



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0057715
(43) 공개일자 2021년05월21일

(51) 국제특허분류(Int. Cl.)
G06N 3/08 (2006.01) G06N 20/00 (2019.01)
G06N 3/04 (2006.01)
(52) CPC특허분류
G06N 3/08 (2013.01)
G06N 20/00 (2019.01)
(21) 출원번호 10-2021-0060053(분할)
(22) 출원일자 2021년05월10일
심사청구일자 2021년05월10일
(62) 원출원 특허 10-2020-0076297
원출원일자 2020년06월23일
심사청구일자 2020년06월23일
(30) 우선권주장
1020190075858 2019년06월25일 대한민국(KR)

(71) 출원인
세종대학교산학협력단
서울특별시 광진구 능동로 209 (군자동, 세종대학교)
(72) 발명자
박기호
서울특별시 노원구 중계로 184, 101동 903호(중계동, 라이프청구신동아아파트)
김준영
인천광역시 서구 봉오재3로 75, 104동 2002호(가정동, 루원 호반베르디움 더센트럴)
(74) 대리인
(뒷면에 계속)
민영준

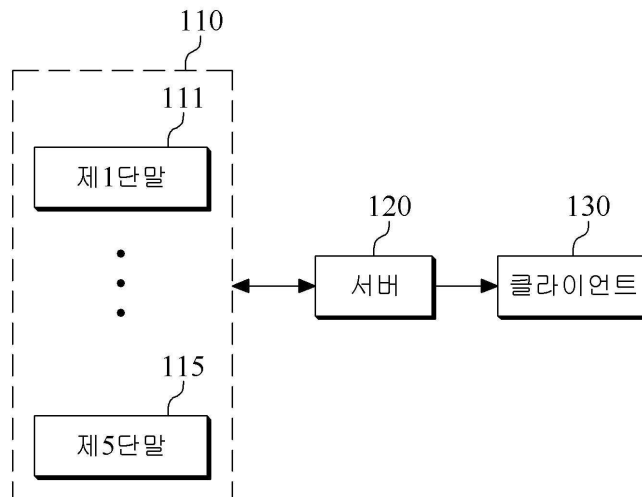
전체 청구항 수 : 총 25 항

(54) 발명의 명칭 분산된 분류 모델을 이용하는 분류 방법

(57) 요약

미리 학습된 분류 모델을 여러 단말에 분산시켜, 타겟 데이터를 특정 클래스로 분류하는 방법이 개시된다. 개시된 분산된 분류 모델을 이용하는 분류 방법은 복수의 단말 각각에 분산된 서브 분류 모델을 통해 생성된, 타겟 데이터에 대한 분류 데이터를 수신하는 단계; 및 상기 분류 데이터를 이용하여, 상기 타겟 데이터에 대한 최종 클래스를 결정하는 단계를 포함하며, 미리 학습된 상기 서브 분류 모델 각각에 할당된 분류를 위한 클래스는, 복수의 타겟 클래스 중 적어도 하나를 포함하며, 상기 서브 분류 모델 각각에 할당된 클래스의 개수는, 상기 타겟 클래스의 개수보다 적다.

대표도 - 도1



(52) CPC특허분류

G06N 3/0454 (2013.01)

(72) 발명자

전중호

경기도 부천시 장말로362번길 9, 402호(심곡동, 신암로즈빌)

기민관

서울특별시 동대문구 서울시립대로 14, 104동 1104호(답십리동, 청계한신휴플러스)

이 발명을 지원한 국가연구개발사업

과제고유번호	2017291010
과제번호	2017K2A9A1A01093012
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	연구교류지원사업(일반)
연구과제명	엣지 컴퓨팅 시스템 구조의 설계 이슈 및 성능 병목에 관한 연구
기 여 율	1/2
과제수행기관명	세종대학교 산학협력단
연구기간	2017.12.20 ~ 2018.12.19

이 발명을 지원한 국가연구개발사업

과제고유번호	1711086645
과제번호	2018R1A2B6002534
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	개인기초연구(과기정통부)(R&D)
연구과제명	차세대 응용을 위한 데이터 중심 가속기 기반 컴퓨팅 시스템 구조 설계
기 여 율	1/2
과제수행기관명	세종대학교 산학협력단
연구기간	2019.03.01 ~ 2020.02.29

명세서

청구범위

청구항 1

복수의 단말 각각에 분산된 서브 분류 모델을 통해 생성된, 타겟 데이터에 대한 분류 데이터를 수신하는 단계;
및

상기 분류 데이터를 이용하여, 상기 타겟 데이터에 대한 최종 클래스를 결정하는 단계를 포함하며,

미리 학습된 상기 서브 분류 모델 각각에 할당된 분류를 위한 클래스는, 복수의 타겟 클래스 중 적어도 하나를 포함하며,

상기 서브 분류 모델 각각에 할당된 클래스의 개수는, 상기 타겟 클래스의 개수보다 적으며,

상기 서브 분류 모델 중 적어도 하나는, 상기 복수의 타겟 클래스가 할당된 메인 분류 모델로부터 경량화된 모델이거나 상기 메인 분류 모델과 독립적으로 학습된 모델인,

분산된 분류 모델을 이용하는 분류 방법.

청구항 2

제 1항에 있어서,

상기 서브 분류 모델 각각에 할당된 클래스는

상기 타겟 클래스 이외의 클래스인 기타 클래스를 더 포함하는

분산된 분류 모델을 이용하는 분류 방법.

청구항 3

제 1항에 있어서,

상기 서브 분류 모델 각각에 할당된 클래스는,

서로 다른 타겟 클래스이거나 모두 동일한 클래스인,

분산된 분류 모델을 이용하는 분류 방법.

청구항 4

제 1항에 있어서,

상기 서브 분류 모델 각각에 할당된 클래스는

서로 다른 타겟 클래스를 포함하는

분산된 분류 모델을 이용하는 분류 방법.

청구항 5

제 1항에 있어서,

상기 서브 분류 모델 각각에 할당되는 클래스는

상기 타겟 클래스의 상위 개념에 따라서 결정되는
분산된 분류 모델을 이용하는 분류 방법.

청구항 6

제 1항에 있어서,
상기 서브 분류 모델 중 제1 및 제2서브 분류 모델 각각에 할당된 클래스는,
적어도 하나의 중복된 타겟 클래스를 포함하는
분산된 분류 모델을 이용하는 분류 방법.

청구항 7

제 6항에 있어서,
상기 분류 데이터는
상기 서브 분류 모델 각각에 할당된 클래스에 대한 신뢰도값을 포함하며,
상기 최종 클래스를 결정하는 단계는
상기 중복된 타겟 클래스의 신뢰도값 중 최대값 또는 상기 중복된 타겟 클래스의 신뢰도값의 평균값을
이용하여, 상기 최종 클래스를 결정하는
분산된 분류 모델을 이용하는 분류 방법.

청구항 8

제 2항에 있어서,
상기 분류 데이터는
상기 서브 분류 모델 각각에 할당된 클래스에 대한 신뢰도값을 포함하며,
상기 최종 클래스를 결정하는 단계는
상기 신뢰도값 중 최대값에 대응되는 클래스를, 상기 최종 클래스로 결정하는
분산된 분류 모델을 이용하는 분류 방법.

청구항 9

제 8항에 있어서,
상기 분류 데이터는
상기 서브 분류 모델 각각에 할당된 클래스 중 상기 타겟 클래스에 대한 신뢰도값을 포함하는
분산된 분류 모델을 이용하는 분류 방법.

청구항 10

제 2항에 있어서,
상기 분류 데이터는

상기 서브 분류 모델 각각에 할당된 클래스에 대한 신뢰도값 중 최대값을 포함하며,
 상기 최종 클래스를 결정하는 단계는
 상기 최대값을 이용하여, 상기 최종 클래스를 결정하는
 분산된 분류 모델을 이용하는 분류 방법.

청구항 11

제 10항에 있어서,
 상기 분류 데이터는
 상기 서브 분류 모델 각각에 할당된 클래스 중 상기 타겟 클래스에 대한 신뢰도값을 포함하는
 분산된 분류 모델을 이용하는 분류 방법.

청구항 12

제 1항에 있어서,
 상기 분류 데이터는
 상기 서브 분류 모델 중 일부에 할당된 클래스에 대한 신뢰도값을 포함하며,
 상기 최종 클래스를 결정하는 단계는
 상기 신뢰도값이 임계값보다 작은 경우, 상기 분류 데이터에 포함되지 않은 타겟 클래스를 상기 최종 클래스로
 결정하는
 분산된 분류 모델을 이용하는 분류 방법.

청구항 13

복수의 단말 각각에 분산된 서브 분류 모델을 통해 생성된, 타겟 데이터에 대한 분류 데이터를 수신하는 단계;
 및
 상기 분류 데이터를 이용하여, 상기 타겟 데이터에 대한 최종 클래스를 결정하는 단계를 포함하며,
 미리 학습된 상기 서브 분류 모델 각각에 할당된 분류를 위한 클래스는, 복수의 타겟 클래스 중 적어도 하나를
 포함하며,
 상기 서브 분류 모델 각각에 할당된 클래스의 개수는, 상기 타겟 클래스의 개수보다 적으며,
 상기 서브 분류 모델 각각에 할당된 클래스는, 서로 다른 타겟 클래스이며,
 상기 서브 분류 모델은, 분류를 위해 상기 복수의 타겟 클래스가 할당된 메인 분류 모델과 상이한 분류 정확도
 를 가지도록 학습된 모델인
 분산된 분류 모델을 이용하는 분류 방법.

청구항 14

복수의 단말 각각에 분산된 서브 분류 모델을 통해 생성된, 타겟 데이터에 대한 분류 데이터를 수신하는 단계;
 및
 상기 분류 데이터를 이용하여, 상기 타겟 데이터에 대한 최종 클래스를 결정하는 단계를 포함하며,

미리 학습된 상기 서브 분류 모델 각각에 할당된 분류를 위한 클래스는, 복수의 타겟 클래스 중 적어도 하나를 포함하며,

상기 서브 분류 모델 중 적어도 하나는, 상기 복수의 타겟 클래스가 할당된 메인 분류 모델로부터 경량화된 모델이거나 상기 메인 분류 모델과 독립적으로 학습된 모델인,

분산된 분류 모델을 이용하는 분류 방법.

청구항 15

제 14항에 있어서,

상기 서브 분류 모델 각각에 할당된 클래스의 개수는,

상기 타겟 클래스의 개수보다 적은 개수인

분산된 분류 모델을 이용하는 분류 방법.

청구항 16

제 14항에 있어서,

상기 서브 분류 모델은,

상기 메인 분류 모델과 상이한 분류 정확도를 가지도록 학습된 모델인

분산된 분류 모델을 이용하는 분류 방법.

청구항 17

복수의 단말의 자원을 모니터링하는 단계; 및

상기 모니터링 결과에 따라서, 상기 단말에 서브 분류 모델을 할당하는 단계를 포함하며,

미리 학습된 상기 서브 분류 모델 각각에 할당된 분류를 위한 클래스는, 복수의 타겟 클래스 중 적어도 하나를 포함하며,

상기 서브 분류 모델 각각에 할당된 클래스의 개수는, 상기 타겟 클래스의 개수보다 적으며,

상기 서브 분류 모델 중 적어도 하나는, 상기 복수의 타겟 클래스가 할당된 메인 분류 모델로부터 경량화된 모델이거나 상기 메인 분류 모델과 독립적으로 학습된 모델인,

분산된 분류 모델을 이용하는 분류 방법.

청구항 18

복수의 단말의 자원을 모니터링하는 단계; 및

상기 모니터링 결과에 따라서, 상기 단말에 서브 분류 모델을 할당하는 단계를 포함하며,

미리 학습된 상기 서브 분류 모델 각각에 할당된 분류를 위한 클래스는, 복수의 타겟 클래스 중 적어도 하나를 포함하며,

상기 서브 분류 모델 각각에 할당된 클래스의 개수는, 상기 타겟 클래스의 개수보다 적으며,

상기 서브 분류 모델 각각에 할당된 클래스는, 서로 다른 타겟 클래스이며,

상기 서브 분류 모델은, 분류를 위해 상기 복수의 타겟 클래스가 할당된 메인 분류 모델과 상이한 분류 정확도를 가지도록 학습된 모델인

분산된 분류 모델을 이용하는 분류 방법.

청구항 19

복수의 단말의 자원을 모니터링하는 단계; 및

상기 모니터링 결과에 따라서, 상기 단말에 서버 분류 모델을 할당하는 단계를 포함하며,

미리 학습된 상기 서버 분류 모델 각각에 할당된 분류를 위한 클래스는, 복수의 타겟 클래스 중 적어도 하나를 포함하며,

상기 서버 분류 모델 중 적어도 하나는, 상기 복수의 타겟 클래스가 할당된 메인 분류 모델로부터 경량화된 모델이거나 상기 메인 분류 모델과 독립적으로 학습된 모델인

분산된 분류 모델을 이용하는 분류 방법.

청구항 20

제 17항 내지 제19항 중 어느 한 항에 있어서,

상기 서버 분류 모델 각각에 할당된 클래스의 개수는,

상기 단말의 가용 자원에 따라서 결정되며,

상기 서버 분류 모델 각각에 할당된 클래스는

상기 타겟 클래스 이외의 클래스인 기타 클래스를 더 포함하는

분산된 분류 모델을 이용하는 분류 방법.

청구항 21

복수의 후보 단말의 자원을 모니터링하는 단계; 및

상기 모니터링 결과에 따라서, 상기 후보 단말 중 서버 분류 모델을 이용하여 타겟 데이터를 분류하는 분류 단말을 결정하는 단계를 포함하며,

미리 학습된 상기 서버 분류 모델 각각에 할당된 분류를 위한 클래스는, 복수의 타겟 클래스 중 적어도 하나를 포함하며,

상기 서버 분류 모델 각각에 할당된 클래스의 개수는, 상기 타겟 클래스의 개수보다 적으며,

상기 서버 분류 모델 중 적어도 하나는, 상기 복수의 타겟 클래스가 할당된 메인 분류 모델로부터 경량화된 모델이거나 상기 메인 분류 모델과 독립적으로 학습된 모델인,

분산된 분류 모델을 이용하는 분류 방법.

청구항 22

복수의 후보 단말의 자원을 모니터링하는 단계; 및

상기 모니터링 결과에 따라서, 상기 후보 단말 중 서버 분류 모델을 이용하여 타겟 데이터를 분류하는 분류 단말을 결정하는 단계를 포함하며,

미리 학습된 상기 서버 분류 모델 각각에 할당된 분류를 위한 클래스는, 복수의 타겟 클래스 중 적어도 하나를 포함하며,

상기 서버 분류 모델 각각에 할당된 클래스의 개수는, 상기 타겟 클래스의 개수보다 적으며,

상기 서브 분류 모델 각각에 할당된 클래스는, 서로 다른 타겟 클래스이며,

상기 서브 분류 모델은, 분류를 위해 상기 복수의 타겟 클래스가 할당된 메인 분류 모델과 상이한 분류 정확도를 가지도록 학습된 모델인

분산된 분류 모델을 이용하는 분류 방법.

청구항 23

복수의 후보 단말의 자원을 모니터링하는 단계; 및

상기 모니터링 결과에 따라서, 상기 후보 단말 중 서브 분류 모델을 이용하여 타겟 데이터를 분류하는 분류 단말을 결정하는 단계를 포함하며,

미리 학습된 상기 서브 분류 모델 각각에 할당된 분류를 위한 클래스는, 복수의 타겟 클래스 중 적어도 하나를 포함하며,

상기 서브 분류 모델 중 적어도 하나는, 상기 복수의 타겟 클래스가 할당된 메인 분류 모델로부터 경량화된 모델이거나 상기 메인 분류 모델과 독립적으로 학습된 모델인

분산된 분류 모델을 이용하는 분류 방법.

청구항 24

제 21항 내지 제23항 중 어느 한 항에 있어서,

상기 모니터링 결과에 따라서, 상기 후보 단말 중 상기 타겟 데이터를 상기 분류 단말로 전송하는 전송 단말을 결정하는 단계를 더 포함하며,

상기 서브 분류 모델 각각에 할당된 클래스는

상기 타겟 클래스 이외의 클래스인 기타 클래스를 더 포함하는

분산된 분류 모델을 이용하는 분류 방법.

청구항 25

제 21항 내지 제23항 중 어느 한 항에 있어서,

상기 분류 단말 각각에 분산된 서브 분류 모델을 통해 생성된, 상기 타겟 데이터에 대한 분류 데이터를 수신하여, 상기 타겟 데이터에 대한 최종 클래스를 결정하는 클래스 결정 단말을, 상기 모니터링 결과에 따라서, 상기 후보 단말 중에서 결정하는 단계

를 더 포함하는 분산된 분류 모델을 이용하는 분류 방법.

발명의 설명

기술 분야

[0001] 본 발명은 분산된 분류 모델을 이용하는 분류 방법에 관한 발명으로서, 보다 상세하게는 미리 학습된 분류 모델을 여러 단말에 분산시켜, 타겟 데이터를 특정 클래스로 분류하는 방법에 관한 것이다.

배경 기술

[0003] 기존의 딥러닝 기반 기계학습 처리는 응용 분야가 지속적으로 확대되고 있으며, 모델의 정확도를 향상시키기 위

하여 모델의 복잡도 및 모델을 수행하기 위한 요구 성능이 지속적으로 증가하고 있다. 이러한 모델의 수행은 주로 고성능 서버나 전용 가속기를 장착한 서버에서 처리함으로써 단일 기기에서 이루어지거나, 말단 기기에서 수집된 데이터를 서버 측에서 모아서 처리하는 방식의 분산처리를 수행하는 방식이다.

[0004] 일반적으로 스마트폰, IoT 단말과 같은 기기들과 같이 성능이 낮은 기기에서는 메모리나 성능 상의 제약으로 인하여, 말단 기기에서 수집한 데이터를 서버에 전송하여 처리를 수행하였다. 최근에는 모델 자체의 크기를 감소시키기 위한 프루닝(pruning) 기법과 같이, 정확도 손실을 동반한 경량화 기법을 적용하여 수행하는 방식을 적용하기도 한다. 프루닝 기법은 신경망의 노드와 연결중에서 결과에 영향을 미치는 중요도가 낮은 노드와 연결을 제거하여, 모델 가중치의 크기를 줄여 모델을 경량화 하는 기법이다. 그러나 이러한 프루닝 기법은 모델의 가중치를 직접 제거하는 방법이기 때문에, 경량화가 진행될수록 모델의 정확도가 크게 하락한다.

[0005] IoT 기술의 발전 등으로 딥러닝에 기반한 지능적인 서비스를 스마트폰, IoT 단말과 같은 상대적으로 저전력, 저성능 기기에서 수행할 필요가 증대하고 있으며, 이들과 같이 성능이 낮은 기기를 다수 활용해 고성능 딥러닝 분류 모델을 분산 처리하는 연구 또한 진행되고 있다. 앞서 기술한 프루닝 기법을 이용한 경량화와 함께, 성능이 낮은 단말 기기에서 딥러닝을 수행하기 위한 다른 방식의 연구로 딥러닝 모델을 구성하는 각 레이어 내부 연산을, 여러 개의 단말 기기에 분할하여 처리하는 방식도 연구되고 있다.

[0006] 관련 선행문헌으로 대한민국 공개특허 제2015-0103447호가 있다.

발명의 내용

해결하려는 과제

[0008] 본 발명은 분류 결과의 정확도를 높이고, 분류 결과를 안정적으로 획득할 수 있는, 분산된 분류 모델을 이용하는 분류 방법을 제공하기 위한 것이다.

[0009] 또한 본 발명은 분류를 위한 연산을 분산시키고, 분류 모델의 경량화 정도를 증가시킬 수 있는, 분산된 분류 모델을 이용하는 분류 방법을 제공하기 위한 것이다.

과제의 해결 수단

[0011] 상기한 목적을 달성하기 위한 본 발명의 일 실시예에 따르면, 복수의 단말 각각에 분산된 서브 분류 모델을 통해 생성된, 타겟 데이터에 대한 분류 데이터를 수신하는 단계; 및 상기 분류 데이터를 이용하여, 상기 타겟 데이터에 대한 최종 클래스를 결정하는 단계를 포함하며, 미리 학습된 상기 서브 분류 모델 각각에 할당된 분류를 위한 클래스는, 복수의 타겟 클래스 중 적어도 하나를 포함하며, 상기 서브 분류 모델 각각에 할당된 클래스의 개수는, 상기 타겟 클래스의 개수보다 적은, 분산된 분류 모델을 이용하는 분류 방법이 제공된다.

[0012] 또한 상기한 목적을 달성하기 위한 본 발명의 다른 실시예에 따르면, 복수의 단말의 자원을 모니터링하는 단계; 및 상기 모니터링 결과에 따라서, 상기 단말에 서브 분류 모델을 할당하는 단계를 포함하며, 미리 학습된 상기 서브 분류 모델 각각에 할당된 분류를 위한 클래스는, 복수의 타겟 클래스 중 적어도 하나를 포함하며, 상기 서브 분류 모델 각각에 할당된 클래스의 개수는, 상기 타겟 클래스의 개수보다 적은, 분산된 분류 모델을 이용하는 분류 방법이 제공된다.

[0013] 또한 상기한 목적을 달성하기 위한 본 발명의 또 다른 실시예에 따르면, 복수의 후보 단말의 자원을 모니터링하는 단계; 및 상기 모니터링 결과에 따라서, 상기 후보 단말 중 서브 분류 모델을 이용하여 타겟 데이터를 분류하는 분류 단말을 결정하는 단계를 포함하며, 미리 학습된 상기 서브 분류 모델 각각에 할당된 분류를 위한 클래스는, 복수의 타겟 클래스 중 적어도 하나를 포함하며, 상기 서브 분류 모델 각각에 할당된 클래스의 개수는, 상기 타겟 클래스의 개수보다 적은, 분산된 분류 모델을 이용하는 분류 방법이 제공된다.

발명의 효과

[0015] 본 발명의 일 실시예에 따르면, 서브 분류 모델을 분산시켜 타겟 데이터를 분류함으로써, 분류 결과의 정확도를

높이고, 분류 결과를 안정적으로 획득할 수 있다.

[0016] 또한 본 발명의 일실시예에 따르면, 일부 서브 분류 모델이 할당된 단말에 장애가 발생하더라도 나머지 서브 분류 모델의 분류 데이터를 이용하여 분류를 수행할 수 있으므로, 보다 안정적인 분류 결과를 획득할 수 있다.

[0017] 또한 본 발명의 일실시예에 따르면, 요구되는 분류 정확도를 만족하는, 보다 경량화된 서브 분류 모델을 이용할 수 있으므로 분류에 소요되는 연산량을 줄일 수 있다.

도면의 간단한 설명

[0019] 도 1은 본 발명의 일실시예에 따른 분류 시스템을 설명하기 위한 도면이다.

도 2는 본 발명의 일실시예에 따른 서브 분류 모델을 설명하기 위한 도면이다.

도 3은 본 발명의 일실시예에 따른 분산된 모델을 이용하는 분류 방법을 설명하기 위한 도면이다.

도 4 및 도 5는 본 발명의 다른 실시예에 따른 서브 분류 모델을 설명하기 위한 도면이다.

도 6은 본 발명의 일실시예에 따른 서브 분류 모델을 이용하는 분류 방법과 메인 분류 모델을 이용하는 분류 방법의 정확도를 설명하기 위한 도면이다.

도 7은 본 발명의 다른 실시예에 따른 분산된 분류 모델을 이용하는 분류 방법을 설명하기 위한 도면이다.

도 8은 본 발명의 다른 실시예에 따른 분산된 분류 모델을 이용하는 분류 방법을 설명하기 위한 도면이다.

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. 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] 타겟 데이터를 특정 클래스로 분류하는 분류 모델의 연산량을 줄이기 위한 방법으로, 분류 모델의 연산을 분할하여 분산 처리를 하는 방법이나, 분류 모델을 프루닝 기법 등을 통해 경량화하는 방법 등이 이용된다.

[0023] 분류 모델의 연산을 분할하여 분산 처리하는 방법은, 예컨대 분산 모델의 레이어 각각의 연산을 서로 다른 장치에서 처리하는 방식으로 이루어질 수 있다. 따라서, 특정 레이어의 연산을 담당하는 장치에 고장이 발생하거나, 네트워크에 문제가 발생하면, 분류 모델의 결과를 획득할 수 없는 문제가 있다.

[0024] 또한 분류 모델을 경량화하는 방법은, 분류 결과의 정확도를 떨어뜨린다는 문제가 있다.

[0025] 이에 본 발명은, 분류 모델의 연산량을 줄이면서도 안정적으로 분류 결과를 획득할 수 있는 분산된 모델을 이용하는 분류 방법을 제안한다.

[0026] 본 발명의 일실시예에 따른 분류 모델은 기계 학습 기반의 분류 모델로서, 미리 학습된 인공 신경망일 수 있다. 또한 본 발명의 일실시예에 따른 분류 모델은 입력 데이터를 복수의 클래스 중 하나의 클래스로 분류하며, 입력 데이터의 특징이나 형태에 따라서 이용되는 신경망은 달라질 수 있다. 입력 데이터는 일실시예로서, 이미지일 수 있으며 이 경우, 컨벌루션 레이어가 포함된 신경망이 분류 모델로 이용될 수 있다.

[0027] 본 발명의 일실시예에 따른 분류 방법은, 프로세서 및 메모리를 포함하는 컴퓨팅 장치에서 수행될 수 있으며, 단말, 서버 등이 컴퓨팅 장치에 포함될 수 있다.

[0028] 이하에서, 본 발명에 따른 실시예들을 첨부된 도면을 참조하여 상세하게 설명한다.

[0030] 도 1은 본 발명의 일실시예에 따른 분류 시스템을 설명하기 위한 도면이며, 도 2는 본 발명의 일실시예에 따른 서브 분류 모델을 설명하기 위한 도면이다.

[0031] 도 1 및 도 2에서는 5개의 단말과 서브 분류 모델이 이용되는 실시예가 설명되나, 실시예에 따라서 분류에 이용

되는 단말과 서버 분류 모델의 개수는 달라질 수 있다.

- [0032] 도 1을 참조하면, 본 발명의 일실시예에 따른 분류 시스템은 복수의 단말 및 서버를 포함한다.
- [0033] 복수의 단말(111 내지 115)은 하나의 로컬 네트워크(110)에 포함될 수 있으나 이에 한정되는 것은 아니다. 단말(111 내지 115) 각각에는 서버 분류 모델이 분산되어 저장되며 일실시예로서, 서버 분류 모델은 서버(120)로부터 제공될 수 있다.
- [0034] 서버 분류 모델은 입력된 타겟 데이터 예컨대, 타겟 이미지에 대한 분류 데이터를 생성하며, 타겟 이미지를 미리 설정된 클래스 중 하나로 분류할 수 있다. 단말에 분산되는 분류 모델은, 타겟 이미지를 복수의 타겟 클래스 중 일부의 클래스로 분류하기 때문에, 서버 분류 모델로 불려질 수 있다.
- [0035] 일실시예로서, 서버 분류 모델 각각에 할당된 분류를 위한 클래스는, 미리 설정된 복수의 타겟 클래스 중 적어도 하나를 포함할 수 있으며, 서버 분류 모델 각각에 할당된 클래스의 개수는 타겟 클래스의 개수보다 적을 수 있다. 또한 실시예에 따라서, 서버 분류 모델 각각에 할당된 클래스는, 타겟 클래스 이외의 클래스인 기타 클래스를 포함할 수 있다. 기타 클래스는, 타겟 클래스 중 어디에도 포함되지 않는 기타의 클래스를 나타낸다.
- [0036] 예컨대 타겟 클래스가 A 내지 J로 총 10개일 경우, 도 2에 도시된 바와 같이, 제1 내지 제5서버 분류 모델(211 내지 215)에 분류를 위한 클래스가 할당될 수 있다. 제1서버 분류 모델(211)에는 A, B 및 기타 클래스(others class), 제2서버 분류 모델(212)에는 C, D 및 기타 클래스, 제3서버 분류 모델(213)에는 E, F 및 기타 클래스, 제4서버 분류 모델(214)에는 G, H 및 기타 클래스, 제5서버 분류 모델(215)에는 I, J 및 기타 클래스가 할당될 수 있다.
- [0037] 또한 서버 분류 모델은 일실시예로서 메인 분류 모델로부터 경량화된 모델일 수 있으며, 메인 분류 모델에 할당된 분류를 위한 클래스는 복수의 타겟 클래스를 포함할 수 있다. 다시 말해, 메인 분류 모델이 프루닝 기법 등을 통해 복수의 서버 분류 모델로 경량화되어 단말 각각에 할당될 수 있다.
- [0038] 분류 모델의 클래스는 학습 과정의 레이블(label)에 대응되는 것으로서, 예컨대, 메인 분류 모델은 A 내지 J 그리고 기타로 레이블링된 훈련 데이터를 통해 학습된 모델일 수 있다.
- [0039] 또는 다른 실시예로서, 서버 분류 모델은 메인 분류 모델과 무관하게 독립적으로 학습된 모델일 수 있다. 그리고 이러한 서버 분류 모델 역시 프루닝 기법 등을 통해 경량화될 수 있다.
- [0040] 도 2에 도시된 바와 같이, 서버 분류 모델 각각에 할당된 클래스는 서로 다른 타겟 클래스를 포함할 수 있으며, 실시예에 따라서 서버 분류 모델 각각에 할당되는 클래스는 달라질 수 있다.
- [0041] 서버 분류 모델들은, 전술된 바와 같이 인공 신경망이며, 서버 분류 모델 각각에 할당된 클래스에 대한 신뢰도 값(confidence value)을, 타겟 이미지에 대한 분류 데이터로 생성할 수 있다. 신뢰도값은 타겟 이미지가 해당 클래스일 확률에 대응되는 값이다.
- [0042] 다시 도 1로 돌아가, 단말(111 내지 115) 각각에 할당되는 서버 분류 모델은 단말 자원의 모니터링 결과에 따라서 결정될 수 있다. 이러한 모니터링은 복수의 단말(111 내지 115) 중 하나에 의해 수행되거나 또는 서버(120)에 의해 수행될 수 있다. 그리고 모니터링되는 자원은 일실시예로서, 단말의 연산량, 메모리 크기, 저장장치 용량, 네트워크 사용량 등일 수 있으며, 단말의 가용 자원량에 따라서 단말(111 내지 115) 각각에 할당되는 서버 분류 모델이 결정될 수 있다.
- [0043] 서버 분류 모델은 입력된 타겟 데이터 예컨대, 타겟 이미지에 대한 분류 데이터를 생성한다. 타겟 이미지는 복수의 단말(111 내지 115) 중 하나로부터 다른 단말로 전송될 수 있으며, 타겟 이미지를 전송하는 단말 역시 단말 자원의 모니터링 결과에 따라서 결정될 수 있다. 실시예에 따라서 타겟 이미지는 서버(120)로부터 복수의 단말(111 내지 115)로 제공될 수 있다.
- [0044] 복수의 단말(111 내지 115) 중 하나는 다른 단말에서 생성된 분류 데이터를 수신하여, 타겟 이미지에 대한 최종 클래스를 결정한다. 최종 클래스를 결정하는 단말은, 단말 자원의 모니터링 결과에 따라서 결정될 수 있으며, 실시예에 따라서 서버(120)가 분류 데이터를 수신하여 최종 클래스를 결정할 수 있다.
- [0045] 이와 같이 결정된 최종 클래스는 클라이언트(120)로 제공된다.
- [0047] 도 3은 본 발명의 일실시예에 따른 분산된 모델을 이용하는 분류 방법을 설명하기 위한 도면이며, 도 4 및 도 5

는 본 발명의 다른 실시예에 따른 서브 분류 모델을 설명하기 위한 도면이다.

[0048] 본 발명의 일실시예에 따른 컴퓨팅 장치는, 복수의 단말 각각에 분산된 서브 분류 모델을 통해 생성된, 타겟 이미지에 대한 분류 데이터를 수신(S310)한다. 컴퓨팅 장치 역시 컴퓨팅 장치에 할당된 서브 분류 모델을 통해 분류 데이터를 생성할 수 있다.

[0049] 서브 분류 모델에 할당된 클래스는 도 2와 같이, 서로 다른 타겟 클래스를 포함하거나, 또는 도 4에 도시된 바와 같이, 서브 분류 모델 중 제1 및 제2서브 분류 모델 각각에 할당된 클래스는, 적어도 하나의 중복된 타겟 클래스를 포함할 수 있다. 또는 서브 분류 모델에 할당된 클래스의 개수는 서로 다를 수 있다.

[0050] 한편, 서브 분류 모델에 할당된 클래스는 타겟 클래스의 상위 개념에 따라서 결정될 수 있다. 상위 개념은, 클래스간 서로 공유되는 특징과, 클래스간 서로 구분되는 특징에 기반하여 결정될 수 있다.

[0051] 예컨대, A 내지 J 클래스의 일예가 [표 1]과 같을 경우, 이러한 클래스는 [표 1]과 같은 상위 개념 그룹으로 그룹핑될 수 있다. 그리고, 도 5에 도시된 바와 같이, 제1 내지 제5서브 분류 모델에 할당되는 클래스는 각각 상위 개념에 따라서 각각 (비행기), (자동차와 트럭), (배), (새), 그리고 (고양이와 사슴과 개와 개구리와 말)을 포함할 수 있다.

표 1

클래스	클래스의 일예	상위 개념
A	비행기	날개를 이용하는 운송 수단
B	자동차	
C	트럭	
D	배	물을 이용하는 운송수단
E	새	날개가 있는 동물
F	고양이	
G	사슴	
H	개	
I	개구리	
J	말	

[0053] 서브 분류 모델은 타겟 이미지를 할당된 클래스 중 하나로 분류하며, 단말에서 이러한 분류 결과를 통해 생성되어 전송되는 분류 데이터는 서브 분류 모델에 할당된 클래스에 대한 신뢰도값을 포함할 수 있다. 일실시예로서, 분류 데이터는 서브 분류 모델 각각에 할당된 클래스 전체에 대한 신뢰도값을 포함하거나, 최종 클래스가 기타 클래스로 결정되는 것은 큰 의미가 없으므로, 기타 클래스는 제외하고 타겟 클래스에 대한 신뢰도값을 포함할 수 있다. 예컨대, 도 5와 같은 서브 분류 모델이 이용될 경우, 분류 데이터는 기타 클래스에 대한 신뢰도값을 포함하지 않고, A 내지 J 클래스에 대한 신뢰도값을 포함할 수 있다.

[0054] 또는 분류 데이터는 서브 분류 모델에 할당된 클래스에 대한 신뢰도값 중 최대값을 포함할 수 있다. 예컨대, 도 4와 같은 제1 내지 제5서브 분류 모델 각각에서의 최대 신뢰도값이 각각 A, C, D, G 및 I 클래스라면 분류 데이터에는 A, C, D, G 및 I 클래스에 대한 신뢰도값이 포함될 수 있다.

[0055] 본 발명의 일실시예에 따른 컴퓨팅 장치는, 이와 같은 분류 데이터를 이용하여, 타겟 데이터에 대한 최종 클래스를 결정(S320)한다.

[0056] 일실시예로서, 컴퓨팅 장치는 신뢰도값 중 최대값에 대응되는 클래스를 타겟 이미지에 대한 최종 클래스로 결정할 수 있다. 예컨대, A 내지 J 클래스에 대한 신뢰도값 중 A에 대한 신뢰도값이 가장 클 경우, 컴퓨팅 장치는 타겟 이미지의 클래스를 A로 결정할 수 있다.

[0057] 한편, 도 4와 같이 서브 분류 모델에 중복된 타겟 클래스가 할당된 경우에는, 서로 다른 서브 분류 모델의 동일한 타겟 클래스에 대한 신뢰도값이 최대값인 경우가 발생할 수 있다. 예컨대, 제1 및 제2서브 분류 모델의 분류 결과, 모두 B 클래스에 대한 신뢰도값이 최대값일 수 있다. 이 경우 컴퓨팅 장치는 B 클래스에 대한 2개의 신뢰도값의 평균값과, 다른 클래스의 신뢰도값 중 최대값에 대응되는 클래스를 최종 클래스로 결정할 수 있다. 또는 B 클래스에 대한 2개의 신뢰도값 중 최대값과, 다른 클래스의 신뢰도값 중에서 가장 큰 신뢰도값에 대응되는 클래스를 최종 클래스로 결정할 수 있다.

[0058] 다른 실시예로서, 컴퓨팅 장치는 신뢰도값과 임계값을 비교하여 최종 클래스를 결정할 수 있다. 예컨대, A 내지

J 클래스에 대한 신뢰도값 중 A에 대한 신뢰도값이 가장 크며, A에 대한 신뢰도값이 임계값 이상이면, 컴퓨팅 장치는 A를 최종 클래스로 결정하지만, A에 대한 신뢰도값이 임계값보다 작으면, 기타 클래스를 최종 클래스로 결정할 수 있다.

- [0059] 또는 서브 분류 모델 중 일부 모델에 할당된 클래스에 대한 신뢰도값이 분류 데이터에 포함되며, 이러한 신뢰도값이 임계값보다 작은 경우, 컴퓨팅 장치는 분류 데이터에 포함되지 않은 타겟 클래스를 최종 클래스로 결정할 수 있다. 예컨대, 도 4와 같은 서브 분류 모델이 이용되며, 컴퓨팅 장치가 제1 및 제2서브 분류 모델의 분류 데이터를 수신하지 못한 상황에서, D 내지 J에 대한 신뢰도값이 모두 임계값 보다 작다면, 컴퓨팅 장치는 A 내지 C 중 하나를 최종 클래스로 결정할 수 있다. D 내지 J에 대한 신뢰도값 중 D에 대한 신뢰도값이 가장 크며 임계값 이상이라면, 컴퓨팅 장치는 D를 최종 클래스로 결정할 수 있다.
- [0060] 제1 및 제2서브 분류 모델의 분류 데이터는 통신 장애나 또는 제1 및 제2서브 분류 모델이 할당된 단말의 장애로 인해 컴퓨팅 장치로 수신되지 못할 수 있는데, 본 발명의 일실시예에 따른 분류 방법은 분산된 분류 모델을 이용함으로써, 이러한 장애 상황에서도 타겟 이미지에 대한 클래스를 결정할 수 있다.
- [0061] 최종 클래스가 결정된 타겟 데이터는, 최종 클래스가 할당된 서브 분류 모델의 학습에 이용될 수 있다.
- [0063] 도 6은 본 발명의 일실시예에 따른 서브 분류 모델을 이용하는 분류 방법과 메인 분류 모델을 이용하는 분류 방법의 정확도를 설명하기 위한 도면이다.
- [0064] 도 6에서는 복수의 타겟 클래스가 할당된 하나의 메인 분류 모델(610), 복수의 타겟 클래스가 분할되어 할당된 2개(620), 5개(630) 및 10개(640)의 서브 분류 모델에 대한 분류 정확도가 도시되며, 여기서 분류 정확도는 메인 분류 모델 및 서브 분류 모델의 경량화 정도에 따른 분류 정확도이다.
- [0065] 도 6에 도시된 바와 같이, 경량화 정도가 70에 이를 때까지는 정확도의 차이가 크지 않으나, 70이 넘어가면서 메인 분류 모델의 정확도가 급격히 낮아짐을 알 수 있다. 그리고 보다 많은 개수의 서브 분류 모델이 이용될수록, 보다 높은 경량화가 이루어지더라도 정확도가 유지됨을 알 수 있다.
- [0066] 따라서, 본 발명의 일실시예에 따르면, 요구되는 분류 정확도를 만족하는, 보다 경량화된 서브 분류 모델을 이용할 수 있으므로 분류에 소요되는 연산량을 줄일 수 있다.
- [0067] 또한 전술된 바와 같이, 본 발명의 일실시예에 따르면, 일부 서브 분류 모델이 할당된 단말에 장애가 발생하더라도 나머지 서브 분류 모델의 분류 데이터를 이용하여 분류를 수행할 수 있으므로, 보다 안정적인 분류 결과를 획득할 수 있다.
- [0069] 도 7은 본 발명의 다른 실시예에 따른 분산된 분류 모델을 이용하는 분류 방법을 설명하기 위한 도면이다.
- [0070] 도 7을 참조하면, 본 발명의 일실시예에 따른 컴퓨팅 장치는 복수의 단말의 자원을 모니터링(S710)하고, 모니터링 결과에 따라서, 단말에 서브 분류 모델을 할당(S720)한다.
- [0071] 일실시예로서, 컴퓨팅 장치는 단말의 가용 자원에 따라서 서브 분류 모델을 단말에 할당할 수 있으며, 서브 분류 모델 각각에 할당된 클래스의 개수는 단말의 가용 자원에 따라서 결정될 수 있다. 가용 자원이 큰 단말에 할당된 서브 분류 모델은, 가용 자원이 적은 단말보다, 많은 클래스가 할당된 모델일 수 있다.
- [0072] 예컨대, 도 5와 같이 서브 분류 모델이 이용되며, 제1 내지 제5단말 중 제1단말의 가용 자원이 가장 크다면, 제1단말에 제5서브 분류 모델이 할당될 수 있다. 그리고 제2단말의 가용 자원이 가장 작다면, 제2단말에는 제1, 제3 및 제4서브 분류 모델 중 하나가 할당될 수 있다.
- [0074] 도 8은 본 발명의 다른 실시예에 따른 분산된 분류 모델을 이용하는 분류 방법을 설명하기 위한 도면이다.
- [0075] 도 8을 참조하면, 본 발명의 일실시예에 따른 컴퓨팅 장치는 복수의 후보 단말의 자원을 모니터링(S810)하고, 모니터링 결과에 따라서, 후보 단말 중 서브 분류 모델을 이용하여 타겟 데이터를 분류하는 분류 단말(S820)을 결정할 수 있다.
- [0076] 단계 S820에서 컴퓨팅 장치는 단말의 가용 자원에 따라서 분류 단말을 결정할 수 있으며, 가용 자원이 많은 순

서대로 분류 단말을 결정할 수 있다.

[0077] 한편, 단말을 통해 타겟 데이터가 전송되는 경우, 본 발명의 일실시예에 따른 컴퓨팅 장치는 모니터링 결과에 따라서, 후보 단말 중 타겟 데이터를 분류 단말로 전송하는 전송 단말을 결정(S830)할 수 있다. 가용 자원이 높은 후보 단말 중 하나가 전송 단말로 결정될 수 있으며, 예컨대 분류 단말 중 하나가 전송 단말로 결정되거나 타겟 데이터를 분류하지 않는 단말이 전송 단말로 결정될 수 있다.

[0078] 또한 단말을 통해 타겟 데이터에 대한 최종 클래스가 결정되는 경우, 본 발명의 일실시예에 따른 컴퓨팅 장치는 모니터링 결과에 따라서, 후보 단말 중 타겟 데이터에 대한 최종 클래스를 결정하는 클래스 결정 단말을 결정(S840)할 수 있다. 클래스 결정 단말은, 분류 단말 각각에 분산된 서브 분류 모델을 통해 생성된, 타겟 데이터에 대한 분류 데이터를 수신하여 최종 클래스를 결정하며, 가용 자원이 높은 후보 단말 중 하나가 클래스 결정 단말로 결정될 수 있다.

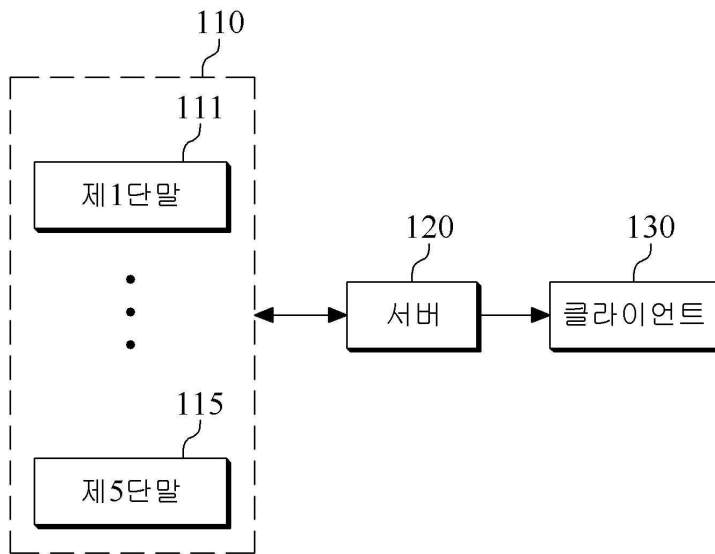
[0079] 단계 S820, S830 및 S840은 모니터링 결과에 기반하여 병렬적으로 그리고 선택적으로 수행될 수 있다.

[0081] 앞서 설명한 기술적 내용들은 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 실시예들을 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. 하드웨어 장치는 실시예들의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

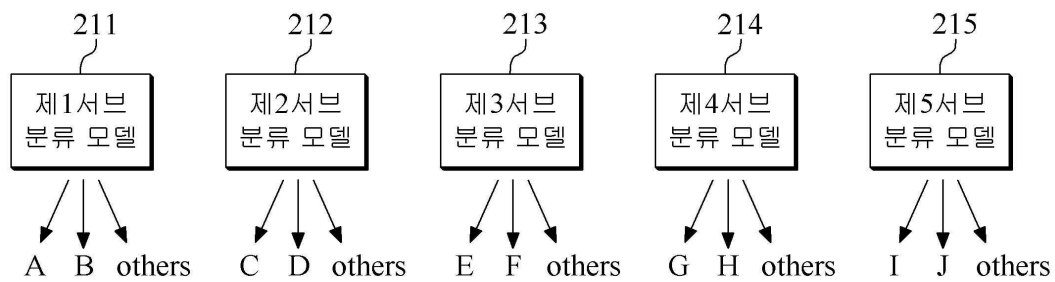
[0083] 이상과 같이 본 발명에서는 구체적인 구성 요소 등과 같은 특정 사항들과 한정된 실시예 및 도면에 의해 설명되었으나 이는 본 발명의 보다 전반적인 이해를 돕기 위해서 제공된 것일 뿐, 본 발명은 상기의 실시예에 한정되는 것은 아니며, 본 발명이 속하는 분야에서 통상적인 지식을 가진 자라면 이러한 기재로부터 다양한 수정 및 변형이 가능하다. 따라서, 본 발명의 사상은 설명된 실시예에 국한되어 정해져서는 아니되며, 후술하는 특허청구범위뿐 아니라 이 특허청구범위와 균등하거나 등가적 변형이 있는 모든 것들은 본 발명 사상의 범주에 속한다고 할 것이다.

도면

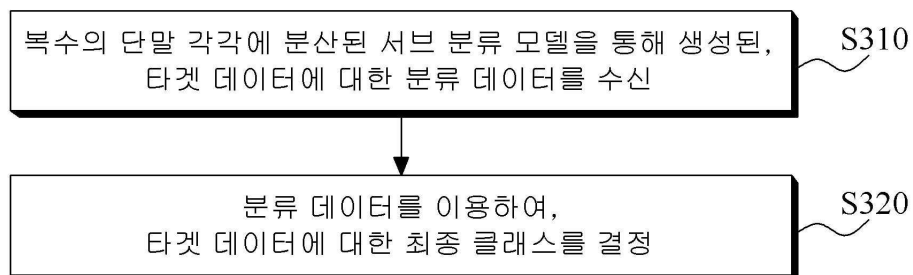
도면1



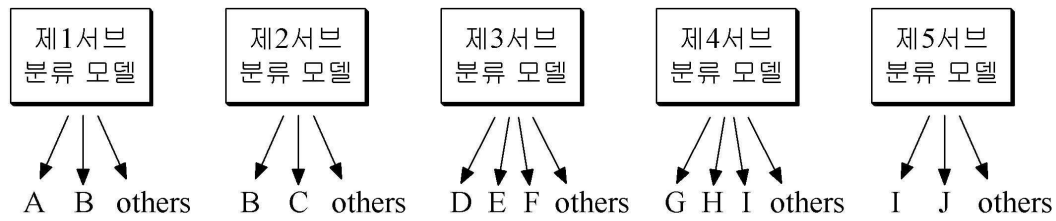
도면2



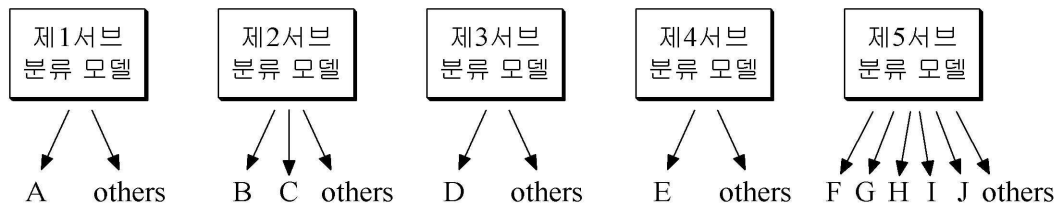
도면3



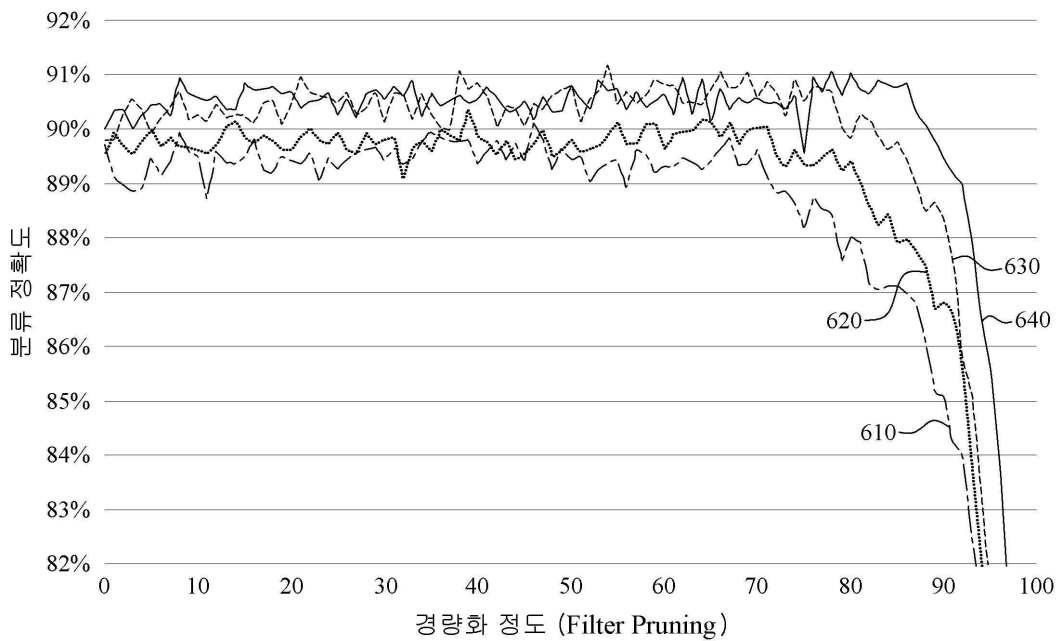
도면4



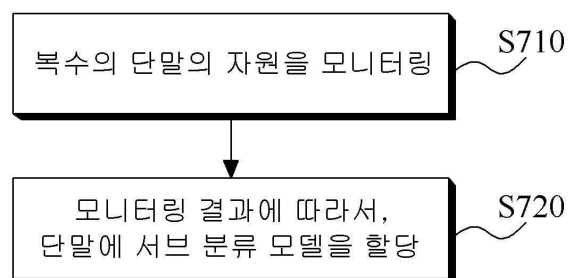
도면5



도면6



도면7



도면8

