

2022년 국가교통조사

모바일통신 빅데이터 구축

10

2022. 12



국토교통부
Ministry of Land, Infrastructure and Transport



한국교통연구원
KOREA TRANSPORT INSTITUTE

제 출 문

국토교통부장관 귀하

본 보고서를 「2022년 국가교통조사」 최종보고서로 제출합니다.

2022년 12월

한국교통연구원

원장 오 재 학

본 『2022년 국가교통조사』는 다음 연구진에 의해
수행되었습니다.

참 여 연 구 진

<한국교통연구원>	
연구책임자	<ul style="list-style-type: none"> ◦ 조종석 연구위원
연구진	<ul style="list-style-type: none"> ◦ 조범철 연구위원 ◦ 김주영, 천승훈, 박용일 연구위원 ◦ 황순연, 장동익, 원민수, 이송봉, 이종우 부연구위원 ◦ 김동호, 신영권 책임전문원 ◦ 김규진, 김정은 주임전문원 ◦ 안덕배 전문연구원 ◦ 가보연, 강국수, 광명신, 권기훈, 김운태, 김 현, 박미란, 박준호, 신유선, 양태양, 오연선, 이선아, 이슬기, 이채영, 채정표, 홍성표 연구원 ◦ 홍연우 연구조원
<한국해양수산개발원>	
연구진	<ul style="list-style-type: none"> ◦ 이호춘, 최건우 부연구위원 ◦ 황수진 전문연구원 ◦ 박일란 선임사무원 ◦ 류희영 연구원
<한국항공협회>	
연구진	<ul style="list-style-type: none"> ◦ 성인영 실장 ◦ 손병일 책임연구원 ◦ 최인영 과장 ◦ 김지한, 김창욱, 김진성 대리

『2022년 국가교통조사』
보고서 구성 및 담당연구진

번 호	과 제 명	연 구 진
제 1권	요약보고서	조종석, 신영권, 가보연
제 2권	전국 여객 O/D 전수화	조종석, 강국수, 박미란, 채정표
제 3권	교통분석용 네트워크 구축	김동호, 이선아, 이슬기
제 4권	항공여객 O/D 조사	한국항공협회
제 5권	전국화물 O/D 본조사	김주영, 황순연, 권기훈, 김정은, 오연선, 김문태
제 6권	전국화물 O/D 보완갱신	김주영, 김정은
제 7권	해상화물 O/D 본조사	한국해양수산개발원
제 8권	KTDB 플랫폼 기반지도 구축	이송봉, 양태양
제 9권	차량 GPS 빅데이터 구축	천승훈, 이종우, 이재영
제10권	모바일통신 빅데이터 구축	원민수, 신우선
제11권	국가교통통계DB구축	박용일, 권명신
제12권	특별교통대책기간 동행실태조사	안락배, 김 현
제13권	교통접근성지표 구축	장동익, 홍성표

『2022년도 국가교통조사』
과제별 공동참여·위탁용역 사업자

【공동사업 참여기관】
<ul style="list-style-type: none"> • 전국 여객O/D 현행화 공동사업 (수도권 부문) <ul style="list-style-type: none"> - 경기연구원, 인천연구원, 서울연구원 • 항공O/D 및 특성 조사 <ul style="list-style-type: none"> - (사)한국항공협회 • 국가교통DB 점검단 <ul style="list-style-type: none"> - 대한교통학회
【위탁용역 사업자】
<ul style="list-style-type: none"> • 전국여객 O/D 전수화 대전세종충청권 <ul style="list-style-type: none"> - ㈜신명이앤씨, ㈜에스트리 • 전국여객 O/D 전수화 제주권 <ul style="list-style-type: none"> - 홍익대학교 산학협력단 • 전국여객 O/D 전수화 대구광역시권 <ul style="list-style-type: none"> - 홍익대학교 산학협력단 • 전국여객 O/D 전수화 부산울산권 <ul style="list-style-type: none"> - ㈜신명이앤씨, ㈜에스트리 • 도로 및 철도 교통분석용 네트워크 보완갱신 <ul style="list-style-type: none"> - 서울시립대학교 산학협력단 • 전국화물 분조사 사업체 및 화물자동차 표본선계 <ul style="list-style-type: none"> - (사)한국교통정책경제학회

【위탁용역 사업자】

- 전국화물 분조사 사업체물류현황조사
- ㈜메트릭스
- 전국화물 분조사 화물자동차 통행실태조사
- ㈜코리아데이터네트워크
- 전국화물 분조사 물류거점조사
- ㈜코리아데이터네트워크
- 영업용화물자동차운행기록계 빅데이터를 이용한 화물 기종점통행량 및 운행특성 분석
- ㈜노트스퀘어
- 모빌리티 빅데이터 DB 구축 및 온라인 서비스 유지보수
- ㈜큐빅웨어
- 빅데이터 플랫폼사업 감리용역
- 악티브
- 모바일통신 원시 데이터를 이용한 통행사슬 DB 구축
- ㈜오픈메이트온
- 모바일통신 원시 데이터 전처리 최적화 및 시스템 연결
- ㈜ 오픈메이트온
- 특별교통통행실태조사 및 이용자 만족도 조사
- ㈜컨슈머인사이트

최종보고서 목차

- 제 1권 요약보고서**
- 제 2권 전국여객 O/D 전수화**
- 제 3권 교통분석용 네트워크 구축**
- 제 4권 항공여객 O/D 조사**
- 제 5권 전국화물 O/D 본조사**
- 제 6권 전국화물 O/D 보완갱신**
- 제 7권 해상화물 O/D 본조사**
- 제 8권 KTDB 플랫폼 기반지도 구축**
- 제 9권 차량 GPS 빅데이터 구축**
- 제 10권 모바일통신 빅데이터 구축**
- 제 11권 국가교통통계DB구축**
- 제 12권 특별교통대책기간 통행실태조사**
- 제 13권 교통접근성지표 구축**

목 차

요 약

제1장 과업의 개요 3

제1절 과업의 배경 및 목적 / 3

제2절 과업의 범위 및 내용 / 5

제2장 모바일 교통 빅데이터 전처리·가공 알고리즘 개발 및 고도화 9

제1절 원시데이터 전처리·가공 알고리즘 고도화 / 9

제2절 체류지 및 통행목적 추정 알고리즘 고도화 / 15

제3절 교통수단 추정 및 전수화 방법론 개발 / 28

제4절 데이터 가공 프로세싱 최적화 및 속도 개선 / 40

제3장 모바일 교통 빅데이터 기반 통행사슬(Trip Chain) DB 구축 및 검증 49

제1절 기준년도 통행사슬(Trip Chain) DB 구축 / 49

제2절 기준년도 구축 DB 검증 / 60

제4장 모바일 교통 빅데이터 기반 교통지표 개발 및 온라인 서비스 제공 79

제1절 기초 DB 재구조화 / 79

제2절 교통지표 개발 및 온라인 서비스 제공 / 105

제3절 온라인 서비스 운영 및 유지보수 / 144

제5장 결론 및 차년도 수행계획 153

제1절 결론 / 153

제2절 차년도 수행계획 / 154

표 목 차

〈표 1〉 모바일 기반 DB 형태	11
〈표 2〉 모바일 통신 빅데이터 DB 테이블 정의서	13
〈표 3〉 체류지 식별 기준	15
〈표 4〉 지역별 통행량 비교 결과	18
〈표 5〉 지역별 출·퇴근 통행량 비교 결과	20
〈표 6〉 지역별 출퇴근 통행시간 비교 결과	21
〈표 7〉 주요 특성 통행지역 분석 그림	23
〈표 8〉 주요통행구간 분석기능의 연도별 표출 형태 변경	29
〈표 2-1〉 체류지 유형별 식별 기준	19
〈표 2-2〉 개선 알고리즘 단계별 코드 구성 내용	25
〈표 2-4〉 지역 간 교통수단 구분 (예시)	28
〈표 2-5〉 개선 알고리즘 단계별 코드 구성 내용	41
〈표 2-6〉 전체 프로세스별 소요시간	41
〈표 2-7〉 산출물 상세 소요시간	42
〈표 3-1〉 모바일 기반 DB 형태	49
〈표 3-2〉 모바일 통신 빅데이터 DB 테이블 정의서	51
〈표 3-3〉 체류지 식별 기준	55
〈표 3-4〉 지역별 통행량 비교 결과	61
〈표 3-5〉 시간대별 통행량 패턴 분석 (모바일 통신 빅데이터 기준)	64
〈표 3-6〉 요일별 통행량 패턴 분석 (모바일 통신 빅데이터 기준)	66
〈표 3-7〉 연령별 통행량 패턴 분석 (모바일 통신 빅데이터 기준)	68
〈표 3-8〉 지역별 출·퇴근 통행량 비교 결과	70
〈표 3-9〉 지역별 출퇴근 통행시간 비교 결과	72
〈표 3-10〉 통행시간 기준별 통근 인구 비율	73
〈표 4-1〉 서비스를 위한 모바일 데이터 가공	80
〈표 4-2〉 격자-교통폴리곤 관계데이터 테이블 정의서	81
〈표 4-3〉 POI 데이터 테이블 정의서	85
〈표 4-4〉 모바일 빅데이터 특징	90

〈표 4-5〉 모바일 데이터 구조	93
〈표 4-6〉 교통폴리곤 기준 전처리 및 집계 통행량 컬렉션의 구성	95
〈표 4-7〉 읍면동-읍면동 단위 통행량 컬렉션의 구성	98
〈표 4-8〉 시군구-시군구 단위 통행량 컬렉션의 구성	99
〈표 4-9〉 시도-시도 단위 통행량 컬렉션의 구성	100
〈표 4-10〉 통행목적별 집계 통행량 컬렉션의 구성	101
〈표 4-11〉 교통폴리곤 기준 통행량 DB 검증 항목	103
〈표 4-12〉 행정구역 단위 통행량 합계 비교 검증 항목	104
〈표 4-13〉 통행목적별 집계 통행량 합계 비교 검증 항목	104
〈표 4-14〉 View-T Light 주요 기능	106
〈표 4-15〉 View-T Expert 분석 도구 및 시뮬레이션	111
〈표 4-16〉 View-T 대시보드 주요 기능 설명	112
〈표 4-17〉 View-T 부가 기능 설명	112
〈표 4-18〉 View-T 웹서비스용 DB 데이터 업데이트 내용	146
〈표 4-19〉 View-T 웹 서비스 성능 테스트	148

그림목차

〈그림 1〉 과업의 수행 절차	5
〈그림 2〉 모바일 통신 빅데이터 가공알고리즘	6
〈그림 3〉 통행자 실제 위치 추정 방법 예시	7
〈그림 4〉 최장체류지 식별 후, 체류시간 aggregation 예시	8
〈그림 5〉 통행지표 산출물 예시	9
〈그림 6〉 모바일 통신 빅데이터 가공 DB 구축 과정	10
〈그림 7〉 기종점 인구 통행구간 분석의 시도 단위 전국 분석	24
〈그림 8〉 고령자 통행 비중 분석의 읍면동 단위 분석 결과 표출	24
〈그림 9〉 고령자 주요 통행 구간 분석의 읍면동 단위 분석 결과 표출	25
〈그림 10〉 통행시간·거리 분석 도구 화면예시	25
〈그림 11〉 출퇴근 통행 접근성 분석의 기준 지역 선택 UI와 범례 색상 적용된 결과 그래프	26
〈그림 12〉 계절별 핫플레이스 분석 결과 (좌)유동인구 통행량 (우)관광목적·연령별 통행량	26
〈그림 13〉 관광지 내·외부 통행 비율 분석 결과	27
〈그림 14〉 (좌) 특별시·광역시를 대상 인접도시 출근 비율 (우) 특별시·광역시 지역 내 통행량 ..	27
〈그림 15〉 통행지표 서비스 대시보드 화면 구성	28
〈그림 16〉 주요통행구간 분석의 전국분석 결과 및 통행량 분포도 그래프	29
〈그림 17〉 출/퇴근 통행 연결성 분석 도구 화면예시	30
〈그림 18〉 유입 기준 폴리곤 형태의 통행 변화 시뮬레이션	31
〈그림 19〉 지역 내 기준 폴리곤 및 희망선 형태의 통행 변화 시뮬레이션	31
〈그림 20〉 관광 특화지역 분석기능 화면	32
〈그림 21〉 도심공동화 심각도 분석화면	32
〈그림 1-1〉 과업의 수행 절차	6
〈그림 2-1〉 모바일 교통 빅데이터 가공알고리즘	9
〈그림 2-2〉 선분 이력 데이터로의 변환 예시	11
〈그림 2-3〉 시간적 데이터 보정 방법 예시	12
〈그림 2-4〉 통행자 실제 위치 추정 방법 예시	13
〈그림 2-5〉 통행자 실제 위치 추정 방법 예시	14
〈그림 2-6〉 시·공간적 데이터 보정(전처리) 적용 전, 후 통행 궤적 비교	14

〈그림 2-7〉 최장체류지 식별 예시	15
〈그림 2-8〉 최장체류지 이외의 체류지에 체류한 시간 중 가장 긴 시간 식별 예시	16
〈그림 2-9〉 첫 번째 주 체류지(최장체류지) 점유시간 산정 예시	16
〈그림 2-10〉 최장체류지 이외 체류지 중 최장체류지 식별	17
〈그림 2-11〉 두 번째 주 체류지 식별 및 점유시간 산정 예시	18
〈그림 2-12〉 통행시간 계산 및 잠재체류 식별 예시	20
〈그림 2-13〉 주체류지 기준 trip type 구분	20
〈그림 2-14〉 일별 데이터에 대한 주체류지 간 통행량 산정 예시	21
〈그림 2-15〉 잠재체류 식별 예시	22
〈그림 2-16〉 Data Lagging 예시	23
〈그림 2-17〉 통행지표 산출물 예시	23
〈그림 2-18〉 지역 간 교통수단 구분 로직 및 결합 흐름도	28
〈그림 2-19〉 서울역, 용산역, 서울고속버스터미널 shp (예시)	29
〈그림 2-20〉 고객 동선 중 항공 이용자 구분 예시	30
〈그림 2-21〉 고객 동선 중 고속철도 이용자 구분 예시	31
〈그림 2-22〉 고객 동선 중 고속버스 이용자 구분 예시	31
〈그림 2-23〉 지역 간 교통수단 구분 결합 전·후 중간 산출물 테이블 정의서 (예시)	33
〈그림 2-24〉 지역 간 교통수단 구분 결합 후 최종 산출물 테이블 (예시)	33
〈그림 2-25〉 시장점유율 기반 전수화 계수 적용 방법	35
〈그림 2-26〉 O 기준 거주지(H)에 따른 전수화 적용(안)	36
〈그림 2-27〉 O 기준 거주지(H)에 따른 전수화 적용 테이블	37
〈그림 2-28〉 개별 통행 OD 데이터 결과 그림	39
〈그림 2-29〉 집계인구 기반에 따른 전수화 적용 안	39
〈그림 2-30〉 기준직경 내 기준체류시간 초과 예시	44
〈그림 2-31〉 이동소요시간이 없는 장기체류지 예시	45
〈그림 3-1〉 선분 이력으로의 변환 예시	52
〈그림 3-2〉 체류 순서 부여 (개인별 이동궤적 형성)	53
〈그림 3-3〉 이동정보와 체류 정보 구분 (예시)	54
〈그림 3-4〉 체류지 유형 구분 (예시)	56
〈그림 3-5〉 위치정보, 시간정보, 연령 정보 변환 (예시)	57
〈그림 3-6〉 출발/도착 구분 및 통행량 집계 (예시)	58

〈그림 3-7〉 기지국 기반 데이터의 통행시간 산출 방법	59
〈그림 3-8〉 지역별 통행 발생량 비교 결과	62
〈그림 3-9〉 지역별 통행 도착량 비교 결과	62
〈그림 3-10〉 시간대별 통행 발생량 패턴 분석 (모바일기지국데이터 기준)	65
〈그림 3-11〉 시간대별 통행 도착량 패턴 분석 (모바일 통신 빅데이터 기준)	65
〈그림 3-12〉 요일별 시간대에 따른 통행 발생량 변화 (모바일 통신 빅데이터 기준)	66
〈그림 3-13〉 요일별 시간대에 따른 통행 도착량 변화 (모바일 통신 빅데이터 기준)	67
〈그림 3-14〉 연령별 통행 발생량 패턴 분석 (모바일 통신 빅데이터 기준)	69
〈그림 3-15〉 연령별 통행 도착량 패턴 분석 (모바일 통신 빅데이터 기준)	69
〈그림 3-16〉 시도별 통근 통행 비율 비교	71
〈그림 3-17〉 시도별 통근 통행시간 비교 (단위 : 분)	73
〈그림 3-18〉 통행시간 기준별 통근 인구 비율 비교	74
〈그림 3-19〉 출근 시간 밀도 그래프	75
〈그림 3-20〉 퇴근 시간 밀도 그래프	76
〈그림 4-1〉 격자-교통폴리곤 간 형상정보 중첩 유형	80
〈그림 4-2〉 격자와 교통폴리곤 중첩 형상	82
〈그림 4-3〉 격자-교통폴리곤 간 면적 비율 산출 방안	83
〈그림 4-4〉 격자로의 OD 통행량 재산출 방안	83
〈그림 4-5〉 격자 체계 기준	84
〈그림 4-6〉 격자 크기별 격자ID (gid) 체계	84
〈그림 4-7〉 국가관심지점(POI) 데이터 형상 예시	85
〈그림 4-8〉 POI_CL_DC 항목의 세부 코드표(관광 검색)	86
〈그림 4-9〉 POI 데이터 내 관광지시설 검색	86
〈그림 4-10〉 용도별 건물공간정보 형상 예시	87
〈그림 4-11〉 건물공간정보 내 관광 관련 데이터 검색	88
〈그림 4-12〉 용도지역 중 도시지역 형상 예시	88
〈그림 4-13〉 용도지역 중 관광진흥/관광지 주제도 데이터 예시	89
〈그림 4-14〉 모바일 빅데이터 자료 형태	90
〈그림 4-15〉 모빌리티 빅데이터 구축 프로세스	91
〈그림 4-16〉 교통폴리곤 기준 통행량 데이터 구조	94
〈그림 4-17〉 집계 통행량 DB 구축 프로세스	96

〈그림 4-18〉 행정구역 단위의 출발/도착 기준 통행량 DB 구조	97
〈그림 4-19〉 격자로의 OD통행량 산출 예시	102
〈그림 4-20〉 View-T 시스템 구성도	105
〈그림 4-21〉 모바일 통신 데이터 통행지표 서비스 화면	113
〈그림 4-22〉 통행지표 서비스 대시보드 화면 구성	114
〈그림 4-23〉 사용자 레이아웃 및 위젯 추가하기 예시 화면	115
〈그림 4-24〉 어두운 테마 / 밝은 테마 비교 화면	116
〈그림 4-25〉 그리드 라인 보기 옵션 ON/OFF 비교 화면	116
〈그림 4-26〉 위젯의 위치 이동 및 크기 변경을 통한 사용자 레이아웃 예시 화면	117
〈그림 4-27〉 위젯 메뉴 버튼 설명 화면	117
〈그림 4-28〉 지도 지역 선택 위젯(좌)과 목록 지역 선택 위젯(우)	118
〈그림 4-29〉 평균 통근시간 위젯 및 전국 값 표시 옵션 예시 화면	119
〈그림 4-30〉 열악 통행자 비율 위젯 및 현재 지역 표시 고정 옵션 예시 화면	119
〈그림 4-31〉 평균 통근시간 비교 위젯 설명	120
〈그림 4-32〉 열악 통행자 비율 비교 위젯 설명	120
〈그림 4-33〉 평균 통근시간 순위 위젯	121
〈그림 4-34〉 열악 통행자 비율 순위 위젯	121
〈그림 4-35〉 통근시간 밀도 위젯	122
〈그림 4-36〉 열악 통행자 비율 분포 위젯	122
〈그림 4-37〉 지역 간 평균 통근시간 위젯	123
〈그림 4-38〉 모빌리티 월간 리포트 화면	124
〈그림 4-39〉 모빌리티 월간 리포트 기간 선택 UI	125
〈그림 4-40〉 모빌리티 월간 리포트의 결과 요약 내용	125
〈그림 4-41〉 HotPlace 분석 도구 화면	126
〈그림 4-42〉 표출 단위 자동 설정 화면	126
〈그림 4-43〉 표출 단위 자동 설정 적용 화면	127
〈그림 4-44〉 관리자 다운로드 현황 화면	128
〈그림 4-45〉 분석 도구 차량 유지보수 전후 비교 화면	129
〈그림 4-46〉 홈페이지 로그인, MyPage 삭제 화면	130
〈그림 4-47〉 데이터 다운로드 단계 축소	130
〈그림 4-48〉 사회경제활동지표 화면예시	131

〈그림 4-49〉 계절별 핫플레이스 분석 도구 화면예시	132
〈그림 4-50〉 주요 통행지역 분석 도구 화면예시	133
〈그림 4-51〉 주요 통행구간 분석 도구 화면예시	134
〈그림 4-52〉 통행시간·거리 분석 도구 화면예시	135
〈그림 4-53〉 출·퇴근 통행시간 분석 도구 화면예시	136
〈그림 4-54〉 출·퇴근 통행량 분석 도구 화면예시	137
〈그림 4-55〉 근무형태에 따른 통근 통행 분석 도구 화면예시	138
〈그림 4-56〉 고용인구 변화에 따른 통근 통행 분석 도구 화면예시	139
〈그림 4-57〉 출·퇴근 연결성 분석 도구 데이터 가공 방법	140
〈그림 4-58〉 출/퇴근 통행 연결성 분석 도구 화면예시	140
〈그림 4-59〉 관광 특화지역 분석 도구 화면예시	141
〈그림 4-60〉 도심 공동화 분석 도구 화면예시	142
〈그림 4-61〉 출·퇴근 개인 속성(고령자) 통행 특성(통행비중) 분석 도구 화면예시	143
〈그림 4-62〉 MongoDB 샤딩 구조	144
〈그림 4-63〉 레인지 샤딩의 구성 화면	145
〈그림 4-64〉 해쉬 샤딩의 구성	145
〈그림 4-65〉 DB 장애 대응 전략	146
〈그림 4-66〉 View-T 데이터 가공 성능 테스트 시나리오	147
〈그림 4-67〉 구글에서 제공하는 웹페이지 속도 테스트 화면	149
〈그림 4-68〉 자체 동작 테스트로 발견된 오류 및 추가 개선사항의 목록	150
〈그림 4-69〉 웹 사이트 동작 테스트 프로세스	150

요 약

요 약

1. 과업의 개요

가. 과업의 배경 및 목적

- 최근 교통 분야에서 각종 모빌리티 정보(예를 들어, 차량GPS 데이터, 대중교통카드 데이터, DTG 데이터, 모바일 통신 빅데이터 등)를 활용하여 개별 통행 정보를 추출하고 각종 통행 관련 분석이 가능해지면서 다양한 방법으로 활용되고 있으나, 각 데이터의 특성에 따라 분석 대상과 범위가 제한적임
- 그러므로, 전 국민의 95% 이상이 사용하고, 이용 교통수단에 구애받지 않고 구독이 가능한 모바일 통신 빅데이터를 이용한 개별통행 DB 구축 및 활용이 요구되고 있음
 - 모바일 기기와 기지국 간의 송·수신 이력인 모바일 통신 빅데이터는 대부분의 국민이 이용하고 있는 모바일 기기를 통해 축적되기 때문에 표본율이 매우 높고, 이동수단에 구애받지 않고(항공 제외), 기기의 전원을 끄지 않는 이상 개인의 이동궤적이 상세히 기록되기 때문
 - 1년 365일 수집되는 모바일 통신 빅데이터는 사회적 현안에 즉각적으로 대응할 수 있는 시의성 있는 모빌리티 정보를 제공할 수 있음
- 그러나, 기존 민간과 공공에서는 다양한 방식으로 모바일 통신 빅데이터를 가공하여 유동인구 형태로 활용하고 있지만 개별통행 행태를 분석하기에는 한계점이 있으며, 이를 극복하기 위한 기술적 DB 구축 및 분석 방법이 요구되고 있음
- 모바일 통신 빅데이터 중 기지국 기반의 데이터를 이용하여 시의성 있는 O/D 기반의 통행정보(통행량, 통행시간) 생성 방법을 정의하고 DB 구축 및 제공
- 통행자별 이용 교통정보 정보가 포함된 기지국 기반 데이터를 이용하여 통행 DB를 구축하고, 이를 기반으로 교통수요를 분석할 수 있는 지표, 기능 등을 개발하여 고도화된 교통 정책 수립 지원에 활용하고자 함

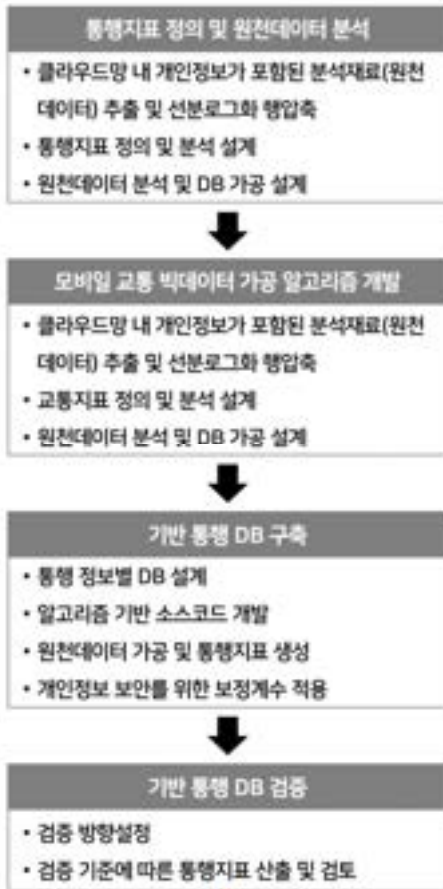
나. 과업의 범위

- 시간적 범위
 - 모바일 빅데이터 가공 알고리즘 개발: 2022년 원시데이터
 - 2022년도 기준 기반 DB 구축: 2022년 1월 1일 ~ 2022년 12월 31일 (1년)
 - 웹 서비스를 위한 각종 분석 도구 개발: 2020년 1월 1일 ~ 2020년 12월 31일
 - 2021년도 연구 성과물인 2020년도 기준 기반 DB 활용
- 공간적 범위: 전국

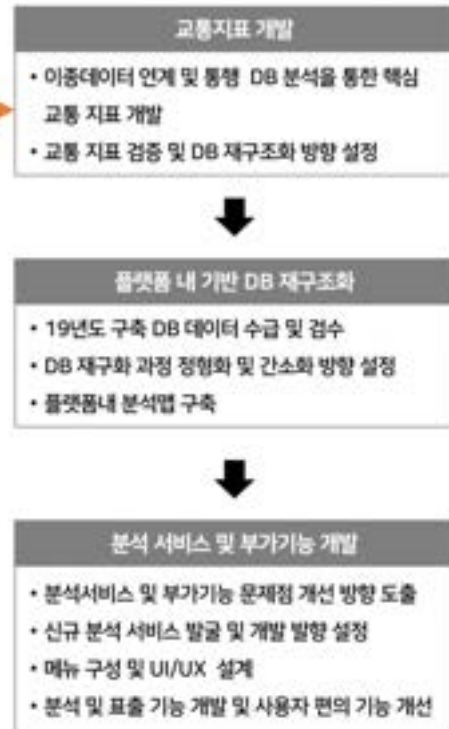
다. 과업의 내용

- 모바일 교통 빅데이터 전처리·가공 알고리즘 개발 및 고도화
 - 원시데이터 전처리·가공 알고리즘 고도화
 - 체류지 및 통행목적 추정 알고리즘 고도화
 - 교통수단 추정, 전수화, 가명화 알고리즘 개발(22년도 사업 중 신규 개발 항목)
 - 데이터 가공 프로세싱 최적화 및 속도 개선
- 모바일 교통 빅데이터 기반 통행사슬(Trip Chain) DB 구축 및 검증
 - 기준년도(22년도) 통행사슬(Trip Chain) DB 구축
 - 기준년도(22년도) 구축 DB 검증
- 모바일 교통 빅데이터 기반 교통지표 개발 및 온라인 서비스 제공
 - 기초 DB 처리 및 재구조화
 - 주요 교통지표 현행화 및 신규 지표 개발
 - 온라인 서비스(View-T) 제공

모바일 교통빅데이터 DB 구축 및 검증 부문



모바일 교통빅데이터 기반 교통지표 및 서비스 개발

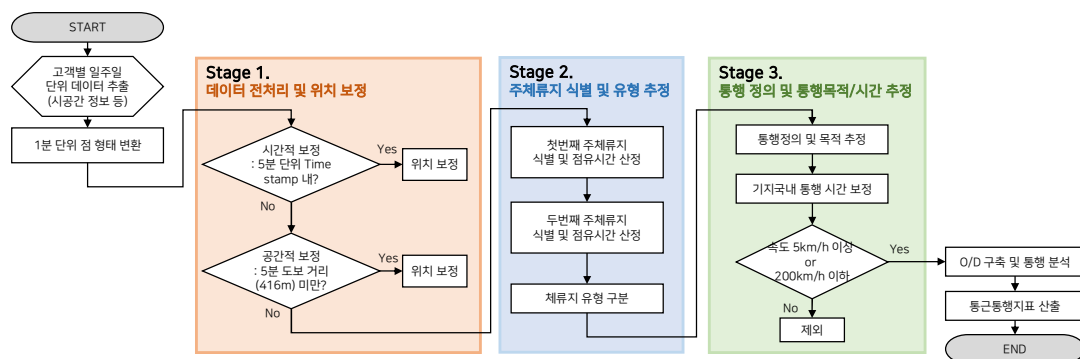


<그림 1> 과업의 수행 절차

2. 모바일 교통 빅데이터 전처리·가공 알고리즘 개발 및 고도화

가. 알고리즘 주요 개념

- 입력데이터(input data)로 각 통행자에 대한 7일 치 원시데이터가 사용되며 각 통행자에 대한 개별통행 특성을 기반으로 주체류지를 식별하고 통행지표를 산출
- 데이터 전처리 단계에서 신호 이상 현상 및 신호 이상치를 정형화할 필요 없이 데이터를 시·공간적으로 보정함으로써 비현실적으로 발생하는 통행자의 이동궤적을 보정
- 기지국 기반 데이터 기록 빈도 및 체류시간을 기반으로 통행자의 개별통행 특성을 고려한 주체류지 식별을 수행하기 때문에 정기 출퇴근 통행뿐만 아니라 야간 근무자를 포함한 비정기 출퇴근 통행도 식별 가능
- 개별 통행자에 대한 주체류지를 식별한 후, 해당 위치에서의 체류시작시간과 체류종료시간을 기반으로 통근시간 및 통행량 등 통행지표 산출



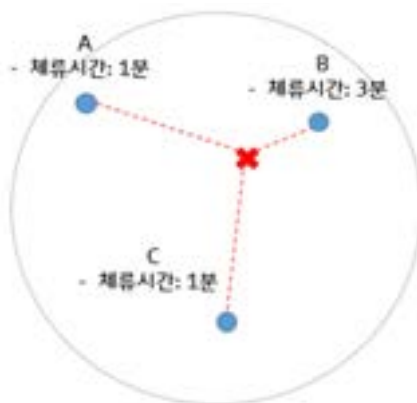
<그림 2> 모바일 통신 빅데이터 가공알고리즘

나. 전처리 알고리즘 개발

- 적절한 공간적 해상도를 바탕으로 통행자의 통행속도를 고려한 이동궤적을 식별하기 위해서 5분이라는 기준 시간을 정하고 해당 시간대에 기록된 기지국의 기록 빈도를 기반으로 통행자의 실제 위치를 추정하고자 함
 - 짧은 시간 동안 발생한 핑퐁 핸드오버 현상에 의한 체류지 식별률 저하 및 비현실적으로 기록된 이동궤적을 보정하기 위하여 위해 특정 시간 단위(5분)별로 데이터를 분할
- 데이터를 분할 한 후, 각 시간 단위(time slice) 내에 기록된 기지국 데이터의 기록빈

도를 기반으로 통행자의 실제 위치 추정

- 통행자의 실제 위치 추정 방법: 각 기지국에서의 체류시간을 기반으로 가중치를 주어 통행자 위치 좌표(X)를 계산



<그림 3> 통행자 실제 위치 추정 방법 예시

- 보행자 평균 도보 속도(5km/h)를 기준으로 5분(시간적 데이터 보정 시 설정한 시간 단위) 거리(약 416m) 범위 내 위치한 체류지들에 대해 각 체류지에서의 체류시간을 기반으로 해당 시간대에서의 통행자 실제 위치를 추정하고자 함
- 원시데이터에 대해 시·공간적으로 신호 이상 기록을 보정해줌으로써, 분 단위로 연속해서 변화하거나 좁은 공간적 범위 내에 산발적으로 기록되는 기지국 위치보정 가능

다. 체류 정보 및 통행목적 추정 알고리즘 개발

- 데이터 전처리 단계에서 추정한 통행자의 실제 위치를 기준으로 통행자가 일주일 동안 각 위치에 몇 분 체류하였는지 계산 후, 최장체류지 식별
- 통행자가 최장체류지 이외의 체류지들에서 연속적으로 체류한 시간을 일별로 계산하여 그중에서 가장 긴 시간을 식별
- 해당 시간 이외에는 통행자가 최장체류지에 체류한 것으로 지정하며, 해당 시간에 속하는 체류지들에 대하여 일주일 기준 체류시간이 가장 긴 체류지 식별(2번째 최장체류지 식별 시, 최장체류지로 지정된 체류지들을 후보에서 제외하고 식별)
- 일별로 2번째 최장체류지가 처음 기록된 시각(t1)과 마지막으로 기록된 시각(t2)를 식별한 후, t1부터 t2까지 통행자가 2번째 최장체류지에 체류한 것으로 보정(단, t1부터 t2 사이에 1번째 최장체류지로 지정된 체류지가 존재하는 경우, 2번째 최장체류지

가 없는 것으로 판단)



<그림 4> 최장체류지 식별 후, 체류시간 aggregation 예시

- 2개의 최장 체류지에 대해서 일주일 중 발생 빈도를 계산하여, 거주지 및 근무지 식별
 - 발생 빈도가 더 많은 것을 거주지, 적은 것을 근무지로 식별하며 발생 빈도가 같은 경우, 일주일 기준 체류시간이 더 긴 것을 거주지, 적은 것을 근무지로 식별
- 근무지로 식별된 체류지에 대해 일주일 중 근무지에 체류한 시간대를 계산하여 동일 시간대에 주 2회, 회당 3시간 미만 체류한 경우는 근무지에서 제외
- 거주지와 근무지 위치는 각 주체류지로 구분된 최장체류지의 위치로 지정
- 원시데이터에 대해 신규 알고리즘을 적용하면 <그림 5>와 같은 형태의 통행지표가 산출되며, 해당 산출물에는 개별 고객에 대한 일자별 통행지표가 포함되어 있음

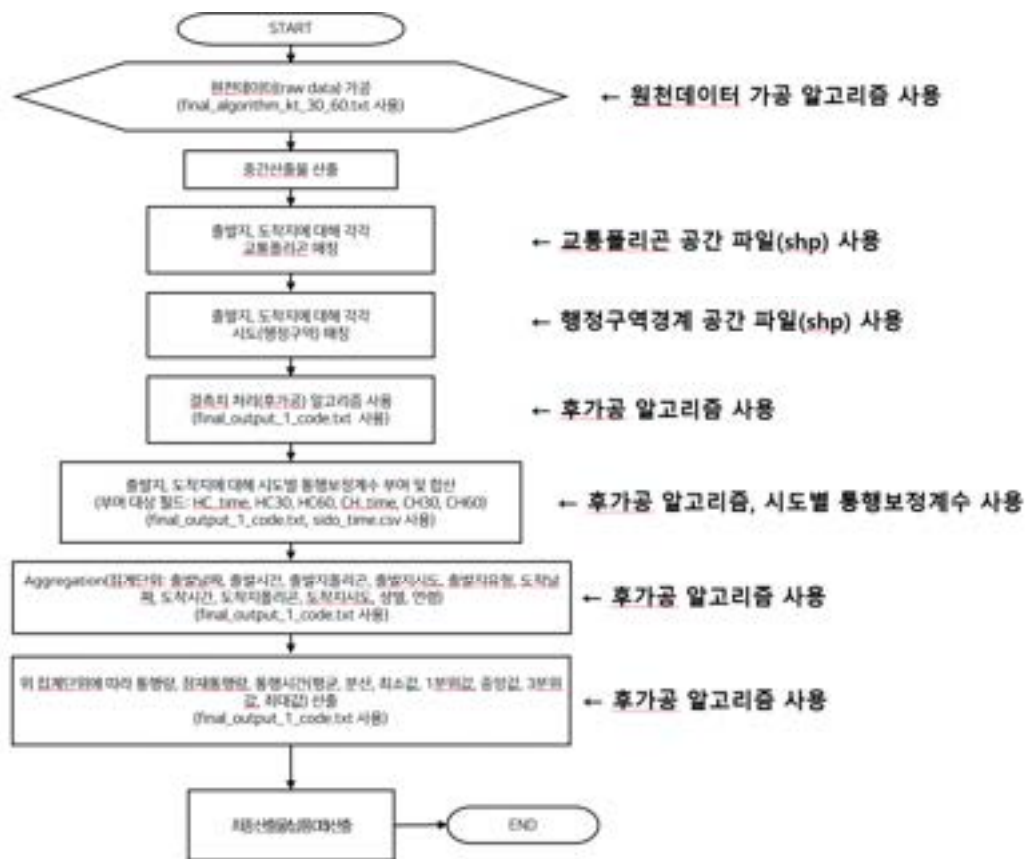
날짜	2019-04-01
고객식별번호	112233
나이	30분
성별	M
출근시작시간	8:30
출근종료시간	9:00
출근소요시간	30분
출근 중 잠재체류 횟수	0회
출근 중 잠재체류 시간	0분
퇴근시작시간	18:27
퇴근종료시간	19:30
퇴근소요시간	63분
퇴근 중 잠재체류 횟수	1
퇴근 중 잠재체류 시간	35
집 위치	127.491, 36.455
회사 위치	127.485, 36.607
집-회사 간 거리	16.9 km
출근시간	30분
퇴근시간	X
출근속도	33.8 km/h
퇴근속도	X

<그림 5> 통행지표 산출물 예시

3. 모바일 교통 빅데이터 기반 통행사슬(Trip Chain) DB 구축 및 검증

가. DB 설계

- 「개인정보보호법」과 「위치정보의 보호 및 이용 등에 관한 법률」에 따라 개인의 위치와 이동궤적을 추적할 수 없도록 기반 DB는 다음 <표 1>과 같이 집계된 형태로 구축하고자 함



<그림 6> 모바일 통신 빅데이터 가공 DB 구축 과정

<표 1> 모바일 기반 DB 형태

출발						도착				
날짜	요일	시간	폴리곤	시도	체류지 유형	날짜	시간	폴리곤	시도	체류지 유형
20180402	월	13	48521351	11	집	20180402	14	65421584	11	회사
20180403	화	06	54754213	21	집	20180403	08	32158421	11	회사
20180404	수	09	15486211	33	집	20180404	09	15486213	33	학교

총통행량	보정계수	성별	연령	잠재통행 고려 안함								
				통행량(비율)				시간			거리 평균	속도 평균
				통행 비율	30분 이상	60분 이상	120분 이상	평균	분산	중앙 값		
10	0.34	M	20	1	0.7	0.5	0.4	35	6.98	26	8.24	29.8
25	0.57	F	50	0.8	0.9	0.8	0.2	33	6.77	24	9.45	11.9
8	0.47	M	10	1	0.9	0.6	0.1	25	4.32	19	10.5	23.5

잠재통행 30분 고려 안함										
통행량(비율)					시간			거리 평균	속도 평균	
통행비율	잠재체류 횟수	30분 이상	60분 이상	120분 이상	평균	분산	중앙값			
0.8	0	0.8	0.6	0.8	35	6.98	26	5.89	36.9	
0.7	3	0.7	0.5	0.2	-	-	-	9.57	10.4	
0.8	2	0.8	0.5	0.1	-	-	-	10.12	20.3	

잠재통행 60분 고려 안함										
통행량(비율)					시간			거리 평균	속도 평균	
통행비율	잠재체류 횟수	30분 이상	60분 이상	120분 이상	평균	분산	중앙값			
0.9	0	0.7	0.6	0.5	35	6.98	26		33.6	
0.8	2	0.8	0.6	0.2	-	-	-		10.9	
0.9	0	0.9	0.6	0.1	25	4.32	19		22.8	

주 *) 요일: 출발 및 도착 요일 (1~7, 일~월)

주 **) 유형: 개인별 출발지 및 도착지 유형 (H: 집, C: 회사, S: 학교, R: 정기 잠재체류지, X: 기타 잠재체류지)

- 개인의 통행 정보가 노출되지 않도록 기점 정보(출발 일자, 출발 시간대, 출발 위치, 출발 체류지 유형), 종점 정보(도착 일자, 도착 시간대, 도착 위치, 도착 체류지 유형), 통행자 정보(성, 연령)를 기준으로 통행 정보가 같은 인구를 집계
- 일자 정보는 1일 단위로 구축하도록 함

- 출발시간 및 도착시간 정보는 집계된 인구가 3명 이하가 될 가능성을 줄이면서, DB 사용자에게 가능한 섬세한 통행 정보를 제공할 수 있도록 한 시간 단위로 입력하도록 함
- 출발지와 도착지 위치는 「위치정보의 보호 및 이용 등에 관한 법률」에 저촉되지 않도록 기지국 좌표 그대로 노출하지 않고 매칭되는 교통폴리곤(교통폴리곤은 기지국의 수신범위를 고려하여 구축한 모바일 기반 DB 분석 단위) ID로 변환하도록 함
- 출발 및 도착 체류지 유형은 평소 통행자가 해당 위치에 얼마나 자주, 오래 머무르는지를 추정할 수 있는 정보를 제공하기 위한 것으로, 「2019년 국가교통조사」에서 설정한 체류지 식별 기준에 따라 출발지(출발 기지국 좌표)와 도착지(도착 기지국 좌표)의 특성을 확인한 후 각각 체류지 유형을 구분하여 입력하도록 함
- 통행자에 대한 정보를 확인할 수 있도록 성, 연령 필드 구성하고, 통행 정보가 지나치게 세분되지 않도록 가능한 통행 특성이 유사한 계층을 묶어 10세 단위로 연령을 구분하여 입력하도록 함
- 통행량 필드에는 전술한 기준에 의해 집계된 인구(단말기 수)를 입력하되, 집계된 통행량이 3 이하의 값을 갖는 경우, 2장에서 개발한 3통행 미만 보정계수를 적용하여 추정된 통행량을 입력하도록 함
- 새로운 지표 개발을 위해 통행시간 필드를 추가
- 통행시간 필드에는 집계된 통행 정보를 기준으로 산출한 평균통행시간 값을 입력
- 5가지 유형으로 구분된 기종점 기준, 통행 도중 인식한 기지국에서의 체류시간에 대해 고려 안함, 60분으로 나누어 통행량 및 통행시간을 산출함

<표 2> 모바일 통신 빅데이터 DB 테이블 정의서

컬럼 ID	컬럼명	Type	비고
o_date	출발 일자	string	yyyymmdd
o_dow	출발 요일	string	0 ~ 6 : 일요일 ~ 토요일
o_time	출발 시간대	string	00~23 (1시간 단위)
o_polycode	출발 폴리곤ID	string	500미터 격자단위
o_type	출발 체류지 유형	string	H: 집, C: 회사, S: 학교, R: 정기 잠재체류지, X: 기타 잠재체류지
d_date	도착 일자	string	yyyymmdd
d_time	도착 시간대	string	00~23 (1시간 단위)
d_polycode	도착 폴리곤ID	string	500미터 격자단위
d_type	도착 체류지 유형	string	H: 집, C: 회사, S: 학교, R: 정기 잠재체류지, X: 기타 잠재체류지
age	연령	integer	0~110 (예: 10: 10세 이상 20세 미만)
gender	성	string	M: 남성, F: 여성
traffic_all	총 통행량	integer	횟수 단위
traffic_time_all_mean	통행시간_평균	integer	1분 단위
traffic_distance_all_mean	통행거리_평균	integer	1분 단위
traffic_speed_all_mean	통행속도_평균	integer	km/h 단위
traffic_gb	교통수단	string	E : 기타, F: 항공, T: 고속철도, B: 고속버스

나. DB 구축 방법

- 원시데이터 수집 및 이상치 제거
 - 2020년 1월 1일부터 2020년 12월 31일까지 기록된 로그를 수집
 - 기지국 좌표가 Null값이거나 행정동 정보가 매칭되지 않는 데이터를 제거
- 선분 이력으로 변환
 - 포인트 단위의(밀리세컨 단위) 로그를 선분 단위의(분 단위) 로그로 변환
 - 닷(dot) 형태로 기록된 로그를 기록된 시간 순서대로 연결한 후, 최초 기록된 시간(이하 ‘체류시작시간’)과 마지막으로 기록된 시간(이하 ‘체류종료시간’) 정보만 추출하여 단말기가 해당 기지국에 식별된 시간(이하 ‘체류시간’)을 산출
 - 체류시간은 체류종료시간에서 체류시작시간을 뺀 값이며, 초(sec) 단위는 생략
- 통행 정보 왜곡 데이터 보완
 - 데이터 보완 범위를 한정하기 위해 로그 기록을 한 달 단위로 구분
 - 통행 정보 왜곡 데이터를 탐색하고 보완하는 범위를 1개월로 한정 (예: 4월 1일부터 4월 30일까지의 로그 기록을 연결하여 탐색하고 보완)
 - 통행 정보를 왜곡할 수 있는 ‘필요 이상의 핸드오버 데이터(Unnecessary Handover)’, ‘핑퐁 핸드오버(Ping-pong Handover)로 인한 데이터’를 본 과업에서 개발한 전처리 기술을 통해 보정함
- 체류 순서 부여
 - 로그 기록 일자, 체류시작시간, 체류종료시간을 기준으로 로그 기록 순서에 따라 개인별 데이터를 정렬하여 체류 순서를 구분
 - 고객식별번호(단말기 구분코드)와 데이터 생성일자를 기준으로 KEY 값을 갖는 필드를 형성한 후(‘기준일-고객식별번호’), 같은 KEY 값을 갖는 데이터 내에서 정렬
- 체류 정보 추출
 - 이동 중에 기록된 로그 기록을 제외하고 체류 중에 기록된 로그 기록만을 추출
 - 체류시간이 15분을 초과하는 로그 기록만 추출
 - 트립타임 필드를 추가하여 체류시간이 15분 이하인 경우는 ‘경로(Pass-by)’로, 15분 초과인 경우는 ‘체류(Stay)’로 구분
- 체류지 유형 구분
 - 체류지 식별 기준(〈표 3〉 참조)에 따라 체류지 유형을 구분하여 필드 추가

<표 3> 체류지 식별 기준

체류지 유형		식별 기준			
		체류특성		통행자 연령	비고
		체류 시간	체류 빈도		
주 체류지	집	알고리즘 기준 최장 체류지 후보 1, 2 중	일주일 기준 최다빈도 체류지	-	알고리즘 기준 제1의 주체류지
	회사	동시간대 3시간 이상	주 2회 이상	30세 이상 60세 미만	알고리즘 기준 제2의 주체류지 중
	학교	동시간대 3시간 이상	주 2회 이상	20세 미만	알고리즘 기준 제2의 주체류지 중
기타 체류지	정기적	3시간 이상	주 2회 이상	-	알고리즘 기준 제 2의 주체류지 중 회사나 학교로 구분되지 않은 체류지 중
	비정기적	3시간 이상	주 2회 미만	-	

○ 위치정보, 시간정보, 연령정보 변환

- 개인의 이동궤적을 추적할 수 없도록 원시데이터에서 출발지와 도착지의 위치 정보를 분석맵의 교통폴리곤 ID로 변경한 후, 시간정보(체류시작시간, 체류종료시간)에서 분 정보를 생략한 시간대 필드를 생성하고, 1세 단위의 연령정보를 10세 단위로 변경함

○ 통행량 집계

- 앞서 부여한 체류순서에 따라 출발과 도착을 구분한 다음, 기점 정보(출발 일자, 출발 시간대, 출발 위치, 출발 체류지 유형), 종점 정보(도착 일자, 도착 시간대, 도착 위치, 도착 체류지 유형), 통행자 정보(성, 연령)가 동일한 로그 기록을 집계함
- 마지막 체류지가 출발 정보로 구분되고, 도착 정보가 없는 경우 이동하지 않은 것으로 간주함

로 보고 ‘무통행’으로 간주함

○ 통행량 보정

- 통행량이 3 이하(0~3통행)인 경우 모두 3통행으로 변환한 후, 3통행 미만 보정계수 (2장 참고)를 적용하여 통행량을 보정
- 개인정보보호법에 의거, 3 통행 이하의 값은 실제 값이 아닌, 추정값을 입력
- 실제 통행 분석에 활용할 때에는 통신업체의 시장점유율을 기준으로 전수화 하여 사용
- 과학기술정보통신부에 고시되고 있는 무선통신서비스 가입자 통계 정보 중 LTE 가입 정보를 활용하여 KT 시장점유율 산출 가능

○ 평균통행시간 산출

- 모바일 통신 데이터는 기지국 기반 로그 기록이므로 통행시간 산출 시, 기종점 부분에서 통행시간이 하향 계산되는 경우가 발생함
- 이와 같은 문제를 해결하고자, 기지국의 평균 수신범위를 계산하여 시도별 기지국 내 이동 시간 산출 후, 통행시간을 보정함
- 각 개인의 통행 정보에서 통행시간(도착시간-출발시간, 분 단위)을 산출한 다음, 기종점 정보(기지국 위·경도 좌표)와 시도(행정구역)정보를 매칭해 기종점 지점에서 시도별 기지국 내 이동 평균 시간을 합산하여 최종 통행시간 계산함
- 개별통행에 대해 기점 정보(출발 일자, 출발 시간대, 출발 위치, 출발 체류지 유형), 종점 정보(도착 일자, 도착 시간대, 도착 위치, 도착 체류지 유형), 통행자 정보(성, 연령)가 동일한 로그 기록의 통행시간을 집계한 후 집계한 통행시간을 산술평균하여 산출함

다. DB 검증 방법

- 모바일 통신 데이터를 이용하여 구축된 통행DB의 신뢰성 및 정확성을 검증하기 위하여 기존 개인통행실태조사자료(2016년 기준, 한국교통연구원)와 인구총조사자료(2015년 기준, 통계청)를 활용
- 모바일 통신 데이터 기반 통행DB와 기존 개인통행실태조사자료와 인구총조사자료의 통행량, 통행시간 등을 비교·분석하고, 단순 비교할 수 없는 모바일 통신 데이터 기반 상세지표(예를 들어, 상세 시간대별, 요일별, 성별, 연령별, 희망통근시간 등)는 통상적인 기대 통행패턴을 벗어나는지 확인
 - 지역별 통행 발생량·도착량을 개인통행실태조사자료와 상대적으로 비교하여, 유사한 패턴을 보이는지 확인
 - 시간대별, 요일별 통행 발생량·도착량을 확인하여, 통상적인 첨두시간 및 평일·주말 통행패턴이 나타나는지 확인
 - 성별, 연령별 통행 발생량·도착량을 확인하여, 통상적인 인구비율 및 연령대별 통행패턴이 나타나는지 확인
 - 지역별 출퇴근 목적 통행량·통행시간을 개인통행실태조사자료, 인구총조사자료와 상대적으로 비교하여, 유사한 패턴을 보이는지 확인
 - 출퇴근 목적 통행시간·통행속도와 희망통근시간(60분)기준 통행패턴을 확인하여, 통상적인 통행패턴이 나타나는지 확인
- 단, 데이터 가공 시간 때문에 1년 치 모든 데이터를 가공하여 과업 기간 내에 검증할 수 없으므로, 통계적으로 충분히 유의한 10만 명의 표본 자료를 이용하여 검증을 시행함
 - 표본 유의성을 분석한 결과, 약 2만 명의 표본 이후 통행시간의 평균 오차는 ± 0.1 분 이하임

라. DB 검증 결과

○ 통행 발생·도착패턴 비교 및 분석 결과

<표 4> 지역별 통행량 비교 결과

(단위: 통행/일 (비율%))

구분	개인통행실태조사자료 기준*		모바일 통신 빅데이터 기준	
	발생 (비율)	도착 (비율)	발생 (비율)	도착 (비율)
서울	17,996,813(20.7)	17,971,878(20.7)	3,532,837(17.5)	3,735,390(18.5)
부산	6,457,538(7.4)	6,458,293(7.4)	1,250,226(6.2)	1,241,939(6.1)
대구	4,333,058(5.0)	4,330,603(5.0)	925,685(4.6)	901,059(4.5)
인천	4,370,318(5.0)	4,433,014(5.1)	930,863(4.6)	887,980(4.4)
광주	3,117,712(3.6)	2,942,575(3.4)	560,475(2.8)	542,603(2.7)
대전	3,091,862(3.6)	3,101,487(3.6)	649,274(3.2)	640,985(3.2)
울산	2,117,733(2.4)	2,109,519(2.4)	460,616(2.3)	461,852(2.3)
세종	389,864(0.4)	391,571(0.5)	106,505(0.5)	100,256(0.5)
경기	18,668,530(21.5)	18,628,440(21.4)	5,303,762(26.3)	5,132,439(25.4)
강원	2,714,015(3.1)	2,715,277(3.1)	733,552(3.6)	737,190(3.6)
충북	2,803,473(3.2)	2,793,476(3.2)	658,793(3.3)	664,843(3.3)
충남	3,871,861(4.5)	3,863,535(4.4)	844,207(4.2)	859,032(4.3)
전북	2,942,438(3.4)	2,943,452(3.4)	713,754(3.5)	715,518(3.5)
전남	2,606,282(3.0)	2,783,178(3.2)	709,859(3.5)	727,290(3.6)
경북	4,431,201(5.1)	4,442,218(5.1)	1,080,468(5.3)	1,108,501(5.5)
경남	5,297,159(6.1)	5,300,768(6.1)	1,300,359(6.4)	1,304,949(6.5)
제주	1,659,446(1.9)	1,660,019(1.9)	443,411(2.2)	442,822(2.2)
계	86,869,303(100)	86,869,303(100)	20,204,646(100)	20,204,646(100)

- 개인통행실태조사자료 기준 1인당 일평균 통행량(발생+도착)은 약 3.4회이며, 모바일 통신데이터 기준 1인당 일평균 통행량(발생+도착)은 약 3.1회 임

- 개인통행실태조사자료 기준 전국 통행 발생·도착량은 1.74통행/일, 1.74통행/일이며, 모바일 통신데이터 기준 전국 통행 발생·도착량은 1.5통행/일, 1.6통행/일임
- 모바일 통신데이터 기준 통행량과 개인통행실태조사자료 기준 통행량 매우 유사하게 나타남
- 하지만 개인통행실태조사자료가 더 높은 통행량을 띄는데, 이러한 차이는 기존의 개인통행실태조사는 출근, 등교, 업무, 쇼핑, 귀가 등 모든 통행목적에 따라 통행량을 집계하지만, 모바일 통신데이터는 체류시간 기준으로 체류지를 구분하기 때문에 기 타잠재체류지에 속하지 않은 이동이 집계되지 않기 때문에 통행량 차이가 과소 추정되었다고 판단됨
- 지역별 통행량 비율을 살펴보면, 평균적으로는 약 0.41%의 차이로 기존 개인통행실태조사자료와 매우 유사한 패턴을 나타냄
- 경기지역 통행 발생량이 약 +2%로 가장 큰 폭으로 과대 추정되었으며, 부산지역 통행 발생량이 약 -1.1%로 가장 큰 폭으로 과소 추정되었으나, 평균적으로 약 $\pm 1\%$ 로 매우 신뢰성 높은 결과를 도출함

○ 목적별(출퇴근) 통행패턴 비교 및 분석 결과

- 지역별 출퇴근 목적 통행 발생량·도착량을 개인통행실태조사자료 · 인구총조사자료와 상대적으로 비교하여, 유사한 패턴을 보이는지 확인함

<표 5> 지역별 출·퇴근 통행량 비교 결과

(단위: 통행/일(비율%))

구분	개인통행실태조사자료 기준		인구총조사자료 기준	모바일 통신 빅데이터 기준	
	출근 (비율)	귀가*	통근**	출근	퇴근
서울	4,516,466(20.7)	8,413,729(22.1)	4,597,322(19.6)	631,221(20.0)	536,024(19.2)
부산	1,492,300(6.8)	3,051,128(8.0)	1,500,101(6.4)	205,559(6.5)	180,838(6.5)
대구	974,423(4.5)	1,801,446(4.7)	1,064,227(4.5)	152,313(4.8)	134,449(4.8)
인천	1,213,843(5.6)	1,794,429(4.7)	1,344,311(5.7)	171,099(5.4)	150,288(5.4)
광주	838,601(3.8)	1,279,355(3.4)	639,412(2.7)	87,253(2.8)	77,097(2.8)
대전	736,843(3.4)	1,313,431(3.4)	672,856(2.9)	97,005(3.1)	86,176(3.1)
울산	514,493(2.4)	995,135(2.6)	538,654(2.3)	71,522(2.3)	62,921(2.3)
세종	89,203(0.4)	149,020(0.4)	92,403(0.4)	17,648(0.6)	16,358(0.6)
경기	5,272,517(24.1)	7,799,702(20.5)	5,745,634(24.4)	852,163(27.1)	752,696(27.0)
강원	644,578(2.9)	1,186,771(3.1)	703,266(3.0)	89,019(2.8)	82,094(2.9)
충북	637,895(2.9)	1,182,200(3.1)	742,399(3.2)	93,268(3.0)	85,818(3.1)
충남	875,177(4.0)	1,616,017(4.2)	1,032,913(4.4)	113,204(3.6)	104,716(3.8)
전북	665,597(3.0)	1,263,474(3.3)	811,899(3.5)	93,693(3.0)	85,406(3.1)
전남	706,911(3.2)	1,241,157(3.3)	864,498(3.7)	87,501(2.8)	81,287(2.9)
경북	1,007,670(4.6)	2,045,073(5.4)	1,280,135(5.4)	143,352(4.6)	133,229(4.8)
경남	1,372,291(6.3)	2,361,671(6.2)	1,560,582(6.6)	191,226(6.1)	174,882(6.3)
제주	291,636(1.3)	581,151(1.5)	309,133(1.3)	52,204(1.7)	47,223(1.7)
계	21,850,443(100)	38,074,889(100)	23,499,745(100)	3,149,247(100)	2,791,504(100)

주: *개인통행실태조사자료는 2016년 기준 출퇴근에 대한 목적통행량이며, 귀가통행 통행목적에 상관없이 집으로 향하는 모든 통행을 의미함

**인구총조사(통계청)자료는 2015년도 기준이며, 출근과 퇴근이 구분되어 있지 않음

- 지역별 출퇴근 목적 통행시간을 개인통행실태조사자료 · 인구총조사자료와 상대적으로 비교하여, 유사한 패턴을 보이는지 확인함

<표 6> 지역별 출퇴근 통행시간 비교 결과

(단위: 분)

구분	개인통행실태조사자료 기준		인구총조사 자료 기준	모바일 통신 빅데이터 기준	
	출근	귀가*	통근**	출근	퇴근
서울	34.8	30.1	45.7	35.8	38.8
부산	30.1	25.6	37.0	30.8	33.0
대구	29.5	28.2	33.4	28.1	30.2
인천	36.6	32.5	43.6	35.9	37.4
광주	20.8	18.3	30.2	25.6	27.5
대전	23.3	21.7	32.3	26.1	27.8
울산	25.6	24.5	28.6	27.1	28.6
세종	15.6	15	29.1	33.0	34.3
경기	41.7	36.3	41.5	36.2	37.6
강원	17.6	20.3	22.1	23.4	24.8
충북	17.9	17.8	25.6	26.1	27.8
충남	15.6	16.7	22.9	27.2	28.7
전북	16	15.4	23.2	25.2	26.9
전남	13.3	13.6	21.3	25.4	27.0
경북	18.8	20	22.1	25.2	26.8
경남	22.5	21.1	25.6	27.0	28.4
제주	20.8	21.5	23.6	25.8	27.5
계	23.6	22.3	35.2	28.5	30.2

주 : *개인통행실태조사자료는 2016년 기준 출근·귀가 통행에 대한 일평균시간이며, 귀가통행 통행목적에 상관없이 집으로 향하는 모든 통행을 의미함

**인구총조사자료는 2015년 기준 통근 통행에 대한 소요 시간별 자료를 일평균시간 형태로 재구성하였으며, 출근과 퇴근이 구분되어 있지 않음

4. 모바일 교통 빅데이터 기반 교통지표 개발 및 온라인 서비스 제공

가. 개요

- 2018년부터 교통 모니터링·데이터 제공·분석 플랫폼(View-T)를 통해 모바일 통신 빅데이터를 기반으로 하는 통행 분석 서비스를 제공해왔음
- 해당 연도에는 2020년 기준 모바일 통신 빅데이터 기반 통행 DB를 기반으로 통행 분석 서비스를 제공하고, 사용자가 시스템에 구축된 기반 데이터를 이종 데이터와 연계하여 분석하거나 다른 방식으로 결과를 표출하는 데 활용할 수 있도록 데이터 다운로드 서비스 제공함

나. 기초 DB 처리 및 재구조화

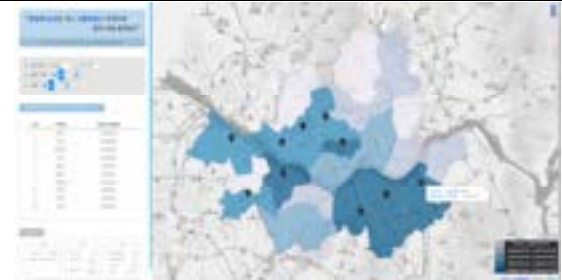
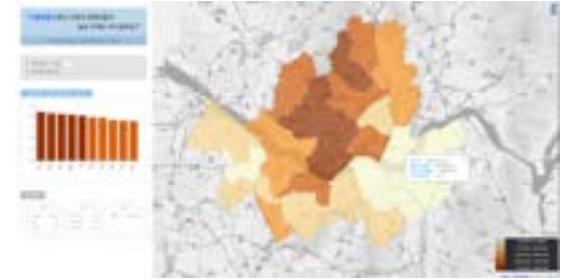


- 수급한 데이터의 분석과 누락된 데이터 또는 이상치데이터 여부에 대한 검증으로 데이터 파악
- View-T 서비스는 분석지역과 표출 단위에 대한 선택이 필수적으로 GIS 분석 최소 단위인 교통폴리곤과 교통폴리곤이 속하는 행정구역의 매칭 정보가 필요함
- 검색 쿼리와 DB 적재 용량 부분을 고려하여 문자형 데이터를 숫자 형식으로 변환 작업이 필요함
- 과대하게 집계된 3 통행에 대해서 보정계수를 적용하여 1 통행과 2 통행 데이터를 생성
- 분석 서비스에 활용하기 위해 출발 교통폴리곤과 도착 교통폴리곤 간의 직선거리를 생성

다. View-T Light

1) 주요통행분석

- 주요 특성 통행지역 분석
 - View-T Light는 총 24개의 기능으로 구분되어 있으며, 기종점 통행 분석, 교통 혼잡 분석 등 차량과 통신 데이터에 관련된 주제별 분석을 수행할 수 있도록 개선

<표 7> 주요 특성 통행지역 분석 그림

주요 특성 통행 분석	핫플레이스 분석		전국에서 선택한 지역으로 오는 인구의 통행량을 분석
	고령자 통행 비중 분석		선택 지역에서 출발하는 고령자 통행비율을 분석
	고령자 주요 통행 구간 분석		선택한 지역 내에서 고령자의 주요 통행구간을 분석
	도심 공동화 심각도 분석		전체 유동인구 대비 상주인구를 분석하여 도심공동화가 심각한 지역을 분석

○ 주요 통행구간 분석

- 기존 지도에서 시군구 단위로 선택할 수 있던 지역 선택의 범위를 콤보박스를 이용하여 시도 단위까지 선택 가능하게 하고, 전국을 선택할 수 있도록 개선
- 기존에는 여성의 데이터로만 분석하였으나 분석 조건에 남성을 선택할 수 있는 UI를 추가하여 성별에 따라 분석할 수 있도록 개선

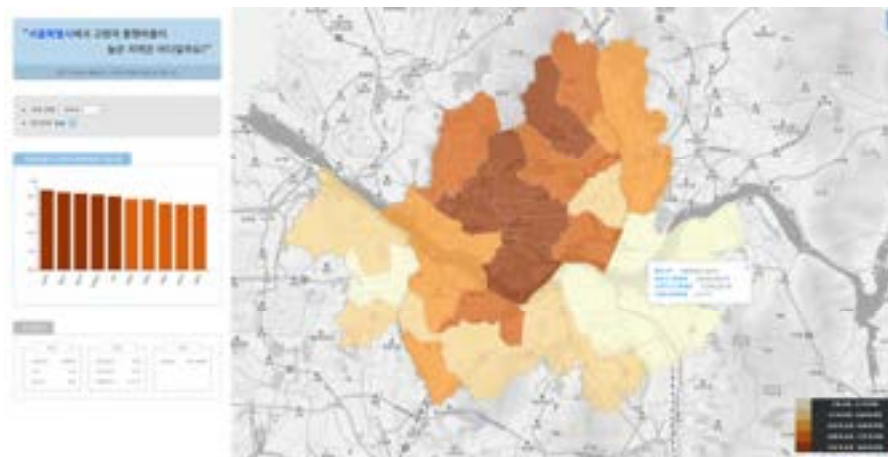


<그림 7> 기존점 인구 통행구간 분석의 시도 단위 전국 분석

2) 고령자 통행 분석

○ 고령자 통행 비중 분석

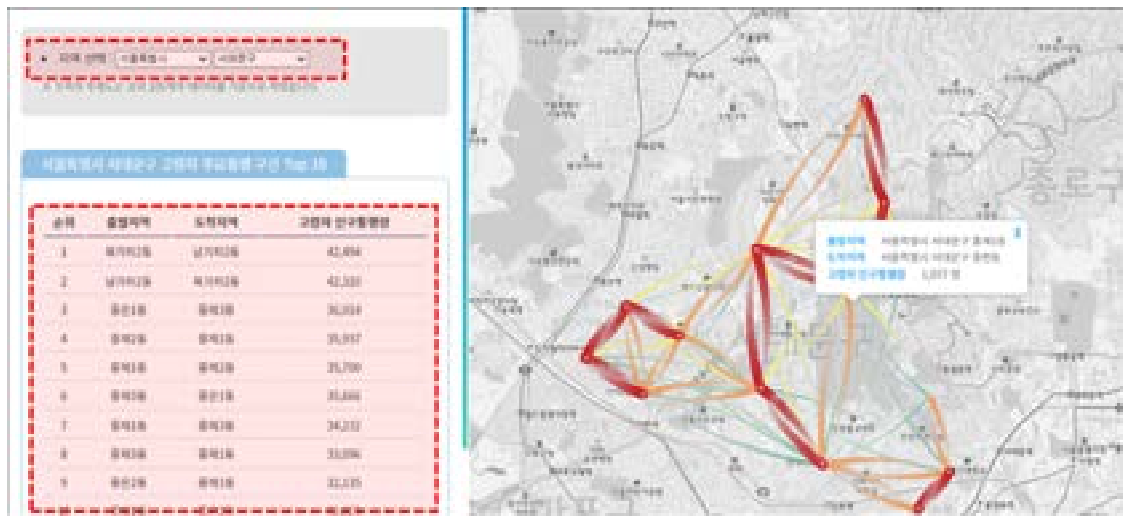
- 기존 지도를 클릭하여 시도 단위로 분석지역을 선택할 수 있던 지역 범위를 시군구 단위까지 선택 가능하게 하고, 읍면동 단위까지의 분석 결과를 표출할 수 있도록 개선



<그림 8> 고령자 통행 비중 분석의 읍면동 단위 분석 결과 표출

○ 고령자 주요 통행 구간 분석

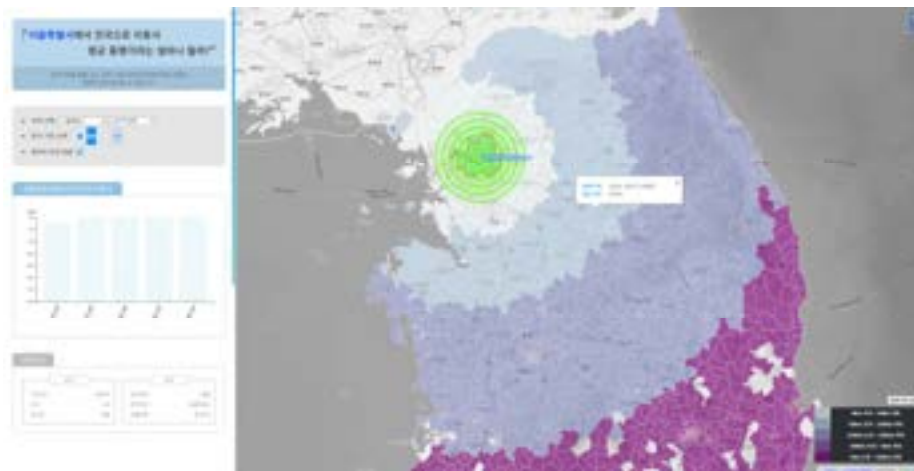
- 기존 지도를 클릭하여 시도 단위로 분석지역을 선택할 수 있던 지역 범위를 시군구 단위까지 선택 가능하게 하고, 읍면동 단위까지의 분석 결과를 표출할 수 있도록 개선



<그림 9> 고령자 주요 통행 구간 분석의 읍면동 단위 분석 결과 표출

3) 출·퇴근 통행 분석

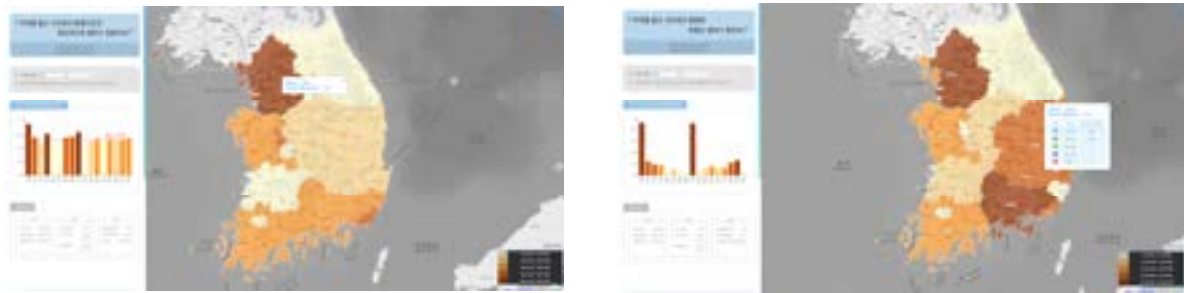
○ 통행시간/거리 분석



<그림 10> 통행시간·거리 분석 도구 화면예시

○ 출·퇴근 시간대 접근성 분석

- 기존 지도를 클릭하여 시도 단위로 선택할 수 있던 기존 지역을 시군구 단위까지 선택 가능하게 개선
- 분석 결과 그래프에 값에 따라 범례의 색상을 적용



<그림 11> 출퇴근 통행 접근성 분석의 기준 지역 선택 UI와 범례 색상 적용된 결과 그래프

4) 관광·연휴 분석

○ 계절별 핫플레이스 분석

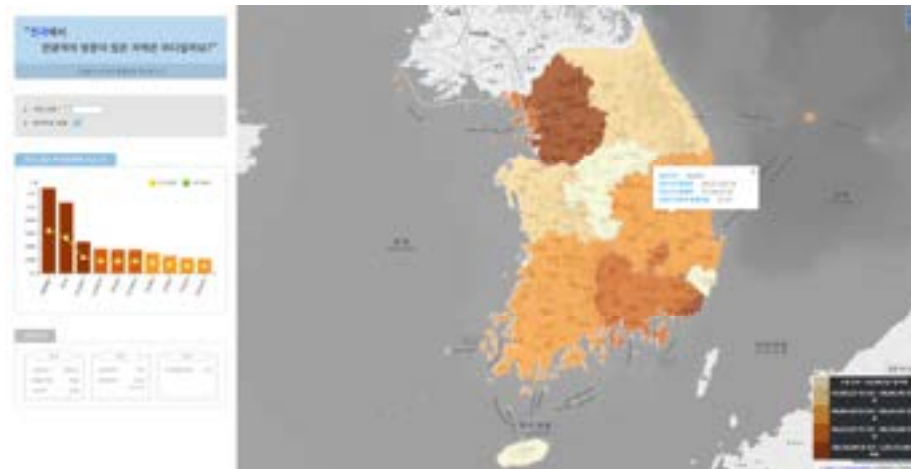
- 계절에 따른 지역별 유동인구 통행량을 분석
- 관광목적의 계절별 통행량과 연령대별 통행량을 분석



<그림 12> 계절별 핫플레이스 분석 결과 (좌)유동인구 통행량 (우)관광목적·연령별 통행량

○ 관광지 내·외부 통행 비율 분석

- 분석 결과 그래프에 표시되는 지역을 순위가 높은 1~10까지로 요약하여 표시하고 각 지역별 인구통행량의 남/여 성별을 구분하여 결과 표출



<그림 13> 관광지 내·외부 통행 비율 분석 결과

5) 경제·사회활동 분석

○ 도심 공동화 심각도 분석

- 분석 결과가 색상 단계로 구분되는 표출 방식에서 결과의 관계 파악 및 확실한 의미 전달을 할 수 있도록 여러 가지 색상을 조합한 색상 스키마를 사용하여 표현 방식을 변경할 수 있는 기능 추가



<그림 14> (좌) 특별시·광역시를 대상 인접도시 출근 비율 (우) 특별시·광역시 지역 내 통행량

라. View-T Expert

1) 사용자 편의성을 위한 UI/UX 개선

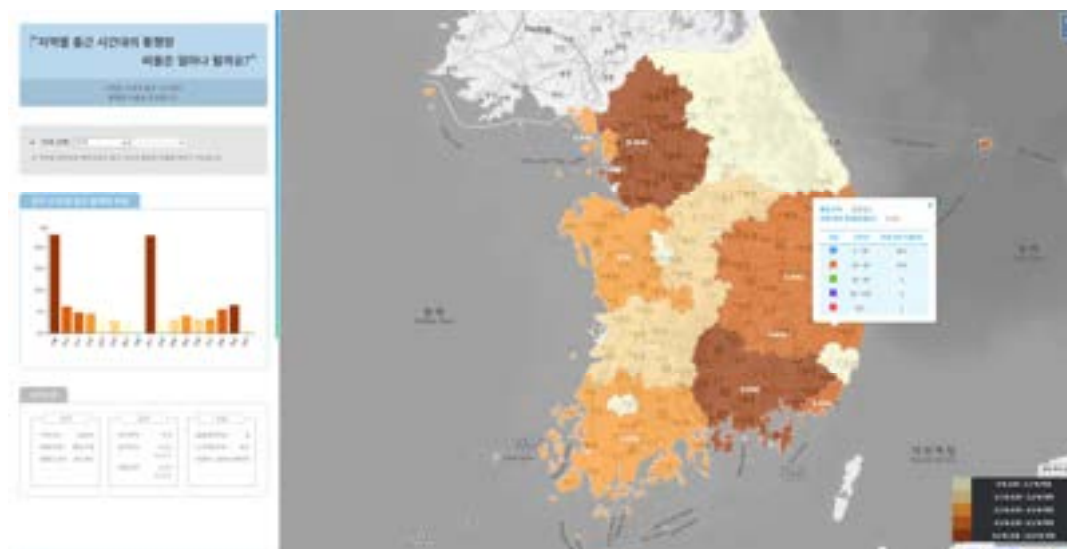


<그림 15> 통행지표 서비스 대시보드 화면 구성

- 사용자 편의성을 위한 다양한 대시보드 옵션 제공
 - 대시보드는 다양한 종류의 위젯을 제공하여 사용자에게 분석 결과를 동적으로 제공
 - 대시보드는 사용자가 위젯을 원하는 형태로 추가, 수정, 설정 등 기능을 제공
- 2) 기존 분석기능 (주요통행구간분석) 개선
 - View-T의 다양한 데이터를 한눈에 알아볼 수 있도록 종합적인 형태로 제공하는 대시보드 서비스 개선
 - 데이터를 기반으로 수집하고 관련 정보를 대시보드 형태로 시각화하여 분석 및 의사 결정을 지원하는 종합적인 지표 서비스 개발
 - 화면은 레이아웃 메뉴 영역, 옵션 설정 영역, 위젯 표출 영역, 위젯 목록 영역으로 구분하여 배치

<표 8> 주요통행구간 분석기능의 연도별 표출 형태 변경

구분	권한	변경 전	변경 후
분석지역 선택	일반사용자	최대 5개 (시군구, 읍면동)	변화 없음
	고급사용자	일반사용자 기능에 더하여 시도 단위는 1개 선택가능	전국 단위 분석 가능 (상세 조건 선택 제한)
표출단위 선택	일반사용자	분석지역 하위 단위 선택 (시군구 → 읍면동)	최소 표출단위 교통폴리곤 → 읍면동
	고급사용자		전국 단위 경우 시도, 시군구 단위 선택



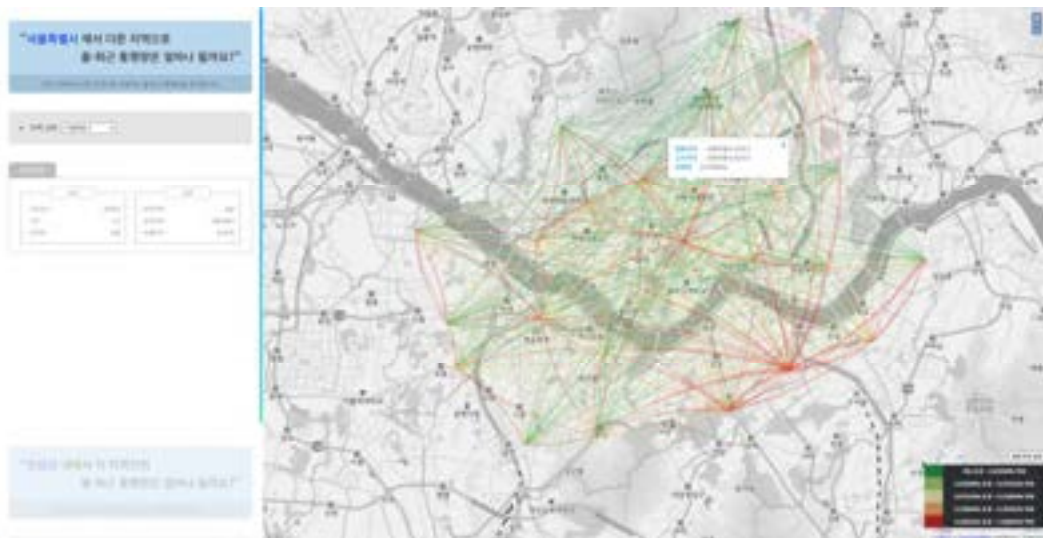
<그림 16> 주요통행구간 분석의 전국분석 결과 및 통행량 분포도 그래프

3) 신규 분석기능 개발

① 심층 분석을 가능케 하는 분석기능 개발

- 여러 분석기능을 선택하고 그에 따른 결과를 동시에 주제도로 표출하여 확인할 수 있는 기능
 - 핫플레이스, 주요통행구간, 주요통행지역, 내부통행지역, 통행거리분석을 연계분석 가능함
- 정보 연계 분석기능 UI

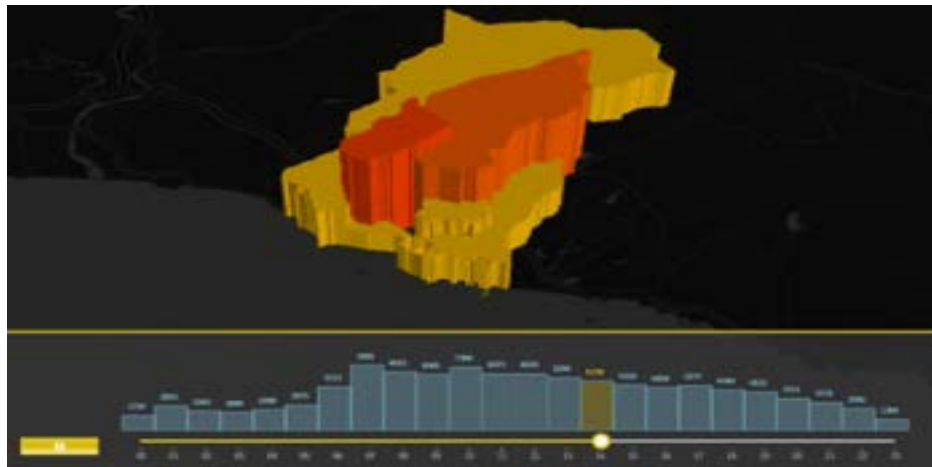
- 각 분석기능을 선택한 후 분석하기 버튼을 클릭하여 연계 분석을 실행함
- 주제도상 클릭하여 연계분석된 데이터를 표출함
- 하단 옵션 버튼을 통하여 연계분석 기능들의 레이어를 끄거나 켜서 개별 주제도를 볼 수 있음
- 희망선도를 통해서 주요 통행구간을 인지할 수 있음
- 희망선도 혹은 지역을 선택하여 해당 위치의 데이터 개괄을 볼 수 있음



<그림 17> 출/퇴근 통행 연결성 분석 도구 화면예시

② 통행 변화 시뮬레이션 기능

- 통행 변화를 시계열로 확인할 수 있는 기능
- 통행 변화 시뮬레이션 분석기능 UI
 - 분석유형(유입, 유출 및 O/D기준), 분석지역 및 분석 기간을 설정해 사용자 의도에 맞는 공간데이터를 시각화하여 통행 변화의 흐름 분석이 가능하도록 개발
 - 시뮬레이션은 시간대별, 요일별, 월별을 1년치 데이터를 기준으로 집계하여 통행의 흐름을 보여줄 수 있도록 개발
 - 예시) 시간대별 - 00시 ~ 24시, 요일별 - 월요일 ~ 일요일, 월별 - 1월 ~12월
 - 시뮬레이터를 통해 시간에 따른 데이터 변화를 제어할 수 있음
 - 재생, 정지, 재생트랙 제어 등



<그림 18> 유입 기준 폴리곤 형태의 통행 변화 시뮬레이션



<그림 19> 지역 내 기준 폴리곤 및 희망선 형태의 통행 변화 시뮬레이션

③ View-T Light의 분석 서비스 중 일부를 Expert에 개발

○ 관광 특화지역 분석기능

- 계절에 따라 통행량 변화가 큰 지역을 확인할 수 있는 기능
 - 도착지 트립타입이 비정기적 잠재체류지인 통행을 지역별, 계절별로 분류 표시
- 관광 특화지역 분석기능 UI
 - 분석지역 단위를 광역시/도, 시/군/구에서 선택
 - 분석하기 원하는 지역을 지도선택 혹은 행정구역 검색을 통하여 선택
 - 표출단위는 분석지역 단위의 하위 지역 중 선택
 - 상위 5개 지역을 주제도 상에 표시

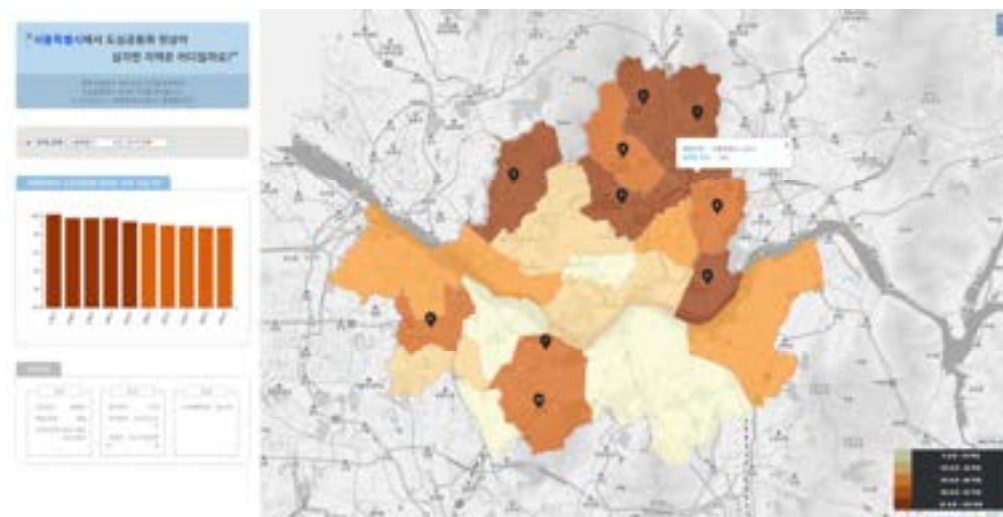
- 해당 지역의 계절별 시간대와 계절별 연령대 통행량의 변화를 순위 그래프로 표시



<그림 20> 관광 특화지역 분석기능 화면

○ 도심공동화 심각도 분석기능

- 도시 중심부의 상주인구가 줄어드는 현상을 확인할 수 있는 기능
 - 심야 시간대 집으로의 통행량과 낮 시간대 통행량의 비율을 통한 상주인구 측정
- 도심공동화 심각도 분석기능 UI
 - 분석지역 단위를 광역시/도, 시/군/구에서 선택
 - 분석하고자 하는 지역을 지도선택 혹은 행정구역 검색을 통하여 선택
 - 표출 단위는 분석지역 단위의 하위 지역으로 고정됨
 - 상위 5개 지역을 주제도 상에 표시
 - 해당 지역의 공동화 심각도를 순위 그래프로 표시



<그림 21> 도심공동화 심각도 분석화면

5. 결론 및 차년도 수행계획

가. 결론

- 본 과업에서는 기존 방법보다 정확하게 통행시간 및 통행량 등 통행지표를 산출하기 위하여 기존 모바일 가공 알고리즘을 보완함
 - 알고리즘은 합리적 수준의 사·공간적 데이터 군집 (Data aggregation and smoothing) 과정을 제시하여 신호 이상 현상을 효과적으로 처리하고, 개별통행 특성 및 패턴을 고려한 주요 통행목적지 추정을 통해 야간 및 비정기 근무 통행을 구분함
 - 또한, 기존 통신폴리곤 기준의 위치 추정 정보를 GPS와 같은 점 단위의 위치 추정 정보로 변환하는 방법을 제안하여 모바일 통신 데이터의 공간적 해상도를 개선하였으며, 이동 경로 및 수단 추정을 위한 개발 알고리즘 적용함
 - 격자형 폴리곤 데이터와 교통폴리곤 데이터 및 행정구역 데이터를 이용해 각각의 폴리곤 형상 레이어를 경량화하는 작업과 폴리곤 중심 위치 보정을 통해 포인트 레이어 구축 작업을 진행함
- 2022년 1월 1일부터 12월 31일까지 생성된 로그 기록에 모바일 통신 빅데이터 가공 알고리즘을 적용하여 기존 DB를 구축하고 검증함
 - 각 개인별 로그 기록은 「개인정보보호」, 「위치정보의 보호 및 이용 등에 관한 법률」에 저촉되지 않도록 집계한 형태의 DB를 설계하였으며, 이를 기준으로 DB를 구축함
 - 모바일 통신 빅데이터 기반으로 구축된 통행DB의 시간대별, 요일별, 성별, 연령별 통행량 및 통행시간을 기존 국가교통DB사업의 ‘개인통행실태조사자료 (2016년)’와 통계청의 ‘인구총조사자료 (2015년)’ 비교하여 검증함
- 모바일 통신 빅데이터 기반 교통지표를 개발하고, 분석 서비스의 성능 및 사용성을 개선하기 위하여 DB를 재구조화하여 서비스를 제공함
 - 또한, 2개 이상의 분석 수행 결과를 주제도로 동시에 확인할 수 있는 ‘정보 연계 분석기능’ 개발하고, 통행 변화를 시계열로 확인할 수 있는 ‘통행 변화 시뮬레이션 분석기능’ 개발함

나. 차년도 수행계획

- 모바일 교통 빅데이터 가공 알고리즘 고도화
 - 데이터 전처리 알고리즘 개선 및 고도화
 - 지역 특성(도시부/비도시부), 인적 특성(연령, 휴대폰 이용률 등)에 따른 로그 특성 분석 및 기 전처리 알고리즘 보완
 - 통행목적 구분 알고리즘 개선 및 고도화
 - 모바일 교통빅데이터 기반 추정 가능 체류지(거주지, 근무지, 여가 등) 선정
 - 거주지 및 근무지 추정 로직 개선
- 격자형 폴리곤 데이터와 교통폴리곤 데이터 및 행정구역 데이터를 이용해 각각의 폴리곤 형상 레이어를 경량화 및 알고리즘 고도화
- 모바일 교통 빅데이터 DB 구축 및 검증
 - 모바일 교통 빅데이터 DB 구축 (2023년도 전국)
 - 기 구축된 DB (2022년도 전국) 검증
 - 개인통행실태조사, 인구총조사 등 기존 조사기반 통계를 이용한 검증
- 모바일 교통빅데이터 기반 교통지표 개발 및 서비스
 - 교통지표 개발
 - 1차 지표 : 모바일 교통빅데이터 DB 기반 산출 가능한 통행지표 개발(통행량, 평균통행시간, 평균통행거리 등)
 - 2차 지표 : 이종 데이터와 모바일 교통빅데이터 DB를 연계·가공하여 산출 가능한 각종 지표 개발(교통안전지수, 도시공동화지수, 교통서비스지수, 경제활동지수 등)
 - 교통빅데이터 플랫폼 분석기능 개발 지원
 - 교통빅데이터 플랫폼을 위한 개별통행 DB 재구조화 지원
 - 교통지표 제공 및 분석 서비스 개발 지원

제1장 과업의 개요

제1장 과업의 개요

제1절 과업의 배경 및 목적

1. 과업의 배경

- 최근 교통 분야에서 각종 모빌리티 정보(예를 들어, 차량 GPS 데이터, 대중교통카드 데이터, DTG 데이터, 모바일 통신 빅데이터 등)를 활용하여 개별 통행 정보를 추출하고 각종 통행 관련 분석이 가능해지면서 다양한 방법으로 활용되고 있으나, 각 데이터의 특성에 따라 분석 대상과 범위가 제한적임
 - 차량에 탑재된 기기(내비게이션, DTG 등)나 도로에 설치된 기기(검지기, DSRC 등)를 통해 차량 통행자에 대한 통행 정보는 취득 가능하나, 자전거·도보 등으로 이동한 통행자에 대한 통행 정보는 표본율이 낮은 개인통행실태조사 결과에만 의존하고 있는 실정
 - 대중교통 이용자에 대한 통행은 교통카드 DB(대중교통 승하차 시 태그 정보)로 대부분 추정 가능하나, 잠재적으로 대중교통을 이용할 가능성이 있는 타 교통수단 이용자(차량, 도보 등)에 대한 통행은 일부만 추정 가능
- 그러므로, 전 국민의 95% 이상이 사용하고, 이용 교통수단에 구애받지 않고 구독이 가능한 모바일 통신 빅데이터를 이용한 개별통행 DB 구축 및 활용이 요구되고 있음
 - 모바일 기기와 기지국 간의 송·수신 이력인 모바일 통신 빅데이터는 대부분의 국민이 이용하고 있는 모바일 기기를 통해 축적되기 때문에 표본율이 매우 높고, 이동수단에 구애받지 않고(항공 제외), 기기의 전원을 끄지 않는 이상 개인의 이동궤적이 상세히 기록되기 때문
 - 1년 365일 수집되는 모바일 통신 빅데이터는 사회적 현안에 즉각적으로 대응할 수 있는 시의성 있는 모빌리티 정보를 제공할 수 있음
- 그러나, 기존 민간과 공공에서는 다양한 방식으로 모바일 통신 빅데이터를 가공하여 유동인구 형태로 활용하고 있지만, 개별통행 행태를 분석하기에는 한계점이 있으며, 이를 극복하기 위한 기술적 DB 구축 및 분석 방법이 요구되고 있음
 - 모바일 통신 데이터는 기존 설문조사 방식의 데이터 수집 방법과 비교하여 개인 및

통행에 대한 정보가 매우 제한적이며, 시간 단위 위치 정보를 이용하여 통행의 유무, 통행목적, 통행시간 등을 추정해야 함

- 또한, 기지국 위치로 대표되는 모바일 기지국 데이터의 공간적 저해상도 문제 때문에 정확한 통행 위치와 정밀한 통행시간을 추정하는 데 어려움이 있음

2. 과업의 목적

- 모바일 통신 빅데이터 중 기지국 기반의 데이터를 이용하여 시의성 있는 O/D 기반의 통행 정보(통행량, 통행시간) 생성 방법을 정의하고 DB 구축 및 제공
 - 기존 모바일 기지국 데이터의 공간적 저해상도를 해결할 수 있는 방안을 제시
 - 모바일 기지국 데이터를 이용한 각종 체류지(집, 회사 등) 추정 방법을 제시
 - 지역별 기지국 간격 등을 고려한 정밀한 통행시간 추정 방법 제시
 - 시의성 있는 전국 500×500m 격자 1시간 단위 연령별/성별 통행량 및 통행시간 제공
- 통행자별 이용 교통정보 정보가 포함된 모바일 통신 빅데이터를 이용하여 통행DB를 구축하고, 이를 기반으로 교통수요를 분석할 수 있는 지표, 기능 등을 개발하여 고도화된 교통 정책 수립 지원에 활용하고자 함

제2절 과업의 범위 및 내용

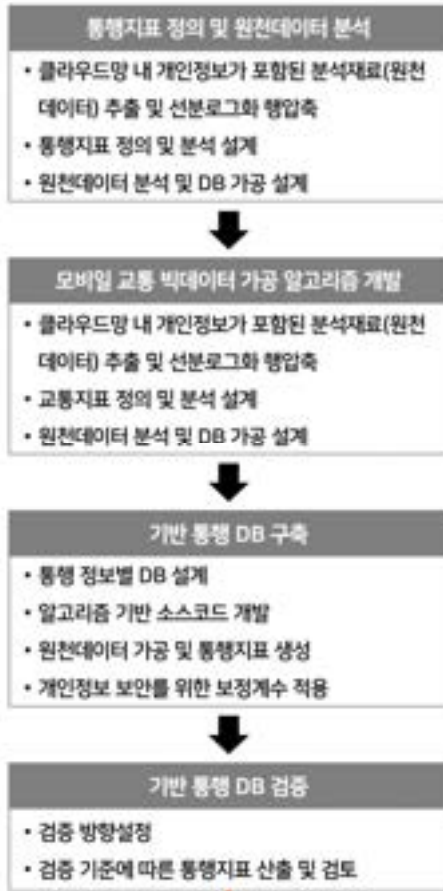
1. 과업의 범위

- 시간적 범위
 - 모바일 빅데이터 가공 알고리즘 개발: 2022년 원시데이터
 - 2022년도 기준 기반 DB 구축: 2022년 1월 1일 ~ 2022년 12월 31일 (1년)
 - 웹 서비스를 위한 각종 분석 도구 개발: 2020년 1월 1일 ~ 2020년 12월 31일
 - 2021년도 연구 성과물인 2020년도 기준 기반 DB 활용
- 공간적 범위: 전국

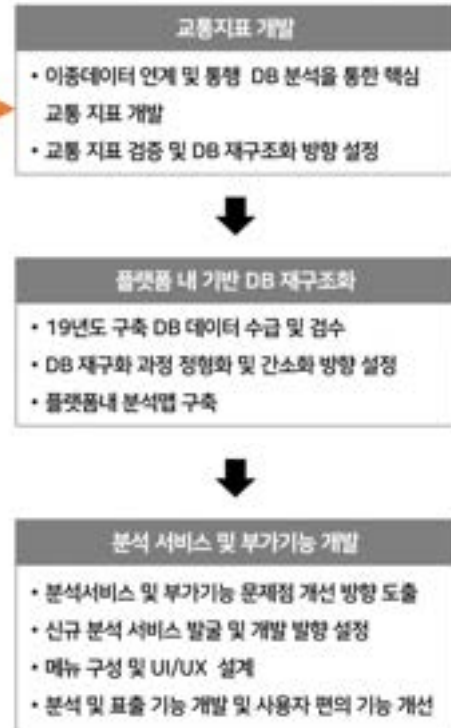
2. 과업의 내용

- 모바일 교통 빅데이터 전처리·가공 알고리즘 개발 및 고도화
 - 원시데이터 전처리·가공 알고리즘 고도화
 - 체류지 및 통행목적 추정 알고리즘 고도화
 - 교통수단 추정, 전수화, 가명화 알고리즘 개발(22년도 사업 중 신규 개발 항목)
 - 데이터 가공 프로세싱 최적화 및 속도 개선
- 모바일 교통 빅데이터 기반 통행사슬(Trip Chain) DB 구축 및 검증
 - 기준년도(22년도) 통행사슬(Trip Chain) DB 구축
 - 기준년도(22년도) 구축 DB 검증
- 모바일 교통 빅데이터 기반 교통지표 개발 및 온라인 서비스 제공
 - 기초 DB 처리 및 재구조화
 - 주요 교통지표 현행화 및 신규 지표 개발
 - 온라인 서비스(View-T) 제공

모바일 교통빅데이터 DB 구축 및 검증 부문



모바일 교통빅데이터 기반 교통지표 및 서비스 개발



<그림 1-1> 과업의 수행 절차

제2장 모바일 교통 빅데이터 전처리·가공 알고리즘 개발 및 고도화

제1절 원시데이터 전처리·가공 알고리즘 개발
및 고도화

제2절 체류지 및 통행목적 추정 알고리즘 고도화

제3절 교통수단 추정 및 전수화 방법론 개발

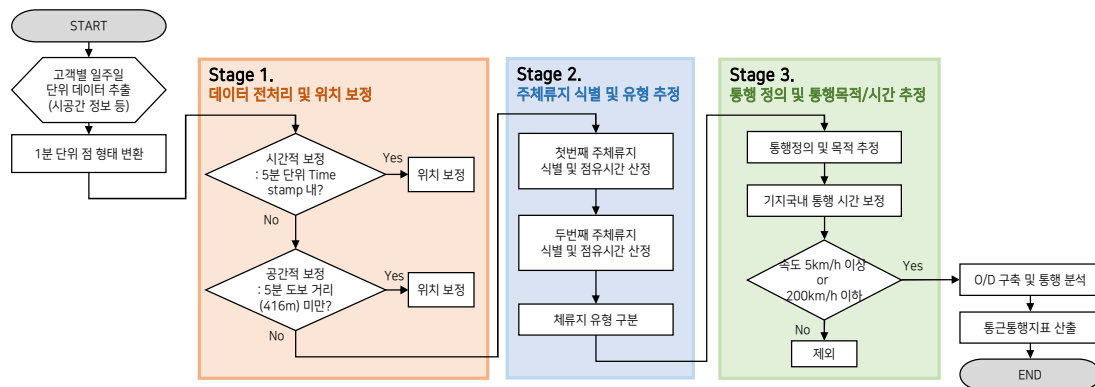
제4절 데이터 가공 프로세스 최적화 및 속도 개선

제2장 모바일 교통 빅데이터 전처리 가공 알고리즘 개발 및 고도화

제1절 원시데이터 전처리 가공 알고리즘 고도화

1. 개요

- 입력데이터(input data)로 각 통행자에 대한 7일 치 원시데이터가 사용되며 각 통행자에 대한 개별통행 특성을 기반으로 주체류지를 식별하고 통행지표를 산출
- 데이터 전처리 단계에서 신호 이상 현상 및 신호 이상치를 정형화할 필요 없이 데이터를 시·공간적으로 보정함으로써 비현실적으로 발생하는 통행자의 이동궤적을 보정
- 기지국 데이터 기록빈도 및 체류시간을 기반으로 통행자의 개별통행 특성을 고려한 주체류지 식별을 수행하기 때문에 정기 출퇴근 통행뿐만 아니라 야간 근무자를 포함한 비정기 출퇴근 통행도 식별 가능
- 개별 통행자에 대한 주체류지를 식별한 후, 해당 위치에서의 체류시작 시간과 체류종료 시간을 기반으로 통근시간 및 통행량 등 통행지표 산출



<그림 2-1> 모바일 교통 빅데이터 가공알고리즘

가. 전처리·가공 알고리즘의 주요 내용

- 기지국과 단말기 간에 발생하는 불규칙적인 신호 이상 현상을 보정함
- 개별통행 특성 및 지역 간 기지국 거리 등을 고려하여 이동/체류를 구분하고 체류지

유형을 추정함

- 통행자에 대한 개별통행 특성을 기반으로 주체류지를 식별하고 통행지표를 산출
- 데이터 전처리 단계에서 신호 이상 및 신호 이상치를 정형화할 필요 없이, 데이터를 시·공간적으로 보정함으로써 비현실적으로 발생하는 통행자의 이동궤적을 보정
- 기지국 데이터 기록빈도 및 체류시간을 기반으로 통행자의 개별통행 특성을 고려한 주체류지 식별을 수행하며 정기 출퇴근뿐만 아니라 야간 근무자를 포함한 비정기 출퇴근 통행도 식별함
- 개별 통행자에 대한 주체류지를 식별한 후, 해당 위치에서의 체류시작시간과 체류종료시간을 기반으로 통근시간 및 통행량 등 통행지표 산출

나. 전처리·가공 알고리즘의 주요 개선 및 고도화 내용

- 모바일 기지국 데이터를 기반으로 한 개인의 이동 및 통행목적 파악할 수 있는 개별 통행사슬(Trip Chain) DB를 구축하기 위한 효율적인 가공알고리즘의 개발이 필요함
- 고도화된 개별 통행사슬 DB를 구축하기 위한 알고리즘 고도화는 크게 데이터 전처리 및 이상치 처리(핸드오버 이슈 해결), 공간적 보정, 주 체류지 식별 및 유형 구분(통행목적) 세분화로 구분하여 보완하였음

2. 전처리·가공 알고리즘의 주요 절차 및 고도화

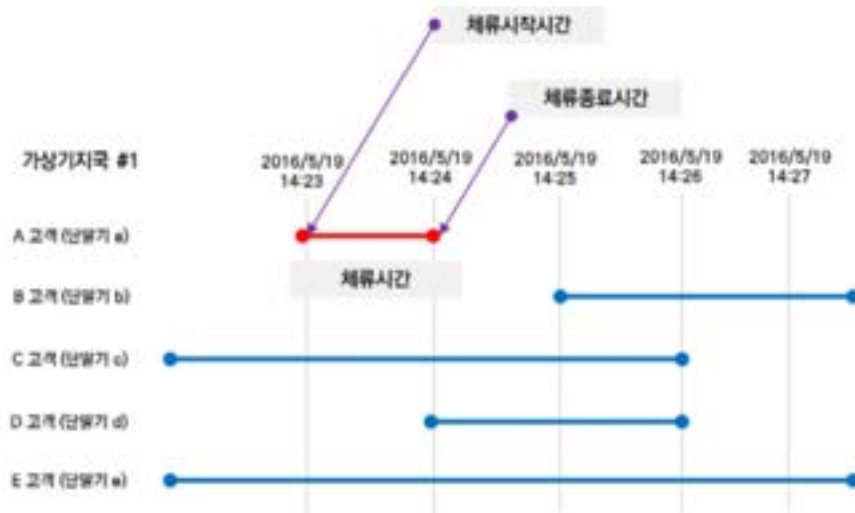
- 적절한 공간적 해상도를 바탕으로 통행자의 통행속도를 고려한 이동궤적을 식별하기 위해서 5분이라는 기준 시간을 정하고 해당 시간대에 기록된 기지국의 기록빈도를 기반으로 통행자의 실제 위치를 추정
- 짧은 시간 동안 발생한 핑퐁 핸드오버 현상에 의한 체류지 식별률 저하 및 비현실적으로 기록된 이동궤적을 보정하기 위하여 위해 특정 시간 단위(5분)별로 데이터를 분할
- 데이터를 분할 한 후, 각 시간 단위(time slice) 내에 기록된 기지국 데이터의 기록빈도를 기반으로 통행자의 실제 위치 추정
- 통행자의 실제 위치 추정 방법: 각 기지국에서의 체류시간을 기반으로 가중치를 주어 통행자 위치 좌표(X)를 계산
- 보행자 평균 도보 속도(5km/h)를 기준으로 5분(시간적 데이터 보정 시 설정한 시간

단위) 거리(약 416m) 범위 내 위치한 체류지들에 대해 각 체류지에서의 체류시간을 기반으로 해당 시간대에서의 통행자 실제 위치를 추정

- 원시데이터에 대해 시·공간적으로 신호 이상 기록을 보정해 줌으로써, 분 단위로 연속해서 변화하거나 좁은 공간적 범위 내에 산발적으로 기록되는 기지국 위치보정 가능

가. 모바일 기지국 원시데이터 전처리

- 포인트 단위의(밀리세컨 단위) 로그를 선분 단위의(분 단위) 로그로 변환
 - 닷(dot) 형태로 기록된 로그를 기록된 시간 순서대로 연결한 후, 최초 기록된 시간(이하 ‘체류시작시간’)과 마지막으로 기록된 시간(이하 ‘체류종료시간’)정보만 추출하여 단말기가 해당 기지국에 식별된 시간(이하 ‘체류시간’)을 산출함
 - 체류시간은 체류종료시간에서 체류시작시간을 뺀 값이며, 초(sec) 단위는 생략함



주: 자료를 참고하여 도식화 함

자료: 박미율·주은정(2017), LTE 시그널 정보를 이용한 위치정보가공 및 이동인구집계 방법, 한국통신학회 2017년도 동계종합학술발표회, p.285.

<그림 2-2> 선분 이력 데이터로의 변환 예시

나. 시간적 데이터 보정

- Liu T. et al.(2018)¹⁾이 제안한 내용에 의거, 적절한 공간적 해상도를 바탕으로 통행자의 통행속도를 고려한 이동궤적을 식별하기 위해서 5분이라는 기준 시간을 정하고 해당 시간대에 기록된 기지국의 기록 빈도를 기반으로 통행자의 실제 위치를 추정하고자 함
- 짧은 시간 동안 발생한 핑퐁 핸드오버 현상에 의한 체류지 식별률 저하 및 비현실적으로 기록된 이동궤적을 보정하기 위하여 위해 특정 시간 단위(5분)별로 데이터를 분할

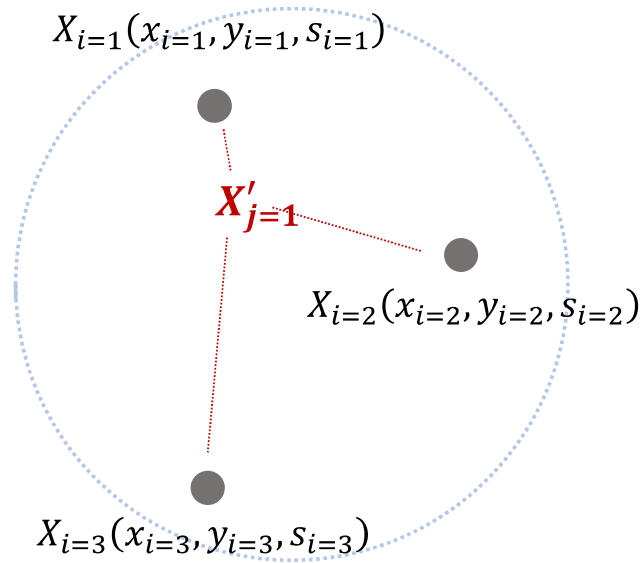
체류시작시간	체류종료시간	기지국	체류시간		체류시간대	통행자 위치
9:00	9:00	A	1	▶	09:00~09:04	$(1A+3B+1C) / 5 = X^*1$
9:01	9:02	B	2			
9:03	9:03	C	1			
9:04	9:07	B	4			
9:08	9:08	A	1		09:05~09:09	$(1A + 4B) / 5 = X^*2$
9:09	10:05	B	57			
10:06	10:07	A	2			
...

<그림 2-3> 시간적 데이터 보정 방법 예시

- 데이터를 분할 한 후, 각 시간 단위(time slice) 내에 기록된 기지국 데이터의 기록 빈도를 기반으로 통행자의 실제 위치 추정
- ※ 통행자의 실제 위치 추정 방법: <그림 2-4>에서 해당 time slice 내에는 3개의 기지국 좌표(A, B, C)가 기록되어 있으며, 통행자의 실제 위치를 추정할 때는 각 기지국에서의 체류시간을 기반으로 가중치를 주어 통행자 위치 좌표(X)를 계산
- 시간 단위 a 로 각 시그널 정보 $X_i(x_i, y_i, s_i)$ 를 그룹화한 후, 각 그룹별 시그널 정보의 위치 (x_i, y_i) 와 체류시간(s_i)을 가중치로 삼각 측량하여 통행자의 실제 위치 $X'_j(x'_j, y'_j, s'_j)$ 를 추정. 단, i 은 각 시간 단위 a 안에 있는 시그널 로그 순서($i = 1, 2, \dots, n$), j 은 보정된 로그 순서($j = 1, 2, \dots, m$).

$$X'_j(x'_j, y'_j) = X'_j\left(\frac{\sum_{i=1}^n x_i \times s_i}{\sum_{i=1}^n s_i}, \frac{\sum_{i=1}^n y_i \times s_i}{\sum_{i=1}^n s_i}\right) \quad \text{for } j = 1, 2, 3, \dots, m \quad \text{식 (1)}$$

¹⁾ T. Liu, Z. Yang, Y. Zhao, C. Wu, Z. Zhou, Y. Liu, "Temporal understanding of human mobility: A multi-time scale anlysis, PLOS ONE, Nov 2018.

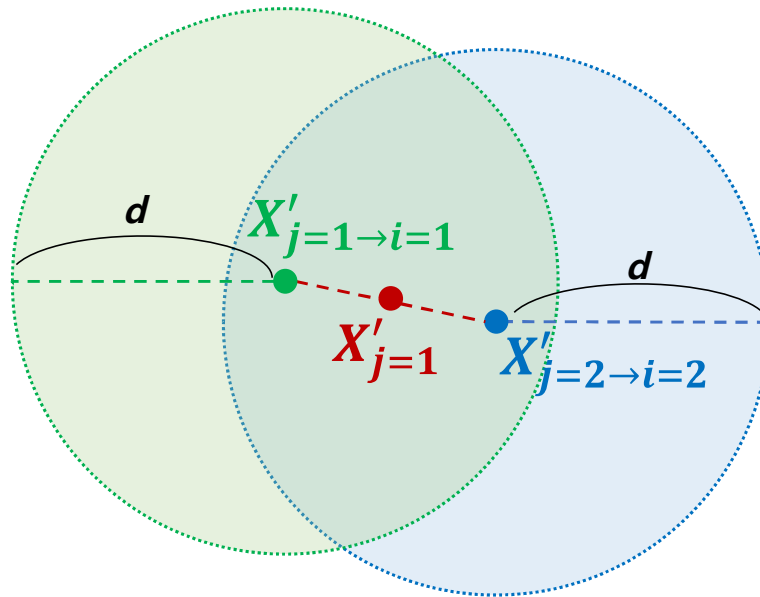


<그림 2-4> 통행자 실제 위치 추정 방법 예시

다. 공간적 데이터 보정

- Browning R. C., Baker E. A., Herron J. A., and Kram R.(2006)²⁾이 제시한 보행자 평균 도보 속도(5km/h)를 기준으로 5분(시간적 데이터 보정 시 설정한 시간 단위) 거리(약 416m) 범위 내 위치한 체류지들에 대해 각 체류지에서의 체류시간을 기반으로 해당 시간대에서의 통행자 실제 위치를 추정하고자 함
- 즉, 특정 시간 단위 a 로 보정된 각 시그널 정보 x'_j 의 위치가 최소 이동 속도 v_{\min} 기준 a 시간 안에 도달할 수 있는 거리 d 보다 작으면, 식(1)을 이용하여 위치를 다시 보정함
 - 원시데이터에 대해 시·공간적으로 신호 이상 기록을 보정해 줌으로써, 분 단위로 연속해서 변화하거나 좁은 공간적 범위 내에 산발적으로 기록되는 기지국 위치보정 가능

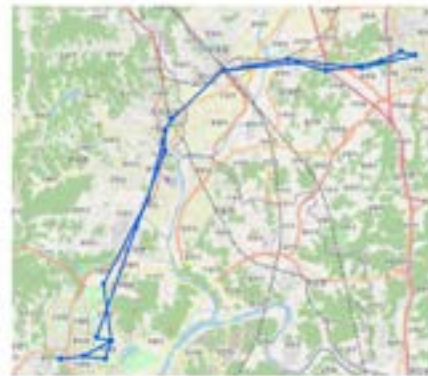
²⁾ R. Browning, E. A. Baker, J. A. Kram, "Effects of obesity and sex on the energetic cost and preferred speed of walking", Journal of Applied Physiology, 2006.



<그림 2-5> 통행자 실제 위치 추정 방법 예시



<데이터 전처리 전 통행 궤적>



<데이터 전처리 후 통행 궤적>

<그림 2-6> 시·공간적 데이터 보정(전처리) 적용 전, 후 통행 궤적 비교

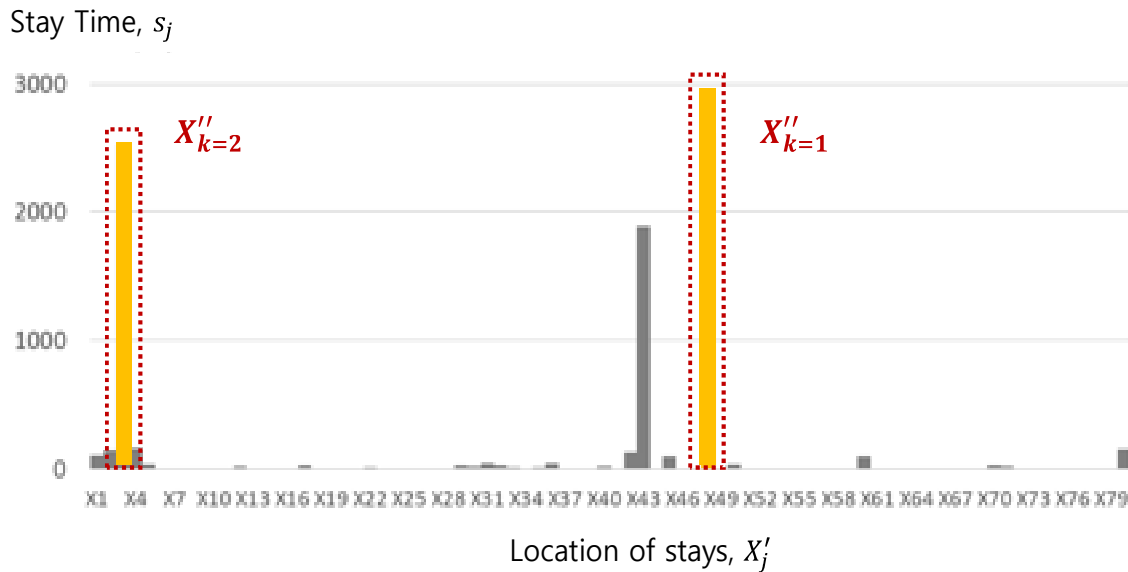
제2절 체류지 및 통행목적 추정 알고리즘 고도화

1. 체류지 식별 및 체류 정보 부여

가. 주체류지 식별 및 점유시간 산정

- 데이터 전처리 단계에서 추정된 통행자의 실제 위치($X'_j(x'_j, y'_j, s'_j)$)를 기준으로 통행자가 일주일 동안 각 위치에 몇 분 체류하였는지 계산 후, 최장체류지($X''_{k=1}(x''_{k=1}, y''_{k=1}, s''_{k=1})$) 식별

$$X''_{k=1}(x''_{k=1}, y''_{k=1}, s''_{k=1}) = [X'_j(x'_j, y'_j, s'_j) \mid \max(s'_j)] \quad \text{식 (2)}$$



<그림 2-7> 최장체류지 식별 예시

- 통행자가 최장체류지 이외의 체류지들에서 연속적으로 체류한 시간을 일별로 계산하여 그중에서 가장 긴 시간을 식별

체류시작시간	체류종료시간	체류시간	통행자 위치
00:00	04:56	04:57	최장 체류지
05:01	05:02	00:02	X'1
05:02	09:03	04:02	최장 체류지
09:04	09:06	00:03	X'2
09:07	09:10	00:04	X'3
09:10	09:15	00:06	X'4
09:20	17:40	08:21	X'5
17:40	17:45	00:06	X'6
17:45	17:50	00:06	최장 체류지
17:50	17:54	00:05	X'7
18:00	21:30	03:31	최장 체류지

→ 시간 1: 6분

← 시간 2가 가장 길

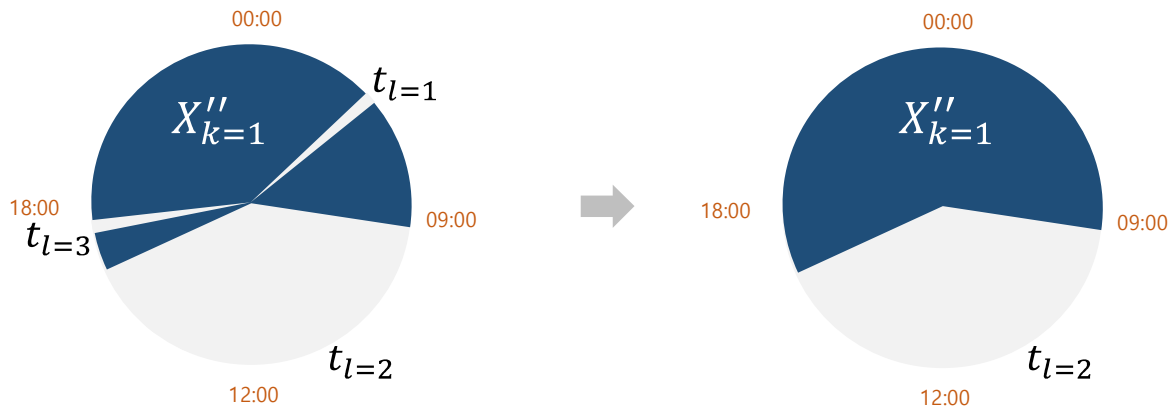
→ 시간 2: 8시간 42분

→ 시간 3: 10분

<그림 2-8> 최장체류지 이외의 체류지에 체류한 시간 중 가장 긴 시간 식별 예시

- 하루 24시간 중, 첫 번째 주 체류지 $X''_{k=1}(x''_{k=1}, y''_{k=1}, s''_{k=1})$ 이외의 점유시간 t 중 가장 점유시간만 남겨놓고 나머지 시간을 첫 번째 주 체류지의 점유시간으로 산정함

$$X''_{k=1}(x''_{k=1}, y''_{k=1}, s''_{k=1}) = X''_{k=1}(x''_{k=1}, y''_{k=1}, s''_{k=1} + \sum_{l=1}^r t_l - \max(t_l)) \quad \text{식 (3)}$$



<그림 2-9> 첫 번째 주 체류지(최장체류지) 점유시간 산정 예시

- 해당 시간 이외에는 통행자가 최장체류지에 체류한 것으로 지정하며, 해당 시간에 속하는 체류지들에 대하여 일주일 기준 체류시간이 가장 긴 체류지 식별 (2번째 최장체류지 식별 시, 최장체류지로 지정된 체류지들을 후보에서 제외하고 식별)

체류시작시간	체류종료시간	체류시간	통행자 위치
00:00	04:56	04:57	max1_X"
05:01	05:02	00:02	X"1
05:02	09:03	04:02	max1_X"
09:04	09:06	00:03	X"2
09:06	13:00	03:55	X"3
13:01	13:10	00:10	X"4
13:11	13:16	00:06	X"5
13:16	13:29	00:14	X"6
13:30	17:39	04:10	X"3
17:40	17:45	00:06	X"7
17:45	17:50	00:06	max1_X"
17:50	17:54	00:05	X"8
18:00	21:30	03:31	max1_X"
...

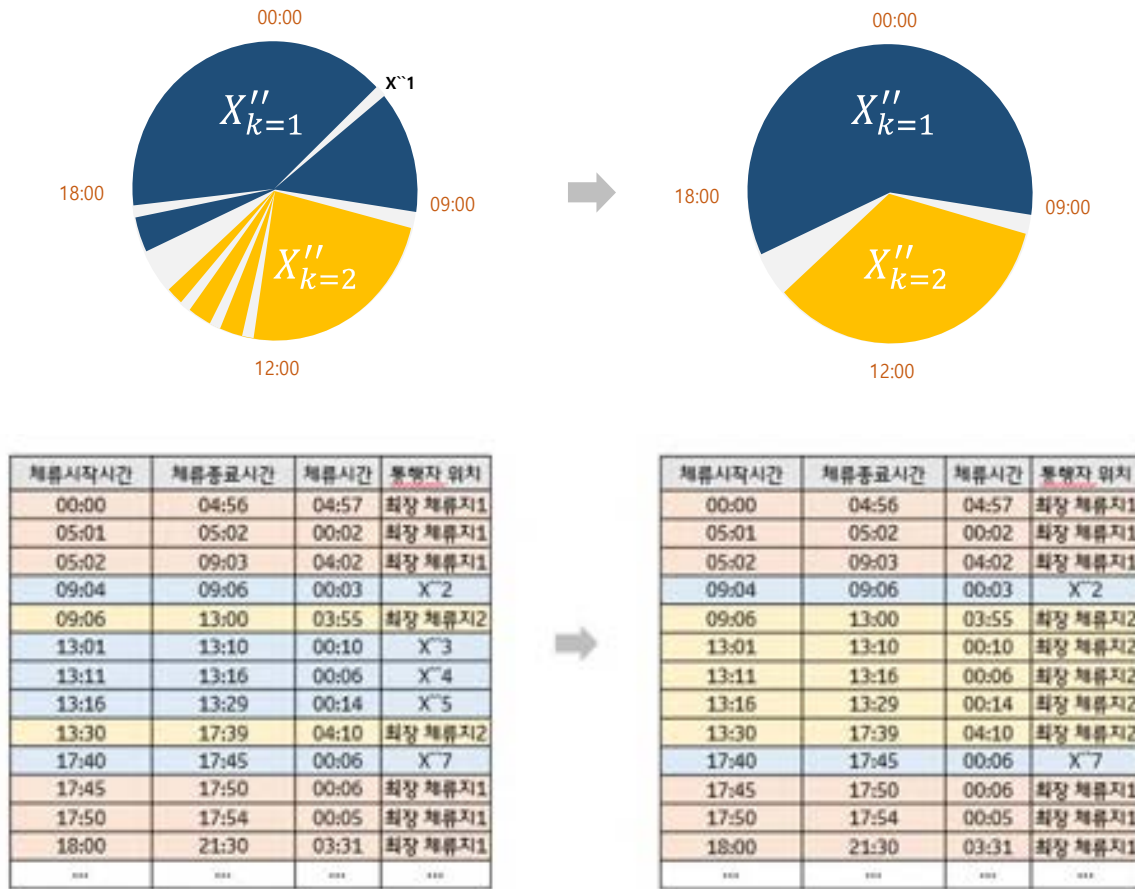
→

체류시작시간	체류종료시간	체류시간	통행자 위치
00:00	04:56	04:57	max1_X"
05:01	05:02	00:02	max1_X"
05:02	09:03	04:02	max1_X"
09:04	09:06	00:03	X"2
09:06	13:00	03:55	X"3
13:01	13:10	00:10	X"4
13:11	13:16	00:06	X"5
13:16	13:29	00:14	X"6
13:30	17:39	04:10	X"3
17:40	17:45	00:06	X"7
17:45	17:50	00:06	max1_X"
17:50	17:54	00:05	max1_X"
18:00	21:30	03:31	max1_X"
...



<그림 2-10> 최장체류지 이외 체류지 중 최장체류지 식별

- 일별로 2번째 최장체류지가 처음 기록된 시각(t1)과 마지막으로 기록된 시각(t2)를 식별한 후, t1부터 t2까지 통행자가 2번째 최장체류지에 체류한 것으로 보정(단, t1부터 t2 사이에 1번째 최장체류지로 지정된 체류지가 존재하는 경우, 2번째 최장체류지가 없는 것으로 판단)



<그림 2-11> 두 번째 주 체류지 식별 및 점유시간 산정 예시

- 2개의 최장체류지에 대해서 일주일 중 발생 빈도를 계산하여, 거주지 및 근무지 식별
 - ※ 발생 빈도가 더 많은 것을 거주지, 적은 것을 근무지로 식별하며 발생 빈도가 같은 경우, 일주일 기준 체류시간이 더 긴 것을 거주지, 적은 것을 근무지로 식별
- 근무지로 식별된 체류지에 대해 일주일 중 근무지에 체류한 시간대를 계산하여 동일 시간대에 주 2회, 회당 3시간 미만 체류한 경우는 근무지에서 제외
- 거주지와 근무지 위치는 각 주체류지로 구분된 최장체류지의 위치로 지정

2. 통행 정보 추정

가. 통행목적 분류

- 정의된 통행에서 출발 체류지 유형과 도착 체류지 유형을 통해 해당 통행에 대한 목적을 정의함
 - 예를 들어, 출발 체류지 유형이 집, 도착 체류지 유형이 회사일 경우, 통행목적은 출근이라고 구분하고, 출발 체류지 유형이 회사, 도착 체류지 유형이 집을 경우, 통행 목적을 퇴근이라고 구분함

<표 2-1> 체류지 유형별 식별 기준

체류지 유형		식별 기준			
		체류 특성		통행자 연령	비고
		체류시간	체류 빈도		
주 체류지	집	알고리즘 기준 최장 체류지 후보 1, 2 중	일주일 기준 최다빈도 체류지	-	첫 번째 주 체류지
	회사	동시간대 3시간 이상	주 2회 이상	30세 이상 60세 미만	두 번째 주 체류지 중
	학교	동시간대 3시간 이상	주 2회 이상	20세 미만	두 번째 주 체류지 중
잠재 체류지	정기적	3시간 이상	주 2회 이상	-	두 번째 주 체류지 중 회사나 학교로 구분되지 않은 체류지 중
	비정기적	3시간 이상	주 2회 미만	-	

나. 통행시간/속도 추정

- 통행시간의 경우, 거주지와 근무지에서의 체류시작시간과 체류종료시간을 기반으로 출근 및 퇴근 시간을 계산
 - ※ <그림 2-12>에서 출근 시간은 09:03부터 09:30까지인 27분으로 계산하며, 퇴근 시간은 17:39부터 18:25까지인 45분으로 계산

채류시작시간	채류종료시간	채류시간	통행자 위치
00:00	04:56	04:57	거주지
05:01	05:02	00:02	거주지
05:02	09:03	04:02	거주지
09:04	09:06	00:03	X"1
09:07	09:29	00:23	X"2
09:30	13:00	03:31	근무지
13:01	13:10	00:10	근무지
13:11	13:16	00:06	근무지
13:16	13:29	00:14	근무지
13:30	17:39	04:10	근무지
17:40	17:45	00:06	X"3
17:45	18:20	00:36	X"4
18:21	18:25	00:05	X"5
18:25	18:30	00:06	거주지
18:30	18:34	00:05	거주지
18:34	21:53	03:20	거주지
...

잠재채류

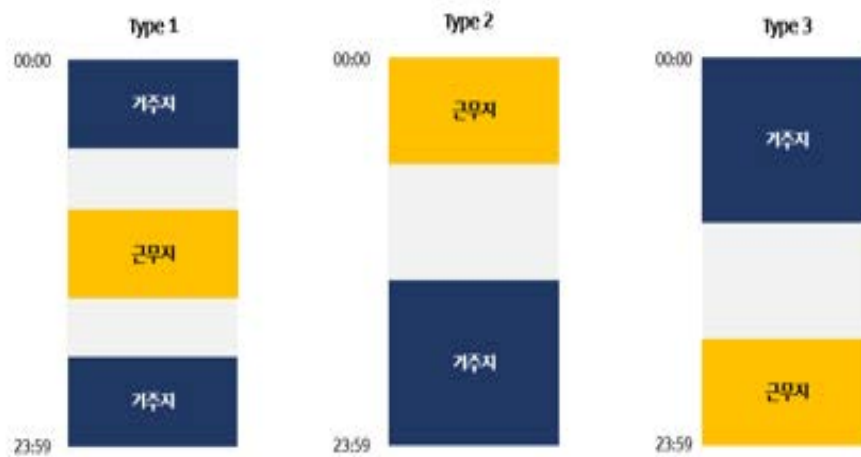
<그림 2-12> 통행시간 계산 및 잠재채류 식별 예시

- 통행량의 경우, 주체류지(거주지 및 근무지)를 기준으로 trip type별 일별 통행량 계산

※ type 1) [거주지-근무지-거주지]인 경우, 일별 통행량=2

type 2) [근무지-거주지]인 경우, 일별 통행량=1

type 3) [근무지-거주지]인 경우, 일별 통행량=1



<그림 2-13> 주체류지 기준 trip type 구분

	순번	고객식별번호	기준일	체류시작시간	체류종료시간	거차국	체류시간	
거주지 체류	206	164	20160525	0010	0158	I	108	
	207	164	20160525	0212	0212	M	0	
	208	164	20160525	0224	0318	I	54	
	209	164	20160525	0413	0413	M	0	
	210	164	20160525	0507	0507	I	0	
	211	164	20160525	0601	0723	M	82	
	212	164	20160525	0808	0808	I	0	
	213	164	20160525	0842	0842	M	0	
	214	164	20160525	0901	0909	I	8	
	215	164	20160525	0910	0911	M	1	
근무지 체류	216	164	20160525	0915	0917	I	2	1통행
	217	164	20160525	0920	0924	N	4	
	218	164	20160525	0925	0929	L	4	
	219	164	20160525	0932	0932	D	0	
	220	164	20160525	0935	1323	B	228	1통행
	221	164	20160525	1329	1343	A	14	
	222	164	20160525	1345	1949	B	364	
	223	164	20160525	2006	2014	L	8	
	224	164	20160525	2015	2019	I	4	
	225	164	20160525	2020	2025	N	5	
거주지 체류	226	164	20160525	2026	2054	I	28	
	227	164	20160525	2058	2108	M	10	
	228	164	20160525	2113	2114	I	1	
	229	164	20160525	2128	2128	M	0	
	230	164	20160525	2135	2143	I	8	
	231	164	20160525	2147	2158	M	11	
	232	164	20160525	2208	2337	I	89	
	233	164	20160525	2343	2343	M	0	
	234	164	20160525	2354	2354	I	0	

<그림 2-14> 일별 데이터에 대한 주체류지 간 통행량 산정 예시

- 출근 시간과 퇴근 시간 사이에 30분 이상 체류한 기록이 있으면 잠재체류로 식별
 - ※ <그림 2-15>에서 출근 시간 사이에 16:31부터 17:16까지 1회 30분 이상 체류한 기록이 있음 → 잠재체류로 판단
- 출근 또는 퇴근 중 잠재체류가 존재하는 경우 출근 시간 및 퇴근 시간 산출 대상에서 제외

	순번	고객식별번호	기준일	체류시작시간	체류종료시간	기지국	체류시간
거주지 체류	17	9939	20160520	0004	0503	T2	300
	18	9939	20160520	0507	0518	P2	12
	19	9939	20160520	0519	0519	K2	1
	20	9939	20160520	0520	0524	F2	5
	21	9939	20160520	0525	0529	Z	5
	22	9939	20160520	0530	0534	T	5
	23	9939	20160520	0539	0540	J	2
	24	9939	20160520	0541	0544	N	4
	25	9939	20160520	0545	1554	P	610
근무지 체류	26	9939	20160520	1555	1558	W	4
	27	9939	20160520	1603	1603	R2	1
	28	9939	20160520	1619	1628	Q2	10
	29	9939	20160520	1631	1716	M2	46
	30	9939	20160520	1716	1720	Q2	5
	31	9939	20160520	1752	2358	T2	367

← 잠재체류지

<그림 2-15> 잠재체류 식별 예시

- 출근 속도의 경우, 거주지와 근무지 간 거리를 출근 시간으로 나누어 계산하고 퇴근 속도의 경우, 거주지와 근무지 간 거리를 퇴근 시간으로 나누어 계산
- 출·퇴근 속도를 기준으로 이상치를 식별하고 출근 시간 및 퇴근 시간 산출 대상에서 제외
 - 출근 속도 또는 퇴근 속도가 평균 도로 속도인 5km/h 미만인 경우, ‘data lagging’ 등 데이터 수집 지연 현상에 의한 이상치로 판단
 - ※ <그림 2-16>은 data lagging이 발생한 예시를 나타내며, 13:28부터 14:11에 해당하는 데이터 로그가 없음. 그림에서 파란색으로 표시된 데이터 로그가 거주지, 빨간색으로 표시된 데이터 로그가 근무지에 해당하는 로그이며 퇴근 시간이 데이터 로그가 없던 시간인 13:28부터 14:11까지인 43분으로 과대 추정될 수 있음
 - 출근 속도 또는 퇴근 속도가 300km/h 이상인 경우, 신호 이상에 의한 체류지 식별 오류로 판단하여 출근 시간 및 퇴근 시간 산출 대상에서 제외
 - ※ KTX 이동속도(300km/h)를 고려하여 임계치 설정

				체류 시작시간	체류 종료시간	경도	위도
21	20190402	387	50	2100	2318	127.0613	37.24505
22	20190403	387	50	12	1126	127.0613	37.24505
23	20190403	387	50	1220	1328	127.067	37.24596
24	20190403	387	50	1411	1730	127.0613	37.24505
25	20190403	387	50	1815	1815	127.0613	37.24505
26	20190403	387	50	1822	2334	127.0613	37.24505
27	20190404	387	50	29	719	127.0613	37.24505

<그림 2-16> Data Lagging 예시

- 원시데이터에 대해 본 알고리즘을 적용하면 <그림 2-17>과 같은 형태의 통행지표가 산출되며, 해당 산출물에는 개별 고객에 대한 일자별 통행지표가 포함되어 있음

날짜	2020-04-01
고객식별번호	112233
나이	30분
성별	M
출근 시작시간	8 : 30
출근 종료시간	9 : 00
출근 소요시간	30분
출근 중 잠재체류 회수	0회
출근 중 잠재체류 시간	0분
퇴근 시작시간	18 : 27
퇴근 종료시간	19 : 30
퇴근 소요시간	63분
퇴근 중 잠재체류 회수	1
퇴근 중 잠재체류 시간	35
집 위치	127.491, 36.455
회사 위치	127.485, 36.607
집/회사 간 거리	16.9 km
출근시간	30분
퇴근시간	X
출근속도	33.8 km/h
퇴근속도	X

<그림 2-17> 통행지표 산출물 예시

3. 격자 단위 공간 매칭

가. 필요성

- 기존 통행사슬 DB의 위상도 체계는 교통폴리곤 기준으로 구성됨
- 행정동 단위보다 더 세밀하게 하여 보고자 하였으나, 지방의 경우 교통폴리곤 체계로는 O/D 지점을 정밀하게 보기엔 어려운 한계점이 존재
- 교통망에 따른 변화도 일어나 데이터의 왜곡을 줄 수 있는 문제점도 존재
- 위 2가지 단점을 보완하기 위하여 기존 교통폴리곤 단위에서 500m 격자 단위로 변경하고자 함
- 지방지역의 경우 교통폴리곤이 읍면동과 큰 차이를 보이지 않아 분석 공간 단위로서 상세함이 떨어짐

나. 국토지리정보원의 국가지점번호 활용 필요성

- 생산시간 유지 및 최적화
 - 16,000여개 폴리곤에서 410,000여 개 500m격자로 단위가 바뀌며 공간 연산량 과 부하가 우려됐으나, 단순한 형태(정사각형 구조)로 인해 규칙 및 규격이 동일해져 표준화가 용이하며, 생산시간 또한 큰 차이가 없음
 - 격자 변경으로 인해 통행사슬 DB 생산 부분에서 기존에 RAM이 과도하게 사용되어 중간에 중지되는 상황이 줄어 최적화 부분 반영
- 교통폴리곤 대비 지방지역 분석 공간 단위 상세함 향상
 - 기존 교통폴리곤 단위 대비 더 낮은 분석 단위로 일정하게 규격화되어 있어, 상세한 분석 공간 단위를 가질 수 있음
- 국가지점번호 격자의 표준성
 - 국가지점번호는 반영구적으로 변하지 않음
 - 행정구역이 다양하게 변한다고 하여도 국가지점번호 격자는 동일하기에 행정정보를 위치로 매핑하면 격자 정보 파악이 용이함
- cpp 파일 재설계
 - KT 내부망 내 R의 sf 패키지 사용 불가(버전 불일치)

- cpp로 별도 function을 구성했으나 활용 불가
- cpp 참조 파일 형태는 오로지 숫자로만 구성됐으며 feature가 3개 필요
- 국가지점번호 500m 격자는 한글 체계이며, feature 2개로 구성되어 cpp 재설계 필요(c++ 체계에서 R 자체 처리 체계로 전환 필요)

다. 500×500m 격자 단위 공간 매칭 과정

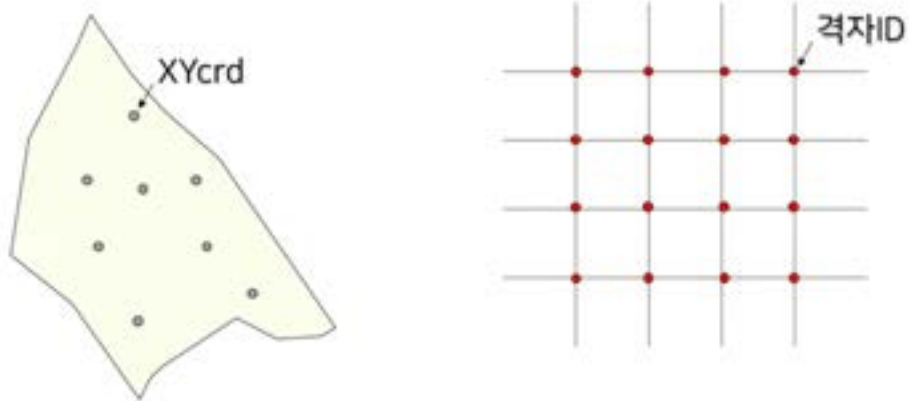
- 국가지점번호 500m 격자 교체 과정은 다음과 같음

<표 2-2> 개선 알고리즘 단계별 코드 구성 내용

구분	순번	내용
Stage 1. 데이터 준비	1	시군구단위 보정속도 테이블 준비(polycodeinfo.tsv)
	2	국가지점번호 500m 격자 준비(국토지리정보원)
Stage 2. 격자별 시군구 단위 속도보정값 join	3	교통폴리곤 단위 시군구 보정속도 값 → 국가지점번호 500m 격자 단위 보정속도 값으로 spatial join
Stage 3 고유좌표값 변환 및 격자ID 좌표와 조인	4	국가지점번호 격자ID는 각 격자의 좌하단 꼭지점을 표시함(미터법에 따름)
	5	격자ID → 좌표화
	6	각 주차별 모바일 신호 포착 고유좌표(uniqueXYcrd)를 격자ID 좌표계와 동일하게 변환
	7	변환된 좌표를 격자 ID 좌표체계와 동일하게 적용
	8	uniqueXYcrd 격자좌표 = 격자ID 좌표와 조인- 각 uniqueXYcrd의 소속 격자 확인
	9	격자에 미리 맵핑된 정보(행정구역정보, 시군구 별 보정속도 값)를 활용하여 처리

< 단계별 프로세스 >

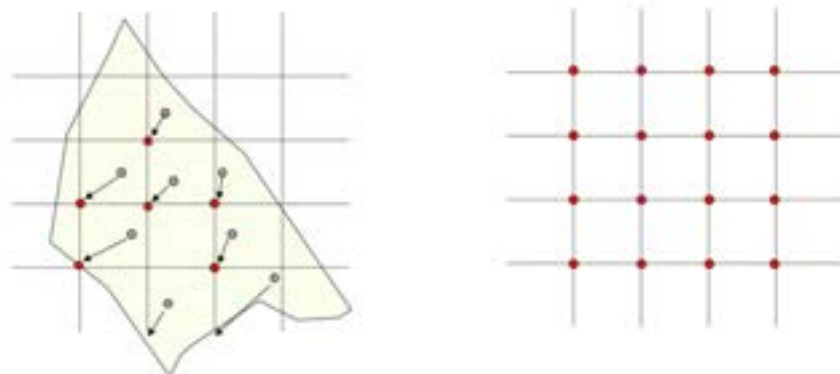
1) stage 1



프로세스 단계별 예시 1

- 각 주차별 모바일 신호 포착 고유좌표(XYcrd) 테이블 준비 및 국가지점번호 500m격자 테이블 준비
- 국가지점번호 격자 ID는 각 격자의 좌 하단 꼭지점을 기준으로 함

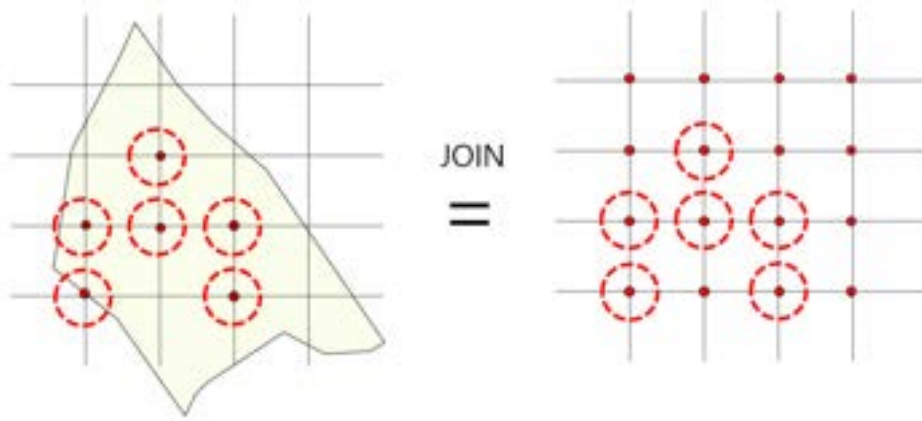
2) stage 2



프로세스 단계별 예시 2

- 모바일 신호 포착 고유좌표(XYcrd)를 격자ID 좌표계와 동일하게 변환
(격자ID와 마찬가지로 좌 하단 꼭지점으로 고유좌표를 변환)

3) stage 3



프로세스 단계별 예시 3

- 변환된 고유좌표(XYcrd) 값과 일치하는 격자 ID를 통해 각 고유좌표값이 소속된 격자를 확인함

제3절 교통수단 추정 및 전수화 방법론 개발

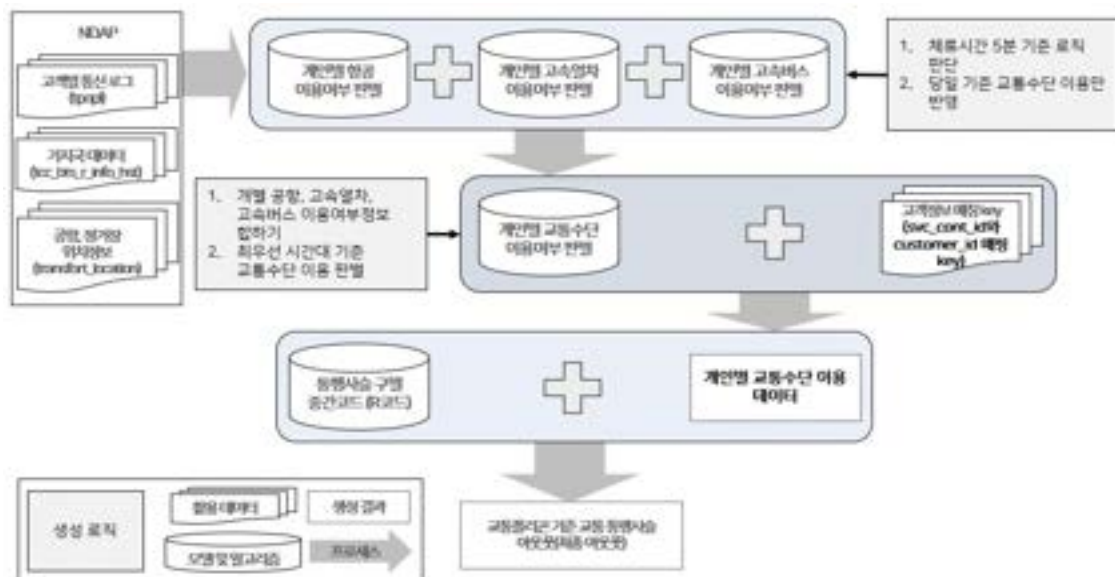
1. 지역 간 교통수단 구분 방법론 개발

가. 개요

- 모바일 통신 데이터에서 추출한 통행 경로와 대중교통 정보를 결합하여 이동이 발생한 통행자의 지역 간 교통수단을 체계적으로 정의
- 기존 통행사슬 DB 알고리즘을 고도화하여, 최종 결과물에 고객의 지역 간 교통수단에 대한 정보를 추가함
- 지역 간 교통수단별 이용 상세 조건은 상이함

<표 2-4> 지역 간 교통수단 구분 (예시)

지역 간 교통수단 유형	세부 내용
항공(F)	국내공항을 경유하여 항공수단을 이용하여 이동이 발생한 경우
고속철도(T)	기차역에 경유하여 고속철도(KTX, SRT)를 이용하여 이동이 발생한 경우
고속버스(B)	버스 터미널에 경유하여 고속버스를 이용하여 이동이 발생한 경우
기타(E)	지역 간 교통수단으로 구분되지 않는 경우



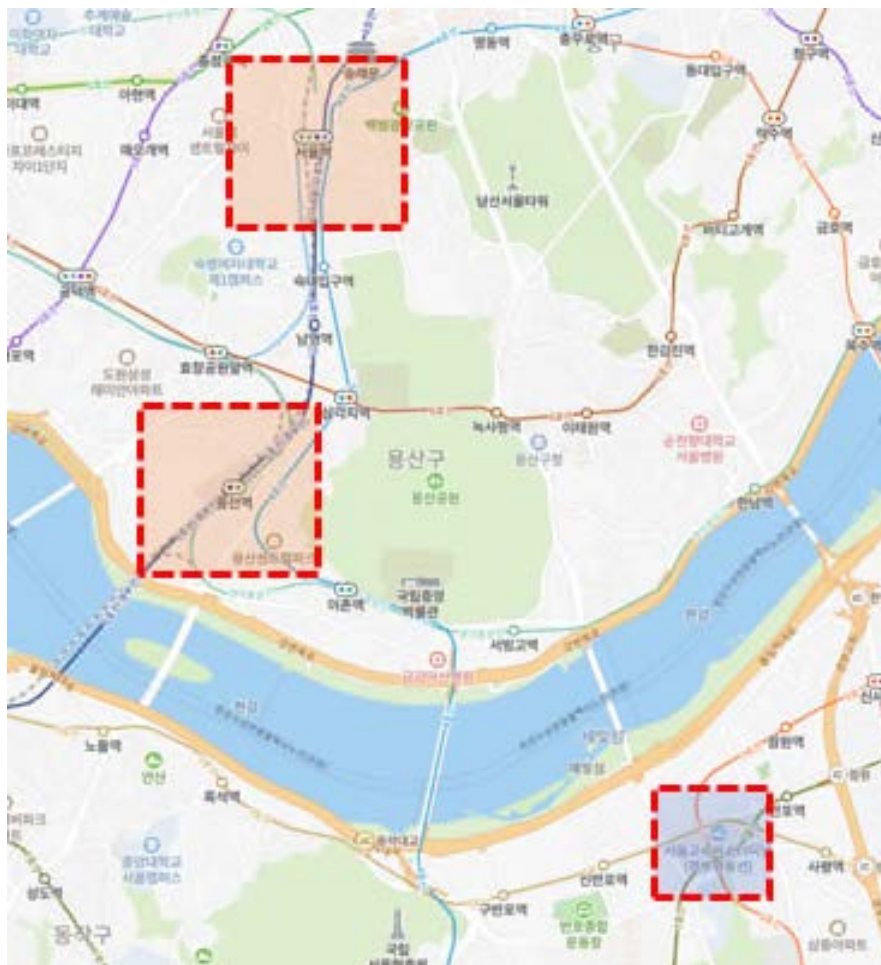
<그림 2-18> 지역 간 교통수단 구분 로직 및 결합 흐름도

나. 지역 간 교통수단 구분 방법론

- 구체적인 지역 간 교통수단 구분 방법론은 다음 프로세스를 적용하여 지역 간 교통수단을 구분함

1) 지역 간 주요 교통시설 설정

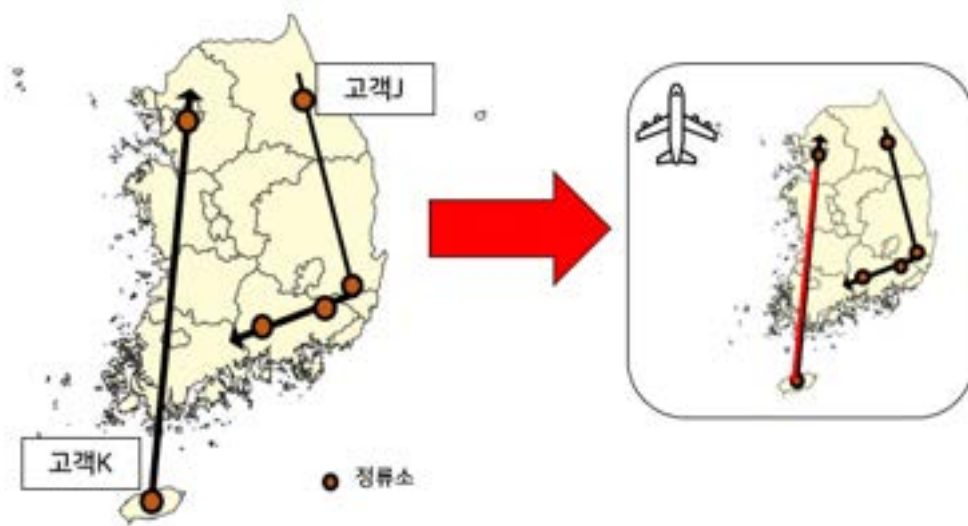
- 국내 공항, 기차역(KTX, SRT), 고속버스터미널 위치 데이터 수집
- 수집된 국내 공항, 기차역(KTX, SRT), 고속버스터미널 위치를 기준으로 버퍼 적용 및 shp 생성
 - 공항 약 5km, 기차역 약 1km, 고속버스터미널 약 500m를 기준으로 버퍼를 적용하여 각 정류소별 기지국 shp을 생성함



<그림 2-19> 서울역, 용산역, 서울고속버스터미널 shp (예시)

2) 체류 정보를 이용한 이용 교통수단 구분

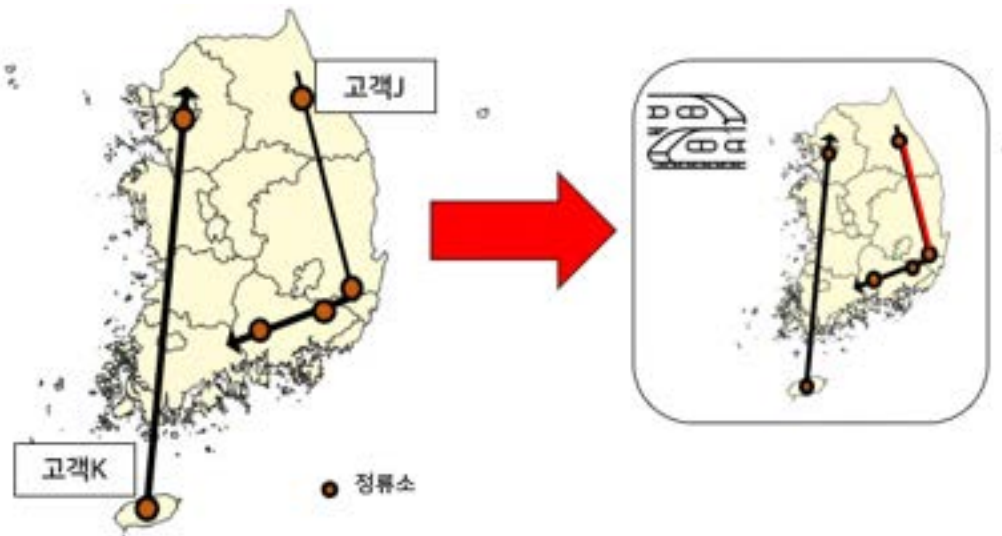
- 고객의 동선 중, 정류소별 기지국 shp 방문 유무를 기준으로 하여, 지역 간 교통수단 이용자로 구분하고 수단별 기준을 달리 적용함
 - 공통기준
 - 당일 고객의 동선 중 각 지역 간 교통수단의 2개 이상의 정류소 shp에 맵핑되는 경우 해당 교통수단 이용자로 구분
 - 각 정류장에 체류한 시간을 계산하여 5분 미만일 경우 정류장을 이용하지 않은 것으로 판단
 - 고객의 동선 중 지역 간 교통수단의 중복이 일어날 경우 가장 처음으로 판별된 교통수단을 적용
 - 항공(F) 이용자 구분 기준



<그림 2-20> 고객 동선 중 항공 이용자 구분 예시

- 최대 이동 시간 2시간 내외(비행시간이 2시간을 넘어서면 비행기를 이용하지 않은 것으로 판별)
- 동일 시도에서의 서로 다른 공항 방문은 비행기를 이용하지 않은 것으로 판별
- 실제 비행기 운행 시간 외의 이용은 결과에서 제외(00:00~05:59)

고속철도(T) 이용자 구분 기준



<그림 2-21> 고객 동선 중 고속철도 이용자 구분 예시

- 최대 이동 시간 6시간 내외(기차 이동 시간이 5시간을 넘어서면 기차를 이용하지 않은 것으로 판별)
- 동일 시도에서의 서로 다른 기차역 방문은 기차를 이용하지 않은 것으로 판별
- 실제 기차 운행 시간 외의 이용은 결과에서 제외(02:00~04:59)

- 고속버스(B) 이용자 구분 기준



<그림 2-22> 고객 동선 중 고속버스 이용자 구분 예시

- 최대 이동 시간 7시간 내외(고속버스 이동 시간이 7시간을 넘어서면 고속버스를 이용하지 않은 것으로 판별)
- 동일한 시도에서의 서로 다른 터미널 방문은 고속버스를 이용하지 않은 것으로 판별(단 광역시도 외의 시도에서는 시도 내 이동을 포함)
- 실제 고속버스 운행 시간 외의 이용은 결과에서 제외(03:00~04:59)
- 각 지역 간 교통수단의 정류장 1곳만 방문한 기록만 존재하는 경우 해당 교통수단을 사용하지 않은 것으로 판단
- 고객의 동선 중 지역 간 교통수단의 중복이 일어날 경우 가장 처음으로 판별된 교통수단을 적용
- 공항, 기차역, 고속버스터미널 간 기지국 버퍼를 적용
 - 기지국에 신호가 100% 수신되지 않으므로, 인근 기지국에서 확인된 신호를 가져올 필요성이 있음 (예시. 항공기의 경우 고도가 높아짐에 따라 지상에 있는 기지국에 신호가 잡히지 않는 경우 발생)
 - 보통 지역 간 교통수단 이용지점은 인구밀집지대로서 기지국 수용량을 초과하는 신호가 다량 발생할 수 있음. 그로 인해 인근 기지국에서 신호가 대신 수신되는 경우가 발생. 따라서 실제 이용자의 통신 신호 파악을 위해 더 넓은 범위에 기지국 신호를 확인해야 함

다. 지역 간 교통수단 정보 테이블 병합

- 중간 산출물 테이블 중의 고객식별 id와 결합할 수 있는 개인 svc_cont_id 값 매칭 key를 생성, 고객 key 값과 일자가 일치하며, 출근종료시각(HC_end), 퇴근시작시간(CH_start) 사이에 고객의 지역 간 교통수단 로그가 존재할 경우 해당 지역 간 교통수단을 이용했다고 판별
- 기존 개발한 모바일 통신 데이터 가공알고리즘의 중간 산출물 테이블과 지역 간 교통수단 구분 테이블을 결합하여 중간 산출물에 이용자의 지역 간 교통수단을 구분하는 정보를 추가, 지역 간 교통수단 정보가 추가된 최종 산출물을 생산
- 집계 시 해당일에 가장 많이 이용한 교통수단을 대표로 표시하며, 이용자 수가 동일할 경우, 가장 상위 교통수단(항공 > 고속열차 > 고속버스 순)으로 표시

코드명	내용	코드명	내용
date	날짜(년/월/일)	date	날짜(년/월/일)
customer id	고객식별 id	customer id	고객식별 id
age	고객 나이대 (ex. 40:40대)	age	고객 나이대 (ex. 40:40대)
gender	고객 성별 (M:남성, F:여성)	gender	고객 성별 (M:남성, F:여성)
type	목적지 유형 (c:회사, s:학교, r:정기잠재체류지, x:기타잠재체류지)	type	목적지 유형 (c:회사, s:학교, r:정기잠재체류지, x:기타잠재체류지)
HC_start	출근시작시각 (분 단위)	HC_start	출근시작시각 (분 단위)
HC_end	출근종료시각 (분 단위)	HC_end	출근종료시각 (분 단위)
HC_time	출근시간 (HC_time = HC_end - HC_start)	HC_time	출근시간 (HC_time = HC_end - HC_start)
HC_ps_num_60	출근시간 사이에 기록된 잠재체류 횟수	HC_ps_num_60	출근시간 사이에 기록된 잠재체류 횟수
HC_ps_time_60	출근시간 사이에 기록된 잠재체류 총 시간 (분 단위)	HC_ps_time_60	출근시간 사이에 기록된 잠재체류 총 시간 (분 단위)
CH_start	퇴근시작시각 (분 단위)	CH_start	퇴근시작시각 (분 단위)
CH_end	퇴근종료시각 (분 단위)	CH_end	퇴근종료시각 (분 단위)
CH_time	퇴근시간 (CH_time = CH_end - CH_start)	CH_time	퇴근시간 (CH_time = CH_end - CH_start)
CH_ps_num_60	퇴근시간 사이에 기록된 잠재체류 횟수	CH_ps_num_60	퇴근시간 사이에 기록된 잠재체류 횟수
CH_ps_time_60	퇴근시간 사이에 기록된 잠재체류 총 시간 (분 단위)	CH_ps_time_60	퇴근시간 사이에 기록된 잠재체류 총 시간 (분 단위)
H_ord	거주지 좌표 (경도, 위도)	H_ord	거주지 좌표 (경도, 위도)
C_ord	근무지 좌표 (경도, 위도)	C_ord	근무지 좌표 (경도, 위도)
HC60	잠재체류 및 이상치를 제외한 출근시간	HC60	잠재체류 및 이상치를 제외한 출근시간
CH60	잠재체류 및 이상치를 제외한 퇴근시간	CH60	잠재체류 및 이상치를 제외한 퇴근시간
distance	거주지와 근무지 간 거리 (km)	distance	거주지와 근무지 간 거리 (km)
hc_speed60	출근속도 (km/h)	hc_speed60	출근속도 (km/h)
ch_speed60	퇴근속도 (km/h)	ch_speed60	퇴근속도 (km/h)
trv60	통행량 (통행 중 잠재 체류가 포함된 경우를 제외한 통행만 계산)	trv60	통행량 (통행 중 잠재 체류가 포함된 경우를 제외한 통행만 계산)
p_trv60	잠재 통행량 (통행 중 잠재 체류가 포함된 통행만 계산)	p_trv60	잠재 통행량 (통행 중 잠재 체류가 포함된 통행만 계산)
		traffic_gp	통행수단구분

<그림 2-23> 지역 간 교통수단 구분 결합 전·후 중간 산출물 테이블 정의서 (예시)

기준일	출발				도착				성별	연령대	통행방향	보정계수	평균 통행 시간	통행속도, 거리, 분 산, 중앙값 등)	통행수단
	요일 *	시간 단위	공간 단위	유형	요일 *	시간 ***	시도	유형***							
20200405	2	7	11010	H	2	8	13010	C	M	20	3	0.3	10	—	-
20201221	2	30	11020	H	2	12	13020	C	F	30	3	0.3	25	—	항공
20200101	6	6	11060	H	6	7	13060	C	M	50	3	0.3	15	—	기차
20200505	3	18	11040	C	3	15	13040	H	M	20	4	0.5	10	—	-
20201302	4	18	23010	C	4	1	23010	H	F	30	6	0.5	25	—	-
20201001	2	3	31023	X	2	5	33023	H	M	50	4	0.5	15	—	버스

<그림 2-24> 지역 간 교통수단 구분 결합 후 최종 산출물 테이블 (예시)

2. 전수화 방법론 개발

가. 개요

1) 필요성

- 데이터 전수화를 통해 KT 데이터에 국한된 정보를 대한민국 전체인구의 속성을 지닌 데이터로 가공하여 전 국민의 정보가 반영되는 데이터로 가공하는 것이 전수화 목적
- 데이터 전수화를 통해 데이터를 KT 이용고객뿐만이 아닌 대한민국 전체인구의 속성이 반영된 데이터를 이용함에 더욱 정교한 분석결과 기대
- 해당 데이터 가공을 통한 대한민국 전체인구의 속성 반영으로 통행사슬 DB의 대중적인 정보를 통해 성별, 연령별, 지역별에 따른 세부 속성의 정보 활용을 통해 의미 있는 결과 도출 기대

2) 기존 전수화 방법의 한계점 및 개선사항

- 기존 통신회사에서는 시장점유율을 이용하여 전수화를 실시함
 - 통행사슬 DB 원시데이터는 KT 내 고객 식별번호, 가상기지국, 체류시작시간, 체류종료시간, 연령대, 성별 등의 정보로 구성되어 있음
 - 기존 통신회사는 이용고객 인원과 시장점유율 정보를 통해 역산하여 전수화 계수를 산정하고 모바일 통신 데이터 기반 통행량을 전수화 함
 - 시장점유율 기반 전수화 계수 적용 방안의 경우 통신사 점유율을 기준으로, [총 통행량 \times (100 / KT 시장점유율)] 수식을 통해 전수화 계수 값 산출

기준일	출발				도착				성별	연령대	총통행량	평균통행 시간
	요일	시간대	공간	유형	요일	시간대	공간	유형				
20201221	2	10	A	집	2	11	B	학교	F	30	3	337
20201221	2	10	B	집	2	13	C	직장	F	30	2	172
20201221	2	10	C	집	2	13	D	기타	F	30	2	114
...

기준일	출발				도착				성별	연령대	총 통행량	평균통행 시간	전수화 계수 (점유율)	전수화 반영 총 통행량
	요일	시간대	공간	유형	요일	시간대	공간	유형						
20201221	2	10	A	집	2	11	B	학교	F	30	3	337	3	9
20201221	2	10	B	집	2	13	C	직장	F	30	2	172	3	6
20201221	2	10	C	집	2	13	D	기타	F	30	2	114	3	6
...

<그림 2-25> 시장점유율 기반 전수화 계수 적용 방법

- 그러나, 기존 시장점유율을 이용한 전수화 기법은 다음과 같은 한계점을 가지고 있음
 - 모바일 통신회사 시장점유율의 경우 성별, 연령별, 지역별 세부적인 점유율 없이 전체 통신 시장점유율에 대한 정보만 존재
 - 시장점유율 기반 전수화 방안을 통해 집계된 데이터는 성별, 연령별, 지역의 경우 통신사 이용고객 수에 따라 과대 혹은 과소 추정되는 문제점 존재
- 따라서, 기존 전수화 보정에서 성별, 연령별, 지역별 등 세부 속성의 보정이 되지 않던 문제를 해당 속성이 포함된 생활인구 계수 적용하여 해결 필요

나. 전수화 방법론

1) 생활인구 기반 전수화 방안

- 통신사 고객의 로그 정보를 기반으로 집계된 생활인구 데이터를 기존 통행사슬 DB 정보와 결합하여 전수화 작업 목표
 - 생활인구 데이터는 지역별, 가상기지국, 연령대, 성별, 숙박인구 등의 정보를 지님
 - 생활인구 데이터는 주민등록인구 기반으로 전수화 가중치 반영이 되어있는 데이터
 - 생활인구 데이터의 숙박인구 정보와 지역별, 성별, 연령별 정보를 활용하여 통행사슬 DB와 결합(join)하여 전수화 보정 작업 진행
 - 해당 전수화 방안 적용시, 단순 결합(join) 작업만 존재하여 전수화 보정 작업이 간

단하며 속도가 빠름

- 또한, 성별, 연령대, 지역의 세부 속성을 모두 반영한 전수화 방법으로 기존 전수화 방안의 대표성 문제를 해결
- 생활인구 기반 전수화 계수 적용 한계점
 - 통행사슬 DB와 생활인구 데이터의 공간 위계 차이로 인한 오차 존재
 - 통행사슬 DB와 생활인구 간의 인구 집계 로직 차이 존재로 인한 오차 존재
- 생활인구 기반 전수화 한계점 해결방안
 - 생활인구에 사용된 주민등록인구 기반의 전수화 보정 방법을 통행사슬 DB에 적용하여 해결

2) 출발지 주민등록인구 기준 전수화 방안

- 생활인구 기반의 전수화 보정 한계점에 따라, 통행사슬 DB에 주민등록인구를 사용하여 전수화 계수 적용 문제를 해결하고자 한 방안
- 통행사슬 DB의 O/D 정보 중 출발지역을 나타내는 O(Origin) 타입이 거주지를 나타내는 H로 식별된 경우, 지역별 주민등록상 인구수를 통한 비율을 구하여 전수화 계수 산출
 - 통행사슬 교통DB는 각 행이 통행별 O/D로 구성되어있기 때문에, 모든 통행의 시작점은 O이고 이는 모든 통행을 반영하기에 O를 기준으로 두고 전수화 작업을 진행



<그림 2-26> O 기준 거주지(H)에 따른 전수화 적용(안)

- O 기준 거주지(H)에 따른 전수화 계수의 경우 성별, 연령, 지역별 (거주지 인구/ 주민등록수)의 수식을 통해 전수화 계수 값 산출하여 전체인구 값 보정

<교통DB 데이터>													<주민등록인구기반 전수화 계수>				
기준일	출발			도착			성별	연령대	통행수단	종통행량	평균통행 시간		공간 (시군구)	성별	연령대	전수화 계수	
	요일	시간대	공간	요일	시간대	공간											
20201221	2	10	A	집	2	11	B	학교	F	30	B	2	337	A	F	30	3.25
20201221	2	10	A	집	2	13	C	직장	M	30	F	2	172	A	M	30	3.12
20201221	2	10	B	집	2	13	D	기타	F	20	T	2	114	B	F	20	2.75
...

기준일	출발			도착			성별	연령대	통행수단	종통행량	평균통행 시간	전수화 계수	전수화 반영 종통행량		
	요일	시간대	공간	요일	시간대	공간									
20201221	2	10	A	집	2	11	B	학교	F	30	B	2	337	3.25	6.50
20201221	2	10	B	집	2	13	C	직장	M	30	F	2	172	3.12	6.24
20201221	2	10	C	집	2	13	D	기타	F	20	T	2	114	2.75	5.50
...

<그림 2-27> O 기준 거주지(H)에 따른 전수화 적용 테이블

- O 기준 거주지(H)에 따른 전수화 적용의 문제점
 - 통행 DB의 O 중 거주지를 나타내는 H는 전체 통행 DB 중 약 10% 존재
 - 전체 개별통행 중 거주지에서 출발하는 통행만 반영되는 한계가 존재
 - 전체 DB 데이터 중 약 10% 정도만 보정되는 한계점으로 인해 대표성을 지니기엔 부족함
 - O 기준 거주지(H)에 따른 전수화 적용의 문제점 해결방안
 - 통행사슬 DB 생성 중간과정 중, 개인별 데이터 집계된 상황에서 보정계수 값 반영 후, 지역별 집계 작업 진행
 - 지역별 집계 이전, 보정계수 적용을 통해 거주지(H) 정보를 최대한 살려 데이터 소실을 줄이는 효과 기대
- 3) 거주지 주민등록인구 기준 전수화 방안
- 주민등록인구 데이터를 통해 기존 KT 사용 고객 인원 중 지역별 거주지 기반 집계인구 데이터의 비율을 통해 전수화 계수를 반영
 - 원시데이터는 고객 식별번호, 가상기지국, 체류시작시간, 체류종료시간, 연령대, 성별, 통계청 집계인구 등의 정보를 이용
 - 성별, 연령대, 지역의 세부 속성을 모두 반영한 전수화 방법으로 기존 생활인구 기반 전수화 방안의 대표성 문제를 해결

- 집계인구 기반 전수화로 기존 O 기준 거주지(H)에 따른 전수화 방안의 한계점인 전체 - 통행 DB 중 약 10%에만 적용되는 한계점도 극복
 - o O 기준 거주지(H)에 따른 전수화 계수 적용 방법에서 집계된 후 전수화 보정 값을 반영하였다면, 집계인구 기반 전수화 방안에서는 집계 이전 개인에게 보정계수 값을 반영 후 집계하는 방안
 - 해당 방안을 통해, 거주지(H) 정보에 대한 데이터 소실을 최소화하여 데이터 정확도 향상

다. 거주지 주민등록인구 기준 전수화 방법론의 적용

1) 전수화 방법론 적용

- o 전수화 계수 추출 방법은 아래 순서에 따라 산출
 - Step-1: 전수화 계수 생성을 위한 주민등록인구 집계(2022년도) 데이터 수집
 - Step-2: KT 통신데이터에 대하여 주민등록인구와 동년도 시군구(거주지)/성별/연령 대별의 인구 집계
 - Step-3: 주민등록인구, KT 통신데이터의 집계 값의 비율을 통한 전수화 계수 값 산출
 - 시군구/성별/연령대별 (주민등록인구 / KT 통신데이터 집계 값)을 통해 전수화 계수 값 산출
 - Step-4: 개인 O/D 정보 중 거주지(H) 식별 및 해당 시군구 지역을 개인의 거주지 역으로 선정
 - Step-5: 개인별 거주지(H)를 기준으로 개인의 모든 개별통행 OD에 전수화 계수 값 반영

< 개별 통행 OD 데이터 >

기준일	개인ID	출발				도착				성별	연령대	통행수단	통행시간	전수화 개수	통행량	전수화 반영 통행량
		요일	시간대	공간	유형	요일	시간대	공간	유형							
20201221	7AAD6431	2	10	A	집	2	11	B	학교	F	10	F	77	3.25	1	3.25
20201221	7AAD6431	2	15	B	학교	2	16	A	집	F	10	F	81	3.25	1	3.25
20201221	6AAK73FH	2	09	B	집	2	10	E	직장	M	30	E	68	2.78	1	2.78
20201221	6AAK73FH	2	18	E	직장	2	19	D	기타	M	30	E	101	2.78	1	2.78
20201221	6AAK73FH	2	21	D	기타	2	23	B	집	M	30	E	152	2.78	1	2.78
20201221	Q3R3GEG3	2	16	E	직장	2	18	G	집	M	50	T	133	2.69	1	2.69
20201221	31GDEAGA	2	16	D	기타	2	19	A	집	F	20	B	211	3.11	1	3.11
20201221	F61233V1	2	19	C	집	2	20	D	기타	F	30	B	114	2.75	1	2.75

<그림 2-28> 개별 통행 OD 데이터 결과 그림

- Step-6: 통행 DB 기준에 따른 500m 단위의 격자 단위로 데이터로 집계
 - 500m 단위의 격자 집계 시, 전수화 반영 후 총통행량 값은 개인별 전수화 보정 후 총통행량 값들의 합산 값으로 집계

2) 전수화 결과

- 전수화 목표 대상을 기준으로 보다 정확한 결과 도출 가능
- 또한, 성별, 연령, 지역 등 세부 속성 정보 누락 없이 반영 가능
- 단순 시장점유율 사용 시 확인하기 어려웠던 속성별 통행량 반영 가능

기준일	개인ID	출발				도착				성별	연령대	통행수단	통행시간	전수화 개수	통행량	전수화 반영 통행량
		요일	시간대	공간	유형	요일	시간대	공간	유형							
20201221	7AAD6431	2	10	A	집	2	11	B	직장	F	30	T	70	3.25	1	3.25
20201221	6AF123AD	2	10	A	집	2	11	B	직장	F	30	T	68	3.25	1	3.25
20201221	F61233V1	2	10	C	집	2	13	D	기타	F	20	F	114	2.75	1	2.75

기준일	출발				도착				성별	연령대	통행수단	평균통행 시간	총통행량	전수화 반영 총통행량
	요일	시간대	공간	유형	요일	시간대	공간	유형						
20201221	2	10	A	집	2	11	B	직장	F	30	T	69	2	6.50
20201221	2	10	B	집	2	13	C	학교	M	10	F	172	3	9.24
20201221	2	10	C	집	2	13	D	기타	F	20	B	114	2	5.50

<그림 2-29> 집계인구 기반에 따른 전수화 적용 안

제4절 데이터 가공 프로세싱 최적화 및 속도 개선

1. 데이터 가공 프로세싱 최적화 및 속도 개선 방안

1) 데이터베이스 최적화 방안

- 데이터베이스는 점차적으로 증가하는 대용량데이터를 고려하여, 안정적이고 효율적으로 관리 할 수 있도록 대용량 데이터베이스 설계
- 모바일 통신 빅데이터의 특성을 고려하여 확장 가능한 설계, 다양한 유형별 자료를 효과적으로 구축할 수 있는 구조로 설계
- 효율적 데이터 처리 방식을 이용하여 디스크 자원을 최소한으로 줄이며, 데이터의 액세스를 분석하여 DB 성능 개선 전략을 수립함
- 데이터베이스 설계 방법론에 입각한 현행 시스템 및 데이터를 분석하고 표준화 방안 등을 마련하여 시스템에서 요구하는 데이터 분석이 가능한 구조로 설계

2) 알고리즘 최적화 방안

- 모바일 통신 빅데이터 가공알고리즘은 대용량 원시 LTE 로그데이터를 사용하여 거주지 및 근무지 추정하고 해당 위치에서의 체류시간을 기반으로 통근시간 및 통행량 등 통행지표 산출하는 과정으로 진행됨
- 원시데이터 전처리 과정에서는 기지국과 단말기 간에 발생하는 불규칙적인 신호 이상 현상을 보정해야 하고, 통행자의 이동궤적을 식별하는 단계에서는 비현실적인 데이터를 시·공간적으로 보정해야 함
- 따라서 효율적인 데이터 산출을 위해서 데이터의 경량화, 알고리즘의 고도화 및 분산 처리 등 데이터 생산 효율을 높이는 것이 주요 이슈임

2. 가공알고리즘의 단계별 세부 개선 내용

- 속도 개선을 위한 단계별 세부 알고리즘 최적화 및 수정내용은 다음과 같음

<표 2-5> 개선 알고리즘 단계별 코드 구성 내용

구분	순번	내용
Stage 1. 데이터 전처리 및 통행자 실제 위치 추정	1	5분 단위 그룹을 만들어 좌표 가중평균화
	2	도보 속도 5km 이하 좌표, 다음 좌표와 체류시간 가중치로 보정 코드 최적화
	3	도보거리 416.7m 이하 좌표, 다음 좌표와 체류시간 가중치로 보정 코드 최적화
Stage 2. 주체류지 식별	4	일주일 기준 최장체류지 1, 2 추출(1차적으로 1을 집으로 정의) 코드 최적화
	5	최장체류지 1, 2와 좌표 간 거리 연산 후 500m 이하 좌표는 최장체류지 1,2로 변경(DBSCAN 개념 차용)
Stage 3. 통행시간 및 통행량 산출	6	30분 이하 체류지는 제외(체류지 시간 최소기준 조작적 정의)
	7	행 압축
	8	최장체류지 1, 2 기준 OD속성 부여 코드 최적화

<표 2-6> 전체 프로세스별 소요시간

순번	항목	내용	소요시간 (일주일 기준)	1년치 소요 시간	비고
1	KOTI_RAW_FIN AL.hql	원시데이터 수집 및 저장	0.8일	41.6일	병렬처리 적용하지 않은 소요시간
2	FINAL_RUN,r	중간 산출물	2.8일	146일	병렬 수행한 결과이며 시간 계산 여유있게 한 결과
3		최종 산출물	3.2시간	7일	실제 수행 시간은 위의 중간산출물 시간에 포함 됨.
4	MASKING,r	개인 식별 위험 데이터 마스킹 처리	30분 미만	21시간 미만	병렬처리 적용하지 않은 소요시간
5	UPLOAD	데이터 반출을 위한 업로드	2.4시간	5.2일	대략적인 시간 기준

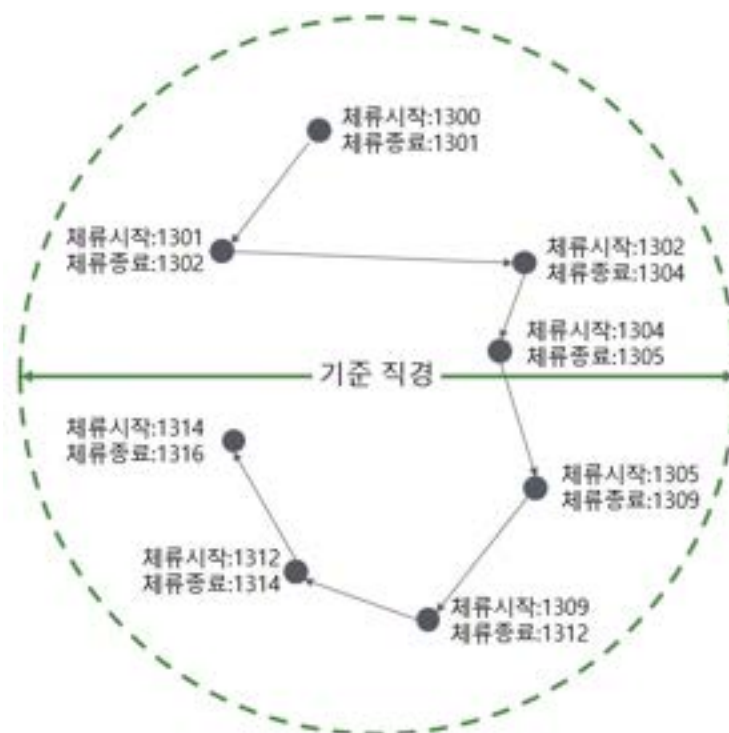
<표 2-7> 산출물 상세 소요시간

순번	내용	1년치 소요 시간	비고
1	주2회 3시간 동일시간대 조건 알고리즘에 사용되는 함수	8.5시간	
2	원시데이터 raw data 필드 맞추기	9.8일	
3	CSID	7시간	
4	고객 1명 7일치 데이터 추출	1.8일	
5	원시데이터를 1분단위로 펼친 데이터(dfs)	1.7일	
6	5분단위 time slice로 데이터를 나누고 각 time slice에서 대표위치 찾은 데이터(dfm)	6.6일	시간보정
7	이웃한 행간 속도 측정 후 5km/h 미만이면 체류시간 기준으로 위치보정	10.5일	속도보정
8	거리보정 (416m)	27.5일	공간보정
9	원시데이터에 대해 시공간적으로 보정+행압축된 데이터(a_1)	4.2일	행압축
10	a_1을 데이터테이블에서 데이터프레임으로 데이터 형식 변경	2일	
11	최장체류지1 찾기	2.5일	
12	변수 이름을 df2에서 a_1으로 다시 변경	1일	
13	max gap time을 찾기 위한 코드	11일	find_max_gap : 일별로 max gap을 찾기 위해 최장체류지가 아닌 체류지에 대한 체류시작시간, 체류종료시간이 기록됨
14	일자별 데이터에 대해 트립 타입 (집-회사-집, 집-회사, 회사-집 등)을 구분하고 max gap에 해당하는 index를 not_h라고 지정	4시간	
15	위와 같은 과정의 결과로 h_a_1과 c_a_1이 생성되며, c_a_1은 일주일동안 max gap에 해당하는 데이터들의 모음	5시간	
16	최장체류지 2를 찾기 위한 과정	3.5일	최장체류지1을 찾기 위한 과정과 동일, input 데이터만 c_a_1으로 바뀜
17	new_a_1 데이터프레임에서 최장체류지 1에 해당하는 데이터를 home='YES', 최장체류지 2에 해당하는 데이터를 home='NO'로 표시	16시간	
18	일별로 가장먼저 기록된 두번째 최장체류지 로그부터 가장 마지막에 기록된 로그까지 최장체류지 2로 지정해주는 코드	1.7일	
19	일주일에 대해 일기록빈도를 기준으로 회사와	2일	기록빈도가 더 많은게 집,

	집을 식별하는 코드		기록빈도 같은경우 체류시간이 더 긴게 집
20	회사와 집 식별 후, 회사에 대해 주2회 3시간(동일시간대) 미만 체류시 회사가 없는 것으로 지정하는 코드	8.5일	
21	출퇴근 시간 및 잠재체류를 계산하기 위한 코드	36일	NEW_HC_Df 가 고객별 산출물에 해당하는 데이터프레임
22	트립타입 구분 코드	14시간	트립 타입1은 집-회사-집 트립타입2는 집-회사 트립타입3은 회사-집
23	1인에 대한 출·퇴근 시간 및 회사 집 위치 정보가 포함된 산출물	9시간	
24	모든고객(iii)에 대한 산출물	2시간	
25	회사나 학교로 이동하는 고객에 대해 출장 통행으로 추정되는 로그 식별 (출근 횟수 2회 미만 & 퇴근 횟수 2회 미만)	1.7일	
26	잠재체류 있는 경우 통행시간 산출에서 제외	3.8일	잠재체류 식별기준: 출퇴근 중 30분 이상 체류한 기록이 있음
27	회사(목적지)-집 거리 구하기	1.5일	
28	출근 속도 및 퇴근 속도 구하기	1시간 미만	시간에 30초 더해주기
29	출근 속도나 퇴근 속도가 5km/h 미만인 경우 출퇴근 시간을 NA 처리	2.5일	
30	출근 속도나 퇴근 속도가 200km/h 초과인 경우 출퇴근 시간을 NA 처리	2.6일	
31	통행량의 경우, 행별로 트립타입이 집-회사-집인 경우는 2, 집-회사 또는 회사-집인 경우는 1로 계산하면 됨	3.3일	HC개수, CH개수 count
32	잠재 통행량의 경우, 행별로 출퇴근 모두 잠재체류가 있으면 2, 출근 또는 퇴근 중 잠재체류가 하나 있으면 1, 둘 다 없으면 0	4일	

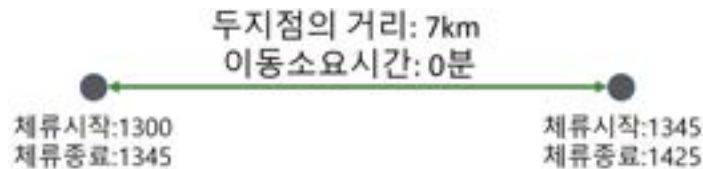
3. 데이터 가공 속도 개선 및 오류 정보 처리

- C++ 언어 변환
 - Rcpp 라이브러리 활용
 - R에서 일부 원 데이터 가공 및 체류지 오류 부분은 C++ 언어로 변환하여 가공 속도 향상 가능
- 숨겨진 체류지 탐색 방안
 - 기준 체류시간미만 기록의 클러스터
 - 기준 직경 내 기준 체류시간을 초과하였을 경우 직경의 중점을 체류지로 지정
 - <그림 2-30>에 보이는 바와 같이 기준 직경 내 체류시간이 16분일 경우, 해당 직경의 중심을 체류지로 설정 (기준체류시간은 15분 가정)



<그림 2-30> 기준직경 내 기준체류시간 초과 예시

- 원데이터 오류 수정 방안
 - 장기 체류시간 이동소요시간이 없는 데이터에 대한 보정
 - 체류시간만큼 가중치를 두어 두 체류 지점의 중점으로 변환 고려



<그림 2-31> 이동소요시간이 없는 장기체류지 예시

4. 가공데이터 용량 최적화

- 데이터 용량 최적화
 - JSON 데이터 형식
 - 가공을 통해 생성된 데이터 중 유심 정보, 성별, 나이, 날짜는 변하지 않는 정보
이므로 1번만 저장하여 데이터 용량 저감효과 기대

```
{
  "usim_no" : "tuitivukUIKQUIA==",
  "etl_ymd" : "20170408",
  "age_itg_cd" : "30",
  "sex_type_itg_cd" : "F",
  "xycrd" : ["YYY.YYYXX.XXX", "YYY.YYYXX.XXX", ...],
  "st_time_cd" : [0000, 1113, 1501, 2013, 2107],
  "fns_time_cd" : [1112, 1500, 2012, 2107, 2335],
  "sty_mnut" : [672, 228, 311, 55, 149]
}
```


제3장 모바일 교통 빅데이터 기반 통행사슬(Trip Chain) DB 구축 및 검증

제1절 기준년도 통행사슬(Trip Chain) DB 구축

제2절 기준년도 구축 DB 검증

제3장 모바일 교통 빅데이터 기반 통행사슬(Trip Chain) DB 구축 및 검증

제1절 기준년도 통행사슬(Trip Chain) DB 구축

1. DB 설계

- 「개인정보보호법」과 「위치정보의 보호 및 이용 등에 관한 법률」에 따라 개인의 위치와 이동궤적을 추적할 수 없도록 기반 DB는 다음 <표 3-1>과 같이 집계된 형태로 구축하고자 함

<표 3-1> 모바일 기반 DB 형태

출발						도착				
날짜	요일	시간	폴리곤	시도	체류지 유형	날짜	시간	폴리곤	시도	체류지 유형
20180402	월	13	48521351	11	집	20180402	14	65421584	11	회사
20180403	화	06	54754213	21	집	20180403	08	32158421	11	회사
20180404	수	09	15486211	33	집	20180404	09	15486213	33	학교

총통행량	보정계수	성별	연령	잠재통행 고려 안함								
				통행량(비율)				시간			거리 평균	속도 평균
				통행 비율	30분 이상	60분 이상	120분 이상	평균	분산	중앙 값		
10	0.34	M	20	1	0.7	0.5	0.4	35	6.98	26	8.24	29.8
25	0.57	F	50	0.8	0.9	0.8	0.2	33	6.77	24	9.45	11.9
8	0.47	M	10	1	0.9	0.6	0.1	25	4.32	19	10.5	23.5

잠재통행 30분 고려 안함										
통행량(비율)					시간			거리 평균	속도 평균	
통행비율	잠재체류 횟수	30분 이상	60분 이상	120분 이상	평균	분산	중앙값			
0.8	0	0.8	0.6	0.8	35	6.98	26	5.89	36.9	
0.7	3	0.7	0.5	0.2	-	-	-	9.57	10.4	
0.8	2	0.8	0.5	0.1	-	-	-	10.12	20.3	

잠재통행 60분 고려 안함									
통행량(비율)					시간			거리 평균	속도 평균
통행비율	잠재체류 횟수	30분 이상	60분 이상	120분 이상	평균	분산	중앙값		
0.9	0	0.7	0.6	0.5	35	6.98	26		33.6
0.8	2	0.8	0.6	0.2	-	-	-		10.9
0.9	0	0.9	0.6	0.1	25	4.32	19		22.8

주 *) 요일: 출발 및 도착 요일 (1~7, 일~월)

주 **) 유형: 개인별 출발지 및 도착지 유형 (H: 집, C: 회사, S: 학교, R: 정기 잠재체류지, X: 기타 잠재체류지)

- 개인의 통행 정보가 노출되지 않도록 기점 정보(출발 일자, 출발 시간대, 출발 위치, 출발 체류지 유형), 종점 정보(도착 일자, 도착 시간대, 도착 위치, 도착 체류지 유형), 통행자 정보(성, 연령)를 기준으로 통행 정보가 같은 인구를 집계
 - 일자 정보는 1일 단위로 구축하도록 함
 - 출발시간 및 도착시간 정보는 집계된 인구가 3명 이하가 될 가능성을 줄이면서, DB 사용자에게 가능한 섬세한 통행 정보를 제공할 수 있도록 한 시간 단위로 입력하도록 함
- * 출발시간은 출발지에서의 체류종료시간을, 도착시간은 도착지에서의 체류시작시간을 의미
 - 출발지와 도착지 위치는 「위치정보의 보호 및 이용 등에 관한 법률」에 저촉되지 않도록 기지국 좌표 그대로 노출하지 않고 매칭되는 500m 격자ID로 변환하도록 함
 - 출발 및 도착 체류지 유형은 평소 통행자가 해당 위치에 얼마나 자주, 오래 머무르는지를 추정할 수 있는 정보를 제공하기 위한 것으로, 「2019년 국가교통조사」에서 설정한 체류지 식별 기준에 따라 출발지(출발 기지국 좌표)와 도착지(도착 기지국 좌표)의 특성을 확인한 후 각각 체류지 유형을 구분하여 입력하도록 함
- * 「2019년 국가교통조사」에서 구분하고 있는 체류지 유형 중에서 집 이외 심야 시간대 주체류지는 제외하고, 총 5가지 유형으로 구분 (집, 회사, 학교, 정기적 잠재체류지, 비정기적 잠재체류지)
 - 통행자에 대한 정보를 확인할 수 있도록 성, 연령 필드 구성하고, 통행 정보가 지나치게 세분되지 않도록 가능한 통행 특성이 유사한 계층을 묶어 10세 단위로 연령을 구분하여 입력하도록 함
 - 통행량 필드에는 전술한 기준에 의해 집계된 인구(단말기 수)를 입력하되, 집계된 통행량이 3 이하의 값을 갖는 경우, 2장에서 개발한 3통행 미만 보정계수를 적용하여 추정된 통행량을 입력하도록 함

- 새로운 지표 개발을 위해 통행시간 필드를 추가
 - 통행시간 필드에는 집계된 통행 정보를 기준으로 산출한 평균 통행시간 값을 입력

<표 3-2> 모바일 통신 빅데이터 DB 테이블 정의서

컬럼 ID	컬럼명	Type	비고
o_date	출발 일자	string	yyyymmdd
o_dow	출발 요일	string	0 ~ 6 : 일요일 ~ 토요일
o_time	출발 시간대	string	00~23 (1시간 단위)
o_polycode	출발 폴리곤ID	string	500미터 격자단위
o_type	출발 체류지 유형	string	H: 집, C: 회사, S: 학교, R: 정기 잠재체류지, X: 기타 잠재체류지
d_date	도착 일자	string	yyyymmdd
d_time	도착 시간대	string	00~23 (1시간 단위)
d_polycode	도착 폴리곤ID	string	500미터 격자단위
d_type	도착 체류지 유형	string	H: 집, C: 회사, S: 학교, R: 정기 잠재체류지, X: 기타 잠재체류지
age	연령	integer	0~110 (예: 10: 10세 이상 20세 미만)
gender	성	string	M: 남성, F: 여성
traffic_all	총 통행량	integer	횟수 단위
traffic_time_all_mean	통행시간_평균	integer	1분 단위
traffic_distance_all_mean	통행거리_평균	integer	1분 단위
traffic_speed_all_mean	통행속도_평균	integer	km/h 단위
traffic_gb	교통수단	string	E : 기타, F: 항공, T: 고속철도, B: 고속버스

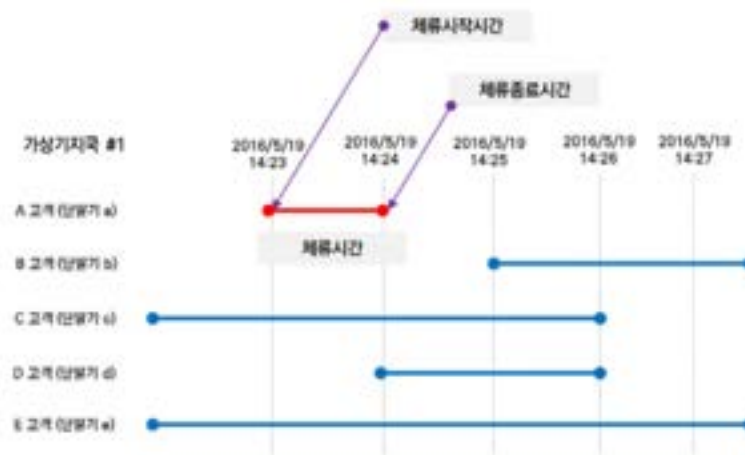
3. DB 구축 방법

1) 원시데이터 수집 및 이상치 제거

- 2022년 1월 1일부터 2022년 12월 31일까지 기록된 로그를 수집
- 기지국 좌표가 Null 값이거나 행정동 정보가 매칭되지 않는 데이터를 제거

2) 선분 이력으로 변환

- 포인트 단위의(밀리세컨 단위) 로그를 선분 단위의(분 단위) 로그로 변환
 - 닷(dot) 형태로 기록된 로그를 기록된 시간 순서대로 연결한 후, 최초 기록된 시간(이하 ‘체류시작시간’)과 마지막으로 기록된 시간(이하 ‘체류종료시간’)정보만 추출하여 단말기가 해당 기지국에 식별된 시간(이하 ‘체류시간’)을 산출
 - 체류시간은 체류종료시간에서 체류시작시간을 뺀 값이며, 초(sec) 단위는 생략



주: 자료를 참고하여 도식화 함

자료: 박미울·주은정(2017), LTE 시그널 정보를 이용한 위치정보가공 및 유동인구집계 방법, 한국통신학회 2017년도 동계종합학술발표회, p.285.

<그림 3-1> 선분 이력으로의 변환 예시

3) 통행 정보 왜곡 데이터 보완

- 데이터 보완 범위를 한정하기 위해 로그 기록을 한 달 단위로 구분
 - 통행 정보 왜곡 데이터를 탐색하고 보완하는 범위를 1개월로 한정 (예: 4월 1일부

터 4월 30일까지의 로그 기록을 연결하여 탐색하고 보완)

- 통행 정보를 왜곡할 수 있는 ‘필요 이상의 핸드오버 데이터(Unnecessary Handover)’, ‘핑퐁 핸드오버(Ping-pong Handover)로 인한 데이터’를 본 과업에서 개발한 전처리 기술을 통해 보정함

4) 체류 순서 부여

- 로그 기록 일자, 체류시작시간, 체류종료시간을 기준으로 로그 기록 순서에 따라 개인별 데이터를 정렬하여 체류 순서를 구분
 - 고객식별번호(단말기 구분 코드)와 데이터 생성일자를 기준으로 KEY 값을 갖는 필드를 형성한 후(‘기준일-고객식별번호’), 같은 KEY 값을 갖는 데이터 내에서 정렬

ID	그룹 번호	KEY 칼럼 (기준일- 고객식별번호)	가상기지국		체류 시작 시간	체류 종료 시간	연 령 대	체류 시간
			x좌표	y좌표				
1	1	20160520-17232311	126.914	37.545	1400	1750	33	240
2	2	20160520-17232311	126.914	37.549	1751	1752	33	1
3	3	20160520-17232311	126.916	37.548	1754	1809	33	15
4	4	20160520-17232311	126.914	37.549	1810	1813	33	3
5	5	20160520-17232311	126.915	37.551	1813	1814	33	1
6	6	20160520-17232311	126.914	37.549	1814	1815	33	1
		⋮	⋮	⋮				
18	18	20160520-17232311	126.815	37.538	1920	2359	33	279
		⋮	⋮	⋮				
453	1	20160520-1843030	126.934	37.556	0805	1041	25	156
454	2	20160520-1843030	126.936	37.556	1041	1042	25	1
455	3	20160520-1843030	126.935	37.554	1043	1044	25	1
456	4	20160520-1843030	126.933	37.552	1045	1050	25	5
457	5	20160520-1843030	126.929	37.547	1051	1052	25	1
458	6	20160520-1843030	126.923	37.547	1053	1055	25	2
459	7	20160520-1843030	126.918	37.547	1056	1059	25	3
460	8	20160520-1843030	126.914	37.545	1100	1200	25	60
		⋮	⋮	⋮				
679	227	20160522-1843030	126.934	37.556	0803	0830	25	507
680	228	20160522-1843030	126.936	37.556	0831	0833	25	2
		⋮	⋮	⋮				
711	259	20160522-1843030	126.815	37.551	0850	1100	25	130

<그림 3-2> 체류 순서 부여 (개인별 이동계적 형성)

5) 체류 정보 추출

- 이동 중에 기록된 로그 기록을 제외하고 체류 중에 기록된 로그 기록만을 추출
 - 체류시간이 15분을 초과하는 로그 기록만 추출
 - 트립 타입 필드를 추가하여 체류시간이 15분 이하인 경우는 ‘경로(Pass-by)’로, 15분 초과인 경우는 ‘체류(Stay)’로 구분

ID	그룹 번호	KEY 칼럼 (기준일- 고객식별번호)	가상기지국		체류 시작 시간	체류 종료 시간	연 령 대	체류 시간	트립 타입
			x좌표	y좌표					
1	1	20160520-17232311	126,914	37,545	1400	1750	33	240	체류
2	2	20160520-17232311	126,914	37,549	1751	1752	33	1	경로
3	3	20160520-17232311	126,916	37,548	1754	1809	33	15	경로
4	4	20160520-17232311	126,914	37,549	1810	1813	33	3	경로
5	5	20160520-17232311	126,915	37,551	1813	1814	33	1	경로
6	6	20160520-17232311	126,914	37,549	1814	1815	33	1	경로
		⋮	⋮	⋮					
18	18	20160520-17232311	126,815	37,538	1920	2359	33	279	체류
		⋮	⋮	⋮					
453	1	20160520-1843030	126,934	37,556	0805	1041	25	156	체류
454	2	20160520-1843030	126,936	37,556	1041	1042	25	1	경로
455	3	20160520-1843030	126,935	37,554	1043	1044	25	1	경로
456	4	20160520-1843030	126,933	37,552	1045	1050	25	5	경로
457	5	20160520-1843030	126,929	37,547	1051	1052	25	1	경로
458	6	20160520-1843030	126,923	37,547	1053	1055	25	2	경로
459	7	20160520-1843030	126,918	37,547	1056	1059	25	3	경로
460	8	20160520-1843030	126,914	37,545	1100	1200	25	60	체류
		⋮	⋮	⋮					
679	227	20160522-1843030	126,934	37,556	0803	0830	25	507	체류
680	228	20160522-1843030	126,936	37,556	0831	0833	25	2	경로
		⋮	⋮	⋮					
711	259	20160522-1843030	126,815	37,561	0850	1100	25	130	체류

<그림 3-3> 이동정보와 체류 정보 구분 (예시)

6) 체류지 유형 구분

- 체류지 식별 기준(〈표 3-3〉 참조)에 따라 체류지 유형을 구분하여 필드 추가

<표 3-3> 체류지 식별 기준

체류지 유형		식별 기준			
		체류특성		통행자 연령	비고
		체류 시간	체류 빈도		
주 체류지	집	알고리즘 기준 최장 체류지 후보 1, 2 중	일주일 기준 최다빈도 체류지	-	알고리즘 기준 제1의 주체류지
	회사	동시간대 3시간 이상	주 2회 이상	30세 이상 60세 미만	알고리즘 기준 제2의 주체류지 중
	학교	동시간대 3시간 이상	주 2회 이상	20세 미만	알고리즘 기준 제2의 주체류지 중
기타 체류지	정기적	3시간 이상	주 2회 이상	-	알고리즘 기준 제 2의 주체류지 중 회사나 학교로 구분되지 않은 체류지 중
	비정기적	3시간 이상	주 2회 미만	-	

ID	그룹 번호	KEY 필명 (기준일- 고객식별번호)	가상기차국		체류 시작 시간	체류 종료 시간	연 행 대	체류 시간	트립 타입	체류지 유형
			X좌표	Y좌표						
1	1	20160520-17232311	126.914	37.545	1400	1750	33	240	체류	집
2	2	20160520-17232311	126.914	37.549	1751	1752	33	1	경로	-
3	3	20160520-17232311	126.915	37.548	1754	1809	33	15	경로	-
4	4	20160520-17232311	126.914	37.549	1810	1813	33	3	경로	-
5	5	20160520-17232311	126.915	37.551	1813	1814	33	1	경로	-
6	6	20160520-17232311	126.914	37.549	1814	1815	33	1	경로	-
18	18	20160520-17232311	126.815	37.538	1920	2359	33	279	체류	회사
453	1	20160520-184000	126.934	37.556	0805	1041	25	156	체류	집
454	2	20160520-184000	126.935	37.556	1041	1042	25	1	경로	-
455	3	20160520-184000	126.935	37.554	1043	1044	25	1	경로	-
456	4	20160520-184000	126.933	37.552	1045	1050	25	5	경로	-
457	5	20160520-184000	126.929	37.547	1051	1052	25	1	경로	-
458	6	20160520-184000	126.923	37.547	1053	1055	25	2	경로	-
459	7	20160520-184000	126.918	37.547	1056	1059	25	3	경로	-
460	8	20160520-184000	126.914	37.545	1100	1200	25	60	체류	기타

<그림 3-4> 체류지 유형 구분 (예시)

7) 위치정보, 시간정보, 연령정보 변환

- 개인의 이동궤적을 추적할 수 없도록 원시데이터에서 출발지와 도착지의 위치정보를 분석맵의 교통폴리곤 ID로 변경한 후, 시간정보(체류시작시간, 체류종료시간)에서 분 정보를 생략한 시간대 필드를 생성하고, 1세 단위의 연령 정보를 10세 단위로 변경함

ID	그룹 번호	개인 ID (가운잡- 고려사항번호)	가상거주지		체류 시작 시간	체류 종료 시간	연 령	체류 시간	통행 목적	체류지 유형	위치 정보 변환 (교통폴리곤 ID)	체류시작 시간대	체류종료 시간대	연령대 변환
			X좌표	Y좌표										
1	1	209020-1722311	126.914	37.546	1400	1750	30	340	체류	집	209001	14	17	30
2	2	209020-1722311	126.914	37.546	1751	1752	30	1	경로	-	-	-	-	-
3	3	209020-1722311	126.916	37.548	1754	1800	30	46	경로	-	-	-	-	-
4	4	209020-1722311	126.914	37.546	1800	1813	30	13	경로	-	-	-	-	-
5	5	209020-1722311	126.915	37.551	1813	1814	30	1	경로	-	-	-	-	-
6	6	209020-1722311	126.914	37.549	1814	1815	30	1	경로	-	-	-	-	-
18	18	209020-1722311	126.815	37.509	1900	2300	30	200	체류	회사	209017	19	23	30
453	1	209020-184000	126.914	37.506	1000	1040	25	40	체류	집	209014	08	10	30
454	2	209020-184000	126.916	37.506	1041	1042	25	1	경로	-	-	-	-	-
455	3	209020-184000	126.916	37.504	1043	1044	25	1	경로	-	-	-	-	-
456	4	209020-184000	126.913	37.512	1045	1050	25	5	경로	-	-	-	-	-
457	5	209020-184000	126.915	37.547	1051	1052	25	1	경로	-	-	-	-	-
458	6	209020-184000	126.913	37.547	1053	1055	25	2	경로	-	-	-	-	-
459	7	209020-184000	126.914	37.547	1056	1059	25	3	경로	-	-	-	-	-
460	8	209020-184000	126.914	37.546	1100	1200	25	60	체류	가정	209015	11	12	30

<그림 3-5> 위치정보, 시간정보, 연령 정보 변환 (예시)

8) 통행량 집계

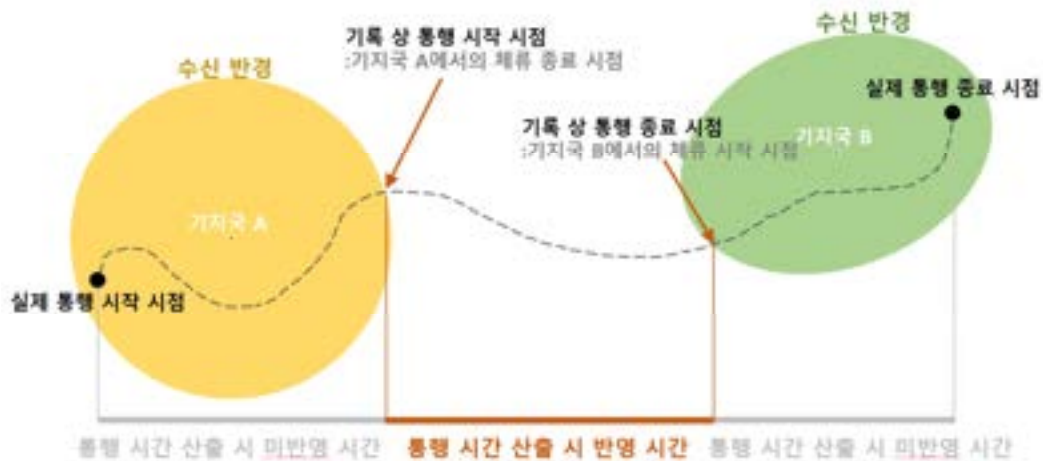
- 앞서 부여한 체류 순서에 따라 출발과 도착을 구분한 다음, 기점 정보(출발 일자, 출발 시간대, 출발 위치, 출발 체류지 유형), 종점 정보(도착 일자, 도착 시간대, 도착 위치, 도착 체류지 유형), 통행자 정보(성, 연령)가 동일한 로그 기록을 집계함
- 마지막 체류지가 출발 정보로 구분되고, 도착 정보가 없는 경우 이동하지 않은 것으로 보고 ‘무(無)통행’으로 간주함

ID	그룹 번호	체류 집계 (기점일- 종점일-종점 지)	가산기차		체류 시작 시간	체류 종료 시간	연 령 대	체류 시간	체류 유형	체류지 유형	위치 정보 변환 (교통플리콘 코드)	체류시작 시간대	체류종료 시간대	연령대 변환	
			X좌표	Y좌표											
1	1	2018020-1720311	08.944	37.540	1400	1750	30	240	체류	집	210007	14	17	30	출발
2	2	2018020-1720311	08.944	37.540	1750	1752	30	1	강도	-	-	-	-	-	1통행 ↓
3	3	2018020-1720311	08.944	37.540	1754	1800	30	15	강도	-	-	-	-	-	
4	4	2018020-1720311	08.944	37.540	1800	1803	30	3	강도	-	-	-	-	-	
5	5	2018020-1720311	08.944	37.540	1803	1804	30	1	강도	-	-	-	-	-	
6	6	2018020-1720311	08.944	37.540	1804	1805	30	1	강도	-	-	-	-	-	
18	18	2018020-1720311	08.944	37.540	1805	2000	30	279	체류	회사	200017	18	20	30	도착
453	1	2018020-1840000	08.944	37.540	0800	1041	25	100	체류	집	311014	08	10	30	출발
454	2	2018020-1840000	08.944	37.540	1041	1042	25	1	강도	-	-	-	-	-	1통행 ↓
455	3	2018020-1840000	08.944	37.540	1042	1044	25	1	강도	-	-	-	-	-	
456	4	2018020-1840000	08.944	37.540	1044	1045	25	5	강도	-	-	-	-	-	
457	5	2018020-1840000	08.944	37.540	1045	1047	25	1	강도	-	-	-	-	-	
458	6	2018020-1840000	08.944	37.540	1047	1048	25	2	강도	-	-	-	-	-	
459	7	2018020-1840000	08.944	37.540	1048	1049	25	3	강도	-	-	-	-	-	도착
460	8	2018020-1840000	08.944	37.540	1049	1050	25	60	체류	가게	312121	11	12	30	

<그림 3-6> 출발/도착 구분 및 통행량 집계 (예시)

9) 평균 통행시간 산출

- 모바일 기지국 데이터는 기지국 기반 로그 기록이므로 통행시간 산출 시, 기종점 부분에서 통행시간이 하향 계산되는 경우가 발생함



<그림 3-7> 기지국 기반 데이터의 통행시간 산출 방법

- 위와 같은 문제를 해결하고자, 기지국의 평균 수신범위를 계산하여 시군별 기지국 내 이동 시간 산출 후, 통행시간을 보정함
 - 시군구별 면적과 기지국 수를 사용하여 시군구별 기지국 수의 밀도 계산
 - 시군구별 기지국의 평균 면적을 계산
 - 기지국의 수신범위가 원이라고 가정하고 원주율(3.14)을 통해 기지국 수신범위의 평균 반경 계산
 - 시군구 내 통행의 평균 통행속도를 구한 뒤, 해당 속도로 시군구별 기지국 수신범위의 평균 반경을 통행하는 시간 산출
- 각 개인의 통행 정보에서 통행시간(도착시간-출발시간, 분 단위)을 산출한 다음, 기종점 정보(기지국 위·경도 좌표)와 시군구(행정구역)정보를 매칭해 기종점 지점에서 시군구별 기지국 내 이동 평균 시간을 합산하여 최종 통행시간 계산함
- 개별통행에 대해 기점 정보(출발 일자, 출발 시간대, 출발 위치, 출발 체류지 유형), 종점 정보(도착 일자, 도착 시간대, 도착 위치, 도착 체류지 유형), 통행자 정보(성, 연령)가 동일한 로그 기록의 통행시간을 집계한 후 집계한 통행시간을 산술평균하여 산출함

제2절 기준년도 구축 DB 검증

1. DB 검증 방법

- 기지국 기반의 데이터를 이용하여 구축된 모바일 통신 빅데이터 통행DB의 신뢰성 및 정확성을 검증하기 위하여 기존 개인통행실태조사자료(2016년 기준, 한국교통연구원)와 인구총조사자료(2015년 기준, 통계청)를 활용
- 기지국 기반의 데이터를 이용하여 모바일 통신 빅데이터 통행 DB와 기존 개인통행실태조사자료와 인구총조사자료의 통행량, 통행시간 등을 비교·분석하고, 단순 비교할 수 없는 모바일 통신 빅데이터 기반 상세지표(예를 들어, 상세 시간대별, 요일별, 성별, 연령별, 희망통근시간 등)는 통상적인 기대 통행패턴을 벗어나는지 확인
 - 지역별 통행 발생량·도착량을 개인통행실태조사자료와 상대적으로 비교하여, 유사한 패턴을 보이는지 확인
 - 시간대별, 요일별 통행 발생량·도착량을 확인하여, 통상적인 첨두시간 및 평일·주말 통행패턴이 나타나는지 확인
 - 성별, 연령별 통행 발생량·도착량을 확인하여, 통상적인 인구 비율 및 연령대별 통행 패턴이 나타나는지 확인
 - 지역별 출퇴근 목적통행량·통행시간을 개인통행실태조사자료, 인구총조사자료와 상대적으로 비교하여, 유사한 패턴을 보이는지 확인
 - 출퇴근 목적 통행시간·통행속도와 희망통근시간(60분)기준 통행패턴을 확인하여, 통상적인 통행패턴이 나타나는지 확인
- 단, 데이터 가공 시간 때문에 1년 치 모든 데이터를 가공하여 과업 기간 내에 검증할 수 없으므로, 통계적으로 충분히 유의한 10만 명의 표본 자료를 이용하여 검증을 시행함
 - 표본 유의성을 분석한 결과, 약 2만 명의 표본 이후 통행시간의 평균 오차는 ± 0.1 분 이하임

2. DB 검증 결과

가. 통행 발생·도착패턴 비교 및 분석 결과

1) 개인통행실태조사자료 기준 지역별 통행량 비교 결과

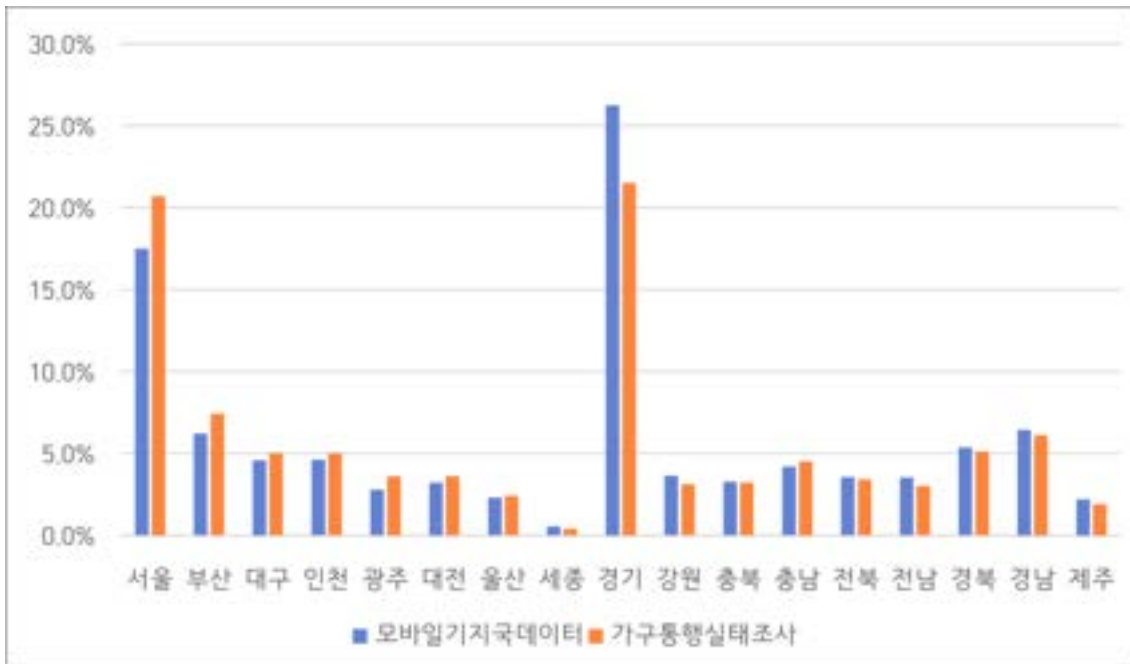
- 지역별 통행 발생량·도착량을 개인통행실태조사 자료와 상대적으로 비교하여, 유사한 패턴을 보이는지 확인

<표 3-4> 지역별 통행량 비교 결과

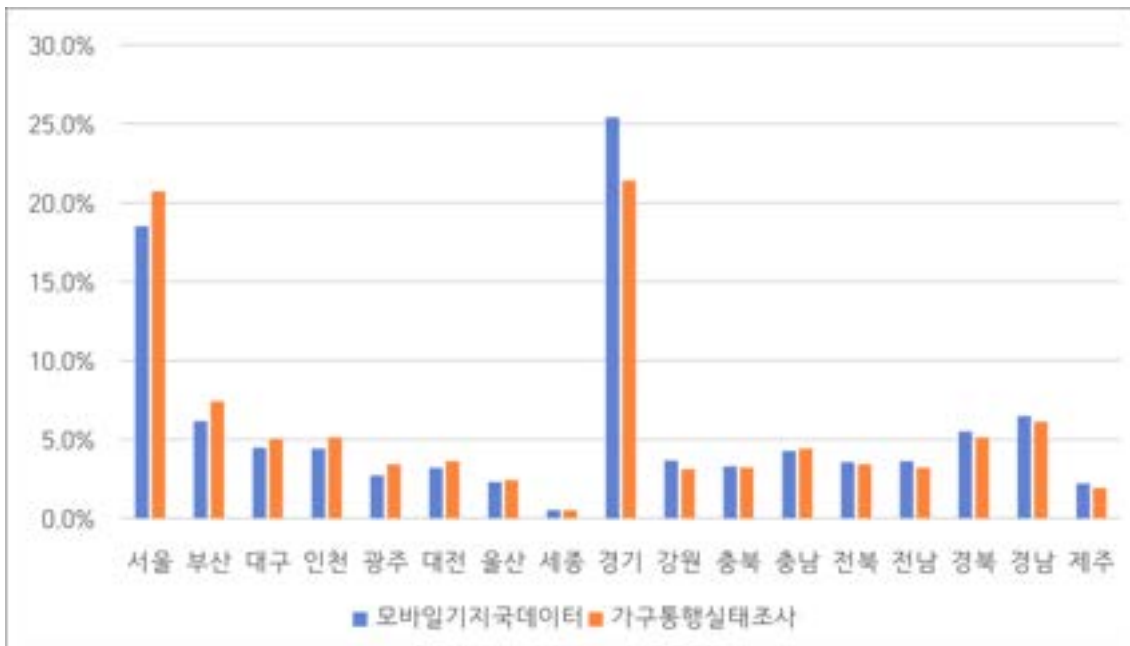
(단위: 통행/일 (비율%))

구분	개인통행실태조사자료 기준*		모바일 통신 빅데이터 기준	
	발생 (비율)	도착 (비율)	발생 (비율)	도착 (비율)
서울	17,996,813(20.7)	17,971,878(20.7)	3,532,837(17.5)	3,735,390(18.5)
부산	6,457,538(7.4)	6,458,293(7.4)	1,250,226(6.2)	1,241,939(6.1)
대구	4,333,058(5.0)	4,330,603(5.0)	925,685(4.6)	901,059(4.5)
인천	4,370,318(5.0)	4,433,014(5.1)	930,863(4.6)	887,980(4.4)
광주	3,117,712(3.6)	2,942,575(3.4)	560,475(2.8)	542,603(2.7)
대전	3,091,862(3.6)	3,101,487(3.6)	649,274(3.2)	640,985(3.2)
울산	2,117,733(2.4)	2,109,519(2.4)	460,616(2.3)	461,852(2.3)
세종	389,864(0.4)	391,571(0.5)	106,505(0.5)	100,256(0.5)
경기	18,668,530(21.5)	18,628,440(21.4)	5,303,762(26.3)	5,132,439(25.4)
강원	2,714,015(3.1)	2,715,277(3.1)	733,552(3.6)	737,190(3.6)
충북	2,803,473(3.2)	2,793,476(3.2)	658,793(3.3)	664,843(3.3)
충남	3,871,861(4.5)	3,863,535(4.4)	844,207(4.2)	859,032(4.3)
전북	2,942,438(3.4)	2,943,452(3.4)	713,754(3.5)	715,518(3.5)
전남	2,606,282(3.0)	2,783,178(3.2)	709,859(3.5)	727,290(3.6)
경북	4,431,201(5.1)	4,442,218(5.1)	1,080,468(5.3)	1,108,501(5.5)
경남	5,297,159(6.1)	5,300,768(6.1)	1,300,359(6.4)	1,304,949(6.5)
제주	1,659,446(1.9)	1,660,019(1.9)	443,411(2.2)	442,822(2.2)
계	86,869,303(100)	86,869,303(100)	20,204,646(100)	20,204,646(100)

주 : * 개인통행실태조사자료는 2016년 기준 승용차(택시), 버스, 철도(일반/고속/지하철), 항공, 해운수단에 대한 목적통행량임.



<그림 3-8> 지역별 통행 발생량 비교 결과



<그림 3-9> 지역별 통행 도착량 비교 결과

- 개인통행실태조사 자료 기준 1인당 일평균 통행량(발생+도착)은 약 3.4회이며, 모바일 통신 빅데이터 기준 1인당 일평균 통행량(발생+도착)은 약 3.1회임
 - 개인통행실태조사 자료 기준 전국 통행 발생·도착량은 1.74통행/일, 1.74통행/일이며, 모바일 통신 빅데이터 기준 전국 통행 발생·도착량은 1.5통행/일, 1.6통행/일임
 - 모바일 통신 빅데이터 기준 통행량과 개인통행실태조사 자료 기준 통행량 매우 유사하게 나타남
 - 하지만 개인통행실태조사 자료가 더 높은 통행량을 띄는데, 이러한 차이는 기존의 개인통행실태조사는 출근, 등교, 업무, 쇼핑, 귀가 등 모든 통행목적에 따라 통행량을 집계하지만, 모바일 통신 빅데이터는 체류시간 기준으로 체류지를 구분하기 때문에 기타잠재 체류지에 속하지 않은 이동이 집계되지 않기 때문에 통행량 차이가 과소 추정되었다고 판단됨
- 지역별 통행량 비율을 살펴보면, 평균적으로는 약 0.41%의 차이로 기존 개인통행실태조사 자료와 매우 유사한 패턴을 나타냄
 - 경기지역 통행 발생량이 약 +4.8%p로 가장 큰 폭으로 과대 추정되었으며, 서울지역 통행 발생량이 약 -3.2%p로 가장 큰 폭으로 과소 추정되었으나, 평균적으로 약 $\pm 1\%$ 로 매우 신뢰성 높은 결과를 도출함

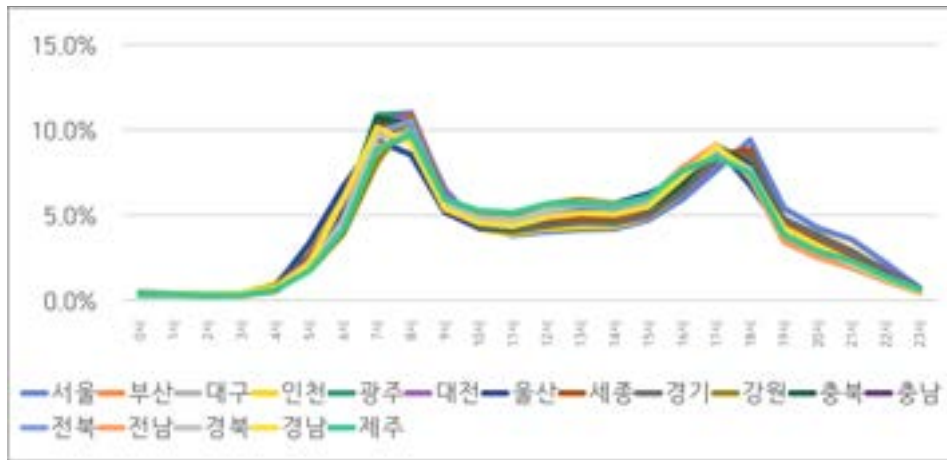
2) 시간대별 통행량 패턴 분석 결과

- 시간대별 통행 발생량·도착량을 확인하여, 통상적인 첨두시간 통행패턴이 나타나는지 확인함

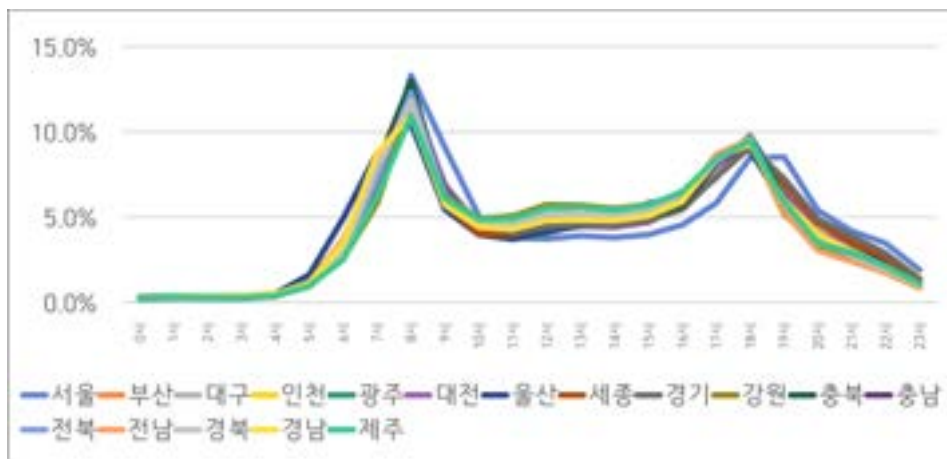
<표 3-5> 시간대별 통행량 패턴 분석 (모바일 통신 빅데이터 기준)

(단위: 통행/일)

구분		새벽		오전		오후		저녁	
		0시~	3시~	6시~	9시~	12시~	15시~	18시~	21시~
발 생	서울	14,713	62,117	450,561	265,127	223,509	329,622	344,294	117,025
	부산	7,015	22,808	147,014	88,259	84,371	123,306	95,707	31,030
	대구	3,586	12,248	113,505	64,921	57,712	85,166	71,511	21,696
	인천	4,908	21,288	126,798	66,854	62,618	94,077	82,771	26,203
	광주	2,216	7,984	65,993	36,063	32,942	48,390	37,714	11,907
	대전	2,747	8,593	71,545	41,647	38,272	53,975	46,496	14,825
	울산	2,604	9,603	51,767	28,359	31,885	47,415	28,890	9,962
	세종	562	1,608	14,955	8,251	8,298	12,028	9,588	2,510
	경기	25,175	97,580	650,711	368,543	346,531	502,740	437,698	134,742
	강원	3,024	8,442	70,794	50,194	53,486	67,878	44,431	11,928
	충북	2,549	9,037	78,751	42,615	45,990	64,001	46,494	11,836
	충남	3,299	13,530	97,354	54,552	61,527	84,367	56,104	14,248
	전북	2,708	8,520	76,206	45,321	48,378	65,411	43,671	12,008
	전남	2,741	10,238	74,055	45,373	50,544	69,225	39,734	10,621
	경북	4,089	13,231	116,703	73,106	80,426	107,629	72,492	19,169
	경남	6,888	19,777	152,574	86,006	90,682	131,294	90,144	25,199
	제주	1,667	4,696	40,805	29,364	30,440	39,609	25,530	7,835
도 착	서울	11,852	31,803	445,874	361,688	230,813	289,367	451,562	192,909
	부산	5,988	12,495	127,085	91,541	81,744	109,178	119,429	44,375
	대구	3,042	6,340	88,518	62,306	53,621	74,713	85,308	30,616
	인천	4,107	10,249	98,098	65,104	58,482	77,729	91,863	35,901
	광주	1,945	4,387	51,269	32,750	29,669	43,565	45,528	15,514
	대전	2,363	4,631	58,679	41,804	36,295	49,182	56,473	20,597
	울산	2,260	5,337	50,983	27,700	29,051	43,418	38,578	14,029
	세종	472	862	11,457	7,054	7,046	9,783	10,731	3,762
	경기	21,308	47,918	509,072	364,582	321,985	420,663	507,808	193,757
	강원	2,577	5,378	63,290	50,646	53,116	64,356	56,509	17,764
	충북	2,051	5,119	73,913	45,135	44,490	57,590	60,666	18,403
	충남	2,579	7,357	95,410	58,414	60,291	76,889	75,808	23,486
	전북	2,238	4,978	68,037	46,640	46,902	60,150	57,447	17,430
	전남	2,230	6,074	75,013	49,229	50,492	64,460	57,309	15,889
	경북	3,314	8,041	117,033	78,357	80,013	101,374	98,901	29,536
	경남	5,904	12,152	140,754	87,602	87,659	117,982	117,274	37,221
	제주	1,473	2,918	35,203	28,510	29,628	36,933	33,763	10,924



<그림 3-10> 시간대별 통행 발생량 패턴 분석 (모바일기지국데이터 기준)



<그림 3-11> 시간대별 통행 도착량 패턴 분석 (모바일 통신 빅데이터 기준)

- 모바일 통신 빅데이터 기준 시간대별 통행패턴 분석 결과, 모든 지역에서 전형적인 첨두시간 통행패턴을 보임
 - 오전 5시를 시작으로 통행량이 점점 증가하여 오전 9시에 감소하는 전형적인 오전 첨두시간 통행패턴이 나타나며, 오후 3시부터 점차 증가하여 오후 6시를 정점으로 오후 10시까지 이어지는 전형적인 오후 첨두시간 통행패턴을 나타냄
 - 퇴근 시간보다 출근 시간에 더 집중되는 통행패턴과 첨두시간 대비 약 1/3에 해당하는 비첨두 통행패턴 등 매우 신뢰성 있는 통행패턴 결과를 도출함
 - 지역별로도 매우 유사한 패턴을 보여, 지역별 편차 없이 신뢰성 있는 결과가 도출됨

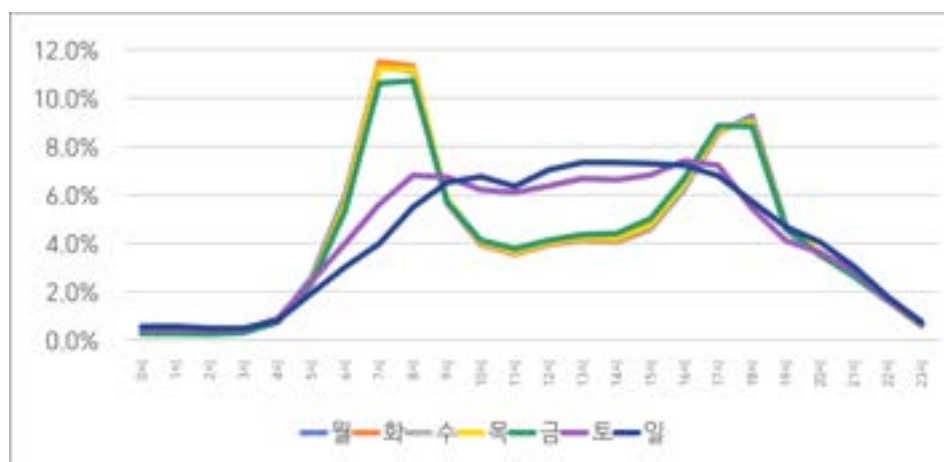
5) 요일별 시간대에 따른 통행량 패턴 분석 결과

- 요일별 통행 발생량·도착량을 확인하여, 통상적인 평일·주말 통행패턴이 나타나는지 확인함

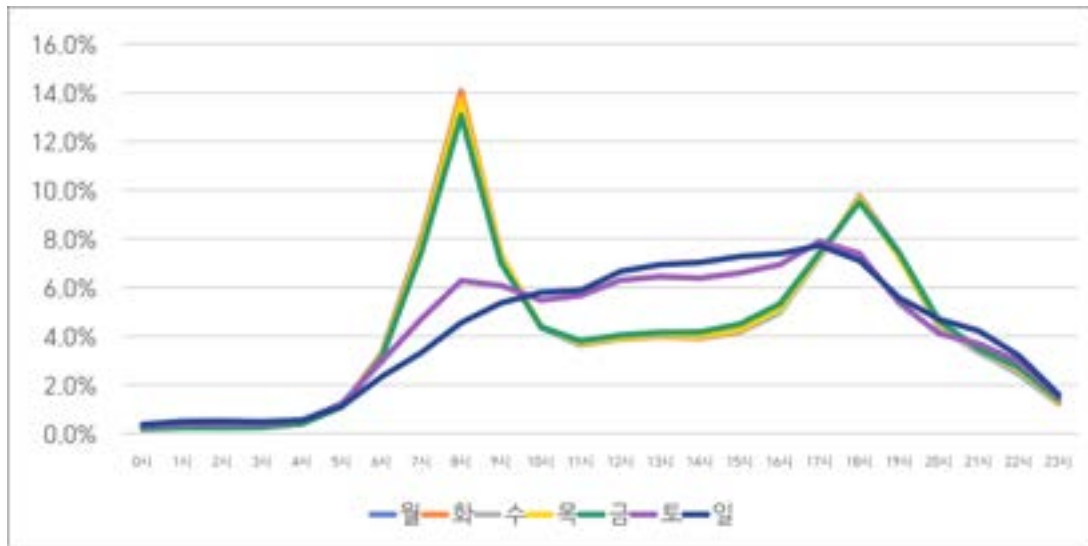
<표 3-6> 요일별 통행량 패턴 분석 (모바일 통신 빅데이터 기준)

(단위: 통행/일)

구분		새벽		오전		오후		저녁	
		0시~	3시~	6시~	9시~	12시~	15시~	18시~	21시~
발 생	월	11,332	54,318	426,274	196,993	181,399	290,245	262,036	72,326
	화	11,974	51,785	427,711	197,516	181,660	293,070	259,918	74,098
	수	12,475	52,084	424,160	202,786	188,089	301,009	258,281	75,287
	목	12,624	51,943	426,002	205,212	189,913	303,421	262,334	78,025
	금	12,921	50,798	401,217	204,853	193,887	309,317	255,199	75,097
	토	14,902	41,988	184,346	213,963	220,947	240,927	147,830	58,604
	일	14,262	28,385	110,381	173,232	191,717	188,142	127,670	49,305
도 착	월	9,526	26,195	379,483	228,794	176,434	244,941	324,200	105,352
	화	9,526	26,195	379,483	228,794	176,434	244,941	324,200	105,352
	수	10,386	26,944	374,849	233,209	182,333	252,514	324,729	109,207
	목	10,601	26,836	376,136	234,998	183,726	256,148	327,007	114,024
	금	10,801	26,581	354,735	228,204	185,818	259,324	323,716	114,111
	토	12,332	24,221	156,795	193,530	215,009	240,721	189,451	91,448
	일	12,026	18,837	89,993	150,273	181,898	197,453	153,178	79,434



<그림 3-12> 요일별 시간대에 따른 통행 발생량 변화 (모바일 통신 빅데이터 기준)



<그림 3-13> 요일별 시간대에 따른 통행 도착량 변화 (모바일 통신 빅데이터 기준)

- 모바일 통신 빅데이터 기준 요일별 통행패턴 분석 결과, 평일과 주말 기준 전형적인 통행패턴을 보임
 - 평일인 월~금요일은 매우 뚜렷한 평일 첨두시간 통행패턴을 보이며, 주말인 토요일~일요일은 낮 시간대 비교적 완만한 전형적인 주말 통행패턴을 보임으로써, 매우 신뢰성 있는 결과가 도출되었다고 판단됨
 - 평일 요일별 통행패턴은 큰 차이가 없으며, 일요일은 토요일에 비해 완만한 낮 시간대 통행패턴을 보임

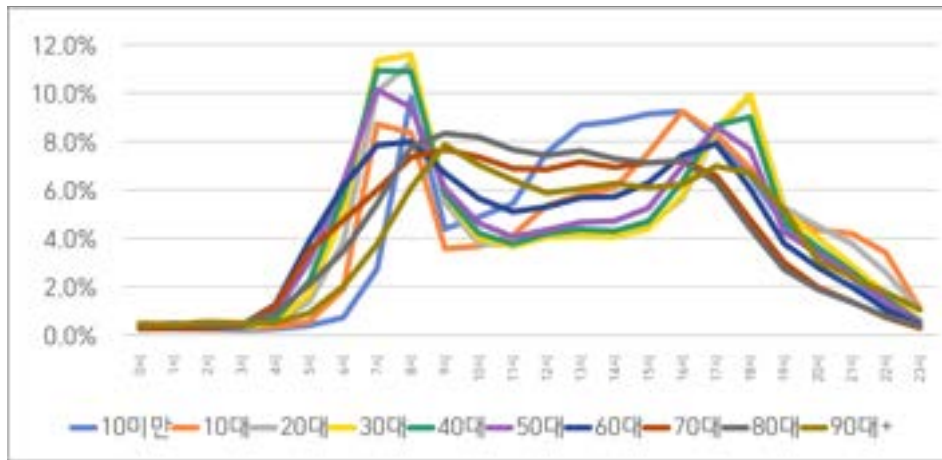
4) 연령별 통행량 패턴 분석 결과

- 연령별 통행 발생량·도착량을 확인하여, 통상적인 인구 비율 및 연령대별 통행패턴이 나타나는지 확인

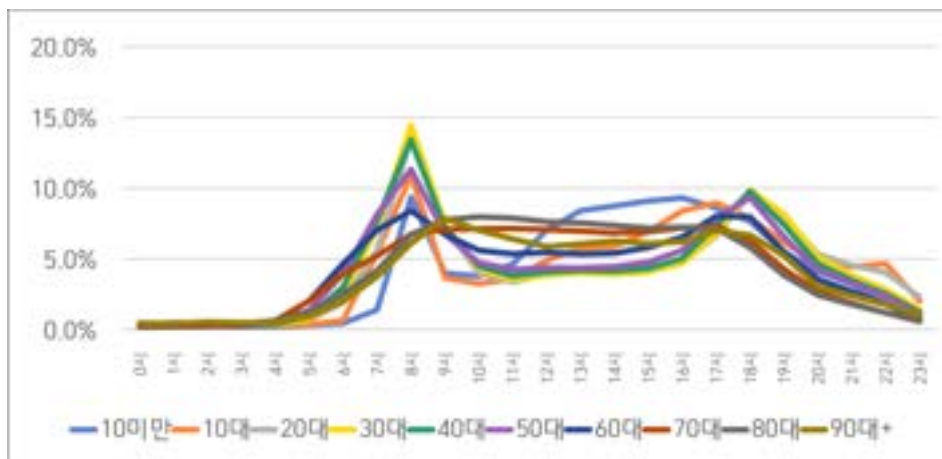
<표 3-7> 연령별 통행량 패턴 분석 (모바일 통신 빅데이터 기준)

(단위: 통행/일)

구분		새벽		오전		오후		저녁	
		0시~	3시~	6시~	9시~	12시~	15시~	18시~	21시~
발 생	0대	542	455	6,916	7,656	13,040	13,766	7,604	2,071
	10대	5,506	5,132	82,370	49,314	74,819	108,830	70,882	37,734
	20대	19,311	34,502	390,070	201,208	199,704	287,656	304,594	114,367
	30대	16,291	42,658	472,446	227,270	206,440	316,119	322,741	88,108
	40대	16,885	61,390	522,448	262,158	243,846	373,572	328,613	89,152
	50대	17,502	94,599	563,865	321,600	299,169	454,517	333,792	96,278
	60대	9,969	69,654	277,337	219,871	210,403	272,850	159,346	43,477
	70대	3,432	20,109	71,072	86,098	82,260	82,116	38,318	9,487
	80대	985	2,642	12,737	18,475	17,067	15,772	6,878	1,913
	90대 ~	68	160	831	906	864	934	499	154
도 착	0대	504	385	5,845	6,459	12,600	14,061	8,998	3,200
	10대	5,036	4,067	73,755	45,353	71,580	105,354	81,351	48,092
	20대	16,645	22,202	328,108	226,979	191,475	241,558	356,378	168,068
	30대	13,626	23,122	406,053	258,047	199,020	262,931	391,919	137,356
	40대	13,756	30,047	458,125	293,825	235,286	317,048	413,390	136,586
	50대	14,006	45,713	511,043	342,814	287,263	397,820	437,123	145,539
	60대	8,149	36,967	253,444	222,863	204,126	256,945	214,236	66,178
	70대	3,034	11,689	62,405	83,781	81,899	84,220	51,726	14,138
	80대	889	1,743	10,232	18,024	17,194	16,469	9,186	2,733
	90대 ~	60	105	679	915	854	928	649	224



<그림 3-14> 연령별 통행 발생량 패턴 분석 (모바일 통신 빅데이터 기준)



<그림 3-15> 연령별 통행 도착량 패턴 분석 (모바일 통신 빅데이터 기준)

- 모바일 통신 빅데이터 기준 연령별 통행패턴 분석 결과, 학업 및 경제활동 연령(0~59세)은 뚜렷한 첨두시간 통행패턴을 보이며, 비경제활동 연령(60세 이상)은 시간대별 다소 완만한 통행패턴을 보임
 - 특히, 0~19세에 해당하는 학업 활동 연령대는 다소 이른 오후 1시부터 통행량이 증가하는 패턴을 보여, 실제 상대적으로 이른 하교 시간을 정확하게 반영하는 것으로 판단됨
 - 주로 비경제활동 인구에 해당하는 60세 이상의 통행패턴은 경제활동 인구에 비해 비교적 완만한 첨두시간 통행패턴을 보여, 실제 연령별 통행패턴과 매우 유사한 결과를 도출한 것으로 판단됨

나. 목적별(출퇴근) 통행패턴 비교 및 분석 결과

1) 개인통행실태조사자료·인구총조사자료 기준 출·퇴근 통행량 비교 결과

- 지역별 출퇴근 목적 통행 발생량·도착량을 개인통행실태조사자료·인구총조사자료와 상대적으로 비교하여, 유사한 패턴을 보이는지 확인함

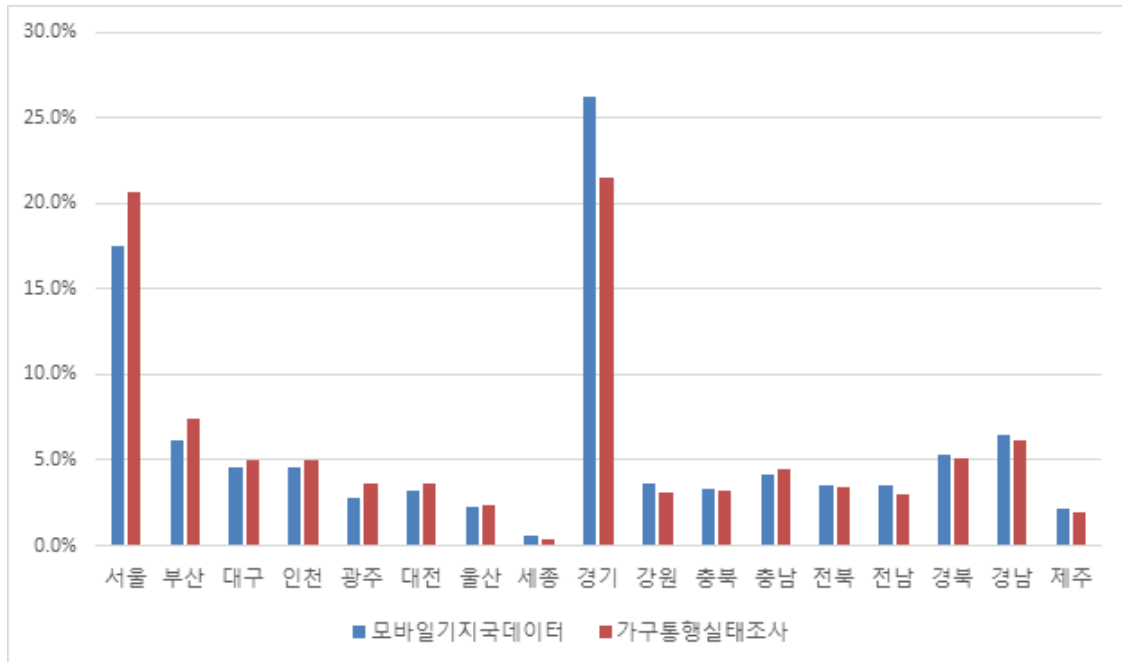
<표 3-8> 지역별 출·퇴근 통행량 비교 결과

(단위: 통행/일 (비율%))

구분	개인통행실태조사자료 기준		인구총조사자료 기준	모바일 통신 빅데이터 기준	
	출근 (비율)	귀가*	통근**	출근	퇴근
서울	4,516,466(20.7)	8,413,729(22.1)	4,597,322(19.6)	631,221(20.0)	536,024(19.2)
부산	1,492,300(6.8)	3,051,128(8.0)	1,500,101(6.4)	205,559(6.5)	180,838(6.5)
대구	974,423(4.5)	1,801,446(4.7)	1,064,227(4.5)	152,313(4.8)	134,449(4.8)
인천	1,213,843(5.6)	1,794,429(4.7)	1,344,311(5.7)	171,099(5.4)	150,288(5.4)
광주	838,601(3.8)	1,279,355(3.4)	639,412(2.7)	87,253(2.8)	77,097(2.8)
대전	736,843(3.4)	1,313,431(3.4)	672,856(2.9)	97,005(3.1)	86,176(3.1)
울산	514,493(2.4)	995,135(2.6)	538,654(2.3)	71,522(2.3)	62,921(2.3)
세종	89,203(0.4)	149,020(0.4)	92,403(0.4)	17,648(0.6)	16,358(0.6)
경기	5,272,517(24.1)	7,799,702(20.5)	5,745,634(24.4)	852,163(27.1)	752,696(27.0)
강원	644,578(2.9)	1,186,771(3.1)	703,266(3.0)	89,019(2.8)	82,094(2.9)
충북	637,895(2.9)	1,182,200(3.1)	742,399(3.2)	93,268(3.0)	85,818(3.1)
충남	875,177(4.0)	1,616,017(4.2)	1,032,913(4.4)	113,204(3.6)	104,716(3.8)
전북	665,597(3.0)	1,263,474(3.3)	811,899(3.5)	93,693(3.0)	85,406(3.1)
전남	706,911(3.2)	1,241,157(3.3)	864,498(3.7)	87,501(2.8)	81,287(2.9)
경북	1,007,670(4.6)	2,045,073(5.4)	1,280,135(5.4)	143,352(4.6)	133,229(4.8)
경남	1,372,291(6.3)	2,361,671(6.2)	1,560,582(6.6)	191,226(6.1)	174,882(6.3)
제주	291,636(1.3)	581,151(1.5)	309,133(1.3)	52,204(1.7)	47,223(1.7)
계	21,850,443(100)	38,074,889(100)	23,499,745(100)	3,149,247(100)	2,791,504(100)

주 : *개인통행실태조사자료는 2016년 기준 출·퇴근에 대한 목적통행량이며, 귀가통행 통행목적에 상관없이 집으로 향하는 모든 통행을 의미함

**인구총조사(통계청)자료는 2015년도 기준이며, 출근과 퇴근이 구분되어 있지 않음



<그림 3-16> 시도별 통근 통행 비율 비교

- 지역별 출퇴근 통행량 비율을 살펴보면, 평균적으로 개인통행실태조사자료 기준 약 0.64%p, 인구총조사자료 기준 약 0.3%p의 차이를 보임
 - 개인통행실태조사자료와 비교하였을 때, 부산지역의 출·퇴근 통행량이 약 2.4%p로 가장 큰 폭으로 과대 추정되었으며, 경기지역의 출퇴근 통행량이 약 3.3%로 가장 큰 폭으로 과소 추정되었으나, 평균적으로 약 0.64%p로 모바일 통신 빅데이터 기준 출퇴근 통행량이 신뢰성 있는 것으로 판단됨
 - 인구총조사자료 기준 대구지역의 출퇴근 통행량이 약 0.4%p로 가장 큰 폭으로 과대 추정되었으며, 전남지역의 출퇴근 통행량이 약 0.6%p로 가장 큰 폭으로 과소 추정되었으나, 평균적으로 약 0.3%p로 모바일 통신 빅데이터 기준 출퇴근 통행량이 신뢰성 있는 것으로 판단됨

2) 개인통행실태조사자료·인구총조사자료 기준 출퇴근 통행시간 비교 결과

- 지역별 출퇴근 목적 통행시간을 개인통행실태조사자료·인구총조사자료와 상대적으로 비교하여, 유사한 패턴을 보이는지 확인함

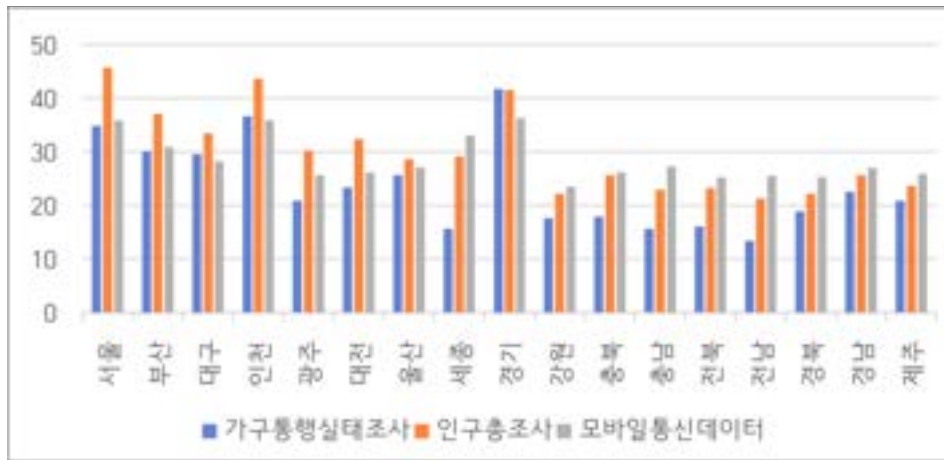
<표 3-9> 지역별 출퇴근 통행시간 비교 결과

(단위: 분)

구분	개인통행실태조사자료 기준		인구총조사 자료 기준	모바일 통신 빅데이터 기준	
	출근	귀가*	통근**	출근	퇴근
서울	34.8	30.1	45.7	35.8	38.8
부산	30.1	25.6	37.0	30.8	33.0
대구	29.5	28.2	33.4	28.1	30.2
인천	36.6	32.5	43.6	35.9	37.4
광주	20.8	18.3	30.2	25.6	27.5
대전	23.3	21.7	32.3	26.1	27.8
울산	25.6	24.5	28.6	27.1	28.6
세종	15.6	15	29.1	33.0	34.3
경기	41.7	36.3	41.5	36.2	37.6
강원	17.6	20.3	22.1	23.4	24.8
충북	17.9	17.8	25.6	26.1	27.8
충남	15.6	16.7	22.9	27.2	28.7
전북	16	15.4	23.2	25.2	26.9
전남	13.3	13.6	21.3	25.4	27.0
경북	18.8	20	22.1	25.2	26.8
경남	22.5	21.1	25.6	27.0	28.4
제주	20.8	21.5	23.6	25.8	27.5
계	23.6	22.3	35.2	28.5	30.2

주 : *개인통행실태조사자료는 2016년 기준 출근·귀가 통행에 대한 일평균시간이며, 귀가통행 통행목적에 상관없이 집으로 향하는 모든 통행을 의미함

**인구총조사자료는 2015년 기준 통근 통행에 대한 소요 시간별 자료를 일평균시간 형태로 재구성하였으며, 출근과 퇴근이 구분되어 있지 않음

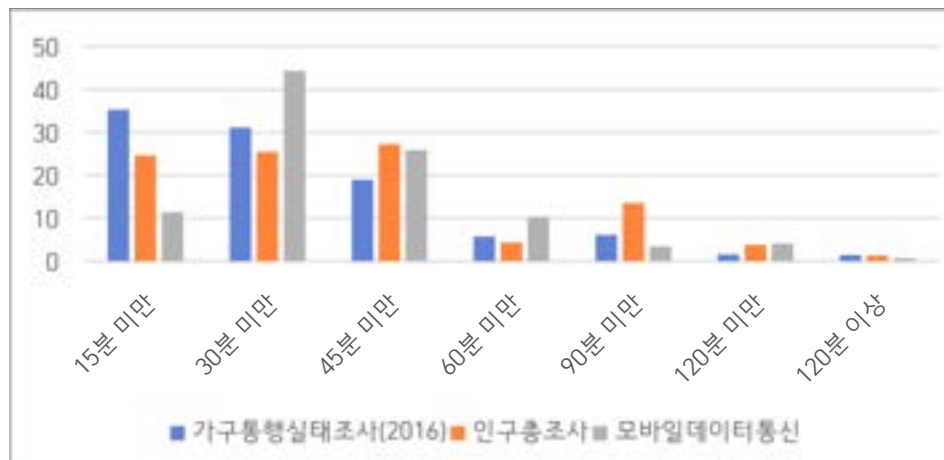


<그림 3-17> 시도별 통근 통행시간 비교 (단위 : 분)

- 개인통행실태조사자료와 인구총조사자료 기준 1인당 평균 출근시간 및 통근시간은 각각 약 23.6분, 35.2분이며, 모바일 통신 빅데이터 기준은 약 28.5분임
- 각 자료별 평균 출퇴근 통행시간을 비교해보면, 인구총조사자료 기준이 가장 길고, 개인통행실태조사자료 기준이 가장 짧은 것으로 나타남
- 지역별 분석결과를 살펴보면, 모바일 통신 빅데이터는 특별시·광역시 지역에서 개인통행실태조사자료와 유사한 출퇴근 통행시간 분포를 띄며, 시도에서는 인구총조사자료와 유사한 출퇴근 통행시간 분포를 띄
- 각각 지역마다 출퇴근 통행시간을 비교해보면 다소 차이가 있다고 해석할 수 있지만, 전체적인 지역적 패턴을 기준으로 보았을 때 매우 유사한 분포를 띄

<표 3-10> 통행시간 기준별 통근 인구 비율

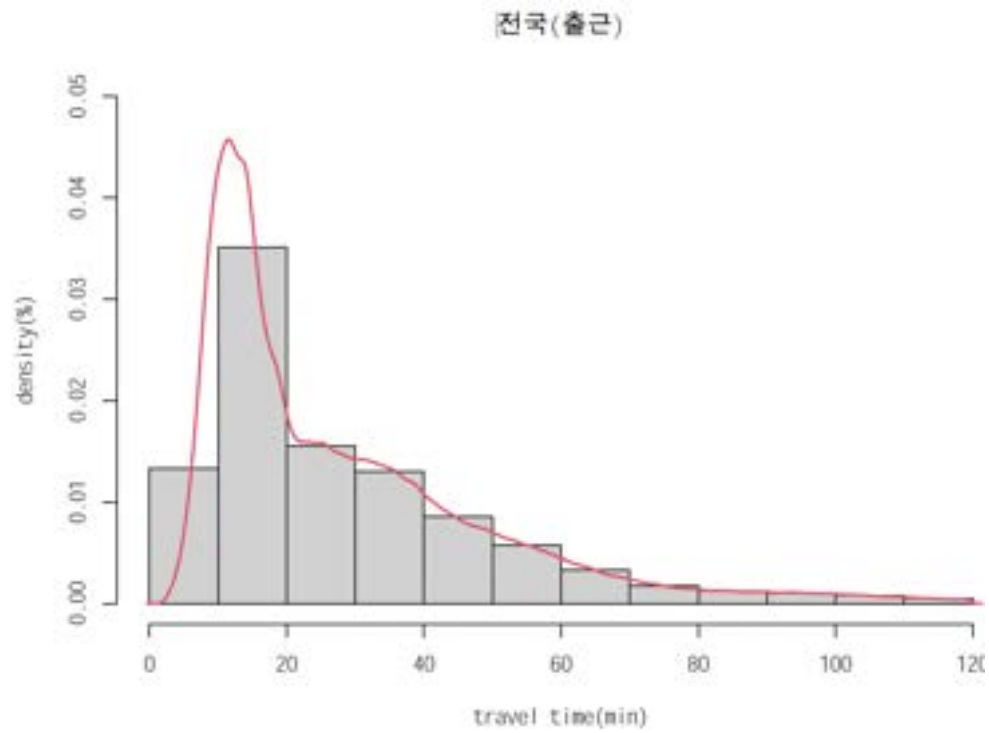
기준	(단위: 통행/(비율%))		
	개인통행실태조사자료 기준 출근 귀가*	인구총조사자료 기준 통근**	모바일 통신 빅데이터 기준 출근 퇴근
통근시간 < 15분	35.3	24.6	11.4
15분 < 통근시간 < 30분	31.1	25.4	44.3
30분 < 통근시간 < 45분	18.9	27.2	25.9
45분 < 통근시간 < 60분	5.7	4.3	10.3
60분 < 통근시간 < 90분	6.1	13.5	3.4
90분 < 통근시간 < 120분	1.5	3.8	4.1
통근시간 > 120분	1.4	1.2	0.6
계	98.6	98.8	99.4



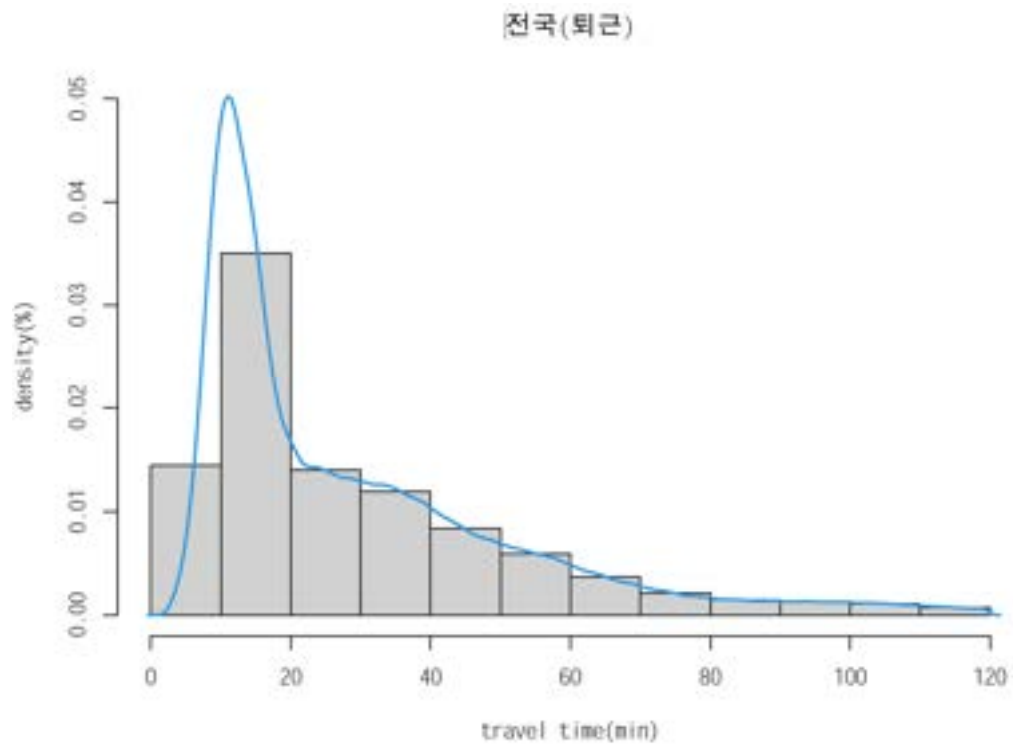
<그림 3-18> 통행시간 기준별 통근 인구 비율 비교

- 각 자료의 통행시간 분포별 통행량을 살펴보면, 특정 시간 분포에서 다소 과대·과소 추정되는 경향이 관찰되나, 전체적으로는 기존자료 대비 매우 유사한 패턴을 보임
 - 15분 미만 출퇴근 통행량은 인구총조사자료 대비 과소 추정되었으나, 개인통행실태조사자료 기준과는 매우 유사한 분포를 띄는 것으로 나타남
 - 15분 이상 30분 미만에서는 과대 추정됨
 - 30분 이상 45분 미만에서 출퇴근 통행량은 인구총조사자료 기준과 매우 유사한 분포를 띄며, 개인통행실태조사자료와 비교하였을 때는 과대 추정된 것으로 나타남
 - 통행 시간대별로 모바일 통신 빅데이터가 과소 또는 과대 추정되어 나타났지만, 출·퇴근시간, 출퇴근 속도 밀도분포를 확인해본 결과, 매우 신뢰성 있는 분포를 보임³⁾
- <그림 3-19>, <그림 3-20>은 출퇴근 시간 밀도 그래프임. 모바일 통신 빅데이터 기준 출퇴근 통행시간은 10분을 정점으로 로그 노말(log-normal) 분포에 가깝게 나타남

³⁾ 노정현, '교통계획: 통행수요이론과 모형', 개정판(2012), 나남.



<그림 3-19> 출근 시간 밀도 그래프



<그림 3-20> 퇴근 시간 밀도 그래프

제4장 모바일 교통 빅데이터 기반 교통지표 개발 및 온라인 서비스 제공

제1절 기초 DB 재구조화

제2절 교통지표 개발 및 온라인 서비스 제공

제3절 온라인 서비스 운영 및 유지보수

제4장 모바일 교통 빅데이터 기반 교통지표 개발 및 온라인 서비스 제공

제1절 기초 DB 재구조화

1. 개요

- KTDB는 인력조사 방식의 교통수요분석 DB 및 통계 생성 체계를 개선하기 위해 모빌리티 데이터와 이를 가공·분석하기 위한 기반 맵을 구축하여 모빌리티 빅데이터 기반 국가교통 DB 생성 체계로의 전환을 준비 중임
- 이에 「모바일 통신빅데이터 구축」 사업을 수행함으로써 모빌리티 빅데이터 기반의 다양한 지표와 분석환경을 구축하고자 함
- 모바일 교통빅데이터 구축
 - 최근 교통 분야에서 각종 모빌리티 정보를 활용하여 각종 통행 관련 분석이 가능해지면서 모빌리티 빅데이터 각광 받고 있지만, 데이터 특성에 따라 분석 대상과 범위가 제한적임
 - 차량에 탑재된 기기(내비게이션, DTG 등)나 도로에 설치된 기기(검지기, DSRC 등)를 통해 차량 통행자에 대한 통행 정보는 취득 가능하나, 자전거·도로 등으로 이동한 통행자에 대한 통행 정보는 표본율이 낮은 개인통행실태조사 결과에만 의존하고 있는 실정
 - 대중교통 이용자에 대한 통행은 교통카드 DB(대중교통 승하차 시 태그 정보)로 대부분 추정 가능하나, 잠재적으로 대중교통을 이용할 가능성이 있는 타 교통수단 이용자(차량, 도보 등)에 대한 통행은 일부만 추정 가능
 - 거의 모든 통행 정보가 기록되어 있는 모바일 빅데이터를 분석 재료로 활용할 필요
 - 모바일 기기와 기지국 간의 송·수신 이력인 모바일 빅데이터는 대부분의 국민이 이용하고 있는 모바일 기기를 통해 축적되기 때문에 표본율이 매우 높고, 이동수단에 구애받지 않고(항공 제외), 기기의 전원을 끄지 않는 이상 개인의 이동 궤적이 상세히 기록되기 때문
 - 또한 국가통합교통체계효율화법 제14조 제1항에 의거 교통수단 및 교통시설의 운영 실태, 통행량 등 교통조사에 정보통신수단(휴대전화)을 활용할 수 있음

2 공간적·시간적 범위

- 공간적 범위 : 전국
- 서비스를 위한 모바일 데이터 가공 : 2020년

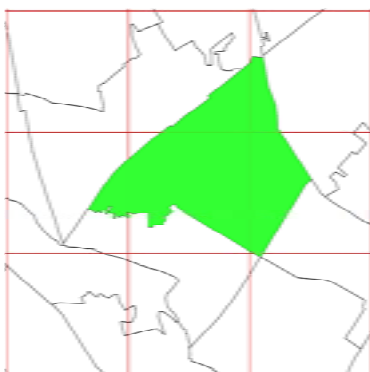
<표 4-1> 서비스를 위한 모바일 데이터 가공

구분		주요내용
데이터 가공	데이터 전처리 및 가공	- 모바일 데이터 재구조화
	1차 가공DB 구축	- 링크-관측교통량 관계 데이터 DB 구축 - 경로 데이터 기반의 속도 프로파일 및 프로브 데이터 구축 - 경로 데이터 기반의 회전통행량 구축
	기초 교통DB 구축	- 전국 추정교통량 및 평균속도 구축 - 모바일 데이터의 이종 데이터간 결합 및 분석 목적별 다중 DB 구성 - 모바일 데이터 행정구역 및 분석 특성별 DB 구축

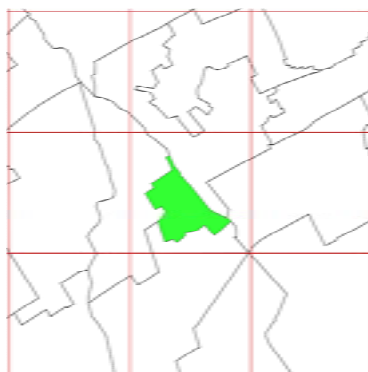
3. 격자-교통폴리곤 관계테이블 구축

- 통신데이터 서비스를 위하여 격자와 교통폴리곤 간의 관계테이블을 구축함. 관계테이블은 수집된 두 공간정보를 중첩시켜 두 데이터 간 공간조인을 통하여 구축함
- 격자와 교통폴리곤의 형상 정보를 중첩시켰을 때, 두 데이터 간의 포함관계는 다음과 같이 크게 세 유형으로 구분할 수 있음

유형1) 격자 : 교통폴리곤 = n개:1개



유형2) 격자 : 교통폴리곤 = n개:1개



유형3) 격자 : 교통폴리곤 = 1개:1개



<그림 4-1> 격자-교통폴리곤 간 형상정보 중첩 유형

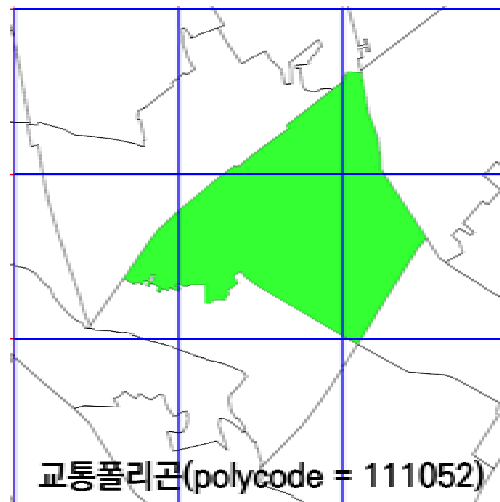
- 격자-교통폴리곤 관계테이블에는 격자 및 교통폴리곤의 기본 속성정보와 함께 격자와 교통폴리곤 간 관계정보가 포함될 수 있도록 구축함
- 격자와 교통폴리곤 간 관계정보에는 크게 면적비율 정보와 개수 정보가 있음
 - 면적비율 정보: ‘격자와 교통폴리곤 공간 데이터의 중첩으로 교차 분할된 면적/교통폴리곤 면적’ 식을 통해 산출
 - 개수 정보: 교통폴리곤 기준 격자 분할 개수
- 격자-교통폴리곤 관계 테이블은 다음과 같이 정의함

<표 4-2> 격자-교통폴리곤 관계테이블 테이블 정의서

No	Column	설명	Type	자리수	코드정보
1	gid	격자ID	VARCHAR	17	
2	polycode	교통폴리곤ID	VARCHAR	100	
3	sido	시도ID	VARCHAR	100	
4	emd	읍면동ID	VARCHAR	8	
5	name	읍면동명	VARCHAR	100	
6	r_bm_p	교통폴리곤 기준 격자 면적비율 정보	VARCHAR	100	
7	p_cnt	교통폴리곤 기준 격자 개수 정보	VARCHAR	100	

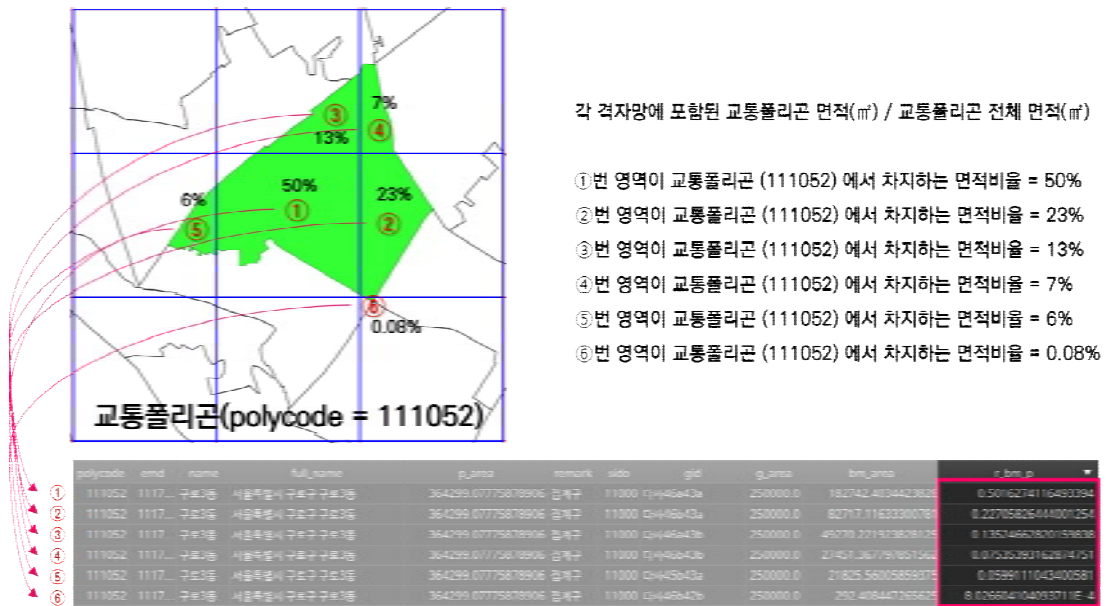
4. 격자-교통폴리곤의 통행량 재산정 방안

- 교통폴리곤 기준으로 가공된 OD 통행량을 500m 격자 단위의 OD 통행량으로 재산정하기 위하여 다음의 방법을 활용하고자 함
- 하기 예시는 격자와 교통폴리곤을 중첩시켰을 때의 형상으로 다음의 예시에서 하나의 폴리곤(교통폴리곤ID = 111052)이 6개의 격자에 의해 분할되었음을 알 수 있음



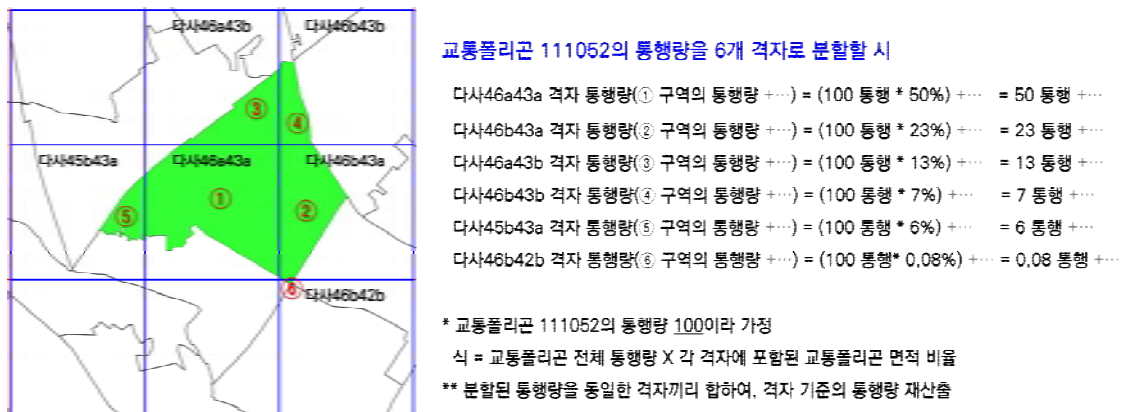
<그림 4-2> 격자와 교통폴리곤 중첩 형상

- 격자와 교통폴리곤 간의 중첩된 공간정보 데이터를 통해 두 공간 데이터 간 교차영역에 대한 면적정보와 교통폴리곤 전체 면적에서의 교차 면적정보간 면적 비율을 산출할 수 있음
- 격자와 교통폴리곤 간 교차되는 ①~⑥번 영역이 ID 111052를 갖는 교통폴리곤에서 차지하는 면적 비율은 각각 약 50%, 23%, 13%, 7%, 6%, 0.08%(전체 비율 합 = 1)임을 계산할 수 있음



<그림 4-3> 격자-교통폴리곤 간 면적 비율 산출 방안

- 교통폴리곤 111052의 OD 통행량을 100이라 가정했을 때, 분할되는 6개 격자 기준의 격자 통행량은 각각 다음의 식으로 정리할 수 있음



<그림 4-4> 격자로의 OD 통행량 재산출 방안

- 위의 개념을 적용하여, 모든 교통폴리곤의 OD 통행량을 격자 기준으로 재산정한 통행량을 동일한 격자끼리 합하여, 최종적으로 각 격자 기준의 통행량을 재산정함
- 격자와 교통폴리곤 간 면적비율을 비롯한 일련의 정보는 격자-교통폴리곤 관계데이터에 적재하여 구축 및 관리함

5. 격자와 연계 공간정보 DB 간 분석을 위한 기초 데이터 분석

가. 국토지리정보원의 격자망 데이터 분석

- 국토교통 분야뿐만 아니라 여러 분야의 이종 데이터와 통신데이터 간의 융합 분석을 위해 전 국토를 표준화된 공간 단위로 구획한 국토지리정보원의 격자망을 활용하고자 함
- 이에 국토지리정보원의 격자망 데이터의 격자 크기별 위계와 격자 ID 체계를 분석함

제2장 격자체계의 위계기준

□ 제5조(위계기준) 격자이란 국토 통계지도를 생산하기 위한 위계기준에 의한다. ① 기준좌표계 : 세계좌표계 기원의 UTM-K

② 격자 원형 : UTM-K 투영원형으로부터 세폭 300m, 남북 700m 지정

□ 제6조(격자체계) 격자이란 국토 통계지도를 생산하기 위한 방법은 다음과 같다. ① 격자 구조 : 상하의 격자는 계층적으로 중첩관계에 속함

② 격자 크기 : 격자는 가로와 세로의 길이를 동일하게 하며, 평면상 면적이 동일하도록 10m, 50m, 100m, 250m, 500m, 1km, 10km, 100km 구획

③ 격자 번호 : 격자크기가 10m, 100m, 1km, 10km, 100km에 대해서는 도로명주소법 시행령 제11조제3항제2호의 부여기준을 따르고, 그 외 격자크기에 대해서는 국토지리정보원 지정기준 사항을 따라야 함

행정안전부, 「행정정보의 격자체계 설정 및 공간정보화 기준」 (시행 2017.12.26.) 발령

<그림 4-5> 격자 체계 기준

- 국토지리정보원의 격자체계 사양 문서 기준에 따르면, 기관 표준에서 정의하는 격자 셀의 크기는 10m, 50m, 100m, 250m, 500m, 1km, 10km, 100km이며, 각 격자의 ID 체계에 관한 내용을 정리하면, 다음과 같이 정리할 수 있음

격자수 목록												
		국가지점번호 (한글 두자리)		가로축 세부번호 (숫자/영소문자, 0~4자리)						세로축 세부번호 (숫자/영소문자, 0~4자리)		비고
격자크기	ID예시	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩	
100km	파사	한	한	=	=	=	=	=	=	=	=	2자리
10km	파사1 0	한	한	0	-	-	-	0	-	-	-	4자리
1km	파사13 67	한	한	0	0	-	-	0	0	-	-	8자리
100m	파사133 677	한	한	0	0	0	-	0	0	0	-	8자리
10m	파사13336777	한	한	0	0	0	0	0	0	0	0	10자리
500m	파사13a 67b	한	한	0	0	소	-	0	0	소	-	8자리
250m	파사13a677b	한	한	0	0	소	소	0	0	소	소	10자리
50m	파사133a677b	한	한	0	0	0	소	0	0	0	소	10자리

한: 한글(가~아), 0: 숫자(0~9), 소: 영소문자(a, b), -: 없음(공란없이)

<그림 4-6> 격자 크기별 격자ID (gid) 체계

- 데이터의 효용성, 분석 및 서비스의 용이성, 격자 단위에 따른 Zero-cell 발생 문제

등을 고려하여, 최종적으로 사용되는 격자 크기는 500m로 설정함

나. 연계 공간정보 DB 분석

1) 국가관심지점(POI) 데이터 분석

- 국가관심지점(이하 POI) 데이터는 국토지리정보원의 국토정보플랫폼에서 수집한 데이터로 관심지점에 대한 정보를 17개의 시도 단위별 포인트 형상의 공간정보 DB임



<그림 4-7> 국가관심지점(POI) 데이터 형상 예시

- POI 데이터의 속성 정보는 다음과 같이 정의됨

<표 4-3> POI 데이터 테이블 정의서

No	Column	설명	Type	자리수	코드정보
1	NF_ID	고유식별자 아이디	VARCHAR	17	
2	POI_NM	관심지점 명칭	VARCHAR	100	
3	POI_CL_DC	관심지점 분류 설명	VARCHAR	100	세부코드표 참조
4	ORIGIN_SE	출처 구분	VARCHAR	8	
5	ENGL_NM	영어 명칭	VARCHAR	100	
6	CHNLNG_NM	중국어 명칭	VARCHAR	100	
7	JPN_NM	일본어 명칭	VARCHAR	100	
8	REFRN_ID	참조 아이디	VARCHAR	17	
9	OBCHG_DT	객체변동 일시	TIMESTAMP		
10	MESRMTH_SE	수정측량방법 구분	VARCHAR	1	
11	RSREG_DT	성과등록 일시	TIMESTAMP		
12	MNENT_NM	제작업체 명	VARCHAR	100	
13	DBREG_DT	데이터베이스등록 일시	TIMESTAMP		

- 예를 들어 관광시설에 대한 POI 정보를 확인하기 위해 POI 속성정보 중 POI_CL_DC (관심지점 분류 설명) 항목에서 관광 관련 코드를 검색함
 - 관광과 관련된 코드를 검색하기 위해 세부코드표의 세부코드명에서 ‘관광’을 키워드로 검색함

행정출처	출처코드	하이라	세부코드 (PQ, CL, DC)	세부코드명
국가기본도DB	AR804000	시설구역장계	FAA4T1	관할지시설
국가기본도DB	BL001000	건물	BDU027	관할표기시설
지명	GM000000	지명	C0003	관공
외부수집POI	KCP00000	외부수집POI	09070006	관공물적확정
외부수집POI	KCP00000	외부수집POI	09020001	관공단지
외부수집POI	KCP00000	외부수집POI	09020002	관공특구
외부수집POI	KCP00000	외부수집POI	09020003	암벽관공지
외부수집POI	KCP00000	외부수집POI	09020004	지형관공지
외부수집POI	KCP00000	외부수집POI	09020400	기타관공지
외부수집POI	KCP00000	외부수집POI	09020401	관공안내소
외부수집POI	KCP00000	외부수집POI	09020501	관공농림
외부수집POI	KCP00000	외부수집POI	09020502	관공목장
외부수집POI	KCP00000	외부수집POI	09020507	특별관공지

(이하 생략)

<그림 4-8> POI_CL_DC 항목의 세부 코드표(관광 검색)

- 세부코드표의 ‘관광’ 키워드 중 세부코드(POI_CL_DC)가 관광지시설(FAA411) 코드를 갖는 데이터를 POI 속성 정보에서 검색

[illegible]

<그림 4-9> POI 데이터 내 관광지·시설 검색

- POI 데이터 분석 결과, 각 시설물 기준의 위치 정보를 포함한 시설물 명칭 정보 등 시설물에 관한 기본적인 정보는 알 수 있으나, 시설물의 면적이나 특정 지역에서의 점유율 등을 알 수 있는 항목은 부재함

2) 용도별 건물공간정보 분석

- 용도별 건물공간정보(이하 건물공간정보)는 국토교통부 국가공간정보포털 오픈 API에서 수집한 데이터로 전국의 건물공간정보를 시군구 단위별 폴리곤 형상의 공간정보 DB임



<그림 4-10> 용도별 건물공간정보 형상 예시

- 건물공간정보의 속성정보는 총 36개의 항목으로 구성되어 있으며, 건축물대장의 층 정보와 전유부분 용도 비율을 기준으로 추출한 건축물의 용도 정보를 포함
- POI와 마찬가지로 관광시설에 관한 정보를 확인하기 위해 건물공간정보의 속성정보 중 A27(세부용도명) 항목에서 관광 관련 코드를 검색함
 - 관광과 관련된 코드를 검색하기 위해 세부코드표의 세부코드 약어명에서 ‘관광’을 키워드로 검색함
 - 세부코드표의 ‘관광’ 키워드 중 세부용도코드(A26)가 관광호텔(15201) 코드를 갖는 데이터를 건물공간정보 속성 정보에서 검색

세부용도코드(α26) 검색조건 "15201(관광호텔)" : 검색조건 지정(세부코드 분류항목 수: 901개) ⇒ 관광시설 분류 가능

<그림 4-11> 건물공간정보 내 관광 관련 데이터 검색

- 건물공간정보 데이터 분석 결과, POI 데이터와는 달리 시설물(건물)의 면적을 포함하고 있으나 검색된 건물이 다양한 용도로 사용되는 경우라도 해당 건물의 주요한 용도 정보만 포함하고 있어 대표성의 문제가 발생함

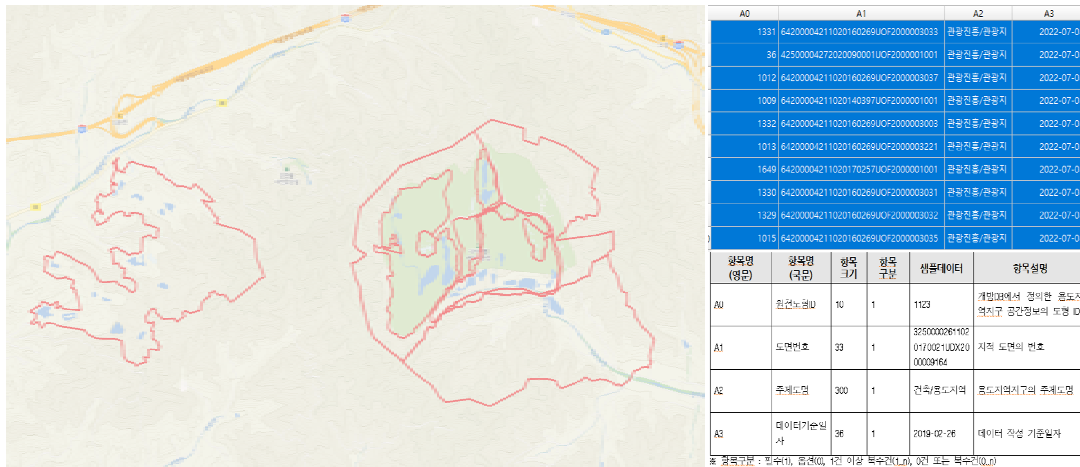
3) 용도지역 데이터 분석

- 용도지역 데이터는 국토교통부 국가공간정보포털 오픈마켓에서 수집한 데이터로 전국의 용도지역 정보를 시도 단위별 폴리곤 형상의 공간정보 DB임



<그림 4-12> 용도지역 중 도시지역 형상 예시

- 용도지역 데이터의 속성정보는 총 4개의 항목으로 구성되어 있으며, 각 형상별 ID, 도면번호, 주제도명, 데이터 작성 기준일자 정보를 포함
- 앞서 분석한 데이터들과 마찬가지로 관광시설에 대한 정보를 확인하기 위해 용도지역 구분 중 '관광진흥/관광지' 관련 데이터를 수집함



<그림 4-13> 용도지역 중 관광진흥/관광지 주제도 데이터 예시

- 용도지역 데이터 분석 결과, 관광지로 계획된 지역의 경계 정보는 알 수 있으나, 실제 관광지로의 사용 여부, 명칭 정보 등과 같은 실질적 정보 불포함

다. 모바일 빅데이터 분석

- 개별 사람의 시공간적 움직임이 휴대폰 신호를 통해 인근 기지국에 송신되어 수집되는 데이터로 가상기지국 단위로 정보가 수집되고 있으나, 데이터 제공 시에는 개인정보보호법과 위치정보의 보호 및 이용 등에 관한 법률 등으로 개인의 위치와 이동 궤적이 추정되지 않는 집계 형태여야 함. 집계 형태는 개인의 위치와 이동 궤적이 추정되지 않도록 집계구의 데이터와 기지국 간의 정보를 이용하여 구축한 교통폴리곤 단위로 변환 후 OD 형태로 데이터가 제공되고 있음. 데이터 제공을 위한 데이터 처리 과정은 아래와 같음

- 원시데이터 전처리

- 원시데이터는 고객 식별번호, 가상기지국, 체류시작시간, 체류종료시간, 연령대, 성별 등의 정보로 구성
- 고객 식별번호와 시간정보를 이용하여, 선을 그린 후 데이터 전환 및 이상치 제거

- 데이터 로그별 이동과 체류 정보 구분

- 체류시간 및 위치정보를 이용하여 사람의 이동인지 체류인지 판단
- 체류의 판단 후 연령대별로 주 몇 회 이상 체류와 체류 시간대에 대한 기준을 두어 통행목적(집/학교/직장/여가 등) 구분

- 위치 정보 변환 및 데이터 집계

- 가상기지국의 위치 정보와 교통폴리곤의 영역과 공간조인 하여 정보 매칭
- 매칭 정보를 이용하여 교통폴리곤 기준으로 데이터 가공

<표 4-4> 모바일 빅데이터 특징

구분	2019년	2020년
제공 파일 구성	1일 단위 텍스트 파일	1일 단위 텍스트 파일
용량(년)	1년 기준 약 307GB	1년 기준 약 1.7TB
이벤트(억/년)	1년 기준 약 24억 건	1년 기준 약 96억 건
위치정보 제공	출퇴근 : 교통폴리곤 단위로 제공 기타통행 : 시군구 단위로 제공	교통폴리곤 단위로 제공
체류지 유형	집(H) 회사(C) 학교(S) 기타 잠재체류지(X)	집(H) 회사(C) 학교(S) 정기적 잠재체류지(R) 비정기적 잠재체류지(X)

- 제공 데이터 자료 예시

출발				도착				연령	성	동행량	기준일
출발폴리곤 ID	출발일자	시간	트립타입	도착폴리곤 ID	도착일자	시간	트립타입				
110006	20160401	0	X	110006	20160401	0	N	20	M	8	20160401
110016	20160401	0	N	110016	20160401	1	X	20	F	15	20160401
110017	20160401	10	D	110716	20160401	12	X	20	F	6	20160401
110033	20160401	12	N	110386	20160401	14	X	60	F	5	20160401
110036	20160401	13	N	110036	20160401	17	D	70	M	3	20160401
110040	20160401	15	N	110628	20160401	19	X	30	F	7	20160401

<그림 4-14> 모바일 빅데이터 자료 형태

6. 모빌리티 빅데이터 구축 프로세스

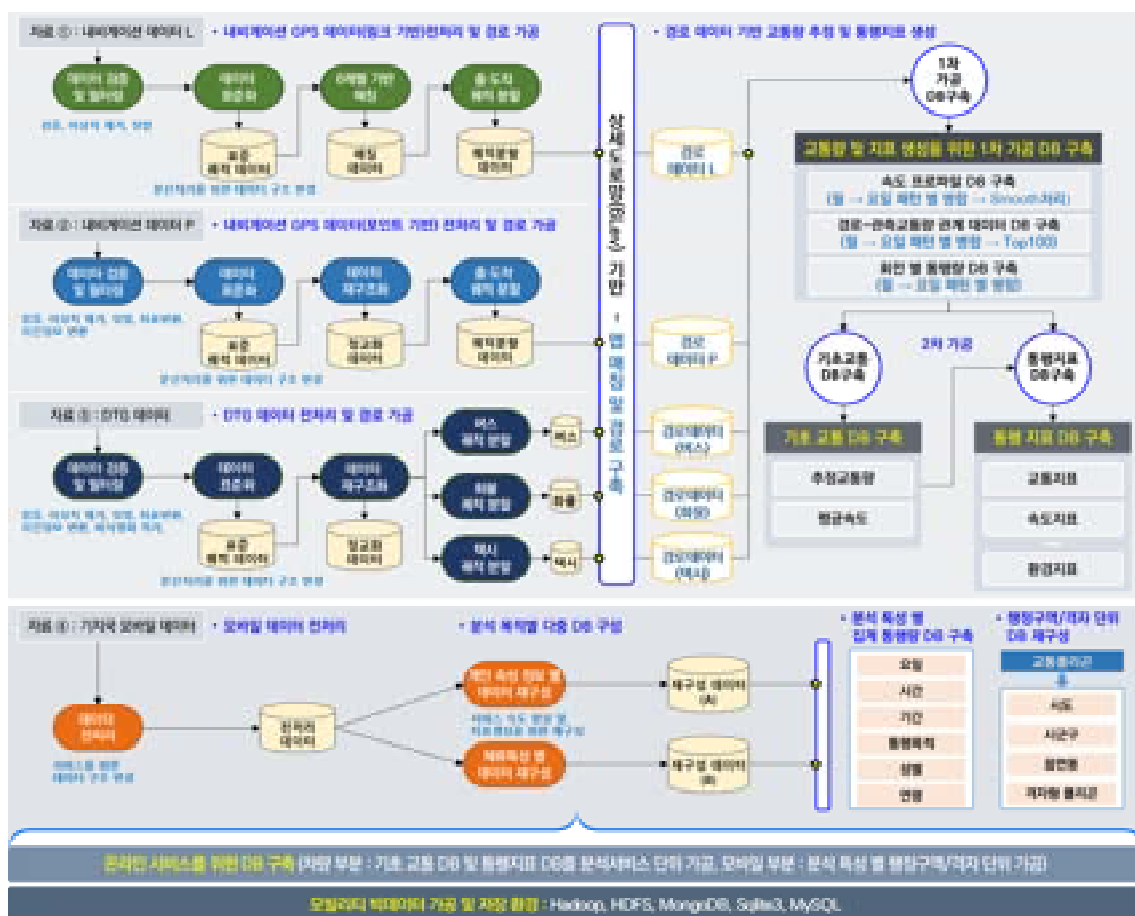
가. 모빌리티 빅데이터 구축 개요

- 본 사업에서 제공되는 모빌리티 빅데이터는 내비게이션 및 차량디지털운영기록을 통하여 수집되는 차량 GPS 데이터와 휴대폰 신호를 통해 가상기지국 단위로 정보가 수집되는 모바일 빅데이터가 있음. 이를 활용하기 위해서는 제공되는 데이터의 특성을 분석하여 각 특성에 맞추어 데이터 가공 및 검증이 필요함
- 모바일 빅데이터는 출발 / 도착 기준으로 데이터가 가공되어 제공. 모바일 빅데이터는

누락 및 이상치 데이터가 있어, 이에 대한 전처리 모듈이 필요함. 또한 데이터 용량이 커서 이로 인한 서비스의 제약(서비스 속도, 지표 생성)이 발생하여, 서비스를 위한 별도의 데이터 가공 필요

나. 모빌리티 빅데이터 구축 프로세스

- 구축 프로세스는 크게 ① 데이터 전처리 → ② 경로 가공(차량) → ③ 지표 생성 → ④ 온라인 서비스를 위한 데이터 구축 순으로 진행



<그림 4-15> 모빌리티 빅데이터 구축 프로세스

7. 모바일 데이터 전처리 및 가공

가. 모바일 데이터 구축 프로세스

- 데이터 구축은 분석 맵 레이어 구축 및 모바일 데이터 전처리와 통행량 데이터 가공으로 구분함
- 분석 맵 레이어 구축은 격자형 폴리곤 데이터와 교통폴리곤 데이터 및 행정구역 데이터를 이용해 각각의 폴리곤 형상 레이어를 경량화하는 작업과 폴리곤 중심 위치 보정을 통해 포인트 레이어 구축 작업을 진행함
- 모바일 데이터 기반의 교통 분석용 데이터를 그대로 서비스 시 많은 데이터를 가져오는 데에 발생하는 처리 속도 이슈로 인하여 행정경계별로 그룹화하여 조회 데이터 건수를 감소시키고 속도를 개선함
 - 모바일 데이터 전처리는 누락 또는 오류 데이터를 분류하고 정제 데이터의 교통폴리곤과 행정구역을 공간적으로 결합시키고 출발교통폴리곤과 도착 교통폴리곤의 중심 포인트 간 직선거리를 추가함
 - 전처리한 모바일 통행량 데이터를 이용해 검색 속도 향상을 위한 인덱스를 생성하고 교통폴리곤 단위로 데이터를 그룹화하여 집계 통행량 DB를 구축함
 - 행정구역 단위별로 서비스하기 때문에 구축한 교통폴리곤 단위 집계 통행량 DB를 이용하여 출발과 도착 읍면동 기준, 시군구 기준, 시도 기준으로 각각 집계 통행량 DB를 구축함
 - 서비스 만족도를 향상시키기 위하여 검색 및 표출 속도를 개선하고자 모바일 데이터 속성 정보를 가지고 분석 특성을 정의 후 행정구역 단위 집계 통행량 DB를 이용해 분석 특성 기준 집계 통행량 DB를 구축함

8. 모바일 데이터 전처리

가. 모바일 데이터 전처리 필요성

- 수급한 모바일 데이터에 대한 속성 누락 및 정보 오류 등을 분석하고, 이에 대한 필터링 및 보정 필요
 - JAVA 기반의 검증시스템을 개발하여 설계 DB를 기준으로 수집된 통신 데이터를 전수 검증함
 - 정보가 누락된 통신 데이터에 대하여 오류 데이터를 생성 후 확인
- View-T 서비스는 분석 지역과 표출 단위에 대한 선택이 필수적으로 GIS 분석 최소 단위인 교통폴리곤과 교통폴리곤이 속하는 행정구역의 매칭 정보가 필요함
- 검색 쿼리와 DB 적재 용량 부분을 고려하여 문자형 데이터를 숫자 형식으로 변환 작업이 필요함
- 분석 서비스에 활용하기 위해 출발 교통폴리곤과 도착 교통폴리곤 간의 직선거리 생성이 필요함

<표 4-5> 모바일 데이터 구조

컬럼 명	비고	컬럼 명	비고
출발 교통폴리곤 ID	-	도착 교통폴리곤 ID	-
출발 일자	yyyymmdd	도착 일자	yyyymmdd
출발 시간대	0 - 23	도착 시간대	0 - 23
출발 체류지 유형	H: 집 C: 회사 S: 학교 R: 정기잠재체류지 X: 기타잠재체류지	도착 체류지 유형	H: 집 C: 회사 S: 학교 R: 정기잠재체류지 X: 기타잠재체류지
연령대	0 - 110	성별	M: 남성 F: 여성
통행량(인구수)	-	보정계수	-

9. 모바일 데이터 행정구역 및 분석 특성별 DB 구축

가. 모바일 데이터를 이용한 교통폴리곤 단위 집계 통행량 DB 구축

- 모바일 전처리 통행량 데이터를 그대로 서비스하게 될 시 1시간 이상의 처리 시간이 소요되는 문제가 발생됨
- 이는 서비스 시 문제로 판단되어 신속한 서비스를 위하여 경량화된 집계 통행량 DB 구축이 필요함
- 통행 유출과 유입 데이터 분석에 용이하도록 출발 기준, 도착 기준 행정구역 단위별로 집계 통행량 DB 구축함
- 집계 기준은 출발 지역과 도착지역, 성별, 연령대, 출발 체류특성, 도착 체류특성의 조합으로 키를 생성하고 집계 기준 키와 일치하는 데이터들은 출발 시간대와 도착 시간대 기준으로 묶어 데이터를 배열 형태로 구성함

필드	Id	CG	A1	E1	A2	E2	A3	E3
Type	object	object 배열	integer	integer	integer	integer	integer	integer
컬럼 구성	"_Id": ["CG": ["A1":	"E1":	"A2":	"E2":	"A3":	"E3":
	"A": 출발 분석구ID, "E": 도착 분석구ID, "J": 성별, "I": 연령대, "D": 출발 체류특성, "H": 도착 체류특성]	{ "C": 출발 시간대, "G": 도착 시간대, "K": [{ "L": 기준일, "V": 통행량, }, { "L": 기준일, "V": 통행량, }] }	출발 읍면동ID	도착 읍면동ID	출발 시군구ID	도착 시군구ID	출발 시도ID	도착 시도ID

<그림 4-16> 교통폴리곤 기준 통행량 데이터 구조

- 출발교통폴리곤 ID 도착교통폴리곤 ID, 체류 특성, 연령대, 성별 필드를 키(필드명: _id)로 생성하여 중복을 제거함
- 키의 조건과 일치하는 데이터들의 시간대를 배열화하여 집계함
- 배열화 된 시간대 그룹(필드명: CG)에 일자별 통행량을 그룹화하여 데이터를 생성함
- 해당 데이터를 생성함과 동시에 해당 교통폴리곤에 해당하는 A1~E3의 출발 도착 행정구역 ID 필드를 삽입하여 차후 행정구역 단위의 출/도착 데이터 생성 시 참조할 수 있도록 구성함

- 출발/도착의 읍면동, 시군구, 시도를 추가함으로써 특정 행정구역에 포함된 교통폴리곤 단위의 통행량 집계 서비스 시 처리 속도가 빠르며 읍면동 단위의 출발/도착 기준 통행량 구축 시 별도로 교통폴리곤과 매칭되는 행정구역들을 찾아서 추가하지 않아도 되기 때문에 구축 시간이 단축됨

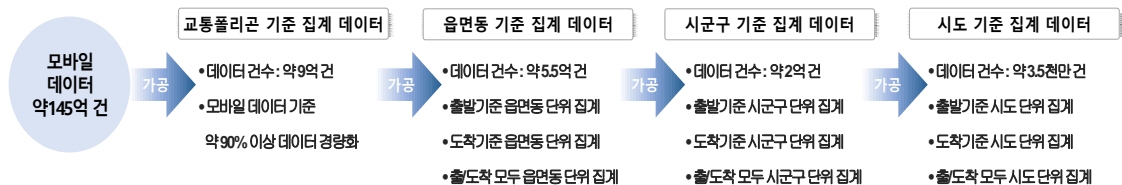
<표 4-6> 교통폴리곤 기준 전처리 및 집계 통행량 컬렉션의 구성

컬렉션명	TMOBILE_2020			
컬렉션 설명	교통폴리곤 기준의 전처리 및 집계 통행량			
데이터 수	27억 건 (2,704,605,308)		데이터 크기	약 362GB
KEY Name	KEY	VALUE TYPE	데이터 구성	비고
식별자	_id	Object	_id : { “A” : 출발분석구코드, “E” : 도착분석구코드, “J” : 성별, “I” : 연령대, “D” : 출발체류특성, “H” : 도착체류특성 }	A, E, D, H, J, I : integer 체류특성 코드 ·H:11, N:12, C:13, S:14, R:15, X: 16 성별 코드 ·F:1, M:2
출도착시간대 인구수	CG	Object 배열	CG : [{ “C” : 출발시간대, “G” : 도착시간대, “K” : [[“B”:출발날짜, “F”:도착날짜, “V”:인구수,], { “B”:출발날짜, “F”:도착날짜, “V”:인구수, }]]	C, G, B, F, V : integer K : object 배열 V : _id, C, G, B, F에 해당하는 원시데이터의 인구수
출발읍면동	A1	INT	“A1” : 출발읍면동코드	
도착읍면동	E1	INT	“E1” : 도착읍면동코드	
출발시군구	A2	INT	“A2” : 출발시군구코드	
도착시군구	E2	INT	“E2” : 도착시군구코드	
출발시도	A3	INT	“A3” : 출발시도코드	
도착시도	E3	INT	“E3” : 도착시도코드	
분석구 간 거리	S	DOUBLE	“S” : 분석구 간 거리	

나. 모바일 데이터를 이용한 행정구역 단위 집계 통행량 DB 구축

- 교통폴리곤 기준 통행량 DB를 이용하여 행정구역 단위의 서비스를 하게 될 시 검색 및 처리 속도가 느려 원활한 서비스에 어려움이 있음. 처리 속도를 향상하고, 서비스를 원활하게 할 수 있도록 행정구역 단위의 DB 가공이 필요함
- 행정구역 단위 집계 통행량 DB 구축은 다음과 같이 진행함
 - 교통폴리곤 기준 통행량 데이터를 이용하여 각 행정구역 단위의 데이터를 생성함
 - 행정구역 단위의 통행량 DB는 GIS 분석 서비스 및 데이터 다운로드 시에 적합하도록 행정구역 단위별로 출발 행정구역 기준 집계 데이터, 도착 행정구역 기준 집계 데이터, 출발과 도착이 모두 행정구역 단위인 집계 데이터를 구축함
 - 출발 행정구역 기준 집계 데이터와 도착 행정구역 기준 집계 데이터는 크게 유출과 유입 데이터 분석에 사용되며 출발과 도착이 모두 행정구역 단위인 데이터는 지역 간, 지역 내 matrix 데이터 분석에 사용됨
 - 구축 순서는 읍면동 -> 시군구 -> 시도로 진행되며 상위 행정구역 단위 통행량 생성 시, 구축 시간을 최소화하기 위하여 집계한 하위 행정구역 단위 통행량을 이용하여 구축함
(읍면동 단위 : 교통폴리곤 기준 통행량, 시군구 단위 : 읍면동 단위 통행량, 시도단위 : 시군구 단위 통행량)
 - 교통폴리곤 기준 통행량과 마찬가지로 MongoDB의 배열 필드를 이용하여 집계함

교통폴리곤 / 행정구역 집계 통행량 DB 구축 과정



집계 통행량 DB 구조 - 출발 읍면동 기준 예시

필드	Object	Array	int	int	int	int	int
컬럼 구성	<pre> _id: { "A1": 출발 읍면동 코드 "E": 도착 교통폴리곤 코드 "J": 성별 "I": 연령대 "D": 출발 체류특성 "H": 도착 체류특성 } </pre>	<pre> CG: [{ "C": 출발 시간대 "G": 도착 시간대 "K": [{ "SD": 출발날짜 "ED": 도착날짜 "V": 통행량 },] }] </pre>	도착 읍면동 코드	출발 시군구 코드	도착 시군구 코드	출발 시도 코드	도착 시도 코드

<그림 4-17> 집계 통행량 DB 구축 프로세스

- 행정구역 단위의 통행량에 대한 데이터 구조는 다음과 같음
 - 각 행정구역 DB에 맞게 출발 도착에 대한 행정구역 ID로 키를 구성

필드	_id					
기준	출발 읍면동	도착 읍면동	출발 시군구	도착 시군구	출발 시도	도착 시도
컬렉션 구성	"_id":{ "L": 기준일 }	"_id":{ "L": 기준일 }	"_id":{ "L": 기준일 }	"_id":{ "L": 기준일 }	"_id":{ "L": 기준일 }	"_id":{ "L": 기준일 }
	"A1": 출발 읍면동ID, "E": 도착 분석구ID	"A": 출발 분석구ID, "E1": 도착 읍면동ID	"A2": 출발 시군구ID, "E": 도착 분석구ID	"A": 출발 분석구ID, "E2": 도착 시군구ID	"A3": 출발 시도ID, "E": 도착 분석구ID	"A": 출발 분석구ID, "E3": 도착 시도ID
	"J": 성별, "I": 연령대, "D": 출발 체류특성, "H": 도착 체류특성	"J": 성별, "I": 연령대, "D": 출발 체류특성, "H": 도착 체류특성	"J": 성별, "I": 연령대, "D": 출발 체류특성, "H": 도착 체류특성	"J": 성별, "I": 연령대, "D": 출발 체류특성, "H": 도착 체류특성	"J": 성별, "I": 연령대, "D": 출발 체류특성, "H": 도착 체류특성	"J": 성별, "I": 연령대, "D": 출발 체류특성, "H": 도착 체류특성
	}	}	}	}	}	}
	}	}	}	}	}	}

<그림 4-18> 행정구역 단위의 출발/도착 기준 통행량 DB 구조

- 구축하고자 하는 행정구역 단위의 출발/도착 별로 기준 키(필드명:_id)의 출발/도착 행정구역을 변경하여 집계함
- 교통플리곤 기준 통행량과 동일하게 출발지 행정구역인 A1(읍면동), A2(시군구), A3(시도)와 도착지 행정구역인 E1(읍면동), E2(시군구), E3(시도)가 존재함
- 행정구역 단위별로 출발, 도착, 출/도착이 모두 행정구역인 데이터를 사용자 분석 조건에 맞게 데이터를 검색할 수 있도록 하여 하나의 DB 사용보다 조금 더 안정된 서비스 제공 가능

<표 4-7> 읍면동-읍면동 단위 통행량 컬렉션의 구성

컬렉션명	TMOBILE_EMD2EMD_2020				
컬렉션 설명	읍면동-읍면동 단위로 그룹화한 통행량 데이터				
데이터 수	7천만 건 (71,189,960)		데이터 크기	약 93GB	
KEY Name	KEY	VALUE TYPE	데이터 구성	비고	
식별자	_id	Object	_id : { "A1" : 출발읍면동코드, "E1" : 도착분석구코드, "J" : 성별, "I": 연령대, "D" : 출발지 체류 유형, "H" : 도착지 체류 유형 }	A1, E1, D, H, J, I : integer 체류특성 코드 ·H:11, N:12, C:13, S:14, R:15, X: 16 성별 코드 ·F:1, M:2	

출도착시간대 통행량	CG	Object 배열	CG : [{ "C" : 출발시간대, "G" : 도착시간대, "K" : [{ "SD":출발날짜, "ED":도착날짜, "SV":총 통행량, }]	C, G, : integer K : object 배열
출발시군구	A2	INT	"A2" : 출발시군구코드	
도착시군구	E2	INT	"E2" : 도착시군구코드	
출발시도	A3	INT	"A3" : 출발시도코드	
도착시도	E3	INT	"E3" : 도착시도코드	

<표 4-8> 시군구-시군구 단위 통행량 컬렉션의 구성

컬렉션명	TMOBILE_SGG2SGG_2020			
컬렉션 설명	시군구-시군구 단위로 그룹화한 통행량 데이터			
데이터 수	3백만 건 (3,929,617)		데이터 크기	약 28GB
KEY Name	KEY	VALUE TYPE	데이터 구성	비고
식별자	_id	Object	_id : { “A2” : 출발시군구코드, “E2” : 도착시군구코드, “J” : 성별, “I” : 연령대, “D” : 출발지 체류 유형, “H” : 도착지 체류 유형 }	A2, E2, D, H, J, I : integer 체류특성 코드 ·H:11, N:12, C:13, S:14, R:15, X: 16 성별 코드 ·F:1, M:2
출도착시간대 통행량	CG	Object 배열	CG : [{ “C” : 출발시간대, “G” : 도착시간대, “K” : [{ “SD”:출발날짜, “ED”:도착날짜, “SV”:총 통행량, }]]]]	C, G, : integer K : object 배열
출발시도	A3	INT	“A3” : 출발시도코드	
도착시도	E3	INT	“E3” : 도착시도코드	

<표 4-9> 시도-시도 단위 통행량 컬렉션의 구성

컬렉션명	TMOBILE_SIDO2SIDO_2020			
컬렉션 설명	시도-시도 단위로 그룹화한 통행량 데이터			
데이터 수	3만 건 (33,119)		데이터 크기	약 3.4GB
KEY Name	KEY	VALUE TYPE	데이터 구성	비고
식별자	_id	Object	_id : { “A3” : 출발시도코드, “E3” : 도착시도코드, “J” : 성별, “I” : 연령대, “D” : 출발지 체류 유형, “H” : 도착지 체류 유형 }	A3, E3, D, H, J, I : integer 체류특성 코드 ·H:11, N:12, C:13, S:14, R:15, X: 16 성별 코드 ·F:1, M:2
출도착시간대 통행량	CG	Object 배열	CG : [{ “C” : 출발시간대, “G” : 도착시간대, “K” : [{ “SD”:출발날짜, “ED”:도착날짜, “SV”:총 통행량, }]]	C, G, : integer K : object 배열

다. 서비스 특성에 따른 집계 통행량 DB 구축

- View-T 서비스 향상을 위하여 검색, 표출 속도를 개선하기 위한 작업이 필요함
- 서비스 속도를 개선하기 위하여 모바일 데이터 속성 정보 및 통행 목적 정보를 기준으로 데이터를 재구성함

1) 통행목적별 집계 통행량 DB 구축

- 통행목적에 따른 분석 및 서비스를 제공하기 위하여 통행목적별로 통행량을 집계한 데이터 구축
 - 통행목적은 출근, 퇴근, 등교, 하교, 귀가, 여가, 관광으로 구분하여, 통행량을 집계

<표 4-10> 통행목적별 집계 통행량 컬렉션의 구성

컬렉션명	TMOBILE_PASSPURPOSE_2020			
컬렉션 설명	통행목적별 통행량을 집계한 데이터			
데이터 수	3천만 건 (31,735,211)		데이터 크기	약 4GB
KEY Name	KEY	VALUE TYPE	데이터 구성	비고
식별자	_id	Object	_id : { “A1” : 출발읍면도코드, “E1” : 도착읍면도코드, “j” : 성별, “P” : 통행목적코드 }	A1, E1, J : integer 통행 목적 코드 코드 ·1: 출근(H→C), 2: 등교(H→S), 3: 퇴근(C→H), 4:하교(S→H), 5:귀가(R→H, X→H), 6: 여가(H→R, C→R, C→X, S→R, R→R, R→X, X→R, X→X), 8: 관광(H→X) 성별 코드 ·F:1, M:2
평일 총 통행량	WDSV	Integer	“WDSV” : 평일 총 통행량	
주말 총 통행량	WESV	Integer	“WESV” : 주말 총 통행량	
평일 최종 통행량	WDFV	Integer	“WDFV” : 평일 최종 통행량	
주말 최종 통행량	WEFV	Integer	“WEFV” : 주말 최종 통행량	
출발 시군구	A2	Integer	“A2” : 출발 시군구 코드	7자리
도착 시군구	E2	Integer	“E2” : 도착 시군구 코드	7자리
출발 시도	A3	Integer	“A3” : 출발 시도 코드	7자리
도착 시도	E3	Integer	“E3” : 도착 시도 코드	7자리
거리	S	Integer	“S” : 거리	직선거리, km 단위
평균시간	T	Integer	“T” : 평균시간	분단위

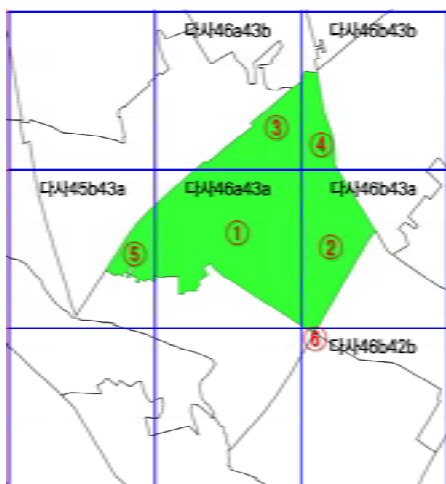
2) 서비스 기능별 DB 구축

- 검색 속도를 높이기 위하여, 각 기능별 목적에 맞추어 DB 구축
 - 기종점 인구통행량 분석 : 출 · 도착 시군구를 기준으로 통행량을 집계한 데이터 구축
 - 고령자 주요 통행 구간 분석 : 70대 이상 고령자의 통행량을 지역 기준으로 그룹화하여 데이터 구축
 - 출근 통행 접근성 분석 : 출근 통행을 읍면동, 시간대 기준으로 그룹화하여 통행량 구축
 - 퇴근 통행 접근성 분석 : 퇴근 통행을 읍면동, 시간대 기준으로 그룹화하여 통행량 구축
 - 연휴기간 통행분석 : 평상시와 연휴 기반의 통행량 차이를 분석할 수 있도록 평상시의 평균 통행량과 연휴기간 평균 통행량 구축

- 출근 통행 접근성 분석 : 출근 통행을 읍면동, 시간대 기준으로 그룹화하여 통행량 구축
- 경제활동 자체 수용도 분석 : 지역 내 출근 통행량 구축
- 도심공동화 심각도 분석 : 읍면동별 집으로 도착하는 통행량을 분석할 수 있도록 데이터 구축
- 경제활동 의존도 분석 : 시군구별 회사로 도착하는 통행량을 분석할 수 있도록 데이터 구축
- 관광지 내/외부 통행 비율 분석 : 시군구별 성별에 따른 진출입 통행량을 분석할 수 있도록 데이터 구축
- 관광 특화 지역 분석 : 읍면동별 연령, 성별, 계절에 따른 통행량을 분석할 수 있도록 데이터 구축
- 근무 형태에 따른 통근 통행 분석 : 전체 통근 통행 중 야간 통근 통행량을 분석할 수 있도록 데이터 구축
- 고용인구 변화에 따른 통근 통행 분석 : 월별 통근통행량을 분석할 수 있도록 데이터 구축

3) 격자 서비스를 위한 DB 구축

- 핫플레이스 분석에서 500m 격자 단위로 정보를 서비스하기 위하여, 격자 단위 통행량 DB 구축
- 기반지도에서 구축한 격자와 교통폴리곤 관계데이터 정보를 이용하여 격자 단위 통행량 DB 구축



교통폴리곤 111052의 통행량을 6개 격자로 분할할 시

- 다사46a43a 격자 통행량(① 구역의 통행량 +...) = (100 통행 * 50%) +... = 50 통행 +...
- 다사46b43a 격자 통행량(② 구역의 통행량 +...) = (100 통행 * 23%) +... = 23 통행 +...
- 다사46a43b 격자 통행량(③ 구역의 통행량 +...) = (100 통행 * 13%) +... = 13 통행 +...
- 다사46b43b 격자 통행량(④ 구역의 통행량 +...) = (100 통행 * 7%) +... = 7 통행 +...
- 다사45b43a 격자 통행량(⑤ 구역의 통행량 +...) = (100 통행 * 6%) +... = 6 통행 +...
- 다사46b42b 격자 통행량(⑥ 구역의 통행량 +...) = (100 통행 * 0.08%) +... = 0.08 통행 +...

* 교통폴리곤 111052의 통행량 100이라 가정

식 = 교통폴리곤 전체 통행량 X 각 격자에 포함된 교통폴리곤 면적 비율

** 분할된 통행량을 동일한 격자끼리 합하여, 격자 기준의 통행량 재산출

<그림 4-19> 격자로의 OD통행량 산출 예시

10. 모바일 데이터 가공 검증

가. 모바일 데이터 가공 검증 목적

- 데이터 가공 과정에서의 데이터 손실을 최소화하기 위하여 데이터 검증 항목을 도출하고 이를 기반으로 데이터를 검증할 수 있도록 함

나. DB별 데이터 검증

- 교통폴리곤 기준 통행량 및 행정구역 단위로 출발 기준, 도착 기준, 출발/도착 기준 통행량 데이터에 대한 검증 항목을 도출 후, 이를 기반으로 검증 진행
- 검증방식은 모바일 데이터 DB와 행정구역이 매칭된 모바일 데이터(전처리 작업 후 데이터)를 비교 검증하여 데이터 손실 여부를 검증함
- 교통폴리곤 기준 통행량 DB 검증 방법
 - 교통폴리곤 기준 통행량 DB를 가공하기 위해 참조한 모바일 데이터 전처리 DB와 항목별 통행량 합계를 비교하여 차이 발생 시 데이터 재가공 및 검증함
 - 모든 항목별 통행량 합계가 동일시 검증 완료
- 교통폴리곤 기준 통행량 DB 검증 항목은 아래와 같음

<표 4-11> 교통폴리곤 기준 통행량 DB 검증 항목

검증 항목	비고
원시 전처리 데이터와 가공 결과 데이터에 대한 출발/도착 교통폴리곤 ID별 통행량 합계 비교	출발/도착 각각의 값으로 비교 검증
원시 전처리 데이터와 가공 결과 데이터에 대한 출발/도착 시간대별 통행량 합계 비교	
원시 전처리 데이터와 가공 결과 데이터에 대한 출발/도착의 체류 특성별 통행량 합계 비교	
원시 전처리 데이터와 가공 결과 데이터에 대한 연령대별 통행량 합계 비교	-
원시 전처리 데이터와 가공 결과 데이터에 대한 성별 통행량 합계 비교	
원시 전처리 데이터와 가공 결과 데이터에 대한 교통폴리곤별 통행량 합계 비교	구축된 DB별 출발 또는 도착의 행정구역 ID로 비교함
원시 전처리 데이터와 가공 결과 데이터에 대한 읍면동별 통행량 합계 비교	
원시 전처리 데이터와 가공 결과 데이터에 대한 시군구별 통행량 합계 비교	
원시 전처리 데이터와 가공 결과 데이터에 대한 시도별 통행량 합계 비교	

- 행정구역 단위 통행량 DB 검증 방법
 - 읍면동 단위 DB를 가공하기 위해 참조한 교통폴리곤 단위 통행량 DB와 항목별 통행량 합계를 비교하여 차이가 발생했는지 검증함
 - 시군구 단위 DB를 가공하기 위해 참조한 읍면동 단위 통행량 DB와 항목별 통행량 합계를 비교하여 차이가 발생했는지 검증함
 - 시도 단위 DB를 가공하기 위해 참조한 시군구 단위 통행량 DB와 항목별 통행량 합계를 비교하여 차이가 발생했는지 검증함
 - 차이가 발생했을 경우, 집계 코드를 수정하여 데이터 재생성 후 다시 검증 진행
 - 모든 항목별 통행량 합계가 동일할 때까지 진행
- 행정구역 단위 통행량 DB 검증 항목은 아래와 같음

<표 4-12> 행정구역 단위 통행량 합계 비교 검증 항목

검증 항목	비고
행정구역이 매칭된 모바일 데이터의 합계와 시도 단위 통행량 합계	출발/도착 각각의 값으로 비교 검증
행정구역이 매칭된 모바일 데이터의 합계와 시군구 단위 통행량 합계	
행정구역이 매칭된 모바일 데이터의 합계와 읍면동 단위 통행량 합계	

- 통행목적별 집계 통행량 DB 검증 방법
 - 통행목적별 집계 통행량은 교통폴리곤 기준 통행량 DB와 행정구역 단위 통행량 DB 항목별 통행량 합계를 비교하여 차이가 발생했는지 검증함
 - 차이가 발생했을 경우, 집계 코드를 수정하여 데이터 재생성 후 다시 검증 진행
 - 모든 항목별 통행량 합계가 동일해질 때까지 진행
- 통행목적별 집계 통행량 DB 검증 항목은 아래와 같음

<표 4-13> 통행목적별 집계 통행량 합계 비교 검증 항목

비교 대상	비교 조건	비고
교통폴리곤 기준 통행량 데이터	<ul style="list-style-type: none"> - 주중/주말 조건 통행량 합계 - 통행목적 조건 통행량 합계 - 성별 통행량 합계 	출발/도착 각각의 값으로 비교 검증
행정구역 단위 출발/도착 통행량 데이터		

제2절 교통지표 개발 및 온라인 서비스 제공

1. View-T 시스템 구성

- View-T 시스템은 다양한 분석 도구와 지표, 및 정보를 이용자에게 제공하는 서비스 기능과 서비스 운영, 모니터링 등을 위한 관리자 기능으로 구성



<그림 4-20> View-T 시스템 구성도


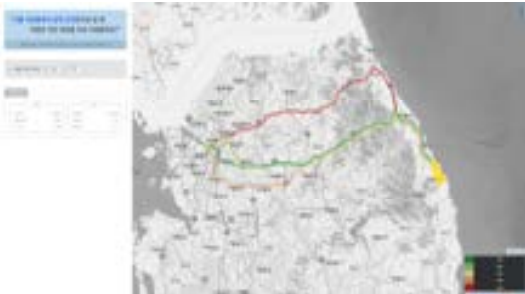
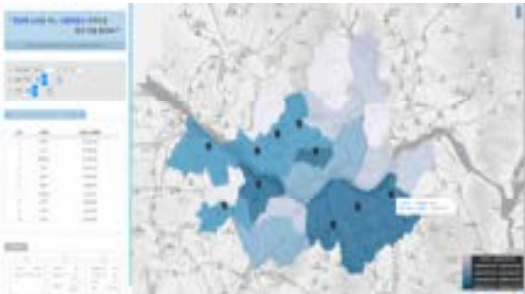
- 서비스 기능 : 교통 및 통신 데이터를 시각적으로 분석할 수 있는 Expert & Light 분석 도구와 대시보드 및 분석 결과와 교통지표 다운로드 서비스 등을 제공
- 관리자 기능 : 회원의 서비스 이용 권한을 관리하며, 회원의 서비스 이용 현황 및 다운로드 데이터 통계, 그리고 회원의 요청이나 문의에 대응하는 시스템을 제공

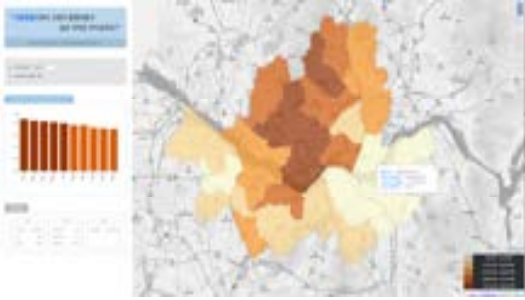




2. View-T 시스템 주요 기능 소개

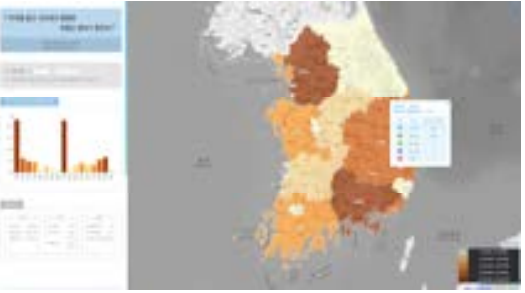
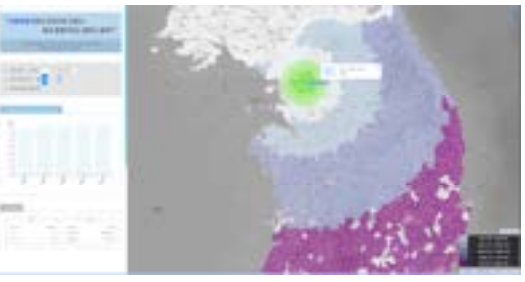
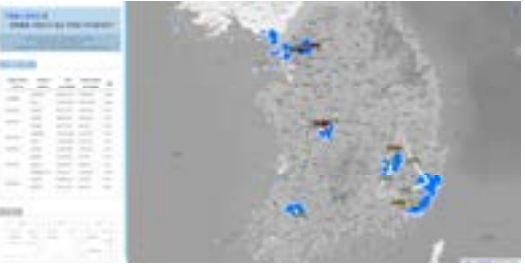

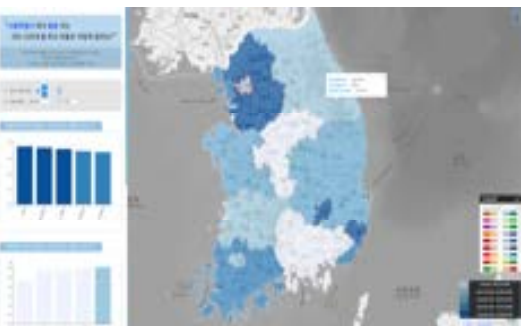
가. View-T Light



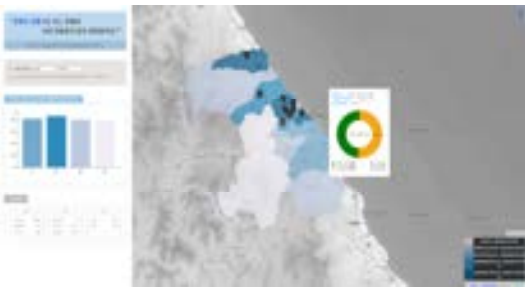


- View-T Light는 일반 사용자들도 손쉽게 모빌리티 데이터의 시공간적 특성을 다양한 분석기능을 통해 빠르고 직관적으로 분석할 수 있게 구축한 서비스임
- View-T Light는 총 24개의 기능으로 구분되어 있으며, 기종점 통행 분석, 교통 혼잡 분석 등 차량과 통신 데이터에 관련된 주제별 분석을 수행할 수 있음

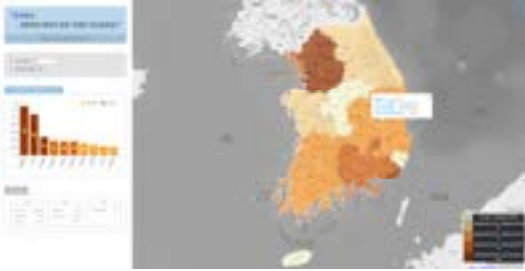
<표 4-14> View-T Light 주요 기능

구분		화면예시	설명
기종점 통행분석	기종점 인구 통행량 분석		출발지에서 다른 지역으로 이동하는 인구 통행량을 분석
	기종점 차량 경로 분석		출발지역에서 도착지역으로 이동하는 차량의 주요경로를 분석
주요 특성 통행 분석	핫플레이스 분석		전국에서 선택한 지역으로 오는 인구의 통행량을 분석

	고령자 통행 비중 분석		선택 지역에서 출발하는 고령자 통행비율을 분석
	고령자 주요 통행 구간 분석		선택한 지역 내에서 고령자의 주요 통행구간을 분석
	도심 공동화 심각도 분석		전체 유동인구 대비 상주인구를 분석하여 도심공동화가 심각한 지역을 분석
출·퇴근 통행 분석	출·퇴근 차량 영향권 분석		선택한 교차로의 진입/진출 영향권을 분석
	출·퇴근 통행 시간 분석		선택한 지역의 출근 시간대의 평균 통행시간을 분석

출·퇴근 통행량 분석		선택한 지역의 출근 시간대의 통행량 비율을 분석
통행 시간· 거리 분석		분석지역을 출발 또는 도착 기준으로 하여 전국단위로 이 동한 통행거리를 분석
경제 활동 의존도 분석		특별시·광역시를 대상으로 경 제활동을 위해 유입되는 인접시도의 출근비율을 분석
경제 활동 자체 수용도 분석		특별시·광역시에서 발생한 출 근 통행량 중 지역 내 통행량 이 차지하는 비율을 분석
근무 형태에 따른 통근 통행 분석		선택 지역에서 출발 또는 도 착하는 야간 통행(17시~23 시)의 출·퇴근 비율을 분석

여가 통행 분석	고용 인구 변화에 따른 통근 통행 분석		선택 지역에서 출발 또는 도착하는 출·퇴근 통행량의 변화를 분석
	출·퇴근 통행 연결성 분석		선택 지역에서 다른 지역으로 이동하는 출·퇴근 통행량을 분석
	관광 특화 지역 분석		관광목적의 계절별 통행량과 연령대별 통행량을 분석
	계절별 핫플레이스 분석		계절에 따른 지역별 유동인구 통행량을 분석
	연휴 기간 통행 분석		연휴기간 지역별 유동인구 통행량을 분석

	관광지 내·외부 통행비율 분석		사람들의 내·외부 통행량을 분석
--	---------------------------	--	----------------------------

나. View-T Expert

- View-T Expert는 빅데이터를 활용하여 상세 분석이 가능한 서비스로 사용자가 직접 분석 조건과 표출 설정을 통해 심도 있는 분석이 가능함
- View-T Expert는 크게 분석의 기초 교통 현황을 모니터링할 수 있는 통행지표와 모빌리티 빅데이터의 시·공간적 특성과 행태를 전국적으로 분석할 수 있는 분석 도구로 구성됨

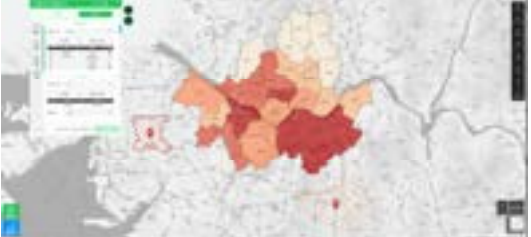





1) 통행지표

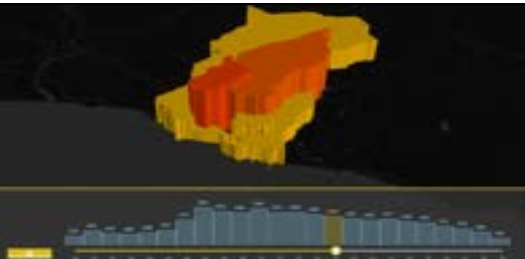
- 통행지표는 모빌리티 데이터를 이용한 다양한 분석방법의 기준이 되는 표지인 기초 정보로 전국 도로망과 공간에 대한 차량 및 인구 이동 정보를 시각화하여 제공

2) 분석 도구

- 분석 도구는 전문 이용자가 다양한 조건을 세밀하게 설정하여 View-T Light 보다 전문적이고 자세한 분석이 가능함
- 분석 사용자가 원하는 방식으로 시각화할 수 있는 다양한 시각화 옵션 및 데이터를 통한 분석이 필요한 다양한 데이터 다운로드 기능 제공

<표 4-15> View-T Expert 분석 도구 및 시뮬레이션

구분		화면예시	설명
기종점 통행 분석	기종점 인구 통행량 분석		출발지에서 다른 지역으로 이동하는 인구 통행량을 분석하는 기능
	기종점 인구 통행 구간 분석		선택한 지역 내에서 이동하는 사람들의 주요 통행구간을 분석하는 기능
주요 특성 통행 분석	핫플레이스 분석		전국에서 선택한 지역으로 오는 인구의 통행량을 분석하는 기능
	고령자 통행 비중 분석		선택 지역의 고령자 통행비율을 분석하는 기능
	고령자 주요 통행 구간 분석		선택한 지역 내에서 고령자의 주요 통행구간을 분석하는 기능
출·퇴근 통행 분석	출퇴근 차량 영향권 분석		선택한 교차로의 진입/진출 영향권을 분석하는 기능

시물레이션	인구 통행 시물 레이션		선택된 행정구역의 유입, 유출, 지역내 인구통행 시물레이션
-------	-----------------------	--	----------------------------------

다. View-T 대시보드

- View-T의 다양한 데이터를 한눈에 알아볼 수 있도록 종합적인 형태로 제공하는 대시보드 서비스



<표 4-16> View-T 대시보드 주요 기능 설명

구분	화면예시	설명
통근시간 분석 서비스		통근(출근, 퇴근)시간을 모바일 통신 데이터를 기반으로 수집하고 관련 정보를 대시보드 형태로 시각화하여 분석 및 의사결정을 지원하는 서비스

라. View-T 데이터 다운로드 기능

- 사용자가 직접 조건을 설정하여 필요한 View-T 데이터를 직접 수집할 수 있도록 하는 서비스

<표 4-17> View-T 부가 기능 설명

구분	화면예시	설명
다운로드		View-T에서 서비스하는 형상정보와 지표 정보를 사용자에게 다운로드 제공
Open API		View-T에서 사용하는 데이터를 Open API 서비스로 사용자에게 제공

3. 모바일 통신 데이터 기반 통행지표 서비스 개발

가. 개발 배경

- 출퇴근 시간은 OECD 웰빙 지표로 사용될 만큼 국민의 삶의 질에 많은 영향을 미치는 요소이며, 정책 의사결정자와 국민들에게 큰 관심사로 작용함
- 도심화가 더욱 심각하게 진행되고 있는 실정으로 향후 통상적인 출퇴근 시간이 증가할 것으로 보이며, 이를 관리하기 위한 의사결정자들에게 더욱 중요한 지표가 될 것으로 예상
- 이에 따라 정기적이며, 직관적으로 정보를 전달할 수 있는 종합적 정보 제공 서비스를 기획하여 개발함

나. 기능 소개

- 국민의 삶의 질에 큰 영향을 미치는 요소인 통근(출근, 퇴근)시간을 모바일 통신 데이터를 기반으로 수집하고 관련 정보를 대시보드 형태로 시각화하여 분석 및 의사결정을 지원하는 종합적인 지표 서비스



<그림 4-21> 모바일 통신 데이터 통행지표 서비스 화면

다. 대시보드 화면 구성

- 대시보드는 다양한 종류의 위젯을 제공하여 사용자에게 분석 결과를 동적으로 제공
- 대시보드는 사용자가 위젯을 원하는 형태로 추가, 수정, 설정 등 기능을 제공

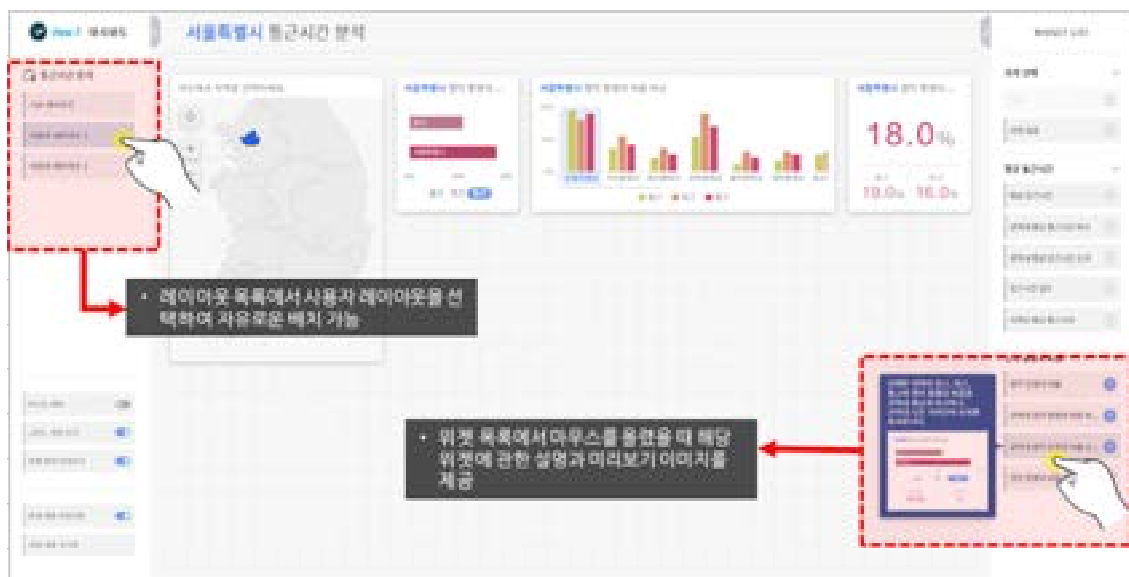


<그림 4-22> 통행지표 서비스 대시보드 화면 구성

- 화면은 레이아웃 메뉴 영역, 옵션 설정 영역, 위젯 표출 영역, 위젯 목록 영역으로 구분하여 배치
- 레이아웃 메뉴 영역 : 레이아웃 목록으로 구성
 - 기본으로 제공하는 레이아웃 및 사용자 정의 레이아웃으로 구성
- 옵션 설정 영역 : 대시보드 테마 변경, 저장된 레이아웃 초기화, 그리드 보기 옵션 ON/OFF 등의 버튼으로 구성
- 위젯 표출 영역 : 위치 변경, 크기 변경, 옵션 설정 등의 기능을 가진 위젯의 배치에 따라 구성이 변경됨
- 위젯 추가 영역 : 대시보드의 추가 가능한 위젯의 목록으로 구성

라. 기능 설명

- 단순한 그래프, 값 표출만이 존재하는 정적인 표현 방식을 탈피하고 화면에 동적인 표현 방식을 도입하여 이용자의 흥미도를 높임
- 종합적 정보를 직관적으로 전달할 수 있는 대시보드 형태로 구성
 - 평균 출근시간 : 기준 지역으로 출근하는 통행자들의 평균 출근시간
 - 평균 퇴근시간 : 기준 지역에서 퇴근하는 통행자들의 평균 퇴근시간
 - 평균 통근시간 : 기준 지역 통행자들의 평균 통근시간
 - 열악 통행자 비율 : 편도 통근시간이 1시간 이상 소요되는 통행자를 열악 통행자로 분류하여 전체 통행자 대비 차지하는 비율
- 위젯 시스템을 도입하여 사용자가 원하는 형태로 화면 배치를 변경 가능
- 위젯 목록을 통해 위젯을 추가하고 중복 생성을 허용



<그림 4-23> 사용자 레이아웃 및 위젯 추가하기 예시 화면

- 사용자 편의성을 위한 다양한 대시보드 옵션 제공
 - 위젯 위치 미리보기 : 위젯을 이동할 때 실제 위젯이 위치하게 될 곳을 미리보기
 - 변경 내용 자동 저장 : 수정된 레이아웃을 페이지가 종료될 때마다 자동으로 저장하며, 사용자가 다시 대시보드 페이지를 열었을 때 마지막으로 저장된 레이아웃이 적용됨
 - 저장내용 초기화 : 저장한 내용을 모두 삭제하고 초기 레이아웃 상태로 돌아감

- 테마 변경 : 사용자 편의성을 고려해 어두운 테마 / 밝은 테마 중 선택이 가능하도록 함



<그림 4-24> 어두운 테마 / 밝은 테마 비교 화면

그리드 라인 보기 : 가상의 그리드 라인을 시각화하여 위젯 이동이나 크기 변경 시 보조



<그림 4-25> 그리드 라인 보기 옵션 ON/OFF 비교 화면

1) 위젯 시스템

- 위젯이란, 독립적으로 수행되는 작은 프로그램이라는 뜻으로 다양한 지표 정보를 사용자의 취향이나 특성에 맞는 형태로 직접 배치할 수 있도록 하기 위해 위젯 시스템을 적용
- 각각의 위젯은 개별적으로 동작하되, 기준 지역을 설정할 수 있는 지역 선택 위젯을 참조
- 위젯을 배치하는 화면을 가상의 격자 형태로 나누어 위젯의 위치와 크기를 격자에 종속시킴

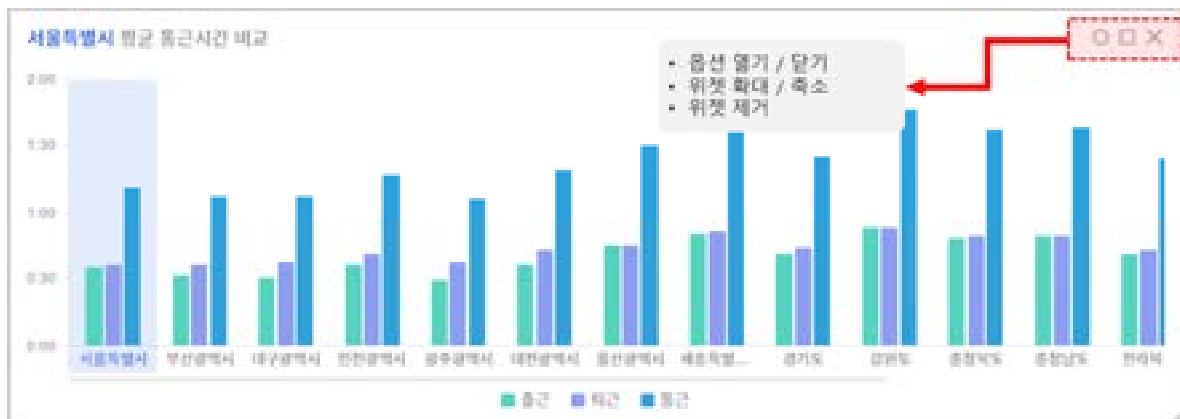
① 위젯의 위치와 크기 변경



<그림 4-26> 위젯의 위치 이동 및 크기 변경을 통한 사용자 레이아웃 예시 화면

② 위젯의 메뉴

- 메뉴 버튼을 통해 위젯을 제거, 확대 보기, 옵션 기능을 실행함



<그림 4-27> 위젯 메뉴 버튼 설명 화면

2) 위젯의 종류

① 지역 선택 위젯

- 지역 선택 위젯을 통해 통행지표 정보를 표출할 기준 지역을 선택
- 지도 지역 선택 위젯과 목록 지역 선택 위젯으로 2가지 종류로 구분



<그림 4-28> 지도 지역 선택 위젯(좌)과 목록 지역 선택 위젯(우)

- 지역 선택 위젯에서 지역을 변경할 때마다 그 외 다른 위젯들의 정보가 즉시 변경됨

② 평균 통근시간 / 열악 통행자 비율

- 지역 선택 위젯에서 선택한 기준 지역의 평균 출근 시간, 퇴근 시간, 통근시간과 출근 시 열악 통행자 비율, 퇴근 시 열악 통행자 비율, 통근 열악 통행자 비율의 정량적인 값을 제공하는 위젯 2종
- 표출 정보를 전국 기준 또는 기준 지역 중 선택하는 옵션과 기준 지역으로 표출 시 지도 위젯의 영향을 받지 않고 고정할 수 있는 옵션을 제공



<그림 4-29> 평균 통근시간 위젯 및 전국 값 표시 옵션 예시 화면



<그림 4-30> 열악 통행자 비율 위젯 및 현재 지역 표시 고정 옵션 예시 화면

③ 평균 통근시간 비교

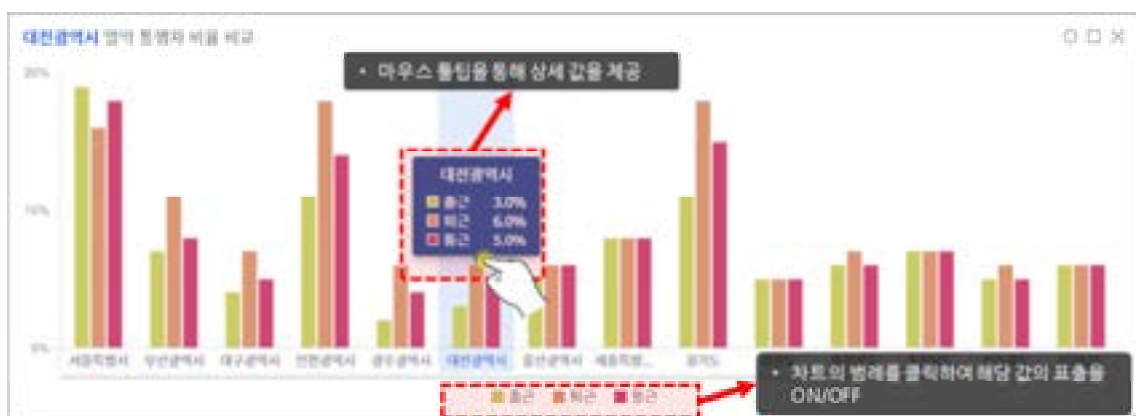
- 지역 선택 위젯에서 선택한 기준 지역을 포함하고 있는 권역 내의 모든 지역의 평균 통근시간을 비교하는 차트 위젯
- 차트의 종류는 묶음형 막대 차트로 지역의 출근/퇴근/통근 막대를 묶어 나열한 형태



<그림 4-31> 평균 통근시간 비교 위젯 설명

④ 열악 통행자 비율 비교

- 지역 선택 위젯에서 선택한 기준 지역을 포함하고 있는 권역 내의 모든 지역의 출근, 퇴근, 통근 시 열악 통행자 비율을 비교하는 차트 위젯
- 차트의 종류는 묶음형 막대 차트로 지역의 출근/퇴근/통근 막대를 묶어 나열한 형태



<그림 4-32> 열악 통행자 비율 비교 위젯 설명

⑤ 평균 통근시간 순위

- 지역 선택 위젯에서 선택한 기준 지역의 상위 권역의 평균 통근시간과 기준 지역의 평균 통근시간 값을 비교하고 권역 내 순위 정보를 제공하는 위젯



<그림 4-33> 평균 통근시간 순위 위젯

⑥ 열악 통행자 비율 순위

- 지역 선택 위젯에서 선택한 기준 지역의 상위 권역의 열악 통행자 비율과 기준 지역의 열악 통행자 비율 값을 비교하고 권역 내 순위 정보를 제공하는 위젯



<그림 4-34> 열악 통행자 비율 순위 위젯

⑦ 통근시간 밀도

- 지역 선택 위젯에서 선택한 기준 지역으로 통근하는 전체 통행자를 통행시간의 분 단위 기준으로 밀도 산정값 시각화하여 제공하는 위젯



<그림 4-35> 통근시간 밀도 위젯

⑧ 열악 통행자 비율 분포

- 지역 선택 위젯에서 선택한 기준 지역으로 통근시간이 편도 1시간 이상 소요되는 열악 통행자가 차지하는 비율 정보를 상세하게 제공하는 위젯



<그림 4-36> 열악 통행자 비율 분포 위젯

⑨ 지역 간 평균 통근시간

- 지역 선택 위젯에서 선택한 기준 지역과 타 시도 단위 지역 간의 평균 통근시간을 종합적으로 제공
 - 기준 지역에서 타 시도 지역으로의 평균 출근, 퇴근 시간
 - 타 시도 지역에서 기준 지역으로의 평균 출근, 퇴근 시간
- 출발 지역과 도착지역 간 통근시간 관계를 파악하기 용이한 형태로 개발한 차트로 표출
 - (중앙) 상대 시도 지역명을 표출
 - (좌) 기준 지역에서 출발하여 상대 지역으로 도착하는 통행자의 통근시간 값 표시
 - (우) 상대 지역에서 출발하여 기준 지역으로 도착하는 통행자의 통근시간 값 표시

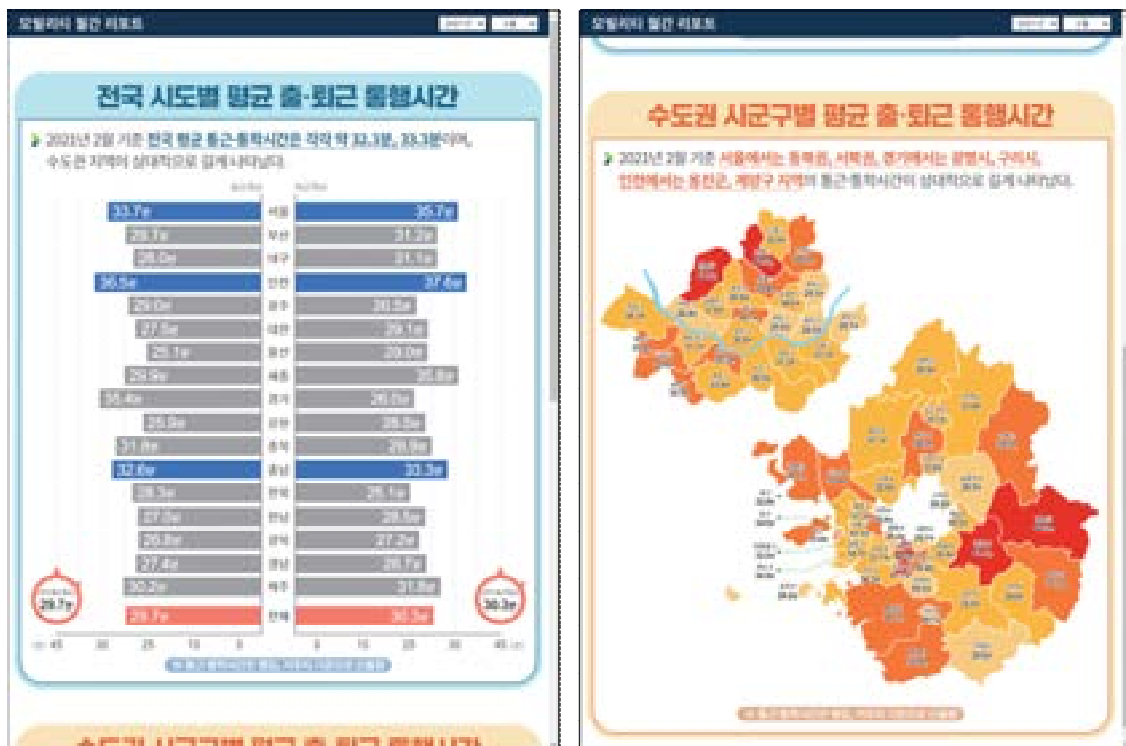


<그림 437> 지역 간 평균 통근시간 위젯

4. 월별 교통 리포트 업로드

가. 기능 소개

- 2021년도 월별 전국 시·도, 수도권 시·군·구별 평균 통근·통학시간 데이터를 한눈에 비교할 수 있는 차트 페이지를 개발
- View-T 홈페이지 상단 메뉴에서 “다운로드 → 모빌리티 월간 리포트” 메뉴를 클릭하면 별도의 팝업창으로 표출
- 전국 시·도별 평균 통근·통학시간의 막대 그래프 차트와 수도권 시·군·구별 평균 통근·통학시간의 지도 차트로 구성되며, 마우스 스크롤 조작을 통해 각각의 내용을 확인



<그림 4-38> 모빌리티 월간 리포트 화면

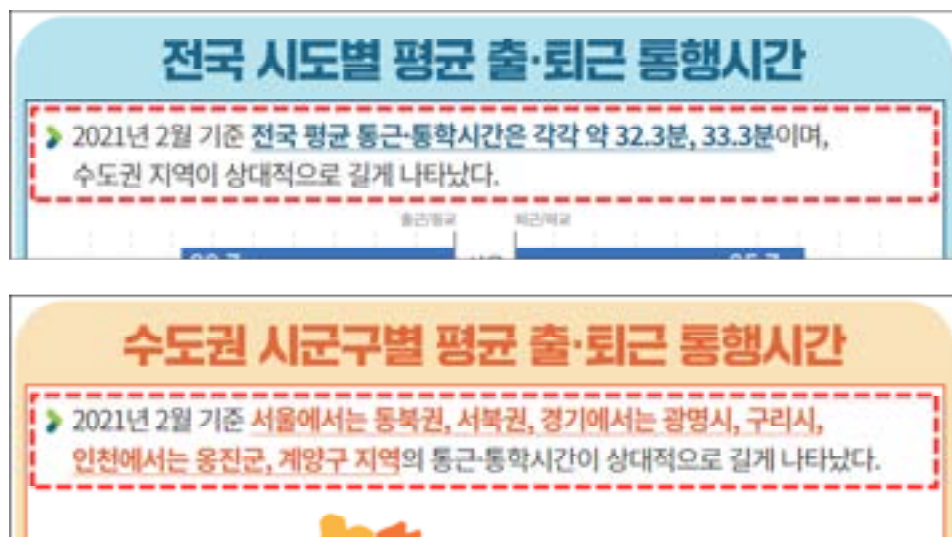
나. 기능 설명

- 표출을 위해 사용되는 2021년도 월별 데이터를 JSON 형태로 작성하여 리포트 페이지에서 확인하고자 하는 기간을 선택하여 동적으로 내용을 확인 가능



<그림 4-39> 모빌리티 월간 리포트 기간 선택 UI

- 전국 시·도별 평균 통근·통학시간은 막대그래프를 통해 각각의 결과값을 비교하기 용이한 형태의 차트로 표현하고, 차트에서 통근·통학시간이 가장 긴 3개 지역의 막대 색상을 강조하여 표현
- 전국 시·도별 평균 통근·통학시간 차트의 마지막 항목에서 전체 데이터의 평균 수치를 계산하여 표출
- 전국 시·군·구별 평균 통근·통학시간은 서울 및 수도권 지역의 지도에 각 지역의 결과값을 표기하고 색상으로 결과값을 수준을 표현 (색상이 붉어질수록 높은 수치)
- 각 전국 시·도, 수도권 시·군·구별 평균 통근·통학시간 차트에서는 표출 결과를 요약한 문구를 표출하여 결과 이해를 도움

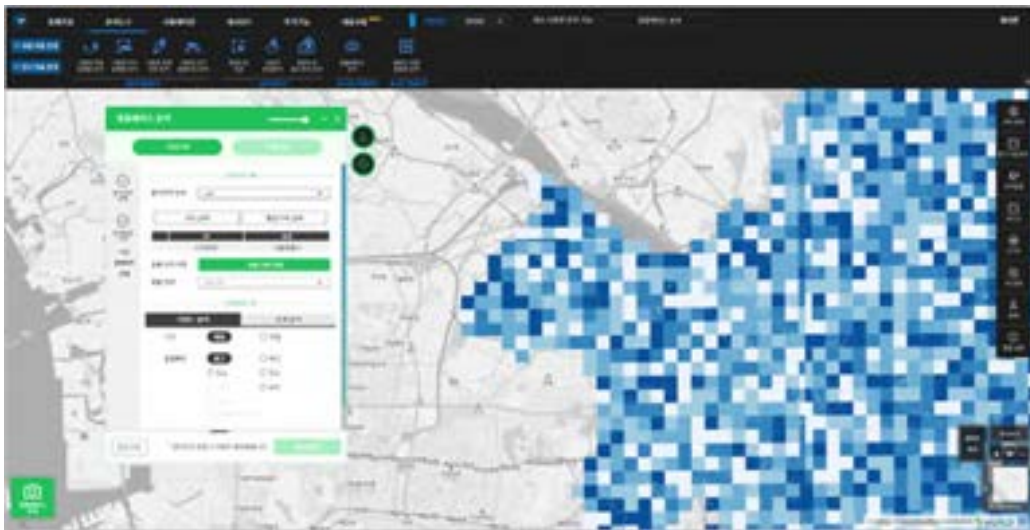


<그림 4-40> 모빌리티 월간 리포트의 결과 요약 내용

5. Hotplace 분석 도구 개선

가. 기능 소개

- 특정 지역으로 유입되는 통행량이 어디가 많은지 확인할 수 있도록 함
- 기존 관광 특화지역 분석, 계절별 핫플레이스 기능을 통합하여 HotPlace의 상세 분석에서 사용할 수 있도록 기능을 변경함



<그림 4-41> HotPlace 분석 도구 화면

나. 기능 설명

- 사용자가 표출지역 단위를 전국, 시도, 시군구를 선택하여 분석하고 하는 지역을 선택할 수 있도록 기능을 제공

ID	명칭
1100000	서울특별시

표출 단위 자동 ☒ 표출 단위 자동

표출 단위

1. 표출 단위 자동 설정은 : On/Off 제공

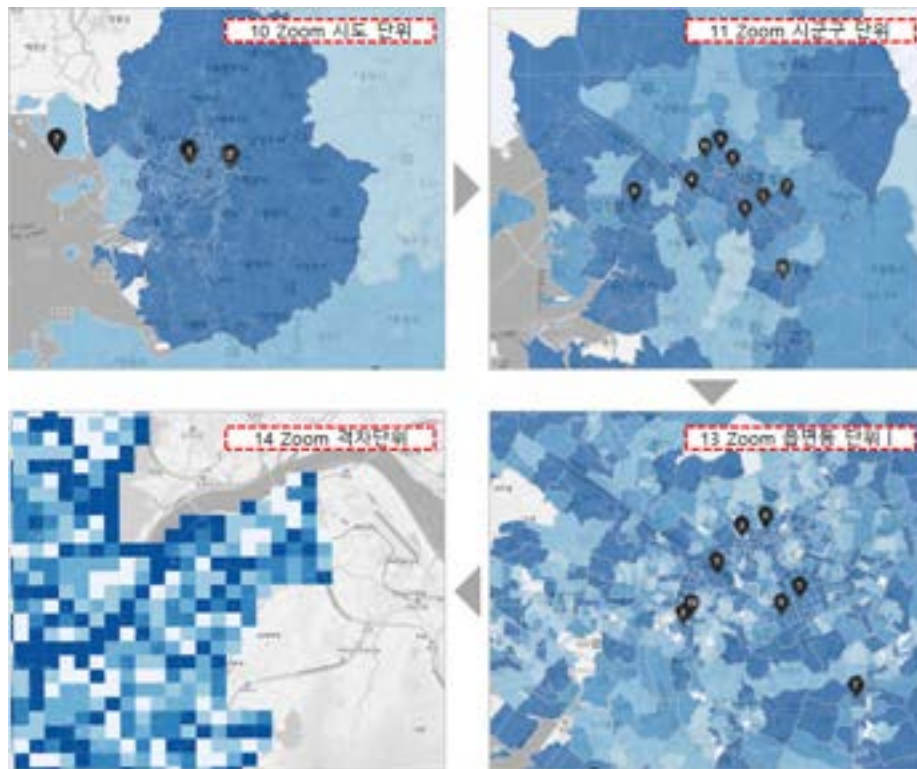
2. 표출 단위는 분석지역 단위 기준으로 정의

3. 자동설정 On 설정시 지도 Zoom에 따라 표출되는 단위가 자동으로 변경됨

4. 시도,시군구,읍면동,국가단위 제공

<그림 4-42> 표출 단위 자동 설정 화면

- 표출 단위 자동 선택을 사용하여 지도 스케일에 따른 표출 단위가 변경되도록 함
- 표출 단위는 시도, 시군구, 읍면동, 격자 단위를 제공하여 상세하게 분석할 수 있도록 함



<그림 4-43> 표출 단위 자동 설정 적용 화면

6. 데이터 다운로드 현황분석 서비스 개발

가. 기능 소개

- View-T에서 제공하는 Faster Indicator, 데이터 다운로드 기능(차량 데이터, 형상정보 데이터)을 사용자의 이용 정보를 확인할 수 있는 관리자 기능 개발함
- 관리자 기능의 데이터 다운로드 현황 페이지는 기능별 다운로드 횟수 및 다운로드 서비스 이용자의 소속 정보를 확인할 수 있도록 개발함

<그림 4-44> 관리자 다운로드 현황 화면

나. 기능 설명

- 관리자 페이지에서 '데이터 다운로드 현황' 버튼을 클릭하여 페이지 이동
- 데이터 다운로드 현황 또는 사용자 이용 현황을 클릭하면 해당 페이지가 호출
- 데이터 다운로드 현황은 '요청 데이터명, 공간적 범위, 시간적 범위'가 같은 내역 정보의 다운로드 횟수의 합계를 호출
- 사용자 이용 현황은 '소속기관, 데이터 활용 목적, 데이터 구분'이 같은 내역 정보의 다운로드 횟수의 합계를 호출

다. 기능 요약

- 관리자가 사용자의 데이터 다운로드 현황 테이블 형태로 확인할 수 있도록 함

7. 분석 도구 차량 대시보드 기능 개선

- 시도별, 시군구별, 연도별 차량 통행지표 대시보드 유지보수
 - 그래프와 주제를 구분하기 쉽도록 색상의 차이를 주어 가독성 및 가시성을 높임
 - 주제부는 그래프의 주제를 알림과 동시에 주제 변경 기능을 가진 메뉴 버튼임을 쉽게 이해할 수 있도록 화살표 아이콘을 추가



<그림 445> 분석 도구 차량 유지보수 전후 비교 화면

- 연도별 차량 대시보드는 기존에 시군구 단위로 서비스를 제공하였으나 시도 단위의 분석을 추가하여 분석의 폭을 넓힘

8. View-T 회원 시스템 개선

- 홈페이지 회원 관련 UI 개선
 - 회원 정보의 수집을 최소화하여 사용자 편의성 개선



<그림 4-46> 홈페이지 로그인, MyPage 삭제 화면

9. 데이터 다운받기 기능 개선

- 데이터 다운받기 단계 축소
 - 기존 단계는 데이터 다운로드 요청 -> 신청서 작성 -> 관리자 데이터 제공 수락 또는 거절 -> MyPage의 데이터 신청 상황 표시 -> 수락된 데이터 다운로드
 - 변경된 단계는 데이터 다운로드 선택 -> 신청서 작성 -> 데이터 다운로드



<그림 4-47> 데이터 다운로드 단계 축소

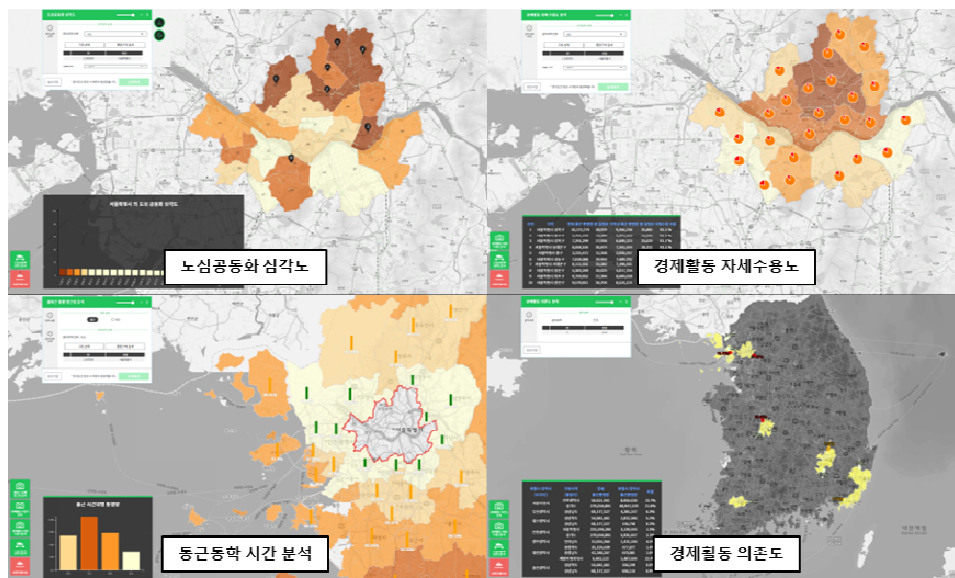
10. 사회경제활동지표 추가 및 데이터 현행화

가. 지표의 종류 및 설명

- 통근·통학 시간 분석 : 통근·통학에 따른 시간대별 통행량과 통행시간 분석지표
- 경제활동 의존도 : 지역 간 출/퇴근 통행량에 경제적 활동 의존성 분석지표
- 경제활동 자체수용도 : 지역 간 출/퇴근 통행량에 따른 경제활동 자체 수용 비율 분석지표
- 도심공동화 심각도 : 유동인구 대비 상주 인구수를 통해 산출한 도심공동화 심각 지수

나. 현행화 내용

- 기존 분석 도구인 출퇴근 통행 접근성 분석을 통근·통학 시간 분석으로 통행지표로 변경
- 기존 분석 도구인 경제활동 의존도 분석을 통행지표로 변경
- 기존 분석 도구인 경제활동 자체수용도 분석을 통행지표로 변경
- 기존 분석 도구인 도심공동화 심각도 분석을 통행지표로 현행화
- 2020년 데이터 현행화



<그림 4-48> 사회경제활동지표 화면예시

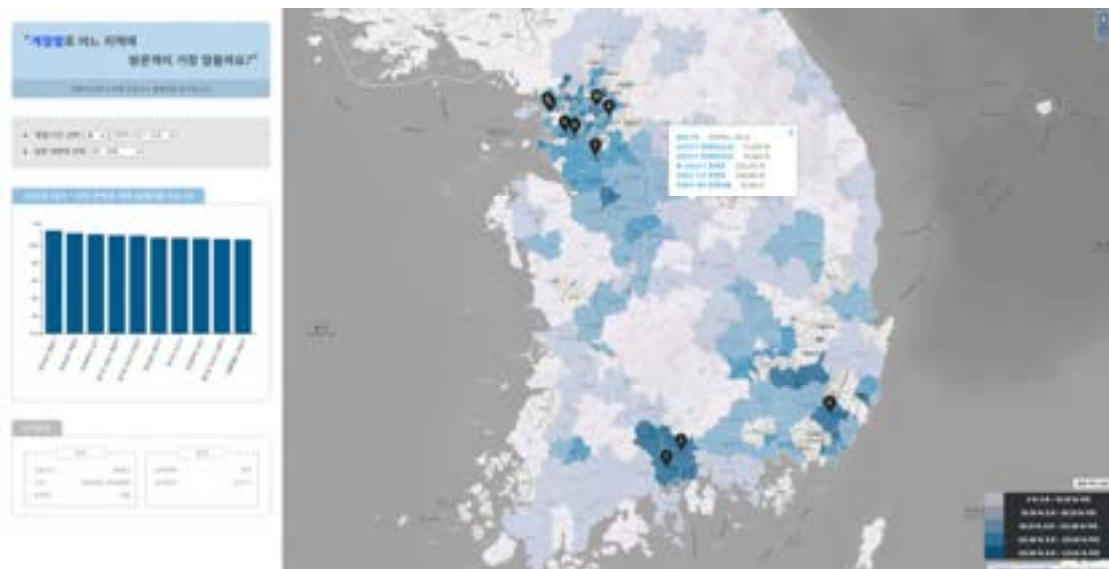
11. 계절별 핫플레이스 분석 도구 현행화

가. 기능 소개

- 계절별로 유동인구량이 많은 핫플레이스 기준 통행행태를 분석하는 기능
- 전국에서 특정 지역으로 유입되는 이동인구와 특정 지역에서 유출되는 이동인구 중 가장 이동이 많은 지역을 분석 및 표출함
 - 계절 구분 및 계절별 통계량 산출
 - 봄 : 3~5월, 여름 : 6~8월, 가을 9~11월, 겨울 : 12월 & 1~2월
 - 일주일 단위로 읍면동별 유입량, 유출량 집계
 - 읍면동별 연평균 유입량 유출량 집계
 - 읍면동별 연평균 유입량 대비 일주일 단위 유입량이 급증하는 지역 1~5위 표출 및 주별 업데이트
 - 우측에는 주제도를 표현하고, 좌측에는 우측 주제도의 조건 및 내용을 설명하고 우측 하단에는 범례 표시

나. 현행화 내용

- 2020년 데이터 현행화 및 유지보수



<그림 4-49> 계절별 핫플레이스 분석 도구 화면예시

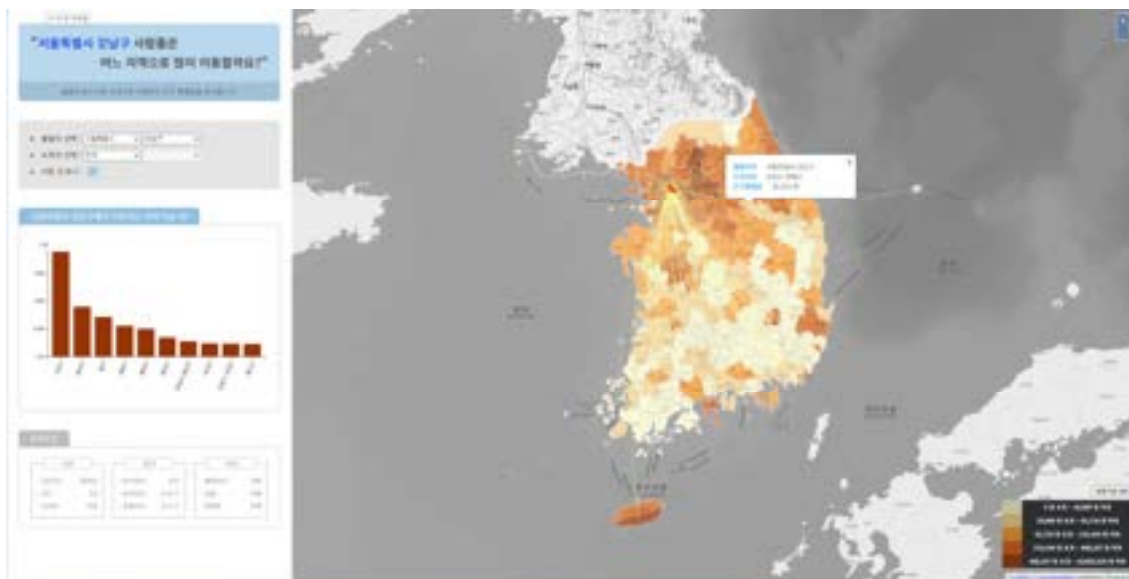
12. 주요 통행지역 분석 도구 현황화

가. 기능 소개

- 모바일 통신 데이터 기반 기종점별 유동인구량(통행량)을 분석하는 기능
- 특정 출발지에서 특정 도착지의 지역별 유동인구 분석 및 표출함
 - 출발지와 도착지를 선택하여 지역별 유동인구수 분석
 - 출발지를 제외한 다른 시/군/구를 선택하면 기존 선택시 시/군/구에서 새로 선택한 시군구로의 유동인구를 표출
 - 선택된 2개의 시/군/구 중 하나를 클릭하면 선택 해제되어 나머지 선택지역에서 전국으로의 유동인구 표출

나. 현황화 내용

- 2020년 데이터 현황화 및 유지보수



<그림 4-50> 주요 통행지역 분석 도구 화면예시

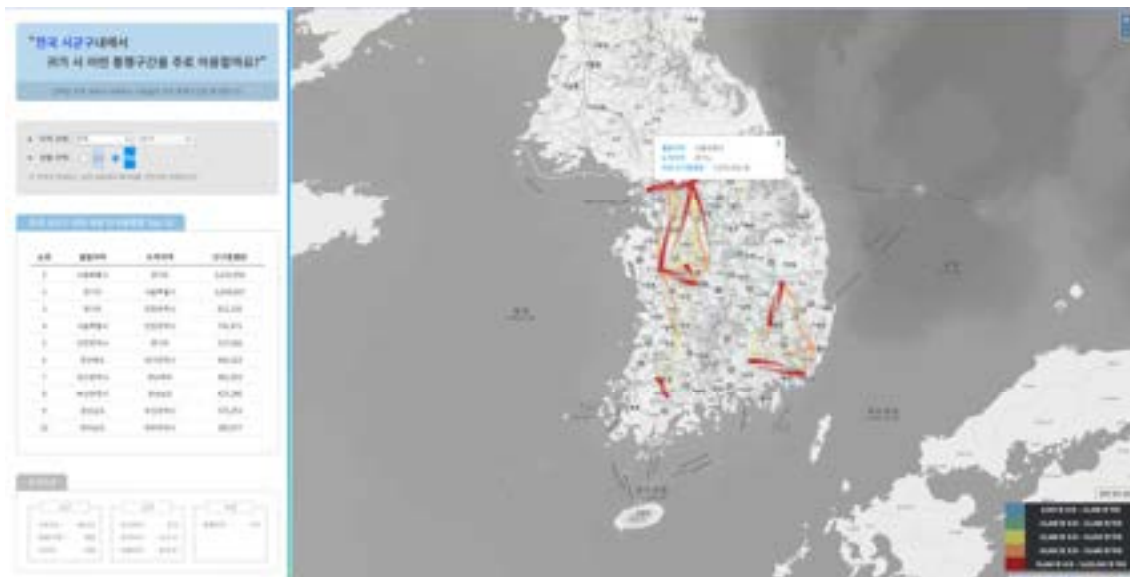
13. 주요 통행구간 분석 도구 현행화

가. 기능 소개

- 모바일 통신 데이터 기반 기종점별 통행 구간을 분석하는 기능
- 통행이 많이 일어나는 구간을 분석함
 - 시도 간 통행이 많이 일어나는 구간을 분석
 - 기본 주제도 출력은 전국의 시도 단위에서 각각 시도에서의 출발 및 도착의 이동인구 수를 희망선도로 표현
 - 주제도 상에 마우스 오버 이벤트로 해당 지역의 정보를 표현
 - 출발지역, 도착지역, 통행량을 표현
 - 좌측에는 우측 주제도의 조건 및 상위 10개 희망선도의 내용 표현
 - 우측 하단에는 범례를 표출

나. 현행화 내용

- 2020년 데이터 현행화 및 유지보수



<그림 4-51> 주요 통행구간 분석 도구 화면예시

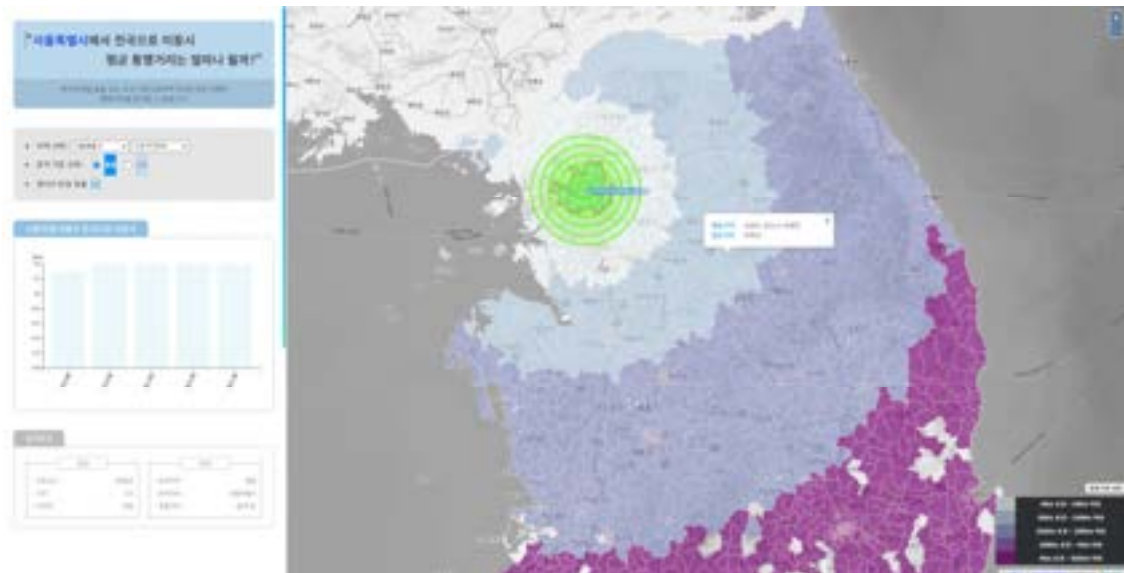
14. 통행시간·거리 분석 도구 현행화

가. 기능 소개

- 모바일 통신 데이터 기반 기종점별 통행시간·거리를 분석하는 기능
- 선택지역을 출발 또는 도착 기준으로 하여 각 지역별로 이동한 평균 통행시간과 통행 거리를 산출함
 - 주제도 상에 마우스 오버 이벤트로 해당 지역의 정보를 표현
 - 행정구역명, 평균 통행시간, 평균 통행 거리를 표현
 - 좌측에는 우측 주제도의 조건을 표출하고 설정 및 분석 결과의 5개 지역을 그래프로 표출
 - 우측 하단에는 범례를 표출

나. 현행화 내용

- 2020년 데이터 현행화 및 유지보수



<그림 4-52> 통행시간·거리 분석 도구 화면예시

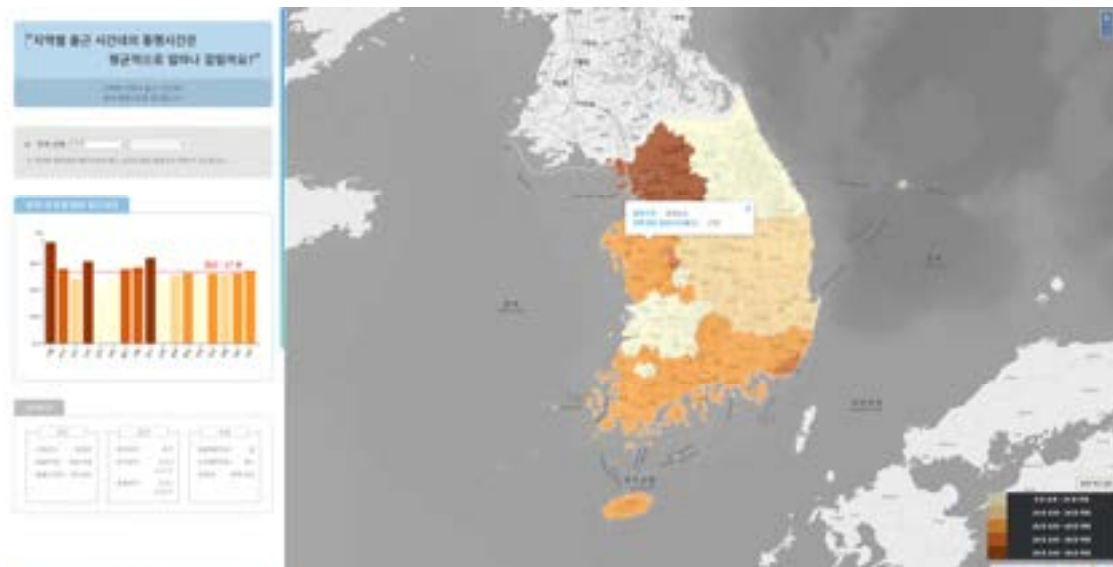
15. 출·퇴근 통행시간 분석 도구 현행화

가. 기능 소개

- 모바일 통신 데이터 기반 출발지역 기준 출·퇴근 통행시간을 분석하는 기능
- 지역별 출·퇴근 통행시간을 분석함
 - 데이터 가공 방법
 - 지역별 출·퇴근 통행을 추출하여 평균 통행시간을 산출
 - 출·퇴근 통행시간 분석 UI
 - 기본 주제도는 전국 시도별 평균 출근시간을 색상으로 표출함과 동시에 각각의 시도별 출근시간 분포를 나타내는 그래프를 추가(예: 30분 이하, 30분 이상, 60분 이하 등의 카테고리별로 나누어 비율 표출)
 - 주제도 상에 마우스 오버 이벤트로 6~9시 각각의 시간대에서의 평균 통행시간을 표출
 - 우측 하단에는 범례 표출

나. 현행화 내용

- 2020년 데이터 현행화 및 유지보수



<그림 4-53> 출·퇴근 통행시간 분석 도구 화면예시

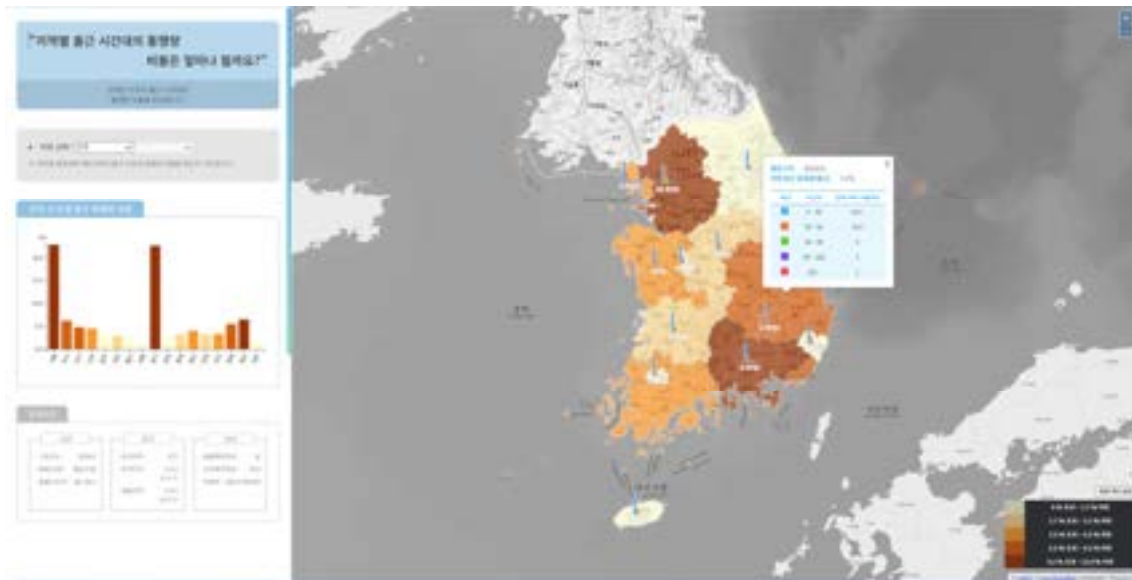
16. 출·퇴근 통행량 분석 도구 현행화

가. 기능 소개

- 모바일 통신 데이터 기반 출발지역 기준 출·퇴근 통행량 비율 분석하는 기능
- 지역별 출·퇴근 통행량을 분석함
 - 데이터 가공 방법
 - 지역별 출·퇴근 통행을 추출하여 지역별 비율을 산출
 - 출·퇴근 통행량 분석 UI
 - 기본 주제도는 전국 시도별 출·퇴근 통행량 비율을 색상으로 표출함
 - 주제도 상에 마우스 오버 이벤트로 6~9시 각각의 시간대에서의 통행량 비율을 표출
 - 우측 하단에는 범례 표출

나. 현행화 내용

- 2020년 데이터 현행화 및 유지보수



<그림 4-54> 출·퇴근 통행량 분석 도구 화면에서

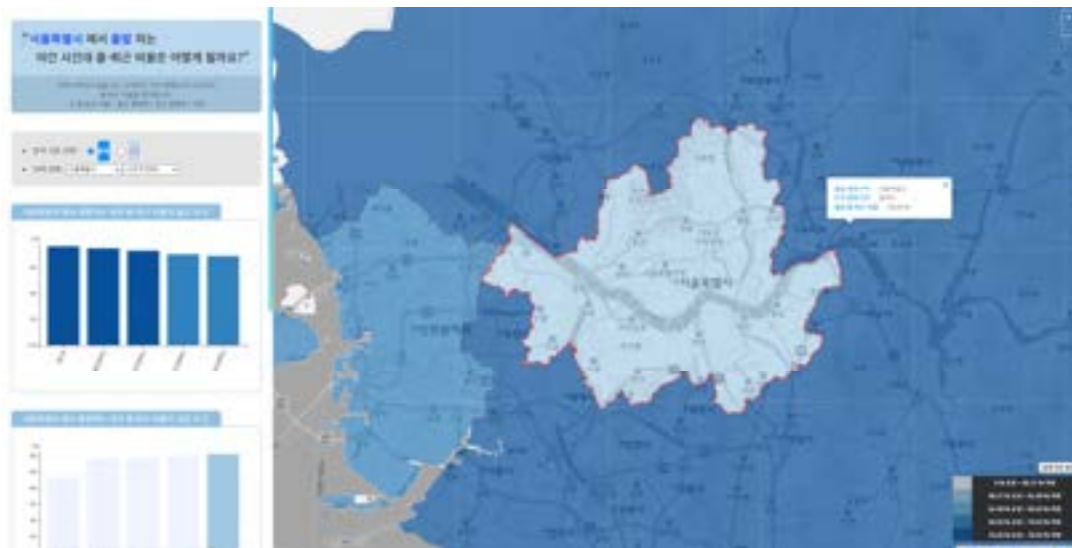
17. 근무형태에 따른 통근 통행 분석 도구 현행화

가. 기능 소개

- 모바일 통신 데이터 기반 지역별·시간대별 통근 통행 비율을 분석하는 기능
- 지역별 야간 출·퇴근 통행 비율을 분석
 - 데이터 가공 방법
 - 월별 통근통행량 변화량을 비교·분석하여 고용인구 변화율 산출
 - $Y(\text{통근통행량}) = X(\text{월}) \times A(\text{고용인구변화율}) \pm \varepsilon$
 - 근무 형태에 따른 통근 통행 분석 UI
 - 기본 주제도는 전국 야간 출·퇴근 통행 비율을 색상으로 표출
 - 주제도 상에 마우스 오버 이벤트로 해당 지역의 심야시간 출·퇴근 통행 비율을 표출
 - 우측에는 주제도, 좌측에는 선택지역을 기준으로 심야시간 출·퇴근 통행 비율이 높은 시도에 대한 상위지역을 표출
 - 좌측 하단에는 전국 시도별 지가지수와 심야시간대 출·퇴근 통행비율 그래프를 표출

나. 현행화 내용

- 2020년 데이터 현행화 및 유지보수



<그림 4-55> 근무형태에 따른 통근 통행 분석 도구 화면예시

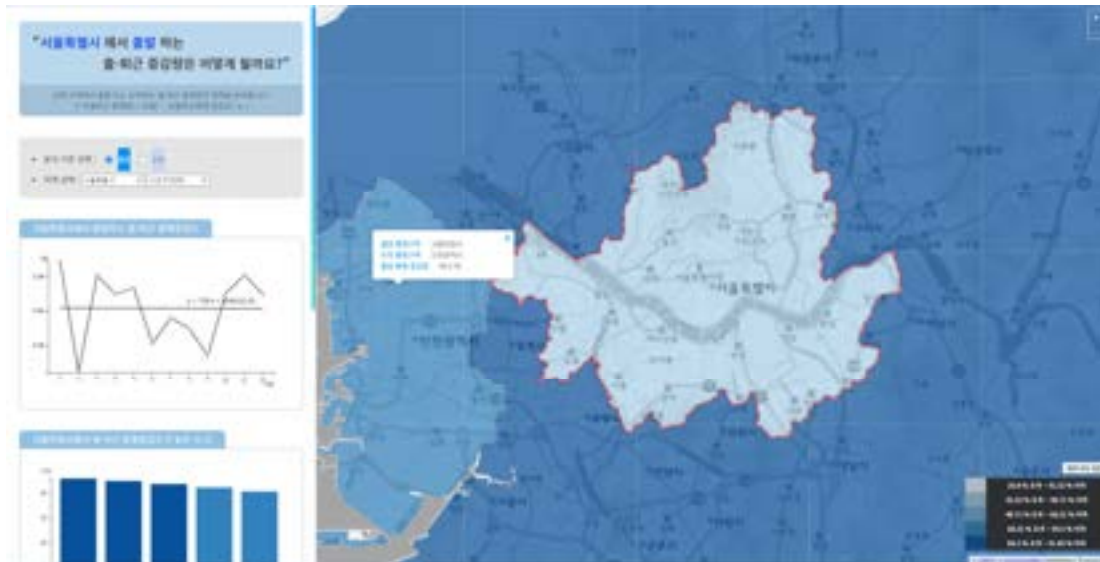
18. 고용인구 변화에 따른 통근 통행 분석 도구 현행화

가. 기능 소개

- 모바일 통신 데이터 기반 지역별·월별 통근 통행 변화 분석하는 기능
- 지역별 출·퇴근 통행량의 변화를 분석함
 - 데이터 가공 방법
 - 월별 통근 통행량을 추출하여 증감도를 산출
 - 고용인구 변화에 따른 통근 통행 분석 UI
 - 기본 주제도는 통근 통행 증감량을 색상으로 표출
 - 주제도 상에 마우스 오버 이벤트로 해당 지역의 통행 증감량을 표출
 - 우측에는 주제도, 좌측에는 월별 통행 증감도 그래프와 선택지역을 출발 기준으로 두고, 출·퇴근 통행증감도가 높은 상위 도착지역 표출
 - 우측 하단에는 범례 표출

나. 현행화 내용

- 2020년 데이터 현행화 및 유지보수



<그림 4-56> 고용인구 변화에 따른 통근 통행 분석 도구 화면예시

19. 출·퇴근 통행 연결성 분석 도구 현행화

가. 기능 소개

- 모바일 통신 데이터 기반 기종점 간 통근 통행 연결성 분석하는 기능
 - O/D간 통행에 대해 그룹화하여 지역 간 통행 연결성을 분석함
- 데이터 가공 방법

Social network analysis_Modularity

$$Q = \frac{1}{2M} \sum_{i,j} (a_{ij} - \frac{k_i k_j}{2M}) \delta[C(i), C(j)]$$

Q	모듈성
M	전체 링크 수
N	전체 노드 수
a _{ij}	i, j간 링크 (있을 경우 1, 없을 경우 0)
k _i	노드 i가 지나는 링크의 개수
C(i)	노드 i가 속하는 커뮤니티
δ(C(i), C(j))	C(i)와 C(j)가 같은 커뮤니티일 때 1, 다를 때 0

<그림 4-57> 출·퇴근 연결성 분석 도구 데이터 가공 방법

나. 현행화 내용

- 2020년 데이터 현행화 및 유지보수



<그림 4-58> 출/퇴근 통행 연결성 분석 도구 화면예시

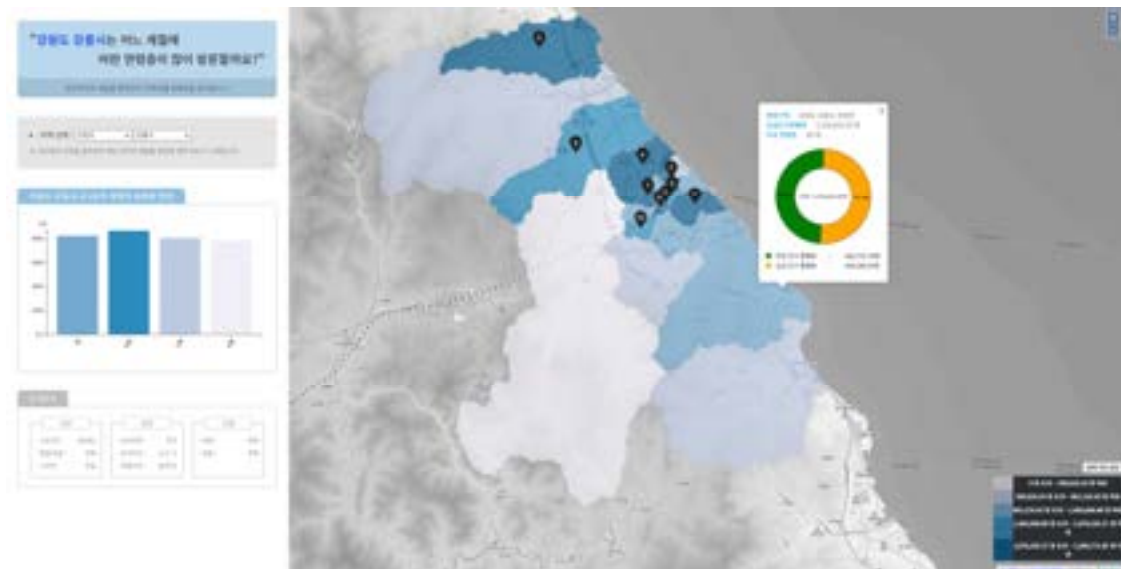
20. 관광 특화지역 분석 도구 현행화

가. 기능 소개

- 모바일 통신 데이터 기반 지역별·계절별·연령별 관광통행 분석하는 기능
- 계절에 따른 통행량 변화를 분석
 - 데이터 가공 방법 : 도착지 타입이 비정기적인 체류지인 통행일 지역별, 계절별로 분류
 - 관광 특화지역 분석 UI
 - 기본 주제도는 강원도 강릉시의 유입량 지표를 색상으로 표출
 - 유입량 지표가 높은 상위 10개 지역의 순위를 표시
 - 주제도 상에 마우스 오버 이벤트로 해당 지역의 유입 인구수를 표현
 - 우측에는 주제도, 좌측에는 우측 주제도의 조건과 지역명, 주요 방문 나이대, 계절별 통행량 차트를 표출
 - 우측 하단에는 범례 표출

나. 현행화 내용

- 2020년 데이터 현행화 및 유지보수



<그림 4-59> 관광 특화지역 분석 도구 화면예시

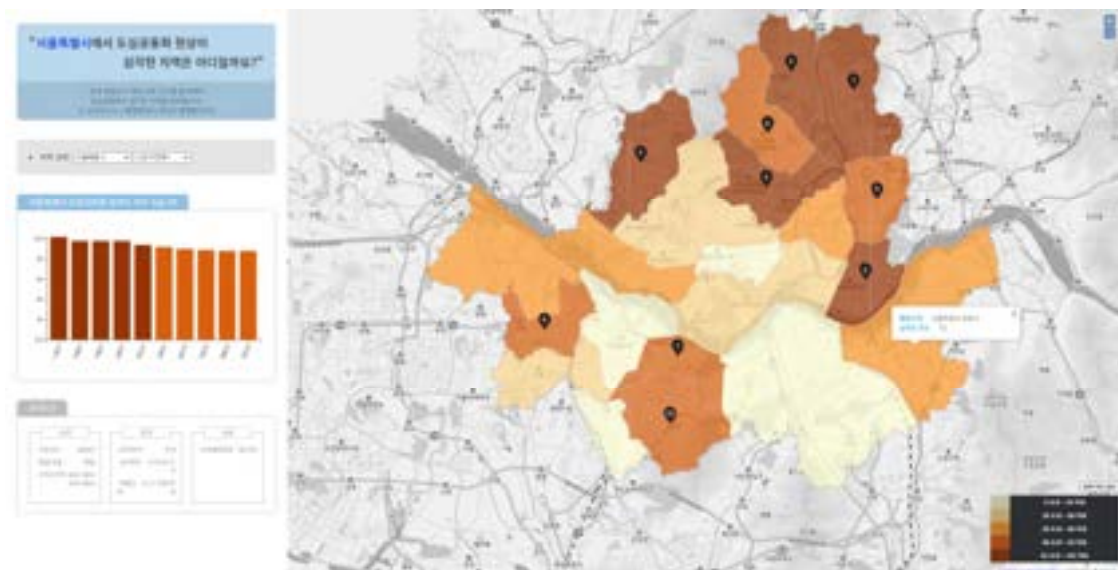
21. 도심 공동화 심각도 분석 도구 현행화

가. 기능 소개

- 모바일 통신 데이터 기반 지역별 유입량·유출량에 따른 도심 공동화 지표 산출 및 분석하는 기능
- 도심 공동화가 나타나는 지역을 분석
 - 데이터 가공 방법
 - 지역별 심야 시간대 도착지 트립타입이 집(H)인 인구 추출...④
 - 지역별 낮 시간대 도착지 트립타입 전체(또는 잠재체류지)인 인구 추출...⑥
 - 도심 공동화 지표 산출 : $(⑥/④) * 100$
 - 심야·낮 시간대에 대한 시간 기준은 한국교통연구원에서 제공
 - 도심 공동화 심각도 분석 UI
 - 기본 주제도는 서울시 도심 공동화 지표를 색상으로 표출
 - 주제도 상에 마우스 오버 이벤트로 해당 지역의 도심 공동화 심각도 지수를 표현
 - 우측에는 주제도, 좌측에는 도심 공동화 심각도에 대한 차트를 표출

나. 현행화 내용

- 2020년 데이터 현행화 및 유지보수



<그림 4-60> 도심 공동화 분석 도구 화면예시

22. 출·퇴근 개인 속성별(성별, 연령별 등) 통행 특성 (통행비율, 구간) 분석 도구 현행화

가. 기능 소개

- 모바일 통신 데이터 기반 지역별 통행 중 고령자가 차지하는 통행 비율 분석하는 기능
- 특정 개인 속성별 통행 발생량을 지역별로 추출하여 통행 비중 분석함
 - 데이터 가공 방법
 - ㉓ 모바일 데이터에서 특정 개인 속성에 대한 통행 발생량을 지역별로 추출
 - ㉔ 모바일 데이터에서 전체 통행 발생량을 지역별로 추출
 - 특정 개인 속성 통행 비중 계산 : ((㉓ / ㉔) * 100)
- 특정 개인 속성에 대한 주요 통행 구간에 대해 기종점별로 분석
 - 데이터 가공 방법
 - 특정 개인 속성에 대한 기종점별 통행을 추출

나. 현행화 내용

- 2020년 데이터 현행화 및 유지보수



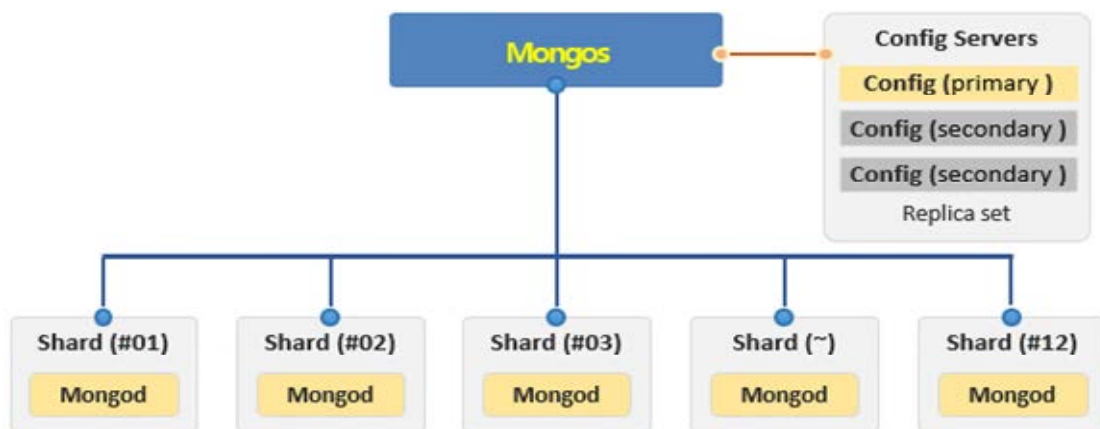
<그림 4-61> 출·퇴근 개인 속성(고령자) 통행 특성(통행비율) 분석 도구 화면예시

제3절 온라인 서비스 운영 및 유지보수

1. 빅데이터 서비스를 위한 최적화 시스템 구성

가. 대용량 통신 데이터 분산시스템 최적화

- 빅데이터 서비스를 위한 최적화의 필요성
 - 대용량의 통신 데이터를 신속하고 안전하게 서비스하기 위해서는, S/W와 H/W에 대한 최적화된 시스템 구성이 필요함
 - 대용량의 데이터를 물리적으로 한 대의 서버에 저장하고 처리하게 된다면 서버의 부하와 하드웨어적 한계가 발생하게 되므로 이를 해결하는 방안이 필요함
- 빅데이터 서비스를 위한 최적화 방안
 - 대용량의 데이터를 빠르게 처리하고 서버와 하드웨어에 부하가 가지 않도록 분산 저장하는 MongoDB의 샤딩 기술을 사용함
- MongoDB 샤딩 정의
 - 샤딩은 다음의 3가지 구성요소로 구성되며 각각의 역할이 부여됨
 - Mongos : 중계자 역할로 Application의 질의를 받아 Shard 서버의 응답을 중계
 - Config : 저장된 데이터가 Shard중에 위치 등의 Shard Meta 정보를 저장
 - Shard : 실제 데이터가 저장되는 서버

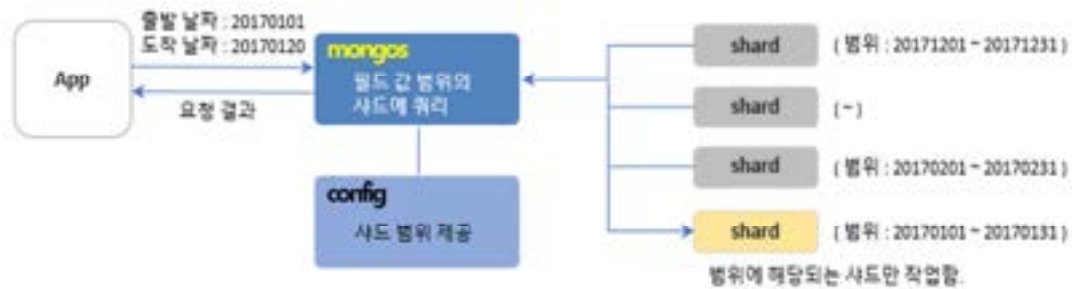


<그림 4-62> MongoDB 샤딩 구조

○ DB 특성별 샤딩 알고리즘 적용

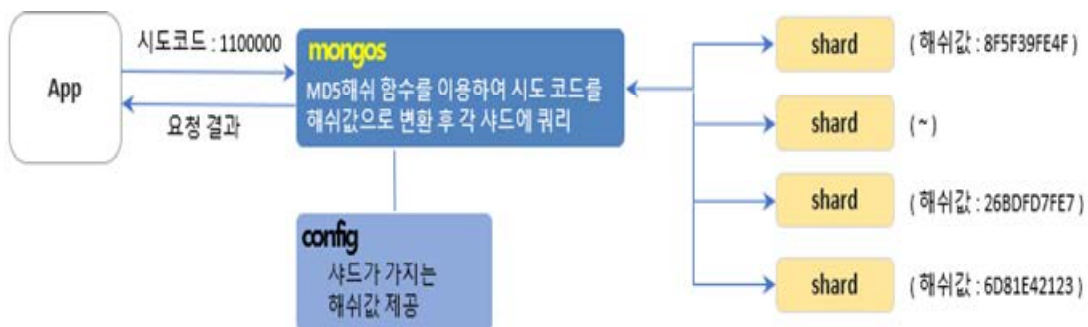
- 행정구역이 매칭된 통신 데이터

- View-T의 서비스 중 데이터 테이블 다운받기에서 사용되며 기간의 범위로 검색하여 서비스하는 데 사용되므로 레인지 샤딩과 지역 기반 샤딩을 사용함



<그림 4-63> 레인지 샤딩의 구성 화면

- 레인지 샤딩을 적용하여 범위 검색 시 해당 범위에 속한 Shard에서만 작업이 이루어지기 때문에 빠르고 서버에 부하를 주지 않음
- 지역 기반 샤딩을 적용하여 월별로 Shard에 분배하기 때문에 분산 시 고르게 분배됨
- 교통폴리곤 기준 통행량, 행정구역 단위의 출발/도착 통행량
 - View-T의 서비스 중 특정 행정구역을 선택 후 사용하는 주요통행지역분석, 주요통행구간분석, HotPlace분석에 사용되며 행정구역으로 검색하여 분석하므로 해쉬 샤딩을 사용함



<그림 4-64> 해쉬 샤딩의 구성

- 해쉬함수를 사용함으로 각 Shard에 데이터 분배가 고르게 되고 특정 행정구역 검색에 대한 처리가 빠르며 기준 필드들을 전부 샤드 키로 구성하여 기준 필드들로 검색하여도 신속한 처리를 할 수 있음

2. 운영 및 DB 유지보수

가. 신규 구축 DB 업데이트

- 구축된 통행지표 DB, 분석 도구 DB를 온라인 서비스 업데이트
 - 업데이트 대상이 되는 데이터는 네트워크 데이터, 웹 분석용 테이블, 형상 정보 및 경로 빅데이터 이관
 - 데이터의 적용 시스템 및 적용 내용은 다음과 같음

<표 4-18> View-T 웹서비스용 DB 데이터 업데이트 내용

대상 데이터	업데이트 대상 시스템	업데이트 내용
네트워크 데이터	RDBMS Server (MySQL)	신규 구축될 네트워크 데이터
웹 분석용 데이터	RDBMS Server (MySQL)	차량 GPS 기반 웹 분석용 데이터
형상 정보	GIS Server (GeoServer)	업데이트된 차량 모빌리티 기반지도 업데이트된 사람 모빌리티 기반지도
경로 데이터	BigData Solution (MongoDB)	차량 GPS 기반 개별차량 이동 궤적 데이터 모바일 데이터 기반 웹 분석용 데이터

- DB 장애 대응 전략
 - DB에 대한 정기적인 백업 정책을 수립하여 항시 데이터 유실을 미리 방지하고자 함

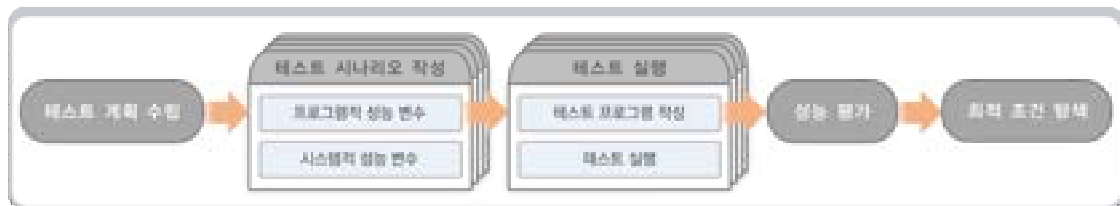


<그림 4-65> DB 장애 대응 전략

3. 성능 테스트

가. 데이터 가공 성능 테스트

- 대용량의 데이터를 빠르게 처리하고 서버와 하드웨어에 부하가 가지 않도록 분산 저장하는 MongoDB의 샤딩 기술을 사용함



<그림 4-66> View-T 데이터 가공 성능 테스트 시나리오

- View-T 서비스에서 이용자 특성을 고려한 고급 사용자에게는 상세 분석이 가능한 교통폴리곤 단위 집계 통행량 DB를 사용하고, 일반 사용자에게는 분석 특성이 있는 분석 특성 집계 통행량 DB를 사용하도록 설계한 내용으로 검색 평균 속도 비교 테스트를 진행
- 조건은 통행목적을 기준으로 출근, 등교, 퇴근, 하교, 점심시간, 귀가, 여가로 하여 크게 시도 단위에서는 서울시와 경기도, 시군구 단위에서는 강남구로 진행한 결과 전체 평균 속도는 교통폴리곤 단위 집계 통행량 DB에서 약 20분, 분석 특성(요일, 기간, 시간, 연령대) 집계 통행량 DB에서 약 5분, 분석 특성(통행목적) 집계 통행량 DB에서 약 1.5분 소요됨

나. 웹 서비스 성능 테스트

- 웹 서비스의 성능 테스트는 시스템 개발 종료 시점에 수행함
- 응답시간과 처리량, 병목 구간 등을 확인하여 서비스나 서비스 시스템의 문제점을 개선함

<표 4-19> View-T 웹 서비스 성능 테스트

테스트 항목	테스트 내용	충족 기준
Load 테스트	부하(Load)를 순차적으로 증가시키면서 테스트 진행 응답시간이 급격히 증가하거나 더는 처리량이 증가하지 않는 구간을 탐색 해당 구간의 CPU, 메모리 등의 점유 값이 기준값을 상회하는 임계 값을 찾아 튜닝 및 테스트 반복	CPU 점유율 90% 미만 메모리 점유율 80% 미만 결과 응답 3초 이내
Stress 테스트	임계 값 이상의 요청이나 비정상적인 요청을 보내 비정상적인 상황의 처리상태를 테스트함 시스템의 최고 성능 한계를 측정하기 위한 테스트	CPU 점유율 100% 인 상태로 3시간 이상 정상 작동
Spike 테스트	갑자기 사용자가 몰렸을 경우를 가정하여 요청이 정상적으로 처리되는지를 테스트함 부하가 줄어들 때 정상적으로 반응하는지를 테스트함	동시접속 100개에서 3000회 접속 시 정상작동 결과 응답 3초 이내
Stability 테스트	긴 시간 동안 테스트를 진행하여 테스트 시간에 따른 시스템의 메모리 증가, 성능 정보의 변화 등을 확인하기 위한 테스트를 진행	일주일간 일 평균 동시접속 50개 테스트 결과 응답 3초 이내

- 페이지에 관한 진단 정보를 수집하고 분석하여 각 카테고리의 점수가 표시됨
 - 90점 이상의 점수는 ‘ 좋음’, 50~89점은 ‘개선이 필요함’, 50점 미만은 ‘ 좋지 않은 것’ 으로 간주함
- PageSpeed Insights에서는 Chrome 사용자 환경 보고서의 데이터를 통합하여 FCP 및 DCL에서 캡처한 사이트의 실제 성능 정보를 제공함



<그림 4-67> 구글에서 제공하는 웹페이지 속도 테스트 화면

다. 웹 사이트 동작 테스트

- 서비스 개선 및 신규 추가 구현된 기능을 중심으로 웹 사이트의 동작 테스트를 자체적으로 진행하여 오류 및 추가 개선점을 해결함
- 사이트 동작 테스트는 기능 구현 담당자 간의 구현 내용을 서로 교환하여(크로스체크)함으로써 세부 내용을 인지하지 못하는 일반 사용자의 관점에서 조작에 불편함이나 의도치 않은 작동 Case를 수집하여 별도로 관리되는 일정에 맞춰 수정을 진행함

제5장 결론 및 차년도 수행계획

제1절 결론

제2절 차년도 수행계획

제5장 결론 및 차년도 수행계획

제1절 결론

- 본 과업에서는 기존 방법보다 정확하게 통행시간 및 통행량 등 통행지표를 산출하기 위하여 기존 모바일 가공 알고리즘을 보완함
 - 알고리즘은 합리적 수준의 사공간적 데이터 군집 (Data aggregation and smoothing) 과정을 제시하여 신호 이상 현상을 효과적으로 처리하고, 개별통행 특성 및 패턴을 고려한 주요 통행목적지 추정을 통해 야간 및 비정기 근무 통행을 구분함
 - 또한, 기존 통신폴리곤 기준의 위치 추정 정보를 GPS와 같은 점단위의 위치 추정 정보로 변환하는 방법을 제안하여 모바일 통신 데이터의 공간적 해상도를 개선하였으며, 이동 경로 및 수단 추정을 위한 개발 알고리즘 적용함
 - 격자형 폴리곤 데이터와 교통폴리곤 데이터 및 행정구역 데이터를 이용해 각각의 폴리곤 형상 레이어를 경량화하는 작업과 폴리곤 중심 위치 보정을 통해 포인트 레이어 구축 작업을 진행함
- 2022년 1월 1일부터 12월 31일까지 생성된 로그 기록에 모바일 통신 빅데이터 가공 알고리즘을 적용하여 기존 DB를 구축하고 검증함
 - 각 개인별 로그 기록은 「개인정보보호」, 「위치정보의 보호 및 이용 등에 관한 법률」에 저촉되지 않도록 집계한 형태의 DB를 설계하였으며, 이를 기준으로 DB를 구축함
 - 모바일 통신 빅데이터 기반으로 구축된 통행 DB의 시간대별, 요일별, 성별, 연령별 통행량 및 통행시간을 기존 국가교통DB사업의 ‘개인통행실태조사자료 (2016년)’와 통계청의 ‘인구총조사자료 (2015년)’ 비교하여 검증함
- 모바일 통신 빅데이터 기반 교통지표를 개발하고, 분석 서비스의 성능 및 사용성을 개선하기 위하여 DB를 재구조화하여 서비스를 제공함
 - 또한, 2개 이상의 분석 수행 결과를 주제도로 동시에 확인할 수 있는 ‘정보 연계 분석기능’ 개발하고, 통행 변화를 시계열로 확인할 수 있는 ‘통행 변화 시뮬레이션 분석기능’ 개발함

제2절 차년도 수행계획

- 모바일 교통빅데이터 가공 알고리즘 고도화
 - 데이터 전처리 알고리즘 개선 및 고도화
 - 지역특성(도시부/비도시부), 인적특성(연령, 휴대폰 이용률 등)에 따른 로그 특성 분석 및 기 전처리 알고리즘 보완
 - 통행목적 구분 알고리즘 개선 및 고도화
 - 모바일 교통빅데이터 기반 추정 가능 체류지(거주지, 근무지, 여가 등) 선정
 - 거주지 및 근무지 추정 로직 개선
 - 격자형 폴리곤 데이터와 교통폴리곤 데이터 및 행정구역 데이터를 이용해 각각의 폴리곤 형상 레이어를 경량화 및 알고리즘 고도화
- 모바일 교통빅데이터 DB 구축 및 검증
 - 모바일 교통빅데이터 DB 구축 (2023년도 전국)
 - 기 구축된 DB (2022년도 전국) 검증
 - 개인통행실태조사, 인구총조사 등 기존 조사 기반 통계를 이용한 검증
- 모바일 교통 빅데이터 기반 교통지표 개발 및 서비스
 - 교통지표 개발
 - 1차 지표 : 모바일 교통빅데이터 DB 기반 산출 가능한 통행지표 개발(통행량, 평균통행시간, 평균통행거리 등)
 - 2차 지표 : 이종 데이터와 모바일 교통빅데이터 DB를 연계·가공하여 산출 가능한 각종 지표 개발(교통안전지수, 도시공동화지수, 교통서비스지수, 경제활동지수 등)
- 교통빅데이터 플랫폼 분석기능 개발 지원
 - 교통빅데이터 플랫폼을 위한 개별통행 DB 재구조화 지원
 - 교통지표 제공 및 분석 서비스 개발 지원

2022년 국가교통조사

10

모바일통신 빅데이터 구축



국토교통부
Ministry of Land, Infrastructure and Transport



한국교통연구원
KOREA TRANSPORT INSTITUTE