.

수학적 식 3

$$p(\mathbf{S}^\pi) = \prod_{i=2}^{|V_O|+|V_M|} p(\mathbf{S}_i^\pi | \mathbf{S}_1^\pi, \dots, \mathbf{S}_{i-1}^\pi).$$

[0072]

위 수학적 식 3의 확률 분포는 그래프 토폴로지들에 대한 확률 분포로서, 특정 토폴로지의 그래프가 학습된 그래프 구조를 따를 확률로 정의할 수도 있을 것이며, 위 수학적 식 3과 같이 조건부 확률의 형태로 구해질 수 있을 것이다.

[0073]

본 발명의 일 실시예에 따르면, 그래프 생성 신경망(500)은 RNN 구조의 네트워크를 포함할 수 있을 것이다.

[0074]

RNN 구조의 그래프 생성 신경망(500)은 그래프 토폴로지와 그래프 토폴로지의 연결 관계에 대한 확률 정보를 학습하는 것이다. 구체적으로, 그래프 생성 신경망은 토폴로지에 관한 f_{trans} 및 연결 관계 확률에 대한 f_{out} 의 두가지 함수를 학습하는 것으로 정의할 수 있으며, f_{trans} 및 f_{out} 은 다음의 수학적 식 4 및 수학적 식 5와 같이 정의할 수 있다.

수학적 식 4

$$\mathbf{h}_i = f_{\text{trans}}(\mathbf{h}_{i-1}, \mathbf{S}_i^\pi).$$

[0076]

수학적 식 5

$$\theta_{i+1} = f_{\text{out}}(\mathbf{h}_i).$$

[0077]

그래프 생성 신경망의 f_{trans} 및 f_{out} 은 학습에 의해 형성되며, 이들 함수의 가중치가 학습에 의해 결정되는 것이다.

[0078]

준거 네트워크를 이용하여 학습된 그래프 생성 신경망으로부터 연결 관계 확률 벡터인 θ 를 획득하고, 본 발명

[0079]

은 획득된 θ 를 이용하여 대상 네트워크를 완성한다.

- [0080] 앞서 설명한 바와 같이, 본 발명은 네트워크 완성 신경망 모듈(510)을 이용하여 네트워크 완성을 수행하며, 네트워크 완성 신경망 모듈(510)로는 연결 관계 확률 벡터인 θ 가 입력된다.
- [0081] 미싱 노드를 포함하는 대상 네트워크(G_0)는 네트워크 완성 신경망 모듈(510)의 신경망과 연결 관계 확률 벡터 θ 를 이용하여 이루어진다.
- [0082] 도 7은 본 발명의 일 실시예에 따른 네트워크 완성 신경망 모듈(510)의 개념적 구조를 나타낸 도면이다.
- [0083] 도 7을 참조하면, 본 발명의 일 실시예에 따른 네트워크 완성 신경망 모듈은 신경망부(700) 및 선정부(710)를 포함한다.
- [0084] 신경망부(700)는 기 학습된 그래프 생성 신경망(500)의 가중치를 따르는 신경망이다. 신경망부(700)로는 미싱 노드를 포함하는 노드 순서가 입력되며, 신경망부(700)는 학습된 가중치에 기초하여 그래프 구조를 완성한다.
- [0085] 도 8은 본 발명의 일 실시예에 따른 신경망부(700)에서의 그래프 완성 과정의 일례를 나타낸 도면이다.
- [0086] 도 8에는 A, B, C 세개의 노드가 가시 노드이고 D, E가 미싱 노드인 경우가 도시되어 있다. 본 발명의 신경망부(700)로는 미싱 노드의 수를 미리 설정하여 미싱 노드를 포함하는 특정 순서의 노드가 입력된다.
- [0087] 도 8에는 페인팅(생성)하는 노드 순서가 {E, D, A, C, B}인 경우가 도시되어 있다. 여기서, E 및 D는 연결 관계를 알 수 없는 미싱 노드이다.
- [0088] 도 8에 도시된 바와 같이, 신경망부(700)는 기 학습된 가중치를 따르는 f_{trans} 및 f_{out} 을 이용하여 S_i^π 를 추론하면서 특정 순서의 네트워크를 완성한다. 도 8에서, 미싱 노드인 D 및 E의 연결 관계는 알 수 없으므로 초기에는 ?로 표시되어 있으나, 학습된 f_{trans} 및 f_{out} 을 이용하여 S_i^π 를 추론하게 된다.
- [0089] 신경망부(700)은 미싱 노드를 포함하는 가능한 모든 순서별로 네트워크 완성을 수행하여 후보 완성 네트워크를 출력한다. 도 8과 같이 3개의 가시 노드와 2개의 미싱 노드가 존재하는 경우 총 5!의 경우의 수가 존재하게 되며, 총 120개의 경우의 수가 존재하게 된다.
- [0090] 선정부(710)는 다수의 후보 완성 네트워크 중 하나를 선정하여 최종적인 네트워크 완성을 수행한다. 선정부(710)는 학습된 그래프 생성 신경망(500)으로부터 획득되는 연결 관계 확률 벡터 θ 를 이용하여 최종적인 완성 네트워크를 선정한다.
- [0091] 선정부(710)는 각각의 후보 완성 네트워크들 중 연결 관계 확률 벡터에 기초할 때 입력되는 대상 네트워크(G_0)와 유사도 확률이 가장 높은 후보 완성 네트워크를 선정한다. 이와 같은 완성 네트워크 선정은 다음의 수학적 식 6과 같이 표현될 수 있다.

수학적 식 6

$$\hat{G} = \operatorname{argmax}_G P(G|G_0, \theta)$$

- [0092]
- [0093] 위 수학적 식 6과 같이, 선정부(710)는 출력되는 후보 완성 네트워크들(G) 중 연결 관계 확률 벡터(θ)에 기초할 때 대상 네트워크의 구조와 확률적으로 가장 유사한 구조를 가지는 특정 순서의 완성 네트워크(\hat{G})를 선정하는 것이다.
- [0094] 도 7 및 도 8을 참조하여 설명한 본 발명의 네트워크 완성은 예시적인 것이며 다양한 방식으로 변경될 수 있을 것이다. 예를 들어, 후보 완성 네트워크는 모든 경우의 수에 대해 출력되지 않을 수도 있을 것이다. 가능성 있는 일부 경우의 수에 대해서만 후보 완성 네트워크가 출력된 후 선정 과정이 이루어질 수도 있을 것이다.
- [0095] 진술한 본 발명의 설명은 예시를 위한 것이며, 본 발명이 속하는 기술분야의 통상의 지식을 가진 자는 본 발명의 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 쉽게 변형이 가능하다는 것을 이해

할 수 있을 것이다.

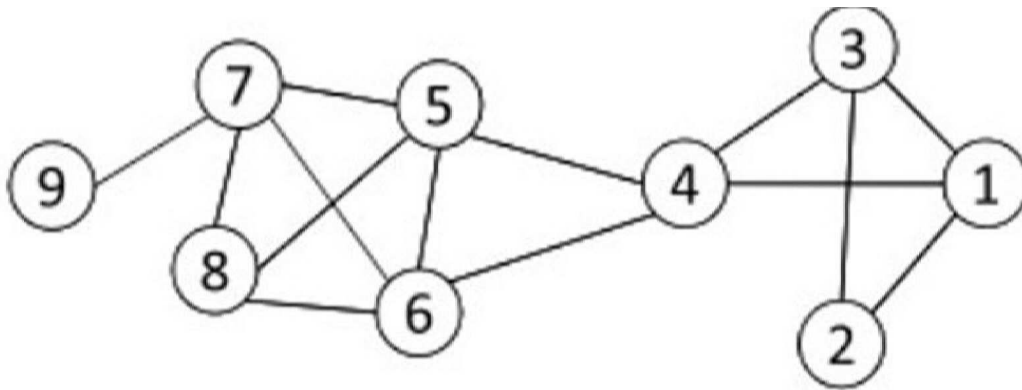
[0096] 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다.

[0097] 예를 들어, 단일형으로 설명되어 있는 각 구성 요소는 분산되어 실시될 수도 있으며, 마찬가지로 분산된 것으로 설명되어 있는 구성 요소들도 결합된 형태로 실시될 수 있다.

[0098] 본 발명의 범위는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

도면

도면1

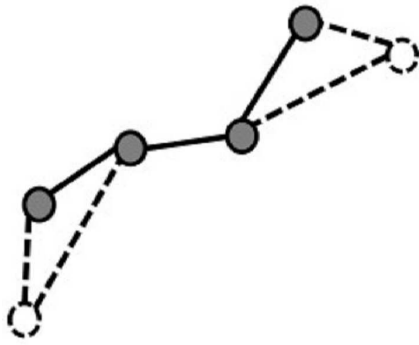


도면2

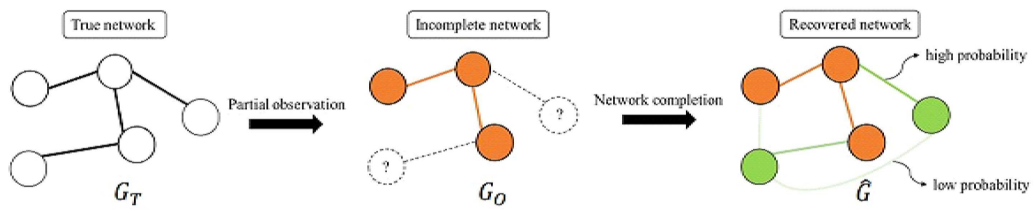
0	1	1	1	0	0	0	0	0
1	0	1	0	0	0	0	0	0
1	1	0	1	0	0	0	0	0
1	0	1	0	1	1	0	0	0
0	0	0	1	0	1	1	1	0
0	0	0	1	1	0	1	1	0
0	0	0	0	1	1	0	1	1
0	0	0	0	1	1	1	0	0
0	0	0	0	0	0	1	0	0

Adjacency matrix (A)

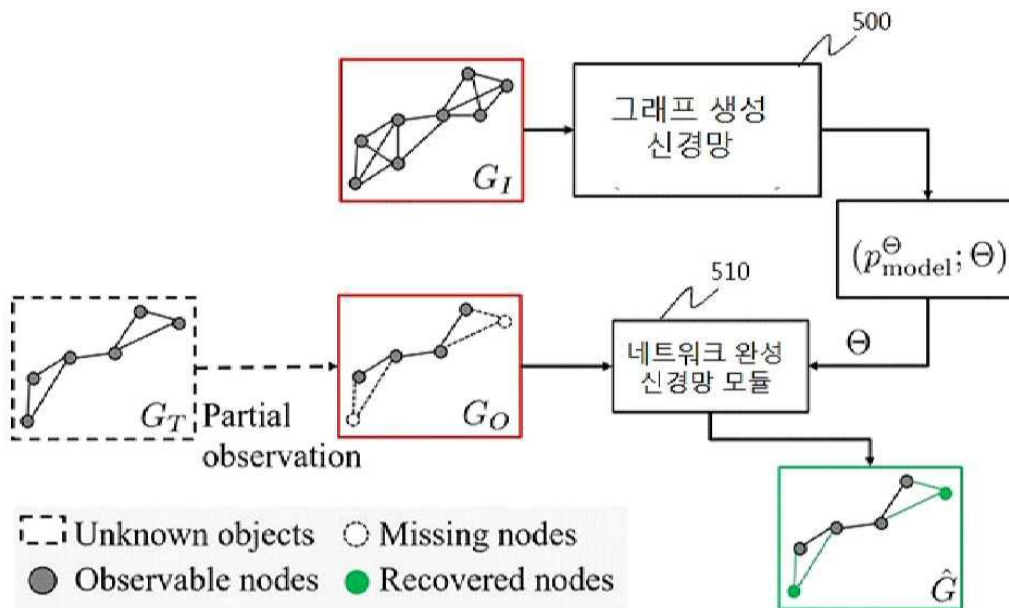
도면3



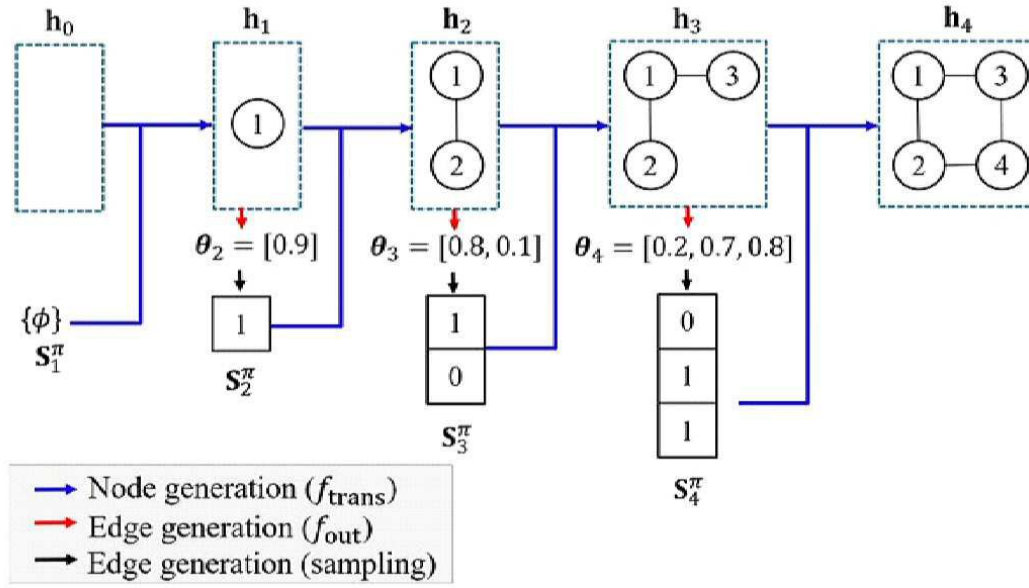
도면4



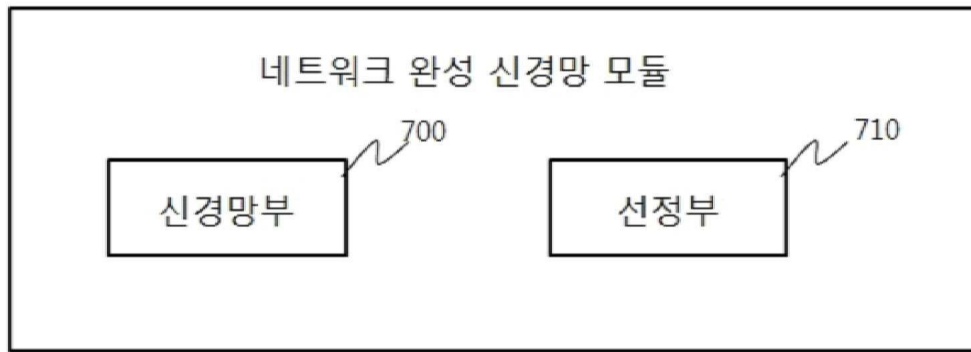
도면5



도면6



도면7



도면8

