(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국

(43) 국제공개일
2022 년 8 월 11 일 (11.08.2022)

WIPO | PCT

(10) 국제공개번호
WO 2022/169060 A1

(51) 국제특허분류:
G06F 16/906 (2019.01) *G06K 9/62* (2006.01)
G06F 16/28 (2019.01) *G06N 20/00* (2019.01)
G06F 40/58 (2020.01)

(21) 국제출원번호: PCT/KR2021/013812

(22) 국제출원일: 2021 년 10 월 7 일 (07.10.2021)

(25) 출원언어: 한국어

(26) 공개언어: 한국어

(30) 우선권정보:
10-2021-0016809 2021 년 2 월 5 일 (05.02.2021) KR

(71) 출원인: 한국기술교육대학교 산학협력단 (**KOREA UNIVERSITY OF TECHNOLOGY AND EDUCATION INDUSTRY-UNIVERSITY COOPERATION FOUNDATION**) [KR/KR]; 31253 충청남도 천안시 동남구 병천면 충절로 1600, Chungcheongnam-do (KR).

(72) 발명자: **오흥선 (OH, Heung-Seon)**; 31253 충청남도 천안시 동남구 병천면 충절로 1600, 공학2관 434호, Chungcheongnam-do (KR). **임상훈 (IM, Sang Hun)**; 31253 충청남도 천안시 동남구 병천면 충절로 1628-12, 205호, Chungcheongnam-do (KR). **조성웅 (JOH, Seong-Ung)**; 17867 경기도 평택시 현신3길 76, 209동 1043호 (용이동, 평택 용이2차 푸르지오), Gyeonggi-do (KR).

(74) 대리인: **특허법인오암 (OAM PATENT & TRADEMARK LAW FIRM)**; 35240 대전시 서구 문예로 69, 201호(둔산동, 오성빌딩), Daejeon (KR).

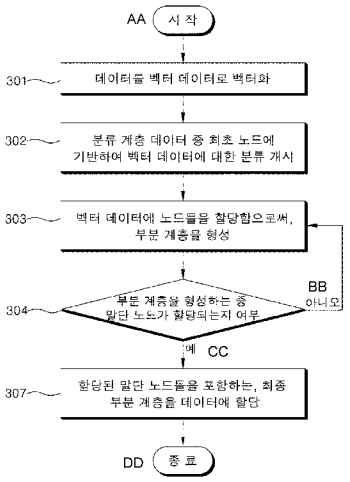
(81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 역내 권리의 보호를 위하여): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 유럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI

WO 2022/169060 A1

(54) **Title:** DATA CLASSIFICATION METHOD AND DEVICE FOR HIERARCHICAL MULTI-LABEL CLASSIFICATION

(54) **발명의 명칭:** 계층적 다중 레이블 분류를 위한 데이터 분류 방법 및 장치



301 ... Vectorize data into vector data
302 ... Start classification of vector data on basis of first node of classification layer data
303 ... Form partial layers by allocating nodes to vector data
304 ... Are terminal nodes allocated during formation of partial layers?
307 ... Allocate, to data, final partial layers including allocated terminal nodes
AA ... Start
BB ... No
CC ... Yes
DD ... End

(57) **Abstract:** A data classification device according to an embodiment of the present invention comprises an encoder for vectorizing data, a decoder that receives the vectorized data from the encoder, and classification data stored in the decoder, wherein the decoder may: start the classification of the vectorized data on the basis of the first node of the classification data; allocate a plurality of nodes included in the classification data to the vectorized data; form, for the vectorized data, partial layers based on the allocation nodes; when terminal nodes or temporary nodes among the nodes are allocated, end allocation to the partial layers to which the terminal nodes or temporary nodes have been allocated; and allocate, to the data, final partial layers formed on the basis of the partial layers.

(57) **요약서:** 본 발명의 다양한 실시 예에 따른 데이터 분류 장치는, 데이터를 벡터화 하는 인코더. 상기 인코더로부터 벡터화된 데이터를 수신하는 디코더 및 상기 디코더에 저장되는 분류 데이터를 포함하고, 상기 디코더는 상기 벡터화된 데이터를 상기 분류 데이터 중 최초 노드에 기반하여 분류를 개시하고, 상기 벡터화된 데이터에 상기 분류 데이터에 포함되는 복수 개의 노드들을 할당하고, 상기 벡터화된 데이터에 대하여, 상기 할당된 노드들에 기반한 부분 계층을 형성하고, 상기 노드들 중 말단 노드 또는 임시 노드가 할당되는 경우, 상기 말단 노드 또는 상기 임시 노드가 할당된 상기 부분 계층에 대한 할당을 종료하고, 상기 부분 계층에 기반하여 형성된 최종 부분 계층을 상기 데이터에 할당할 수 있다.

[다음 쪽 계속]

(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

공개:

— 국제조사보고서와 함께 (조약 제21조(3))

명세서

발명의 명칭: 계층적 다중 레이블 분류를 위한 데이터 분류 방법 및 장치

기술분야

[1] 본 발명의 다양한 실시 예들은 계층적 데이터 분류를 위한 데이터 분류 방법 및 데이터 분류 장치에 관한 것이다.

배경기술

[2] 최근 들어 머신러닝, 딥러닝 등의 다양한 학습 방법을 활용하여 데이터 분류를 자동화 하는 연구가 많이 진행되고 있다. 특히 분류에 활용되는 분류 데이터가 계층적 구조를 가지고 있을 때의 자동 분류와 하나의 데이터에 여러 분류가 할당되는 다중 분류에 대한 연구가 진행되고 있다.

[3] 일례로, 데이터의 분류가 계층적 구조(hierarchical structure)를 가지는 경우, 이러한 데이터에 대하여 데이터를 벡터(vector)로 표현하고 벡터로 표현된 데이터와 분류 데이터를 활용해 계층적 분류를 진행한다.

[4] 이러한 연구에 더해 인코더-디코더 구조를 이용한 기계요약, 기계번역 또는 생성하는 기술이 활발히 이루어 지고 있다.

[5] 일례로, 기계 번역의 경우, 특정 언어를 다른 언어로 번역하기 위하여 인코더 구조를 통해 특정 언어의 데이터를 벡터로 표현하고 벡터로 표현된 특정 언어 데이터를 디코더를 통해 목표 언어로 번역을 진행한다.

발명의 상세한 설명

기술적 과제

[6] 종래에 계층적 구조를 갖는 데이터를 분류하는 방법은 두가지를 포함할 수 있다.

[7] 첫번째로, 데이터 분류의 계층적 구조를 무시하고 플랫(flat)한 분류체계를 가정하고 분류하는 방식으로서, 이러한 방식을 이용하는 경우, 단일한 분류 알고리즘을 이용하여 분류할 수 있으나 데이터 분류의 계층적 구조를 고려하는데 제약이 있을 수 있다.

[8] 두번째는, 데이터 분류의 계층적 구조에서 각 계층 별로 별도의 분류 알고리즘을 이용하여 분류하는 방식으로서, 이러한 방식을 이용하는 경우, 데이터 분류의 계층적 구조를 고려할 수 있으나, 다수의 분류 알고리즘 또는 장치를 필요로 하므로, 분류를 수행하기에 상대적으로 데이터 및 시간의 소모가 증가할 수 있다.

[9] 따라서, 부분 계층을 활용하는 분류 방법을 통해, 데이터 분류의 계층적 구조를 고려하면서 단일한 알고리즘을 이용하여 계층적 다중 레이블 분류를 수행할 수 있다.

과제 해결 수단

[10] 본 발명의 다양한 실시 예에 따른 데이터 분류 장치는, 데이터를 벡터화 하는 인코더. 상기 인코더로부터 벡터화된 데이터를 수신하는 디코더 및 상기 디코더에 저장되는 분류 데이터를 포함하고, 상기 디코더는 상기 벡터화된 데이터를 상기 분류 데이터 중 최초 노드에 기반하여 분류를 개시하고, 상기 벡터화된 데이터에 상기 분류 데이터에 포함되는 복수 개의 노드들을 할당하고, 상기 벡터화된 데이터에 대하여, 상기 할당된 노드들에 기반한 부분 계층을 형성하고, 상기 노드들 중 말단 노드 또는 임시 노드가 할당되는 경우, 상기 말단 노드 또는 상기 임시 노드가 할당된 상기 부분 계층에 대한 할당을 종료하고, 상기 부분 계층에 기반하여 형성된 최종 부분 계층을 상기 데이터에 할당할 수 있다.

[11] 본 발명의 다양한 실시 예에 따른 데이터 분류 방법은 데이터를 벡터 데이터로 벡터화 하는 동작, 기 저장된 분류 데이터 중 최초 노드에 기반하여 상기 벡터 데이터에 대한 분류를 개시하는 동작, 상기 벡터 데이터에 대하여 상기 분류 데이터에 포함되는 노드들을 할당함으로써, 부분 계층을 형성하는 동작, 상기 부분 계층을 형성하는 동작을 수행하는 도중, 상기 노드들 중 말단 노드 또는 임시 노드가 할당되는 경우, 상기 분류를 종료하는 동작 및 상기 분류에 의해 형성된 최종 부분 계층을 상기 데이터에 할당하는 동작을 포함할 수 있다.

발명의 효과

[12] 본 발명은 인코더 및 디코더가 포함하는 단일한 알고리즘을 통해 계층적 분류 구조를 갖는 데이터를 분류하는 방법 및 장치에 관한 것이다.

[13] 본 발명은, 단일한 알고리즘을 통해 계층적 분류 구조를 갖는 데이터를 분류함으로써, 데이터 분류에 소모되는 시간 및 비용을 절약할 수 있다.

[14] 또한 본 발명은, 계층적 분류 구조를 갖는 데이터에 대하여 부분 계층을 활용하여 분류함으로써, 다중 계층 구조를 갖는 데이터에 대한 분류에서 계층 구조를 명확히 하는 분류를 수행할 수 있다.

[15] 본 발명의 다양한 실시 예들에서 얻을 수 있는 효과는 이상에서 언급한 효과들로 제한되지 않으며, 언급하지 않은 또 다른 효과들은 아래의 기재로부터 본 개시가 속하는 기술 분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

[16] 도 1은 데이터 분류를 위한 분류 데이터를 도시한다.

[17] 도 2는 일 실시 예에 따른 데이터 분류 장치의 구성 및 동작을 도시한다.

[18] 도 3은 일 실시 예에 따르면, 데이터 분류 방법을 나타내는 흐름도이다.

[19] 도 4는, 도 2 및 도 3을 통해 생성되는 출력을 도시한다.

발명의 실시를 위한 형태

[20] 본 발명의 실시 예들에 대한 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 상세하게 후술되어 있는 실시 예들을 참조하면 명확해질

것이다. 그러나 본 발명은 이하에서 개시되는 실시 예들에 한정되는 것이 아니라 서로 다른 다양한 형태로 구현될 수 있으며, 단지 본 실시 예들은 본 발명의 개시가 완전하도록 하고, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 발명의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 발명은 청구항의 범주에 의해 정의될 뿐이다. 명세서 전체에 걸쳐 동일 참조 부호는 동일 구성 요소를 지칭한다.

[21] 본 발명의 실시예들을 설명함에 있어서 공지 기능 또는 구성에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략할 것이다. 그리고 후술되는 용어들은 본 발명의 실시 예에서의 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.

[22] 이하, 첨부된 도면을 참조하여 본 발명의 실시예를 상세히 설명하기로 한다.

[23] 도 1은 데이터 분류를 위한 분류 데이터를 도시한다.

[24] 도 1을 참조하면, 데이터 분류를 위한 분류 데이터(100)는 각 계층에 따른 다양한 노드들(예: 최초 노드(101), 말단 노드(131, 132, 133))를 포함할 수 있다.

[25] 일 실시 예에 따르면, 분류 데이터(100)는 분류를 개시하기 위한 최초 노드(101)(또는 루트 노드(root node))를 포함할 수 있다. 일 실시 예에 따르면, 데이터에 최초 노드(101)가 할당됨에 따라, 데이터에 대한 분류가 개시될 수 있다.

[26] 일 실시 예에 따르면, 분류 데이터(100)는 최초 노드(101)에 대한 하위 노드(또는 자식 노드(child node))로서, 제1 계층 노드들(110)을 포함할 수 있다. 일 실시 예에 따르면, 제1 계층 노드들(110)은 각각 제1-1 노드(111), 제1-2 노드(112) 및 제1-3 노드(113)를 포함할 수 있다. 일 실시 예에 따르면, 최초 노드(101)는 제1 계층 노드들(110)에 대하여 상위 노드(또는 부모 노드(parent node)))로 참조될 수 있다.

[27] 일 실시 예에 따르면, 분류 데이터(100)는 제1 계층 노드들(110)에 대한 하위 노드(또는 자식 노드(child node))로서, 제2 계층 노드들(120)을 포함할 수 있다. 일 실시 예에 따르면, 제2 계층 노드들(120)은 각각 제2-1 노드(121), 제2-2 노드(122), 제2-3 노드(123) 및 제2-4 노드(124)를 포함할 수 있다. 일 실시 예에 따르면, 제1 계층 노드들(110)은 제2 계층 노드들(120)에 대하여 상위 노드(또는 부모 노드(parent node)))로 참조될 수 있다.

[28] 일 실시 예에 따르면, 분류 데이터(100)는 제2 계층 노드들(120)에 대한 하위 노드(또는 자식 노드(child node))로서, 제3 계층 노드들(130)을 포함할 수 있다. 일 실시 예에 따르면, 제3 계층 노드들(130)은 각각 제3-1 노드(131), 제3-2 노드(132) 및 제3-3 노드(133)를 포함할 수 있다. 일 실시 예에 따르면, 제2 계층 노드들(120)은 제3 계층 노드들(130)에 대하여 상위 노드(또는 부모 노드(parent node)))로 참조될 수 있다.

[29] 일 실시 예에 따르면, 제3 계층 노드들(130)은 말단 노드로 참조될 수 있다. 일 실시 예에 따라 제3 계층 노드들(130)(또는 말단 노드)이 데이터에 할당되는 경우, 분류가 종료될 수 있다. 이에 대한 구체적인 설명은 후술한다.

[30] 다만, 분류 데이터(100)의 계층은 상술한 구성(예: 제1 계층 노드들(110), 제2 계층 노드들(120) 및 제3 계층 노드들(130))에 한정되는 것은 아니고, 다양한 계층을 포함할 수 있다. 다른 실시 예(미도시)에 따르면 상술한 구성 중 일부(예: 제2 계층 노드들(120))가 생략될 수 있다. 또 다른 실시 예(미도시)에 따르면, 제3 계층 노드들(130)에 대한 하위 노드로서 제4 계층 노드들(미도시)이 추가될 수 있고, 제4 계층 노드들이 말단 노드로 참조될 수 있다.

[31] 도 2는 일 실시 예에 따른 데이터 분류 장치의 구성 및 동작을 도시한다.

[32] 도 2를 참조하면, 일 실시 예에 따른 데이터 분류 장치(200)는 인코더(210) 및 디코더(220)를 포함할 수 있다. 일 실시 예에 따르면, 데이터 분류 장치(200)는 인코더(210) 및 디코더(220)를 이용함으로써 데이터의 분류를 할당할 수 있다.

[33] 일 실시 예에 따르면, 인코더(210)는 데이터를 벡터화 할 수 있다. 일 실시 예에 따르면, 인코더(210)에 포함되는 데이터 전처리부(211)는 수신한 데이터를 벡터화 하기 위한 전처리를 수행할 수 있다. 예를 들어, 데이터를 전처리하는 동작은 데이터를 수정 또는 가공하는 동작을 포함할 수 있으나, 이에 한정되는 것은 아니다. 일 실시 예에 따르면, 인코더(210)에 포함되는 데이터 표상화 부(212)는 전처리된 데이터를 벡터 표현으로 변환할 수 있다.

[34] 일 실시 예에 따르면, 인코더(210)는 벡터화 된 데이터를 디코더(220)로 전송할 수 있다. 일 실시 예에 따르면, 디코더(220)는 인코더(210)로부터 벡터화 된 데이터를 수신할 수 있다. 일 실시 예에 따르면, 디코더(220)는 기 생성된 부분 계층을 데이터 분류 장치(200)로부터 수신할 수 있다.

[35] 일 실시 예에 따르면, 디코더(220)는 부분 계층 전처리부(221), 부분 계층 표상화 부(222), 부모-자식 노드 의존 관계 파악부(223), 데이터-부분 계층 관계 파악부(224), 하위 부분 계층 판별부(225) 및 부분 계층 확장부(226)를 포함할 수 있다. 다만, 상술한 디코더(220)의 구성 중 일부(예: 부모-자식 노드 의존 관계 파악부(223))는 생략되거나, 다른 구성(예: 부분 계층 전처리부(221))와 일체로 형성될 수 있고, 디코더(220)의 구성이 상술한 구성에 한정되는 것은 아니다.

[36] 일 실시 예에 따르면, 디코더(220)는 부분 계층 전처리부(221)를 통해, 분류 데이터(230)(예: 도 1의 분류 데이터(100))의 노드들 간의 의존 관계(예: 부모-자식 노드, 상위-하위 노드)를 정의할 수 있다. 일 실시 예에 따르면, 부분 계층 전처리부(221)는 부분 계층을, 벡터화 하기 위하여, 전처리할 수 있다.

[37] 일 실시 예에 따르면, 디코더(220)는 부분 계층 표상화 부(222)를 통해, 노드들이 정의된 분류 데이터(230) 및 이에 포함되는 부분 계층을 벡터화 할 수 있다. 예를 들어, 부분 계층 표상화 부(222)는, 부분 계층 전처리부(221)에서 노드들이 정의된 분류 데이터(230) 및 부분 계층을 벡터 표현으로 변환할 수 있다.

[38] 일 실시 예에 따르면, 부모-자식 노드 의존 관계 파악부(223)는 부분 계층 전처리부(221)에서 정의된 노드들 간의 의존 관계에 기반하여, 부분 계층의 말단 노드(예: 도 1의 제3 계층 노드들(130))를 정의할 수 있다. 일 실시 예에 따르면, 데이터-부분 계층 관계 파악부(224)는 부모-자식 노드 의존 관계 파악부(223)에서 정의된 말단 노드와 인코더(210)로부터 수신한 벡터화 된 데이터의 관계를 정의할 수 있다. 일 실시 예에 따르면, 데이터-부분 계층 관계 파악부(224)는 인코더(210)로부터 수신한 벡터화 된 데이터에 기반하여, 말단 노드를 재 정의할 수 있다.

[39] 일 실시 예에 따르면, 하위 부분 계층 판별부(225)는 벡터화 된 데이터 및 말단 노드에 기반하여, 추가 확장될 하위 노드(또는 자식 노드)를 판별할 수 있다. 일 실시 예에 따르면, 부분 계층 확장부(226)는 하위 부분 계층 판별부(225)에서 판별된 하위 노드(또는 자식 노드)를 수신한 부분 계층에 추가할 수 있다. 일 실시 예에 따르면, 부분 계층 확장부(226)는 하위 부분 계층 판별부(225)에서 판별된 하위 노드를 기 수신한 부분 계층에 추가함으로써, 부분 계층을 확장할 수 있다.

[40] 일 실시 예에 따르면, 디코더(220)는 확장된 부분 계층에 대하여 확장될 수 있는 하위 노드가 존재하는 경우, 확장된 부분 계층을 디코더(220)에 대한 상술한 과정을 반복할 수 있다. 이에 대한 구체적인 설명은 후술한다.

[41] 일 실시 예에 따르면, 디코더(220)는 확장된 부분 계층에 더 이상 확장될 수 없는 말단 노드가 할당된 경우, 확장된 부분 계층을 최종 부분 계층으로서 분류 할당부(240)에 전송할 수 있다. 일 실시 예에 따르면, 분류 할당부(240)는 수신한 최종 부분 계층을 데이터에 할당할 수 있다.

[42] 도 3은 일 실시 예에 따르면, 데이터 분류 방법을 나타내는 흐름도다.

[43] 도 1, 도 2 및 도 3을 함께 참조하면, 데이터 분류 장치(200)는 데이터를 수신하고, 이러한 데이터에 대하여 부분 계층을 생성하고 할당할 수 있다.

[44] 일 실시 예에 따르면, 동작 301에서 인코더(210)는 데이터를 수신하고, 수신한 데이터를 벡터 데이터로 벡터화 할 수 있다. 일 실시 예에 따르면, 동작 302에서 디코더(220)는 분류 데이터(230)에 기반하여 수신한 벡터 데이터에 대한 분류를 개시할 수 있다. 예를 들어, 디코더(220)는 분류 데이터(230) 중 최초 노드(101)가 할당됨에 따라, 벡터 데이터에 대한 분류를 개시할 수 있다.

[45] 일 실시 예에 따른 벡터 데이터에 대한 분류는 디코더(220)에 기 저장된 문류 데이터(230)에 기반하여 수행될 수 있다.

[46] 일 실시 예에 따르면, 동작 303에서 디코더(220)는 기 저장된 부분 계층에 노드들을 할당함으로써, 부분 계층을 형성할 수 있다. 다른 실시 예에 따르면, 동작 303에서 디코더(220)는 노드들에 기반하여, 부분 계층을 생성할 수 있다. 예를 들어, 디코더(220)는 벡터 데이터에 대해 제1-1 노드(111), 제2-1 노드(121) 및 제3-2 노드(132)를 할당함으로써, 부분 계층을 형성할 수 있다. 다른 예를 들어, 디코더(220)는 벡터 데이터에 대해 제1-2 노드(112) 및 제2-3 노드(123)를 할당함으로써, 부분 계층을 형성할 수 있다. 또한, 디코더(220)는 상술한

예시들과 같은 부분 계층을 동시에 복수 개로 형성할 수 있다.

[47] 일 실시 예에 따르면, 디코더(220)는 동작 304에서 부분 계층에 말단 노드가 할당되었는지 여부를 판단할 수 있다. 예를 들어, 동작 303을 통해 부분 계층을 형성하는 동작 중, 동작 304에서 디코더(220)는 부분 계층에 말단 노드(예: 도 1의 제3-1 노드(131))가 할당되었는지 여부를 판단할 수 있다. 일 실시 예에 따라 동작 304에서 부분 계층에 말단 노드가 할당되지 않는 것으로 판단된 경우, 디코더(220)는 동작 303 및 동작 304를 반복적으로 수행함으로써 부분 계층을 형성할 수 있다.

[48] 일 실시 예에 따라, 부분 계층에 말단 노드가 할당된 것으로 판단된 경우, 디코더(220)는 동작 307에서, 부분 계층에 할당된 말단 노드들을 포함하는 최종 부분 계층을 데이터에 할당할 수 있다. 예를 들어, 디코더(220)는 부분 계층에 할당된 말단 노드들(예: 제3-1 노드(131), 제2-3 노드(123))을 포함하는 최종 부분 계층을 생성하고, 데이터에 할당할 수 있다.

[49] 일 실시 예에 따르면, 디코더(220) 또는 분류 데이터(230)는 부분 계층 형성에 대한 종료 여부 판단을 위한 말단 노드 또는 임시 노드를 포함할 수 있다. 예를 들어, 하위 노드로서 제2-4 노드(124) 및 제3-3 노드(133)를 포함하는 제1-3 노드(113)를 최종 부분 계층으로 할당하기 위해서, 제1-3 노드(113)는 가상의 임시 노드를 하위 노드로 포함할 수 있다. 예를 들어, 제1-3 노드(113)의 하위 노드로 임시 노드가 할당되는 경우, 제1-3 노드(113)를 말단 노드로 포함하는 최종 부분 계층을 생성할 수 있다. 다만 상술한 임시 노드는 생략될 수 있고, 다른 실시 예에 따르면 임시 노드는 최초 노드(101) 및 다른 노드들에 대해서도 생성될 수 있다.

[50] 도 4는, 도 2 및 도 3을 통해 생성되는 출력을 도시한다.

[51] 도 1, 도 2, 도 3 및 도 4를 함께 참조하면, 일 실시 예에 따른 디코더(220)는 부분 계층들(401, 402, 403)을 형성함으로써, 분류 결과를 획득할 수 있다.

[52] 일 실시 예에 따르면, 디코더(220)는 도 3의 과정을 통해 제1 부분 계층(401), 제2 부분 계층(402) 및 제3 부분 계층(403)을 형성할 수 있다. 일 실시 예에 따르면, 디코더(220)는 제1 부분 계층(401), 제2 부분 계층(402) 및 제3 부분 계층(403)을 함께 형성할 수 있다. 일 실시 예에 따르면, 디코더(220)는 단일한 알고리즘을 통해 제1 부분 계층(401), 제2 부분 계층(402) 및 제3 부분 계층(403)을 형성할 수 있다. 단, 상술한 부분 계층들 중 일부(예: 제3 부분 계층(403))은 생략될 수 있고, 다른 부분 계층이 추가될 수 있다.

[53] 일 실시 예에 따르면, 디코더(220)는 제1 부분 계층(401), 제2 부분 계층(402) 및 제3 부분 계층(403)으로부터 벡터화 된 최종 부분 계층을 획득할 수 있다. 일 실시 예에 따르면, 디코더(220)는 제1 부분 계층(401), 제2 부분 계층(402) 및 제3 부분 계층(403)의 각각의 말단 노드들(132, 123, 113)을 포함하는 분류 결과(또는 최종 부분 계층)을 획득할 수 있다.

[54] 일 실시 예에 따른 데이터 분류 장치는, 데이터를 벡터화 하는 인코더. 상기

인코더로부터 벡터화된 데이터를 수신하는 디코더 및 상기 디코더에 저장되는 분류 데이터를 포함하고, 상기 디코더는 상기 벡터화된 데이터를 상기 분류 데이터 중 최초 노드에 기반하여 분류를 개시하고, 상기 벡터화된 데이터에 상기 분류 데이터에 포함되는 복수 개의 노드들을 할당하고, 상기 벡터화된 데이터에 대하여, 상기 할당된 노드들에 기반한 부분 계층을 형성하고, 상기 노드들 중 말단 노드 또는 임시 노드가 할당되는 경우, 상기 말단 노드 또는 상기 임시 노드가 할당된 상기 부분 계층에 대한 할당을 종료하고, 상기 부분 계층에 기반하여 형성된 최종 부분 계층 상기 데이터에 할당할 수 있다.

[55] 일 실시 예에 따르면, 상기 인코더는, 상기 데이터를 전처리하는 데이터 전처리부 및 상기 전처리된 데이터를 벡터 표현으로 변환하는 데이터 표상화부를 포함할 수 있다.

[56] 일 실시 예에 따른 디코더는, 상기 분류 데이터에 기반하여 상기 노드들의 의존 관계를 정의하는 부분 계층 전처리부, 상기 노드들을 벡터 표현으로 변환하는 부분 계층 표상화부 및 벡터로 표현된 상기 노드들과 상기 벡터화된 데이터의 관계를 파악하는 데이터-부분 계층 관계 파악부를 포함할 수 있다.

[57] 일 실시 예에 따르면, 데이터 분류 장치는 정의된 의존 관계에 기반하여 상기 벡터화된 데이터에 대한 상기 부분 계층을 확장하는, 부분 계층 확장부를 포함할 수 있다.

[58] 일 실시 예에 따른 분류 데이터는, 상기 최초 노드에 기반한 분류에서 종료 여부를 판단하기 위한, 임시 노드를 포함할 수 있다.

[59] 일 실시 예에 따르면, 상기 최종 부분 계층은 상기 형성된 부분 계층 중 상기 말단 노드에 해당하는 노드들의 데이터를 포함할 수 있다.

[60] 일 실시 예에 따르면, 데이터 분류 장치는 메모리를 포함하고, 디코더는 최종 부분 계층이 할당된 데이터를 메모리에 저장할 수 있다.

[61] 일 실시 예에 따른 데이터 분류 방법은 데이터를 벡터 데이터로 벡터화 하는 동작, 기 저장된 분류 데이터 중 최초 노드에 기반하여 상기 벡터 데이터에 대한 분류를 개시하는 동작, 상기 벡터 데이터에 대하여 상기 분류 데이터에 포함되는 노드들을 할당함으로써, 부분 계층을 형성하는 동작, 상기 부분 계층을 형성하는 동작을 수행하는 도중, 상기 노드들 중 말단 노드가 할당되는 경우, 상기 분류를 종료하는 동작 및 상기 분류에 의해 형성된 최종 부분 계층을 상기 데이터에 할당하는 동작을 포함할 수 있다.

[62] 일 실시 예에 따르면, 데이터 분류 방법은 상기 데이터를 상기 벡터 데이터로 벡터화 하는 동작은, 상기 데이터를 전처리하는 동작 및 상기 전처리된 데이터를 벡터 표현으로 변환하는 동작을 포함할 수 있다.

[63] 일 실시 예에 따르면, 상기 부분 계층을 형성하는 동작은, 상기 노드들 중 상위 노드에 기반하여 하위 노드를 할당하는 동작을 포함할 수 있다.

[64] 일 실시 예에 따르면, 상기 최초 노드에 기반하여 상기 벡터 데이터에 대한 분류를 개시하는 동작은, 상기 최초 노드에 기반한 분류에서 종료 여부를

판단하기 위하여 임시 노드를 할당하는 동작을 포함할 수 있다.

[65] 일 실시 예에 따르면 데이터 분류 방법은 분류 데이터에 포함되는 상기 노드들의 의존 관계를 정의하는 동작 및 상기 노드들을 벡터 표현으로 변환하는 동작을 포함할 수 있다.

[66] 일 실시 예에 따르면, 상기 최종 부분 계층은 상기 형성된 부분 계층 중 상기 말단 노드에 해당하는 노드들의 데이터를 포함할 수 있다.

[67] 일 실시 예에 따르면, 데이터 분류 방법은 상기 최종 부분 계층이 할당된 상기 데이터를 메모리에 저장하는 동작을 포함할 수 있다.

청구범위

[청구항 1] 데이터 분류 장치에 있어서,
데이터를 벡터화 하는 인코더;
상기 인코더로부터 벡터화된 데이터를 수신하는 디코더; 및
상기 디코더에 저장되는 분류 데이터를 포함하고,
상기 디코더는:
상기 벡터화된 데이터를 상기 분류 데이터 중 최초 노드에 기반하여 분류를 개시하고,
상기 벡터화된 데이터에 상기 분류 데이터에 포함되는 복수 개의 노드들을 할당하고,
상기 벡터화된 데이터에 대하여, 상기 할당된 노드들에 기반한 부분 계층을 형성하고,
상기 노드들 중 말단 노드 또는 임시 노드가 할당되는 경우, 상기 말단 노드 또는 상기 임시 노드가 할당된 상기 부분 계층에 대한 할당을 종료하고,
상기 부분 계층에 기반하여 형성된 최종 부분 계층을 상기 데이터에 할당하는, 데이터 분류 장치.

[청구항 2] 제1 항에 있어서,
상기 인코더는, 상기 데이터를 전처리하는 데이터 전처리부 및 상기 전처리된 데이터를 벡터 표현으로 변환하는 데이터 표상화부를 포함하는, 데이터 분류 장치.

[청구항 3] 제1 항에 있어서,
상기 디코더는,
상기 분류 데이터에 기반하여 상기 노드들의 의존 관계를 정의하는 부분 계층 전처리부,
상기 노드들을 벡터 표현으로 변환하는 부분 계층 표상화부 및
벡터로 표현된 상기 노드들과 상기 벡터화된 데이터의 관계를 파악하는 데이터-부분 계층 관계 파악부를 포함하는, 데이터 분류 장치.

[청구항 4] 제3 항에 있어서,
상기 의존 관계에 기반하여 상기 벡터화된 데이터에 대한 상기 부분 계층을 확장하는, 부분 계층 확장부를 포함하는, 데이터 분류 장치.

[청구항 5] 제1 항에 있어서,
상기 말단 노드 및 상기 임시 노드는 상기 최초 노드에 기반한 분류에서 상기 분류의 종료 여부를 판단하기 위한 노드인, 데이터 분류 장치.

[청구항 6] 제1 항에 있어서,
상기 최종 부분 계층은 상기 형성된 부분 계층 중 상기 말단 노드에 해당하는 노드들의 데이터를 포함하는, 데이터 분류 장치.

[청구항 7] 제1 항에 있어서,
상기 데이터 분류 장치는 메모리를 포함하고,
상기 디코더는 상기 최종 부분 계층이 할당된 상기 데이터를 상기 메모리에 저장하는, 데이터 분류 장치.

[청구항 8] 데이터 분류 방법에 있어서,
데이터를 벡터 데이터로 벡터화 하는 동작;
기 저장된 분류 데이터 중 최초 노드에 기반하여 상기 벡터 데이터에 대한 분류를 개시하는 동작;
상기 벡터 데이터에 대하여 상기 분류 데이터에 포함되는 노드들을 할당함으로써, 부분 계층을 형성하는 동작;
상기 부분 계층을 형성하는 동작을 수행하는 도중, 상기 노드들 중 말단 노드 또는 임시 노드가 할당되는 경우, 상기 분류를 종료하는 동작; 및
상기 분류에 의해 형성된 최종 부분 계층을 상기 데이터에 할당하는 동작을 포함하는, 데이터 분류 방법.

[청구항 9] 제8 항에 있어서,
상기 데이터를 상기 벡터 데이터로 벡터화 하는 동작은, 상기 데이터를 전처리하는 동작 및 상기 전처리된 데이터를 벡터 표현으로 변환하는 동작을 포함하는, 데이터 분류 방법.

[청구항 10] 제8 항에 있어서,
상기 부분 계층을 형성하는 동작은, 상기 노드들 중 상위 노드에 기반하여 하위 노드를 할당하는 동작을 포함하는, 데이터 분류 방법.

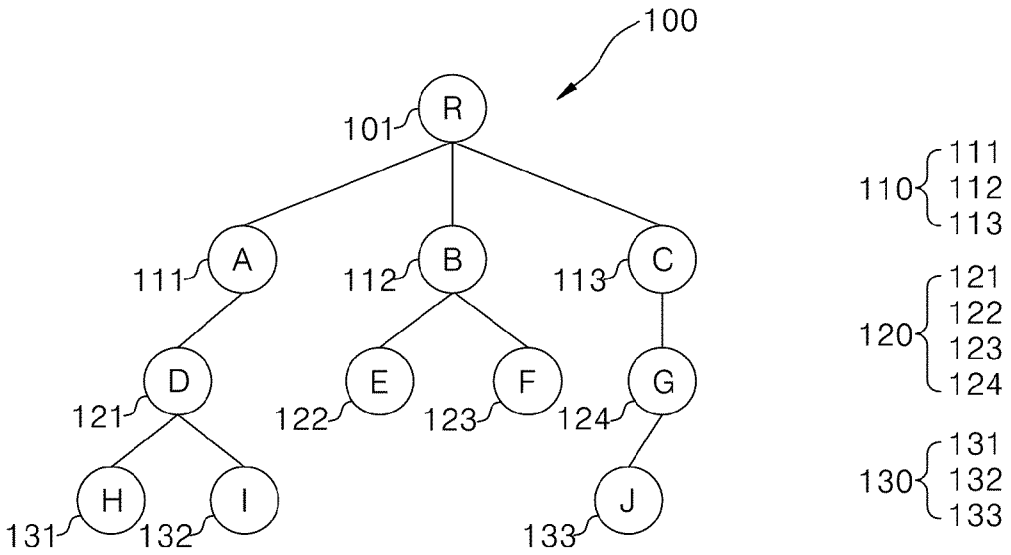
[청구항 11] 제8 항에 있어서,
상기 최초 노드에 기반하여 상기 벡터 데이터에 대한 분류를 개시하는 동작은, 상기 최초 노드에 기반한 분류에서 종료 여부를 판단하기 위하여 상기 임시 노드를 할당하는 동작을 포함하는, 데이터 분류 방법.

[청구항 12] 제8 항에 있어서,
상기 분류 데이터에 포함되는 상기 노드들의 의존 관계를 정의하는 동작 및 상기 노드들을 벡터 표현으로 변환하는 동작을 포함하는, 데이터 분류 방법.

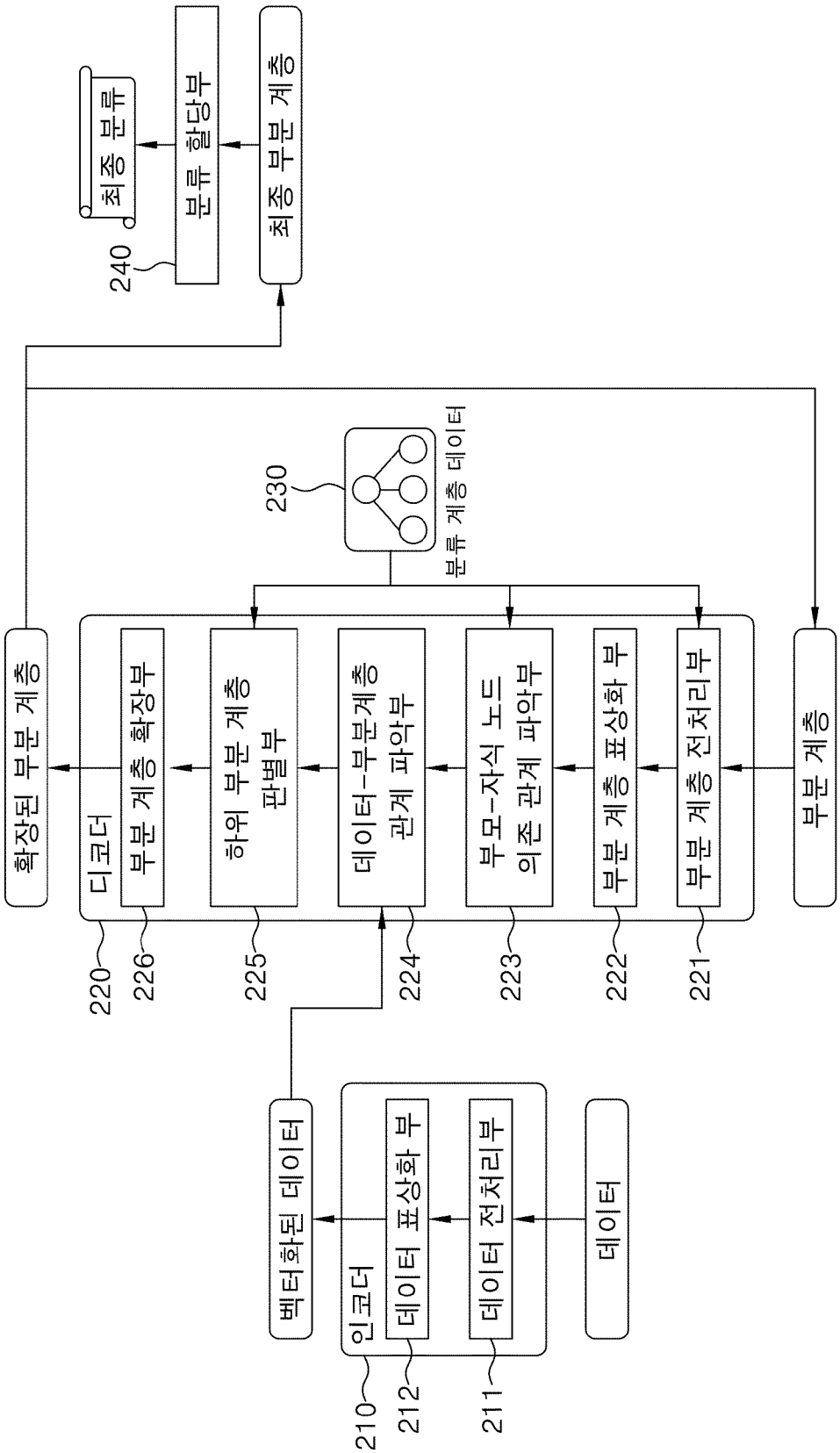
[청구항 13] 제8 항에 있어서,
상기 최종 부분 계층은 상기 형성된 부분 계층 중 상기 말단 노드에 해당하는 노드들의 데이터를 포함하는, 데이터 분류 방법.

[청구항 14] 제8 항에 있어서,
상기 최종 부분 계층이 할당된 상기 데이터를 메모리에 저장하는 동작을 포함하는, 데이터 분류 방법.

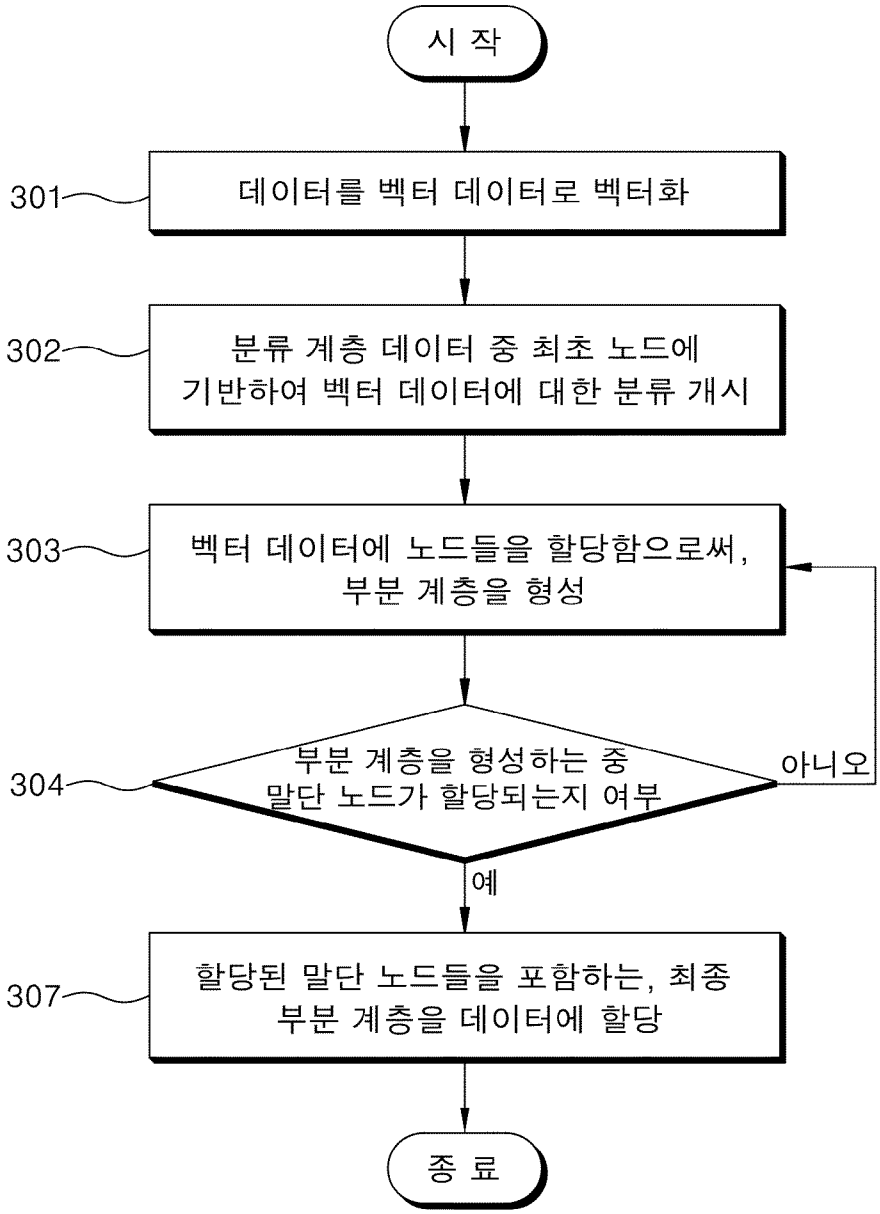
[도1]


100
101
R
111
A
112
B
113
C
D
E
F
G
121
122
123
124
H
I
J
131
132
133
110
111
112
113
120
121
122
123
124
130
131
132
133

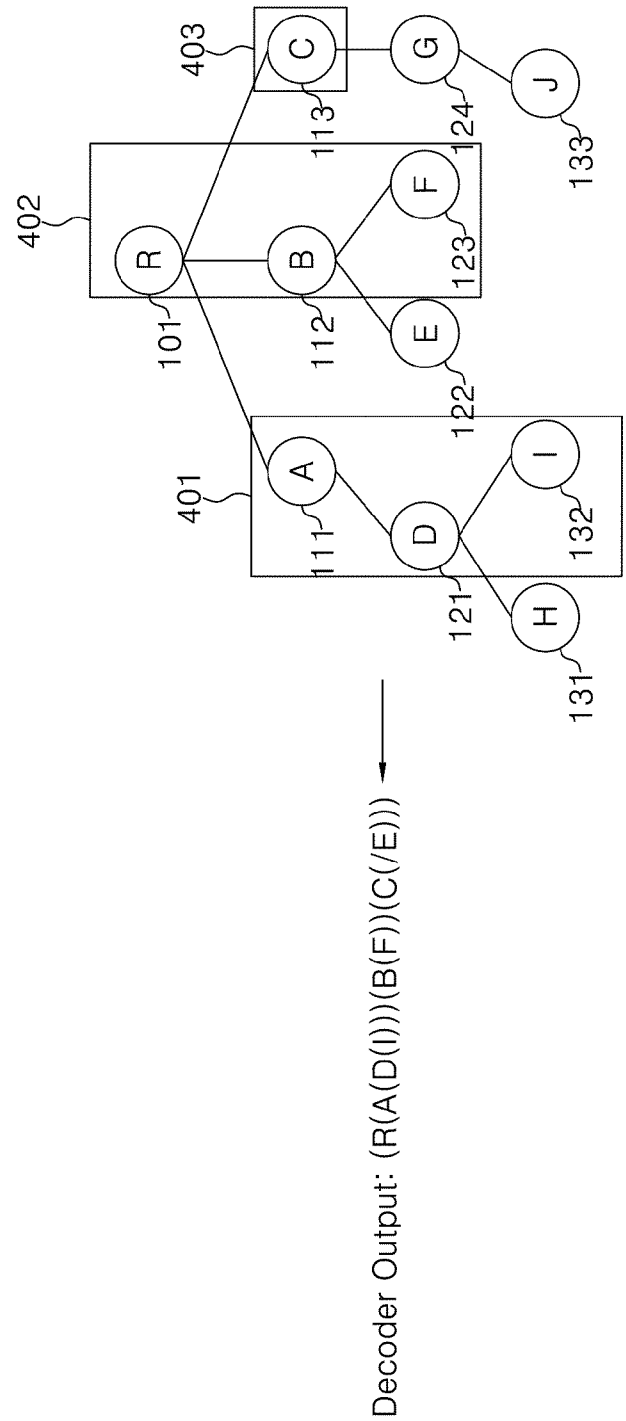
[도2]


최종 분류
분류 할당부
최종 부분 계층
240
230
분류 계층 데이터
확장된 부분 계층
디코더
부분 계층 확장부
하위 부분 계층 판별부
데이터-부분계층 관계 파악부
부모-자식 노드 의존 관계 파악부
부분 계층 표상화 부
부분 계층 전처리부
부분 계층
220
226
225
224
223
222
221
벡터화된 데이터
인코더
데이터 표상화 부
데이터 전처리부
데이터
210
212
211

[도3]


시 작
301
데이터를 벡터 데이터로 벡터화
302
분류 계층 데이터 중 최초 노드에
기반하여 벡터 데이터에 대한 분류 개시
303
벡터 데이터에 노드들을 할당함으로써,
부분 계층을 형성
304
부분 계층을 형성하는 중
말단 노드가 할당되는지 여부
아니오
예
307
할당된 말단 노드들을 포함하는, 최종
부분 계층을 데이터에 할당
종 료

[도4]


403
C
G
J
113
124
133
402
R
B
F
123
101
112
E
122
401
A
D
I
111
132
121
H
131
Decoder Output: (R(A(D(I)))(B(F))(C(/E)))

INTERNATIONAL SEARCH REPORT

International application No.
PCT/KR2021/013812

A. CLASSIFICATION OF SUBJECT MATTER

G06F 16/906(2019.01)i; **G06F 16/28**(2019.01)i; **G06F 40/58**(2020.01)i; **G06K 9/62**(2006.01)i; **G06N 20/00**(2019.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F 16/906(2019.01); G06F 12/00(2006.01); G06F 17/30(2006.01); G06F 40/20(2020.01); G06F 40/58(2020.01); G06K 9/00(2006.01); G06K 9/46(2006.01)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean utility models and applications for utility models: IPC as above
Japanese utility models and applications for utility models: IPC as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS (KIPO internal) & keywords: 계층적(hierarchical), 다중 레이블(multi-lable), 분류(classification), 벡터화(vectorize), 인코더(encoder), 디코더(decoder), 노드(node), 할당(allocate)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	KR 10-2002-0042939 A (CORELOGIX) 08 June 2002 (2002-06-08) See paragraphs [0013]-[0064]; and figures 1-6.	1-14
A	박영선. 다중계층구조를 갖는 데이터의 구조적 분류 및 분석. 이화여자대학교 과학기술대학원 2004학년도 석사학위 논문. pp. 1-57, January 2004 (PARK, Youngsun. Classification and analysis of the structure of data with multiple hierarcies. thesis 2004, Department of Computer Science & Engineering, Ewha Institute of Science and Technology.) See pages 9-20.	1-14
A	KR 10-2019-0135129 A (INCHEON NATIONAL UNIVERSITY RESEARCH & BUSINESS FOUNDATION) 06 December 2019 (2019-12-06) See paragraphs [0024]-[0070]; and figure 2.	1-14
A	KR 10-2019-0123372 A (GACHON UNIVERSITY OF INDUSTRY-ACADEMIC COOPERATION FOUNDATION) 01 November 2019 (2019-11-01) See paragraphs [0009] and [0031]-[0042]; and figure 1.	1-14

☑ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"D" document cited by the applicant in the international application
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
06 January 2022	**07 January 2022**

Name and mailing address of the ISA/KR	Authorized officer
Korean Intellectual Property Office **Government Complex-Daejeon Building 4, 189 Cheongsa-ro, Seo-gu, Daejeon 35208**	
Facsimile No. **+82-42-481-8578**	Telephone No.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 08-153121 A (HITACHI LTD.) 11 June 1996 (1996-06-11) See paragraphs [0027] and [0030]-[0035]; and figure 1.	1-14
PX	KR 10-2277643 B1 (KOREA UNIVERSITY OF TECHNOLOGY AND EDUCATION INDUSTRY-UNIVERSITY COOPERATION FOUNDATION) 14 July 2021 (2021-07-14) See claims 1-14. * This document is a published earlier application that serves as a basis for claiming priority of the present international application.	1-14

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/KR2021/013812

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
KR 10-2002-0042939 A	08 June 2002	KR 10-0427603 B1	30 April 2004
KR 10-2019-0135129 A	06 December 2019	KR 10-2069621 B1	23 January 2020
KR 10-2019-0123372 A	01 November 2019	KR 10-2138657 B1	28 July 2020
JP 08-153121 A	11 June 1996	DE 69526168 D1	08 May 2002
		DE 69526168 T2	28 November 2002
		EP 0704810 A1	03 April 1996
		EP 0704810 B1	03 April 2002
		JP 3669016 B2	06 July 2005
		US 5832470 A	03 November 1998
KR 10-2277643 B1	14 July 2021	None	

국 제 조 사 보 고 서

국제출원번호
PCT/KR2021/013812

A. 발명이 속하는 기술분류(국제특허분류(IPC))

G06F 16/906(2019.01)i; **G06F 16/28**(2019.01)i; **G06F 40/58**(2020.01)i; **G06K 9/62**(2006.01)i; **G06N 20/00**(2019.01)i

B. 조사된 분야

조사된 최소문헌(국제특허분류를 기재)

G06F 16/906(2019.01); G06F 12/00(2006.01); G06F 17/30(2006.01); G06F 40/20(2020.01); G06F 40/58(2020.01); G06K 9/00(2006.01); G06K 9/46(2006.01)

조사된 기술분야에 속하는 최소문헌 이외의 문헌

한국등록실용신안공보 및 한국공개실용신안공보: 조사된 최소문헌란에 기재된 IPC
일본등록실용신안공보 및 일본공개실용신안공보: 조사된 최소문헌란에 기재된 IPC

국제조사에 이용된 전산 데이터베이스(데이터베이스의 명칭 및 검색어(해당하는 경우))

eKOMPASS(특허청 내부 검색시스템) & 키워드: 계층적(hierarchical), 다중 레이블(multi-lable), 분류(classification), 벡터화(vectorize), 인코더(encoder), 디코더(decoder), 노드(node), 할당(allocate)

C. 관련 문헌

카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
A	KR 10-2002-0042939 A ((주)코어로직스) 2002.06.08 단락 [0013]-[0064]; 및 도면 1-6	1-14
A	박영선, `다중계층구조를 갖는 데이터의 구조적 분류 및 분석`, 이화여자대학교 과학기술대학원 2004학년도 석사학위 논문, 페이지 1-57, 2004.01 페이지 9-20	1-14
A	KR 10-2019-0135129 A (인천대학교 산학협력단) 2019.12.06 단락 [0024]-[0070]; 및 도면 2	1-14
A	KR 10-2019-0123372 A (가천대학교 산학협력단) 2019.11.01 단락 [0009], [0031]-[0042]; 및 도면 1	1-14
A	JP 08-153121 A (HITACHI LTD.) 1996.06.11 단락 [0027], [0030]-[0035]; 및 도면 1	1-14

☑ 추가 문헌이 C(계속)에 기재되어 있습니다.

☑ 대응특허에 관한 별지를 참조하십시오.

* 인용된 문헌의 특별 카테고리:

"A" 특별히 관련이 없는 것으로 보이는 일반적인 기술수준을 정의한 문헌
"D" 본 국제출원에서 출원인이 인용한 문헌
"E" 국제출원일보다 빠른 출원일 또는 우선일을 가지나 국제출원일 이후에 공개된 선출원 또는 특허 문헌
"L" 우선권 주장에 의문을 제기하는 문헌 또는 다른 인용문헌의 공개일 또는 다른 특별한 이유(이유를 명시)를 밝히기 위하여 인용된 문헌
"O" 구두 개시, 사용, 전시 또는 기타 수단을 언급하고 있는 문헌
"P" 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌
"T" 국제출원일 또는 우선일 후에 공개된 문헌으로, 출원과 상충하지 않으며 발명의 기초가 되는 원리나 이론을 이해하기 위해 인용된 문헌
"X" 특별한 관련이 있는 문헌. 해당 문헌 하나만으로 청구된 발명의 신규성 또는 진보성이 없는 것으로 본다.
"Y" 특별한 관련이 있는 문헌. 해당 문헌이 하나 이상의 다른 문헌과 조합하는 경우로 그 조합이 당업자에게 자명한 경우 청구된 발명은 진보성이 없는 것으로 본다.
"&" 동일한 대응특허문헌에 속하는 문헌

국제조사의 실제 완료일	국제조사보고서 발송일
2022년01월06일(06.01.2022)	**2022년01월07일(07.01.2022)**

ISA/KR의 명칭 및 우편주소

대한민국 특허청
(35208) 대전광역시 서구 청사로 189, 4동 (둔산동, 정부대전청사)

팩스 번호 **+82-42-481-8578**

심사관

양정록

전화번호 **+82-42-481-5709**

국 제 조 사 보 고 서

국제출원번호
PCT/KR2021/013812

C. 관련 문헌

카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
PX	KR 10-2277643 B1 (한국기술교육대학교 산학협력단) 2021.07.14 청구항 1-14 * 위 문헌은 본 국제출원의 우선권주장의 기초가 되는 선출원의 공개된 공보임.	1-14

국 제 조 사 보 고 서
대응특허에 관한 정보

국제출원번호
PCT/KR2021/013812

국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
KR 10-2002-0042939 A	2002/06/08	KR 10-0427603 B1	2004/04/30
KR 10-2019-0135129 A	2019/12/06	KR 10-2069621 B1	2020/01/23
KR 10-2019-0123372 A	2019/11/01	KR 10-2138657 B1	2020/07/28
JP 08-153121 A	1996/06/11	DE 69526168 D1	2002/05/08
		DE 69526168 T2	2002/11/28
		EP 0704810 A1	1996/04/03
		EP 0704810 B1	2002/04/03
		JP 3669016 B2	2005/07/06
		US 5832470 A	1998/11/03
KR 10-2277643 B1	2021/07/14	없음	