(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0105854
(43) 공개일자 2022년07월28일

(51) 국제특허분류(Int. Cl.)
G06F 21/62 (2013.01) *G06N 20/00* (2019.01)
(52) CPC특허분류
G06F 21/6245 (2013.01)
G06N 20/00 (2021.08)
(21) 출원번호 10-2021-0008725
(22) 출원일자 2021년01월21일
심사청구일자 2021년01월21일

(71) 출원인
인하대학교 산학협력단
인천광역시 미추홀구 인하로 100(용현동, 인하대학교)
충남대학교산학협력단
대전광역시 유성구 대학로 99 (궁동, 충남대학교)
(72) 발명자
홍성은
인천광역시 미추홀구 인하로 100(용현동, 인하대학교)
조동현
대전광역시 유성구 대학로 99, 충남대학교 공학2호관 407호(궁동)
김영은
서울특별시 서대문구 통일로34길 43(홍제동, 홍제원현대아파트)
(74) 대리인
양성보

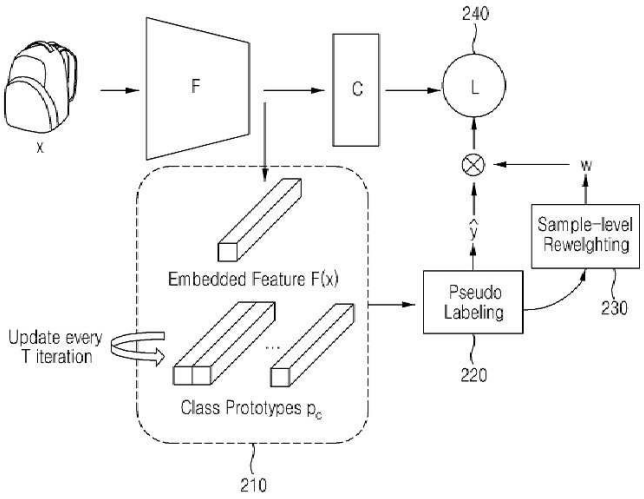
전체 청구항 수 : 총 8 항

(54) 발명의 명칭 **개인 정보 보호 도메인 적응 방법 및 장치**

(57) 요 약

개인 정보 보호 도메인 적응 방법 및 장치가 제시된다. 본 발명에서 제안하는 개인 정보 보호 도메인 적응 방법은 클래스 프로토타입 업데이트부를 통해 각 클래스에 대하여 미리 정해진 자체 엔트로피 값보다 낮은 스몰 자가 엔트로피(small self-entropy)가 있는 샘플을 주기적으로 프로토타입으로 업데이트하는 단계, 유사 레이블 할당부가 각 업데이트 단계에 대하여 타겟 샘플과 클래스 프로토타입 간의 유사성 비교를 통해 가장 유사성이 높은 유사 레이블을 할당하는 단계, 필터링부를 통해 유사 레이블들 중 신뢰도에 따라 샘플을 추가로 필터링하기 위해 샘플 레벨 재가중 방식을 수행하는 단계 및 필터링된 유사 레이블에 기초하여 학습부를 통해 타겟 샘플을 학습하는 단계를 포함한다.

대 표 도 - 도2



이 발명을 지원한 국가연구개발사업

과제고유번호	1711116910
과제번호	2020-0-01389-001
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	인공지능융합선도프로젝트(R&D)
연구과제명	인공지능융합연구센터지원(인하대학교)
기 여 율	1/1
과제수행기관명	인하대학교 산학협력단
연구기간	2020.04.01 ~ 2020.12.31

명 세 서

청구범위

청구항 1

클래스 프로토타입 업데이트부를 통해 각 클래스에 대하여 미리 정해진 자체 엔트로피 값보다 낮은 스몰 자가 엔트로피(small self-entropy)가 있는 샘플을 주기적으로 프로토타입으로 업데이트하는 단계;

유사 레이블 할당부가 각 업데이트 단계에 대하여 타겟 샘플과 클래스 프로토타입 간의 유사성 비교를 통해 가장 유사성이 높은 유사 레이블을 할당하는 단계;

필터링부를 통해 유사 레이블들 중 신뢰도에 따라 샘플을 추가로 필터링하기 위해 샘플 레벨 재가중 방식을 수행하는 단계; 및

필터링된 유사 레이블에 기초하여 학습부를 통해 타겟 샘플을 학습하는 단계

를 포함하는 개인 정보 보호 도메인 적응 방법.

청구항 2

제1항에 있어서,

클래스 프로토타입 업데이트부를 통해 각 클래스에 대하여 미리 정해진 자체 엔트로피 값보다 낮은 스몰 자가 엔트로피(small self-entropy)가 있는 샘플을 주기적으로 프로토타입으로 업데이트하는 단계는,

모든 타겟 샘플을 특성 추출기에 공급함으로써 초기 유사 레이블들을 획득하고, 모든 초기 유사 레이블들 중 클래스 프로토타입에 대해 미리 정해진 자체 엔트로피 값보다 낮은 스몰 자가 엔트로피(small self-entropy)가 있는 샘플에 기초하여 수정된 유사 레이블만을 업데이트하는

개인 정보 보호 도메인 적응 방법.

청구항 3

제2항에 있어서,

미리 정해진 자체 엔트로피 값보다 낮은 스몰 자가 엔트로피가 있는 샘플을 선택하는 방법은 모든 타겟 샘플의 표준화된 자가 엔트로피를 계산하고, 표준화된 자가 엔트로피로부터 각 클래스의 엔트로피 집합을 정의하고, 각 클래스의 엔트로피 집합의 최하위 샘플을 선택하고, 타겟 샘플에 내장된 특성을 평균화하여 최하위 샘플이 포함된 클래스 프로토타입 집합을 정의하는

개인 정보 보호 도메인 적응 방법.

청구항 4

제1항에 있어서,

필터링부를 통해 유사 레이블들 중 신뢰도에 따라 샘플을 추가로 필터링하기 위해 샘플 레벨 재가중 방식을 수행하는 단계는,

각 업데이트 단계에 대하여 타겟 샘플과 클래스 프로토타입 간의 유사성 비교를 통해 가장 유사성이 높은 유사 레이블 및 두 번째로 유사성이 높은 유사 레이블을 찾고, 가장 유사성이 높은 유사 레이블 및 두 번째로 유사성이 높은 유사 레이블을 이용하여 유사 레이블을 세분화하기 위한 샘플 레벨 가중치를 정의하는

개인 정보 보호 도메인 적응 방법.

청구항 5

각 클래스에 대하여 미리 정해진 자체 엔트로피 값보다 낮은 스몰 자가 엔트로피(small self-entropy)가 있는 샘플을 주기적으로 프로토타입으로 업데이트하는 클래스 프로토타입 업데이트부;

각 업데이트 단계에 대하여 타겟 샘플과 클래스 프로토타입 간의 유사성 비교를 통해 가장 유사성이 높은 유사 레이블을 할당하는 유사 레이블 할당부;

유사 레이블들 중 신뢰도에 따라 샘플을 추가로 필터링하기 위해 샘플 레벨 재가중 방식을 수행하는 필터링부; 및

필터링된 유사 레이블에 기초하여 타겟 샘플을 학습하는 학습부

를 포함하는 개인 정보 보호 도메인 적응 장치.

청구항 6

제5항에 있어서,

클래스 프로토타입 업데이트부는,

모든 타겟 샘플을 특성 추출기에 공급함으로써 초기 유사 레이블들을 획득하고, 모든 초기 유사 레이블들 중 클래스 프로토타입에 대해 미리 정해진 자체 엔트로피 값보다 낮은 스몰 자가 엔트로피(small self-entropy)가 있는 샘플에 기초하여 수정된 유사 레이블만을 업데이트하는

개인 정보 보호 도메인 적응 장치.

청구항 7

제6항에 있어서,

클래스 프로토타입 업데이트부는,

미리 정해진 자체 엔트로피 값보다 낮은 스몰 자가 엔트로피가 있는 샘플을 선택하는 방법은 모든 타겟 샘플의 표준화된 자가 엔트로피를 계산하고, 표준화된 자가 엔트로피로부터 각 클래스의 엔트로피 집합을 정의하고, 각 클래스의 엔트로피 집합의 최하위 샘플을 선택하고, 타겟 샘플에 내장된 특성을 평균화하여 최하위 샘플이 포함된 클래스 프로토타입 집합을 정의하는

개인 정보 보호 도메인 적응 장치.

청구항 8

제5항에 있어서,

필터링부는,

각 업데이트 단계에 대하여 타겟 샘플과 클래스 프로토타입 간의 유사성 비교를 통해 가장 유사성이 높은 유사 레이블 및 두 번째로 유사성이 높은 유사 레이블을 찾고, 가장 유사성이 높은 유사 레이블 및 두 번째로 유사성이 높은 유사 레이블을 이용하여 유사 레이블을 세분화하기 위한 샘플 레벨 가중치를 정의하는

개인 정보 보호 도메인 적응 장치.

발명의 설명

기 술 분 야

[0001] 본 발명은 개인 정보 보호 도메인 적응 방법 및 장치에 관한 것이다.

배 경 기 술

[0002] 비감독 도메인 적응(Unsupervised Domain Adaptation; UDA)[1]은 레이블이 부착된 소스 도메인의 특정 사실을 레이블이 부착되지 않은 타겟 도메인으로 전송할 수 있어 많은 관심을 끌었다. UDA의 주요 과제는 학습 단계에서는 타겟 영역의 레이블에 접근할 수 없지만 차별적인 샘플 분포를 생성하는 것이다. 초기 UDA 방법은 도메인 내 공간을 학습하거나 얕은 체제 하에서 인스턴스 수준의 공통성을 추정한다[2]-[7]. 한편, 최근의 방법들은 심층 신경망을 활용하여 이동 가능한 표현을 얻는데, 이것은 두 가지 범주로 나눌 수 있다. 분포 기반 방법[8]-[10], [10], [11]은 소스 도메인과 타겟 도메인의 특성 분포를 직접적으로 일치시키는 반면, 도메인 적대적 방법[1], [12]-[15]은 특성 추출기와 도메인 판별기 사이의 최소-최대 게임을 이용하여 도메인 내 변동을 나타낸

다. 전반적으로, UDA가 레이블이 부착되지 않은 타겟 도메인과 레이블이 부착된 소스 도메인을 성공적으로 정렬할 수 있다는 것을 수많은 연구에서 입증했다[16]-[21].

[0003] 그럼에도 불구하고, 특히 소스 샘플의 레이블에 민감한 속성이 포함되어 있는 경우, 기존 UDA 접근방식은 데이터 개인 정보보호 문제를 야기할 가능성이 높다. 예를 들어, 원본 도메인이 고품질 ID 사진 갤러리인 반면 타겟 도메인은 보안 감시 카메라에서 캡처한 저품질 프로브 세트인 얼굴 도메인 적응 시나리오를 고려해보라[22]. 각 소스 이미지와 결합할 때, 레이블은 개인의 식별자 역할을 하므로, 소스 데이터의 부적절한 액세스는 잘못된 개인정보 보호 위반을 야기할 수 있다[23]-[27]. 이로부터, 개인정보 보호 시나리오에 대한 도메인 적응 과정 중 소스 데이터의 직접적인 이용을 피할 필요가 있다.

[0004] 이러한 개인정보 문제로 인한 문제에서는 소스 샘플에 접근할 수 없으므로, 소스 데이터를 직접 사용하지 않고 소스 도메인과 타겟 도메인을 맞추기 위해서는, 기존의 분포 일치 또는 도메인 적대적 접근법 대신 새로운 접근법이 필요하다.

발명의 내용

해결하려는 과제

[0005] 본 발명이 이루고자 하는 기술적 과제는 기존 도메인 적응방식의 학습 단계에서 소스와 타겟 샘플에 접근할 수 있는 강한 제약을 가지고 있는 잠재적인 데이터 개인정보 문제를 해결하기 위한 개인정보 보호 도메인 적응(Privacy-Preserving Domain Adaptation; PDA) 방법 및 장치를 제공하는데 있다.

과제의 해결 수단

[0006] 일 측면에 있어서, 본 발명에서 제안하는 개인 정보 보호 도메인 적응 방법은 클래스 프로토타입 업데이트부를 통해 각 클래스에 대하여 미리 정해진 자체 엔트로피 값보다 낮은 스몰 자가 엔트로피(small self-entropy)가 있는 샘플을 주기적으로 프로토타입으로 업데이트하는 단계, 유사 레이블 할당부가 각 업데이트 단계에 대하여 타겟 샘플과 클래스 프로토타입 간의 유사성 비교를 통해 가장 유사성이 높은 유사 레이블을 할당하는 단계, 필터링부를 통해 유사 레이블들 중 신뢰도에 따라 샘플을 추가로 필터링하기 위해 샘플 레벨 재가중 방식을 수행하는 단계 및 필터링된 유사 레이블에 기초하여 학습부를 통해 타겟 샘플을 학습하는 단계를 포함한다.

[0007] 클래스 프로토타입 업데이트부를 통해 각 클래스에 대하여 미리 정해진 자체 엔트로피 값보다 낮은 스몰 자가 엔트로피(small self-entropy)가 있는 샘플을 주기적으로 프로토타입으로 업데이트하는 단계는 모든 타겟 샘플을 특성 추출기에 공급함으로써 초기 유사 레이블들을 획득하고, 모든 초기 유사 레이블들 중 클래스 프로토타입에 대해 미리 정해진 자체 엔트로피 값보다 낮은 스몰 자가 엔트로피(small self-entropy)가 있는 샘플에 기초하여 수정된 유사 레이블만을 업데이트한다.

[0008] 미리 정해진 자체 엔트로피 값보다 낮은 스몰 자가 엔트로피가 있는 샘플을 선택하는 방법은 모든 타겟 샘플의 표준화된 자가 엔트로피를 계산하고, 표준화된 자가 엔트로피로부터 각 클래스의 엔트로피 집합을 정의하고, 각 클래스의 엔트로피 집합의 최하위 샘플을 선택하고, 타겟 샘플에 내장된 특성을 평균화하여 최하위 샘플이 포함된 클래스 프로토타입 집합을 정의한다.

[0009] 필터링부를 통해 유사 레이블들 중 신뢰도에 따라 샘플을 추가로 필터링하기 위해 샘플 레벨 재가중 방식을 수행하는 단계는 각 업데이트 단계에 대하여 타겟 샘플과 클래스 프로토타입 간의 유사성 비교를 통해 가장 유사성이 높은 유사 레이블 및 두 번째로 유사성이 높은 유사 레이블을 찾고, 가장 유사성이 높은 유사 레이블 및 두 번째로 유사성이 높은 유사 레이블을 이용하여 유사 레이블을 세분화하기 위한 샘플 레벨 가중치를 정의한다.

[0010] 또 다른 일 측면에 있어서, 본 발명에서 제안하는 개인 정보 보호 도메인 적응장치는 각 클래스에 대하여 미리 정해진 자체 엔트로피 값보다 낮은 스몰 자가 엔트로피(small self-entropy)가 있는 샘플을 주기적으로 프로토타입으로 업데이트하는 클래스 프로토타입 업데이트부, 각 업데이트 단계에 대하여 타겟 샘플과 클래스 프로토타입 간의 유사성 비교를 통해 가장 유사성이 높은 유사 레이블을 할당하는 유사 레이블 할당부, 유사 레이블들 중 신뢰도에 따라 샘플을 추가로 필터링하기 위해 샘플 레벨 재가중 방식을 수행하는 필터링부 및 필터링된 유사 레이블에 기초하여 타겟 샘플을 학습하는 학습부를 포함한다.

발명의 효과

[0011] 본 발명의 실시예들에 따르면 데이터-개인정보 문제를 해결하는, 특히 레이블이 부착된 소스 영역에 대한 개인정보 보호 도메인 적응(PPDA)이라는 UDA에 대해 새로운 패러다임을 도입한다. 제안하는 방법은 소스 도메인에서 사전 학습된 모델을 활용하고, 타겟 모델을 자체 학습 방식으로 업데이트하며 그 결과 소스 데이터 자유 도메인 적응을 가져온다. 제안하는 PPDA 모델은 소스 데이터에 직접 액세스하지 않고도 기존 도메인 적응방식의 학습 단계에서 소스와 타겟 샘플에 접근할 수 있는 강한 제약을 가지고 있는 잠재적인 데이터 개인정보 문제를 해결할 수 있다.

도면의 간단한 설명

[0012] 도 1은 본 발명의 일 실시예에 따른 오피스-홈(Office-Home)의 Ar ⇒ Cl에서 사전 학습된 소스 모델(ResNet-50)로 전달되는 대상 샘플의 정규화된 자체 엔트로피 통계를 시각화한 도면이다.

도 2는 본 발명의 일 실시예에 따른 개인정보 보호 도메인 적응 장치의 구성을 나타내는 도면이다.

도 3은 본 발명의 일 실시예에 따른 개인정보 보호 도메인 적응 방법을 설명하기 위한 흐름도이다.

도 4는 본 발명의 일 실시예에 따른 특성 시각화를 종래기술과 비교한 도면이다.

도 5는 본 발명의 일 실시예에 따른 각 단계에서의 성능 변화를 나타내는 도면이다.

발명을 실시하기 위한 구체적인 내용

[0013] 본 발명은 잠재적인 데이터 개인정보 문제를 해결할 수 있는 새로운 도메인 적응 패러다임을 제시한다. 기존 도메인 적응방식의 장점에도 불구하고, 학습 단계에서 소스와 타겟 샘플에 접근할 수 있는 강한 제약을 가지고 있다. 그러나 소스 샘플의 직접 사용은 특히 소스 도메인의 각 레이블이 생체정보와 같은 개인의 식별자 역할을 하는 경우 데이터 개인정보 문제를 야기할 수 있다. 기존 도메인 적응에서 데이터 개인정보 보호 문제를 해결하기 위해 개인정보 보호 도메인 적응(Privacy-Preserving Domain Adaptation; PDA)을 제안한다. 본 발명의 주요한 가설은 사전 학습된 소스 모델에서 초기화된 제안하는 목표 모델을 스스로 학습하는 방식으로 학습한다면, 레이블이 부착된 소스 도메인에서 레이블이 부착되지 않은 타겟 도메인으로 성공적으로 특정 사실을 이전할 수 있다는 것이다. 예비 연구에서는 사전 학습된 소스 모델에서 측정된 스몰 자가 엔트로피를 가진 타겟 샘플이 충분히 높은 정확도를 달성하는 것을 관찰한다. 이 핵심 관찰로부터 우선 자가 엔트로피를 바탕으로 믿을 수 있는 샘플을 선별하여 클래스 프로토타입으로 정의한다. 이후, 타겟 샘플과 클래스 프로토타입 사이의 유사성을 통해 타겟 샘플에 유사 레이블을 할당한다. 유사 레이블링 프로세스의 불확실성을 더욱 줄이기 위해 샘플 레벨 재가중 방식도 도입한다. 제안하는 PPDA 모델은 소스 데이터에 직접 액세스하지 않더라도 공공 데이터셋에서 기존의 도메인 적응 방법을 능가하는 것을 확인하였다.

[0015] 도 1은 본 발명의 일 실시예에 따른 오피스-홈(Office-Home)의 Ar ⇒ Cl에서 사전 학습된 소스 모델(ResNet-50)로 전달되는 대상 샘플의 정규화된 자체 엔트로피 통계를 시각화한 도면이다.

[0016] 도 1(a)와 같이 자체 엔트로피 값이 낮은 대상 샘플은 올바르게 분류 될 수 있다. 하지만, 도 1(b)와 같이 모든 대상 샘플 중에서 자가 엔트로피가 낮은 샘플은 거의 존재하지 않는다.

[0017] 사전 학습된 소스 모델과 신뢰성의 척도로 자가 엔트로피[28]를 사용하여 타겟 샘플의 신뢰할 수 있는 유사 레이블을 얻으면 자체 학습 방식으로 목표 모델을 학습할 수 있다.

[0018] 본 발명에서의 새로운 접근법의 실현 가능성을 조사하기 위해, 우선 레이블이 부착된 소스 데이터를 사용하여 모델을 학습시킨다. 이후, 사전 학습된 소스 모델을 사용하여 타겟 샘플의 정규화된 자가 엔트로피를 측정한다. 도 1(a)에서 충분히 스몰 자가 엔트로피(예를 들어, 자가 엔트로피가 0.2 이하인 샘플)를 가진 타겟 샘플의 정확도가 상당히 높다는 것을 알 수 있다. 하지만, 그러한 신뢰할 수 있는 타겟 샘플은 도 1(b)에 나타낸 것과 같이 전체 타겟 샘플의 극히 일부에 불과하다.

[0019] 본 발명에서는 개인정보 보호 도메인 적응(Privacy-Preserving Domain Adaptation; PPDA)이라고 불리는 새로운 UDA 패러다임을 제안한다. 제안하는 방법은 자가 엔트로피를 통해 소수의 신뢰할 수 있는 타겟 샘플을 클래스별 프로토타입으로 사용하면 충분히 차별적인 타겟 분포를 만들 수 있다는 것이다. 이하, 본 발명의 실시 예를 첨부된 도면을 참조하여 상세하게 설명한다.

[0021] 도 2는 본 발명의 일 실시예에 따른 개인정보 보호 도메인 적응 장치의 구성을 나타내는 도면이다.

[0022] 제안하는 개인정보 보호 도메인 적응 장치는 클래스 프로토타입 업데이트부(210), 유사 레이블 할당부(220), 필터링부(230) 및 학습부(240)를 포함한다.

[0023] 클래스 프로토타입 업데이트부(210)는 각 클래스에 대하여 미리 정해진 자체 엔트로피 값보다 낮은 스몰 자가 엔트로피(small self-entropy)가 있는 샘플을 주기적으로 프로토타입으로 업데이트한다.

[0024] 클래스 프로토타입 업데이트부(210)는 모든 타겟 샘플을 특성 추출기에 공급함으로써 초기 유사 레이블들을 획득하고, 모든 초기 유사 레이블들 중 클래스 프로토타입에 대해 미리 정해진 자체 엔트로피 값보다 낮은 스몰 자가 엔트로피(small self-entropy)가 있는 샘플에 기초하여 수정된 유사 레이블만을 업데이트한다.

[0025] 클래스 프로토타입 업데이트부(210)는 미리 정해진 자체 엔트로피 값보다 낮은 스몰 자가 엔트로피가 있는 샘플을 선택하는 방법은 먼저 모든 타겟 샘플의 표준화된 자가 엔트로피를 계산한다. 표준화된 자가 엔트로피로부터 각 클래스의 엔트로피 집합을 정의하고, 각 클래스의 엔트로피 집합의 최하위 샘플을 선택한다. 이후, 타겟 샘플에 내장된 특성을 평균화하여 최하위 샘플이 포함된 클래스 프로토타입 집합을 정의한다.

[0026] 유사 레이블 할당부(220)는 각 업데이트 단계에 대하여 타겟 샘플과 클래스 프로토타입 간의 유사성 비교를 통해 가장 유사성이 높은 유사 레이블을 할당한다.

[0027] 필터링부(230)는 유사 레이블들 중 신뢰도에 따라 샘플을 추가로 필터링하기 위해 샘플 레벨 재가중 방식을 수행한다.

[0028] 필터링부(230)는 각 업데이트 단계에 대하여 타겟 샘플과 클래스 프로토타입 간의 유사성 비교를 통해 가장 유사성이 높은 유사 레이블 및 두 번째로 유사성이 높은 유사 레이블을 찾고, 가장 유사성이 높은 유사 레이블 및 두 번째로 유사성이 높은 유사 레이블을 이용하여 유사 레이블을 세분화하기 위한 샘플 레벨 가중치를 정의한다.

[0029] 학습부(240)는 필터링된 유사 레이블에 기초하여 학습부를 통해 타겟 샘플을 학습한다.

[0030] 제안하는 타겟 모델은 사전 학습된 소스 모델에서 출발하지만 클래스 프로토타입에 기초한 유사 레이블을 통해 점진적으로 진화한다. 각 클래스에 대한 스몰 자가 엔트로피(small self-entropy)가 있는 샘플을 주기적으로 프로토타입으로 업데이트하고 각 단계에 대한 타겟 샘플과 프로토타입 간의 유사성 비교를 통해 유사 레이블을 할당한다는 점에 유의한다. 유사 레이블에는 잘못된 레이블이 붙은 샘플을 포함할 수 있으므로 신뢰할 수 있는 샘플을 추가로 필터링하기 위해 샘플 레벨 재가중 방식을 수행한다. 마지막으로 개선된 유사 레이블을 기반으로 타겟 모델을 자체 학습 방식으로 학습하고 이전 과정을 반복한다.

[0031] 기존 UDA와 PPDA의 차이점은 제안된 PPDA가 소스 샘플의 직접 사용으로부터 도메인 적응 프로세스를 분리할 수 있다는 것이며, 이는 데이터 개인정보 보호 문제를 야기할 수 있다는 것이다. 이를 통해 레이블이 부착된 소스 데이터 샘플에서 발생하는 개인정보 문제에 대한 다양한 문제점을 완화할 수 있다. 제안하는 PPDA는 사전 학습된 모델을 통해 소스 도메인의 특정 사실을 타겟 도메인으로 이전한다는 점에서 미세 조정[29], [30]과도 관련이 있다. 미세 조정 기법은 다양한 분야[12], [31], [32]에서 우수한 성능을 보였지만, 일반적으로 타겟 도메인의 레이블 정보가 필요하다. 그러나 타겟 도메인의 레이블링 비용은 시간이 많이 걸릴 뿐만 아니라 비용도 많이 든다.

[0032] 본 발명의 특징은 다음과 같이 요약할 수 있다: (1) 본 발명의 실시예에 따르면 데이터-개인정보 문제를 해결하는, 특히 레이블이 부착된 소스 영역에 대한 개인정보 보호 도메인 적응(PPDA)이라는 UDA에 대해 새로운 패러다임을 도입한다. (2) 제안하는 방법은 소스 도메인에서 사전 학습된 모델을 활용하고, 타겟 모델을 자체 학습 방식으로 업데이트하며 그 결과 소스 데이터 자유 도메인 적응을 가져온다. (3) 제안하는 PPDA 모델은 소스 데이터에 직접 액세스하지 않더라도 공공 데이터셋에서 기존 UDA 방법을 능가한다.

[0034] 도 3은 본 발명의 일 실시예에 따른 개인정보 보호 도메인 적응 방법을 설명하기 위한 흐름도이다.

[0035] 제안하는 개인정보 보호 도메인 적응 방법은 클래스 프로토타입 업데이트부를 통해 각 클래스에 대하여 미리 정해진 자체 엔트로피 값보다 낮은 스몰 자가 엔트로피(small self-entropy)가 있는 샘플을 주기적으로 프로토타

입으로 업데이트하는 단계(310), 유사 레이블 할당부가 각 업데이트 단계에 대하여 타겟 샘플과 클래스 프로토타입 간의 유사성 비교를 통해 가장 유사성이 높은 유사 레이블을 할당하는 단계(320), 필터링부를 통해 유사 레이블들 중 신뢰도에 따라 샘플을 추가로 필터링하기 위해 샘플 레벨 재가중 방식을 수행하는 단계(330) 및 필터링된 유사 레이블에 기초하여 학습부를 통해 타겟 샘플을 학습하는 단계(340)를 포함한다.

[0036] 단계(310)에서, 클래스 프로토타입 업데이트부를 통해 각 클래스에 대하여 미리 정해진 자체 엔트로피 값보다 낮은 스몰 자가 엔트로피(small self-entropy)가 있는 샘플을 주기적으로 프로토타입으로 업데이트한다.

[0037] 클래스 프로토타입 업데이트부는 모든 타겟 샘플을 특성 추출기에 공급함으로써 초기 유사 레이블들을 획득하고, 모든 초기 유사 레이블들 중 클래스 프로토타입에 대해 미리 정해진 자체 엔트로피 값보다 낮은 스몰 자가 엔트로피(small self-entropy)가 있는 샘플에 기초하여 수정된 유사 레이블만을 업데이트한다.

[0038] 클래스 프로토타입 업데이트부는 미리 정해진 자체 엔트로피 값보다 낮은 스몰 자가 엔트로피가 있는 샘플을 선택하는 방법은 먼저 모든 타겟 샘플의 표준화된 자가 엔트로피를 계산한다. 표준화된 자가 엔트로피로부터 각 클래스의 엔트로피 집합을 정의하고, 각 클래스의 엔트로피 집합의 최하위 샘플을 선택한다. 이후, 타겟 샘플에 내장된 특성을 평균화하여 최하위 샘플이 포함된 클래스 프로토타입 집합을 정의한다.

[0039] 단계(320)에서, 유사 레이블 할당부가 각 업데이트 단계에 대하여 타겟 샘플과 클래스 프로토타입 간의 유사성 비교를 통해 가장 유사성이 높은 유사 레이블을 할당한다.

[0040] 단계(330)에서, 필터링부를 통해 유사 레이블들 중 신뢰도에 따라 샘플을 추가로 필터링하기 위해 샘플 레벨 재가중 방식을 수행한다.

[0041] 필터링부는 각 업데이트 단계에 대하여 타겟 샘플과 클래스 프로토타입 간의 유사성 비교를 통해 가장 유사성이 높은 유사 레이블 및 두 번째로 유사성이 높은 유사 레이블을 찾고, 가장 유사성이 높은 유사 레이블 및 두 번째로 유사성이 높은 유사 레이블을 이용하여 유사 레이블을 세분화하기 위한 샘플 레벨 가중치를 정의한다.

[0042] 단계(340)에서, 필터링된 유사 레이블에 기초하여 학습부를 통해 타겟 샘플을 학습한다. 제안하는 본 발명의 일 실시예에 따른 개인정보 보호 도메인 적응 방법 및 장치에 대하여 아래에서 더욱 상세히 설명한다.

[0044] 제안하는 PPDA의 주요 목표는 사전 학습된 소스 모델과 레이블이 부착되지 않은 타겟 샘플 간의 불일치를 최소화하는 것이다. 이를 위한 첫 번째 단계로서, 모든 타겟 샘플을 특성 추출기 F(도 2 참조)에 공급함으로써 초기 유사 레이블을 얻을 수 있다. 그러나 대부분의 타겟 유사 레이블은 소스와 타겟 도메인 사이의 공변량(covariate) 이동[33]에 의한 도메인 격차 때문에 정확하지 않을 수 있다. 따라서 모든 초기 유사 레이블을 사용하는 대신에 클래스 프로토타입이라고 하는 각 클래스에 대해 소수의 신뢰할 수 있는 샘플을 바탕으로 수정된 유사 레이블을 사용한다.

[0045] 사전 학습된 소스 모델에서 측정된 스몰 자가 엔트로피 값을 가진 샘플은 숫자가 작지만 충분히 높은 정확도를 보이는 것을 관찰한다(도 1 참조). NC-ways 분류 시나리오의 경우 먼저 표준화된 자가 엔트로피를 다음과 같이 계산한다. $H(x) = -\frac{1}{\log N_c}\sum l(x)\log(l(x))$, 여기서 $l(x)$는 소프트맥스 함수를 가진 분류기 C(도 2 참조)로부터의 예측 확률을 나타낸다. 모든 타겟 샘플의 $H(x)$로부터 클래스 c의 엔트로피 집합을 정의한다.

[0046] $$E_c = \{H(x)|x \in X_c\}, \qquad (1)$$

[0047] 여기서 X_c는 분류기 C에 의해 클래스 c로 예측된 샘플 집합을 나타낸다. 그 후, 각 클래스 엔트로피 집합 E_c의 K 최하위 샘플을 선택하고, 내장된 특성을 평균하여 클래스 프로토타입을 얻고, 이 샘플을 다음과 같이 정의할 수 있다.

[0048] $$p_c = \frac{1}{K}\sum_{x_i \in B_K} F(x_i). \qquad (2)$$

[0049] 여기서 B_K는 클래스 엔트로피 집합 E_c의 최하위 샘플 K(bottom-K)가 포함된 집합이다. 그런 다음 클래스 프로토타입 집합 $P = \{p_c|c \in C\}$를 정의한다. 점차 진화하는 타겟 모델을 통한 타겟 분포를 반영하기 위해 학습 단계에서 주기적으로 P를 업데이트한다는 점에 유의한다.

[0050] 프로토타입 집합 P가 확보되면, 각 타겟 내장 특성과 클래스 프로토타입 사이의 유사성을 계산하는데, 이 유사성은 다음과 같이 공식화할 수 있다.

[0051] $$s(x, p_c) = \frac{p_c^T F(x)}{\|p_c\|_2 \, \|F(x)\|_2}. \qquad (3)$$

[0052] 그 후, 유사 레이블로서 가장 유사한 클래스를 선택한다. 즉, $\hat{y} = \text{argmax}_c s(x, p_c)$ 이다. 도메인의 불일치에도 불구하고 제안하는 모델은 신뢰할 수 있는 샘플, 즉 클래스 프로토타입에 기반한 유사 레이블링으로 우수한 성능을 달성할 수 있다.

[0053] 신뢰할 수 있는 클래스 프로토타입을 기반으로 유사 레이블을 지정하지만, 레이블이 잘못된 샘플은 여전히 존재한다. 유사 레이블을 더욱 세분화하기 위해 샘플 레벨 재가중 방식을 제안한다. 본 발명에서는 신뢰성 있는 샘플과 가장 유사한 클래스 프로토타입 사이의 거리가 두 번째 유사한 클래스의 거리보다 더 가까워야 한다는 가설을 세운다. 타겟 샘플이 주어지면, 먼저 가장 유사한 클래스 p_{t1}과 두 번째 유사한 클래스 p_{t2}의 클래스 프로토타입을 찾는다. 그런 다음 샘플 레벨 가중치를 다음과 같이 정의한다.

[0054] $$w(x) = \begin{cases} 1, & \text{if } d(F(x), p_{t1}) + m < d(F(x), p_{t2}), \\ 0, & \text{otherwise.} \end{cases} \qquad (4)$$

[0055] 여기서 $d(\alpha, \beta) = 1 - \frac{\alpha \cdot \beta}{\|\alpha\|_2 \|\beta\|_2}$ 이고 m은 마진을 나타낸다. 코사인 유사성은 거리를 [-1, 1]로 제한하므로 L-2 거리와 같은 다른 메트릭스과 비교했을 때 임계값을 쉽게 선택할 수 있다는 점에 유의한다. 결과적으로, 제안하는 재가중 방식은 노이즈 있는 유사 레이블에 의해 야기되는 셀프-바이어스(self-bias)문제를 상당히 완화시킨다.

[0056]
Algorithm 1: Optimization Process.

Input: pre-trained source model with weights of (θ_F, θ_C), unlabeled target dataset $(D_t = \{x^i\}_{i=1}^{N_t})$
Output: updated target networks $\theta = \{\theta_F, \theta_C\}$

begin
for $t \leftarrow 1$ to max_iter **do**
  if t % update_period $== 0$ **then**
    $E_c \leftarrow (F, C, D_t)$ // Eq. (1)
    $P \leftarrow (F, E_c, D_t)$ // Eq. (2)
  end if
  $(\hat{y}, w) \leftarrow (x, F, P)$ // Eq. (3) & Eq. (4)
  $\mathcal{L} \leftarrow (x, \hat{y}, w)$ // Eq. (5)
  $\theta^{(t+1)} \leftarrow \theta^{(t)} - \xi \nabla \mathcal{L}(MiniBatch, \theta^{(t)})$
end for

[0057] 결과적으로, 제안하는 모델은 재가중된 유사 레이블을 사용하여 교차 엔트로피 손실에 의해 학습된다. PPDA의 전체 목표는 다음과 같이 정의할 수 있다.

[0058] $$\mathcal{L}(D_t) = -\mathbb{E}_{x \sim D_t} \sum_{c=1}^{N_c} w(x) \mathbb{1}_{[c=\hat{y}]} log(\sigma(C(F(x)))), \qquad (5)$$

[0059] 여기서 D_t는 타겟 데이터셋을 가리키며, $\sigma(\cdot)$는 소프트맥스 활성화 함수이다. 전체적으로 알고리즘 1은 제안된 PPDA 프레임워크의 전체 절차를 요약한다.

[0060] 본 발명의 실시예에 따른 실험에서, 두 가지 공공 데이터셋, 즉 Office-Home[34]과 VisDA-C[35]에서 제안된 PPDA를 평가한다. Office-Home은 총 15,500개의 이미지, 즉 Artistic(Ar), Clipart(Cl), Product(Pr), Real-World(Rw) 등 총 65개의 범주를 가진 4개의 다른 도메인으로 구성되어 있다. VisDA-C는 합성 대 실제 시나리오를 고려하여 설계되었다. 이 대규모 데이터셋에는 152,397개의 합성(S) 이미지와 55,388개의 실제(R) 이미지가 포함되어 있다. 본 발명의 실시예에서는 모든 데이터셋에 걸친 기존 접근방식과의 공정한 비교를 위해 공식 UDA 프로토콜을 따른다. 기존의 UDA 접근 방식과는 달리, 제안하는 PPDA는 소스 데이터에 직접 접근하기보다는 사전

학습된 모델을 사용한다.

[0061] 본 발명의 실시예에 따르면, ImageNet[37]에서 사전 학습한 ResNet-50 또는 ResNet-101[36]을 기본 특성 추출기로 설정했다. 32의 배치 크기는 모든 실험에 사용되며, Office-Home의 경우 최대 반복 단계를 20,000, VisDA-C의 경우 15,000으로 설정했다. 기본 학습률을 10^{-3}으로 설정했지만, 미세 조정된 계층의 경우 10^{-4}로 설정했다. 가중치 감소는 5×10^{-4}이고 모멘텀이 0.9인 SGD를 옵티마이저로 사용한다.

[0062] 제안된 PPDA 방법을 ResNet-50(소스 전용)[36], DAN[9], DANN[1], MSTN[38], CDAN[39], MCD[11], DSBN[40] 등의 이전 UDA 방법과 비교한다. 초기 UDA 작업(예를 들어, ResNet-50, DAN, DANN, MSTN) 이후 Caffe 구현을 기반으로 성능을 보고하는 반면, 최근 보고된 방법은 PyTorch 구현을 기반으로 한다. 구현 프레임워크의 차이로 인해, 초기 방법의 성과가 최근 작업의 성과에 비해 크게 떨어진다. 공정한 비교를 위해 PyTorch를 이용한 DAN, DANN, MSTN 등 Caffe구현 초기 방법을 재실행하고 있다.

[0063] 표 I에 나타난 바와 같이, 제안하는 PPDA 방식을 Office-Home의 이전 도메인 적응 방법과 비교한다.

[0064] <표 1>

[0065]

Method	Office-Home												
	Ar⇒Cl	Ar⇒Pr	Ar⇒Rw	Cl⇒Ar	Cl⇒Pr	Cl⇒Rw	Pr⇒Ar	Pr⇒Cl	Pr⇒Rw	Rw⇒Ar	Rw⇒Cl	Rw⇒Pr	Avg
ResNet (source only) [34]	42.5	65.8	74.1	57.4	63.7	67.7	55.7	39.1	72.9	66.1	46.3	76.9	60.7
DAN [8]	45.4	65.8	73.9	56.9	61.4	65.9	56.1	43.0	73.2	67.5	50.4	78.5	61.5
DANN [1]	45.7	66.6	73.4	58.0	64.9	68.3	55.9	42.6	74.1	66.5	49.7	77.5	62.0
MSTN [36]	45.8	67.9	74.1	57.9	65.1	68.2	56.7	43.2	74.8	67.0	50.4	79.0	62.5
CDAN [37]	50.7	70.6	76.0	57.6	70.0	70.0	57.4	50.9	77.3	70.9	56.7	81.6	65.8
PPDA w.o. Rw (ours)	40.5	69.5	72.6	62.5	67.3	70.5	61.5	41.6	74.7	66.5	44.5	72.6	62.0
PPDA (ours)	48.5	71.3	75.6	63.9	69.0	72.1	62.4	43.5	76.0	70.4	50.1	76.1	64.9

[0066] 학습 단계에서는 어떤 소스 샘플에도 접근하지 않지만, PPDA는 최첨단 UDA 방식으로 비교 가능한 성과를 달성한다. 또한 제안된 샘플 레벨 재가중의 효과를 입증한다. 표에서 'PPDA'는 최종 모델을 나타내며 'PPDA w.o. Rw'는 샘플 레벨 재가중 방식이 없는 PPDA를 나타낸다. 표 I의 마지막 두 행에서 보듯이, 재가중 방식은 성능을 2.9% 향상시켜 자체 학습 기반 프레임워크에서 필터링 체계의 중요성을 강조한다.

[0067] 표 II는 실제 및 대규모 시나리오를 고려한 VisDA-C의 비교 결과를 제시한다.

[0068] <표 2>

[0069]

Method	VisDA-C												
	aero	bicycle	bus	car	horse	knife	motor	person	plant	skate	train	train	Avg
ResNet (source only) [34]	88.8	56.1	67.0	69.3	92.3	30.3	87.9	53.1	81.9	40.0	82.9	24.8	64.5
DANN [1]	89.6	72.6	72.9	57.9	89.6	51.6	88.0	78.3	85.0	30.5	81.7	37.0	69.6
MSTN [36]	89.7	72.6	75.6	57.4	91.1	46.5	88.9	77.4	85.3	39.2	81.1	37.8	70.2
MCD [10]	87.0	60.9	83.7	64.0	88.9	79.6	84.7	76.9	88.6	40.3	83.0	25.8	71.9
DSBN [38]	94.7	86.7	76.0	72.0	95.2	75.1	87.9	81.3	91.1	68.9	88.3	45.5	80.2
PPDA w.o. Rw (ours)	82.8	74.9	50.6	35.8	87.8	84.1	72.3	57.1	72.9	47.4	71.1	36.5	64.4
PPDA (ours)	81.5	79.4	80.3	61.8	92.3	91.9	84.5	82.7	86.5	58.4	74.2	43.5	76.4

[0070] 표에서 PPDA는 소스 데이터셋에서 학습된 ResNet 기준에 비해 11.9%의 상당한 성능 향상을 보여 준다. 흥미롭게도, PPDA는 칼, 사람, 열차와 같은 정밀도가 낮은 클래스에 대해 큰 개선을 보여주는데, 이는 낮은 엔트로피를 가진 신뢰할 수 있는 샘플을 선택하는 것은, 이러한 어려운 클래스에서도 효과가 있음을 암시한다.

[0072] 도 4는 본 발명의 일 실시예에 따른 특성 시각화를 종래기술과 비교한 도면이다.

[0073] 도 4는 VisDA-C 데이터셋에서 ResNet-50(도 4(a))과 PPDA(도 4(b)) 간의 특성 공간의 t-SNE를 비교하는 도면이다.

[0074] 제안된 방법의 효과를 시각화하기 위해 VisDA-C 데이터셋에서 ResNet-50과 PPDA 간의 특성 공간[41]의 t-SNE를 비교한다. 도 4와 같이 소스 도메인과 타겟 도메인이 PPDA에 의해 더 잘 정렬되어 있음을 명확히 관찰할 수 있다. 이 결과는 이전에 다루지 않았던 소스 데이터에 직접 액세스하지 않고도 도메인 불일치를 줄일 수 있는 가능성을 보여준다.

[0076] 도 5는 본 발명의 일 실시예에 따른 각 단계에서의 성능 변화를 나타내는 도면이다.

[0077] 도 5(a)는 프로토 타입 업데이트 기간에 대한 성능 변화, 도 5(b)는 프로토 타입 선택을 위한 하이퍼 파라미터

K에 대한 성능 변화, 도 5(c)는 샘플 레벨 재가중의 마진 m에 대한 성능 변화를 나타내는 도면이다.

[0078] 본 발명의 실시예에 따르면, 점진적으로 업데이트된 모델에서 보다 신뢰할 수 있는 프로토타입을 얻기 위해 정기적으로 클래스 프로토타입을 업데이트한다. 도 5(a)는 서로 다른 업데이트 기간을 가진 제안하는 방법의 정확성을 보여준다. 비교적 짧은 기간이 고성능으로 이어지지만, 많은 계산 비용이 필요하다. 반대로 업데이트 기간을 늘리면 계산 복잡성은 감소하지만 성능은 저하된다. 경험적으로, Office-home의 경우 100, VisDA-C의 경우 400으로 업데이트 기간을 설정했다.

[0079] 각 클래스에 대해 낮은 엔트로피 샘플 K(식 2)의 평균을 구한다. 도 5(b)에서는 K의 민감도를 분석한다. K에 대해 많은 수의 샘플을 선택하면 다수의 샘플이 노이즈 있는 샘플을 포함할 가능성이 높기 때문에 성능이 저하된다. 그럼에도 불구하고, 제안하는 방법은 평균 샘플 수와 관련하여 견실한 성능을 보여준다. 이는 낮은 엔트로피 값을 가진 샘플에서 프로토타입을 얻을 경우, 제안하는 PPDA 프레임워크는 평균 샘플의 수에 크게 영향을 받지 않음을 의미한다. 경험적으로 모든 데이터셋에서 K를 3으로 설정했다.

[0080] 본 발명의 실시예에 따른 샘플 레벨 재가중은 표 I과 표 II에 나타난 불완전한 유사 레이블의 불확실성을 효과적으로 다룬다. 제안하는 재가중 방식을 보다 자세히 조사하기 위해, 도 5(c)에 나타낸 것과 같이 마진 m 값(식 4)에 따라 성능 변화를 평가한다. 마진이 높을수록 필터링의 임계값이 높아지므로 제안하는 모델은 학습을 위해 유효한 샘플을 거의 활용할 수 없다. 따라서 마진이 클수록 제안하는 PPDA의 정확도는 낮아진다. 모든 데이터셋에서 마진을 0.1로 설정했다. 일반성의 관점에서, 제안하는 재가중 방식은 유사 레이블을 이용하는 다른 자체 학습 방법에 적용될 수 있다.

[0081] 본 발명의 실시예에 따른 Office-Home에서 Ar⇒Cl에 대한 실험을 한다. 제안하는 PPDA를 DAN과 DANN을 포함한 종래기술들과 비교한다. 이 모든 방법들(예를 들어, PPDA, DAN, DANN)은 네트워크를 2만번 반복해서 학습시키기 때문에 단지 하나의 반복에 대한 계산 시간만을 비교한다. 표 III에서 PPDA는 메모리 업데이트를 고려하지 않고 1회 반복 계산 시간이 훨씬 짧다는 것을 알 수 있다.

[0082] <표 3>

[0083]

Method	1 iteration(s)	Memory update(s)	Total time(s)
DAN [8]	0.4506	-	0.4506
DANN [1]	0.4625	-	0.4625
PPDA (ours)	0.1923	30.20	0.4943 (= $0.1923+\frac{30.20}{100}$)

[0084] 제안된 샘플 레벨 재가중(식 4)은 적은 수의 매트릭스 승수와 임계값 함수로 구현될 수 있기 때문에 병렬 컴퓨팅의 장점을 활용할 수 있다. 한편, DAN은 GPU에서는 구현하기 어려운 복수의 연산을 하고 있으며, DANN은 판별자 아키텍처를 위한 추가 연산을 요구한다. 그러나 PPDA는 약 30초가 필요한 100회의 반복마다 메모리를 업데이트한다. 전체적으로 세 가지 방법에 대한 총 시간을 계산하고, 그 결과를 보면 본 발명의 실시예에 따른 PPDA는 다른 방법과 계산 비용이 비슷하다는 것을 알 수 있다.

[0086] 이상에서 설명된 장치는 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성요소는, 예를 들어, 프로세서, 콘트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPA(field programmable array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

[0087] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로

(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치에 구체화(embody)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

[0088] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 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 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.

[0089] 이상과 같이 실시예들이 비록 한정된 실시예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

[0090] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

[0092] <참조문헌>

[0093] [1] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in Proc. Int. Conf. Mach. Learn., 2015, pp. 1180-1189.

[0094] [2] J. Huang, A. Gretton, K. Borgwardt, B. Scholkopf, and A. J. Smola, "Correcting sample selection bias by unlabeled data," in Proc. Neural Inf. Process. Syst., 2007, pp. 601-608.

[0095] [3] H. Daume III, "Frustratingly easy domain adaptation," in Proc. Annu. Meeting Assoc. Comput. Linguistics, 2017, pp. 256-263.

[0096] [4] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," IEEE Trans. Neural Netw., vol. 22, no. 2, pp. 199-210, Feb. 2011.

[0097] [5] B. Gong, K. Grauman, and F. Sha, "Connecting the dots with landmarks: Discriminatively learning domain-invariant features for unsupervised domain adaptation," in Proc. Int. Conf. Mach. Learn., 2013, pp. 222-230.

[0098] [6] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars, "Unsupervised visual domain adaptation using subspace alignment," in Proc. Int. Conf. Comput. Vis., 2013, pp. 2960-2967.

[0099] [7] J. Hoffman, E. Rodner, J. Donahue, T. Darrell, and K. Saenko, "Efficient learning of domain-invariant image representations," in Proc. Int. Conf. Learn. Rep., 2013.

[0100] [8] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, "Deep domain confusion: Maximizing for domain invariance," 2014, arXiv:1412.3474.

[0101] [9] M. Long, Y. Cao, J.Wang, and M. I. Jordan, "Learning transferable features with deep adaptation networks," in Proc. Int. Conf. Mach. Learn., 2015, pp. 97-105.

[0102] [10] W. Zellinger, T. Grubinger, E. Lughofer, T. Natschläger, and S. Saminger-Platz, "Central moment

discrepancy (CMD) for domaininvariant representation learning," in Proc. Int. Conf. Learn. Rep., 2017, arXiv:1702.08811.

[0103] [11] K. Saito, K. Watanabe, Y. Ushiku, and T. Harada, "Maximum classifier discrepancy for unsupervised domain adaptation," in Proc. Comput. Vis. Pattern Recognit., 2018, pp. 3723-3732.

[0104] [12] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in Proc. Comput. Vis. Pattern Recognit., 2017, pp. 7167-7176.

[0105] [13] K. Bousmalis, N. Silberman, D. Dohan, D. Erhan, and D. Krishnan, "Unsupervised pixel-level domain adaptation with generative adversarial networks," in Proc. Comput. Vis. Pattern Recognit., 2017, pp. 3722-3731.

[0106] [14] J. Hoffman et al., "CYCADA: Cycle-consistent adversarial domain adaptation," in Proc. Int. Conf. Mach. Learn., 2018, pp. 1989-1998.

[0107] [15] S. Hong and J. Ryu, "Attention-guided adaptation factors for unsupervised facial domain adaptation," Electron. Lett., vol. 56, no. 16, pp. 816-818, Aug. 2020.

[0108] [16] Y.-H. Tsai, W.-C. Hung, S. Schulter, K. Sohn, M.-H. Yang, and M. Chandraker, "Learning to adapt structured output space for semantic segmentation," in Proc. Comput. Vis. Pattern Recognit., 2018, pp. 7472-7481.

[0109] [17] S. Hong and J. Ryu, "Unsupervised face domain transfer for low-resolution face recognition," IEEE Signal Process. Lett., vol. 27, pp. 156-160, 2020.

[0110] [18] S. Choi, S. Lee, Y. Kim, T. Kim, and C. Kim, "Hi-CMD: Hierarchical crossmodality disentanglement for visible-infrared person re-identification," in Proc. Comput. Vis. Pattern Recognit., 2020, pp. 10 257-10 266.

[0111] [19] Y. Kim, S. Choi, T. Kim, S. Lee, and C. Kim, "Learning to align multicamera domains using part-aware clustering for unsupervised video person re-identification," 2019, arXiv:1909.13248.

[0112] [20] J. Ryu, G. Kwon, M.-H. Yang, and J. Lim, "Generalized convolutional forest networks for domain generalization and visual recognition," in Proc. Int. Conf. Learn. Rep., 2019.

[0113] [21] Y. Kim, S. Hong, S. Yang, S. Kang, Y. Jeon, and J. Kim, "Associative partial domain adaptation," 2020, arXiv:2008.03111.

[0114] [22] S. Hong, W. Im, J. Ryu, and H. S. Yang, "SPP-DAN: Deep domain adaptation network for face recognition with single sample per person," in Proc. IEEE Int. Conf. Image Process., 2017, pp. 825-829.

[0115] [23] V. Smith, C.-K. Chiang, M. Sanjabi, and A. S. Talwalkar, "Federated multitask learning," in Proc. Neural Inf. Process. Syst., 2017, pp. 4424-4434.

[0116] [24] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," Trans. Intell. Syst. Technol., vol. 10, no. 2, pp. 1-19, 2019.

[0117] [25] S. J. Oh, R. Benenson, M. Fritz, and B. Schiele, "Faceless person recognition: Privacy implications in social media," in Proc. Eur. Conf. Comput. Vis., 2016, pp. 19-35.

[0118] [26] D. Reich et al., "Privacy-preserving classification of personal text messages with secure multi-party computation," in Proc. Neural Inf. Process. Syst., 2019, pp. 3757-3769.

[0119] [27] J. Chen, X. Kang, Y. Liu, and Z. J. Wang, "Median filtering forensics based on convolutional neural networks," IEEE Signal Process. Lett., vol. 22, no. 11, pp. 1849-1853, Nov. 2015.

[0120] [28] J. Lin, "Divergence measures based on the shannon entropy," IEEE Trans. Inf. Theory, vol. 37, no. 1, pp. 145-151, Jan. 1991.

[0121] [29] Y. Guo, H. Shi, A. Kumar, K. Grauman, T. Rosing, and R. Feris, "Spottune: Transfer learning

through adaptive fine-tuning," in Proc. Comput. Vis. Pattern Recognit., 2019, pp. 4805-4814.

[0122] [30] M. Long, H. Zhu, J. Wang, and M. I. Jordan, "Deep transfer learning with joint adaptation networks," in Proc. Int. Conf. Mach. Learn., 2017, pp. 2208-2217.

[0123] [31] A. Karpathy and L. Fei-Fei, "Deep visual-semantic alignments for generating image descriptions," in Proc. Comput. Vis. Pattern Recognit., 2015, pp. 3128-3137.

[0124] [32] S. Hong, J. Ryu, W. Im, and H. S. Yang, "D3: Recognizing dynamic scenes with deep dual descriptor based on key frames and key segments," Neurocomputing, vol. 273, pp. 611-621, 2018.

[0125] [33] M. Sugiyama and A. J. Storkey, "Mixture regression for covariate shift," in Proc. Neural Inf. Process. Syst., 2007, pp. 1337-1344.

[0126] [34] H. Venkateswara, J. Eusebio, S. Chakraborty, and S. Panchanathan, "Deep hashing network for unsupervised domain adaptation," in Proc. Comput. Vis. Pattern Recognit., 2017, pp. 5018-5027.

[0127] [35] X. Peng, B. Usman, N. Kaushik, J. Hoffman, D. Wang, and K. Saenko, "VISDA: The visual domain adaptation challenge," 2017, arXiv:1710.06924.

[0128] [36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. Comput. Vis. Pattern Recognit., 2016, pp. 770-778.

[0129] [37] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in Proc. IEEE Comput. Vis. Pattern Recognit., 2009, pp. 248-255.

[0130] [38] S. Xie, Z. Zheng, L. Chen, and C. Chen, "Learning semantic representations for unsupervised domain adaptation," in Proc. Int. Conf. Mach. Learn., 2018, pp. 5423-5432.

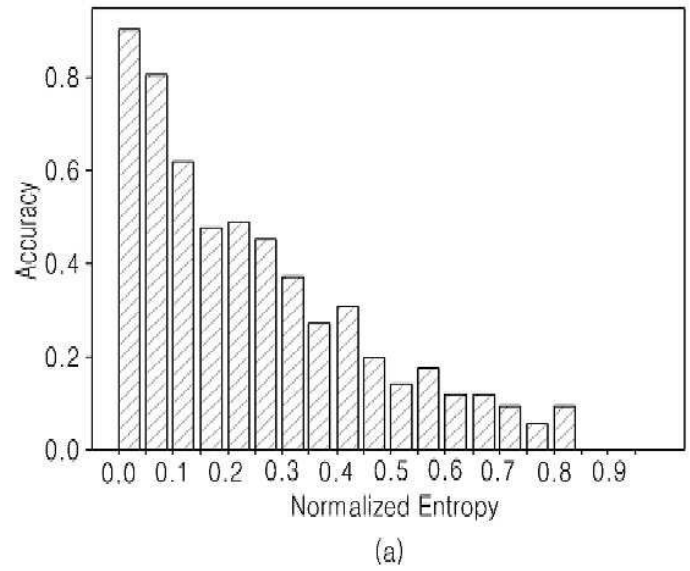
[0131] [39] M. Long, Z. Cao, J. Wang, and M. I. Jordan, "Conditional adversarial domain adaptation," in Proc. Adv. Neural Inf. Process. Syst., 2018, pp. 1640-1650.

[0132] [40] W.-G. Chang, T. You, S. Seo, S. Kwak, and B. Han, "Domain-specific batch normalization for unsupervised domain adaptation," in Proc. Comput. Vis. Pattern Recognit., 2019, pp. 7354-7362.

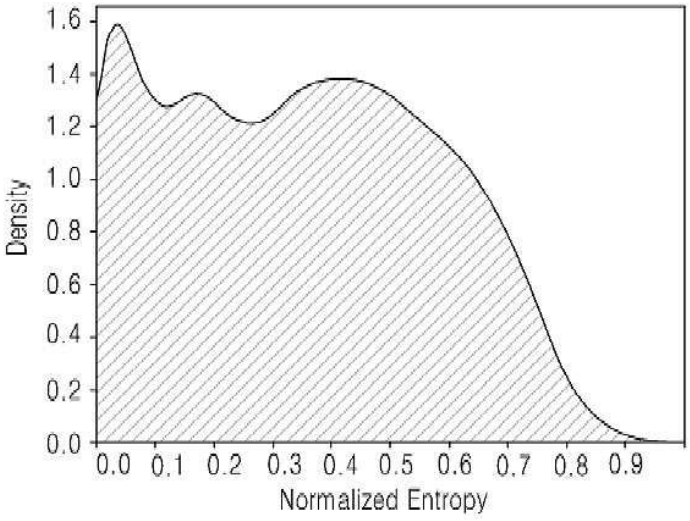
[0133] [41] L. V. D. Maaten and G. Hinton, "Visualizing data using T-SNE," J.Mach. Learn. Res., vol. 9, pp. 2579-2605, 2008.

도면

도면1

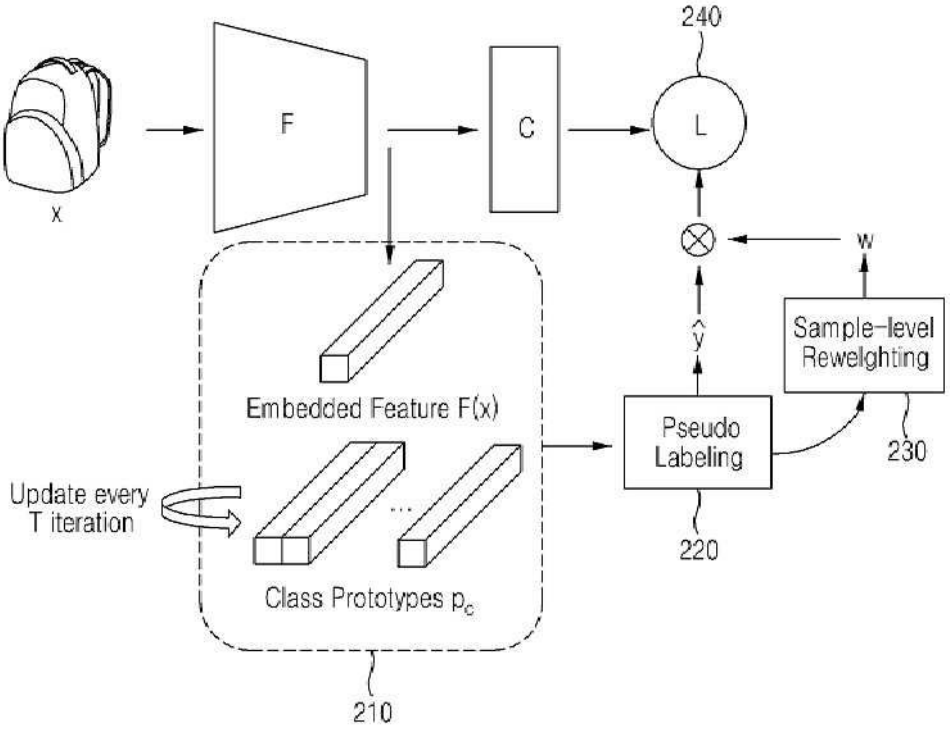

Accuracy
0.0
0.2
0.4
0.6
0.8
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
Normalized Entropy

(a)

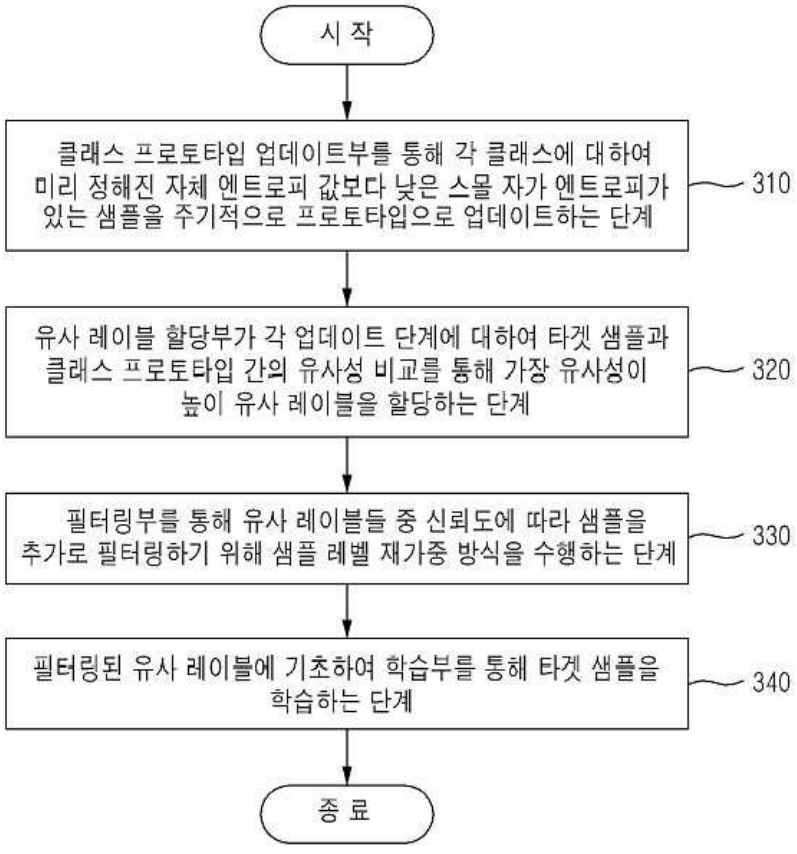

Density
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
Normalized Entropy

(b)

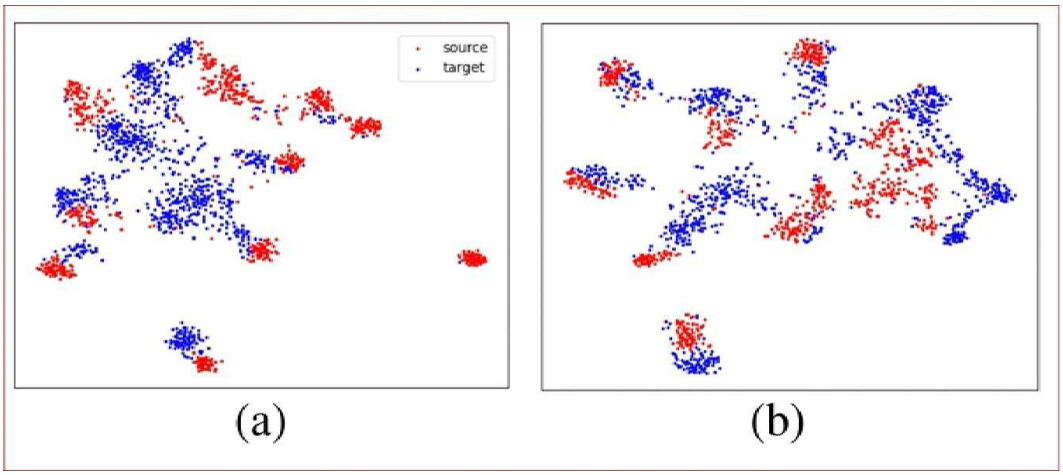
도면2


240
F
C
L
x
w
ŷ
Sample-level Rewelghting
Embedded Feature F(x)
Pseudo Labeling
230
Update every T iteration
220
Class Prototypes p_c
210

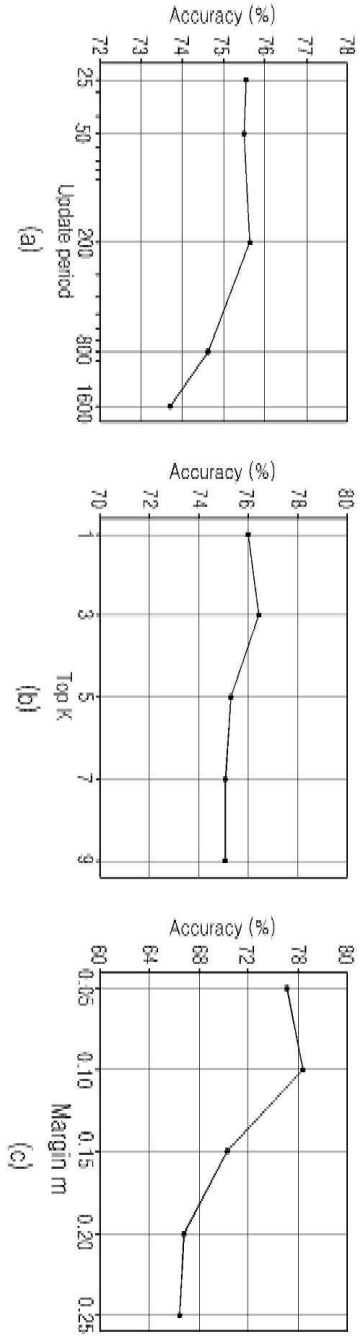
도면3


시 작
클래스 프로토타입 업데이트부를 통해 각 클래스에 대하여 미리 정해진 자체 엔트로피 값보다 낮은 스몰 자가 엔트로피가 있는 샘플을 주기적으로 프로토타입으로 업데이트하는 단계
310
유사 레이블 할당부가 각 업데이트 단계에 대하여 타겟 샘플과 클래스 프로토타입 간의 유사성 비교를 통해 가장 유사성이 높이 유사 레이블을 할당하는 단계
320
필터링부를 통해 유사 레이블들 중 신뢰도에 따라 샘플을 추가로 필터링하기 위해 샘플 레벨 재가중 방식을 수행하는 단계
330
필터링된 유사 레이블에 기초하여 학습부를 통해 타겟 샘플을 학습하는 단계
340
종 료

도면4



도면5


Accuracy (%)
72
73
74
75
76
77
78
25
50
200
800
1600
Update period
(a)
Accuracy (%)
70
72
74
76
78
80
1
3
5
7
9
Top K
(b)
Accuracy (%)
60
64
68
72
78
80
0.05
0.10
0.15
0.20
0.25
Margin m
(c)