

교통 혼잡 원인과 영향을 분석하기 위한 시각적 분석 기술

(A Visual Analytics Technique for Analyzing the Cause
and Influence of Traffic Congestion)

피 민 규 [†]
(Mingyu Pi)

연 한 벌 ^{††}
(Hanbyul Yeon)

손 혜 숙 ^{†††}
(Hyesook Son)

장 윤 ^{††††}
(Yun Jang)

요 약 본 논문에서는 교통 흐름 이론을 기반으로 교통 혼잡의 원인을 분석하는 기술을 제시한다. 우리는 GPS 궤적 및 차량 감지기 데이터(VD)와 같은 교통 데이터에서 차량의 흐름을 추출한다. 또한 우리는 교통 데이터에 정보이론의 엔트로피를 사용하여 차량의 흐름 변화를 식별한다. 그런 다음 혼잡 지역의 차량 흐름을 정량화할 수 있는 누적 차량 수 커브(N-curve)를 추출한다. 교통 흐름 이론에 따르면 혼잡 유형에 따라 고유한 N-curve 패턴을 관찰할 수 있다. 우리는 N-curve를 네 가지의 혼잡 패턴으로 분류할 수 있는 합성곱 신경망을 설계한다. 교통 혼잡의 원인과 영향을 분석하는 것은 어렵고 상당한 경험과 지식이 필요하다. 따라서 논문에서는 교통 혼잡의 원인과 영향을 분석하기 위한 일련의 프로세스를 효율적으로 수행할 수 있는 시각적 분석 시스템을 제시한다. 논문에서는 두 가지의 사례 연구를 통해 교통 혼잡의 원인을 분석할 수 있는 시스템을 평가한다.

키워드: 교통 혼잡 원인 분석, 시각적 분석, 교통 흐름 이론, 합성곱 신경망

Abstract In this paper, we present a technique to analyze the causes of traffic congestion based on the traffic flow theory. We extracted vehicle flows from the traffic data, such as GPS trajectory and Vehicle Detector data. Also, vehicle flow changes were identified by utilizing the entropy from the information theory. Then, we extracted cumulative vehicle count curves (N-curve) that can quantify the vehicle flows in the congestion area. According to the traffic flow theory, unique N-curve patterns can be observed depending on the congestion type. We build a convolution neural network classifier that can classify N-curve into four different congestion patterns. Analyzing the cause and influence of congestion is difficult and requires considerable experience and knowledge. Apparently, we present a visual analytics system that can efficiently perform a series of processes to analyze the cause and influence of traffic congestion. Through case studies, we have evaluated our system that can analyze the cause of traffic congestion.

Keywords: traffic congestion causes, visual analytics, traffic flow theory, convolutional neural network

· 본 연구는 2019년도 과학기술정보통신부의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (2019-0-00136, 스마트시티 산업 생산성 혁신을 위한 AI 융합 기술 개발), (No.2019-0-00242, 빅데이터 신뢰성과 활용성 극대화를 위한 빅데이터 증강분석 프로파일링 플랫폼 개발)

[†] 학생회원 : 세종대학교 컴퓨터공학과 학생
pmg9405@naver.com

^{††} 학생회원 : 세종대학교 컴퓨터공학과 연구원
hbyeon109@gmail.com

^{†††} 비 회 원 : 세종대학교 컴퓨터공학과 연구원
atieer@naver.com

^{††††} 종신회원 : 세종대학교 컴퓨터공학과 교수(Sejong Univ.)
jangy@sejong.edu

(Corresponding author임)

논문접수 : 2019년 10월 14일

(Received 14 October 2019)

논문수정 : 2019년 11월 18일

(Revised 18 November 2019)

심사완료 : 2019년 11월 19일

(Accepted 19 November 2019)

Copyright©2020 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.
정보과학회논문지 제47권 제2호(2020. 2)

1. 서론

교통 혼잡은 도로 사용량이 증가함에 따라 발생하는 운송 네트워크의 현상으로 느린 속도, 긴 소요시간 및 증가된 차량 대기열과 같은 특징이 나타난다[1]. 미국 연방 고속도로 관리국에 의하면 교통 혼잡 원인은 병목 현상(40%), 교통사고(25%), 기상 악화(15%), 도로 공사(10%), 잘못된 신호 체계(5%), 축제, 스포츠 경기와 같은 행사(5%) 등이 있다[2]. 교통 정체로 인하여 차량이 도로를 통과하는데 소요되는 시간이 길어지면서 대기 오염과 연료 손실과 같은 부작용이 발생한다. CEBR의 보고서에 따르면 미국, 영국, 프랑스, 독일의 교통 혼잡 비용은 2020년 2,395억 달러에서 2030년에는 2,931억 달러로 증가할 것으로 예상된다[3].

교통 혼잡을 해결할 수 있는 방법은 도로 인프라 확장 또는 효율성 향상이 있다. 도로 인프라를 확장하는 것은 물리적인 한계가 있다. 도로의 효율성을 높이기 위한 방법은 교통량을 예측하고 특정 도로에 집중되어 있는 차량을 다른 도로로 분배할 대체 경로를 제시하는 것이다. 대부분의 연구자들은 교통 혼잡 지역을 식별하고 혼잡 정도를 예측하고자 했다. 하지만 근본적인 문제를 해결하기 위해서는 교통 혼잡 현황을 이해할 수 있어야 한다. 교통 혼잡 현황은 혼잡 원인과, 혼잡으로 인한 차량의 영향을 포함한다. 도시의 교통 정체를 이해하려면 먼저 교통 혼잡 원인을 식별해야 한다. 그리고 교통 혼잡으로 인한 통행 지연 시간, 혼잡을 겪고 있는 차량의 수를 식별해야 한다. 그리고 같은 유형의 교통 혼잡이 어느 지역에 언제 발생하는지 분석해야 한다. 이러한 이유로 최근 시각적 분석 커뮤니티에서는 전반적인 교통 흐름을 이해하기 위한 많은 연구가 수행되었다. 그들은 교통 상황을 모니터링하고, 시공간 클러스터링을 사용하여 혼잡 지역을 식별했다. 그리고 그들은 혼잡이 언제 발생할 것인지 예측하는 모델을 개발했다. 하지만 교통 혼잡 원인과 영향을 조사한 연구는 찾기 어렵다[4].

본 논문에서는 교통 혼잡 원인과 영향을 분석할 수 있는 새로운 기법을 제안한다. 교통 혼잡은 일반적인 차량의 흐름을 벗어나는 현상이며 교통 혼잡 탐지는 이상치 탐지와 유사하다. 따라서 우리는 정보 이론의 엔트로피를 사용하여 교통 혼잡 지역을 식별한다. 정보 엔트로피와 평균속도를 사용하면 차량의 평균속도만 사용하는 것 보다 더 많은 차량 흐름을 분류할 수 있다. 왜냐하면 데이터를 분류하는데 사용되는 변수가 평균속도 하나에서 평균속도와 정보엔트로피를 사용한 두 개로 늘어나기 때문이다. 그림 1 (a1)-(a3)에서 도로를 통과하는 차량의 평균속도는 약 20km/h로 동일하다. 평균 속도만 이용하는 경우 (a1)-(a3)과 같은 차량의 흐름을 분류할 수 없다. 반면 정보엔트로피와 평균 속도를 같이 이용하

는 경우 차량의 흐름은 (a1)과 (a2)-(a3)과 같이 두 가지의 차량 흐름으로 분류할 수 있다. 혼잡 지역을 식별한 다음 차량의 흐름을 정량화 할 수 있는 누적 차량 수 커브(N-curve)를 추출한다. 교통 흐름 이론에 따르면 교통 혼잡 원인에 따라 고유한 N-curve를 관찰할 수 있다. 우리는 N-curve를 4가지의 교통 혼잡 유형으로 분류할 수 있는 합성곱 신경망(Convolutional Neural Network)을 설계했다. 교통 혼잡 원인을 식별한 다음 차량 지연 시간과 정체에 영향을 받는 차량 수와 같은 혼잡 영향을 분석한다. 또한 우리는 교통 혼잡 원인 분석 프로세스를 지원하는 시각적 분석 시스템을 개발했다. 이 논문의 주요 공헌은 다음과 같다.

- 교통 흐름 이론을 기반으로 교통 혼잡의 원인과 영향을 이해할 수 있는 시각적 분석 시스템을 제시한다.
- 정보 엔트로피를 사용하여 교통 혼잡 지역을 탐지할 수 있는 방법을 제시한다.
- N-curve에서 교통 혼잡 유형을 분류할 수 있는 Convolutional Neural Network을 제시한다.

2. 관련 연구

시각적 분석 커뮤니티에서는 지난 수년간 교통 데이터에 대한 많은 연구가 발표되었다[4]. 관련 연구에서 사용된 교통 데이터는 GPS trajectory[5], CCTV[6], Origin-Destination[7], Vehicle Detector[8]가 있다. 교통 분석을 위한 시각적 분석 시스템은 교통 상황 모니터링, 교통 패턴 발견, 상황 인식 및 예측, 도시 계획 및 경로 추천으로 분류할 수 있다.

교통 상황 모니터링 연구는 교통사고, 교통량, 속도 및 도로 점유율과 같은 교통 상황을 실시간으로 분석한다[9]. 교통 상황을 모니터링하기 위해 실시간 스트림 데이터와 CCTV 이미지를 사용한다[10]. 교통 패턴 발견 연구에서는 시간 경과에 따른 교통 패턴의 변화를 분석하고 도시 교통 생태계를 이해하고자 한다[11]. 교통 패턴의 변화를 분석하기 위해 차량의 흐름과 유사한 패턴을 클러스터링 할 수 있는 방법을 제시한 연구가 있다[12]. 교통 생태계를 이해하기 위해 스마트 교통 카드 데이터에 반영된 이동 패턴을 이용하여 집과 직장을 식별하는 연구가 있다[13]. 그리고 교차로에서 대규모 교통 흐름을 분석할 수 있는 시각화 기술을 제안한 연구가 있다[14].

GPS trajectory 데이터는 도시에서 사람들의 움직임 패턴을 직접 나타내기 때문에 적절한 분석 모델과 함께 다양한 분야에 적용될 수 있다. 예를 들어, 이동 패턴 검색[15], 경로 추천[16], 도시 계획[17], 교통 정체 감지[18]와 같은 다양한 연구가 있다. Veloso et al.[19]는 택시 데이터를 사용하여 사람들의 생활 패턴을 분석한다.

Yuan et al.는 택시 운전자가 승객을 빨리 찾을 수 있는 경로를 제안한다[20]. Ferreira et al.는 시각적 질의 모델을 사용하여 도시에서 이동 패턴을 탐색할 수 있는 기술을 제안한다[11]. Liu et al.은 교통 데이터를 이용하여 적합한 광고 위치를 계산하는 방법을 제안한다[21]. 상황 인식 및 예측 연구에서는 교통 상황이 미래에 어떠한 영향을 미치는지 연구한다[22]. 대표적으로 교통 상황을 이해할 수 있는 질의응답 인터페이스를 개발한 연구가 있다[23].

도시에서 교통 혼잡의 원인을 분석하기 위해서는 교통 상황을 이해해야 한다. 현재 많은 교통 혼잡 추정 알고리즘은 차량 수, 도로의 평균 속도 및 시공간 차량 밀도를 사용한다. Pattara-Atikom et al.은 GPS 속도의 가중치 수 이동 평균을 사용하여 교통 정체를 추정한다[24]. Kong et al.은 교통 혼잡을 추정하기 위해 여러 교통량 데이터를 이용하여 도로 구간의 평균 속도를 계산했다[25]. 위의 접근법은 교통 혼잡 정도를 계산하기 위해 평균 속도, 주행 시간 또는 차량 밀도와 같은 특정 매개변수 하나만을 사용한다. 그러나 다양한 상황이 결합되어 교통 혼잡이 발생하기 때문에 위에서 제안한 연구는 혼잡 지역을 추정하는데 많은 불확실성이 있다.

이 문제를 해결하기 위해 퍼지 논리[26], 유전자 알고리즘[27], 신경망[28]을 이용하여 교통 혼잡 지역을 분석하는 연구가 있다. Shankar et al.은 도로에서 차량의 밀도와 차량 속도를 이용하여 혼잡을 식별할 수 있는 퍼지 추론 시스템을 제시했다[29]. Zhou et al.는 SVM과 회귀 모형을 이용하여 교통 혼잡을 예측하는 모델을 제안한다[30]. 더불어 교통 신호체계를 최적화하여 교통 혼잡을 줄이기 위한 연구도 발표되었다.

3. 교통 흐름 데이터 및 전처리 과정

본 절에서는 연구에 사용된 데이터를 설명하고 불완전한 GPS 데이터를 필터링하고 교차로 위치를 추출하고 차량의 방향과 속도를 계산하는 방법을 소개한다.

3.1 교통 흐름 데이터

본 논문에서는 두 가지 형태의 교통 흐름 데이터와 Open Street Map[33] 데이터를 사용하여 교통 혼잡 원인을 분석한다. 교통 흐름 데이터는 GPS trajectory와 Vehicle Detector (VD)[34] 데이터를 사용한다. GPS trajectory 데이터는 베이징, 샌프란시스코에서 운행하는 택시 데이터를 사용한다. GPS trajectory 데이터는 표 1과 같이 차량 ID, 날짜, 시간, 위도 및 경도로 구성된다. 시애틀의 Vehicle Detector(VD) 데이터는 도로의 시작과 끝에 설치되어 24시간 동안 차량의 통행량을 감지한다. VD 데이터는 VD의 GPS 위치, 차량 ID, VD로 차량이 감지된 시간, 주행 시간 및 차량의 방향으로

표 1 GPS trajectory 데이터

Table 1 GPS trajectory data

City	Time	# of vehicles	# of samples
Beijing [31]	2009.05.01. - 2009.05.09.	8,602	128,914,190
San Francisco [32]	2008.05.17. - 2008.07.10.	536	11,219,419

표 2 Open Street Map 데이터

Table 2 Open Street Map data

City	GPS rance	# of nodes	# of ways
Beijing	[39.7785N, 40.1038N] [116.1389E, 116.6216E]	423,013	76,614
San Francisco	[37.7099N, 37.811N] [122.5151W, 122.3554W]	1,768,088	189,177

구성된다. 그리고 표 2와 같이 2개 도시의 GPS 범위 내에 있는 Open Street Map 데이터에서 교차로의 위치를 추출한다.

3.2 GPS 데이터 전처리

교통 혼잡 원인 분석을 위한 전처리 작업으로 비현실적인 좌표, 긴 샘플링 주기와 중복된 GPS를 필터링 한다. GPS 필터링은 Wang et al.이 제시한 방법을 사용한다[18]. 비현실적인 좌표는 표 2에서 제시한 범위를 벗어나는 GPS 데이터를 의미한다. 두 개 이상의 중복된 데이터는 첫 번째 GPS 기록만 사용한다. GPS 샘플링 간격이 5분 이상인 데이터는 사용하지 않는다. 차량이 20분 이상 50m 반경 내에 있으면 차량이 주차했다고 가정하여 차량의 GPS 데이터를 삭제한다. 기록 빈도가 낮은 데이터는 차량의 속도를 정확하게 예측하기 어렵기 때문에 논문에서는 km당 최소 6개 이상의 데이터가 기록된 차량의 GPS 데이터를 사용한다.

3.3 교통 혼잡 지역 정의

교통 혼잡 원인 분석을 위한 시각적 분석에서 그리드(grid), 브러시(brush) 영역 및 교차점을 포함하는 세 가지 영역을 정의한다. 그리드는 도시 수준에서 교통 흐름 변화가 발생한 지역을 식별하는데 사용된다. 하나의 그리드는 1km × 1km로 구성된다. 브러시 영역은 사용자가 지도에서 선택할 수 있으며 특정 도로에서 교통 정체 원인을 분석할 때 사용된다. 교차점은 엔트로피, 속도, 및 부피와 같은 차량의 흐름을 정량화 할 때 사용한다.

3.4 차량의 방위각과 거리 계산 방법

차량의 방향은 진북 방위각을 이용하여 계산한다. 진북은 지구의 북극점 또는 북극점을 향한 지구 표면의 방향을 의미한다. 진북 방위각은 0도인 북극점을 기준으로 시계 방향으로 측정한 수평각을 의미한다. 시간 t에

서 $t+1$ 까지 이동한 차량의 방위각(θ)은 다음과 같은 수식을 이용하여 계산한다.

$$\theta = \arctan2(\sin \Delta\lambda \cdot \cos \phi_2, \cos \phi_1 \cdot \sin \phi_2 - \sin \phi_1 \cdot \cos \phi_2 \cdot \cos \Delta\lambda)$$

ϕ_1 와 ϕ_2 는 시간 t 와 $t+1$ 에서 라디안 단위의 위도(latitude)를 의미한다. $\Delta\lambda$ 는 시간 t 와 $t+1$ 사이의 라디안 단위의 경도(longitude) 차이를 의미한다. 진북 방위각을 이용하여 차량의 방향을 분류한다. 진북 방위각이 $\theta < \pi/4$ 또는 $\theta \geq 7\pi/4$ 이면 차량의 방향은 북쪽으로 분류한다. 이와 같이 방위각이 $\pi/4 \leq \theta < 3\pi/4$ 이면 차량의 방향은 동쪽으로, 방위각이 $3\pi/4 \leq \theta < 5\pi/4$ 이면 남쪽으로, 방위각이 $5\pi/4 \leq \theta < 7\pi/4$ 이면 차량의 방향은 서쪽으로 분류한다. 시간 t 에서 $t+1$ 까지 차량의 속도는 거리를 시간으로 나누어서 계산한다. 두 GPS 지점 간의 거리는 다음과 같은 Haversine 공식을 이용한다.

$$distance = 2r \cdot \arcsin(\sin^2(\frac{\phi_2 - \phi_1}{2}) + \cos(\phi_1) \cos(\phi_2) \sin^2(\frac{\lambda_2 - \lambda_1}{2}))$$

ϕ_1 와 ϕ_2 는 시간 t 와 $t+1$ 의 GPS 위도를 나타낸다. λ_1 와 λ_2 는 시간 t 와 $t+1$ 의 GPS 경도를 나타낸다. r 은 지구의 반지름을 의미한다.

4. 교통 혼잡 분석

본 절에서는 교통 흐름 이론을 기반으로 교통 혼잡의 원인과 영향을 분석하는 방법을 소개한다.

4.1 차량 흐름 감지

일반적으로 차량의 평균 속도만 이용하여 교통 혼잡 정도를 계산한다. 하지만 특정 시간 동안 차량의 평균 속도만 이용하는 경우 차량의 속도 변화를 감지하는 것은 어렵다. 그림 1의 (a1)~(a2)에서 차량은 평균속도 20km/h로 교통 혼잡 구간을 통과한다. (a1)에서 차량은 모든 구간을 평균속도 20km/h로 통과한다. (a2)에서 차량은 평균속도 10km/h로 감속한 다음 25km/h로 이동한다. 두 흐름에서 차량은 같은 평균속도로 혼잡 구간을

통과하지만 차량의 흐름 패턴은 다르다. 더불어 평균 속도만을 이용하여 교통 혼잡을 추정할 때 (a3)과 같이 신호등으로 인한 교통 흐름 변화는 감지하기 어렵다. 논문에서는 (a4)와 같이 차량이 평균속도 40km/h 이상으로 통과하면 교통 혼잡은 발생하지 않는다고 가정한다. 위에서 언급한 문제를 해결하기 위해 본 논문에서는 정보이론에서 엔트로피를 이용하여 교통 혼잡지역에서 차량의 흐름 변화를 분류한다.

엔트로피는 불확실성이나 무질서의 정도를 나타내는 확률모델이다[35]. 엔트로피의 핵심은 드물게 발생하는 이벤트의 정보는 자주 발생하는 이벤트보다 더 유익한 정보라고 가정한다. 이벤트의 발생 확률이 높으면 다음 조건을 예측하기 쉽기 때문에 새로운 정보는 거의 없다고 볼 수 있다. 반대로 사건이 일어날 가능성이 없다면 새로운 정보는 매우 많다고 볼 수 있다. Shannon은 이산 확률 변수 X 의 엔트로피 H 를 다음과 같이 정의한다.

$$H(X) = - \sum_{i=1}^n P(x_i) \log_b P(x_i)$$

$P(x_i)$ 는 이벤트가 발생할 확률을 의미한다. 논문에서는 3.3절에서 설명한 바와 같이 도시 지역을 여러 개의 그리드로 나눈다. 각 그리드에 포함된 차량의 trajectory 데이터를 로드하고 차량의 속도 값에 대한 히스토그램을 추출한다. 히스토그램의 간격은 5km/h이며 각 히스토그램 막대의 속도 값에 해당하는 차량의 수는 x_i 로 설정한다. 따라서 $P(x_i)$ 는 특정 속도가 그리드에 나타날 확률이며 각 막대의 차량 수를 그리드에 포함된 차량의 수로 나누어 확률을 계산할 수 있다. 위에서 소개한 엔트로피 수식에 $P(x_i)$ 를 대입하여 그리드의 속도 엔트로피 $H(X)$ 를 계산한다. 그리드 또는 혼잡 영역에서 교통 흐름 변화가 빈번하면 높은 엔트로피 값이 도출된다. 예를 들어 그림 1 (a2)는 (a1)보다 속도 변화가 크기 때문에 높은 엔트로피를 갖는다.

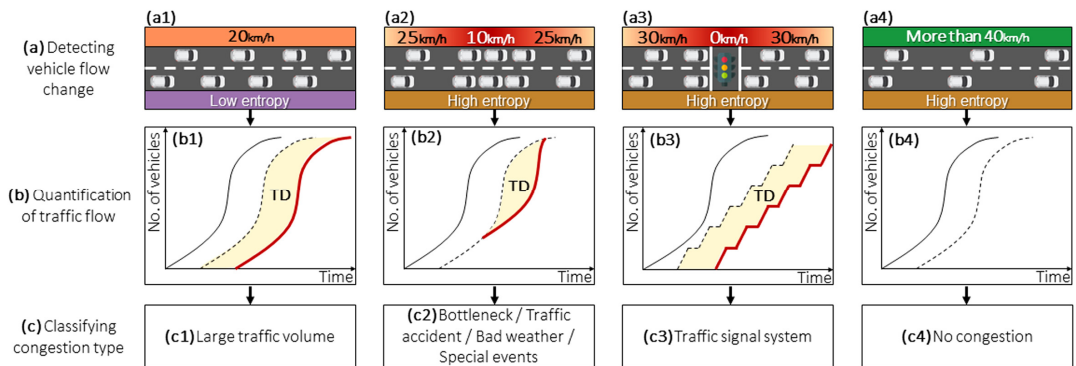


그림 1 교통 혼잡 원인 분류 절차

Fig. 1 Classification for the cause of traffic congestion

4.2 차량 흐름 정량화

N-curve는 교통 흐름 이론에서 출발 차량과 도착 차량의 누적 횟수를 나타낸다[36]. N-curve는 시간 지연 및 지연된 차량 수를 포함하여 교통 혼잡 시간, 위치, 원인 및 혼잡으로 인한 영향에 관한 정보를 제공한다. 그림 2에서 차량이 X_1 에서 X_2 로 이동한다고 가정했을 때, N_1 과 N_2 는 차량이 특정 위치 X_1 과 X_2 를 통과하는 차량의 N-curve이다. TT_{ff} ($Travel Time_{free flow}$)는 교통 혼잡이 없는 상황에서 차량이 X_1 부터 X_2 까지 이동하는데 소요되는 시간을 의미한다. 교통 혼잡이 없는 상황에서 X_1 을 지나는 차량의 N-curve는 N_1 과 같다. 시간 t_1 과 t_2 사이에 교통 혼잡이 발생하면, X_1 을 지나는 차량의 N-curve는 N_2 와 같이 변경된다. TT_{delay} 는 차량이 혼잡으로 인해 발생한 지연 시간을 의미한다. 차량 v_n 이 교통 혼잡 구간을 통과하는데 소요되는 시간은 $TT_{ff} + TT_{delay}$ 로 계산할 수 있다. 교통 혼잡을 겪은 차량의 수는 $v_2 - v_1$ 이다. 그리고 차량의 총 지연 시간은 TD (Total Delay)이다. N-curve는 교통 체증 원인에 따라 고유한 패턴을 가지고 있다. 차량이 그림 1의 (a1)과 같이 일정한 속도로 천천히 움직일 때, TT_{delay} 는 (b1)과 같이 시간이 지남에 따라 일정하게 유지된다. 그림 1의 (a2)와 같이 차량이 감속한 다음 다시 가속하면 (b2)와 같은 패턴을 얻을 수 있다. 도로 중간에 신호등이 있는 경우 (b3)과 같은 독특한 계단 형태의 TD 패턴이 관찰된다.

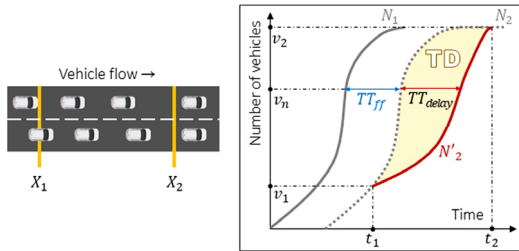


그림 2 누적 차량 수 커브(N-curve)

Fig. 2 Cumulative vehicle count curve (N-curve)

4.3 교통 혼잡 패턴 분류

N-curve는 혼잡 지역의 교통 흐름을 이해하는데 유용하다. 특히 전체 지연 시간 (TD)은 정체 원인을 파악하는데 중요하다. 그림 1 (c1-c4)에 제시된 4가지 혼잡 유형은 (b1-b4)에서 볼 수 있는 고유한 TD 패턴을 갖는다[37]. 그림 3은 네 가지의 혼잡 유형에 따라 관찰되는 TD 패턴의 예를 나타낸다. 교통 혼잡 원인 분류를 위한 CNN 모델은 표 3과 같이 2개의 convolution 레이어,

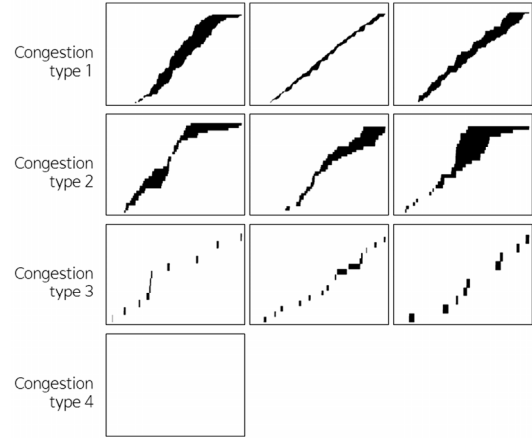


그림 3 교통 혼잡 유형에 따른 TD 패턴의 예

Fig. 3 Examples of TD Patterns based on Traffic Congestion Type

표 3 교통 혼잡 원인을 분류하기 위한 CNN 모델 설계
Table 3 CNN design to classify the causes of traffic congestion

Layer	Image size	Channels
Conv 1	I: 112×112 O: 110×110	I: 1 O: 32
Pooling 1	I: 110×110 O: 55×55	I: 32 O: 32
Conv 2	I: 55×55 O: 53×53	I: 32 O: 32
Pooling 2	I: 53×53 O: 26×26	I: 32 O: 32
Flatten	I: 26×26 O: 21,632 neurons	I: 32 O: -
FC1	I: 21,632 neurons O: 128 neurons	I: - O: -
FC2	I: 128 neurons O: 4 congestion types	I: - O: -

2개의 pooling 레이어 및 2개의 fully connected 레이어로 구성된다. Convolution layer는 입력 데이터에서 합성곱 연산을 이용하여 이미지의 공간 정보를 유지하면서 인접 이미지와의 특징을 표현할 수 있는 feature map을 추출하는데 사용된다. Convolution layer에서 합성곱 연산에 사용되는 커널의 크기는 3×3 이다.

Pooling layer는 convolution 레이어의 출력 크기를 줄이는데 활용된다. Pooling layer에서 2×2 크기의 window size와 stride를 사용하며, Max pooling 방식으로 feature map을 요약한다. 훈련 데이터의 과적합을 방지하기 위해 fully connected 레이어 사이에 drop out 레이어를 배치한다. Rectified Linear Unit (ReLU) 활성화함수는 두 개의 convolution layer와 첫 번째 fully connected layer에서 사용된다. 그리고 두 번째 fully connected

표 4 CNN 모델의 Confusion 행렬
Table 4 Confusion matrix of CNN model

CNN		True condition			
		Congestion type 1	Congestion type 2	Congestion type 3	Congestion type 4
Predicted condition	Congestion type 1	2695	16	6	27
	Congestion type 2	74	2783	25	27
	Congestion type 3	37	10	2778	54
	Congestion type 4	9	5	6	2708

layer에서는 softmax 활성화수를 사용한다. 마지막으로 손실함수는 cross entropy loss를 사용한다. 논문에서 제안한 CNN의 분류 성능은 표 4와 같다.

4.4 교차로에서 차량 흐름을 식별하기 위한 시각화

특정 지역의 교통 정체는 인접한 지역과 관련되어 있다. 교통 혼잡 원인을 이해하기 위해서는 시간과 공간의 제약을 받는 일시적인 상황 또는 흐름이 아닌 전체적인 상황 또는 흐름을 분석해야 한다. 교차로는 두 개 이상의 도로가 연결된 곳으로서 도로에서 발생한 교통 혼잡을 이해하기 위해서는 교차로를 통과하는 차량의 흐름을 직관적으로 이해할 수 있어야 한다. 우리는 교차로를 통과하는 차량의 흐름을 시각화하기 위해 교차로를 통과하는 교통량과 평균속도, 교통 흐름 패턴의 변화, 그리고 교차로에 연결된 도로의 유입과 유출 비율을 시각화하는데 중점을 두었다.

그림 4는 교차로에서 교통 흐름을 시각화 하는 새로운

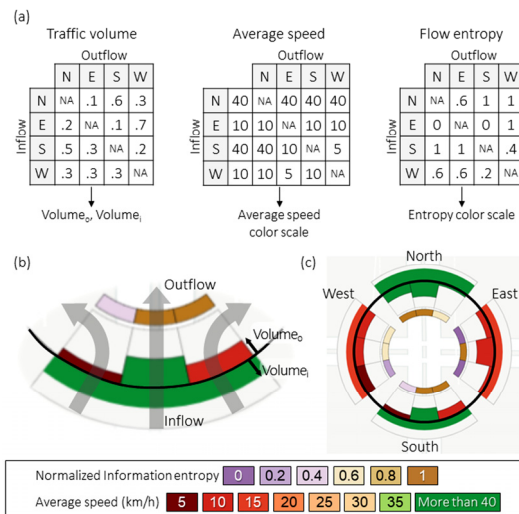


그림 4 교차로에서 차량의 흐름을 시각화 하는 방법
Fig. 4 Vehicle flow visualization at the intersection

클리프 디자인을 나타낸다. 하나의 클리프는 교차로의 각 도로에서 차량의 흐름을 나타내는 일련의 하위 클리프로 구성된다. 교차점에서 차량의 교통량, 평균 속도 및 엔트로피는 그림 4(a)와 같은 행렬로 표현할 수 있다. 그림 4(a) 행렬은 (c)와 같이 인코딩 된다.

행렬의 대각 원소는 NA이므로, U 턴 궤적이 없다고 가정한다. 하나의 하위 클리프는 그림 4(b)와 같이 차량의 유입과 유출을 표현한다. $Volume_o$ 와 $Volume_i$ 는 각각 유출과 유입의 교통량을 나타낸다. 차량 속도는 녹색에서 붉은색으로 인코딩한다. 엔트로피는 보라색에서 붉은색 컬러 스케일로 인코딩한다. 하위 클리프에서 원호의 크기는 도로 수로 조정되며 각 도로의 차량 방향은 하위 클리프의 방향을 결정한다.

5. 시각적 분석 시스템

본 연구에서 제시한 시각적 분석 시스템은 그림 5와 같다. 시각적 분석 시스템은 교통 혼잡 감지, 혼잡 원인 분류, 혼잡 영향 분석 및 혼잡 전과 분석과 같은 네 가지의 분석 절차를 지원한다.

교통 데이터 분석의 첫 번째 단계는 초기 매개 변수를 설정하는 것이다. 사용자는 그림 5(a)와 같이 대시보드에서 도시, 날짜, 시간 및 기타 여러 조건을 선택할 수 있다. 시간 집계는 평균 속도와 엔트로피를 계산하는 시간 단위를 의미한다. 시스템에서 60분, 30분, 10분의 시간 집계가 지원된다. 차량 방향 필터를 사용하면 사용자는 특정 방향으로만 움직이는 차량의 교통 혼잡 패턴을 분석할 수 있다.

두 번째 단계는 도시의 차량 흐름 변화를 분석하는 것이다. 사용자는 그림 1 (a1-a4)에서 소개한 네 가지 차량 흐름을 식별하고 교통 혼잡 분석이 필요한 후보 지역을 선택한다. 이 절차에서는 그림 5(b)의 평행 좌표 플롯(Parallel coordinates plot), (c)의 산포도(Scatter plot) 및 (d)의 Grid map을 사용한다. Parallel coordinates plot은 차량 흐름 변화를 시간에 따라 분석할 수 있다. x 축은 시간을 나타내고, y 축은 차량 흐름 엔트로피를 나타낸다. Parallel coordinates plot에서 하나의 라인은 그림 5(d)에 나타난 각 그리드에서 시간 경과에 따른 엔트로피 값을 나타낸다. Grid에서 차량의 엔트로피와 평균 속도를 계산하기 위해 다음과 같은 방법을 사용한다. GPS 필터링 규칙을 사용하여 오프로드 GPS 문제를 해결한다. 차량이 두 개 이상의 그리드를 움직이면 각 그리드는 GPS 위치가 그리드 영역에만 있는 차량만 사용한다. 제한한 시스템은 Parallel coordinates plot에서 시계열 클러스터링을 지원한다. 시계열 클러스터링은 일정 시간 동안 유사한 트래픽 흐름을 갖는 그리드를 탐색할 수 있도록 지원한다. 대시보드에서 클리

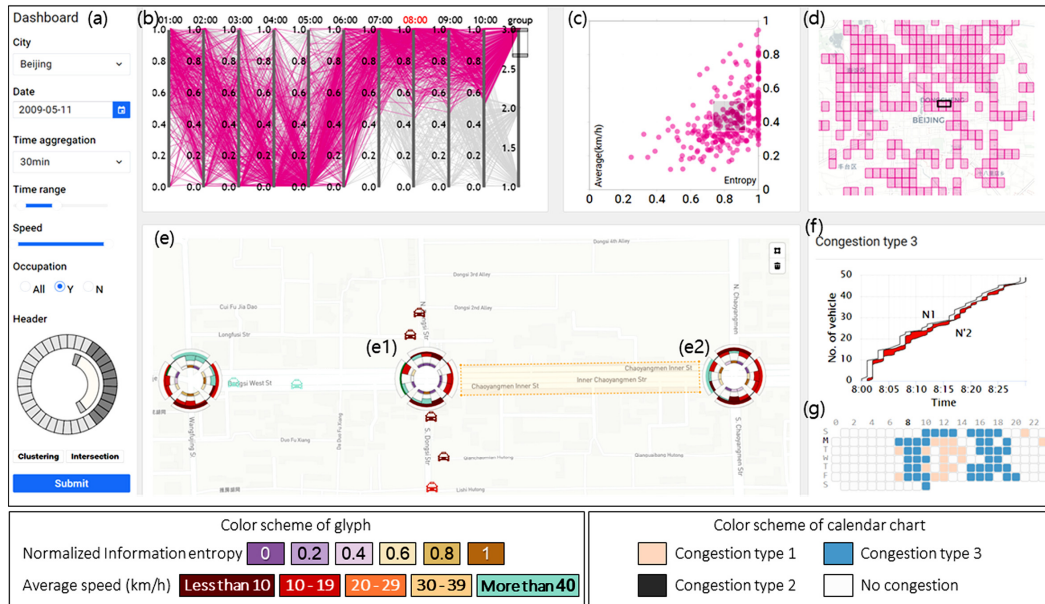


그림 5 교통 혼잡 원인 분석을 위한 시각적 분석 시스템

Fig. 5 Visual analytics system for the analysis of cause of traffic congestion

스터링 버튼을 클릭하면 클러스터 번호, 클러스터 색상 및 시간 범위를 선택하는 팝업 창이 나타난다. Scatter plot은 차량 속도와 엔트로피 간의 관계를 제공한다. Parallel coordinates plot은 엔트로피만으로 차량 흐름 변화를 분류할 수 있지만 그림 1 (a1-a5)에 나와 있는 모든 차량 흐름 패턴을 분류할 수 없다. 따라서 그림 1 (a1-a5)에 있는 모든 패턴을 분류하기 위해서 parallel coordinates plot과 scatter plot을 같이 사용한다. Scatter plot의 x 축과 y 축은 정규화된 엔트로피와 평균 속도를 나타낸다. Scatter plot에서 하나의 점은 그리드를 나타낸다. Grid map은 Parallel coordinates plot과 Scatter plot에서 브러시된 그리드를 시각화한다.

세 번째 분석 절차에서는 혼잡이 발생할 것으로 예상되는 후보 지역에서 실제 교통 혼잡이 발생한 도로를 식별한다. 이 절차에서는 도로의 차량 흐름을 분석하기 위해 시스템은 사용자에게 그림 5(e)의 Map view와 글리프를 제공한다. 교차로는 두 개 이상의 도로가 만나는 지점이며, 차량은 일반적으로 교차로에서 이동 방향을 변경한다. 따라서 각 교차점에 글리프를 표시하여 도로에서 다양한 차량 흐름을 시각적으로 요약한다. 글리프는 차량의 방향, 속도, 볼륨 및 엔트로피 정보를 포함한다. 글리프에서 차량의 방향과 양에 따라 도로 상태를 예측할 수 있다. 차량 속도 및 엔트로피는 그림 1 (a1-a5)에 표시된 차량 흐름 패턴을 추론하는 데 사용된다. 세 번째 분석 절차에서 그림 5(e)의 주황색 박스 영역과 같

이 추가 정체 원인 분석에 필요한 도로를 선택할 수 있다. 주황색 영역은 사용자가 지도에서 도로 경계를 지정하여 설정할 수 있다. 교통 혼잡 원인 분석을 위해 지도에서 경계가 설정되면 시스템은 그림 5(a)에 제시된 헤더 범위 내에서만 주황색 영역의 차량 정보를 추출한다.

네 번째 분석 절차에서는 교통 혼잡 원인 and 영향을 분석다. 지도에서 정체 도로를 선택하면 시스템은 그림 5(f)의 N-curve와 (g)의 캘린더 차트를 통해 혼잡 원인 및 영향 정보를 제공한다. N-curve는 차량 흐름을 정량화한다. N-curve에서 x 축은 시간을 나타내고 y 축은 누적 차량 수를 나타낸다. N-curve의 붉은색 영역은 교통 정체(TD)를 의미한다. 시스템은 TD를 이용하여 교통 혼잡으로 인한 이동 시간 지연 및 정체를 경험하는 차량의 수와 같은 정보를 추출한다. 캘린더 차트는 사용자가 선택한 도로에 대한 과거의 정체 정보를 제공한다. 캘린더 차트를 통해 사용자는 교통 정체가 일시적인지 반복적인지를 판단 할 수 있다. 시스템은 지난 7일 동안 한 시간 간격으로 선택된 도로에서 N-curve를 추출하고 Convolutional Neural Network를 이용하여 그림 1에서 소개한 네 가지 교통 혼잡 유형으로 분류한다. 각 교통 정체 원인은 고유한 색상으로 맵핑되어 그림 5(g)와 같이 표현된다.

마지막 분석 과정에서는 이전에 분석한 교통 정체 도로와 관련된 새로운 혼잡 도로를 탐색한다. 한 지역의 교통 혼잡은 인접 지역과 관련되어 있기 때문에 정체가

발생한 도로와 주변 도로를 함께 분석해야한다. 시스템은 그림 5(e)에 표시된 맵과 (g)의 캘린더 차트에서 혼잡 전파와 관련된 정보를 제공한다. 지도상의 차량 아이콘은 혼잡 지역에 진입하는 차량의 위치 및 속도 정보를 포함한다. 속도는 그림 5와 같이 색상 스케일을 사용하여 차량 아이콘에 인코딩된다. 차량 아이콘을 통해 특정 지역의 혼잡으로 인한 주변에 있는 다른 정체 도로를 식별 할 수 있다. 캘린더 차트에서 셀을 클릭하면 현재 분석중인 도로의 다른 시간 창에서 정체 상태를 분석 할 수 있다. 이 경우 첫 번째 분석 절차를 건너뛰고 두 번째 분석 절차에서 교통 정체 상황을 분석할 수 있다. 최종 분석 과정에서 추가로 분석해야하는 정체 도로를 선택할 수 있다. 또한 특정 도로의 교통 정체가 주변 도로에 미치는 영향을 설명 할 수 있다.

교통 혼잡 원인 분석 시스템은 APM (Apache, PHP, Mysql) 프레임 워크에서 개발된 웹 기반 응용 프로그램이다. 시각적 분석 모듈은 D3.js와 highcharts로 구현되었다. 백엔드에서는 R과 Python을 이용하여 차량의 방향과 속도, 엔트로피와 같은 교통 정체 계산 모듈과 합성곱 신경망 분류기를 구현했다.

6. 교통 혼잡 분석

이 절에서는 세 가지의 사례를 통해 교통 혼잡 원인 분석을 위한 시각적 분석 시스템의 효율성을 소개한다.

6.1 베이징에서의 교통 혼잡 분석

Beijing 사례에서는 GPS trajectory 데이터를 사용한다. 사용자는 그림 5(a)와 같이 2009년 5월 11일 월요일 오전 1시부터 오전 10시까지 베이징에서 교통 혼잡 패턴을 분석한다. 사용자는 차량 방향 필터를 사용하여 동쪽으로 이동하는 차량을 조사한다. 그림 5(a)의 대시 보드에서 데이터 분석 조건을 선택하면 시스템은 (b)의 parallel coordinates plot (PCP), (c)의 scatter plot 및 (d)의 grid map에서 교통 혼잡 정보를 제공한다. PCP에서 사용자는 도시의 하위 그리드를 오전 7시부터 오전 10시까지 비슷한 엔트로피 변화를 갖는 세 그룹으로 클러스터링 한다. 그림 5(b)에서 분홍색은 오전 7시부터 오전 10시까지 엔트로피가 높은 그룹을 나타냅니다. 이 그룹의 차량 흐름은 그림 1 (a2), (a3)의 차량 흐름 패턴 중 하나로 분류된다. 그림 5(c)의 scatter plot과 (d)의 grid map은 (b)의 PCP에서 선택된 오전 8시의 교통 상태를 표시한다. Scatter plot에서 사용자는 엔트로피가 높고 속도가 느린 영역을 선택한다. 선택된 영역은 그림 1 (a2) 또는 (a3)에 제시된 혼잡 패턴 중 하나일 가능성이 있다. Grid map은 사용자가 scatter plot에서 선택한 혼잡 패턴이 나타난 위치가 표시된다.

시스템에서 제공하는 대부분의 시각화 모듈은 대략적

인 분석만 허용하지만 N-curve는 교통 혼잡의 시간, 원인 및 영향에 대한 자세한 정보를 제공한다. 그림 5(e)에서, 주황색 영역은 교차로 (e1)에서 (e2)까지의 경로를 나타낸다. 주황색 영역이 나타내는 경로에서 오전 8시부터 오전 8시 25분까지 교통 혼잡이 감지된다. 교통 혼잡이 발생하지 않으면 차량이 경로를 통과하는 데 평균 1분이 소요된다. 그러나 교통 혼잡으로 인해 차량은 평소보다 1~3분 지연되었다. 또한 시스템은 N-curve 패턴을 그림 1 (b3)과 같이 신호 시스템으로 인하여 교통 혼잡이 발생했음을 나타내는 Congestion type 3으로 분류한다.

캘린더 차트는 사용자가 분석하고 있는 도로의 과거 혼잡 원인에 대한 정보를 제공한다. 사용자는 그림 5(d)의 캘린더 차트를 활용하여 도로에서 발생한 혼잡 원인이 일시적인지 또는 주기적으로 발생하는 것인지 판단할 수 있다. 캘린더 차트에서 평일 오전 8시부터 오전 10시 사이에 Congestion type 3이 발생함을 보여 주며, 이는 교통 신호가 오전에 정기적으로 교통 체증을 유발함을 나타낸다. 사용자는 그림 5(e)의 지도에서 차량 아이콘을 탐색하여 혼잡 전파 패턴을 분석 할 수 있다. 지도상의 차량 아이콘은 교차로 (e1)의 북쪽과 남쪽에서 진입하는 차량의 평균속도가 낮은 것을 볼 수 있다.

6.2 병목현상으로 인한 교통 혼잡

San Francisco 사례에서는 GPS trajectory 데이터를 사용한다. 사용자는 San Francisco에서 도로 설계로 인한 병목 현상이 두 도로에 어떻게 영향을 미치는지 분석한다. 그림 6에서 Route 1은 San Francisco에서 Oakland Bay Bridge까지 연결된 도로이며, Route 2는 South San Francisco에서 Oakland Bay Bridge까지의 연결된 도로이다. 그림 5(a)는 2008년 5월 17일 토요일 오후 3시 Route 1의 교통 혼잡 상황을 보여준다. 그림 6의 지도에서 cyan 영역은 교통 혼잡 분석을 위해 사용자가 브러쉬한 영역을 나타낸다. 교통 체증이 발생하지 않으면 차량이 Route 1을 통과하는데 평균 2분이 소요된다. 그러나 교통 혼잡으로 인해 오후 3시에서 3시 35분 사이에 차량의 이동 시간은 3분에서 7분이 소요된다. 오후 3시 35분에 출발하는 차량은 Route 1을 통과하는데 10분에서 15분이 소요된다. 그림 6(b)에서 오후 3시부터 Route 2에서 교통 혼잡이 발생한다. 교통 혼잡이 없다면 차량은 Route 2를 통과하는 데 평균 2분이 소요된다. 그러나 교통 정체로 인해 오후 3시에서 3시 50분 사이에 출발하는 차량의 운행 시간은 6~7분으로 증가한다. Convolutional Neural Network은 Route 1의 N-curve 패턴을 혼잡 유형 2로 분류한다. 병목 현상, 교통사고, 악천후 및 특별 이벤트가 위 유형에 속한다. 또한 Route 2는 도로를 통과하는 차량이 많다는 것을



그림 6 샌프란시스코에서 병목현상으로 인한 교통 혼잡
Fig. 6 Traffic congestion due to Bottleneck in San Francisco

의미하는 Congestion type 1로 분류된다. Route 1은 고속도로 인터체인지에서 Route 2에 합류된 다음 두 Route는 Oakland Bay Bridge로 향한다. 이 도로는 San Francisco에서 Oakland까지 이동하는 유일한 경로이기 때문에 병목 현상이 이 지역의 교통 혼잡을 크게 유발할 수 있다고 추론할 수 있다.

6.3 교통사고로 인한 교통 혼잡

Seattle 사례에서는 VD(Vehicle Detector) 데이터를 이용한다. VD는 도로의 시작과 끝 부분에 설치되어 24 시간 동안 차량의 통행량을 감지한다. 데이터는 VD의 GPS 위치, 차량 ID, VD에 차량이 감지된 시간, 주행 시간 및 차량 방향으로 구성된다.

사용자는 이른 아침에 발생한 교통사고가 Seattle의 두 도로에 어떻게 영향을 미치는지 분석한다. 그림 7에서 Route 1은 West Seattle에서 도심까지의 경로이며 Route 2는 South Seattle에서 도심까지의 또 다른 경로이다. Route 1은 고속도로 인터체인지에서 Route 2로 병합되며 두 Route 모두 Seattle 도심으로 향한다. 그림 7의 지도에서 cyan 영역은 교통 혼잡 분석을 위해 사용자가 브러시한 영역을 나타낸다. 그림 7(a)는 2011년 9월 6일 화요일 오전 6시에 Route 1의 교통 혼잡 상황을 보여준다. 교통 혼잡이 발생하지 않으면 차량이 Route 1을 통과하는데 평균 2분이 소요된다. 그러나 교통 혼잡으로 인해 오전 6시 20분에서 오전 6시 25분 사이 15대의 차량 운행 시간은 4분에서 7분으로 증가한다. 오전 6시 30분에 출발하는 차량조차도 Route 1을 통과하는데 11분에서 13분이 소요될 것으로 예상된다. 그림 7(b)를 보면 오전 6시 5분부터 Route 2에서 교통 정체가 발생한다. 교통 혼잡이 없는 경우, 차량은 Route 2를 통과하는데 평균 3분이 필요하다. 그러나 교통 혼잡

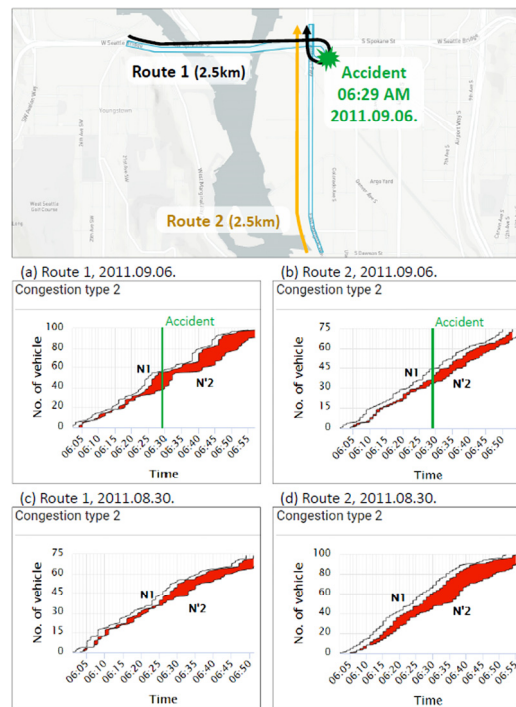


그림 7 시애틀에서 교통사고로 인한 교통 혼잡
Fig. 7 Traffic congestion caused by car accidents in Seattle

으로 인해 오전 6시 5분에서 6시 30분 사이에 출발하는 35 대의 차량은 Route 2를 통과하는데 4~5분이 소요된다. 오전 6시 30분에서 오전 7시 사이에 출발하는 차량 50대는 Route 2를 통과하는데 6분에서 8분이 소요된다. 그림 7(a)와 (b)는 Route 1과 Route 2의 여행 시간을 나타내며 시간이 지남에 따라 오전 6시부터 오전 7시까지 여행 시간이 증가하는 것을 관찰할 수 있다. 그러나 그림 7(a)에서 알 수 있듯이 Route 1에서 오전 6시 30분부터 여행 시간이 갑자기 증가한다. 그러므로 사용자는 여행 시간이 갑자기 증가한 이유를 분석하기 위해 혼잡 원인을 분석해야 한다.

먼저 Convolutional Neural Network은 Route 1의 N-curve 패턴을 Congestion type 2로 분류한다. 병목 현상, 교통사고, 악천후 및 특별 이벤트가 이 유형에 속한다. 고속도로 인터체인지에서 Route 1이 Route 2로 병합되고, Seattle에서는 오전 6시부터 러시아워가 시작되기 때문에 Cyan영역에서 나타난 교통 혼잡 원인은 병목 현상으로 추정할 수 있다. 일반적인 병목 현상의 경우 들어오는 차량의 수가 증가하면 여행 시간이 점차적으로 증가한다. 그러나 그림 7(a)에서 Route 1은 이동 시간이 갑자기 증가하여 외부 요인이 교통 혼잡에 영향을 미칠 수 있음을 나타낸다. 외부 요인을 검색하기 위해 사용자

는 날씨 및 교통 정보를 모니터링 한다. 기상정보에 따르면 2011년 9월 6일, 시애틀의 날씨는 17°C-27°C로 맑고 화창했다. 따라서 사용자는 교통 혼잡은 날씨와 관련이 없다는 것을 추론 할 수 있습니다. 또한 고속도로 인터체인지에서는 축제, 스포츠행사와 같은 특별 이벤트가 필요하지 않는다. Seattle 경찰에 따르면 Route 1에서 Route 2로 병합되는 지점에서 교통사고가 발생했다[38]. 따라서 교통사고로 인하여 Route 1의 소요 시간이 급격히 증가했다고 추론 할 수 있다. 이를 확인하기 위해 교통사고 발생 일주일 전인 2011년 8월 30일 오전 6시부터 오전 7시까지의 N-Curve를 비교한다. 그림 7(c)에서 2011년 8월 30일 Route 1에서 오전 6시에 교통 체증이 시작된다. 오전 6시부터 오전 6시 30분까지 40대의 차량이 Route 1을 통과하는데 평균 3분에서 5분이 소요된다. 오전 6시 30분부터 오전 7시까지 25대의 차량이 Route 1을 통과하는데 7분에서 10분이 소요된다. 8월 30일에 차량은 9월 6일보다 1~3 분 정도 더 빨리 Route 1을 통과할 수 있다. 또한 그림 7(c)에서 볼 수 있듯이 오전 6시 30분부터 오전 7시까지 Route 1을 통과하는데 필요한 시간은 꾸준히 증가하고 있으며, 9월 6일의 사례와 달리 여행 시간이 급격하게 증가하지 않는다. 그림 7(d)는 Route 2에서 8월 30일 오전 6시부터 교통 혼잡이 발생한다는 것을 보여준다. 흥미롭게도 Route 2에서는 8월 30일의 교통 정체가 9월 6일의 교통 혼잡보다 더 심한 것을 볼 수 있다. 차량이 Route 1에서 발생한 사고로 인해 시간당 Route 2에 진입할 수 있는 차량의 수가 줄어들어서 발생한 것으로 해석할 수 있다. Route 2의 평균적인 N-curve는 9월 6일의 패턴을 제외하고는 그림 7(d)의 패턴과 비슷하다.

7. 한계점

본 논문에서는 교통 혼잡의 원인과 영향을 분석하기 위한 시각적 분석 시스템을 소개했다. 평균 속도와 속도 엔트로피를 이용하여 교통 흐름의 변화를 식별하고, 합성곱 신경망을 사용하여 N-curve를 4가지의 혼잡 유형으로 분류했다. 혼잡 원인 분석의 유용성은 사례 연구를 통해 입증되었다고 생각한다. 그럼에도 불구하고 이 기술이 교통 상황을 이해하는 데 어떻게 적절한 지에 대한 심층적인 토론이 필요하다. 교통 정체 원인은 일반적으로 날씨, 특정 지역의 사건, 사고, 도로 상태, 교통 신호 제어 및 운전자의 잘못된 습관과 같은 다양한 요소의 조합으로 발생한다. 시각적 분석 시스템을 통해 사용하는 교통 신호 시스템, 병목 현상 및 너무 많은 차량으로 인한 정체 원인을 분석할 수 있다. 외부 데이터와 연계하여 교통사고 및 악천후로 인한 교통 혼잡을 분류할 수 있다. 교통사고, 악천후 및 지역 행사와 같은 비정상

적인 혼잡 상황은 많은 외부 데이터와 함께 분석해야 한다.

혼잡은 한 지역에만 국한되지 않고 주변 지역의 영향을 받기 때문에 교통 혼잡은 다양한 관점에서 분석되어야 한다. 그러므로 인과 관계와 교통 흐름 간의 상관관계를 분석 할 필요가 있다. 또한 하나의 혼잡 지역뿐만 아니라 서로 연결된 다른 지역과의 관계도 분석하는 것이 중요하다. 제안된 시스템은 혼잡 지역과 연결된 주변 지역과의 교통 흐름을 분석할 수 있다. 향후에 교통 혼잡이 주변 지역에 미치는 영향을 자동으로 계량화할 수 있는 모델을 연구할 예정이다.

본 연구에서는 GPS trajectory 데이터를 이용하여 교통 혼잡의 원인을 분석했다. 베이징 택시 데이터에서 km당 최소 6번 이상 GPS 레코드가 기록된 차량의 trajectory 데이터를 사용한다. 그럼에도 불구하고, 택시 및 버스 데이터를 사용하는 N-curve는 통계적 대표성에 관한 문제가 발생할 수 있다. 제시한 교통 혼잡 분석 시스템은 도로상의 모든 차량을 포함하여 고품질의 데이터를 활용할 때 신뢰할만한 결과를 제공 할 수 있다. 불행하게도 모든 차량 움직임이 기록 된 데이터 세트를 얻는데 어려움이 있었다. 대신 차량 탐지기(VD) 데이터를 사용하여 제시한 혼잡 원인 분석의 접근 방식을 검증했다. VD 데이터에는 도로를 통과하는 모든 차량의 주행 기록이 포함된다. Seattle 사례 연구에서 볼 수 있듯이 N-curve를 통해 교통 정체 상황을 완벽하게 분석 할 수 있다.

8. 결론

본 논문에서는 교통 흐름 이론을 바탕으로 혼잡 원인과 영향을 분석할 수 있는 방법을 제안했다. 제안된 방법은 차량의 흐름 정보가 기록된 GPS trajectory와 Vehicle Detector 데이터에서 엔트로피와 속도를 이용하여 차량의 교통 혼잡이 발생한 것으로 예상되는 지역을 탐지한다. 교차로에서 차량의 흐름 정보를 요약할 수 있는 glyph를 이용하여 교통 혼잡이 발생한 도로를 식별한다. 교통 흐름 이론에서 사용되는 누적 차량 수 커브(N-curve)를 이용하여 혼잡도로를 통과하는 차량의 흐름을 정량화한다. Convolutional Neural network(CNN)를 이용하여 N-curve를 네 가지의 혼잡 유형으로 분류한다. Calendar chart와 지도를 이용하여 혼잡 영향을 분석한다. 시각적 분석 시스템은 여러 시각화와 상호작용 인터랙션을 이용하여 교통 혼잡 지역을 식별하고 원인과 영향을 분석할 수 있도록 지원한다. 사례연구에서 최적화 되지 못한 병목 현상 및 너무 많은 차량으로 인한 교통 혼잡을 감지할 수 있다. 기상정보 및 교통사고 기록과 같은 외부 데이터 셋을 활용하여 교통사

고 및 악천후와 같은 비정상적인 교통 혼잡 원인을 식별할 수 있었다. 향후 연구로 TD 패턴을 분류할 수 있는 최적화된 아키텍처를 연구할 예정이다. 더불어 교통과 관련된 다양한 매개 변수를 사용하여 심층학습 기반 교통 혼잡 분류 모델을 개발할 계획이다. 또한 도시 수준에서 동적으로 최적의 교통 신호를 제어할 수 있는 방법을 연구할 계획이다.

References

- [1] Wikipedia - traffic congestion [Online] Available: https://en.wikipedia.org/wiki/Traffic_congestion (downloaded 2019)
- [2] Federal highway administration [Online] Available: <https://www.fhwa.dot.gov/policy/2013> (downloaded 2013)
- [3] CEBR, "The future economic and environmental costs of gridlock in 2030," Centre for Economics and Business Research, 2014.
- [4] W. Chen, F. Guo, and F.-Y. Wang, "A survey of traffic data visualization," *IEEE Transactions on Intelligent Transportation Systems*, Vol. 16, No. 6, pp. 2970-2984, 2015.
- [5] R. Kruger, D. Thom, M. Worner, H. Bosch, and T. Ertl, "Trajectorylenses-a set-based filtering and exploration technique for long-term trajectory data," *Computer Graphics Forum*, Vol. 32, No. 3, pp. 451-460, 2013.
- [6] H. Piringier, M. Buchetics, and R. Benedik, "Alvis: Situation awareness in the surveillance of road tunnels," *Proc. of the IEEE Conference On Visual Analytics Science and Technology*, 2012.
- [7] W. Zeng, C.-W. Fu, S. Muller Arisona, A. Erath, and H. Qu, "Visualizing waypoints-constrained origin-destination patterns for massive transportation data," *Computer Graphics Forum*, Vol. 35, No. 8, pp. 95-107, 2016.
- [8] Y.C. Chen, Y.-S. Wang, W.-C. Lin, W.-X. Huang, I.-C. Lin et al., "Interactive visual analysis for vehicle detector data," *Computer Graphics Forum*, Vol. 34, No. 3, pp. 171-180, 2015.
- [9] M. L. Pack, K. Wongsuphasawat, M. VanDaniker, and D. Filippova, "Ice-visual analytics for transportation incident datasets," *Proc. of the IEEE Conference on Information Reuse&Integration*, 2009.
- [10] M. VanDaniker, "Visualizing real-time and archived traffic incident data," *Proc. of the IEEE Conference on Information Reuse&Integration*, 2009.
- [11] N. Ferreira, J. Poco, H. T. Vo, J. Freire, and C. T. Silva, "Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 19, No. 12, pp. 2149-2158, 2013.
- [12] L. Xin, D. Yang, Y. Chen, and Z. Li, "Traffic flow characteristic analysis at intersections from multi-layer spectral clustering of motion patterns using raw vehicle trajectory," *Proc. of the IEEE Conference on Intelligent Transportation Systems*, 2011.
- [13] L. Yu, W. Wu, X. Li, G. Li, W. S. Ng, S.-K. Ng, Z. Huang, A. Arunan, and H. M.Watt, "iviztrans: Interactive visual learning for home and work place detection from massive public transportation data," *Proc. of the IEEE Conference on Visual Analytics Science and Technology*, 2015.
- [14] W. Zeng, C.-W. Fu, S. M. Arisona, and H. Qu, "Visualizing interchange patterns in massive movement data," *Computer Graphics Forum*, Vol. 32, No. 3, pp. 271-280, 2013.
- [15] C. Peng, X. Jin, K.-C. Wong, M. Shi, and P. Li'o, "Collective human mobility pattern from taxi trips in urban area," *PloS one*, Vol. 7, No. 4, 2012.
- [16] M. Lu, C. Lai, T. Ye, J. Liang, and X. Yuan, "Visual analysis of multiple route choices based on general gps trajectories," *IEEE Transactions on Big Data*, Vol. 3, No. 2, pp. 234-247, 2017.
- [17] X. Huang, Y. Zhao, C. Ma, J. Yang, X. Ye, and C. Zhang, "Trajgraph: A graph-based visual analytics approach to studying urban network centralities using taxi trajectory data," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 22, No. 1, pp. 160-169, 2016.
- [18] Z. Wang, M. Lu, X. Yuan, J. Zhang, and H. Van De Wetering, "Visual traffic jam analysis based on trajectory data," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 19, No. 12, pp. 2159-2168, 2013.
- [19] M. Veloso, S. Phithakitnukoon, and C. Bento, "Urban mobility study using taxi traces," *International workshop on Trajectory data mining and analysis*, 2011.
- [20] J. Yuan, Y. Zheng, L. Zhang, X. Xie, and G. Sun, "Where to find my next passenger," *Proc. of the International conference on Ubiquitous computing*, 2011.
- [21] D. Liu, D. Weng, Y. Li, J. Bao, Y. Zheng, H. Qu, and Y. Wu, "Smartadp: Visual analytics of large-scale taxi trajectories for selecting billboard locations," *IEEE transactions on visualization and computer graphics*, Vol. 23, No. 1, pp. 1-10, 2017.
- [22] Y.C. Chen, Y.-S. Wang, W.-C. Lin, W.-X. Huang, I.-C. Lin et al., "Interactive visual analysis for vehicle detector data," *Computer Graphics Forum*, Vol. 34, No. 3, pp. 171-180, 2015.
- [23] G. Di Lorenzo, M. Sbodio, F. Calabrese, M. Berlingerio, F. Pinelli, and R. Nair, "Allaboard: visual exploration of cellphone mobility data to optimise public transport," *IEEE transactions on visualization and computer graphics*, Vol. 22, No. 2, pp. 1036-1050, 2016.
- [24] W. Pattara-Atikom, P. Pongpaibool, and S. Thajachayapong, "Estimating road traffic congestion

- using vehicle velocity," *Proc. of the IEEE Conference on Intelligent Transportation Systems*, 2006.
- [25] Q.-J. Kong, Q. Zhao, C. Wei, and Y. Liu, "Efficient traffic state estimation for large-scale urban road networks," *IEEE Transactions on Intelligent Transportation Systems*, Vol. 14, No. 1, pp. 398-407, 2013.
- [26] J. Lu and L. Cao, "Congestion evaluation from traffic flow information based on fuzzy logic," *Proc. of the IEEE conference on Intelligent Transportation Systems*, 2003.
- [27] X. Zhang, E. Onieva, A. Perallos, E. Osaba, and V. C. Lee, "Hierarchical fuzzy rule-based system optimized with genetic algorithms for short term traffic congestion prediction," *Transportation Research Part C: Emerging Technologies*, Vol. 43, pp. 127-142, 2014.
- [28] W. Zheng, D.-H. Lee, and Q. Shi, "Short-term freeway traffic flow prediction: Bayesian combined neural network approach," *Journal of transportation engineering*, Vol. 132, No. 2, pp. 114-121, 2006.
- [29] H. Shankar, P. Raju, and K. R. M. Rao, "Multi model criteria for the estimation of road traffic congestion from traffic flow information based on fuzzy logic," *Journal of Transportation Technologies*, Vol. 2, No. 1, 2012.
- [30] X. Zhou, W. Wang, and L. Yu, "Traffic flow analysis and prediction based on gps data of floating cars," *Proc. of the 2012 International Conference on Information Technology and Software Engineering*, 2013.
- [31] Beijing taxi data [Online]. Available: <http://sensor.ee.tsinghua.edu.cn/datasets.html> (downloaded 2017)
- [32] San Francisco taxi data [Online] Available: <https://crawdad.org/epfl/mobility/20090224/> (downloaded 2018)
- [33] OSM extended API [Online] Available: <https://wiki.openstreetmap.org/wiki/Xapi> (downloaded 2019)
- [34] Seattle Traffic Detectors [Online] Available: [https://data.seattle.gov/Transportation /Traffic-Detectors/f6h6-kgmz](https://data.seattle.gov/Transportation/Traffic-Detectors/f6h6-kgmz) (downloaded 2019)
- [35] C. E. Shannon, "A mathematical theory of communication," *Bell system technical journal*, Vol. 27, No. 3, pp. 379-423, 1948.
- [36] M. J. Cassidy and R. L. Bertini, "Some traffic features at freeway bottlenecks," *Transportation Research Part B: Methodological*, Vol. 33, No. 1, pp. 25-42, 1999.
- [37] L. Elefteriadou, "An Introduction to Traffic Flow Theory," Springer, New York, 2014.
- [38] Seattle police department 911 incident response [Online] Available: <https://catalog.data.gov/dataset/seattle-police-department-911-incident-response-52779> / (downloaded 2019)



피민규

2016년, 2018년 세종대학교 컴퓨터공학과 졸업(학사, 석사). 2018년~현재 세종대학교 컴퓨터공학과 박사과정. 관심분야는 Visual analytics, 기계학습



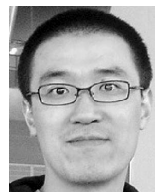
연한별

2012년, 2014년, 2019년 세종대학교 컴퓨터공학과 졸업(학사, 석사, 박사). 2019년~현재 세종대학교 컴퓨터공학과 박사후연구원. 관심분야는 데이터 시각화, Visual Analytics, 기계학습



손혜숙

2019년 세종대학교 응용통계학과 학사 졸업. 2019년~현재 세종대학교 컴퓨터공학과 학생연구원. 관심분야는 기계학습, 예측모델링, 데이터시각화



장윤

2000년 서울대학교 전기공학부 졸업(학사) 2002년, 2007년 미국 Purdue University ECE 졸업(석사, 박사). 2007년~2009년 Swiss National Supercomputing Center 박사 후 연구원. 2009년~2011년 Swiss ETH Zurich 박사 후 연구원. 2012년~현재 세종대학교 컴퓨터공학과 부교수. 관심분야는 데이터 시각화, Visual Analytics, 빅데이터 분석 및 시각화, 컴퓨터 그래픽스