

< 응용 논문 >

계층적 지리공간 인덱싱 기반 차량 경로 자동군집화 시스템 개발

김용진·김용현*

(주)한스네트워크

Development of a Vehicle Route Automatic Clustering System Based on Hierarchical Geospatial Indexing

Yongjin Kim · Yonghyun Kim*

HANS NETWORK Co., Ltd., 303, Business Incubation Center, 13-13, Hayang-ro, Gyeongsan-si, Gyeongbuk 38430, Korea
(Received 19 December 2025 / Revised 13 January 2026 / Accepted 29 January 2026)

Abstract : This study proposed a route clustering system that utilizes GPS-based driving data to fairly compare vehicle performance indicators. The accumulated data was clustered using Valhalla map matching, H3 hexagonal partitioning, Hausdorff distance, and DBSCAN algorithm. Computational efficiency was enhanced through H3-based representative data extraction. Validation results using real bus data demonstrate that the recommended system effectively compares fuel economy, electric energy efficiency, safe driving scores, and carbon emissions by reflecting driving conditions. A more generalized and performance-enhanced system based on this research can be developed by considering additional factors, such as deep learning-based high-speed clustering, temporal dynamics, and unstructured routes.

Key words : GPS(범지구위치결정시스템), H3 hexagonal hierarchical geospatial indexing(H3 육각형 계층적 지리공간 인덱싱), Clustering(군집화), Unsupervised learning(비지도 학습), DBSCAN(밀도 기반 클러스터링 비모수적 알고리즘), Hausdorff distance(Hausdorff 거리)

1. 서론

1.1 연구 배경: 지속 가능한 교통과 데이터 분석의 필요성

최근 기후변화와 에너지 위기 문제가 심화됨에 따라, 교통 부문의 탄소 배출 저감과 에너지 효율 향상은 사회 전반의 중요한 과제가 되고 있다. 1990년부터 이산화탄소 배출량은 매년 평균 7.2-7.6%의 높은 증가율을 보이고 있으며¹⁾ 내연기관 자동차의 꾸준한 증가는 CO₂의 배출량 증가와 온실효과의 가속화를 초래할 것으로 예상된다. 기후 변화에 대응하기 위한 국제적인 노력이 가속화됨에 따라, 운송 산업은 내연기관 차량의 효율 향상뿐만 아니라 전기차, 하이브리드 차량 및 수소연료 전지차로의 전환이라는 거대한 패러다임의 변화를 겪고 있다.²⁾

또한 온실가스 배출의 주요 원인 중 하나인 도로 교통

부문에서 개인 차량 이용을 줄이고 대중교통을 활성화하는 것은 지속 가능한 도시 교통 체계 구축의 핵심 전략으로 주목받고 있다. 대중교통은 일정한 경로를 반복적으로 운행한다는 특성을 지니며, 이에 따라 축적되는 GPS 기반 운행 데이터는 운행 행태를 정량적으로 분석할 수 있는 중요한 자료로 활용될 수 있다.

이러한 배경 속에서, 차량의 주행 데이터를 수집하고 분석하여 운영 최적화를 도모하는 텔레매틱스(Telematics) 기술의 중요성은 날로 증대되고 있다. 차량에 부착된 센서와 GPS 장비를 통해 수집되는 대규모의 주행 데이터는 운전자의 습관, 차량의 상태, 그리고 도로 환경에 대한 풍부한 정보를 담고 있다. 하지만, 이러한 데이터의 양적 팽창에도 불구하고, 이를 질적인 인사이트로 변환하는 과정에는 여전히 기술적 장벽이 존재한다. 가장 큰 문제는 데이터의 비정형성과 환경 변수의 다양성

*Corresponding author, E-mail: mrhans74@hansnetwork.co.kr

¹⁾This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium provided the original work is properly cited.

이다. 차량의 연비나 전비, 그리고 안전운전 점수와 같은 성능 지표는 주행하는 도로의 기하학적 구조(경사도, 커브), 교통 흐름(정체, 신호 대기), 기상 상황 등에 절대적인 영향을 받는다. 예를 들어, 평탄하고 교통량이 적은 고속화 도로를 주행하는 차량과, 경사가 심하고 가다 서다를 반복하는 도심 혼잡 구간을 운행하는 차량의 연비를 단순히 절대적인 수치로 비교하는 것은 통계적으로 타당하지 않다. 이러한 비교는 운수사의 노선 운영 전략 수립이나 운전자에 대한 공정한 인센티브 제공을 저해하는 요인이 된다.³⁾ 따라서 차량 간, 혹은 운전자 간의 성능을 객관적이고 공정하게 비교하기 위해서는 주행 조건이 유사한 데이터끼리 그룹화하여 비교군을 형성하는 과정, 즉 경로 군집화(Trajectory clustering)가 선행되어야 한다. 이 작업을 통해 다수의 운행 경로를 효과적으로 군집화하여 주요 노선을 식별하고, 반복 운행 행태를 정량화할 수 있을 것으로 기대한다.

1.2 문제 정의 및 기술적 난제

기존의 경로 군집화 연구들은 기존 군집화 모델의 단점을 보완하는 새로운 모델을 제시하는 연구,⁴⁾ 경로 전체가 아닌 공통된 하위 경로를 기반으로 군집화를 수행하는 연구⁵⁾ 등이 이루어졌지만, 이들은 이미 단일 주행 단위로 분할된 데이터셋을 전제로 수행되는 경우가 대부분이다. 그러나 실제 운행 환경에서 수집되는 일일 GPS 데이터는 동일한 경로가 임의의 횟수만큼 반복 주행된 연속 시퀀스로 구성되며, 이를 자동으로 분할할 수 있는 일반화된 알고리즘이 충분히 제시되지 않았다. 이러한 한계는 경로 분석 및 군집화 과정의 실용성을 저해하는 요인으로 작용해 왔다.

GPS 기반의 경로 데이터를 군집화하여 비교 가능한 모집단을 구축하는 것은 이론적으로는 명확해 보이나, 실무적 구현 단계에서는 다음과 같은 복합적인 기술적 난제에 직면하게 된다.

첫째, 데이터의 불확실성과 오차(Uncertainty and error) 문제이다. 도심 지역의 고층 빌딩 숲(Urban canyon)이나 지하차도, 터널 등은 GPS 신호의 수신을 방해하거나 다중 경로 오차(Multi-path error)를 유발하여, 차량의 위치가 실제 도로를 벗어난 곳에 기록되게 한다.⁶⁾ 이러한 노이즈가 섞인 데이터를 그대로 분석에 활용할 경우, 경로 간의 유사도 측정에 심각한 왜곡이 발생한다. 따라서 원시 GPS 데이터를 실제 도로 네트워크 위에 정확히 정합시키는 맵 매칭(Map matching) 기술과 위치 보정 알고리즘의 고도화가 필수적이다.

둘째, 대규모 데이터 처리의 연산 복잡도(Computational complexity) 문제이다. 상용차 한 대가 하루에 생성하는

GPS 포인트는 수만 개에 달하며, 수백, 수천 대의 차량이 수년 동안 축적한 데이터는 페타바이트급에 이른다. 기존의 동적 시간 워핑(DTW: Dynamic Time Warping)과 같은 정밀한 유사도 측정 알고리즘은 두 시계열 데이터 길이의 곱에 비례하는 연산 비용($O(N^2)$)을 요구하기 때문에, 대규모 데이터셋에 대해 실시간으로 적용하기에는 한계가 있다.⁷⁾ 따라서 데이터의 기하학적 특성을 보존하면서도 차원을 효과적으로 축소하는 기술과, 대용량 데이터를 효율적으로 처리할 수 있는 아키텍처가 요구된다.

셋째, 경로의 방향성 및 세부 패턴 식별(Directionality and micro-patterns) 문제이다. 동일한 도로 구간이라 하더라도 상행선과 하행선은 교통 흐름, 신호 체계, 그리고 도로의 구배(Omsby)가 전혀 다를 수 있다. 단순히 공간적인 겹침(Spatial overlap)만을 기준으로 군집화를 수행할 경우, 서로 반대 방향으로 주행한 데이터를 동일한 그룹으로 오분류할 위험이 크다.⁸⁾ 이는 연비 분석이나 안전 운전 평가에 있어 치명적인 오류를 초래할 수 있으므로, 경로의 형상뿐만 아니라 벡터 기반의 방향성까지 고려한 정교한 군집화 방법론이 필요하다.

1.3 본 연구의 목적

본 보고서는 상기한 문제점들을 해결하고, 대규모 차량 주행 데이터를 기반으로 차량 성능 지표를 공정하게 비교할 수 있는 ‘계층적 경로 군집화 시스템(Hierarchical trajectory clustering system)’을 제안하고 이를 상세히 기술한다. 본 시스템을 통해 어떠한 형태의 일일 GPS 데이터가 주어지더라도 이를 독립적인 단일 주행 경로로 자동 분할할 수 있으며, 나아가 분할된 각 주행 경로는 H3 기반 공간 격자를 활용하여 합리적인 공간 해상도에서 좌표를 추출·압축함으로써, 경로 군집화에 필요한 표본의 크기를 효과적으로 축소할 수 있다.

본 연구는 Valhalla 엔진을 이용한 정밀 맵 매칭, H3 육각형 공간 인덱싱을 통한 데이터 압축, Hausdorff 거리와 DBSCAN을 결합한 1차 형상 군집화, 그리고 벡터 기반의 2차 방향성 군집화로 이어지는 일련의 파이프라인을 구축하였다. 또한, 연 단위의 대규모 배치 처리와 일 단위의 실시간 데이터 처리를 결합한 람다 아키텍처(Lambda architecture)를 도입하여 시스템의 확장성과 실용성을 확보하였다.

이와 같은 접근은 기존 연구가 전제로 삼았던 사전 분할 데이터의 한계를 극복하고, 실제 운행 환경에서 수집되는 원시 GPS 데이터로부터 데이터 분할-압축-군집화에 이르는 전 과정을 자동화한다는 점에서 중요한 의미를 지닌다. 나아가 이는 반복 운행 특성을 갖는 대중교통

뿐만 아니라 다양한 형태의 이동 데이터를 포괄할 수 있는 범용 경로 분석 프레임워크로 확장될 수 있는 기반을 제공한다. Fig. 1은 구축된 시스템 전체의 플로우 차트를 간략하게 그린 것이다.

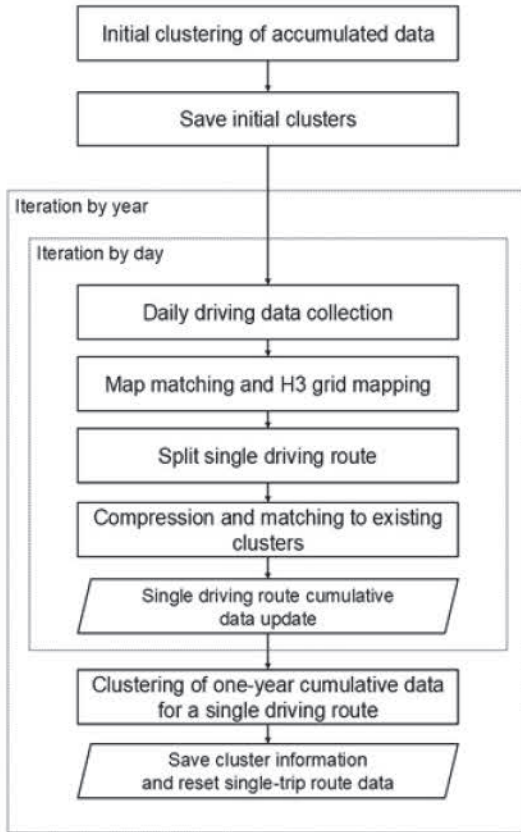


Fig. 1 Flow chart of the entire system

2. 이론적 배경 및 관련 연구

2.1 맵 매칭(Map Matching) 및 위치 보정 기술

차량 궤적 데이터 분석의 첫 단추는 불완전한 GPS 좌표를 디지털 지도상의 도로 링크(Link) 또는 노드(Node)에 정확히 대응시키는 맵 매칭 과정이다. 이는 단순한 좌표의 이동이 아니라, 차량의 주행 궤적, 속도, 방향, 도로 네트워크의 위상(Topology) 등을 종합적으로 고려하여 가장 개연성 높은 경로를 추론하는 확률적 과정이다.

본 연구에서 채택한 Valhalla는 오픈 소스 기반의 고성능 라우팅 및 맵 매칭 엔진으로, 타일 기반의 계층적 데이터 구조(Tiled hierarchical data structure)를 사용하여 메모리 효율성과 연산 속도를 극대화하였다.⁹⁾ Valhalla는 내부적으로 Meili라는 라이브러리를 통해 맵 매칭을 수행하는데, 이는 은닉 마르코프 모델(HMM: Hidden Markov Model)을 기반으로 작동한다. HMM은 관측된 GPS 좌표

(Emission)와 실제 도로 네트워크 상의 상태(State) 간의 전이 확률(Transition probability)을 계산하여, 가장 가능성이 높은 상태의 시퀀스, 즉 비터비 경로(Viterbi path)를 찾아낸다.

Valhalla의 차별점은 단순한 거리 기반 매칭을 넘어, 도로의 속성(일방통행 여부, 회전 제한, 차량 등급 제한 등)과 비용 모델(Costing models)을 통합하여 매칭의 정확도를 높인다는 점이다. 또한, Thor 라이브러리는 그래프 타일 계층을 통해 경로를 생성하고, Odin은 이를 바탕으로 내비게이션 안내(Narrative)를 생성하는데, 이러한 모듈화 된 구조는 대규모 데이터 처리 파이프라인에 유연하게 통합될 수 있는 장점을 제공한다.

한편, 도심 환경에서는 위성 신호가 고층 건물에 반사되는 다중 경로 효과나 신호 차폐로 인해 GPS 위치가 수십 미터 이상 튀는 현상이 빈번하다. 이러한 오차는 맵 매칭의 실패나 엉뚱한 도로로의 매칭을 유발할 수 있다. 이를 보정하기 위해 전통적으로 칼만 필터(Kalman filter)나 확장 칼만 필터(EKF)가 사용되어 왔다.¹⁰⁾ 칼만 필터는 이전 상태의 추정값과 현재의 측정값을 결합하여 시스템의 상태(위치, 속도)를 최적으로 추정하는 알고리즘으로, GPS 오차의 공분산을 모델링하여 노이즈를 평활화(Smoothing)하는 데 효과적이다.¹¹⁾ 본 연구에서는 Valhalla의 전처리 단계에서 중앙값 필터와 함께 이러한 확률적 보정 기법을 적용하여 입력 데이터의 품질을 확보하였다.

2.2 H3 육각형 계층적 공간 인덱싱

대규모 궤적 데이터를 효율적으로 관리하고 연산하기 위해서는 공간을 이산적인 격자(Grid)로 나누어 인덱싱하는 것이 필수적이다. Uber에서 개발하여 오픈 소스로 공개한 H3은 지구 표면을 육각형 셀로 분할하는 계층적 인덱싱 시스템이다. 기존의 사각형(Latitude/longitude bounding box) 그리드나 Geohash와 비교할 때, 육각형 그리드는 결정적인 장점을 가진다.

첫 번째로, 중심점에서 모든 6개의 인접한 이웃 셀까지 거리가 동일하다. 반면 사각형은 대각성 방향의 이웃과 수직/수평방향의 이웃 간 거리가 다르다. 이러한 특성은 이동 경로의 근사화나 반경 검색(Radius query), 그리고 평활화 연산에 있어 왜곡을 최소화한다.

두 번째로, H3는 16단계의 해상도(Resolution)를 제공하여 분석 목적에 따라 공간 분할의 정밀도를 조절할 수 있다. 예를 들어 Resolution 9는 한 변의 길이가 약 174 m, 영역 넓이가 약 0.1 km²인 육각형으로, 도심 내 도로 구간을 식별하기에 적절한 크기를 제공한다.

2.3 궤적 유사도 측정 방법론의 비교

두 경로가 얼마나 유사한지를 정량적으로 판단하는 척도(Distance measure)는 군집화의 품질을 결정짓는 가장 중요한 요소이다. 학계에서는 다양한 척도가 제안되었으며, 각각의 장단점은 다음과 같다.

- 1) Euclidean 거리(Euclidean distance): 동일 시점의 점 간 거리 평균으로 계산이 매우 빠르고 직관적이다. 그러나 두 궤적의 길이가 같아야 하며, 시간 축의 이동을 반영하지 못해 본 연구에 적합하지 못하다.
- 2) 동적 시간 워핑(Dynamic time warping): 시계열의 길이를 늘이거나 줄여 최적의 매칭을 수행함으로써 산출되는 척도이다. 속도가 다르거나 길이가 다른 궤적 비교에 탁월하지만 연산 복잡도가 $O(N^2)$ 로 높고 노이즈에 민감할 수 있다. 소규모 데이터에 적합하다.
- 3) LCSS(Longest Common Subsequence): 경로들의 공통된 부분 경로의 길이를 측정한다. 노이즈와 이상치(Outlier)에 매우 강건하지만 비교 대상인 두 궤적의 전체적인 형상보다 부분적인 일치에 초점을 맞춘다.
- 4) Hausdorff 거리(Hausdorff distance): 두 점 집합 간의 최대-최소거리로 나타난다. 점의 순서나 개수에 무관하며 궤적의 전체적인 공간적 형상(Shape) 유사도 측정에 최적인 척도이다. 그러나 시간 정보와 방향성을 무시하기 때문에 실제 경로 상 주행 방향을 구분할 수 없다.
- 5) Fréchet 거리(Fréchet distance): 경로의 순서를 고려한다는 점에서 Hausdorff 거리보다 엄밀한 척도이다. 그러나 연산 비용이 매우 높아 대규모 데이터를 처리하기에 적합하지 않다.

본 연구에서는 대규모 데이터에 대한 연산 효율성을 최우선으로 고려하여, 1차적으로는 집합 간 거리를 빠르게 계산할 수 있는 Hausdorff 거리를 채택하였다. 그러나 앞서 언급된 Hausdorff 거리의 한계, 즉 방향성 상실 문제를 보완하기 위해 2차 단계에서 벡터 기반의 방향성 필터링을 추가하는 하이브리드 접근 방식을 설계하였다.

2.4 군집화 알고리즘 선택

유사도 행렬이 도출된 후, 이를 바탕으로 데이터를 그룹화하는 알고리즘의 선택 역시 중요하다.

K-Means 알고리즘은 가장 대중적인 알고리즘이나, 군집의 개수를 사전에 지정해야 하고, 군집의 형태가 같은 크기를 가진 구형(Spherical)일 것이라고 가정한다는 한계가 있다. 도로망을 따르는 궤적 데이터는 길게 뻗은 형태나 임의의 모양을 가지므로 K-Means는 본 연구에 부

적합할 수 있다.¹²⁾

DBSCAN(밀도 기반 클러스터링 비모수적 알고리즘, Density-Based Spatial Clustering of Applications with Noise) 알고리즘은 데이터의 밀도를 기반으로 군집을 형성한다. 임의의 형상을 가진 군집을 잘 찾아내며, 무엇보다 이상치(Outlier)를 별도의 군집으로 분류해내는 능력이 탁월하다. 버스 노선 데이터의 경우 정규 노선을 이탈한 경로를 탐지하는 것이 중요하므로 DBSCAN이 보다 적합한 모델로 판단된다.¹³⁾

2.5 차량 성능 지표와 정규화

본 연구의 궁극적인 목표는 단순한 경로 군집화가 아니라, 이를 통한 차량 성능의 공정한 비교이다. 차량의 연료 소비는 차량의 중량, 엔진 효율 뿐만 아니라 도로의 구배, 교통 부하, 정지 빈도 등에 의해 결정된다.¹⁴⁾ CMEM(Comprehensive Modal Emission Model)과 같은 미시적 배출 모델은 이러한 변수들을 고려하여 연료 소모를 예측한다.¹⁵⁾

유사한 맥락에서, 운전자의 안전 운전 점수(예: 급가속, 급감속 횟수) 역시 주행 환경에 따라 보정되어야 한다. 미국의 CSA(Compliance, Safety, Accountability) 프로그램은 운수사의 안전도를 평가할 때, 노선의 난이도나 운행 거리에 따른 노출도를 고려하여 점수를 정규화한다.¹⁶⁾ 본 시스템은 이러한 정규화를 위한 ‘동일 조건의 모집단’을 제공하는 기반 인프라로서 기능한다.

3. 방법론

본 연구는 대규모 차량 주행 데이터를 기반으로 경로의 군집화 및 신규 주행 경로의 자동 분류를 수행하기 위한 일련의 처리 절차로 구성된다. 전체 프로세스는 크게 (1) 맵 매칭(Map matching), (2) 단일 주행 분할, (3) 1차 군집화, (4) 2차 군집화의 네 단계로 이루어진다.

먼저, 주행 궤적 데이터는 Valhalla 엔진을 이용하여 실제 도로망과의 정합을 수행하고, 이후 H3 육각형 계층적 지리공간 인덱싱(Hexagonal hierarchical geospatial indexing)을 적용하여 각 GPS 포인트를 고정된 해상도의 육각형 격자 단위로 변환한다. 이 과정을 통해 연속된 격자 시퀀스로 표현된 경로를 얻을 수 있으며, 이를 토대로 특정 노선을 여러 차례 왕복 주행한 동일 데이터로부터 개별 단일 주행 경로를 효과적으로 분리하고, 군집화에 활용할 표본의 크기를 효율적으로 축소할 수 있다.

이후 H3 인덱스를 기반으로 각 차량의 일일 주행 데이터를 단일 주행 구간 단위로 분할하고, 경로의 공간적 유사성을 정량적으로 비교하기 위해 Hausdorff 거리(Hausdorff distance)를 계산한다. 이를 입력 특성으로 사

용하여 DBSCAN(Density-Based Spatial Clustering of Applications with Noise) 알고리즘을 수행함으로써, 1차적으로 유사한 경로 집단을 도출한다.

마지막으로, 1차 군집 내의 경로 간 방향성 및 세부 이동 패턴을 벡터 형태로 표현하고, 이 벡터를 입력 특성으로 사용하는 2차 군집화를 수행하여 상행·하행 또는 외선·내선순환 등 방향성이 상이한 세부 경로를 구분한다. 이러한 다단계 군집화 접근은 GPS 오차가 존재하는 실제 교통 데이터 환경에서도 안정적으로 경로 유형을 분류할 수 있는 구조적 강점을 제공한다.

3.1 핵심 방법론 I: 데이터 전처리 및 차원 축소

본 절에서는 원시 데이터가 군집화 알고리즘에 입력되기 전까지 거치는 전처리 과정을 기술한다. 이 단계의 품질이 최종 분석 결과의 신뢰도를 좌우한다. Fig. 2는 맵 매칭 및 단일 주행 추출 과정을 포함하는 전처리 과정을 나타낸 플로우 차트이다.

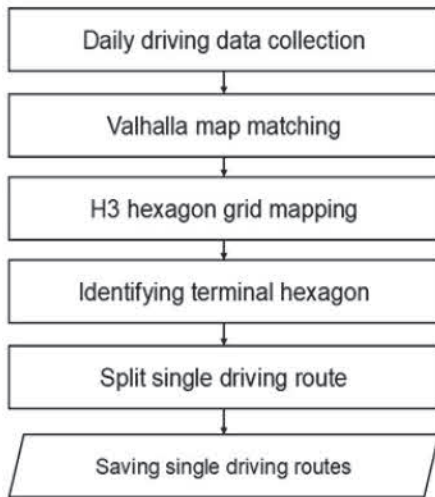


Fig. 2 Flow chart of the pre-processing process

3.1.1 GPS 노이즈 필터링 및 맵 매칭 구현

시내버스의 일일 주행 데이터는 일정한 노선을 하루 동안 여러 차례 왕복 운행하며 기록된 GPS 좌표(위도·경도 쌍)로 구성된다. 원본 GPS 데이터는 측정 오차를 포함하고 있기 때문에, 우선 중앙값 필터를 적용하여 노이즈를 제거하였다. 이는 짧은 시간 동안 위치가 급격히 튀는 임펄스 노이즈(Impulse noise)를 제거하는 데 효과적이며, 이동 평균 필터에 비해 경로의 엣지(Edge, 회전 구간 등)를 뭉개지 않고 보존하는 특성이 있다.

이후 Valhalla 엔진의 trace_attributes API를 사용하여 맵 매칭을 수행한다.¹⁷⁾ 입력값으로는 타임스탬프와 위경

도 좌표 쌍의 리스트를 제공하며, Valhalla는 내부의 비용 함수를 통해 차량이 주행했을 가능성이 가장 높은 도로 링크들을 반환한다. 이 과정에서 단순히 도로 위에 점을 찍는 것을 넘어, 유턴 금지 구역이나 진입 금지 구역을 역주행하는 것과 같은 물리적으로 불가능한 경로를 배제함으로써 데이터의 논리적 정합성을 확보한다. 매칭 결과는 보정된 위경도 좌표뿐만 아니라, 매칭된 도로의 ID(Edge ID), 도로 등급, 제한 속도 등의 메타데이터를 포함한다. 이러한 정보는 추후 탄소 배출량 산정 시 도로 유형(고속도로 vs 시내도로)에 따른 가중치를 적용하는데 활용될 수 있다. Fig. 3에 맵 매칭 전과 후의 경로를 비교하여 나타내었다.

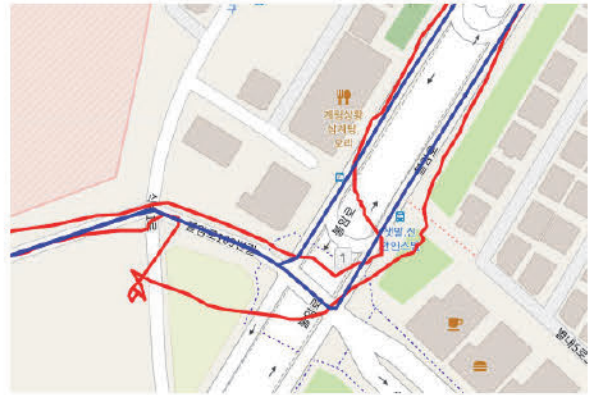


Fig. 3 Route before(red) and after(blue) map matching using valhalla and median filter

3.1.2 H3 기반 공간 분할 및 단일 주행 추출

버스와 같은 상용차는 하루 종일 운행하며 여러 번의 회차를 반복한다. 따라서 일일(Daily) 데이터를 단일 운행(Single trip) 단위로 분할(Segmentation)하는 것이 필요하다. 이를 위해 본 연구는 H3 그리드를 활용한 체류 기반 분할 기법을 적용하였다.

- 1) H3 인덱싱: 맵 매칭된 궤적의 모든 포인트를 H3 Resolution 9~10 수준의 셀로 변환한다.
- 2) 체류 이벤트 감지: 차량이 특정 H3 셀(예: 차고지, 회차지) 내에서 일정 시간(예: 10분) 이상 머무르거나, 시동이 꺼지는 이벤트를 감지한다.
- 3) 경로 분할: 체류 이벤트를 기준으로 긴 시계열 데이터를 여러 개의 짧은 트립으로 자른다.

정합된 경로는 일정 해상도로 H3 육각형 격자에 매핑하였다. 동일한 격자를 이탈 후 재방문하는 경우도 고려하기 위해 특정 격자를 방문하여 연속적으로 체류하는 이벤트를 별도로 인덱싱하여 동일한 격자를 다시 방문하는 경우를 구분하였다. 인덱싱한 방문 이벤트를 기반

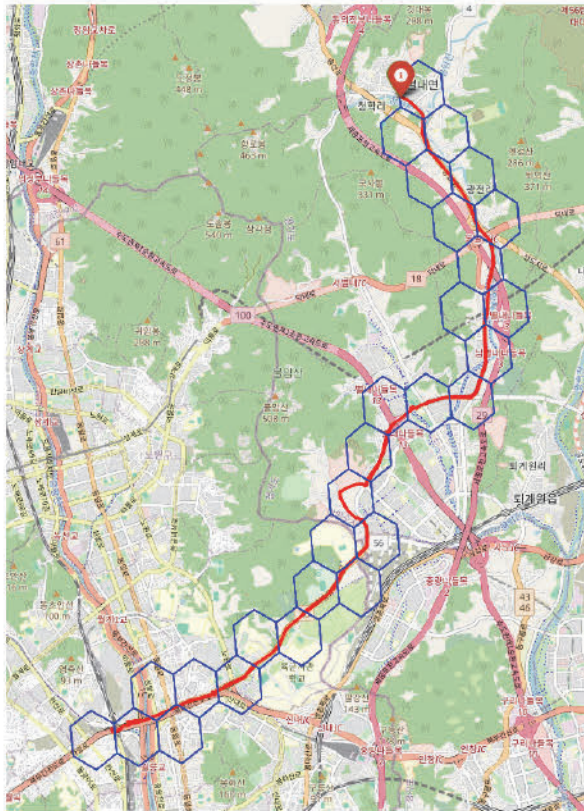


Fig. 4 Result after H3 indexing. Route aligned (map-matched) with the actual road(red) and hexagonal grids mapped onto the route(blue)

으로, 미리 설정한 시간 임계 값 이상 체류한 격자를 기준으로 삼아 하나의 일일 주행 데이터를 여러 개의 단일 주행 데이터로 분할하였다. Fig. 4는 맵 매칭 후의 경로가 지나가는 H3 셀을 나타낸다.

3.1.3 대표 데이터 추출을 통한 차원 축소(Compression)

군집화의 연산 속도를 높이기 위해 궤적 데이터를 압축한다. 수천 개의 GPS 포인트로 구성된 궤적을 그대로 Hausdorff 거리 계산에 사용하면 연산량이 과다하다. 본 연구에서는 공간적 형태를 보존하는 압축을 위해 H3 기반의 대표점 추출 방식을 사용하였다.

- 1) 차량이 통과한 H3 셀들의 유니크한 집합(Set)을 추출한다.
- 2) 각 H3 셀 내부를 지나간 GPS 포인트들의 평균 좌표(Centroid)를 계산하여, 해당 셀을 대표하는 하나의 점으로 대체한다.¹⁸⁾
- 3) 결과적으로, 궤적은 차량이 지나간 H3 셀의 개수만큼의 점들로 구성된 단순화된 폴리라인(Polyline)으로 변환된다.

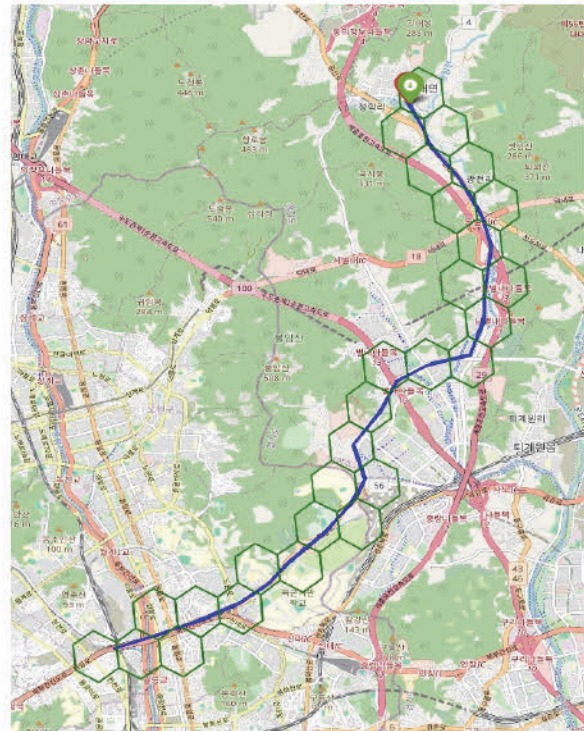


Fig. 5 Hexagonal grids mapped onto an aligned route(green) and the route compressed to one point per grid(blue)

이 방식은 Douglas-peucker 알고리즘과 같은 기하학적 단순화 기법과 달리, 공간 인덱스를 기반으로 하므로 데이터베이스 검색과 연동이 용이하고, 노이즈에 의해 경로의 형상이 왜곡되는 것을 방지할 수 있다. 실험 결과, 데이터 포인트의 개수를 원본 대비 약 1/100 수준(평균 9,000개 -> 70개)으로 줄이면서도 경로의 거시적인 패턴은 완벽하게 유지함을 확인하였다. Fig. 5는 경로가 지나가는 셀 하나 당 평균 좌표 하나를 계산하여 압축된 경로를 나타낸다. Fig. 6은 맵 매칭 직후의 압축되지 않은 경로와 압축된 경로를 비교하여 나타낸다.

3.2 핵심 방법론 II: 계층적 경로 군집화 알고리즘

전처리된 데이터를 바탕으로 수행되는 2단계 계층적 군집화(Two-stage hierarchical clustering)는 본 시스템의 핵심 기술이다. 1단계에서는 형상(Shape)을, 2단계에서는 방향(Direction)을 기준으로 분류하여 정밀도를 극대화한다.

3.2.1 1차 군집화: Hausdorff 거리 기반 형상 분류

1차 군집화의 목표는 기하학적으로 유사한 경로들을 대략적인 그룹(Coarse-grained cluster)으로 묶는 것이다. Fig. 7은 1차 군집화 과정을 나타내는 플로우 차트이다.

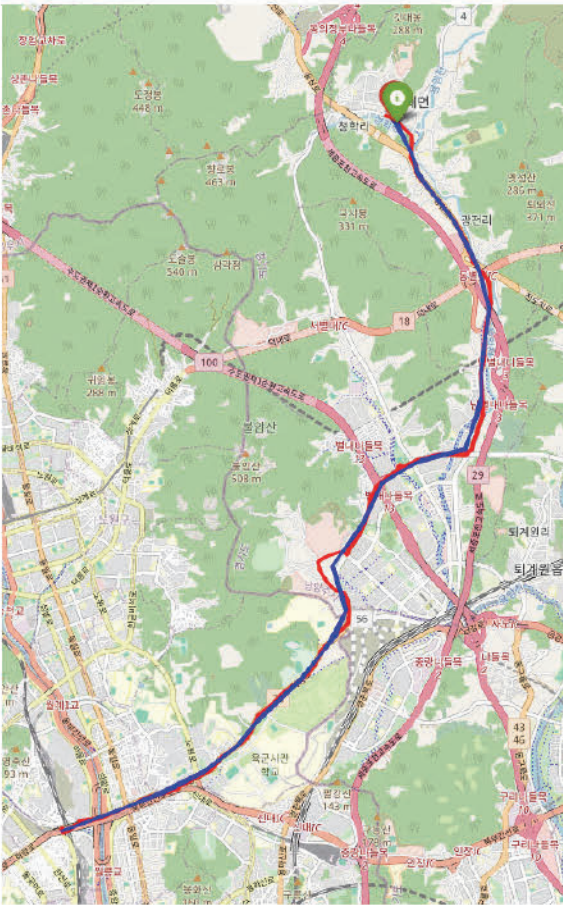


Fig. 6 Comparison of map-matched(red) and compressed (blue) routes

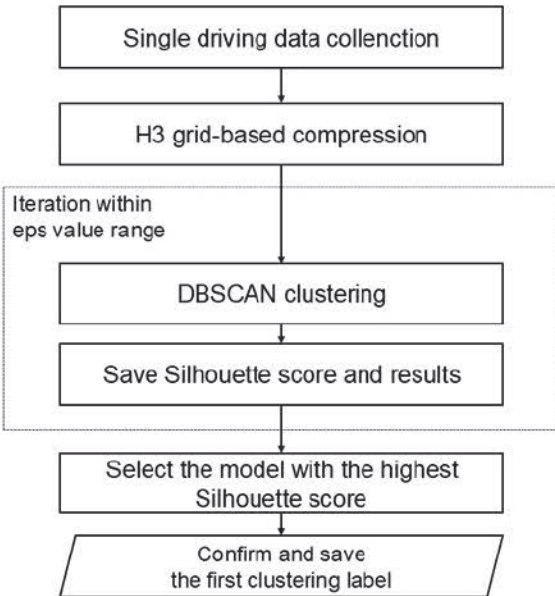


Fig. 7 Flow chart of the first clustering process

경로 간 유사도를 정량적으로 평가하기 위하여 압축된 각 경로 쌍에 대해 Hausdorff 거리를 계산하였다. Hausdorff 거리는 두 점 집합 사이의 최대 거리를 정의하는 척도이다.

압축된 두 경로 $A = \{a_1, \dots, a_m\}$ 와 $B = \{b_1, \dots, b_n\}$ 사이의 Hausdorff 거리 $d_H(A, B)$ 는 다음과 같이 정의된다.

$$d_H(A, B) = \max \left\{ \sup_{a \in A} \inf_{b \in B} |a - b|, \sup_{b \in B} \inf_{a \in A} |b - a| \right\} \quad (1)$$

여기서 $|a - b|$ 는 두 점 a 와 b 사이의 유클리드 거리이며, Sup은 상한, Inf는 하한을 의미한다.

이 정의에 따라 Hausdorff 거리는 점들의 이산적인 집합으로 표현되는 경로 데이터에 적용할 수 있으며, 집합의 크기나 경로의 길이에 영향을 받지 않고 집합 간 공간적 유사성을 수치화할 수 있다는 점에서 본 연구에 적합하다.

또한 이 척도는 “경로 A의 모든 점들이 경로 B로부터 얼마나 멀리 떨어져 있는가?” 중 최악의 경우(Worst-case)를 측정한다. 즉, 두 경로가 대부분 일치하더라도 어느 한 구간에서 크게 벗어난다면 Hausdorff 거리는 커지게 된다. 이는 버스 노선이 잠시 우회하거나 이탈한 경우를 민감하게 잡아낼 수 있어, 엄격한 노선 관리에 적합하다.

계산된 $N \times N$ 거리 행렬을 기반으로, DBSCAN 알고리즘을 적용하여 군집화를 수행하였다. DBSCAN은 두 데이터 간 거리가 임계 값(ϵ 또는 Eps) 이하인 경우를 이웃으로 정의하고, 이웃의 수가 최소 이웃 수(Min_samples) 이상인 데이터를 핵심 포인트(Core point)로 하여 군집을 확장해 나간다. 이때 Eps와 Min_sample은 사용자가 지정하는 상수이다. 또한 최적의 Eps 값을 자동으로 탐색하기 위해, 다양한 Eps 값에 대한 군집화 모델을 시험하고 그중 실루엣 점수(Silhouette score)가 최대가 되는 모델을 선택하도록 하였다.

실루엣 점수는 각 데이터 포인트와 주위 데이터 포인트들과의 거리 계산을 통해 구해지는 값으로, 군집 안에 있는 데이터들이 잘 모여있는지와 군집끼리 잘 구분되는지 여부를 판단함으로써 군집화 작업을 평가하는 척도이다. 먼저 군집 C_i 로 분류된 데이터 i 와 같은 군집으로 분류된 다른 데이터들 간의 거리 평균을 아래와 같이 정의한다.

$$a(i) = \frac{1}{|C_i| - 1} \sum_{j \in C_i, i \neq j} d(i, j) \quad (2)$$

이어서 데이터 i 가 속하지 않은 군집 중 가장 가까운 군집 내 데이터와의 평균 거리를 아래와 같이 정의한다.

$$b(i) = \min_{j \neq i} \frac{1}{|C_j|} \sum_{j \in C_j} d(i, j) \quad (3)$$

위에서 계산된 두 값을 가지고 데이터 i 의 실루엡 값을 아래와 같이 정의한다.

$$s(i) = \begin{cases} \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}, & \text{if } |C_i| > 1 \\ 0, & \text{if } |C_i| = 0 \end{cases} \quad (4)$$

위 정의에 따라 데이터의 실루엡 값은 -1과 1 사이의 값을 갖는다. 분류된 군집의 크기가 1일 경우 $a(i)$ 가 정의되지 않는 대신, 해당 구간의 중심인 0으로 실루엡 값을 정의한다. 실루엡 값이 1에 가까울수록 데이터가 소속 군집으로부터 떨어진 거리보다 이웃 군집으로의 거

리가 크다는 뜻이므로 군집화가 잘 이루어졌음을 시사한다. 데이터셋 내의 모든 데이터의 실루엡 값의 평균값을 모델의 실루엡 점수로 정의한다. 마찬가지로 실루엡 점수가 1에 가까울수록 종합적인 모델의 군집화 성능이 높다고 해석할 수 있다.

Fig. 8은 샘플 데이터에 1차 군집화를 적용한 결과이다.

3.2.2 2차 군집화: 벡터 기반 방향성 세분화

앞선 1차 군집화 단계에서는 Hausdorff 거리를 활용하여 주행 경로의 전체적인 형태를 기준으로 노선을 분류하였다. 이 과정은 경로 간의 길이나 세부적인 점 분포에 영향을 받지 않고 노선의 전반적인 유사도를 평가할 수 있다는 장점이 있으나, 데이터가 점들의 순서를 고려하

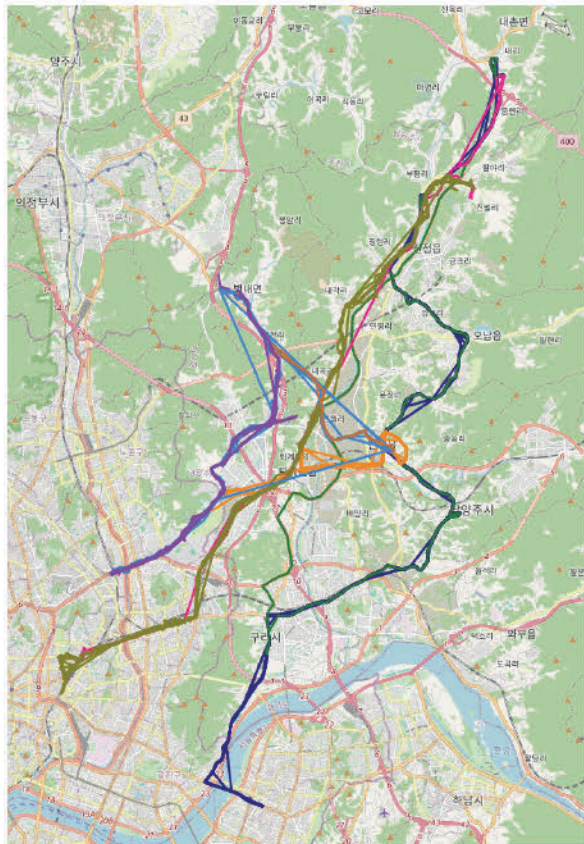


Fig. 8 Example of the results of the first clustering based on Hausdorff distance. The sample data was classified into eight clusters

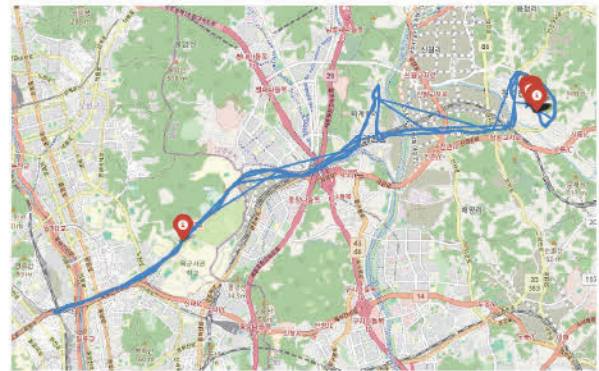


Fig. 9 Routes classified into the same cluster in Fig. 8. Although the starting points (red markers) are distributed differently, they are classified into the same cluster

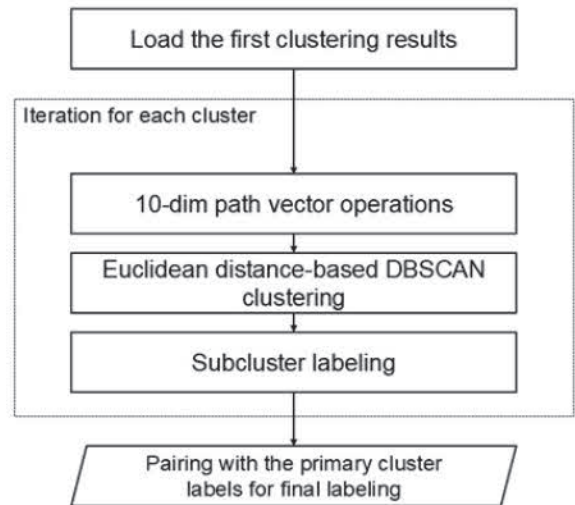


Fig. 10 Flow chart of the second clustering process

지 않는 집합으로 표현되기 때문에 주행 방향 및 횡수에 따른 차이를 반영하지 못한다는 한계가 존재한다. 즉, 동일한 노선 내에서 상행과 하행 등 주행 방향의 구분이 이루어지지 않는 문제가 발생한다. Fig. 9는 1차 군집화에서 같은 군집으로 분류된 경로들의 출발점을 나타낸 것으로, 각 경로들이 다른 지점에서 출발하고 나아가 다른 방향으로 주행하는 경로임을 의미한다. 이를 해결하기 위해 2차 군집화를 수행한다. Fig. 10은 2차 군집화 과정을 나타내는 플로우 차트이다.

2차 군집화 단계에서는 분할된 단일 주행 경로의 전반부 내에 등간격으로 위치한 5개 지점(시작점, 10%, 20%, 30%, 40%)의 경도 및 위도로 구성된 10차원 특성 벡터를 정의하고, 이 10차원 벡터 공간에서 Euclidean 거리를 계산하여 추가적인 군집화를 수행한다. Fig. 11은 Fig. 9의 군집으로 분류된 경로 하나를 가져와 벡터화에 사용될 5개 지점을 표시한 것이다.

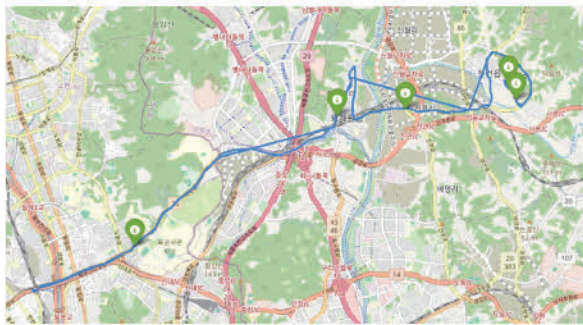


Fig. 11 Five points(green markers) used to form a 10-dimensional vector on one path within the cluster in Fig. 9

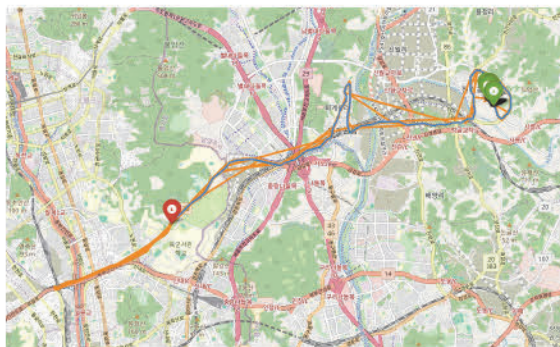


Fig. 12 The results of secondary clustering of the clusters in Fig. 9. The starting point and driving direction are indicated by an orange path starting from a green marker and a light blue path starting from a red marker, respectively

해당 작업은 앞서 이루어진 1차 군집화 결과로 도출된 각각의 군집에 대해 한 번씩 이루어지며, 1차 군집화에서 부여된 Label과 해당 군집 내에서 분류된 하위 군집의 Label의 쌍을 기반으로 최종 Label을 출력하여 전체 군집화 작업을 마무리한다.

Fig. 12는 Fig. 9의 군집에 2차 군집화를 적용한 결과로, 출발점 및 주행 방향에 따라 2개의 하위 군집으로 분리되었음을 의미한다. 이 과정을 통해 동일한 노선 내에서도 상행과 하행을 효과적으로 분리할 수 있었으며, 여기에 더해 도착지에서 장시간 체류하지 않고 곧바로 출발지로 복귀하는 경우에는 왕복 노선으로 인식되어 별개의 군집으로 분류되는 결과를 도출할 수 있었다.

4. 군집화 실험 방법 및 결과

본 시스템의 유효성을 검증하기 위한 실험은 Python 환경에서 Valhalla, H3-py, Scikit-learn 라이브러리를 활용하여 수행되었으며, PostGIS가 설치된 PostgreSQL 데이터베이스와 연동되었다. 실험에 사용된 데이터는 익명의 단일 운수사에 소속된 35개 차량의 2025년 10월 1일부터 10일까지의 주행 데이터이다. 수집된 개별 차량의 일일 주행 데이터는 총 276건이다.

H3 격자 기반 분할 작업을 거친 결과, 단일 주행 데이터는 983건이 도출되었다. 해당 경로들의 1초 당 수집되는 원본 GPS 데이터는 총 8,886,570개이며, 경로 하나 당 개수는 약 9,040.25개이다.

여기에 H3 격자 방문 이벤트 당 하나의 데이터로 압축한 결과 총 GPS 데이터 71,303개, 경로 하나 당 약 72.54개로 데이터 수를 축소하였다.

4.1 정량적 평가: 군집 타당성 지표 분석

군집화는 정답 레이블이 없는 비지도 학습이므로, 내부 평가 지표를 통해 성능을 검증하였다. Table 1은 각기 다른 Eps 값을 가지는 모델들이 얼마의 Silhouette score를 기록했고 데이터를 몇 개의 군집으로 분류했는지를 나타낸다. 1차 군집화 결과, Eps 0.025에서 최대의 Silhouette score 0.8912를 기록했으며 해당 세팅의 모델은 데이터를 18개 군집으로 분류하였다.

실루엣 점수는 -1에서 1 사이의 값을 가지며, 1에 가까울수록 군집 내 거리는 가깝고 군집 간 거리는 멀다는 것을 의미한다. 0.8912라는 수치는 본 시스템이 경로들을 매우 명확하고 밀집된 형태로 분리했음을 시사한다.

Fig. 13은 위에서 선택한 모델을 통해 1차 군집화를 수행하여 분류된 18개 군집을 나타낸다. 군집 당 하나의 경로씩 지도에 표시했다.

Table 1 Changes in silhouette score according to eps

ϵ (eps)	Clusters	Silhouette Score	Notes
0.005	93	0.4119	Excessive segmentation
0.010	33	0.7789	
0.015	24	0.8359	
0.020	20	0.8689	
0.025	18	0.8912	Selected as the optimal model
0.030	15	0.7788	Cluster merging occurs
0.035	11	0.7759	Under-clustering

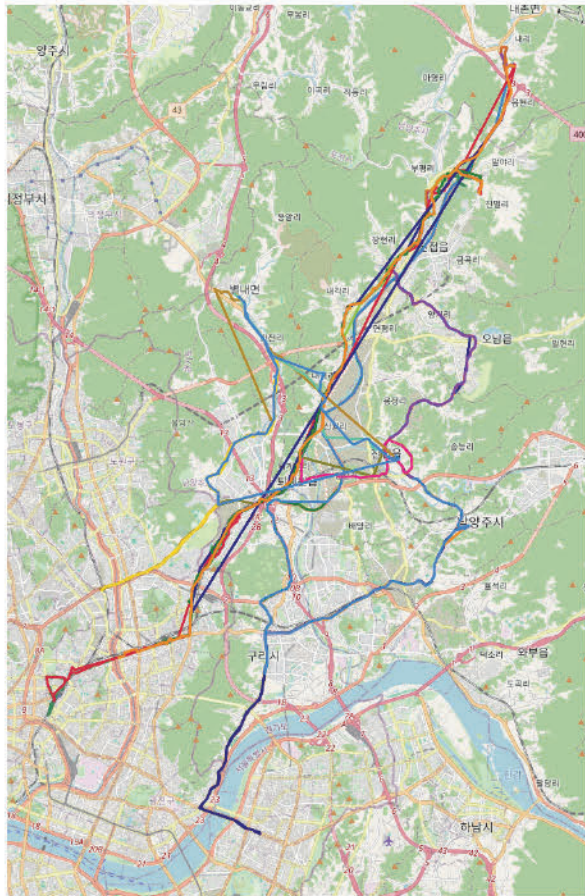


Fig. 13 The first clustering results were classified into 18 clusters. One route was displayed for each cluster

4.2 정성적 평가 및 시각화

이후 각 군집에 대해 2차 군집화를 거쳐 경로 데이터들을 더 세밀하게 분류하였다. Fig. 14 및 Fig. 15는 1차 군집화에서 10번째 군집으로 분류된 경로들을 대상으로

2차 군집화를 수행한 결과를 지도 상에 시각화한 것이다. 경로 상의 마커는 붉은색, 주황색, 녹색, 푸른색, 보라색 순서대로 차량의 이동 방향을 나타낸다.

Fig. 14의 북쪽 차고지에서 출발하여 남쪽 회차지를 돌아오는 경로(상행)와, Fig. 15의 남쪽에서 출발하여 북쪽으로 복귀하는 경로(하행)가 기하학적으로는 거의 겹쳐 있음에도 불구하고, 2차 군집화를 통해 서로 다른 하위 군집으로 완벽하게 분리되었다.

이는 동일 노선이라도 상행/하행의 도로 경사도(오르막/내리막)가 다를 수 있음을 감안할 때, 연비 분석의 정밀도를 획기적으로 높일 수 있는 기반이 된다. 예를 들어, 오르막 구간이 많은 상행선의 연비가 하행선보다 나쁘게 나오는 것은 자연스러운 현상이므로, 이를 구분하지 않고 평균을 내면 데이터의 왜곡이 발생한다. 본 시스템은 이러한 왜곡을 원천 차단한다.

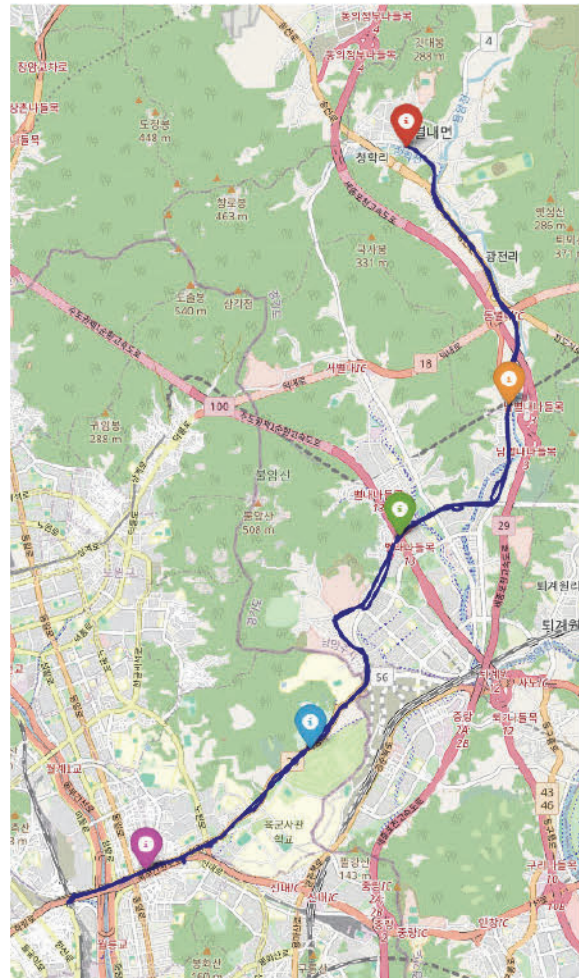


Fig. 14 The first subcluster classified by applying secondary clustering to the 10th cluster. This is the route of vehicles moving from north to south

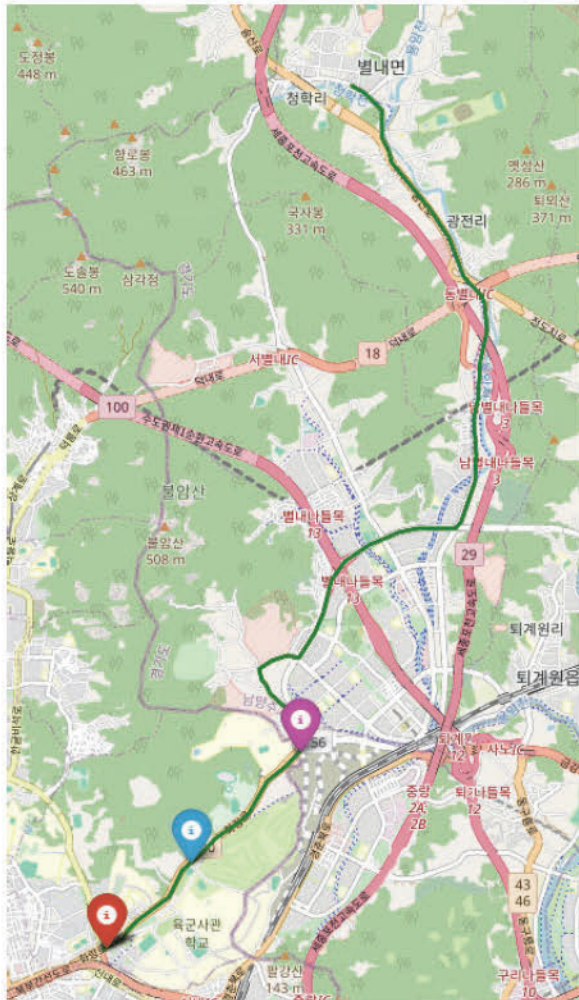


Fig. 15 The second subcluster classified by applying secondary clustering to the 10th cluster. This is the route of vehicles moving from south to north

5. 논의: 응용 분야 및 시사점

본 연구에서 구축한 경로 군집화 시스템은 단순한 데이터 정리를 넘어, 교통 운영 관리와 정책 수립에 있어 다음과 같은 심도 있는 응용 가치를 제공한다.

5.1 공정한 운전자 평가 모델

운전자의 안전 운전 습관을 평가할 때, ‘운행 환경’이라는 맥락을 고려하는 것은 필수적이다. 본 시스템을 통해 동일 노선을 운행한 운전자 그룹을 식별할 수 있다. 이를 바탕으로 특정 운전자의 급가속 횟수가 단순히 높은 것이 아니라, 해당 노선을 운행한 다른 운전자들의 평균보다 상위 몇 %인지를 평가하는 상대 평가 모델로 전환할 수 있다. 이는 운전자들의 수용성을 높이고, 공정한 인센티브 시스템을 구축하는 데 기여한다.

5.2 에너지 효율 및 탄소 배출량 분석의 고도화

차량의 탄소 배출량은 도로의 구배와 교통 상황에 민감하다. 37톤 트럭의 경우, 브레이크 열손실로 인해 소비되는 에너지가 전체의 50%에 달한다는 연구 결과도 있다.¹⁹⁾ 본 시스템을 통해 군집별(노선별)로 표준 연비 프로파일을 생성할 수 있다.

1) 이상징후 탐지: 특정 차량의 연비가 소속 군집의 평균 범위를 크게 벗어난다면, 이는 차량의 정비 불량(엔진, 타이어 공기압 등)이나 운전자의 비효율적인 운전 습관을 의심해 볼 수 있는 강력한 시그널이 된다.

2) 전기차 전환 전략: 각 노선의 에너지 소모량 분포를 분석하여, 해당 노선에 투입될 전기 버스의 적정 배터리 용량을 산출하거나, 충전 인프라의 최적 위치를 선정하는데 활용할 수 있다.

5.3 이상 경로 탐지를 통한 운영 효율화

이상치로 분류된 데이터들은 단순한 노이즈일 수도 있지만, 실제 운영상의 문제를 암시할 수도 있다. 운전자가 임의로 경로를 이탈하여 사적인 용무를 보았거나, 도로 공사로 인해 장기간 우회 운행이 발생하고 있음에도 노선 정보가 업데이트되지 않았을 수 있다. 이러한 이상 경로를 자동으로 탐지하고 관리자에게 알림을 제공함으로써, 운수사는 노선 운영의 투명성을 확보하고 불필요한 연료 낭비를 줄일 수 있다.

5.4 개선 사항 및 향후 발전 방향

그러나 본 연구는 여전히 발전의 여지를 남겨두고 있다. 향후 연구에서는 다음과 같은 과제들이 다루어져야 한다.

1) 딥러닝 기반 경로 임베딩(Deep trajectory embedding): 현재의 Hausdorff 거리 계산 방식은 여전히 $O(N^2)$ 의 연산 복잡도를 가진다. 이를 개선하기 위해 Autoencoder나 t-SNE와 같은 딥러닝 기법을 활용하여 경로를 고정된 크기의 잠재 벡터(Latent vector)로 변환하고, 벡터 공간에서 고속 군집화를 수행하는 연구가 필요하다.²⁰⁾

2) 시간적 역동성(Temporal dynamics)의 반영: 현재 시스템은 공간적 경로만을 기준으로 군집화한다. 하지만 동일한 노선이라도 출퇴근 시간대와 심야 시간대의 주행 패턴(소요 시간, 정차 빈도)은 판이하게 다르다. 향후에는 시간적 변수를 군집화의 축으로 추가하여, 시공간(Spatio-temporal) 군집화 모델로 확장함으로써 교통 혼잡 패턴까지 반영한 정교한 분석을 수행해야 한다.²¹⁾

3) 다양한 차종 및 비정형 경로로의 확장: 버스뿐만 아

나라 택시, 화물차 등 노선이 고정되지 않은(Free-floating) 차량에 대해서도 본 시스템을 적용하여, 유사한 통행 패턴을 보이는 이동 수요(OD)를 발굴하고 승차 공유(Ride-sharing)나 물류 최적화에 활용하는 방안을 모색할 수 있다.²²⁾

6. 결론

본 연구는 GPS 기반의 대규모 차량 주행 데이터를 효과적으로 분석하기 위한 계층적 경로 군집화 시스템을 제안하였다. Valhalla를 이용한 정밀 맵 매칭, H3를 활용한 효율적인 차원 축소, 그리고 Hausdorff 거리와 방향 벡터를 결합한 2단계 군집화 알고리즘은 실제 버스 운행 데이터에 적용되어 높은 군집 타당성(Silhouette score 0.8912)과 실용성을 입증하였다. 특히 람다 아키텍처를 도입하여 시스템의 확장성을 확보하고, 상/하행 노선의 분리와 같은 실무적 요구사항을 충족시킨 점은 본 연구의 주요한 기여라 할 수 있다.

본 시스템은 운전자 평가 모델 구축, 노선별 탄소 배출량 분석, 노선 운영 효율화 등의 실무 분야에 적용될 수 있다. 또한 딥러닝 기반의 고속 클러스터링, 시간적 동적 특성, 비정형 경로 등의 과제에 대해 추가 연구를 진행하면 일반성 및 성능 면에서 더 향상된 시스템을 구축할 수 있을 것으로 기대된다.

결론적으로, 본 시스템은 데이터 기반의 교통 운영 관리와 지속 가능한 모빌리티 생태계 구축을 위한 핵심적인 기반 기술로서, 향후 스마트 시티와 자율주행 관제 시스템의 중요한 구성 요소로 자리 잡을 것으로 기대된다.

후 기

이 연구는 2025년도 산업통상자원부 및 한국산업기술기술평가원(KEIT) 연구비 지원에 의한 연구임(RS-2024-00506825).

References

- 1) J. Yoo, D. Kim, Y. Yoo, M. Eum, J. Kim, S. Lee and D. Baik, "Study on the Characteristics of Carbon Dioxide Emissions Factors from Passenger Cars," Transactions of KSAE, Vol.17, No.4, pp.10-15, 2009.
- 2) K. Han, "Optimized Speed and Gearshift Trajectories Planning for Autonomous Electric Vehicles," Transactions of KSAE, Vol.28, No.10, pp.669-676, 2020.
- 3) A. Pirayre, P. Michel, S. S. Rodriguez and A. Chasse, "Driving Behavior Identification and Real-World Fuel Consumption Estimation with Crowdsensing Data," IEEE Transactions on Intelligent Transportation Systems, Vol.23, No.10, pp.18378-18391, 2022.
- 4) S. Dutta, A. Da and B. K. Patra, "CLUSTMOSA: Clustering for GPS Trajectory Data Based on Multi-objective Simulated Annealing to Develop Mobility Application," Applied Soft Computing, Vol.130, Paper No.109655, 2022.
- 5) J. Lee, J. Han and K. Whang, "Trajectory Clustering: A Partition-and-group Framework," SIGMOD '07: Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data, pp.593-604, 2007.
- 6) S. Syed, Development of Map Aided GPS Algorithms for Vehicle Navigation in Urban Canyons, M.S. Thesis, University of Calgary, Calgary, Canada, 2005.
- 7) Y. Chang, E. Tanin, G. Cong, C. S. Jensen and J. Qi, "Trajectory Similarity Measurement: An Efficiency Perspective," Proceedings of the VLDB Endowment, Vol.17, No.9, pp.2293-2306, 2024.
- 8) K. Xu, R. Wang, Q. Gao and J. Tan, "A Novel Method for Vessel Trajectory Clustering Based on Three-Dimensional Triangulation Division Algorithm," IEEE Access, Vol.13, pp.51388-51405, 2025.
- 9) Map Matching API – Valhalla Docs, <https://valhalla.github.io/valhalla/api/map-matching/api-reference/>, 2025.
- 10) W. Y. Ochieng, M. Quddus and R. B. Noland, "Map-matching in Complex Urban Road Networks," Brazilian Journal of Cartography, Vol.55, No.2, pp.1-14, 2003.
- 11) M. Cossaboom, J. Georgy, T. Karamat and A. Noureldin, "Augmented Kalman Filter and Map Matching for 3D RISS/GPS Integration for Land Vehicles," International Journal of Navigation and Observation, Vol.2012, Paper No.576807, p.16, 2012.
- 12) Z. Zhang, G. Ni and Y. Xu, "Comparison of Trajectory Clustering Methods Based on K-means and DBSCAN," 2020 IEEE International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA), pp.557-561, 2020.
- 13) K. Xu, R. Wang, Q. Gao and J. Tan, "A Novel Method for Vessel Trajectory Clustering Based on Three-Dimensional Triangulation Division Algorithm," IEEE Access, Vol.13, pp.51388-51405, 2025.

- 2025.
- 14) C. Nitoiu, C. Corafu and M. Popescu, "Influence of Road and Traffic Conditions on Emissions and Fuel Consumption of Light Vehicles in a Real Urban Driving Cycle," *Environmental Science and Pollution Research*, Vol.32, pp.15150-15165, 2025.
 - 15) A. Capiello, I. Chabini, E. K. Nam, A. Luè and M. Abou Zeid, "A Statistical Model of Vehicle Emissions and Fuel Consumption," *The IEEE 5th International Conference on Intelligent Transportation Systems*, pp.801-809, 2002.
 - 16) U.S. Department of Transportation, "Safety Measurement System (SMS) Methodology," <https://csa.fmcsa.dot.gov/documents/smsmethodology.pdf>, 2025.
 - 17) Map Matching API – Valhalla Docs, <https://valhalla.github.io/valhalla/api/map-matching/api-reference/>, 2025.
 - 18) J. P. Figueira, "Map-Matching for Trajectory Prediction," *Medium*, <http://medium.com/data-sci> ence/map-matching-for-trajectory-prediction-be307a1547f0, 2025.
 - 19) H. Ferreira, C. M. Rodrigues and C. Pinho, "Impact of Road Geometry on Vehicle Energy Consumption and CO2 Emissions: An Energy-efficiency Rating Methodology," *Energies*, Vol.13, No.1, Paper No.119, 2019.
 - 20) D. Lee and J. Kim, "Development of HTC-DBSCAN: A Hierarchical Trajectory Clustering Algorithm with Automated Parameter Tuning," *Applied Sciences*, Vol.14, No.23, Paper No.10995, 2024.
 - 21) T. Coelho da Silva, K. Zeitouni and J. Macedo, "Online Clustering of Trajectory Data Stream," *2016 17th IEEE International Conference on Mobile Data Management*, pp.112-121, 2016.
 - 22) X. Qin, Z. Li, K. Zhang, F. Mao and X. Jin, "Vehicle Trajectory Prediction via Urban Network Modeling," *Sensors*, Vol.23, No.10, Paper No.4893, 2023.