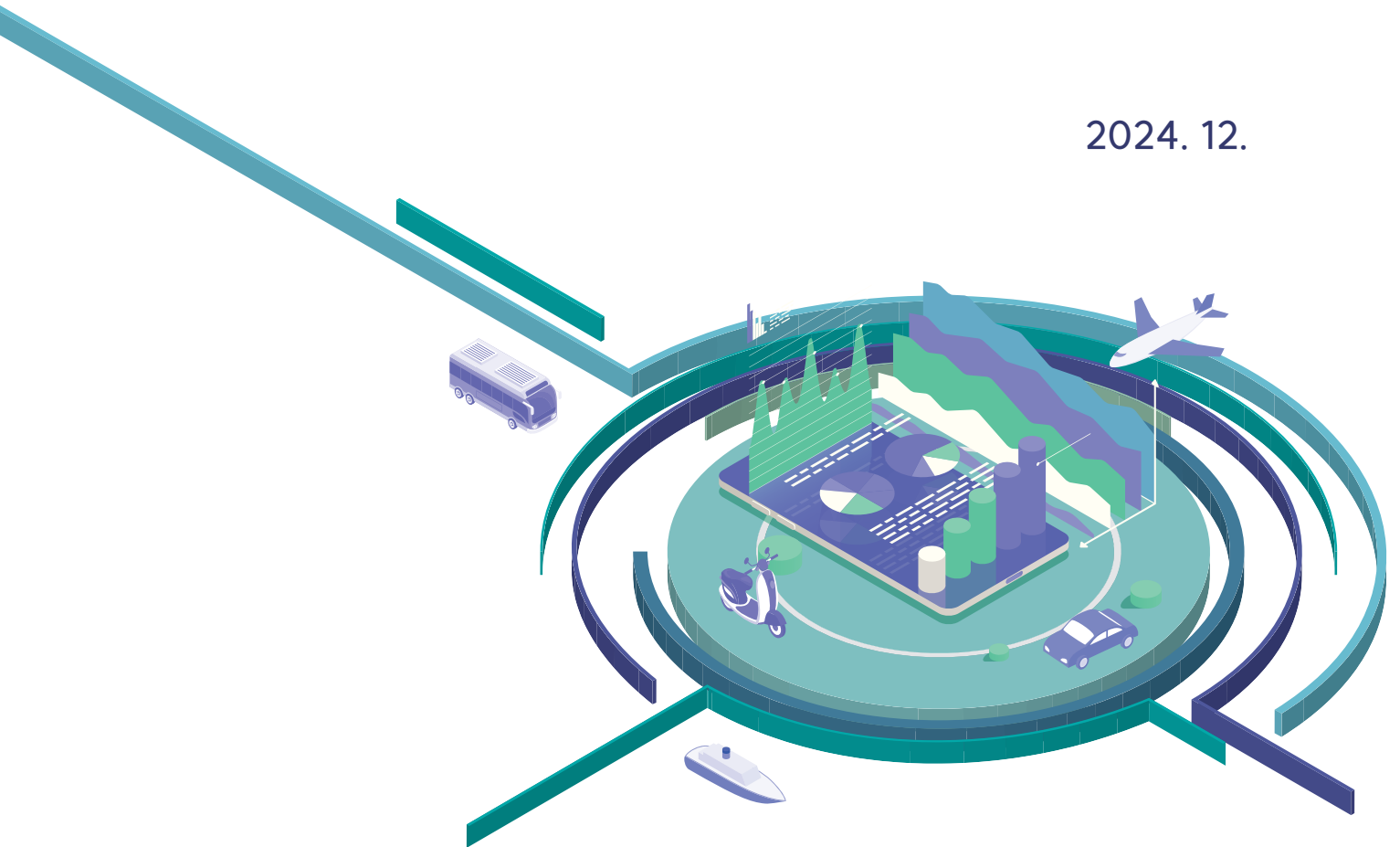


2024년 국가교통조사 및 분석

차량 GPS 빅데이터 구축

10

2024. 12.



국토교통부



한국교통연구원
KOREA TRANSPORT INSTITUTE

제 출 문

국토교통부 장관 귀하

본 보고서를 “2024년 국가교통조사 및 분석”의
최종보고서로 제출합니다.

2024년 12월

한국교통연구원

원장 김 영 찬

본 『2024년 국가교통조사 및 분석』은 다음 연구진에 의해
수행되었습니다.

참여연구진

<한국교통연구원>	
연구책임자	◦ 김주영 선임연구위원
연구진	◦ 조범철, 김주영 선임연구위원 ◦ 천승훈, 박용일 연구위원 ◦ 황순연, 장동익, 원민수, 이송봉, 이종우 부연구위원 ◦ 김동호, 신영권 책임전문원 ◦ 김규진 주임전문원 ◦ 가보연, 강국수, 광명신, 권순옥, 김관용, 김수아, 김승주, 김호용, 박미란, 박준호, 오연선, 이동엽, 이새봄, 이재영, 이해선, 정정호, 채운혁, 홍성표 연구원 ◦ 조지윤 연구조원
<한국해양수산개발원>	
연구진	◦ 이호춘, 최건우, 황수진 부연구위원 ◦ 류희영 전문연구원 ◦ 박일란 선임사무원
<한국항공협회>	
연구진	◦ 성인영 책임연구원 ◦ 최인영, 김지한, 김창욱, 김진성, 박다영 연구원

『2024년 국가교통조사 및 분석』

보고서 구성 및 담당연구진

번호	과제명	연구진
제 1권	요약보고서	김주영, 신영권, 가보연
제 2권	전국 여객 기종점통행량 보완갱신	김동호, 김관용, 강국수, 박미란
제 3권	주말 및 침두·비침두 기종점통행량 시범 구축	김동호, 채운혁
제 4권	교통분석용 네트워크 구축	김동호, 김관용, 이동엽, 이새봄
제 5권	항공여객 O/D 조사	한국항공협회
제 6권	전국화물 기종점통행량 보완갱신	원민수, 오연선, 이해선
제 7권	물류거점 화물실태조사	황순연, 김승주, 김호용
제 8권	연안화물 기종점통행량 구축	한국해양수산개발원
제 9권	KTDB 모빌리티 기반지도 구축	천승훈, 정정호
제10권	차량 GPS 빅데이터 구축	이승봉, 이채영
제11권	모바일통신 빅데이터 구축	조범철, 이종우, 곽명신
제12권	국가교통통계DB구축	박용일, 김수아
제13권	특별교통대책기간 통행실태조사	김주영, 김관용, 권순욱
제14권	교통접근성지표 구축	장동익, 박준호

『2024년 국가교통조사 및 분석』
과제별 공동참여·위탁용역 사업자

【공동사업 참여기관】

- 전국 여객O/D 현행화 공동사업(수도권 부문)
 - 경기연구원, 인천연구원, 서울연구원

- 항공O/D 및 특성 조사
 - (사)한국항공협회

【위탁용역 사업자】

- 전국여객 O/D 현행화(대구광역권)
 - 홍익대학교 산학협력단

- 전국여객 O/D 현행화(대전세종충청권)
 - 신명이앤씨(주)

- 전국여객 O/D 현행화(제주권)
 - 홍익대학교 산학협력단

- 주말 및 첨두·비첨두 O/D 기초데이터 구축 및 분석
 - (주)모비크리에티브

- 교통부문 네트워크 갱신을 위한 GIS 기반 교통망 기초자료 구축
 - 서울시립대학교 산학협력단

- 물류거점화물실태조사
 - (주)코리아데이터네트워크

【위탁용역 사업자】

- 화물 기종점 통행량 보완·갱신을 위한 교통물류 실증 데이터 수집·가공·전처리 용역
- ㈜노트스퀘어
- 모빌리티 빅데이터를 활용한 KTDB 기반지도 및 차량통행 데이터셋 구축
- ㈜큐빅웨어, ㈜유아이네트웍스
- 특별교통대책기간 통행실태조사
- (주)컨슈머인사이트
- 대중교통 GTFS기반 네트워크 구축
- ㈜아로정보기술
- Open-source 기반의 교통접근성 산정 최적화 방법론 연구
- (주)아로정보기술

최종보고서 목차

- 제 1권 요약보고서
- 제 2권 전국 여객 기종점통행량 보완갱신
- 제 3권 주말 및 첨두·비첨두 기종점통행량 시범 구축
- 제 4권 교통분석용 네트워크 구축
- 제 5권 항공여객 O/D 조사
- 제 6권 전국화물 기종점통행량 보완갱신
- 제 7권 물류거점 화물실태조사
- 제 8권 연안화물 기종점통행량 구축
- 제 9권 KTDB 모빌리티 기반지도 구축
- 제 10권 차량 GPS 빅데이터 구축
- 제 11권 모바일통신 빅데이터 구축
- 제 12권 국가교통통계DB구축
- 제 13권 특별교통대책기간 통행실태조사
- 제 14권 교통접근성지표 구축

● 목 차

요약

제1장 과업의 개요	3
제1절 과업의 배경 및 목적	3
제2절 과업의 범위 및 내용	5
제2장 차량 GPS 빅데이터 수집 및 분석	8
제1절 차량 GPS 데이터 수집 개요	8
제2절 차량 GPS 원시데이터 수집 및 분석	17
제3장 차량 GPS 데이터 가공 및 DB구축	27
제1절 차량 DB 가공 및 구축 프로세스	27
제2절 차량 GPS 데이터 전처리 및 경로 가공	29
제3절 기초 교통 DB 구축	48
제4절 차량 통행지표 구축	62
제4장 차량 기반 분석·활용시스템 개선	79
제1절 차량 기반 분석·활용 시스템 개요	79
제2절 기능 개선	81
제5장 결론 및 향후계획	87



● 표목차

〈표 2-1〉 차량 GPS 데이터 특징	9
〈표 2-2〉 데이터 수집 목록 및 기준연도	10
〈표 2-3〉 수집 데이터의 시간적 범위	10
〈표 2-4〉 링크 기반 내비게이션 데이터 형식	12
〈표 2-5〉 포인트 기반 내비게이션 데이터 형식	13
〈표 2-6〉 내비게이션 데이터 차종 정보	13
〈표 2-7〉 내비게이션 데이터 유종 정보	14
〈표 2-8〉 DTG 데이터의 포맷	16
〈표 2-9〉 DTG 데이터의 자동차 유형 코드	16
〈표 2-10〉 데이터의 용량 및 개별 차량의 링크 발생 건수	17
〈표 2-11〉 티맵 내비게이션 데이터 차종 분석 결과	18
〈표 2-12〉 티맵 내비게이션 데이터 유종 분석 결과	18
〈표 2-13〉 연도별 내비게이션 데이터 프로파일링 비교	19
〈표 2-15〉 DTG 데이터 현황	20
〈표 3-1〉 개별차량 경로DB 테이블 정의서	35
〈표 3-2〉 DTG 데이터 오류 발생 필터링 비율	37
〈표 3-3〉 DTG 전처리 데이터 DB	37
〈표 3-4〉 DTG 필터링 및 보정 데이터 표준 DB	40
〈표 3-5〉 티맵 상세도로망 기반 경로 데이터 표준 DB	43
〈표 3-6〉 DTG 상세도로망 기반 경로 데이터 표준 DB	43
〈표 3-7〉 경로 데이터 압축 및 통합 경로 데이터 NoSQL DB 포맷	45
〈표 3-8〉 속도 프로파일 DB 포맷	48
〈표 3-9〉 속도 프로파일 스무싱 DB 형태	49
〈표 3-10〉 연도별 도로등급별 평균속도	49
〈표 3-11〉 연도별 시도별 평균속도	50
〈표 3-12〉 기종점 통행량 DB 포맷	52
〈표 3-13〉 교통량 추정을 위한 입력 데이터 DB	55
〈표 3-14〉 교통량 추정을 위한 입력 데이터 포맷	58
〈표 3-15〉 교통망 성능평가지표 DB 생성 기준	60
〈표 3-16〉 차량 교통망 성능평가지표	61
〈표 3-17〉 머신러닝 기반 추정교통량 모형 구축 프로세스	64
〈표 3-18〉 추정교통량 DB 테이블 구성	65
〈표 3-19〉 도로등급별 추정교통량 추이	68
〈표 3-20〉 차종별 총 주행거리(관용+자가용 기준)	69
〈표 3-21〉 이산화탄소 배출량의 배출계수	70
〈표 3-22〉 일산화탄소 배출량의 배출계수	71

〈표 3-23〉 휘발성 유기화합물 배출량의 배출계수	71
〈표 3-24〉 질소산화물 배출량의 배출계수	72
〈표 3-25〉 미세먼지 배출량의 배출계수	72
〈표 4-1〉 데이터 수집 현황 로그 항목	77
〈표 4-2〉 데이터 전처리 로그 항목	77
〈표 5-4〉 데이터 제공 및 분석 지원 주요 사례	85
〈표 5-5〉 데이터 분석·활용 시스템 활용 실적	86

● 그림목차

〈그림 2-1〉 티맵 네트워크와 상세도로망 네트워크 상세도 비교	16
〈그림 2-2〉 시도별 연평균 일 내비 프로브 샘플율	22
〈그림 3-1〉 차량 GPS 데이터 가공 및 구축 프로세스	26
〈그림 3-2〉 링크 기반 데이터 전처리 및 경로 데이터 가공 프로세스	27
〈그림 3-3〉 포인트 기반 데이터 전처리 및 경로 데이터 가공 프로세스	27
〈그림 3-4〉 차량 GPS 데이터를 활용한 개별차량 경로DB 구축 프로세스	28
〈그림 3-5〉 데이터 적재	29
〈그림 3-6〉 자료형 변환	29
〈그림 3-7〉 데이터 포맷 변환	29
〈그림 3-8〉 주행 시작일 필드와 링크 진입시간 필드 날짜 불일치 오류 예시	30
〈그림 3-9〉 링크 진입시간 필드와 링크 진출시간 필드 시간 차이 오류 예시	30
〈그림 3-10〉 링크 중복 경로 예시	30
〈그림 3-11〉 DTG 데이터 오류 검토	31
〈그림 3-12〉 이동궤적(포인트) 데이터 보정	32
〈그림 3-13〉 통행 분리되지 않은 차종별 궤적 예시	33
〈그림 3-14〉 포인트-링크 맵매칭	34
〈그림 3-15〉 포인트와 도로망 네트워크 맵 매칭 예시	34
〈그림 3-16〉 경로 분할 및 통합 기준	35
〈그림 3-17〉 포인트 정보와 매칭된 링크의 통행속도 산출	36
〈그림 3-18〉 좌표오류 및 시간오류 필터링 방법론	37
〈그림 3-19〉 통행분리 유형 판단을 위한 데이터 분석	39
〈그림 3-20〉 지하차도 음영구간의 데이터 재생성	40
〈그림 3-21〉 터널 음영구간의 데이터 재생성	41
〈그림 3-22〉 정지 시 꼬임구간의 데이터 재생성	41
〈그림 3-23〉 맵매칭 프로세스	42
〈그림 3-24〉 링크 매칭 정보를 통한 통행속도 산출	43
〈그림 3-25〉 통행병합 예시1	43
〈그림 3-26〉 통행병합 예시2	43
〈그림 3-27〉 압축된 경로 데이터 형태	45
〈그림 3-28〉 기초교통 DB 구축 프로세스	47
〈그림 3-29〉 속도 프로파일 구축 프로세스	48
〈그림 3-30〉 속도 프로파일 스무싱 예시	49
〈그림 3-31〉 기종점 통행량 DB 구축 프로세스	52
〈그림 3-32〉 기점과 종점의 시군구에 따른 집계 방식 차이	52
〈그림 3-33〉 기점과 종점의 시군구에 따른 집계 방식 차이	54
〈그림 3-34〉 교통량 추정 입력 데이터 구성	55

〈그림 3-35〉 교통량 추정을 위한 데이터 구성	58
〈그림 3-36〉 네트워크 상의 시공간적 연결성 변수 형태 예시	58
〈그림 3-37〉 교통량 추정 및 구축 프로세스	61
〈그림 3-38〉 도로등급별 신뢰도 분석 결과	65
〈그림 3-39〉 교통량 수준별 신뢰도 분석 결과	65
〈그림 3-40〉 교통량 추정 및 차량 통행지표 구축 프로세스	66
〈그림 4-1〉 모니터링 프로세스	77
〈그림 4-2〉 모니터링 메뉴 구조도	78
〈그림 4-3〉 데이터 수집 현황 모니터링 화면	79
〈그림 4-4〉 데이터 전처리 모니터링 화면	79
〈그림 4-5〉 기초 데이터 모니터링 화면	80



후
야

1. 과업의 개요

가. 과업의 배경 및 목적

1) 과업의 배경

- 전국 단위 공공의 기초 교통데이터 수집 및 지점 데이터 분석의 한계
 - 교통 분야의 기초 데이터인 교통량, 속도 데이터는 한국건설기술연구원, 한국도로공사, 광역지자체 등에서 조사하고 있으나, 데이터의 공간적 수집 범위가 매우 낮아 전국의 교통 현황을 분석하기에 한계가 있음
 - 공공에서 수집하고 있는 데이터는 지점 정보 중심의 데이터로 단편적 정보만 확인할 수 있어 시공간적으로 연결된 교통의 흐름을 파악하기엔 한계가 있음
- 전국 단위 일관된 교통DB 및 통행지표 부재
 - 데이터 수집기관, 수집방식 등 기관별로 상이하게 수집되는 데이터 수집 체계는 전국을 일관된 기준으로 분석하기엔 한계가 있음
- 빅데이터 전처리·가공 환경 및 빅데이터 기반 차량통행 분석시스템 부재
 - 차량 GPS 빅데이터를 효율적으로 전처리·가공하기 위한 환경이 부족하며, 과거에는 차량의 시·공간적 통행패턴과 현황을 분석하기 위해 모형 중심의 프로그램을 이용하여 분석을 수행하였음
- 국가통합교통체계효율화법 시행령 제8조1)에 의거하여 교통물류활동으로 발생

하는 교통혼잡 등 교통 관련 사회적 외부비용을 국가교통물류에 관한 지표를 설정하여 고시하도록 규정하고 있음

2) 과업의 목적

- 국가통합교통체계효율화법²⁾에 의거 국가교통데이터베이스 구축 및 운영을 위한 기초교통DB 및 통행지표 생성
 - 개별 차량 기반의 교통량, 속도, 차량주행거리 등 교통분야의 핵심 기초교통 DB 구축(공공 조사교통량 약 1만 개 지점, 전국 도로 62만 개 링크)
 - 차량 이용 및 교통물류 활동으로 인한 교통혼잡, 환경오염, 온실가스 배출 등 다양한 측면의 도로 교통망 성능평가 지표 구축
 - 지자체별로 상이한 방식으로 수집되고 구축되는 교통DB·통행지표의 한계를 개선하기 위한 전국단위 일관된 교통DB·통행지표 구축 필요
 - 여객 O/D구축 및 교통수요예측을 위한 기반데이터(승용차 O/D 데이터 등) 제공
- 국가교통 데이터 제공 및 교통 분야 주요 정책 지원
 - 이용자의 요구에 대응할 수 있는 데이터 제공 및 분석지원
 - 탄소공간지도 사업, 환경소음측정망 선정 사업, 지자체 실증사업 등 주요 정책 제언을 위한 데이터 및 분석 지원(매년 요청 및 제공)

나. 과업의 범위

1) 시간적 범위

- 기준연도 : 2023년
- 데이터 수집 : 2023년, 2024년
- 데이터 가공 및 DB구축 : 2023년, 2024년 1-6월

1) 국가교통조사에는 다음 각 호 사항이 포함되어야 한다 - 교통물류활동으로 발생하는 교통혼잡, 교통사고, 환경오염, 온실가스 배출 등 교통 관련 사회적 비용, 그 밖에 교통 관련 정책 및 계획의 수립, 교통시설 투자분석 및 평가에 필요한 사항

2) 국가통합교통체계효율화법 제12조(국가교통조사), 제14조(정보통신수단 등을 통한 교통조사) 및 동법 시행령 제8조(국가교통조사의 실시)

2) 공간적 범위

- 전국 2차로 이상 도로 및 주요 도로³⁾

3) 내용적 범위

- 차량 원시 GPS 빅데이터 전처리 및 DB가공·구축
 - 차량 GPS 데이터, DTG 데이터 수집 및 검토
 - 원시 데이터 전처리 및 경로 데이터 가공
 - 기초 교통 DB구축 및 검증
- 교통량 추정 알고리즘 개선
 - 네트워크 연결성을 고려한 교통량 추정 알고리즘 개선
 - 기초 교통DB 구축을 위한 교통량 추정 신뢰도 분석 및 연도별 패턴 분석
- 차량 통행지표 가공 알고리즘 개선 및 DB구축
 - 통행지표 가공 알고리즘 개선 및 고도화
 - 목적별 전국 통행지표 산출 및 DB 구축
- 데이터 제공 및 분석지원

2. 차량 GPS 빅데이터 수집 및 분석

가. 차량 GPS 빅데이터 수집 개요

1) 데이터 수집 대상

- 본 과업의 기준연도는 2023년 기준으로 차량 GPS 빅데이터 기반 기초교통 및 교통망 성능평가지표DB를 구축하기 위해 다음과 같은 데이터를 수집함

3) 주요 도로 : 편도 1차로 도로 중 데이터 수집의 안정성이 확보된 도로

- 링크 기반 내비게이션 데이터는 티맵의 내비게이션 사용자의 주행 정보, 주행 정보를 도로구간(링크)별로 나타내는 도로 네트워크 데이터를 수집함
- 포인트 기반 내비게이션 데이터는 티맵의 내비게이션 사용자의 주행 정보, 주행 정보를 GPS 단위로 데이터를 수집하므로 초 단위 데이터를 수집함
- 포인트 기반 사업용 차량 운행기록계(DTG) 데이터는 한국교통안전공단에서 제공하며, 사업용 차량에 부착된 운행기록계로 수집되는 버스, 화물, 택시의 주행 정보 데이터를 수집함

〈표 1〉 데이터 수집 목록 및 기준연도

구분	데이터 목록	수집처	수집연도
링크 기반	내비게이션 도로 네트워크	티맵	2023년 1-12월, 2024년 1-6월
	내비게이션 주행 정보		
포인트 기반	사업용 차량 운행기록계 버스 주행 정보	한국교통안전공단	2023년 1-12월, 2024년 1-6월
	사업용 차량 운행기록계 화물 주행 정보		
	사업용 차량 운행기록계 택시 주행 정보		

- 시의성 있는 데이터 제공을 위해 당해연도인 2024년 반기 데이터를 하반기부터 수집하여 빠르게 가공 및 구축 가능한 지표를 산출함
 - 링크 기반 내비게이션 데이터는 2023년, 2024년 12주, 명절을 포함하여 월별 추세를 파악할 수 있는 기간으로 총 179일을 수집하며, 2024년까지 데이터를 수집하였음
 - 포인트 기반 내비게이션 데이터는 2024년 4주, 여름과 겨울 추세를 파악할 수 있는 기간으로 총 31일을 수집할 예정이며, 데이터 오류 발견으로 재요청함
 - 포인트 기반 내비게이션 데이터의 수집 시기는 데이터 수집기관의 서버 문제로 1월-5월까지 데이터가 삭제되었기 때문에 2024년 1월, 4월, 7월, 10월 데이터를 받기로 했으나, 수집 기간을 조정하였음

〈표 2〉 수집 데이터의 시간적 범위

구분	링크 기반 내비게이션		포인트 기반 내비게이션
	2023년	2024년	2024년
1월	25일 ~ 29일 (6일)	22일 ~ 28일 (7일)	-
2월	13일 ~ 19일 (7일)	13일 ~ 18일 (7일)	-
3월	13일 ~ 19일 (7일)	11일 ~ 17일 (7일)	-
4월	17일 ~ 23일 (7일)	15일 ~ 21일 (7일)	-

6월			17일 ~ 23일(7일)
7월			8일 ~ 14일(7일)

9월	20일, 22일 ~ 27일 (7일)	19일 ~ 25일 (7일)	-
10월	16일 ~ 22일 (7일)	14일 ~ 20일 (7일)	21일 ~ 27일(7일)
11월	13일 ~ 19일 (7일)	11일 ~ 17일 (7일)	-
12월	11일 ~ 17일 (7일)	16일 ~ 22일 (7일)	9일 ~ 15일(7일)
명절	설날 : 1월 21일 ~ 1월 24일 (4일)	설날 : 2월 09일 ~ 2월 12일 (4일)	추석 : 9월 16일 ~ 9월 18일 (3일)
	추석 : 9월 28일 ~ 9월 30일 (3일)	추석 : 9월 16일 ~ 9월 18일 (3일)	
합계	89일	90일	31일

2) 데이터 수집 대상의 특징 및 형식

- 링크 기반 내비게이션 데이터는 링크의 진입시간, 진출 시간, 속도 등의 정보가 포함되어 있으며, 차종과 유종, 목적지 정보가 포함되어 제공받음
- 포인트 기반 내비게이션 데이터는 GPS좌적 X좌표, Y좌표에 따른 주행 시간, 속도, 고도값 등의 정보가 있으며, 주행차량의 배기량과 연식을 포함하여 제공받아 링크 기반 내비게이션 데이터와 차이가 있음
- 사업용 운행기록계(DTG) 데이터는 사업용 차량의 운행정보를 실시간으로 저장하여 변화하는 차량운행 상황을 기록하는 디지털 운행기록계를 통해 수집되는 데이터로 자동차의 순간속도, GPS, 분당 엔진회전수, 가속도, 자동차 유형 등의 정보가 제공되며, 1초 단위의 GPS 좌표를 수집함

나. 차량 GPS 원시데이터 분석

- 2023년 티맵 내비게이션 데이터는 2022년 대비 약 12.8% 증가된 총 793억 건의 개별 차량 주행 발생 건수(이벤트 수)가 수집되었음
- DTG데이터의 개별 차량 발생 건수는 전년 대비 버스 약 0.2% 감소하였고, 화물차 약 11% 증가, 택시 48% 증가하였음
 - 택시 데이터의 추출 오류로 인해 5개월치 데이터를 다시 받았고, 사업자에서 제출한 데이터가 늘어난 만큼 개별 차량 발생 건수가 증가한 것으로 분석됨

〈표 3〉 데이터의 용량 및 개별 차량의 링크 발생 건수

데이터명	용량(TB)		이벤트 수(천 건)	
	2022년	2023년	2022년	2023년
링크 기반 내비게이션 데이터	19.4	22.3	70,349,051	79,364,038
DTG 버스 데이터	9.64	9.22	772,320,634	770,774,514
DTG 화물차 데이터	3.93	4.32	317,497,577	355,365,816
DTG 택시 데이터	5.91	10.24	491,244,500	953,355,667

- 티맵 내비게이션 데이터의 차종과 유종 정보는 티맵 전체 가입자가 아닌 자동차보험 할인을 위해 “운전습관 가입자”가 제공하는 정보이므로 100% 포함하지 않는 한계점이 존재함
- 티맵 내비게이션 데이터를 통한 차종 정보 분석 결과
 - 내비게이션 데이터의 주행 고유 ID별로 차종 정보가 포함된 비율을 분석한 결과, 75% 이상의 데이터에서 차종 정보가 확인되었음
 - 승용차가 약 68%로 가장 많이 발생하였고, 2022년과 비교하여 14% 증가한 것으로 분석됨

〈표 4〉 티맵 내비게이션 데이터 차종 분석 결과

차종 정보	2022년	2023년
승용차	55.2%	68.6%
경차	2.4%	3.0%
중형승합차	1.7%	1.8%
대형승합차	0.5%	0.6%
소형화물차	0.5%	0.8%
대형화물차	0.03%	0.1%
특수화물차	0.02%	0.3%
정보 없음	39.6%	24.9%
합계	100%	100%

○ 티맵 내비게이션 데이터를 통한 유종 정보 분석

- 내비게이션 데이터의 주행 고유 ID별로 유종 정보가 포함된 비율을 분석한 결과, 약 49% 이상의 데이터에서 유종 정보가 확인되었음
- 휘발유에서 약 51%로 가장 많이 발생하였고 2022년과 비교했을 때 10% 증가한 것으로 분석됨

〈표 5〉 티맵 내비게이션 데이터 유종 분석 결과

유종 정보	2022년	2023년
휘발유	41.4%	51.3%
경유	16.6%	20.1%
LPG	4.7%	5.6%
고급휘발유	1.1%	1.3%
전기	1.1%	1.5%
정보 없음	35.3%	20.3%
합계	100%	100%

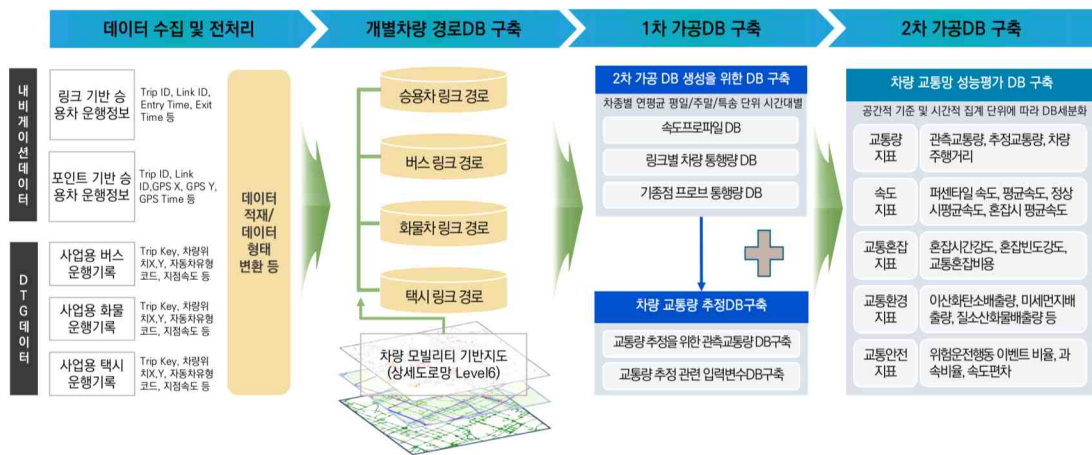
○ DTG데이터의 자동차 유형 정보를 이용하여 차종 분석

- DTG데이터의 약 32%는 버스, 약 49%는 택시, 약 19%는 화물차로 분석되었으며, 시내버스와 개인택시, 일반화물의 차종이 높은 비율로 분포하고 있음
- DTG데이터는 세분화된 차종을 포함하고 있으나, 교통량 추정할 때는 버스, 화물차에 대한 정보를 집계하여 활용함
- 이는 공공에서 수집하고 있는 관측교통량 데이터의 차종 구분이 세분화되지 않아 차종을 집계하여 활용함

3. 차량 GPS 데이터 가공 및 DB구축

가. 차량 DB가공 및 구축 방법론 설정

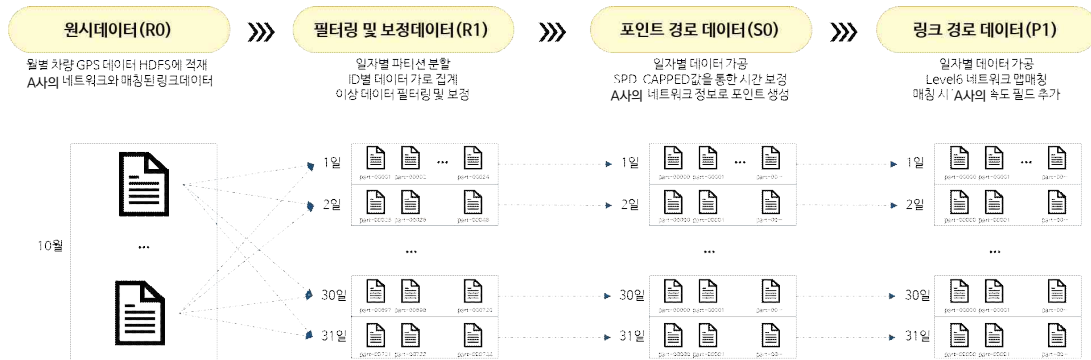
- 본 사업에서 활용하고 있는 차량 GPS 데이터는 내비게이션 및 사업용 DTG 데이터로, 이를 활용하기 위해서는 제공되는 데이터의 특성을 분석하여 각 특성에 맞춰 데이터 가공 및 검증이 필요함
- 티맵 내비게이션 데이터는 링크 기반 데이터이고, DTG 데이터는 포인트 기반 데이터로 각 데이터의 제공되는 데이터 형식 및 수집 주기가 다르기 때문에 특성에 맞는 전처리 및 가공 모듈 적용을 수행함
- 차량 GPS 데이터 가공 및 구축 프로세스는 데이터 전처리 → 개별차량 경로 DB구축 → 1차 가공DB(기초교통DB) 구축 → 2차 가공DB(교통망 성능평가지표DB) 구축 순으로 진행



〈그림 1〉 차량GPS 데이터 가공 및 구축 프로세스

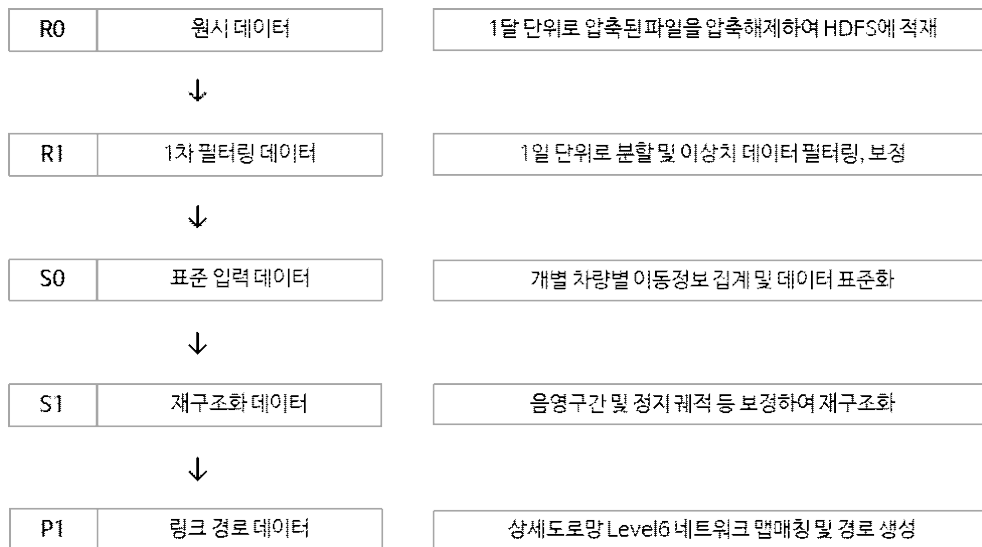
- 티맵 내비게이션 데이터는 링크 기반 데이터로 링크를 1초 단위 궤적(포인트) 정보로 생성하여 표준 포인트 경로 데이터를 구축한 뒤 Level6 네트워크와 맵 매칭하여 링크 경로 데이터를 생성함
- 티맵 네트워크의 링크는 Level6 네트워크보다 상세하므로 데이터의 공간적

범위에 맞게 구축하기 위해 궤적(포인트)정보를 생성함



〈그림 2〉 링크 기반 데이터 전처리 및 경로 데이터 가공 프로세스

- DTG 데이터는 포인트 기반 데이터로 포인트 기반 데이터 맵매칭 및 경로 가공 모듈을 이용하여 가공 및 구축함(해당 모듈은 R&D사업을 통해 개발된 것임)

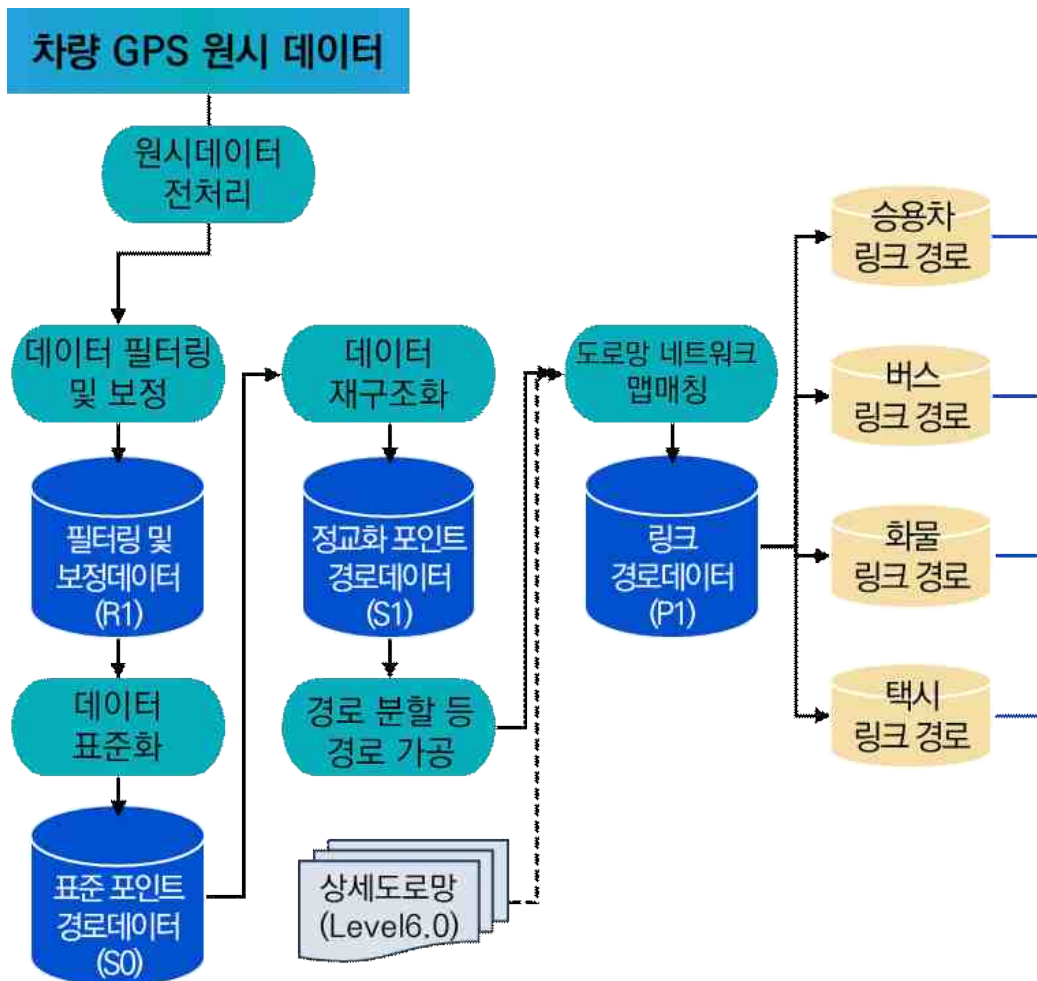


〈그림 3〉 포인트 기반 데이터 전처리 및 경로 데이터 가공 프로세스

나. 차량 GPS 데이터 전처리 및 경로 가공

1) 개별차량 경로DB 구축 프로세스

- 차량 GPS 데이터를 활용하여 개별차량 경로DB를 구축하는 프로세스는 다음과 같은 과정으로 수행됨
 - 링크 기반 데이터는 맵 정보와 매칭되어 있어 궤적정보를 정교화하는 데이터 재구조화 과정이 없음
 - 포인트 기반 데이터는 GPS 수신이 불안정한 궤적에 대한 보정을 수행하는 재구조화 과정이 포함됨



<그림 4> 차량 GPS 데이터를 활용한 개별차량 경로DB 구축 프로세스

2) 원시데이터 전처리

○ 원시데이터 적재 및 전처리

- 대용량의 압축 파일을 빅데이터 가공 시스템으로 처리하기 위해 원시 데이터를 분할하여 HDFS에 적재함



〈그림 5〉 데이터 적재

- 원시데이터를 일자별로 분리하고, 불필요한 속성정보를 제거하여 데이터 용량을 축소시킴



〈그림 6〉 자료형 변환

- 빅데이터 환경의 처리속도 향상을 위해 데이터 포맷을 변환



〈그림 7〉 데이터 포맷 변환

○ 내비게이션 데이터 필터링 및 보정

- 아래의 오류 유형에 따라 필터링 및 보정을 수행함

- (시간 오류) 주행 시작일 필드 오류 유형 : 주행 시작일(R_P_REQDATE)과 링크 진입 시간(ENTRY_TIME) 필드의 일자가 맞지 않는 경우 링크 진입시간을 기준으로 주행 일자에 대해 정렬하여 일자별로 분할함

DATE	TRIP_INIT_SESSION_ID	LINK_SEQ	MESH_CD	LINK_ID	LINK_DIR	LENGTH	DRIVING_TIME	ENTRY_TIME	EXIT_TIME
20210208	105520210208235734644500	1	4787	16033	0	63.894627	6.790106	20210209000226	20210209000232
20210208	105720210208235718506500	1	5631	114	0	121	44.565998	20210209000653	20210209000737
20210208	108120210208232709308500	1	4777	3744	1	26	2.8363636	20210209073336	20210209073339
20210209	105020210209175213265501	1	7413	28853	1	84	27.49091	20210210030545	20210210030559
20210209	105020210209210128425500	1	4775	9789	1	231.51741	11.883713	20210210084021	20210210084032

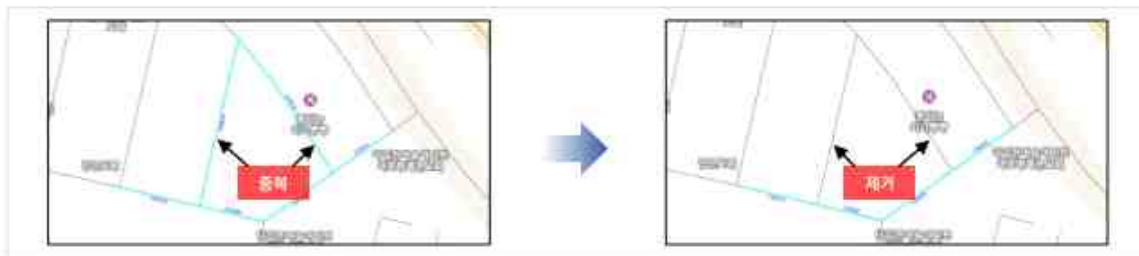
<그림 8> 주행 시작일 필드와 링크 진입시간 필드 날짜 불일치 오류 예시

- (시간 오류) 링크 시퀀스 필드 오류 유형 : 링크 진입시간(ENTRY_TIME)과 링크 진출 시간(EXIT_TIME) 필드의 시간이 링크 시퀀스(LINK_SEQ) 순서와 다를 경우 진입시간, 진출시간, 링크 시퀀스 순서로 정렬함

DATE	TRIP_INIT_SESSION_ID	LINK_SEQ	MESH_CD	LINK_ID	LINK_DIR	LENGTH	DRIVING_TIME	ENTRY_TIME	EXIT_TIME
20210208	105520210208235734644500	1	4787	16033	0	63.894627	6.790106	20210209000226	20210209000232
20210208	105720210208235718506500	1	5631	114	0	121	44.565998	20210209000653	20210209000737
20210208	108120210208232709308500	1	4777	3744	1	26	2.8363636	20210209073336	20210209073339
20210209	105020210209175213265501	1	7413	28853	1	84	27.49091	20210210030545	20210210030559
20210209	105020210209210128425500	1	4775	9789	1	231.51741	11.883713	20210210084021	20210210084032

<그림 9> 링크 진입시간 필드와 링크 진출시간 필드 시간 차이 오류 예시

- (필드 오류) 링크 중복 오류 유형 : 동일 시간에서 발생한 링크의 중복 경로를 보정함



<그림 10> 링크 중복 경로 예시

- (속도 오류) 속도 필드 오류 유형 : 속도 제한이 없는 상태에서 속도를 계산 시 이상값이 나타나는 경우(GPS가 미존재하여 실제 속도 값이 계산되지 않는 경우가 있어 제한이 걸린 속도를 사용함에도 200km/h 이상이 나타난 속도값에 대해 필터링함

○ 필터링 및 보정 데이터를 표준DB 형태로 변환 및 저장

3) 경로 데이터 구축 및 경로 압축(통합)

○ 맵매칭된 링크 경로 데이터를 통해 링크 통행속도 산출

- 링크와 매칭된 포인트 정보의 도로구간 길이와 시간정보를 이용하여 링크별 통행속도를 산출함

〈그림 11〉 포인트 정보와 매칭된 링크의 통행속도 산출

○ 개별 차량 경로DB 구축

- 링크와 매칭된 경로 데이터를 통행지표 생성을 위한 기초 DB로 활용됨
 - 단말기 ID별로 1일 단위의 개별 차량의 경로 데이터 구축

〈표 6〉 개별차량 경로DB 테이블 정의서

컬럼명	설명
단말기 ID	단말기 가상키로 개별 차량을 의미하며, 1일 단위로 단말기ID가 변경됨
경로그룹번호	통행이 발생한 이동 경로를 그룹화하였으며, 5분 이상 정차 시 그룹분리, 경로병합 적용
차종	Auto : 티맵 내비게이션 Bus : DTG 버스 Truck : DTG 화물 Taxi : DTG 택시
순서	경로 그룹 내 이동이 발생한 순서
날짜	차량의 이동궤적이발생한 날짜 및 시간
링크ID	상세도로망 네트워크의 링크ID
진입시간	링크 내 차량의 이동궤적이발생한 처음 시간이며, 시분초로구성됨
진출시간	링크 내 차량의 이동궤적이발생한 끝 시간이며, 시분초로구성됨
링크길이	상세도로망 네트워크의 링크 길이
링크속도	상세도로망 네트워크의 링크 길이로 계산된 속도

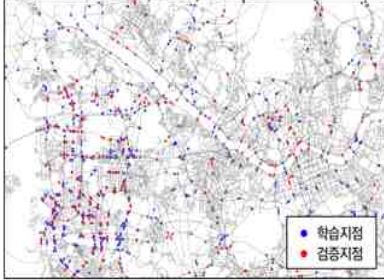
다. 교통량 추정 알고리즘 개선

1) 교통량 추정 입력 데이터 구축 프로세스

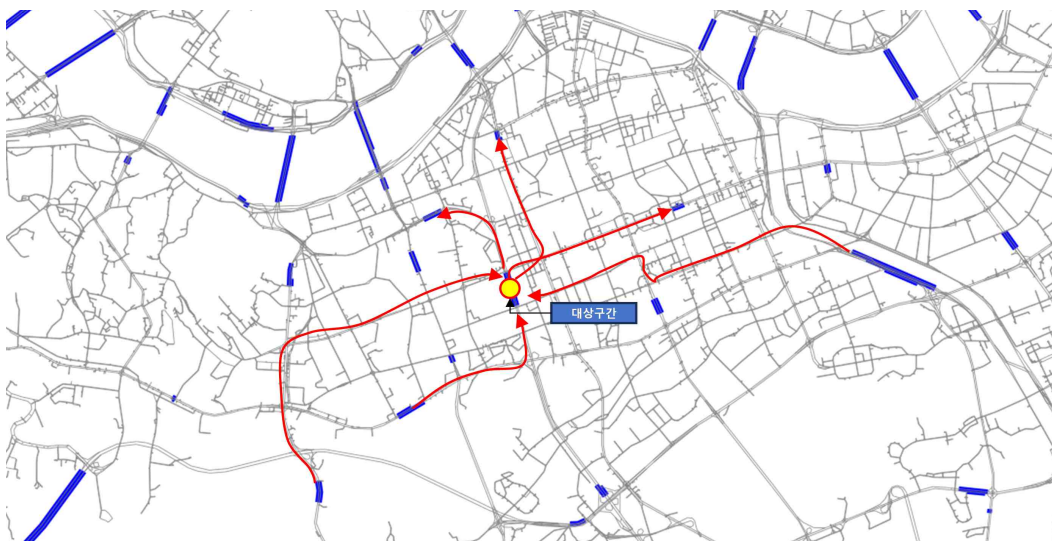
- 교통량을 추정하기 위한 데이터로써 입력 데이터 구축
 - 상세도로망 Level6 네트워크 기반의 경로 데이터로 구축한 속도 프로파일 DB를 활용하여 교통량 추정 입력 데이터 DB를 구축
- 차종별 링크의 속도 프로파일 데이터에서 평균속도 정보를 가져오고, 차종별 링크의 통행량 데이터에서 통행량 정보를 가져와 링크 전체 지점의 차종별 통행량 및 속도 데이터 생성함
- 관측 링크와 전체 링크의 교통량 추정을 위한 입력 데이터 가공
 - 관측교통량이 존재하는 링크를 대상으로 머신러닝의 학습 데이터로 사용될 관측지점의 교통량 추정 입력 데이터를 가공
 - 학습된 모델을 통해 관측지점을 포함하여 전체지점의 교통량을 추정하기 위해 전체 링크의 교통량 추정 입력 데이터를 가공

2) 교통량 추정 입력 데이터 고도화

- 교통량 추정을 위한 학습데이터는 약 1만여개의 관측교통량을 학습지점과 검증지점으로 구분하여 약 8,500개 관측지점을 학습데이터로 활용함
- 티맵, DTG 경로 데이터의 연속 주행 기록 정보를 통해 대상 관측지점과 연결된 다른 관측지점 정보를 추가로 입력 데이터를 구축해 도로의 기능적 특성을 고려할 수 있도록 하고자 함
 - 대상 관측지점을 지나는 개별차량 경로를 이용하여 추정 대상 도로구간과 가까운 통행량 및 차선수, 도로등급 등을 고려하여 최대 10개까지 생성한 뒤 연결성이 높은 3개 지점을 네트워크 연결성 변수로 구축함

구분	학습데이터(80%)	검증데이터(20%)
교통량 추정을 위한 데이터 구성	<ul style="list-style-type: none"> ☑ 학습 데이터 누적방식의 교통량 추정모형 개발 ✓ 일반적으로 학습 데이터의 양이 늘어날수록 모형의 일반화 성능과 강건성 향상 ✓ 단, 정갈하게 엄선된 학습 데이터 필요하며 추가되는 데이터의 품질이 매우 중요 ✓ 관측교통량 이상치 제거 필요 	<ul style="list-style-type: none"> ☑ 학습 데이터와 완전히 독립된 데이터로 선정 (Data Leakage 방지) ✓ 일관성 확보를 위해 매년 동일한 지점을 신뢰도 평가지점으로 선정
	<p>2020년 학습 데이터 (8,554개 지점)</p> <p>2021년 학습 데이터 (8,768개 지점)</p> <p>2022년 학습 데이터 (8,959개 지점)</p> <p>2023년 학습 데이터 (9,051개 지점)</p>	<p>2023년 검증 데이터 (2,009개 지점)</p> 

<그림 12> 교통량 추정을 위한 데이터 구성



<그림 13> 네트워크 상의 시공간적 연결성 변수 형태 예시

3) 교통량 추정

- 전국에 조사된 교통량은 약 1만 개 지점으로 상세도로망 Level6네트워크의 링크 갯수(약 62만 개)에 비해 매우 낮은 샘플율을 가지고 있음
- 각 지자체에서 조사한 관측교통량은 차종이 구분되어 있지 않거나, 24시간 전체 시간대에 조사하지 않는 경우가 많아 차종별 시간대별 교통량 추정을 위해 상시지점의 차종을 추정함
- 본 과업을 통해 교통량 추정 알고리즘을 개선하여 과거연도(2019~2022년)데이터를 재산정 및 2023년 교통량을 추정하였음
- 기초교통 DB 기반으로 전국 2차로 이상 도로에 대한 추정교통량DB 구축함
 - 기본 구축범위는 차종(승용차, 버스, 트럭), 평일, 시간(전일, 시간대별)단위, Level6 링크 단위 구축
 - 집계단위 : 상세도로망, 주요도로망, 표준노드링크, 행정구역(시도/시군구/읍면동)
- 한국개발연구원(2008)에 따르면 추정교통량의 신뢰성을 평가하기 위해 조사교통량과의 오차율 기준으로 추정교통량에 대한 평가를 수행하고 있으며, 본 연구에서 분석하고 있는 조사지점의 교통량 수준을 고려하면 ±15% 이내의 오차율이 유의한 추정 범위라고 볼 수 있음
- 추정교통량의 신뢰도 평가지표는 평균절대비율오차(Mean Absolute Percent Error, MAPE, %)를 활용함

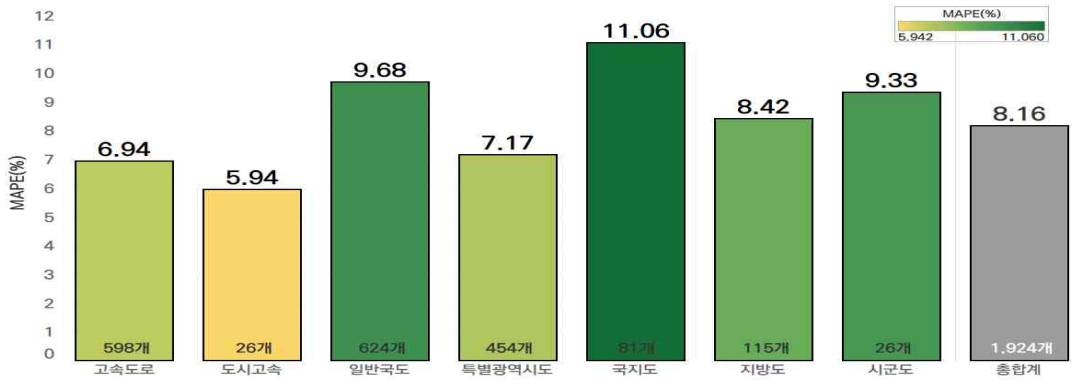
$$MAPE(\%) = \frac{\sum_{i=1}^n \left(\frac{|q_i - \hat{q}_i|}{q_i} \right) \times 100}{n},$$

여기서, q_i = 지점 i 의 조사교통량(대/일)

\hat{q} = 지점 i 의 추정교통량(대/일)

n = 전체 지점 개수(개)

- 검증용 데이터를 활용하여 도로등급별 신뢰도 분석 결과, 고속도로 6.9%, 일반국도 9.7%, 특별광역시도 7.2%로 유의미한 결과가 나타남

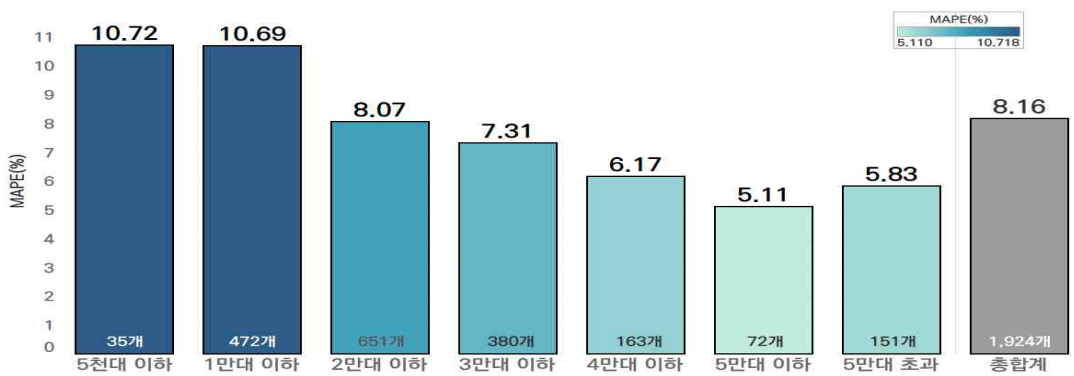


※ 참고1. 평가지점은 ITS 성능평가 기준(1시간 기준 200대 이상인 도로를 대상)으로 선정함

참고2. MAPE 목표치는 한국개발연구원의 직접 영향권 범위 참고함

참고3. 2023년 추정교통량 기준 분석결과

〈그림 14〉 도로등급별 신뢰도 분석 결과



※ 참고1. 평가지점은 ITS 성능평가 기준(1시간 기준 200대 이상인 도로를 대상)으로 선정함

참고2. MAPE 목표치는 한국개발연구원의 직접 영향권 범위 참고함

참고3. 2023년 추정교통량 기준 분석결과

〈그림 15〉 교통량 수준별 신뢰도 분석 결과

라. 기초교통 DB구축

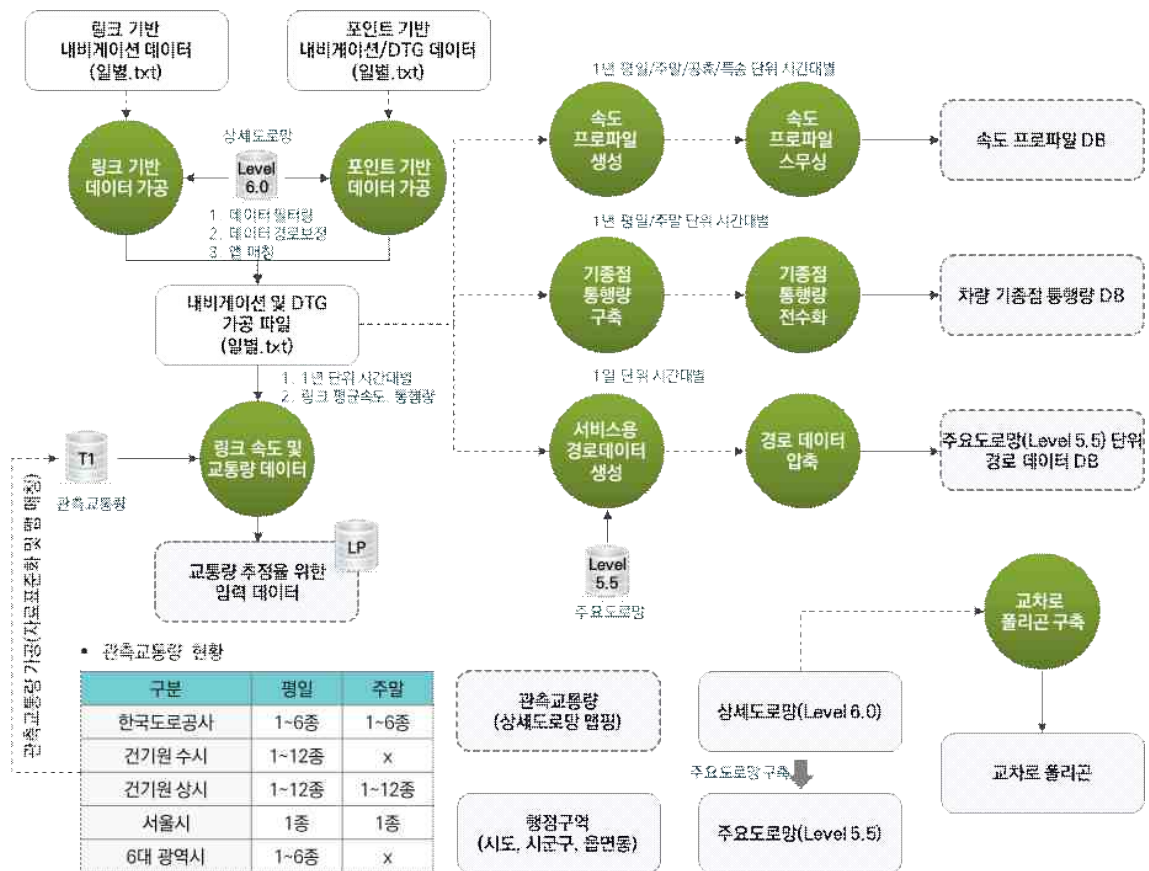
1) 기초교통 DB 구축 개요

○ 전국 교통량 전수화 및 통행지표를 구축하기 위해 경로 데이터 및 관측교통량을 이용하여 1차 가공한 기초교통 DB 구축

- 경로 데이터의 속도 정보를 이용하여 구축한 속도 프로파일 구축

- 차량 기종점 통행량 DB 구축

- 링크 통행량과 속도로 구성된 교통량 추정을 위한 입력 데이터 구축

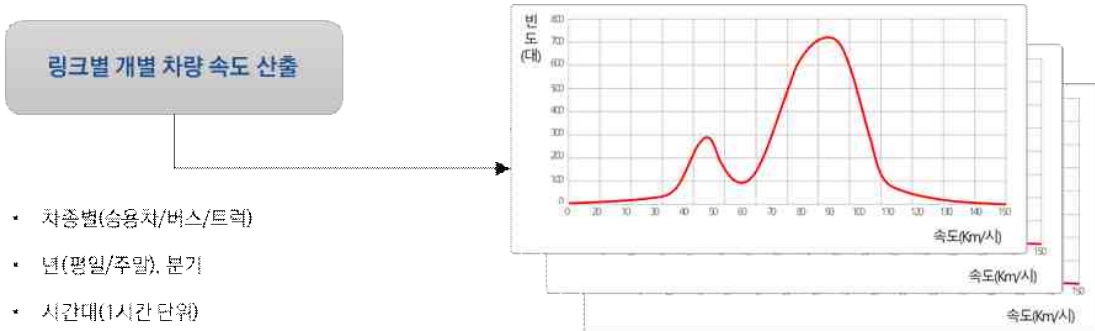


<그림 16> 기초교통 DB 구축 프로세스

2) 속도 프로파일 DB 구축

○ 링크별 속도 프로파일 집계

- 교통량 추정 DB와 통행지표 구축을 위하여 속도 프로파일 DB 구축
 - 속도지표, 혼잡지표, 환경지표 등 다양한 통행지표 등에 활용
- 링크에 매칭된 궤적 정보를 이용하여 개별 차량의 속도 산출 및 산출된 속도 범위에 따른 집계를 수행
 - 링크별로 1km/h 단위의 통행속도(1km/h ~ 150km/h, 초과시 150km/h)의 통행량을 집계
- 개별 링크의 년/분기/평일주말/1시간 단위의 속도 프로파일 구축



- 차종별(승용차/버스/트럭)
- 년(평일/주말), 분기
- 시간대(1시간 단위)

〈그림 17〉 속도 프로파일 구축 프로세스

○ 도로등급별 평균속도 비교분석

- 2022년 대비 2023년 평균속도는 모든 도로등급에서 감소함

〈표 7〉 연도별 도로등급별 평균속도

단위 : km/h

도로등급	평균속도				'20년 대비	'21년 대비	'22년 대비
	2020년	2021년	2022년	2023년	'21년 증감율	'22년 증감율	'23년 증감율
고속도로	97.62	95.29	93.87	91.09	-2.4%	-1.5%	-3.0%
도시고속도로	65.87	65.19	65.42	63.89	-1.0%	0.4%	-2.3%
일반국도	62.44	61.21	60.86	60.32	-2.0%	-0.6%	-0.9%
특별광역시도	26.04	25.24	25.11	24.79	-3.1%	-0.5%	-1.3%
국가지원지방도	51.24	50.45	50.03	49.59	-1.5%	-0.8%	-0.9%
지방도	49.73	48.82	48.41	47.91	-1.8%	-0.8%	-1.0%
시군도	34.09	33.6	33.24	32.94	-1.4%	-1.1%	-0.9%
총계	38.21	37.49	37.12	36.75	-1.9%	-1.0%	-1.0%

○ 시도별 평균속도 비교분석

- 연도별 평균속도를 분석했을 때 2020년 38.2km/h, 2021년 37.5km/h, 2022년 37.1km/h, 2023년 36.8km/h로 매년 약 1%씩 속도가 감소하고 있음
- 2022년 대비 2023년 평균속도가 가장 크게 증가한 지역은 대구로 약 6.1%이고, 그 다음 제주에서 약 0.2% 증가하였음
- 나머지 시도에서 2022년 대비 2023년 평균속도가 감소한 것으로 분석되었음

〈표 8〉 연도별 시도별 평균속도

단위 : km/h

시도	평균속도				'20년 대비 '21년 증감율	'21년 대비 '22년 증감율	'22년 대비 '23년 증감율
	2020년	2021년	2022년	2023년			
서울	24.07	22.98	23.07	22.77	-4.5%	0.4%	-1.3%
부산	29.34	28.89	28.86	28.31	-1.5%	-0.1%	-1.9%
대구	30.14	29.55	29.58	31.37	-2.0%	0.1%	6.1%
인천	31.09	30.12	29.72	29.24	-3.1%	-1.3%	-1.6%
광주	28.52	28.08	27.64	27.48	-1.5%	-1.6%	-0.6%
대전	30.39	29.31	29.1	28.66	-3.6%	-0.7%	-1.5%
울산	33.9	33.53	33.23	32.76	-1.1%	-0.9%	-1.4%
세종	36.98	35.99	35.41	35.15	-2.7%	-1.6%	-0.7%
경기	33.57	32.79	32.35	32.03	-2.3%	-1.3%	-1.0%
강원	42.05	41.63	41.24	41.13	-1.0%	-0.9%	-0.3%
충북	41.81	41.26	40.56	40.25	-1.3%	-1.7%	-0.8%
충남	42.83	41.9	41.14	40.59	-2.2%	-1.8%	-1.3%
전북	42.23	41.38	41.07	40.72	-2.0%	-0.7%	-0.9%
전남	43.28	42.87	42.61	42.12	-0.9%	-0.6%	-1.1%
경북	43.24	42.68	42.26	41.68	-1.3%	-1.0%	-1.4%
경남	39.59	38.94	38.68	38.3	-1.6%	-0.7%	-1.0%
제주	37.52	36.84	36.34	36.43	-1.8%	-1.4%	0.2%
총계	38.21	37.49	37.12	36.75	-1.9%	-1.0%	-1.0%

* 연단위 평일 기준 도로등급 연결로를 제외한 평균속도를 분석함

3) 교통량 추정 입력데이터 DB 구축

○ 교통량 추정을 위한 입력 데이터 정의

- 관측링크와 전체링크의 입력 데이터는 형태는 동일하며, 머신러닝의 학습에 영향을 주는 정보를 선택하여 구축함
- 머신러닝 학습에 영향을 주는 정보는 교통량 추정 알고리즘 개발 시 탐색적 분석을 통해 선별한 변수임

○ 네트워크 연결성 변수 구축

- 네트워크 연결성 변수는 미관측 대상 링크와 가장 연관성이 있는 주변 도로의 통행 정보를 사용하는 것으로 미관측 대상 링크를 지나는 차량의 통행량과 관측교통량을 그룹으로 묶어 입력데이터로 활용됨
- 미관측 대상 링크를 지나는 모든 경로를 추출하여 해당 경로와 연결된 관측 교통량, 통행량 정보를 구축한 뒤 연결성이 높은 데이터만을 사용함
- 미관측 대상 링크 주변도로는 거리에 따라 다양한 도로등급이 존재하므로 미관측 대상 링크와 교통량이 유사한 동일한 도로등급 그룹을 사용됨
- 도로등급 그룹은 고속도로와 도시고속도로, 일반국도와 특별광역시도, 국가지원지방도, 지방도, 시군도로 구분해서 2개의 그룹으로 구분함
- 미관측 대상 링크를 지나는 차량의 경로의 거리를 기준으로 최단거리, 동일한 도로등급 그룹인 정보를 사용하여 네트워크 연결성 변수를 구축함

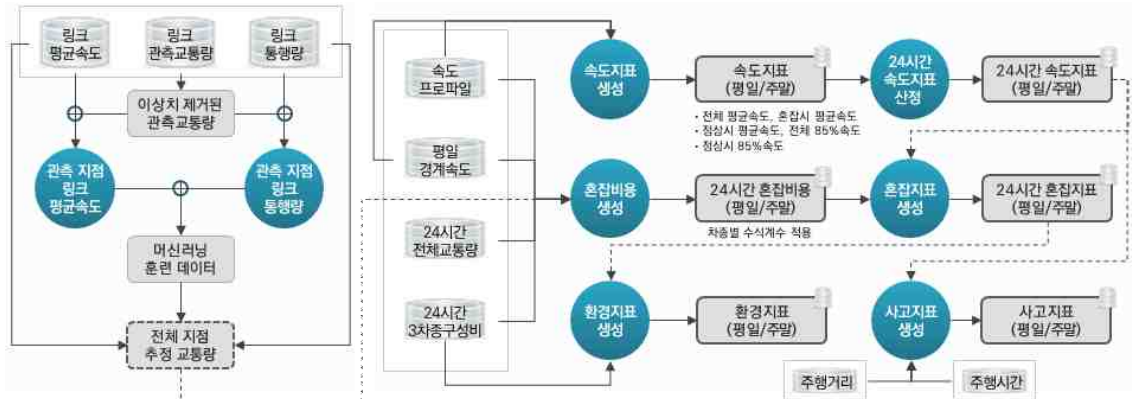
〈표 9〉 교통량 추정을 위한 입력 데이터 DB

순번	필드명(코드명)	필드명(한글)	설명
1	v_link_id	가상링크ID	상세도로망 네트워크의 가상링크ID
2	week_code	주중주말코드	주중, 주말을 구분하는 코드
3	time_code	시간코드	하루 24시간을 1시간 단위로 구분한 코드
4	total_probe	전체 프로브 통행량	승용, 버스, 트럭 프로브 통행량의 합
5	auto_probe	승용 프로브 통행량	시간대별 일 평균 승용 프로브 통행량
6	bus_probe	버스 프로브 통행량	시간대별 일 평균 버스 프로브 통행량
7	truck_probe	트럭 프로브 통행량	시간대별 일 평균 트럭 프로브 통행량
8	auto_spd	승용 속도	링크를 통과한 승용의 평균 속도
9	length	링크 길이	상세도로망 네트워크의 가상링크 길이
10	road_rank	도로등급	상세도로망 네트워크의 가상링크 도로등급
11	lanes	차선수	상세도로망 네트워크의 가상링크 차선수
12	k	K	승용 프로브 통행량 / 승용 속도
13	emd_type	동 여부	해당 링크가 동인지 아닌지 코드
14	sido_id	시도ID	상세도로망 네트워크의 가상링크 시도ID
15	traffic_type	조사기관	관측교통량을 조사한 기관
16	v_total_wd	전체 관측교통량	승용, 버스, 트럭 관측교통량의 합(차종추정버전)
17	v_auto_wd	승용 관측교통량	시간대별 일 평균 승용 관측교통량(차종추정버전)
18	v_bus_wd	버스 관측교통량	시간대별 일 평균 버스 관측교통량(차종추정버전)
19	v_truck_wd	트럭 관측교통량	시간대별 일 평균 트럭 관측교통량(차종추정버전)
20	total_scale	전체 프로브 통행량 스케일	승용, 버스, 트럭 프로브 통행량의 합의 min-max 스케일값
21	auto_scale	승용 프로브 통행량 스케일	시간대별 일 평균 승용 프로브 통행량 min-max 스케일값
22	bus_scale	버스 프로브 통행량 스케일	시간대별 일 평균 버스 프로브 통행량 min-max 스케일값
23	truck_scale	트럭 프로브 통행량 스케일	시간대별 일 평균 트럭 프로브 통행량 min-max 스케일값
24	v_total_wd_1	전체 관측교통량_1	승용, 버스, 트럭 관측교통량의 합_1
25	v_auto_wd_1	승용 관측교통량_1	시간대별 일 평균 승용 관측교통량_1
26	v_bus_wd_1	버스 관측교통량_1	시간대별 일 평균 버스 관측교통량_1
27	v_truck_wd_1	트럭 관측교통량_1	시간대별 일 평균 트럭 관측교통량_1
28	auto_scale_1	승용 프로브 통행량 스케일_1	시간대별 일 평균 승용 프로브 통행량 min-max 스케일값_1
29	v_total_wd_2	전체 관측교통량_2	승용, 버스, 트럭 관측교통량의 합_2
30	v_auto_wd_2	승용 관측교통량_2	시간대별 일 평균 승용 관측교통량_2
31	v_bus_wd_2	버스 관측교통량_2	시간대별 일 평균 버스 관측교통량_2
32	v_truck_wd_2	트럭 관측교통량_2	시간대별 일 평균 트럭 관측교통량_2
33	auto_scale_2	승용 프로브 통행량 스케일_2	시간대별 일 평균 승용 프로브 통행량 min-max 스케일값_2
34	v_total_wd_3	전체 관측교통량_3	승용, 버스, 트럭 관측교통량의 합_3
35	v_auto_wd_3	승용 관측교통량_3	시간대별 일 평균 승용 관측교통량_3
36	v_bus_wd_3	버스 관측교통량_3	시간대별 일 평균 버스 관측교통량_3
37	v_truck_wd_3	트럭 관측교통량_3	시간대별 일 평균 트럭 관측교통량_3
38	auto_scale_3	승용 프로브 통행량 스케일_3	시간대별 일 평균 승용 프로브 통행량 min-max 스케일값_3

마. 차량 교통망 성능평가지표 구축

1) 차량 교통망 성능평가지표 구축 개요

- 차량 교통망 성능평가지표 DB는 1차 가공된 속도 프로파일과 추정된 교통량을 기반으로 전국 혼잡지표, 속도지표, 환경지표, 안전지표를 구축함



〈그림 18〉 교통량 추정 및 차량 통행지표 구축 프로세스

- 교통망 성능평가지표 구축을 위한 기초교통 DB는 다음과 같이 정의된 범위가 존재하여 범위에 따라 지표를 구축함

〈표 10〉 교통망 성능평가지표 DB 생성 기준

데이터명	연	분기	평일 주말	시간	차종	상세 도로망	주요 도로망	행정 구역
속도 빈도분포	○	○	○	○	○	○	○	○
교통량지표	○		평일	○	○	○	○	○
속도지표	○		○	○		○	○	○
혼잡지표	○		평일			○	○	○
환경지표	○		평일		○	○	○	○
안전지표	○		평일			○	○	

- 차량 교통망 성능평가지표는 도로망의 현황과 성능을 평가하기 위한 지표를 의미하며, 약 17종의 차량 통행지표를 구축

〈표 11〉 차량 교통망 성능평가지표

지표명		지표설명
교통량 지표	관측교통량	공공에서 조사하는 교통량으로 수집기관, 조사기간 등이 상이하여 동일한 기준으로 도로망 네트워크에 매칭하여 나타낸 연평균 일단위 조사교통량
	추정 교통량	특정 시간대 동안 해당 도로구간을 통과하는 차량 대수
	차량 주행거리	추정 교통량 기준의 차량주행거리
속도지표	평균속도	전체 차량의 속도를 평균한 값
	정상시 평균속도	전체 차량 중 교통 혼잡을 경험한 차량들의 평균속도
	혼잡시 평균속도	전체 차량 중 교통 혼잡을 경험하지 않은 차량들의 평균속도
혼잡지표	혼잡시간강도	전체 차량의 총 통행시간 중 교통 혼잡을 경험한 차량들의 총 통행시간 비율
	혼잡빈도강도	전체 차량의 총 통행시간 중 교통 혼잡을 경험한 차량들의 총 통행량 비율
	교통혼잡비용	차량들이 도로상에서 교통혼잡이 발생하여 정상속도 이하로 운행하게 됨으로써 추가로 발생하게 되는 비용
환경지표	이산화탄소배출량	차량들로 주행으로 발생하는 대기오염물질 배출량
	미세먼지배출량	
	일산화탄소배출량	
	휘발성 유기화합물배출량	
	질소산화물배출량	
안전지표	과속비율	과속기준치를 초과하는 차량의 비율
	속도 표준편차	개별 차량들의 속도의 퍼짐 정도
	위험운전행동 이벤트 비율	위험운전행동기준을 개별 차량 속도기준에 따라 위험운전행동이벤트 발생 건수를 비율로 나타냄

2) 차량 교통망 성능평가지표 활용

- 차량 교통망 성능평가지표를 View-T 웹서비스를 통해 '통행지표'와 '데이터 다운로드' 등의 기능에 업로드하여 데이터를 제공하고 있음
- 교통량지표 중 차량주행거리는 통계청의 실험적 통계에서 차량 GPS 기반 도로 유형별 주행거리 통계로 활용되고 있음



〈그림 19〉 통계청의 실험적 통계에 활용되는 차량 교통망 성능평가지표

- 교통혼잡지표 중 교통혼잡비용은 통계청의 e-나라지표에서 도로교통 혼잡비용 통계로 활용되고 있음
- 교통환경지표 중 이산화탄소배출량은 국토교통부 탄소공간지도에서 수송부문 배출량 데이터로 활용되고 있음

4. 차량 GPS 데이터 구축·분석시스템 개선

가. 차량 데이터 구축·분석 시스템 개요

- 데이터 수집과 가공 각 단계를 모니터링하여 데이터가 정상적으로 처리되고 있는지 실시간으로 확인할 수 있는 서비스 지원
- KTDB 기반 지도 및 차량 통행 데이터셋을 체계적으로 관리하고 분석에 활용할 수 있는 시스템 구축
- 데이터셋 가공·구축 시스템 관리
 - 서버 상태 모니터링 환경 구축을 통해 실시간 상태 확인 및 장애 예방 가능
 - 데이터 가공 모듈 개선 시 서버 업데이트 및 개발 정보 지속 관리
- 분석 시스템 관리
 - 수집 현황과 데이터 전처리 상태를 제공하기 위한 대시보드 서비스 제공
 - 공간 단위별로 구축된 교통망 성능 평가 지표 및 통행 분석 DB 업로드 및 버전 관리 지원
 - 데이터 작업 현황에 따라 데이터베이스(MySQL) 저장 및 관리

나. 데이터 가공 모니터링 기능 개발

- 원시 데이터 수집 및 데이터 전처리에서 이상치 제거, 정렬, 좌표 변환, 시간 정보 변환 등의 전처리 과정을 각 단계별로 수행하며, 일자별 오류 유형, 가공 용량, OBU 등 수집된 로그 기반으로 모니터링 기능 개발

1) 기능 구성

① 데이터 수집 현황 분석기능

- 원시 데이터의 파일 유효성 검사 정보를 테이블 및 그래프로 제공

- 주요 항목 : 파일 수, 오류 파일 수, 총용량 등
- 년/월/일 별 데이터 용량, 건수를 시각적으로 표현

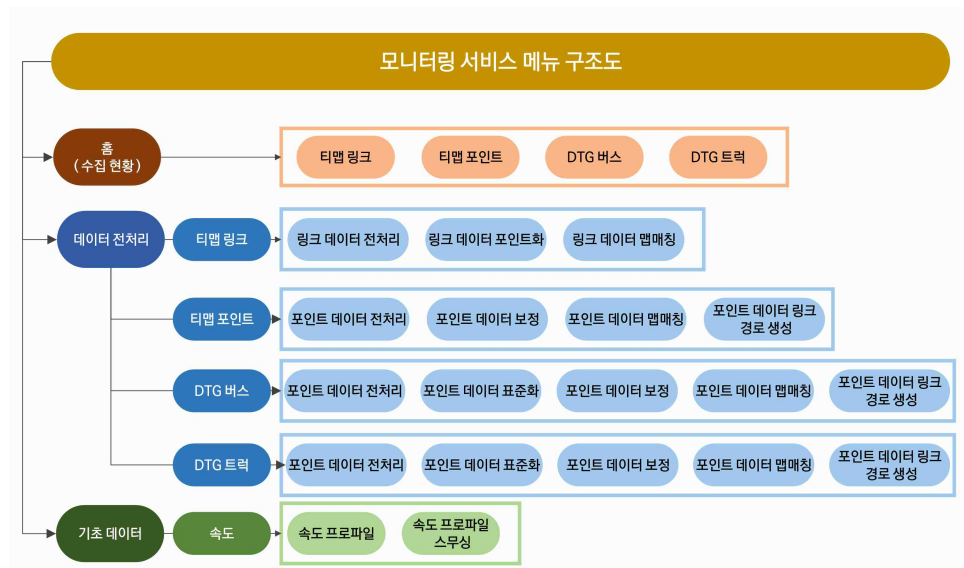
② 데이터 전처리 현황 분석기능

- 이상치 제거, 정렬, 좌표 변환, 시간 정보 변환 등의 데이터 전처리 작업에서 발생하는 로그 정보를 테이블 및 그래프로 제공

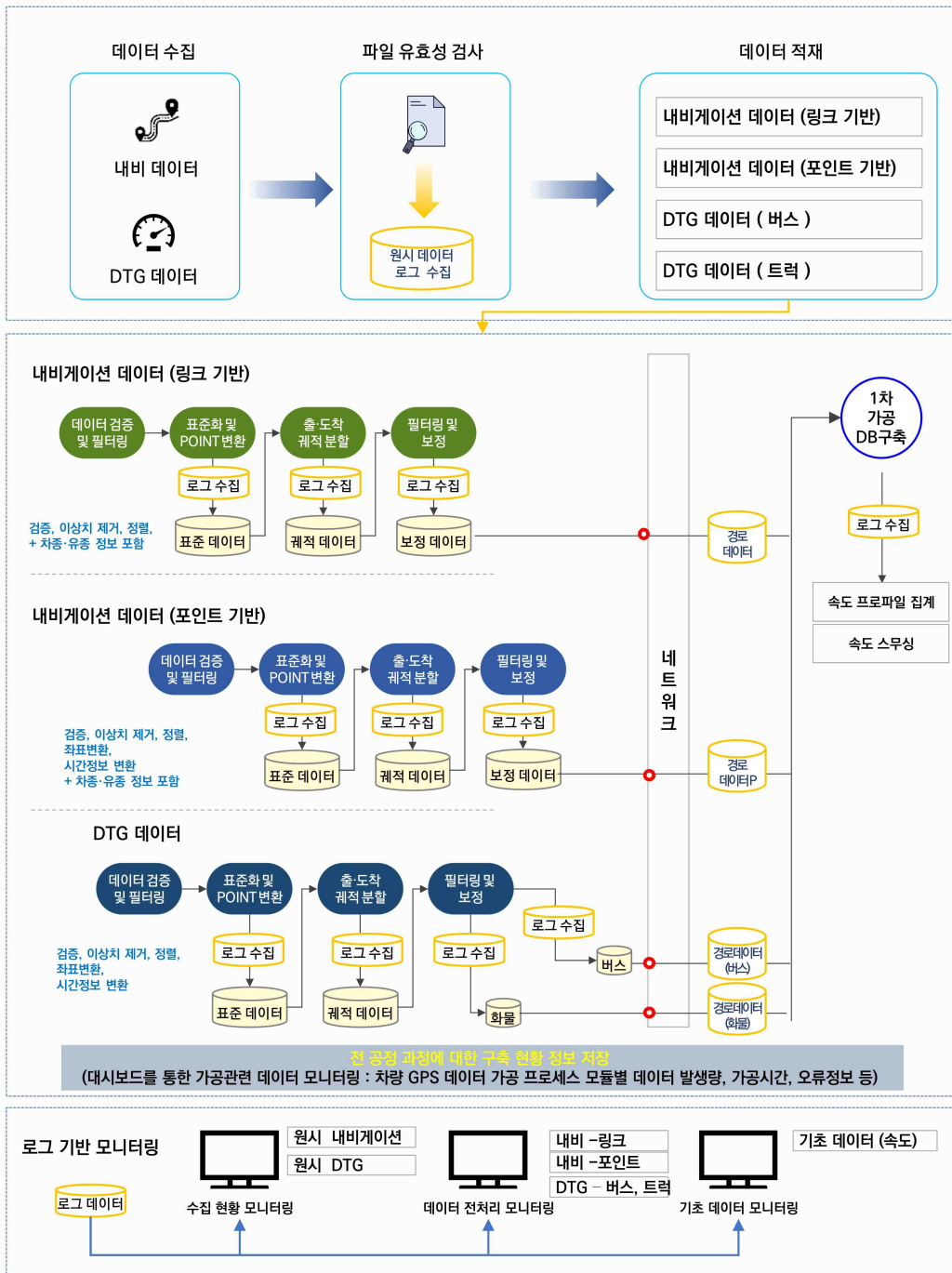
- 주요 항목 : 작업별 처리 데이터 수, 처리 성공 및 오류 수 등
- 전처리 과정에서 데이터의 상태, 정보 등 각 전처리 과정에 맞는 정보 시각적으로 표현

③ 기초 데이터 현황 분석기능

- 전처리 데이터와 속도 프로브 데이터에서 생성되는 기초 데이터의 로그 정보를 테이블 및 그래프로 제공



<그림 20> 모니터링 메뉴 구조도



〈그림 21〉 모니터링 프로세스

2) 기능 구축 결과

- 선택한 년도, 기간에 따라 이벤트 건수, 파일 수, 용량 및 데이터 현황 등을 테이블을 통해 사용자는 데이터를 구체적으로 확인할 수 있도록 함



〈그림 22〉 데이터 수집 현황 모니터링 화면

- 선택한 표출 컬럼에 따라 테이블 데이터가 구성되며, 각 전처리 단계에서 발생하는 수집된 로그 정보를 자세하게 확인할 수 있도록 함



〈그림 23〉 데이터 전처리 모니터링 화면



제1장 과업의 개요

제1절 과업의 배경 및 목적

제2절 과업의 범위 및 내용

제1장 과업의 개요

제1절 과업의 배경 및 목적

1. 과업의 배경

- 전국 단위 공공의 기초 교통데이터 수집 및 지점 데이터 분석의 한계
 - 교통 분야의 기초 데이터인 교통량, 속도 데이터는 한국건설기술연구원, 한국도로공사, 광역지자체 등에서 조사하고 있으나, 데이터의 공간적 수집 범위가 매우 낮아 전국의 교통 현황을 분석하기에 한계가 있음
 - 공공에서 수집하고 있는 데이터는 지점 정보 중심의 데이터로 단편적 정보만 확인할 수 있어 시공간적으로 연결된 교통의 흐름을 파악하기엔 한계가 있음
- 전국 단위 일관된 교통DB 및 통행지표 부재
 - 데이터 수집기관, 수집방식 등 기관별로 상이하게 수집되는 데이터 수집 체계는 전국을 일관된 기준으로 분석하기엔 한계가 있음
- 빅데이터 전처리·가공 환경 및 빅데이터 기반 차량통행 분석시스템 부재
 - 차량 GPS 빅데이터를 효율적으로 전처리·가공하기 위한 환경이 부족하며, 과거에는 차량의 시·공간적 통행패턴과 현황을 분석하기 위해 모형 중심의 프로그램을 이용하여 분석을 수행하였음
- 국가통합교통체계효율화법 시행령 제8조1)에 의거하여 교통물류활동으로 발생

1) 국가교통조사에는 다음 각 호 사항이 포함되어야 한다 - 교통물류활동으로 발생하는 교통혼잡, 교통사고, 환경오염, 온실가스

하는 교통혼잡 등 교통 관련 사회적 외부비용을 국가교통물류에 관한 지표를 설정하여 고시하도록 규정하고 있음

2. 과업의 목적

- 국가통합교통체계효율화법²⁾에 의거 국가교통데이터베이스 구축 및 운영을 위한 기초교통DB 및 통행지표 생성
 - 개별 차량 기반의 교통량, 속도, 차량주행거리 등 교통분야의 핵심 기초교통 DB 구축(공공 조사교통량 약 1만 개 지점, 전국 도로 62만 개 링크)
 - 차량 이용 및 교통물류 활동으로 인한 교통혼잡, 환경오염, 온실가스 배출 등 다양한 측면의 도로 교통망 성능평가 지표 구축
 - 지자체별로 상이한 방식으로 수집되고 구축되는 교통DB·통행지표의 한계를 개선하기 위한 전국단위 일관된 교통DB·통행지표 구축 필요
 - 여객 O/D구축 및 교통수요예측을 위한 기반데이터(승용차 O/D 데이터 등) 제공
- 국가교통 데이터 제공 및 교통 분야 주요 정책 지원
 - 이용자의 요구에 대응할 수 있는 데이터 제공 및 분석지원
 - 탄소공간지도 사업, 환경소음측정망 선정 사업, 지자체 실증사업 등 주요 정책 제언을 위한 데이터 및 분석 지원(매년 요청 및 제공)

배출 등 교통 관련 사회적 비용, 그 밖에 교통 관련 정책 및 계획의 수립, 교통시설 투자분석 및 평가에 필요한 사항

2) 국가통합교통체계효율화법 제12조(국가교통조사), 제14조(정보통신수단 등을 통한 교통조사) 및 동법 시행령 제8조(국가교통조사의 실시)

제2절 과업의 범위 및 내용

1. 과업의 범위

- 공간적 범위
 - 전국 2차로 이상 도로 및 주요 도로³⁾
- 시간적 범위
 - 기준연도 : 2023년
 - 데이터 수집 : 2023년, 2024년 1-10월
 - 데이터 가공 및 DB구축 : 2023년, 2024년 1-6월

2. 과업의 내용

- 차량 원시 GPS 빅데이터 전처리 및 DB가공·구축
 - 차량 GPS 데이터, DTG 데이터 수집 및 검토
 - 원시 데이터 전처리 및 경로 데이터 가공
 - 기초 교통 DB구축 및 검증
- 교통량 추정 알고리즘 개선
 - 네트워크 연결성을 고려한 교통량 추정 알고리즘 개선
 - 기초 교통DB 구축을 위한 교통량 추정 신뢰도 분석 및 연도별 패턴 분석
- 차량 통행지표 가공 알고리즘 개선 및 DB구축
 - 통행지표 가공 알고리즘 개선 및 고도화
 - 목적별 전국 통행지표 산출 및 DB 구축

3) 주요 도로 : 편도 1차로 도로 중 데이터 수집의 안정성이 확보된 도로



제2장 차량 GPS 빅데이터 수집 및 분석

제1절 차량 GPS 빅데이터 수집 개요

제2절 차량 원시데이터 분석

제2장 차량 GPS 빅데이터 수집 및 분석

제1절 차량 GPS 빅데이터 수집 개요

- 차량 GPS 데이터는 차량의 이동경로에 대해 시간 연속성과 공간 연결성이 동시에 수집되는 데이터로 차량의 위도, 경도, 고도 등 개별 차량의 속성 정보를 포함하고 있음
- 본 사업에서 수집한 차량 GPS 데이터는 티맵의 내비게이션 데이터⁴⁾와 한국교통안전공단의 DTG데이터이며, 데이터 수집 기관마다 활용되는 단말기 차이로 인해 데이터 수집 방식, 수집 주기 등 수집되는 속성 정보의 차이가 있음
 - 내비게이션 데이터는 경로탐색 시 차량의 운행정보가 생성되어 시간, 위치, 속도 정보를 포함하고 있음
 - DTG데이터는 차량운행기록계를 통해 사업용 차량의 운행정보를 기록하여 시간, 위치, 속도, 가속도 등을 포함하며, 사업용 차량은 크게 버스, 화물, 택시로 구분됨
- 차량 GPS 데이터는 도로속성 정보가 연결된 링크 기반 데이터와 X,Y 위치 좌표가 수집되는 포인트 기반 데이터로 구분됨
 - 내비게이션 데이터는 링크 기반 데이터와 포인트 기반 데이터로 구분됨
 - 링크 기반 내비게이션 데이터는 포인트 기반으로 구축된 정보를 도로속성 정보를 포함하는 맵 링크ID로 매칭되어 제공되므로 위치좌표마다 생성된 가속도,

4) 내비게이션 데이터는 데이터 신뢰도를 높이기 위해 2022년 사업부터 샘플율이 높은 티맵 내비게이션 데이터를 수집하여 2020년 데이터부터 원천 데이터가 변경됨

경사도 등 정보가 누락되어 수집되었음

- 금년도부터 포인트 기반 데이터도 수집할 수 있도록 협의하여 가감속, 경사도 등 정보를 활용한 상세한 교통DB구축을 수행할 예정임

〈표 2-1〉 차량 GPS 데이터 특징

구분	링크 기반 내비게이션 데이터	포인트 기반 내비게이션 데이터	DTG 데이터
주행 ID 생성기준	경로 탐색 시 ID 생성	경로 탐색 시 ID 생성	차량등록번호와 주행시작일로 ID 생성
속성 정보 생성기준	내비게이션 경로 탐색 시	내비게이션 경로 탐색 시	차량 운행 시
데이터 형태	SK맵 링크 단위	1초 ~ 3초 단위 포인트	1초 단위 포인트
좌표체계	Korea 2000	Korea 2000	WGS84

1. 차량 GPS 빅데이터 수집

가. 데이터 수집 대상

- 본 과업의 기준연도는 2023년 기준으로 차량 GPS 빅데이터 기반 기초교통 및 교통망 성능평가지표DB를 구축하기 위해 다음과 같은 데이터를 수집함
 - 링크 기반 내비게이션 데이터는 티맵의 내비게이션 사용자의 주행 정보, 주행 정보를 도로구간(링크)별로 나타내는 도로 네트워크 데이터를 수집함
 - 포인트 기반 내비게이션 데이터는 티맵의 내비게이션 사용자의 주행 정보, 주행 정보를 GPS 단위로 데이터를 수집하므로 초 단위 데이터를 수집함
 - 포인트 기반 사업용 차량 운행기록계(DTG) 데이터는 한국교통안전공단에서 제공하며, 사업용 차량에 부착된 운행기록계로 수집되는 버스, 화물, 택시의 주행 정보 데이터를 수집함

〈표 2-2〉 데이터 수집 목록 및 기준연도

구분	데이터 목록	수집처	수집연도
링크 기반	내비게이션 도로 네트워크	티맵	2023년 1-12월, 2024년 1-6월
	내비게이션 주행 정보		
포인트 기반	내비게이션 주행 GPS 정보	티맵	2024년 6, 7월, 10월, 12월
	사업용 차량 운행기록계 버스 주행 정보	한국교통안전공단	2023년 1-12월, 2024년 1-6월
	사업용 차량 운행기록계 화물 주행 정보		
	사업용 차량 운행기록계 택시 주행 정보		

○ 시의성 있는 데이터 제공을 위해 당해연도인 2024년 반기 데이터를 하반기부터 수집하여 빠르게 가공 및 구축 가능한 지표를 산출함

- 링크 기반 내비게이션 데이터는 2023년, 2024년 12주, 명절을 포함하여 월별 추세를 파악할 수 있는 기간으로 총 179일을 수집하며, 2024년까지 데이터를 수집하였음
- 포인트 기반 내비게이션 데이터는 2024년 4주, 여름과 겨울 추세를 파악할 수 있는 기간으로 총 31일을 수집할 예정이며, 데이터 오류 발견으로 재요청함
- 포인트 기반 내비게이션 데이터의 수집 시기는 데이터 수집기관의 서버 문제로 1월-5월까지 데이터가 삭제되었기 때문에 2024년 1월, 4월, 7월, 10월 데이터를 받기로 했으나, 수집 기간을 조정하였음

〈표 2-3〉 수집 데이터의 시간적 범위

구분	링크 기반 내비게이션		포인트 기반 내비게이션
	2023년	2024년	2024년
1월	25일 ~ 29일 (6일)	22일 ~ 28일 (7일)	-
2월	13일 ~ 19일 (7일)	13일 ~ 18일 (7일)	-
3월	13일 ~ 19일 (7일)	11일 ~ 17일 (7일)	-
4월	17일 ~ 23일 (7일)	15일 ~ 21일 (7일)	-

6월			17일 ~ 23일(7일)
7월			8일 ~ 14일(7일)

9월	20일, 22일 ~ 27일 (7일)	19일 ~ 25일 (7일)	-
10월	16일 ~ 22일 (7일)	14일 ~ 20일 (7일)	21일 ~ 27일(7일)
11월	13일 ~ 19일 (7일)	11일 ~ 17일 (7일)	-
12월	11일 ~ 17일 (7일)	16일 ~ 22일 (7일)	9일 ~ 15일(7일)
명절	설날 : 1월 21일 ~ 1월 24일 (4일)	설날 : 2월 09일 ~ 2월 12일 (4일)	추석 : 9월 16일 ~ 9월 18일 (3일)
	추석 : 9월 28일 ~ 9월 30일 (3일)	추석 : 9월 16일 ~ 9월 18일 (3일)	
합계	89일	90일	31일

나. 데이터 수집 대상의 특징 및 형식

1) 차량 내비게이션 데이터

- 링크 기반 내비게이션 데이터는 링크의 진입시간, 진출 시간, 속도 등의 정보가 포함되어 있으며, 차종과 유종, 목적지 정보가 포함되어 제공받음
- 링크 기반 내비게이션 데이터는 티맵에서 제공하는 도로 네트워크와 맵매칭해야 경로 위치를 파악할 수 있으며, 특정 기간마다 도로 네트워크의 링크ID가 변경되어 데이터 수집 기간별로 도로 네트워크가 존재함
- 티맵 내비게이션 데이터의 주행 ID는 목적지를 설정하고 경로 탐색 시 생성되어 동일한 내비게이션 사용자라도 주행마다 주행ID가 다르게 부여되므로 개인의 이동 경로를 추적할 수 없음
- 휴게소나 주정차 시 내비게이션 종료 상태가 아니라면 정차 상태로 데이터가 수집되지만, 종료 상태라면 경로를 임의로 생성하거나 보정하지 않음
- 출발지명은 GPS 정보가 있음에도 NULL인 경우는 사용자가 출발하는 위치가 법정동에 매칭이 안되는 경우로 사용자 단말기에서 서버로 전송하는 과정에서 발생함
- “운전습관 가입자”는 사용자 단말기 앱에서 사용자가 설정하는 것으로 100% 정보가 포함되어있지 않으며, 차종을 대상으로 분석했을 때 전체 주행 고유ID 대비 약 70~80% 범위 내에서 차종 정보가 존재함
- 목적지명은 경로 탐색을 위해 사용자가 입력하는 정보로 건물명, 사회시설, 기관명 등이 포함되어 있으며, 주소로 검색했을 경우 목적지 중분류 정보가 NULL로 나타남
- 목적지 중분류 정보는 티맵 POI 데이터로부터 수집되는 정보로 공공기관, 공공기업, 교육기관, 교통시설 등 다양한 분류 정보가 포함되어 있으며, 해당 정보를 통해 개별 차량의 경로 목적을 파악할 수 있음
- 목적지 정보를 통해 주행 목적을 파악할 수 있으나, 주행이 끊긴 경우 새로운 주행 고유 ID가 부여되므로 장거리 통행에서는 휴게소, 주차 등으로 인해 목적지명이 중복으로 집계될 수 있으므로 목적지 정보를 주행 목적별로 분류하기 어려운 것으로 분석됨

〈표 2-4〉 링크 기반 내비게이션 데이터 형식

구분	내용	예시
P_REQDATE	주행시작일(YYYYMMDD)	20230209
TRIP_INIT_SESSION_ID	주행 고유 ID	108820220209171330540S00
LINK_SEQ	주행 내 링크 시퀀스	1
MESH_CD	MESH 코드	5716
LINK_ID	LINK 코드	512
LINK_DIR	LINK 방향 (0 : 정방향 / 1 :역방향)	1
LENGTH	LINK 길이	49
DRIVING_TIME	LINK 주행시간 (초)	41
ENTRY_TIME	링크 진입시간(연월일시분초)	20240209171332
EXIT_TIME	링크 진출시간(연월일시분초)	20240209171413
SPEED_CAPPED	CAP이 씌워진 상태로 계산된 속도값	/N
SPEED_GPS	단말기에서 올려준 속도값의 평균	4.9333
SPEED_REFINED	보완된 최종 속도값	4.9333
CAPPED_YN	CAP 쓰워짐 여부	N
DEPART_NAME	출발지명(읍면동 기준, GPS 문제시 NULL)	서울특별시 강남구 역삼1동
DEPART_X	출발지 X좌표(BESSEL)	127.039416
DEPART_Y	출발지 Y좌표(BESSEL)	37.6938611
DEST_NAME	목적지명	성남시청 음악분수
DEST_POIID	목적지 POI ID	1000670
DEST_X	목적지 X좌표(BESSEL)	127.1278333
DEST_Y	목적지 Y좌표(BESSEL)	37.4176111
CB_NAME	목적지 중분류 정보(주소 검색 등의 경우 NULL)	구역내시설물
CAR_MODEL	차종("운전습관 가입자" 대상)	CT_NORMAL
CAR_FUEL	유종("운전습관 가입자" 대상)	FT_LPG

- 포인트 기반 내비게이션 데이터는 GPS궤적 X좌표, Y좌표에 따른 주행 시간, 속도, 고도값 등의 정보가 있으며, 주행차량의 배기량과 연식을 포함하여 제공받아 링크 기반 내비게이션 데이터와 차이가 있음
 - GPS 수집 주기는 기본적으로 1초 단위로 수집되나 경로이탈, 맵매칭 오류 등으로 1초 이상의 수집 주기가 발생할 수 있음
 - GPS속도, 고도값이 음수인 경우 GPS가 잡히지않아 오류가 발생한 것으로 분류되며, 임의로 보정하지 않으므로 필터링이 필요함
 - 배기량과 연식 정보는 교통환경지표 산출을 위한 변수로 사용할 수 있으나, CP사를 통해 수집된 정보로 차종과 유종에 비해 샘플율이 부족하여 활용하기 어려움

〈표 2-5〉 포인트 기반 내비게이션 데이터 형식

구분	내용	예시
P_REQDATE	주행시작일(YYYYMMDD)	20240113
TRIP_INIT_SESSION_ID	주행 고유 ID	105020220113081947967S00
MESH_CD	MESH 코드	5616
LINK_ID	LINK ID	15656
LINK_DIR	LINK 방향 (0 : 정방향/ 1 :역방향)	1
GPS_X	GPS 궤적 X좌표	127.1137778
GPS_Y	GPS 궤적 Y좌표	36.79916111
GPS_TM	해당 GPS좌표 시 주행시간(연월일시분초)	2.02201E+13
SPEED_GPS	GPS 속도(GPS 오류 일 경우 음수값존재,m/s)	0
ALTITUDE	단말기로부터 수집된 상대 고도값(단위°)	0
ANGLE	단말기로부터 수집된 차량 진행 방향각(0~359°)	273
DEPART_NAME	출발지명(읍면동기준, GPS 문제시 NULL)	경상북도 문경시 마성면
DEST_NAME	목적지명	양천구보건소
DEST_POIID	목적지 POI ID	459268
DEST_X	목적지 X좌표(BESSEL)	126.8680833
DEST_Y	목적지 Y좌표(BESSEL)	37.51502777
CB_NAME	목적지 중분류 정보(주소 검색 등의 경우 NULL)	의료시설
CAR_MODEL	차종("운전습관 가입자" 대상)	CT_NORMAL
CAR_FUEL	유종("운전습관 가입자" 대상)	FT_GAS
CAR_DISPLACEMENT	주행 차량 배기량("운전습관 가입자" 대상)	2199
CAR_MODELYEAR	주행 차량 연식("운전습관 가입자" 대상)	2016

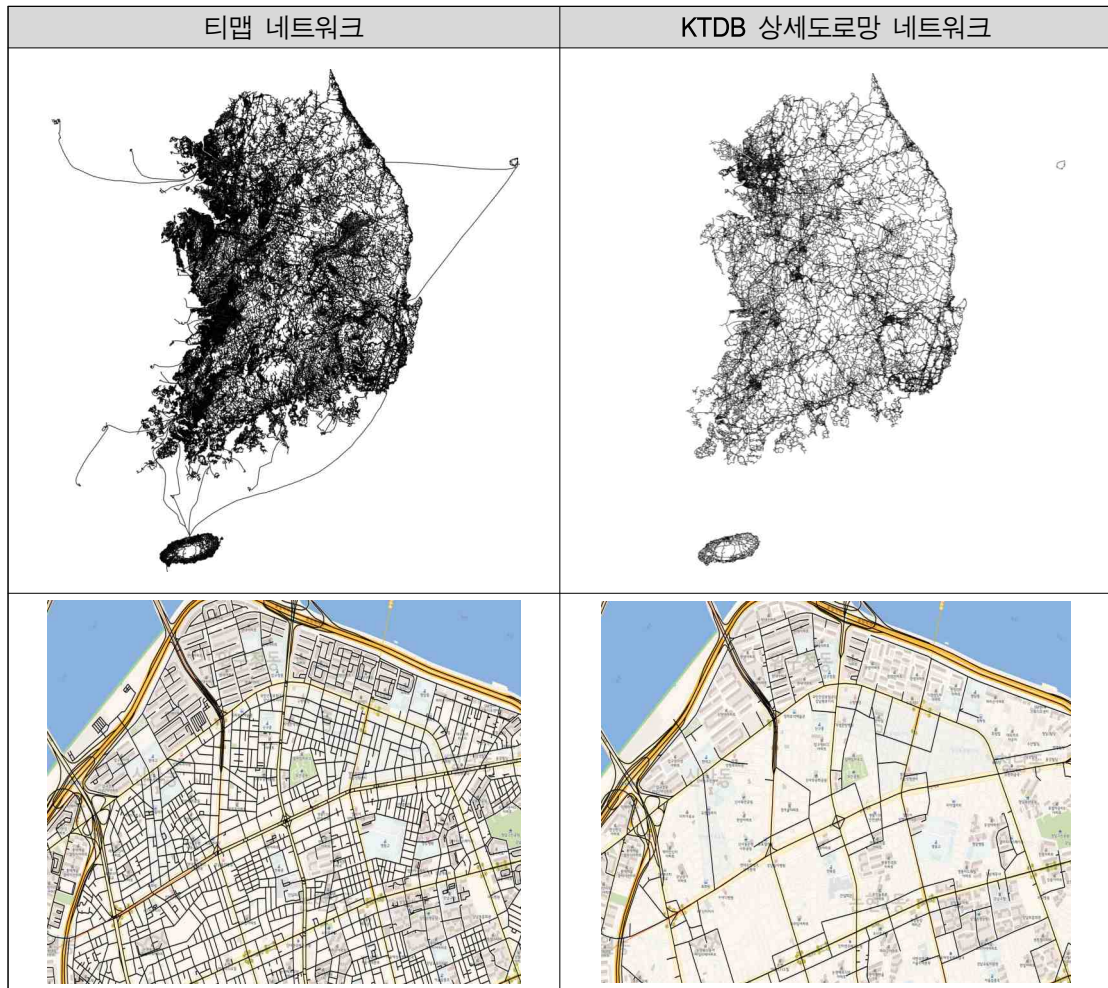
〈표 2-6〉 내비게이션 데이터 차종 정보

구분	차종 정보	차종 구분 기준
CT_COMPACT	경차	경차
CT_NOMAL	승용차	16인승 이하 승합차
CT_MIDDLE	중형승합차	17인승 이상 32인승 이하 승합차, 2.5톤 이상 5.5톤 이하 화물차
CT_LARGE	대형승합차	33인승 이상 승합차, 5.5톤 초과 10톤 미만 화물차
CT_TRUCK	소형화물차	2.5톤 미만 화물차
CT_SPECIAL	대형화물차	10톤 이상 20톤 미만 화물차, 3축 차량
CT_SAMLL_TRUCK	특수화물차	20톤 이상 화물차, 4축 이상 차량

〈표 2-7〉 내비게이션 데이터 유종 정보

구분	유종 정보
FT_GAS	휘발유
FT_DIESEL	경유
FT_LPG	LPG
FT_GASPM	고급휘발유
FT_EV	전기

- 티맵 내비게이션 데이터는 티맵의 도로 네트워크의 링크 기반으로 위치정보를 나타냄
 - 본 과업에서는 상세도로망 네트워크로 기초교통 및 통행지표DB를 구축하기 때문에 티맵 도로 네트워크 링크 기반으로 제공되는 경로를 상세도로망 링크 기반으로 경로 재생성이 필요함
 - 티맵 도로 네트워크의 단방향 링크 연장은 447,610km(2021년 12월 기준)로 KTDB 상세도로망 네트워크 연장인 126,040km(2021년 12월 기준)에 비해 3.5배 상세한 수준인 것으로 분석됨



〈그림 2-1〉 티맵 네트워크와 상세도로망 네트워크 상세도 비교

2) 사업용 운행기록계 데이터

- 사업용 운행기록계(DTG) 데이터는 사업용 차량의 운행정보를 실시간으로 저장하여 변화하는 차량운행 상황을 기록하는 디지털 운행기록계를 통해 수집되는 데이터로 한국교통안전공단으로부터 수집함
 - DTG 데이터는 자동차의 순간속도, GPS, 분당 엔진회전수, 가속도, 자동차 유형 등의 정보가 제공되며, 1초 단위의 GPS 좌표로 수집됨
 - DTG 데이터의 자동차 유형 정보를 이용하여 세분화된 차종의 차량의 이동경로, 통행에 대한 분석을 수행할 수 있음

〈표 2-8〉 DTG 데이터의 포맷

필드명	설명	자료형
TRIP_KEY	등록번호 & 정보발생 일시	string
DTG_MDL_NM	단말기 모델명	string
CHASSIS_NO	차량 고유번호 1	string
CAR_TP_CD	차량 유형 구분	string
CAR_REG_NO	차량 고유번호 2	string
BIZ_REG_NO	운송사업자 등록번호	string
DRIVER_CODE	운전자 코드	string
DLY_DRIV_DIST	일일 주행거리	double
ACCM_DRIV_DIST	누적 주행거리	double
OPT_SPD	차량속도	double
RPM	분당 엔진 회전수	double
BREAK_SIG	1: 유, 2: 무	integer
GPS_X	차량위치 X (WGS84 경위도)	double
GPS_Y	차량위치 Y (WGS84 경위도)	double
GPS_AGL	지점별 방위각	double
ACCEL_VX	횡가속	double
ACCEL_VY	종가속	double
CONT_CODE	통신상태코드	integer
AREA_CODE	행정기관코드, 대존코드	integer
OPT_Time	YYMMDDHHMMSSSS	string

〈표 2-9〉 DTG 데이터의 자동차 유형 코드

구분 코드	자동차 유형	구분 코드	자동차 유형
00	미분류	17	장외차량
11	시내버스	21	일반택시
12	농어촌버스	22	개인택시
13	마을버스	31	일반화물
14	시외버스	32	개인화물
15	고속버스	41	비사업용차량
16	전세버스	-	-

제2절 차량 GPS 원시데이터 분석

가. 데이터 프로파일링

- 2023년 티맵 내비게이션 데이터는 2022년 대비 약 12.8% 증가된 총 793억 건의 개별 차량 주행 발생 건수(이벤트 수)가 수집되었음
- DTG데이터의 개별 차량 발생 건수는 전년 대비 버스 약 0.2% 감소하였고, 화물차 약 11% 증가, 택시 48% 증가하였음
 - 택시 데이터의 추출 오류로 인해 5개월치 데이터를 다시 받았고, 사업자에서 제출한 데이터가 늘어난 만큼 개별 차량 발생 건수가 증가한 것으로 분석됨

〈표 2-10〉 데이터의 용량 및 개별 차량의 링크 발생 건수

데이터명	용량(TB)		이벤트 수(천 건)	
	2022년	2023년	2022년	2023년
링크 기반 내비게이션 데이터	19.4	22.3	70,349,051	79,364,038
DTG 버스 데이터	9.64	9.22	772,320,634	770,774,514
DTG 화물차 데이터	3.93	4.32	317,497,577	355,365,816
DTG 택시 데이터	5.91	10.24	491,244,500	953,355,667

- 티맵 내비게이션 데이터의 차종과 유종 정보는 티맵 전체 가입자가 아닌 자동차보험 할인을 위해 “운전습관 가입자”가 제공하는 정보이므로 100% 포함하지 않는 한계점이 존재함
- 티맵 내비게이션 데이터를 통한 차종 정보 분석 결과
 - 내비게이션 데이터의 주행 고유 ID별로 차종 정보가 포함된 비율을 분석한 결과, 75% 이상의 데이터에서 차종 정보가 확인되었음
 - 승용차가 약 68%로 가장 많이 발생하였고, 2022년과 비교하여 14% 증가한 것으로 분석됨

〈표 2-11〉 티맵 내비게이션 데이터 차종 분석 결과

차종 정보	2022년	2023년
승용차	55.2%	68.6%
경차	2.4%	3.0%
중형승합차	1.7%	1.8%
대형승합차	0.5%	0.6%
소형화물차	0.5%	0.8%
대형화물차	0.03%	0.1%
특수화물차	0.02%	0.3%
정보 없음	39.6%	24.9%
합계	100%	100%

○ 티맵 내비게이션 데이터를 통한 유종 정보 분석

- 내비게이션 데이터의 주행 고유 ID별로 유종 정보가 포함된 비율을 분석한 결과, 약 49% 이상의 데이터에서 유종 정보가 확인되었음
- 휘발유에서 약 51%로 가장 많이 발생하였고 2022년과 비교했을 때 10% 증가한 것으로 분석됨

〈표 2-12〉 티맵 내비게이션 데이터 유종 분석 결과

유종 정보	2022년	2023년
휘발유	41.4%	51.3%
경유	16.6%	20.1%
LPG	4.7%	5.6%
고급휘발유	1.1%	1.3%
전기	1.1%	1.5%
정보 없음	35.3%	20.3%
합계	100%	100%

- 차량 내비게이션 데이터의 용량 및 경로 수 현황
 - 데이터의 총 용량은 20.28TB이며, 총 793억 건의 이벤트 수가 수집되어 개별 차량 경로 수는 총 7.1억 건의 경로가 수집되었음
 - 2022년 데이터와 비교하여 데이터 용량은 약 8TB 증가, 이벤트 수는 약 90억 건 증가, 경로 수는 약 0.9억 건 증가한 것으로 분석됨

〈표 2-13〉 연도별 내비게이션 데이터 프로파일링 비교

구분	2021년	2022년	2023년	'21년 대비 '22년 증감율	'22년 대비 '23년 증감율
데이터 용량	7.2TB	19.4TB	20.28TB	145.1%	4.5%
이벤트 건수	586억 건	703억 건	793억 건	35.0%	12.8%
경로 수	4.5억 건	6.2억 건	7.1억 건	15.1%	14.9%

○ DTG데이터의 자동차 유형 정보를 이용하여 차종 분석

- DTG데이터는 세분화된 차종을 포함하고 있으나, 교통량 추정할 때는 버스, 화물차에 대한 정보를 집계하여 활용함
- 이는 공공에서 수집하고 있는 관측교통량 데이터의 차종 구분이 세분화되지 않아 차종을 집계하여 활용함
- DTG 데이터의 자동차 유형 정보를 이용하여 차종을 분석함
 - 택시 데이터는 하반기 데이터가 오류로 인해 재요청하여 다시 받았으며, 개인택시의 비율이 작년과 비교하여 61%나 증가한 것으로 나타났음
 - 택시 데이터는 공공에서 조사하는 관측교통량 데이터에는 택시로 차종이 구분된 정보가 없어 택시를 교통량 추정 알고리즘에 사용하고 있지 않음

〈표 2-14〉 DTG데이터 차종 분석 결과

차종 정보		주행 고유 ID 대수 (천대)	비율
버스	시내버스	61,389	44.9%
	농어촌버스	12,994	9.5%
	마을버스	12,504	9.1%
	시외버스	4,141	3.0%
	고속버스	1,636	1.2%
	전세버스	44,100	32.2%
	계	136,764	100%
택시	일반택시	78,025	96.6%
	개인택시	2,758	1.3%
	계	80,782	100%
화물	일반화물	73,715	35.6%
	개인화물	133,381	64.4%
	계	207,096	100%

- DTG데이터의 용량 및 이벤트 건수 현황

- DTG데이터의 용량이 과다하여 압축해제 후 가공하는 과정에서 시간 소요가 많아 2023년 데이터부터 압축 해제 없이 가공할 수 있도록 여러 개로 분할된 gz파일로 데이터 추출 요청함
- 압축데이터의 총 용량은 23.7TB이며, 총 191억 건의 궤적 이벤트 수가 수집되었음
- 택시 데이터는 하반기로 갈수록 데이터의 이벤트 건수가 많은 것으로 나타났고, 버스와 화물은 월별로 비슷한 이벤트 건수가 발생하였음

〈표 2-15〉 DTG 데이터 현황

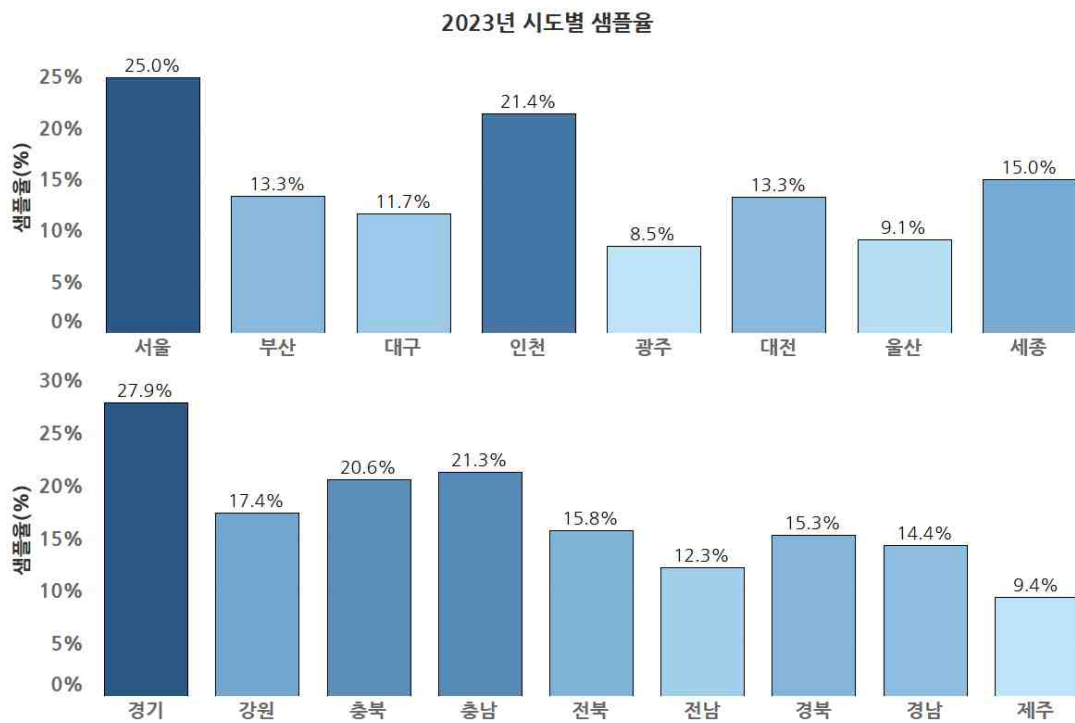
구분	버스		화물		택시	
	용량	이벤트 수(억 건)	용량	이벤트 수(억건)	용량	이벤트 수(억건)
1월	798.3 GB	62.7	328.0 GB	26.3	496.1 GB	40.8
2월	750.4 GB	58.6	324.8 GB	25.8	476.5 GB	39.1
3월	826.2 GB	64.1	349.3 GB	27.4	673.6 GB	54.8
4월	839.3 GB	64.9	379.1 GB	29.6	542.7 GB	81.3
5월	858.7 GB	66.0	401.3 GB	31.4	904.0 GB	72.5
6월	803.5 GB	62.0	392.3 GB	31.1	911.8 GB	74.1
7월	650.4 GB	73.7	371.0 GB	29.9	996.2 GB	93.2
8월	841.6 GB	66.1	387.6 GB	34.1	1119.0 GB	93.8
9월	750.8 GB	61.7	334.1 GB	26.7	1146.9 GB	110.5
10월	854.0 GB	72.9	384.7 GB	30.6	1186.8 GB	110.1
11월	760.5 GB	61.6	390.0 GB	31.4	862.5 GB	83.6
12월	711.8 GB	56.6	379.1 GB	31.1	1173.2 GB	99.6
합계	9.22 TB	770.8	4.32 TB	355.4	10.24 TB	953.4

-

나. 데이터 샘플을 분석

○ 티맵 내비게이션 데이터를 통한 관측지점 샘플을 분석 결과

- 내비게이션 데이터를 활용하여 시도별 연평균 일 프로브 샘플을 분석함
 - 공공에서 수집되는 관측교통량 데이터와 연평균 일 통행량 데이터를 활용하여 샘플을 산출하였음
 - 상세도로망 네트워크 기반 2023년 관측교통량 DB를 활용하여 분석하였음
 - 시도별로 샘플을 분석했을 때, 수도권(경기, 서울, 인천)과 충청권, 세종시가 상대적으로 높은 샘플을 보이는 것으로 분석됨



〈그림 2-2〉 시도별 연평균 일 내비 프로브 샘플



제3장 차량 GPS 데이터 가공 및 DB구축

제1절 차량 DB 가공 및 구축 프로세스

제2절 차량 GPS 데이터 전처리 및 경로 가공

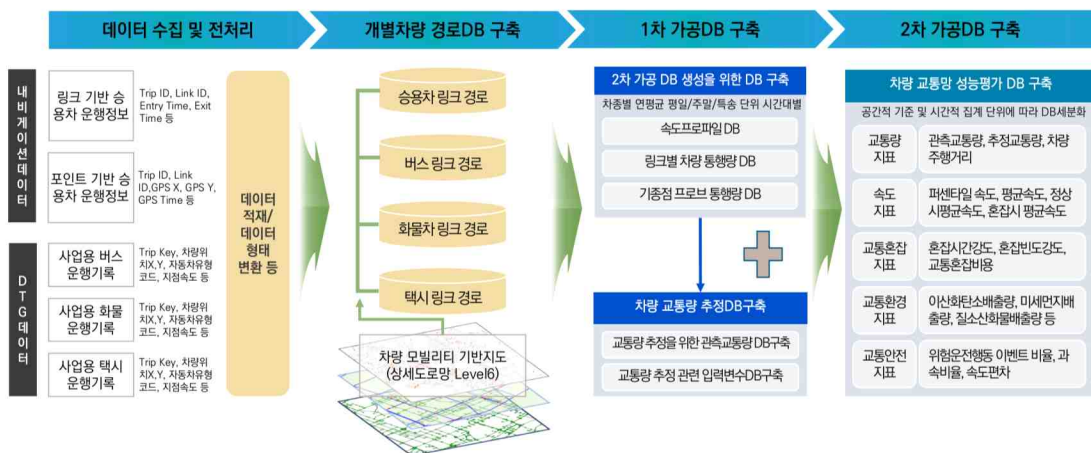
제3절 기초교통 DB 구축

제4절 차량 통행지표 구축

제3장 차량 GPS 데이터 가공 및 DB구축

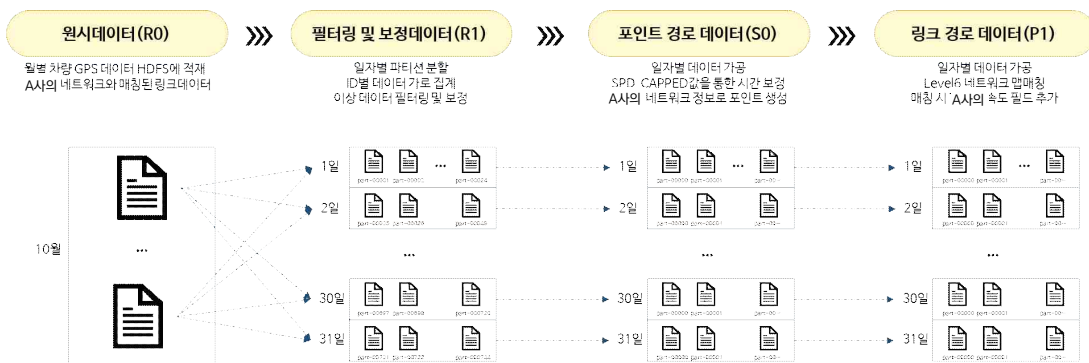
제1절 차량 DB 가공 및 구축 프로세스

- 본 사업에서 활용하고 있는 차량 GPS 데이터는 내비게이션 및 사업용 DTG 데이터로, 이를 활용하기 위해서는 제공되는 데이터의 특성을 분석하여 각 특성에 맞춰 데이터 가공 및 검증이 필요함
- 티맵 내비게이션 데이터은 링크 기반 데이터이고, DTG 데이터는 포인트 기반 데이터로 각 데이터의 제공되는 데이터 형식 및 수집 주기가 다르기 때문에 특성에 맞는 전처리 및 가공 모듈 적용을 수행함
- 차량 GPS 데이터 가공 및 구축 프로세스는 데이터 전처리 → 개별차량 경로 DB구축 → 1차 가공DB(기초교통DB) 구축 → 2차 가공DB(교통망 성능평가지표DB) 구축 순으로 진행



<그림 3-1> 차량 GPS 데이터 가공 및 구축 프로세스

- 티맵 내비게이션 데이터는 링크 기반 데이터로 링크를 1초 단위 궤적(포인트) 정보로 생성하여 표준 포인트 경로 데이터를 구축한 뒤 Level6 네트워크와 맵 매칭하여 링크 경로 데이터를 생성함
- 티맵 네트워크의 링크는 Level6 네트워크보다 상세하므로 데이터의 공간적 범위에 맞게 구축하기 위해 궤적(포인트)정보를 생성함



〈그림 3-2〉 링크 기반 데이터 전처리 및 경로 데이터 가공 프로세스

- DTG 데이터는 포인트 기반 데이터로 포인트 기반 데이터 맵매칭 및 경로 가공 모듈을 이용하여 가공 및 구축함(해당 모듈은 R&D사업을 통해 개발된 것임)



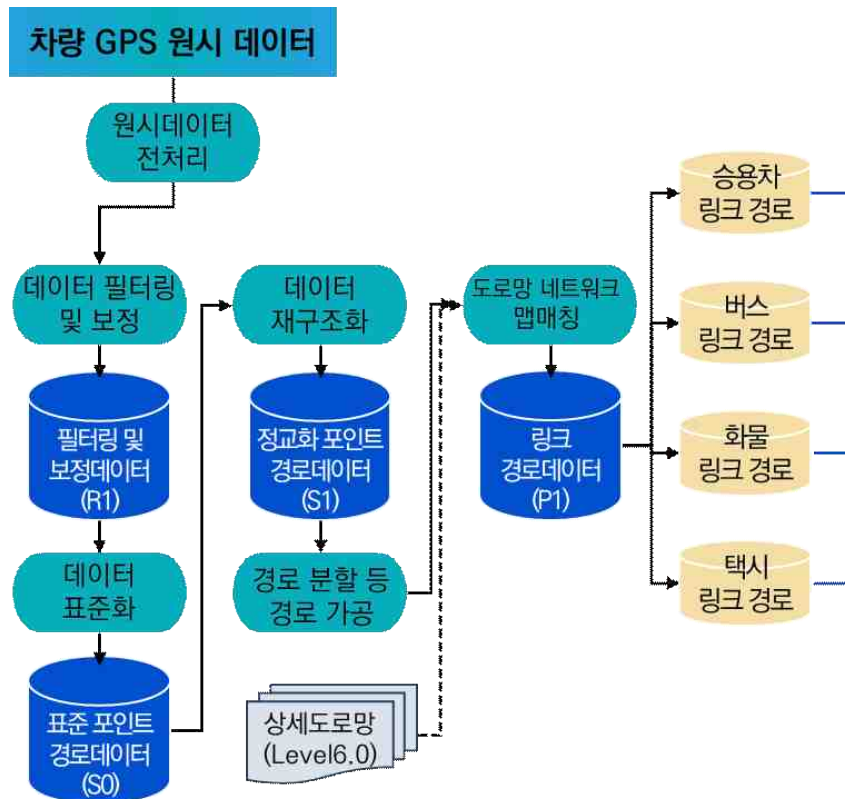
〈그림 3-3〉 포인트 기반 데이터 전처리 및 경로 데이터 가공 프로세스

제2절 차량 GPS 데이터 전처리 및 경로 가공

1. 링크 단위 티맵 내비게이션 데이터 전처리 및 경로 가공

가. 개별차량 경로DB 구축 프로세스

- 차량 GPS 데이터를 활용하여 개별차량 경로DB를 구축하는 프로세스는 다음과 같은 과정으로 수행됨
 - 링크 기반 데이터는 맵 정보와 매칭되어 있어 궤적정보를 정교화하는 데이터 재구조화 과정이 없음
 - 포인트 기반 데이터는 GPS 수신기 불안정한 궤적에 대한 보정을 수행하는 재구조화 과정이 포함됨



<그림 3-4> 차량 GPS 데이터를 활용한 개별차량 경로DB 구축 프로세스

나. 원시데이터 전처리

○ 원시데이터 적재 및 전처리

- 대용량의 압축 파일을 빅데이터 가공 시스템으로 처리하기 위해 원시 데이터를 분할하여 HDFS에 적재함



〈그림 3-5〉 데이터 적재

- 원시데이터를 일자별로 분리하고, 불필요한 속성정보를 제거하여 데이터 용량을 축소시킴



〈그림 3-6〉 자료형 변환

- 빅데이터 환경의 처리속도 향상을 위해 데이터 포맷을 변환



〈그림 3-7〉 데이터 포맷 변환

다. 데이터 필터링 및 보정

○ 내비게이션 데이터 필터링 및 보정

- 아래의 오류 유형에 따라 필터링 및 보정을 수행함

- (시간 오류) 주행 시작일 필드 오류 유형 : 주행 시작일(R_P_REQDATE)과 링크 진입 시간(ENTRY_TIME) 필드의 일자가 맞지 않는 경우 링크 진입시간을 기준으로 주행 일자에 대해 정렬하여 일자별로 분할함

DATE	TRIP_INIT_SESSION_ID	LINK_SEQ	MESH_CD	LINK_ID	LINK_DIR	LENGTH	DRIVING_TIME	ENTRY_TIME	EXIT_TIME
20210208	105520210208235734644500 ...	1	4787	16033	0	63.894627	6.790106	20210209000226	20210209000232
20210208	105720210208235718506500	1	5631	114	0	121	44.565998	20210209000653	20210209000737
20210208	108120210208232709308500	1	4777	3744	1	26	2.8363636	20210209073336	20210209073339
20210209	105020210209175213265501	1	7413	28853	1	84	27.49091	20210210030545	20210210030559
20210209	105020210209210128425500	1	4775	9789	1	231.51741	11.883713	20210210084021	20210210084032

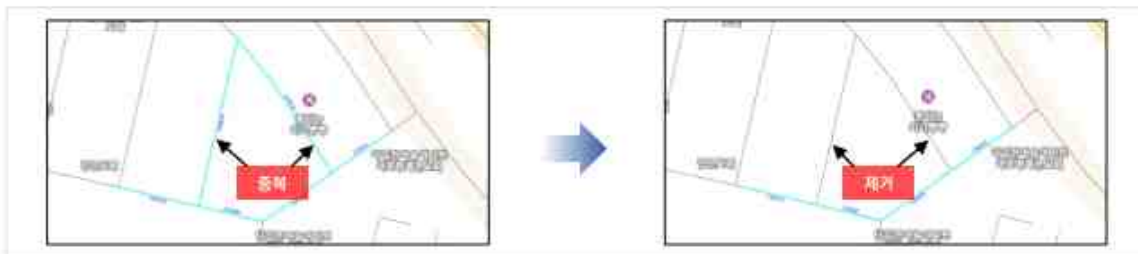
<그림 3-8> 주행 시작일 필드와 링크 진입시간 필드 날짜 불일치 오류 예시

- (시간 오류) 링크 시퀀스 필드 오류 유형 : 링크 진입시간(ENTRY_TIME)과 링크 진출 시간(EXIT_TIME) 필드의 시간이 링크 시퀀스(LINK_SEQ) 순서와 다를 경우 진입시간, 진출시간, 링크 시퀀스 순서로 정렬함

DATE	TRIP_INIT_SESSION_ID	LINK_SEQ	MESH_CD	LINK_ID	LINK_DIR	LENGTH	DRIVING_TIME	ENTRY_TIME	EXIT_TIME
20210208	105520210208235734644500 ...	1	4787	16033	0	63.894627	6.790106	20210209000226	20210209000232
20210208	105720210208235718506500	1	5631	114	0	121	44.565998	20210209000653	20210209000737
20210208	108120210208232709308500	1	4777	3744	1	26	2.8363636	20210209073336	20210209073339
20210209	105020210209175213265501	1	7413	28853	1	84	27.49091	20210210030545	20210210030559
20210209	105020210209210128425500	1	4775	9789	1	231.51741	11.883713	20210210084021	20210210084032

<그림 3-9> 링크 진입시간 필드와 링크 진출시간 필드 시간 차이 오류 예시

- (필드 오류) 링크 중복 오류 유형 : 동일 시간에서 발생한 링크의 중복 경로를 보정함



<그림 3-10> 링크 중복 경로 예시

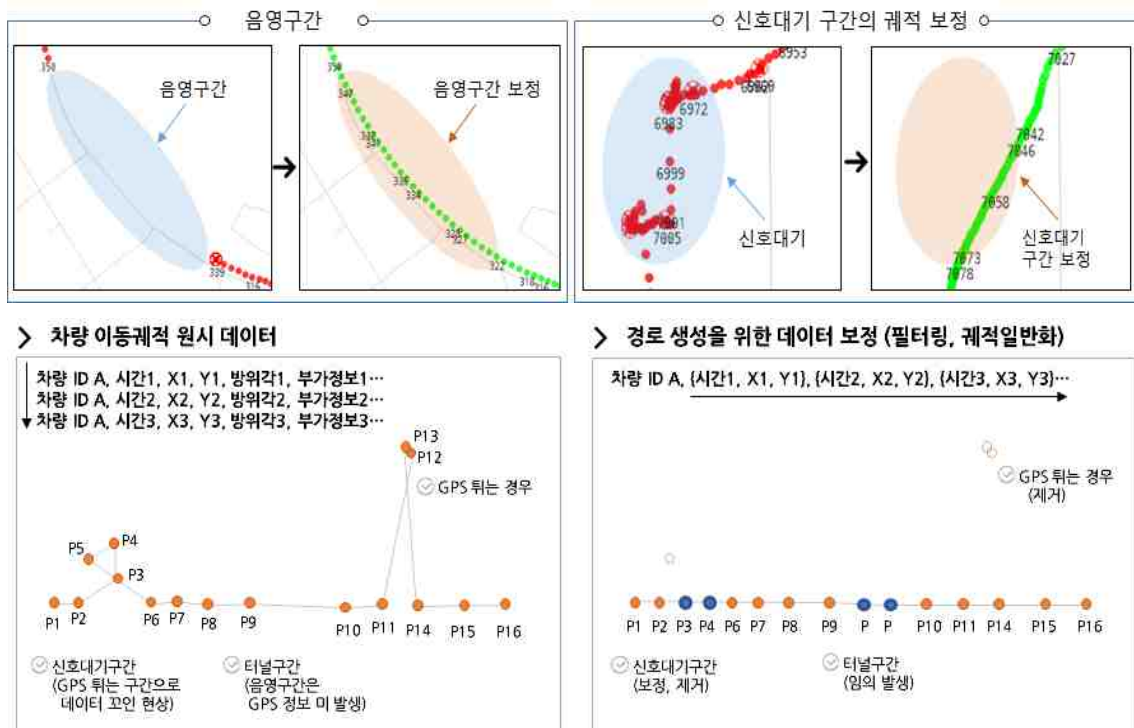
- (속도 오류) 속도 필드 오류 유형 : 속도 제한이 없는 상태에서 속도를 계산 시 이상값이 나타나는 경우(GPS가 미존재하여 실제 속도 값이 계산되지 않는 경우가 있어 제한

구조로 DB를 재구성함

마. 데이터 재구조화

○ DTG 데이터의 이상치 판단 및 오차구간 보정

- DTG 데이터는 음영구간(지하차도, 터널, 고가 밑 등) 및 고층 빌딩 주변, 신호대기 상태에서는 GPS 수신에 불안정하여 좌표가 튀는 형태가 발생할 수 있어, 재구조화 알고리즘을 적용하여 GPS 궤적에 대한 이상치 여부를 판단 후 데이터의 오차구간을 보정하는 정교화 과정을 진행함
- 국부적 오차 보정 : 차량 GPS 궤적에 대한 이상치 판단 및 연결성이 끊어진 궤적에 대해 보정작업을 수행함
- 음영구간 판단 및 재구성 : 유형별 음영구간 진입 여부 판단 후 재구성 작업을 수행



<그림 3-12> 이동궤적(포인트) 데이터 보정

- 재구성된 GPS 궤적에 대한 분리 : 차량의 정차시간, 거리차, 회차 등에 대한 조건에 따라 기준을 정의하여 GPS 궤적을 분리하는 작업을 수행함

○ DTG 데이터의 출·도착 구분 및 분할

- DTG데이터는 단말기가 종료되지 않으면 공간좌표정보가 계속 수집되므로 1일 기준으로 출발과 도착을 구분할 수 있는 분할 작업이 필요함



〈그림 3-13〉 통행 분리되지 않은 차종별 궤적 예시

- 출·도착 분할에 대한 작업을 위해 데이터를 분석한 결과, 목적지까지 정차 없이 도착하는 경우와 통행 중에 경유지를 거쳐 목적지에 도착하는 경우, 장거리 운행 중 휴게소에 정차나 주유, 배송업무 등으로 인해 반복 정차하는 경우로 나타났음
- 특정 시간 이상 위치정보를 수집하지 않을 때 해당 구간을 기준으로 통행을 분리함
- GPS 위치 전 후를 비교하여 좌표가 동일할 때 GPS 위치 후반으로부터 50m 이내의 GPS 궤적 개수를 합하여 300개 이상(5분 이상)이면 그 위치로부터 통행을 분리함
- 버스, 트럭 데이터는 노선을 반복해서 통행하거나 출발지로 회귀하는 특성이 존재하므로 차고지와 회차지를 판별하고 통행을 분리하기 위해 통행분리 지점을 집계하여 시간과 공간이 일치하는 조건에 따라 통행 분리 지점을 판단하여 분리함

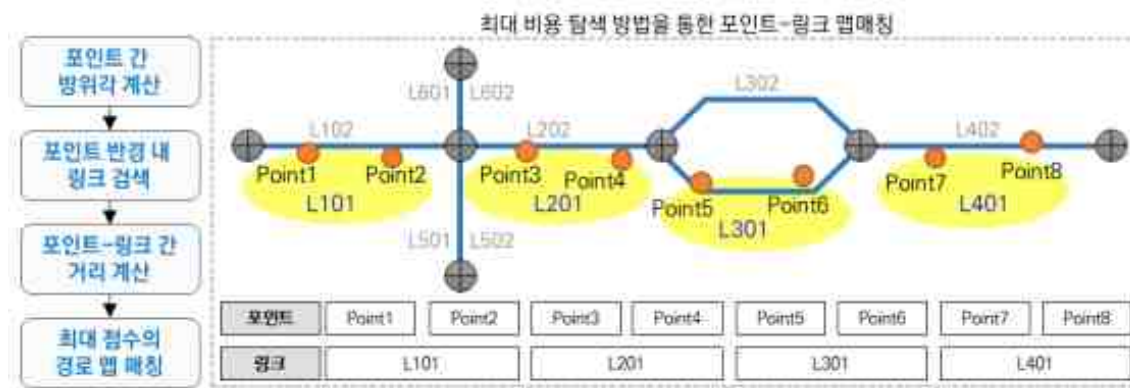
바. 경로 가공 및 맵매칭

○ 포인트 데이터의 경로 가공 및 맵매칭

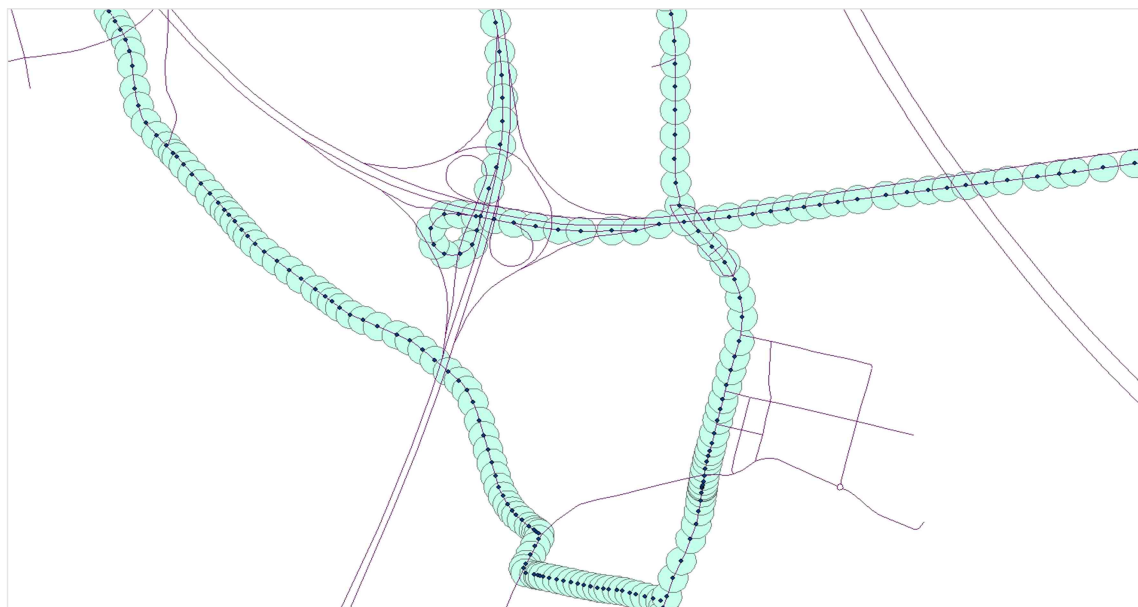
- 링크 기반 내비게이션을 상세도로망 네트워크와 매칭시키기 위해 1초 단위 포인트 데이터로 변환 후 도로망 네트워크와 맵매칭함
- 본 사업의 공간적 범위인 상세도로망 Level6네트워크 기준으로 맵 매칭 후 링크

경로 데이터를 생성함

- 차량 GPS 데이터의 위치 정보를 기준으로 각 포인트의 진행 방향각 정보를 계산함. 방향각은 진북을 기준으로 시계방향인 우측 방향 각도를 계산
- 포인트 자료 주변 도로 네트워크(링크) 검색. 검색된 링크를 대상으로 포인트와 링크와의 최단 거리, 링크에서의 방위각 계산, 링크를 따라서 이동한 거리 정보를 계산함
- 모든 포인트는 주변에 검색된 링크를 대상으로 거리, 방위각 비용, 현재 링크와 다음 링크와의 전환 점수를 계산하게 되며 모든 점수가 최대 비용을 나타내는 경로를 선정 및 적재



<그림 3-14> 포인트-링크 맵매칭

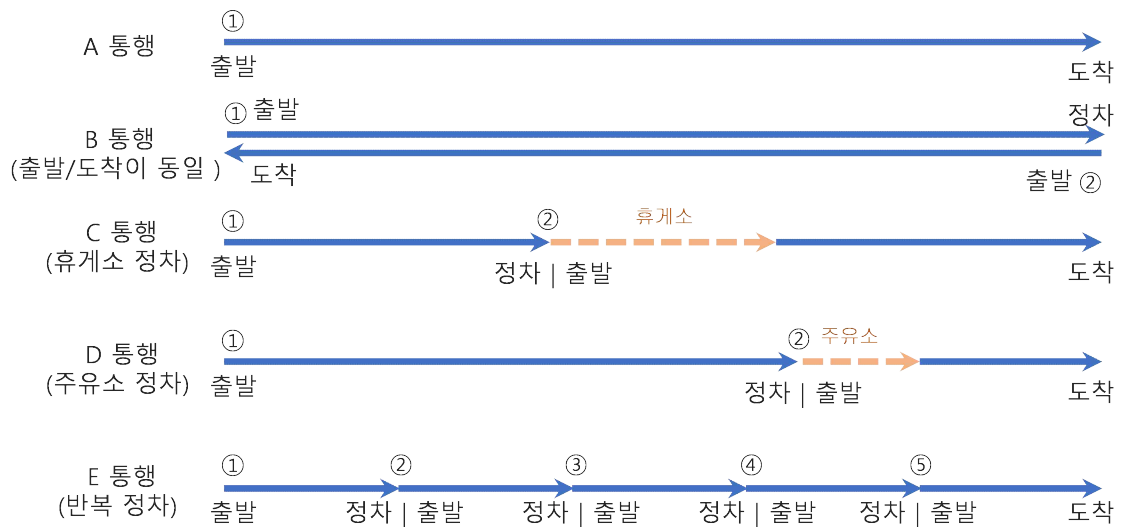


<그림 3-15> 포인트와 도로망 네트워크 맵 매칭 예시

- 사용자가 휴게소에서의 내비게이션 종료 또는 실행 시 휴게소를 기점으로 통행이

분리될 수 있기 때문에 통행의 시종점으로 적합하지 않는 구간은 연속된 통행을 유지할 수 있도록 처리함

- 통행이 병합되는 경우는 휴게소-휴게소, 고속도로-고속도로, 도시고속도로-고속도로 조합의 경우 발생
- 통행 병합 시 시간 정보 및 구분 코드 정보를 추가하여, 데이터 분석 및 검증 시 활용할 수 있도록 함



〈그림 3-16〉 경로 분할 및 통합 기준

○ 맵매칭된 링크 경로 데이터를 통해 링크 통행속도 산출

- 링크와 매칭된 포인트 정보의 도로구간 길이와 시간정보를 이용하여 링크별 통행속도를 산출함

선택	순번	포인트시간	위도	경도	링크ID	링크출발거리	링크남은거리	GIS...	GIS...	GPS...	GPS...
원	125	2019/05/15 20:35:13	967962.028	1951976.268	57276491201	88.49	182.19	230.68	87.53	195.56	88.00
원	126	2019/05/15 20:35:14	967985.671	1951984.697	57276491201	93.49	137.20	230.68	87.53	195.56	88.00
원	127	2019/05/15 20:35:15	968009.154	1951991.78	57276491201	118.00	112.68	230.68	87.53	195.56	88.00
원	128	2019/05/15 20:35:16	968032.637	1951997.807	57276491201	170.00	61.98	230.68	87.53	195.56	88.00
원	129	2019/05/15 20:35:17	968056.120	1952001.929	57276491201	222.00	11.28	230.68	87.53	195.56	88.00
원	130	2019/05/15 20:35:18	968079.603	1952005.186	57276491201	274.00	41.68	230.68	87.53	195.56	88.00
원	131	2019/05/15 20:35:19	968103.086	1952007.398	57276491201	326.00	18.62	230.68	87.53	195.56	88.00
원	132	2019/05/15 20:35:20	968126.569	1952008.604	57276491201	378.00	234.82	230.68	87.53	195.56	88.00

<그림 3-17> 포인트 정보와 매칭된 링크의 통행속도 산출

○ 개별 차량 경로DB 구축

- 링크와 매칭된 경로 데이터를 통행지표 생성을 위한 기초 DB로 활용됨
- 단말기 ID별로 1일 단위의 개별 차량의 경로 데이터 구축

<표 3-1> 개별차량 경로DB 테이블 정의서

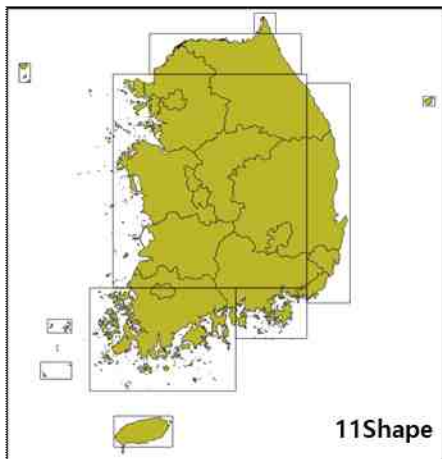
컬럼명	설명
단말기 ID	단말기 가상키로 개별 차량을 의미하며, 1일 단위로 단말기ID가 변경됨
경로그룹번호	통행이 발생한 이동 경로를 그룹화하였으며, 5분 이상 정차 시 그룹분리, 경로병합 적용
차종	Auto : 티맵 내비게이션 Bus : DTG 버스 Truck : DTG 화물 Taxi : DTG 택시
순서	경로 그룹 내 이동이 발생한 순서
날짜	차량의 이동궤적이발생한 날짜 및 시간
링크ID	상세도로망 네트워크의 링크ID
진입시간	링크 내 차량의 이동궤적이발생한 처음 시간이며, 시분초로구성됨
진출시간	링크 내 차량의 이동궤적이발생한 끝 시간이며, 시분초로구성됨
링크길이	상세도로망 네트워크의 링크 길이
링크속도	상세도로망 네트워크의 링크 길이로 계산된 속도

2. 포인트 단위 DTG 데이터 전처리 및 경로 가공

가. DTG 데이터 전처리 및 표준화

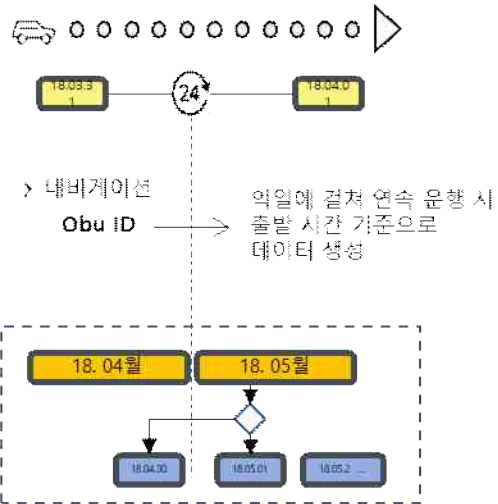
- DTG 원시데이터를 압축해제 및 데이터의 이상유무를 판별하고, HDFS에 적재
- DTG 원시데이터의 오류 유형을 파악하고, 유형별로 보정 방식을 적용하여 전처리
 - 좌표오류 : 수집된 포인트 데이터가 대한민국 육지 영역을 벗어나는 데이터 → 오류 좌표 데이터를 제거하여 보정
 - 시간오류 : 제공 기준 일자와 다른 시간이 입력된 데이터 → 오류 시간 데이터를 제거하여 보정

- ⊙ 좌표이상 데이터 제거
 - 내륙 및 섬 지역 영역 설정
 - 설정영역 이외의 데이터 제거



> 1일 약 2만 건(0.00003%) 공간이상 데이터 발생

- ⊙ 날짜이상 데이터 제거
 - 1일 단위 데이터 생성 - 출발 시간 기준



> 설정한 수집시간 외의 데이터는 제거

〈그림 3-18〉 좌표오류 및 시간오류 필터링 방법론

- 중복오류 : 동일 시간에 여러 데이터가 발생한 경우 → 중복 데이터를 제거하여 보정

- 2023년 DTG 데이터의 전체 이벤트 수 대비 오류 이벤트 수의 비율은 약 6.3%로 확인됨
 - 2022년 대비 2023년에 전체 이벤트 건수가 증가하였고, 오류 이벤트 건수도 증가하여 그에 따라 필터링됨
 - 링크 단위 내비게이션 데이터보다 오류 비율이 높은 이유는 포인트의 GPS 좌표 수집 시 오류 발생 확률이 더 높기 때문

〈표 3-2〉 DTG 데이터 오류 발생 필터링 비율

연도	전체 이벤트 수	오류 이벤트 수	필터링 비율
2021년	1,520,275,548,153	83,114,871,158	5.4%
2022년	1,581,062,712,184	76,782,907,716	4.8%
2023년	2,079,495,998,909	130,371,787,790	6.3%

- 통행정보 이외의 불필요 컬럼을 제거하고, 차종 정보를 포함하여 전처리함으로써 용량 축소
 - 2023년 DTG 원시데이터 전처리에 약 28일 소요되었으며, 작년에는 약 92일 10시간 소요된 것과 비교하여 전처리를 위한 자원 증가를 통해 약 1/4 감소시켰음
 - 이 중 데이터 적재에 약 80%의 시간이 소요됨
 - 전처리 과정을 통해 용량이 약 66%로 줄어듦
- 개인식별 정보 제거 및 전처리 결과 데이터를 DB 형태로 변환 및 저장

〈표 3-3〉 DTG 전처리 데이터 DB

K-V	순번	필드명	필드설명	자료형	비고
K	1	OBU_ID	주행 고유 ID	string	
	2	VEH_TYPE	차종	string	
V	3	TIME	포인트 시간	long	Unix Time
	4	X	X좌표	double	UTM-K
	5	Y	Y좌표	double	UTM-K

나. DTG데이터 필터링 및 보정 프로세스

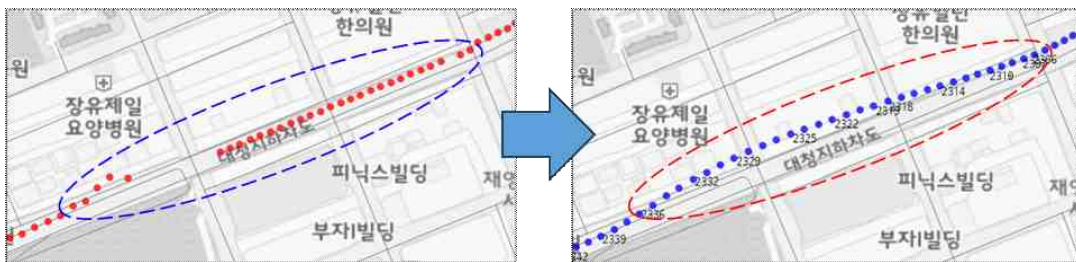
- 전처리 결과로 생성된 DB로부터 궤적별 차량의 출발/도착을 구분하여 분할
 - 차량의 궤적 데이터는 1일 기준으로 데이터가 연속 수집되기 때문에 출·도착 기준의 통행분리가 필요함
 - 포인트 단위인 DTG 데이터의 경우 링크 단위 데이터에 비해 좀 더 상세하게 통행분리 유형을 판별할 수 있음
- 사업용 차량의 출발·경유·도착에 대한 판단 및 통행분리를 위해 차량 유형별로 데이터를 분석했을 때 다음과 같이 구분됨
 - 목적지까지 정차없이 도착하는 통행
 - 장거리 운행 중 휴게소에 정차 및 주유하기 위한 정차나 배송업무 등으로 인해 반복 정차하는 경우 등 경유지를 거쳐 목적지에 도착하는 통행
 - 차고지에서 시동을 끄지않고 정차 및 차고지를 계속 주행하는 경우는 운반한 물건을 적재하는 과정의 통행
 - 회차지에서 정차하지 않고 주행하거나 몇 시간 이상 정차하는 등의 통행



〈그림 3-19〉 통행분리 유형 판단을 위한 데이터 분석

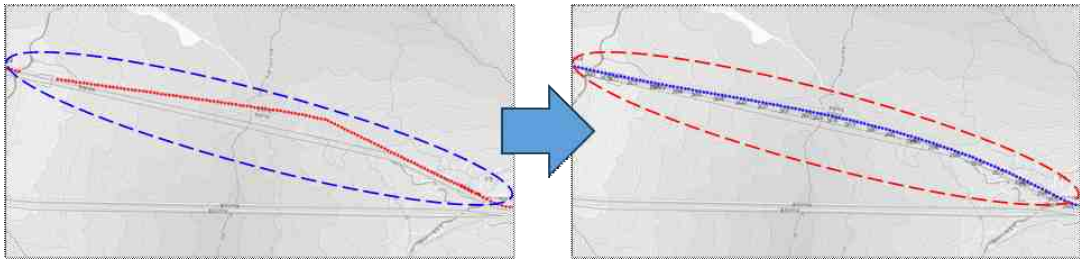
- 사업용 차량의 통행을 목적별로 구분하기에 다양한 유형이 존재하므로 특정 기준에 해당되면 통행을 분리함
 - 이전 포인트와 현재 포인트를 비교하여 좌표가 동일할 시, 현재 포인트로부터 50m 이내의 포인트 개수를 카운트. 카운트된 포인트의 개수가 300개 이상 (5분 이상)일 시, 해당 지점으로부터 통행을 분리
 - 버스, 트럭 데이터는 노선을 반복해서 통행하거나, 출발지로 회귀하는 특성이 존재하므로 차고지, 회차지를 판별 및 통행분리를 위해 시간/공간적 조건을 통한 통행분리 지점을 집계하여 집계 결과를 이용한 통행분리 지점 판단 및 분리

- GPS 포인트의 오차 범위 및 주변 환경에 따라 발생하는 오류를 줄여 통행 정보, 속도 등의 지표 산출 시 오차를 줄이기 위해 보정 작업을 진행
 - (지하차도) 지하차도와 같은 음영 구간에서는 GPS 데이터가 수신되지 않을 수 있으므로, 이 경우 내비게이션은 인위적인 더미 데이터(왼쪽)를 생성함. 이는 실제 차량의 움직임과 다른 데이터이므로 더미 데이터를 제거 후 음영 구간의 데이터를 재생성(오른쪽)하여야 함



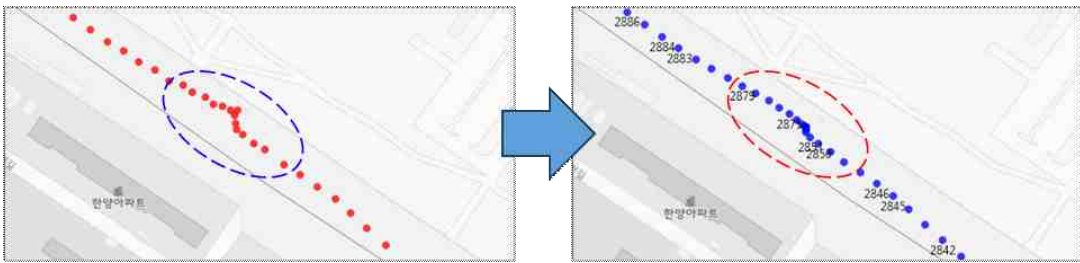
〈그림 3-20〉 지하차도 음영구간의 데이터 재생성

- (터널 및 고가 밑) 터널에서는 GPS 데이터가 수신되지 않을 수 있으므로, 이 경우 내비게이션은 인위적인 더미 데이터(왼쪽)를 생성함. 이는 실제 차량의 움직임과 다른 데이터이므로 더미 데이터를 제거 후 음영 구간의 데이터를 재생성(오른쪽)하여야 함



〈그림 3-21〉 터널 음영구간의 데이터 재생성

- 차량이 신호대기 임시 정차 시 GPS 좌표가 불안정하게 되는 형태가 발생하는 데 튀는 형태는, 데이터가 없거나, 한 곳에서 빙글빙글 도는 패턴 등을 가짐. 이는 차량의 움직임과 다르므로 차량의 감속/정지/가속 단계로 정지 구간 데이터를 재생성(오른쪽)하여야 함



〈그림 3-22〉 정지 시 꼬임구간의 데이터 재생성

- 2022년 DTG 데이터 필터링 및 보정에 약 9일 소요되었으며, 약 12배 용량이 증가함
- 필터링 및 보정 데이터를 표준DB 형태로 변환 및 저장

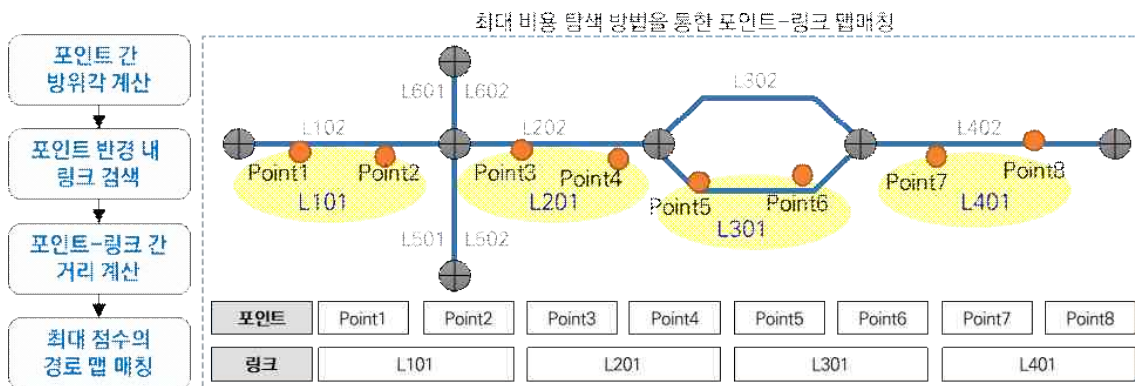
〈표 3-4〉 DTG 필터링 및 보정 데이터 표준 DB

K-V	순번	필드명	필드설명	자료형	비고
K	1	OBUID	통행 ID	string	
	2	VEH_TYPE	차종	string	
	3	AVG_GAP	음영구간을 제외한 원본-Recon 포인트 간 거리 오차의 평균	double	
	4	RELIABILITY1	이상치 판단 후 궤적 신뢰성	double	
	5	RELIABILITY2	필터링 후 궤적 신뢰성	double	
V	6	TIME	포인트 시간	long	Unix Time
	7	X	경도	double	UTM-K
	8	Y	위도	double	UTM-K
	9	SPD	포인트 속도	double	
	10	ACC	포인트 가속도	double	
	11	TYPE	포인트 유형	double	

3. 경로 데이터 구축 및 경로 압축(통합)

가. 상세도로망 기반 경로 데이터 가공 및 DB 구축

- 티맵과 DTG의 상세도로망 기반 경로 데이터 가공 및 DB 구축 프로세스는 동일함
- 본 사업에서 구축하는 기초 교통 및 통행지표 DB의 공간적 범위는 상세도로망 네트워크 기준으로 필터링 및 보정된 표준DB와 도로망 네트워크를 맵매칭 후 링크 기준의 경로 데이터 생성 필요
 - 차량 GPS 데이터의 위치 정보를 기준으로 각 포인트의 진행 방향각 정보 계산. 방향각은 진북을 기준으로 시계방향인 우측 방향각도를 계산
 - 포인트 자료 주변 도로 네트워크(링크) 검색. 검색된 링크를 대상으로 포인트와 링크와의 최단 거리, 링크에서의 방위각 차이, 링크를 따라서 이동한 거리 정보를 계산하여 점수를 부여
 - 모든 포인트는 주변에 검색된 링크를 대상으로 거리 비용, 방위각 비용, 전환 비용을 통해 모든 점수가 최대 비용을 나타내는 경로를 선정 및 적재



〈그림 3-23〉 맵매칭 프로세스

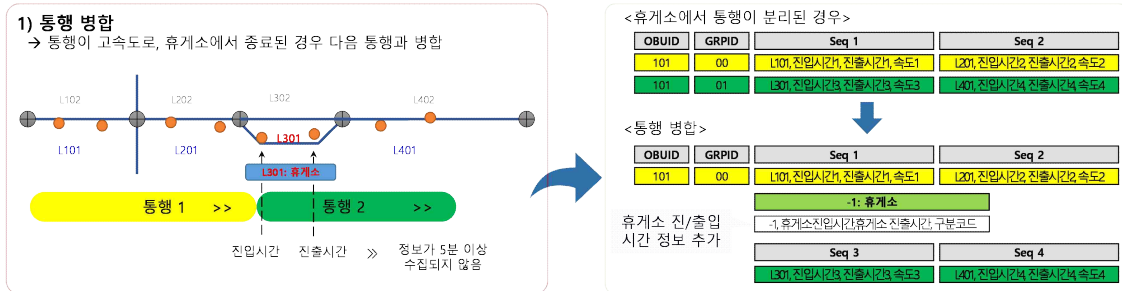
- 궤적별 링크 통행속도 산출
 - 링크 통과시 포인트의 좌표 정보를 통해 통행속도를 산출
 - 궤적별 포인트-링크 간 매칭된 페어 정보를 통하여 개별 차량의 링크 통행속도를 산출

선택	순번	포인트시간	위도	경도	리...	링크ID	링크운행거리	링크남은거리	GIS...	GIS...	GPS...	GPS...
☑	125	2019/05/15 20:35:13	967962.028	1951976.266	2..	57276491201	68.49	162.19	230.68	87.53	195.56	88.00
☑	126	2019/05/15 20:35:14	967985.671	1951984.697	2..	57276491201	93.49	137.20	230.68	87.53	195.56	88.00
☑	127	2019/05/15	968009.154	1951991.78	2..	57276491201		112.68	230.68			00
☑	128	2019/05/15						88.98	230.68			00
☑	129	2019/05/15	968055.856	1952001.929	2..	57276491201		65.26	230.68			00
☑	130	2019/05/15 20:35:18	968079.209	1952005.186	2..	57276491201	189.00	41.68	230.68	87.53	195.56	88.00
☑	131	2019/05/15 20:35:19	968102.594	1952007.398	2..	57276491201	212.06	18.62	230.68	87.53	195.56	88.00
☑	132	2019/05/15 20:35:20	968126.022	1952008.604	2..	57276491001	4.73	234.83	239.56	83.95	233.34	84.00

〈그림 3-24〉 링크 매칭 정보를 통한 통행속도 산출

- 휴게소 또는 고속도로에서 내비게이션 종료 또는 실행 시 통행의 시종점으로 적합하지 않기에, 해당 구간은 연속된 통행을 유지할 수 있도록 통행병합
 - 통행이 병합되는 경우는 휴게소-휴게소, 고속도로-고속도로, 도시고속도로-고속도로 조합의 경우 발생
 - 통행 병합 시 시간 정보 및 구분 코드 정보를 추가하여, 데이터 분석 및 검증 시 활용할 수 있도록 함

• 통행 병합 -> 휴게소, (도시)고속도로에서의 연속된 통행 유지



〈그림 3-25〉 통행병합 예시1



〈그림 3-26〉 통행병합 예시2

○ 티맵 상세도로망 기반 경로 데이터 가공 및 표준DB 형태로 변환 및 저장

〈표 3-5〉 티맵 상세도로망 기반 경로 데이터 표준 DB

K-V	순번	필드명	필드설명	자료형	비고
K	1	OBU_ID	주행 고유 ID	string	
	2	GRP_ID	통행 그룹 ID	integer	
	3	CAR_MODEL	차종("운전습관 가입자" 대상)	string	
	4	CAR_FUEL	유종("운전습관 가입자" 대상)	string	
	5	DEST_NAME	목적지명	string	
	6	DB_NAME	목적지 중분류 정보(POI가 아닌 주소 검색 등의 경우null)	string	
V	7	vLinkId	가상링크ID	long	
	8	inTime	링크 진입시간	long	
	9	outTime	링크 진출시간	long	
	10	lev6Len	6레벨 링크 길이	double	
	11	lev6Spd	6레벨 링크 속도	double	
	12	gpsLen	링크에 매칭된 포인트 간 길이의 합	double	
	13	tmapSpd	tmap 링크 속도	double	

○ DTG 상세도로망 기반 경로 데이터 가공 및 표준DB 형태로 변환 및 저장

〈표 3-6〉 DTG 상세도로망 기반 경로 데이터 표준 DB

K-V	순번	필드명	필드설명	자료형	비고
K	1	OBU_ID	주행 고유 ID	string	
	2	GRP_ID	통행 그룹 ID	integer	
	3	VEH_TYPE	차종	double	
	4	AVG_GAP	음영구간을 제외한 원본-Recon 포인트 간 거리 오차의 평균	double	
	5	RELIABILITY1	이상치 판단 후 궤적 신뢰성	double	
	6	RELIABILITY2	필터링 후 궤적 신뢰성	double	
	7	MATCHING_RATE	포인트-링크 매칭률	double	
	8	CONNECTING_RATE	링크 연결률	double	
V	9	vLinkId		long	
	10	inTime	링크 진입시간	long	
	11	outTime	링크 진출시간	long	
	12	lev6Len	6레벨 링크 길이	double	
	13	lev6Spd	6레벨 링크 속도	double	
	14	gpsLen	링크에 매칭된 포인트 간 길이의 합	double	
	15	tmapSpd	tmap 링크 속도	double	

나. 경로 데이터 압축 및 통합 경로 데이터 구축

- 티맵과 DTG의 경로 데이터 압축 및 통합 경로 데이터 구축 프로세스는 동일함
- 경로 데이터 압축 및 통합 경로 데이터 구축은 온라인 서비스를 위한 목적으로 수행하였음
- 상세도로망 단위의 경로 데이터는 대용량 데이터의 검색속도의 한계로 서비스가 어렵기 때문에 경로 데이터를 주요도로망 Level5.5 단위로 압축하여 온라인 서비스용으로 통합 경로데이터를 구축함
- 상세도로망 단위 경로 데이터 압축 및 통합 경로 데이터 구축 시 610GB에서 150GB로 데이터 용량을 압축하여 빠른 서비스가 가능
- 경로 데이터 압축 및 통합 경로 데이터를 NoSQL 형태로 변환 및 저장



<그림 3-27> 압축된 경로 데이터 형태

〈표 3-7〉 경로 데이터 압축 및 통합 경로 데이터 NoSQL DB 포맷

순번	필드명	필드설명	자료형	비고
1	_id	경로ID	NumberLong	
2	PL	경로별 분석맵 링크 구성	Array(NumberInt)	
3	PT	경로별 차종 정보	NumberInt	
4	PM	경로별 월 정보	NumberInt	
5	PD	경로별 일 정보	NumberInt	
6	PW	경로별 평일/주말 정보	NumberInt	
7	DW	경로별 요일 정보	NumberInt	
8	IT	경로별 진입 시간	Array(NumberInt)	
9	OT	경로별 진출 시간	Array(NumberInt)	
10	DT	행정구역_읍면동	Array(NumberInt)	
11	DT_SGG	행정구역_시군구	Array(NumberInt)	
12	DT_SD	행정구역_시도	Array(NumberInt)	

제3절 교통량 추정 알고리즘 개선

1. 교통량 추정 알고리즘 개선


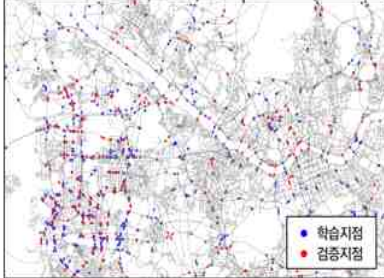
가. 교통량 추정 입력 데이터 구축 프로세스

- 교통량을 추정하기 위한 데이터로써 입력 데이터 구축
 - 상세도로망 Level6 네트워크 기반의 경로 데이터로 구축한 속도 프로파일 DB를 활용하여 교통량 추정 입력 데이터 DB를 구축
- 차종별 링크의 속도 프로파일 데이터에서 평균속도 정보를 가져오고, 차종별 링크의 통행량 데이터에서 통행량 정보를 가져와 링크 전체 지점의 차종별 통행량 및 속도 데이터 생성함
- 관측 링크와 전체 링크의 교통량 추정을 위한 입력 데이터 가공
 - 관측교통량이 존재하는 링크를 대상으로 머신러닝의 학습 데이터로 사용될 관측지점의 교통량 추정 입력 데이터를 가공
 - 학습된 모델을 통해 관측지점을 포함하여 전체지점의 교통량을 추정하기 위해 전체 링크의 교통량 추정 입력 데이터를 가공

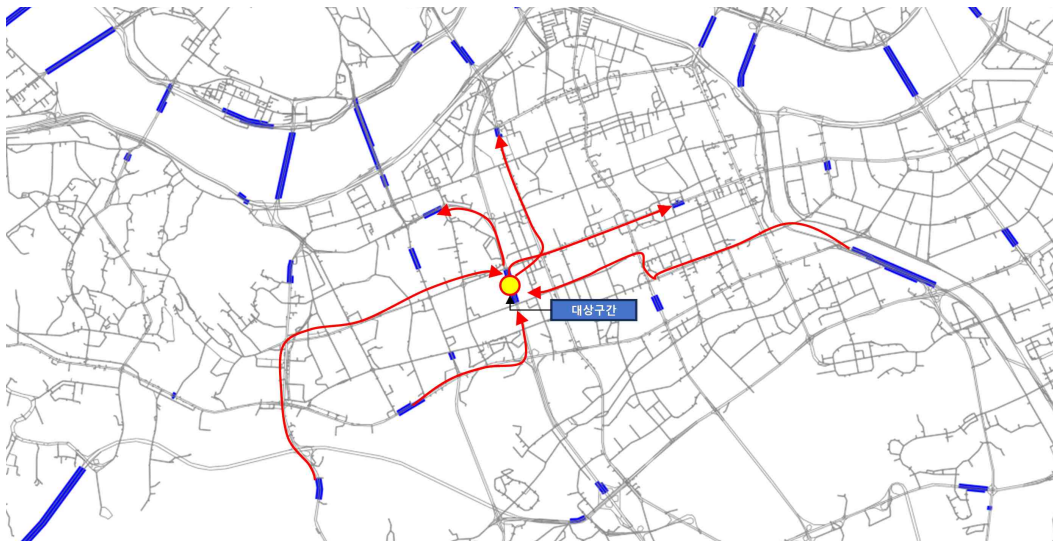
나. 교통량 추정 입력 데이터 고도화

- 교통량 추정을 위한 학습데이터는 약 1만여개의 관측교통량을 학습지점과 검증지점으로 구분하여 약 8,500개 관측지점을 학습데이터로 활용함
- 티맵, DTG 경로 데이터의 연속 주행 기록 정보를 통해 대상 관측지점과 연결된 다른 관측지점 정보를 추가로 입력 데이터를 구축해 도로의 기능적 특성을 고려할 수 있도록 하고자 함
 - 대상 관측지점을 지나는 개별차량 경로를 이용하여 추정 대상 도로구간과 가까

은 통행량 및 차선수, 도로등급 등을 고려하여 최대 10개까지 생성한 뒤 연결성이 높은 3개 지점을 네트워크 연결성 변수로 구축함

구분	학습데이터(80%)	검증데이터(20%)
교통량 추정을 위한 데이터 구성	<ul style="list-style-type: none"> ☑ 학습 데이터 누적방식의 교통량 추정모형 개발 <ul style="list-style-type: none"> ✓ 일반적으로 학습 데이터의 양이 늘어날수록 모형의 일반화 성능과 강건성 향상 ✓ 단, 정갈하게 엄선된 학습 데이터 필요하며 추가되는 데이터의 품질이 매우 중요 ✓ 과측교통량 이상치 제거 필요 ☑ 교통량 추정을 위한 학습 데이터가 부족한 한계를 개선하기 위해 활용 중 	<ul style="list-style-type: none"> ☑ 학습 데이터와 완전히 독립된 데이터로 선정 (Data Leakage 방지) <ul style="list-style-type: none"> ✓ 일관성 확보를 위해 매년 동일한 지점을 신뢰도 평가지점으로 선정
		

<그림 3-28> 교통량 추정을 위한 데이터 구성



<그림 3-29> 네트워크 상의 시공간적 연결성 변수 형태 예시

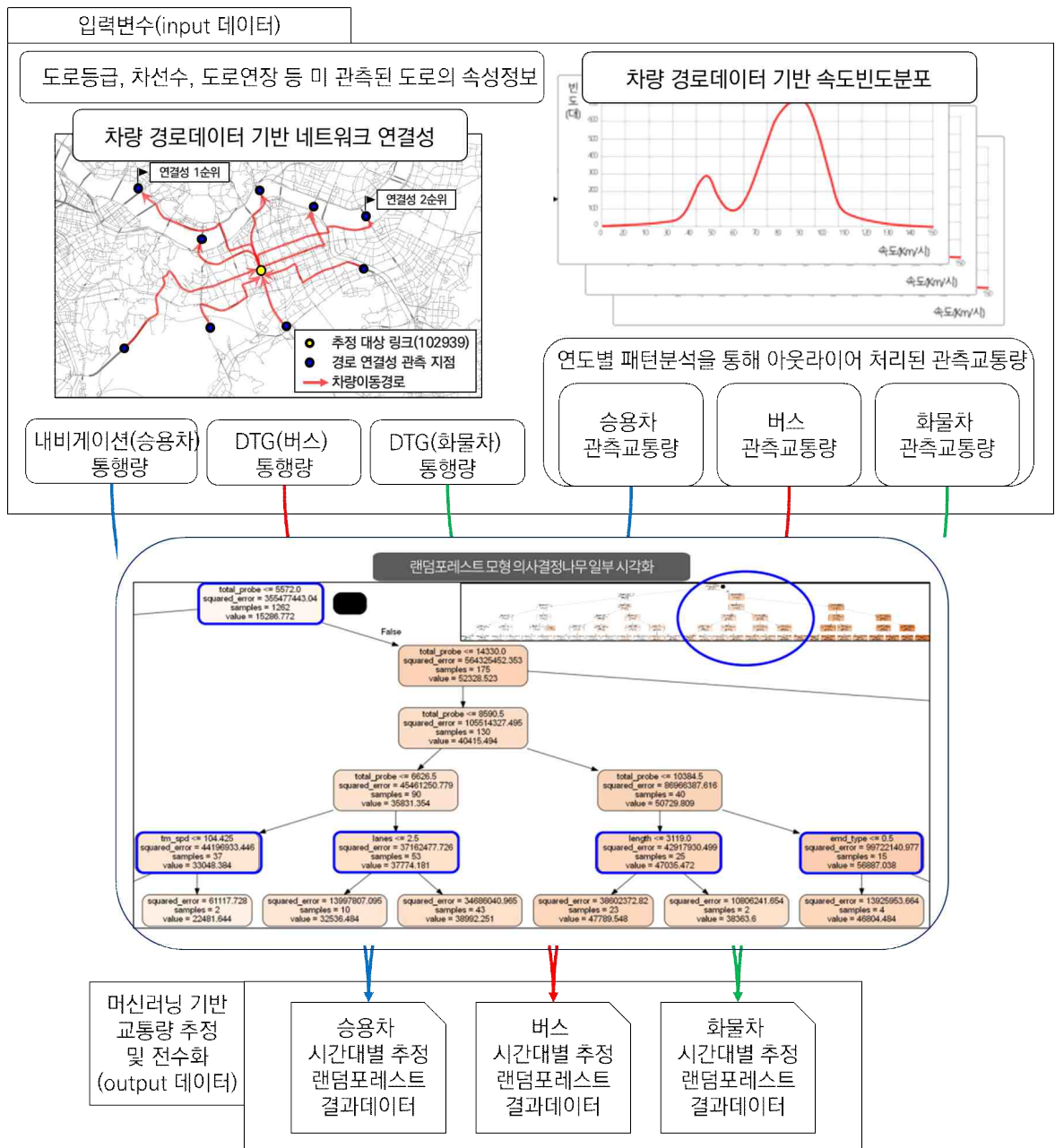
4. 2. 교통량 추정

가. 교통량 추정 목적

- 전국에 조사된 교통량은 약 1만 개 지점으로 상세도로망 Level6네트워크의 링크 갯수(약 62만 개)에 비해 매우 낮은 샘플율을 가지고 있음
- 각 지자체에서 조사한 관측교통량은 차종이 구분되어 있지 않거나, 24시간 전체 시간대에 조사하지 않는 경우가 많아 차종별 시간대별 교통량 추정을 위해 상시지점의 차종을 추정함

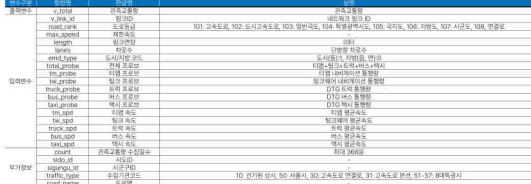
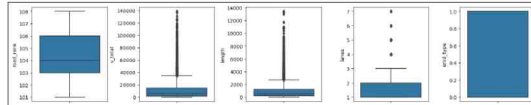
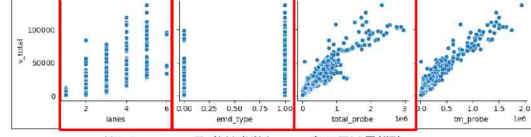
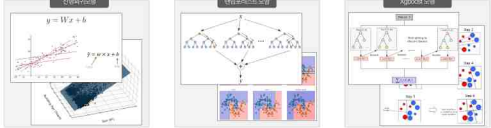
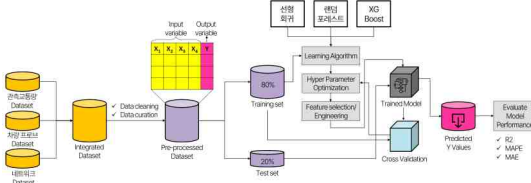
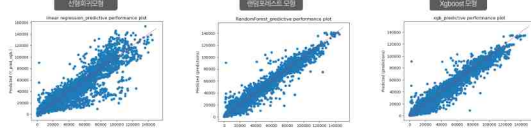
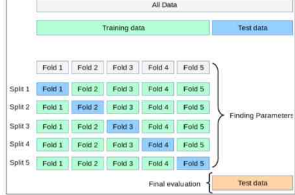

나. 교통량 추정 프로세스

- 교통량 추정은 다음과 같은 프로세스로 수행됨
 - 차량 프로브 통행량 및 관측교통량 데이터 입력
 - 관측교통량 상시지점의 차종이 없는 링크를 대상으로 차종 추정
 - 머신러닝 입력변수 구축
 - 머신러닝 기반 차종별·시간대별 교통량 추정 모형구축
 - 연도별 비교 및 관측지점의 교통량과 추정된 교통량 결과를 비교하여 데이터 신뢰도 평가 및 보완
 - 최종 시간대별·차종별 교통량 추정



<그림 3-30> 교통량 추정 및 구축 프로세스

〈표 3-8〉 머신러닝 기반 추정교통량 모형 구축 프로세스

구분	설명	예시																																
1차가공DB를 활용한 입력변수 구축	관측교통량, 차량 프로브 통행량, 평균속도 등 교통량 추정을 위한 입력 변수 구축																																	
데이터 전처리	수집된 데이터의 이상치 검토 및 전처리																																	
탐색적 데이터분석	변수별 상관관계 분석을 통한 입력 변수 선정																																	
입력변수 선정	도로등급, 제한속도, 도로연장, 차로수, 도시부/지방부, 차량프로브 통행량, 평균속도	<table border="1" data-bbox="779 975 1307 1136"> <thead> <tr> <th>번호</th> <th>변수명</th> <th>연관성</th> <th>설명</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>road_rank</td> <td>도로등급</td> <td>고속도로, 도시고속도로, 일반도로, 특별광역시도, 국사도, 지방도, 시도도, 연경로</td> </tr> <tr> <td>2</td> <td>max_speed</td> <td>제한속도</td> <td></td> </tr> <tr> <td>3</td> <td>length</td> <td>링크연장</td> <td>미터</td> </tr> <tr> <td>4</td> <td>lanes</td> <td>차로수</td> <td>단방향 차로수</td> </tr> <tr> <td>5</td> <td>emd_type</td> <td>도시/지방 코드</td> <td>도시(검은), 지방(흰), 면(노)</td> </tr> <tr> <td>6</td> <td>total_probe</td> <td>전체 프로브</td> <td>티엔 차량+링크웨어 차량+트릭 차량+버스 차량+택시 차량</td> </tr> <tr> <td>7</td> <td>tm_spd</td> <td>티엔 속도</td> <td>티엔 연평균 평균 링크 속도 티엔 연평균 평균 속도</td> </tr> </tbody> </table>	번호	변수명	연관성	설명	1	road_rank	도로등급	고속도로, 도시고속도로, 일반도로, 특별광역시도, 국사도, 지방도, 시도도, 연경로	2	max_speed	제한속도		3	length	링크연장	미터	4	lanes	차로수	단방향 차로수	5	emd_type	도시/지방 코드	도시(검은), 지방(흰), 면(노)	6	total_probe	전체 프로브	티엔 차량+링크웨어 차량+트릭 차량+버스 차량+택시 차량	7	tm_spd	티엔 속도	티엔 연평균 평균 링크 속도 티엔 연평균 평균 속도
번호	변수명	연관성	설명																															
1	road_rank	도로등급	고속도로, 도시고속도로, 일반도로, 특별광역시도, 국사도, 지방도, 시도도, 연경로																															
2	max_speed	제한속도																																
3	length	링크연장	미터																															
4	lanes	차로수	단방향 차로수																															
5	emd_type	도시/지방 코드	도시(검은), 지방(흰), 면(노)																															
6	total_probe	전체 프로브	티엔 차량+링크웨어 차량+트릭 차량+버스 차량+택시 차량																															
7	tm_spd	티엔 속도	티엔 연평균 평균 링크 속도 티엔 연평균 평균 속도																															
모형 검토	선형회귀모형, 랜덤포레스트모형, XGboost 모형 선정																																	
교통량 추정모형 구축	전처리된 데이터와 선정된 모형을 활용하여 교통량 추정모형 구축																																	
교통량 추정 및 평가	구축된 교통량 추정의 추정교통량 신뢰도 분석																																	
교차검증	최종 선정된 랜덤포레스트 모형에 대한 교차검증을 통해 일반화 성능 검토																																	
변수 분석	교통량 추정에 영향을 끼치는 변수별 영향 분석																																	

다. 추정교통량 구축

- 추정교통량 구축 범위는 다음과 같음
 - 기초교통 DB 기반으로 전국 2차로 이상 도로에 대한 추정교통량DB 구축
 - 기본 구축범위는 차종(승용차, 버스, 트럭), 요일패턴별(평일, 주말), 시간(전일, 시간대별)단위, Level6 링크 단위 구축
 - 집계단위 : 상세도로망, 주요도로망, 표준노드링크, 행정구역(시도/시군구/읍면동)
 - 관측지점을 대상으로 추정된 결과 비교 및 연도별 추이 분석을 통한 검증 진행

〈표 3-9〉 추정교통량 DB 테이블 구성

No	Column	Type	설명	코드	코드정보
1	V_link_id	Integer	가상링크ID	-	-
2	Week_code	Integer	요일코드	0, 1	주말, 평일
3	time_code	Integer	시간코드	1-24	24시간
4	total_veh	Double	교통량(전체)	-	-
5	auto	Double	교통량(승용차)	-	-
6	bus	Double	교통량(버스)	-	-
7	truck	Double	교통량(화물차)	-	-

라. 교통량 추정 검증 결과

- 교통량의 신뢰도를 평가하기 위해서는 조사된 교통량 자료가 존재하는 지점에서 교통량 추정을 통해 현실을 어느 정도로 구현하고 있는지에 대한 검토가 필요함
- 한국개발연구원(2008)에 따르면 추정교통량의 신뢰성을 평가하기 위해 조사교통량과의 오차율 기준으로 추정교통량에 대한 평가를 수행하고 있다. 본 연구에서 분석하고 있는 조사지점의 교통량 수준을 고려하면 ±15% 이내의 오차율이 유의한 추정 범위라고 볼 수 있음
- 추정교통량의 신뢰도 평가지표는 평균절대비율오차(Mean Absolute Percent Error, MAPE, %)를 활용함

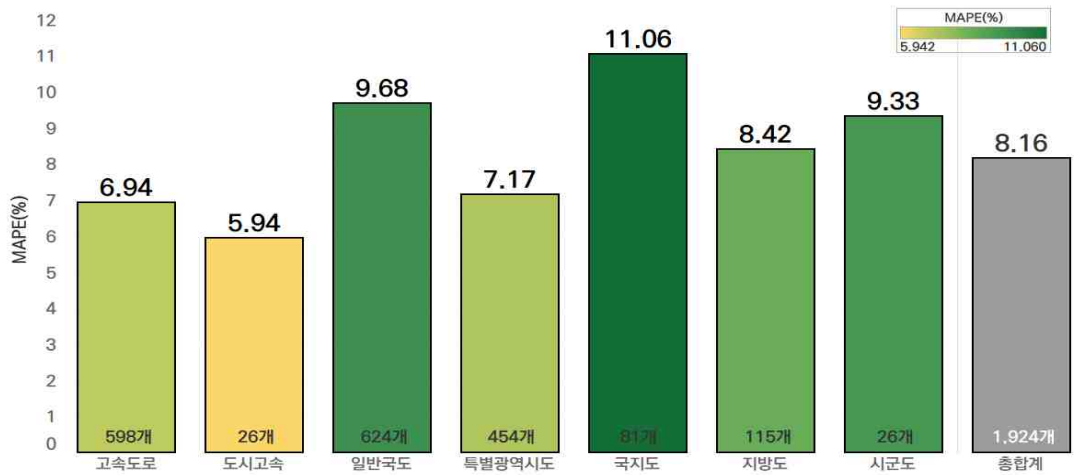
$$MAPE(\%) = \frac{\sum_{i=1}^n \left(\frac{|q_i - \hat{q}_i|}{q_i} \right) \times 100}{n},$$

여기서, q_i = 지점 i 의 조사교통량(대/일)

\hat{q} = 지점 i 의 추정교통량(대/일)

n = 전체 지점 개수(개)

- 검증용 데이터를 활용하여 도로등급별 신뢰도 분석 결과, 고속도로 6.9%, 일반국도 9.7%, 특별광역시도 7.2%로 유의미한 결과가 나타남
- 교통량이 많을수록 오차율이 낮게 나타나는 것으로 분석됨

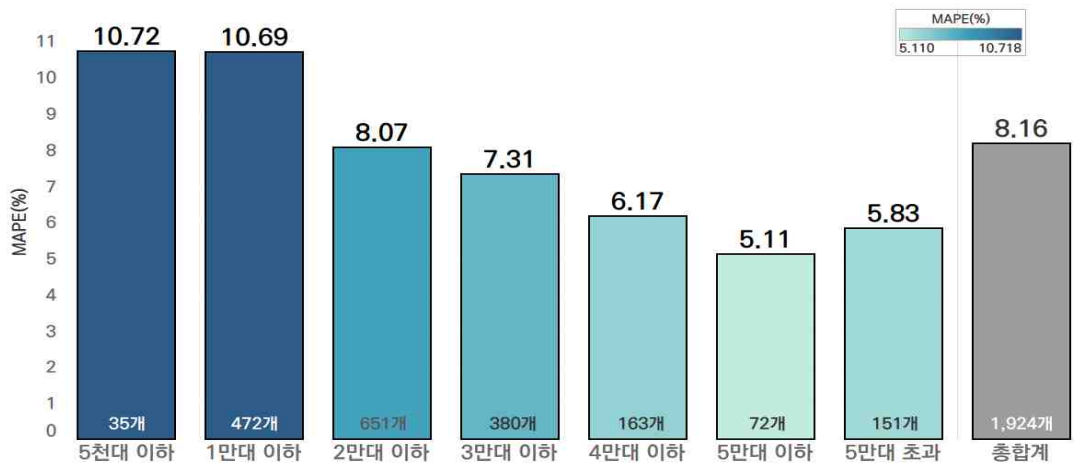


※ 참고1. 평가지점은 ITS 성능평가 기준(1시간 기준 200대 이상인 도로를 대상)으로 선정함

참고2. MAPE 목표치는 한국개발연구원의 직접 영향권 범위 참고함

참고3. 2023년 추정교통량 기준 분석결과

<그림 3-31> 도로등급별 신뢰도 분석 결과



※ 참고1. 평가지점은 ITS 성능평가 기준(1시간 기준 200대 이상인 도로를 대상)으로 선정함

참고2. MAPE 목표치는 한국개발연구원의 직접 영향권 범위 참고함

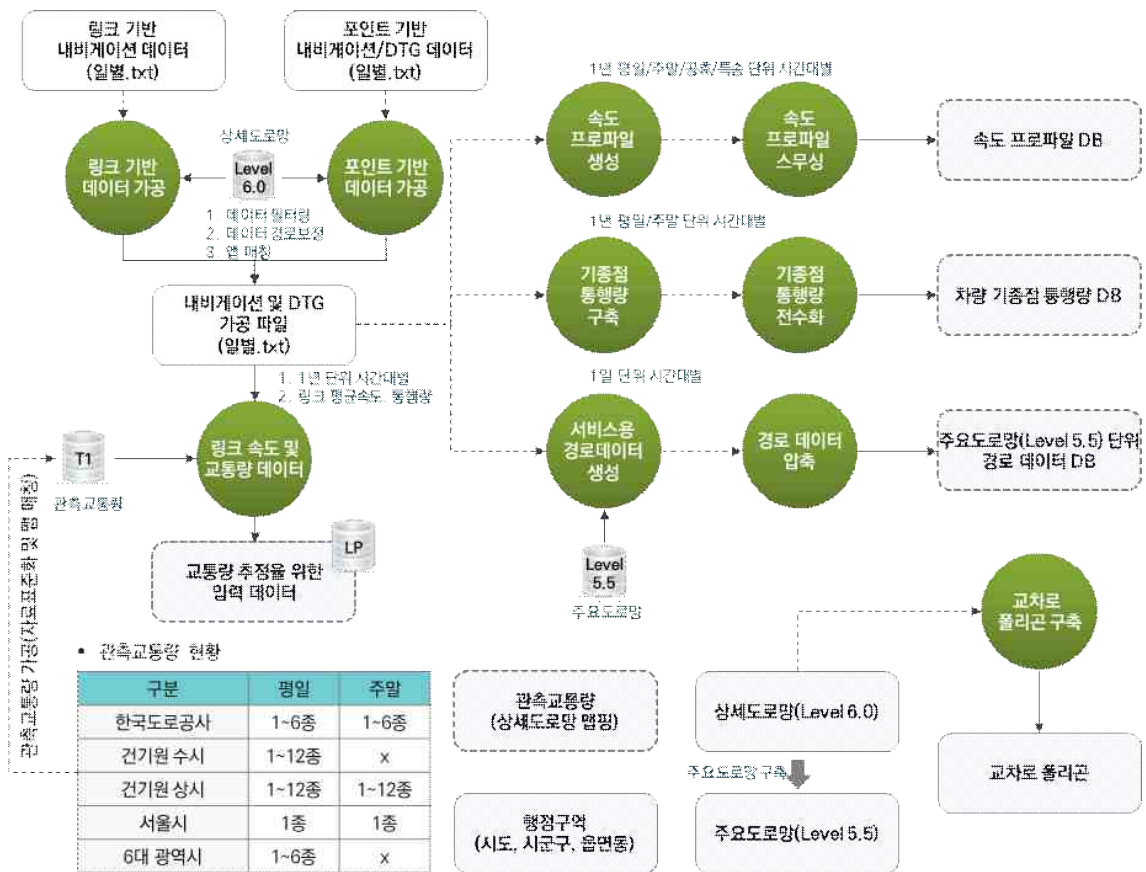
참고3. 2023년 추정교통량 기준 분석결과

<그림 3-32> 교통량 수준별 신뢰도 분석 결과

제4절 기초교통 DB 구축

1. 기초교통 DB 구축 개요

- 전국 교통량 전수화 및 통행지표를 구축하기 위해 경로 데이터 및 관측교통량을 이용하여 1차 가공한 기초교통 DB 구축
 - 경로 데이터의 속도 정보를 이용하여 구축한 속도빈도분포 구축
 - 차량 기종점 통행량 DB 구축
 - 링크 통행량과 속도로 구성된 교통량 추정을 위한 입력 데이터 구축



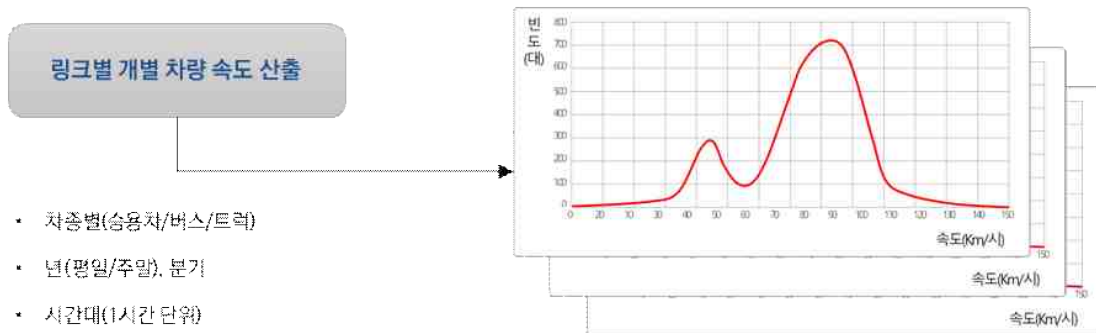
〈그림 3-33〉 기초교통 DB 구축 프로세스

2. 속도빈도분포 DB 구축

가. 속도빈도분포 구축 프로세스

○ 링크별 속도 프로파일 집계

- 교통량 추정 DB와 통행지표 구축을 위하여 속도빈도분포 DB 구축
 - 속도지표, 혼잡지표, 환경지표 등 다양한 통행지표 등에 활용
- 링크에 매칭된 궤적 정보를 이용하여 개별 차량의 속도 산출 및 산출된 속도 범위에 따른 집계를 수행
 - 링크별로 1km/h 단위의 통행속도(1km/h ~ 150km/h, 초과시 150km/h)의 통행량을 집계
- 개별 링크의 연/평일주말/1시간 단위의 속도빈도분포 구축



〈그림 3-34〉 속도빈도분포 구축 프로세스

나. 속도빈도분포 구축 결과

○ 속도빈도분포 DB 형태

〈표 3-10〉 속도빈도분포 DB 포맷

순번	필드명	필드설명	자료형	비고
1	v_link_id	상세도로망 가상링크ID	long	
2	quarter_code	분기코드	int	1 ~4
3	week_code	주중코드	int	1: 평일 2: 주말
4	time_code	시간코드	int	0: 전일 1~24: 시간대
5	road_name	도로명	string	
6	road_rank	도로등급	int	
7	lanes	차선수	int	
8	length	도로길이	double	
9	emd_id	읍면동ID	int	
10	avg_spd	평균속도	double	
11	med_spd	중위속도	double	
12	probe	통행량	int	
13	spd_0kmh	0km/h 속도 통행량	int	
14	spd_1kmh	1km/h 이하 속도 통행량	int	
...	
163	spd_150kmh	149km/h 속도 초과 통행량	int	

○ 도로등급별 평균속도 비교분석

- 2022년 대비 2023년 평균속도는 모든 도로등급에서 감소함

〈표 3-11〉 연도별 도로등급별 평균속도

단위 : km/h

도로등급	평균속도				'20년 대비	'21년 대비	'22년 대비
	2020년	2021년	2022년	2023년	'21년 증감율	'22년 증감율	'23년 증감율
고속도로	97.62	95.29	93.87	91.09	-2.4%	-1.5%	-3.0%
도시고속도로	65.87	65.19	65.42	63.89	-1.0%	0.4%	-2.3%
일반국도	62.44	61.21	60.86	60.32	-2.0%	-0.6%	-0.9%
특별광역시도	26.04	25.24	25.11	24.79	-3.1%	-0.5%	-1.3%
국가지원지방도	51.24	50.45	50.03	49.59	-1.5%	-0.8%	-0.9%
지방도	49.73	48.82	48.41	47.91	-1.8%	-0.8%	-1.0%
시군도	34.09	33.6	33.24	32.94	-1.4%	-1.1%	-0.9%
총계	38.21	37.49	37.12	36.75	-1.9%	-1.0%	-1.0%

* 연단위 평일 기준 도로등급 연결로를 제외한 평균속도를 분석함

○ 시도별 평균속도 비교분석

- 연도별 평균속도를 분석했을 때 2020년 38.2km/h, 2021년 37.5km/h, 2022년 37.1km/h, 2023년 36.8km/h로 매년 약 1%씩 속도가 감소하고 있음
- 2022년 대비 2023년 평균속도가 가장 크게 증가한 지역은 대구로 약 6.1%이고, 그 다음 제주에서 약 0.2% 증가하였음
- 나머지 시도에서 2022년 대비 2023년 평균속도가 감소한 것으로 분석되었음

〈표 3-12〉 연도별 시도별 평균속도

단위 : km/h

시도	평균속도				'20년 대비 '21년 증감율	'21년 대비 '22년 증감율	'22년 대비 '23년 증감율
	2020년	2021년	2022년	2023년			
서울	24.07	22.98	23.07	22.77	-4.5%	0.4%	-1.3%
부산	29.34	28.89	28.86	28.31	-1.5%	-0.1%	-1.9%
대구	30.14	29.55	29.58	31.37	-2.0%	0.1%	6.1%
인천	31.09	30.12	29.72	29.24	-3.1%	-1.3%	-1.6%
광주	28.52	28.08	27.64	27.48	-1.5%	-1.6%	-0.6%
대전	30.39	29.31	29.1	28.66	-3.6%	-0.7%	-1.5%
울산	33.9	33.53	33.23	32.76	-1.1%	-0.9%	-1.4%
세종	36.98	35.99	35.41	35.15	-2.7%	-1.6%	-0.7%
경기	33.57	32.79	32.35	32.03	-2.3%	-1.3%	-1.0%
강원	42.05	41.63	41.24	41.13	-1.0%	-0.9%	-0.3%
충북	41.81	41.26	40.56	40.25	-1.3%	-1.7%	-0.8%
충남	42.83	41.9	41.14	40.59	-2.2%	-1.8%	-1.3%
전북	42.23	41.38	41.07	40.72	-2.0%	-0.7%	-0.9%
전남	43.28	42.87	42.61	42.12	-0.9%	-0.6%	-1.1%
경북	43.24	42.68	42.26	41.68	-1.3%	-1.0%	-1.4%
경남	39.59	38.94	38.68	38.3	-1.6%	-0.7%	-1.0%
제주	37.52	36.84	36.34	36.43	-1.8%	-1.4%	0.2%
총계	38.21	37.49	37.12	36.75	-1.9%	-1.0%	-1.0%

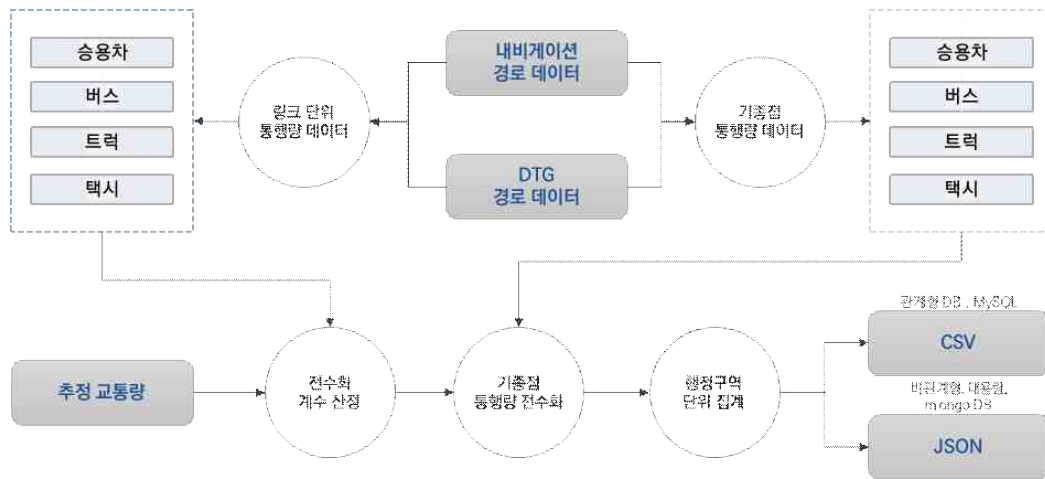
* 연단위 평일 기준 도로등급 연결로를 제외한 평균속도를 분석함

3. 기종점 통행량 DB 구축

가. 기종점 통행량 DB 구축 프로세스

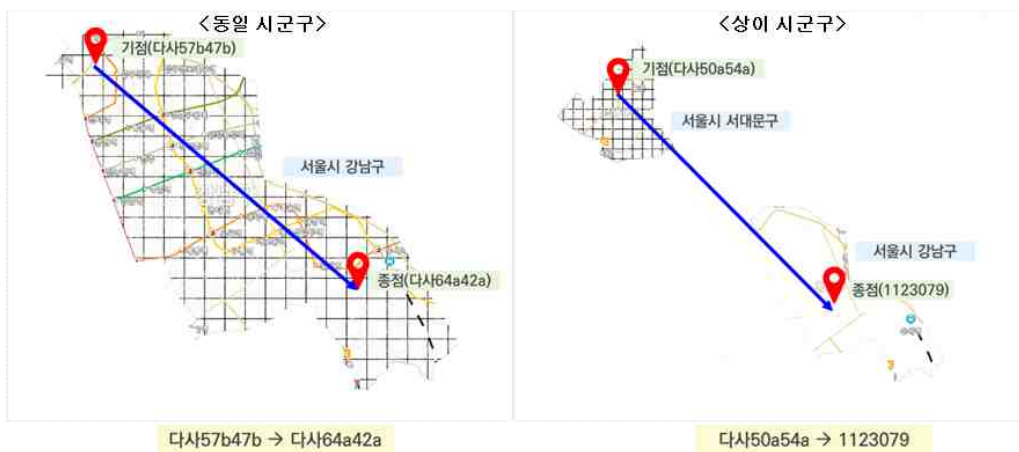
○ 차종별 기종점 통행량 DB를 구축

- 티맵, DTG 경로 데이터를 활용해 차종별 기종점 통행량 DB를 구축



〈그림 3-35〉 기종점 통행량 DB 구축 프로세스

- 기종점의 시군구가 동일하면 격자-격자 통행량, 상이하면 격자-읍면동 통행량을 집계



〈그림 3-36〉 기점과 종점의 시군구에 따른 집계 방식 차이

○ 기종점 단위 통행량 데이터 집계

- 티맵, DTG 경로 데이터를 통해 기점과 종점의 시군구가 동일할 경우 격자-격자 통행량을, 상이할 경우 격자-읍면동 통행량을 집계

나. 기종점 통행량 DB 구축 결과

○ 기종점 통행량 DB구축 결과

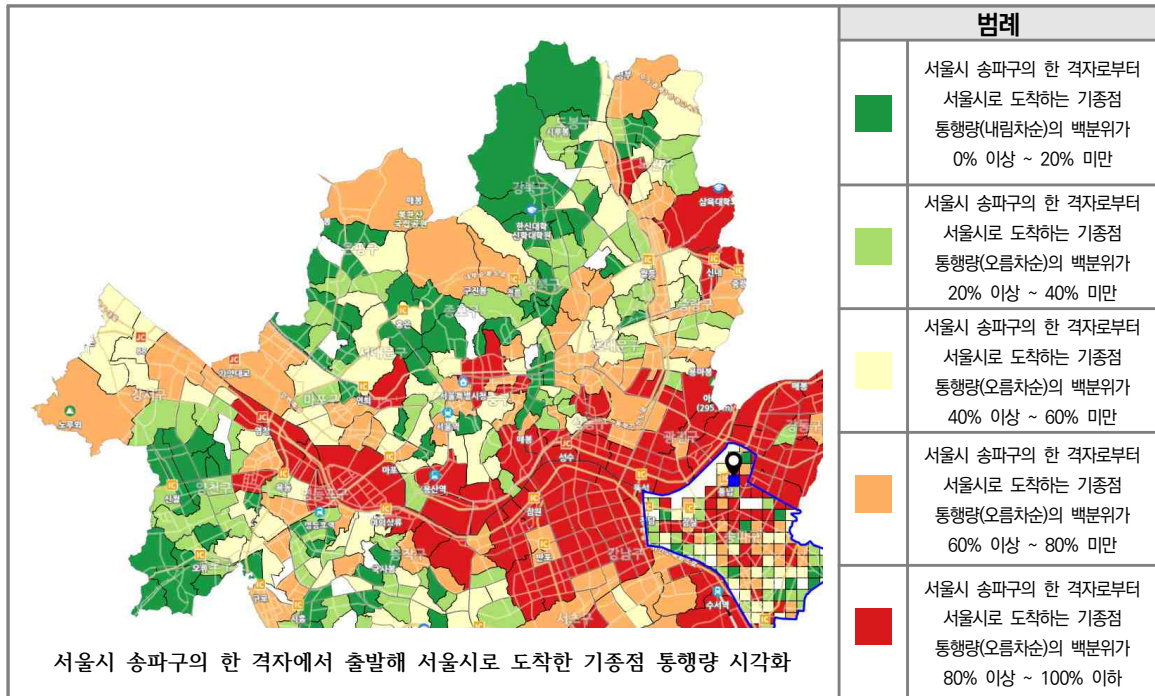
- 링크 통행량 집계 시 읍면동 코드 라인에 대한 판별 및 집계가 이뤄지기에 다른 과정에 비해 상대적으로 시간이 오래 걸림

〈표 3-13〉 기종점 통행량 DB 포맷

순번	필드명	필드설명	자료형	비고
1	DATA_STRD_YM	기준년월	string	
2	DPTR_LTTC_ID	출발 격자ID	string	
3	DPTR_LEGALDONG_CD	출발 법정동코드	int	
4	DPTR_SIGNGU_CD	출발 시군구코드	int	
5	DPTR_CTPR_CD	출발 시도코드	int	
6	ARVL_PYN_ID	도착 폴리곤ID	string	
7	ARVL_LEGALDONG_CD	도착 법정동코드	int	
8	ARVL_SIGNGU_CD	도착 시군구코드	int	
9	ARVL_CTPR_CD	도착 시도 코드	int	
10	WEEK_CD	주 코드	int	
11	TM_CD	시간 코드	int	
12	OD_TYPE_CD	OD 유형 코드	int	
13	DLMN_TRFC_QTY	연평균 통행량	double	

○ 기종점 통행량 결과 시각화 예시

- 동일 시군구 내에서는 비교적 골고루 통행량이 분포함을 확인
- 상이 시군구의 가까운 읍면동으로 통행량이 많은 것으로 확인



〈그림 3-37〉 기점과 종점의 시군구에 따른 집계 방식 차이

4. 교통량 추정 입력 데이터 DB 구축

가. 교통량 추정 입력 데이터 구축 프로세스

- 교통량을 추정하기 위한 데이터로써 입력 데이터 구축
 - 상세도로망 Level6 네트워크 기반의 경로 데이터로 구축한 속도 프로파일 DB를 활용하여 교통량 추정 입력 데이터 DB를 구축



〈그림 3-38〉 교통량 추정 입력 데이터 구성

- 관측 링크와 전체 링크의 교통량 추정을 위한 입력 데이터 가공
 - 관측교통량이 존재하는 링크를 대상으로 머신러닝의 학습 데이터로 사용될 관측지점의 교통량 추정 입력 데이터를 가공
 - 학습된 모델을 통해 관측지점을 포함하여 전체지점의 교통량을 추정하기 위해 전체 링크의 교통량 추정 입력 데이터를 가공

나. 교통량 추정을 위한 입력 데이터 구축 결과

○ 교통량 추정을 위한 입력 데이터

- 관측링크와 전체링크의 입력 데이터는 형태는 동일하며, 머신러닝의 학습에 영향을 주는 정보를 선택하여 구축함
- 머신러닝 학습에 영향을 주는 정보는 교통량 추정 알고리즘 개발 시 탐색적 분석을 통해 선별한 변수임

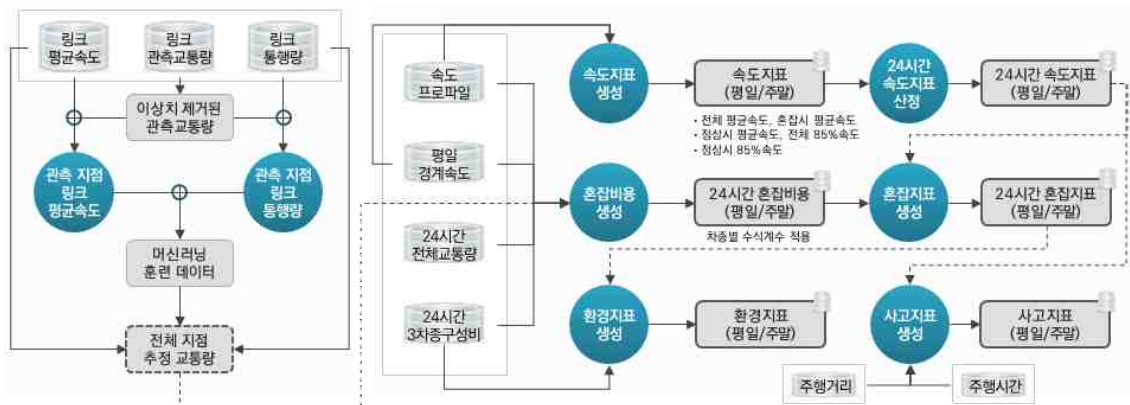
〈표 3-14〉 교통량 추정을 위한 입력 데이터 DB

순번	필드명	필드설명	자료형
1	v_link_id	가상링크ID	Long
2	up_dw	기관 코드	integer
3	traffic_type	도로등급	integer
4	spot_id	도로등급명	string
5	road_rank	조사지점ID	string
6	time_info	조사지점명	string
7	time_code	시간코드	integer
8	v_total_wd	전체 교통량	double
9	v_auto_wd	승용 교통량	double
10	v_bus_wd	버스 교통량	double
11	v_truck_wd	트럭 교통량	double
12	total_probe	전체 통행량	integer
13	tm_probe	티맵 통행량	integer
14	truck_probe	버스 통행량	integer
15	bus_probe	트럭 통행량	integer
16	taxi_probe	택시 통행량	integer
17	tm_spd	티맵 속도	double
18	truck_spd	버스 속도	double
19	bus_spd	트럭 속도	double
20	taxi_spd	택시 속도	double
21	emd_type	행정구역 동 여부	Integer
22	lanes	차선수	integer
23	length	도로길이	double
24	vkt	총도로연장	double
25	k	티맵 K값	double
26	sido_id	시도ID	Integer
27	tm_scale	승용 스케일	double
28	bus_scale	버스 스케일	double
29	truck_scale	트럭 스케일	double

제4절 차량 교통망 성능평가지표 구축

1. 차량 교통망 성능평가지표 구축 개요

- 차량 통행지표 DB는 1차 가공된 속도 프로파일과 추정된 교통량을 기반으로 전국 혼잡지표, 속도지표, 환경지표, 안전지표를 구축함



〈그림 3-39〉 교통량 추정 및 차량 통행지표 구축 프로세스

- 통행지표 구축을 위한 기초교통 DB는 다음과 같이 정의된 범위가 존재하여 범위에 따른 통행지표를 구축함

〈표 3-15〉 교통망 성능평가지표 DB 생성 기준

데이터명	연	분기	평일 주말	시간	차종	상세 도로망	주요 도로망	행정 구역
속도 프로파일	○	○	○	○	○	○	○	○
추정 교통량	○		○	○	○	○	○	○
혼잡비용	○		○	○	○	○	○	○
속도지표	○		○	○		○	○	○
혼잡지표	○		○			○	○	○
환경지표	○		○		○	○	○	○
사고지표	○		○			○	○	

- 차량 통행지표는 도로망의 현황과 성능을 평가하기 위한 지표를 의미하며, 약 21종의 차량 통행지표를 구축

〈표 3-16〉 차량 교통망 성능평가지표

지표명		지표설명
교통량	관측교통량	공공에서 조사하는 교통량으로 수집기관, 조사기간 등이 상이하여 동일한 기준으로 도로망 네트워크에 매칭하여 나타낸 연평균 일단위 조사교통량
	추정교통량	특정 시간대 동안 해당 도로구간을 통과하는 차량 대수
	차량주행거리	추정 교통량 기준의 차량주행거리
속도지표	평균속도	전체 차량의 속도를 평균한 값
	정상시 평균속도	전체 차량 중 교통 혼잡을 경험한 차량들의 평균속도
	혼잡시 평균속도	전체 차량 중 교통 혼잡을 경험하지 않은 차량들의 평균속도
혼잡지표	혼잡시간강도	전체 차량의 총 통행시간 중 교통 혼잡을 경험한 차량들의 총 통행시간 비율
	혼잡빈도강도	전체 차량의 총 통행시간 중 교통 혼잡을 경험한 차량들의 총 통행량 비율
	교통혼잡비용	차량들이 도로상에서 교통혼잡이 발생하여 정상속도 이하로 운행하게 됨으로써 추가로 발생하게 되는 비용
환경지표	이산화탄소배출량	차량들로 주행으로 발생하는 도로이동오염물질 배출량
	미세먼지배출량	
	일산화탄소배출량	
	휘발성 유기화합물배출량	
	질소산화물배출량	
안전지표	과속비율	과속기준치를 초과하는 차량의 비율
	속도 표준편차	개별 차량들의 속도의 퍼짐 정도
	위험운전행동 이벤트 비율	위험운전행동기준을 개별 차량 속도기준에 따라 위험운전행동이벤트 발생 건수를 비율로 나타냄

2. 교통환경지표 산출 모듈 개선

가. 교통환경지표 산출 방법론 정의

- 교통환경지표는 도로이동오염원 배출량 중 이산화탄소배출량, 일산화탄소배출량, 질소산화물배출량, 미세먼지배출량, 휘발성유기화합물배출량으로 구성되어 있으며, 도로이동오염원 배출량 산정을 위한 방법은 기본적으로 IPCC (Intergovernmental Panel on Climate Change) Guideline⁵⁾에서 제시하고 있음
- IPCC Guideline 보고서에서는 배출량 산정을 위한 방법으로 Tier 1, Tier 2, Tier 3을 제시하고 있으며, 본 과업에서는 가장 정확한 산정을 할 수 있는 Tier 3 방법론을 기반으로 도로구간별(링크별) 차량의 배출량을 산정하였음

〈표 3-17〉 IPCC Guideline 온실가스 추정방법 비교

방법	산정식	특징
Tier 1	$Emissions = \sum(Fuel_a \times EF_a)$ <p><i>Emissions</i> : 배출량(kg) <i>Fuel_a</i> : 연료 a 소비량(TJ), <i>EF_a</i> : 배출계수(kg/TJ) <i>a</i> : 연료 종류(예, 휘발유, 경유, LPG 등)</p>	Top-down 방식 연료별 IPCC 배출계수 적용, 차량주행거리는 고려되지 않음
Tier 2	$Emissions = \sum(Fuel_{a,b,c} \times EF_{a,b,c})$ <p><i>Emissions</i> : 배출량(kg) <i>Fuel_{a,b,c}</i> : 이동 배출원 활동도에 대한 연료 소비량(TJ) <i>EF_{a,b,c}</i> : 배출계수(kg/TJ) <i>a</i> : 연료 종류(예, 휘발유, 경유, LPG 등), <i>b</i> : 차종 <i>c</i> : 배출제어기술(제어장치의 미장착, 촉매변환장치 등)</p>	Bottom-up 방식 차종별, 배출제어기술에 따른 배출계수 적용, 차종별 에너지 소비량 자료 필요, 차량주행거리는 고려되지 않음
Tier 3	$Emissions = \sum(Distance_{a,b,c,d} \times EF_{a,b,c,d}) + \sum C_{a,b,c,d}$ <p><i>Emissions</i> : 배출량(kg) <i>EF_{a,b,c,d}</i> : 배출계수(kg/km) <i>Distance_{a,b,c,d}</i> : 주어진 이동 배출원 활동도에 대해 열적으로 안정된 엔진 운전 단계 동안의 차량주행거리(VKT)(km) <i>C_{a,b,c,d}</i> : 예열 중 배출량(cold start)(kg) <i>a</i> : 연료 종류(예, 휘발유, 경유, LPG 등), <i>b</i> : 차종 <i>c</i> : 배출제어기술(제어장치의 미장착, 촉매변환장치 등) <i>d</i> : 운전 조건(예, 도심 또는 교외의 도로종류, 기타 환경적 요인)</p>	Bottom-up 방식 차종별, 배출제어기술에 따른 배출계수 적용, 배출량을 가장 정확히 산정하는 수준이기 때문에 많은 자료가 요구됨

자료 : 국가 온실가스 인벤토리 작성을 위한 2006 IPCC 가이드라인, 환경부(2008)

5) 국가 온실가스 인벤토리 작성을 위한 2006 IPCC 가이드라인, 환경부(2008)

- Tier3 방법론에서 배출제어기술, 운전 조건에 대한 자료는 차량 GPS 데이터를 통해 확인하기 어려우므로 제외하였으며, 차종별 연료별에 대한 배출계수 및 차량주행거리 자료를 이용하여 배출량을 산정함
 - 여기서 차량주행거리 자료는 교통량 기반 차량주행거리이며, KTDB 상세도로망 level6 네트워크⁶⁾의 링크별 교통량을 통해 산정된 링크별 차량주행거리 데이터를 이용함
 - 차종별 규모별 연료별 네트워크 점유율은 링크별 도로를 주행하는 차량들의 차종, 규모, 연료별로 구성하는 비율을 의미하며, 교통안전공단에서 제공하는 자동차 주행거리 자료를 바탕으로 도로상에 점유하고 있는 차량의 비율을 산정함

$$Emissions = \sum (Distance_{a,b} \times EF_{a,b} \times NOR_{a,b})$$

$$Distance_{a,b} = Volume_a \times RoadLength$$

Emissions : 배출량 (kg)

EF_{a,b} : 배출계수 (kg/km)

Distance_{a,b} : 차량주행거리 (VKT) (km)

NOR_{a,b} : 차종별 규모별 연료별 네트워크 점유율 (%)

a : 차종 (승용차, 버스, 화물차)

b : 연료 (휘발유, 경유, LPG, HEV)

- 한국교통안전공단의 자동차주행거리는 자동차검사 시 차량 등록지와 실제 운행지가 다를 수 있기 때문에 지역별 차량주행거리가 실제 차량주행거리와 차이가 있으므로 교통량 기반 차량주행거리 데이터를 이용함
- 기존 도로이동오염원 배출량 산정을 위한 배출계수는 아래의 문헌에서 제공하고 있는 배출계수를 이용하였으나, 최신 배출계수를 반영하고 있지않기 때문에 현재 제시되고 있는 배출계수와 차이가 있음
 - 수송부문 온실가스 기후변화대응 시스템 구축(II) - 자동차 온실가스 Bottom-up 배출계수 개발의 온실가스(CO₂) 배출계수와 국가 대기오염물질 배출량 산정방법 편람(III)의 배출계수를 이용하였음

6) 도로망 네트워크는 2023 도로업무편람의 2022년 111,420km 도로연장을 기준으로 약 97% 도로를 커버하는 상세도로망 level6 네트워크 데이터를 활용함

- 본 사업에서 사용된 최신 배출계수는 환경부의 국가온실가스종합센터와 국가미세먼지정보센터에서 제공하는 배출계수를 이용하였음
 - 이산화탄소배출량의 배출계수는 “국가온실가스종합정보센터”에서 제공하고 있는 자료이며, 도로수송 부문에서 가장 최신 자료인 “2021년 승인 국가 온실가스 배출·흡수계수”의 1A3b도로수송의 배출계수를 이용함
 - 일산화탄소배출량, 질소산화물배출량, 미세먼지배출량, 휘발성유기화합물배출량의 배출계수는 “국가미세먼지정보센터”에서 제공하고 있는 “국가 대기오염물질 배출량 산정방법 편람(VI)”의 도로이동오염원의 차종별 연료별 배출계수를 이용함
- 기존 활용된 배출계수는 '06년 - '08년 연식에 적용된 계수값으로 최근 배출계수를 반영이 필요하며, 차종별 규모별 연료별 네트워크 점유율도 각 연도에 맞는 값으로 반영해야하므로 교통환경지표 모듈 개선이 필요함
 - 최신화된 배출계수와 차종별 규모별 연료별 네트워크 점유율을 산정하여 교통환경지표 모듈을 개선함
- 차종별 규모별 연료별 네트워크 점유율은 도로를 주행하는 차량의 공간적 해상도를 할당하기 위해 적용되는 것으로 한국교통안전공단의 차량의 차종/규모/연료별 연간주행거리와 국토교통 통계누리의 자동차등록대수를 통해 산정함

〈표 3-18〉 차종별 유형별 규모별 전국 연간주행거리

차종		등록대수(천대)	연간주행거리(백만km)	비율	
승용차 (기타형 제외)	경형	2,000	16,277	6.4%	100%
	소형	250	1,655	0.7%	
	중형	12,617	154,681	61.2%	
	대형	6,329	80,312	31.8%	
버스	시내버스	41,466	-	39.7%	100%
	시외버스	24,047	-	60.3	
	고속버스	15	-		
	전세버스	38,808	-		
화물차 (특수용도형 제외)	소형	2,519	33,886	66.9%	100%
	중형	349	6,374	12.6%	
	대형	220	10,393	20.5%	

자료1 : 2023년 자동차 주행거리통계, 교통안전정보관리시스템, 교통안전공단

자료2 : 2023년 12월 자동차 등록자료 통계, 국토교통 통계누리, 국토교통부

〈표 3-19〉 차종별 연료별 전국 연간주행거리

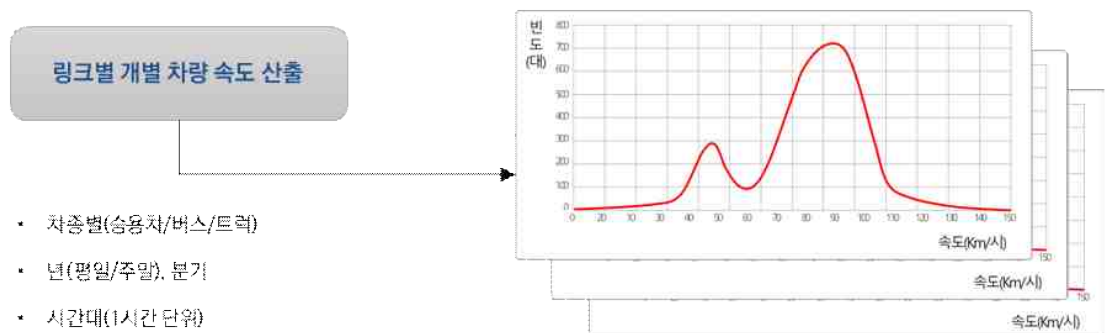
차종		등록대수(천대)	연간주행거리(백만km)	비율	
승용차	휘발유	12,198	122,754	48.5%	100%
	경유	5,592	77,026	30.5%	
	LPG	1,673	27,252	10.8%	
	기타	1,734	25,893	10.2%	
승합차	휘발유	7	72	0.5%	100%
	경유	588	10,305	73.5%	
	LPG	64	909	6.5%	
화물차	기타	50	2,743	19.6%	100%
	휘발유	28	342	0.5%	
	경유	3,331	57,853	90.1%	

자료1 : 2023년 자동차 주행거리통계, 교통안전정보관리시스템, 교통안전공단

자료2 : 2023년 12월 자동차 등록자료 통계, 국토교통 통계누리, 국토교통부

○ 이산화탄소배출량의 배출계수

- 2021년 승인 국가 온실가스 배출·흡수계수의 1A3b도로수송의 배출계수에서 차종 버스를 제공하지 않고, 승합 대형을 전세광역과 시내로 구분하여 제공하므로 해당 항목을 버스의 배출계수*로 이용함
- * 제시되지 않는 계수 중 차종·연료가 동일한 자동차의 규모별 항목이 없는 경우에 상위 규모의 동일 차종·연료 차량을 적용
- 배출계수의 산출식에 나타난 차속은 개별차량 경로데이터를 이용하여 구축한 링크별 차량의 속도빈도분포 데이터를 이용함



〈그림 3-40〉 속도빈도분포 형태

<표 3-20> 이산화탄소 배출량의 배출계수

차종	연료	배출계수		
		차속구분	산출식 (y: 배출량(g/km), x:차속(km/h))	
승용차	경형	휘발유	65.4km/h 미만	$y = 1065.1722x^{-0.5889}$
			65.4km/h 이상	$y = 0.0225x^2 - 3.3075x + 212.8460$
		LPG	65.4km/h 미만	$y = 989.9413x^{-0.5937}$
			65.4km/h 이상	$y = 0.0172x^2 - 2.3601x + 167.3842$
	소형	휘발유	65.4km/h 미만	$y = 1256.0382x^{-0.5914}$
			65.4km/h 이상	$y = 0.0252x^2 - 3.7270x + 245.9051$
		경유	65.4km/h 미만	$y = 1037.3974x^{-0.5800}$
			65.4km/h 이상	$y = 0.0133x^2 - 1.3612x + 129.4859$
		LPG	65.4km/h 미만	$y = 1223.8670x^{-0.6046}$
			65.4km/h 이상	$y = 0.0188x^2 - 2.7902x + 203.7804$
	중형	휘발유	65.4km/h 미만	$y = 1446.3728x^{-0.5793}$
			65.4km/h 이상	$y = 0.0343x^2 - 5.4212x + 339.8479$
		경유	65.4km/h 미만	$y = 1153.5685x^{-0.5507}$
			65.4km/h 이상	$y = 0.0226x^2 - 3.0857x + 225.8804$
		LPG	65.4km/h 미만	$y = 1513.8104x^{-0.6075}$
			65.4km/h 이상	$y = 0.0245x^2 - 3.6654x + 257.7428$
		HEV	65.4km/h 미만	$y = 211.9807x^{-0.1884}$
			65.4km/h 이상	$y = 0.0205x^2 - 2.8635x + 190.4598$
	대형	휘발유	65.4km/h 미만	$y = 2022.6604x^{-0.6183}$
			65.4km/h 이상	$y = 0.0374x^2 - 5.9783x + 385.8791$
		경유	65.4km/h 미만	$y = 1149.2206x^{-0.5313}$
			65.4km/h 이상	$y = 0.0246x^2 - 3.3168x + 239.5643$
		LPG	65.4km/h 미만	$y = 1967.2719x^{-0.6616}$
			65.4km/h 이상	$y = 0.0295x^2 - 4.6079x + 301.8248$
HEV		65.4km/h 미만	$y = 522.2199x^{-0.3855}$	
		65.4km/h 이상	$y = 0.0205x^2 - 2.8635x + 190.4598$	
버스	시내	CNG	$y = 5054.5880x^{-0.4910}$	
	전세 광역	경유	65.4km/h 미만	$y = 4317.2386x^{-0.5049}$
			65.4km/h 이상	$y = 0.1829x^2 - 29.8145x + 1670.8962$
	CNG	65.4km/h 미만	$y = 5011.6276x^{-0.5579}$	
65.4km/h 이상		$y = 0.1122x^2 - 17.5798x + 1141.5327$		
화물차	소형	경유	65.4km/h 미만	$y = 1250.4831x^{-0.4630}$
			65.4km/h 이상	$y = 0.0292x^2 - 2.9530x + 258.3205$
	중형	경유	65.4km/h 미만	$y = 1385.8860x^{-0.4184}$
			65.4km/h 이상	$y = 1.6720x + 141.2224$
	대형	경유	$y = 3351.2892x^{-0.4407}$	

○ 일산화탄소배출량의 배출계수

- 도로이동오염원 배출계수로 실적용연식은 '15년 이후에 해당됨

〈표 3-21〉 일산화탄소 배출량의 배출계수

차종	연료	네트워크 점유율		배출계수					
		21'	22'	차속구분	산출식				
승용차	경형	휘발유	6.7%	5.9%	45.0km/h 미만	$Y=4.164xV^{(-0.8461)}$			
					45.0km/h 이상	$Y=-8.3360E-05xV^2+1.9089E-02xV-4.7524E-01$			
	소형	휘발유	0.8%	1.3%	-	$Y=8.3360E-05xV^2-5.9186E-03xV+1.8714E-01$			
					경유	0.7%	0.4%	-	$Y=0.4574xV^{(-0.5215)}$
	중형	휘발유	0.3%	0.3%	-	$Y=1.9089E-05xV^2-1.3588E-03xV+4.8599E-02$			
					경유	0.1%	0.1%	-	$Y=0.4574xV^{(-0.5215)}$
					LPG	36.1%	29.2%	79.6km/h 미만	$Y=0.6374xV^{(-0.7666)}$
	대형	휘발유	17.7%	20.1%	79.6km/h 이상	$Y=4.1425E-16xV^{(7.2766)}$			
LPG					4.1%	6.4%	65.4km/h 미만	$Y=1.1739xV^{(-0.7728)}$	
버스	대형(시내버스)	CNG	2.0%	4.7%	65.4km/h 이상	$Y=6.6688E-05xV^2-1.0587E-02xV+4.7940E-01$			
	대형(고속버스)	경유	19.7%	16.6%	79.6km/h 미만	$Y=0.6374xV^{(-0.7666)}$			
화물차	소형	경유	9.7%	11.4%	79.6km/h 이상	$Y=4.1425E-16xV^{(7.2766)}$			
	중형	경유	2.2%	3.7%	-	$Y=0.935$			
	대형	경유	39.5%	39.9%	-	$Y=11.4415xV^{(-0.8036)}$			

○ 휘발성유기화합물배출량의 배출계수

〈표 3-22〉 휘발성 유기화합물 배출량의 배출계수

차종	연료	네트워크 점유율		배출계수					
		21'	22'	차속구분	산출식				
승용차	경형	휘발유	6.7%	5.9%	-	$Y=0.2343xV^{(-0.7830)}$			
					휘발유	0.8%	1.3%	65.4km/h 미만	$Y=0.0501xV^{(-1.0484)}$
	소형	경유	0.7%	0.4%	65.4km/h 이상	$Y=1.0500E-06xV^2-1.4890E-04xV+6.0984E-03$			
					휘발유	0.3%	0.3%	-	$Y=0.1300xV^{(-0.7265)}$
	중형	경유	0.1%	0.1%	65.4km/h 미만	$Y=0.0501xV^{(-1.0484)}$			
					LPG	36.1%	29.2%	65.4km/h 이상	$Y=1.0500E-06xV^2-1.4890E-04xV+6.0984E-03$
					휘발유	17.7%	20.1%	-	$Y=0.1300xV^{(-0.7265)}$
	대형	LPG	4.1%	6.4%	79.6km/h 미만	$Y=0.0879xV^{(-1.0745)}$			
79.6km/h 이상					$Y=8.2667E-16xV^{(6.2696)}$				
버스	대형(시내버스)	CNG	2.0%	4.7%	79.6km/h 이상	$Y=8.2667E-16xV^{(6.2696)}$			
	대형(고속버스)	경유	19.7%	16.6%	65.4km/h 미만	$Y=0.0501xV^{(-1.0484)}$			
화물차	소형	경유	9.7%	11.4%	65.4km/h 이상	$Y=1.0500E-06xV^2-1.4890E-04xV+6.0984E-03$			
	중형	경유	2.2%	3.7%	79.6km/h 미만	$Y=0.0879xV^{(-1.0745)}$			
	대형	경유	39.5%	39.9%	79.6km/h 이상	$Y=8.2667E-16xV^{(6.2696)}$			

○ 질소산화물배출량의 배출계수

〈표 3-23〉 질소산화물 배출량의 배출계수

차종	연료	네트워크 점유율		배출계수		
		21'	22'	차속구분	산출식	
승용차	경형	휘발유	6.7%	5.9%	-	$Y=0.4106xV^{(-0.9198)}$
	소형	휘발유	0.8%	1.3%	-	$Y=-2.9823E-06xV^2+2.8119E-04xV+9.5434E-03$
		경유	0.7%	0.4%	-	$Y=2.7144xV^{(-0.3437)}$
	중형	휘발유	0.3%	0.3%	-	$Y=2.7702xV^{(-0.3869)}$
		경유	0.1%	0.1%	-	$Y=2.7241xV^{(-0.2743)}$
			-	-	-	$Y=-2.9823E-06xV^2+2.8119E-04xV+9.5434E-03$
		LPG	36.1%	29.2%	-	$Y=2.7144xV^{(-0.3437)}$
	대형	휘발유	17.7%	20.1%	-	$Y=2.7702xV^{(-0.3869)}$
LPG		4.1%	6.4%	-	$Y=2.7241xV^{(-0.2743)}$	
버스	대형(시내 버스)	CNG	2.0%	4.7%	-	$Y=-3.4453E-06xV^2+5.1680E-04xV+4.7373E-03$
	대형(고속 버스)	경유	19.7%	16.6%	-	$Y=-2.9823E-06xV^2+2.8119E-04xV+9.5434E-03$
화물차	소형	경유	9.7%	11.4%	-	$Y=0.0003xV^2-0.041xV+1.4756$
			-	-	-	$Y=112.1229xV^{(-1.6393)}$
	중형	경유	2.2%	3.7%	65.4km/h 미만	$Y=2.0217xV^{(-0.2645)}$
	대형	경유	39.5%	39.9%	65.4km/h 이상	$Y=0.0271xV^{(0.7596)}$

○ 미세먼지배출량의 배출계수

〈표 3-24〉 미세먼지 배출량의 배출계수

차종	연료	네트워크 점유율		배출계수		
		21'	22'	차속구분	산출식	
승용차	경형	휘발유	6.7%	5.9%		
	소형	휘발유	0.8%	1.3%		
		경유	0.7%	0.4%	65.4km/h 미만	$Y=0.0225xV^{(-0.7264)}$
	중형	휘발유	0.3%	0.3%	65.4km/h 이상	$Y=0.0009xV^{(0.0416)}$
		경유	0.1%	0.1%		
	LPG	36.1%	29.2%	65.4km/h 미만	$Y=0.0225xV^{(-0.7264)}$	
대형	휘발유	17.7%	20.1%	65.4km/h 이상	$Y=0.0009xV^{(0.0416)}$	
	LPG	4.1%	6.4%	85km/h 미만	$Y=0.0002$	
버스	대형(시내 버스)	CNG	2.0%	4.7%	85km/h 이상	$Y=0.0005$
	대형(고속 버스)	경유	19.7%	16.6%	85km/h 미만	$Y=0.0002$
화물차	소형	경유	9.7%	11.4%	85km/h 이상	$Y=0.0005$
	중형	경유	2.2%	3.7%		
	대형	경유	39.5%	39.9%	-	$Y=0.0363xV^{(-0.4727)}$



제4장 차량 GPS 데이터 구축·분석 시스템 개선

제1절 차량 데이터 구축·분석 시스템 개요

제2절 차량 데이터 가공 모니터링 기능 개발

제3절 기능 구현 결과

제4장 차량 GPS 데이터 구축·분석 시스템개선

제1절 차량 데이터 구축·분석 시스템 개요

- 데이터 수집과 가공 각 단계를 모니터링하여 데이터가 정상적으로 처리되고 있는지 실시간으로 확인할 수 있는 서비스 지원
- KTDB 기반 지도 및 차량 통행 데이터셋을 체계적으로 관리하고 분석에 활용할 수 있는 시스템 구축
- 데이터셋 가공·구축 시스템 관리
 - 서버 상태 모니터링 환경 구축을 통해 실시간 상태 확인 및 장애 예방 가능
 - 데이터 가공 모듈 개선 시 서버 업데이트 및 개발 정보 지속 관리
- 분석 시스템 관리
 - 수집 현황과 데이터 전처리 상태를 제공하기 위한 대시보드 서비스 제공
 - 공간 단위별로 구축된 교통망 성능 평가 지표 및 통행 분석 DB 업로드 및 버전 관리 지원
 - 데이터 작업 현황에 따라 데이터베이스(MySQL) 저장 및 관리

1. 차량 데이터 가공 이력 항목

가. 데이터 수집 현황

- 전처리 단계 이전에 원시 데이터의 파일 정보, 용량, 오류 등 기초적인 상태를 사전에 확인할 수 있는 정보 수집

〈표 4-1〉 데이터 수집 현황 로그 항목

구분		설명
원시 데이터	티맵 링크	원시 데이터의 이벤트 건수, 파일 건수, 용량 등 기초 정보의 오류 정보 수집
	티맵 포인트	
	DTG 버스, 트럭	

나. 데이터 전처리 정보

- 원시 데이터의 이상치 제거, 정렬, 좌표 변환, 시간 정보 변환 등의 전처리 과정을 각 단계별로 수행하며 일자별 오류 유형, 가공 용량, OBU 정보 등 전처리 관련 정보를 수집

〈표 4-2〉 데이터 전처리 로그 항목

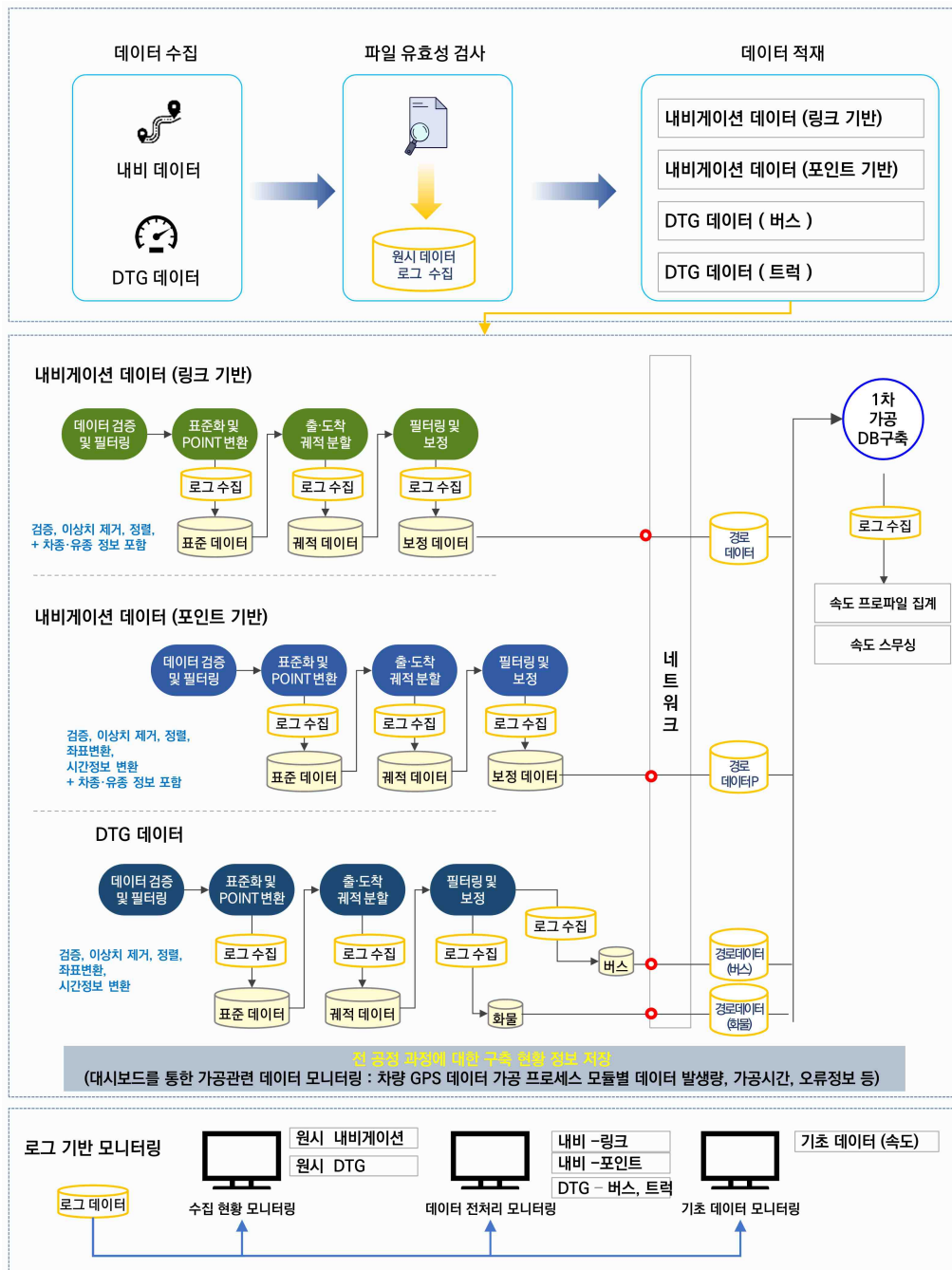
구분		설명
티맵 링크	링크 데이터 전처리	데이터 검증 및 필터링 정보
	링크 데이터 포인트화	표준화 및 Point 변환 정보
	링크 데이터 맵매칭	데이터 맵매칭 정보
티맵 포인트	포인트 데이터 전처리	데이터 검증 및 필터링 정보
	포인트 데이터 보정	표준화 및 Point 변환
	포인트 데이터 맵매칭	데이터 맵매칭 정보
	포인트 데이터 링크 경로 생성	데이터 링크 경로 생성 정보
DTG 버스, 트럭	포인트 데이터 전처리	포인트 데이터 전처리 정보
	포인트 데이터 표준화	포인트 데이터 표준화 정보
	포인트 데이터 보정	포인트 데이터 보정 정보
	포인트 데이터 맵매칭	포인트 데이터 맵매칭 정보
	포인트 데이터 링크 경로 생성	포인트 데이터 링크 경로 생성 정보

다. 기초 교통DB 구축 정보

- 데이터 전처리를 통해 가공된 데이터와 속도 프로파일이 집계, 스무싱 처리된 정보를 수집
 - 속도프로파일 집계 정보 및 이상치 정규화 분포 정보

제2절 차량 데이터 가공 모니터링 기능 개발

- 원시 데이터 수집 및 데이터 전처리에서 이상치 제거, 정렬, 좌표 변환, 시간 정보 변환 등의 전처리 과정을 각 단계별로 수행하며, 일자별 오류 유형, 가공 용량, OBU 등 수집된 로그 기반으로 모니터링 기능 개발



〈그림 4-1〉 모니터링 프로세스

가. 기능 구성

1) 데이터 수집 현황

- 원시 데이터의 파일 유효성 검사 정보를 테이블 및 그래프로 제공
 - 주요 항목 : 파일 수, 오류 파일 수, 총용량 등
 - 년/월/일 별 데이터 용량, 건수를 시각적으로 표현

2) 데이터 전처리

- 이상치 제거, 정렬, 좌표 변환, 시간 정보 변환 등의 데이터 전처리 작업에서 발생하는 로그 정보를 테이블 및 그래프로 제공
 - 주요 항목 : 작업별 처리 데이터 수, 처리 성공 및 오류 수 등
 - 전처리 과정에서 데이터의 상태, 정보 등 각 전처리 과정에 맞는 정보 시각적으로 표현

3) 기초 데이터

- 전처리 데이터와 속도 프로브 데이터에서 생성되는 기초 데이터의 로그 정보를 테이블 및 그래프로 제공



〈그림 4-2〉 모니터링 메뉴 구조도

나. 기능 구축 결과

- 선택한 년도, 기간에 따라 이벤트 건수, 파일 수, 용량 및 데이터 현황 등을 테이블을 통해 사용자는 데이터를 구체적으로 확인할 수 있도록 함



<그림 4-3> 데이터 수집 현황 모니터링 화면

- 선택한 표출 컬럼에 따라 테이블 데이터가 구성되며, 각 전처리 단계에서 발생하는 수집된 로그 정보를 자세하게 확인할 수 있도록 함



<그림 4-4> 데이터 전처리 모니터링 화면

- 조건 설정에 따라 각 데이터 기초에서 발생하는 수집된 로그 정보를 자세하게 확인할 수 있도록 함



〈그림 4-5〉 기초 데이터 모니터링 화면



제5장 결론 및 향후 계획

제5장 결론 및 향후 계획

1. 사업결과 요약

- 본 과업은 전국 상세도로망 네트워크 기준 교통망 성능평가지표 구축 및 지표 구축을 위한 기초교통DB를 가공·구축하였음
- 2023년 기준 전국 개별차량 경로DB 구축, 기초교통 DB 구축, 교통량 추정, 교통망 성능평가지표에 대해 구축하였음
- 2024년 1-6월 기준 월단위 퍼센타일 속도지표에 대해 구축하였음
- 2023년 기준 교통망 성능평가지표를 산정하기 위한 교통량 추정 입력데이터를 고도화하여 추정교통량DB의 신뢰도를 개선하였음
 - 특히, 2023년 교통량 추정을 위한 데이터는 2020년부터 2023년까지 누적모형의 입력데이터로 본 과업을 통해 2023년과 동일한 조건 및 기준으로 2020년 경로데이터부터 재가공함
 - 고도화된 교통량 추정 알고리즘은 학습 데이터 누적방식의 교통량 추정모형으로 입력데이터의 품질이 중요하므로 관측교통량과 프로브 통행량의 이상치를 제거하였음
 - 즉, 2020년부터 2023년까지 고도화된 교통량 추정 알고리즘을 이용하여 재산출하였으므로 기초교통DB 및 교통망 성능평가지표DB를 고도화하였음
- 2023년 기준 교통망 성능평가지표를 활용하여 통계청의 “차량 GPS 기반 도로 유형별 주행거리”를 실험적 통계로 공표할 예정임

2. 사업결과의 활용방향

- 데이터 분석·활용시스템을 이용하여 이용자 문의나 공문을 통해 이용자 맞춤형 데이터를 제공함
 - 특히, 타 기관에서 운영하는 시스템이나 사업의 기반데이터로 이용되어 매년 제공하고 있음
- 주로 중앙부처 및 지자체의 교통운영 및 계획을 위한 정책 수립 근거자료로 사용되었으며, 최근 빅데이터의 관심이 커지면서 학계·개인의 연구 목적으로도 폭넓게 활용되고 있음
- 서울대, 한양대, 인하대 등 대학교에서 개인연구를 위한 데이터 요청 건이 많음
 - 데이터 분석·활용 시스템인 View-T를 통해 교통망 성능평가지표 데이터 및 경로데이터를 제공하였으며, 데이터 제공 실적이 8,696에서 13,744건으로 전년 대비 약 37% 증가함

〈표 5-1〉 데이터 분석·활용 시스템 활용 실적

기능	설명	데이터 건수
링크 다운로드	이용자 선택 도로 구간(링크)별로 통행지표 데이터를 다운받음	7,122
교통망 성능평가지표 데이터 다운로드	통행지표를 행정구역별, 도로별, 시간대별 등을 선택하여 데이터를 제공함	2,528
월단위 데이터 다운로드	가장 최신의 월 단위 데이터를 받을 수 있으며, 퍼센타일 속도지표를 다운받음	1,195
분석도구 다운로드	지도화면의 분석도구 결과데이터를 다운받음	2,899
	총계	13,744

〈표 5-2〉 데이터 제공 및 분석 지원 주요 사례

1차 가공DB (속도분포, 경로데이터 등)	<ul style="list-style-type: none"> ■ 법정계획 및 정책 수립 지원 <ul style="list-style-type: none"> · 국가-지자체 도로정책 수립 지원 연구 	수시
	<ul style="list-style-type: none"> ■ 데이터 관련 사업 지원 <ul style="list-style-type: none"> · 부동산 빅데이터 플랫폼 구축사업 · 서울시 개인차량 이동행태 분석 사업 	상시 수시
	<ul style="list-style-type: none"> ■ 민간 및 학술적 연구 지원 <ul style="list-style-type: none"> · 온실가스 저감을 위한 국토도시공간 계획 및 관리기술 연구 · 영업용 화물자동차의 충전인프라 구축 연구 · 전기차 배터리 수명개선용 BMS 알고리즘 연구 · 국토30호선 차량 과속관련 교통안전 연구 	수시
	<ul style="list-style-type: none"> ■ 법정계획 및 정책 수립 지원 <ul style="list-style-type: none"> · 통계청의 차량 GPS 기반 도로유형별 주행거리 실험적 통계 지원 · 도로건설계획(2026-2030) 수립 연구 · 지방 교통시설 국고보조금 관리 방안 수립 연구 	상시 수시
기초지표 (교통량, 속도, 차량주행거리)	<ul style="list-style-type: none"> ■ 데이터 관련 사업 지원 <ul style="list-style-type: none"> · 교통 접근성 지표 산출을 위한 기초데이터 지원 · 교통 빅데이터 플랫폼 구축사업 · 환경소음측정망 적정 측정 지점 선정 사업 · 탄소공간지도 기반 계획지원 기술개발 사업 · COMPAS 국토도시데이터 분석과제 공모전 	상시
	<ul style="list-style-type: none"> · 머신러닝을 이용한 소음지도 작성 사업 · 교통 빅데이터 활성화를 위한 지자체 지원사업 · 경기도 공공투자사업 투자심사 사전검토 	수시
	<ul style="list-style-type: none"> ■ 민간 및 학술적 연구 지원 <ul style="list-style-type: none"> · 국토교통부 주관 데이터 공모전 · 도심항공 UAM 운용을 위한 연구 	상시
	<ul style="list-style-type: none"> · 대도시권 광역교통 주관 경진대회 · 울릉도 교통현안 분석 · 탄소중립 수송부문 감축 고도화 기술 개발 연구 	수시
성능평가지표 (혼잡지표, 환경지표, 안전지표 등)	<ul style="list-style-type: none"> ■ 법정계획 및 정책 수립 지원 <ul style="list-style-type: none"> · e-나라지표의 도로교통 혼잡비용 지원 · 행정중심복합도시 장래 도로계획 수립 	상시 수시
	<ul style="list-style-type: none"> ■ 데이터 관련 사업 지원 <ul style="list-style-type: none"> · 대도시권 광역교통 통계 분석 · 교통 빅데이터 플랫폼 구축사업 · COMPAS 국토도시데이터 분석과제 공모전 	상시
	<ul style="list-style-type: none"> · 교통 빅데이터 활성화를 위한 지자체 지원사업 	수시
	<ul style="list-style-type: none"> ■ 민간 및 학술적 연구 지원 <ul style="list-style-type: none"> · 국토교통부 주관 데이터 공모전 	상시
	<ul style="list-style-type: none"> · 대도시권 광역교통 주관 경진대회 · 인천광역시 보통교부세 진단 및 보정수요발굴 연구 · 해운대구 탄소중립 녹색성장 기본계획 수립연구 	수시

* 혼잡지표(교통혼잡비용, 혼잡시간강도, 혼잡빈도강도, 혼잡시평균속도),
환경지표(이산화탄소배출량, 일산화탄소배출량, 미세먼지배출량, 질소산화물배출량,
휘발성유기화합물배출량),
안전지표(운전자피로도, 과속비율, 속도편차)

3. 향후 계획

- 티맵 포인트 기반 내비게이션 데이터의 특징을 분석하여 링크 기반 내비게이션 데이터와의 일관성을 위한 데이터 가공 방법론 검토
- 전국 2020년부터 2023년 교통망 성능평가지표 산정결과를 View-T 웹서비스를 통해 공표 예정
- 2023년 기준 교통망 성능평가지표를 활용하여 통계청의 “차량 GPS 기반 도로 유형별 주행거리”를 실험적 통계로 공표 예정

2024
국가교통조사
및 분석

10



차량 GPS 빅데이터 구축