

동적 타임 워핑 거리 기반 비 계층적 군집분석을 활용한 TOD 시간분할 최적화

Optimize TOD Time-Division with Dynamic Time Warping Distance-based Non-Hierarchical Cluster Analysis

황재연* · 박민주** · 김영호*** · 강우진****

* 주저자 : 한남대학교 빅데이터 응용학과 연구원
 ** 교신저자 : 한남대학교 빅데이터 응용학과 조교수
 *** 공저자 : 한국교통연구원 광역교통연구본부 본부장, 연구위원장
 **** 공저자 : 한국교통연구원 광역교통연구본부 연구원

Jae-Yeon Hwang* · Minju Park** · Yongho Kim*** · Woojin Kang****

* Researcher, Univ. of Hannam
 ** Assistant Professor, Dept. of Big Data Application, Univ. of Hannam
 *** Chief Director, Research Fellow, Dept. of Metropolitan Transport, Korea Transport Institute
 **** Researcher, Dept. of Metropolitan Transport, Korea Transport Institute

† Corresponding author : Minju Park, parkmj@hnu.kr

Vol.20 No.5(2021)

October, 2021
pp.113~129

pISSN 1738-0774
eISSN 2384-1729
<https://doi.org/10.12815/kits.2021.20.5.113>

Received 14 July 2021
Revised 28 July 2021
Accepted 11 October 2021

© 2021. The Korea Institute of
Intelligent Transport Systems. All
rights reserved.

요약

최근 수도권 중심의 생활권역 확장과 대도시로의 인구 집중으로 도시 내의 교통 혼잡이 지속적으로 증가하고 있다. 도심지의 땅값 상승과 한정된 부지로 인해 새로운 도로 건설은 불가능하게 되었고, 데이터 기반의 효율적인 도로 운영의 중요성이 점점 부각되고 있다. 효율적인 도로 운영을 위해서는 교통상황의 변화에 따른 적절한 TOD 시간분할과 TOD 시간분할을 통한 최적의 신호 운영 방안이 필수적이다. 본 연구에서는 최적의 TOD 시간 분할을 위해 연속된 교차로에서 수집된 교통량과 속도 데이터에 시계열 데이터의 군집 분석을 위한 동적 타임 워핑 모델을 적용하였다. 시간 분할을 위해 활용된 데이터별 군집의 특성을 분석하여 최적의 신호 운영 시나리오를 구성하기 위한 시간 분할 방법론을 제안하고자 한다.

핵심어 : 교통신호제어, 교통 시나리오, TOD, 군집 분석, 동적 타임 워핑

ABSTRACT

Recently, traffic congestion in the city is continuously increasing due to the expansion of the living area centered in the metropolitan area and the concentration of population in large cities. New road construction has become impossible due to the increase in land prices in downtown areas and limited sites, and the importance of efficient data-based road operation is increasingly emerging. For efficient road operation, it is essential to classify appropriate scenarios according to changes in traffic conditions and to operate optimal signals for each scenario. In this study, the Dynamic Time Warping model for cluster analysis of time series data was applied to traffic volume and speed data collected at continuous intersections for optimal scenario classification. We propose a methodology for composing an optimal signal operation scenario by analyzing the characteristics of the scenarios for each data used for classification.

Key words : Traffic signal control, Signal scenario, TOD, Cluster analysis, Dynamic Time Warping

I. 서 론

수도권 중심의 생활권역 확장과 대도시로의 인구 집중은 도시 내 교통혼잡 비용의 지속적인 증가를 이끌어왔다. 2015년 기준 전체 교통 혼잡비용은 33조 4천억 원으로 우리나라 전체 GDP의 2.16% 수준이며, 시군 도로를 포함한 방식으로 산출한 2017년 전국의 교통 혼잡비용은 총 59조 6천억 원으로 전체 GDP의 3.4% 규모에 달하는 것으로 추정된다.

이러한 문제를 해결하기 위해 정부에서는 새로운 도로 신설을 통해 교통량을 분산시키고자 하였지만, 공급되는 도로용량 대비 높은 교통 수요의 증가로 인해 정체는 점점 악화되고 있다. 또한, 도시부에서는 높은 토지 보상비용의 증가와 한정된 부지로 인해 무한정 도로시설을 확장할 수 없게 되었고, 한정된 도로 용량 속에서 효율적인 도로 운영의 중요성이 점점 부각되고 있다.

90년대 이후 시작된 지능형 교통체계(ITS) 사업을 통해 대다수의 교통시설 운영주체(지자체, 도로공사, 국토관리청 등)는 도로 위에서 다양한 센서로 부터 데이터를 수집하게 되었으며, 효율적인 도로 운영 방안을 마련하기 위해 수집된 데이터를 활용해왔다. 하지만, 수집된 데이터의 활용은 실시간 도로 상황 정보(속도, 통행시간)를 알려주는 1차원적인 활용에 그치고 있다.

도심지 교통 혼잡문제를 해결하기 위해서는 수집된 데이터를 기반으로 교통 상황의 변화에 맞추어 교통 운영조건을 최적화시켜야 한다. 현재 서울특별시를 비롯한 대부분의 지방자치단체에서는 실시간 신호 운영 시스템을 운영하지 않고 TOD(Time Of Day)제어를 활용하고 있다. 그리고 하나의 교차로에 신호를 따로 적용하는 방식이 아닌 하나의 신호그룹(Sub Area)에 동일한 신호 운영을 적용시켜 작동하는 방식으로 운영하고 있다.

TOD 제어는 비슷한 교통 패턴 또는 교통량을 가진 시간대를 그룹으로 분류해서 같은 신호 운영 방안을 부여해 교통 정체를 완화시키고 교차로를 효율적으로 운영하는 신호제어 방법으로 국내에서 가장 많이 활용되고 있다. 실시간 신호 운영 시스템은 AI를 기반으로 실시간 교통량에 따라 신호를 부여하는 방식으로 개발 중에 있는데, 신호 운영 시간이 최소전이시간을 고려하지 않고, 실시간 교통량에 따라 신호가 변화하다 보니 잦은 신호의 변경이 발생할 수 있다. 이는 교차로의 신호그룹을 이용하는 많은 운전자들의 사고 위험성을 높이고, 혼란을 야기할 가능성이 많다. TOD 제어를 활용하며, 실시간 신호 운영 시스템을 개발해 나가는 현 시점에서 신호 제어를 위한 TOD 시간계획을 각 시간대의 특성을 고려해 분류해야 하고, 잘 분류된 TOD 시간계획으로 각 교통 시나리오별 최적의 신호운영을 활용할 필요가 있다.

하지만, 최적의 TOD 운영 시나리오를 선정하는 과정에서 대부분 운영자의 주관적인 요소를 기반으로 첨부, 비첨두 시간에 각기 다른 운영방안을 적용해왔다. 효율적인 신호 운영을 위해서는 계속적으로 변하는 도로 상황에 적합한 시간 계획을 시나리오로 분류한 뒤 각 시나리오별 최적 운영방안이 적용되어야 할 것이다.

본 논문에서 기존의 주관적으로 TOD 시간을 분할하는 방법이 아닌 기계 학습을 활용해 데이터 간 비슷한 특성을 가진 데이터를 Group화시키는 군집을 생성하고, 군집 개수를 최적화시켜 TOD 시간을 분할하는 방법을 제안한다. 특히 도로 구간의 교통량과 평균 속력을 하나의 시계열 데이터로 활용해 시계열 군집 분석으로 진행되는 방법을 중점적으로 제안하고자 한다. 본 논문의 구성은 다음과 같다. 2장에서는 선행 연구에 대한 고찰과 본 논문에서 활용한 연구 방법들에 대한 이론적인 개념을 소개하고, 3장에서는 활용된 데이터가 어떤 데이터인지, 데이터를 어떻게 수집하고 활용했는지에 대한 내용을 소개한다. 4장에서는 데이터를 통해 얻은 분석 결과를 평균 속도, 교통량과 정규화 된(평균 속도+교통량) 3가지 기준으로 분석하고, Heatmap 시각화를 통해 요일별/시간별 시간 계획 분류 결과를 바탕으로 적합한 분류 방안을 소개한다. 마지막으로 5장에서 결론과 향후 연구에 대해 정리하였다.

II. 선행 연구 및 연구 방법

1. 선행 연구

오랜 시간 동안 신호 제어에 활용하기 위해 TOD 시간을 분할하는 다양한 방법론들이 제시되었다. Hua and Faghri(2005)는 교통을 통제하는 적절한 시간 간격을 결정하는 방법 중 하나의 접근법인 동적 프로그래밍(Dynamic Programming)과 인공 신경망(Artificial Neural Networks)을 비교하며 교통 패턴 분류 문제에 대해 성공적인 결과를 보여준 Adaptive Resonance Theory(ART)라는 인공 신경망 구조를 개발했다. 이는 시간 분할을 작은 단위의 시간에도 적용해 교통량의 차이가 적을 때에도 분리된 그룹을 제시하는 문제가 있어 분석 시간 대가 길어질 경우 활용하기 어려울 수 있지만, 기존의 관습적으로 사용하는 방법들에 비해 더 효율적인 방법을 Deep Learning과 결합해 제시했다.

Jia et al.(2006)은 Genetic algorithm을 기반으로 한 계층적 군집분석을 TOD 시간 분할에 활용했다. 이는 군집화 된 데이터의 중복된 정보들을 효율적으로 줄이는 효과가 있었고, 조정 가능한 매개 변수가 많아 범위를 합리적으로 설정할 수 있다. 하지만 이러한 매개변수가 너무 많은 경우 이상적이지 않은 과도한 설정으로 인해 군집 분석에 영향을 미칠 수 있다.

Jeong et al.(2008)은 독립교차로를 기준으로 비 계층적 군집 분석 중 대표적으로 사용되는 K-means 군집분석을 활용해 TOD의 시간을 분할하는 방법을 제안했다. 최소전이시간을 통한 VISSIM 모형을 사용해 군집 분석을 조정했고, 운영적 측면에서 활용할 수 있는 TOD 시간 분할을 제안해 교차로의 지체를 감소시켰으나, 군집화의 가장 기초가 되는 군집의 개수에 대한 최적화를 주관적 직관에 의존했다.

Lee et al.(2011)은 TOD 시간을 분할하는 방법으로 Genetic Algorithm을 제안해 Greedy Search 방법보다 더 나은 성과를 나타냈다. 또한 전환비용을 고려해 운영 측면에서 실제적으로 활용할 수 있는 TOD 시간 계획을 구분하는 방법을 제시했다.

Lee et al.(2011)은 K-means 군집 분석에 VPLUSKO라는 새로운 지표를 활용해 교통량으로 군집을 분석하는데 모호한 부분들을 개선했다. 그리고 현장적용과 결과를 통해 교차로의 교통 정체를 완화시키는 성과를 얻었으나, 군집의 개수에 대한 최적화를 위한 추가적인 연구가 필요하다고 언급했다.

일반적으로 최적화된 TOD의 시간을 분할하기 위한 군집 분석은 계층적 군집 분석보다 군집의 개수를 지정할 수 있는 비 계층적 군집 분석과 Genetic Algorithm과 같은 Machine Learning을 활용하였다. 하지만 군집의 개수를 최적화하는 방법에 대한 새로운 방법이 필요하다.

2. 최적화된 TOD 시나리오 추출을 위해 적용된 방법론

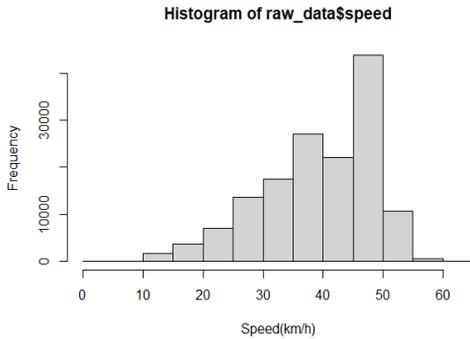
1) 정규화를 통한 데이터 변형

데이터는 각각의 변수에 다양한 단위를 가지고 있다. 교통 데이터에서 예시로 보면 평균 속도와 교통량은 같은 숫자를 활용하나 그에 대한 의미는 다르다. 평균 속도는 km/h, 교통량은 해당 도로를 지나간 차량의 미하는데, 평균 속도와 교통량을 비교했을 때 <Fig. 1>과 <Fig. 2>에서 보이는 것과 같이 평균 속도보다 교통량이 훨씬 값이 크고 편차도 심하다.

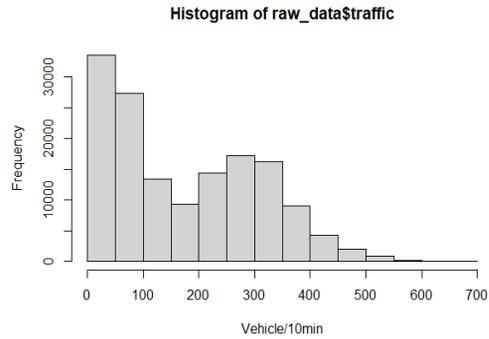
이러한 평균 속도와 교통량을 동일하게 비교하기 위해서 단위를 없애줘야 한다. 두 변수가 단위에 영향을 받지 않도록 변형시키기 위해 데이터 정규화(Data Scaling)를 사용한다. 데이터 정규화에는 다양한 방법들이 있지만 본 논문에서는 교통량이 각 도로에 따라 편차가 심해 이상치가 존재할 수 있다고 판단해 이상치 값

에 영향을 덜 받는 Z-점수 정규화(Z-Score Normalization)를 사용한다. Z-점수 정규화는 데이터가 평균 0, 표준편차가 1이 되도록 변경하는 방법으로 기존 데이터 값에서 전체 평균을 빼주고, 표준 편차로 나누는 것을 의미하며, 수식으로 표현하면 (1)과 같다.(Jang, 2020)

$$Z = \frac{x_i - \bar{x}}{\sigma} \dots\dots\dots (1)$$



<Fig. 1> Histogram of speed data



<Fig. 2> Histogram of traffic data

2) 기계 학습을 활용한 교통 데이터 분석

교통 데이터는 같은 시간일지라도 날짜별로 일반적이지 않은 복합적인 요소(도로 공사, 교통사고, 날씨)들에 의해 평균 속도나, 교통량이 매번 다르다. 시간별로 group을 만들기 위해서는 각각의 시간에 데이터들이 가지는 특성을 파악해 각 특성에 맞는 값들을 연결시켜주는 작업들이 필요하나, 현대 사회에서 급증하는 방대한 양의 데이터를 보고 해당 시간들이 가지는 특징을 일일이 파악하기에는 어려움이 있다.

이러한 어려움을 해결하기 위해 오랜 시간 발전해온 기계 학습은 group을 구분하기 위해 군집분석(Clustering)을 활용하고 있다. 군집 분석은 어떠한 정답과 정보도 없이 순전한 데이터만을 가지고 정보를 찾아주는 지식의 발전에 주로 활용되고 있다. 군집 분석은 대량의 데이터를 훨씬 적은 수의 군집으로 묶어 복잡성은 줄여주고 데이터 내에 존재하는 패턴에 대해서는 통찰력을 제공하는 측면에서 이상 패턴과 불필요한 데이터들을 구분시켜 분석 데이터를 질적으로 향상시키는 역할을 한다. 이러한 군집 분석을 통해 복합적인 요인들을 축소시키고 비슷한 특징을 가지는 데이터들을 군집으로 분류한다.(Lantz, 2020)

3) DTW(Dynamic Time Warping) 모델을 활용한 군집 분석

DTW 모델은 시계열 데이터 간의 유사도를 구하기 위해 활용되는 군집 분석 알고리즘이다. 군집 분석은 주로 거리를 기반으로 Euclidean Distance를 활용해 군집 간의 거리를 구하는데, Euclidean Distance는 데이터 간의 길이가 맞지 않거나, 시계열 데이터의 경우 적합하지 않다.

DTW는 Euclidean Distance가 시계열 데이터에 활용하지 못하는 부분을 개선하고자 시계열 데이터에서 짝을 이루는 군집 간의 거리를 최소화 시키는 방법으로, 시계열 데이터에 대해 일부를 확장 또는 압축시켜 서로 길이가 다른 두 시계열 간의 거리도 최적으로 구할 수 있다. DTW 알고리즘에서 시계열 데이터 간 거리

를 구하는 식은 아래의 (2)와 같이 정의된다.(Lee et al., 2019)

$$d_{DTW}(X_{a_i}, Y_{b_j}) = \min_{r \in m} \left(\sum_{i=1, \dots, m} |X_{a_i} - Y_{b_j}| \right) \dots \dots \dots (2)$$

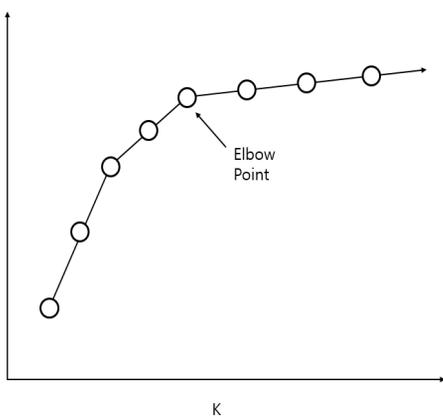
4) 군집 개수 k의 최적화

군집 분석은 사전에 군집의 개수를 지정하는지의 여부에 따라 계층적 군집 분석(Hierarchical Clustering)과 비 계층적 군집 분석(Non-Hierarchical Clustering)으로 구분하는데, 비 계층적 군집 분석의 경우 임의의 군집 개수인 k를 사전에 지정해줘야 한다. 군집의 개수는 몇 개를 지정하나에 따라 결과가 많이 달라지기 때문에 군집화를 이뤄나가는 과정에서도 가장 중요하게 작용한다. 군집의 개수는 주관적인 요소가 많이 포함되지만, 사전 지식이 없는 경우 군집의 개수 k를 데이터 개수인 n을 2로 나눈 값의 제곱근으로 사용하는 것이 일반적이다.(Lantz, 2020)

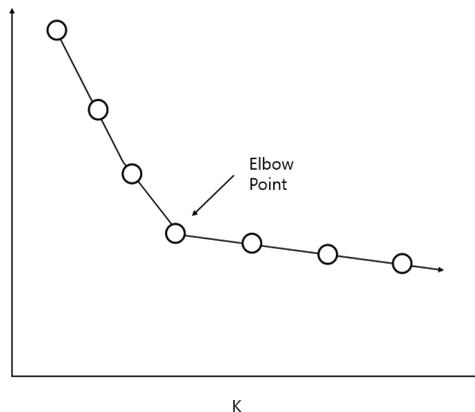
$$k = \sqrt{n/2} \dots \dots \dots (3)$$

하지만 (3)과 같이 일반적인 군집 개수를 지정하는 경우 대량의 데이터에 적합하지 않다. 대량의 데이터와 더불어 군집의 개수도 무분별하게 커지기 때문이다. 그러므로 데이터의 양에 구애받지 않고 군집의 개수를 최적화시키기 위해 Elbow Method를 본 논문에서는 사용했다.

Elbow Method는 군집의 개수 k의 값이 변화됨에 따라 군집 내의 동질성은 <Fig. 3>과 같이 군집이 늘어날수록 계속해서 증가, 이질성은 <Fig. 4>와 같이 군집이 늘어날수록 계속해서 감소할 것으로 가정하고, 이 동질성과 이질성이 어떻게 변화되는지를 측정한다. Elbow Method 기법은 동질성은 최대, 이질성은 최소인 k를 찾는 것이 아닌, 특정 군집을 넘어가면 결과가 크게 안 좋아지는 k를 찾는 것이다. 이러한 k를 그래프 상으로 나타냈을 때 팔꿈치 모습과 비슷하다 하여 Elbow Method라고 하고, 이 k 값을 Elbow Point라고 하며 최적의 군집 개수 k라고 한다.(Lantz, 2020)



<Fig. 3> Elbow point on homogeneity in cluster



<Fig. 4> Elbow point for out-of-cluster heterogeneity

Ⅲ. 데이터 전처리

1. 데이터 수집

본 논문에서는 데이터를 얻기 위해 특정 구간의 교차로에 설치된 RSE(Road Side Equipment)를 활용했다. RSE는 교차로에 설치된 장비이면서 DSRC 기술이 탑재되어 단거리 통신이 가능한 장비로, Hi-pass를 사용하는 차량에 장착된 OBU(On Board Unit)가 해당 교차로를 지나갈 때 RSE와 통신하고 또 다른 RSE가 설치된 지역을 지나갈 때의 통신 시간을 기록해 두 지역 간 통행시간과 속도 정보를 수집한다.

<Fig. 5>와 같이 대전 대덕대로에서 하나의 신호그룹으로 분류된 지역(대덕대교 네거리, 정부청사역네거리, 큰 마을 네거리)을 연구 대상으로 선정했다. 11개의 신호 교차로에 설치된 RSE에 포함되는 구간은 44개 존재하고, 동서남북에 해당하는 4개의 방향에 각각 11개씩 존재한다. 남과 북의 방향은 대덕대로 주 방향을 의미하고, 동서 방향은 부 방향을 의미한다.

RSE에서 수집된 교통량 데이터는 OBU가 설치된 차량에서만 수집되기 때문에, 11개 교차로에 카메라를 설치하여 하루 동안의 방향별 교통량을 전수 조사하였다. 이 정보를 기준으로 대덕대로의 마이크로시물레이션(AIMSUN)을 구축하였으며, 대전시에서 제공받은 링크별 속도 값과 VDS(Vehicle Detection System)에서 수집한 교통량 정보를 기반으로, 시물레이션을 통해 약 한 달 동안(19년 8월 18일부터 19년 9월 21일)의 10분 단위 평균속도 및 교통량, 혼잡도를 기록한 데이터를 수집하였다.

2. 데이터 변환

<Table 1>은 수집된 데이터의 종류와 시간적, 공간적 범위를 나타낸다. 해당 구간의 시간 계획을 분류하기 위해서는 주어진 데이터를 기반으로 해당 구간의 시간별 속성 정보가 필요하다. 이를 위해 <Table 3>과 같이 모든 구간에 시간별 평균 속도와 교통량을 각각 하나의 벡터로 변환하였다.

<Table 1> Raw data Information

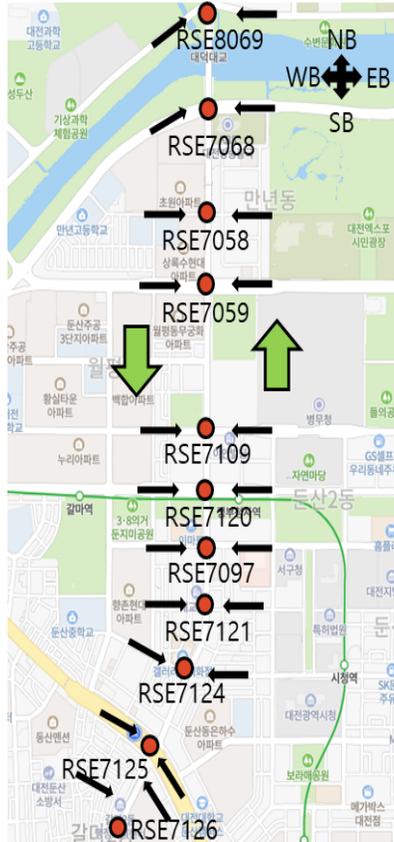
Variable	Configuration
Week	1 ~ 5
Date	19 .8 .18 ~ 19 .9 .21
Day	Monday ~ Sunday
Section	1 ~ 44 (1 ~ 22 : Main boulevard north-south 1 ~ 11 : North 12~22 : South 23 ~ 44 : Small boulevard east-west 23 ~ 33 : West, 34 ~ 44 : East)
Route	SB, NB, EB, WB
Time	From 6:00 to 21:50 on a 10-minute units
Speed	Average speed
Traffic	Traffic
Density	Traffic density

<Table 2> Example of data before conversion

Date	Day	Section	Time	speed	traffic
8.18	Sun	1	6:00	37.47	95
8.18	Sun	1	6:10	36.29	106
8.18	Sun	1	6:20	30.9	98
8.18	Sun	1	6:30	32.45	125

<Table 3> Example of data after conversion

Date	Day	Time	1	2	...
8.18	Sun	6:00	37.47	37.85	...
8.18	Sun	6:10	36.29	38.68	...
8.18	Sun	6:20	30.9	36.14	...
8.18	Sun	6:30	32.45	36.93	...



<Fig. 5> Daedeokdaero DSRC (RSE) location, direction, and major intersection

IV. 분석결과

<Fig. 6>에서와 같이 TOD 시나리오의 적절한 분할을 위해 수집된 모든 링크의 속도와 교통량 정보를 10분 단위의 시계열 데이터로 변환하였다. 최적의 시계열 군집 분석을 위해 아래의 3가지 기준을 바탕으로 분류하고자 한다. c)의 경우인 평균 속도와 교통량은 서로 단위가 다르다. DTW 모델은 Input Data를 list형태로 받기 때문에 평균 속도 값과 교통량 값을 단위로 구분하지 않고 같은 정수로 인식한다. 그래서 데이터 속도와 교통량에 대해 각각 Z-점수 정규화 방법으로 값들을 변경시키고 단위를 통일시켜 나머지 경우와 비교하였다.

- a) 평균 속도
- b) 교통량
- c) 정규화된(평균 속도 + 교통량)

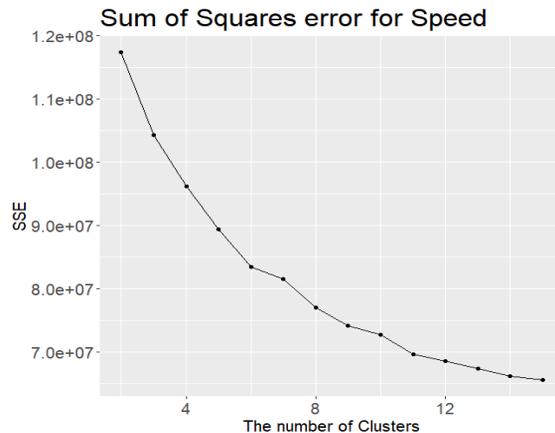
	NB			SB			WB			EB			
Speed	37.47	37.85	...	39.94	46.72	...	24.6	49.59	...	53.8	39.4	...	06:00
	36.29	38.68	...	40.28	50.66	...	26.19	50.96	...	53.19	29.93	...	06:10
	30.9	36.14	...	46.93	54.01	...	25.98	49.16	...	51.34	38.92	...	06:20
Traffic	95	133	...	37	33	...	23	50	...	18	18	...	06:00
	106	125	...	32	27	...	20	48	...	14	23	...	06:20
	98	116	...	39	37	...	26	45	...	26	21	...	06:30

<Fig. 6> Input data example

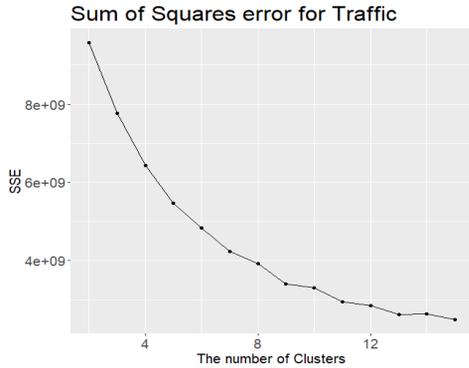
1. Elbow method를 활용한 최적의 군집 개수 k

본 논문은 도로에서 수집되는 데이터를 활용하여 효율적으로 신호를 운영하기 위한 시간을 분할하고자 한다. 이를 위해 데이터를 일정한 간격으로 배열시킨 시계열 데이터로 변환하고, 변환한 데이터를 활용해 비 계층적 군집 분석을 진행하였다. 비 계층적 군집 분석의 최적화된 군집 개수를 구하기 위해 오차 제곱 합(Sum of squares error)을 활용한 Elbow method를 사용한다.

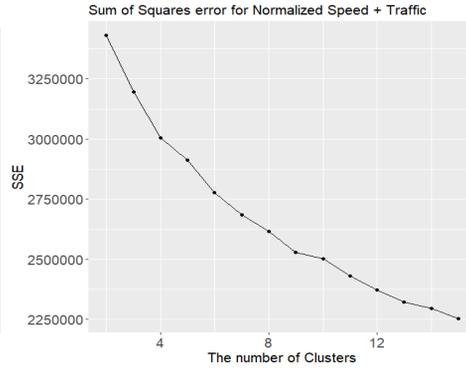
<Fig. 7>은 평균 속도 데이터를 기준으로 분석된 결과 그래프이다. 비 계층적 군집 분석이 무작위하게 집단을 분류해주는 과정에서 매번 동일한 군집으로 분류해주지 않기 때문에 이를 보완하고자 100회 반복 실행하였다. x축은 군집의 개수, y축은 오차 제곱 합(SSE)으로, 군집의 개수에 대한 오차 제곱 합은 반복 시 생성되는 각각의 군집에 대해 오차 제곱 합의 전체 평균을 나타낸 값이다. 결과적으로 오차 제곱 합의 이질성에 대한 평균값을 살펴볼 때, <Fig. 7>과 같은 결과가 나왔다. <Fig. 7>에서 오차 제곱 합이 줄어드는 폭이 서서히 감소되는 지점은 k가 8을 넘어가는 부분으로, Elbow point는 중간에 위치한 8로 볼 수 있고, 이를 최적의 군집 개수로 본다.



<Fig. 7> Sum of squares error for speed



<Fig. 8> Sum of squares error for traffic



<Fig. 9> Sum of squares error for normalized speed, traffic

<Fig. 8>은 교통량에 대한 군집 개수에 따른 오차 제곱 합을 나타내며, <Fig. 9>는 정규화된 평균 속도와 교통량에 분석된 결과를 보여준다. 시계열 데이터 간의 차이가 큰 교통량의 오차 제곱 합 편차가 가장 크게 나타나며, 정규화된 평균 속도와 교통량 데이터의 오차 제곱 합 편차는 상대적으로 작게 나타났다. Elbow method를 적용했을 때, <Fig. 8>과 <Fig. 9>에서도 최적의 군집 개수는 오차 제곱 합이 줄어드는 폭이 감소되는 지점인 8개로 본다.

2. 분석 결과 시각화

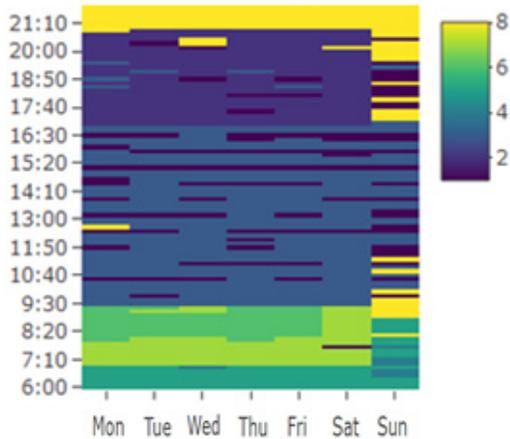
1) Heatmap 시각화

군집 분석의 결과를 비교하기 위해 각 요일별 시간에 해당하는 군집의 색을 동일하게 구분해주는 Heatmap 그래프를 활용했다. 그래프는 전체 기간 중 마지막 주에 해당하는 데이터를 앞서와 마찬가지로 3가지 기준을 관점으로 비교하였다.

<Fig. 10 ~ 12>에 해당하는 3가지 경우 모두 침투 시간의 전체 특징이 반영되어 군집을 분류한 것으로 볼 수 있다. 평균 속도가 반영된 <Fig. 10>과 <Fig. 12>는 침투 시간에 대한 분류가 잘되어 있으나, 비침투 시간(9:30부터 16:00까지)에는 군집의 반복적인 변화가 심한 것으로 나타난다. 일반적으로 침투 시간으로 분류되는 평일 아침 출근 시간인 7시부터 9시, 퇴근 시간인 오후 5시부터 오후 8시에 대해서는 군집 변화가 뚜렷하게 나타난다.

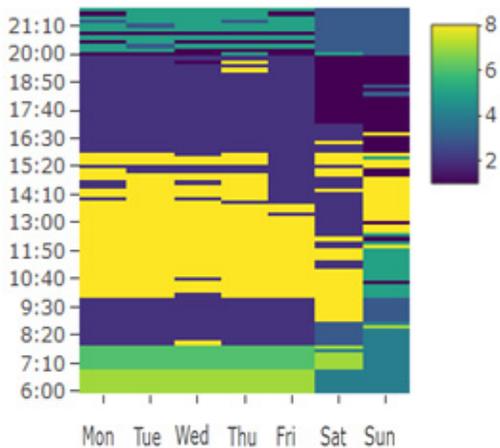
교통량 기반으로 분류한 <Fig. 11>은 다른 경우보다 평일과 주말에서 군집 형성의 차이가 있음을 보여준다. 또한, 교통량이 밀집되는 침투 시간의 군집 범위는 더 넓게 형성되며, 전체적인 군집이 시간대 별로 통일성 있고 뚜렷하게 나타난다. 하지만, 침투 시간을 같은 군집으로 분류하여 침투 시간에 방향별 교통량의 차이를 제대로 반영하지 못하고 있다.

분석에 활용한 10분 단위의 평균 속도와 교통량 데이터는 순간적인 교통 상황 변화를 반영할 수 있는 장점도 있지만, 작은 변화에도 민감하게 반응할 수 있다. 이는 교통의 상황 변화가 일시적으로 커져 동 시간대의 군집의 변화가 많이 발생할 수 있고, 잦은 군집의 변화는 시간 경계를 결정하기 어렵게 만들 뿐 아니라 잦은 신호시간의 변경을 일으켜 운전자들에게 혼란을 야기할 가능성이 높다.(Lee et al., 2011) 이러한 특성을 보완해 10분 단위의 데이터를 30분 단위의 데이터로 가공하여 위와 동일한 분석을 실행하였다.

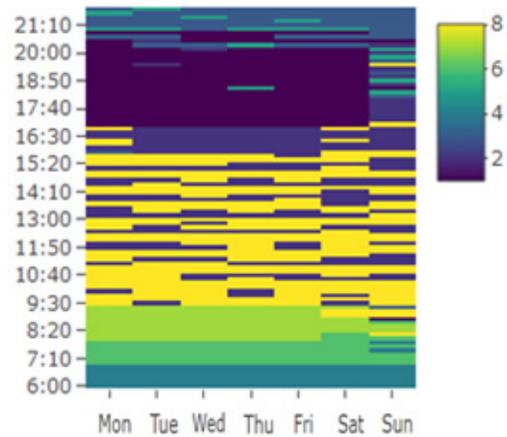


<Fig. 10> Heatmap graph of speed data clustering results in 10-minute units

- a) 평균 속도
- b) 교통량
- c) 정규화 된(평균 속도 + 교통량)



<Fig. 11> Heatmap graph of traffic data clustering results in 10-minute units



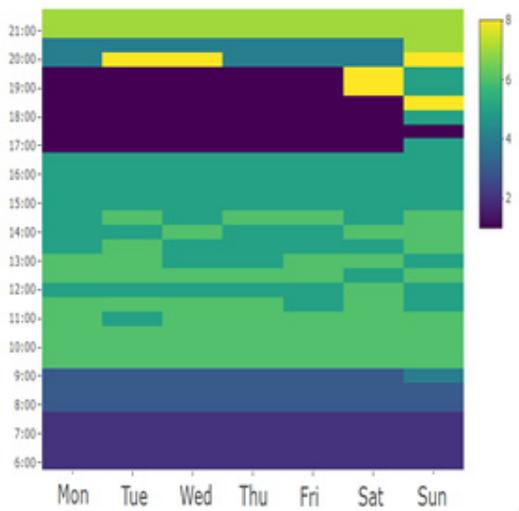
<Fig. 12> Heatmap graph of normalized speed and traffic data clustering results in 10-minute units

2) Time gap, 군집의 유동성 비교

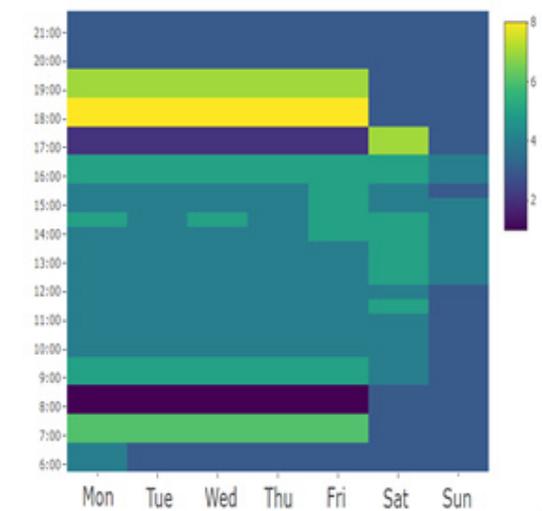
2-1) Time gap

데이터의 시간적 범위를 30분 단위로 늘려서 10분 단위의 데이터와 군집 분류 결과를 비교하고자 한다. <Fig. 13>, <Fig. 14>는 30분 단위의 데이터를 앞서 기술한 3가지 기준 중에서 군집을 잘 형성한 두 가지(교

통량과 정규화한(속도+교통량)경우에 대해 군집 분석한 결과를 Heatmap으로 나타낸 결과이다. <Fig. 13>은 30분 단위로 정규화한 평균 속도와 교통량으로 군집 분석을 한 결과로 10분 단위로 살펴본 <Fig. 12>와 비교했을 때 오후 시간대의 군집이 지속적으로 변화하는 모습이 눈에 띄게 줄었다. 이는 10분의 짧은 시간에서의 교통량과 평균 속도가 지속적인 변화가 있었지만, 30분 단위로 늘렸을 때 지속적인 변화가 어느 정도 감소되었음을 나타낸다. <Fig. 14>는 30분 단위로 교통량 데이터만을 군집 분석한 결과로 평일과 주말간의 구분이 뚜렷하게 나타나고, 평일의 침투 시간도 각각 다른 군집으로 분류한 것으로 나타난다. 또 다른 특징은 출근 시간대(07:00~09:30)에도 시간대별로 3개의 다른 군집을 제시해준다. 이는 교통량의 증가, 최고(Peak), 감소 단계로 3개의 군집으로 분류됨을 확인할 수 있다. 퇴근시간대 역시 시간대별로 3개의 다른 군집을 제시해준다.



<Fig. 13> Heatmap graph of normalized speed and traffic data clustering results in 30-minute units

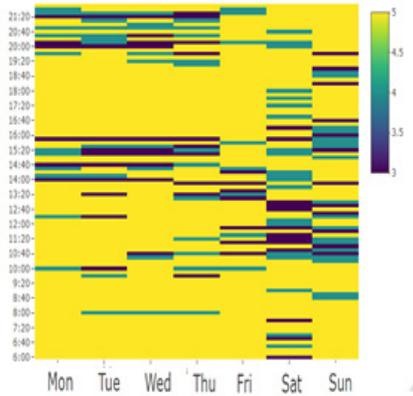


<Fig. 14> Heatmap graph of traffic data clustering results in 30-minute units

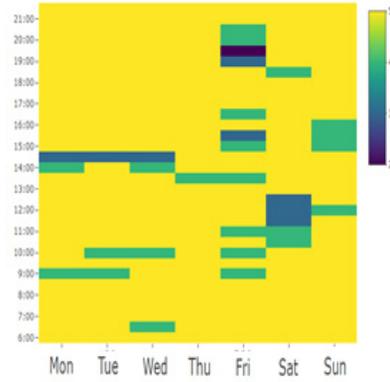
2-2) 군집의 유동성

앞서 전체 기간 중 마지막 주에 대한 군집 분석의 결과를 Heatmap으로 나타냈다. 전체 기간을 두고 Heatmap으로 나타내기에는 각 주마다 복합적인 요인들이 존재할 수 있기 때문에 군집이 항상 고정적이지 않고, 데이터의 기간이 길지 않기 때문에 각각의 주에 영향을 받기 쉽다.

<Fig. 15>를 통해 10분 단위 교통량 데이터를 요일별로 군집이 변화되는 값들을 비교하기 위해 5주 동안의 데이터에서 각 요일의 각 시간마다 똑같은 군집으로 지정된 횟수를 나타냈다. 노란색으로 나타나는 부분은 모든 기간에서 같은 군집으로 지정됨을 의미하고, 청록색은 5주 중 4주는 같은 군집으로 지정, 남색은 5주 중 3주를 같은 군집으로 분류하고 나머지 2번은 다른 군집으로 분류했음을 의미한다. 이를 통해 10분 단위 데이터의 경우 평일 새벽, 침투 시간은 비교적 군집이 일정하게 형성되지만 그 외의 시간들은 군집이 유동적으로 변화됨을 의미한다.



<Fig. 15> Flow comparison graphs for 10-minute clusters

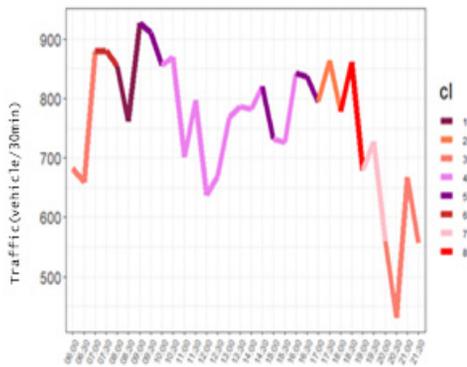


<Fig. 16> Flow comparison graphs for 30-minute clusters

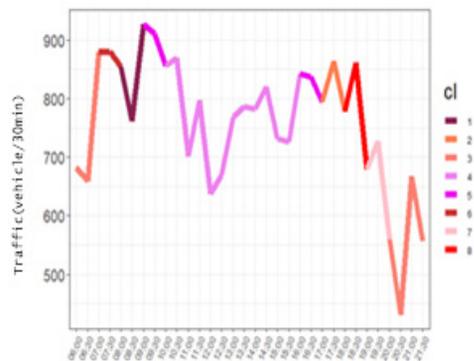
다음으로는 30분 단위의 교통량 데이터를 가지고 군집의 유동성에 대해 <Fig. 16>으로 나타냈다. <Fig. 15>와 어느 정도 비슷한 경향을 보이나, 군집의 유동성이 작용하는 부분을 10분 단위로 비교했을 때 많이 감소되는 것을 알 수 있고, 군집의 유동성이 10분 단위 데이터와 비교했을 때 크지 않음을 알 수 있다.

Heatmap 그래프와 군집의 유동성에 대한 그래프로 10분 단위와 30분 단위의 Time Gap을 비교해봤을 때, 10분 단위보다 30분 단위가 더 나은 것으로 나타났고, 그 중 교통량 데이터를 통해 군집 분석을 했을 때 가장 좋은 결과를 나타냈다. 10분 단위로 신호 주기를 살펴봤을 때, 신호 주기가 너무 짧아 교통량의 변동이 심하고 비슷한 시간대의 군집이 다르게 형성될 수 있다는 것을 <Fig. 10> ~ <Fig. 12>를 통해 확인했다. 또한 신호가 너무 자주 바뀌게 되면 신호 체계의 운영적인 부분에서도 짧은 시간의 지속적인 변화를 활용하기 어렵고, 운전자들에게도 혼란을 야기할 수 있다.

<Fig. 17>은 교통량 데이터의 군집 분석결과를 통해 하루의 데이터를 시계열로 표시하고 각 시간별로 군집의 색을 다르게 구분한 그래프이다. 그래프를 통해 살펴봤을 때 짧은 시간 동안에 군집이 형성되는 부분들이 있음을 확인할 수 있다.



<Fig. 17> Cluster results before minimum transition time applied



<Fig. 18> Cluster results after minimum transition time applied

<Table 4> Weekdays TOD time boundary

Cluster	Time
1	8~9 o'clock
2	17~18 o'clock
3	6~7 o'clock, 20~22 o'clock
4	10~16 o'clock
5	9~10 o'clock, 16~17 o'clock
6	7~8 o'clock
7	19~20 o'clock
8	18~19 o'clock

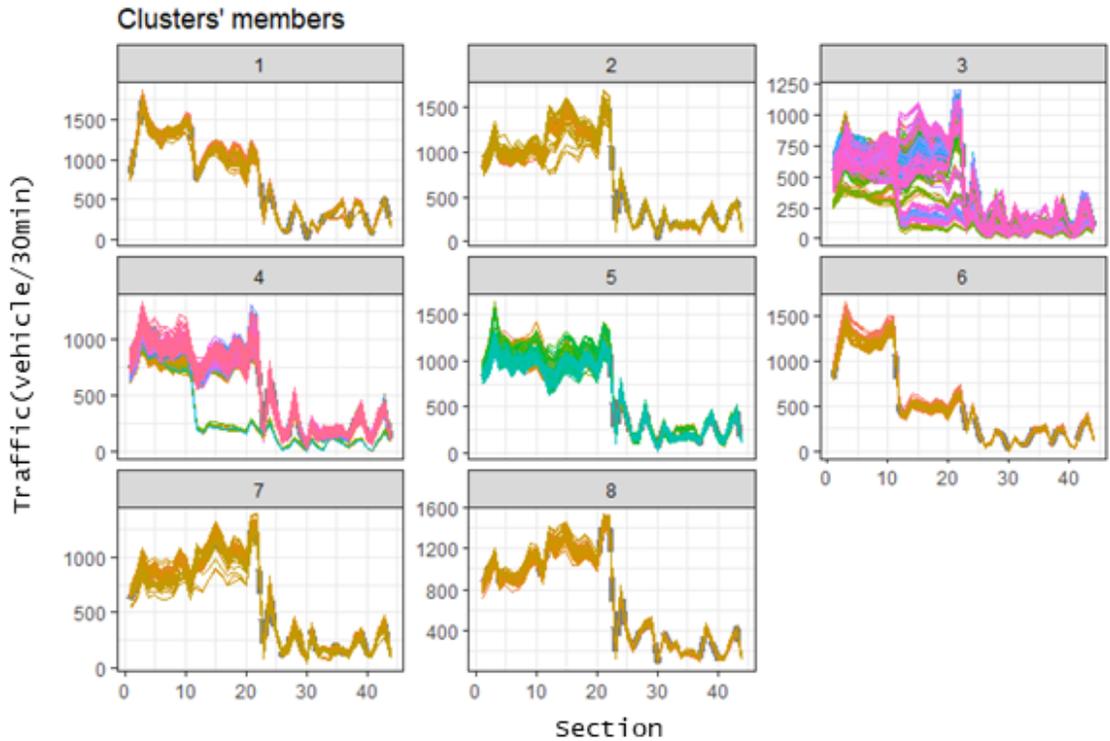
군집 분석을 통해 TOD의 시간 계획 분류 결과를 현장에 적용하기 위해서는 신호가 변경되기 위해 필요한 최소 전이시간이 필요하다. 신호 주기 변화가 자주 일어나면 교통에 많은 영향을 미칠 수 있고, 운영적 인 면에서도 활용할 수 없기 때문에 신호 주기의 변화를 최소 30분에서 1시간으로 둔다. 최소 전이시간을 1시간으로 두기 위해, 군집 분석의 결과를 보정하는 작업이 필요하다. 해당 작업은 1시간 미만의 군집은 인접한 군집 중 교통 패턴이 비슷한 군집으로 보정하는 작업을 뜻한다. <Fig. 17>에서는 오후 2시부터 2시 30분까지 약 30분에 해당하는 5번 군집을 오후 시간대에 주로 분포하는 4번 군집으로 보정해 활용하고자 한다. 보정한 결과는 <Fig. 18>과 같다. 평일을 기준으로 그려진 그래프를 확인했을 때, 주로 차량이 밀집되는 첨두 시간을 기준으로 1시간 전, 1시간 후, 그리고 나머지 시간으로 구분됨을 확인할 수 있고 표로 정리했을 때, <Table 4>와 같다.

2-3) 군집 특성 시각화

여러 기준 중 가장 좋은 군집 분석의 결과를 나타낸 기준은 30분 단위의 교통량 데이터이다. 이에 대한 전체 군집에 대한 특성은 <Fig. 19>에 나타나고 군집 별 분포 수는 <Table 5>를 통해 나타난다. 군집 3, 4, 5번을 제외한 나머지 분포 수가 상대적으로 적은데, 5개의 군집은 첨두 시간인 출·퇴근 시간에 주로 분포한다. 군집 1번과 6번은 첨두 시간 중 출근 시간에 해당하는 오전 7시에서 9시의 구간을 나타낸다. 그래프를 통해 살펴봤을 때 북쪽에 해당하는 1~11 구간에 교통량이 다른 구간에 비해 많이 이용하고 있음을 나타낸다. 오전 출근 시간이 시작되는 7시에서 8시에 분포한 6번 군집의 경우 북쪽으로 향하는 차량들이 월등히 많음을 확인할 수 있다.

다음으로 군집 2, 7, 8번은 퇴근 시간에 해당하는 17시에서 20시에 분포한 군집으로 남쪽에 해당하는 12~22 구간이 다른 구간에 비해 교통량이 많은 것으로 나타난다. 2번과 7번의 경우 비슷한 모습을 띄고 있는데, 이는 가장 교통이 정체되는 18시부터 19시를 기준으로 전과 후를 나타낸다.

다른 군집에 비해 비교적 분포 수가 많은 3, 4, 5번 군집은 비첨두 시간에 주로 분포한 군집으로, 3개의 군집 중 가장 적은 분포수를 가진 군집 5번의 경우는 출근 시간이 끝나는 9시 이후와 퇴근이 시작되는 17시 이전에 주로 분포하고 있으며, 금요일의 경우 좀 더 이른 시간인 2시, 토요일의 경우 오후 시간에 주로 분포하고 있는 것으로 확인된다. 그리고 가장 많은 분포 수를 가진 4번 군집은 주로 오후 시간대에 분포해 있으며, 군집 5번은 출근 시간 이전과 퇴근 시간 이후에 분포되어 있다. 군집 3, 4, 5번은 북쪽과 남쪽 구간에 해당하는 1~22 구간이 고루 분포되어 있음을 확인할 수 있다.

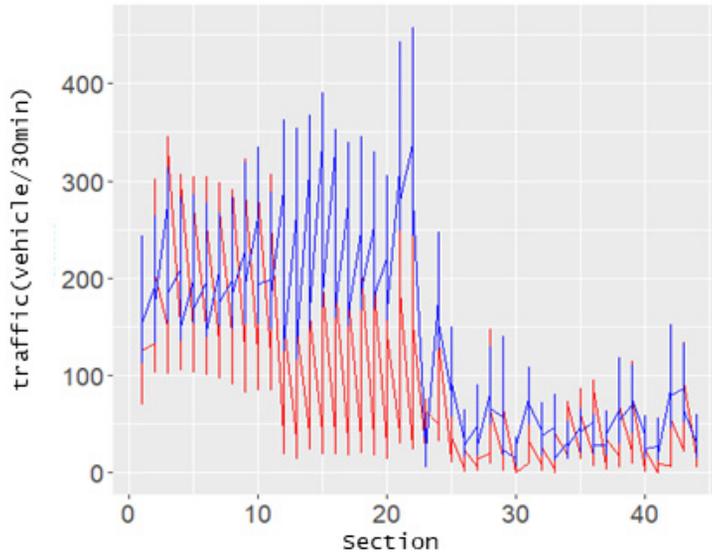


<Fig. 19> Overall Cluster properties for Traffic Data

<Table 5> Each cluster distribution

Cluster	Distribution
1	53
2	39
3	323
4	357
5	171
6	50
7	57
8	50

<Fig. 20>은 교통량 데이터에 대한 군집 중 새벽 시간대에 분포하며 가장 편차가 큰 3번 군집을 세분화한 그래프이다. x축은 도로 구간, y축은 교통량을 나타내며, 빨간 시계열 그래프는 오전 6시부터 12시, 파란 시계열 그래프는 17시부터 22시까지를 나타낸다. 구간별 교통량을 확인했을 때 3번 군집의 경우 오전 시간대에는 북쪽으로 향하는 차량이 다른 방향에 비해서 많고, 야간 시간대에는 남쪽으로 향하는 차량이 다른 방향에 비해 많다. 결과적으로 야간 시간대에 차량이 주간 시간대에 비해 많음을 그래프를 통해 확인할 수 있다.



<Fig. 20> Cluster 3 time series graph

V. 결론 및 향후 연구

본 논문에서는 도시 내 교통혼잡 비용의 지속적인 증가를 해결하기 위해 효율적인 도로 운영의 중요성이 부각되고, AI 기반의 실시간 신호 운영 시스템을 개발하는 현시점에서 가장 많이 활용되고 있는 신호제어 방법인 TOD 제어를 활용하기 위해 최적화된 TOD 시간 분할 방법에 대해서 제안하고 있다. 비 계층적 군집 분석과 계층적 군집 분석에서 활용되는 Euclidean Distance 기반이 아닌 시계열 군집 간의 거리를 데이터의 일부 확장 또는 압축시키는 DTW basic 거리 기반의 DTW 모델을 활용해 TOD 시간을 분할하는 방법을 사용했다. 또한 기존의 군집 분석에서 가장 중요하게 여기는 군집의 개수(k)를 정하는 과정에서 오랜 경험과 가정을 통한 주관적인 견해로 군집의 개수를 정하는 것이 아닌, 군집의 개수가 증가함에 따라 오차 제곱 합의 이질성을 그래프로 살펴봄에 이질성 값의 폭이 줄어드는 지점을 Elbow Point로 지정해 Elbow Point에 해당하는 군집의 개수를 최적의 군집으로 정했다.

최적의 TOD 시간 분할을 위해 대전광역시 대덕대로의 11개 교차로를 대상으로 수집된 교통량과 속도 데이터를 활용했다. TOD 시간 분할을 하는 데 있어 가장 교통 상황을 잘 반영하는 데이터를 알아보기 위해 평균 속도, 교통량, 정규화 된(평균 속도 + 교통량)의 3가지 기준을 나누어 각각의 기준을 비 계층적 군집 분석 결과로 비교했다. 그 결과, 평균 속도를 포함한 경우보다 교통량을 포함한 경우의 군집 분석 결과가 교통 상황 변화와 도로 특징을 더 잘 반영한 결과를 보여준다. 이는 교통량 데이터가 평균 속도 데이터에 비해 값의 편차가 크기 때문에 교통 상황 변화의 특징을 보다 잘 나타내 이런 결과가 나온 것으로 볼 수 있다.

또한, 10분 단위로만 데이터를 살펴보는 것이 아닌 30분 단위로 확장해 각각의 결과를 비교 분석했고, 10분 단위로 군집 분석을 진행한 결과보다 30분 단위로 진행한 군집 분석의 결과가 더 나은 결과를 보여주었다. 10분 단위의 데이터를 사용해 군집 분석을 진행하면 데이터의 잦은 변화로 인해 군집이 자주 바뀌게 되고, 실제 현장에서 10분 단위로 변경되는 신호 운영을 적용할 시 운전자에게 많은 혼란을 야기할 수 있으며,

신호 변경을 위한 최소 전이시간 또한 확보하기 어렵다. 잦은 신호의 변경보다는 신호 변경 시간이 현실적으로 반영되었을 때 더 효율적인 운영이 가능할 것으로 보인다.

해당 논문에서 사용한 데이터는 19년도에 일정 시간 동안 수집된 교통 데이터를 토대로 시뮬레이션을 활용해 추정교통량과 추정 평균 속력을 통한 데이터로 실제와는 차이가 있을 수 있다. 시뮬레이션을 활용한 추정량이기에 실제와 동일하다고 볼 수 없으며, 데이터의 기간이 5주정도로 짧아 각 주에 복합적인 요인들이 발생할 경우 해당 요인들에 영향을 받기가 쉽다. 또한 구간이 전체적인 구간이 아닌 대덕대로를 지나는 특정 구간으로 한정해 분석한 결과물로 모든 교차로 구간에 활용할 수 있는 일반성을 띠는다고 보기에는 어려움이 있다.

현재 정부에서 추진 중인 ITS 사업을 통해 일부 교차로는 영상 또는 라이다 기반의 방향별 교통량과 속도 정보가 수집되고 있다. 현장에서는 교통 상황과 미래 예측 정보를 기반으로 시나리오를 선택해 신호를 운영할 경우, 정체 개선을 위한 효율적인 도로 운영이 가능해질 것으로 보인다.

향후 분석에 정체가 극심한 주요 도심 교차로에 1년 이상의 방향별 교통량 데이터가 충분히 수집된다면, 데이터를 기반으로 다양한 TOD 시간 분할을 최적화하고, 각 시간 계획별 최적 신호 운영 방안을 마련할 수 있을 것이다. 신호를 운영하기 위해서는 TOD 시간을 분할하고, 분할된 TOD 시간 계획에 신호를 최적화시키는 작업이 필요하다. 신호를 최적화시키기 위해 현장에 직접적인 신호 변화를 주기 전과 후를 직접적으로 비교해 교통 정체가 완화되었는지를 살펴보는 검증이 필요하다. 본 논문에서는 신호를 최적화시키는 것보다 TOD 시간 분할 최적화를 진행하는 작업에 초점을 맞추고 방법론을 제시하고 있다. 향후에 DTW 거리 기반의 비 계층적 군집 분석 방법론을 통해 최적화시킨 TOD 시간 분할을 가지고, 신호를 최적화시켜 현장에 직접 적용한 후에 교통 정체가 얼마나 개선되었는지 확인하는 것이 향후 연구가 될 것이다.

ACKNOWLEDGEMENTS

이 논문은 2021년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2020R1I1A3074217).

REFERENCES

- Cho J., Jung H., Nam Bui K. H. and Yi H.(2020), "Image Generation-based Traffic Density Classification for Large-Scale Urban Road Networks," *KIISE Transactions on Computing Practices*, vol. 26, no. 11, pp.499-506.
- Hua J. and Faghri A.(2005), "Dynamic Traffic Pattern Classification Using Artificial Neural Networks," *Transportation Research Record*, vol. 1399, pp.14-19.
- Jang C.(2020), *From machine learning with Python optimization concepts learned with linear algebra and statistics to deep learning using tensorflow*, Bjpulic, p.138.
- Jeong Y., Kim D., Kim Y., Hwang K. and Kim Y.(2008), "Determination of TOD Plan Breaking Points Using a Clustering Analysis," *Korean Society of Road Engineers*, pp.341-347.
- Jia L., Yang L., Kong Q. and Lin S.(2006), "Study of Artificial Immune Clustering Algorithm and Its

Applications to Urban Traffic Control,” *International Journal of Information Technology*, vol. 12, no. 3, pp.1-6.

Lantz B.(2020), *Machine Learning with R*, Acorn, pp.388-389, p.400.

Lee I., Lee H. and Kim Y.(2011), “Optimization of TIME-OF-DAY and Estimation on the Field Application for Arterial Road,” *Journal of Korean Society of Transportation*, vol. 29, no. 4, pp.113-123.

Lee J., Kim J. and Park B.(Brian)(2011), “A genetic algorithm-based procedure for determining optimal time-of-day break points for coordinated actuated traffic signal systems,” *KSCE Journal of Civil Engineering*, vol. 15, pp.197-203.

Lee Y., Ko M. and Choi D.(2019), “Pattern and Energy Intensity Analysis of Monthly Gas Energy Consumption in Apartment Using Dynamic Time Warping Hierarchical Clustering,” *The Korean Society of Living Environmental System*, vol. 26, no. 1, pp.134-139.