



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0172831
(43) 공개일자 2024년12월10일

(51) 국제특허분류(Int. Cl.)

G06V 10/764 (2022.01) G06N 3/08 (2023.01)
G06V 10/70 (2022.01) G06V 10/82 (2022.01)
G06V 10/98 (2022.01)

(52) CPC특허분류

G06V 10/764 (2023.08)
G06N 3/08 (2023.01)

(21) 출원번호 10-2023-0071375

(22) 출원일자 2023년06월02일

심사청구일자 2023년06월02일

(71) 출원인

계명대학교 산학협력단

대구광역시 달서구 달구벌대로 1095, 계명대학교
산학협력관 201호(신당동)

(72) 발명자

고병철

대구광역시 수성구 범어천로 107

김상원

대구광역시 수성구 지범로22길 20

(74) 대리인

김건우

전체 청구항 수 : 총 16 항

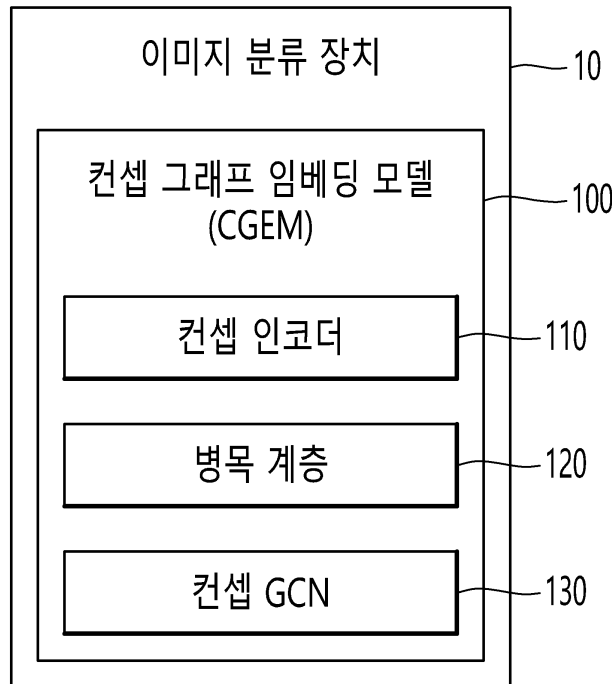
(54) 발명의 명칭 컨셉 그래프 임베딩 모델을 이용한 해석 가능한 이미지 분류 장치 및 방법

(57) 요약

본 발명은 컨셉 그래프 임베딩 모델을 이용한 해석 가능한 이미지 분류 장치에 관한 것으로서, 보다 구체적으로는 이미지 분류 장치로서, 백본 네트워크로부터 입력 이미지에서 추출된 최종 피쳐 맵을 입력받고, 디포머블 이분(Deformable Bi-partite) GCN(Graph Convolution Neural Networks)을 포함해 컨셉을 임베딩하여 컨셉 노드를

(뒷면에 계속)

대표도 - 도1



출력하는 컨셉 인코더; 상기 컨셉 노드와 노드 회귀 모듈을 이용해 컨셉 스코어를 생성하는 병목 계층; 및 상기 컨셉 스코어를 이용해 상기 컨셉 노드 사이의 관련성을 학습하는 컨셉 GCN을 포함하여 구성되는 컨셉 그래프 임베딩 모델(Concept Graph Embedding Model, CGEM)을 이용해 이미지를 분류하는 것을 그 구성상의 특징으로 한다.

또한, 본 발명은 컨셉 그래프 임베딩 모델을 이용한 해석 가능한 이미지 분류 방법에 관한 것으로서, 보다 구체적으로는 컴퓨터에서 각 단계가 수행되는 이미지 분류 방법으로서, 컨셉 인코더, 병목 계층 및 컨셉 GCN을 포함하여 구성되는 컨셉 그래프 임베딩 모델을 이용해 이미지를 분류하며, (1) 상기 컨셉 인코더에서, 백본 네트워크로부터 입력 이미지에서 추출된 최종 피쳐 맵을 입력받고, 디포머블 이분 GCN을 포함해 컨셉을 임베딩하여 컨셉 노드를 출력하는 단계; (2) 상기 병목 계층에서, 상기 컨셉 노드와 노드 회귀 모듈을 이용해 컨셉 스코어를 생성하는 단계; 및 (3) 상기 컨셉 GCN에서, 상기 컨셉 스코어를 이용해 상기 컨셉 노드 사이의 관련성을 학습하는 단계를 포함하는 것을 그 구성상의 특징으로 한다.

본 발명에서 제안하고 있는 컨셉 그래프 임베딩 모델을 이용한 해석 가능한 이미지 분류 장치 및 방법에 따르면, 디포머블 이분 그래프 컨볼루션을 사용해 컨셉의 동질성을 유지하면서도 다양한 왜곡에 대한 견고성을 갖는 컨셉 인코더와, 인간과 높은 상호작용을 가능하게 하는 동시에 인간의 오류로 인해 항상 완벽하지 않을 수 있는 잘못된 컨셉 개입에도 강력한 성능을 보이는 컨셉 GCN을 포함하는 컨셉 그래프 임베딩 모델을 이용해 이미지를 분류함으로써, 분류 문제에서 해석성과 정확도 간의 트레이드 오프를 해결할 수 있고, 잘못된 컨셉 개입에 대한 강력한 견고성을 보여주면서도 인간과 높은 상호작용이 가능하게 할 수 있다.

(52) CPC특허분류

G06V 10/82 (2022.01)

G06V 10/87 (2023.08)

G06V 10/987 (2023.08)

이 발명을 지원한 국가연구개발사업

과제고유번호	1345363225
과제번호	2022R1I1A3058128
부처명	교육부
과제관리(전문)기관명	한국연구재단
연구사업명	지역대학우수과학자지원사업
연구과제명	자기-지도학습 기반의 경량 시공간 그래프 트랜스포머 연구
기 여 율	1/1
과제수행기관명	계명대학교
연구기간	2023.03.01 ~ 2024.02.29

명세서

청구범위

청구항 1

이미지 분류 장치(10)로서,

백본 네트워크로부터 입력 이미지에서 추출된 최종 피쳐(feature) 맵을 입력받고, 디포머블 이분(Deformable Bi-partite) GCN(Graph Convolution Neural Networks)을 포함해 컨셉을 임베딩하여 컨셉 노드를 출력하는 컨셉 인코더(110);

상기 컨셉 노드와 노드 회귀 모듈을 이용해 컨셉 스코어를 생성하는 병목 계층(120); 및

상기 컨셉 스코어를 이용해 상기 컨셉 노드 사이의 관련성을 학습하는 컨셉 GCN(130)을 포함하여 구성되는 컨셉 그래프 임베딩 모델(Concept Graph Embedding Model, CGEM)(100)을 이용해 이미지를 분류하는 것을 특징으로 하는, 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 장치(10).

청구항 2

제1항에 있어서, 상기 컨셉 인코더(110)는,

백본 네트워크로부터 입력 이미지의 최종 피쳐 맵을 입력받아, 컨볼루션 레이어를 통해 N개의 노드 피쳐로 구성된 평탄화된 피쳐 맵을 생성하는 피쳐 맵 생성부(111);

N개의 노드 피쳐로 구성된 평탄화된 피쳐 맵을 고려해, 중심 노드를 찾는 중심 노드 임베딩부(112); 및

디포머블 이분 GCN을 통해, 상기 중심 노드 임베딩부(112)에서 찾은 중심 노드와 관련이 높은 이웃 노드를 탐색하고 최종 컨셉 노드를 유도하는 디포머블 이분 GCN부(113)를 포함하는 것을 특징으로 하는, 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 장치(10).

청구항 3

제2항에 있어서, 상기 중심 노드 임베딩부(112)는,

차수 기반 중심점 샘플링을 이용해 상기 중심 노드를 찾는 것을 특징으로 하는, 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 장치(10).

청구항 4

제3항에 있어서, 상기 중심 노드 임베딩부(112)는,

선택된 노드들 사이의 연결성을 결정하기 위해 노드 유사도에 기반하여 정의된 인접 행렬을 이용해, 각 노드의 연결 정도를 고려해 가중치 합을 계산하여 중심 노드를 선택하는 것을 특징으로 하는, 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 장치(10).

청구항 5

제2항에 있어서, 상기 디포머블 이분 GCN부(113)는,

학습 가능한 파라미터를 사용해 디포머블한 이웃 노드들의 변위 집합을 예측하는 오프셋 네트워크를 이용해, 상기 중심 노드의 2차원 좌표를 기준점으로 하여 상기 오프셋 네트워크에서 예측된 변위를 사용해 디포머블 이웃 노드를 탐색하는 것을 특징으로 하는, 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 장치

(10).

청구항 6

제5항에 있어서, 상기 디포머블 이분 GCN부(113)는,

미분 가능한 노드 선택을 위해, 탐색한 디포머블 이웃 노드 중에서 공일차내삽법(bilinear interpolation)을 이용해 이웃 노드를 선택하며, 근사 GCN 연산을 사용해 상기 중심 노드와 선택한 이웃 노드로부터 최종 컨셉 노드를 유도하는 것을 특징으로 하는, 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 장치(10).

청구항 7

제1항에 있어서, 상기 병목 계층(120)은,

상기 노드 회귀 모듈이 상기 컨셉 인코더(110)에서 출력된 컨셉 노드로부터, 노드 회귀를 통해 각 컨셉 노드의 기여도를 스칼라 값으로 표현해 시그모이드 정규화된 컨셉 스코어를 생성하는 것을 특징으로 하는, 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 장치(10).

청구항 8

제1항에 있어서, 상기 컨셉 GCN(130)은,

상기 컨셉 스코어를 사용해 각 컨셉 간의 상대적인 영향을 정량화해 방향성 인접 행렬을 생성하되, 정방향 인접 행렬과 역방향 인접 행렬을 생성하는 상대적 컨셉 매핑부(131); 및

상기 정방향 인접 행렬을 이용해 컨셉 노드 사이의 관련성을 학습하는 제1 GCN과 상기 역방향 인접 행렬을 이용해 컨셉 노드 사이의 관련성을 학습하는 제2 GCN을 학습하며, 상기 제1 GCN과 제2 GCN의 결과를 결합해 최종 클래스 레이블을 예측하는 컨셉 상호 방향성 학습부(132)를 포함하는 것을 특징으로 하는, 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 장치(10).

청구항 9

컴퓨터에서 각 단계가 수행되는 이미지 분류 방법으로서,

컨셉 인코더(110), 병목 계층(120) 및 컨셉 GCN(Graph Convolution Neural Networks)(130)을 포함하여 구성되는 컨셉 그래프 임베딩 모델(Concept Graph Embedding Model, CGEM)(100)을 이용해 이미지를 분류하며,

(1) 상기 컨셉 인코더(110)에서, 백본 네트워크로부터 입력 이미지에서 추출된 최종 피쳐(feature) 맵을 입력받고, 디포머블 이분(Deformable Bi-partite) GCN을 포함해 컨셉을 임베딩하여 컨셉 노드를 출력하는 단계;

(2) 상기 병목 계층(120)에서, 상기 컨셉 노드와 노드 회귀 모듈을 이용해 컨셉 스코어를 생성하는 단계; 및

(3) 상기 컨셉 GCN(130)에서, 상기 컨셉 스코어를 이용해 상기 컨셉 노드 사이의 관련성을 학습하는 단계를 포함하는 것을 특징으로 하는, 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 방법.

청구항 10

제9항에 있어서, 상기 단계 (1)은,

(1-1) 백본 네트워크로부터 입력 이미지의 최종 피쳐 맵을 입력받아, 컨볼루션 레이어를 통해 N개의 노드 피쳐로 구성된 평탄화된 피쳐 맵을 생성하는 단계;

(1-2) N개의 노드 피쳐로 구성된 평탄화된 피쳐 맵을 고려해, 중심 노드를 찾는 단계; 및

(1-3) 디포머블 이분 GCN을 통해, 상기 단계 (1-2)에서 찾은 중심 노드와 관련이 높은 이웃 노드를 탐색하고 최종 컨셉 노드를 유도하는 단계를 포함하는 것을 특징으로 하는, 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 방법.

청구항 11

제10항에 있어서, 상기 단계 (1-2)에서는,

차수 기반 중심점 샘플링을 이용해 상기 중심 노드를 찾는 것을 특징으로 하는, 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 방법.

청구항 12

제11항에 있어서, 상기 단계 (1-2)에서는,

선택된 노드들 사이의 연결성을 결정하기 위해 노드 유사도에 기반하여 정의된 인접 행렬을 이용해, 각 노드의 연결 정도를 고려해 가중치 합을 계산하여 중심 노드를 선택하는 것을 특징으로 하는, 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 방법.

청구항 13

제10항에 있어서, 상기 단계 (1-3)에서는,

학습 가능한 파라미터를 사용해 디포머블한 이웃 노드들의 변위 집합을 예측하는 오프셋 네트워크를 이용해, 상기 중심 노드의 2차원 좌표를 기준으로 하여 상기 오프셋 네트워크에서 예측된 변위를 사용해 디포머블 이웃 노드를 탐색하는 것을 특징으로 하는, 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 방법.

청구항 14

제13항에 있어서, 상기 단계 (1-3)에서는,

미분 가능한 노드 선택을 위해, 탐색한 디포머블 이웃 노드 중에서 공일차내삽법(bilinear interpolation)을 이용해 이웃 노드를 선택하며, 근사 GCN 연산을 사용해 상기 중심 노드와 선택한 이웃 노드로부터 최종 컨셉 노드를 유도하는 것을 특징으로 하는, 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 방법.

청구항 15

제9항에 있어서, 상기 단계 (2)에서는,

상기 노드 회귀 모듈이 상기 컨셉 인코더(110)에서 출력된 컨셉 노드로부터, 노드 회귀를 통해 각 컨셉 노드의 기여도를 스칼라 값으로 표현해 시그모이드 정규화된 컨셉 스코어를 생성하는 것을 특징으로 하는, 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 방법.

청구항 16

제9항에 있어서, 상기 단계 (3)은,

(3-1) 상기 컨셉 스코어를 사용해 각 컨셉 간의 상대적인 영향을 정량화해 방향성 인접 행렬을 생성하되, 정방향 인접 행렬과 역방향 인접 행렬을 생성하는 단계; 및

(3-2) 상기 정방향 인접 행렬을 이용해 컨셉 노드 사이의 관련성을 학습하는 제1 GCN과 상기 역방향 인접 행렬

을 이용해 컨셉 노드 사이의 관련성을 학습하는 제2 GCN을 학습하며, 상기 제1 GCN과 제2 GCN의 결과를 결합해 최종 클래스 레이블을 예측하는 단계를 포함하는 것을 특징으로 하는, 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 방법.

발명의 설명

기술 분야

[0001] 컨셉 그래프 임베딩 모델을 이용한 해석 가능한 이미지 분류 장치 및 방법{INTERPRETABLE IMAGE CLASSIFICATION APPARATUS AND METHOD USING CONCEPT GRAPH EMBEDDING MODEL}

배경 기술

[0002] 이 부분에 기술된 내용은 단순히 본 발명의 일실시예에 대한 배경 정보를 제공할 뿐 종래기술을 구성하는 것은 아니다.

[0004] 의료, 자율주행, 군사, 금융, 보안과 같이 고도의 책임을 요구하는 분야에서는 딥러닝 모델의 분류 정확도뿐만 아니라 분류 결과의 원인을 해석하고 설명하는 것이 매우 중요하다. 다양한 데이터 해석 방법 중, 영상 해석 방법은 히트맵(heat map)과 같이 주로 하위 수준의 픽셀이나 세그먼트의 중요성을 추정하였으나, 인간이 이해하고 판단하는 데 가장 중요한 것은 상위 수준의 컨셉(개념)이다.

[0006] 이러한 문제점을 해결하기 위해 최근 컨셉 기반의 해석 방법이 활발히 연구되고 있다. 이미지 분석에서 컨셉이란, 모델이 생성한 결과에서 인간이 직관적으로 이해할 수 있는 특징 또는 사용자가 직접 정의한 특징을 의미하며, 예를 들어 새 이미지에서 ‘몸 색깔’, ‘부리 모양’, ‘날개 무늬’ 등이 컨셉이 될 수 있다. 컨셉 기반의 해석 방법은 딥러닝 모델이 예측이나 결정을 내리는 데 사용되는 내부 컨셉을 추출함으로써, 예측의 투명성, 신뢰성 및 해석 가능성을 높이고 결정 과정을 이해하고 설명하는 데 도움이 된다.

[0008] 그러나 기존의 컨셉 기반 모델은 주로 모델의 해석 가능성에 초점을 두고 있다. 그런데 해석 가능성과 분류 정확도 사이에는 일반적으로 트레이드 오프가 존재하는 문제가 있다.

[0010] 진술한 배경 기술은 발명자가 본 발명의 도출을 위해 보유하고 있었거나, 본 발명의 도출 과정에서 습득한 기술 정보로서, 반드시 본 발명의 출원 전에 일반 공중에게 공개된 공지 기술이라 할 수는 없다.

선행기술문헌

비특허문헌

[0011] (비특허문헌 0001) P. W. Koh, T. Nguyen, Y. S. Tang, S. Mussamann, E. Pierson and B. Kim, “Concept Bottleneck Model,” in Proc. of International Conference on Machine Learning, pp. 5338-5348, Virtual, July 2020.

(비특허문헌 0002) M. Espinosa Zarlenga, P. Barbiero, G. Ciravegna, G. Marra, F. Giannini, M. Diligenti and M. Jannik, “Concept Embedding Models: Beyond the Accuracy-Explainability Trade-Off,” in Proc. of Advances in Neural Information Processing Systems, pp. 21400-21413, New Orleans, LA, USA, December 2022.

(비특허문헌 0003) B. Kim, M. Wattenberg, J. Gilmer, C. Cai, J. Wexler and F. Viegas, “Interpretability beyond feature attribution: Quantitative testing with concept activation vectors (tcav),” in Proc. of International Conference on Machine Learning, pp. 2668-2677, Stockholm, Sweden, July 2018.

(비특허문헌 0004) A. Ghorbani, J. Wexler, J. Y. Zou and B. Kim, “Towards Automatic Concept-based

Explanations,” in Proc. of Advances in Neural Information Processing Systems, pp. 1-10, Vancouver, Canada, December 2019.

(비특허문헌 0005) P. Welinder, S. Branson, T. Mita and P. Perona, Caltech-UCSD birds 200, Caltech, 2010.

(비특허문헌 0006) Z. Liu, P. Luo, X. Wang and T. Xiaoou, “Deep Learning Face Attributes in the Wild,” in Proc. of International Conference on Computer Vision, pp. 3730-3738, Santiago, Chile, December 2015.

(비특허문헌 0007) A. Farhadi, I. Endres, D. Hoiem, and D.A. Forsyth, “Describing Objects by their Attributes,” in Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1778-1785, Miami, FL, USA, June 2009.

발명의 내용

해결하려는 과제

[0012] 본 발명은 기존에 제안된 방법들의 상기와 같은 문제점들을 해결하기 위해 제안된 것으로서, 디포머블 이분(Deformable Bi-partite) 그래프 컨볼루션(Graph Convolution Neural Networks: GCN)을 사용해 컨셉의 동질성을 유지하면서도 다양한 왜곡에 대한 견고성을 갖는 컨셉 인코더와, 인간과 높은 상호작용을 가능하게 하는 동시에 인간의 오류로 인해 항상 완벽하지 않을 수 있는 잘못된 컨셉 개입에도 강력한 성능을 보이는 컨셉 GCN을 포함하는 컨셉 그래프 임베딩 모델(Concept Graph Embedding Model, CGEM)을 이용해 이미지를 분류함으로써, 분류 문제에서 해석성과 정확도 간의 트레이드 오프를 해결할 수 있고, 잘못된 컨셉 개입에 대한 강력한 견고성을 보여주면서도 인간과 높은 상호작용이 가능하게 하는, 컨셉 그래프 임베딩 모델을 이용한 해석 가능한 이미지 분류 장치 및 방법을 제공하는 것을 그 목적으로 한다.

[0014] 다만, 본 발명이 이루고자 하는 기술적 과제는 상기한 바와 같은 기술적 과제로 한정되지 않으며, 또 다른 기술적 과제들이 존재할 수 있고, 명시적으로 언급하지 않더라도 과제의 해결수단이나 실시 형태로부터 파악될 수 있는 목적이나 효과도 이에 포함됨은 물론이다.

과제의 해결 수단

[0015] 상기한 목적을 달성하기 위한 본 발명의 특징에 따른 컨셉 그래프 임베딩 모델을 이용한 해석 가능한 이미지 분류 장치는,

[0016] 이미지 분류 장치로서,

[0017] 백본 네트워크로부터 입력 이미지에서 추출된 최종 피쳐(feature) 맵을 입력받고, 디포머블 이분 GCN을 포함해 컨셉을 임베딩하여 컨셉 노드를 출력하는 컨셉 인코더;

[0018] 상기 컨셉 노드와 노드 회귀 모듈을 이용해 컨셉 스코어를 생성하는 병목 계층; 및

[0019] 상기 컨셉 스코어를 이용해 상기 컨셉 노드 사이의 관련성을 학습하는 컨셉 GCN을 포함하여 구성되는 컨셉 그래프 임베딩 모델을 이용해 이미지를 분류하는 것을 그 구성상의 특징으로 한다.

[0021] 바람직하게는, 상기 컨셉 인코더는,

[0022] 백본 네트워크로부터 입력 이미지의 최종 피쳐 맵을 입력받아, 컨볼루션 레이어를 통해 N개의 노드 피쳐로 구성된 평탄화된 피쳐 맵을 생성하는 피쳐 맵 생성부;

[0023] N개의 노드 피쳐로 구성된 평탄화된 피쳐 맵을 고려해, 중심 노드를 찾는 중심 노드 임베딩부; 및

[0024] 디포머블 이분 GCN을 통해, 상기 중심 노드 임베딩부에서 찾은 중심 노드와 관련이 높은 이웃 노드를 탐색하고 최종 컨셉 노드를 유도하는 디포머블 이분 GCN부를 포함할 수 있다.

- [0026] 더욱 바람직하게는, 상기 중심 노드 임베딩부는,
- [0027] 차수 기반 중심점 샘플링을 이용해 상기 중심 노드를 찾을 수 있다.
- [0029] 더더욱 바람직하게는, 상기 중심 노드 임베딩부는,
- [0030] 선택된 노드들 사이의 연결성을 결정하기 위해 노드 유사도에 기반하여 정의된 인접 행렬을 이용해, 각 노드의 연결 정도를 고려해 가중치 합을 계산하여 중심 노드를 선택할 수 있다.
- [0032] 더욱 바람직하게는, 상기 디포머블 이분 GCN부는,
- [0033] 학습 가능한 파라미터를 사용해 디포머블한 이웃 노드들의 변위 집합을 예측하는 오프셋 네트워크를 이용해, 상기 중심 노드의 2차원 좌표를 기준으로 하여 상기 오프셋 네트워크에서 예측된 변위를 사용해 디포머블 이웃 노드를 탐색할 수 있다.
- [0035] 더더욱 바람직하게는, 상기 디포머블 이분 GCN부는,
- [0036] 미분 가능한 노드 선택을 위해, 탐색한 디포머블 이웃 노드 중에서 공일차내삽법(bilinear interpolation)을 이용해 이웃 노드를 선택하며, 근사 GCN 연산을 사용해 상기 중심 노드와 선택한 이웃 노드로부터 최종 컨셉 노드를 유도할 수 있다.
- [0038] 바람직하게는, 상기 병목 계층은,
- [0039] 상기 노드 회귀 모듈이 상기 컨셉 인코더에서 출력된 컨셉 노드로부터, 노드 회귀를 통해 각 컨셉 노드의 기여도를 스칼라 값으로 표현해 시그모이드 정규화된 컨셉 스코어를 생성할 수 있다.
- [0041] 바람직하게는, 상기 컨셉 GCN은,
- [0042] 상기 컨셉 스코어를 사용해 각 컨셉 간의 상대적인 영향을 정량화해 방향성 인접 행렬을 생성하되, 정방향 인접 행렬과 역방향 인접 행렬을 생성하는 상대적 컨셉 매핑부; 및
- [0043] 상기 정방향 인접 행렬을 이용해 컨셉 노드 사이의 관련성을 학습하는 제1 GCN과 상기 역방향 인접 행렬을 이용해 컨셉 노드 사이의 관련성을 학습하는 제2 GCN을 학습하며, 상기 제1 GCN과 제2 GCN의 결과를 결합해 최종 클래스 레이블을 예측하는 컨셉 상호 방향성 학습부를 포함할 수 있다.
- [0045] 상기한 목적을 달성하기 위한 본 발명의 특징에 따른 컨셉 그래프 임베딩 모델을 이용한 해석 가능한 이미지 분류 방법은,
- [0046] 컴퓨터에서 각 단계가 수행되는 이미지 분류 방법으로서,
- [0047] 컨셉 인코더, 병목 계층 및 컨셉 GCN을 포함하여 구성되는 컨셉 그래프 임베딩 모델을 이용해 이미지를 분류하며,
- [0048] (1) 상기 컨셉 인코더에서, 백본 네트워크로부터 입력 이미지에서 추출된 최종 피쳐 맵을 입력받고, 디포머블 이분 GCN을 포함해 컨셉을 임베딩하여 컨셉 노드를 출력하는 단계;
- [0049] (2) 상기 병목 계층에서, 상기 컨셉 노드와 노드 회귀 모듈을 이용해 컨셉 스코어를 생성하는 단계; 및
- [0050] (3) 상기 컨셉 GCN에서, 상기 컨셉 스코어를 이용해 상기 컨셉 노드 사이의 관련성을 학습하는 단계를 포함하는 것을 그 구성상의 특징으로 한다.

- [0052] 바람직하게는, 상기 단계 (1)은,
- [0053] (1-1) 백본 네트워크로부터 입력 이미지의 최종 피쳐 맵을 입력받아, 컨볼루션 레이어를 통해 N개의 노드 피쳐로 구성된 평탄화된 피쳐 맵을 생성하는 단계;
- [0054] (1-2) N개의 노드 피쳐로 구성된 평탄화된 피쳐 맵을 고려해, 중심 노드를 찾는 단계; 및
- [0055] (1-3) 디포머블 이분 GCN을 통해, 상기 단계 (1-2)에서 찾은 중심 노드와 관련이 높은 이웃 노드를 탐색하고 최종 컨셉 노드를 유도하는 단계를 포함할 수 있다.
- [0057] 더욱 바람직하게는, 상기 단계 (1-2)에서는,
- [0058] 차수 기반 중심점 샘플링을 이용해 상기 중심 노드를 찾을 수 있다.]
- [0060] 더더욱 바람직하게는, 상기 단계 (1-2)에서는,
- [0061] 선택된 노드들 사이의 연결성을 결정하기 위해 노드 유사도에 기반하여 정의된 인접 행렬을 이용해, 각 노드의 연결 정도를 고려해 가중치 합을 계산하여 중심 노드를 선택할 수 있다.
- [0063] 더욱 바람직하게는, 상기 단계 (1-3)에서는,
- [0064] 학습 가능한 파라미터를 사용해 디포머블한 이웃 노드들의 변위 집합을 예측하는 오프셋 네트워크를 이용해, 상기 중심 노드의 2차원 좌표를 기준으로 하여 상기 오프셋 네트워크에서 예측된 변위를 사용해 디포머블 이웃 노드를 탐색할 수 있다.
- [0066] 더더욱 바람직하게는, 상기 단계 (1-3)에서는,
- [0067] 미분 가능한 노드 선택을 위해, 탐색한 디포머블 이웃 노드 중에서 공일차내삽법을 이용해 이웃 노드를 선택하며, 근사 GCN 연산을 사용해 상기 중심 노드와 선택한 이웃 노드로부터 최종 컨셉 노드를 유도할 수 있다.
- [0069] 바람직하게는, 상기 단계 (2)에서는,
- [0070] 상기 노드 회귀 모듈이 상기 컨셉 인코더에서 출력된 컨셉 노드로부터, 노드 회귀를 통해 각 컨셉 노드의 기여도를 스칼라 값으로 표현해 시그모이드 정규화된 컨셉 스코어를 생성할 수 있다.
- [0072] 바람직하게는, 상기 단계 (3)은,
- [0073] (3-1) 상기 컨셉 스코어를 사용해 각 컨셉 간의 상대적인 영향을 정량화해 방향성 인접 행렬을 생성하되, 정방향 인접 행렬과 역방향 인접 행렬을 생성하는 단계; 및
- [0074] (3-2) 상기 정방향 인접 행렬을 이용해 컨셉 노드 사이의 관련성을 학습하는 제1 GCN과 상기 역방향 인접 행렬을 이용해 컨셉 노드 사이의 관련성을 학습하는 제2 GCN을 학습하며, 상기 제1 GCN과 제2 GCN의 결과를 결합해 최종 클래스 레이블을 예측하는 단계를 포함할 수 있다.

발명의 효과

- [0075] 본 발명에서 제안하고 있는 컨셉 그래프 임베딩 모델을 이용한 해석 가능한 이미지 분류 장치 및 방법에 따르면, 디포머블 이분 그래프 컨볼루션을 사용해 컨셉의 동질성을 유지하면서도 다양한 왜곡에 대한 견고성을 갖는 컨셉 인코더와, 인간과 높은 상호작용을 가능하게 하는 동시에 인간의 오류로 인해 항상 완벽하지 않을 수 있는 잘못된 컨셉 개입에도 강력한 성능을 보이는 컨셉 GCN을 포함하는 컨셉 그래프 임베딩 모델을 이용해 이미지를 분류함으로써, 분류 문제에서 해석성과 정확도 간의 트레이드 오프를 해결할 수 있고, 잘못된 컨셉 개입에

대한 강력한 견고성을 보여주면서도 인간과 높은 상호작용이 가능하게 할 수 있다.

[0077] 더불어, 본 발명의 다양하면서도 유익한 장점과 효과는 상술한 내용에 한정되지 않으며, 본 발명의 구체적인 실시 형태를 설명하는 과정에서 보다 쉽게 이해될 수 있을 것이다.

도면의 간단한 설명

[0078] 도 1은 본 발명의 일실시예에 따른 컨셉 그래프 임베딩 모델을 이용한 해석 가능한 이미지 분류 장치의 구성을 도시한 도면.

도 2는 본 발명의 일실시예에 따른 컨셉 그래프 임베딩 모델을 이용한 해석 가능한 이미지 분류 장치에서, 컨셉 그래프 임베딩 모델(CGEM)의 전체적인 구조를 도시한 도면.

도 3은 본 발명의 일실시예에 따른 컨셉 그래프 임베딩 모델을 이용한 해석 가능한 이미지 분류 장치에서, 컨셉 인코더의 세부적인 구성을 도시한 도면.

도 4는 본 발명의 일실시예에 따른 컨셉 그래프 임베딩 모델을 이용한 해석 가능한 이미지 분류 장치에서, 컨셉 인코더의 구조를 도시한 도면.

도 5는 본 발명의 일실시예에 따른 컨셉 그래프 임베딩 모델을 이용한 해석 가능한 이미지 분류 장치에서, 컨셉 GCN의 세부적인 구성을 도시한 도면.

도 6은 본 발명의 일실시예에 따른 컨셉 그래프 임베딩 모델을 이용한 해석 가능한 이미지 분류 장치에서, 컨셉 GCN의 구조를 도시한 도면.

도 7은 본 발명의 일실시예에 따른 컨셉 그래프 임베딩 모델을 이용한 해석 가능한 이미지 분류 방법의 흐름을 도시한 도면.

도 8은 본 발명의 일실시예에 따른 컨셉 그래프 임베딩 모델을 이용한 해석 가능한 이미지 분류 방법에서, 단계 S110의 세부적인 흐름을 도시한 도면.

도 9는 본 발명의 일실시예에 따른 컨셉 그래프 임베딩 모델을 이용한 해석 가능한 이미지 분류 방법에서, 단계 S130의 세부적인 흐름을 도시한 도면.

도 10은 여러 분류 모델에 대해, 컨셉 개입 시 다운스트림 분류 정확도의 변화를 비교하여 나타낸 도면.

발명을 실시하기 위한 구체적인 내용

[0079] 아래에서는 첨부한 도면을 참조하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 본 발명의 실시예를 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며, 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

[0081] 명세서 전체에서, 어떤 부분이 다른 부분과 "연결"되어 있다고 할 때, 이는 "직접적으로 연결"되어 있는 경우뿐 아니라, 그 중간에 다른 소자를 사이에 두고 "간접적으로 연결"되어 있는 경우도 포함한다. 또한 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 의미하며, 하나 또는 그 이상의 다른 특징이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0083] 이하의 실시예는 본 발명의 이해를 돕기 위한 상세한 설명이며, 본 발명의 권리 범위를 제한하는 것이 아니다. 따라서 본 발명과 동일한 기능을 수행하는 동일 범위의 발명 역시 본 발명의 권리 범위에 속할 것이다.

- [0085] 또한, 본 발명의 각 실시예에 포함된 각 구성, 과정, 공정 또는 방법 등은 기술적으로 상호간 모순되지 않는 범위 내에서 공유될 수 있다.
- [0087] 또한, 본 발명에 있어서 단말, 장치 또는 디바이스가 수행하는 것으로 기술된 동작이나 기능 중 일부는 해당 단말, 장치 또는 디바이스와 연결된 서버에서 대신 수행될 수 있다. 마찬가지로, 서버가 수행하는 것으로 기술된 동작이나 기능 중 일부도 해당 서버와 연결된 단말, 장치 또는 디바이스에서 수행될 수도 있다.
- [0089] 특히, 본 발명의 각 실시예에 따른 시스템을 실행시키기 위한 수단으로는 애플리케이션(Application), 또는 웹 서버일 수 있으며, 이 애플리케이션, 또는 웹 서버를 기록한 기록매체를 읽을 수 있는 수단인 단말로는, 일반적인 데스크톱이나 노트북 등의 일반 PC뿐만 아니라, 스마트 폰, 태블릿 PC, 등의 모바일 단말기를 포함할 수 있다.
- [0091] 이하, 첨부된 도면을 참고하여 본 발명의 실시예들을 상세히 설명하도록 한다.
- [0093] 도 1은 본 발명의 일실시예에 따른 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 장치(10)의 구성을 도시한 도면이다. 도 1에 도시된 바와 같이, 본 발명의 일실시예에 따른 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 장치(10)는, 컨셉 인코더(110), 병목 계층(120) 및 컨셉 GCN(130)을 포함하여 구성되는 컨셉 그래프 임베딩 모델(CGEM)(100)을 이용해 이미지를 분류할 수 있다. 본 발명에서 제안하는 CGEM(100)은 복잡한 컨셉 간 의존성 및 구조를 반영할 수 있는데, 임베딩된 컨셉에 기반해 사람과 높은 상호작용을 가능하게 한다. CGEM(100)을 통해 정확도와 해석 가능성의 균형을 유지하고 컨셉 간 의존성 및 복잡한 구조를 반영하여 분류 문제를 해결할 수 있다.
- [0095] 본 발명의 기여는 다음과 같이 정리할 수 있다: 1) 본 발명에서는, 컨셉 인코더(110)를 위해 디포머블 이분 그래프 컨볼루션 기반의 CGEM(100)을 제안하며, 설계된 컨셉 인코더(110)는 다양한 실세계 왜곡에 강인성을 유지하면서 컨셉의 일관성을 향상할 수 있다. 2) 제안된 컨셉 GCN(130)은 임베딩된 컨셉을 기반으로 부정확한 컨셉 개입에 대해 강한 내성을 유지하면서 높은 인간 상호작용을 가능하게 할 수 있다. 3) 다양한 실험에서 CGEM(100)은 개선된 강인성을 기반으로 정확도와 해석력 간의 균형을 효과적으로 유지함을 확인할 수 있다.
- [0097] 컨셉 보틀넥 (bottleneck) 디자인은 복잡한 입력 데이터를 매우 작은 차원(예: 본 발명에서는 16)의 컨셉으로 매핑하여, 모델이 더 추상적인 정보에 집중하도록 만들 수 있다. 이 추상화된 컨셉들은 인간이 이해하기 쉬운 컨셉 레이블과 대응되며, 이 작은 컨셉들만 사용하여 클래스 레이블을 예측하도록 모델을 구성할 수 있다. 그러나 지나치게 컨셉을 추상화하면 모델의 표현력이 제한되어 입력의 다양성을 반영하기 어려워지며, 다양한 데이터 왜곡에 대한 강건성도 감소할 수 있다. 이전의 컨셉 보틀넥 디자인(비특허문헌 0001 및 0002 참조)은 전역 평균 풀링 벡터나 최종 로짓(logit)을 사용하여 추상화를 시도해왔다. 본 발명에서는, 이 한정된 차원에서 가능한 한 다양한 성격을 모델에 반영하고 왜곡에 대해 강건하게 하는 새로운 컨셉 인코더(110)로 디포머블 이분 GCN을 제안한다.
- [0099] 도 2는 본 발명의 일실시예에 따른 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 장치(10)에서, 컨셉 그래프 임베딩 모델(CGEM)(100)의 전체적인 구조를 도시한 도면이다. 도 2에 도시된 바와 같이, 본 발명의 일실시예에 따른 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 장치(10)의 CGEM(100)은, 백본 네트워크($f(\cdot)$)의 최종 피쳐 맵(h)이 컨셉 인코더(110)에 입력되고 컨셉 인코더(110)는 k 개의 컨셉을 나타내는 컨셉 노드($\{c_1, \dots, c_k\} \in \mathbf{C}$)를 포함할 수 있다. 컨셉 노드 $\mathbf{C}$ 와 노드 회귀 모듈 $p(\cdot)$ 를 이용하여 컨셉 스코어($\{s_1, \dots, s_k\} \in \mathbf{S}$)를 생성하고 이를 기반으로 노드들의 관련성을 학습하는 컨셉 GCN(130) $g(\cdot)$ 을 수행할 수 있다.

- [0101] 이하에서는, 도 1 및 도 2를 참고하여, 본 발명의 일실시예에 따른 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 장치(10)에서, CGEM(100)의 각 구성에 대해 상세히 설명하도록 한다.
- [0103] 컨셉 인코더(110)는, 백본 네트워크로부터 입력 이미지에서 추출된 최종 피쳐 맵을 입력받고, 디포머블 이분 GCN을 포함해 컨셉을 임베딩하여 컨셉 노드를 출력할 수 있다. 보다 구체적으로, 입력 이미지에 대해, 백본 네트워크(f)의 최종 피쳐 맵($h \in \mathbb{R}^{512 \times H \times W}$)을 사용하여 컨셉 추상화를 수행할 수 있다. 피쳐 맵은 컨셉 인코더(110)의 1×1 컨볼루션 레이어(w_k)를 통해 추상화되어, 평탄화된 피쳐 맵($\hat{h}_k \in \mathbb{R}^{16 \times W}$)이 생성될 수 있다. 본 발명의 컨셉 인코더(110)는 N개의 노드 피쳐로 이루어진 $\hat{h}_k$ 를 고려하여, 가장 중요한 중심 노드($c_k \in \mathbb{R}^{16}$)를 찾고, 이와 관련이 높은 이웃 노드를 디포머블 이분 GCN으로 임베딩할 수 있다.
- [0105] 도 3은 본 발명의 일실시예에 따른 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 장치(10)에서, 컨셉 인코더(110)의 세부적인 구성을 도시한 도면이다. 도 3에 도시된 바와 같이, 본 발명의 일실시예에 따른 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 장치(10)에서 컨셉 인코더(110)는, 피쳐 맵 생성부(111), 중심 노드 임베딩부(112) 및 디포머블 이분 GCN부(113)를 포함하여 구성될 수 있다.
- [0107] 도 4는 본 발명의 일실시예에 따른 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 장치(10)에서, 컨셉 인코더(110)의 구조를 도시한 도면이다. 이하에서는, 도 3 및 도 4를 참조하여 본 발명의 일실시예에 따른 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 장치(10)에서, 컨셉 인코더(110)의 각 구성에 대해 상세히 설명하도록 한다.
- [0109] 피쳐 맵 생성부(111)는, 백본 네트워크로부터 입력 이미지의 최종 피쳐 맵을 입력받아, 컨볼루션 레이어를 통해 N개의 노드 피쳐로 구성된 평탄화된 피쳐 맵을 생성할 수 있다. 즉, 피쳐 맵 생성부(111)는, 백본 네트워크(f)의 최종 피쳐 맵(h)을 1×1 컨볼루션 레이어(w_k)를 통해 추상화하여, 평탄화된 피쳐 맵($\hat{h}_k$)을 생성할 수 있다.
- [0111] 중심 노드 임베딩부(112)는, N개의 노드 피쳐로 구성된 평탄화된 피쳐 맵을 고려해, 중심 노드를 찾을 수 있다. 여기서, 중심 노드 임베딩부(112)는, 차수 기반 중심점 샘플링을 이용해 중심 노드를 찾을 수 있다. 보다 구체적으로, 중심 노드 임베딩부(112)는, 선택된 노드들 사이의 연결성을 결정하기 위해 노드 유사도에 기반하여 정의된 인접 행렬을 이용해, 각 노드의 연결 정도를 고려해 가중치 합을 계산하여 중심 노드를 선택할 수 있다.
- [0113] 즉, 본 발명에서는, 노드 피쳐들에서 가장 관련성 높은 노드를 찾기 위해 간단하면서도 효과적인 차수 기반 중심점 샘플링을 제안한다. 도 4에서처럼 모든 미연결 노드에 대한 잠재 공간의 인접 행렬 ($H_k \in \mathbb{R}^{N \times N}$)을 다음 수학적 식 1과 같이 정의할 수 있다.

수학식 1

$$H_k^{(i,j)} = \exp\left(\frac{\hat{h}_k^{(i)} \cdot \hat{h}_k^{(j)}}{\|\hat{h}_k^{(i)}\| \cdot \|\hat{h}_k^{(j)}\|} - 1\right)$$

[0115]

수학식 2

$$c_k^* = \sum_i \frac{\exp(D_k^{(i)})}{\sum_j \exp(D_k^{(j)})} \hat{h}_k^{(i)} ; D_k^{(i)} = \sum_j^N H_k^{(i,j)}$$

[0117]

[0119] 위의 수학식 2와 같이, 선택된 노드들 사이의 연결성을 결정하기 위해 노드 유사도에 기반한 인접 행렬을 정의하고, 중심 노드(c_k^*)를 선택하기 위해 각 노드의 연결 정도를 고려하여 가중치 합을 계산할 수 있다. 연결 정도가 높은 노드가 중앙 노드로 선택되기 쉽게 함으로써 모델은 노이즈를 줄이고 더 일관된 노드 표현 학습을 할 수 있다.

[0121] 디포머블 이분 GCN부(113)는, 디포머블 이분 GCN을 통해, 중심 노드 임베딩부(112)에서 찾은 중심 노드와 관련이 높은 이웃 노드를 탐색하고 최종 컨셉 노드를 유도할 수 있다. 보다 구체적으로, 디포머블 이분 GCN부(113)는, 학습 가능한 파라미터를 사용해 디포머블한 이웃 노드들의 변위 집합을 예측하는 오프셋 네트워크를 이용해, 중심 노드의 2차원 좌표를 기준으로 하여 오프셋 네트워크에서 예측된 변위를 사용해 디포머블 이웃 노드를 탐색할 수 있다. 이때, 디포머블 이분 GCN부(113)는, 미분 가능한 노드 선택을 위해, 탐색한 디포머블 이웃 노드 중에서 공일차내삽법(bilinear interpolation)을 이용해 이웃 노드를 선택하며, 근사 GCN 연산을 사용해 중심 노드와 선택한 이웃 노드로부터 최종 컨셉 노드를 유도할 수 있다.

[0123] 이전 연구들(비특허문헌 0003 및 0004 참고)은 주로 모델의 해석 가능성에 초점을 두고 있었으며, 실세계에서 발생할 수 있는 다양한 컨셉의 왜곡을 고려하지 않았다. 이에 따라, 본 발명에서는 모델이 실세계에서 자연스럽게 존재할 수 있는 비정상적인 컨셉에 강건성을 가질 수 있도록 디포머블 이분 GCN을 설계했다. 이 방법은 컨셉 인코더(110)에서 동작하며, 앞서 구해진 중심 노드의 2차원 좌표($\text{pos}(c_k^*) \in \mathbb{R}^N$)를 기준으로 하여 디포머블 노드를 탐색할 수 있다. 디포머블 이분 GCN의 전체 프로세스는 아래 수학식 3 내지 6을 따른다.

수학식 3

$$\Delta \mathbf{d} = \text{off}(c_k^*; \theta_k); \left\{ \left(d_x^{(1)}, d_y^{(1)} \right), \dots, \left(d_x^{(\tilde{n})}, d_y^{(\tilde{n})} \right) \right\} \in \Delta \mathbf{d}$$

[0125]

수학식 4

[0127]

$$(p_x^{(\tilde{n})}, p_y^{(\tilde{n})}) = \tanh \left(\text{pos}(c_k^*) + (d_x^{(\tilde{n})}, d_y^{(\tilde{n})}) \right)$$

수학식 5

[0129]

$$\tilde{h}_k^{(\tilde{n})} = d_k \left(\hat{h}_k; (p_x^{(\tilde{n})}, p_y^{(\tilde{n})}) \right)$$

수학식 6

[0131]

$$d_k \left(\hat{h}_k; (p_x^{(\tilde{n})}, p_y^{(\tilde{n})}) \right) = \sum_{(x,y) \in (W,H)} \hat{h}_k^{(x,y)} b(p_x^{(\tilde{n})}, x) b(p_y^{(\tilde{n})}, y), \text{ with } \hat{h}_k \in \mathbb{R}^{16 \times H \times W}$$

[0133]

오프셋 네트워크 $\text{off}(c_k^*; \theta_k)$ 는 학습 가능한 파라미터 θ_k 를 사용하여 디포머블한 이웃 노드들의 변위 집합 $\Delta d \in \mathbb{R}^{2 \times \tilde{n}}$ 을 예측하는 MLP이다. 하이퍼파라미터 $\tilde{n}$ 은 조절 가능한 디포머블 커널의 크기를 나타낸다. 예측된 변위와 기준점 좌표를 사용해 수학식 4를 따라 디포머블 이웃 노드를 탐색할 수 있다. 여기에서, 미분 가능한 노드 선택을 위해 bilinear interpolation ($b(\cdot)$)을 이용해 해당 위치의 노드를 선택할 수 있다. 여기서, (x,y) 는 $H \times W$ 크기의 원래 피쳐맵 $\hat{h}_k$ 상의 위치를 나타낸다. 선택된 이웃 노드($\tilde{h}_k$)는 수학식 6을 통해 잠재 공간에서 컨셉 k 의 중요한 표현을 포함할 수 있다. 따라서 다음 수학식 7에서 근사 GCN 연산을 사용하여 최종 컨셉 노드 $c_k \in \mathbb{R}^{16}$ 을 유도할 수 있다.

수학식 7

[0135]

$$c_k = c_k^* + \sum_{\tilde{n}} \tilde{h}_k^{(\tilde{n})}$$

[0137]

병목 계층(120)은, 컨셉 노드와 노드 회귀 모듈을 이용해 컨셉 스코어를 생성할 수 있다. 보다 구체적으로, 병목 계층(120)은, 컨셉 인코더(110)에서 출력된 컨셉 노드로부터, 노드 회귀 모듈이 노드 회귀를 통해 각 컨셉 노드의 기여도를 스칼라 값으로 표현해 시그모이드 정규화된 컨셉 스코어를 생성할 수 있다.

[0139]

기존 방법들(비특허문헌 0001 및 0002 참조)은 컨셉 스코어를 기반으로 컨셉의 존재 여부만 고려하는 반면, 본 발명에서 제안하는 방법은 예측된 컨셉 간의 관계를 이해하고 활용하여 높은 성능과 강건성을 달성할 수 있다. 도 2에서 노드 회귀 모듈 $p(\cdot)$ 은 컨셉 인코더(110)에서 예측된 컨셉 노드로부터 시그모이드 정규화된 컨셉 스코어 $S \in \mathbb{R}^{|C| \times 1}$ 를 예측할 수 있다. 노드 회귀를 통해 병목 계층(120)은 각 컨셉 노드의 기여도를 연속적인 스칼라로 표현할 수 있다.

[0141] 컨셉 GCN(130)은, 컨셉 스코어를 이용해 컨셉 노드 사이의 관련성을 학습할 수 있다.

[0143] 도 5는 본 발명의 일실시예에 따른 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 장치(10)에서, 컨셉 GCN(130)의 세부적인 구성을 도시한 도면이고, 도 6은 본 발명의 일실시예에 따른 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 장치(10)에서, 컨셉 GCN(130)의 구조를 도시한 도면이다.

[0145] 이전 컨셉 보틀넥 디자인(비특허문헌 0001 및 0002 참조)은 예측된 컨셉에 개입을 통해 조작함으로써 모델의 성능을 개선할 수 있다는 것을 증명했다. 이는 해석 가능한 모델이 잘못된 컨셉을 예측할 때 사람이 컨셉을 조작하여 모델이 올바른 예측을 하도록 유도할 수 있다는 것을 의미한다. 그러나 이전 방법들은 인간 개입이 항상 완벽하지 않기 때문에 잘못된 컨셉 조작으로 인한 성능 감소 가능성에 대해 주의를 기울이지 않았다. 본 발명에서는, 이 문제를 해결하기 위해 도 6과 같이 컨셉 간 상대적인 관계에 기초한 새로운 컨셉 GCN(130)을 제안한다. 이 접근 방식은 긍정적인 개입과 높은 상호작용 능력을 제공하면서도 잘못된 컨셉 개입에 대한 강건성을 유지할 수 있다.

[0147] 도 5에 도시된 바와 같이, 본 발명의 일실시예에 따른 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 장치(10)에서, 컨셉 GCN(130)은, 상대적 컨셉 매핑부(131) 및 컨셉 상호 방향성 학습부(132)를 포함하여 구성될 수 있다. 이하에서는, 도 5 및 도 6을 참조하여 본 발명의 일실시예에 따른 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 장치(10)의 컨셉 GCN(130)의 각 구성에 대해 상세히 설명하도록 한다.

[0149] 상대적 컨셉 매핑부(131)는, 컨셉 스코어를 사용해 각 컨셉 간의 상대적인 영향을 정량화해 방향성 인접 행렬을 생성하되, 정방향 인접 행렬과 역방향 인접 행렬을 생성할 수 있다. 보다 구체적으로, 상대적 컨셉 매핑부(131)는, 컨셉 스코어를 사용하여 다음 수학적 식 8을 통해 각 컨셉 간의 상대적인 영향을 정량화할 수 있다.

수학적 식 8

$$\vec{A}_{i,j} = \begin{cases} \frac{s_j - s_i + \gamma}{2\gamma}, & i \neq j \\ 0, & i = j \end{cases}$$

[0153] 여기서, $\vec{A} \in \mathbb{R}^{|C| \times |C|}$ 는 컨셉 노드 i 에서 j 로의 방향성 인접 행렬을 나타낸다. 스무딩 팩터인 $\gamma \in [1, \infty)$ 은 컨셉 노드 간의 방향성 강도를 조정할 수 있다.

[0155] 컨셉 상호 방향성 학습부(132)는, 정방향 인접 행렬을 이용해 컨셉 노드 사이의 관련성을 학습하는 제1 GCN과 역방향 인접 행렬을 이용해 컨셉 노드 사이의 관련성을 학습하는 제2 GCN을 학습하며, 제1 GCN과 제2 GCN의 결과를 결합해 최종 클래스 레이블을 예측할 수 있다.

[0157] CGEM(100)에서는 잘못된 컨셉 개입에 대한 강건성과 긍정적인 컨셉 개입에서 인간과 높은 상호작용을 가능하게 하는 것을 목표로 한다. 이를 위해 본 발명에서는, GCN 레이어를 디자인해 정방향 인접 행렬($\vec{A}$)에서 역행렬을

언어 역방향 인접 행렬($\vec{A}$)의 방향성을 함께 학습하도록 할 수 있다. 따라서, 각 노드가 다른 컨셉을 나타내는 이중 그래프에서 양방향성을 학습하여, 추상 컨셉 노드로부터 풍부한 표현을 얻고 잘못된 컨셉 개입에 대한 강건성을 보장할 수 있다. 이 과정은 도 6에 나타나 있으며 다음 수학식 9로 표현될 수 있다.

수학식 9

$$y = FC(\vec{C}); \quad \vec{C} = C + FC(\vec{D}^{-1}\vec{A}\vec{C}\vec{W} + \vec{D}^{-1}\vec{A}\vec{C}\vec{W})$$

[0159]

여기에서 차수행렬 $\vec{D}$ 는 정방향 인접 행렬 $\vec{A}$ 에 대해 비대칭 정규화를 수행할 수 있다. 컨셉 상호 방향성 학습부(132)는 두 개의 유도된 GCN에 대해 학습 가능한 매개변수 $\vec{W}$ 및 $\vec{W}$ 를 적용하고 결과를 결합하여 최종 클래스 레이블 y 를 예측할 수 있다.

[0161]

도 7은 본 발명의 일실시예에 따른 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 방법의 흐름을 도시한 도면이다. 도 7에 도시된 바와 같이, 본 발명의 일실시예에 따른 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 방법은, 컨셉 인코더(110), 병목 계층(120) 및 컨셉 GCN(130)을 포함하여 구성되는 컨셉 그래프 임베딩 모델(100)을 이용해 이미지를 분류하며, 컨셉 인코더(110)에서, 백본 네트워크로부터 입력 이미지에서 추출된 최종 피쳐 맵을 입력받고, 디포머블 이분 GCN을 포함해 컨셉을 임베딩하여 컨셉 노드를 출력하는 단계(S110), 병목 계층(120)에서, 컨셉 노드와 노드 회귀 모듈을 이용해 컨셉 스코어를 생성하는 단계(S120) 및 컨셉 GCN(130)에서, 컨셉 스코어를 이용해 컨셉 노드 사이의 관련성을 학습하는 단계(S130)를 포함하여 구현될 수 있다.

[0163]

도 8은 본 발명의 일실시예에 따른 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 방법에서, 단계 S110의 세부적인 흐름을 도시한 도면이다. 도 8에 도시된 바와 같이, 본 발명의 일실시예에 따른 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 방법의 단계 S110은, 백본 네트워크로부터 입력 이미지의 최종 피쳐 맵을 입력받아, 컨볼루션 레이어를 통해 N개의 노드 피쳐로 구성된 평탄화된 피쳐 맵을 생성하는 단계(S111), N개의 노드 피쳐로 구성된 평탄화된 피쳐 맵을 고려해 중심 노드를 찾는 단계(S112) 및 디포머블 이분 GCN을 통해 중심 노드와 관련이 높은 이웃 노드를 탐색하고 최종 컨셉 노드를 유도하는 단계(S113)를 포함하여 구현될 수 있다.

[0165]

도 9는 본 발명의 일실시예에 따른 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 방법에서, 단계 S130의 세부적인 흐름을 도시한 도면이다. 도 9에 도시된 바와 같이, 본 발명의 일실시예에 따른 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 방법의 단계 S130은, 컨셉 스코어를 사용해 각 컨셉 간의 상대적인 영향을 정량화해 방향성 인접 행렬을 생성하는 단계(S131) 및 정방향 인접 행렬을 이용해 컨셉 노드 사이의 관련성을 학습하는 제1 GCN과 역방향 인접 행렬을 이용해 컨셉 노드 사이의 관련성을 학습하는 제2 GCN을 학습하며, 제1 GCN과 제2 GCN의 결과를 결합해 최종 클래스 레이블을 예측하는 단계(S132)를 포함하여 구현될 수 있다.

[0167]

각각의 단계들과 관련된 상세한 내용들은, 앞서 본 발명의 일실시예에 따른 본 발명의 일실시예에 따른 개선된 시공간 교차 주의 트랜스포머를 이용한 인간 행동 인식 시스템(10)과 관련하여 충분히 설명되었으므로, 상세한 설명은 생략하기로 한다.

[0169]

[0171] 실험

[0173] 본 발명에서 제안한 CGEM(100)은 비전 데이터셋에서 이미지 특징과 추상화된 컨셉의 관계적 학습을 통해 컨셉 해석을 가능하게 한다. 따라서 애트리뷰트를 제공하는 네 가지 데이터셋 CUB(비특허문헌 0005), CelebA(비특허문헌 0006), aPY(비특허문헌 0007), 그리고 Travelling Birds(비특허문헌 0002)를 사용하여 모델 학습과 평가를 수행했다. 해석 가능한 모델의 정량적 성능 비교를 위해 분류 정확도(Task Accuracy)와 컨셉 정렬도(Concept Alignment)를 사용했다. 여기서, 컨셉 정렬도는 CEM(비특허문헌 0002)에서 제안한 메트릭으로, 각 컨셉 인코더(110)가 생성하는 추상 컨셉(C)의 일관성을 정량화한다. 이를 통해 컨셉 인코더(110)의 일관성을 이해하는데 도움이 된다.

[0175] 1. 분류 정확도와 해석력 간의 트레이드 오프

[0177] 표 1은 세 가지 주요 데이터셋에서 컨셉 병목 설계 모델 간 분류 정확도(%)와 컨셉 정렬도(%)를 비교한 것이다. 표 1은 세 가지 데이터 세트 CUB, CelebA, aPY에서 제안된 CGEM(100)과 다른 개념 병목 설계 방법인 CEM(비특허문헌 0002), Bool-CBM(비특허문헌 0001) 및 Fuzzy-CBM(비특허문헌 0001)의 분류 정확도와 해석력 간의 트레이드 오프(trade-off)를 비교한다. 여기서, “No Concepts”는 제안된 모델이 컨셉 스코어를 업데이트하지 않고 훈련된 경우를 나타낸다.

표 1

	Method	Task Accuracy (%)	Concept Alignment (%)
CUB ^[12]	No Concepts (Ours)	75.04	62.86
	Bool-CBM ^[10]	67.11	81.18
	Fuzzy-CBM ^[10]	72.98	80.79
	CEM ^[11]	77.11	86.14
	CGEM (Ours)	80.61	91.19
CelebA ^[13]	No Concepts (Ours)	37.76	69.48
	Bool-CBM ^[10]	31.95	83.18
	Fuzzy-CBM ^[10]	32.85	82.70
	CEM ^[11]	40.29	83.62
	CGEM (Ours)	41.75	88.48
aPY ^[14]	No Concepts (Ours)	38.85	63.89
	Bool-CBM ^[10]	33.03	56.32
	Fuzzy-CBM ^[10]	35.61	56.29
	CEM ^[11]	40.17	70.94
	CGEM (Ours)	44.76	71.96

[0179]

[0181] 실험 결과, 본 발명에서 제안된 CGEM(100)이 모든 데이터셋에서 가장 완화된 트레이드 오프를 달성한다는 것을 보여준다. 구체적으로, 본 발명에서 제안한 CGEM(100)은 CUB와 CelebA 데이터셋에서 분류 정확도와 해석력을 모두 크게 향상시킨다. CEM도 다른 기존 방법들과 비교해 좋은 성능을 보였지만, 본 발명에서 제안한 CGEM(100)은 트레이드 오프 문제를 해결하는 능력에서 가장 큰 향상을 보였다. No Concepts의 경우, 분류 정확도가 Bool-CBM 또는 Fuzzy-CBM보다 높지만, 해석력은 현저하게 낮다는 것을 알 수 있다. 그러나 컨셉 스코어 업데이트를 함께 수행하면 최상의 성능을 달성할 수 있다. aPY 데이터셋의 경우, 본 발명에서 제안한 CGEM(100)은 트레이드 오프를 해결하는 능력이 가장 좋았지만, 해석력은 다소 제한적이었다. 이는 부정확한 컨셉 레이블링과 심각한 컨셉 불균형이 원인으로 보여진다. 제안된 CGEM(100)은 컨셉 간의 관계에 기반하여 학습하기 때문에 이러한 문제를 해결하는 능력이 다소 제한될 수 있다. 컨셉 레이블링 문제를 무시하고 훈련할 경우(*i.e.*, No concepts), 비교적 높은 수준의 분류 정확도와 해석력을 달성할 수 있다.

[0183] 2. 컨셉 개입

[0185] 도 10은 여러 분류 모델에 대해, 컨셉 개입 시 다운스트림 분류 정확도의 변화를 비교하여 나타낸 도면이다. 도 10은 CEM, Bool-CBM, Fuzzy-CBM 및 제안된 CGEM(100)에 대한 각 데이터 세트에서 (a) 잘못된 컨셉 개입과 (b) 긍정적인 컨셉 개입에 의한 작업 정확도의 변화를 관찰한 결과를 보여준다. 상호 컨셉을 여러 그룹으로 그룹화하고 더 많은 컨셉 그룹에 점진적으로 개입함에 따라 모델이 얼마나 강력하게 작동하는지 테스트했다. 'R'은 무작위 긍정 컨셉 개입으로 훈련된 모델을 의미한다. 모든 결과는 평균과 95% 신뢰 구간으로 보고된다.

[0187] 2-1. 잘못된 컨셉 개입

[0189] 다양한 데이터셋에서 제안된 방법이 인간의 잘못된 개입에 얼마나 강건한지 비교하고 검증했다. 이를 위해 먼저, 각 데이터셋에서 상관관계가 높은 컨셉들을 여러 개의 그룹으로 나눈다. CelebA의 경우 컨셉이 6개뿐이므로 개별 컨셉을 하나의 그룹으로 고려하였다. 이 실험에서는 잘못된 컨셉 그룹 수가 증가함에 따라 분류 정확도의 변화를 관찰할 수 있다.

[0191] 도 10의 (a)에서 제안된 방법 (CGEM 및 CGEM+R)은 잘못된 컨셉 개입이 발생하더라도 상대적으로 높은 성능을 유지한다. CUB 데이터셋의 경우, 제안된 방법은 잘못된 컨셉 개입의 수가 증가함에 따라 일관된 성능을 보여준다. 반면 CelebA 및 aPY 데이터셋의 경우 성능 저하가 발생하지만, 제안된 방법은 여전히 가장 높은 성능을 유지한다. 이는 컨셉의 수가 매우 적은 경우 (e.g., CelebA) 또는 레이블이 불완전한 경우 (e.g., aPY)에는 성능 저하를 피하는 것이 어렵다는 것을 알 수 있다. 그러나 제안된 CGEM(100)은 이러한 경우에도 상대적으로 높은 성능을 유지한다.

[0193] 2-2. 긍정 컨셉 개입

[0195] 도 10의 (b)는 모델이 컨셉 그룹 수준에서 긍정적으로 개입될 때 분류 정확도가 어떻게 변하는지를 관찰할 수 있다. 제안된 CGEM(100)모델은 전체 데이터셋에서 분류 정확도가 향상되었다. CUB 및 aPY 데이터셋의 경우, 긍정적으로 개입된 컨셉 수가 증가함에 따라 분류 정확도가 일관되게 향상되었다. 학습에 6개의 컨셉만 사용된 CelebA 데이터셋의 경우, 정확도 향상률이 경우에 따라 감소하는 경향이 있었으나 긍정적 개입 비율이 높아질수록 정확도가 크게 향상되었다.

[0197] 이러한 결과로부터, 본 발명의 일 실시예에 따른 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 장치(10) 및 방법은 전반적으로 잘못된 컨셉 개입에 대한 강력한 견고성을 보여주면서도 인간과 높은 상호작용이 가능함을 확인할 수 있다.

[0199] 본 발명에서는, 분류 문제에서 해석성과 정확도 간의 트레이드 오프를 해결할 수 있는 CGEM(100)을 제안했다. CGEM(100)은 컨셉 간의 관계 학습을 통해 높은 인간 상호작용을 가능하게 하고, 잘못된 컨셉에 대한 강력한 성능을 유지한다. 기존의 해석 방법과 비교해 제안된 CGEM(100)은 벤치마크 데이터셋에서 우수한 성능과 향상된 해석성을 나타내었다. 또한, 다양한 실험에서 제안된 CGEM(100)은 강력한 인간 상호작용을 가능하게 하고, 컨셉간의 관계 학습을 통해 잘못된 컨셉 개입으로 인한 오류를 보완할 수 있다는 것을 증명했다.

[0201] 전술한 바와 같이, 본 발명에서 제안하고 있는 컨셉 그래프 임베딩 모델(100)을 이용한 해석 가능한 이미지 분류 장치(10) 및 방법에 따르면, 디포머블 이분 그래프 컨볼루션을 사용해 컨셉의 동질성을 유지하면서도 다양한

왜곡에 대한 견고성을 갖는 컨셉 인코더(110)와, 인간과 높은 상호작용을 가능하게 하는 동시에 인간의 오류로 인해 항상 완벽하지 않을 수 있는 잘못된 컨셉 개입에도 강력한 성능을 보이는 컨셉 GCN(130)을 포함하는 컨셉 그래프 임베딩 모델(100)을 이용해 이미지를 분류함으로써, 분류 문제에서 해석성과 정확도 간의 트레이드 오프를 해결할 수 있고, 잘못된 컨셉 개입에 대한 강력한 견고성을 보여주면서도 인간과 높은 상호작용이 가능하게 할 수 있다.

[0203] 한편, 본 발명은 다양한 통신 단말기로 구현되는 동작을 수행하기 위한 프로그램 명령을 포함하는 컴퓨터에서 판독 가능한 매체를 포함할 수 있다. 예를 들어, 컴퓨터에서 판독 가능한 매체는, 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media) 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치를 포함할 수 있다.

[0205] 이와 같은 컴퓨터에서 판독 가능한 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 이때, 컴퓨터에서 판독 가능한 매체에 기록되는 프로그램 명령은 본 발명을 구현하기 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 예를 들어, 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해 실행될 수 있는 고급 언어 코드를 포함할 수 있다.

[0207] 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 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다. 예를 들어, 단일형으로 설명된 각 구성요소는 분산되어 실시될 수도 있으며, 마찬가지로 분산된 것으로 설명된 구성 요소들도 결합된 형태로 실시될 수 있다.

[0209] 본 발명의 범위는 상기 상세한 설명보다는 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.

부호의 설명

[0210] 10: 이미지 분류 장치

100: 컨셉 그래프 임베딩 모델 (CGEM)

110: 컨셉 인코더

111: 피쳐 맵 생성부

112: 중심 노드 임베딩부

113: 디포머블 이분 GCN부

120: 병목 계층

130: 컨셉 GCN

131: 상대적 컨셉 매핑부

132: 컨셉 상호 방향성 학습부

S110: 컨셉 인코더에서, 백본 네트워크로부터 입력 이미지에서 추출된 최종 피쳐 맵을 입력받고, 디포머블 이분 GCN을 포함해 컨셉을 임베딩하여 컨셉 노드를 출력하는 단계

S111: 백본 네트워크로부터 입력 이미지의 최종 피쳐 맵을 입력받아, 컨볼루션 레이어를 통해 N개의 노드 피쳐

로 구성된 평탄화된 피쳐 맵을 생성하는 단계

S112: N개의 노드 피쳐로 구성된 평탄화된 피쳐 맵을 고려해 중심 노드를 찾는 단계

S113: 디포머블 이분 GCN을 통해 중심 노드와 관련이 높은 이웃 노드를 탐색하고 최종 컨셉 노드를 유도하는 단계

S120: 병목 계층에서, 컨셉 노드와 노드 회귀 모듈을 이용해 컨셉 스코어를 생성하는 단계

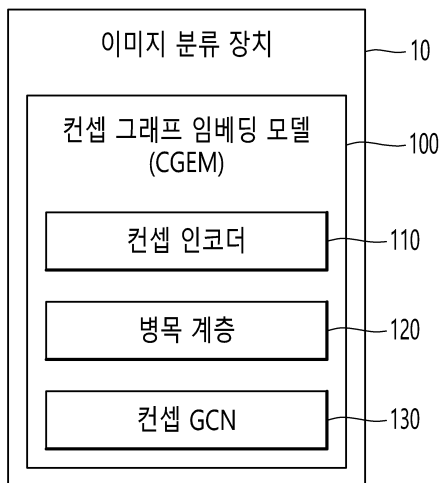
S130: 컨셉 GCN에서, 컨셉 스코어를 이용해 컨셉 노드 사이의 관련성을 학습하는 단계

S131: 컨셉 스코어를 사용해 각 컨셉 간의 상대적인 영향을 정량화해 방향성 인접행렬을 생성하는 단계

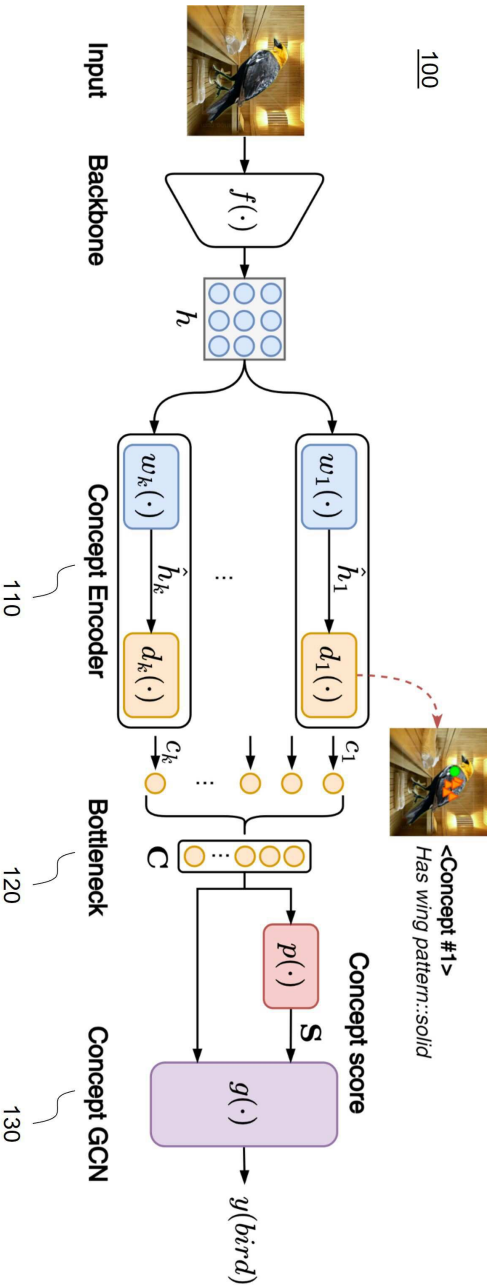
S132: 정방향 인접행렬을 이용해 컨셉 노드 사이의 관련성을 학습하는 제1 GCN과 역방향 인접행렬을 이용해 컨셉 노드 사이의 관련성을 학습하는 제2 GCN을 학습하며, 제1 GCN과 제2 GCN의 결과를 결합해 최종 클래스 레이블을 예측하는 단계

도면

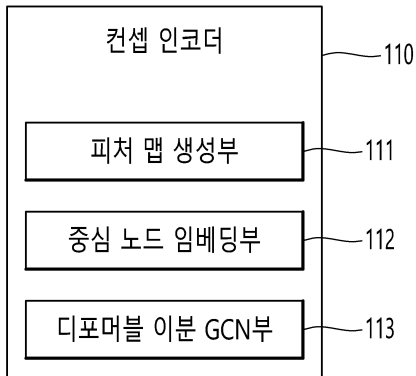
도면1



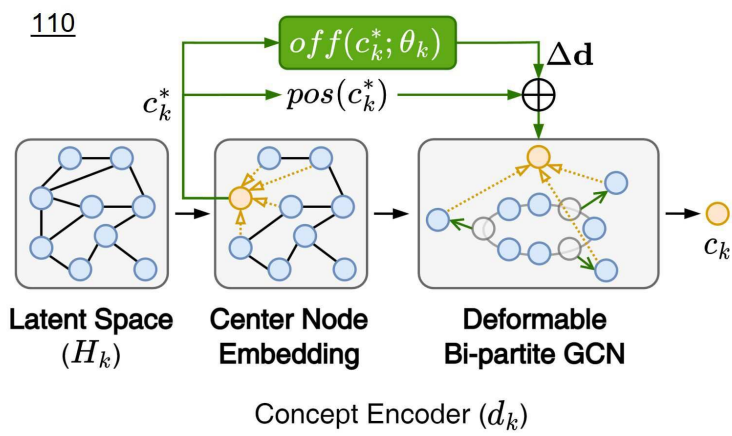
도면2



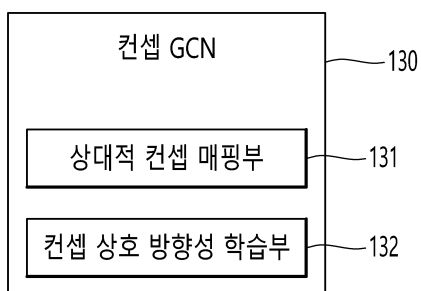
도면3



도면4

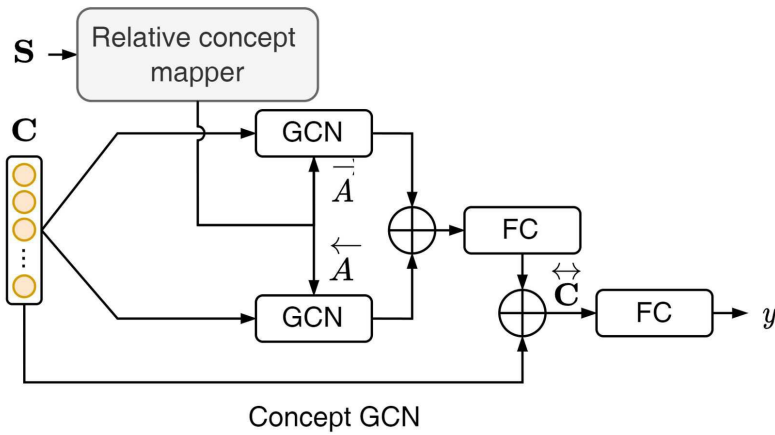


도면5

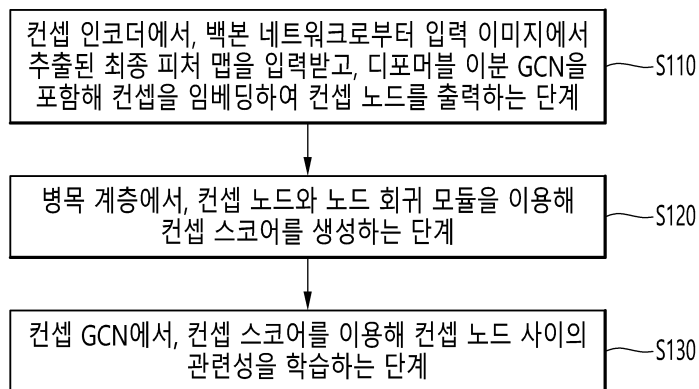


도면6

130

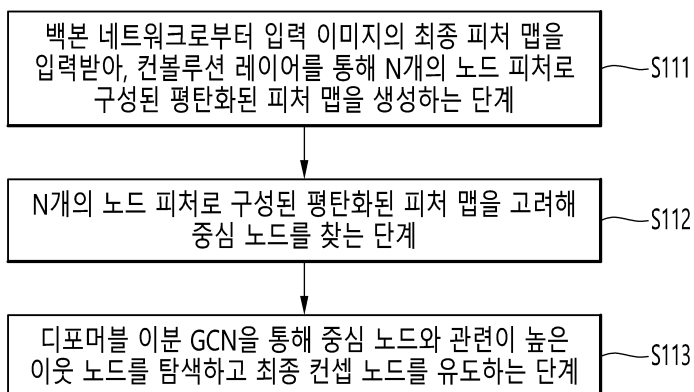


도면7



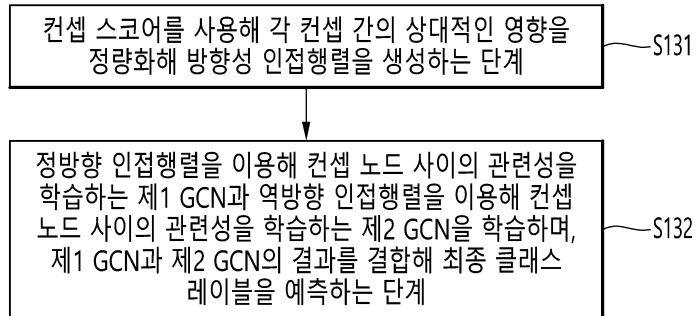
도면8

S110



도면9

S130



도면10

