(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2022년02월18일
(11) 등록번호 10-2365658
(24) 등록일자 2022년02월16일

(51) 국제특허분류(Int. Cl.)
H04L 47/00 (2022.01) *G06N 20/20* (2019.01)
H04L 43/00 (2022.01)
(52) CPC특허분류
H04L 47/2441 (2013.01)
G06N 20/20 (2021.08)
(21) 출원번호 10-2019-0160530
(22) 출원일자 2019년12월05일
심사청구일자 2020년09월10일
(65) 공개번호 10-2021-0070597
(43) 공개일자 2021년06월15일
(56) 선행기술조사문헌
KR1020100125076 A*
KR1020190048004 A*
KR1020190117724 A*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
포항공과대학교 산학협력단
경상북도 포항시 남구 청암로 77 (지곡동)
(72) 발명자
홍원기
경상북도 포항시 남구 지곡로 319, 328동 304호
유재형
서울특별시 송파구 올림픽로 135, 211동 1303호
홍지범
경기도 안산시 상록구 삼리로 24, 101동 805호
(74) 대리인
특허법인이상

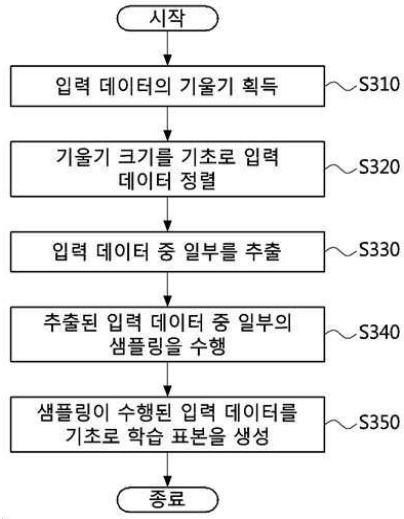
전체 청구항 수 : 총 18 항 심사관 : 김대성

(54) 발명의 명칭 **트래픽 분류 방법 및 장치**

(57) 요 약

트래픽 분류 방법 및 장치가 개시된다. 본 발명의 일 실시예에 따른 트래픽 분류 방법은 플로우에 대한 정보를 포함하는 플로우 데이터를 수신하는 단계, 상기 플로우 데이터에 대한 스케일링을 수행하는 단계, 상관 관계를 기초로 상기 스케일링이 수행된 플로우 데이터 중 중복되는 데이터를 제거하여 입력 데이터를 생성하는 단계 및 상기 입력 데이터를 기초로 네트워크 트래픽을 분류하는 단계를 포함할 수 있다.

대 표 도 - 도3



(52) CPC특허분류
H04L 43/026 (2013.01)
H04L 43/50 (2013.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 1711090692
부처명 과학기술정보통신부
과제관리(전문)기관명 포항공과대학교 산학협력단
연구사업명 방송통신산업기술개발
연구과제명 인공지능 기반 가상 네트워크 관리기술 개발
기 여 율 1/2
과제수행기관명 포항공과대학교 산학협력단
연구기간 2019.01.01 ~ 2019.12.31
이 발명을 지원한 국가연구개발사업
과제고유번호 1711093074
부처명 과학기술정보통신부
과제관리(전문)기관명 정보통신기술진흥센터
연구사업명 방송통신산업기술개발
연구과제명 글로벌 SDN/NFV 공개소프트웨어 핵심 모듈/기능 개발
기 여 율 1/2
과제수행기관명 광주과학기술원
연구기간 2019.01.01 ~ 2019.12.31

명 세 서

청구범위

청구항 1

SDN(software defined networking) 컨트롤러에 연결되며 프로세서를 구비하는 트래픽 분류 장치에 의해 수행되는 트래픽 분류 방법으로서,

네트워크의 플로우에 대한 정보를 포함하는 플로우 데이터를 수신하는 단계-여기서 상기 플로우 데이터는 상기 SDN 컨트롤러에 의해 트래픽 정보의 레이블링이 수행된 것임-;

상기 플로우 데이터에 대한 전처리에서 상기 플로우 데이터에 대한 스케일링을 수행하는 단계;

상기 전처리에서 상기 스케일링이 수행된 플로우 데이터 중 중복되는 데이터를 상기 플로우 데이터 간의 상관 관계를 기초로 제거하여 입력 데이터를 생성하는 단계; 및

상기 입력 데이터를 기초로 기계 학습을 통해 네트워크 트래픽을 분류하는 단계를 포함하며,

상기 기계 학습을 수행하기 위해, 상기 입력 데이터 각각의 기울기를 획득하는 단계; 상기 기울기의 크기를 기초로 상기 입력 데이터를 정렬하는 단계; 상기 정렬된 입력 데이터 중 일부를 미리 설정한 비율에 따라 추출하는 단계; 상기 추출된 입력 데이터 중 일부를 샘플링하는 단계; 및 상기 샘플링이 수행된 입력 데이터를 기초로 상기 학습 표본을 추출하는 단계를 더 포함하는 트래픽 분류 방법.

청구항 2

청구항 1에 있어서,

상기 입력 데이터를 생성하는 단계는, 주성분 분석 방식을 사용하여 획득한 상기 상관 관계를 기초로 상기 플로우 데이터를 분류하고 분류된 플로우 데이터를 플로우 특성 별로 묶는 것을 포함하는, 트래픽 분류 방법.

청구항 3

청구항 2에 있어서,

상기 플로우 특성은, 플로우가 도달한 시각에 대한 플로우 특성(time stamp), 플로우의 시작 위치에 대한 플로우 특성(source ip, source port), 플로우의 목적지에 대한 플로우 특성(destination IP, destination port), 플로우가 사용하는 프로토콜에 대한 플로우 특성, 데이터 패킷들의 번호(packet count)에 대한 플로우 특성, 데이터 패킷들의 바이트 번호(byte count)에 대한 플로우 특성, 데이터 패킷들이 도달하는 시간 사이의 간격(inter-arrival time)에 대한 플로우 특성 중 하나 이상을 포함하는, 트래픽 분류 방법.

청구항 4

청구항 1에 있어서,

상기 입력 데이터를 생성하는 단계 후에, GOSS(Gradient-based One-Side Sampling) 방식을 사용하여 상기 기계 학습을 수행하기 위한 학습 표본을 생성하는 단계를 더 포함하고,

상기 GOSS 방식은 기울기를 기반으로 하며, 상기 기울기는 상기 입력 데이터에 대한 손실함수의 기울기이고, 상기 손실함수는 상기 입력 데이터에 대한 출력 예상 값 및 실체 출력 값 사이의 차이에 관한 함수를 포함하는, 트래픽 분류 방법.

청구항 5

청구항 1에 있어서,

상기 기계 학습은, 앙상블 학습 기반으로 분류 시스템을 통해 분류자를 학습하고, 학습된 분류자를 통해 상기 트래픽을 분류하는 것을 포함하는, 트래픽 분류 방법.

청구항 6

청구항 1에 있어서,

상기 기계 학습은,

상기 추출된 학습 표본을 기초로 학습 데이터를 생성하는 단계; 및

상기 학습 데이터를 기초로 기계 학습을 수행하는 단계를 더 포함하는, 트래픽 분류 방법.

청구항 7

청구항 1에 있어서,

상기 추출된 입력 데이터 중 일부를 샘플링하는 단계는, 상기 입력 데이터 중 상기 기울기의 크기가 상위 a%에 해당하는 입력 데이터 및 상기 기울기의 크기가 하위 b%에 해당하는 입력 데이터 중 상기 하위 b%에 해당하는 입력 데이터를 샘플링하는 것을 포함하고,

상기 트래픽 분류 방법은, 상기 기울기의 크기가 상위 a%에 해당하는 입력 데이터 및 상기 기울기의 크기가 하위 b%에 해당하는 입력 데이터를 기초로 정보 이득을 연산하는 단계를 더 포함하며,

상기 기울기의 크기가 하위 b%에 해당하는 입력 데이터는 소정 가중치만큼 증폭된 입력 데이터이고, 상기 정보 이득은 분산으로 측정되고, 상기 분산은 상기 상관 관계를 기초로 상기 플로우 데이터를 분류하여 얻은 플로우 특성으로 정의되는 노드에서의 상기 입력 데이터에 대한 분산인, 트래픽 분류 방법.

청구항 8

청구항 6에 있어서,

상기 학습 데이터를 생성하는 단계는,

상기 학습 표본에 대한 번들링을 수행하는 단계; 및

상기 번들링이 수행된 학습 표본을 병합하여 상기 학습 데이터를 생성하는 단계를 포함하며,

상기 번들링을 수행하는 단계는 그래프를 기초로 상기 학습 표본의 번들링을 수행하고,

상기 그래프는, 상기 상관 관계를 기초로 상기 플로우 데이터를 분류하여 얻은 플로우 특성이 상호간에 0이 아닌 값을 동시에 갖지 않는 경우를 상호 배타적인 것으로 판단할 때, 상호 배타적이지 않은 노드들 사이에 엣지를 연결하여 상기 학습 표본의 표본들 사이의 관계를 나타내는 그래프를 포함하는, 트래픽 분류 방법.

청구항 9

청구항 6에 있어서,

상기 기계 학습을 수행하는 단계는,

상기 학습 데이터를 기초로 상기 기계 학습을 수행하기 위한 학습 모델을 생성하기 위해, 상기 학습 데이터를 트레이닝 셋, 검증 셋 및 테스트 셋으로 분류하는 단계;

상기 트레이닝 셋을 기초로 미리 준비된 학습 모델의 가중치를 획득하는 단계;

상기 검증 셋을 통해 상기 학습 모델을 검증하는 단계; 및

상기 테스트 셋을 통해 상기 검증된 학습 모델을 테스트하는 단계를 포함하는, 트래픽 분류 방법.

청구항 10

프로세서(processor); 및

상기 프로세서에 의해 실행되는 하나 이상의 명령들이 저장된 메모리(memory)를 포함하며,

상기 하나 이상의 명령들에 의해 상기 프로세서는,

SDN(Software Defined Networking) 컨트롤러로부터 플로우에 대한 정보를 포함하는 플로우 데이터를 수신하는

단계;

상기 플로우 데이터에 대한 전처리에서 상기 플로우 데이터에 대한 스케일링을 수행하는 단계;

상기 전처리에서 상기 스케일링이 수행된 플로우 데이터 중 중복되는 데이터를 상기 플로우 데이터 간의 상관 관계를 기초로 제거하여 입력 데이터를 생성하는 단계;

상기 입력 데이터를 기초로 기계 학습을 통해 네트워크 트래픽을 분류하는 단계를 수행하며,

여기서 상기 플로우 데이터는 상기 SDN 컨트롤러에 의해 트래픽 정보의 레이블링이 수행된 것이고,

상기 프로세서는, 상기 기계 학습을 위해, 상기 입력 데이터 각각의 기울기를 획득하는 단계; 상기 기울기의 크기를 기초로 상기 입력 데이터를 정렬하는 단계; 상기 정렬된 입력 데이터 중 일부를 미리 설정한 비율에 따라 추출하는 단계; 상기 추출된 입력 데이터 중 일부를 샘플링하는 단계; 및 상기 샘플링이 수행된 입력 데이터를 기초로 상기 학습 표본을 추출하는 단계를 더 실행하는, 트래픽 분류 장치.

청구항 11

청구항 10에 있어서,

상기 프로세서는, 상기 입력 데이터의 생성하는 단계에서, 주성분 분석 방식을 사용하여 획득한 상기 상관 관계를 기초로 상기 플로우 데이터를 분류하고 분류된 플로우 데이터를 플로우 특성별로 묶는, 트래픽 분류 장치.

청구항 12

청구항 11에 있어서,

상기 플로우 특성은, 플로우가 도달한 시각에 대한 플로우 특성(time stamp), 플로우의 시작 위치에 대한 플로우 특성(source ip, source port), 플로우의 목적지에 대한 플로우 특성(destination IP, destination port), 플로우가 사용하는 프로토콜에 대한 플로우 특성, 데이터 패킷들의 번호(packet count)에 대한 플로우 특성, 데이터 패킷들의 바이트 번호(byte count)에 대한 플로우 특성, 데이터 패킷들이 도달하는 시간 사이의 간격(inter-arrival time)에 대한 플로우 특성 중 하나 이상을 포함하는, 트래픽 분류 장치.

청구항 13

청구항 10에 있어서,

상기 프로세서는, 상기 입력 데이터를 생성하는 단계 후에, GOSS(Gradient-based One-Side Sampling) 방식을 사용하여 상기 기계 학습을 수행하기 위한 학습 표본을 생성하는 단계를 더 실행하도록 구성되고,

상기 GOSS 방식은 기울기를 기반으로 하며, 상기 기울기는 상기 입력 데이터에 대한 손실함수의 기울기이고, 상기 손실함수는 상기 입력 데이터에 대한 출력 예상 값 및 실체 출력 값 사이의 차이에 관한 함수를 포함하는, 트래픽 분류 장치.

청구항 14

청구항 10에 있어서,

상기 프로세서는, 상기 기계 학습에서, 앙상블 학습 기반으로 분류 시스템을 통해 분류자를 학습하고, 학습된 분류자를 통해 상기 트래픽을 분류하는, 트래픽 분류 장치.

청구항 15

청구항 10에 있어서,

상기 프로세서는, 상기 기계 학습에서,

상기 추출된 학습 표본을 기초로 학습 데이터를 생성하는 단계; 및

상기 학습 데이터를 기초로 기계 학습을 수행하는 단계를 더 포함하는, 트래픽 분류 장치.

청구항 16

청구항 10에 있어서,

상기 프로세서는, 상기 추출된 입력 데이터 중 일부를 샘플링하는 단계에서, 상기 입력 데이터 중 상기 기울기의 크기가 상위 a%에 해당하는 입력 데이터 및 상기 기울기의 크기가 하위 b%에 해당하는 입력 데이터 중 상기 하위 b%에 해당하는 입력 데이터를 샘플링하고,

상기 샘플링하는 단계 후에, 상기 기울기의 크기가 상위 a%에 해당하는 입력 데이터 및 상기 기울기의 크기가 하위 b%에 해당하는 입력 데이터를 기초로 정보 이득을 연산하는 단계를 더 실행하며,

여기서 상기 기울기의 크기가 하위 b%에 해당하는 입력 데이터는 소정 가중치만큼 증폭된 입력 데이터이고, 상기 정보 이득은 분산으로 측정되고, 상기 분산은 상기 상관 관계를 기초로 상기 플로우 데이터를 분류하여 얻은 플로우 특성으로 정의되는 노드에서의 상기 입력 데이터에 대한 분산인, 트래픽 분류 장치.

청구항 17

청구항 15에 있어서,

상기 프로세서는, 상기 학습 데이터를 생성하는 단계에서,

상기 학습 표본에 대한 번들링을 수행하는 단계; 및

상기 번들링이 수행된 학습 표본을 병합하여 상기 학습 데이터를 생성하는 단계를 실행하며,

여기서 상기 번들링의 수행은, 그래프를 기초로 상기 학습 표본의 번들링을 수행하는 것을 포함하고,

상기 그래프는, 상기 상관 관계를 기초로 상기 플로우 데이터를 분류하여 얻은 플로우 특성이 상호간에 0이 아닌 값을 동시에 갖지 않는 경우를 상호 배타적인 것으로 판단할 때, 상호 배타적이지 않은 노드들 사이에 엣지를 연결하여 상기 학습 표본의 표본들 사이의 관계를 나타내는 그래프를 포함하는 트래픽 분류 장치.

청구항 18

청구항 15에 있어서,

상기 프로세서는, 상기 기계 학습을 수행하는 단계에서,

상기 기계 학습을 수행하기 위한 학습 모델을 생성하는 단계;

상기 학습 데이터를 트레이닝 셋, 검증 셋 및 테스트 셋으로 분류하는 단계;

상기 트레이닝 셋을 기초로 상기 학습 모델의 가중치를 획득하는 단계;

상기 검증 셋을 통해 상기 학습 모델을 검증하는 단계; 및

상기 테스트 셋을 통해 상기 검증된 학습 모델을 테스트하는 단계를 실행하는 트래픽 분류 장치.

발명의 설명

기 술 분 야

[0001] 본 발명은, 트래픽 분류 방법 및 장치에 관한 것으로 더욱 상세하게는 기계 학습을 이용한 트래픽 분류 방법 및 장치에 관한 것이다.

배 경 기 술

[0002] 네트워크 트래픽 분류는 네트워크에서 사용되는 여러 응용 서비스와 프로토콜을 식별하기 위한 것일 수 있다. 네트워크 트래픽 분류는 네트워크 관리 및 보안의 중요한 요소 중 하나일 수 있다.

[0003] 예를 들어 네트워크 트래픽 분류는 QoS(Quality of Service) 제어를 위한 메커니즘은 제한된 대역폭(bandwidth)에서 서로 다른 응용 서비스의 처리 우선 순위를 결정하기 위해 사용될 수 있다.

[0004] 네트워크 트래픽 분류 방법에는 포트 기반(port-based) 트래픽 분류 방법, 페이로드 기반 (payload-based) 트래픽 분류 방법 및 플로우 통계 정보 기반(flow statistics-based) 트래픽 분류 방법이 있을 수 있다.

[0005] 포트 기반 트래픽 분류 방법은 응용 프로그램에서 사용하는 표준 포트 번호를 확인하여 트래픽을 분류할 수 있

다. 다만, 모든 응용 서비스가 표준 포트 번호를 사용하는 것은 아니기 때문에 포트 기반 트래픽 분류 방법은 트래픽 분류에 실패하는 문제점이 발생할 수 있다.

[0006] 페이로드 기반 트래픽 분류 방법은 IP(Internet Protocol) 패킷의 페이로드에 포함되어 있는 응용 서비스의 시그니처 (signature)를 확인하여 트래픽을 분류할 수 있다. 다만, 페이로드 기반 트래픽 분류 방법은 암호화된 트래픽이 점차 증가함에 따라 트래픽 분류에 실패하는 경우가 빈번해지는 문제점이 발생할 수 있다.

[0007] 플로우 통계 정보 기반 네트워크 트래픽 분류 방법은 네트워크 플로우의 플로우 특성을 분석하여 트래픽을 분류할 수 있다. 다만, 플로우 통계 정보 기반 네트워크 트래픽 분류 방법은 새로운 응용 서비스의 트래픽이 발생하고 트래픽 혼잡도가 증가하는 경우 네트워크 트래픽을 정확하게 분류하지 못하는 문제점이 발생할 수 있다.

발명의 내용

해결하려는 과제

[0008] 상기와 같은 문제점을 해결하기 위한 본 발명의 목적은, 앙상블 학습 기법을 사용하는 트래픽 분류 방법 및 장치를 제공하는 데 있다.

[0009] 상기와 같은 문제점을 해결하기 위한 본 발명의 다른 목적은, 복잡한 네트워크 환경에서 높은 분류 정확도를 가지는 트래픽 분류 방법 및 장치를 제공하는 데 있다.

과제의 해결 수단

[0010] 상기와 같은 문제점을 해결하기 위한 본 발명의 일 실시예에 따른 트래픽 분류 방법은 플로우에 대한 정보를 포함하는 플로우 데이터를 수신하는 단계, 상기 플로우 데이터에 대한 스케일링을 수행하는 단계, 상기 플로우 데이터 간의 상관 관계를 기초로 상기 스케일링이 수행된 플로우 데이터 중 중복되는 데이터를 제거하여 입력 데이터를 생성하는 단계 및 상기 입력 데이터를 기초로 네트워크 트래픽을 분류하는 단계를 포함할 수 있다.

[0011] 여기서, 상기 상관 관계는, 주성분 분석 방식을 사용하여 획득할 수 있다.

[0012] 여기서, 상기 입력 데이터는 플로우 특성을 포함할 수 있고, 상기 플로우 특성은, 상기 플로우가 도달한 시각에 대한 정보, 상기 플로우의 시작 위치에 대한 정보, 상기 플로우의 목적지에 대한 정보 중 하나 이상을 포함할 수 있다.

[0013] 여기서, 상기 네트워크 트래픽을 분류하는 단계는, 미리 기계 학습이 수행된 트래픽 분류 장치에 의해 네트워크 트래픽을 분류하는 단계를 포함할 수 있다.

[0014] 여기서, 상기 기계 학습은, 앙상블 학습 기법을 사용하여 수행될 수 있다.

[0015] 여기서, 상기 기계 학습은, 레이블링이 수행된 플로우 데이터를 기초로 입력 데이터를 생성하는 단계, 상기 입력 데이터를 기초로 학습 표본을 추출하는 단계, 상기 추출된 학습 표본을 기초로 학습 데이터를 생성하는 단계 및 상기 학습 데이터를 기초로 상기 기계 학습을 수행하는 단계를 통해 수행될 수 있다.

[0016] 여기서, 상기 학습 표본을 추출하는 단계는, 상기 입력 데이터 각각의 기울기를 획득하는 단계, 상기 입력 데이터의 기울기 크기를 기초로 상기 입력 데이터를 정렬하는 단계, 상기 입력 데이터 중 일부를 미리 설정한 비율에 따라 추출하는 단계 및 상기 샘플링이 수행된 입력 데이터를 기초로 상기 학습 표본을 추출하는 단계를 포함할 수 있다.

[0017] 여기서, 상기 학습 데이터를 생성하는 단계는 상기 학습 표본에 대한 번들링을 수행하는 단계, 상기 번들링이 수행된 학습 표본을 병합하여 상기 학습 데이터를 생성하는 단계를 포함할 수 있다.

[0018] 여기서, 상기 기계 학습을 수행하는 단계는, 상기 기계 학습을 수행하기 위한 학습 모델을 생성하는 단계, 상기 학습 데이터를 트레이닝 셋, 검증 셋 및 테스트 셋으로 분류하는 단계, 상기 트레이닝 셋을 기초로 상기 학습 모델의 가중치를 획득하는 단계, 상기 검증 셋을 통해 상기 학습 모델을 검증하는 단계 및 상기 테스트 셋을 통해 상기 검증된 학습 모델을 테스트하는 단계를 포함할 수 있다.

[0019] 상기와 같은 문제점을 해결하기 위한 본 발명의 다른 실시예에 따른 트래픽 분류 장치는 프로세서(processor) 및 상기 프로세서에 의해 실행되는 하나 이상의 명령들이 저장된 메모리(memory)를 포함할 수 있고, 상기 하나 이상의 명령들은, SDN(Software Difine Network) 컨트롤러로부터 플로우에 대한 정보를 포함하는 플로우 데이터를 수신하고, 상기 플로우 데이터에 대한 스케일링을 수행하고, 상기 플로우 데이터 간의 상관 관계를 기초로

상기 스케일링이 수행된 플로우 데이터 중 중복되는 데이터를 제거하여 입력 데이터를 생성하고 그리고 상기 입력 데이터를 기초로 네트워크 트래픽을 분류하도록 실행될 수 있다.

[0020] 여기서, 상기 상관 관계를 획득하는 경우, 상기 하나 이상의 명령들은, 주성분 분석 방식을 사용하여 상기 상관 관계를 획득하도록 실행될 수 있다.

[0021] 여기서, 상기 입력데이터는 플로우 특성을 포함할 수 있고, 상기 플로우 특성은, 상기 플로우가 도달한 시각에 대한 플로우 특성, 상기 플로우의 시작 위치에 대한 플로우 특성, 상기 플로우의 목적지에 대한 플로우 특성 중 하나 이상을 포함할 수 있다.

[0022] 여기서, 상기 네트워크 트래픽을 분류 하는 경우, 상기 하나 이상의 명령들은, 기계 학습이 수행된 트래픽 분류 장치에 의해 상기 네트워크 트래픽을 분류하도록 실행될 수 있다.

[0023] 여기서, 상기 트래픽 분류 장치는, 앙상블 학습 기법을 사용하여 상기 기계 학습을 수행할 수 있다.

[0024] 여기서, 상기 트래픽 분류 장치는, 레이블링이 수행된 플로우 데이터를 기초로 입력 데이터를 생성하고, 상기 입력 데이터를 기초로 학습 표본을 추출하고, 상기 추출된 학습 표본을 기초로 학습 데이터를 생성하고 그리고 상기 학습 데이터를 기초로 상기 기계 학습을 수행할 수 있다.

[0025] 여기서, 상기 트래픽 분류 장치는, 기 입력 데이터 각각의 기울기를 획득하고, 상기 입력 데이터의 기울기 크기를 기초로 상기 플로우 특성을 정렬하고, 상기 입력 데이터 중 일부를 미리 설정한 비율에 따라 추출하고 그리고 상기 샘플링이 수행된 입력 데이터를 기초로 상기 학습 표본을 추출할 수 있다.

[0026] 여기서, 상기 트래픽 분류 장치는, 상기 학습 표본에 대한 번들링을 수행하고 그리고 상기 번들링이 수행된 학습 표본을 병합하여 상기 학습 데이터를 생성할 수 있다.

[0027] 여기서, 상기 트래픽 분류 장치는, 상기 기계 학습을 수행하기 위한 학습 모델을 생성하고, 상기 학습 데이터를 트레이닝 셋, 검증 셋 및 테스트 셋으로 분류하고, 상기 트레이닝 셋을 기초로 상기 학습 모델의 가중치를 획득하고, 상기 검증 셋을 통해 상기 학습 모델을 검증하고 그리고 상기 테스트 셋을 통해 상기 검증된 학습 모델을 테스트 할 수 있다.

발명의 효과

[0028] 본 발명에 의하면, 기계 학습을 통해 네트워크 트래픽을 분류함으로써 트래픽 분류의 자동화를 이룰 수 있다.

[0029] 또한, 본 발명에 의하면, 기계 학습을 통해 네트워크 트래픽을 분류함으로써 트래픽 분류의 정확도가 향상될 수 있다.

도면의 간단한 설명

[0030] 도 1은 본 발명의 일 실시예에 따른 소프트웨어 정의 네트워크를 도시한 블록도이다.

도 2은 본 발명의 일 실시예에 따른 기계 학습 방법을 설명하기 위한 흐름도이다.

도 3는 본 발명의 일 실시예에 따른 학습 표본을 생성하는 방법을 도시한 흐름도이다.

도 4는 본 발명의 일 실시예에 따른 학습 데이터를 생성하는 방법을 도시한 흐름도이다.

도 5는 본 발명의 일 실시예에 따른 트래픽 분류 방법을 도시한 순서도이다.

도 6은 소프트웨어 정의 네트워크를 구성하는 통신 노드의 일 실시예를 도시한 블록도이다.

발명을 실시하기 위한 구체적인 내용

[0031] 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.

[0032] 제1, 제2 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다. 및/또는 이라는 용어는 복수의 관련된 기재된 항목들의

조합 또는 복수의 관련된 기재된 항목들 중의 어느 항목을 포함한다.

[0033] 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.

[0034] 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 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0035] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가지고 있다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥 상 가지는 의미와 일치하는 의미를 가진 것으로 해석되어야 하며, 본 출원에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.

[0036] 이하, 첨부한 도면들을 참조하여, 본 발명의 바람직한 실시예를 보다 상세하게 설명하고자 한다. 본 발명을 설명함에 있어 전체적인 이해를 용이하게 하기 위하여 도면상의 동일한 구성요소에 대해서는 동일한 참조부호를 사용하고 동일한 구성요소에 대해서 중복된 설명은 생략한다.

[0037] 도 1은 본 발명의 일 실시예에 따른 소프트웨어 정의 네트워크를 도시한 블록도이다.

[0038] 도 1을 참조하면, 본 발명의 일 실시예에 따른 SDN 기반의 네트워크는 스위치(110), SDN 컨트롤러(software define network controller, 120) 및 트래픽 분류 장치(130)를 포함할 수 있다.

[0039] 스위치(110)는 프로그래머블 스위치(Programmable Switch)일 수 있다. 프로그래머블 스위치는 스위치가 데이터 패킷을 처리하는 방식을 프로그래밍할 수 있는 스위치를 의미할 수 있다. 프로그래밍 스위치는 상위 수준 언어(DSL, Domain-Specific Language)를 이용하여 스위치가 데이터 패킷을 처리하는 방식을 프로그래밍할 수 있다. 예를 들어, 상위 수준 언어는 P4(Programming Protocol-independent Packet Processors)일 수 있다. 한편 도면에는 1 개의 스위치(110)만을 도시하였으나, 스위치(110)는 복수일 수 있다.

[0040] 스위치(110)는 복수의 출발지(예를 들어, 단말, 컴퓨터)로부터 데이터 패킷들을 수신할 수 있다. 스위치(110)는 데이터 패킷들 각각으로부터 네트워크 트래픽 분류를 위해 필요한 정보인 트래픽 정보를 획득할 수 있다. 여기에서 트래픽 정보는 데이터 패킷이 스위치에 도달한 시간을 나타내는 타임 스탬프(time stamp), 데이터 패킷의 출발지 IP(Internet Protocol) 주소, 데이터 패킷의 출발지 포트, 데이터 패킷의 목적지 IP 주소, 데이터 패킷의 목적지 포트 주소, 데이터 패킷이 사용하는 프로토콜, 데이터 패킷의 번호, 데이터 패킷의 바이트 번호, 데이터 패킷의 크기를 포함할 수 있다. 스위치(110)는 트래픽 정보를 포함하는 패킷 정보 메시지를 생성할 수 있다. 스위치(110)는 패킷 정보 메시지를 SDN 컨트롤러(120)에 전송할 수 있다.

[0041] SDN 컨트롤러(120)는 패킷 정보 메시지를 스위치(110)로부터 수신할 수 있다. SDN 컨트롤러(120)는 패킷 정보 메시지로부터 데이터 패킷들 각각의 트래픽 정보를 획득할 수 있다. SDN 컨트롤러(120)는 트래픽 정보를 플로우별로 분류하고 트래픽 정보의 평균을 연산하는 방식으로 플로우 데이터를 획득할 수 있다. 여기에서 플로우는 출발지와 목적지가 동일한 데이터 패킷들의 집합일 수 있다.

[0042] SDN 컨트롤러(120)는 타임 스탬프를 기초로 데이터 패킷들 각각이 스위치(110)에 도달한 시점을 미리 설정한 시간 단위에 따라 분류하여 플로우를 획득할 수 있다. SDN 컨트롤러(120)는 플로우에 포함된 데이터 패킷들의 트래픽 정보의 평균을 연산하는 방식으로 플로우 데이터를 획득할 수 있다.

[0043] 또한, SDN 컨트롤러(120)는 플로우 데이터를 포함하는 플로우 메시지를 생성할 수 있다. SDN 컨트롤러(120)는 플로우 메시지를 트래픽 분류 장치(130)에 전송할 수 있다.

[0044] 트래픽 분류 장치(130)는 플로우 메시지를 SDN 컨트롤러(120)로부터 수신할 수 있다. 트래픽 분류 장치(130)는 서버일 수 있으며, 복수의 분류기들을 포함할 수 있다. 트래픽 분류 장치(130)는 플로우 메시지로부터 플로우 데이터를 획득할 수 있다. 트래픽 분류 장치(130)는 플로우 데이터에 대한 전처리(pre-processing)를 수행하여

플로우 특성(feature)을 획득할 수 있다. 플로우 특성은 네트워크 트래픽 분류를 위해 필요한 값일 수 있다.

[0045] 트래픽 분류 장치(130)는 플로우 데이터에 대한 스케일링(scaling)을 수행할 수 있다. 트래픽 분류 장치(130)는 플로우 데이터에 대한 표준 값을 연산하는 방식으로 스케일링을 수행할 수 있다. 트래픽 분류 장치(130)는 다음 수학식 1에 따라 플로우 데이터에 대한 표준 값을 연산할 수 있다.

수학식 1

[0046] $$\text{표준 값 } (x') = \frac{x-\bar{x}}{\sigma(x)}$$

[0047] 수학식 1에서 $\bar{x}$ 는 플로우 데이터의 평균 값일 수 있고, $\sigma(x)$ 는 플로우 데이터의 표준 편차일 수 있다.

[0048] 트래픽 분류 장치(130)는 상관 관계를 기초로 중복되는 플로우 데이터를 하나의 플로우 특성으로 묶을 수 있다. 트래픽 분류 장치(130)는 주성분 분석(Principal Component Analysis, PCA) 방식을 사용하여 플로우 데이터에 대한 상관 관계를 획득할 수 있다. 트래픽 분류 장치(130)는 상관 관계를 기초로 플로우 데이터를 분류하고, 분류된 플로우 데이터를 하나의 플로우 특성으로 묶을 수 있다.

[0049] 예를 들어, 트래픽 분류 장치(130)는 플로우 데이터 중 출발지 아이피 주소 및 출발지 포트 정보의 상관 관계가 높은 것으로 판단할 수 있고, 출발지 아이피 주소 및 출발지 포트 정보를 플로우의 출발지 위치에 대한 플로우 특성으로 묶을 수 있다. 트래픽 분류 장치(130)는 플로우 데이터 중 목적지 아이피 주소 및 목적지 포트 정보의 상관 관계가 높은 것으로 판단할 수 있고, 목적지 아이피 주소 및 목적지 포트 정보를 플로우의 목적지에 대한 플로우 특성으로 묶을 수 있다.

[0050] 즉, 트래픽 분류 장치(130)는 플로우 데이터를 상관 관계를 통해 분류하고 분류된 플로우 데이터를 플로우가 도달한 시각에 대한 플로우 특성(time stamp), 플로우의 시작 위치에 대한 플로우 특성(source ip, source port), 플로우의 목적지에 대한 플로우 특성(destination IP, destination port), 플로우가 사용하는 프로토콜에 대한 플로우 특성, 데이터 패킷들의 번호(packet count)에 대한 플로우 특성, 데이터 패킷들의 바이트 번호(byte count)에 대한 플로우 특성, 데이터 패킷들이 도달하는 시간 사이의 간격(inter-arrival time)에 대한 플로우 특성으로 묶을 수 있다. 트래픽 분류 장치(130)는 트래픽 데이터 및 플로우 특성에 대한 인스턴스를 수행하여, 입력 데이터를 생성할 수 있다.

[0051] 트래픽 분류 장치(130)는 입력 데이터를 기초로 플로우의 트래픽을 분류할 수 있다. 트래픽 분류 장치(130)는 미리 기계 학습이 수행될 수 있다. 트래픽 분류 장치(130)는 기계 학습 수행 결과를 기초로 플로우의 트래픽을 분류할 수 있다. 트래픽 분류 장치(130)는 플로우의 트래픽을 분류하기 위한 복수의 분류기들을 포함할 수 있다. 예를 들어, 복수의 분류기들은 n개일 수 있다.

[0052] 트래픽 분류 장치(130)는 기계 학습 프로그래밍 언어로 파이썬(python) 2 or 파이썬 3를 사용할 수 있다. 트래픽 분류 장치(130)는 텐서플로우(Tensorflow)와 같은 프레임워크를 사용하여 기계 학습을 수행할 수 있다. 트래픽 분류 장치(130)는 앙상블 학습 기법 기반으로 기계 학습을 수행할 수 있다. 트래픽 분류 장치(130)는 앙상블 학습 기법의 알고리즘으로 LightGBM(Light Gradient Boosting Machine)을 사용할 수 있다. LightGBM은 부스팅(boosting)을 기반으로 하는 LightGBM 알고리즘일 수 있다. 한편, 트래픽 분류 장치(130)는 다음과 같은 방법으로 기계 학습을 수행할 수 있다.

[0053] 도 2는 본 발명의 일 실시예에 따른 기계 학습 방법을 설명하기 위한 흐름도이다.

[0054] 도 2를 참조하면, 트래픽 분류 장치(예를 들어, 도 1의 트래픽 분류 장치(130))는 입력 데이터를 기초로 기계 학습을 수행하기 위한 학습 표본을 생성할 수 있다(S210).

[0055] 트래픽 분류 장치는 SDN 컨트롤러(예를 들어, 도 1의 SDN 컨트롤러(120))으로부터 플로우 데이터를 포함하는 플로우 메시지를 수신할 수 있다. 플로우 데이터는 SDN 컨트롤러에 의해 트래픽 정보의 레이블링이 수행된 것일 수 있다. 트래픽 분류 장치는 트래픽 정보에 대한 전처리 및 인스턴스를 수행하여 입력 데이터를 생성할 수 있다.

[0056] 트래픽 분류 장치는 GOSS(Gradient-based One-Side Sampling) 방식을 사용하여 기계 학습을 수행하기 위한 학

습 표본을 생성할 수 있다. GOSS 방식은 입력 데이터의 기울기를 기반으로 학습 표본을 생성하는 방식일 수 있다. 여기에서 입력 데이터의 기울기는 입력 데이터에 대한 손실함수의 기울기일 수 있다. 손실 함수는 입력 데이터에 대한 출력 예상 값 및 실체 출력 값 사이의 차이에 관한 함수일 수 있다. 트래픽 분류 장치가 GOSS 방식을 사용하여 학습 표본을 생성하는 방법은 다음과 같을 수 있다.

[0057] 도 3은 본 발명의 일 실시예에 따른 학습 표본을 생성하는 방법을 도시한 흐름도이다.

[0058] 도 3을 참조하면, 트래픽 분류 장치는 입력 데이터 각각의 기울기를 획득할 수 있다(S310). 트래픽 분류 장치는 입력 데이터각각을 concave 함수 또는 convex 함수로 나타낼 수 있고, 각각의 함수를 미분하여 기울기를 획득할 수 있다. 트래픽 분류 장치는 기울기 크기를 기초로 입력 데이터를 정렬할 수 있다(S320). 예를 들어, 트래픽 분류 장치는 기울기 크기가 큰 순서대로 입력 데이터를 정렬할 수 있다.

[0059] 트래픽 분류 장치는 입력 데이터 중 일부를 추출할 수 있다(S330). 트래픽 분류 장치는 입력 데이터 중 기울기의 크기가 상위 a%에 해당하는 입력 데이터를 추출할 수 있고, 기울기의 크기가 하위 b%에 해당하는 입력 데이터를 추출할 수 있다. 트래픽 분류 장치는 추출된 입력 데이터 중 일부의 샘플링을 수행할 수 있다(S340). 트래픽 분류 장치는 하위 b%에 해당하는 입력 데이터에 대한 샘플링을 수행할 수 있다. 트래픽 분류 장치는 기울기의 크기가 하위 b%에 해당하는 입력 데이터를 가중치만큼 증폭시킬 수 있다. 트래픽 분류 장치는 아래 수학식 2에 따라 가중치 c를 연산할 수 있다.

수학식 2

[0060] $$c = \frac{1-a}{b}$$

[0061] 한편, 트래픽 분류 장치는 기울기의 크기가 상위 a%에 해당하는 입력 데이터에 대해서는 샘플링을 수행하지 않을 수 있다. 트래픽 분류 장치는 샘플링이 수행된 입력 데이터를 기초로 학습 표본을 생성할 수 있다(S350).

[0062] 트래픽 분류 장치는 기울기의 크기가 상위 a%에 해당하는 입력 데이터 및 기울기의 크기가 하위 b%에 해당하는 입력 데이터를 기초로 정보 이득을 연산할 수 있다. 여기에서 정보 이득(information gain)은 분산으로 측정될 수 있고, 기울기의 크기가 하위 b%에 해당하는 입력 데이터는 가중치 c만큼 증폭된 입력 데이터일 수 있다. 예를 들어, 노드 d에서 입력 데이터 j에 대한 분산 획득은 다음 수학식 3에 따라 연산될 수 있다.

수학식 3

[0063] $$\tilde{V}_j(d) = \frac{1}{n}\left(\frac{\left(\sum_{i\in A_l} g_i + \frac{1-a}{b}\sum_{i\in B_l} g_i\right)^2}{n_l(d)} + \frac{\left(\sum_{i\in A_r} g_i + \frac{1-a}{b}\sum_{i\in B_r} g_i\right)^2}{n_r(d)}\right)$$

[0064] 여기에서, $\tilde{V}_j(d)$는 기울기의 크기가 상위 a%에 해당하는 입력 데이터 및 기울기의 크기가 하위 b%에 해당하는 입력 데이터를 기초로 획득한 정보 이득일 수 있다. $n_l(d)$는 기울기의 크기가 상위 a%에 해당하는 입력 데이터일 수 있고, $n_r(d)$는 기울기의 크기가 하위 b%에 해당하는 입력 데이터일 수 있다. N은 기울기의 크기가 상위 a%에 해당하는 입력 데이터의 개수 및 기울기의 크기가 하위 b%에 해당하는 입력 데이터의 개수의 총합일 수 있다. i는 플로우 특성 각각의 기울기일 수 있다. g_i는 입력 데이터 각각에 대한 손실 함수(loss function)의 기울기를 의미할 수 있다. 또한, 이 식에서 A는 기울기가 큰 상위 a% 개의 입력 데이터의 집합일 수 있고, B는 기울기의 크기가 하위 b%에 해당하는 입력 데이터 집합일 수 있다. 한편, GOSS 방식의 근사 비율은 다음 수학식 4와 같을 수 있다.

수학식 4

[0065] $$O\left(\frac{1}{n_l(d)}+\frac{1}{n_r(d)}+\frac{1}{\sqrt{n}}\right)$$

[0066] 한편, 플로우 특성을 기초로 정보 이득을 연산하는 방법은 다음 수학식 5와 같을 수 있다,

수학식 5

[0067] $$V_j(d)=\frac{1}{n}\left(\frac{\left(\sum_{\{i\le d\}} g_i\right)^2}{n_l(d)}+\frac{\left(\sum_{\{i>d\}} g_i\right)^2}{n_r(d)}\right)$$

[0068] $V_j(d)$ 는 입력 데이터의 모든 플로우 특성을 기초로 획득한 정보 이득일 수 있다. 즉, 수학식 3에 의해 획득되는 정보 이득은 정보 이득의 연산에 사용되는 입력 데이터의 개수가 증가할수록, 수학식 5에 의해 획득되는 정보 이득과 동일해질 수 있다.

[0069] 다시 도 2를 참조하면, 트래픽 분류 장치는 추출된 학습 표본을 번들링(bundling) 및 병합(merging)하여 학습 데이터를 생성할 수 있다(S220). 트래픽 분류 장치는 EFB(Exclusive Feature Bundling) 방식을 사용하여 학습 표본에 대한 번들링 및 병합을 수행할 수 있다. EFB 방식은 학습 표본의 플로우 특성을 기초로 학습 표본의 번들링 및 병합을 수행하는 방식일 수 있다. 트래픽 분류 장치가 EFB 방식을 사용하여 학습 표본을 번들링 및 병합하여 학습 데이터를 생성하는 방법은 다음과 같을 수 있다.

[0070] 도 4는 본 발명의 일 실시예에 따른 학습 데이터를 생성하는 방법을 도시한 흐름도이다.

[0071] 도 4를 참조하면, 트래픽 분류 장치는 학습 표본간의 관계를 나타내는 그래프를 생성할 수 있다(S410). 트래픽 분류 장치는 플로우 특성을 노드(node)로 하여, 플로우 특성 각각이 상호 배타적인지 확인할 수 있다. 플로우 특성이 상호간에 0이 아닌 값을 동시에 갖지 않을 경우, 트래픽 분류 장치는 플로우 특성이 상호 배타적이라고 판단할 수 있다. 트래픽 분류 장치는 상호 배타적이지 않은 각각의 노드들 사이에 엣지를 연결하여 표본 사이의 관계를 나타내는 그래프를 획득할 수 있다.

[0072] 트래픽 분류 장치는 그래프를 기초로 학습 표본의 번들링을 수행할 수 있다(S420). 트래픽 분류 장치는 노드의 차수에 따라서 표본 각각을 내림차순으로 정렬할 수 있다. 트래픽 분류 장치는 획득한 그래프를 기초로 표본을 내림차순으로 정렬할 수 있다. 트래픽 분류 장치는 내림차순으로 정렬된 플로우 특성을 번들에 할당할 수 있다. 트래픽 분류 장치는 플로우 특성과 상호 배타적이지 않은 표본을 포함하는 번들이 존재하는 경우, 해당 플로우 특성을 미리 존재하는 번들에 할당할 수 있다. 트래픽 분류 장치는 표본과 상호 배타적인 표본을 포함하는 번들만이 존재하는 경우 새로운 번들을 생성할 수 있다. 트래픽 분류 장치는 해당 표본들을 새로운 번들에 할당할 수 있다.

[0073] 트래픽 분류 장치는 번들을 병합하여 학습 데이터를 생성할 수 있다(S430). 트래픽 분류 장치는 번들 각각에 포함된 표본들의 값을 확인할 수 있다. 트래픽 분류 장치는 표본들의 값에 오프셋을 더할 수 있다. 예를 들어 표본 A가 [0, 10]의 값을 가지고 표본 B는 [0, 20]의 값을 가진다고 할 때, 표본 B에 오프셋 (offset)을 더하여 [10, 30]에서 값을 가지게 할 수 있다. 그 다음 표본 A와 B를 병합하고 [0, 30]에서 값을 취하게 하여 표본 B를 표본 A로 대체할 수 있다. 즉, 표본 B를 표본 A에 병합하는 방식으로 학습 데이터를 생성할 수 있다.

[0074] 다시 도 2를 참조하면, 트래픽 분류 장치는 학습 데이터를 기초로 기계 학습을 수행할 수 있다(S230).

[0075] 트래픽 분류 장치는 학습 데이터를 기초로 학습 모델을 생성할 수 있다. 트래픽 분류 장치는 k-fold 교차 검증(cross validation) 방식을 통해 기계 학습을 수행할 수 있다. 트래픽 분류 장치는 학습 데이터에 대한 셔플을 수행할 수 있다. 트래픽 분류 장치는 셔플이 수행된 학습 데이터를 테스트 셋(test set), 검증 셋(validation set) 및 트레이닝 셋(training set)으로 분류할 수 있다. 트래픽 분류 장치는 학습 데이터 중 일부를 테스트 셋

으로 분류할 수 있다. 트래픽 분류 장치는 나머지 학습 데이터를 k 개의 데이터 셋들로 분류할 수 있다. 나머지 학습 데이터는 학습 데이터 중 테스트 셋을 제외한 학습 데이터일 수 있다. 예를 들어, k는 5일 수 있고 트래픽 분류 장치는 나머지 학습 데이터를 제1 데이터 셋 내지 제5 데이터 셋으로 분류할 수 있다.

[0076] 트래픽 분류 장치는 k 개의 데이터 셋들 가운데 하나의 데이터 셋을 검증 셋으로 분류할 수 있고, 나머지 데이터 셋을 트레이닝 셋으로 분류할 수 있다. 예를 들어, 트래픽 분류 장치는 제1 데이터 셋을 검증 셋으로 제2 데이터 셋 내지 제5 데이터 셋을 트레이닝 셋으로 각각 분류할 수 있고, 제2 데이터 셋을 검증 셋으로 제1 데이터 셋 및 제3 데이터 셋 내지 제5 데이터 셋을 트레이닝 셋으로 각각 분류할 수 있으며, 제3 데이터 셋을 검증 셋으로 제1 데이터 셋 내지 제2 데이터 셋 및 제4 데이터 셋 내지 제5 데이터 셋을 트레이닝 셋으로 각각 분류할 수 있다. 또한 트래픽 분류 장치는 제4 데이터 셋을 검증 셋으로 제1 데이터 셋 내지 제3 데이터 셋 및 제5 데이터 셋을 트레이닝 셋으로 각각 분류할 수 있고, 제5 데이터 셋을 검증 셋으로 제1 데이터 셋 내지 제4 데이터 셋을 트레이닝 셋으로 각각 분류할 수 있다.

[0077] 트래픽 분류 장치는 트레이닝 셋을 입력하여 출력 값을 획득할 수 있고, 검증 셋을 기초로 출력 값을 검증하는 방식으로 기계 학습을 수행할 수 있다. 트래픽 분류 장치는 트레이닝 셋을 입력하여 획득한 출력 값을 기초로 학습 모델에 대한 가중치를 획득할 수 있다. 트래픽 분류 장치는 획득한 학습 모델에 검증 셋을 입력할 수 있고, 검증 셋의 출력과 레이블링 값을 비교하여 학습 결과를 검증할 수 있다. 즉, 트래픽 분류 장치는 트레이닝 셋을 통해 획득한 학습 모델의 가중치를 검증 셋을 통해 검증할 수 있다.

[0078] 예를 들어, 트래픽 분류 장치는 제2 데이터 셋 내지 제5 데이터 셋을 입력하여 출력 값을 획득할 수 있고, 제1 데이터 셋을 기초로 출력 값을 검증할 수 있으며, 제1 데이터 셋 및 제3 데이터 셋 내지 제5 데이터 셋을 입력하여 출력 값을 획득할 수 있고, 제2 데이터 셋을 기초로 출력 값을 검증할 수 있으며, 제1 데이터 셋 내지 제2 데이터 셋 및 제4 데이터 셋 내지 제5 데이터 셋을 입력하여 출력 값을 획득할 수 있고, 제3 데이터 셋을 기초로 출력 값을 검증할 수 있다. 또한. 트래픽 분류 장치는 제1 데이터 셋 내지 제3 데이터 셋 및 제5 데이터 셋을 입력하여 출력 값을 획득할 수 있고, 제4 데이터 셋을 기초로 출력 값을 검증할 수 있으며, 제1 데이터 셋 내지 제4 데이터 셋을 입력하여 출력 값을 획득할 수 있고, 제5 데이터 셋을 기초로 출력 값을 검증할 수 있다. 트래픽 분류 장치는 이와 같은 과정을 n회 반복할 수 있다. 여기에서 n은 10일 수 있다. 이후 트래픽 분류 장치는 테스트 셋을 통해 기계 학습의 수행 결과를 판단할 수 있다. 트래픽 분류 장치는 테스트 셋을 입력하여 출력 값을 획득할 수 있고, 출력 값을 테스트 셋에 포함된 데이터의 레이블링 값과 비교하여 학습 수행 결과를 테스트할 수 있다.

[0079] 한편, 이와 같은 과정은 각각의 분류기들 마다 수행될 수 있다. 트래픽 분류 장치는 제1 분류기의 테스트 셋의 출력 값과 레이블링 값이 다른 경우, 트레이닝 셋 중 출력의 오류가 발생한 트레이닝 셋을 추출할 수 있다. 트래픽 분류 장치는 추출한 트레이닝 셋을 제2 분류기에 입력할 수 있고, 전술한 학습 과정을 반복 수행하여 트레이닝 셋의 가중치를 조절할 수 있다. 트래픽 분류 장치는 트레이닝 셋에 대한 출력 값에 오류가 발생하지 않을 때까지 이러한 학습 과정을 반복하여 수행할 수 있다.

[0080] 도 5는 본 발명의 일 실시예에 따른 트래픽 분류 방법을 도시한 순서도이다.

[0081] 도 5를 참조하면, 스위치, SDN 컨트롤러 및 트래픽 분류 장치는 도 1의 스위치(110), SDN 컨트롤러(120) 및 트래픽 분류 장치(130)와 동일하거나 유사하게 구성될 수 있다.

[0082] 스위치는 복수의 출발지들로부터 데이터 패킷들을 수신할 수 있다(S505). 스위치는 데이터 패킷들을 기초로 패킷 정보 메시지를 생성할 수 있다(S510). 스위치는 데이터 패킷들 각각으로부터 트래픽 정보를 획득할 수 있다. 트래픽 정보는 네트워크 트래픽 분류를 위해 필요한 정보일 수 있다. 스위치는 트래픽 정보를 포함하는 패킷 정보 메시지를 생성할 수 있다. 스위치는 패킷 정보 메시지를 SDN 컨트롤러에 전송할 수 있다(S515).

[0083] SDN 컨트롤러는 패킷 정보 메시지를 스위치로부터 수신할 수 있다(S515). SDN 컨트롤러는 패킷 정보 메시지를 기초로 각각의 데이터 패킷들의 트래픽 정보를 획득할 수 있다. SDN 컨트롤러는 트래픽 정보를 기초로 플로우 메시지를 생성할 수 있다(S520). SDN 컨트롤러는 트래픽 정보를 플로우 별로 분류할 수 있다. 즉, SDN 컨트롤러는 트래픽 정보 중 타임 스탬프를 기초로 트래픽 정보를 플로우 별로 분류할 수 있다. SDN 컨트롤러는 플로우 별로 트래픽 정보의 평균을 연산하는 방식으로 플로우 데이터를 획득할 수 있다. SDN 컨트롤러는 플로우 데이터를 포함하는 플로우 메시지를 생성할 수 있다. SDN 컨트롤러는 플로우 메시지를 트래픽 분류 장치에 전송할 수 있다(S525).

[0084] 트래픽 분류 장치는 플로우 메시지를 SDN 컨트롤러로부터 수신할 수 있다(S525). 트래픽 분류 장치는 플로우 메

시지로부터 플로우 데이터를 획득할 수 있다. 트래픽 분류 장치는 플로우 특성을 추출할 수 있다(S530). 트래픽 분류 장치는 플로우 데이터를 기초로 플로우 특성을 추출할 수 있다. 트래픽 분류 장치는 플로우 데이터 각각의 전처리를 수행하여 플로우 특성을 획득할 수 있다. 트래픽 분류 장치는 플로우 데이터 각각의 스케일링을 수행하고, 스케일링이 수행된 플로우 데이터를 상관 관계에 따라 분류하는 방식으로 플로우 데이터 각각에 대한 전처리를 수행하여 플로우 특성을 추출할 수 있다.

[0085] 트래픽 분류 장치는 추출된 플로우 특성을 네트워크 트래픽을 분류할 수 있다(S535). 트래픽 분류 장치는 미리 수행된 기계 학습 결과를 기초로 네트워크 트래픽을 분류할 수 있다. 즉, 트래픽 분류 장치는 각각의 플로우 특성에 대한 출력 값을 획득하여 네트워크 트래픽을 분류할 수 있다.

[0086] 한편, 도 1의 스위치(110), SDN 컨트롤러(120) 및 트래픽 분류 장치(130)는 통신 노드일 수 있고, 통신 노드는 다음과 같이 구성될 수 있다.

[0087] 도 6은 본 발명의 일 실시예에 따른 소프트웨어 정의 네트워크를 구성하는 통신 노드의 일 실시예를 도시한 블록도이다.

[0088] 도 6을 참조하면, 통신 노드(600)는 적어도 하나의 프로세서(610), 메모리(620) 및 네트워크와 연결되어 통신을 수행하는 송수신 장치(630)를 포함할 수 있다. 또한, 통신 노드(600)는 입력 인터페이스 장치(640), 출력 인터페이스 장치(650), 저장 장치(660) 등을 더 포함할 수 있다. 통신 노드(600)에 포함된 각각의 구성 요소들은 버스(bus)(670)에 의해 연결되어 서로 통신을 수행할 수 있다. 다만, 통신 노드(600)에 포함된 각각의 구성요소들은 공통 버스(670)가 아니라, 프로세서(610)를 중심으로 개별 인터페이스 또는 개별 버스를 통하여 연결될 수도 있다. 예를 들어, 프로세서(610)는 메모리(620), 송수신 장치(630), 입력 인터페이스 장치(640), 출력 인터페이스 장치(650) 및 저장 장치(660) 중에서 적어도 하나와 전용 인터페이스를 통하여 연결될 수도 있다.

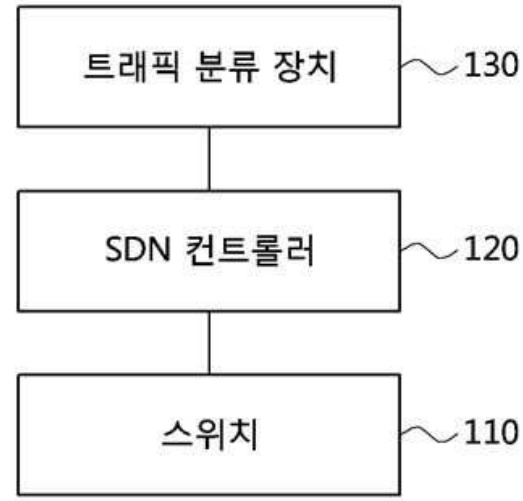
[0089] 프로세서(610)는 메모리(620) 및 저장 장치(660) 중에서 적어도 하나에 저장된 프로그램 명령(program command)을 실행할 수 있다. 프로세서(610)는 중앙 처리 장치(central processing unit, CPU), 그래픽 처리 장치(graphics processing unit, GPU), 또는 본 발명의 실시예들에 따른 방법들이 수행되는 전용의 프로세서를 의미할 수 있다. 메모리(620) 및 저장 장치(660) 각각은 휘발성 저장 매체 및 비휘발성 저장 매체 중에서 적어도 하나로 구성될 수 있다. 예를 들어, 메모리(620)는 읽기 전용 메모리(read only memory, ROM) 및 랜덤 액세스 메모리(random access memory, RAM) 중에서 적어도 하나로 구성될 수 있다.

[0090] 본 발명에 따른 방법들은 다양한 컴퓨터 수단을 통해 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 컴퓨터 판독 가능 매체에 기록되는 프로그램 명령은 본 발명을 위해 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다.

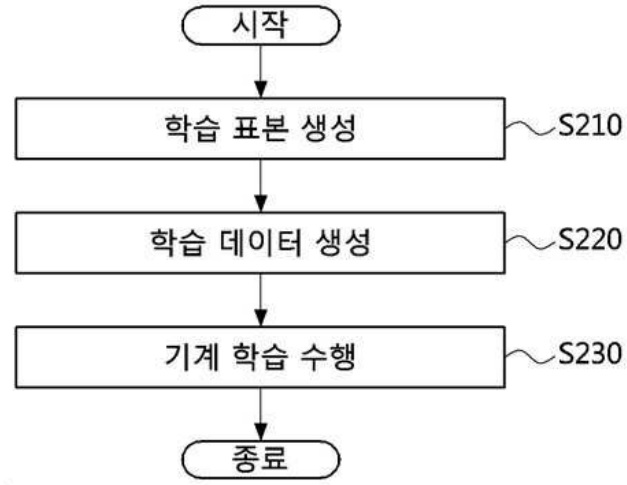
[0091] 컴퓨터 판독 가능 매체의 예에는 롬(rom), 램(ram), 플래시 메모리(flash memory) 등과 같이 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러(compiler)에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터(interpreter) 등을 사용해서 컴퓨터에 의해 실행될 수 있는 고급 언어 코드를 포함한다. 상술한 하드웨어 장치는 본 발명의 동작을 수행하기 위해 적어도 하나의 소프트웨어 모듈로 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

도면

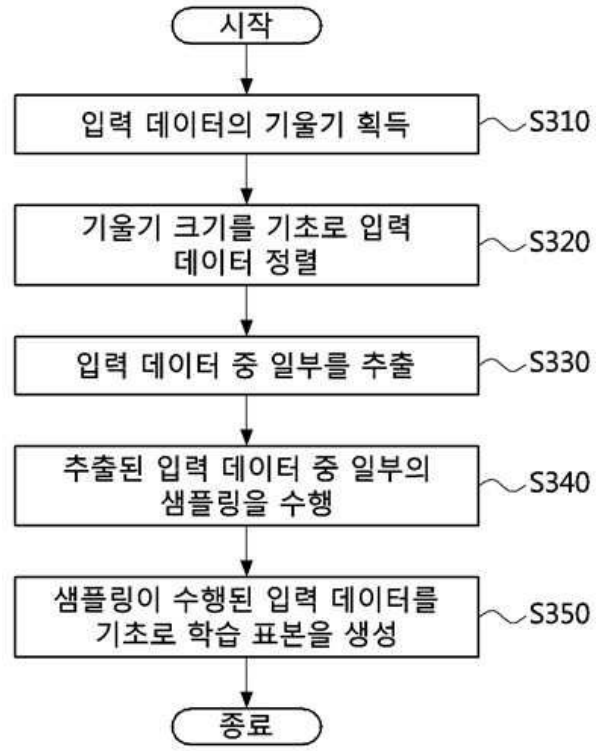
도면1


트래픽 분류 장치
130
SDN 컨트롤러
120
스위치
110

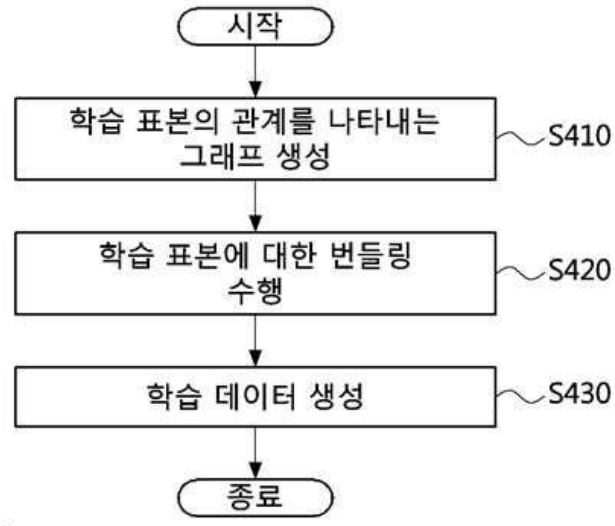
도면2


시작
학습 표본 생성
S210
학습 데이터 생성
S220
기계 학습 수행
S230
종료

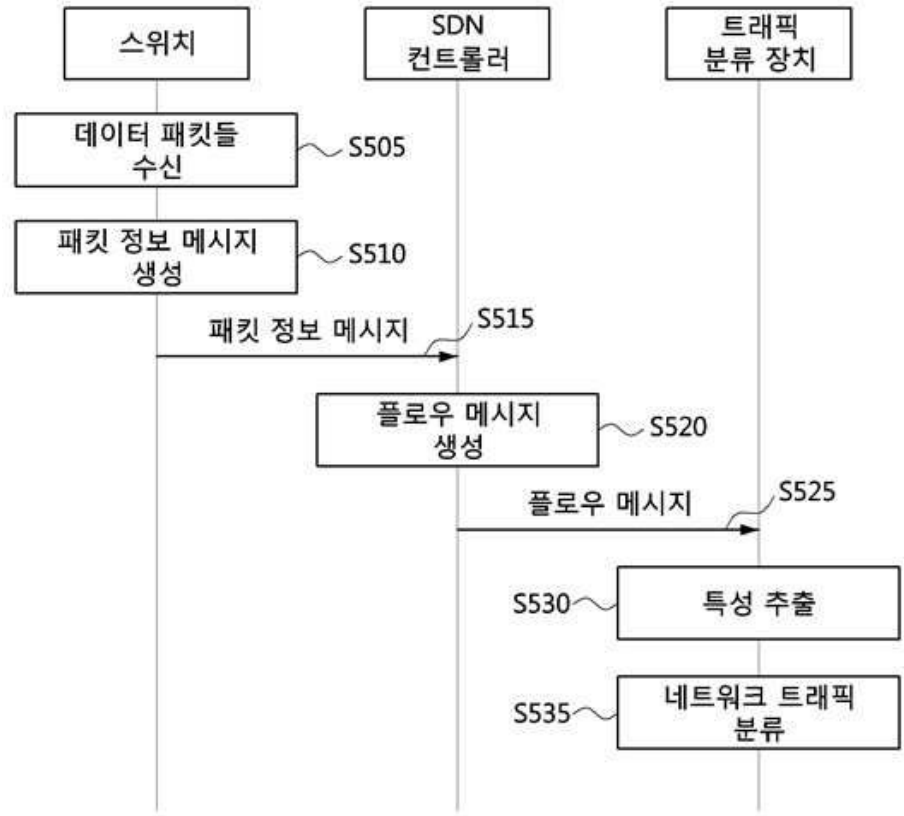
도면3


시작
입력 데이터의 기울기 획득
S310
기울기 크기를 기초로 입력 데이터 정렬
S320
입력 데이터 중 일부를 추출
S330
추출된 입력 데이터 중 일부의 샘플링을 수행
S340
샘플링이 수행된 입력 데이터를 기초로 학습 표본을 생성
S350
종료

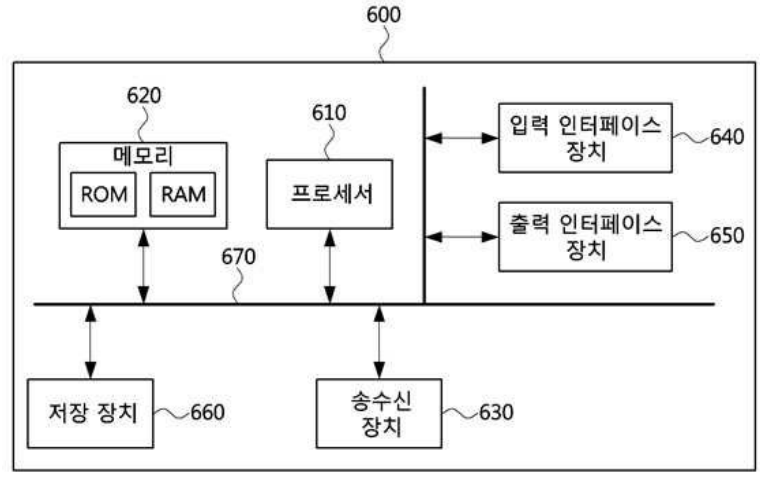
도면4


시작
학습 표본의 관계를 나타내는 그래프 생성
S410
학습 표본에 대한 번들링 수행
S420
학습 데이터 생성
S430
종료

도면5


스위치
SDN 컨트롤러
트래픽 분류 장치
데이터 패킷들 수신
S505
패킷 정보 메시지 생성
S510
패킷 정보 메시지
S515
플로우 메시지 생성
S520
플로우 메시지
S525
S530
특성 추출
S535
네트워크 트래픽 분류

도면6


600
620
메모리
ROM
RAM
610
프로세서
670
입력 인터페이스 장치
640
출력 인터페이스 장치
650
저장 장치
660
송수신 장치
630