

# 요인분석 및 군집분석을 활용한 교통상황 유형 분류분석

## A Study of Classification Analysis about Traffic Conditions Using Factor Analysis and Cluster Analysis

정수환\* · 한경희\*\* · 소재현\*\*\* · 이철기\*\*\*\*

\* 주저자 : 아주대학교 교통공학과 석사과정  
 \*\* 공저자 : 아주대학교 교통공학과 박사과정  
 \*\*\* 교신저자 : 아주대학교 교통시스템공학과 교수  
 \*\*\*\* 공저자 : 아주대학교 교통시스템공학과 교수

Su-hwan Jeong\* · Kyeong-hee Han\* · Jaehyun (Jason) So\*\* · Choul-ki Lee\*\*

\* Dept. of Transportation Eng., Univ. of Ajou  
 \*\* Dept. of Transportation System Eng., Univ. of Ajou

† Corresponding author : Jaehyun (Jason) So, jso@ajou.ac.kr

Vol. 22 No.1(2023)  
 February, 2023  
 pp.65~80

pISSN 1738-0774  
 eISSN 2384-1729  
<https://doi.org/10.12815/kits.2023.22.1.65>

Received 4 November 2022  
 Revised 29 November 2022  
 Accepted 15 December 2022

© 2023. The Korea Institute of  
 Intelligent Transport Systems. All  
 rights reserved.

### 요약

본 연구에서는 교통상황을 대변할 수 있는 주요 요인변수 도출을 목적으로 교통상황 유형에 대해 분류분석을 수행하였다. TTI(Travel Time Index)를 교통상황 판단 기준으로 사용하였고 VDS에서 일반적으로 검지되는 데이터를 활용하여 분석을 수행하였다. 먼저 요인분석을 통해 교통상황에 영향을 주는 주요인을 선정하였고, 주요인에 대하여 군집분석을 통해 교통상황을 군집화하였다. 그 후 TTI를 기준으로 각 군집별 분산분석을 실시하고 유사한 군집을 병합하여, 교통상황 유형을 분류하였다. 분석 결과 교통상황을 대변할 수 있는 주요 요인변수로 최대대기행렬길이와 점유율을 도출하였다. 본 연구 방법론을 통해 교통상황에 영향을 미치는 주요 요인변수만을 활용하여 효율적인 교통혼잡 관리가 가능할 것으로 기대된다.

핵심어 : VDS, 교통상황 분류, 요인분석, 군집분석, 분산분석

### ABSTRACT

In this study, a classification analysis was performed based on the type of traffic situation. The purpose was to derive the major variable factors that could represent the traffic situation. The TTI(Travel Time Index) was used as a criterion for determining traffic conditions, and analysis was performed using data generally detected by the Vehicle Detecting System(VDS). First, the major factors influencing the traffic situation were selected through factor analysis, and traffic conditions were clustered through a cluster analysis of the major factors. After that, variance analysis for each cluster was performed based on the TTI, and similar clusters were merged to categorize the type of traffic situation. The analysis derived, the maximum queue length and occupancy as major factors that could represent the traffic situation. Through this study, it is expected that efficient management of traffic congestion would be possible by just concentrating on the main variable factors that affect the traffic situation.

Key words : VDS, Traffic type classify, Factor analysis, Clustering analysis, Analysis of variance

## I. 서론

교통혼잡비용은 2016년 55.86조원, 2017년 59.62조원, 2018년 67.76조원, 2019년 70.62조원으로 매년 증가하고 있는 추세를 보이고 있으며, GDP(Gross Domestic Product) 대비 약 3.2 ~ 3.5%의 비중을 차지하고 있다.(The Korea Transport Institute, 2021a) 비록 2020년 교통혼잡비용이 전년대비 6.5% 감소하였지만, 이는 코로나 19로 인한 교통량 감소와 속도 증가의 영향으로 분석되었다.(The Korea Transport Institute, 2021b) 이렇듯 교통혼잡으로 발생하는 손실이 계속해서 증가하고 있으며, GDP의 높은 비중을 차지할 정도로 국가 경제에 큰 영향을 미치고 있어 사회·경제적 측면에서 교통혼잡의 지속적인 관리는 필수적이다. 따라서 혼잡의 정도를 판단하는 것이 매우 중요하다.

기존에 도로의 혼잡을 분석하는 기법으로 도로용량편람(Highway Capacity Manual, HCM)의 서비스수준(Level of Service, LOS) 판단 효과척도(Measure of Effectiveness, MOE)를 주로 사용하고 있으며 HCM은 국내 외로 도로체계의 계획, 설계, 분석, 운영에 사용되는 핵심적인 기준이 된다. 현재 국내에서는 「도로용량편람 2013」을 활용하고 있으며, 도로 유형에 따라 밀도, 제어지체 등을 효과척도로 제시하고 있다. 그러나 일부 도로 시설의 효과척도가 신호조건 및 구간길이에 영향을 받고, 다양한 교통류에 적합한 차로군 분류를 하지 못하는 등 서비스 분석 결과가 합리적이지 않아 수정 및 보완이 필요하다는 의견이 제기되고 있다.(The Korea Transport Institute, 2021c; Kim et al., 2014)

이에 따라 국내외적으로 통행시간, 통행속도, 교통량, 대기행렬 등 검지 데이터를 기반으로 혼잡 판단을 위한 지표 개발 등 다양한 연구들이 활발히 진행되고 있다. 기존 연구들에서는 개별 검지 데이터만을 활용하였으나, 최근에는 다양한 검지 데이터를 함께 고려하여 혼잡지표 개발 및 관련 연구들을 수행하고 있다. 많은 종류의 검지 데이터를 고려할수록 다양한 교통상황을 반영할 수 있지만, 서로 상관관계가 높은 데이터들 간의 다중공선성이 존재하여 불필요한 연산이 수반될 수 있다. 특히 AI(Artificial Intelligence)와 빅데이터 시대가 도래하면서 교통 분야에서는 C-ITS(Cooperative Intelligent Transport Systems) 및 자율주행기술이 등장하였으며, 이러한 기술들을 통해 더욱 많은 양과 다양한 종류의 데이터가 수집되고 있다. 또한, 실시간으로 정보를 주고받는 이러한 시스템 속에서 무조건적으로 많은 종류의 데이터를 활용하는 것은 비효율적이다.

따라서 본 연구의 목적은 교통혼잡 판단에 필요한 주요 교통 데이터를 선정하고자, 다양한 검지 데이터에 따른 교통상황 유형을 분류하고 각 유형들이 어떠한 영향을 미치는지 분석하는 것이다. 주요 교통 데이터만을 선정하며 검지되는 다양한 빅데이터 차원을 축소함으로써, 불필요한 변수의 저장을 하지 않아도 되므로 데이터 분석 시간의 감소가 기대된다. 또한 혼잡지표 개발에 많은 종류의 데이터를 활용할 경우, 모델의 복잡도가 증가하여 차원의 저주라고 불리는 과적합이 발생할 수 있다. 이를 위해 VDS(Vehicle Detecting System)에서 수집되는 교통 데이터 중 교통상황에 영향을 주는 주요인을 선정하였고, 교통혼잡을 기준으로 교통상황 유형을 분류하였다. 그리고 신호조건과 같이 기존 혼잡지표의 한계점과 관련된 요소를 포함하고자, 본 연구의 공간적 범위를 단속류로 설정하였다.

## II. 관련 이론 및 연구 고찰

### 1. VDS 수집 데이터 관련 고찰

VDS의 검지기로 루프검지기와 영상검지기가 있다. 루프검지기는 검지장치에서 공급하는 10~200KHz의 주

과수를 갖는 에너지로 인해 도로 위에 매설된 루프에 균일한 인덕턴스를 가진 교번자장이 형성되며, 루프 위를 차량이 통과할 때의 인덕턴스 변화를 검지하는 방식이다. 이러한 원리로 루프검지기는 일반적으로 교통량, 지점속도, 점유율, 대기행렬길이 등을 수집한다. 영상검지기는 영상처리기술을 활용하며, 기본적인 검지 원리는 루프검지기와 유사하다. 영상검지기는 일반적으로 교통량, 차량군별 지점속도, 점유율, 대기행렬길이 등을 수집한다.(KoROAD, 2014)

## 2. 혼잡지표 관련 고찰

Lee et al.(2008)에서는 도로 용량을 초과하는 과도한 교통수요, 도로 기하구조 문제, 상층, 교통사고 등 다양한 원인에 의해 발생하는 차량의 지·정체 및 긴 대기행렬을 일반적으로 교통 혼잡이라 기술하였다. 즉, 정상적인 자유흐름(free-flow) 통행상태에서 소요되는 통행시간과 지체상태에서 소요되는 통행시간을 비교하였을 때 나타나는 통행시간이나 지체의 증분으로 설명할 수 있다는 것이다. 또한 기존에 많이 사용하던 혼잡측정지표에 대하여 <Table 1>과 같이 제시하였다. 통행시간 기반의 지표로, TTI(Travel Time Index)는 자유류 통행시간과 침두시 통행시간의 비율을 의미한다. 또한 TRI(Travel Rate Index)는 침두시간 통행률을 비침두시간 통행률로 나눈 값이다. 교통량 기반의 지표로, RCI(Roadway Congestion Index)는 단위 도로연장 당 일 교통량을 사용하여 산출한다. 또한 CSI(Congestion Severity Index)는 LOS D 이상인 상태에서 발생하는 총 지체시간을 통행량으로 나눈 값이다. 그리고 혼잡시간 기반의 지표로, LMDI(Lane-Mile Duration Index)는 각 도로구간 별로 혼잡한 도로연장에 혼잡체류시간을 곱하여 합계를 낸 값이다. 또한 FCI(Freeway Congestion Index)는 LMDI가 도로 일정구간의 혼잡 크기를 나타낼 수는 있으나 서로 다른 도로간의 상호 비교가 불가능한 단점을 보완하기 위해 LMDI를 전체 도로연장으로 나눔으로써 정규화한 값이다. 마지막으로 밀도 기반의 지표로, OCI(Occupancy Congestion Index)는 포화통행밀도(jam density)의 역수에 100을 곱한 값에 실제 통행밀도를 곱하여 혼잡의 강도를 나타낸다.

<Table 1> Characteristics by Traffic Condition Index

Base Traffic Data	Traffic Condition Index	Characteristics
Travel Time	TTI (Travel Time Index)	It has the advantage of being easy to obtain input data and simple calculation because it can be applied arbitrarily to the desired analysis time zone.
	TRI (Travel Rate Index)	Because only congestion during peak hours is considered, congestion at non-peak times is not considered.
Traffic Volume	RCI (Roadway Congestion Index)	Since traffic volume is used as a criterion for determining congestion, the more severe the congestion, the less the traffic volume may decrease.
	CSI (Congestion Severity Index)	Since it has a one-year temporal range and a spatial range of the entire city, it is difficult to use it without extensive basic data.
Congestion Time	LMDI (Lane-Mile Duration Index)	Comparison of relative congestion between different road routes is not possible.
	FCI (Freeway Congestion Index)	Developed to compensate for the shortcomings of LMDI, it is an indicator that allows relative comparison of congestion, but congestion intensity is not considered.
Density	OCI (Occupancy Congestion Index)	Unlike other congestion indicators considering only the temporal and spatial extent of congestion, congestion indicators considering the intensity of congestion together.

Kong et al.(2015)는 교통 혼잡지표로 TTI를 선정하였고 택시 GPS 데이터를 활용하여 계산하였다. TTI를

지표로 제시한 근거로는 다른 혼잡지표 대비 두 가지 장점으로 설명하였다. 첫 번째로 혼잡도와 도로 시설 조건을 구분할 수 있다. 평균통행속도를 활용하는 베이징의 혼잡 평가 시스템은 각각의 도로마다 조건이 다르므로 속도 뿐만 아니라 도로 조건 또한 조사해야 한다. 그러나 TTI는 자유흐름 상태의 평균속도를 파라메타로 고려하고 있기 때문에 계산식에 도로 조건을 포함하고 있다. 두 번째로 TTI는 직관적으로 이해하기 용이하다. 예를 들어 TTI가 1.2라면 자유흐름 상태에서보다 20% 더 긴 통행시간이 소요됨을 의미한다. 또한 앞서 언급한대로 각 도로 조건을 포함하고 있으므로 각 도로 구간의 혼잡도를 비교하기에도 용이하다.

Cho(2017)는 도로 이동성의 증대를 위한 사업추진에 대해 분석 및 모니터링이 이루어지지 못하고 교통 혼잡 이동성 평가지표가 부재한 문제를 해결하기 위해, 교통류와 혼잡 상태를 판단할 수 있는 지표를 제시하였다. 지표로는 국내 실정에 부합하고 다양한 분석에 활용할 수 있는 TTI를 선정하였다. TTI를 산출하는 과정에서 자유흐름상태에서의 통행시간으로 제한속도나 설계속도, 비첨두시 85퍼센타일 속도를 사용한다. TTI를 지표로 하여 성능을 검토한 결과, 활용 가능성과 적용성이 상당한 수준으로 도출되었다.

통행시간지표(TTI)는 자유통행시간 대비 실제 통행시간의 비로 산출한다. 자유통행시간은 자유속도(free flow speed)로 해당 구간을 주행할 때 소요되는 시간이고, 실제 통행시간은 해당 시간대에 실제로 주행한 (actual speed) 통행시간이다.(Han and Kim, 2017)

$$TTI = \frac{TT_{actual\ speed}}{TT_{free\ flow\ speed}} \dots\dots\dots (1)$$

where,  $TTI$  = Travel Time Index,  
 $TT_{observed}$  = observed travel time,  
 $TT_{free\ flow\ speed}$  = free-flow travel time,

### 3. 교통상황 유형 분류 관련 고찰

Sim and Chou(2009)는 RWIS(Road Weather Information System)를 통해 수집되는 다양한 기상요인의 영향 정도에 따른 속도변화를 분석하여 고속도로의 교통상황을 분류하였다. 모든 기상요인을 고려할 경우 분류되는 집단의 수가 너무 많아져 불필요한 연산을 추가적으로 수행해야 한다. 따라서 기상요인에 대해 요인분석을 실시하여 교통상황에 영향을 주는 주요인만을 추출하고, 이를 활용하여 분산분석을 통해 교통상황을 분류하였다. 요인분석에는 주성분 분석과 공통요인분석 2가지 방법이 있다. 공통요인분석은 측정변수 간 상관관계를 설명하는 공통요인을 추출하는 기법이다. 반면, 주성분 분석은 측정변수의 선형결합을 통해 주성분을 추출하므로 해당 연구에서는 주성분 분석을 활용하였다. 온도, 습도, 풍속, 시정거리 등 기상요인에 대한 요인분석 결과 3개의 주요 성분(시정거리, 온도, 날씨)으로 분류되었다. 또한 분산분석은 먼저 등분산성 검정을 통해 분석기법을 결정하고, 각 집단의 평균 간의 차이가 있는지 분석한다. 분석 결과로 각 집단의 평균에 차이가 있다고 도출되면, 사후검증을 통해 어떤 집단 간 어떠한 차이가 있는지 비교한다. 해당 연구에서 분류된 요인을 대상으로 분산분석을 수행한 결과 날씨에 따라 맑음과 강우, 온도는 5℃를 기준으로 분류되었으며, 시정거리는 강우 시에만 10km를 기준으로 분류되었다. 각 분류 기준에 의해 교통상황은 5가지로 분류되었다.

Yuan et al.(2017)은 다변량 통계분석을 활용하여 교통사고의 주요 영향 요인을 분석하였다. 2008년부터

2014년까지 중국에서 10명 이상의 사망자를 발생시킨 심각한 교통사고 데이터를 수집하여 요인분석과 군집 분석을 실시하였다. 먼저 군집분석을 통해 수집한 교통사고 데이터에 대하여 인적·차량·도로 등 23개 요인을 도출하였다. 그 후 요인분석을 통해 23개 변수에서 공통요인을 추출하였다. 요인분석 기법 중 주성분 분석을 통해 각 변수의 요인 적재 행렬을 계산하고, 분산의 합계를 최대화함으로써 행렬을 단순화하는 베리맥스(VARIMAX) 요인회전을 통해 요인점수를 계산하였다. 마지막으로 산출한 요인점수를 활용하여 23개 요인에 대한 순위를 가렸다. 분석 결과 과속(8.0%), 램프(6.3%), 날씨(6.1%), 노면(5.9%), 운전경력(5.6%) 등의 순으로 교통사고에 영향을 미치는 것으로 도출되었다. 이를 통해 교통안전에 관한 정책적 제언을 제시하였으며, 향후 빅데이터를 교통에 적용하는 데 역할을 할 것으로 기대하였다.

#### 4. 분류 분석 기법 고찰

Cho and Kim(2007)은 특수법인 경영지표를 세분화하여 전략방향을 도출하기 위해, 목적변수가 존재하지 않을 경우의 군집화를 수행하였다. 많은 변수에 대해 군집분석만을 활용하여 세분화할 경우, 군집별 특성이 어렵게 된다. 따라서 군집의 특성을 잘 반영할 수 있는 대표적인 주요인을 요인분석을 통해 추출하고, 이를 바탕으로 군집분석을 실시하였다.

Hong et al.(2020)은 소비자별 구매 품목 간 연관성을 파악하기 위해 소비자의 상품 구매 데이터를 군집화하여 소비자의 유형을 구분하는 장바구니 분석을 수행하였다. 군집화는 K-Means Clustering 방법을 활용하였으며, 군집분석에 앞서 17개의 많은 데이터 유형의 차원 축소를 위해 요인분석 기법들을 검토 및 비교하였다. 변수의 개수가 많은 경우 군집 분류에 고려되지 않을 가능성이 있으며, 이러한 변수들은 군집구조를 찾아내는 것을 방해할 가능성이 있다는 점을 근거로 군집분석에 앞서 요인분석을 수행하였다.

요인분석은 변수들 간 상호 관련성(공분산 또는 공분산을 표준화한 상관계수)을 분석해서 이들 간에 공통적으로 작용하는 요인을 추출하여 정보의 손실을 최소화 한 상태로 변수를 축약하는 기법이다. 이러한 요인 분석을 활용하면 다수의 변수를 핵심적인 요인으로 축약하여 정보에 대한 이해력을 높일 수 있다. 요인 분석을 실시하기 위해서는 먼저 관련 변수들이 연속형 데이터여야 하고, 관련 변수들 간 어느 정도의 상관관계가 있어야 한다. 일반적으로 변수들이 모두 상관관계가 없음을 귀무가설로 하는 바틀렛 검정(bartlett test)과 상관관계가 얼마나 작은지를 확인하는 KMO(Kaiser-Meyer-Olkin) 검정을 활용한다. 바틀렛 검정의 경우 대립가설이 채택되는지, KMO 검정의 경우 KMO 값이 최소 0.5 이상인지 확인하여야 한다. 그리고 요인의 고유값(eigenvalue)을 기준으로 요인수를 결정하여야 한다. 고유값이 1 이상일 때 요인이 최소한 변수 하나의 분산을 설명하므로, 해당 기준으로 요인수를 결정한다. 다음으로 추출된 요인들과 변수들 간의 관계를 보다 명확하게 하기 위해 요인을 회전시킨다. 일반적으로 베리맥스(VARIMAX) 직각 회전 방법을 활용한다. 마지막으로 산출한 요인 적재량)을 바탕으로 변수를 정제한다. 요인 회전 결과로 산출한, 요인 적재량이 가장 높은 변수를 요인으로 선정한다.(Choi and You, 2017)

K-Means Clustering은 각 군집의 중심점(centroid)과 군집에 포함되는 각 개체의 거리 제곱의 합을 비용함수로 설정하고, 그 비용함수를 최소화시키는 것을 목적으로 각 개체(데이터)의 소속 군집을 업데이트하는 군집분석 기법이다. 먼저, 군집수(K)가 사전에 결정되어야 한다. 다음으로, 모든 객체에 대하여 임의로 설정된 K개의 초기 중심점과의 거리를 계산하여 가장 거리가 짧은 군집에 배정한다. 배정된 결과를 기준으로 각 군집별로 평균값을 계산하여 중심점을 갱신한다. 각 군집과의 거리 계산과 중심점 갱신의 과정을 중심점의 변

1) 요인 적재량 : 각 변수와 해당 요인 간의 상관계수

화가 충분히 작아질 때까지 반복하여, 객체 별로 소속하는 군집을 배정한다.(Lee et al., 2020)

## 5. 시사점 및 연구 차별성

본 연구에서는 VDS에서 일반적으로 검지되는 교통량, 지점속도, 점유율, 평균대기행렬길이, 최대대기행렬 길이를 활용하여 분석을 수행하였다. 교통상황을 나타내는 값으로 선행연구에서 자주 활용된 혼잡지표인 TTI를 사용하였고, 군집화를 수행하여 교통상황 유형을 분류하였다. TTI는 계산식에 도로 조건을 포함하고 있어 각 도로 구간의 혼잡도를 비교하기에 용이하고 그 수치가 직관적으로 이해하기 쉬워, 많은 선행연구에 활용하고 있다. 그러나 다른 혼잡지표의 경우 도로 조건을 포함하고 있지 않아, 도로 조건을 함께 고려해야 하므로 범용성이 낮다. 또한 TTI는 관측통행시간(observed travel time)을 자유통행시간(free-flow travel time)으로 나누어 산출하였다. 자유통행시간은 비첨두 시 85퍼센타일 속도를 사용하였다. 군집화는 요인분석을 통해 추출한 주요인에 대해 군집분석을 실시하여, 군집분석만을 활용할 때의 단점을 보완하였다. 군집분석만을 실시하여 군집화를 수행할 경우, 너무 많은 변수를 활용하여 군집화에 고려되지 않는 변수들이 존재할 가능성이 있으므로 군집별 특성화가 어렵다. 요인분석은 교통상황 유형의 주요인을 추출하기 위해 주성분 분석(Principal Component Analysis, PCA)을 사용하였고 군집분석은 K-Means Clustering 기법을 사용하였으며, 요인분석에 앞서 상관분석을 수행하여 요인분석의 가능성을 정성적으로 판단하였다. 마지막으로 분산분석을 통해 요인분석 및 군집분석을 통해 수행된 군집화에 대하여, 서로 동일한 군집이 있는지 최종적으로 점검하였다.

교통 혼잡지표를 개발하는 연구들은 활발히 진행되고 있다. 그러나 혼잡지표 개발을 위해 활용하는 검지 데이터에 대한 연구는 부족한 실정이다. 본 연구는 교통혼잡 판단에 필요한 주요 교통 데이터를 선정하여, 모든 검지 데이터를 활용하지 않고도 효율적으로 혼잡지표를 개발할 수 있도록 하였다. 기상상황에 대하여 교통상황에 영향을 주는 주요인을 선정하는 연구 사례가 있었지만, 본 연구에서는 검지된 교통 데이터에 대해 분석하므로 차별성이 있다.

## Ⅲ. 연구 방법론

### 1. 연구 범위

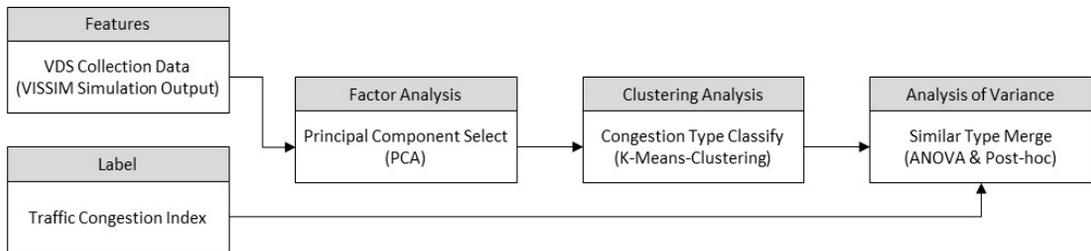
본 연구는 단속류 구간인 대치역·도곡역 부근 8개 교차로의 영상검지기를 통해, 2021년 9월 1일부터 2021년 11월 11일까지 수집한 15분 단위의 데이터를 사용하여 수행하였다. 루프검지기에서 수집할 수 있는 데이터는 영상검지기로 모두 수집이 가능하므로, 더욱 다양한 데이터를 고려하고자 영상검지기를 채택하였다. 해당 교차로의 영상검지기에서 교통량, 속도, 대기행렬길이를 수집하였으나, 속도의 경우 다수의 데이터가 누락되었다. 그러나 일반적으로 영상검지기에서 교통량과 대기행렬길이 뿐만 아니라 속도와 점유율, 통행 시간 또한 수집할 수 있다. 따라서 대치역·도곡역 부근 8개 교차로의 영상검지기에서 미검지 및 누락된 데이터 항목을 보완하기 위해, 수집한 교통량, 대기행렬길이와 micro-simulation인 VISSIM을 활용하여 시뮬레이션을 구축하였다.

구축한 단속류 구간의 시뮬레이션 환경을 기반으로 대치역과 도곡역을 연결하는 링크에 대하여, 현재 VDS에서 수집하는 일반적인 교통 데이터를 추출하여 본 연구에 활용하였다. VISSIM에서 분석 결과를 추출하는 기능으로 “Data Collection Points”, “Queue Counter”, “Vehicle Travel Time Measurements” 등이 있다. “Data

Collection Points” 기능을 활용하여 교통량과 속도, 점유율을 추출하였으며, “Queue Counter” 기능을 활용하여 평균 및 최대대기행렬길이를 추출하였고, “Vehicle Travel Time Measurements” 기능을 활용하여 통행시간을 추출하였다. 시뮬레이션은 대치역과 도곡역을 연결하는 링크 양방향에 대하여 5회 수행하였으며, 2021년 9월 1일부터 2021년 11월 11일까지 오전점두 및 오후점두와 주간 비점두 및 새벽시간대의 데이터에 대하여 결측 데이터를 제외하고 수행하였다.

## 2. 분석 방법론

본 연구에서는 먼저 일반적으로 VDS에서 수집되는 교통 데이터를 시뮬레이션을 통해 도출하였고, 교통상황을 대변할 수 있는 혼잡지표를 설정하였다. 그리고 train과 test을 수행하기 위해 각각의 데이터 셋으로 분리하였다. 그 후, train 데이터 셋을 활용하여 요인분석과 군집분석, 분산분석을 수행하여 교통상황 유형을 분류하였다. 마지막으로 test 데이터 셋을 통해 분석 결과에 대하여 검증은 수행하였다.



<Fig. 1> Schematic of the analysis procedure about Traffic Congestion Type

본 연구의 주요 분석 내용인 교통상황 유형 분류 과정은 <Fig. 1>과 같다. VDS에서 수집되는 교통 데이터에 대한 요인분석을 실시하여 교통상황에 영향을 주는 주요인을 선정하였고, 선정된 주요인에 대하여 군집분석을 통해 교통상황을 군집화하였다. 그리고 혼잡지표를 기준으로 군집별 분산분석을 실시하여 유사한 군집을 병합하여, 최종적으로 교통상황 유형을 분류하였다.

## IV. VDS 수집 데이터에 따른 교통상황 분류

### 1. 데이터 셋 분리 및 상관분석

#### 1) 기초통계분석

요인분석, 군집분석, 분산분석과 같은 응용통계분석에 앞서, 본 연구에서 활용할 데이터의 특징을 파악하기 위해 <Table 2>와 같이 기초통계분석을 수행하였다. 기초통계분석은 Python의 Pandas 라이브러리에서 평균 및 표준편차, 백분위수(Percentile)를 요약하여 출력해주는 “Describe()” 메서드를 활용하여 수행하였다. 분석 결과, 각각의 VDS 데이터 측정 단위가 다르고, 그 범위가 서로 큰 차이를 보였다. 이에 따라 각각의 VDS 데이터를 직접적으로 비교할 수 없으므로, 표준화와 같은 데이터 스케일링이 필요하다.

<Table 2> Descriptive statistics analysis about TTI and VDS collection data

	TTI	Volume (veh/hour)	Time Mean Speed(km/h)	Occupancy	Mean Queue Length(m)	Max Queue Length(m)
Count	2420	2420	2420	2420	2420	2420
Mean	1.360	314.296	31.180	0.463	59.231	122.823
Standard Deviation	0.646	74.062	4.732	0.137	49.163	77.160
Minimum	0.558	36.000	17.469	0.009	1.371	18.566
25% Percentile	0.807	272.000	27.980	0.368	28.916	74.412
50% Percentile	1.264	310.000	31.668	0.429	38.325	92.699
75% Percentile	1.646	355.000	34.840	0.573	71.497	156.977
Maximum	5.230	527.000	54.948	0.828	341.173	510.183

### 2) train 및 test 데이터 셋 분리

본 연구는 교통상황 유형을 분류분석하고 이를 검증하는 과정으로 진행하였다. 분류분석 시에는 train 데이터 셋, 검증 시에는 test 데이터 셋을 사용하였다. 모델 개발을 위해 train 데이터와 test 데이터는 일반적으로 80%대 20%로 분리한다. 따라서 train 데이터 셋과 test 데이터 셋 각각을 80%대 20%로 분리하였으며, 데이터 셋 분리는 Python의 Pandas 라이브러리를 활용하였다.

### 3) 데이터 간 상관관계 분석

교통상황 분류분석에 앞서 교통혼잡지표로 사용할 TTI와 VDS 수집 데이터에 대하여 상관분석을 수행하였다. 데이터 표본 개수가 충분히 크므로 모수적 방법인 피어슨(Pearson) 상관분석을 활용하였다. 상관분석 결과는 <Table 3>과 같이 나타났다. TTI와 VDS 수집 데이터들 간의 상관관계는 높은 것으로 분석되어, 교통상황을 대변하는 데 적절함을 보였다. 특히 평균대기행렬길이와 최대대기행렬길이는 양의 상관관계가 크게 나타났으며, 지점평균속도와 점유율은 음의 상관관계가 크게 나타났다. 이와 같이 각 교통 데이터 간에 높은 상관관계를 가지는 경우가 있으므로, 여러 개의 관련있는 변수들을 하나의 요인으로 축소하는 요인분석을 실시하였다.

<Table 3> Pearson' correlation analysis results about TTI and VDS collection data

	TTI	Volume (veh/hour)	Time Mean Speed(km/h)	Occupancy	Mean Queue Length(m)	Max Queue Length(m)
TTI	1.000	-0.090	-0.768	0.560	0.702	0.681
Volume	-0.090	1.000	-0.215	0.102	0.373	0.390
Speed	-0.768	-0.215	1.000	-0.712	-0.603	-0.593
Occupancy	0.560	0.102	-0.712	1.000	0.169	0.173
Mean Queue Length	0.702	0.373	-0.603	0.169	1.000	0.976
Max Queue Length	0.681	0.390	-0.593	0.173	0.976	1.000

## 2. 요인 분석

### 1) 요인분석 가능성 평가

요인분석(주성분 분석)을 수행하기 전, 분석 대상 데이터 셋에 대하여 요인분석 모형의 적합성 여부를 판

단해야 한다. 즉, 요인분석 가능성 평가는 데이터 셋에서 요인을 찾을 수 있는가를 의미한다. 이를 위해 바틀렛 검정(bartlett test)과 KMO 검정(Kaiser-Meyer-Olkin test)을 Python의 “factor-analyzer” 패키지를 활용하여 실시하였다. 바틀렛 검정 결과, P-value가 0.05보다 작게 도출되어 요인분석에 적합한 데이터 셋으로 판별되었다. 또한 KMO 검정 결과, KMO 값이 약 0.62로 도출되었다. 일반적으로 KMO 값이 0.6 이상일 경우 요인분석에 적합한 데이터 셋임을 의미하므로, KMO 검정 결과 또한 해당 데이터 셋이 요인분석에 적합한 것으로 판별되었다.

2) 데이터 표준화 처리

각각의 VDS 데이터는 측정 단위가 모두 동일하지 않다. 교통량은 시간당 차량 대수, 지점평균속도는 km/h, 점유율은 백분율, 평균 및 최대대기행렬길이는 m를 단위로 한다. 해당 데이터를 그대로 사용하여 분석할 경우, 분석 결과에 영향을 미칠 수 있다.

표준화는 기존 변수의 범위를 정규 분포로 변환하는 것으로, 모든 수치의 평균을 0, 분산을 1로 조정하여 각 데이터의 범위를 동일하게 하는 기법이다. 따라서 각 VDS 수집 데이터에 대하여 표준화 처리하였으며, <Table 4>와 같이 측정 단위가 상이한 상태의 데이터를 <Table 5>와 같이 표준화 값으로 변환하였다.

<Table 4> VDS data before Standardization

Volume (veh/hour)	Time Mean Speed(km/h)	Occupancy	Mean Queue Length(m)	Max Queue Length(m)
292	25.372	0.640	63.277	156.930
223	40.060	0.216	16.245	56.054
...	...	...	...	...
229	31.557	0.511	30.830	72.695
359	35.295	0.392	29.522	84.062

<Table 5> VDS data after Standardization

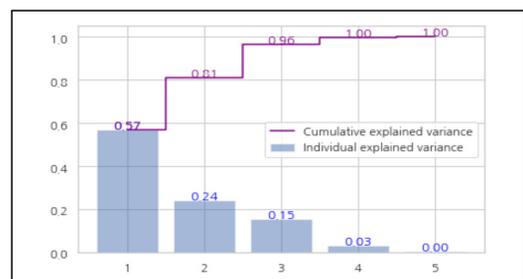
Volume	Time Mean Speed	Occupancy	Mean Queue Length	Max Queue Length
-0.318	-1.233	1.307	0.070	0.428
-1.258	1.876	-1.820	-0.871	-0.866
...	...	...	...	...
-1.176	0.076	0.356	-0.579	-0.653
0.594	0.867	-0.522	-0.605	-0.507

3) 주성분 개수 선정

추출할 주성분의 개수를 선정하기 위해 고유값(eigenvalue) 및 설명된 분산 비율(explained variance ratio)을 산출하였다. 고유값은 각 성분(component)이 설명할 수 있는 분산의 정도를 의미한다. 즉, 고유값이 1이라면 해당 성분이 한 개의 변수의 분산을 설명하고 있음을 의미한다. 일반적으로 최소 변수 한 개 이상의 분산을 설명할 수 있는 성분을 주성분으로써 선정하기 때문에, 고유값이 1 이상인 성분의 개수를 주성분의 개수로 선정한다. 고유값 및 설명된 분산 비율의 산출 결과는 <Table 6> 및 <Fig. 2>와 같으며, 고유값이 1 이상인

<Table 6> Eigenvalue and explained variance

component index	eigenvalue	explained variance ratio	cumulative explained variance ratio
1	2.846554	0.57	0.57
2	1.204197	0.24	0.81
3	0.769496	0.15	0.96
4	0.158332	0.03	1.00
5	0.024005	0.00	1.00



<Fig. 2> Individual and cumulative explained variance

성분이 두 개이므로 주성분의 개수는 두 개로 선정하였다. 이 때 첫 번째 주성분이 전체 분산의 57%를 설명하는 것으로 나타났으며, 선정한 두 개의 주성분이 전체 분산의 81%를 설명함을 보여준다.

#### 4) 요인 회전

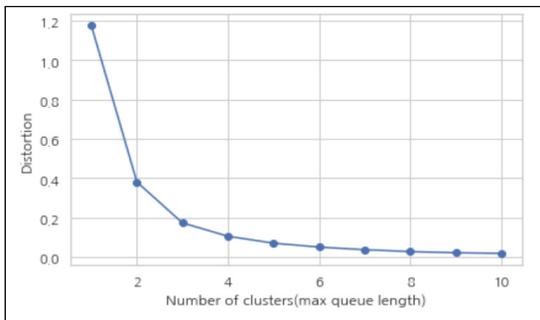
선정한 두 개의 주성분에 대해 직각 회전을 수행하여, 각각의 주성분과 가장 상관관계가 높은 VDS 수집 데이터를 주성분의 변수로 선택하였다. 각각의 VDS 수집 데이터가 여러 주성분에 비슷한 요인 적재량을 나타낼 우려가 있으므로, 요인 회전을 통해 각 데이터가 어느 한 주성분에 치우치도록 하여 요인 구조를 명확하게 하는 것이다. 요인 회전을 가장 대중적인 배리맥스(VARIMAX) 직각 회전을 수행하였다. 배리맥스 직각 회전은 하나의 요인에 대해 각 변수가 가지는 요인 적재량 제공의 분산이 최대가 되도록 변환하여 하나의 요인에 높게 적재하는 변수의 수를 줄이는 기법이다. 배리맥스 직각 회전 결과 <Table 7>과 같이 도출되었으며, 1성분과 2성분의 변수로 각각 가장 상관계수가 큰 최대대기행렬길이와 점유율을 선택하였다. 이 때 초기 고유값이 높은 1성분(최대대기행렬길이)이 2성분(점유율)보다 교통상황에 더 많은 영향을 미치는 중요한 변수가 된다.

<Table 7> PCA results by VARIMAX orthogonal rotation

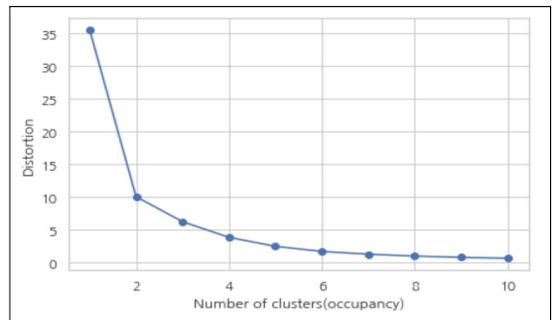
	Component	
	1	2
Volume	0.371894	0.078040
Time Mean Speed	-0.497858	-0.701434
Occupancy	0.022829	<b>*0.997540</b>
Mean Queue Length	0.975357	0.153224
Max Queue Length	<b>*0.981290</b>	0.150209

### 3. 군집 분석

#### 1) 군집수(K) 선정



<Fig. 3> Elbow graph of max queue length using SSE



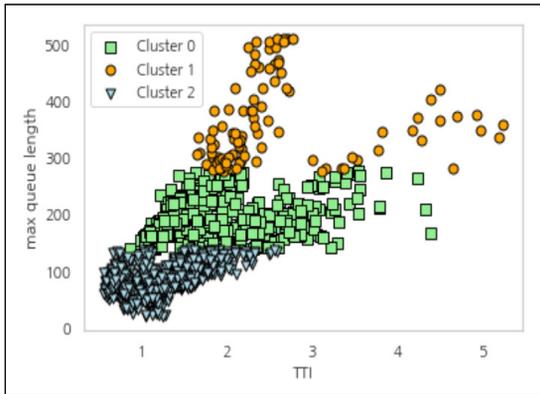
<Fig. 4> Elbow graph of occupancy using SSE

주성분으로 선택한 두 변수에 대해 군집분석을 수행하여 교통상황을 군집화하기 위해, 먼저 군집의 개수

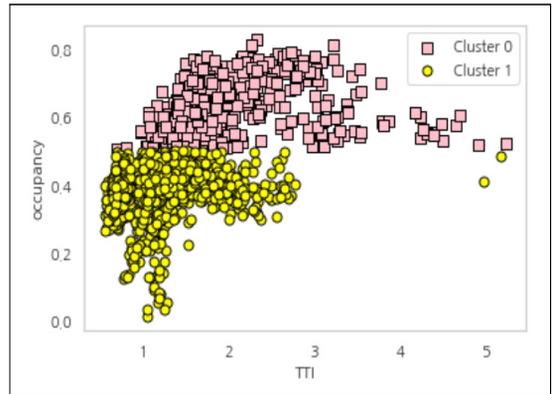
2) 요인 구조 : 어떤 변수가 어떤 요인에서 높은 상관계수를 가지는지 나타내는 요인 행렬

를 결정해야 한다. 군집의 개수는 오차제곱합(Sum of Square for Error, SSE)의 값이 빠르게 증가하는 지점의 K값을 찾는 엘보우(elbow) 기법을 사용하였다. 1성분인 최대대기행렬길이에 대한 오차제곱합 그래프, <Fig. 3>에서 볼 수 있듯이 K값이 3일 때 엘보우가 나타났다. 그리고 2성분인 점유율에 대한 오차제곱합 그래프, <Fig. 4>에서는 K값이 2일 때 엘보우가 나타났다. 이에 따라, 1성분(최대대기행렬길이)과 2성분(점유율)의 군집수는 각각 3과 2로 결정하였다.

## 2) 군집화(Clustering)



<Fig. 5> Cluster results about max queue length



<Fig. 6> Cluster results about occupancy

선정한 군집수(K)를 기반으로 K-Means Clustering 기법을 활용하여 최대대기행렬길이와 점유율에 대하여 군집화를 수행하였다. K-Means Clustering은 사전에 결정한 군집수(K)에 따라, 각각의 개체는 하나의 중심(centroid)에 할당되어 군집을 형성하는 기법이다. 최대대기행렬길이와 점유율의 군집수(K)는 앞서 선정했듯이 각각 3과 2로 설정하여 K-Means Clustering 분석을 수행하였다. 분석 후 주성분으로 선택한 두 변수와 교통혼잡지표로서 교통상황을 대변하는 TTI와의 관계를 알아보기 위해, 두 변수를 Y축 그리고 TTI를 X축으로 표현하여 분석 결과를 그래프로 나타내었다. 그 그래프는 <Fig. 5>와 <Fig. 6>과 같다. 먼저 최대대기행렬길이의 경우, 분류된 군집 및 객체가 TTI와 양의 선형관계를 강하게 보이고 있다. 이는 최대대기행렬길이 TTI를 잘 설명하고 있음을 의미한다. 점유율의 경우도 마찬가지로 분류된 군집 및 객체가 양의 선형관계를 보이고 있어, 점유율이 TTI를 잘 설명하고 있음을 알 수 있다. 또한 점유율보다 최대대기행렬길이 TTI가 더 강한 선형관계를 보이고 있으므로, 두 주성분 중 최대대기행렬길이 TTI가 교통상황에 더 많은 영향을 미치는 변수임을 다시 확인할 수 있었다.

교통상황 유형 분류를 위해 최대대기행렬길이와 점유율에 대해 분류된 각 군집의 최대·최소값을 산출하였으며, 그 결과는 각각 <Table 8>과 <Table 9>에서 나타난다. 또한 최대대기행렬길이와 점유율의 각 군집별 최대·최소값을 활용하여, 교통상황 유형을 분류하는 기준값을 도출하였다. 이 때 기준값은 연속된 군집의 최대값과 최소값의 평균값을 사용한다. 이렇게 <Table 10>과 같이 요인분석 및 군집분석을 활용하여 교통상황 유형을 총 6개로 군집화하였다.

<Table 8> Maximum and minimum value in cluster about max queue length

Cluster Classification	Minimum value in cluster	Maximum value in cluster
Cluster 2	18.566	137.836
Cluster 0	138.194	275.353
Cluster 1	276.724	510.183

<Table 9> Maximum and minimum value in cluster about occupancy

Cluster Classification	Minimum value in cluster	Maximum value in cluster
Cluster 1	0.009	0.498
Cluster 0	0.499	0.828

<Table 10> Criteria value about max queue length and occupancy by Traffic congestion type

Congestion type	a	b	c	d	e	f
Max queue length criteria value	< 138.015		< 276.039		≥ 276.039	
occupancy criteria value	< 0.499	≥ 0.499	< 0.499	≥ 0.499	< 0.499	≥ 0.499

#### 4. 분산 분석

##### 1) 등분산성 검정

요인분석 및 분산분석을 활용하여 군집화한 6개의 교통상황 유형에 대해, 각각의 군집별로 다른 교통상황을 나타내는지 비교하기 위해 TTI를 기준으로 분산분석 및 사후검증을 수행하였다. 분산분석을 통해 6개의 유형별로 TTI 값의 차이(교통상황의 차이가)가 있음을 확인하였고, 사후검증을 통해 각각의 유형 간에 어떠한 차이가 있는지 분석하였다.

분산분석 및 사후검증을 수행하기 전, 먼저 등분산성 검정을 통해 어떠한 기법을 활용하여 분석을 수행할지 판별하였다. 검정 대상이 6개의 교통상황 유형으로 2개 이상이므로, 다집단 등분산성 검정 중 하나인 Levene's test를 수행하였다. 검정 결과는 <Table 11>과 같으며, P-value가 0.05보다 작으므로 등분산성을 만족하지 않는다. 이에 따라 분산분석으로 Welch's ANOVA, 사후검증으로 Games Howell 기법을 활용하였다.

<Table 11> Levene's equal-variance test result

Equal-variance test	W (test statistics)	P-value	test result
levene	94.941889	1.371420e-89	False

##### 2) 분산 분석(Welch's ANOVA) 및 사후 검정(Games Howell)

6개의 교통상황 유형별 평균 TTI 값의 차이가 유의미한지 알아보기 위해 Welch's ANOVA 분산분석을 수행하였다. 그 결과 <Table 12>와 같이 P-value가 0.05보다 작으므로 유형별 평균 TTI 값의 차이가 통계적으로 유의미함을 확인하였다.

<Table 12> Welch's ANOVA analysis result

Source	ddof1	ddof2	F (test statistics)	P-value	analysis result
Congestion type	5	157.242404	698.268904	2.187949e-105	False

이어서 각각의 6개의 교통상황 유형 간 평균 TTI 값의 차이가 어디서 발생하는지 알아보기 위해 Games Howell 사후검증을 수행하였으며, 검증 결과는 <Table 13>과 같이 분석되었다. 거의 대부분이 P-value가 0.05 보다 작아 교통상황 유형 간 평균 TTI 값의 차이가 통계적으로 유의미하며, 요인분석 및 군집분석에 따른 유형 분류가 적절하게 이루어졌음을 확인하였다. 그러나 교통상황 유형 d와 유형 e의 경우 P-value값이 0.05보다 크므로, 평균 TTI 값이 서로 동일하다고 분석되었다. 이에 따라 유형 d와 유형 e를 병합하여, <Table 14>와 같이 기존에 분류한 6개의 교통상황 유형을 5개로 갱신하였다. 유형 A의 경우 최대대기행렬길이와 점유율이 가장 작으며, 유형 E로 갈수록 그 값은 커진다. 따라서 유형 A가 가장 교통흐름이 원활한 상태이고, 유형 E가 가장 혼잡한 상태를 의미한다.

<Table 13> Games Howell Post-Hoc test result

group 1	group 2	mean(group 1)	mean(group 2)	P-value	test result
type a	type b	0.896771	1.539201	4.773959e-14	False
type a	type c	0.896771	1.456392	0.000000e+00	False
type a	type d	0.896771	2.329411	0.000000e+00	False
type a	type e	0.896771	2.265388	1.221245e-14	False
type a	type f	0.896771	3.911600	4.491740e-12	False
type b	type c	1.539201	1.456392	6.521676e-03	False
type b	type d	1.539201	2.329411	0.000000e+00	False
type b	type e	1.539201	2.265388	2.220446e-16	False
type b	type f	1.539201	3.911600	2.954381e-10	False
type c	type d	1.456392	2.329411	1.221245e-14	False
type c	type e	1.456392	2.265388	8.548717e-15	False
type c	type f	1.456392	3.911600	1.345403e-10	False
<b>type d</b>	<b>type e</b>	<b>2.329411</b>	<b>2.265388</b>	<b>9.582671e-01</b>	<b>True</b>
type d	type f	2.329411	3.911600	1.465921e-07	False
type e	type f	2.265388	3.911600	5.076170e-08	False

<Table 14> Renew Criteria value about max queue length and occupancy by Traffic congestion type

New congestion type	A	B	C	D		E
Old congestion type	a	b	c	d	e	f
Max queue length criteria value	< 138.015		< 276.039		≥ 276.039	
occupancy criteria value	< 0.499	≥ 0.499	< 0.499	≥ 0.499	< 0.499	≥ 0.499

### 5. 교통상황 분류분석 검증

이렇게 본 연구에서 수행한 교통상황 분류분석에 대해, test 데이터 셋을 활용하여 검증하였다. 검증은 교

통상황 유형을 분류하는 기준값에 따라 train 데이터 셋과 test 데이터 셋에 대해 유형을 분류하였다. 그 후 각 교통상황 유형별로 train 데이터 셋과 test 데이터 셋을 독립표본 t 검정을 통해, 평균 TTI 값을 기준으로 두 집단의 동일성을 검증하였다. 검증 결과는 <Table 15>와 같으며, 모든 교통상황 유형에서 P-value가 0.05보다 크므로 train 데이터 셋과 test 데이터 셋의 평균 TTI 값의 차이가 유의하지 않아 서로 동일함을 보였다. 따라서 요인분석·군집분석·분산분석을 활용하여, 본 연구에서 수행한 교통상황 유형 분류분석은 적절하다고 볼 수 있다.

<Table 15> Independent two sample t-test result

congestion type	group 1	group 2	T (test statistics)	P-value	test result
type A	Train data set	Test data set	0.203367	0.838968	True
type B	Train data set	Test data set	1.754275	0.080922	True
type C	Train data set	Test data set	0.253698	0.800412	True
type D	Train data set	Test data set	1.448122	0.150043	True
type E	Train data set	Test data set	0.248305	0.842878	True

## V. 결론 및 향후 연구과제

본 연구는 교통혼잡 판단에 필요한 주요 교통 변수만을 추출하기 위해, 일반적으로 VDS를 통해 수집되는 교통 데이터를 분석하여 교통상황을 분류하였다. 이를 위해 요인분석 및 군집분석을 수행하였으며, 주요 교통변수를 선택하고 각 변수에 대해 군집을 분류하였다. 더불어 혼잡지표인 TTI를 기준으로 분산분석을 통해, 최종적으로 교통상황 유형을 분류하고 각 유형들이 어떠한 영향을 미치는 지 분석하였다.

요인분석 및 군집분석, 그리고 분산분석을 통해 교통상황을 5개의 유형으로 분류한 최종 결과는 다음과 같다. 최대대기행렬길이가 138.015m 미만이고 점유율이 0.499 미만인 경우 유형 A로, 0.499 이상인 경우 유형 B로 구분하였다. 그리고 최대대기행렬길이가 276.039m 미만이고 점유율이 0.499 미만인 경우 유형 C로, 0.499 이상인 경우 유형 D로 구분하였다. 특히 최대대기행렬길이가 276.039m 이상이고 점유율이 0.499 미만인 경우도 유형 D에 포함된다. 마지막으로 최대대기행렬길이가 276.039m 이상이고 점유율이 0.499 이상인 경우 유형 E로 구분하였다. 이렇게 분류한 본 연구의 방법론에 대해 미리 분리해 놓은 test 데이터 셋을 통해, TTI를 기준으로 독립표본 t 검정을 활용하여 검증을 수행하였다. 그 결과 모든 유형에서 train 데이터 셋과 test 데이터 셋의 TTI 값이 동일함을 보였고, 본 연구의 방법론 또한 적절하다고 판단하였다.

HCM과 KHCM의 서비스 수준 평가 시 활용하는 효과적도인 제어지체는 차로수, 차로폭, 경사, 도류화 및 버스정류장 위치와 같은 도로조건과, 각 접근로의 교통량, 차종, 속도 등의 교통조건, 신호, 회전 및 주차제한 등과 같은 통제조건을 모두 고려하여 산출해야 한다. 그러나 본 연구의 방법론에 따라, 기존의 많은 조건들을 고려할 필요 없이 교통상황에 영향을 미치는 주 요인변수만을 활용하여 혼잡을 판단할 수 있다. 따라서 추후 여러 종류의 교통 데이터를 고려하여 혼잡지표 개발 시, 본 연구의 방법론을 통해 효율적인 교통혼잡 관리가 가능할 것으로 기대된다.

그러나 본 연구의 한계 또한 존재한다. 본 연구는 대치역·도곡역 부근 8개 교차로에 대해서만 수행되었다. 도로 조건을 포함한 TTI를 기준으로 교통상황의 주 요인변수를 추출하였으나, 도로 기하구조와 같은 교통소통정보 외 요소의 검토가 필요하다. 따라서 다양한 종류의 도로를 대상으로 검토하여, 소통정보 외의 요인

과 교통상황의 관계를 분석할 필요가 있다. 또한 C-ITS 및 자율주행기술을 통해 더욱 많은 종류의 데이터에 대해 분석한다면, 보다 더 명확하게 교통상황 유형을 분류할 수 있고 주 요인변수를 선택할 수 있을 것으로 사료된다.

## ACKNOWLEDGEMENTS

이 논문은 2022년도 정부(경찰청)의 재원으로 과학치안진흥센터의 지원을 받아 수행된 연구임(092021C29S01000, 네트워크 제어를 위한 교통정체 및 혼잡 운영관리 기술 개발). 본 논문은 2022년 한국ITS 학회 춘계학술대회에 게재되었던 논문을 수정·보완하여 작성하였습니다.

## REFERENCES

- Cho, E. H.(2017), *A study on estimation and application of travel time index for comparison and analysis in Seoul*, Master's Thesis, University of Seoul, <https://dl.nanet.go.kr/>
- Cho, Y. J. and Kim, Y. H.(2007), "A strategy through segmentation using factor and cluster analysis: Focusing on corporations having a special status", *The Korean Journal of Applied Statistics*, vol. 20, no. 1, pp.23-38.
- Choi, C. H. and You, Y. Y.(2017), "The study on the comparative analysis of EFA and CFA", *Journal of Digital Convergence*, vol. 15, no. 10, pp.103-111.
- Han, Y. H. and Kim, Y. C.(2017), "A study of measuring traffic congestion for urban network using average link travel time based on DTG big data", *The Journal of The Korea Institute of Intelligent Transport System*, vol. 16, no. 5, pp.72-84.
- Hong, J. H., Oh, M. J., Cho, Y. B., Lee, K. H. and Cho, W. S.(2020), "A dimensional reduction method in cluster analysis for multidimensional data: Principal component analysis and factor analysis comparison", *The Korea Journal of BigData*, vol. 5, no. 2, pp.135-143.
- Kim, S. G., Yun, I. S., Oh, Y. T., Ahn, H. K., Kwon, K. A. and Hong, D. P.(2014), "Feasibility evaluation of lane grouping methods for signalized intersection performance index analysis in KHCM", *The Journal of the Korea Institute of Intelligent Transport System*, vol. 10, no. 1, pp.109-126.
- Kong, X., Yang, J. W. and Yang, Z. G.(2015), "Measuring traffic congestion with taxi GPS data and travel time index", *COTA International Conference of Transportation Professionals (CICTP) 2015*, pp.3751-3762.
- Korea Road Traffic Authority(KoROAD)(2014), *A study on the establishment data quality diagnosis and evaluation system for traffic information management system*, pp.14-16.
- Lee, I. M., Min, J. H., Kim, K. T. and Kho, S. Y.(2020), "Generating travel patterns of public transportation users using a K-means clustering based on smart card data", *Journal of the Korean Society for Railway*, vol. 23, no. 3, pp.204-215.
- Lee, S. J., Kim, T. Y., Ko, H. G. and Bok, K. C.(2008), "The evaluation of existing congestion

- indices' applicability for development of traffic condition index”, *Korean Society of Road Engineers*, vol. 10, no. 3, pp.119-128.
- Sim, S. W. and Chou, K. C.(2009), “Classification of freeway traffic condition by the impacts of road weather factors”, *Journal of the Korean Society of Civil Engineers*, vol. 29, no. 6, pp.685-691.
- The Korea Transport Institute(2021a), *Traffic policy evaluation index survey project brief 2020*, pp.6-8.
- The Korea Transport Institute(2021b), *2021 National transportation cost forecast-congestion cost, logistics cost, accident cost*, pp.2-6.
- The Korea Transport Institute(2021c), *A basic study on the revision of the Highway Capacity Manual*, pp.5-15.
- Yuan, Q., Li, X., Wang, C., Li, Y. and Gao, Y.(2017), “Cluster and factor analysis on data of fatal traffic crashes in China”, *2017 4th International Conference on Transportation Information and Safety (ICTIS)*, pp.211-224.