(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0068941
(43) 공개일자 2023년05월18일

(51) 국제특허분류(Int. Cl.)
G06N 3/08 (2023.01)
(52) CPC특허분류
G06N 3/08 (2023.01)
G06N 3/084 (2023.01)
(21) 출원번호 10-2022-0008262
(22) 출원일자 2022년01월20일
심사청구일자 2022년01월20일
(30) 우선권주장
1020210154864 2021년11월11일 대한민국(KR)

(71) 출원인
창원대학교 산학협력단
경상남도 창원시 의창구 창원대학로 20(퇴촌동)
(72) 발명자
차정원
경상남도 창원시 성산구 동산로 115, 105동 1407호 (상남동, 대동아파트)
성수진
경상남도 창원시 마산회원구 내서읍 광려천동로 22, 512호 (서광맨션)
(74) 대리인
김종선, 이형석

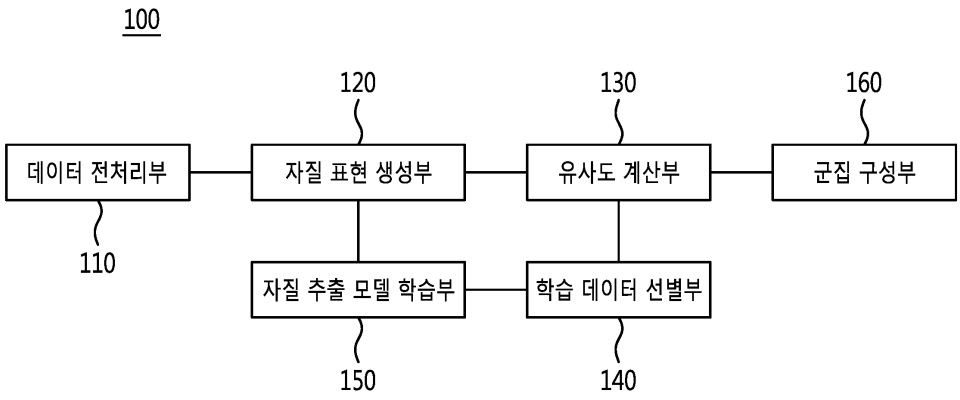
전체 청구항 수 : 총 7 항

(54) 발명의 명칭 **딥러닝 학습 기법을 이용하는 유사도 기반 클러스터링 장치 및 그 방법**

(57) 요 약

본 발명은 딥러닝 학습 기법을 이용하는 유사도 기반 클러스터링 장치 및 그 방법에 관한 것으로, 군집화 대상 데이터에 대하여 전처리를 수행하는 데이터 전처리부, 전처리가 수행된 군집화 대상 데이터에 대한 자질 표현 벡터를 생성하는 자질 표현 생성부, 생성된 자질 표현 벡터를 기반으로 전처리가 수행된 군집화 대상 데이터간의 유사도를 계산하는 유사도 계산부, 계산된 유사도를 토대로 자질 추출 모델을 학습하기 위한 학습 데이터를 선별하는 학습 데이터 선별부, 선별된 학습 데이터를 이용하여 자질 추출 모델의 가중치를 학습하고 업데이트하는 자질 추출 모델 학습부 및 자질 추출 모델의 학습 결과를 토대로 전처리가 수행된 군집화 대상 데이터에 대하여 군집을 구성하는 군집 구성부를 포함한다.

대 표 도 - 도1



이 발명을 지원한 국가연구개발사업

과제고유번호	1711148998
과제번호	K-21-L01-C03
부처명	과학기술정보통신부
과제관리(전문)기관명	한국과학기술정보연구원
연구사업명	Data/AI 기반 문제해결 체계 구축
연구과제명	과학기술 논문 조사 분석을 위한 딥러닝 기술 적용 연구
기 여 율	1/1
과제수행기관명	창원대학교 산학협력단
연구기간	2021.07.15 ~ 2021.11.15

명 세 서

청구범위

청구항 1

군집화 대상 데이터에 대하여 전처리를 수행하는 데이터 전처리부;

전처리가 수행된 군집화 대상 데이터에 대한 자질 표현 벡터를 생성하는 자질 표현 생성부;

생성된 자질 표현 벡터를 기반으로 전처리가 수행된 군집화 대상 데이터간의 유사도를 계산하는 유사도 계산부;

계산된 유사도를 토대로 자질 추출 모델을 학습하기 위한 학습 데이터를 선별하는 학습 데이터 선별부;

선별된 학습 데이터를 이용하여 자질 추출 모델의 가중치를 학습하고 업데이트하는 자질 추출 모델 학습부; 및

자질 추출 모델의 가중치 학습 및 업데이트 결과를 토대로 전처리가 수행된 군집화 대상 데이터에 대하여 군집을 구성하는 군집 구성부;

를 포함하는 것을 특징으로 하는 딥러닝 학습 기법을 이용하는 유사도 기반 클러스터링 장치.

청구항 2

제1항에 있어서,

상기 자질 표현 생성부는 전처리가 수행된 군집화 대상 데이터의 자질의 내적(dot product)을 나타내는 자질 표현 벡터를 생성하고, 자질 표현 벡터의 차원은 사용자가 기 지정한 군집의 개수와 동일하며, 최종 군집이 구성될 때까지 자질 표현 벡터는 상기 자질 추출 모델 학습부를 통해 갱신되는 것을 특징으로 하는 딥러닝 학습 기법을 이용하는 유사도 기반 클러스터링 장치.

청구항 3

제1항에 있어서,

상기 유사도 계산부는 코사인 유사도, 내적 유사도 및 유클리드 거리를 토대로 전처리가 수행된 군집화 대상 데이터간의 유사도를 계산하는 것을 특징으로 하는 딥러닝 학습 기법을 이용하는 유사도 기반 클러스터링 장치.

청구항 4

제1항에 있어서,

상기 학습 데이터 선별부는 선별된 학습 데이터 중 기 설정된 초기 상한선 기준 이상의 유사도를 갖는 학습 데이터와 기 설정된 초기 하한선 기준 이하의 유사도를 갖는 학습 데이터를 선별하여 레이블을 구축하는 것을 특징으로 하는 딥러닝 학습 기법을 이용하는 유사도 기반 클러스터링 장치.

청구항 5

제1항에 있어서,

상기 자질 추출 모델 학습부는 선별된 학습 데이터를 토대로 구축된 레이블 이용하여 지도 학습 방식으로 자질 추출 모델의 가중치를 학습하는 것을 특징으로 하는 딥러닝 학습 기법을 이용하는 유사도 기반 클러스터링 장치.

청구항 6

제1항에 있어서,

상기 군집 구성부는 상기 자질 추출 모델 학습부에 의해 업데이트 진행에 따라 상한선 유사도와 하한선 유사도가 동일한 값을 가질 때, 군집화 대상 데이터에 대하여 군집을 구성하는 것을 특징으로 하는 딥러닝 학습 기법을 이용하는 유사도 기반 클러스터링 장치.

청구항 7

데이터 전처리부에 의해, 군집화 대상 데이터에 대하여 전처리를 수행하는 단계;

자질 표현 생성부에 의해, 전처리가 수행된 군집화 대상 데이터에 대한 자질 표현 벡터를 생성하는 단계;

유사도 계산부에 의해, 생성된 자질 표현 벡터를 기반으로 전처리가 수행된 군집화 대상 데이터간의 유사도를 계산하는 단계;

학습 데이터 선별부에 의해, 계산된 유사도를 토대로 자질 추출 모델을 학습하기 위한 학습 데이터를 선별하는 단계;

자질 추출 모델 학습부에 의해, 선별된 학습 데이터를 이용하여 자질 추출 모델의 가중치를 학습하고 업데이트하는 단계; 및

군집 구성부에 의해, 자질 추출 모델의 가중치 학습 및 업데이트 결과를 토대로 전처리가 수행된 군집화 대상 데이터에 대하여 군집을 구성하는 단계;

를 포함하는 것을 특징으로 하는 딥러닝 학습 기법을 이용하는 유사도 기반 클러스터링 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 딥러닝 학습 기법을 이용하는 유사도 기반 클러스터링 장치 및 그 방법에 관한 것으로, 보다 자세하게는 지속적으로 자질 추출 모델의 가중치를 업데이트하여 군집화 대상 데이터의 특성을 보다 잘 표현하는 자질 표현 벡터 값을 구하고 이를 통해 개선된 군집 결과물을 획득할 수 있는 딥러닝 학습 기법을 이용하는 유사도 기반 클러스터링 장치 및 그 방법에 관한 것이다.

배 경 기 술

[0002] 인공신경망을 기반으로 하는 딥러닝이 다양한 분야에서 기존의 기계 학습을 능가하는 우수한 성능을 보여주고 있으나, 신경망 모델이 특정 작업을 수행하기 위해서 모델 가중치의 학습이 선행되어야 한다.

[0003] 머신러닝 및 딥러닝은 주로 모델의 예측값을 입력 샘플에 대한 정답과 비교하여 학습하는 지도 학습 방법으로 학습한다.

[0004] 지도 학습을 수행하기 위해서는 학습에 사용하기 위한 모든 입력 샘플에 대하여 정답이 부착되어 있어야 한다.

[0005] 정답이 없는 데이터에 대하여 유의미한 자질을 자동으로 추출하기 위해 비지도 학습 방법이 사용된다. 비지도 학습 방법 중 군집화가 있으며 서로 동일하거나 유사한 특징을 가진 데이터 군집을 구하여 새로운 분류 기준을 정의하는데 사용될 수 있다.

[0006] 기존의 군집화 방법에서는 군집의 구성을 학습할 뿐 입력 샘플에서 자질을 추출하는 방법 및 결과값은 군집 생성 이전에 정의된 후 업데이트되지 않기 때문에 입력 샘플의 특성을 잘 나타내는 방법을 선정하는 것이 매우 중요하다.

[0007] 이와 관련하여, 한국공개특허 제2021-0074023호는 "딥러닝 기반의 문서 유사도 측정 모델을 이용한 중복 문서 탐지 방법 및 시스템"에 관하여 개시하고 있다.

발명의 내용

해결하려는 과제

[0008] 본 발명은 상기와 같은 문제점을 해결하기 위해 발명된 것으로서, 전처리가 수행된 군집화 대상 데이터에 대한 자질 표현 벡터를 생성하고, 생성된 자질 표현 벡터를 기반으로 전처리가 수행된 군집화 대상 데이터간의 유사도를 계산하는 딥러닝 학습 기법을 이용하는 유사도 기반 클러스터링 장치 및 그 방법을 제공하는데 그 목적이 있다.

[0009] 또한, 본 발명은 선별된 학습 데이터를 이용하여 자질 추출 모델의 가중치를 학습하고 업데이트하고, 자질 추출 모델의 학습 및 업데이트 결과를 토대로 전처리가 수행된 군집화 대상 데이터에 대하여 군집을 구성하는 딥러닝

학습 기법을 이용하는 유사도 기반 클러스터링 장치 및 그 방법을 제공하는데 그 목적이 있다.

과제의 해결 수단

[0010] 상기의 목적을 달성하기 위한 본 발명에 따른 딥러닝 학습 기법을 이용하는 유사도 기반 클러스터링 장치는 군집화 대상 데이터에 대하여 전처리를 수행하는 데이터 전처리부; 전처리가 수행된 군집화 대상 데이터에 대한 자질 표현 벡터를 생성하는 자질 표현 생성부; 생성된 자질 표현 벡터를 기반으로 전처리가 수행된 군집화 대상 데이터간의 유사도를 계산하는 유사도 계산부; 계산된 유사도를 토대로 자질 추출 모델을 학습하기 위한 학습 데이터를 선별하는 학습 데이터 선별부; 선별된 학습 데이터를 이용하여 자질 추출 모델의 가중치를 학습하고 업데이트하는 자질 추출 모델 학습부; 및 자질 추출 모델의 가중치 학습 및 업데이트 결과를 토대로 전처리가 수행된 군집화 대상 데이터에 대하여 군집을 구성하는 군집 구성부;를 포함한다.

[0011] 또한, 상기 자질 표현 생성부는 전처리가 수행된 군집화 대상 데이터의 자질의 내적(dot product)을 나타내는 자질 표현 벡터를 생성하고, 자질 표현 벡터의 차원은 사용자가 기 지정한 군집의 개수와 동일하며, 최종 군집이 구성될 때까지 자질 표현 벡터는 상기 자질 추출 모델 학습부를 통해 갱신되는 것을 특징으로 한다.

[0012] 또한, 상기 유사도 계산부는 코사인 유사도, 내적 유사도 및 유클리드 거리를 토대로 전처리가 수행된 군집화 대상 데이터간의 유사도를 계산하는 것을 특징으로 한다.

[0013] 또한, 상기 학습 데이터 선별부는 선별된 학습 데이터 중 기 설정된 초기 상한선 기준 이상의 유사도를 갖는 학습 데이터와 기 설정된 초기 하한선 기준 이하의 유사도를 갖는 학습 데이터를 선별하여 레이블을 구축하는 것을 특징으로 한다.

[0014] 또한, 상기 자질 추출 모델 학습부는 선별된 학습 데이터를 토대로 구축된 레이블 이용하여 지도 학습 방식으로 자질 추출 모델의 가중치를 학습하는 것을 특징으로 한다.

[0015] 또한, 상기 군집 구성부는 상기 자질 추출 모델 학습부에 의해 업데이트 진행에 따라 상한선 유사도와 하한선 유사도가 동일한 값을 가질 때, 군집화 대상 데이터에 대하여 군집을 구성하는 것을 특징으로 한다.

[0016] 상기의 목적을 달성하기 위한 본 발명에 따른 딥러닝 학습 기법을 이용하는 유사도 기반 클러스터링 방법은 데이터 전처리부에 의해, 군집화 대상 데이터에 대하여 전처리를 수행하는 단계; 자질 표현 생성부에 의해, 전처리가 수행된 군집화 대상 데이터에 대한 자질 표현 벡터를 생성하는 단계; 유사도 계산부에 의해, 생성된 자질 표현 벡터를 기반으로 전처리가 수행된 군집화 대상 데이터간의 유사도를 계산하는 단계; 학습 데이터 선별부에 의해, 계산된 유사도를 토대로 자질 추출 모델을 학습하기 위한 학습 데이터를 선별하는 단계; 자질 추출 모델 학습부에 의해, 선별된 학습 데이터를 이용하여 자질 추출 모델의 가중치를 학습하고 업데이트하는 단계; 및 군집 구성부에 의해, 자질 추출 모델의 가중치 학습 및 업데이트 결과를 토대로 전처리가 수행된 군집화 대상 데이터에 대하여 군집을 구성하는 단계;를 포함한다.

발명의 효과

[0017] 본 발명은 딥러닝 학습 기법을 이용하는 유사도 기반 클러스터링 장치 및 그 방법은 전처리가 수행된 군집화 대상 데이터에 대한 자질을 함축적으로 나타내는 자질 표현 벡터를 생성하고, 생성된 자질 표현 벡터를 기반으로 전처리가 수행된 군집화 대상 데이터간의 유사도를 계산함으로써, 보다 쉽게 유사도를 추출하여 딥러닝 학습 기능을 향상시킬 수 있는 효과가 있다.

[0018] 또한, 본 발명은 선별된 학습 데이터를 이용하여 자질 추출 모델의 가중치를 학습하고 업데이트하고, 자질 추출 모델의 학습 및 업데이트 결과를 토대로 전처리가 수행된 군집화 대상 데이터에 대하여 군집을 구성함으로써, 개선된 군집 결과물을 획득할 수 있는 효과가 있다.

도면의 간단한 설명

[0019] 도 1은 본 발명에 따른 딥러닝 학습 기법을 이용하는 유사도 기반 클러스터링 장치의 구성을 설명하기 위한 도면이다.

도 2는 본 발명에 따른 딥러닝 학습 기법을 이용하는 유사도 기반 클러스터링 방법의 순서를 설명하기 위한 순서도이다.

발명을 실시하기 위한 구체적인 내용

[0020] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세하게 설명하고자 한다.

[0021] 그러나 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다. 각 도면을 설명하면서 유사한 참조부호를 유사한 구성요소에 대해 사용하였다.

[0022] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소가 다른 구성요소에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는, 중간에 다른 구성요소가 존재하지 않는 것으로 이해되어야 할 것이다.

[0023] 본 출원에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 출원에서, "포함하다" 또는 "가지다" 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능선을 미리 배제하지 않는 것으로 이해되어야 한다.

[0024] 이하, 첨부한 도면들을 참조하여, 본 발명의 바람직한 실시예를 보다 상세하게 설명하고자 한다. 이하, 도면상의 동일한 구성요소에 대해서는 동일한 참조부호를 사용하고 동일한 구성요소에 대해서 중복된 설명은 생략한다.

[0026] 도 1은 본 발명에 따른 딥러닝 학습 기법을 이용하는 유사도 기반 클러스터링 장치의 구성을 설명하기 위한 도면이다.

[0027] 도 1을 참조하여 설명하면, 본 발명에 따른 딥러닝 학습 기법을 이용하는 유사도 기반 클러스터링 장치(100)는 크게 데이터 전처리부(110), 자질 표현 생성부(120), 유사도 계산부(130), 학습 데이터 선별부(140), 자질 추출 모델 학습부(150) 및 군집 구성부(160)를 포함한다.

[0028] 데이터 전처리부(110)는 군집화 대상 데이터에 대하여 사용자가 사전에 정의한 방법으로 전처리를 수행한다. 여기서, 군집화 대상 데이터는 유사한 특성을 가진 데이터들을 요소로 갖는 군집을 구하고자 하는 모든 분야 및 형식의 데이터로, 자연어로 작성된 문서, 이미지, 음성 데이터 등을 포함할 수 있다.

[0029] 데이터 전처리부(110)는 군집화 대상 데이터를 자질 추출 모델에 입력하기에 앞서 사용자가 사전에 정의한 처리를 수행하고 자질 추출 모델이 처리할 수 있는 형식으로 변경한다. 이때, 사용자는 사전에 데이터의 특성을 파악하여 불필요한 정보가 자질 추출 모델에 반영되지 않도록 규칙을 작성하고, 데이터 전처리부(110)는 작성된 규칙을 따라 전처리를 수행한다.

[0030] 자질 표현 생성부(120)는 전처리가 수행된 군집화 대상 데이터에 대한 자질 표현 벡터를 생성한다.

[0031] 이때, 자질 표현 벡터는 전처리가 수행된 군집화 대상 데이터의 자질의 내적(dot product)을 나타내고, 유사한 자질을 가진 군집화 대상 데이터는 유사한 자질 벡터 표현을 갖는다. 그리고 자질 표현 벡터의 차원은 사용자가 기 지정한 군집의 개수와 동일하며 이후 군집화 대상 데이터가 포함되는 군집을 결정하는데 사용된다. 이러한 자질 표현을 생성하기 위한 모델은 입력데 대한 벡터 값을 출력하며 학습을 통해 모델 가중치를 업데이트할 수 있어야 한다.

[0032] 유사도 계산부(130)는 생성된 자질 표현 벡터를 기반으로 전처리가 수행된 군집화 대상 데이터간의 유사도를 계산한다.

[0033] 유사도 계산부(130)는 코사인 유사도, 내적 유사도 및 유클리드 거리를 토대로 전처리가 수행된 군집화 대상 데이터간의 유사도를 계산할 수 있다.

[0034] 학습 데이터 선별부(140)는 계산된 유사도를 토대로 자질 추출 모델을 학습하기 위한 학습 데이터를 선별한다.

[0035] 학습 데이터 선별부(140)는 선별된 학습 데이터 중 기 설정된 초기 상한선 기준 이상의 유사도를 갖는 학습 데이터와 기 설정된 초기 하한선 기준 이하의 유사도를 갖는 학습 데이터를 선별하여 레이블을 구축할 수 있다. 여기서, 지정된 범위 내에 포함되지 않는 샘플은 학습 데이터에서 제외된다.

[0036] 학습 데이터 선별부(140)는 계산된 유사도를 토대로 자질 추출 모델을 학습하기 위한 학습 데이터를 선별하는 수학식 1은 하기와 같다.

수학식 1

[0037]
$$\min\ E(w)_w = \sum_{ij} L(r_{ij}, g(x_i, x_j; w))$$
$$g(x_i, x_j; w) = f(x_i; w) \cdot f(x_j; w) = 1_i \cdot 1_j$$
$$\forall_i,\ \ |\ |1_i|\ |_2 = 1,\ h_{ih} \geq 0, h = 1, \cdots, k$$
$$r_{ij} = \begin{cases} 1, & \text{if } 1_i \cdot 1_j \geq u(\lambda), \\ 0, & \text{if } 1_i \cdot 1_j \leq l(\lambda),\ i, j = 1, \cdots n \\ None, & \text{otherwise} \end{cases}$$

[0038] 여기서, 상한선 기준과 하한선 기존의 초기값은 사용자에 의해 지정되지만 자질 추출모델의 업데이트가 진행되며 그 차이가 감소하여 최종적으로 상한선과 하한선이 동일한 값을 가지고, 이때 모든 학습 데이터에 대한 군집 결과가 정의된다.

[0039] 자질 추출 모델 학습부(150)는 선별된 학습 데이터를 이용하여 자질 추출 모델의 가중치를 학습하고 업데이트한다.

[0040] 자질 추출 모델 학습부(150)는 선별된 학습 데이터를 토대로 구축된 레이블 이용하여 지도 학습 방식으로 자질 추출 모델의 가중치를 학습하며, 자질 표현 생성부(120)에서 구한 결과값을 정답 레이블로 사용한다.

[0041] 군집 구성부(160)는 자질 추출 모델의 가중치 학습 및 업데이트 결과를 토대로 전처리가 수행된 군집화 대상 데이터에 대하여 군집을 구성한다.

[0042] 군집 구성부(160)는 자질 추출 모델 학습부(150)에 의해 업데이트 진행에 따라 상한선 유사도와 하한선 유사도가 동일한 값을 가질 때, 군집화 대상 데이터에 대하여 군집을 구성한다. 이때, 자질 추출 모델의 결과 벡터의 차원 수가 군집 개수와 동일하기 때문에 가장 높은 값을 갖는 차원이 동일한 군집화 대상 데이터들을 같은 군집에 할당한다.

[0044] 도 2는 본 발명에 따른 딥러닝 학습 기법을 이용하는 유사도 기반 클러스터링 방법의 순서를 설명하기 위한 순서도이다.

[0045] 도 2를 참조하여 설명하면, 본 발명에 따른 딥러닝 학습 기법을 이용하는 유사도 기반 클러스터링 방법은 딥러닝 학습 기법을 이용하는 유사도 기반 클러스터링 장치를 이용하는 것으로, 이하 중복되는 설명은 생략하기로 한다.

[0046] 먼저, 입력된 군집화 대상 데이터에 대하여 사용자가 사전에 정의한 방법으로 전처리를 수행한다(S100). 여기서, 군집화 대상 데이터는 유사한 특성을 가진 데이터들을 요소로 갖는 군집을 구하고자 하는 모든 분야 및 형식의 데이터로, 자연어로 작성된 문서, 이미지, 음성 데이터 등을 포함할 수 있다.

[0047] S100 단계는 군집화 대상 데이터를 자질 추출 모델에 입력하기에 앞서 사용자가 사전에 정의한 처리를 수행하고 자질 추출 모델이 처리할 수 있는 형식으로 변경한다.

[0048] 다음, 전처리가 수행된 군집화 대상 데이터에 대한 자질 표현 벡터를 생성한다(S200).

[0049] 다음, 생성된 자질 표현 벡터를 기반으로 전처리가 수행된 군집화 대상 데이터간의 유사도를 계산한다(S300).

[0050] S300 단계는 코사인 유사도, 내적 유사도 및 유클리드 거리를 토대로 전처리가 수행된 군집화 대상 데이터간의 유사도를 계산할 수 있다.

[0051] 다음, 계산된 유사도를 토대로 자질 추출 모델을 학습하기 위한 학습 데이터를 선별한다(S400).

[0052] S400 단계는 선별된 학습 데이터 중 기 설정된 초기 상한선 기준 이상의 유사도를 갖는 학습 데이터와 기 설정된 초기 하한선 기준 이하의 유사도를 갖는 학습 데이터를 선별하여 레이블을 구축하고, 상한선 기준과 하한선

기존의 초기값은 사용자에 의해 지정되지만 자질 추출 모델의 업데이트가 진행되며 그 차이가 감소하여 최종적으로 상한선과 하한선이 동일한 값을 가지게 된다.

[0053] 다음, 선별된 학습 데이터의 유사도 상한선과 유사도 하한선이 동일한 값을 갖는지 여부를 파악한다(S500).

[0054] S500 단계에서 선별된 학습 데이터의 유사도 상한선과 유사도 하한선이 동일한 값을 갖지 않은 경우, 선별된 학습 데이터를 이용하여 자질 추출 모델의 가중치를 학습하고 업데이트한다(S600).

[0055] S600 단계는 선별된 학습 데이터를 토대로 구축된 레이블 이용하여 지도 학습 방식으로 자질 추출 모델의 가중치를 학습한다.

[0056] S500 단계에서 선별된 학습 데이터의 유사도 상한선과 유사도 하한선이 동일한 값을 갖은 경우, 자질 추출 모델의 가중치 학습 및 업데이트 결과를 토대로 전처리가 수행된 군집화 대상 데이터에 대하여 군집을 구성한다(S700).

[0057] S700 단계에서 자질 추출 모델의 가중치 학습 및 업데이트 진행에 따라 상한선 유사도와 하한선 유사도가 동일한 값을 가질 때, 군집화 대상 데이터에 대하여 군집을 구성한다.

[0059] 이상 본 명세서에서 설명한 기능적 동작과 본 주제에 관한 실시형태들은 본 명세서에서 개시한 구조들 및 그들의 구조적인 등가물을 포함하여 디지털 전자 회로나 컴퓨터 소프트웨어, 펌웨어 또는 하드웨어에서 또는 이들 중 하나 이상이 조합에서 구현 가능하다.

[0060] 본 명세서에서 기술하는 주제의 실시형태는 하나 이상이 컴퓨터 프로그램 제품, 다시 말해 데이터 처리 장치에 의한 실행을 위하여 또는 그 동작을 제어하기 위하여 유형의 프로그램 매체 상에 인코딩되는 컴퓨터 프로그램 명령에 관한 하나 이상이 모듈로서 구현될 수 있다. 유형의 프로그램 매체는 전파형 신호이거나 컴퓨터로 판독 가능한 매체일 수 있다. 전파형 신호는 컴퓨터에 의한 실행을 위하여 적절한 수신기 장치로 전송하기 위한 정보를 인코딩하기 위하여 생성되는 예컨대 기계가 생성한 전기적, 광학적 또는 전자기 신호와 같은 인공적으로 생성된 신호이다. 컴퓨터로 판독 가능한 매체는 기계로 판독 가능한 저장장치, 기계로 판독 가능한 저장 기판, 메모리 장치, 기계로 판독 가능한 전파형 신호에 영향을 미치는 물질의 조합 또는 이들 중 하나 이상이 조합일 수 있다.

[0061] 부가적으로, 본 특허문헌에서 기술하는 논리 흐름과 구조적인 블록도는 개시된 구조적인 수단의 지원을 받는 대응하는 기능과 단계의 지원을 받는 대응하는 행위 및/또는 특정한 방법을 기술하는 것으로, 대응하는 소프트웨어 구조와 알고리즘과 그 등가물을 구축하는 데에도 사용 가능하다.

[0062] 그리고 본 기술한 설명은 본 발명의 최상의 모드를 제시하고 있으며, 본 발명을 설명하기 위하여, 그리고 당업자가 본 발명을 제작 및 이용할 수 있도록 하기 위한 예를 제공하고 있다. 이렇게 작성된 명세서는 그 제시된 구체적인 용어에 본 발명을 제한하는 것이 아니다.

[0063] 따라서, 상술한 예를 참조하여 본 발명을 상세하게 설명하였지만, 당업자라면 본 발명의 범위를 벗어나지 않으면서도 본 예들에 대한 개조, 변경 및 변형을 가할 수 있다. 요컨대 본 발명이 의도하는 효과를 달성하기 위해 도면에 도시된 모든 기능 블록을 별도로 포함하거나 도면에 도시된 모든 순서를 도시된 순서 그대로 따라야만 하는 것은 아니며, 그렇지 않더라도 얼마든지 청구항에 기재된 본 발명의 기술적 범위에 속할 수 있음에 주의한다.

부호의 설명

[0064] 100 : 딥러닝 학습 기법을 이용하는 유사도 기반 클러스터링 장치

110 : 데이터 전처리부

120 : 자질 표현 생성부

130 : 유사도 계산부

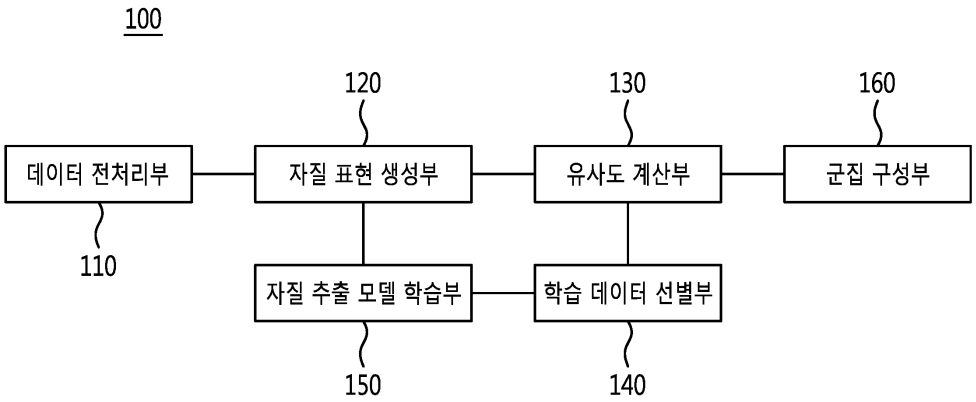
140 : 학습 데이터 선별부

150 : 자질 추출 모델 학습부

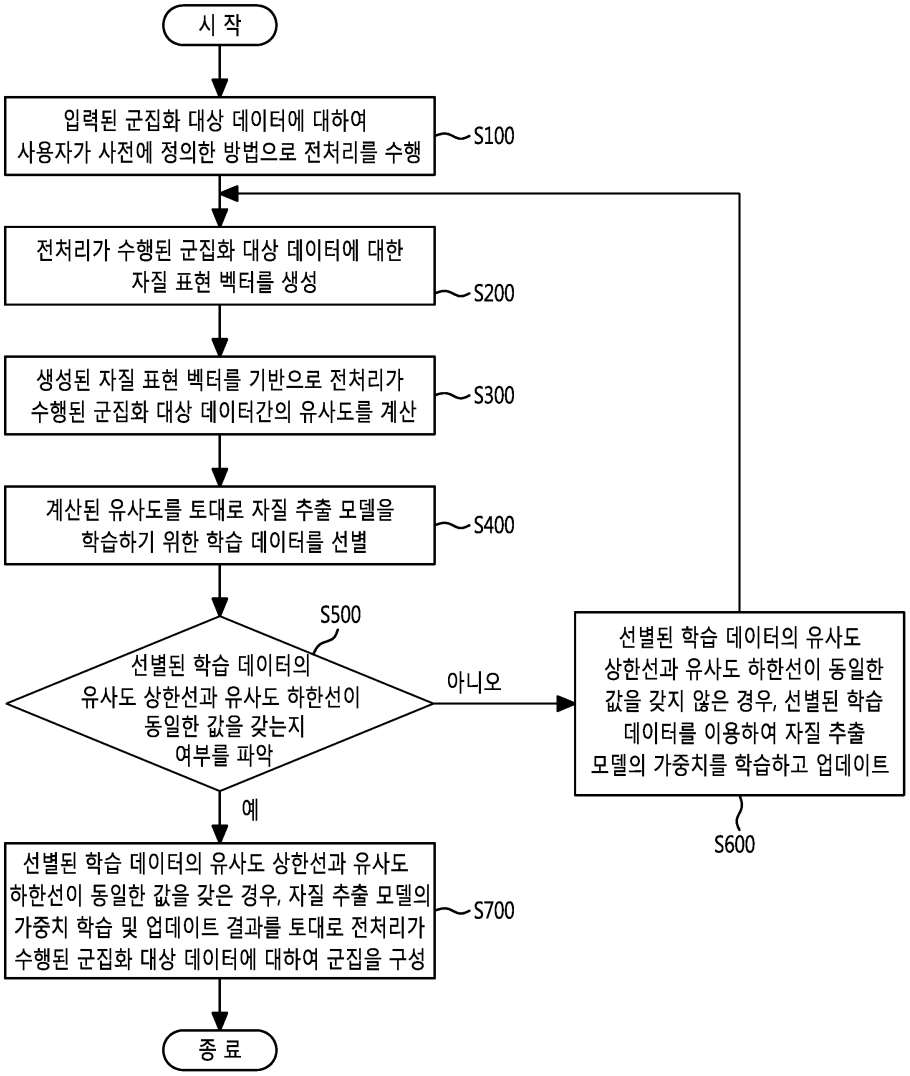
160 : 군집 구성부

도면

도면1



도면2


시 작
입력된 군집화 대상 데이터에 대하여
사용자가 사전에 정의한 방법으로 전처리를 수행
S100
전처리가 수행된 군집화 대상 데이터에 대한
자질 표현 벡터를 생성
S200
생성된 자질 표현 벡터를 기반으로 전처리가
수행된 군집화 대상 데이터간의 유사도를 계산
S300
계산된 유사도를 토대로 자질 추출 모델을
학습하기 위한 학습 데이터를 선별
S400
S500
선별된 학습 데이터의
유사도 상한선과 유사도 하한선이
동일한 값을 갖는지
여부를 파악
아니오
예
선별된 학습 데이터의 유사도
상한선과 유사도 하한선이 동일한
값을 갖지 않은 경우, 선별된 학습
데이터를 이용하여 자질 추출
모델의 가중치를 학습하고 업데이트
S600
선별된 학습 데이터의 유사도 상한선과 유사도
하한선이 동일한 값을 갖은 경우, 자질 추출 모델의
가중치 학습 및 업데이트 결과를 토대로 전처리가
수행된 군집화 대상 데이터에 대하여 군집을 구성
S700
종 료