(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2022년05월30일
(11) 등록번호 10-2403376
(24) 등록일자 2022년05월25일

(51) 국제특허분류(Int. Cl.)
H04L 47/2441 (2022.01) *G06N 5/00* (2022.01)
H04L 43/062 (2022.01) *H04L 43/067* (2022.01)
H04L 43/16 (2022.01)
(52) CPC특허분류
H04L 47/2441 (2022.05)
G06N 5/003 (2019.01)
(21) 출원번호 10-2022-0023416
(22) 출원일자 2022년02월23일
심사청구일자 2022년02월23일
(56) 선행기술조사문헌
KR1020140064149 A*
KR1020210070597 A*
KR1020210121253 A*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
경희대학교 산학협력단
경기도 용인시 기흥구 덕영대로 1732 (서천동, 경희대학교 국제캠퍼스내)
(72) 발명자
홍인기
경기도 성남시 분당구 수내로 181, 313동 301호(분당동, 샛별마을)
김영준
서울특별시 강남구 삼성로63길 36-10, 202호(대치동)
(74) 대리인
김연권

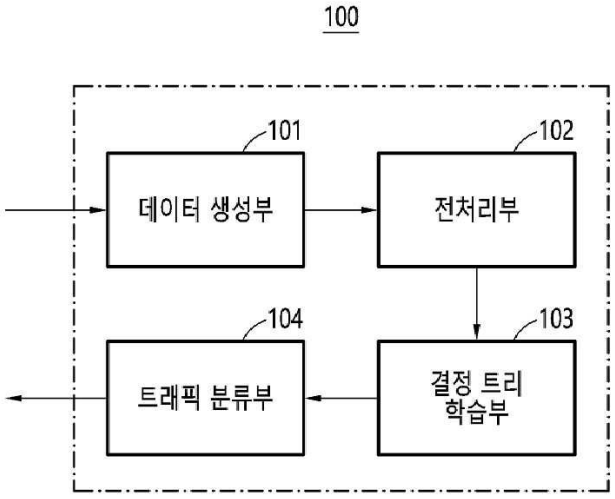
전체 청구항 수 : 총 9 항 심사관 : 김대성

(54) 발명의 명칭 **결정 트리에 기반하는 트래픽 분류 장치 및 그 방법**

(57) 요 약

본 발명은 결정 트리에 기반하는 트래픽 분류 장치 및 그 방법에 관한 것으로서, 일실시예에 따른 트래픽 분류 장치는 기설정된 영역에서 발생되는 트래픽 데이터에 기반하여 제1 트래픽 데이터 세트 및 제2 트래픽 데이터 세트를 생성하는 데이터 생성부와, 기설정된 타겟 지역에 대응되는 지역 특성 정보 및 시간 특성 정보에 기반하여 제1 트래픽 데이터 세트에 포함된 트래픽 데이터 각각을 전처리하는 전처리부와, 타겟 지역에 대응되는 트래픽을 분류하는 결정 트리 알고리즘을 전처리된 트래픽 데이터에 기반하여 학습하는 결정 트리 학습부, 및 제2 트래픽 데이터 세트를 학습된 결정 트리 알고리즘에 입력하여 타겟 지역에 대응되는 적어도 하나의 트래픽 데이터를 도출하는 트래픽 분류부를 포함한다.

대 표 도 - 도1



(52) CPC특허분류
H04L 43/062 (2022.05)
H04L 43/067 (2022.05)
H04L 43/16 (2022.05)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711139229
과제번호	2021-0-02046-001
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신기술인력양성/대학ICT연구센터육성지원
연구과제명	6H 차세대 이동통신 기술 개발
기 여 율	1/1
과제수행기관명	경희대학교 산학협력단
연구기간	2021.07.01 ~ 2028.12.31

명 세 서

청구범위

청구항 1

기설정된 영역에서 발생되는 트래픽 데이터에 기반하여 제1 트래픽 데이터 세트 및 제2 트래픽 데이터 세트를 생성하는 데이터 생성부;

기설정된 타겟 지역에 대응되는 지역 특성 정보 및 시간 특성 정보에 기반하여 상기 제1 트래픽 데이터 세트에 포함된 트래픽 데이터 각각을 전처리하는 전처리부;

상기 타겟 지역에 대응되는 트래픽을 분류하는 결정 트리 알고리즘을 상기 전처리된 트래픽 데이터에 기반하여 학습하는 결정 트리 학습부 및

상기 제2 트래픽 데이터 세트를 상기 학습된 결정 트리 알고리즘에 입력하여 상기 타겟 지역에 대응되는 적어도 하나의 트래픽 데이터를 도출하는 트래픽 분류부

를 포함하고,

상기 전처리부는,

상기 제1 트래픽 데이터 세트에 포함된 트래픽 데이터 중 적어도 하나의 트래픽 데이터에 상기 타겟 지역에 대응되는 트래픽 데이터를 식별하기 위한 인덱스 정보를 반영하는

트래픽 분류 장치.

청구항 2

기설정된 영역에서 발생되는 트래픽 데이터에 기반하여 제1 트래픽 데이터 세트 및 제2 트래픽 데이터 세트를 생성하는 데이터 생성부;

기설정된 타겟 지역에 대응되는 지역 특성 정보 및 시간 특성 정보에 기반하여 상기 제1 트래픽 데이터 세트에 포함된 트래픽 데이터 각각을 전처리하는 전처리부;

상기 타겟 지역에 대응되는 트래픽을 분류하는 결정 트리 알고리즘을 상기 전처리된 트래픽 데이터에 기반하여 학습하는 결정 트리 학습부 및

상기 제2 트래픽 데이터 세트를 상기 학습된 결정 트리 알고리즘에 입력하여 상기 타겟 지역에 대응되는 적어도 하나의 트래픽 데이터를 도출하는 트래픽 분류부

를 포함하고,

상기 데이터 생성부는,

상기 기설정된 영역을 기설정된 크기를 갖는 복수의 그리드(grids)로 분할하고, 상기 분할된 그리드 마다 기설정된 시간 단위로 수집되는 트래픽 데이터에 기반하는 상기 제1 트래픽 데이터 세트 및 상기 제2 트래픽 데이터 세트를 생성하는

트래픽 분류 장치.

청구항 3

삭제

청구항 4

제1항에 있어서,

상기 전처리부는,

상기 지역 특성 정보에 따른 상기 제1 트래픽 데이터 세트의 기지국 정보에 기반하여 상기 제1 트래픽 데이터

세트에 포함된 트래픽 데이터 중 적어도 하나의 트래픽 데이터에 상기 인덱스 정보를 반영하는

트래픽 분류 장치.

청구항 5

제1항에 있어서,

상기 전처리부는,

상기 제1 트래픽 데이터 세트에 포함된 트래픽 데이터 중 상기 시간 특성 정보에 대응되는 시간에 수집된 트래픽 데이터를 추출하는

트래픽 분류 장치.

청구항 6

제5항에 있어서,

상기 전처리부는,

상기 타겟 지역이 대학 지역, 상업 지역 및 주거 지역 중 적어도 하나인 경우, 상기 제1 트래픽 데이터 세트에 포함된 트래픽 데이터 중 평일에 수집된 트래픽 데이터를 추출하는

트래픽 분류 장치.

청구항 7

제5항에 있어서,

상기 전처리부는,

상기 타겟 지역이 상업 지역인 경우, 상기 제1 트래픽 데이터 세트에 포함된 트래픽 데이터 중 10시 내지 16시에 수집된 트래픽 데이터를 추출하고,

상기 타겟 지역이 주거 지역인 경우, 18시 내지 24시에 수집된 트래픽 데이터를 추출하는

트래픽 분류 장치.

청구항 8

제1항에 있어서,

상기 결정 트리 학습부는,

상기 전처리된 트래픽 데이터 중 상기 타겟 지역에 대응되는 트래픽 데이터의 개수에 대응하여 상기 결정 트리 알고리즘의 리프 노드 각각의 최소 샘플 수를 조절하는

트래픽 분류 장치.

청구항 9

제8항에 있어서,

상기 결정 트리 학습부는,

상기 전처리된 트래픽 데이터 중 상기 타겟 지역에 대응되는 트래픽 데이터의 개수가 기설정된 임계 개수 이상인 경우, 상기 최소 샘플 수를 제1 데이터 값으로 조절하고,

상기 전처리된 트래픽 데이터 중 상기 타겟 지역에 대응되는 트래픽 데이터의 개수가 상기 임계 개수 미만인 경우, 상기 최소 샘플 수를 상기 제1 데이터 값보다 큰 제2 데이터 값으로 조절하는

트래픽 분류 장치.

청구항 10

데이터 생성부에서, 기설정된 영역에서 발생되는 트래픽 데이터에 기반하여 제1 트래픽 데이터 세트를 생성하는 단계;

전처리부에서, 기설정된 타겟 지역에 대응되는 지역 특성 정보 및 시간 특성 정보에 기반하여 상기 제1 트래픽 데이터 세트에 포함된 트래픽 데이터 각각을 전처리하는 단계;

결정 트리 학습부에서, 상기 타겟 지역에 대응되는 트래픽을 분류하는 결정 트리 알고리즘을 상기 전처리된 트래픽 데이터에 기반하여 학습하는 단계;

상기 데이터 생성부에서, 상기 기설정된 영역에서 발생되는 트래픽 데이터에 기반하여 제2 트래픽 데이터 세트를 생성하는 단계 및

트래픽 분류부에서, 상기 제2 트래픽 데이터 세트를 상기 학습된 결정 트리 알고리즘에 입력하여 상기 타겟 지역에 대응되는 적어도 하나의 트래픽 데이터를 도출하는 단계

를 포함하고,

상기 전처리하는 단계는,

상기 전처리부에서, 상기 제1 트래픽 데이터 세트에 포함된 트래픽 데이터 중 적어도 하나의 트래픽 데이터에 상기 타겟 지역에 대응되는 트래픽 데이터를 식별하기 위한 인덱스 정보를 반영하는

트래픽 분류 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 트래픽 분류 장치 및 그 방법에 관한 것으로, 보다 상세하게는 결정 트리 알고리즘에 기반하여 트래픽을 분류하는 기술적 사상에 관한 것이다.

배 경 기 술

[0002] 현재, 효율적인 무선 자원 할당을 위해 딥러닝을 이용한 모바일 트래픽 예측에 대한 기술연구가 활발하게 진행되고 있다.

[0003] 그러나, 기존의 딥러닝을 이용한 모바일 트래픽 예측 기술은 특정 특징을 갖는 트래픽 데이터뿐만 아니라 전체 트래픽 데이터를 딥러닝 학습하고 이를 통해 모바일 트래픽을 분류하여, 트래픽 분류의 정확도가 떨어진다는 문제가 있다.

선행기술문헌

특허문헌

[0004] (특허문헌 0001) 한국등록특허 제10-1110956호, "트래픽 분류 방법 및 트래픽 분류 설비"

발명의 내용

해결하려는 과제

[0005] 본 발명은 트래픽 데이터 세트를 입력으로 하는 결정 트리 알고리즘을 이용하여 타겟 지역에 대응되는 트래픽 데이터를 추출할 수 있는 트래픽 분류 장치 및 그 방법을 제공하고자 한다.

[0006] 또한, 본 발명은 타겟 지역에 대응되는 지역 특성 및 시간 특성을 고려한 전처리 과정을 통해 타겟 지역에 대응되는 트래픽 데이터를 보다 정확하게 추출할 수 있는 트래픽 분류 장치 및 그 방법을 제공하고자 한다.

[0007] 또한, 본 발명은 전처리된 트래픽 데이터에 기반하여 결정 트리 알고리즘의 리프 노드를 최적화하여 타겟 지역에 대응되는 트래픽 데이터를 보다 정확하게 추출할 수 있는 트래픽 분류 장치 및 그 방법을 제공하고자 한다.

과제의 해결 수단

[0008] 본 발명의 일실시예에 따른 트래픽 분류 장치는 기설정된 영역에서 발생되는 트래픽 데이터에 기반하여 제1 트래픽 데이터 세트 및 제2 트래픽 데이터 세트를 생성하는 데이터 생성부와, 기설정된 타겟 지역에 대응되는 지역 특성 정보 및 시간 특성 정보에 기반하여 제1 트래픽 데이터 세트에 포함된 트래픽 데이터 각각을 전처리하는 전처리부와, 타겟 지역에 대응되는 트래픽을 분류하는 결정 트리 알고리즘을 전처리된 트래픽 데이터에 기반하여 학습하는 결정 트리 학습부 및 제2 트래픽 데이터 세트를 학습된 결정 트리 알고리즘에 입력하여 타겟 지역에 대응되는 적어도 하나의 트래픽 데이터를 도출하는 트래픽 분류부를 포함할 수 있다.

[0009] 일측에 따르면, 데이터 생성부는 기설정된 영역을 기설정된 크기를 갖는 복수의 그리드(grids)로 분할하고, 분할된 그리드 마다 기설정된 시간 단위로 수집되는 트래픽 데이터에 기반하는 제1 트래픽 데이터 세트 및 제2 트래픽 데이터 세트를 생성할 수 있다.

[0010] 일측에 따르면, 전처리부는 제1 트래픽 데이터 세트에 포함된 트래픽 데이터 중 적어도 하나의 트래픽 데이터에 타겟 지역에 대응되는 트래픽 데이터를 식별하기 위한 인덱스 정보를 반영할 수 있다.

[0011] 일측에 따르면, 전처리부는 지역 특성 정보에 따른 제1 트래픽 데이터 세트의 기지국 정보에 기반하여 제1 트래픽 데이터 세트에 포함된 트래픽 데이터 중 적어도 하나의 트래픽 데이터에 인덱스 정보를 반영할 수 있다.

[0012] 일측에 따르면, 전처리부는 제1 트래픽 데이터 세트에 포함된 트래픽 데이터 중 시간 특성 정보에 대응되는 시간에 수집된 트래픽 데이터를 추출할 수 있다.

[0013] 일측에 따르면, 전처리부는 타겟 지역이 대학 지역, 상업 지역 및 주거 지역 중 적어도 하나인 경우, 제1 트래픽 데이터 세트에 포함된 트래픽 데이터 중 평일에 수집된 트래픽 데이터를 추출할 수 있다.

[0014] 일측에 따르면, 전처리부는 타겟 지역이 상업 지역인 경우, 제1 트래픽 데이터 세트에 포함된 트래픽 데이터 중 10시 내지 16시에 수집된 트래픽 데이터를 추출하고, 타겟 지역이 주거 지역인 경우, 18시 내지 24시에 수집된 트래픽 데이터를 추출할 수 있다.

[0015] 일측에 따르면, 결정 트리 학습부는 전처리된 트래픽 데이터 중 타겟 지역에 대응되는 트래픽 데이터의 개수에 대응하여 결정 트리 알고리즘의 리프 노드 각각의 최소 샘플 수를 조절할 수 있다.

[0016] 일측에 따르면, 결정 트리 학습부는 전처리된 트래픽 데이터 중 타겟 지역에 대응되는 트래픽 데이터의 개수가 기설정된 임계 개수 이상인 경우, 최소 샘플 수를 제1 데이터 값으로 조절하고, 전처리된 트래픽 데이터 중 타겟 지역에 대응되는 트래픽 데이터의 개수가 임계 개수 미만인 경우, 최소 샘플 수를 제1 데이터 값보다 큰 제2 데이터 값으로 조절할 수 있다.

[0017] 본 발명의 일실시예에 따른 트래픽 분류 방법은 데이터 생성부에서 기설정된 영역에서 발생되는 트래픽 데이터에 기반하여 제1 트래픽 데이터 세트를 생성하는 단계와, 전처리부에서 기설정된 타겟 지역에 대응되는 지역 특성 정보 및 시간 특성 정보에 기반하여 제1 트래픽 데이터 세트에 포함된 트래픽 데이터 각각을 전처리하는 단계와, 결정 트리 학습부에서 타겟 지역에 대응되는 트래픽을 분류하는 결정 트리 알고리즘을 전처리된 트래픽 데이터에 기반하여 학습하는 단계와, 데이터 생성부에서 기설정된 영역에서 발생되는 트래픽 데이터에 기반하여 제2 트래픽 데이터 세트를 생성하는 단계 및 트래픽 분류부에서 제2 트래픽 데이터 세트를 학습된 결정 트리 알고리즘에 입력하여 타겟 지역에 대응되는 적어도 하나의 트래픽 데이터를 도출하는 단계를 포함할 수 있다.

발명의 효과

[0018] 일실시예에 따르면, 본 발명은 트래픽 데이터 세트를 입력으로 하는 결정 트리 알고리즘을 이용하여 타겟 지역에 대응되는 트래픽 데이터를 추출할 수 있다.

[0019] 일실시예에 따르면, 본 발명은 타겟 지역에 대응되는 지역 특성 및 시간 특성을 고려한 전처리 과정을 통해 타겟 지역에 대응되는 트래픽 데이터를 보다 정확하게 추출할 수 있다.

[0020] 일실시예에 따르면, 본 발명은 전처리된 트래픽 데이터에 기반하여 결정 트리 알고리즘의 리프 노드를 최적화하여 타겟 지역에 대응되는 트래픽 데이터를 보다 정확하게 추출할 수 있다.

도면의 간단한 설명

[0021] 도 1은 일실시예에 따른 트래픽 분류 장치를 설명하는 도면이다.

도 2는 일실시예에 따른 트래픽 분류 장치에서 트래픽 데이터 세트를 생성하는 예시를 설명하는 도면이다.

도 3a 내지 도 3b는 타겟 지역에 따른 시간 특성 정보를 설명하기 위한 도면이다.

도 4는 일실시예에 따른 결정 트리 알고리즘을 설명하는 도면이다.

도 5a 내지 도 5d는 일실시예에 따른 트래픽 분류 장치에서 리프 노드의 최소 샘플 수를 최적화하는 예시를 설명하는 도면이다.

도 6은 일실시예에 따른 트래픽 분류 방법을 설명하는 도면이다.

발명을 실시하기 위한 구체적인 내용

[0022] 본 명세서에 개시되어 있는 본 발명의 개념에 따른 실시예들에 대해서 특정한 구조적 또는 기능적 설명들은 단지 본 발명의 개념에 따른 실시예들을 설명하기 위한 목적으로 예시된 것으로서, 본 발명의 개념에 따른 실시예들은 다양한 형태로 실시될 수 있으며 본 명세서에 설명된 실시예들에 한정되지 않는다.

[0023] 본 발명의 개념에 따른 실시예들은 다양한 변경들을 가할 수 있고 여러 가지 형태들을 가질 수 있으므로 실시예들을 도면에 예시하고 본 명세서에 상세하게 설명하고자 한다. 그러나, 이는 본 발명의 개념에 따른 실시예들을 특정한 개시형태들에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 변경, 균등물, 또는 대체물을 포함한다.

[0024] 제1 또는 제2 등의 용어를 다양한 구성요소들을 설명하는데 사용될 수 있지만, 구성요소들은 용어들에 의해 한정되어서는 안 된다. 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만, 예를 들면 본 발명의 개념에 따른 권리 범위로부터 이탈되지 않은 채, 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소는 제1 구성요소로도 명명될 수 있다.

[0025] 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. 구성요소들 간의 관계를 설명하는 표현들, 예를 들면 "~사이에"와 "바로~사이에" 또는 "~에 직접 이웃하는" 등도 마찬가지로 해석되어야 한다.

[0026] 본 명세서에서 사용한 용어는 단지 특정한 실시예들을 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 설시된 특징, 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것이 존재함으로 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0027] 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 가진다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 의미를 갖는 것으로 해석되어야 하며, 본 명세서에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.

[0029] 이하, 실시예들을 첨부된 도면을 참조하여 상세하게 설명한다. 그러나, 특허출원의 범위가 이러한 실시예들에 의해 제한되거나 한정되는 것은 아니다. 각 도면에 제시된 동일한 참조 부호는 동일한 부재를 나타낸다.

[0031] 도 1은 일실시예에 따른 트래픽 분류 장치를 설명하는 도면이다.

[0032] 도 1을 참조하면, 일실시예에 따른 트래픽 분류 장치(100)는 트래픽 데이터 세트를 입력으로 하여 기설정된 타겟 지역에 대응되는 트래픽을 분류하는 결정 트리 알고리즘을 이용하여 타겟 지역에 대응되는 트래픽 데이터를 추출할 수 있다.

[0033] 또한, 트래픽 분류 장치(100)는 타겟 지역에 대응되는 지역 특성 및 시간 특성을 고려한 전처리 과정을 통해 타겟 지역에 대응되는 트래픽 데이터를 보다 정확하게 추출할 수 있다.

[0034] 또한, 트래픽 분류 장치(100)는 전처리된 트래픽 데이터에 기반하여 결정 트리 알고리즘의 리프 노드를 최적화

하여 타겟 지역에 대응되는 트래픽 데이터를 보다 정확하게 추출할 수 있다.

[0035] 구체적으로, 트래픽 분류 장치(100)는 특정 장소(타겟 지역)에서 발생하는 트래픽 데이터가 가지는 고유 특성을 추출하고, 추출된 고유 특성에 기반하여 타겟 지역에 대응되는 트래픽 데이터를 분류할 수 있다.

[0036] 예를 들면, 트래픽 데이터는 타겟 지역이 대학 구내인 경우, 대학생들의 평균 취침시간 이후에는 트래픽 발생량이 줄어들고 평균 기상시간 이후에는 트래픽 발생이 늘어나는 등의 특징적인 패턴을 나타내며, 트래픽 분류 장치(100)는 전체 트래픽 데이터 중 이러한 고유 특성에 따른 패턴을 갖는 트래픽 데이터만을 이용하여 결정 트리 알고리즘을 학습함으로써, 보다 정확하게 트래픽 데이터를 분류할 수 있다.

[0037] 이를 위해, 트래픽 분류 장치(100)는 데이터 생성부(101), 전처리부(102), 결정 트리 학습부(103) 및 트래픽 분류부(104)를 포함할 수 있다.

[0038] 일실시예에 따른 데이터 생성부(101)는 기설정된 영역에서 발생되는 트래픽 데이터에 기반하여 제1 트래픽 데이터 세트 및 제2 트래픽 데이터 세트를 생성할 수 있다.

[0039] 예를 들면, 트래픽 데이터는 기설정된 영역에서 발생되는 모바일 트래픽 데이터일 수 있다.

[0040] 또한, 제1 트래픽 데이터 세트는 결정 트리 알고리즘을 학습시키기 위한 학습 데이터를 포함하고, 제2 트래픽 데이터 세트는 학습된 결정 트리 알고리즘을 이용하여 트래픽을 분류하고자 하는 타겟 데이터를 포함할 수 있다.

[0041] 일측에 따르면, 데이터 생성부(101)는 동일한 영역에서 서로 다른 시점에 발생되는 트래픽 데이터에 기반하여 제1 트래픽 데이터 세트 및 제2 트래픽 데이터 세트를 생성할 수 있다.

[0042] 또한, 데이터 생성부(101)는 제1 영역에서 발생되는 트래픽 데이터에 기반하여 제1 트래픽 데이터 세트를 생성하고, 제2 영역에서 발생되는 트래픽 데이터에 기반하여 제2 트래픽 데이터 세트를 생성할 수도 있다.

[0043] 일측에 따르면, 제1 트래픽 데이터 세트는 학습된 결정 트리 알고리즘을 테스트하는 테스트 데이터를 더 포함할 수도 있다. 예를 들면, 제1 트래픽 데이터는 10,000개의 트래픽 데이터로 구성될 수 있으며, 이 중 8,000개는 학습 데이터이고, 2,000개는 테스트 데이터일 수 있다.

[0044] 일측에 따르면, 데이터 생성부(101)는 기설정된 영역을 기설정된 크기를 갖는 복수의 그리드(grids)로 분할하고, 분할된 그리드 마다 기설정된 시간 단위로 수집되는 트래픽 데이터에 기반하는 제1 트래픽 데이터 세트 및 제2 트래픽 데이터 세트를 생성할 수 있다.

[0045] 즉, 제1 트래픽 데이터 세트 및 제2 트래픽 데이터 세트에 포함된 트래픽 데이터 각각은 그리드 단위 각각에서 시계열적으로 누적된 트래픽 정보를 의미할 수 있다.

[0046] 일실시예에 따른 전처리부(102)는 타겟 지역에 대응되는 지역 특성 정보 및 시간 특성 정보에 기반하여 제1 트래픽 데이터 세트에 포함된 트래픽 데이터 각각을 전처리할 수 있다.

[0047] 예를 들면, 전처리부(102)는 타겟 지역에 대응되는 시간 특성 정보에 기반하여 제1 트래픽 데이터 세트를 1차 전처리하고, 1차 전처리된 트래픽 데이터 세트를 지역 특성 정보에 기반하여 제2 전처리하여 최종적으로 전처리된 트래픽 데이터를 출력할 수 있다.

[0048] 또한, 전처리부(102)는 타겟 지역에 대응되는 지역 특성 정보에 기반하여 제1 트래픽 데이터 세트를 1차 전처리하고, 1차 전처리된 트래픽 데이터 세트를 시간 특성 정보에 기반하여 제2 전처리하여 최종적으로 전처리된 트래픽 데이터를 출력할 수도 있다.

[0049] 일측에 따르면, 전처리부(102)는 제1 트래픽 데이터 세트에 포함된 트래픽 데이터 중 적어도 하나의 트래픽 데이터에 타겟 지역에 대응되는 트래픽 데이터를 식별하기 위한 인덱스 정보를 반영할 수 있다.

[0050] 다시 말해, 전처리부(102)는 결정 트리 알고리즘의 지도 학습을 위해 데이터를 인덱싱할 수 있다.

[0051] 예를 들면, 전처리부(102)는 사용자가 사전에 인지하고 있는 정보에 기반하여 제1 트래픽 데이터 세트에 포함된 트래픽 데이터 중 타겟 지역에 대응되는 트래픽 데이터에는 'True'를 인덱싱하고, 타겟 지역에 대응되지 않는 트래픽 데이터에는 'False'를 인덱싱할 수 있다.

[0052] 일측에 따르면, 전처리부(102)는 지역 특성 정보에 따른 제1 트래픽 데이터 세트의 기지국 정보에 기반하여 제1 트래픽 데이터 세트에 포함된 트래픽 데이터 중 적어도 하나의 트래픽 데이터에 인덱스 정보를 반영할 수 있다.

[0053] 예를 들면, 전처리부(102)는 제1 트래픽 데이터 세트에 포함된 트래픽 데이터 각각을 수신하는 기지국의 위치 정보에 기반하여, 제1 트래픽 데이터 세트에 포함된 트래픽 데이터 중 타겟 지역에 위치한 기지국에서 수신한 트래픽 데이터에는 'True'를 인덱싱하고, 타겟 지역 이외의 지역에 위치한 기지국에서 수신한 트래픽 데이터에는 'False'를 인덱싱할 수 있다.

[0054] 또한, 전처리부(102)는 사용자가 사전에 인지하고 있는 정보에 기반하여 'False'로 인덱싱된 트래픽 데이터 중 타겟 지역에 위치한 기지국에서 수신한 트래픽 데이터를 'True'로 다시 인덱싱할 수도 있다.

[0055] 일측에 따르면, 전처리부(102)는 제1 트래픽 데이터 세트에 포함된 트래픽 데이터 중 시간 특성 정보에 대응되는 시간에 수집된 트래픽 데이터를 추출할 수 있다.

[0056] 구체적으로, 전처리부(102)는 타겟 지역이 대학 지역, 상업 지역 및 주거 지역 중 적어도 하나인 경우, 제1 트래픽 데이터 세트에 포함된 트래픽 데이터 중 평일에 수집된 트래픽 데이터를 추출할 수 있다.

[0057] 또한, 전처리부(102)는 타겟 지역이 상업 지역인 경우, 제1 트래픽 데이터 세트에 포함된 트래픽 데이터 중 10시 내지 16시에 수집된 트래픽 데이터를 추출하고, 타겟 지역이 주거 지역인 경우, 18시 내지 24시에 수집된 트래픽 데이터를 추출할 수 있다.

[0058] 바람직하게는, 전처리부(102)는 타겟 지역이 상업 지역인 경우, 제1 트래픽 데이터 세트에 포함된 트래픽 데이터 중 평일 10시 내지 16시에 수집된 트래픽 데이터만을 시간 특성 정보에 대응하여 전처리된 트래픽 데이터로 추출할 수 있다.

[0059] 또한, 전처리부(102)는 타겟 지역이 주거 지역인 경우, 제1 트래픽 데이터 세트에 포함된 트래픽 데이터 중 평일 18시 내지 24시에 수집된 트래픽 데이터만을 시간 특성 정보에 대응하여 전처리된 트래픽 데이터로 추출할 수 있다.

[0060] 일실시예에 따른 결정 트리 학습부(103)는 결정 트리 알고리즘을 전처리된 트래픽 데이터에 기반하여 학습할 수 있다.

[0061] 구체적으로, 결정 트리 알고리즘은 입력되는 데이터들 속에서 규칙을 찾아내 트리 기반의 분류 규칙을 만들어 가지가 뻗어나가는 형태로 데이터들을 분류하는 알고리즘으로, 최초 루트 노드에서 새로운 규칙 조건이 발생할 때마다 가지치기(pruning)를 수행하여 서브 트리를 생성하며, 마지막으로 더 이상 가지치기를 하지 않는 리프 노드에서 결정된 분류 값을 출력할 수 있다.

[0062] 일측에 따르면, 결정 트리 학습부(103)는 전처리된 트래픽 데이터 중 타겟 지역에 대응되는 트래픽 데이터의 개수에 대응하여 결정 트리 알고리즘의 리프 노드 각각의 최소 샘플 수를 조절할 수 있다.

[0063] 예를 들면, 결정 트리 학습부(103)는 전처리된 트래픽 데이터 중 'True'로 인덱싱된 트래픽 데이터의 개수에 기반하여 리프 노드 각각의 최소 샘플 수를 조절할 수 있다.

[0064] 여기서, 리프 노드의 최소 샘플 수는 리프 노드에 포함되는 트래픽 데이터(그리드)의 개수를 의미하며, 결정 트리 알고리즘에 포함된 노드 각각은 트래픽 데이터의 개수가 최소 샘플 수 이하가 될때까지 가지치기를 수행할 수 있다.

[0065] 예를 들면, 결정 트리 학습부(103)는 타겟 지역이 대학 지역이고, 최소 샘플 수가 '20'개로 설정되며, 결정 트리 알고리즘의 가지치기 동작 과정에서 어느 하나의 노드에서 분류된 트래픽 데이터(그리드)의 수가 '대학 지역인 데이터: 16개, 대학 지역이 아닌 데이터: 4개'이면, 어느 하나의 노드를 리프 노드로 설정하여 해당 노드에서의 가지치기를 중단할 수 있다.

[0066] 일측에 따르면, 결정 트리 학습부(103)는 전처리된 트래픽 데이터 중 타겟 지역에 대응되는 트래픽 데이터의 개수가 기설정된 임계 개수 이상인 경우, 최소 샘플 수를 제1 데이터 값으로 조절할 수 있다.

[0067] 또한, 결정 트리 학습부(103)는 전처리된 트래픽 데이터 중 타겟 지역에 대응되는 트래픽 데이터의 개수가 임계 개수 미만인 경우, 최소 샘플 수를 제1 데이터 값보다 큰 제2 데이터 값으로 조절할 수 있다.

[0068] 예를 들면, 기설정된 임계 개수는 100개, 제1 데이터 값은 3개 내지 10개, 제2 데이터 값은 20개 내지 40개일 수 있다.

[0069] 한편, 결정 트리 학습부(103)는 타겟 지역이 대학 지역이고, 최소 샘플 수가 '20'개로 설정되며, 결정 트리 알고리즘의 가지치기 동작 과정에서 어느 하나의 노드에서 분류된 트래픽 데이터(그리드)의 수가 '대학 지역인 데

이터: 30개, 대학 지역이 아닌 데이터: 0개'이면, 어느 하나의 노드는 완전히 분류가 종료된 노드이므로 최소 샘플 수와 상관없이 분류가 완료된 어느 하나의 노드를 리프 노드로 설정하여 해당 노드에서의 가지치기를 중단할 수도 있다.

[0070] 즉, 결정 트리 학습부(103)는 오버피팅(overfitting)을 방지하고 분류의 정확도를 향상시키기 위해 타겟 지역에 대응되는 트래픽 데이터의 개수에 대응하여 결정 트리 알고리즘에서의 가지치기 수준을 최적화할 수 있다.

[0071] 일실시예에 따른 트래픽 분류부(104)는 제2 트래픽 데이터 세트를 학습된 결정 트리 알고리즘에 입력하여 타겟 지역에 대응되는 적어도 하나의 트래픽 데이터를 도출할 수 있다.

[0072] 다시 말해, 트래픽 분류부(104)는 학습된 결정 트리 알고리즘에 기반하여 제2 트래픽 데이터 세트에 포함된 트래픽 데이터 각각을 타겟 지역에 대응되는 트래픽 데이터와 타겟 지역에 대응되지 않는 트래픽 데이터로 분류할 수 있다.

[0074] 도 2는 일실시예에 따른 트래픽 분류 장치에서 트래픽 데이터 세트를 생성하는 예시를 설명하는 도면이다.

[0075] 도 2를 참조하면, 트래픽 분류 장치는 도면부호 200에 도시된 바와 같이 기설정된 영역에서 발생되는 트래픽 데이터에 기반하여 트래픽 데이터 세트를 생성할 수 있다.

[0076] 일측에 따르면, 트래픽 분류 장치는 기설정된 영역을 기설정된 크기를 갖는 복수의 그리드(grids)로 분할하고, 분할된 그리드 마다 기설정된 시간 단위로 수집되는 트래픽 데이터에 기반하는 트래픽 데이터 세트를 생성할 수 있다.

[0077] 도면부호 200의 예시에 따르면, 트래픽 분류 장치는 이탈리아 밀라노 지역을 100 x 100 그리드(즉, 10,000개의 그리드)로 분할하고, 2013년 11월 1일부터 2014년 1월 1일까지 62일간 1시간 단위로 누적시킨 트래픽 데이터를 포함하는 트래픽 데이터 세트를 형성할 수 있다.

[0078] 또한, 트래픽 분류 장치는 사용자가 사전에 인지하고 있는 정보에 기반하여 10,000개의 트래픽 데이터(그리드) 중 타겟 지역인 대학 지역에 대응되는 104개의 트래픽 데이터를 도출하였으며, 이를 통해 104개의 트래픽 데이터에는 'True'를 인덱싱하고, 나머지 트래픽 데이터에는 'False'를 인덱싱할 수 있다.

[0080] 도 3a 내지 도 3b는 타겟 지역에 따른 시간 특성 정보를 설명하기 위한 도면이다.

[0081] 도 3a 내지 도 3b를 참조하면, 도면부호 310은 주거 지역 및 상업 지역에서 평일 및 주말에 발생되는 업링크(Uplink) 트래픽량(Data Traffic Volum)의 분석 결과를 도시하고, 도면부호 320은 주거 지역 및 상업 지역에서 하루(0시 내지 24시) 동안 발생되는 트래픽량(Data Traffic Volum)의 분석 결과를 도시한다.

[0082] 도면부호 310에 따르면, 주거 지역 및 상업 지역 모두 고유 특성(트래픽 특징)이 주말 구간(도면부호 311) 보다는 평일 구간에서 보다 뚜렷하게 드러나는 것으로 나타났다.

[0083] 한편, 도 2를 통해 설명한 밀라노 지역의 트래픽 데이터 중 대학교 트래픽을 평일과 주말로 나눠서 분석한 결과, 주말은 대학생들의 생활 패턴의 영향이 약해 고유 특성이 드러나지 않는 것으로 나타났다.

[0084] 또한, 평일 및 주말에 수집된 트래픽 데이터 10,000개로 구성된 트래픽 데이터 세트(평일+주말 데이터 세트)와, 평일에 수집된 트래픽 데이터 10,000개로 구성된 트래픽 데이터 세트(평일 데이터 세트) 각각을 이용하여 결정 트리 알고리즘을 학습한 결과, '평일+주말 데이터 세트'로 학습된 결정 트리 알고리즘의 분류 정확도는 약 98.12%, '평일 데이터 세트'로 학습된 결정 트리 알고리즘의 분류 정확도는 약 98.45%로, 평일에 수집된 트래픽 데이터만을 이용하여 학습을 진행하였을 때 분류의 정확도가 0.3% 향상된 것으로 나타났다.

[0085] 다시 말해, 주거 지역, 상업 지역 및 대학 지역은 주말 보다는 평일에 수집된 트래픽 데이터에서 고유 특성이 드러나는 것으로 나타났다.

[0086] 다시 말해, 타겟 지역이 주거 지역, 상업 지역 및 대학 지역 중 적어도 하나인 경우에는 평일에 수집된 트래픽 데이터만을 추출하여 결정 트리 알고리즘을 학습하면 보다 정확하게 트래픽을 분류할 수 있다.

[0087] 다음으로, 도면부호 320에 따르면, 주거 지역 및 상업 지역은 서로 트래픽이 최대가 되는 지점이 다르게 나타났으며(주거 지역: 20시, 상업 지역: 13시), 상업 지역은 하루 중 10시 내지 16시, 주거 지역은 하루 중 18시 내지 24시에 수집된 트래픽 데이터에서 고유 특성이 보다 뚜렷하게 드러나는 것으로 나타났다.

[0088] 다시 말해, 상업 지역은 10시 내지 16시, 주거 지역은 18시 내지 24시에 수집된 트래픽 데이터만을 추출하여 결

정 트리 알고리즘을 학습하면 보다 정확하게 트래픽을 분류할 수 있다.

[0090] 도 4는 일실시예에 따른 결정 트리 알고리즘을 설명하는 도면이다.

[0091] 도 4를 참조하면, 일실시예에 따른 트래픽 분류 장치는 전처리된 트래픽 데이터를 이용하여 도면부호 400에 도시된 바와 같이 결정 트리 알고리즘을 학습하고, 학습된 결정 트리 알고리즘에 기반하여 타겟 트래픽 데이터를 분류할 수 있다.

[0092] 도면부호 400에 도시된 'samples'는 대응되는 노드에 포함된 트래픽 데이터(그리드)의 개수, 'Value = [A, B]'에서 'A'는 대응되는 노드에 포함된 트래픽 데이터 중 결과 값이 False인 트래픽 데이터(즉, 타겟 지역에 대응되지 않는 트래픽 데이터)의 개수, 'B'는 대응되는 노드에 포함된 트래픽 데이터 중 결과 값이 True인 트래픽 데이터(즉, 타겟 지역에 대응되는 트래픽 데이터)의 개수를 의미한다.

[0093] 또한, 최상단에 위치한 노드(즉, 루트 노드)를 예시로, '471 <= 1789.217'은 471번째 데이터, 즉 471번째 시간 트래픽의 양이 1790.217 보다 작은 경우(즉, 규칙 조건)을 의미하며, 'gini'는 지니 계수로 불균형 지수를 의미한다.

[0094] 구체적으로, 결정 트리 알고리즘은 도면부호 410에 도시된 바와 같이 8,000개의 트래픽 데이터를 구비하는 루트 노드를 형성할 수 있고, 도면부호 420에 도시된 바와 같이 1차 가지치기 동작을 수행하여 2개의 노드(7,519개 및 481개의 트래픽 데이터를 구비하는 노드)를 형성할 수 있으며, 도면부호 430에 2차 가지치기 동작을 통해 4개의 노드(7495개, 24개, 6개, 475개의 트래픽 데이터를 구비하는 노드)를 형성할 수 있다.

[0095] 또한, 결정 트리 알고리즘은 도면부호 440에 도시된 바와 같이 분류가 최종 완료될 때까지, 즉 리프 노드의 형성이 완료될 때까지 가지치기 동작을 복수 회 반복 수행할 수 있다.

[0096] 한편, 일실시예에 따른 트래픽 분류 장치는 전처리된 트래픽 데이터 중 타겟 지역에 대응되는 트래픽 데이터의 개수에 대응하여 결정 트리 알고리즘의 리프 노드 각각의 최소 샘플 수를 조절할 수 있다.

[0097] 예를 들면, 트래픽 분류 장치는 전처리된 트래픽 데이터 중 타겟 지역에 대응되는 트래픽 데이터의 개수에 대응하여 최소 샘플 수를 10개로 설정할 수 있다.

[0098] 즉, 트래픽 분류 장치는 도면부호 430에 도시된 바와 같이 최소 샘플 수(10개) 보다 적은 6개의 트래픽 데이터를 구비하는 노드를 리프 노드로 설정하여 해당 노드에서의 가지치기를 중단할 수 있다.

[0100] 도 5a 내지 도 5d는 일실시예에 따른 트래픽 분류 장치에서 리프 노드의 최소 샘플 수를 최적화하는 예시를 설명하는 도면이다.

[0101] 도 5a 내지 도 5d를 참조하면, 도면부호 510 및 520은 타겟 지역에 대응되는 트래픽 개수가 임계 개수 이상인 서로 다른 전처리된 트래픽 데이터를 입력으로 하여 결정 트리 알고리즘을 학습할 때, 리프 노드의 최소 샘플 수의 변화(spit 3, 5, 10, 20, 40)에 따른 학습된 결정 트리 알고리즘의 분류 정확도 특성을 도시한다.

[0102] 또한, 도면부호 530 및 540은 타겟 지역에 대응되는 트래픽 개수가 임계 개수 미만인 서로 다른 전처리된 트래픽 데이터를 입력으로 하여 결정 트리 알고리즘을 학습할 때, 리프 노드의 최소 샘플 수의 변화(spit 3, 5, 10, 20, 40)에 따른 학습된 결정 트리 알고리즘의 분류 정확도 특성을 도시한다.

[0103] 구체적으로, 도면부호 510 및 520은 2,000개의 전처리된 트래픽 데이터 중 104개의 트래픽 데이터가 대학 지역에 대응되는 트래픽 데이터일 때의 결과를 도시하고, 도면부호 530 및 540은 2,000개의 전처리된 트래픽 데이터 중 40개의 트래픽 데이터가 대학 지역에 대응되는 트래픽 데이터일 때의 결과를 도시한다.

[0104] 도면부호 510 및 520에 따르면, 대학 지역에 대응되는 트래픽 데이터가 104개인 전처리된 트래픽 데이터를 입력으로 수신하는 결정 트리 알고리즘은 최소 샘플 수가 3개(split 3) 내지 10개(split 3)로 설정되어 학습이 되는 경우에는 93% 이상의 분류 정확도를 보이나, 최소 샘플 수가 20개(split 20) 내지 40개(split 40)으로 설정되어 학습이 되는 경우에는 분류 정확도가 93% 아래로 감소하는 것을 확인할 수 있다.

[0105] 또한, 대학 지역에 대응되는 트래픽 데이터가 104개인 전처리된 트래픽 데이터를 입력으로 수신하는 결정 트리 알고리즘은 최소 샘플 수가 작을수록 분류 정확도가 향상되는 것을 확인할 수 있다.

[0106] 다음으로, 도면부호 530 및 540에 따르면, 대학 지역에 대응되는 트래픽 데이터가 40개인 전처리된 트래픽 데이터를 입력으로 수신하는 결정 트리 알고리즘은 통상적으로 최소 샘플 수가 3개(split 3) 내지 10개(split 3)일 때보다 20개(split 20) 내지 40개(split 40)일 때 분류의 정확도가 향상되는 것을 확인할 수 있다.

[0107] 다시 말해, 일실시예에 따른 결정 트리 알고리즘은 분류하고 싶은 인덱스 데이터가 적절한 경우(즉, 타겟 지역에 대응되는 트래픽 개수가 임계 개수 이상인 경우)에는 최소 샘플 수가 클수록 분류의 경확도가 향상되고, 분류하고 싶은 인덱스 데이터가 부족한 경우(즉, 타겟 지역에 대응되는 트래픽 개수가 임계 개수 이하인 경우)에는 최소 샘플 수가 작을수록 분류의 정확도가 향상되는 것으로 나타났다.

[0109] 도 6은 일실시예에 따른 트래픽 분류 방법을 설명하는 도면이다.

[0110] 다시 말해, 도 6은 도 1 내지 도 5d를 통해 설명하는 일실시예에 따른 트래픽 분류 장치의 동작 방법을 설명하는 도면으로, 이하에서 도 6을 통해 설명하는 내용 중 도 1 내지 도 5d를 통해 설명한 내용과 중복되는 설명은 생략하기로 한다.

[0111] 도 6을 참조하면, 610 단계에서 일실시예에 따른 트래픽 분류 방법은 데이터 생성부에서 기설정된 영역에서 발생되는 트래픽 데이터에 기반하여 제1 트래픽 데이터 세트를 생성할 수 있다.

[0112] 일측에 따르면, 610 단계에서 일실시예에 따른 트래픽 분류 방법은 데이터 생성부에서 기설정된 영역을 기설정된 크기를 갖는 복수의 그리드(grids)로 분할하고, 분할된 그리드 마다 기설정된 시간 단위로 수집되는 트래픽 데이터에 기반하는 제1 트래픽 데이터 세트를 생성할 수 있다.

[0113] 다음으로, 620 단계에서 일실시예에 따른 트래픽 분류 방법은 전처리부에서 기설정된 타겟 지역에 대응되는 지역 특성 정보 및 시간 특성 정보에 기반하여 제1 트래픽 데이터 세트에 포함된 트래픽 데이터 각각을 전처리할 수 있다.

[0114] 일측에 따르면, 620 단계에서 일실시예에 따른 트래픽 분류 방법은 전처리부에서 제1 트래픽 데이터 세트에 포함된 트래픽 데이터 중 적어도 하나의 트래픽 데이터에 타겟 지역에 대응되는 트래픽 데이터를 식별하기 위한 인덱스 정보를 반영할 수 있다.

[0115] 일측에 따르면, 620 단계에서 일실시예에 따른 트래픽 분류 방법은 전처리부에서 지역 특성 정보에 따른 제1 트래픽 데이터 세트의 기지국 정보에 기반하여 제1 트래픽 데이터 세트에 포함된 트래픽 데이터 중 적어도 하나의 트래픽 데이터에 인덱스 정보를 반영할 수 있다.

[0116] 또한, 620 단계에서 일실시예에 따른 트래픽 분류 방법은 전처리부에서 제1 트래픽 데이터 세트에 포함된 트래픽 데이터 중 시간 특성 정보에 대응되는 시간에 수집된 트래픽 데이터를 추출할 수 있다.

[0117] 구체적으로, 620 단계에서 일실시예에 따른 트래픽 분류 방법은 타겟 지역이 대학 지역, 상업 지역 및 주거 지역 중 적어도 하나인 경우, 전처리부에서 제1 트래픽 데이터 세트에 포함된 트래픽 데이터 중 평일에 수집된 트래픽 데이터를 추출할 수 있다.

[0118] 또한, 620 단계에서 일실시예에 따른 트래픽 분류 방법은 타겟 지역이 상업 지역인 경우, 전처리부에서 제1 트래픽 데이터 세트에 포함된 트래픽 데이터 중 10시 내지 16시에 수집된 트래픽 데이터를 추출할 수 있다.

[0119] 또한, 620 단계에서 일실시예에 따른 트래픽 분류 방법은 타겟 지역이 주거 지역인 경우, 전처리부에서 18시 내지 24시에 수집된 트래픽 데이터를 추출할 수 있다.

[0120] 다음으로, 630 단계에서 일실시예에 따른 트래픽 분류 방법은 결정 트리 학습부에서 타겟 지역에 대응되는 트래픽을 분류하는 결정 트리 알고리즘을 전처리된 트래픽 데이터에 기반하여 학습할 수 있다.

[0121] 일측에 따르면, 630 단계에서 일실시예에 따른 트래픽 분류 방법은 결정 트리 학습부에서 전처리된 트래픽 데이터 중 타겟 지역에 대응되는 트래픽 데이터의 개수에 대응하여 결정 트리 알고리즘의 리프 노드 각각의 최소 샘플 수를 조절할 수 있다.

[0122] 구체적으로, 630 단계에서 일실시예에 따른 트래픽 분류 방법은 전처리된 트래픽 데이터 중 타겟 지역에 대응되는 트래픽 데이터의 개수가 기설정된 임계 개수 이상인 경우, 결정 트리 학습부에서 최소 샘플 수를 제1 데이터 값으로 조절할 수 있다.

[0123] 또한, 630 단계에서 일실시예에 따른 트래픽 분류 방법은 전처리된 트래픽 데이터 중 타겟 지역에 대응되는 트래픽 데이터의 개수가 임계 개수 미만인 경우, 결정 트리 학습부에서 최소 샘플 수를 제1 데이터 값보다 큰 제2 데이터 값으로 조절할 수 있다.

[0124] 다음으로, 640 단계에서 일실시예에 따른 트래픽 분류 방법은 기설정된 영역에서 발생되는 트래픽 데이터에 기반하여 제2 트래픽 데이터 세트를 생성할 수 있다.

[0125] 일측에 따르면, 640 단계에서 일실시예에 따른 트래픽 분류 방법은 데이터 생성부에서 기설정된 영역을 기설정된 크기를 갖는 복수의 그리드(grids)로 분할하고, 분할된 그리드 마다 기설정된 시간 단위로 수집되는 트래픽 데이터에 기반하는 제2 트래픽 데이터 세트를 생성할 수 있다.

[0126] 다음으로, 650 단계에서 일실시예에 따른 트래픽 분류 방법은 트래픽 분류부에서 제2 트래픽 데이터 세트를 학습된 결정 트리 알고리즘에 입력하여 타겟 지역에 대응되는 적어도 하나의 트래픽 데이터를 도출할 수 있다.

[0128] 결국, 본 발명을 이용하면, 트래픽 데이터 세트를 입력으로 하는 결정 트리 알고리즘을 이용하여 타겟 지역에 대응되는 트래픽 데이터를 추출할 수 있다.

[0129] 또한, 본 발명을 이용하면, 타겟 지역에 대응되는 지역 특성 및 시간 특성을 고려한 전처리 과정을 통해 타겟 지역에 대응되는 트래픽 데이터를 보다 정확하게 추출할 수 있다.

[0130] 또한, 본 발명을 이용하면, 전처리된 트래픽 데이터에 기반하여 결정 트리 알고리즘의 리프 노드를 최적화하여 타겟 지역에 대응되는 트래픽 데이터를 보다 정확하게 추출할 수 있다.

[0132] 이상과 같이 실시예들이 비록 한정된 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들면, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 장치, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

[0133] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

부호의 설명

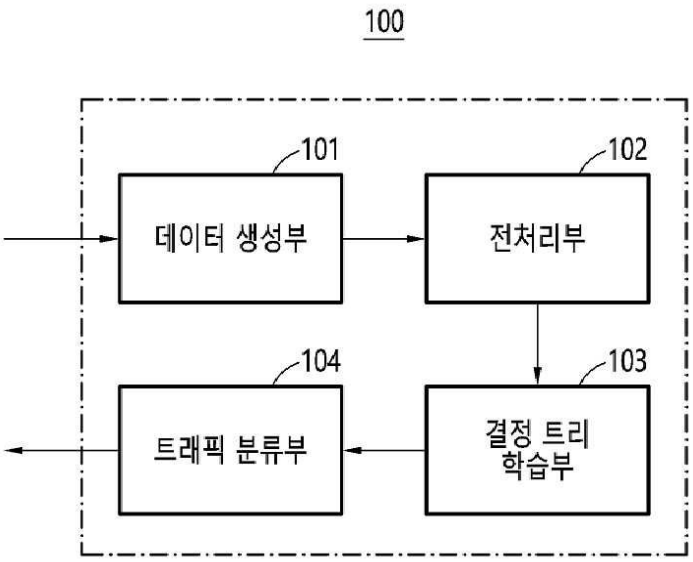
[0134] 100: 트래픽 분류 장치 101: 데이터 생성부

102: 전처리부 103: 결정 트리 학습부

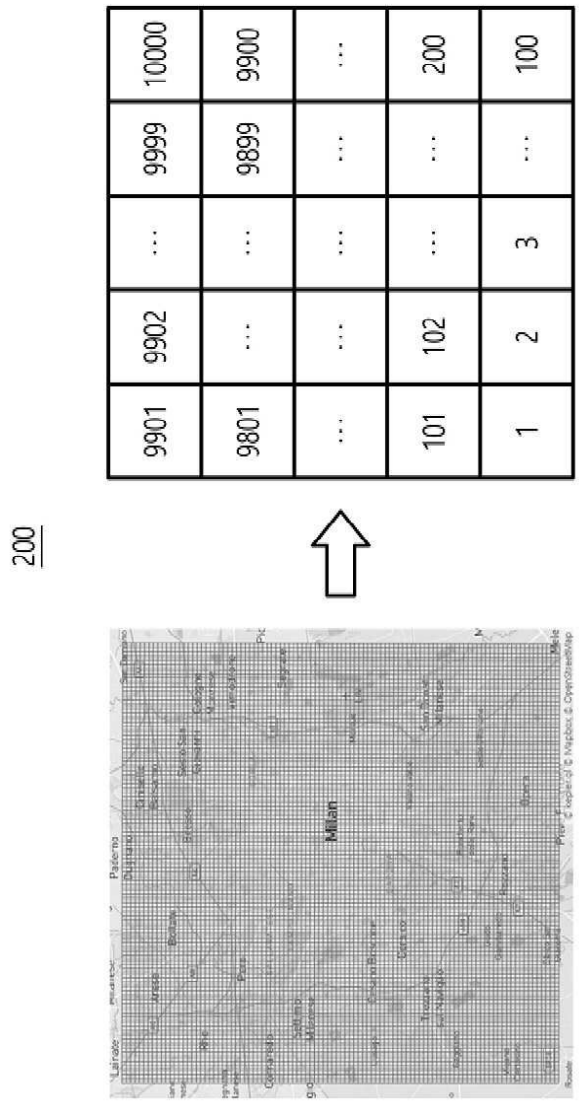
104: 트래픽 분류부

도면

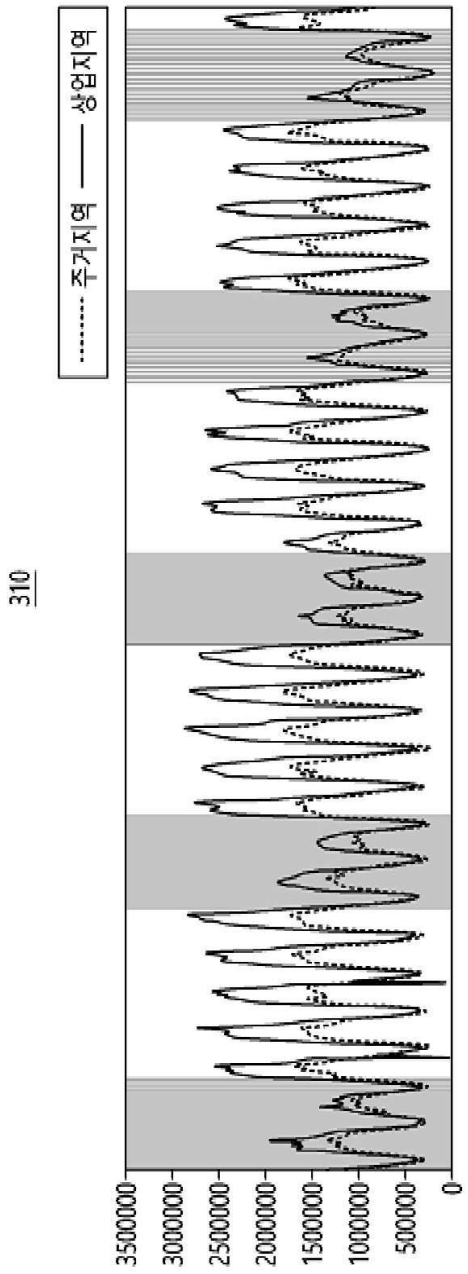
도면1



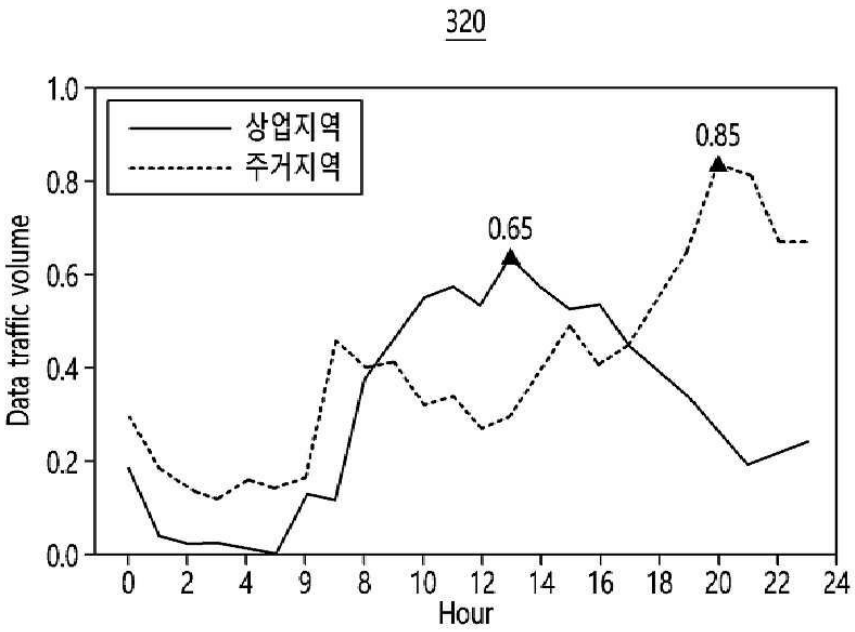
도면2


200
9901
9902
...
9801
...
...
...
101
102
...
1
2
3
...
9999
9899
...
10000
9900
...
200
100
Milan

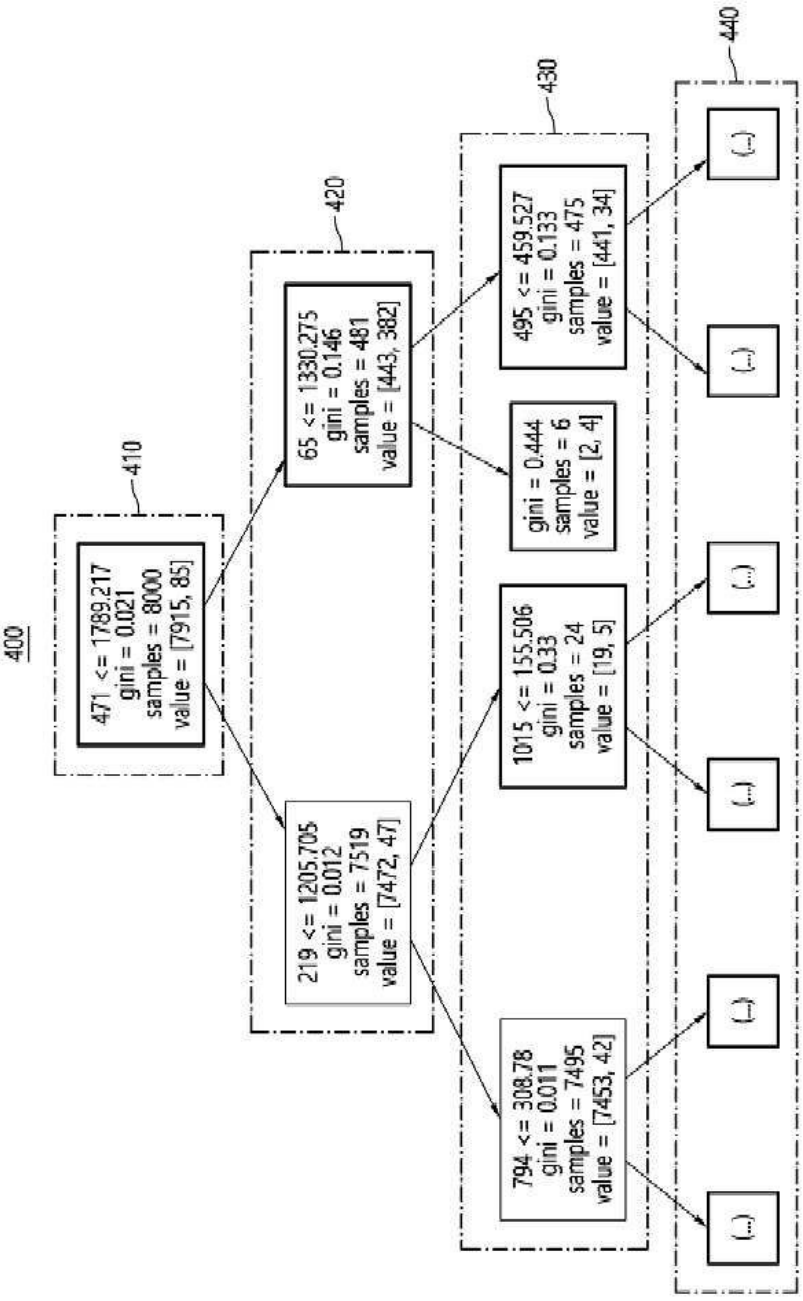
도면3a


310
주거지역
상업지역
3500000
3000000
2500000
2000000
1500000
1000000
500000
0

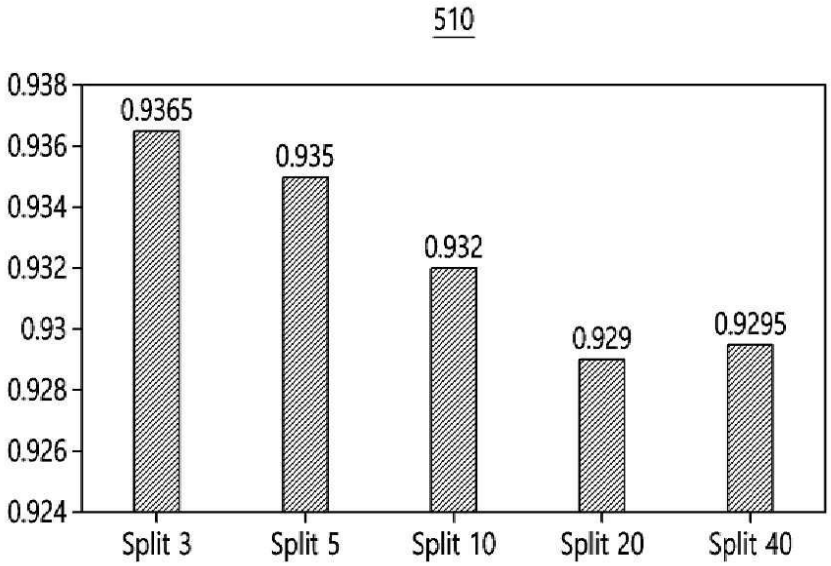
도면3b


320
상업지역
주거지역
0.85
0.65
Data traffic volume
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
9
8
10
12
14
16
18
20
22
24
Hour

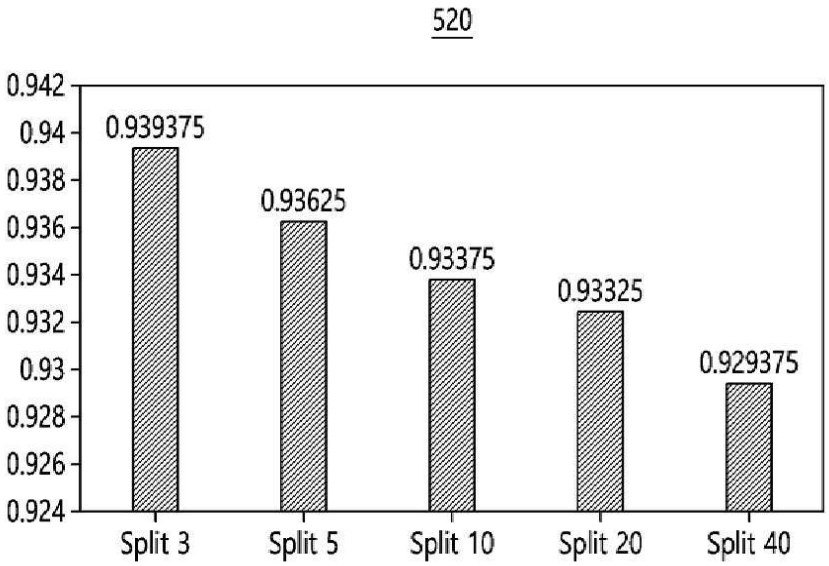
도면4


400
410
471 <= 1789.217
gini = 0.021
samples = 8000
value = [7915, 85]
420
65 <= 1330.275
gini = 0.146
samples = 481
value = [443, 382]
219 <= 1205.705
gini = 0.012
samples = 7519
value = [7472, 47]
430
495 <= 459.527
gini = 0.133
samples = 475
value = [441, 34]
gini = 0.444
samples = 6
value = [2, 4]
1015 <= 155.506
gini = 0.33
samples = 24
value = [19, 5]
794 <= 308.78
gini = 0.011
samples = 7495
value = [7453, 42]
440
(...)
(...)
(...)
(...)
(...)
(...)

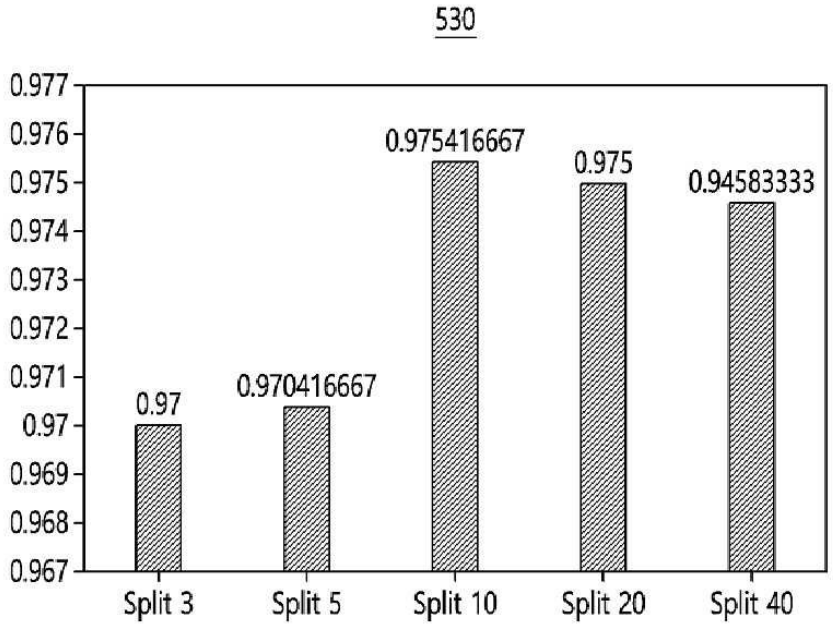
도면5a


510
0.938
0.936
0.934
0.932
0.93
0.928
0.926
0.924
0.9365
0.935
0.932
0.929
0.9295
Split 3
Split 5
Split 10
Split 20
Split 40

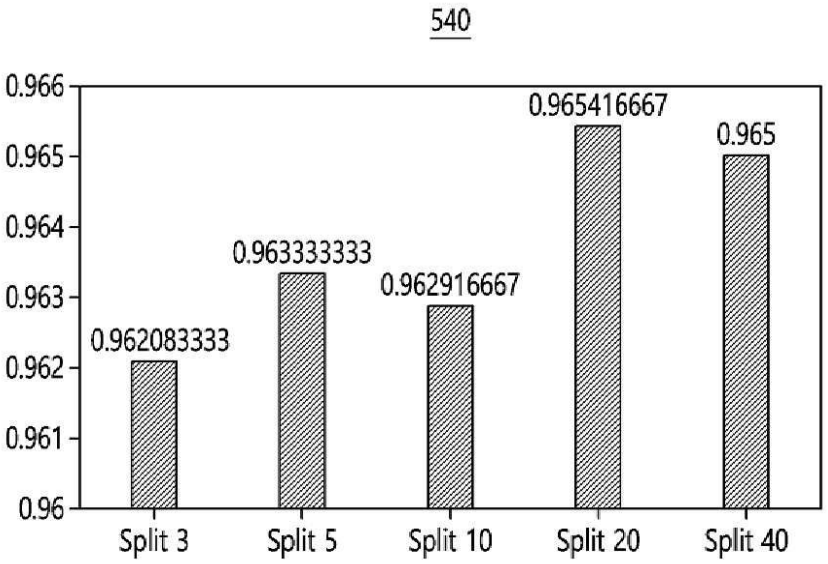
도면5b


520
0.942
0.94
0.938
0.936
0.934
0.932
0.93
0.928
0.926
0.924
0.939375
0.93625
0.93375
0.93325
0.929375
Split 3
Split 5
Split 10
Split 20
Split 40

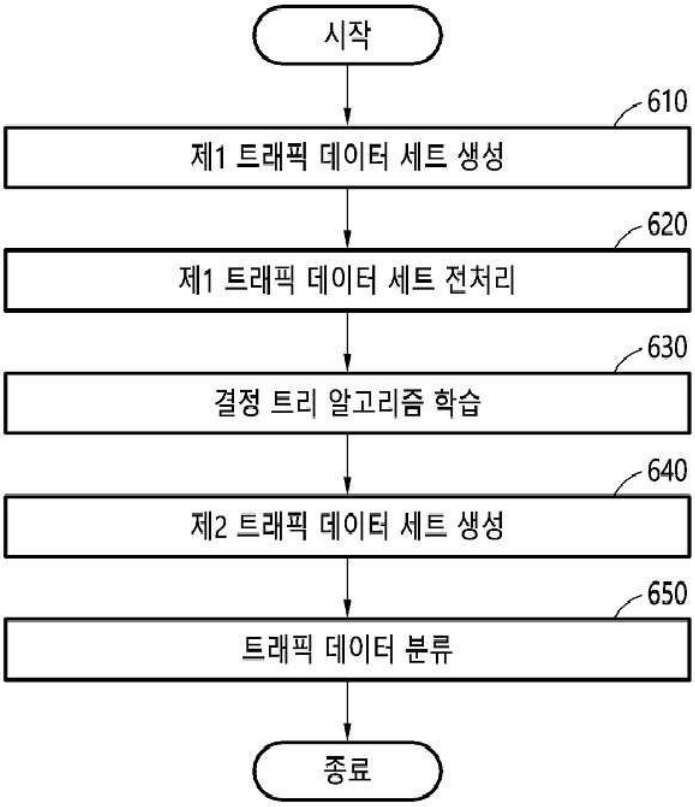
도면5c


530
0.977
0.976
0.975
0.974
0.973
0.972
0.971
0.97
0.969
0.968
0.967
0.97
0.970416667
0.975416667
0.975
0.94583333
Split 3
Split 5
Split 10
Split 20
Split 40

도면5d


540
0.966
0.965
0.964
0.963
0.962
0.961
0.96
0.962083333
0.963333333
0.962916667
0.965416667
0.965
Split 3
Split 5
Split 10
Split 20
Split 40

도면6


시작
610
제1 트래픽 데이터 세트 생성
620
제1 트래픽 데이터 세트 전처리
630
결정 트리 알고리즘 학습
640
제2 트래픽 데이터 세트 생성
650
트래픽 데이터 분류
종료