



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2019-0095847
(43) 공개일자 2019년08월16일

(51) 국제특허분류(Int. Cl.)
G06N 99/00 (2019.01) G06K 9/62 (2006.01)
(52) CPC특허분류
G06N 20/00 (2019.01)
G06K 9/6267 (2013.01)
(21) 출원번호 10-2018-0015341
(22) 출원일자 2018년02월07일
심사청구일자 2018년02월07일

(71) 출원인
계명대학교 산학협력단
대구광역시 달서구 달구벌대로 1095, 계명대학교
산학협력관 201호(신당동)
(72) 발명자
고병철
대구광역시 수성구 범어천로 107 (범어동, 범어
STX KAN) 103동 701호
남재열
대구광역시 수성구 수성로 412 수성보성타운 107
동 503호
(뒷면에 계속)
(74) 대리인
김건우

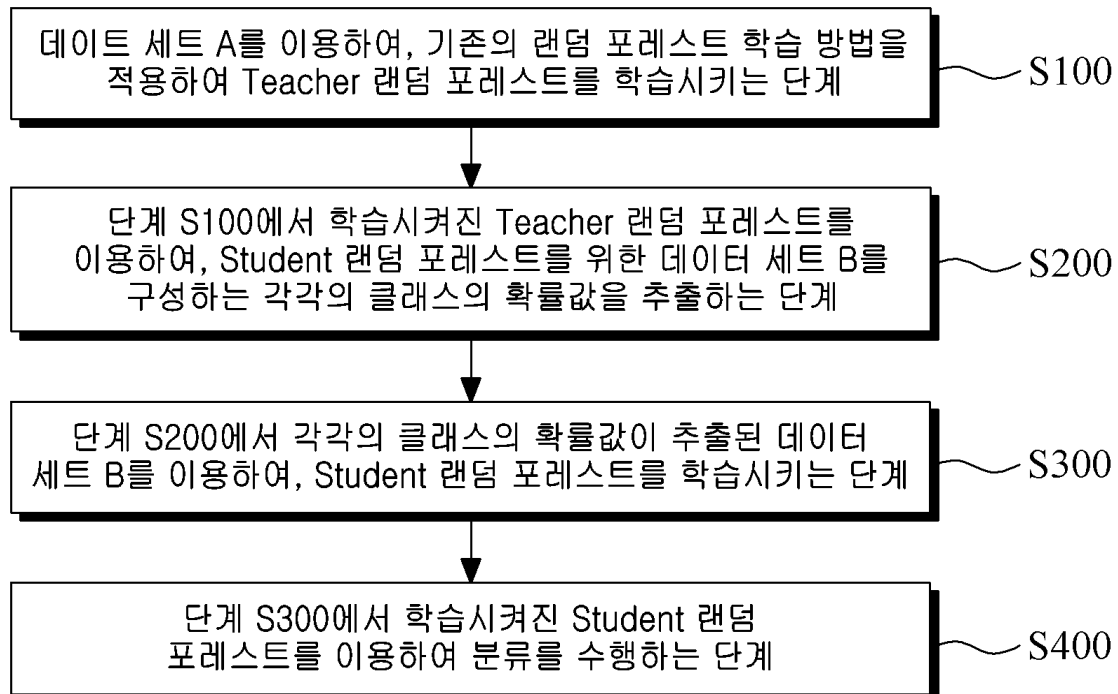
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 소프트웨어 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법 및 이를 이용한 분류기

(57) 요약

본 발명은 소프트웨어 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법에 관한 것으로서, 보다 구체적으로는 랜덤 포레스트 분류 방법으로서, (1) 데이터 세트 A를 이용하여, 기존의 랜덤 포레스트 학습 방법을 적용하여 Teacher 랜덤 포레스트를 학습시키는 단계; (2) 상기 단계 (1)에서 학습시켜진 Teacher 랜덤 포레스트를 이 (뒷면에 계속)

대표도 - 도2



용하여, Student 랜덤 포레스트를 위한 데이터 세트 B를 구성하는 각각의 클래스의 확률값을 추출하는 단계; (3) 상기 단계 (2)에서 각각의 클래스의 확률값이 추출된 데이터 세트 B를 이용하여, Student 랜덤 포레스트를 학습시키는 단계; 및 (4) 상기 단계 (3)에서 학습시켜진 Student 랜덤 포레스트를 이용하여 분류를 수행하는 단계를 포함하는 것을 그 구성상의 특징으로 한다.

또한, 본 발명은 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법을 이용한 분류기에 관한 것으로서, 보다 구체적으로는 랜덤 포레스트 분류 방법을 이용한 분류기로서, (1) 데이터 세트 A를 이용하여, 기존의 랜덤 포레스트 학습 방법을 적용하여 Teacher 랜덤 포레스트를 학습시키는 Teacher 랜덤 포레스트 학습 모듈; (2) 상기 Teacher 랜덤 포레스트 학습 모듈에서 학습시켜진 Teacher 랜덤 포레스트를 이용하여, Student 랜덤 포레스트를 위한 데이터 세트 B를 구성하는 각각의 클래스의 확률값을 추출하는 클래스 확률값 추출 모듈; (3) 상기 클래스 확률값 추출 모듈에서 각각의 클래스의 확률값이 추출된 데이터 세트 B를 이용하여, Student 랜덤 포레스트를 학습시키는 Student 랜덤 포레스트 학습 모듈; 및 (4) 상기 Student 랜덤 포레스트 학습 모듈에서 학습시켜진 Student 랜덤 포레스트를 이용하여 분류를 수행하는 분류 모듈을 포함하는 것을 그 구성상의 특징으로 한다.

본 발명에서 제안하고 있는 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법 및 이를 이용한 분류기에 따르면, 기존의 랜덤 포레스트 학습 방법을 이용하여 Teacher 랜덤 포레스트를 학습시키고, 이렇게 학습시켜진 Teacher 랜덤 포레스트를 이용하여 Student 랜덤 포레스트를 위한 데이터 세트를 구성하는 각각의 클래스의 확률값을 추출한 후, 이렇게 각각의 클래스의 확률값이 추출된 데이터 세트를 이용하여 Student 랜덤 포레스트를 학습시킴으로써, 랜덤 포레스트의 성능은 유지하면서도 랜덤 포레스트의 트리의 수를 줄여 처리 시간 및 메모리량을 대폭 줄일 수 있다.

(72) 발명자

정미라

대구광역시 달서구 호산로 125 삼성명가타운 220동 1207호

김상준

대구광역시 서구 문화로 156 B동 101호 (평리동, 목련빌라)

이 발명을 지원한 국가연구개발사업

과제고유번호 B20170130

부처명 한국전자통신연구원

연구관리전문기관 한국전자통신연구원

연구사업명 국내위탁연구

연구과제명 심층 신경망을 이용한 밀집/비밀집 복합환경에서의 피플 카운팅 및 추적 기술 연구

기 여 율 1/1

주관기관 계명대학교 산학협력

연구기간 2017.06.01 ~ 2017.11.30

명세서

청구범위

청구항 1

랜덤 포레스트 분류 방법으로서,

- (1) 데이터 세트 A를 이용하여, 기존의 랜덤 포레스트 학습 방법을 적용하여 Teacher 랜덤 포레스트를 학습시키는 단계;
- (2) 상기 단계 (1)에서 학습시켜진 Teacher 랜덤 포레스트를 이용하여, Student 랜덤 포레스트를 위한 데이터 세트 B를 구성하는 각각의 클래스의 확률값을 추출하는 단계;
- (3) 상기 단계 (2)에서 각각의 클래스의 확률값이 추출된 데이터 세트 B를 이용하여, Student 랜덤 포레스트를 학습시키는 단계; 및
- (4) 상기 단계 (3)에서 학습시켜진 Student 랜덤 포레스트를 이용하여 분류를 수행하는 단계를 포함하는 것을 특징으로 하는, 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법.

청구항 2

제1항에 있어서, 상기 Student 랜덤 포레스트는,

상기 Teacher 랜덤 포레스트의 트리의 수보다 작은 값의 트리의 수를 갖는 것을 특징으로 하는, 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법.

청구항 3

제1항에 있어서, 상기 단계 (1)에서,

상기 데이터 세트 A는 클래스 레이블이 포함되는 것을 특징으로 하는, 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법.

청구항 4

제1항에 있어서, 상기 단계 (2)에서,

상기 데이터 세트 B는 클래스 레이블이 포함되지 않는 것을 특징으로 하는, 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법.

청구항 5

제1항에 있어서, 상기 단계 (3)은,

- (3-1) 상기 단계 (2)에서 각각의 클래스의 확률값이 추출된 데이터 세트 B에서, 임의의 데이터를 선택하여 서브 세트를 구성하는 단계;
- (3-2) 상기 단계 (3-1)에서 구성된 서브 세트에 대하여, 각각의 내부 노드는 특징 벡터 x 에서 파라미터 θ_i 를 무작위로 선택하고, 그 중에서 가장 높은 성능을 갖는 분할 함수 f 와 θ 를 결정한 후, $f(\theta, x)$ 를 사용하여 트리를 $left(S_l)$, $right(S_r)$ 로 분할하는 과정을, 최대 허용 깊이까지 반복하여 트리를 생성하는 단계; 및
- (3-3) 상기 단계 (3-2)에서 생성된 트리에 대하여, 평가 기준을 만족하는 트리를 찾는 과정을, 트리의 개수까지

반복하여 수행하는 단계를 포함하는 것을 특징으로 하는, 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법.

청구항 6

제5항에 있어서, 상기 단계 (3-2)에서,

성능 평가 기준은, 상기 Teacher 랜덤 포레스트에 의해 추출된 클래스 확률 값을 이용하여 계산되는, 분할 함수 f에 대한 엔트로피 값을 이용하는 것을 특징으로 하는, 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법.

청구항 7

제6항에 있어서, 상기 단계 (3-2)에서,

상기 성능 평가 기준은, 다음 수학적식을 이용하는 것을 특징으로 하는, 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법.

$$\Delta E = \frac{|S_l|}{|S_n|} E(S_l) + \frac{|S_r|}{|S_n|} E(S_r)$$

상기 수학적식에서, ΔE 는 엔트로피의 합, S_n 은 n번째 노드에 도달하는 데이터의 집합, S_l , S_r 은 각각 n번째 노드에서 왼쪽 혹은 오른쪽의 자식 노드로 들어가는 데이터의 집합, $|\cdot|$ 은 데이터 집합에 속한 데이터의 개수를 나타내며, $E(S)$ 는 다음 수학적식으로 계산됨.

$$E(S) = -\frac{1}{N} \sum_{i=1}^N P(C_i) \log(P(C_i))$$

상기 수학적식에서, $P(C_i)$ 는 각각의 클래스에 대한 확률 질량 함수를 나타냄.

청구항 8

제5항에 있어서, 상기 단계 (3-3)에서,

상기 성능 평가 기준은, 상기 Teacher 랜덤 포레스트와 상기 Student 랜덤 포레스트에 의해 각각 추출된 클래스 확률값의 교차-엔트로피(cross-entropy)를 이용하는 것을 특징으로 하는, 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법.

청구항 9

제8항에 있어서, 상기 단계 (3-3)에서,

상기 성능 평가 기준은, 다음 수학적식을 이용하는 것을 특징으로 하는, 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법.

$$cross-entropy = -\sum_{n=1}^N \sum_{i=1}^C P_T(C_i|x(n)) \cdot \log P_S(C_i|x(n))$$

상기 수학적식에서, $P_T(C_i|x(n))$, $P_S(C_i|x(n))$ 는 각각 상기 Teacher 랜덤 포레스트와 상기 Student 랜덤 포레스트에 대한, 특징 벡터 $x(n)$ 에서의 각각의 클래스에 대한 확률 질량 함수를 나타냄.

청구항 10

제9항에 있어서, 상기 단계 (3-3)에서,

상기 계산된 교차-엔트로피(cross-entropy)가 미리 지정된 임계값보다 작을 경우 해당 트리를 사용하되, 그렇지 않을 경우 트리를 재생성하는 것을 특징으로 하는, 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법.

청구항 11

랜덤 포레스트 분류 방법을 이용한 분류기로서,

(1) 데이터 세트 A를 이용하여, 기존의 랜덤 포레스트 학습 방법을 적용하여 Teacher 랜덤 포레스트를 학습시키는 Teacher 랜덤 포레스트 학습 모듈;

(2) 상기 Teacher 랜덤 포레스트 학습 모듈에서 학습시켜진 Teacher 랜덤 포레스트를 이용하여, Student 랜덤 포레스트를 위한 데이터 세트 B를 구성하는 각각의 클래스의 확률값을 추출하는 클래스 확률값 추출 모듈;

(3) 상기 클래스 확률값 추출 모듈에서 각각의 클래스의 확률값이 추출된 데이터 세트 B를 이용하여, Student 랜덤 포레스트를 학습시키는 Student 랜덤 포레스트 학습 모듈; 및

(4) 상기 Student 랜덤 포레스트 학습 모듈에서 학습시켜진 Student 랜덤 포레스트를 이용하여 분류를 수행하는 분류 모듈을 포함하는 것을 특징으로 하는, 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법을 이용한 분류기.

청구항 12

제11항에 있어서, 상기 Student 랜덤 포레스트는,

상기 Teacher 랜덤 포레스트의 트리의 수보다 작은 값의 트리의 수를 갖는 것을 특징으로 하는, 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법을 이용한 분류기.

청구항 13

제11항에 있어서,

상기 데이터 세트 A는 클래스 레이블이 포함되는 것을 특징으로 하는, 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법을 이용한 분류기.

청구항 14

제11항에 있어서,

상기 데이터 세트 B는 클래스 레이블이 포함되지 않는 것을 특징으로 하는, 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법을 이용한 분류기.

청구항 15

제11항에 있어서, 상기 Student 랜덤 포레스트 학습 모듈은,

(3-1) 상기 클래스 확률값 추출 모듈에서 각각의 클래스의 확률값이 추출된 데이터 세트 B에서, 임의의 데이터를 선택하여 서브 세트를 구성하는 단계;

(3-2) 상기 단계 (3-1)에서 구성된 서브 세트에 대하여, 각각의 내부 노드는 특징 벡터 x 에서 파라미터 θ_i 를 무작위로 선택하고, 그 중에서 가장 높은 성능을 갖는 분할 함수 f 와 θ 를 결정한 후, $f(\theta, x)$ 를 사용하여 트

리를 $\text{left}(S_l)$, $\text{right}(S_r)$ 로 분할하는 과정을, 최대 허용 깊이까지 반복하여 트리를 생성하는 단계; 및

(3-3) 상기 단계(3-2)에서 생성된 트리에 대하여, 평가 기준을 만족하는 트리를 찾는 과정을, 트리의 개수까지 반복하여 수행하는 단계를 포함하여, Student 랜덤 포레스트를 학습시키는 것을 특징으로 하는, 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법을 이용한 분류기.

청구항 16

제15항에 있어서, 상기 단계 (3-2)에서,

성능 평가 기준은, 상기 Teacher 랜덤 포레스트에 의해 추출된 클래스 확률 값을 이용하여 계산되는, 분할 함수 f에 대한 엔트로피 값을 이용하는 것을 특징으로 하는, 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법을 이용한 분류기.

청구항 17

제16항에 있어서, 상기 단계 (3-2)에서,

상기 성능 평가 기준은, 다음 수학적식을 이용하는 것을 특징으로 하는, 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법을 이용한 분류기.

$$\Delta E = \frac{|S_l|}{|S_n|} E(S_l) + \frac{|S_r|}{|S_n|} E(S_r)$$

상기 수학적식에서, ΔE 는 엔트로피의 합, S_n 은 n번째 노드에 도달하는 데이터의 집합, S_l , S_r 은 각각 n번째 노드에서 왼쪽 혹은 오른쪽의 자식 노드로 들어가는 데이터의 집합, $|\cdot|$ 은 데이터 집합에 속한 데이터의 개수를 나타내며, $E(S)$ 는 다음 수학적식으로 계산됨.

$$E(S) = -\frac{1}{N} \sum_{i=1}^N P(C_i) \log(P(C_i))$$

상기 수학적식에서, $P(C_i)$ 는 각각의 클래스에 대한 확률 질량 함수를 나타냄.

청구항 18

제15항에 있어서, 상기 단계 (3-3)에서,

상기 성능 평가 기준은, 상기 Teacher 랜덤 포레스트와 상기 Student 랜덤 포레스트에 의해 각각 추출된 클래스 확률값의 교차-엔트로피(cross-entropy)를 이용하는 것을 특징으로 하는, 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법을 이용한 분류기.

청구항 19

제18항에 있어서, 상기 단계 (3-3)에서,

상기 성능 평가 기준은, 다음 수학적식을 이용하는 것을 특징으로 하는, 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법을 이용한 분류기.

$$\text{cross-entropy} = -\sum_{n=1}^N \sum_{i=1}^C P_T(C_i|x(n)) \cdot \log P_S(C_i|x(n))$$

상기 수학적식에서, $P_T(C_i|x(n))$, $P_S(C_i|x(n))$ 는 각각 상기 Teacher 랜덤 포레스트와 상기 Student 랜덤 포레스트에 대한, 특징 벡터 $x(n)$ 에서의 각각의 클래스에 대한 확률 질량 함수를 나타냄.

청구항 20

제19항에 있어서, 상기 단계 (3-3)에서,

상기 계산된 교차-엔트로피(cross-entropy)가 미리 지정된 임계값보다 작을 경우 해당 트리를 사용하되, 그렇지 않을 경우 트리를 재생성하는 것을 특징으로 하는, 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법을 이용한 분류기.

발명의 설명**기술 분야**

[0001] 본 발명은 랜덤 포레스트 분류 방법 및 분류기에 관한 것으로서, 보다 구체적으로는 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법 및 분류기에 관한 것이다.

배경 기술

[0002] 기계 학습(Machine Learning)에서의 랜덤 포레스트(Random Forest)는 분류, 회귀 분석 등에 사용되는 앙상블 학습 방법의 일종으로서, 훈련 과정에서 구성한 다수의 결정 트리로부터 부류(분류) 또는 평균 예측치(회귀 분석)를 출력함으로써 동작한다. 랜덤 포레스트는 여러 개의 결정 트리들을 임의적으로 학습하는 방식의 앙상블 방법이다. 랜덤 포레스트 방법은 크게 다수의 결정 트리를 구성하는 학습 단계와, 입력 벡터가 들어왔을 때 분류하거나 예측하는 테스트 단계로 구성되어 있다. 랜덤 포레스트는 검출, 분류, 그리고 회귀 등 다양한 애플리케이션으로 활용되고 있다. 도 1은 일반적인 랜덤 포레스트 분류 방법의 플로차트를 도시한 도면이다.

[0003] 랜덤 포레스트의 가장 핵심적인 특징은 임의성(randomness)에 의해 서로 조금씩 다른 특성을 갖는 트리들로 구성된다는 점이다. 이 특징은 각각의 트리들의 예측(prediction)들이 비상관화(decorrelation) 되게 하며, 결과적으로 일반화(generalization) 성능을 향상시킨다. 또한, 임의화(randomization)는 포레스트가 노이즈가 포함된 데이터에 대해서도 강인하게 만들어 준다. 임의화는 각각의 트리들의 훈련 과정에서 진행되며, 가장 널리 쓰이는 두 가지 방법으로는 임의 학습 데이터 추출 방법을 이용한 앙상블 학습법인 배깅(bagging)과 임의 노드 최적화(randomized node optimization)가 있다. 이 두 가지 방법은 서로 동시에 사용되어 임의화 특성을 더욱 증진시킬 수 있다.

[0004] 랜덤 포레스트에서 가장 큰 영향을 미치는 매개변수들은 포레스트의 크기(트리의 개수)와 최대 허용 깊이 등이 다. 이 중, 포레스트의 크기(트리의 개수)는, 총 포레스트를 몇 개의 트리로 구성할지를 결정하는 매개변수이다. 포레스트의 크기가 작으면, 즉 트리의 개수가 적으면 트리들을 구성하고 테스트하는데 걸리는 시간이 짧은 대신, 일반화 능력이 떨어져 임의의 입력 데이터 포인트에 대해 틀린 결과를 내놓을 확률이 높다. 반면에, 포레스트의 크기가 크면, 즉 트리의 개수가 많으면 높은 성능을 보장하지만, 훈련과 테스트 시간이 길어지고 메모리량이 증가하는 단점이 있다. 따라서, 높은 성능은 보장하면서도, 처리 시간 및 메모리량을 줄일 수 있는 개선된 랜덤 포레스트 방법을 개발할 필요성이 있다.

[0005] 랜덤 포레스트 분류 방법과 관련된 선행특허로서는, 특허 제10-1237089호(발명의 명칭: 랜덤 포레스트 분류 기법을 이용한 산불연기 감지 방법)와 특허 제10-1697183호(발명의 명칭: 인공위성 영상과 랜덤포레스트 분류기 결합을 이용한 자동 하천 검출 시스템 및 방법) 등이 있다.

발명의 내용**해결하려는 과제**

[0006] 본 발명은 기존에 제안된 방법들의 상기와 같은 문제점들을 해결하기 위해 제안된 것으로서, 기존의 랜덤 포레스트 학습 방법을 이용하여 Teacher 랜덤 포레스트를 학습시키고, 이렇게 학습시켜진 Teacher 랜덤 포레스트를

이용하여 Student 랜덤 포레스트를 위한 데이터 세트를 구성하는 각각의 클래스의 확률값을 추출한 후, 이렇게 각각의 클래스의 확률값이 추출된 데이터 세트를 이용하여 Student 랜덤 포레스트를 학습시킴으로써, 랜덤 포레스트의 성능은 유지하면서도 랜덤 포레스트의 트리의 수를 줄여 처리 시간 및 메모리량을 대폭 줄일 수 있는, 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법 및 이를 이용한 분류기를 제공하는 것을 그 목적으로 한다.

과제의 해결 수단

- [0007] 상기한 목적을 달성하기 위한 본 발명의 특징에 따른 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법은,
- [0008] 랜덤 포레스트 분류 방법으로서,
- [0009] (1) 데이터 세트 A를 이용하여, 기존의 랜덤 포레스트 학습 방법을 적용하여 Teacher 랜덤 포레스트를 학습시키는 단계;
- [0010] (2) 상기 단계 (1)에서 학습시켜진 Teacher 랜덤 포레스트를 이용하여, Student 랜덤 포레스트를 위한 데이터 세트 B를 구성하는 각각의 클래스의 확률값을 추출하는 단계;
- [0011] (3) 상기 단계 (2)에서 각각의 클래스의 확률값이 추출된 데이터 세트 B를 이용하여, Student 랜덤 포레스트를 학습시키는 단계; 및
- [0012] (4) 상기 단계 (3)에서 학습시켜진 Student 랜덤 포레스트를 이용하여 분류를 수행하는 단계를 포함하는 것을 그 구성상의 특징으로 한다.
- [0013] 바람직하게는, 상기 Student 랜덤 포레스트는,
- [0014] 상기 Teacher 랜덤 포레스트의 트리의 수보다 작은 값의 트리의 수를 가질 수 있다.
- [0015] 바람직하게는, 상기 단계 (1)에서,
- [0016] 상기 데이터 세트 A는 클래스 레이블이 포함될 수 있다.
- [0017] 바람직하게는, 상기 단계 (2)에서,
- [0018] 상기 데이터 세트 B는 클래스 레이블이 포함되지 않을 수 있다.
- [0019] 바람직하게는, 상기 단계 (3)은,
- [0020] (3-1) 상기 단계 (2)에서 각각의 클래스의 확률값이 추출된 데이터 세트 B에서, 임의의 데이터를 선택하여 서브 세트를 구성하는 단계;
- [0021] (3-2) 상기 단계 (3-1)에서 구성된 서브 세트에 대하여, 각각의 내부 노드는 특징 벡터 x 에서 파라미터 θ_i 를 무작위로 선택하고, 그 중에서 가장 높은 성능을 갖는 분할 함수 f 와 θ 를 결정한 후, $f(\theta, x)$ 를 사용하여 트리를 $\text{left}(S_l)$, $\text{right}(S_r)$ 로 분할하는 과정을, 최대 허용 깊이까지 반복하여 트리를 생성하는 단계; 및
- [0022] (3-3) 상기 단계 (3-2)에서 생성된 트리에 대하여, 평가 기준을 만족하는 트리를 찾는 과정을, 트리의 개수까지 반복하여 수행하는 단계를 포함할 수 있다.
- [0023] 더욱 바람직하게는, 상기 단계 (3-2)에서,
- [0024] 성능 평가 기준은, 상기 Teacher 랜덤 포레스트에 의해 추출된 클래스 확률 값을 이용하여 계산되는, 분할 함수 f 에 대한 엔트로피 값을 이용할 수 있다.

[0025] 더욱더 바람직하게는, 상기 단계 (3-2)에서,

[0026] 상기 성능 평가 기준은, 다음 수학적식을 이용할 수 있다.

$$\Delta E = \frac{|S_l|}{|S_n|} E(S_l) + \frac{|S_r|}{|S_n|} E(S_r)$$

[0028] 상기 수학적식에서, ΔE 는 엔트로피의 합, S_n 은 n 번째 노드에 도달하는 데이터의 집합, S_l , S_r 은 각각 n 번째 노드에서 왼쪽 혹은 오른쪽의 자식 노드로 들어가는 데이터의 집합, $|\cdot|$ 은 데이터 집합에 속한 데이터의 개수를 나타내며, $E(S)$ 는 다음 수학적식으로 계산될 수 있다.

$$E(S) = -\frac{1}{N} \sum_{i=1}^N P(C_i) \log(P(C_i))$$

[0030] 상기 수학적식에서, $P(C_i)$ 는 각각의 클래스에 대한 확률 질량 함수를 나타낸다.

[0031] 더욱 바람직하게는, 상기 단계 (3-3)에서,

[0032] 상기 성능 평가 기준은, 상기 Teacher 랜덤 포레스트와 상기 Student 랜덤 포레스트에 의해 각각 추출된 클래스 확률값의 교차-엔트로피(cross-entropy)를 이용할 수 있다.

[0033] 더욱더 바람직하게는, 상기 단계 (3-3)에서,

[0034] 상기 성능 평가 기준은, 다음 수학적식을 이용할 수 있다.

$$cross-entropy = -\sum_{n=1}^N \sum_{i=1}^C P_T(C_i|x(n)) \cdot \log P_S(C_i|x(n))$$

[0036] 상기 수학적식에서, $P_T(C_i|x(n))$, $P_S(C_i|x(n))$ 는 각각 상기 Teacher 랜덤 포레스트와 상기 Student 랜덤 포레스트에 대한, 특징 벡터 $x(n)$ 에서의 각각의 클래스에 대한 확률 질량 함수를 나타낸다.

[0037] 더 더욱 바람직하게는, 상기 단계 (3-3)에서,

[0038] 상기 계산된 교차-엔트로피(cross-entropy)가 미리 지정된 임계값보다 작을 경우 해당 트리를 사용하되, 그렇지 않을 경우 트리를 재생성할 수 있다.

[0039] 또한, 상기한 목적을 달성하기 위한 본 발명의 특징에 따른 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법을 이용한 분류기는,

[0040] 랜덤 포레스트 분류 방법을 이용한 분류기로서,

[0041] (1) 데이터 세트 A를 이용하여, 기존의 랜덤 포레스트 학습 방법을 적용하여 Teacher 랜덤 포레스트를 학습시키는 Teacher 랜덤 포레스트 학습 모듈;

[0042] (2) 상기 Teacher 랜덤 포레스트 학습 모듈에서 학습시켜진 Teacher 랜덤 포레스트를 이용하여, Student 랜덤 포레스트를 위한 데이터 세트 B를 구성하는 각각의 클래스의 확률값을 추출하는 클래스 확률값 추출 모듈;

[0043] (3) 상기 클래스 확률값 추출 모듈에서 각각의 클래스의 확률값이 추출된 데이터 세트 B를 이용하여, Student 랜덤 포레스트를 학습시키는 Student 랜덤 포레스트 학습 모듈; 및

[0044] (4) 상기 Student 랜덤 포레스트 학습 모듈에서 학습시켜진 Student 랜덤 포레스트를 이용하여 분류를 수행하는 분류 모듈을 포함하는 것을 그 구성상의 특징으로 한다.

- [0045] 바람직하게는, 상기 Student 랜덤 포레스트는,
- [0046] 상기 Teacher 랜덤 포레스트의 트리의 수보다 작은 값의 트리의 수를 가질 수 있다.
- [0047] 바람직하게는,
- [0048] 상기 데이터 세트 A는 클래스 레이블이 포함될 수 있다.
- [0049] 바람직하게는,
- [0050] 상기 데이터 세트 B는 클래스 레이블이 포함되지 않을 수 있다.
- [0051] 바람직하게는, 상기 Student 랜덤 포레스트 학습 모듈은,
- [0052] (3-1) 상기 단계 (2)에서 각각의 클래스의 확률값이 추출된 데이터 세트 B에서, 임의의 데이터를 선택하여 서브 세트를 구성하는 단계;
- [0053] (3-2) 상기 단계 (3-1)에서 구성된 서브 세트에 대하여, 각각의 내부 노드는 특징 벡터 x 에서 파라미터 θ_i 를 무작위로 선택하고, 그 중에서 가장 높은 성능을 갖는 분할 함수 f 와 θ 를 결정한 후, $f(\theta, x)$ 를 사용하여 트리를 $\text{left}(S_l)$, $\text{right}(S_r)$ 로 분할하는 과정을, 최대 허용 깊이까지 반복하여 트리를 생성하는 단계; 및
- [0054] (3-3) 상기 단계 (3-2)에서 생성된 트리에 대하여, 평가 기준을 만족하는 트리를 찾는 과정을, 트리의 개수까지 반복하여 수행하는 단계를 포함하여, Student 랜덤 포레스트를 학습시킬 수 있다.
- [0055] 더욱 바람직하게는, 상기 단계 (3-2)에서,
- [0056] 성능 평가 기준은, 상기 Teacher 랜덤 포레스트에 의해 추출된 클래스 확률 값을 이용하여 계산되는, 분할 함수 f 에 대한 엔트로피 값을 이용할 수 있다.
- [0057] 더욱더 바람직하게는, 상기 단계 (3-2)에서,
- [0058] 상기 성능 평가 기준은, 다음 수학적식을 이용할 수 있다.
- [0059]
$$\Delta E = \frac{|S_l|}{|S_n|} E(S_l) + \frac{|S_r|}{|S_n|} E(S_r)$$
- [0060] 상기 수학적식에서, ΔE 는 엔트로피의 합, S_n 은 n 번째 노드에 도달하는 데이터의 집합, S_l , S_r 은 각각 n 번째 노드에서 왼쪽 혹은 오른쪽의 자식 노드로 들어가는 데이터의 집합, $|\cdot|$ 은 데이터 집합에 속한 데이터의 개수를 나타내며, $E(S)$ 는 다음 수학적식으로 계산될 수 있다.
- [0061]
$$E(S) = -\frac{1}{N} \sum_{i=1}^N P(C_i) \log(P(C_i))$$
- [0062] 상기 수학적식에서, $P(C_i)$ 는 각각의 클래스에 대한 확률 질량 함수를 나타낸다.
- [0063] 더욱 바람직하게는, 상기 단계 (3-3)에서,
- [0064] 상기 성능 평가 기준은, 상기 Teacher 랜덤 포레스트와 상기 Student 랜덤 포레스트에 의해 각각 추출된 클래스 확률값의 교차-엔트로피(cross-entropy)를 이용할 수 있다.
- [0065] 더욱더 바람직하게는, 상기 단계 (3-3)에서,

[0066] 상기 성능 평가 기준은, 다음 수학적식을 이용할 수 있다.

$$[0067] \text{cross-entropy} = - \sum_{n=1}^N \sum_{i=1}^C P_T(C_i|x(n)) \cdot \log P_S(C_i|x(n))$$

[0068] 상기 수학적식에서, $P_T(C_i|x(n))$, $P_S(C_i|x(n))$ 는 각각 상기 Teacher 랜덤 포레스트와 상기 Student 랜덤 포레스트에 대한, 특징 벡터 $x(n)$ 에서의 각각의 클래스에 대한 확률 질량 함수를 나타낸다.

[0069] 더 더욱 바람직하게는, 상기 단계 (3-3)에서,

[0070] 상기 계산된 교차-엔트로피(cross-entropy)가 미리 지정된 임계값보다 작을 경우 해당 트리를 사용하되, 그렇지 않을 경우 트리를 재생성할 수 있다.

발명의 효과

[0071] 본 발명에서 제안하고 있는 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법 및 이를 이용한 분류기에 따르면, 기존의 랜덤 포레스트 학습 방법을 이용하여 Teacher 랜덤 포레스트를 학습시키고, 이렇게 학습시켜진 Teacher 랜덤 포레스트를 이용하여 Student 랜덤 포레스트를 위한 데이터 세트를 구성하는 각각의 클래스의 확률값을 추출한 후, 이렇게 각각의 클래스의 확률값이 추출된 데이터 세트를 이용하여 Student 랜덤 포레스트를 학습시킴으로써, 랜덤 포레스트의 성능은 유지하면서도 랜덤 포레스트의 트리의 수를 줄여 처리 시간 및 메모리량을 대폭 줄일 수 있다.

도면의 간단한 설명

[0072] 도 1은 일반적인 랜덤 포레스트 분류 방법의 플로차트를 도시한 도면.

도 2는 본 발명의 일실시예에 따른 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법의 구성을 도시한 도면.

도 3은 본 발명의 일실시예에 따른 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법에서, Student 랜덤 포레스트를 학습시키는 단계 S300의 세부구성을 도시한 도면.

도 4는, 본 발명의 일실시예에 따른 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법을 알고리즘으로 구현한 것을 나타내는 도면.

도 5 및 도 6은 본 발명의 일실시예에 따른 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법에 대한 전체적인 과정을 그림으로 표현한 것으로서, 도 5는 본 발명의 일실시예에 따른 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법에서, Teacher 랜덤 포레스트의 학습 과정을 그림으로 표현한 도면이며, 도 6은 본 발명의 일실시예에 따른 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법에서, Student 랜덤 포레스트의 학습 과정을 그림으로 표현한 도면.

도 7은, 본 발명의 일실시예에 따른 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법을 이용한 분류기의 구성을 도시한 도면.

발명을 실시하기 위한 구체적인 내용

[0073] 이하, 첨부된 도면을 참조하여 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 본 발명을 용이하게 실시할 수 있도록 바람직한 실시예를 상세히 설명한다. 다만, 본 발명의 바람직한 실시예를 상세하게 설명함에 있어, 관련된 공지 기능 또는 구성에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략한다. 또한, 유사한 기능 및 작용을 하는 부분에 대해서는 도면 전체에 걸쳐 동일한 부호를 사용한다.

[0074] 덧붙여, 명세서 전체에서, 어떤 부분이 다른 부분과 ‘연결’되어 있다고 할 때, 이는 ‘직접적으로 연결’되어 있는 경우뿐만 아니라, 그 중간에 다른 소자를 사이에 두고 ‘간접적으로 연결’되어 있는 경우도 포함한다. 또한, 어떤 구성요소를 ‘포함’한다는 것은, 특별히 반대되는 기재가 없는 한 다른 구성요소를 제

외하는 것이 아니라 다른 구성요소를 더 포함할 수 있다는 것을 의미한다.

- [0075] 도 2는 본 발명의 일실시예에 따른 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법의 구성을 도시한 도면이다. 도 2에 도시된 바와 같이, 본 발명의 일실시예에 따른 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법은, 데이터 세트 A를 이용하여, 기존의 랜덤 포레스트 학습 방법을 적용하여 Teacher 랜덤 포레스트를 학습시키는 단계(S100); 단계 S100에서 학습시켜진 Teacher 랜덤 포레스트를 이용하여, Student 랜덤 포레스트를 위한 데이터 세트 B를 구성하는 각각의 클래스의 확률값을 추출하는 단계(S200); 단계 S200에서 각각의 클래스의 확률값이 추출된 데이터 세트 B를 이용하여, Student 랜덤 포레스트를 학습시키는 단계(S300); 및 단계 S300에서 학습시켜진 Student 랜덤 포레스트를 이용하여 분류를 수행하는 단계(S400)를 포함하여 구성될 수 있다.
- [0076] 이때, Student 랜덤 포레스트는, Teacher 랜덤 포레스트의 트리의 수보다 작은 값의 트리의 수를 가지도록 함으로써, 랜덤 포레스트의 경량화를 달성할 수 있다. 또한, 단계 S100에서 Teacher 랜덤 포레스트를 학습시키는데 사용되는 데이터 세트 A는 클래스 레이블이 포함되도록 설정하고, 단계 S200 및 단계 S300에서 Student 랜덤 포레스트를 학습시키는데 사용되는 데이터 세트 B는 클래스 레이블이 포함되지 않도록 설정할 수 있다.
- [0077] 단계 S100에서는, 데이터 세트 A를 이용하여, 기존의 랜덤 포레스트 학습 방법을 적용하여 Teacher 랜덤 포레스트를 학습시킨다. 본 단계에서는, 기존의 랜덤 포레스트 학습 방법을 적용함으로써, 종래와 동일한 방식의 랜덤 포레스트가 적용된 일반적인 랜덤 포레스트를 얻게 된다. 본 단계는 사전 처리가 가능한 학습 단계로서, 실질적인 분류 단계의 처리 시간과 무관하므로, 기존의 랜덤 포레스트 학습 방법을 적용한, 충분한 트리의 수를 갖는 랜덤 포레스트를 생성하는 것이 바람직하다. 한편, 데이터 세트 A는, 앞서 언급한 바와 같이, 클래스 레이블이 포함되도록 설정할 수 있다.
- [0078] 단계 S200에서는, 단계 S100에서 학습시켜진 Teacher 랜덤 포레스트를 이용하여, Student 랜덤 포레스트를 위한 데이터 세트 B를 구성하는 각각의 클래스의 확률값을 추출한다. 본 단계에서는, 충분한 트리의 수를 갖도록 생성된 Teacher 랜덤 포레스트를 이용하여, 실질적인 분류 단계에서 사용될 Student 랜덤 포레스트를 학습시키는데 사용될 데이터 세트 B를 구성하는 각각의 클래스의 확률값을 추출하게 된다. 데이터 세트 B는, 앞서 언급한 바와 같이, 클래스 레이블이 포함되지 않도록 설정되며, 그 결과 각각의 클래스의 확률값이 추출될 수 있다.
- [0079] 단계 S300에서는, 단계 S200에서 각각의 클래스의 확률값이 추출된 데이터 세트 B를 이용하여, Student 랜덤 포레스트를 학습시킨다. 본 단계에서는, Teacher 랜덤 포레스트를 이용하여 각각의 클래스의 확률값이 추출된 데이터 세트 B를 이용하여, Student 랜덤 포레스트를 학습시키게 된다.
- [0080] 도 3은 본 발명의 일실시예에 따른 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법에서, Student 랜덤 포레스트를 학습시키는 단계 S300의 세부구성을 도시한 도면이다. 도 3에 도시된 바와 같이, 단계 S300은, 단계 S200에서 각각의 클래스의 확률값이 추출된 데이터 세트 B에서, 임의의 데이터를 선택하여 서브 세트를 구성하는 단계(S310); 단계 S310에서 구성된 서브 세트에 대하여, 각각의 내부 노드는 특징 벡터 x 에서 파라미터 θ_i 를 무작위로 선택하고, 그 중에서 가장 높은 성능을 갖는 분할 함수 f 와 θ 를 결정한 후, $f(\theta, x)$ 를 사용하여 트리를 $left(S_i)$, $right(S_i)$ 로 분할하는 과정을, 최대 허용 깊이까지 반복하여 트리를 생성하는 단계(S320); 및 단계 S320에서 생성된 트리에 대하여, 평가 기준을 만족하는 트리를 찾는 과정을, 트리의 개수까지 반복하여 수행하는 단계(S330)를 포함하여 구성될 수 있다.
- [0081] 단계 S320에서는, 성능 평가 기준으로서, Teacher 랜덤 포레스트에 의해 추출된 클래스 확률 값을 이용하여 계산되는, 분할 함수 f 에 대한 엔트로피 값을 이용할 수 있으며, 이는 다음 수학적 식 1과 같이 계산될 수 있다.

수학식 1

$$\Delta E = \frac{|S_l|}{|S_n|} E(S_l) + \frac{|S_r|}{|S_n|} E(S_r)$$

[0082]

[0083] 상기 수학식 1에서, ΔE 는 엔트로피의 합, S_n 은 n 번째 노드에 도달하는 데이터의 집합, S_l , S_r 은 각각 n 번째 노드에서 왼쪽 혹은 오른쪽의 자식 노드로 들어가는 데이터의 집합, $|\cdot|$ 은 데이터 집합에 속한 데이터의 개수를 각각 나타내며, $E(S)$ 는 다음 수학식 2와 같이 계산될 수 있다.

수학식 2

$$E(S) = -\frac{1}{N} \sum_{i=1}^N P(C_i) \log(P(C_i))$$

[0084]

[0085] 상기 수학식 2에서, $P(C_i)$ 는 각각의 클래스에 대한 확률 질량 함수를 나타낸다. 여기서, 확률 질량 함수(probability mass function, pmf)는, 이산 확률 변수에서 특정 값에 대한 확률을 나타내는 함수로서, 연속 확률 변수에서의 확률 밀도 함수와 대응된다.

[0086] 단계 S330에서는, 성능 평가 기준으로서, Teacher 랜덤 포레스트와 Student 랜덤 포레스트에 의해 각각 추출된 클래스 확률값의 교차-엔트로피(cross-entropy)를 이용할 수 있으며, 이는 다음 수학식 3과 같이 계산될 수 있다.

수학식 3

$$cross-entropy = -\sum_{n=1}^N \sum_{i=1}^C P_T(C_i|x(n)) \cdot \log P_S(C_i|x(n))$$

[0087]

[0088] 상기 수학식 3에서, $P_T(C_i|x(n))$, $P_S(C_i|x(n))$ 는 각각 Teacher 랜덤 포레스트와 Student 랜덤 포레스트 대한, 특징 벡터 $x(n)$ 에서의 각각의 클래스에 대한 확률 질량 함수를 나타낸다.

[0089] 단계 S330에서는, 이렇게 수학식 3을 통해 계산된 교차-엔트로피(cross-entropy)가 미리 지정된 임계값보다 작을 경우 해당 트리를 사용하되, 그렇지 않을 경우 트리를 재생성할 수 있다.

[0090] 도 4는, 본 발명의 일실시예에 따른 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법을 알고리즘으로 구현한 것을 나타내는 도면이다. 도 4를 통해, 본 발명의 일실시예에 따른 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법을 어떠한 방식으로 알고리즘으로 구현할 수 있는지 확인할 수 있다.

[0091] 도 5 및 도 6은 본 발명의 일실시예에 따른 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법에 대한 전체적인 과정을 그림으로 표현한 것으로서, 도 5는 본 발명의 일실시예에 따른 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법에서, Teacher 랜덤 포레스트의 학습 과정을 그림으로 표현한 도면이며, 도 6은 본 발명의 일실시예에 따른 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법에서, Student 랜덤 포레스트의 학습 과정을 그림으로 표현한 도면이다.

[0092] 도 7은, 본 발명의 일실시예에 따른 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법을 이용한 분류기(10)의 구성을 도시한 도면이다. 도 7에 도시된 바와 같이, 본 발명의 일실시예에 따른 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법을 이용한 분류기(10)는, 데이터 세트 A를 이용하여, 기존의 랜덤 포레스트 학습 방법을 적용하여 Teacher 랜덤 포레스트를 학습시키는 Teacher 랜덤 포레스트 학습 모듈(100); Teacher 랜덤 포레스트 학습 모듈(100)에서 학습시켜진 Teacher 랜덤 포레스트를 이용하여, Student 랜덤 포레스트를 위한 데이터 세트 B를 구성하는 각각의 클래스의 확률값을 추출하는 클래스 확률값 추출 모듈(200); 클래스 확률값 추출 모듈(200)에서 각각의 클래스의 확률값이 추출된 데이터 세트 B를 이용하여, Student 랜덤 포레스트를 학습시키는 Student 랜덤 포레스트 학습 모듈(300); 및 Student 랜덤 포레스트 학습 모듈(300)에서 학습시켜진 Student 랜덤 포레스트를 이용하여 분류를 수행하는 분류 모듈(400)을 포함하여 구성될 수 있다.

[0093] 본 발명의 일실시예에 따른 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법을 이용한 분류기를 구성하는 각각의 구성요소들, 즉 Teacher 랜덤 포레스트 학습 모듈(100), 클래스 확률값 추출 모듈(200), Student 랜덤 포레스트 학습 모듈(300), 및 분류 모듈(400)의 구체적인 역할 및 구성에 대해서는, 도 2 내지 도 6을 통해 이루어진, 본 발명의 일실시예에 따른 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법에 대한 설명에서 충분히 확인할 수 있으므로, 상세한 설명은 생략하도록 한다.

[0094] 상술한 바와 같이, 본 발명에서 제안하고 있는 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법 및 이를 이용한 분류기에 따르면, 기존의 랜덤 포레스트 학습 방법을 이용하여 Teacher 랜덤 포레스트를 학습시키고, 이렇게 학습시켜진 Teacher 랜덤 포레스트를 이용하여 Student 랜덤 포레스트를 위한 데이터 세트를 구성하는 각각의 클래스의 확률값을 추출한 후, 이렇게 각각의 클래스의 확률값이 추출된 데이터 세트를 이용하여 Student 랜덤 포레스트를 학습시킴으로써, 랜덤 포레스트의 성능은 유지하면서도 랜덤 포레스트의 트리의 수를 줄여 처리 시간 및 메모리량을 대폭 줄일 수 있다.

[0095] 이상 설명한 본 발명은 본 발명이 속한 기술분야에서 통상의 지식을 가진 자에 의하여 다양한 변형이나 응용이 가능하며, 본 발명에 따른 기술적 사상의 범위는 아래의 특허청구범위에 의하여 정해져야 할 것이다.

부호의 설명

[0096] 10: 본 발명의 일실시예에 따른 소프트 타겟 학습 방법을 적용하여 경량화된 랜덤 포레스트 분류 방법을 이용한 분류기

100: Teacher 랜덤 포레스트 학습 모듈

200: 클래스 확률값 추출 모듈

300: Student 랜덤 포레스트 학습 모듈

400: 분류 모듈

S100: 데이터 세트 A를 이용하여, 기존의 랜덤 포레스트 학습 방법을 적용하여 Teacher 랜덤 포레스트를 학습시키는 단계

S200: 단계 S100에서 학습시켜진 Teacher 랜덤 포레스트를 이용하여, Student 랜덤 포레스트를 위한 데이터 세트 B를 구성하는 각각의 클래스의 확률값을 추출하는 단계

S300: 단계 S200에서 각각의 클래스의 확률값이 추출된 데이터 세트 B를 이용하여, Student 랜덤 포레스트를 학습시키는 단계

S310: 단계 S200에서 각각의 클래스의 확률값이 추출된 데이터 세트 B에서, 임의의 데이터를 선택하여 서브 세트를 구성하는 단계

S320: 단계 S310에서 구성된 서브 세트에 대하여, 각각의 내부 노드는 특징 벡터 x 에서 파라미터 θ_i 를 무작위로 선택하고, 그 중에서 가장 높은 성능을 갖는 분할 함수 f 와 θ 를 결정한 후, $f(\theta, x)$ 를 사용하여 트리를

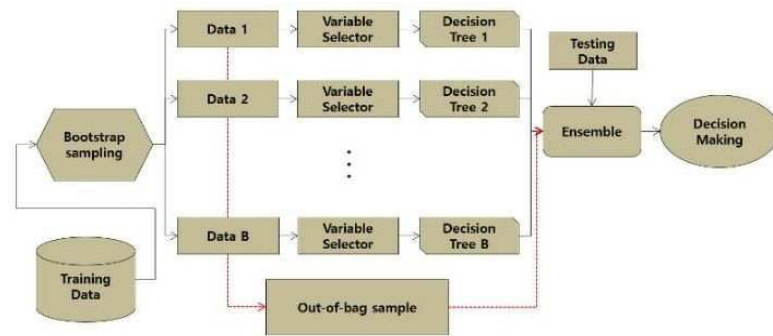
$left(S_l)$, $right(S_r)$ 로 분할하는 과정을, 최대 허용 깊이까지 반복하여 트리를 생성하는 단계

S330: 단계 S320에서 생성된 트리에 대하여, 평가 기준을 만족하는 트리를 찾는 과정을, 트리의 개수까지 반복하여 수행하는 단계

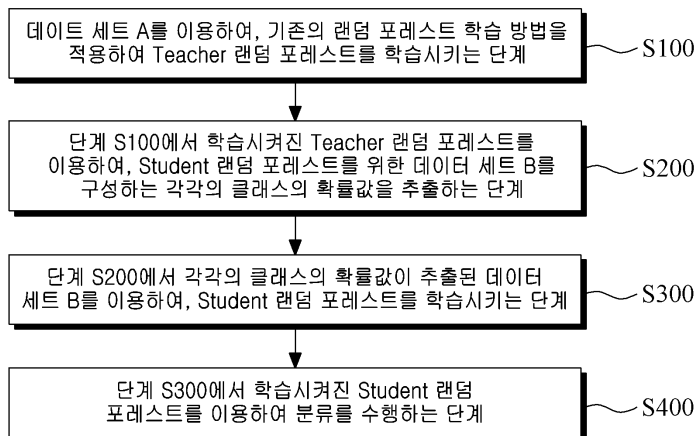
S400: 단계 S300에서 학습시켜진 Student 랜덤 포레스트를 이용하여 분류를 수행하는 단계

도면

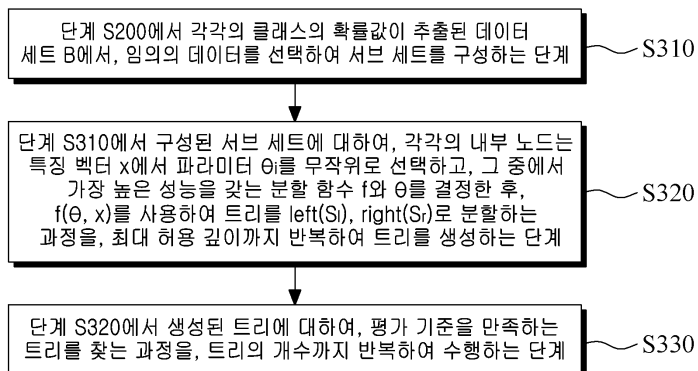
도면1



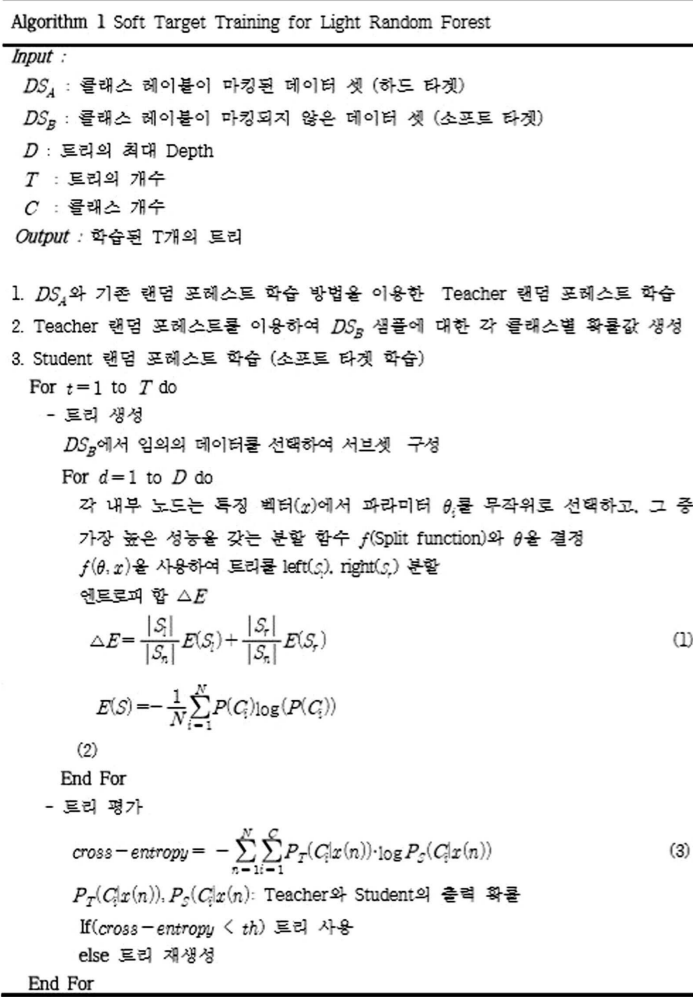
도면2



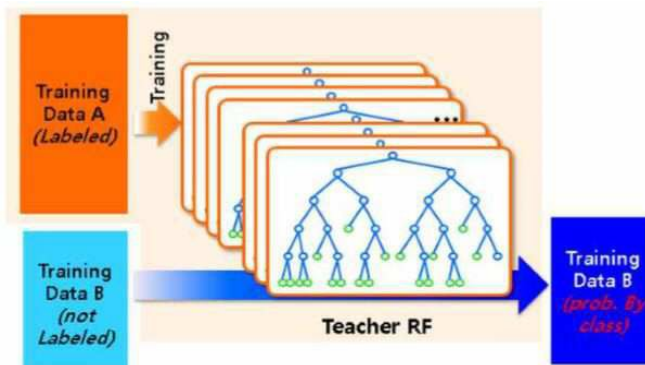
도면3



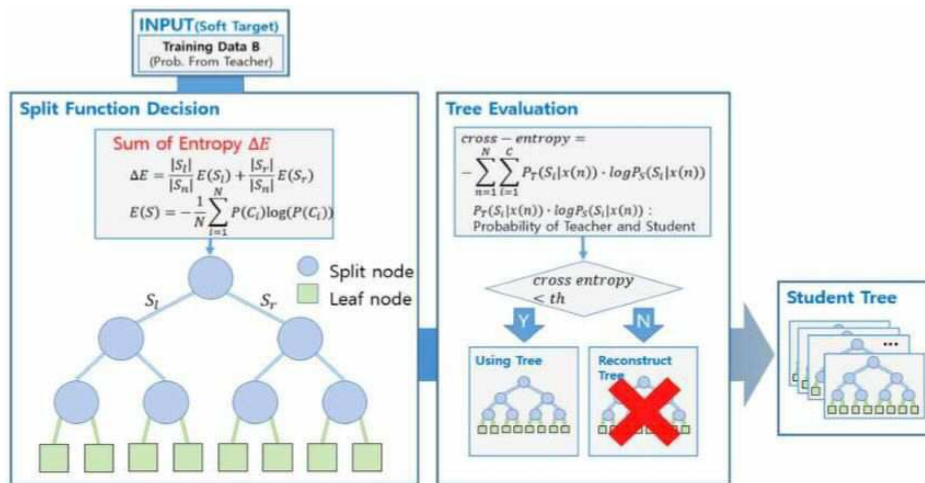
도면4



도면5



도면6



도면7

