

2020년 「국가교통조사·DB시스템 운영 및
유지보수」

대 중 교 통 정 책 지 원
고 도 화 를 위 한 모 바 일
빅 데이터 DB 구축

14

제 출 문

국토교통부장관 귀하

본 보고서를 「2020년도 국가교통조사 및 DB시스템 운영 및 유지 보수」 최종보고서로 제출합니다.

2020년 12월

한국교통연구원

원장 오 재 학

**본 『2020년도 국가교통조사 및 DB시스템 운영 및
유지보수』는 다음 연구진에 의해 수행되었습니다.**

참 여 연 구 진

<한국교통연구원>	
연구책임자	◦ 김주영 연구위원
연 구 진	<ul style="list-style-type: none"> ◦ 박인기, 최정민, 조종석, 천승훈 연구위원 ◦ 박용일, 황순연, 장동익, 성홍모, 원민수, 김병관, 우왕희 부연구위원 ◦ 신영권, 김동호, 김규진, 김정은 주임전문원, 이종우 전문연구원 ◦ 가보연, 강국수, 강명제, 곽명신, 김관용, 김성민, 김운태, 김은미, 김 현, 박미란, 박준호, 오연선, 이선아, 이슬기, 이채영, 이해선, 조용훈, 채정표, 홍성표 연 구 원 ◦ 강아라 연구조원
<한국해양수산개발원>	
연 구 진	<ul style="list-style-type: none"> ◦ 전형진, 이종필 부연구위원 ◦ 류희영 연구원
<한국항공협회>	
연 구 진	<ul style="list-style-type: none"> ◦ 성인영 실장 ◦ 강영주 과장, 김창욱 대리

『2020년도 국가교통조사 및 DB구축사업』

보고서 구성 및 담당연구진

번 호	과 제 명	연 구 진
제 1권	요약 보고서	김주영, 최정민, 신영권, 박준호
제 2권	전국 여객 O/D 보완 갱신	조종석, 강국수, 박미란
제 3권	빅데이터 분석 지원 사업	장동익, 김동호, 홍성표, 우왕희
제 4권	여객 O/D 신뢰도 제고 예비조사	조종석, 김동호, 채정표, 김병관
제 5권	항공여객 O/D 및 특성조사	한국항공협회
제 6권	전국 화물 O/D 보완 갱신	박인기, 김정은, 조용훈, 가보연, 김운태
제 7권	해상 화물 O/D 보완 갱신	한국해양수산개발원
제 8권	빅데이터 기반 화물 O/D 신뢰도 제고 연구	박인기, 성홍모, 김정은, 강명제
제 9권	교통분석용 네트워크 구축	최정민, 이선아, 이슬기
제10권	KTDB 플랫폼 기반지도 구축	김동호, 김관용
제11권	국가 교통통계 DB 구축	박용일, 곽명신
제12권	특별교통대책 기간 통행실태조사	우왕희, 김은미
제13권	교통혼잡 지도 DB 구축	천승훈, 김성민, 이채영
제14권	대중교통 정책지원 고도화를 위한 모바일 빅데이터 DB구축	원민수, 이해선, 이종우, 백현진
제15권	교통유발원 단위 조사	황순연, 오연선, 김현
제16권	국가교통물류 경쟁력지표 조사연구	장동익, 홍성표
별 권	DB시스템 운영 및 유지보수	신영권, 김규진, 박준호

『2020년도 국가교통조사 및 DB구축사업』

과제별 공동참여·위탁용역 사업자

【공동사업 참여기관】

- 전국 여객O/D 현행화 공동사업 (제주특별자치도 부문)
 - 홍익대학교산학협력단
- 전국 여객O/D 현행화 공동사업 (부산·울산권 부문)
 - ㈜두운엔지니어링, 경성대학교산학협력단
- 전국 여객O/D 현행화 공동사업 (대전·세종·충청권 부문)
 - ㈜신명이앤씨
- 항공O/D 및 특성 조사
 - (사)한국항공협회

【위탁용역 사업자】

- 전국 일반버스 도로기반 교통 네트워크 구축
 - ㈜아로정보기술
- 빅데이터 기반 O/D검증을 위한 스마트폰 기반 GPS조사
 - ㈜컨슈머인사이트
- 가구통행실태조사 예비조사
 - ㈜코리아데이터네트워크
- 영업용 화물차 운행기록계 빅데이터를 이용한 화물 기종점통행량 및 운행특성 분석연구
 - ㈜노트스퀘어
- 도로 및 철도 교통분석용 네트워크 보완갱신
 - ㈜올포랜드, ㈜엔토포스

【위탁용역 사업자】

- KTDB 교통빅데이터 플랫폼 (View-T) 기반맵 구축
 - ㈜큐빅웨어
- 국가교통DB Breif발간대행
 - ㈜우공이산
- 특별교통통행실태조사 및 이용자 만족도 조사
 - ㈜컨슈머인사이트
- View-T 서비스 제공을 위한 차량 모빌리티 데이터 구축 및 기능 개선
 - 큐빅웨어/ ITS학회
- 교통유발원단위 조사
 - ㈜아이로드테크, ㈜지알아이리서치, 주식회사 시그널웍스
- 모바일 데이터 기반 교통 분석용 DB 구축 및 View-T 2.0 분석 서비스 개발
 - 충북대학교산학협력단, ㈜큐빅웨어, ㈜넷케이티아이

최종보고서 목차

제 1권 요약 보고서

제 2권 전국 여객 O/D 보완 갱신

제 3권 빅데이터 분석 지원 사업

제 4권 여객 O/D 신뢰도 제고 예비조사

제 5권 항공여객 O/D 및 특성조사

제 6권 전국 화물 O/D 보완 갱신

제 7권 해상 화물 O/D 보완 갱신

제 8권 빅데이터 기반 화물 O/D 신뢰도 제고 연구

제 9권 교통분석용 네트워크 구축

제 10권 KTDB 플랫폼 기반지도 구축

제 11권 국가 교통통계 DB 구축

제 12권 특별교통대책 기간 통행실태조사

제 13권 교통혼잡 지도 DB 구축

제 14권 대중교통 정책지원 고도화를 위한 모바일 빅데이터 DB 구축

제 15권 교통유발원 단위 조사

제 16권 국가교통물류 경쟁력지표 조사연구

별 권 DB 시스템 운영 및 유지보수

목 차

요 약

제1장 과업의 개요 1

제1절 배경 및 목적 / 3

제2절 범위 및 내용 / 4

제2장 모바일통신 빅데이터 가공 알고리즘 개발 7

제1절 데이터 전처리 알고리즘 고도화 / 9

제2절 체류정보 및 통행목적 추정 알고리즘 고도화 / 15

제3절 이동수단 추정 알고리즘 개발 / 22

제3장 모바일통신 빅데이터 기반 통행 DB 구축 및 검증 45

제1절 기준년도 통행DB 구축 / 47

제2절 구축 DB 검증 / 60

제4장 모바일통신 빅데이터 기반 통행 분석 서비스 개발 79

제1절 개요 / 81

제2절 분석 서비스 고도화 / 87

제3절 기반 DB 현행화 및 속도 개선 / 113

제4절 데이터 다운로드 서비스 개선 / 133

제5장 결론 및 차년도 수행계획 135

제1절 결론 / 137

제2절 차년도 수행계획 / 138

표 목 차

〈표 1〉 모바일 기반 DB 형태	12
〈표 2〉 모바일 DB 테이블 정의서	14
〈표 3〉체류지 식별 기준	16
〈표 4〉지역별 통행량 비교 결과	19
〈표 5〉 지역별 출퇴근 통행량 비교 결과	21
〈표 6〉 지역별 출퇴근 통행시간 비교 결과	22
〈표 7〉 2019년 기준 모바일 통신 빅데이터 기반 통행 분석 서비스	23
〈표 8〉 View-T Light 분석서비스 - 기존 분석 기능의 View-T Light 버전 개발 (5가지) ...	26
〈표 9〉 View-T Light 분석서비스 -신규 통행 분석 아이템의 View-T Ligh (8가지)	27
〈표 10〉 교통폴리곤 기준 집계 통행량 DB 구축 효과 (도큐먼트 수 변화)	30
〈표 11〉 행정구역 기준 집계 통행량 DB 구축 효과	31
〈표 12〉 분석특성별(통행목적별) 집계 통행량 DB 구축 효과	31
〈표 13〉 2020년 기준 다운로드 정책	32
〈표 2-1〉 지역간 통행 관련 문헌고찰	24
〈표 2-2〉 Unlinked Trip 관련 문헌고찰	24
〈표 2-3〉 수단추정 알고리즘 관련 문헌고찰(규칙 기반 학습법)	25
〈표 2-4〉 수단추정 알고리즘 관련 문헌고찰(기계 학습법)	27
〈표 2-5〉 2016년 5월 일주일 간 세종시 수단별 Aggregated O/D	43
〈표 3-1〉 모바일 기반 DB 형태	48
〈표 3-2〉 모바일 DB 테이블 정의서	50
〈표 3-3〉 체류지 식별 기준	54
〈표 3-4〉 KT 시장점유율	58
〈표 3-5〉 지역별 통행량 비교 결과	61
〈표 3-6〉 시간대별 통행량 패턴 분석 (모바일기지국데이터 기준)	64
〈표 3-7〉 요일별 통행량 패턴 분석 (모바일기지국데이터 기준)	67
〈표 3-8〉 연령별 통행량 패턴 분석 (모바일기지국데이터 기준)	69
〈표 3-9〉 지역별 출퇴근 통행량 비교 결과	72
〈표 3-10〉 지역별 출퇴근 통행시간 비교 결과	74

〈표 4-1〉 2019년 기준 모바일 통신 빅데이터 기반 통행 분석 서비스	81
〈표 4-2〉 Light version과 Expert version 구분	83
〈표 4-3〉 분석 기능 사용방법 안내 방식 변경	84
〈표 4-4〉 기반 DB 업데이트	85
〈표 4-5〉 2020년 기준 다운로드 정책	86
〈표 4-6〉 View-T Portal 사이트 구성	87
〈표 4-7〉 주제별 키워드 사전 조건	105
〈표 4-8〉 주요통행구간 분석 기능의 연도별 표출형태 변경	107
〈표 4-9〉 2017년, 2018년 통신 데이터 현황	113
〈표 4-10〉 모바일 통신 빅데이터 기반 통행정보 DB 컬럼명 및 데이터 타입	113
〈표 4-11〉 교통폴리곤 기준 통행량 DB 테이블 정의서	117
〈표 4-12〉 교통폴리곤 기준 집계 통행량 DB 구축 효과 (도큐먼트 수 변화)	117
〈표 4-13〉 행정구역 단위의 출발/도착 기준 통행량 DB 테이블 정의서	118
〈표 4-14〉 행정구역 기준 집계 통행량 DB 구축 효과	119
〈표 4-15〉 출발과 도착 모두 행정구역 단위인 통행량 DB 테이블 정의서	119
〈표 4-16〉 출발과 도착 모두 행정구역 단위인 집계 통행량 DB 구축 효과	120
〈표 4-17〉 통행목적별 집계 통행량 데이터 DB 테이블 정의서	120
〈표 4-18〉 분석특성별(통행목적별) 집계 통행량 DB 구축 효과	121
〈표 4-19〉 교통폴리곤 기준 통행량 DB 검증 항목 예시	121
〈표 4-20〉 행정구역 단위 통행량 합계 비교 검증 항목	123
〈표 4-21〉 통행목적별 집계 통행량 합계 비교 검증 항목	126
〈표 4-22〉 교통폴리곤 레이어 필드	127
〈표 4-23〉 VectorTile, WMS의 지도 일반화	130
〈표 4-24〉 공공용 다운로드 서비스에서 제공되는 데이터 형태	133

그림목차

〈그림 1〉 신규 모바일통신 빅데이터 가공 알고리즘	6
〈그림 2〉 통행자 실제 위치 추정 방법 예시	7
〈그림 3〉 최장체류지 식별 후, 체류시간 aggregation 예시	8
〈그림 4〉 통행지표 산출물 예시	9
〈그림 5〉 수단추정 알고리즘 종합	10
〈그림 6〉 신규 모바일통신 빅데이터 가공 DB 구축 과정	11
〈그림 7〉 View-T Portal 메인 화면	25
〈그림 8〉 View-T Light 화면 구성	26
〈그림 9〉 분석기능 튜토리얼 UI	28
〈그림 10〉 View-T Expert 신규분석 기능 개발 - 정보 연계 분석 기능	29
〈그림 11〉 View-T Expert 신규분석 기능 개발 - 유입 기준 폴리곤 형태의 통행변화 시뮬레이션	29
〈그림 12〉 View-T Expert 신규분석 기능 개발 - Light 분석서비스 중 일부 구현	30
〈그림 13〉 모바일 분석 맵 DB 구축 (예시: 교통폴리곤)	31
〈그림 1-1〉 과업의 수행절차	5
〈그림 2-1〉 신규 모바일통신 빅데이터 가공 알고리즘	11
〈그림 2-2〉 시간적 데이터 보정 방법 예시	12
〈그림 2-3〉 통행자 실제 위치 추정 방법 예시	13
〈그림 2-4〉 공간적 데이터 보정 방법 예시	14
〈그림 2-5〉 시·공간적 데이터 보정(전처리) 적용 전, 후 통행 궤적 비교	14
〈그림 2-6〉 최장 체류지 식별 예시	15
〈그림 2-7〉 최장 체류지 이외의 체류지에 체류한 시간 중 가장 긴 시간 식별 예시	15
〈그림 2-8〉 최장체류지 이외 체류지 중 최장체류지 식별	16
〈그림 2-9〉 최장체류지 식별 후, 체류시간 aggregation 예시	17
〈그림 2-10〉 통행지표 산출물 예시	18
〈그림 2-11〉 통행시간 계산 및 잠재체류 식별 예시	19
〈그림 2-12〉 주체류지 기준 trip type 구분	20
〈그림 2-13〉 일별 데이터에 대한 주체류지 간 통행량 산정 예시	20
〈그림 2-14〉 data lagging 예시	21
〈그림 2-15〉 연구의 기본 방향	23

〈그림 2-16〉 알고리즘 적용 DB 구축 예시	30
〈그림 2-17〉 모바일 데이터와 기반맵 연계 Flow Chart	31
〈그림 2-18〉 교통수단 구분 기본 알고리즘	32
〈그림 2-19〉 항공수단 구분 지표	33
〈그림 2-20〉 항공수단 추정 알고리즘	34
〈그림 2-21〉 철도(고속)수단 구분 지표	35
〈그림 2-22〉 철도(고속)수단 추정 알고리즘	36
〈그림 2-23〉 버스(고속)수단 구분 지표	37
〈그림 2-24〉 버스(고속)수단 추정 알고리즘	38
〈그림 2-25〉 유클리드 거리에 의한 방향별 터미널 Sequence 설정	39
〈그림 2-26〉 버스(시외)수단 구분 지표	40
〈그림 2-27〉 버스(시외)수단 추정 알고리즘	41
〈그림 2-28〉 수단추정 알고리즘 종합	42
〈그림 2-29〉 2016년 5월 일주일 간 세종시 일별 수단별 통행량 분포	43
〈그림 3-1〉 신규 모바일통신 빅데이터 가공 DB 구축 과정	47
〈그림 3-2〉 선분 이력으로의 변환	51
〈그림 3-3〉 체류순서 부여 (개인별 이동궤적 형성)	52
〈그림 3-4〉 이동정보와 체류정보 구분 (예시)	53
〈그림 3-5〉 체류지 유형 구분 (예시)	55
〈그림 3-6〉 위치정보, 시간정보, 연령정보 변환 (예시)	56
〈그림 3-7〉 출발/도착 구분 및 통행량 집계 (예시)	57
〈그림 3-8〉 기지국 기반 데이터의 통행 시간 산출 방법	59
〈그림 3-9〉 지역별 통행 발생량 비교 결과	62
〈그림 3-10〉 지역별 통행 도착량 비교 결과	62
〈그림 3-11〉 시간대별 통행 발생량 패턴 분석 (모바일기지국데이터 기준)	65
〈그림 3-12〉 시간대별 통행 도착량 패턴 분석 (모바일기지국데이터 기준)	65
〈그림 3-13〉 요일별 시간대에 따른 통행량 변화 (발생 기준)	67
〈그림 3-14〉 요일별 시간대에 따른 통행량 변화 (도착 기준)	68
〈그림 3-15〉 연령별 통행 발생량 패턴 분석 (모바일기지국데이터 기준)	70
〈그림 3-16〉 연령별 통행 도착량 패턴 분석 (모바일기지국데이터 기준)	70
〈그림 3-17〉 지역별 출퇴근 통행량 분포 비교 결과	73
〈그림 3-18〉 시도별 통근 통행시간 비교(단위 : 분)	75
〈그림 3-19〉 통행시간 기준별 통근 인구 비율 비교	76
〈그림 3-20〉 출근시간 밀도그래프	77

〈그림 3-21〉 퇴근시간 밀도그래프	77
〈그림 3-22〉 출근속도 밀도그래프	78
〈그림 3-23〉 퇴근속도 밀도그래프	78
〈그림 4-1〉 View-T Portal 메인 화면	87
〈그림 4-2〉 View-T Light와 View-T Expert 직접 진입 방법	88
〈그림 4-3〉 View-T Light와 View-T Expert 연계 진입 방법	89
〈그림 4-4〉 View-T Light 화면 구성	90
〈그림 4-5〉 View-T Light Hotplace 분석 화면	91
〈그림 4-6〉 View-T Light 주요통행지역 분석 출/도착지 변경 화면	92
〈그림 4-7〉 View-T Light 주요통행지역 분석 화면	92
〈그림 4-8〉 View-T Light 주요통행구간 분석 화면	93
〈그림 4-9〉 View-T Light 내부통행지역 분석 화면	94
〈그림 4-10〉 View-T Light 통행시간/거리 분석 화면	95
〈그림 4-11〉 관광 특화 지역 분석 화면	96
〈그림 4-12〉 도심공동화 심각도 분석 화면	97
〈그림 4-13〉 SDE로 그려진 타원체 예시	98
〈그림 4-14〉 생활권분석 화면	99
〈그림 4-15〉 어린이 취약구간 분석 화면	100
〈그림 4-16〉 고령자 통행비중 분석 화면	101
〈그림 4-17〉 고령자 주요통행구간 분석 화면	102
〈그림 4-18〉 출퇴근시간대 접근성 분석 화면	103
〈그림 4-19〉 코로나19 위험지역 분석 화면	104
〈그림 4-20〉 분석 조건값 UI 변경	105
〈그림 4-21〉 범례설정 UI 변경	106
〈그림 4-22〉 분석기능 튜토리얼 UI	107
〈그림 4-23〉 주요통행구간 분석의 전국분석 결과 및 통행량 분포도 그래프	108
〈그림 4-24〉 정보 연계 분석 기능 화면	109
〈그림 4-25〉 유입 기준 폴리곤 형태의 통행변화 시뮬레이션	110
〈그림 4-26〉 지역 내 기준 폴리곤 및 희망선 형태의 통행변화 시뮬레이션	110
〈그림 4-27〉 관광 특화지역 분석 기능 화면	111
〈그림 4-28〉 도심공동화 심각도 분석화면	112
〈그림 4-29〉 통신데이터의 형식변환 및 거리산출	115
〈그림 4-30〉 교통폴리곤 기준 통행량 데이터 구조	115
〈그림 4-31〉 행정구역 단위의 출발/도착 기준 통행량 DB 구조	116

〈그림 4-32〉 통행목적별 집계 통행량 데이터 조건	116
〈그림 4-33〉 교통폴리곤 기준 통행량 DB 검증 쿼리와 결과 예시(시도별 통행량 확인) ..	122
〈그림 4-34〉 통신데이터 전처리DB와 교통폴리곤 단위 집계 통행량 DB 비교검증 결과 ..	122
〈그림 4-35〉 교통폴리곤 단위 집계 통행량 DB와 읍면동 단위 집계 통행량 DB 비교검증 결과	123
〈그림 4-36〉 읍면동 단위 집계 통행량 DB와 시군구 단위 집계 통행량 DB 비교검증 결과	124
〈그림 4-37〉 시도 단위 출발기준 집계 통행량 DB와 출발/도착이 모두 시도 단위인 집계 통행량 DB 비교검증 결과	124
〈그림 4-38〉 행정구역 단위 집계 통행량 DB 검증 쿼리와 결과 예시	125
〈그림 4-39〉 교통폴리곤 단위 집계 통행량DB와 통행목적별 집계 통행량 DB 비교검증 결과(2017년 기준)	126
〈그림 4-40〉 교통폴리곤의 형상정보	127
〈그림 4-41〉 교통폴리곤 형상 데이터	128
〈그림 4-42〉 읍면동 형상 데이터	129
〈그림 4-43〉 노드 위치 보정 작업 예시	130
〈그림 4-44〉 기반 DB 구축 과정	131

요약

요 약

1. 과업의 개요

가. 과업의 배경 및 목적

- 대중교통 이용 활성화 및 정책을 지원하기 위해서는, 대중교통 이용자뿐만 아니라 잠재적으로 대중교통을 이용할 가능성이 높은 통행자에 대한 분석이 필수적임
 - 대중교통 이용자에 대한 통행은 교통카드 DB(대중교통 승하차 시 태그 정보)로 대부분 추정할 수 있으나, 잠재적으로 대중교통을 이용할 가능성이 있는 타 교통수단 이용자(차량, 도보 등)에 대한 통행은 일부만 추정 가능
 - 차량에 탑재된 기기(내비게이션, DTG 등)나 도로에 설치된 기기(검지기, DSRC 등)를 통해 차량 통행자에 대한 통행 정보는 취득할 수 있으나, 자전거·도보 등으로 이동한 통행자에 대한 통행 정보는 표본율이 낮은 가구통행실태조사 결과에만 의존하고 있는 실정
- 스마트폰 보급률 99%에 육박하는 모바일통신 빅데이터를 이용하면, 국민의 거의 모든 통행 정보를 분석할 수 있음
 - 모바일 기기와 기지국 간의 송·수신 이력인 모바일통신 빅데이터는 대부분 국민이 이용하고 있는 모바일 기기를 통해 축적되기 때문에 표본율이 매우 높고, 이동 수단에 구애받지 않고 개별통행에 대한 정보를 상세하게 포함하고 있음
- 통행자별 이용 교통수단 정보가 포함된 모바일통신 빅데이터를 이용하여 통행DB를 구축하고, 이를 기반으로 교통수요를 분석할 수 있는 지표, 기능 등을 개발하여 고도화된 대중교통 정책 수립 지원에 활용하고자 함

나. 과업의 범위

- 시간적 범위
 - 모바일 빅데이터 가공 알고리즘 개발: 2019년 4월, 2020년 4월의 원천데이터
 - 2019년도 기준 기반 DB 구축: 2019년 1월 1일 ~ 2019년 12월 31일 (1년)
 - 웹 서비스를 위한 각종 분석도구 개발: 2018년 1월 1일 ~ 2018년 12월 31일
 - 2019년도 연구 성과물인 2018년도 기준 기반 DB 활용
- 공간적 범위: 전국

다. 과업의 내용

- 모바일통신 빅데이터 가공 알고리즘 개발
 - 데이터 전처리 알고리즘 고도화
 - 체류정보 및 통행목적 추정 알고리즘 고도화
 - 이동수단 추정 알고리즘 검토
- 모바일통신 빅데이터 기반 통행 DB
 - 기준년도 통행DB 구축 및 검증
 - 주요 교통지표 생성 및 분석
- 모바일통신 빅데이터 기반 통행 분석 서비스 개발
 - 분석 서비스 고도화
 - 기반DB 현행화 및 속도 개선
 - 데이터 다운로드 서비스 개선

2. 모바일통신 빅데이터 가공 알고리즘 개발

가. 데이터 전처리 알고리즘 고도화

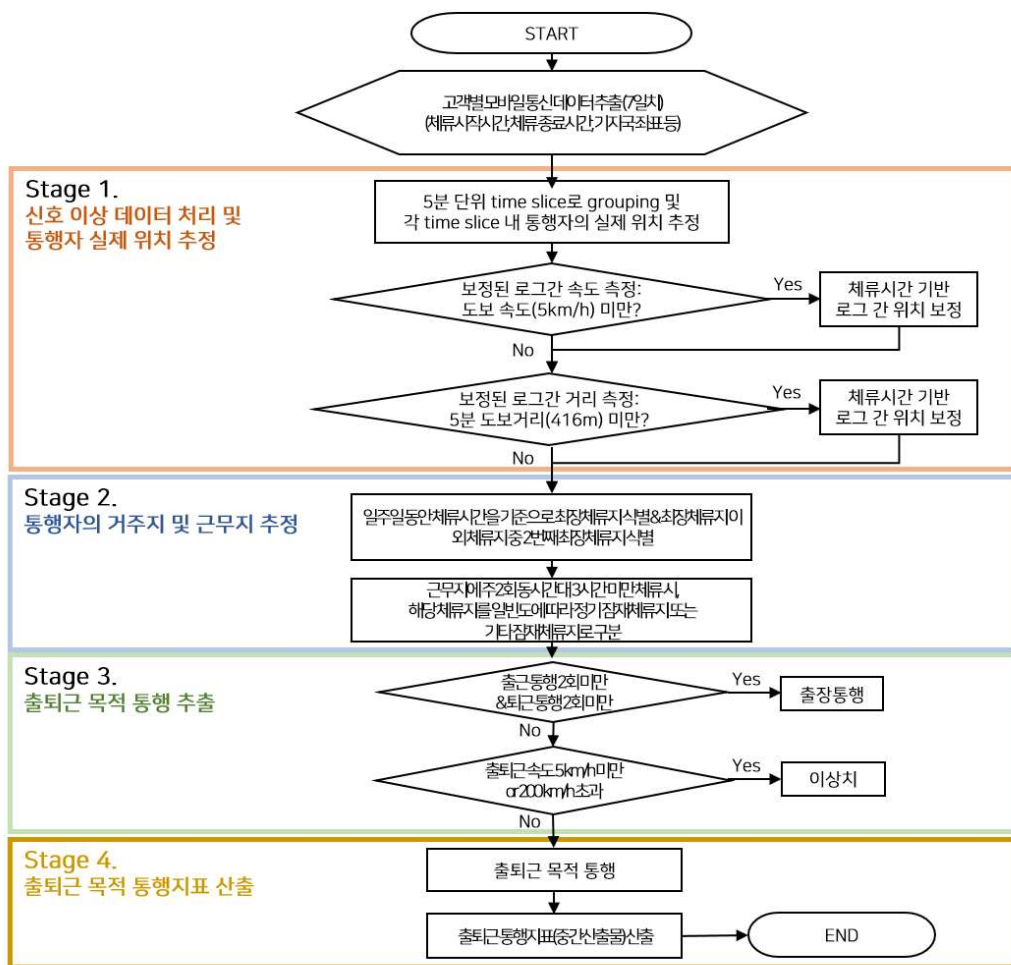
1) 기존 가공 알고리즘의 보완 및 개선 필요성

- 기존 알고리즘의 경우, 속도기반 방법을 사용하여 시그널점프 현상에 의해 발생한 이상치를 보정하고, 패턴 기반 방법을 사용하여 핑퐁핸드오버 현상에 의해 발생한 이상치를 보정하는 등 신호 이상 현상별로 다른 알고리즘을 적용
 - 시그널점프 데이터로 식별되어 제거된 데이터 중에서 핑퐁핸드오버 데이터가 포함되어 있을 수 있음
 - 핑퐁핸드오버 패턴이 다양하고 실제로는 핑퐁핸드오버 현상이 무수히 많이 발생할 수 있기 때문에 3회 반복되는 $A \rightarrow B \rightarrow A \rightarrow B$ 기준만으로는 모든 핑퐁 핸드오버 데이터를 보정하는 것이 어려움
- 기존 알고리즘의 경우, 특정 시간(25분)을 기준으로 데이터 로그별 체류/이동 여부를 판단하고 주요 통행 시간대나 연령 등을 기준으로 체류지를 5-6개의 유형으로 구분
 - 대표 통행자(통근·통학 위주)에 대한 통행 시간대를 미리 가정한 후, 통계청에서 공표하고 있는 인구수(통근·통학 인구)와 가장 유사한 값을 갖는 통행 패턴 조건을 찾는 방식으로, 대표 통행자 외의 통행자(자영업자, 비정기적 통행자 등)는 배제됨
 - 사전에 정의한 통행시간대의 변화에 따라 통행 패턴 조건이 달라질 수 있음

2) 신규 가공 알고리즘의 주요 고려사항 및 개념

- 신규 가공 알고리즘은 기지국과 단말기 간에 발생하는 불규칙적인 신호 이상 현상을 효과적으로 보정 할 수 있어야 함
- 또한, 개별 통행 특성 및 지역 간 기지국 거리 등을 고려하여 이동/체류를 구분하고 체류지 유형을 추정하여야 함
- 입력데이터(input data)로 각 통행자에 대한 7일 치 원천데이터가 사용되며 각 통행자에 대한 개별 통행 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 이동 궤적을 보정

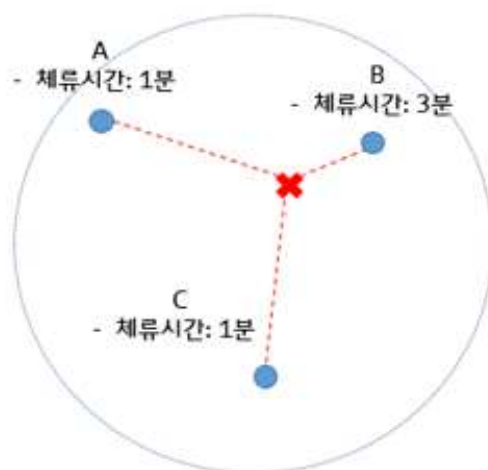
- 기지국 데이터 기록빈도 및 체류시간을 기반으로 통행자의 개별 통행 특성을 고려한 주체류지 식별을 수행하기 때문에 정기 출퇴근 통행뿐만 아니라 야간 근무자를 포함한 비정기 출퇴근 통행도 식별 가능
- 개별 통행자에 대한 주체류지를 식별한 후, 해당 위치에서의 체류시작시간과 체류종료시간을 기반으로 통근시간 및 통행량 등 통행지표 산출



<그림 1> 신규 모바일통신 빅데이터 가공 알고리즘

나. 신규 데이터 전처리 알고리즘 개발

- 적절한 공간적 해상도를 바탕으로 통행자의 통행속도를 고려한 이동 궤적을 식별하기 위해서 5분이라는 기준 시간을 정하고 해당 시간대에 기록된 기지국의 기록빈도를 기반으로 통행자의 실제 위치를 추정하고자 함
 - 짧은 시간 동안 발생한 핑퐁 핸드오버 현상에 의한 체류지 식별률 저하 및 비현실적으로 기록된 이동 궤적을 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

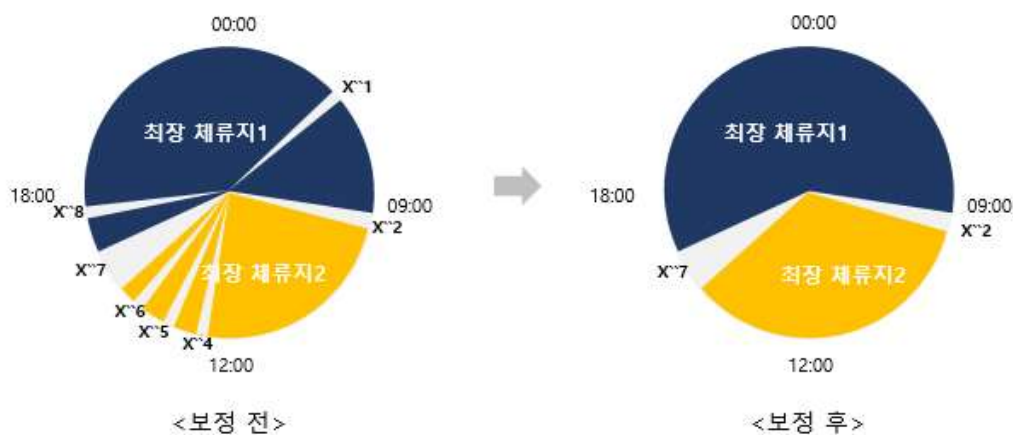


<그림 2> 통행자 실제 위치 추정 방법 예시

- 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
 - 원천데이터에 대해 시·공간적으로 신호 이상 기록을 보정해줌으로써, 분 단위로 연속해서 변화하거나 좁은 공간적 범위 내에 산발적으로 기록되는 기지국 위치 보정 가능

다. 체류정보 및 통행목적 추정 알고리즘 개발

- 데이터 전처리 단계에서 추정한 통행자의 실제 위치를 기준으로 통행자가 일주일 동안 각 위치에 몇 분 체류하였는지 계산 후, 최장 체류지 식별
- 통행자가 최장 체류지 이외의 체류지들에서 연속적으로 체류한 시간을 일별로 계산하여 그중에서 가장 긴 시간을 식별
- 해당 시간 이외에는 통행자가 최장 체류지에 체류한 것으로 지정하며, 해당 시간에 속하는 체류지들에 대하여 일주일 기준 체류시간이 가장 긴 체류지 식별(2번째 최장 체류지 식별 시, 최장 체류지로 지정된 체류지들을 후보에서 제외하고 식별)
- 일별로 2번째 최장 체류지가 처음 기록된 시각(t_1)과 마지막으로 기록된 시각(t_2)를 식별한 후, t_1 부터 t_2 까지 통행자가 2번째 최장 체류지에 체류한 것으로 보정(단, t_1 부터 t_2 사이에 1번째 최장체류지로 지정된 체류지가 존재하는 경우, 2번째 최장 체류지가 없는 것으로 판단)



<그림 3> 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
- 원천데이터에 대해 신규 알고리즘을 적용하면 <그림 4>와 같은 형태의 통행지표가 산출되며, 해당 산출물에는 개별 고객에 대한 일자별 통행지표가 포함되어 있음

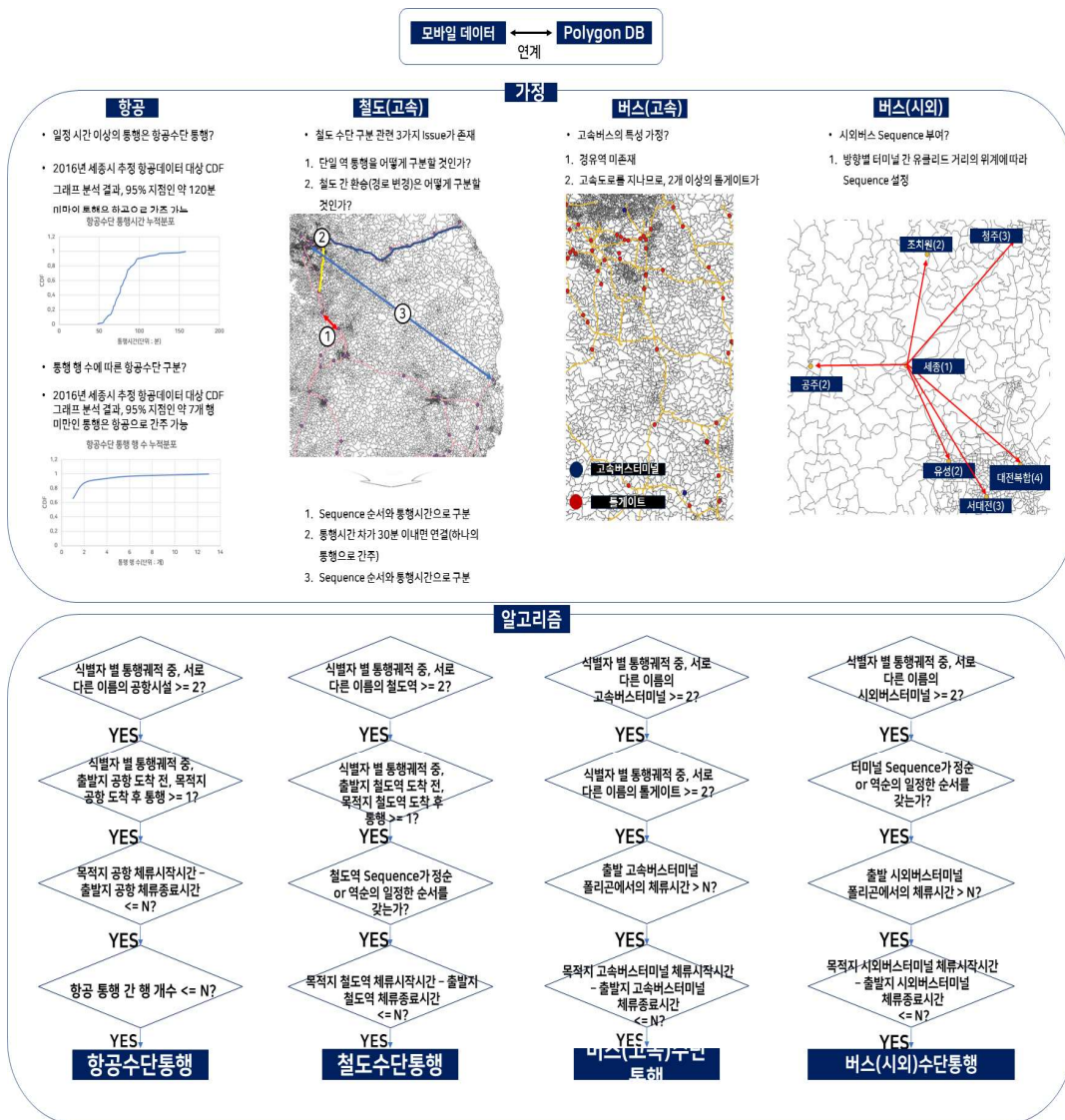
날짜	2019-04-01
고객식별번호	112233
나이	30분
성별	M
출근시작시간	8:30
출근종료시간	9:00
출근소요시간	30분
출근 중 잠재체류 횟수	0회
출근 중 잠재체류 시간	0분
퇴근시작시간	18:27
퇴근종료시간	19:30
퇴근소요시간	63분
퇴근 중 잠재체류 횟수	1
퇴근 중 잠재체류 시간	35
집 위치	127.491, 36.455
회사 위치	127.485, 36.607
집-회사 간 거리	16.9 km
출근시간	30분
퇴근시간	X
출근속도	33.8 km/h
퇴근속도	X

<그림 4> 통행지표 산출물 예시

마. 이동수단 추정 알고리즘 개발

- 본 연구에서 개발한 알고리즘은 새로운 통행분석 지표인 Unlinked Trip 기반 지역간 이동수단을 추정하는 목적
- 기존 모바일 데이터 기반맵 자료에 각 수단별 교통시설 공간정보를 매칭시켜 교통수단 분석용 모바일 데이터 기반맵을 구축
- 교통폴리곤, 공항, 철도역, 철도망, 도로망, 고속버스터미널, 시외버스터미널 데이터의 속성 정보를 결합, 해당 폴리곤 내 교통시설 존재의 유무에 따라 각 교통시설 컬럼값을 0 또는 1 이상의 정수형 값으로 표시

- 구축된 교통수단 분석용 기반맵과 모바일 데이터의 속성을 결합
- 매칭된 최종 데이터를 기반으로, 통행수단별 분류 알고리즘을 제시
- 한 사람을 기준으로 하루의 통행패턴 중, 서로 다른 이름을 가진 교통유발시설이 2개 이상 존재하고, 통행시간(도착행 도착시간 - 출발행 출발시간)과 철도망, 도로망 정보가 상당 부분 일치할 경우 이를 Unlinked Trip 기반 지역간 통행으로 간주
- 분류 순서는 항공 -> 철도 -> 고속버스 -> 시외버스 -> 그 외 수단 등 5단계로 설정



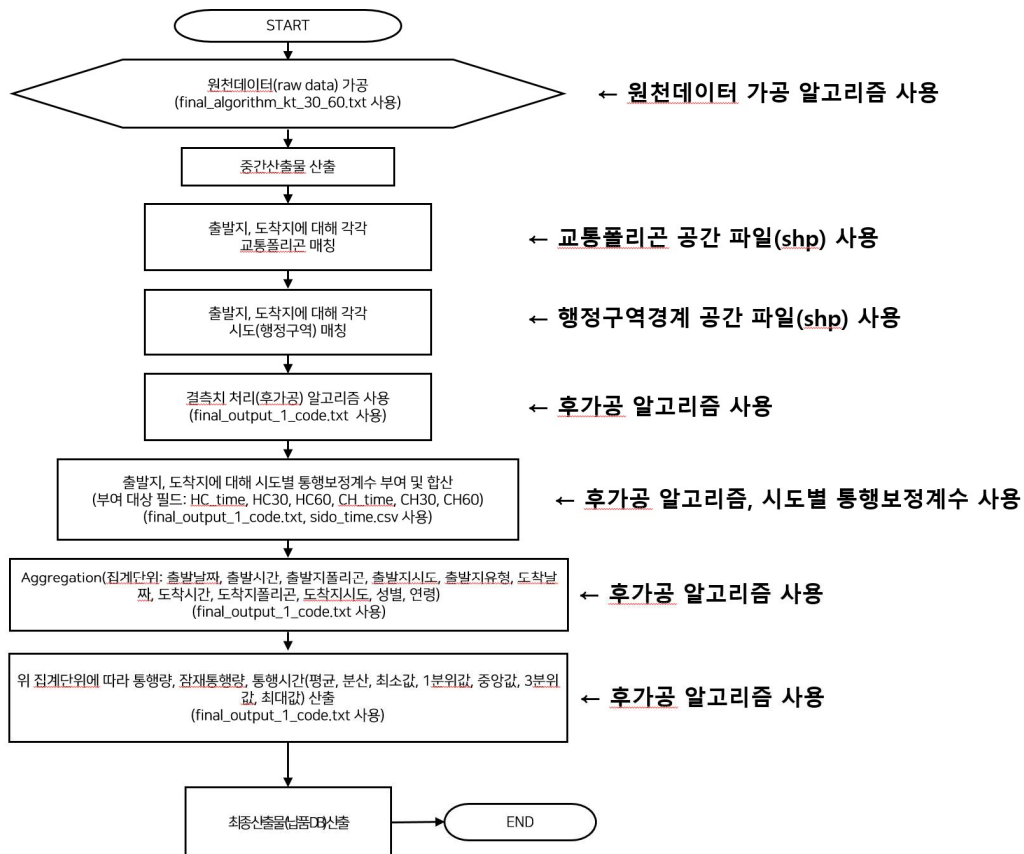
<그림 5> 수단추정 알고리즘 종합

3. 모바일통신 빅데이터 기반 통행 DB 구축 및 검증

가. 기준년도 통행DB 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 1>과 같이 집계된 형태로 구축하고자 함



<그림 6> 신규 모바일통신 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	중앙값		
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	120분이 상	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	120분이 상	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

- 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
- 출발 시간 및 도착 시간 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 매칭 되는 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 머무르는 지를 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 DB 테이블 정의서

컬럼 ID	컬럼명	Type	비고
o_base_ymd	출발 일자	string	yyyymmdd
o_base_dow	출발 요일	string	0 : 일요일 1 : 월요일 2 : 화요일 3 : 수요일 4 : 목요일 5 : 금요일 6 : 토요일
o_timezn_cd	출발 시간대	string	00-23 (1시간 단위)
o_polygon	출발 폴리곤ID	string	-
o_sido	출발 시도ID	string	-
o_trip_type	출발 체류지 유형	string	H: 집 C: 회사 S: 학교 R: 정기 잠재체류지 X: 기타 잠재체류지
d_base_ymd	도착 일자	string	yyyymmdd
d_timezn_cd	도착 시간대	string	00-23 (1시간 단위)
d_polygon	도착 폴리곤ID	string	-
d_sido	도착 시도ID	string	-
d_trip_type	도착 체류지 유형	string	H: 집 C: 회사 S: 학교 R: 정기 잠재체류지 X: 기타 잠재체류지 -: 이동정보 없음
age_itg_cd	연령	integer	0~110 (예: 10: 10세 이상 20세 미만)
sex_type_itg_cd	성	string	M: 남성, F: 여성
total	통행량	integer	보정계수 적용 후 일의 자리에서 올림 처리
potential_travle	잠재통행량	integer	기종점 기준 통행 도중, 60분 이상 체류한 지점으로의 통행
travel_time	통행시간	integer	1분 단위

2) DB 구축 방법

- 원천데이터 수집 및 이상치 제거
 - 2019년 1월 1일부터 2020년 12월 31일까지 기록된 로그를 수집
 - 기지국 좌표가 Null값이거나 행정동 정보가 매칭 되지 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
- 체류순서 부여
 - 로그 기록 일자, 체류시작시간, 체류종료시간을 기준으로 로그 기록 순서에 따라 개별 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
 - 트립타입 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		
주상주 지역	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
잠재 활동 지역	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 원천데이터에서 출발지와 도착지의 위치 정보를 분석 맵의 교통폴리곤 ID로 변경한 후, 시간 정보(체류시작시간, 체류종료시간)에서 분 정보를 생략한 시간대 필드를 생성하고, 1세 단위의 연령 정보를 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 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, 추정 값을 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

- 모바일통신 데이터는 기지국 기반 로그 기록이므로 통행시간 산출 시, 기종점 부분에서 통행시간이 하향계산되는 경우가 발생함
- 이와 같은 문제를 해결하고자, 기지국의 평균 수신 범위를 계산하여 시도별 기지국 내 이동 시간 산출 후, 통행 시간을 보정함
- 각 개인의 통행정보에서 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
- 개별 통행에 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

나. 구축 DB 검증

1) DB 검증 방법

- 모바일통신 데이터를 이용하여 구축된 통행DB의 신뢰성 및 정확성을 검증하기 위하여 기존 가구통행실태조사자료(2016년 기준, 한국교통연구원)와 인구총조사자료(2015년 기준, 통계청)를 활용
- 모바일통신 데이터 기반 통행DB와 기존 가구통행실태조사자료와 인구총조사자료의 통행량, 통행시간 등을 비교·분석하고, 단순 비교할 수 없는 모바일통신 데이터 기반 상세지표(예를 들어, 상세 시간대별, 요일별, 성별, 연령별, 희망통근시간 등)는 통상적인 기대 통행패턴을 벗어나는지 확인
 - － 지역별 통행 발생량·도착량을 가구통행실태조사자료와 상대적으로 비교하여, 유사한 패턴을 보이는지 확인
 - － 시간대별, 요일별 통행 발생량·도착량을 확인하여, 통상적인 침두시간 및 평일·주말 통행패턴이 나타나는지 확인
 - － 성별, 연령별 통행 발생량·도착량을 확인하여, 통상적인 인구비율 및 연령대별 통행패턴이 나타나는지 확인
 - － 지역별 출퇴근 목적 통행량·통행시간을 가구통행실태조사자료, 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

○ 통행 발생·도착패턴 비교 및 분석 결과

<표 4>지역별 통행량 비교 결과

(단위: 통행/일(비율%))

구분	가구통행실태조사자료 기준		모바일기지국데이터 기준	
	발생 (비율)	도착 (비율)	발생 (비율)	도착 (비율)
서울	17,996,813 (20.7)	17,971,878 (20.7)	12,578 (21.2)	12,891 (21.7)
부산	6,457,538 (7.4)	6,458,293 (7.4)	3,847 (6.5)	3,872 (6.5)
대구	4,333,058 (5.0)	4,330,603 (5.0)	2,987 (5.0)	2,964 (5.0)
인천	4,370,318 (5.0)	4,433,014 (5.1)	2,913 (5.0)	2,879 (4.8)
광주	3,117,712 (3.6)	2,942,575 (3.4)	1,673 (2.8)	1,644 (2.8)
대전	3,091,862 (3.6)	3,101,487 (3.6)	1,876 (3.2)	1,855 (3.1)
울산	2,117,733 (2.4)	2,109,519 (2.4)	1,341 (2.3)	1,343 (2.2)
세종	389,864 (0.4)	391,571 (0.5)	325 (0.5)	325 (0.5)
경기	18,668,530 (21.5)	18,628,440 (21.4)	14,155 (24.0)	13,877 (23.4)
강원	2,714,015 (3.1)	2,715,277 (3.1)	1,820 (3.1)	1,812 (3.1)
충북	2,803,473 (3.2)	2,793,476 (3.2)	1,987 (3.4)	1,996 (3.3)
충남	3,871,861 (4.5)	3,863,535 (4.4)	2,270 (3.8)	2,293 (3.9)
전북	2,942,438 (3.4)	2,943,452 (3.4)	1,975 (3.3)	1,973 (3.3)
전남	2,606,282 (3.0)	2,783,178 (3.2)	1,832 (3.1)	1,860 (3.1)
경북	4,431,201 (5.1)	4,442,218 (5.1)	3,202 (5.4)	3,225 (5.4)
경남	5,297,159 (6.1)	5,300,768 (6.1)	3,556 (6.0)	3,529 (6.0)
제주	1,659,446 (1.9)	1,660,019 (1.9)	862 (1.5)	861 (1.5)
계	86,869,303 (100)	86,869,303 (100)	59,199 (100)	59,199 (100)

- 가구통행실태조사자료 기준 1인당 일평균 통행량(발생+도착)은 약 3.4회이며, 모바일통신데이터 기준 1인당 일평균 통행량(발생+도착)은 약 3.1회 임
- 가구통행실태조사자료 기준 전국 통행 발생·도착량은 1.74통행/일, 1.74통행/일이며, 모바일통신데이터 기준 전국 통행 발생·도착량은 1.5통행/일, 1.6통행/일 임
- 모바일통신데이터 기준 통행량과 가구통행실태조사자료 기준 통행량 매우 유사하게 나타남
- 하지만 가구통행실태조사자료가 더 높은 통행량을 띄는데, 이러한 차이는 기존의 가구통행실태조사는 출근, 등교, 업무, 쇼핑, 귀가 등 모든 통행목적에 따라 통행량을 집계하지만, 모바일통신데이터는 체류시간 기준으로 체류지를 구분하기 때문에 기타 잠재체류지에 속하지 않은 이동이 집계되지 않기 때문에 통행량차이가 과소추정되었다고 판단됨
- 지역별 통행량 비율을 살펴보면, 평균적으로는 약 0.41%의 차이로 기존 가구통행실태조사자료와 매우 유사한 패턴을 나타냄
- 경기지역 통행 발생량이 약 +2%로 가장 큰 폭으로 과대추정되었으며, 부산지역 통행 발생량이 약 -1.1%로 가장 큰 폭으로 과소추정되었으나, 평균적으로 약 $\pm 1\%$ 로 매우 신뢰성 높은 결과를 도출함

○ 목적별(출퇴근) 통행패턴 비교 및 분석 결과

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

<표 5> 지역별 출퇴근 통행량 비교 결과

(단위: 통행/일(비율%))

구분	가구통행실태조사자료 기준		인구총조사자료 기준	모바일기저국데이터 기준	
	출근 (비율)	귀가*	통근**	출근	퇴근
서울	4,516,466(20.7)	8,413,729(22.1)	4,597,322(19.6)	23,531(20.0)	22,979(22.3)
부산	1,492,300(6.8)	3,051,128(8.0)	1,500,101(6.4)	7,534(6.4)	6,358(6.2)
대구	974,423(4.5)	1,801,446(4.7)	1,064,227(4.5)	6,053(5.2)	4,616(4.5)
인천	1,213,843(5.6)	1,794,429(4.7)	1,344,311(5.7)	6,116(5.2)	4,828(4.7)
광주	838,601(3.8)	1,279,355(3.4)	639,412(2.7)	3,424(2.9)	2,668(2.6)
대전	736,843(3.4)	1,313,431(3.4)	672,856(2.9)	3,617(3.1)	2,952(2.9)
울산	514,493(2.4)	995,135(2.6)	538,654(2.3)	2,791(2.4)	2,518(2.4)
세종	89,203(0.4)	149,020(0.4)	92,403(0.4)	693(0.6)	514(0.5)
경기	5,272,517(24.1)	7,799,702(20.5)	5,745,634(24.4)	30,438(25.9)	24,143(23.4)
강원	644,578(2.9)	1,186,771(3.1)	703,266(3.0)	3,274(2.8)	2,951(2.9)
충북	637,895(2.9)	1,182,200(3.1)	742,399(3.2)	3,696(3.1)	3,433(3.3)
충남	875,177(4.0)	1,616,017(4.2)	1,032,913(4.4)	4,236(3.6)	4,175(4.1)
전북	665,597(3.0)	1,263,474(3.3)	811,899(3.5)	3,935(3.4)	3,463(3.4)
전남	706,911(3.2)	1,241,157(3.3)	864,498(3.7)	3,506(3.0)	3,328(3.2)
경북	1,007,670(4.6)	2,045,073(5.4)	1,280,135(5.4)	5,774(4.9)	5,981(5.8)
경남	1,372,291(6.3)	2,361,671(6.2)	1,560,582(6.6)	7,103(6.1)	6,591(6.4)
제주	291,636(1.3)	581,151(1.5)	309,133(1.3)	1,663(1.4)	1,479(1.4)
계	21,850,443(100)	38,074,889(100)	23,499,745(100)	117,384(100)	102,974(100)

주 : *가구통행실태조사자료는 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 가구통행실태조사자료·인구총조사자료와 상대적으로 비교하여, 유사한 패턴을 보이는지 확인함

<표 6> 지역별 출퇴근 통행시간 비교 결과

(단위: 분)

구분	가구통행실태조사자료 기준		인구총조사 자료 기준	모바일기지국데이터 기준	
	출근	귀가*	통근**	출근	퇴근
서울	34.8	30.1	45.7	35.4	41.6
부산	30.1	25.6	37.0	30.5	31.2
대구	29.5	28.2	33.4	30.1	26.8
인천	36.6	32.5	43.6	32.1	35.2
광주	20.8	18.3	30.2	26.9	25.4
대전	23.3	21.7	32.3	26.8	26.2
울산	25.6	24.5	28.6	28.1	30.4
세종	15.6	15	29.1	33.0	31.8
경기	41.7	36.3	41.5	37.4	37.0
강원	17.6	20.3	22.1	24.0	26.1
충북	17.9	17.8	25.6	24.9	27.1
충남	15.6	16.7	22.9	25.2	29.6
전북	16	15.4	23.2	25.7	26.3
전남	13.3	13.6	21.3	26.0	30.2
경북	18.8	20	22.1	25.5	30.0
경남	22.5	21.1	25.6	28.0	30.8
제주	20.8	21.5	23.6	32.8	34.6
계	23.6	22.3	35.2	32.1	34.0

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

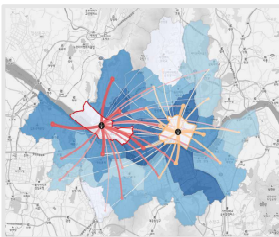
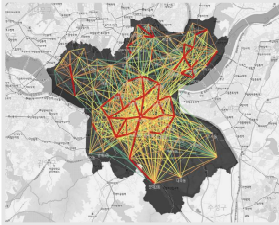
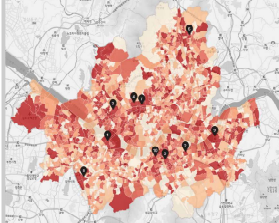
**인구총조사자료는 2015년 기준 통근 통행에 대한 소요 시간별 자료를 일평균시간 형태로 재구성하였으며, 출근과 퇴근이 구분되어 있지 않음

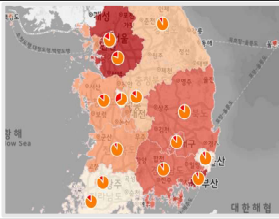
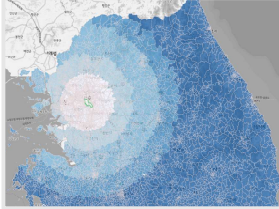
4. 모바일통신 빅데이터 기반 통행 분석 서비스 개발

가. 개요

- 2018년부터 교통 모니터링·데이터 제공·분석 플랫폼(View-T)를 통해 모바일 통신 빅데이터를 기반으로 하는 통행분석 서비스를 제공해왔음
- 2017년 기준 모바일통신 빅데이터 기반 통행 DB를 기반으로 총 5가지의 통행분석 서비스(주요통행지역분석, 주요통행구간분석, Hot place 분석, 내부통행지역분석, 통행시간·거리분석)를 제공하고, 사용자가 시스템에 구축된 기반 데이터를 이중 데이터와 연계하여 분석하거나 다른 방식으로 결과를 표출하는데 활용할 수 있도록 데이터 다운로드 서비스 제공함

<표 7> 2019년 기준 모바일 통신 빅데이터 기반 통행 분석 서비스

명칭	기능 설명	2019년 개발 내용
주요 통행 지역 분석	 <p>선택 지역에 대한 주요 출발지와 주요 도착지를 확인·비교할 수 있는 기능</p>	<ul style="list-style-type: none"> - 2018년에 개발된 기능 개선 <ul style="list-style-type: none"> · 기존보다 더 많은 지역을 비교 분석할 수 있도록 분석 대상 지역을 확대 (2개 지역 선택 가능 → 5개 지역 선택 가능) · 표출 방법 개선 (선택지역별 분석결과를 각기 다른 색상으로 표현) · 행정구역 ON/OFF 기능 추가 · 상세결과를 파이차트 형식으로 제공
주요 통행 구간 분석	 <p>통행이 많은 구간을 확인할 수 있는 기능</p>	<ul style="list-style-type: none"> - 2018년에 개발된 기능 개선 <ul style="list-style-type: none"> · 지역간 통행 정보를 확인할 수 있도록 분석 대상 지역을 확대 (1개 지역 선택 가능 → 5개 지역 선택 가능) · 표출 단위 확대 (교통폴리곤 단위로 표출 → 교통폴리곤, 읍면동, 시군구 단위로 표출)
Hot Place 분석	 <p>통행수요가 높은 출발지 또는 도착지를 확인할 수 있는 기능</p>	<ul style="list-style-type: none"> - 2018년에 개발된 기능 개선 <ul style="list-style-type: none"> · 표출 단위 확대 (교통폴리곤 단위로 표출 → 교통폴리곤, 읍면동, 시군구 단위로 표출) · 다양한 통행 정보를 표출 (총 통행량 표출 → 총 통행량, 일평균 통행량, 면적대비 통행량 표출)

내부 통행 지역 분석		내부 통행량이 많은 지역(전체 통행량 대비 내부 통행량 비율)을 확인할 수 있는 기능	- 신규 기능 개발
통행 시간 거리 분석		지역별 접근성(평균 통행시간, 평균 통행거리)을 확인할 수 있는 기능	- 신규 기능 개발

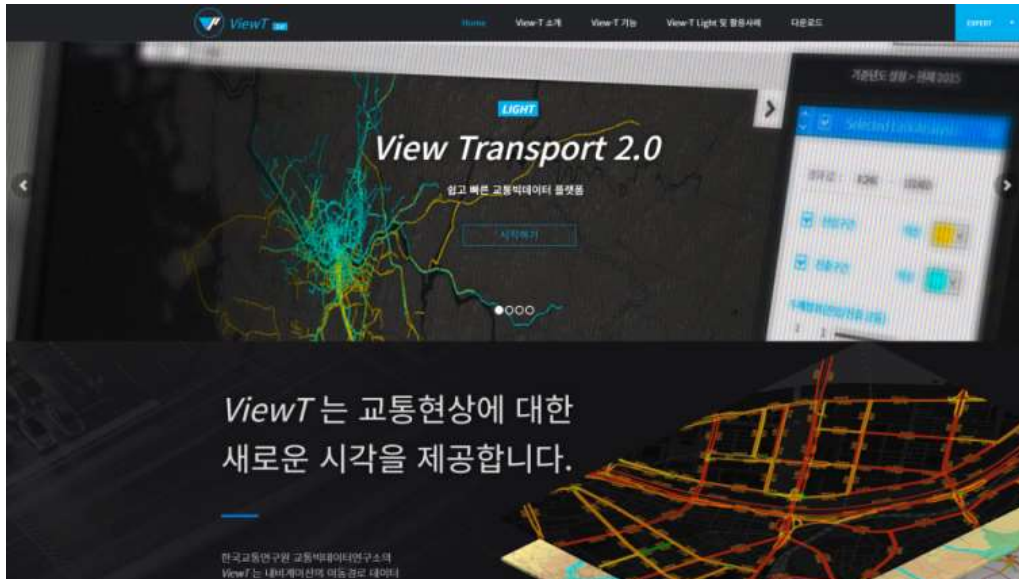
- 본 과업에서는 기 개발된 통행 분석 서비스를 이용자의 편의성 측면을 고려하여 개선하고, 이용자가 최신의 통행 정보를 확인할 수 있도록 전년도 과업에서 구축한 2018년 기준의 통행 DB와 기반 맵 DB로 업데이트 함
 - View-T 및 모바일 통신 데이터 기반 통행 분석에 대한 이해를 돕기 위해 View-T Portal 사이트를 제작
 - 일반 국민도 쉽고 빠르게 통행 정보를 확인할 수 있도록 일반사용자용 분석서비스(View-T Light)를 별도로 개발
 - 기존에 제공했던 분석서비스는 고급사용자용 분석서비스(View-T Expert)로 전환하고, 교통관련 전문가가 더욱 심층적인 분석을 수행할 수 있도록 신규 분석 기능을 추가로 개발
 - 사용자가 기존보다 신속하게 분석 결과를 확인할 수 있도록 2018년 기준 통행 DB는 교통폴리곤별, 행정구역별, 분석조건별로 집계하여 구축
 - View-T 내 구축된 통행 DB를 사용자가 타 연구에도 활용할 수 있도록, 3가지 다운로드 서비스를 제공

나. 분석 서비스 고도화

1) View-T Portal

- View-T를 처음 접한 사용자를 위한 서비스로, 시스템 전반의 이해를 돕기 위해 “View-T 소개, View-T 기능, View-T Light 및 활용사례, 다운로드” 네 가지 메뉴

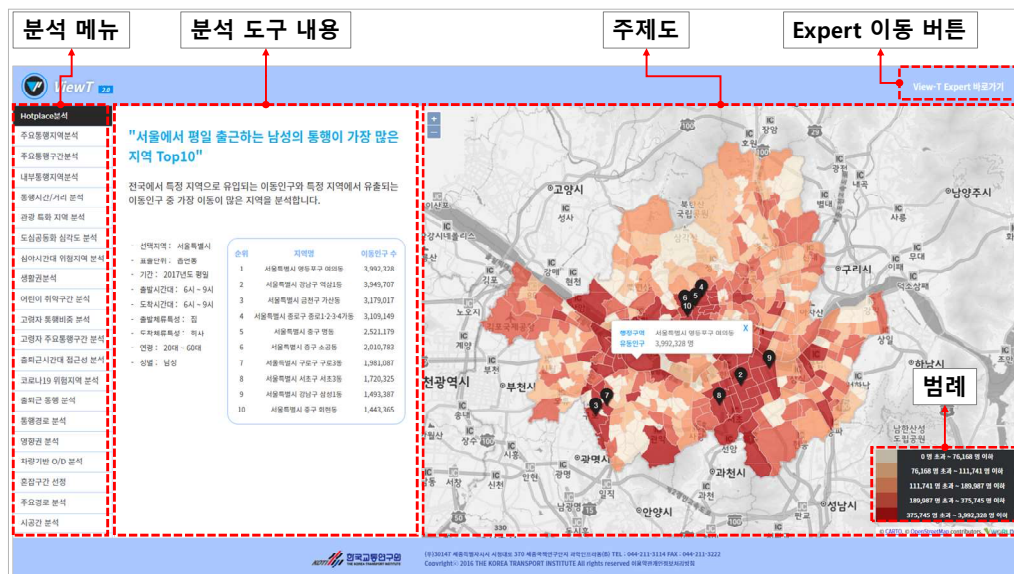
- 를 구성하고, 각 메뉴에서 View-T의 구성요소에 대한 설명 및 활용사례를 소개
- View-T Light와 View-T Expert로 진입 가능



<그림 7> View-T Portal 메인 화면

2) View-T Light

- 교통 비전문가, 일반 국민도 쉽고 빠르게 통행 정보를 확인할 수 있도록 개발한 통행 분석서비스로, 출근, 관광 등 일반 국민에게 친숙한 주제를 기반으로 분석 기능을 개발
 - 사용자가 간단한 마우스 조작만으로도 분석 결과를 확인할 수 있도록 기능을 개발
 - 사용자가 각 분석 기능을 쉽게 이해할 수 있도록 대표적인 캐치프레이즈, 분석 내용에 대한 간단한 설명과 함께 분석에 사용된 기본 설정 값을 좌측에 표시
 - 사용자가 직관적으로 분석 결과를 해석할 수 있도록 주제도, 테이블 차트, 그래프 등을 시각화하여 제공



<그림 8> View-T Light 화면 구성

- 기존 View-T에서 제공했던 5가지 분석 기능과 8가지 새로운 통행 분석 아이템을 주제로 View-T Light 분석서비스를 구성

<표 8> View-T Light 분석서비스 - 기존 분석 기능의 View-T Light 버전 개발 (5가지)

분석서비스		분석 및 표출 조건
명칭	설명	
Hot place 분석	전국에서 특정 지역으로 유입되는 이동인구와 특정 지역에서 유출되는 이동인구 중 가장 이동이 많은 지역을 간단하게 분석	서울에서 평일 출근하는 남성의 통행이 가장 많은 지역 Top10을 표출
주요통행 지역 분석	특정 출발지에서 특정 도착지의 지역별 통행량을 간단하게 분석 (마우스 클릭만으로 출발지와 도착지를 선택하여 유동인구를 파악할 수 있는 기능)	강남구를 출발지로 하는 전국 시군구로의 유동인구를 표출
주요통행 구간 분석	통행이 많이 일어나는 구간을 간단하게 분석	전국 시도간 통행이 많이 일어나는 Top 10 구간을 표출
내부통행 지역 분석	출발지와 도착지가 같은 내부통행량을 주제도로 보여주고, 파이차트를 이용하여 내부통행량과 외부통행량의 비율을 간단하게 분석	경제활동 자체 수용 비율이 가장 높은 시/도를 표출
통행시간/거리 분석	특정 분석지역을 출발 또는 도착 기준으로 하여 전국단위로 이동한 통행시간과 통행거리를 간단하게 분석	세종시에서 출발한 통행에 대한 평균통행시간과 거리를 표출

<표 9> View-T Light 분석서비스 - 신규 통행 분석 아이템의 View-T Light (8가지)

분석서비스		분석 및 표출 조건
명칭	설명	
관광 특화 지역 분석	계절에 따라 통행량 변화 큰 지역을 확인할 수 있는 기능	특정 계절에 여가 목적 통행의 통행량 변화가 큰 지역을 표출
도심공동화 심각도 분석	도심공동화가 나타나는 지역을 분석할 수 있는 기능	낮 시간대(10~18시) 통행량 대비 심야시간대(22~06시) 통행량이 매우 낮은 지역을 표출
생활권 분석	주요활동(통행) 권역을 확인할 수 있는 기능	시군구별 기종점 point를 기준으로 SDE(Standard Deviation Ellips) 분석을 통해 중첩되는 비율을 측정한 후, 기준치보다 높은 중첩 비율을 보이는 지역을 하나의 생활권역으로 표출
어린이 취약구간 분석	어린이 교통사고가 발생할 가능성이 높은 지역을 확인할 수 있는 기능	어린이 교통사고 가능성(학교를 기준으로 반경 300m내 통학시간대 어린이 주요통행구간을 모바일 통신데이터로 추출한 후, 내비게이션 데이터를 기반으로 차량 통행이 많은 구간을 추출)이 높은 지역을 시각화
고령자 통행비중 분석	고령자의 통행비중이 높은 지역을 확인할 수 있는 기능	전체 통행 중에서 고령자(70~110대 연령 통행자)의 통행 비중이 높은 지역을 표출
고령자 주요통행구간 분석	고령자가 주로 통행하는 구간을 확인할 수 있는 기능	고령자(70~110대 연령 통행자)의 통행량이 많은 구간을 1위부터 10위까지 표출
출퇴근 시간대 접근성 분석	출퇴근시간대 목적지까지 접근성을 확인할 수 있는 기능	통행시간이 1시간 이상 소요되는 지역을 표출
코로나19 위험지역 분석	코로나19에 노출될 가능성이 높은 지역을 확인할 수 있는 기능	코로나19 확진자 이동 동선과 통행 발생량이 높은 지역이 일치하는 지역을 위험지역으로 선정하여 표출

3) View-T Expert

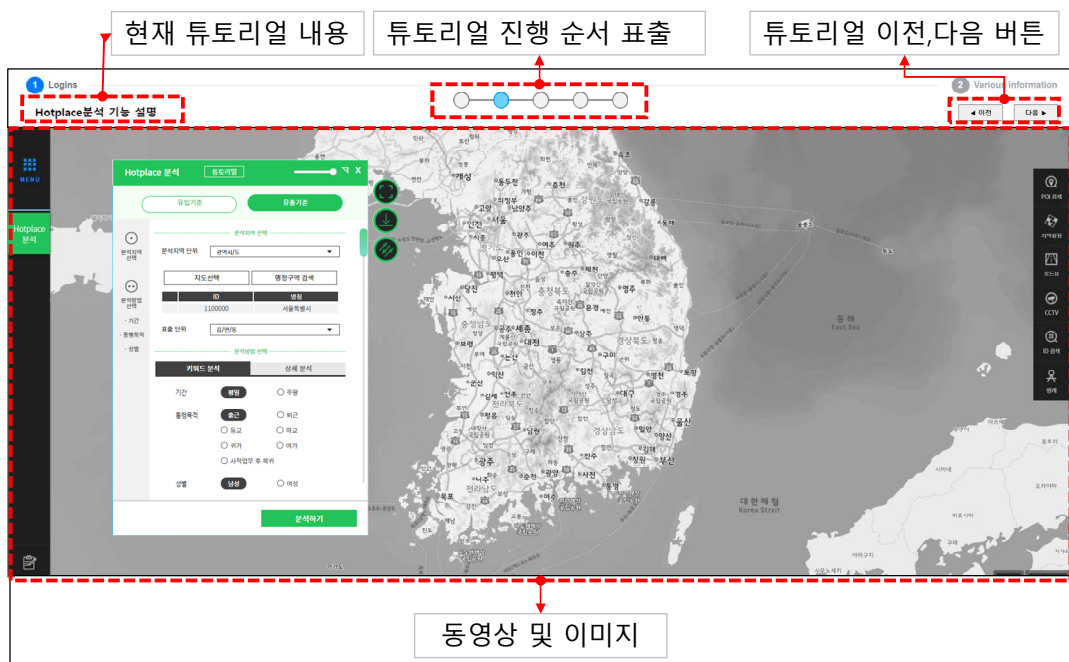
- View-T Light에서 제공하는 분석서비스보다 심층적인 분석을 수행할 수 있도록 개발한 분석서비스로, 사용자가 이용하기 편리하도록 기존 View-T 분석서비스를 개선하고, 신규 분석 기능을 추가함
- 분석 조건 설정, 범례 설정 간소화
 - “키워드 분석” 옵션을 조건 설정 부분에 추가하여, 사용자가 상세하게 조건을 입력하

지 않아도 분석을 수행할 수 있도록 함

- 필수입력 조건이 아닌 범례 및 표출정의를 별도의 아이콘으로 배치하여 레이아웃의 여유 공간을 확보하고, 요소 전체 UI를 개선해 사용자가 정보를 식별하는데 수월할 수 있도록 개선

○ 분석 기능 튜토리얼 제작

- 분석도구 사용이 익숙하지 않은 사용자에게 동영상 또는 그림을 통하여 통신 주요분석기능의 기능 및 분석 결과 설명, 분석 도구의 사용 방법을 제공



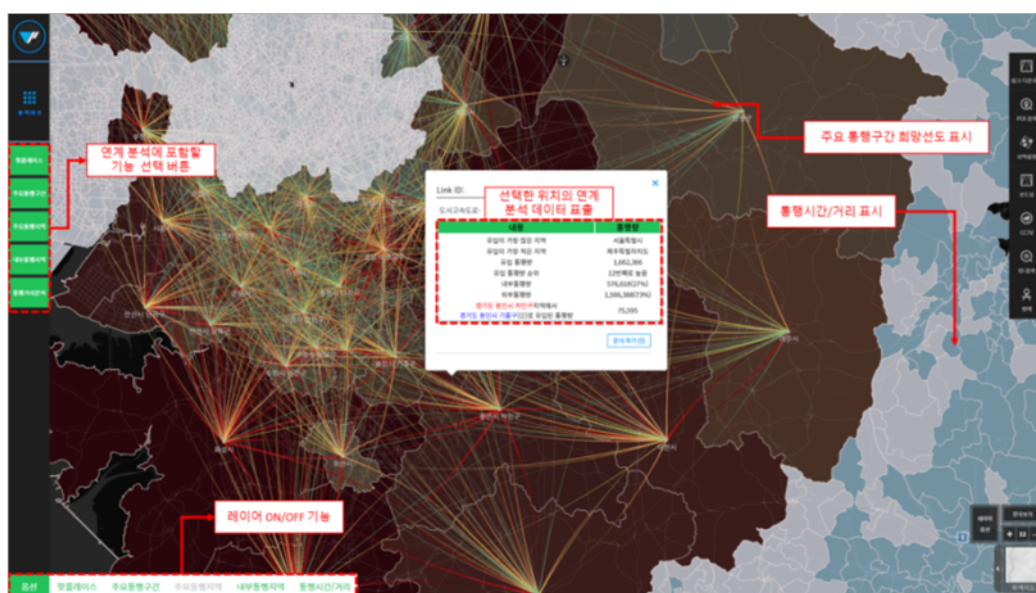
<그림 9> 분석기능 튜토리얼 UI

○ 기존 분석기능(주요통행구간분석) 개선

- 시도, 시군구 단위 간 구간 분석이 가능하도록 개선하여 단거리 구간 통행분석뿐만 아니라 장거리 구간 통행분석도 수행할 수 있도록 함

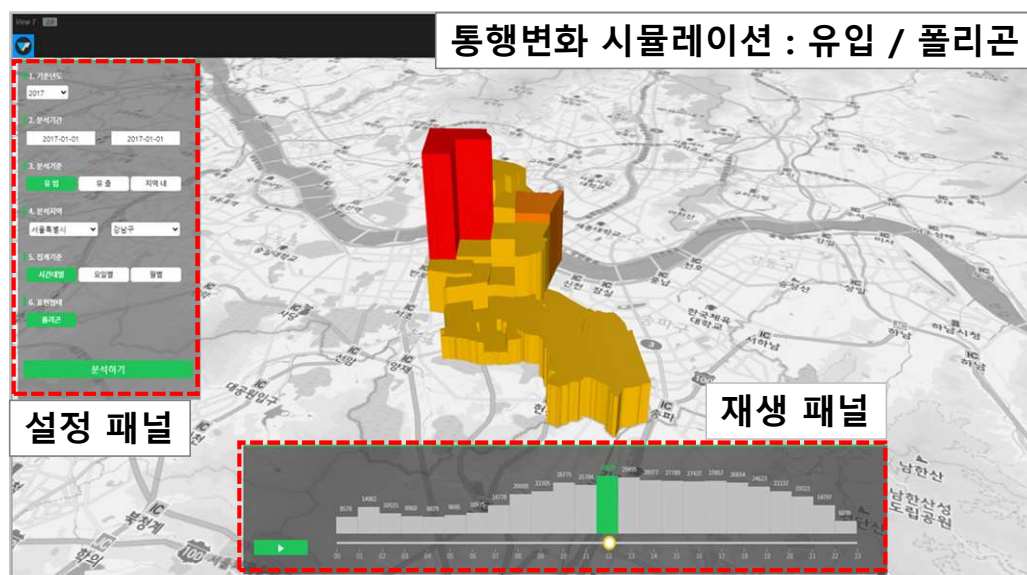
○ 신규 분석 기능 개발

- 심층 분석을 가능케 하는 분석 기능 탑재
- 2개 이상의 분석 수행 결과를 주제도로 동시에 확인할 수 있는 “정보 연계 분석 기능” 개발



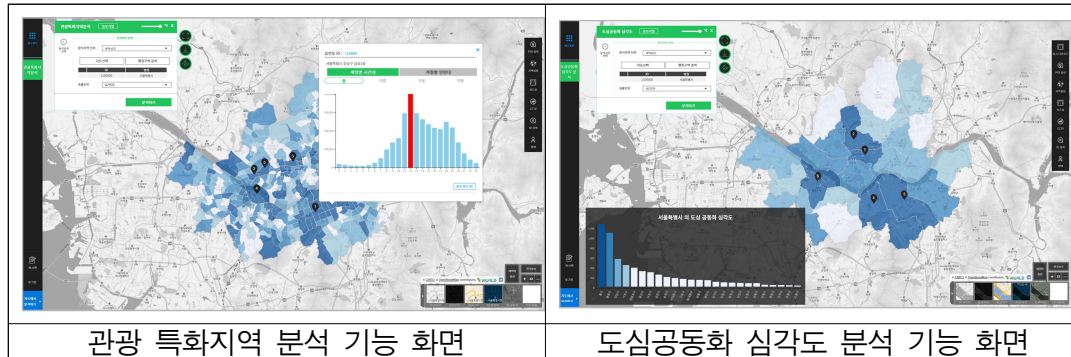
<그림 10> View-T Expert 신규분석 기능 개발 - 정보 연계 분석 기능

- 통행변화를 시계열로 확인할 수 있는 “통행변화 시뮬레이션 분석 기능” 개발



<그림 11> View-T Expert 신규분석 기능 개발 - 유입 기준 폴리곤 형태의
통행변화 시뮬레이션

- View-T Light의 분석서비스 중 일부(관광 특화지역 분석, 도심공동화 심각도 분석)를 Expert에 개발



<그림 12> View-T Expert 신규분석 기능 개발 - Light 분석서비스 중 일부 구현

다. 기반 DB 현행화 및 속도 개선

1) 모바일 통신 빅데이터 기반 통행정보 DB

- 검색·표출 시간을 단축할 수 있도록 약 145억 건의 2018년 기준 통행정보 DB를 3가지 형태로 집계하여 구축
 - 교통폴리곤 단위별 출발/도착에 대한 통행량을 집계
 - 출발 교통폴리곤ID, 도착 교통폴리곤ID, 체류특성, 연령대, 성별 필드를 키(Key)로 구성하고 키 값이 동일한 데이터들을 그룹화

<표 10> 교통폴리곤 기준 집계 통행량 DB 구축 효과 (도큐먼트 수 변화)

구분	가공 전	가공 후	감소율
교통폴리곤 기준 집계	약 145억 건	약 9억 건	약 94%

- 행정구역 단위별 출발/도착에 대한 통행량을 집계
 - 1단계에서 매칭한 행정구역 정보를 활용하여 출발 행정구역 기준 집계 데이터, 도착 행정구역 기준 집계 데이터, 출발과 도착이 모두 행정구역 단위인 집계 데이터를 구축

<표 11> 행정구역 기준 집계 통행량 DB 구축 효과

구분	가공 전	가공 후	감소율
읍면동 기준 집계	약 9억 건	약 3억 건	약 30%
시군구 기준 집계	약 9억 건	약 1,300만 건	약 98%
시도 기준 집계	약 9억 건	약 12만 건	약 99%

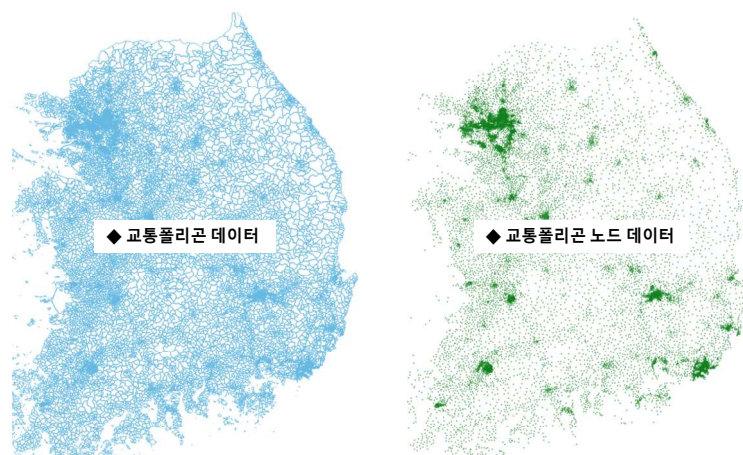
- 사용자가 자주 검색하는 분석 조건을 기준으로 통행 정보를 집계
- 교통폴리곤 기준 통행량 DB를 통행목적 조건에 부합하는 형태로 데이터를 압축(경량화)

<표 12> 분석특성별(통행목적별) 집계 통행량 DB 구축 효과

구분	가공 전	가공 후	감소율
통행목적 기준 집계	약 9억 건	약 1.7억 건	약 79%

2) 모바일 분석 맵 DB

- 교통폴리곤, 읍면동, 시군구, 시도 총 4가지의 맵을 폴리곤과 노드 데이터(대표포인트) 형상으로 시스템에 구축
- 교통폴리곤(2018년 기준): 기지국 수신범위(셀 반경)와 집계구를 매칭시켜 모바일 통신 빅데이터 기반 통행 정보(통행량)를 집계하는데 용이한 형태로 구축된 맵
- 읍면동, 시군구, 시도(2017년 기준): 시도 17개, 시군구 250개, 읍면동 3,500개로 구성



<그림 13> 모바일 분석 맵 DB 구축 (예시: 교통폴리곤)

라. 데이터 다운로드 서비스 개선

- 데이터 제공처(통신업체), 데이터 수요처(국민, 민간기업 등)의 요구를 반영하여 다운로드 서비스를 “공공용”, “인가용”, “주문형” 세 가지로 구분

<표 13> 2020년 기준 다운로드 정책

구분	공공용	인가용	주문형
목적	사용자(일반 국민) 대기시간 최소화	데이터 제공처의 데이터 판매사업과의 상충 문제 해소	사용자(민간기업, 지자체)의 분석 목적 반영
데이터 형태	시/도 단위의 통행 정보	시군구 이하 단위의 통행 정보	시군구 이하 단위의 통행 정보 (이종 데이터 연계 고려)
절차	데이터 신청서 작성(기본 정보만 작성) → 데이터 다운로드	데이터 신청서 작성(상세 정보까지 작성) → 데이터 제공처(통신업체) 승인 → 데이터 다운로드	데이터 신청서 작성(상세 정보까지 작성) → 데이터 제공처(통신업체) 승인 → 데이터 가공(KOTI) → 데이터 다운로드

5. 결론 및 차년도 수행계획

가. 결론

- 본 과업에서는 기존 방법보다 정확하게 통행시간 및 통행량 등 통행지표를 산출하기 위하여 '19년에 개발한 모바일 가공 알고리즘을 보완함
 - 기존의 이동속도 및 패턴기반 기지국 신호 보정 알고리즘을 보완하여, 데이터의 이상치를 효과적으로 보정하기 위한 알고리즘을 제시함
 - 신규 알고리즘은 합리적 수준의 시·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 점단위의 위치 추정 정보로 변환하는 방법을 제안하여 모바일통신 데이터의 공간적 해상도를 개선하였으며, 향후 이동 경로 및 수단 추정을 위한 개발 알고리즘의 적용방안을 검토함
- 2019년 1월 1일부터 12월 31일까지 생성된 로그 기록에 새로 개발한 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
 - 모바일통신 빅데이터기반으로 구축된 통행DB의 시간대별, 요일별, 성별, 연령별 통행량 및 통행시간을 기존 국가교통DB사업의 '가구통행실태조사자료 (2016년)'와 통계청의 '인구총조사자료 (2015년)'비교하여 검증함
- 모바일통신 빅데이터 기반 통행 분석 서비스의 성능 및 사용성을 개선하기 위하여 DB를 재구조화하고 이용자 맞춤형 'Light version'을 추가 개발함
 - 일반 국민도 쉽고 빠르게 통행 정보를 확인할 수 있도록 일반사용자용 'Light version'의 분석서비스를 추가로 개발하고, 기존에 제공했던 분석서비스는 'Expert version'으로 명명하고 전문가에게 친숙한 UI/UX로 변경함
 - 또한, 2개 이상의 분석 수행 결과를 주제도로 동시에 확인할 수 있는 '정보 연계 분석 기능' 개발하고, 통행변화를 시계열로 확인할 수 있는 '통행변화 시뮬레이션 분석 기능' 개발함

나. 차년도 수행계획

- 모바일통신 빅데이터 가공 로직 개발
 - － 기 구축된 DB (전국, 2019년도) 검증
 - 국가교통DB 등 기존 조사기반 통계를 이용한 검증
 - GPS 정보 등을 활용한 신뢰성 검증 (신규)
 - － 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
 - 모바일통신 빅데이터 기반 추정 가능 체류지(거주지, 근무지, 여가 등) 선정
 - 거주지 및 근무지 추정 로직 개선
 - － 통행경로 및 수단 구분 알고리즘 개발 (신규)
 - 개별통행에 대한 이동경로 추정 및 보정 로직 개발 (신규)
 - 대중교통(철도, 버스) / 승용차 / 기타 이용자 구분 로직 개발 (신규)
- 모바일통신 빅데이터기반 교통지표 개발
 - － 교통지표 개발
 - 모바일통신 빅데이터 DB 기반 산출 가능한 통행지표 개발 (통행량, 평균통행시간, 평균통행거리 등)
 - 이종 데이터와 모바일통신 빅데이터 DB를 연계·가공하여 산출 가능한 각종 지표 개발 (교통안전지수, 도시공동화지수, 교통서비스지수, 경제활동지수 등) (신규)
 - － 교통빅데이터 플랫폼 분석기능 개발 지원
 - 교통빅데이터 플랫폼을 위한 개별통행 DB 재구조화 지원
 - 교통지표 제공 및 분석 서비스 개발 지원

제1장 과업의 개요

제1절 배경 및 목적

제2절 범위 및 내용

제1장 과업의 개요

제1절 배경 및 목적

1. 배경

- 대중교통 이용 활성화 및 정책을 지원하기 위해서는, 대중교통 이용자뿐만 아니라 잠재적으로 대중교통을 이용할 가능성이 높은 통행자에 대한 분석이 필수적임
 - － 대중교통 이용자에 대한 통행은 교통카드 DB(대중교통 승하차 시 태그 정보)로 대부분 추정할 수 있으나, 잠재적으로 대중교통을 이용할 가능성이 있는 타 교통수단 이용자(차량, 도보 등)에 대한 통행은 일부만 추정 가능
 - － 차량에 탑재된 기기(내비게이션, DTG 등)나 도로에 설치된 기기(검지기, DSRC 등)를 통해 차량 통행자에 대한 통행 정보는 취득할 수 있으나, 자전거·도보 등으로 이동한 통행자에 대한 통행 정보는 표본율이 낮은 가구통행실태조사 결과에만 의존하고 있는 실정
- 스마트폰 보급률 99%에 육박하는 모바일통신 빅데이터를 이용하면, 국민의 거의 모든 통행 정보를 분석할 수 있음
 - － 모바일 기기와 기지국 간의 송·수신 이력인 모바일통신 빅데이터는 대부분 국민이 이용하고 있는 모바일 기기를 통해 축적되기 때문에 표본율이 매우 높고, 이동 수단에 구애받지 않고 개별통행에 대한 정보를 상세하게 포함하고 있음

2. 목적

- 통행자별 이용 교통수단 정보가 포함된 모바일통신 빅데이터를 이용하여 통행DB를 구축하고, 이를 기반으로 교통 수요를 분석할 수 있는 지표, 기능 등을 개발하여 고도화된 대중교통 정책 수립 지원에 활용하고자 함

제2절 범위 및 내용

1. 범위

- 시간적 범위
 - － 모바일 빅데이터 가공 알고리즘 개발: 2019년 4월, 2020년 4월의 원천데이터
 - － 2019년도 기준 기반 DB 구축: 2019년 1월 1일 ~ 2019년 12월 31일 (1년)
 - － 웹 서비스를 위한 각종 분석도구 개발: 2018년 1월 1일 ~ 2018년 12월 31일
 - 2019년도 연구 성과물인 2018년도 기준 기반 DB 활용
- 공간적 범위: 전국

2. 내용

- 모바일통신 빅데이터 가공 알고리즘 개발
 - － 데이터 전처리 알고리즘 고도화
 - － 체류정보 및 통행목적 추정 알고리즘 고도화
 - － 이동수단 추정 알고리즘 개발
- 모바일통신 빅데이터 기반 통행 DB
 - － 기준년도 통행DB 구축 및 검증
 - － 주요 교통지표 생성 및 분석
- 모바일통신 빅데이터 기반 통행 분석 서비스 개발
 - － 분석 서비스 고도화
 - － 기반DB 현행화 및 속도 개선
 - － 데이터 다운로드 서비스 개선

DB 구축 부문 (모바일 데이터 기반 통행지표)

통행지표(추출목표) 정의 및 원천데이터 분석

- 분석자료(원천데이터) 추출 및 선분로그화 행압축
- 통행지표(추출목표) 정의 및 분석 방향 설정
- 원천데이터 분석 및 가공 방향 설정
(시공간적, 인적특성별 로그 기록 특성 파악 → 가공 방향 설정)



가공 알고리즘 개발

- 가공 알고리즘 설계 및 구현
- 설계(안)별 결과 비교 및 최적안 선정
- 알고리즘별 기준 로직 및 가공 순서 정의



모바일 데이터 기반 통행지표 구축

- 통행지표별 DB 설계
- 알고리즘 기반 소스코드 개발
- 원천데이터 가공 및 통행지표 생성
- 전수화를 위한 보정계수 적용



모바일 데이터 기반 통행지표 검증

- 검증 방향 설정
- 검증 기준에 따른 통행지표 검토
(이상치 발견시 단계별 가공결과 확인 및 로직 수정)

분석서비스 개발 부문

플랫폼(View-T) 내 기반 DB 구축

- '18년도/19년도 통행지표·분석맵 데이터 수집 및 검수
- DB 구축 과정 정형화 및 간소화 방향 설정
- 플랫폼 내 통행지표 데이터 구축
(데이터 경량화 → 카테고리별 데이터 집계 → 인덱스 생성)
- 플랫폼 내 분석맵 구축
(형상정보 가공 → 속성정보 가공 → 데이터 변환 → GeoServer 레이어 구축)



분석 서비스 개발

- 기존 분석서비스 문제점 및 개선방향 도출
- 신규 분석서비스 발굴 및 개발방향 설정
- 메뉴 구성 및 UI/UX 설계
- 분석 및 표출 기능 개발
- 분석서비스 테스트



부가기능 개발

- 기존 부가기능 문제점 및 개선방향 도출
- 사용자 편의 기능 개선
- 다운로드 기능 개선
- 부가기능 테스트

'19년도 통행지표
검증 완료

<그림 1- 1> 과업의 수행절차

제2장 모바일통신 빅데이터 가공 알고리즘 개발

제1절 데이터 전처리 알고리즘 고도화

제2절 체류정보 및 통행목적 추정 알고리즘
고도화

제3절 이동수단 추정 알고리즘 개발

제2장 모바일통신 빅데이터 가공 알고리즘 개발

제1절 데이터 전처리 알고리즘 고도화

1. 기존 가공 알고리즘의 보완 및 개선 필요성

- 기존 알고리즘의 경우, 속도기반 방법을 사용하여 시그널점프 현상에 의해 발생한 이상치를 보정하고, 패턴 기반 방법을 사용하여 핑퐁핸드오버 현상에 의해 발생한 이상치를 보정하는 등 신호 이상 현상별로 다른 알고리즘을 적용
 - 시그널점프 데이터로 식별되어 제거된 데이터 중에서 핑퐁핸드오버 데이터가 포함되어 있을 수 있음
 - 핑퐁핸드오버 패턴이 다양하고 실제로는 핑퐁핸드오버 현상이 무수히 많이 발생할 수 있기 때문에 3회 반복되는 $A \rightarrow B \rightarrow A \rightarrow B$ 기준만으로는 모든 핑퐁 핸드오버 데이터를 보정하는 것이 어려움
- 기존 알고리즘의 경우, 특정 시간(25분)을 기준으로 데이터 로그별 체류/이동 여부를 판단하고 주요 통행 시간대나 연령 등을 기준으로 체류지를 5-6개의 유형으로 구분
 - 대표 통행자(통근·통학 위주)에 대한 통행 시간대를 미리 가정한 후, 통계청에서 공표하고 있는 인구수(통근·통학 인구)와 가장 유사한 값을 갖는 통행 패턴 조건을 찾는 방식으로, 대표 통행자 외의 통행자(자영업자, 비정기적 통행자 등)는 배제됨
 - 사전에 정의한 통행시간대의 변화에 따라 통행 패턴 조건이 달라질 수 있음

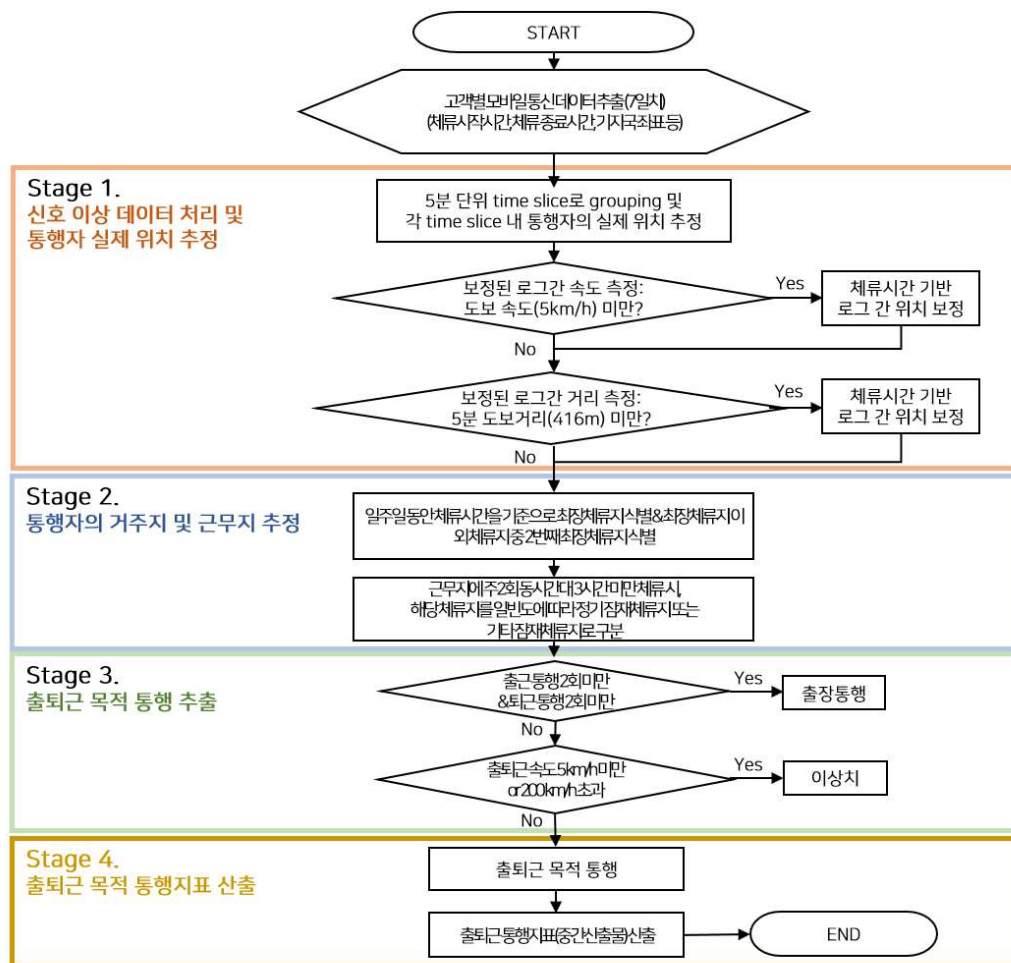
2. 신규 가공 알고리즘의 주요 고려사항 및 개념

가. 신규 가공 알고리즘의 주요 고려사항

- 신규 가공 알고리즘은 기지국과 단말기 간에 발생하는 불규칙적인 신호 이상 현상을 효과적으로 보정 할 수 있어야 함
- 또한, 개별통행 특성 및 지역 간 기지국 거리 등을 고려하여 이동/체류를 구분하고 체류지 유형을 추정하여야 함

나. 신규 가공 알고리즘의 주요 개념

- 입력데이터(input data)로 각 통행자에 대한 7일 치 원천데이터가 사용되며 각 통행자에 대한 개별 통행 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 이동 궤적을 보정
- 기지국 데이터 기록빈도 및 체류시간을 기반으로 통행자의 개별 통행 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- 1> 신규 모바일통신 빅데이터 가공 알고리즘

3. 신규 데이터 전처리 알고리즘 개발

가. 시간적 데이터 보정

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

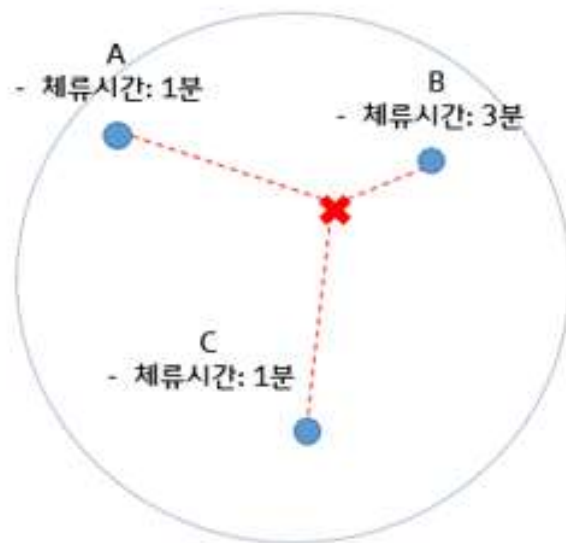


<그림 2- 2> 시간적 데이터 보정 방법 예시

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

- 데이터를 분할 한 후, 각 시간 단위(time slice) 내에 기록된 기지국 데이터의 기록빈도를 기반으로 통행자의 실제 위치 추정

※ 통행자의 실제 위치 추정 방법: <그림 2-3>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



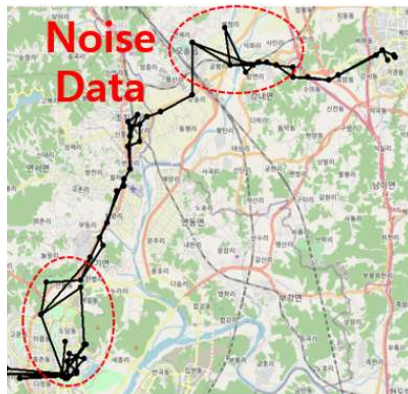
<그림 2- 3> 통행자 실제 위치 추정 방법 예시

나. 공간적 데이터 보정

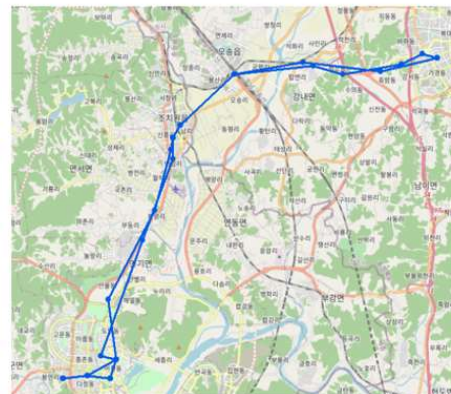
- Browning R. C., Baker E. A., Herron J. A., and Kram R.(2006)²⁾이 제시한 보행자 평균 도보 속도(5km/h)를 기준으로 5분(시간적 데이터 보정 시 설정한 시간 단위) 거리(약 416m) 범위 내 위치한 체류지들에 대해 각 체류지에서의 체류시간을 기반으로 해당 시간대에서의 통행자 실제 위치를 추정하고자 함
- 원천데이터에 대해 시·공간적으로 신호 이상 기록을 보정해줌으로써, 분 단위로 연속해서 변화하거나 좁은 공간적 범위 내에 산발적으로 기록되는 기지국 위치 보정 가능



<그림 2- 4> 공간적 데이터 보정 방법 예시



<데이터 전처리 전 통행 궤적>



<데이터 전처리 후 통행 궤적>

<그림 2- 5> 시·공간적 데이터 보정(전처리) 적용 전, 후 통행 궤적 비교

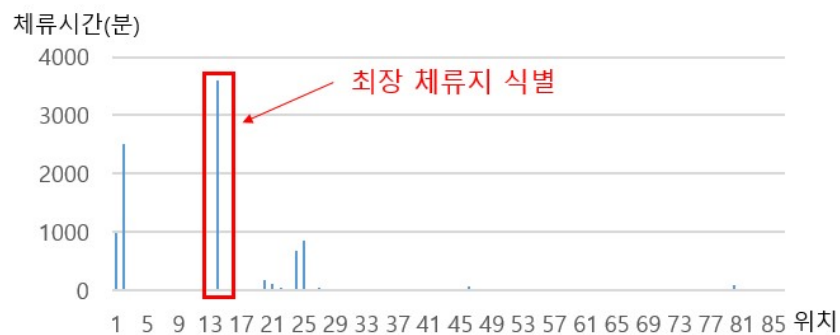
²⁾ R. Browning, E. A. Baker, J. A. Kram, "Effects of obesity and sex on the energetic cost and preferred speed of walking", Journal of Applied Physiology, 2006

제2절 체류정보 및 통행목적 추정 알고리즘 고도화

1. 신규 체류정보 및 통행목적 추정 알고리즘 개발

가. 이동/체류 구분 및 체류유형 추정

- 데이터 전처리 단계에서 추정한 통행자의 실제 위치를 기준으로 통행자가 일주일 동안 각 위치에 몇 분 체류하였는지 계산 후, 최장 체류지 식별



<그림 2- 6> 최장 체류지 식별 예시

- 통행자가 최장 체류지 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	→ 시간 1: 6분
05:02	09:03	04:02	최장 체류지	
09:04	09:06	00:03	X`2	← 시간 2가 가장 김
09:07	09:10	00:04	X`3	
09:10	09:15	00:06	X`4	→ 시간 2: 8시간 42분
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	→ 시간 3: 10분
18:00	21:30	03:31	최장 체류지	
...	

<그림 2- 7> 최장 체류지 이외의 체류지에 체류한 시간 중 가장 긴 시간 식별 예시

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

체류시작시간	체류종료시간	체류시간	동행자 위치
00:00	04:56	04:57	max1 X ¹
05:01	05:02	00:02	X ¹
05:02	09:03	04:02	max1 X ¹
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 ¹
17:50	17:54	00:05	X ⁸
18:00	21:30	03:31	max1 X ¹
***	***	***	***

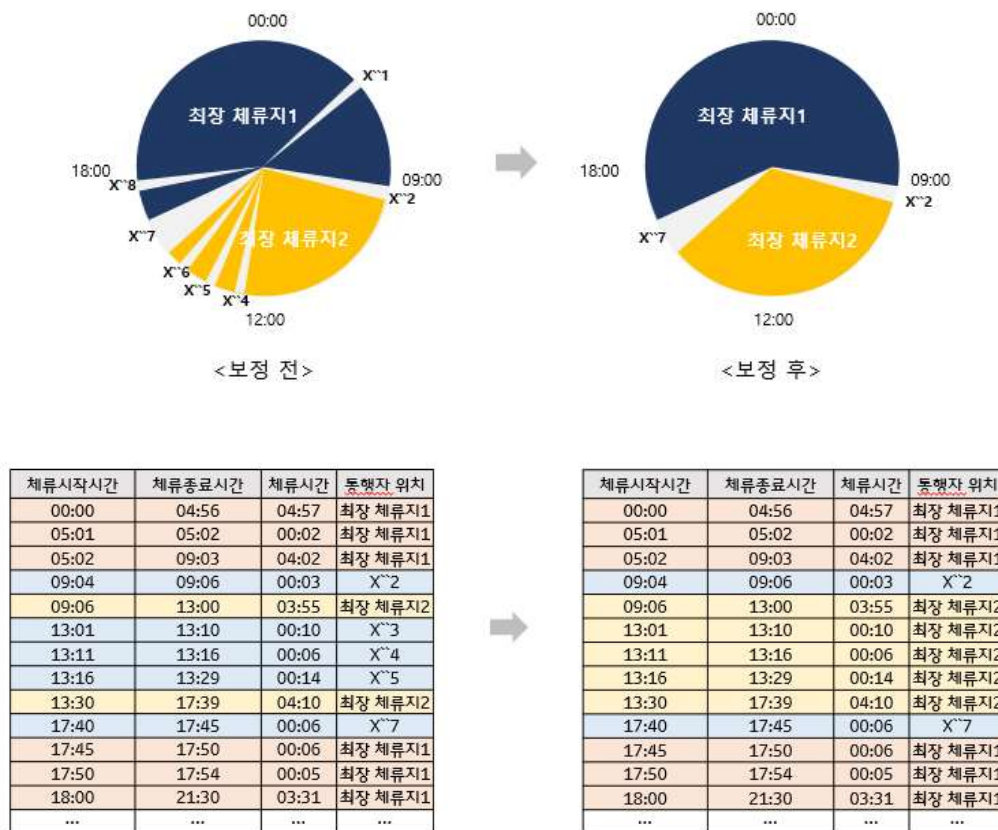
→

체류시작시간	체류종료시간	체류시간	동행자 위치
00:00	04:56	04:57	max1 X ¹
05:01	05:02	00:02	max1 X ¹
05:02	09:03	04:02	max1 X ¹
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 ¹
17:50	17:54	00:05	max1 X ¹
18:00	21:30	03:31	max1 X ¹
***	***	***	***



<그림 2- 8> 최장체류지 이외 체류지 중 최장체류지 식별

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



<그림 2- 9> 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

- 원천데이터에 대해 신규 알고리즘을 적용하면 <그림 2-10>과 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 2- 10> 통행지표 산출물 예시

나. 통행목적 및 통행시간/거리 추정

- 통행시간의 경우, 거주지와 근무지에서의 체류시작시간과 체류종료시간을 기반으로 출근 및 퇴근 시간을 계산

※ <그림 2-11>에서 출근시간은 09:03부터 09:30까지인 27분으로 계산하며, 퇴근시간은 17:39부터 18:25까지인 45분으로 계산



체류시작시간	체류종료시간	체류시간	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
...

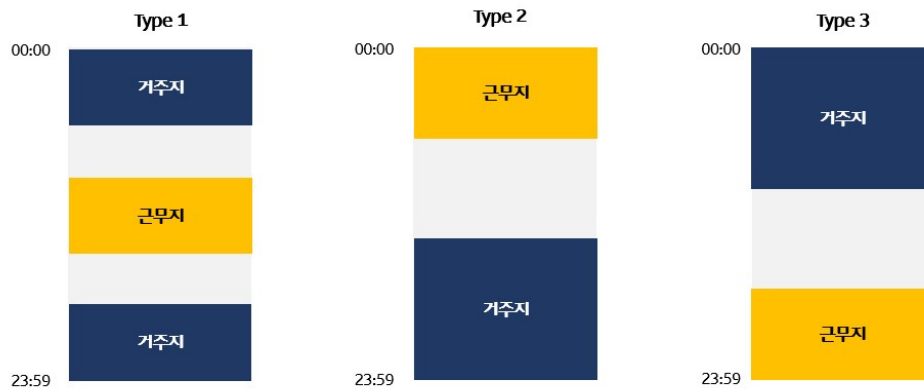
<그림 2- 11> 통행시간 계산 및 잠재체류 식별 예시

○ 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- 12> 주체류지 기준 trip type 구분

	순번	고객식별번호	기준일	체류시작시간	체류종료시간	기지국	체류시간		
거주지 체류	206	164	20160525	0010	0158	I	108	1통행	
	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		
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			
근무지 체류	220	164	20160525	0935	1323	B		228	
	221	164	20160525	1329	1343	A		14	
	222	164	20160525	1345	1949	B		364	
거주지 체류	223	164	20160525	2006	2014	L		8	1통행
	224	164	20160525	2015	2019	I		4	
	225	164	20160525	2020	2025	N		5	
	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- 13> 일별 데이터에 대한 주체류지 간 통행량 산정 예시

- 출근속도의 경우, 거주지와 근무지 간 거리를 출근시간으로 나누어 계산하고 퇴근속도의 경우, 거주지와 근무지 간 거리를 퇴근시간으로 나누어 계산
- 출·퇴근속도를 기준으로 이상치를 식별하고 출근시간 및 퇴근시간 산출 대상에서 제외
 - 출근속도 또는 퇴근속도가 평균 도보속도인 5km/h 미만인 경우, 'data lagging'등 데이터 수집 지연 현상에 의한 이상치로 판단
- ※ <그림 2-15>는 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 퇴근시간이 데이터 로그가 없던 시간인 13:28부터 14:11까지인 43분으로 과대추정될 수 있음
 - 출근속도 또는 퇴근속도가 300km/h 이상인 경우, 신호 이상에 의한 체류지 식별 오류로 판단하여 출근시간 및 퇴근시간 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- 14> data lagging 예시

제3절 이동수단 추정 알고리즘 개발

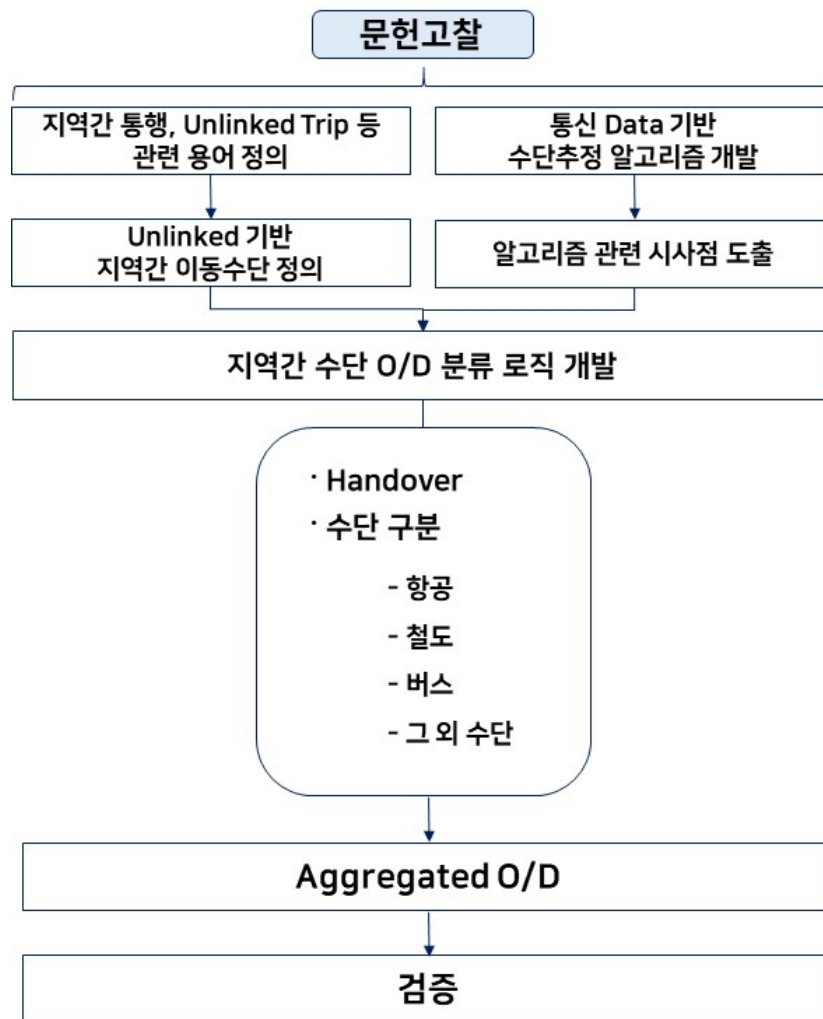
1. 기본 방향

가. 과업의 배경

- 표본조사 대상자의 개별통행 이용수단을 조사 후 전수화하고 대중교통, 전세버스, 택시 수송실적 등의 자료를 이용하여 보정하는 것이 기존 가구통행실태조사에서의 통행수단 추정 방법론
- 이에 반해 통신데이터, 대중교통카드 데이터, GPS 데이터 등으로 대두되는 ‘빅데이터’의 경우, 해당 데이터만으로 사람의 개별 통행수단을 구체적으로 파악하는 것은 불가능에 가까움
- 통행수단추정 방법론의 신뢰도를 높이기 위해, 교통시설, 교통 네트워크, 대중교통 운행정보 등 다양한 이종 데이터와의 시간적/공간적 결합을 통해 사람의 개별 통행수단을 추정하는 연구가 국내·외에서 활발히 진행 중

나. 연구의 방향

- 본 연구는 2019년 KT 기지국 원천데이터를 활용, Unlinked Trip 기반 지역간 이동수단을 유형을 정의하고 이를 추정하는 로직을 개발하는 것을 목적으로 함
- 본 연구는 다음과 같은 절차를 통해 빅데이터 기반 지역간(Unlinked Trip) 이동수단을 추정
 - － 지역간 통행, Unlinked Trip에 관련된 문헌고찰을 통해 Unlinked Trip 기반 지역간 이동의 개념과 이에 적합한 수단 유형을 정의
 - － 원천데이터 구조 탐구 및 통행수단 추정 관련 문헌고찰을 통해 DB 가공방안을 정립
 - － 정의된 교통수단(항공, 철도, 고속·시외버스, 자동차 등)의 특성을 분석하고 수단별로 구분할 수 있는 논리를 개발



<그림 2- 15> 연구의 기본 방향

2. 관련 문헌고찰

가. 지역간 통행

<표 2- 1> 지역간 통행 관련 문헌고찰

구분	저자	특성 or 정의	분류 기준
1	Makoto et al. (2009).	지역간 통행은 트립 발생 빈도가 낮고, 빈번한 서비스가 적으며 이동시간이 긴 것이 특징	트립 발생 빈도, 통행시간
2	Tokuya et al. (2016).	지역간 통행은 기본적으로 현 경계를 넘는 통행으로 정의되나, 3대 대도시권 내의 통행은 예외로, 사이타마 -> 도쿄 등의 통행은 현 경계를 넘더라도 지역내 통행으로 분류	지역 경계 통과 여부
3	Oskar et al. (2013).	지역간 통행은 100km 이상을 이동한 통행으로 정의되며, 주 목적은 여가와 사업	통행거리, 통행목적
4	한국교통연구원. (2005).	지역간 통행은 행정구역 크기와 무관하게 단순히 시, 군 행정경계를 통과하는 통행으로 정의, 비정기적이며 장거리 특성을 지님	지역 경계 통과 여부, 통행분포, 통행거리
5	국가교통 DB.	지역간 통행은 최초 출발지와 최종 목적지가 서로 다른 지역(시, 도)에 있는 통행	출발지, 목적지 정보

나. Unlinked Trip

<표 2- 2> Unlinked Trip 관련 문헌고찰

구분	저자	특성 or 정의	비 고
1	Craig A. McGeoch. (2011).	Unlinked Trip은 통행 방법과 차량 내 인원수를 변경하지 않고 한 장소에서 다른 장소로 이동하는 것	예를 들어, 열차로 목적지까지 가기 위해 집에서 A역까지 자동차로 이동할 경우, 자동차 Unlinked Trip의 목적은 '교통수단을 바꾸기 위한 것'
2	Tadahiro et al. (2010).	Unlinked Trip은 Linked Trip에 포함되는 요소 여행	Linked Trip은 목적이 있으며, Unlinked Trip은 목적 달성을 위한 중간 다리 역할

3	김형철. (2015).	Unlinked Trip은 목적통행을 분석목적에 따라 교통수단 또는 요금체계 등 동질의 통행특성으로 구분하여 각각 별개의 통행으로 취급하는 것	목적이 있는 통행만이 Unlinked Trip이 될 수 있음
4	신강원. (2016).	Unlinked Trip은 통행연계거리(연속된 통행들의 승차지점과 잠재하차지점 간 최소거리)가 2km를 넘는 통행	통행연계거리가 0.8km 이상일 때 추정 통행량에 큰 변화가 없는 것으로 밝혀짐
5	국토교통부, 한국교통연구원	Unlinked Trip은 수단통행, 즉 각각의 교통수단에 의한 통행을 별개의 통행으로 취급하는 것	-

다. 수단추정 알고리즘

<표 2- 3> 수단추정 알고리즘 관련 문헌고찰(규칙 기반 학습법)

구분	저자	추정수단	추가데이터	분류 특성	공간분석 기법	검증데이터
1	Hui et al. (2017)	지역간 수단 : 항공, 육상	-	통행 시간	-	항공 데이터
2	Calabrese et al. (2011a)	동력, 비동력	-	통행 속도	-	-
3	Smoreda et al. (2013)	지역간 수단 : 항공, 철도, 자동차	철도/도로 네트워크 공간데이터, 공항 위치데이터	네트워크 근접성, 통행속도	네트워크와의 유클리드 거리	-
4	Horn and Kern. (2015)	철도	철도역 위치데이터	철도역과의 근접성	철도역과의 유클리드 거리	-
5	Qu et al. (2015)	통근 수단 : 승용차, 대중교통, 도보	대중교통 노선 및 도로 네트워크 데이터	통행속도, 네트워크와의 근접성, 통행거리	네트워크와의 유클리드 거리	통행자의 수단 분포 통계
6	Yamada et al. (2016)	철도, 자동차	철도 및 도로 네트워크, 철도 운행시간표	네트워크 및 운행시각과의 근접성	네트워크와의 유클리드 거리	-
7	Phithakkinukoon et al. (2017)	통근 수단 : 대중교통, 승용차	구글 지도에서 얻은 경로	구글 지도 경로와의 근접성	네트워크와의 유클리드 거리	통행자의 수단 분포 통계

8	Schlaich et al. (2010)	지역간 수단 : 자동차, 트럭, 기타 수단	도로 네트워크 데이터	도로 네트워크와의 근접성, 통행 시간	특정 기지국과의 연결	-
9	Doyle et al. (2011)	지역간 수단 : 철도, 자동차	철도 및 도로 네트워크 데이터	네트워크 근접성	특정 기지국과의 연결	-
10	Wu et al. (2013)	지역간 수단 : 자동차(고속도로)	고속도로 네트워크 데이터	고속도로와의 근접성	특정 기지국과의 연결	-
11	Holleczech et al. (2015)	지하철	지하철 네트워크 데이터, 지하철 전용 기지국 목록	네트워크 근접성	특정 기지국과의 연결	수동 카운팅
12	Larijani et al. (2015)	지하철, 통근열차, 자동차	교통 네트워크 데이터, 지하철 전용 기지국 목록	네트워크 근접성	특정 기지국과의 연결	-
13	Garcia et al. (2016)	지역간 수단 : 철도, 항공, 도로(자동차 등)	교통 및 도로 네트워크 데이터	네트워크 근접성, 통행 시간	특정 기지국과의 연결	통행자의 수단 분포 통계
14	Li et al. (2017)	지하철	지하철 네트워크 데이터, 지하철역 간 통행시간	네트워크 근접성, 통행 시간	특정 기지국과의 연결	라벨링 데이터
15	Asgari. (2016)	철도(기차, 지하철, 트램), 도로	교통 및 도로 네트워크 데이터	통행 거리, 네트워크 근접성	지도 매칭	라벨링 데이터
16	Poonawala et al. (2016)	지하철	지하철 네트워크 데이터	네트워크 근접성	지도 매칭	OD 매트릭스 데이터
17	Danafar et al. (2017)	자동차, 버스, 트램, 기차, 자전거, 도보	교통 및 도로 네트워크 데이터	통행 속도, 최단 거리 근접성	지도 매칭	-

<표 2- 4> 수단추정 알고리즘 관련 문헌고찰(기계 학습법)

구분	저자	추정수단	추가데이터	분류 특성	공간분석 기법	검증데이터
1	Wang et al. (2010)	대중교통, 승용차	구글 지도로 얻은 통행시간	통행 시간	K-means	통행자의 수단 분포 통계
2	Kalatian and Shafahi. (2016)	도보, 승용차, 버스	-	통행 속도	K-means	통행자의 수단 분포 통계
3	Hui. (2017)	지역간 수단: 항공, 육상	-	통행 시간	Hierarchical agglomerative clustering	항공 데이터
4	Xu et al. (2011)	동력 수단, 자전거, 도보	-	통행속도의 평균, 표준편차	Hidden Markov Model	라벨링 데이터

라. 요약 및 시사점

- 지역간 통행 및 Unlinked Trip의 정의는 학자마다 상이하며, 구체적인 방법론을 통해 그 정의를 도출한 것이 아닌 임의로 정의한 것
- 지역간 교통수단 추정 관련 문헌 역시 지역간 통행을 별도로 정의하지 않고, 임의적으로 정의
- 통행속도에 따른 분류 기법은 교통수단을 보다 미시적으로 분류하는 데 한계를 보임
- 대부분의 수단추정 알고리즘 관련 연구에서 교통 네트워크와의 매칭을 통한 지리적 접근법이 사용됨
- 네트워크 근접성은 수단추정에 가장 많이 사용된 분류 특성으로, 보조 Data와의 결합이 필수적
- 실측 데이터의 부족으로 대부분의 연구는 그 결과를 검증하지 않거나, 간단한 집계 데이터를 검증에 사용
- 기계학습법은 통신데이터를 이용한 수단추정에 잘 사용되지 않는 경향을 보이는데, 이는 라벨링 데이터의 부족에서 기인

마. Unlinked Trip 기반 지역간 이동수단 정의

- 본 연구에서는 이동수단을 중심으로 Unlinked Trip 기반 지역간 통행을 정의
- 항공의 경우, 그 통행목적이 장거리 이동에 해당하므로 지역간 통행수단에 부합
- 철도의 경우, 항공에 비해 장거리를 이동한다 할 수 없으나 지역 경계를 넘어 타 지역으로 이동할 때 사용되는 교통수단이므로, 지역간 통행수단에 부합
- 고속/시외버스의 경우, 철도와 마찬가지로 지역 경계를 넘어 타 지역으로 이동할 때 사용되는 교통수단이므로, 지역간 통행수단에 부합
- 자동차의 경우, 고속도로 T/G를 이용한다면 지역 경계를 넘어 타 지역으로 이동한다고 볼 수 있으므로 T/G를 이용하는 경우에 한하여 지역간 통행수단에 부합
- 항공, 철도, 고속·시외버스, 자동차 수단의 공통점은 지역간 통행의 기·종점이 공항, 철도역, 버스터미널, T/G 등의 교통시설이라는 것이며, 교통시설 간의 통행을 하나의 Unlinked Trip으로 볼 수 있음
- 따라서, 본 연구에서의 Unlinked 기반 지역간 통행은 ‘항공, 철도, 고속·시외버스, 자동차(T/G 이용) 등의 수단을 이용한 통행’으로 정의할 수 있음

바. 연구의 차별성 및 기대효과

1) 연구의 차별성

- 본 연구에서는 Unlinked 기반 지역간 통행과 이에 해당하는 수단을 상호 보완적으로 새롭게 정의함
- 모바일 데이터의 공간적 분석에 특화된 ‘모바일 데이터 기반맵’ 자료와의 매칭을 통해 보다 신뢰도 높은 수단추정 로직을 개발
- 기존 모바일 데이터 기반맵 자료에 교통수단별 교통유발시설 자료를 결합하여 교통수단 분석용 기반맵 자료를 새롭게 개발

2) 기대효과

- 전수에 가까운 빅데이터를 통한 수단추정 로직 개발을 통한 교통수요 예측 신뢰도의 제고
- 기존 가구통행실태조사에 소요되는 조사비용의 절감
- 새로운 분석 기반맵 자료의 개발을 통한 통신 빅데이터의 활용성 증대

3. 교통수단 구분 알고리즘 구축 방안

가. 알고리즘의 개념 및 목적

- 교통정책지원을 위한 플랫폼인 View-T의 분석 서비스를 고도화하기 위해, 새로운 통행분석 지표를 개발해야 할 필요성 존재
- 18년부터 모바일 데이터를 교통 수요 분석 목적에 맞게 가공하여 View-T에 구축 중
- 본 연구에서 개발한 알고리즘은 새로운 통행분석 지표인 Unlinked Trip 기반 지역간 이동수단을 추정하는 목적

년월일	고객 식별번호	X좌표	Y좌표	체류 시작시간	체류 종료시간	연령	체류시간	Trip Type	Trip Trajectory	Trip Purpose	Trip Mode
2019 0401	765	128,071	36,743	1424	1425	30	2	경로	1	여가	BUS
2019 0401	765	128,071	36,743	1426	1511	30	46	체류	2	여가	-
2019 0401	765	128,041	36,752	1512	1513	30	2	경로	3	여가	BUS
2019 0401	765	128,096	36,743	1514	1517	30	4	경로	4	여가	BUS
2019 0401	765	128,101	36,735	1518	1659	30	102	체류	5	여가	-
2019 0401	1024	126,549	37,431	0537	0604	40	28	체류	1	출근	-
2019 0401	1024	126,549	37,435	0605	0619	40	15	경로	2	출근	KTX
2019 0401	1024	126,557	37,468	0620	0633	40	14	경로	3	출근	KTX
2019 0401	1024	126,558	37,489	0634	0636	40	3	경로	4	출근	KTX
2019 0401	1024	126,558	37,489	0637	0837	40	121	체류	5	출근	-

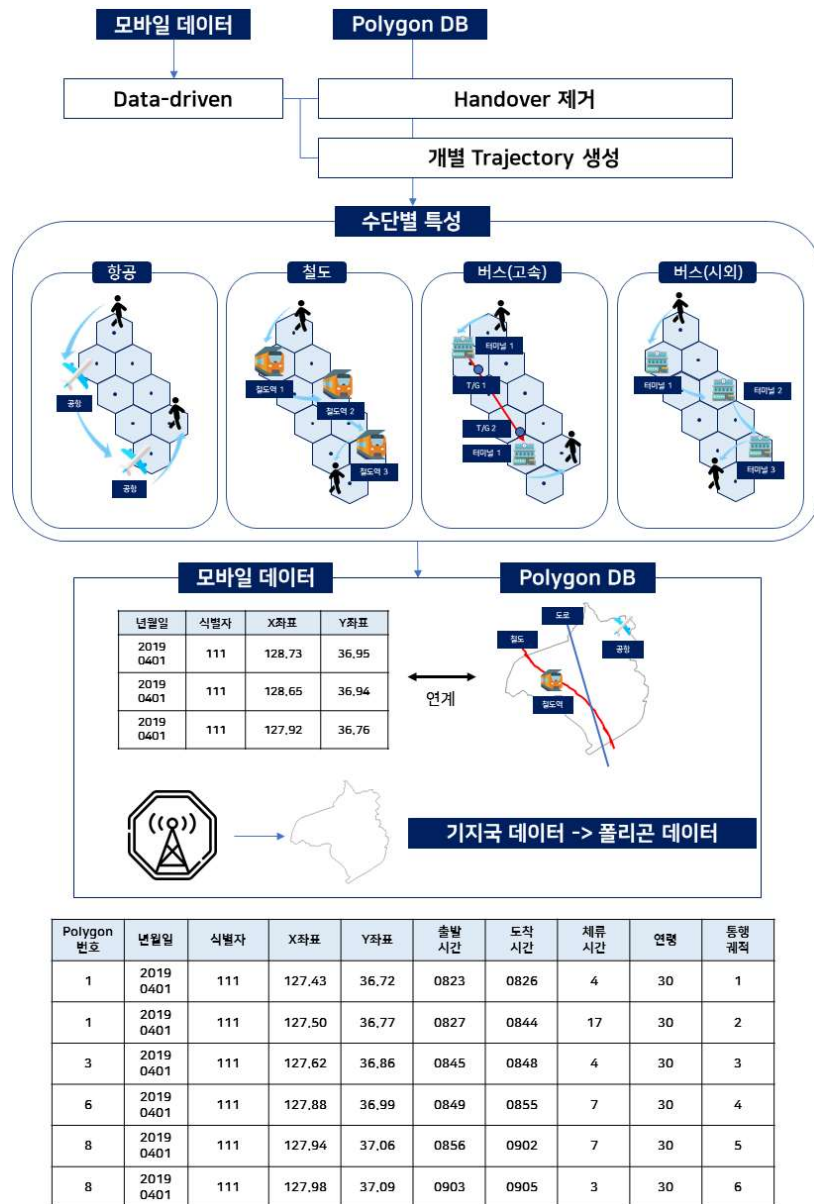
<그림 2- 16> 알고리즘 적용 DB 구축 예시

나. 교통수단 분석용 기반맵 구축

- 기존 모바일 데이터 기반맵 자료에 각 수단별 교통시설 공간정보를 매칭시켜 교통수단 분석용 모바일 데이터 기반맵을 구축
- 교통폴리곤, 공항, 철도역, 철도망, 도로망, 고속버스터미널, 시외버스터미널 데이터의 속성 정보를 결합, 해당 폴리곤 내 교통시설 존재의 유무에 따라 각 교통시설 컬럼값을 0 또는 1 이상의 정수형 값으로 표시

다. 모바일 데이터와 교통수단 분석용 기반맵 연계

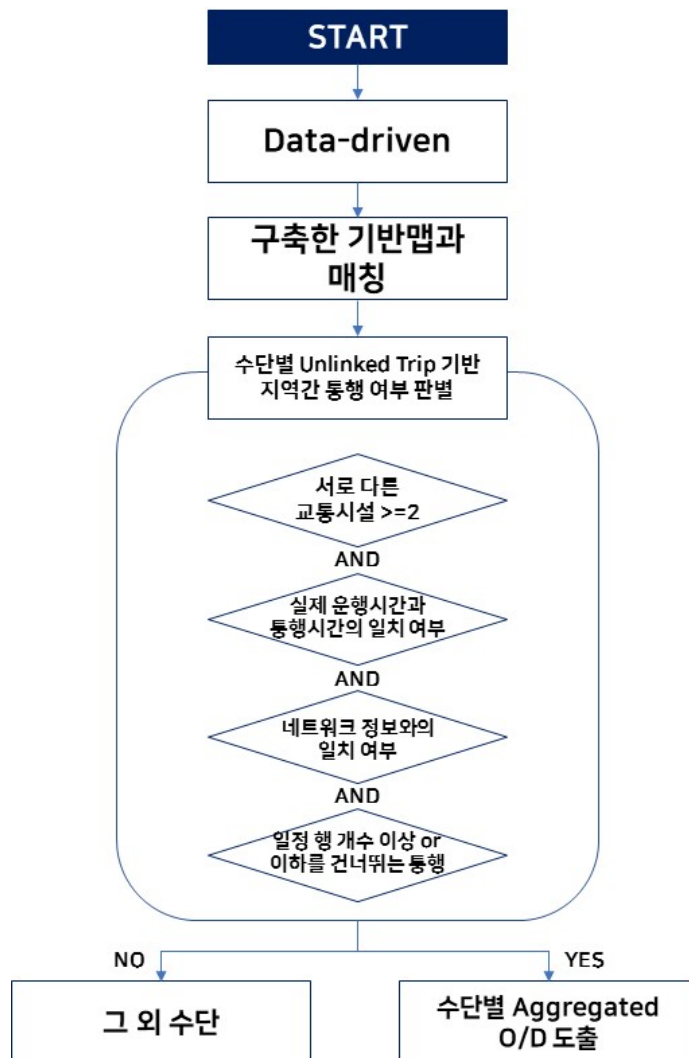
- 구축된 교통수단 분석용 기반맵과 모바일 데이터의 속성을 결합
- 결합된 데이터는 개별통행 Trajectory에 따른 교통수단별 네트워크 유무 정보를 보유하고 있음



<그림 2- 17> 모바일 데이터와 기반맵 연계 Flow Chart

라. 교통수단 구분 기본 알고리즘

- 매칭된 최종 데이터를 기반으로, 통행수단별 분류 알고리즘을 제시
- 한 사람을 기준으로 하루의 통행패턴 중, 서로 다른 이름을 가진 교통유발시설이 2개 이상 존재하고, 통행시간(도착행 도착시간 - 출발행 출발시간)과 철도망, 도로망 정보가 상당 부분 일치할 경우 이를 Unlinked Trip 기반 지역간 통행으로 간주
- 분류 순서는 항공 -> 철도 -> 고속버스 -> 시외버스 -> 그 외 수단 등 5단계로 설정



<그림 2- 18> 교통수단 구분 기본 알고리즘

4. 수단별 알고리즘 개발

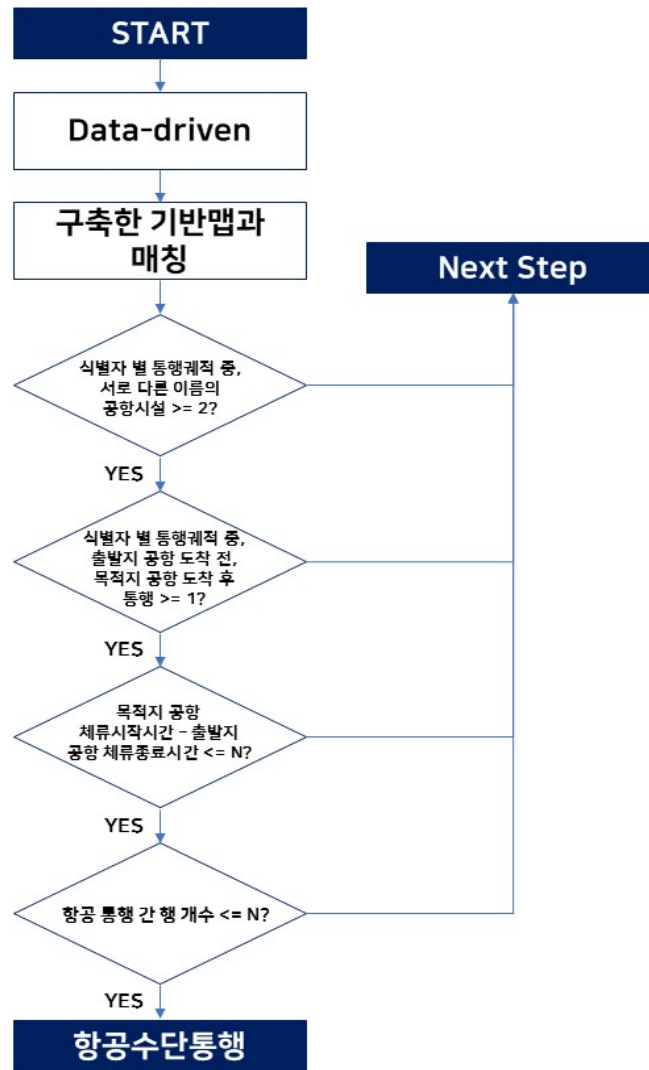
가. 항공

- 항공 수단은 DB 내에 공항 외 별도의 추정을 위한 네트워크가 존재하지 않음
- 분석 결과의 신뢰성을 높이기 위해 통행시간, 행 개수 등의 총량 제약조건을 활용
 - 개인의 하루 통행 중, 서로 다른 이름의 공항시설이 2개 이상인 통행(즉, ‘공항 이름’ 행의 고유값은 3개여야 함(EX : 청주공항, 제주공항, 0))
 - 목적지 공항 도착 행의 체류시작시간 - 출발지 공항 도착 행의 체류종료시간(통행시간)이 일정 시간 내에 이뤄진 통행
 - 출발지 공항 도착 행과 목적지 공항 도착행 사이 행 개수의 누적분포그래프를 그려 95% 지점 이상의 값은 항공수단 통행에서 제외

폴리곤 번호	년월일	식별자	X좌표	Y좌표	출발 시간	도착 시간	체류 시간	연령	통행 궤적	공항	공항 이름
459	2019 0401	111	128.7 3	36.95	1347	1348	2	30	1	0	0
664	2019 0401	111	128.6 5	36.94	1348	1410	22	30	2	1	김포
782	2019 0401	111	127.9 2	36.76	1411	1423	13	30	3	0	0
1245	2019 0401	111	127.5 3	36.55	1424	1435	12	30	4	0	0
2727	2019 0401	111	127.3 0	36.31	1436	1455	20	30	5	0	0
3994	2019 0401	111	126.8 8	36.11	1455	1503	9	30	6	0	0
6389	2019 0401	111	126.4 5	35.94	1503	1511	9	30	7	1	제주



<그림 2- 19> 항공수단 구분 지표



<그림 2- 20> 항공수단 추정 알고리즘

나. 철도(고속)

- 개인의 하루 통행 중, 서로 다른 이름의 고속철도역이 2개 이상인 통행
- 철도 수단은 기·종점 지역에 따라 여러 개의 경로로 구분되며, 경로별로 철도역이 순차적으로 존재
- 2018년 철도망 DB를 활용, 철도 노선 경로(ex : KTX 경부선, SRT 호남선 등) 별로 철도역에 정수형 Sequence를 부여
- 통행궤적에 따라 철도역 Sequence가 순차적으로 증가하거나 감소한다면, 고속철도 수단으로 정의(ex : 1 → 2 → 4 or 5 → 3 → 2 → 1)
- 경로 환승 등을 고려하여 일정 통행시간 미만의 통행만을 고속철도 수단으로 정의

폴리곤 번호	년월일	식별자	X좌표	Y좌표	출발 시간	도착 시간	체류 시간	연령	통행 궤적	철도역 Sequence	공항 이름
459	2019 0401	111	128.7 3	36.95	1347	1348	2	30	1	0	0
664	2019 0401	111	128.6 5	36.94	1348	1410	22	30	2	1	행신
782	2019 0401	111	127.9 2	36.76	1411	1423	13	30	3	0	0
1245	2019 0401	111	127.5 3	36.55	1424	1435	12	30	4	2	서울
2727	2019 0401	111	127.3 0	36.31	1436	1455	20	30	5	0	0
3994	2019 0401	111	126.8 8	36.11	1455	1503	9	30	6	0	0
6389	2019 0401	111	126.4 5	35.94	1503	1511	9	30	7	3	광명



통행시간

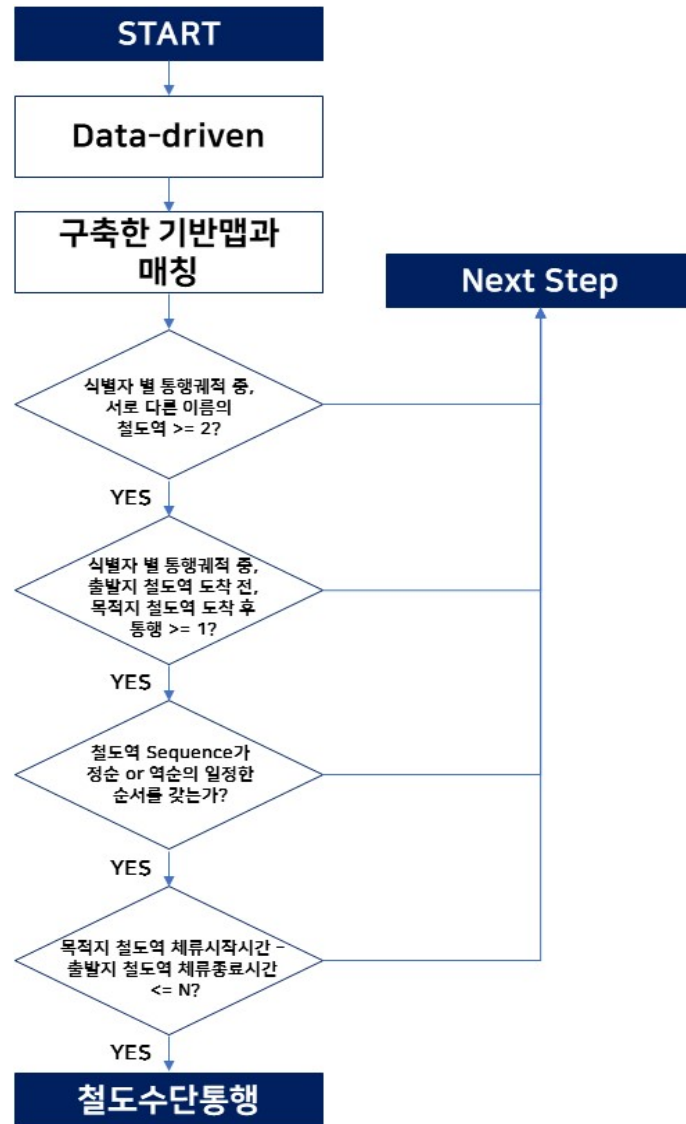


철도역 이름



철도역 Sequence

<그림 2- 21> 철도(고속)수단 구분 지표



<그림 2- 22> 철도(고속)수단 추정 알고리즘

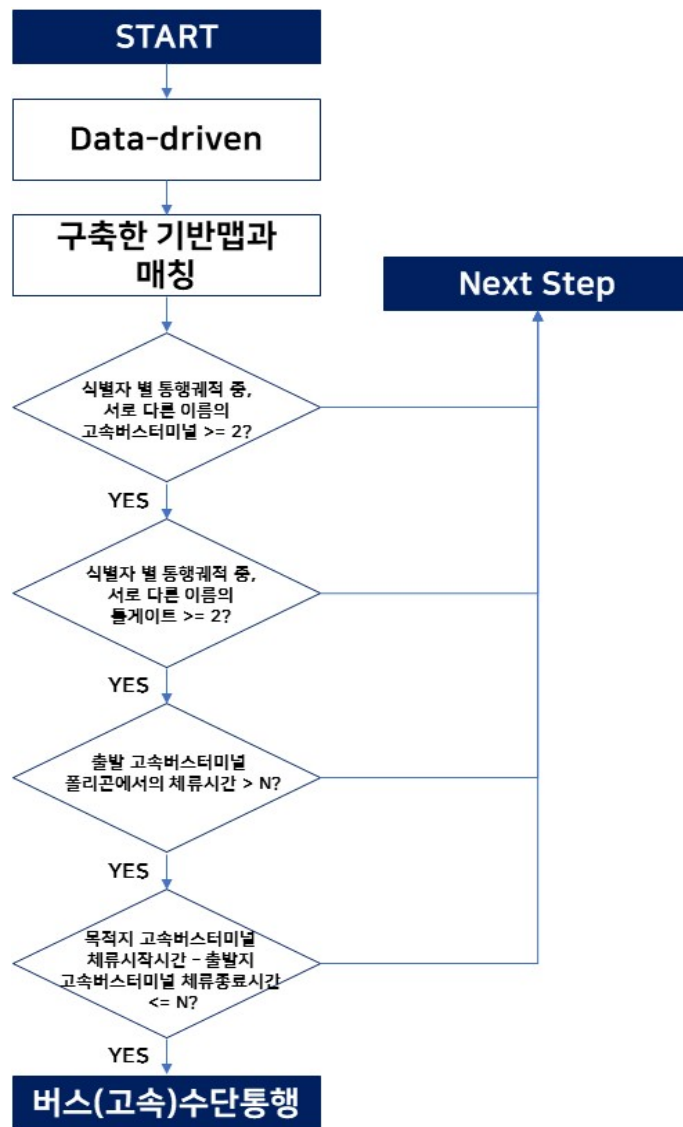
다. 버스(고속)

- 개인의 하루 통행 중, 서로 다른 이름의 고속버스터미널이 2개 이상인 통행
- 대부분의 고속버스 노선이 중간 경유지 없이 기·종점 터미널만을 연결하며, 국내선 항공과 비슷한 수단 특성을 가지고 있어 항공 수단 추정 알고리즘과 유사한 방법론으로 접근
- 고속버스는 수단 특성상 대부분의 경로가 고속도로를 통과하므로, 통행 중 기·종점에 인접한 톨게이트를 지남
- 전국 표준노드링크의 노드 정보를 활용하여 톨게이트를 DB화하며, 개별 통행 중 서로 다른 이름의 톨게이트와 고속버스터미널이 2개 이상 포함되어 있으면 고속버스 통행으로 간주
- 또한, 버스 대기시간을 고려하여 터미널 폴리곤 총 체류시간이 일정 시간 이상일 경우에만 고속버스 통행으로 간주

폴리곤 번호	년월일	식별자	X좌표	Y좌표	출발 시간	도착 시간	체류 시간	연령	통행 목적	고속버스 터미널	터미 널명	T/G	T/G명
459	2019 0401	111	128.73	36.95	1347	1348	2	30	1	1	서울	0	0
664	2019 0401	111	128.65	36.94	1348	1410	22	30	2	0	0	1	서울
782	2019 0401	111	127.92	36.76	1411	1423	13	30	3	0	0	0	0
1245	2019 0401	111	127.53	36.55	1424	1435	12	30	4	0	0	0	0
2727	2019 0401	111	127.30	36.31	1436	1455	20	30	5	0	0	0	0
3994	2019 0401	111	126.88	36.11	1455	1503	9	30	6	0	0	1	청주
6389	2019 0401	111	126.45	35.94	1503	1511	9	30	7	1	청주	0	0



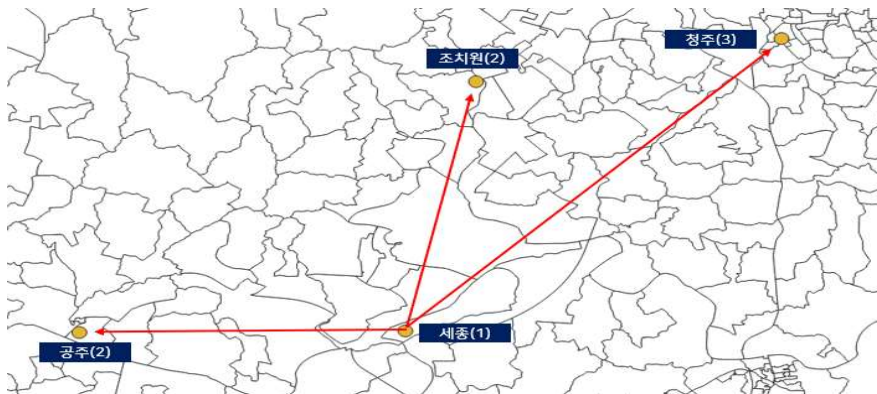
<그림 2- 23> 버스(고속)수단 구분 지표



<그림 2- 24> 버스(고속)수단 추정 알고리즘

라. 버스(시외)

- 개인의 하루 통행 중, 서로 다른 이름의 시외버스터미널이 2개 이상인 통행
- 시외버스는 고속버스와 달리 기·종점 간에 여러 정류장들을 중간 경유지로 두고 있어, 철도수단과 같이 경로별 터미널 Sequence를 부여할 수 있음
- 시외버스는 철도망 DB와 같은 구축된 터미널 Sequence 자료가 존재하지 않으므로, 터미널 간 유클리드 거리의 위계를 Sequence 부여 기준으로 선정



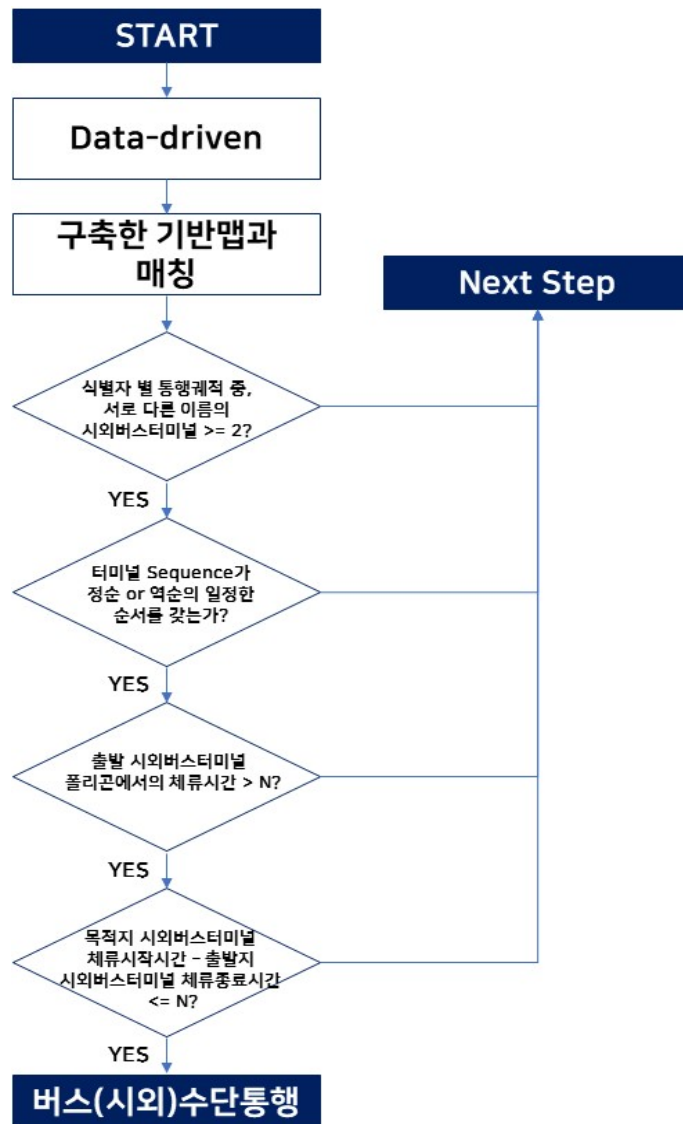
<그림 2- 25> 유클리드 거리에 의한 방향별 터미널 Sequence 설정

- 뿐만 아니라, 시외버스 경로 중에서는 국도를 경유하는 노선이 일부 존재하므로, 톨게이트를 추정 지표로 사용 불가
- 따라서, 시외버스 수단 추정 방법론은 철도 수단 추정 방법론과 유사하게 접근

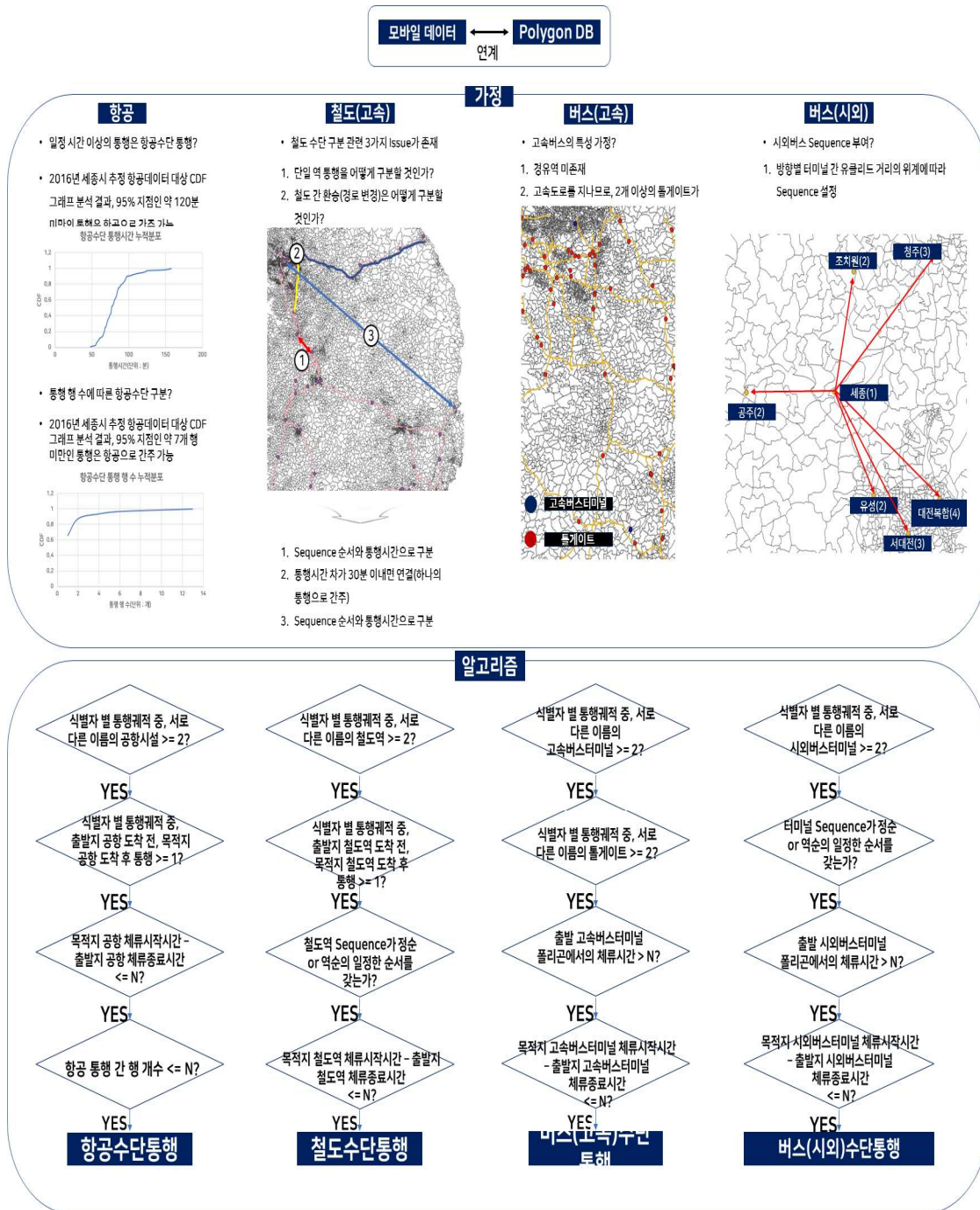
플리곤 번호	년월일	식별자	X좌표	Y좌표	출발 시간	도착 시간	체류 시간	연령	통행 궤적	터미널 Sequence	터미 널명
459	2019 0401	111	128.73	36.95	1347	1348	2	30	1	1	세종
664	2019 0401	111	128.65	36.94	1348	1410	22	30	2	0	0
782	2019 0401	111	127.92	36.76	1411	1423	13	30	3	0	0
1245	2019 0401	111	127.53	36.55	1424	1435	12	30	4	2	초치 원
2727	2019 0401	111	127.30	36.31	1436	1455	20	30	5	0	0
3994	2019 0401	111	126.88	36.11	1455	1503	9	30	6	0	0
6389	2019 0401	111	126.45	35.94	1503	1511	9	30	7	3	청주



<그림 2- 26> 버스(시외)수단 구분 지표



<그림 2- 27> 버스(시외)수단 추정 알고리즘



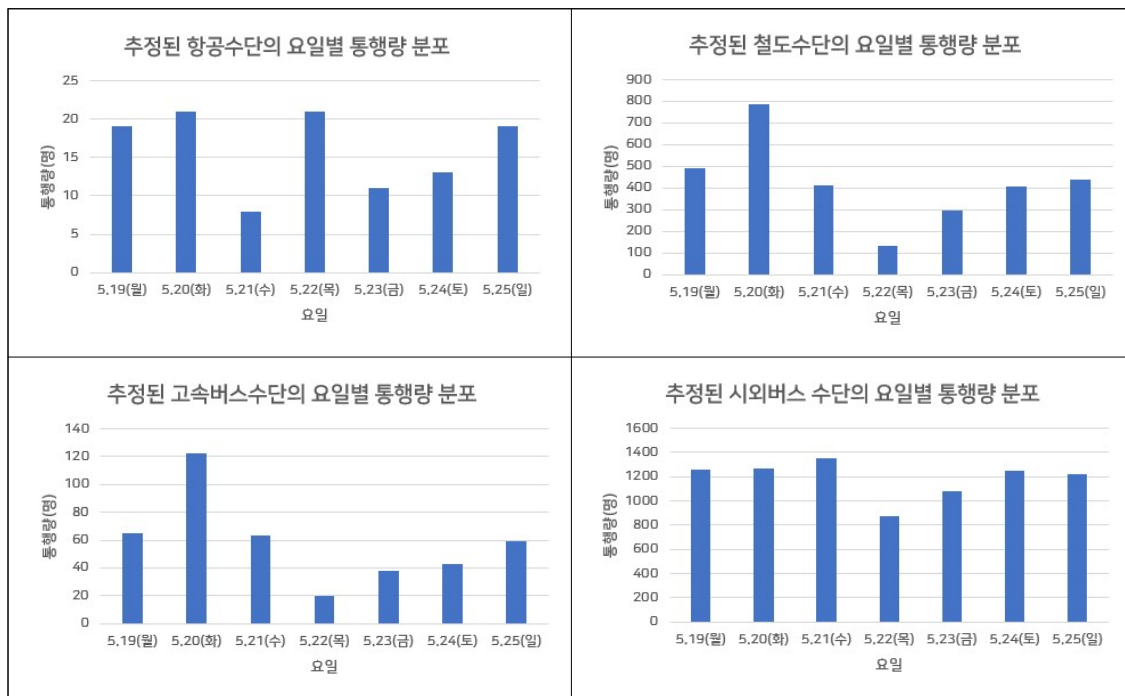
<그림 2- 28> 수단추정 알고리즘 종합

5. 개발 알고리즘 기반 수단별 Aggregated O/D 산출

- 2016년 5월 19일~25일 일주일치 세종시 모바일 데이터를 대상으로 본 연구의 알고리즘을 적용하여 수단별 Aggregated O/D를 산출

<표 2- 5> 2016년 5월 일주일 간 세종시 수단별 Aggregated O/D

구분	항공	고속철도	고속버스	시외버스
5/19(월)	19	493	65	1,259
5/20(화)	21	785	122	1,264
5/21(수)	8	414	63	1,355
5/22(목)	21	134	20	870
5/23(금)	11	295	38	1,078
5/24(토)	13	408	43	1,247
5/25(일)	19	438	59	1,218
총합	112	2,967	410	8,291



<그림 2- 29> 2016년 5월 일주일 간 세종시 일별 수단별 통행량 분포

제3장 모바일통신 빅데이터 기반 통행 DB 구축 및 검증

제1절 기준년도 통행DB 구축

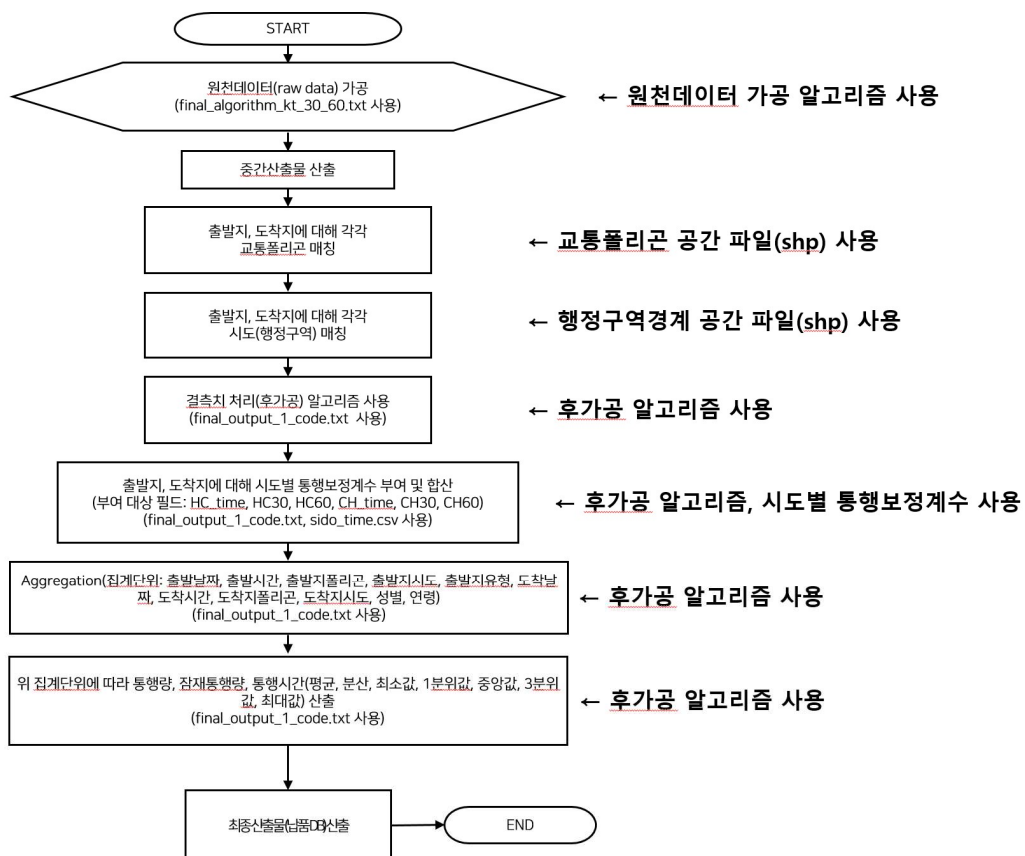
제2절 구축 DB 검증

제3장 모바일통신 빅데이터 기반 통행 DB 구축 및 검증

제1절 기준년도 통행DB 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



<그림 3- 1> 신규 모바일통신 빅데이터 가공 DB 구축 과정

<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	120분이 상	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	120분이 상	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

- 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 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
- * 「2019년 국가교통조사」에서 구분하고 있는 체류지 유형 중에서 집 이외 심야시간대 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 평균통행시간 값을 입력

– 5가지 유형으로 구분된 기종점 기준, 통행 도중 인식한 기지국에서의 체류시간에 대해 고려 안함, 60분으로 나누어 통행량 및 통행시간을 산출함

- 고려 안함 : 통행 도중 인식한 기지국 체류시간을 고려하지 않고, 해당 통행에 대해 출발 시점과 도착 시점의 차이로 통행시간을 계산함
- 60분 기준 : 통행 도중 인식한 모든 기지국 체류시간이 60분 미만인 경우만 통행으로 인정하고, 출발 시점과 도착 시점의 차이로 통행시간을 계산함

<표 3- 2> 모바일 DB 테이블 정의서

컬럼 ID	컬럼명	Type	비고
o_base_ymd	출발 일자	string	yyyymmdd
o_base_dow	출발 요일	string	0 : 일요일 1 : 월요일 2 : 화요일 3 : 수요일 4 : 목요일 5 : 금요일 6 : 토요일
o_timezn_cd	출발 시간대	string	00~23 (1시간 단위)
o_polygon	출발 폴리곤ID	string	-
o_sido	출발 시도ID	string	-
o_trip_type	출발 체류지 유형	string	H: 집 C: 회사 S: 학교 R: 정기 잠재체류지 X: 기타 잠재체류지
d_base_ymd	도착 일자	string	yyyymmdd
d_timezn_cd	도착 시간대	string	00~23 (1시간 단위)
d_polygon	도착 폴리곤ID	string	-
d_sido	도착 시도ID	string	-
d_trip_type	도착 체류지 유형	string	H: 집 C: 회사 S: 학교 R: 정기 잠재체류지 X: 기타 잠재체류지 -: 이동정보 없음
age_itg_cd	연령	integer	0~110 (예: 10: 10세 이상 20세 미만)
sex_type_itg_cd	성	string	M: 남성, F: 여성
total	통행량	integer	보정계수 적용 후 일의 자리에서 올림 처리
potential_travle	잠재통행량	integer	기종점 기준 통행 도중, 60분 이상 체류한 지점으로의 통행
travel_time	통행시간	integer	1분 단위

2. DB 구축 방법

1) 원천데이터 수집 및 이상치 제거

- 2019년 1월 1일부터 2020년 12월 31일까지 기록된 로그를 수집
- 기지국 좌표가 Null값이거나 행정동 정보가 매칭되지 않는 데이터를 제거

2) 선분 이력으로 변환

- 포인트 단위(밀리세컨 단위) 로그를 선분 단위(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: 박미율·주은정 (2017), LTE 시그널 정보를 이용한 위치정보가공 및 유동인구집계 방법, 한국통신학회 2017년도 동계종합학술발표회, p. 285.

<그림 3- 2> 선분 이력으로의 변환

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 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

4) 체류순서 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

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- 3> 체류순서 부여 (개인별 이동궤적 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 컬럼 (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	0003	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- 4> 이동정보와 체류정보 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		
주상주 지역	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
잠재 활동 지역	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 컬럼 (기준일- 고객식별번호)	가상기지국		체류 시작 시간	체류 종료 시간	연 령 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	126.916	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	회사
453	1	20160520-1843030	126.934	37.556	0805	1041	25	156	체류	집
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	기타

<그림 3- 5> 체류지 유형 구분 (예시)

7) 위치정보, 시간정보, 연령정보 변환

- 개인의 이동궤적을 추적할 수 없도록 원천데이터에서 출발지와 도착지의 위치 정보를 분석 맵의 교통폴리곤 ID로 변경한 후, 시간 정보(체류시작시간, 체류종료시간)에서 분 정보를 생략한 시간대 필드를 생성하고, 1세 단위의 연령 정보를 10세 단위로 변경함

ID	그룹 번호	KEY 컬럼 (기준일- 고객식별번호)	가상거주국		체류 시작 시간	체류 종료 시간	연 령 대	체류 시간	트립 타입	체류지 유형	위치 정보 변환 (교통폴리곤 ID)	체류시작 시간대	체류종료 시간대	연령대 변환
			x좌표	y좌표										
1	1	20160520-17232311	126.914	37.545	1400	1750	33	240	체류	집	210001	14	17	30
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	회사	230017	19	23	30
453	1	20160520-1843030	126.934	37.556	0805	1041	25	156	체류	집	311014	08	10	20
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	기타	312123	11	12	20

<그림 3- 6> 위치정보, 시간정보, 연령정보 변환 (예시)

8) 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 보고 ‘무통행’으로 간주함

ID	그룹 번호	KEY 컬럼 (기준일- 고객식별번호)	가상기지국		체류 시작 시간	체류 종료 시간	연 령 대	체류 시간	트립 타입	체류지 유형	위치 정보 변환 (교통플라곤 ID)	체류시작 시간대	체류종료 시간대	연령대 변환	
			X좌표	Y좌표											
1	1	20160520-17232311	126.914	37.545	1400	1750	33	240	체류	집	210001	14	17	30	출발
2	2	20160520-17232311	126.914	37.549	1751	1752	33	1	경로	-	-	-	-	-	1통행 ↓
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	회사	230017	19	23	30	도착
453	1	20160520-1843030	126.934	37.556	0805	1041	25	156	체류	집	311014	08	10	20	출발
454	2	20160520-1843030	126.936	37.556	1041	1042	25	1	경로	-	-	-	-	-	1통행 ↓
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	-	-	-	-	-	도착
460	8	20160520-1843030	126.914	37.545	1100	1200	25	60	체류	기타	312123	11	12	20	

<그림 3- 7> 출발/도착 구분 및 통행량 집계 (예시)

9) 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, 추정 값을 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

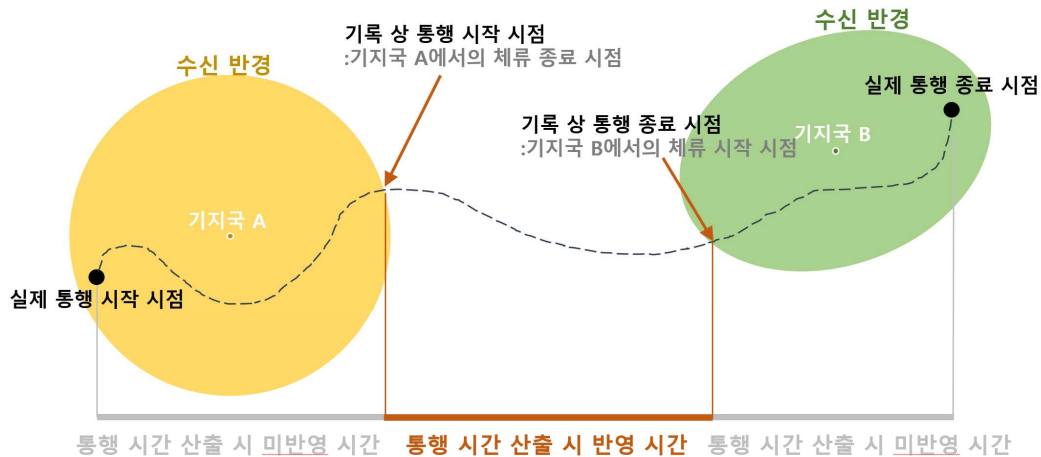
<표 3- 4> KT 시장점유율

구분		가입자 수	비율
LTE	SKT	24,205,134	43.9%
	KT	15,259,806	27.7%
	LGU+	12,647,000	22.9%
	MVNO	3,021,741	5.5%
	소계	55,133,681	100.0%

자료 : 과학기술정보통신부 “무선 통신서비스 통계 현황 (2018년 12월)”,
[https://www.msit.go.kr/\(20.03.25\)](https://www.msit.go.kr/(20.03.25)).

10) 평균통행시간 산출

- 모바일통신 데이터는 기지국 기반 로그 기록이므로 통행시간 산출 시, 기종점 부분에서 통행시간이 하향계산되는 경우가 발생함



<그림 3- 8> 기지국 기반 데이터의 통행 시간 산출 방법

- 위와 같은 문제를 해결하고자, 기지국의 평균 수신 범위를 계산하여 시군별 기지국 내 이동 시간 산출 후, 통행 시간을 보정함
 - 시군구별 면적과 기지국 수를 사용하여 시군구별 기지국 수의 밀도 계산
 - 시군구별 기지국의 평균 면적을 계산
 - 기지국의 수신 범위가 원이라고 가정하고 원주율(3.14)을 통해 기지국 수신 범위의 평균 반경 계산
 - 시군구 내 통행의 평균 통행 속도를 구한 뒤, 해당 속도로 시군구별 기지국 수신 범위의 평균 반경을 통행하는 시간 산출
- 각 개인의 통행정보에서 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
- 개별 통행에 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 구축 DB 검증

1. DB 검증 방법

- 모바일통신 데이터를 이용하여 구축된 통행DB의 신뢰성 및 정확성을 검증하기 위하여 기존 가구통행실태조사자료(2016년 기준, 한국교통연구원)와 인구총조사자료(2015년 기준, 통계청)를 활용
- 모바일기지국 데이터 기반 통행DB와 기존 가구통행실태조사자료와 인구총조사자료의 통행량, 통행시간 등을 비교·분석하고, 단순 비교할 수 없는 모바일통신 데이터 기반 상세지표(예를 들어, 상세 시간대별, 요일별, 성별, 연령별, 희망통근시간 등)는 통상적인 기대 통행패턴을 벗어나는지 확인
 - 지역별 통행 발생량·도착량을 가구통행실태조사자료와 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 가구통행실태조사자료, 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) 가구통행실태조사자료 기준 지역별 통행량 비교 결과

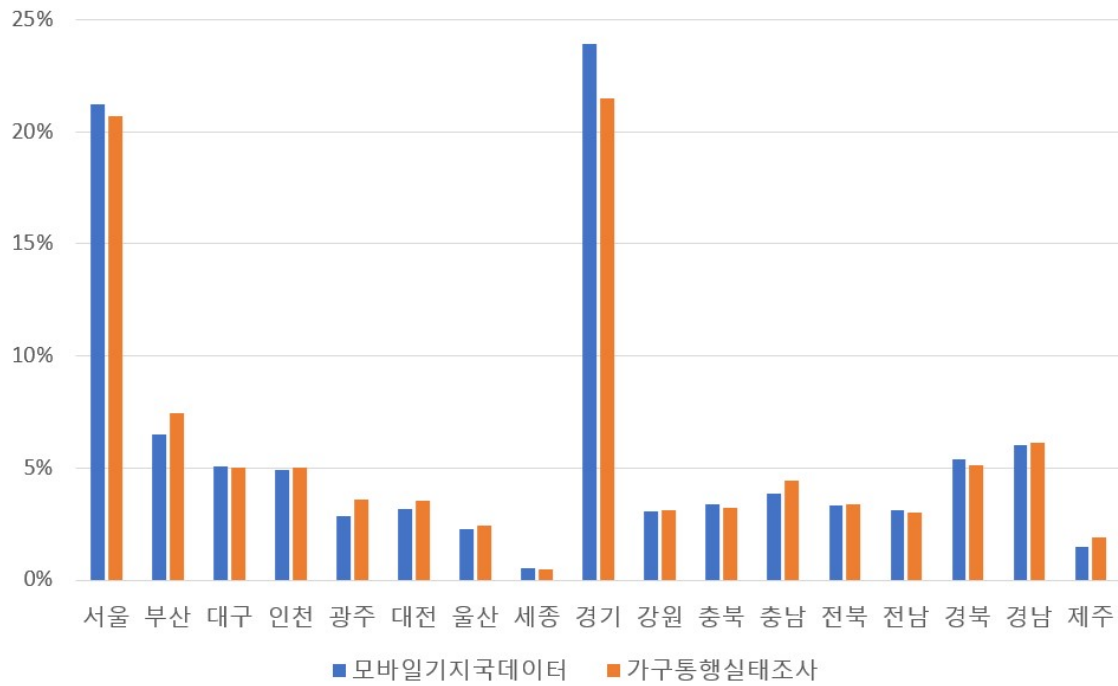
- 지역별 통행 발생량·도착량을 가구통행실태조사자료와 상대적으로 비교하여, 유사한 패턴을 보이는지 확인

<표 3- 5> 지역별 통행량 비교 결과

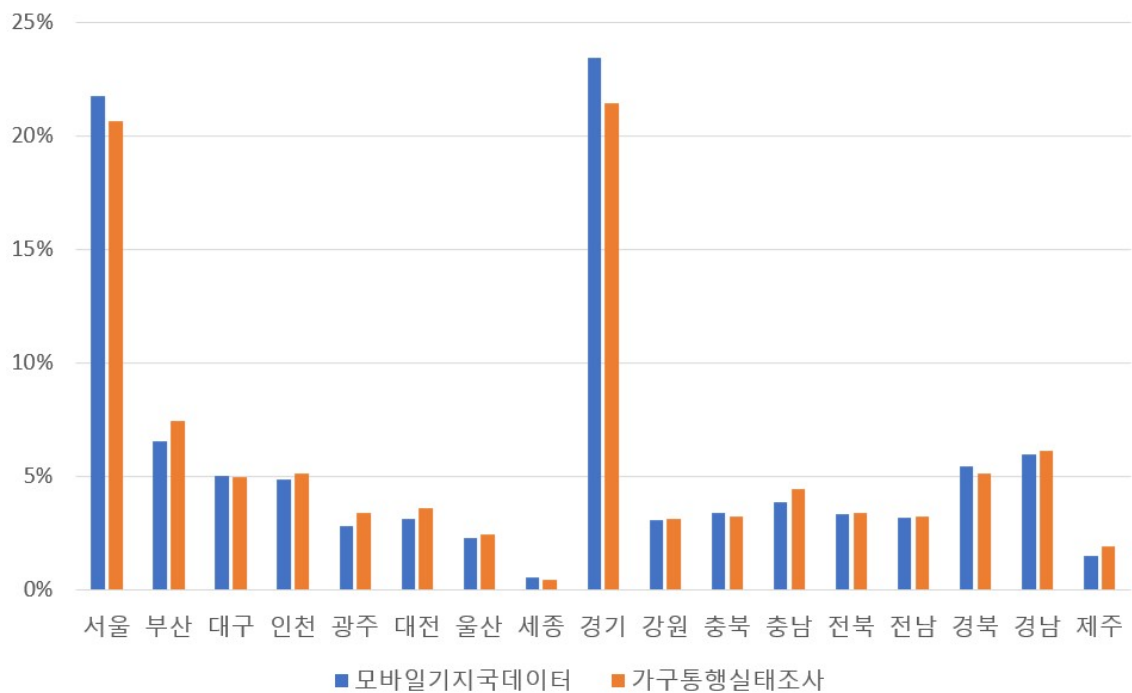
(단위: 통행/일(비율%))

구분	가구통행실태조사자료 기준*		모바일기지국데이터 기준	
	발생 (비율)	도착 (비율)	발생 (비율)	도착 (비율)
서울	17,996,813 (20.7)	17,971,878 (20.7)	12,578 (21.2)	12,891 (21.7)
부산	6,457,538 (7.4)	6,458,293 (7.4)	3,847 (6.5)	3,872 (6.5)
대구	4,333,058 (5.0)	4,330,603 (5.0)	2,987 (5.0)	2,964 (5.0)
인천	4,370,318 (5.0)	4,433,014 (5.1)	2,913 (5.0)	2,879 (4.8)
광주	3,117,712 (3.6)	2,942,575 (3.4)	1,673 (2.8)	1,644 (2.8)
대전	3,091,862 (3.6)	3,101,487 (3.6)	1,876 (3.2)	1,855 (3.1)
울산	2,117,733 (2.4)	2,109,519 (2.4)	1,341 (2.3)	1,343 (2.2)
세종	389,864 (0.4)	391,571 (0.5)	325 (0.5)	325 (0.5)
경기	18,668,530 (21.5)	18,628,440 (21.4)	14,155 (24.0)	13,877 (23.4)
강원	2,714,015 (3.1)	2,715,277 (3.1)	1,820 (3.1)	1,812 (3.1)
충북	2,803,473 (3.2)	2,793,476 (3.2)	1,987 (3.4)	1,996 (3.3)
충남	3,871,861 (4.5)	3,863,535 (4.4)	2,270 (3.8)	2,293 (3.9)
전북	2,942,438 (3.4)	2,943,452 (3.4)	1,975 (3.3)	1,973 (3.3)
전남	2,606,282 (3.0)	2,783,178 (3.2)	1,832 (3.1)	1,860 (3.1)
경북	4,431,201 (5.1)	4,442,218 (5.1)	3,202 (5.4)	3,225 (5.4)
경남	5,297,159 (6.1)	5,300,768 (6.1)	3,556 (6.0)	3,529 (6.0)
제주	1,659,446 (1.9)	1,660,019 (1.9)	862 (1.5)	861 (1.5)
계	86,869,303 (100)	86,869,303 (100)	59,199 (100)	59,199 (100)

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



<그림 3- 9> 지역별 통행 발생량 비교 결과



<그림 3- 10> 지역별 통행 도착량 비교 결과

- 가구통행실태조사자료 기준 1인당 일평균 통행량(발생+도착)은 약 3.4회이며, 모바일기지국 데이터 기준 1인당 일평균 통행량(발생+도착)은 약 3.1회 임
 - 가구통행실태조사자료 기준 전국 통행 발생·도착량은 1.74통행/일, 1.74통행/일이며, 모바일기지국데이터 기준 전국 통행 발생·도착량은 1.5통행/일, 1.6통행/일 임
 - 모바일기지국데이터 기준 통행량과 가구통행실태조사자료 기준 통행량 매우 유사하게 나타남
 - 하지만 가구통행실태조사자료가 더 높은 통행량을 띄는데, 이러한 차이는 기존의 가구통행실태조사는 출근, 등교, 업무, 쇼핑, 귀가 등 모든 통행목적에 따라 통행량을 집계하지만, 모바일기지국데이터는 체류시간 기준으로 체류지를 구분하기 때문에 기타잠재체류지에 속하지 않은 이동이 집계되지 않기 때문에 통행량차이가 과소추정되었다고 판단됨
- 지역별 통행량 비율을 살펴보면, 평균적으로는 약 0.41%의 차이로 기존 가구통행실태조사자료와 매우 유사한 패턴을 나타냄
 - 경기지역 통행 발생량이 약 +2.2%p로 가장 큰 폭으로 과대추정되었으며, 부산지역 통행 발생량이 약 -1.1%p로 가장 큰 폭으로 과소추정되었으나, 평균적으로 약 $\pm 1\%$ 로 매우 신뢰성 높은 결과를 도출함

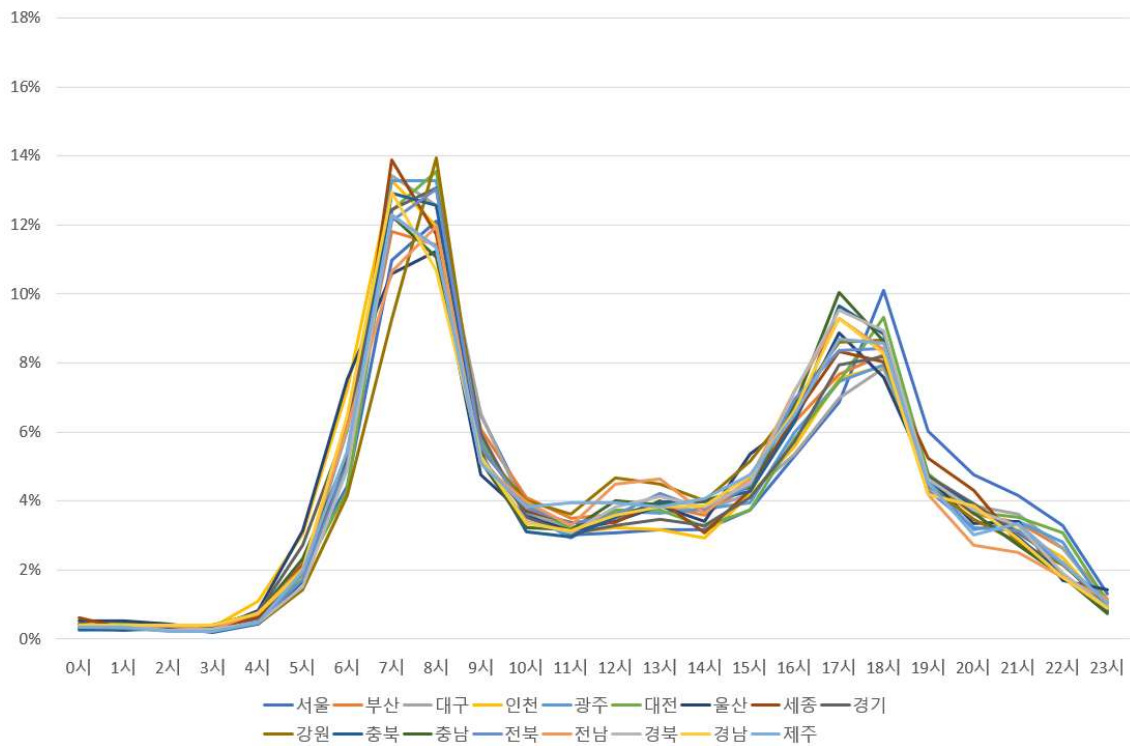
4) 시간대별 통행량 패턴 분석 결과

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

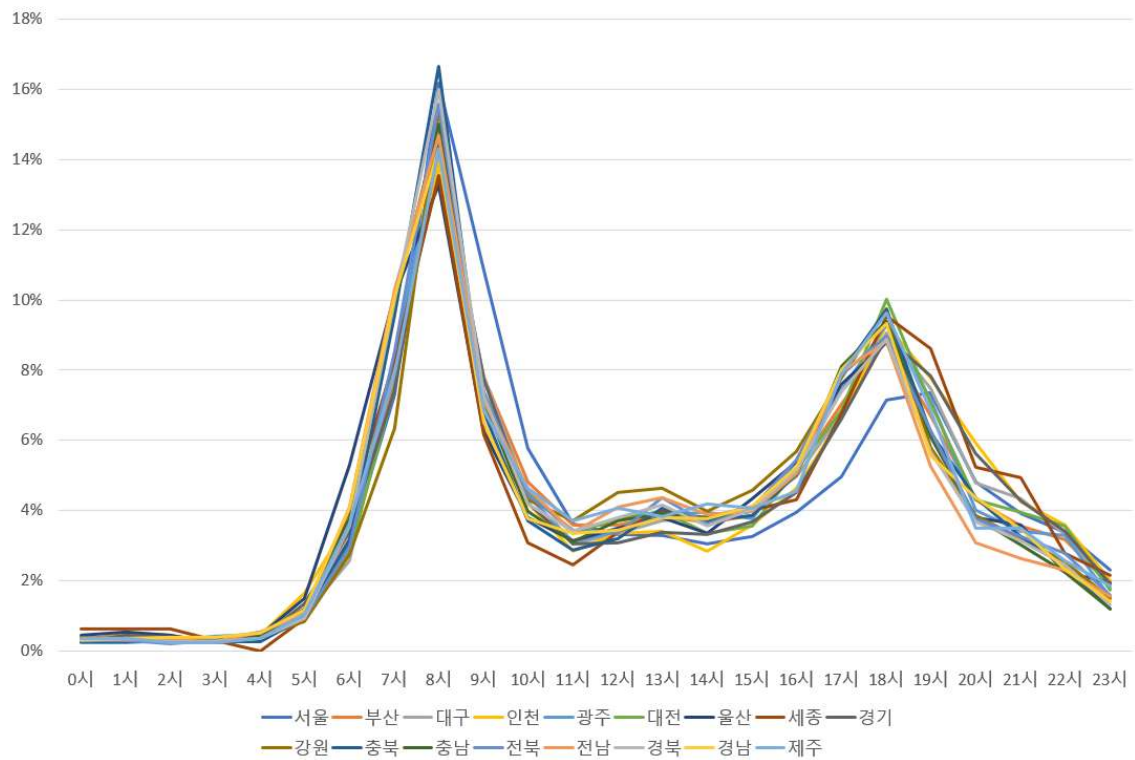
<표 3- 6> 시간대별 통행량 패턴 분석 (모바일기지국데이터 기준)

(단위: 통행/일)

구분		새벽		오전		오후		저녁	
		0시~	3시~	6시~	9시~	12시~	15시~	18시~	21시~
발 생	서울	128	390	3,460	1,680	1,186	2,003	2,628	1,102
	부산	47	128	1,100	528	427	711	632	276
	대구	30	90	936	400	332	495	486	218
	인천	36	131	946	363	271	504	474	187
	광주	14	47	545	207	187	291	263	119
	대전	22	50	568	239	202	319	333	144
	울산	20	57	393	154	144	276	209	88
	세종	4	9	99	40	35	63	57	18
	경기	159	541	4,485	1,773	1,425	2,516	2,386	871
	강원	19	39	499	238	240	375	302	109
	충북	17	46	609	228	228	402	345	113
	충남	22	78	649	262	270	481	386	120
	전북	18	49	597	251	228	390	315	126
	전남	20	58	526	226	234	387	281	99
	경북	32	73	914	377	378	679	555	194
	경남	42	110	1,070	419	401	734	580	199
	제주	8	23	250	111	102	173	138	57
도 착	서울	114	213	3,413	2,607	1,243	1,568	2,494	1,238
	부산	45	74	973	627	429	620	769	334
	대구	28	55	728	454	322	468	631	279
	인천	34	68	743	386	277	427	662	283
	광주	13	33	429	241	185	273	337	134
	대전	18	29	461	283	207	289	397	171
	울산	19	30	384	177	143	232	253	104
	세종	6	4	84	38	36	50	76	32
	경기	153	282	3,513	2,083	1,358	2,050	3,107	1,332
	강원	18	26	445	276	238	327	352	131
	충북	16	30	586	268	219	335	405	137
	충남	21	46	657	326	262	392	443	146
	전북	17	31	533	301	223	342	375	151
	전남	18	35	541	281	229	315	320	121
	경북	29	50	953	480	371	534	587	221
	경남	38	72	1,004	482	387	614	679	253
	제주	8	13	222	132	104	142	172	67



<그림 3- 11> 시간대별 통행 발생량 패턴 분석 (모바일기지국데이터 기준)



<그림 3- 12> 시간대별 통행 도착량 패턴 분석 (모바일기지국데이터 기준)

- 모바일기지국데이터 기준 시간대별 통행패턴 분석결과, 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
 - － 퇴근시간보다 출근시간에 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

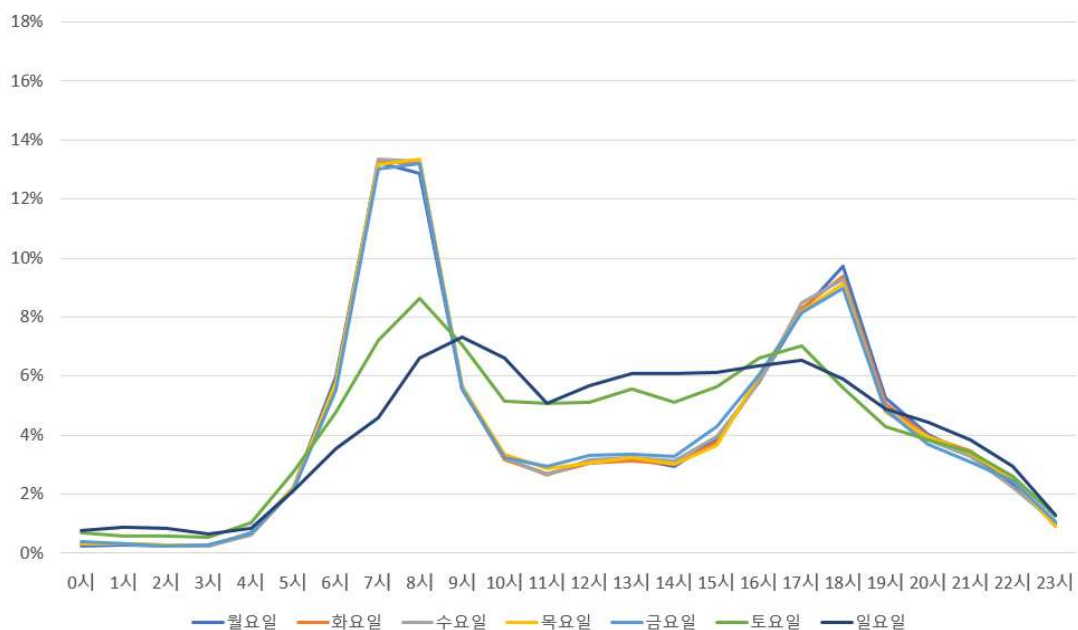
5) 요일별 시간대에 따른 통행량 패턴 분석 결과

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

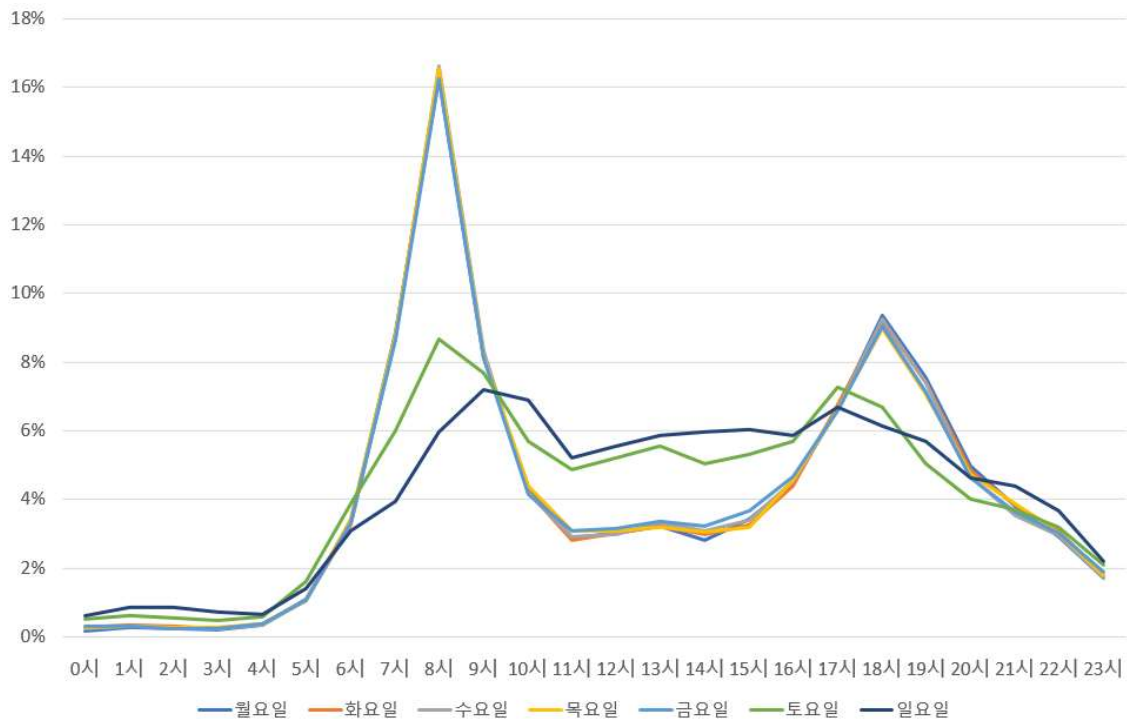
<표 3- 7> 요일별 통행량 패턴 분석 (모바일기지국데이터 기준)

(단위: 통행/일)

구분		새벽		오전		오후		저녁	
		0시~	3시~	6시~	9시~	12시~	15시~	18시~	21시~
발생 시간 기준	월	74	308	3,173	1,132	906	1,767	1,879	654
	화	93	305	3,225	1,144	924	1,795	1,844	681
	수	86	305	3,222	1,156	951	1,818	1,796	651
	목	88	312	3,192	1,172	920	1,765	1,762	671
	금	90	296	3,065	1,133	958	1,779	1,687	638
	토	102	239	1,144	957	876	1,069	760	405
	일	106	154	624	803	754	804	645	340
도착 시간 기준	월	95	149	1,029	1,014	879	1,015	872	499
	화	99	115	549	814	736	786	696	434
	수	68	162	2,816	1,520	896	1,435	2,161	835
	목	87	168	2,854	1,527	926	1,440	2,142	867
	금	80	164	2,862	1,541	938	1,449	2,131	821
	토	82	170	2,839	1,543	918	1,422	2,051	856
	일	84	162	2,719	1,480	938	1,435	2,004	823



<그림 3- 13> 요일별 시간대에 따른 통행량 변화 (발생 기준)



<그림 3- 14> 요일별 시간대에 따른 통행량 변화 (도착 기준)

- 모바일기^{지국}데이터 기준 요일별 통행패턴 분석결과, 평일과 주말 기준 전형적인 통행 패턴을 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 보다 완만한 낮 시간대 통행패턴을 보임

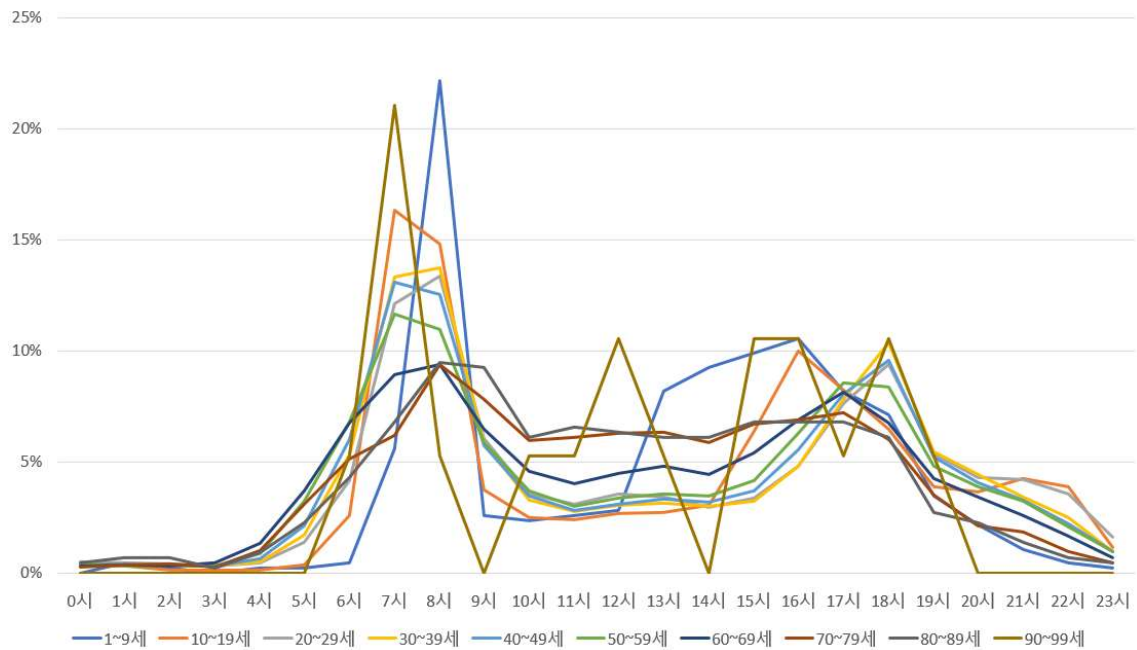
6) 연령별 통행량 패턴 분석 결과

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

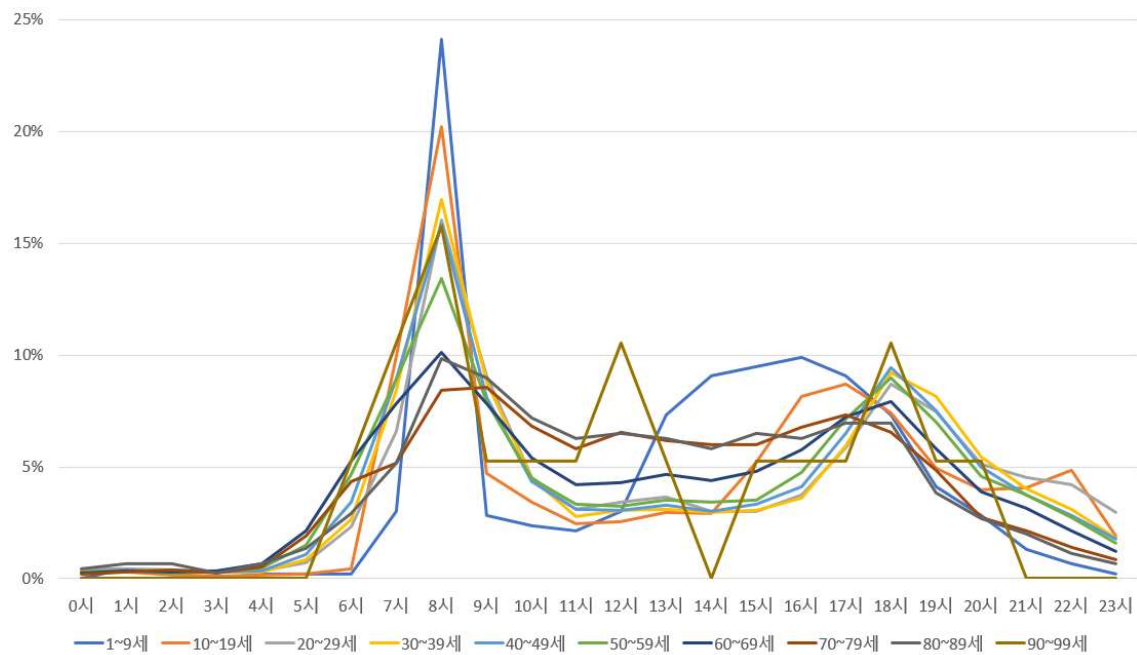
<표 3- 8> 연령별 통행량 패턴 분석 (모바일기지국데이터 기준)

(단위: 통행/일)

구분		새벽		오전		오후		저녁	
		0시~	3시~	6시~	9시~	12시~	15시~	18시~	21시~
발생	0대	3	2	131	34	94	133	59	8
	10대	31	26	1,374	352	344	1,002	571	377
	20대	138	219	2,991	1,275	997	1,594	1,919	949
	30대	119	266	3,455	1,270	982	1,690	2,164	732
	40대	127	363	3,775	1,429	1,146	2,060	2,248	773
	50대	120	549	3,631	1,551	1,281	2,343	2,106	775
	60대	69	375	1,718	1,034	937	1,403	988	340
	70대	23	101	473	454	424	475	265	75
	80대	8	16	91	97	82	91	49	11
	90대	1	0	6	2	3	4	3	1
도착	0대	3	2	126	34	91	132	66	10
	10대	30	19	1,250	430	343	902	665	440
	20대	131	140	2,514	1,685	1,020	1,269	2,147	1,175
	30대	112	156	2,990	1,715	975	1,343	2,432	957
	40대	118	204	3,365	1,850	1,117	1,662	2,611	994
	50대	109	281	3,335	1,943	1,256	1,900	2,536	996
	60대	63	217	1,594	1,193	917	1,221	1,212	448
	70대	21	62	410	485	429	460	324	99
	80대	7	11	80	101	82	88	60	16
	90대	1	0	6	2	3	4	4	1



<그림 3- 15> 연령별 통행 발생량 패턴 분석 (모바일기~~지~~국데이터 기준)



<그림 3- 16> 연령별 통행 도착량 패턴 분석 (모바일기~~지~~국데이터 기준)

- 모바일기지국데이터 기준 연령별 통행패턴 분석결과, 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 학업활동 연령대는 다소 이른 오후 1시부터 통행량이 증가하는 패턴을 보여, 실제 상대적으로 이른 하교시간을 정확하게 반영하는 것으로 판단됨
 - 주로 비경제활동 인구에 해당하는 60세 이상의 통행패턴은 경제활동인구에 비해 비교적 완만한 첨두시간 통행패턴을 보여, 실제 연령별 통행패턴과 매우 유사한 결과를 도출한 것으로 판단됨

나. 목적별(출퇴근) 통행패턴 비교 및 분석 결과

1) 가구통행실태조사자료·인구총조사자료 기준 출퇴근 통행량 비교 결과

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

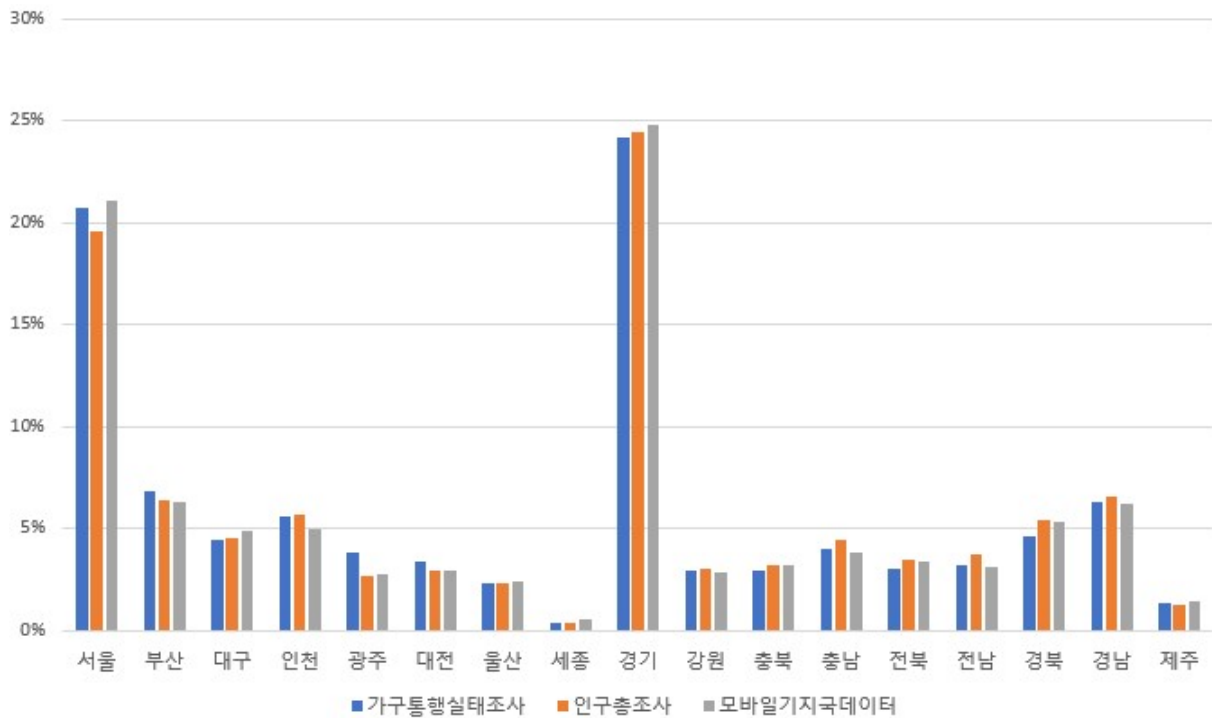
<표 3- 9> 지역별 출퇴근 통행량 비교 결과

(단위: 통행/일(비율%))

구분	가구통행실태조사자료 기준		인구총조사자료 기준	모바일기지국데이터 기준	
	출근 (비율)	귀가*	통근**	출근	퇴근
서울	4,516,466(20.7)	8,413,729(22.1)	4,597,322(19.6)	23,531(20.0)	22,979(22.3)
부산	1,492,300(6.8)	3,051,128(8.0)	1,500,101(6.4)	7,534(6.4)	6,358(6.2)
대구	974,423(4.5)	1,801,446(4.7)	1,064,227(4.5)	6,053(5.2)	4,616(4.5)
인천	1,213,843(5.6)	1,794,429(4.7)	1,344,311(5.7)	6,116(5.2)	4,828(4.7)
광주	838,601(3.8)	1,279,355(3.4)	639,412(2.7)	3,424(2.9)	2,668(2.6)
대전	736,843(3.4)	1,313,431(3.4)	672,856(2.9)	3,617(3.1)	2,952(2.9)
울산	514,493(2.4)	995,135(2.6)	538,654(2.3)	2,791(2.4)	2,518(2.4)
세종	89,203(0.4)	149,020(0.4)	92,403(0.4)	693(0.6)	514(0.5)
경기	5,272,517(24.1)	7,799,702(20.5)	5,745,634(24.4)	30,438(25.9)	24,143(23.4)
강원	644,578(2.9)	1,186,771(3.1)	703,266(3.0)	3,274(2.8)	2,951(2.9)
충북	637,895(2.9)	1,182,200(3.1)	742,399(3.2)	3,696(3.1)	3,433(3.3)
충남	875,177(4.0)	1,616,017(4.2)	1,032,913(4.4)	4,236(3.6)	4,175(4.1)
전북	665,597(3.0)	1,263,474(3.3)	811,899(3.5)	3,935(3.4)	3,463(3.4)
전남	706,911(3.2)	1,241,157(3.3)	864,498(3.7)	3,506(3.0)	3,328(3.2)
경북	1,007,670(4.6)	2,045,073(5.4)	1,280,135(5.4)	5,774(4.9)	5,981(5.8)
경남	1,372,291(6.3)	2,361,671(6.2)	1,560,582(6.6)	7,103(6.1)	6,591(6.4)
제주	291,636(1.3)	581,151(1.5)	309,133(1.3)	1,663(1.4)	1,479(1.4)
계	21,850,443(100)	38,074,889(100)	23,499,745(100)	117,384(100)	102,974(100)

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

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



<그림 3- 17> 지역별 출퇴근 통행량 분포 비교 결과

- 지역별 출퇴근 통행량 비율을 살펴보면, 평균적으로 가구통행실태조사자료 기준 약 0.37%p, 인구총조사자료 기준 약 0.33%p의 차이를 보임
 - 가구통행실태조사자료와 비교하였을 때, 경북지역의 출퇴근 통행량이 약 0.72%p로 가장 큰 폭으로 과대추정되었으며, 광주지역의 출퇴근 통행량이 약 1.26%p로 가장 큰 폭으로 과소추정되었으나, 평균적으로 약 0.37%p로 모바일기지국데이터 기준 출퇴근 통행량이 신뢰성 있는 것으로 판단됨
 - 인구총조사자료 기준 서울지역의 출퇴근 통행량이 약 1.51%p로 가장 큰 폭으로 과대추정되었으며, 인천지역의 출퇴근 통행량이 약 0.73%p로 가장 큰 폭으로 과소추정되었으나, 평균적으로 약 0.33%p로 모바일기지국데이터 기준 출퇴근 통행량이 신뢰성 있는 것으로 판단됨

2) 가구통행실태조사자료·인구총조사자료 기준 출퇴근 통행시간 비교 결과

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

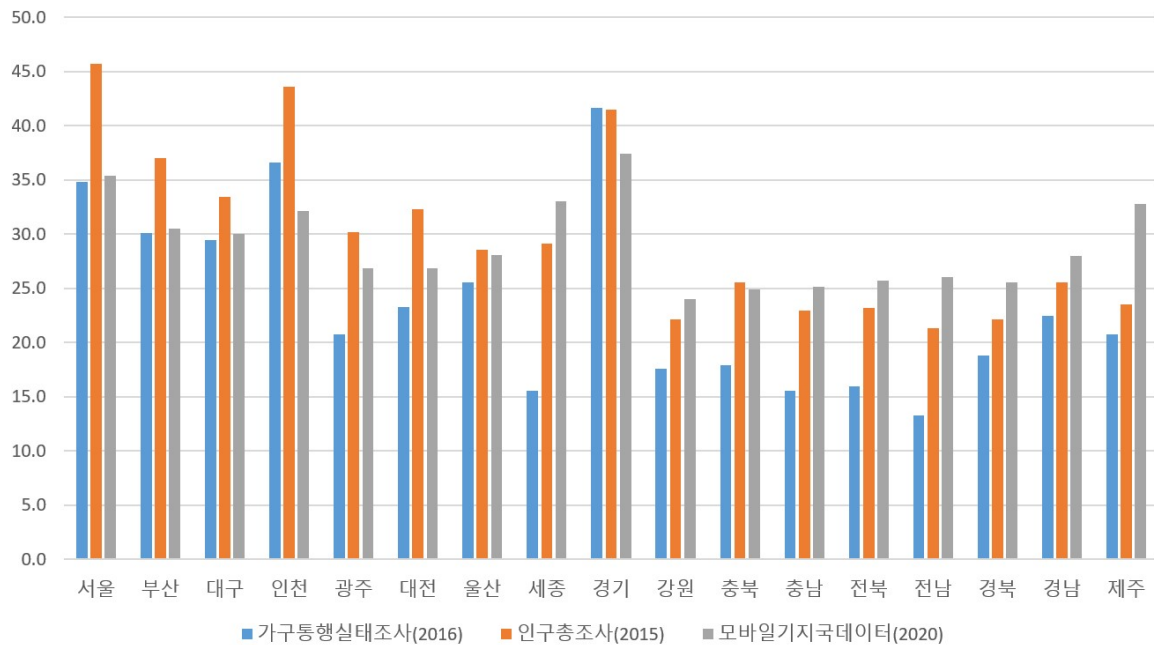
<표 3- 10> 지역별 출퇴근 통행시간 비교 결과

(단위: 분)

구분	가구통행실태조사자료 기준		인구총조사 자료 기준	모바일기지국데이터 기준	
	출근	귀가*	통근**	출근	퇴근
서울	34.8	30.1	45.7	35.4	41.6
부산	30.1	25.6	37.0	30.5	31.2
대구	29.5	28.2	33.4	30.1	26.8
인천	36.6	32.5	43.6	32.1	35.2
광주	20.8	18.3	30.2	26.9	25.4
대전	23.3	21.7	32.3	26.8	26.2
울산	25.6	24.5	28.6	28.1	30.4
세종	15.6	15	29.1	33.0	31.8
경기	41.7	36.3	41.5	37.4	37.0
강원	17.6	20.3	22.1	24.0	26.1
충북	17.9	17.8	25.6	24.9	27.1
충남	15.6	16.7	22.9	25.2	29.6
전북	16	15.4	23.2	25.7	26.3
전남	13.3	13.6	21.3	26.0	30.2
경북	18.8	20	22.1	25.5	30.0
경남	22.5	21.1	25.6	28.0	30.8
제주	20.8	21.5	23.6	32.8	34.6
계	23.6	22.3	35.2	32.1	34.0

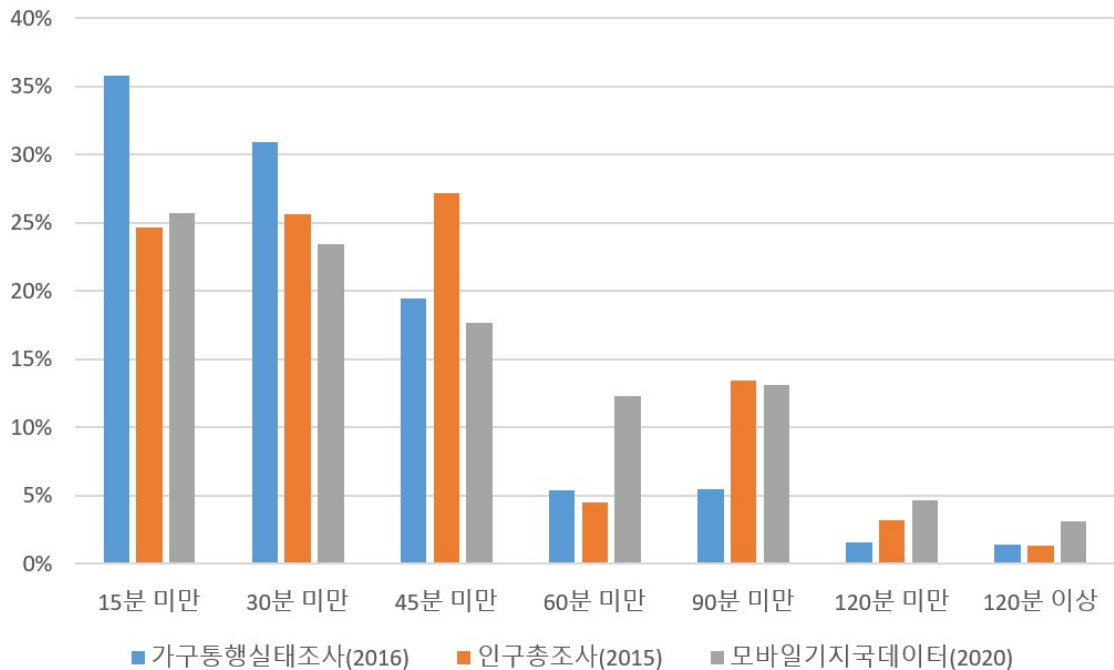
주 : *가구통행실태조사자료는 2016년 기준 출근·귀가 통행에 대한 일평균시간이며, 귀가통행 통행목적에 상관 없이 집으로 향하는 모든 통행을 의미함

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



<그림 3- 18> 시도별 통근 통행시간 비교(단위 : 분)

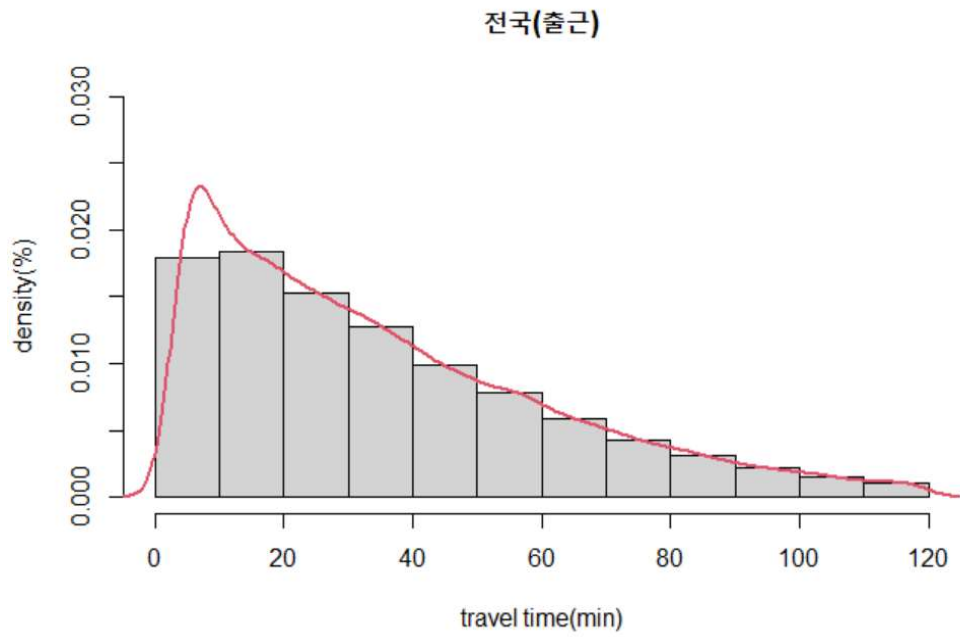
- 가구통행실태조사자료와 인구총조사자료 기준 1인당 평균 출근시간 및 통근시간은 각각 약 23.6분, 35.2분이며, 모바일기지국데이터 기준은 약 32.1분임
- 각 자료별 평균 출퇴근 통행시간을 비교해보면, 인구총조사자료 기준이 가장 길고, 가구통행실태조사자료 기준이 가장 짧은 것으로 나타남
- 지역별 분석결과를 살펴보면, 모바일기지국데이터는 특별시·광역시 지역에서 가구통행실태조사자료와 유사한 출퇴근 통행시간 분포를 띄며, 시도에서는 인구총조사자료와 유사한 출퇴근 통행시간 분포를 띄
- 각각 지역마다 출퇴근 통행시간을 비교해보면 다소 차이가 있다고 해석할 수 있지만, 전체적인 지역적 패턴을 기준으로 보았을 때 매우 유사한 분포를 띄



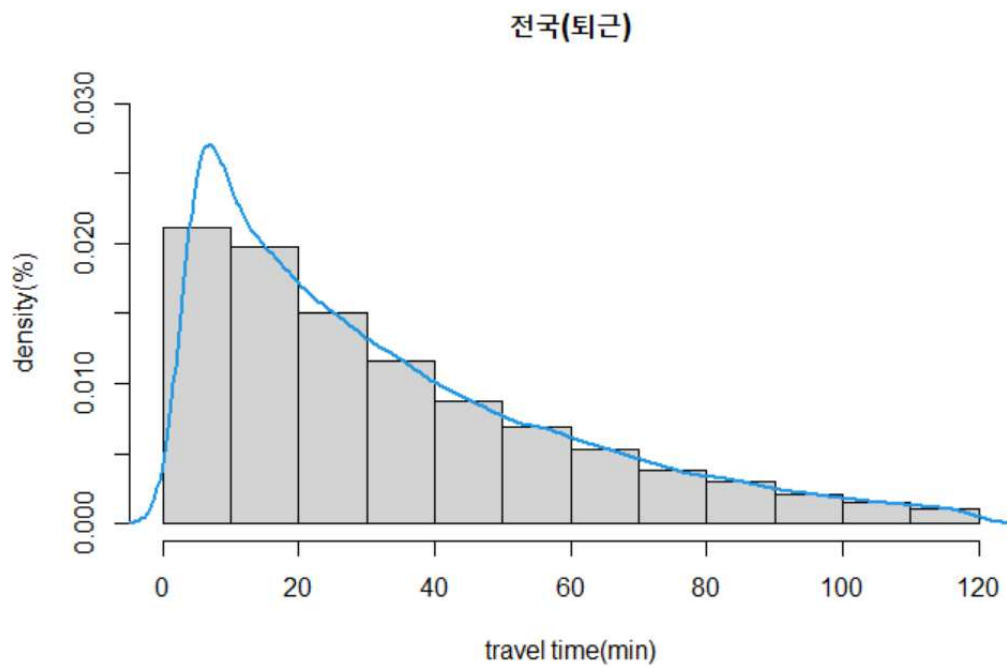
<그림 3- 19> 통행시간 기준별 통근 인구 비율 비교

- 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 가구통행실태조사자료 대비 다소 과소 추정되었으나, 인구총조사자료 기준과는 매우 유사한 분포를 띄는 것으로 나타남
 - 30분 이상 45분 미만 출퇴근 통행량은 가구통행실태조사 기준과 매우 유사한 분포를 띄며, 인구총조사자료와 비교하였을때는 과소 추정된 것으로 나타남
 - 45분 이상 60분 미만에서는 다소 과대추정됨
 - 통행시간대별로 모바일기지기국데이터가 과소 또는 과대추정되어 나타났지만, 출퇴근 시간, 출퇴근 속도 밀도분포를 확인해본 결과, 매우 신뢰성 있는 분포를 보임³⁾
 - <그림 3-20>, <그림 3-21>은 출퇴근 시간 밀도그래프임. 모바일기지기국데이터기준 출퇴근 통행시간은 8분을 정점으로 음이항 지수(negative exponential) 분포에 가깝게 나타남
 - <그림 3-22>와 <그림 3-23>은 출퇴근 속도 밀도그래프임. 출퇴근 통행속도는 12km/h를 정점으로 로그 노말(log-normal) 분포에 가깝게 나타남

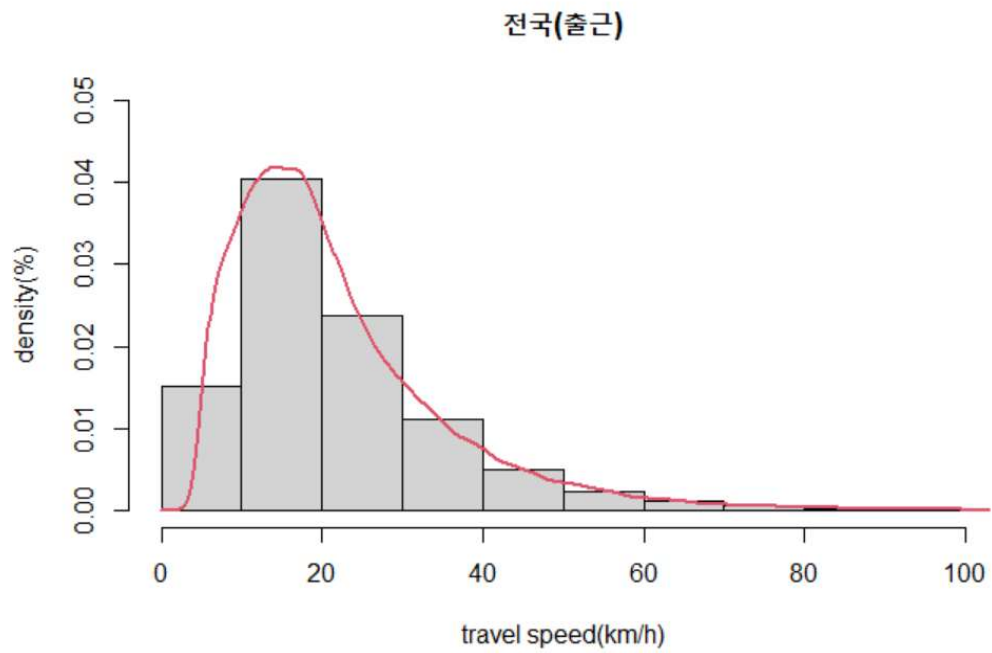
3) 노정현, '교통계획: 통행수요이론과 모형', 개정판(2012), 나남.



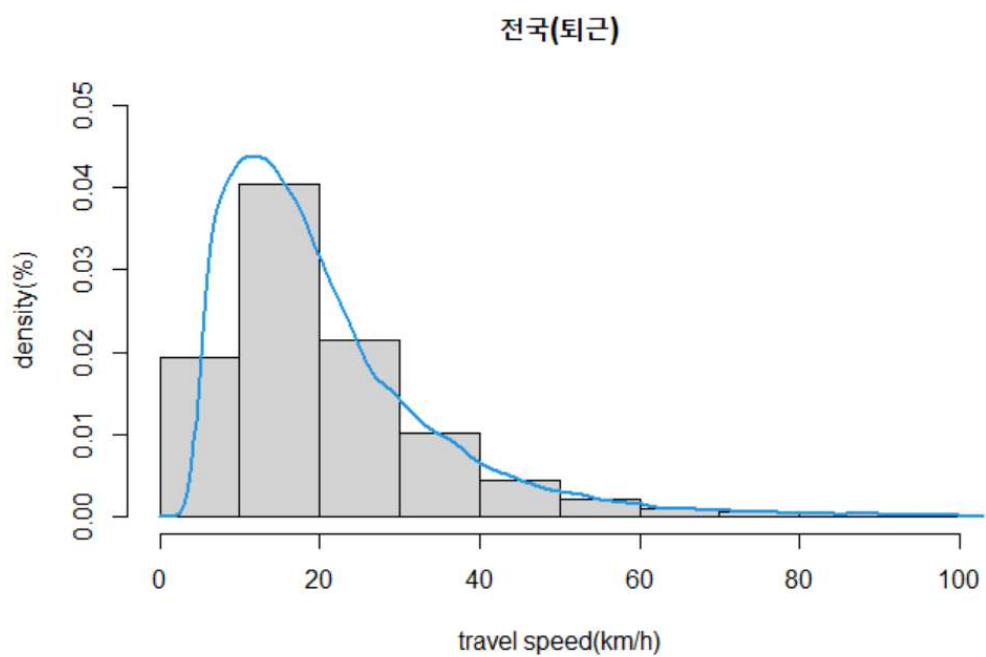
<그림 3- 20> 출근시간 밀도그래프



<그림 3- 21> 퇴근시간 밀도그래프



<그림 3- 22> 출근속도 밀도그래프



<그림 3- 23> 퇴근속도 밀도그래프

제4장 모바일통신 빅데이터 기반 통행 분석 서비스 개발

제1절 개요

제2절 분석 서비스 고도화

제3절 기반 DB 현행화 및 속도 개선

제4절 데이터 다운로드 서비스 개선

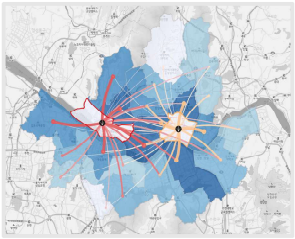
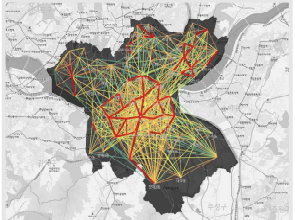
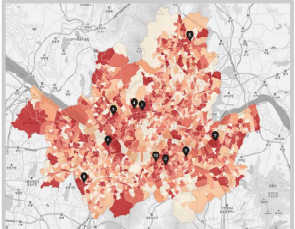
제4장 모바일통신 빅데이터 기반 통행 분석 서비스 개발

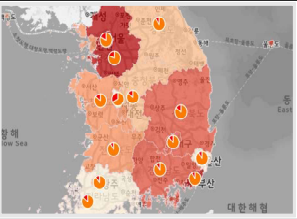
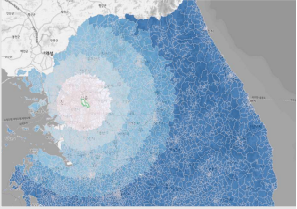
제1절 개요

1. 기존 분석 서비스 (2019년 개발 기준)

- 교통 모니터링·데이터 제공·분석 플랫폼(View-T)를 통해 모바일 통신 빅데이터를 기반으로 하는 5가지의 통행분석 서비스를 제공
- － 2018년에 개발한 3종의 통행 분석 기능(주요통행지역분석, 주요통행구간분석, Hot place 분석)을 개선하고, 2종의 통행 분석 기능(내부통행지역분석, 통행시간·거리분석)을 추가로 탑재

<표 4- 1> 2019년 기준 모바일 통신 빅데이터 기반 통행 분석 서비스

명칭	기능 설명	2019년 개발 내용
주요 통행 지역 분석	 <p>선택 지역에 대한 주요 출발지와 주요 도착지를 확인·비교할 수 있는 기능</p>	<ul style="list-style-type: none"> － 2018년에 개발된 기능 개선 <ul style="list-style-type: none"> · 기존보다 더 많은 지역을 비교 분석할 수 있도록 분석 대상 지역을 확대 (2개 지역 선택 가능 → 5개 지역 선택 가능) · 표출 방법 개선 (선택지역별 분석결과를 각기 다른 색상으로 표현) · 행정구역 ON/OFF 기능 추가 · 상세결과를 파이차트 형식으로 제공
주요 통행 구간 분석	 <p>통행이 많은 구간을 확인할 수 있는 기능</p>	<ul style="list-style-type: none"> － 2018년에 개발된 기능 개선 <ul style="list-style-type: none"> · 지역간 통행 정보를 확인할 수 있도록 분석 대상 지역을 확대 (1개 지역 선택 가능 → 5개 지역 선택 가능) · 표출 단위 확대 (교통폴리곤 단위로 표출 → 교통폴리곤, 읍면동, 시군구 단위로 표출)
Hot Place 분석	 <p>통행수요가 높은 출발지 또는 도착지를 확인할 수 있는 기능</p>	<ul style="list-style-type: none"> － 2018년에 개발된 기능 개선 <ul style="list-style-type: none"> · 표출 단위 확대 (교통폴리곤 단위로 표출 → 교통폴리곤, 읍면동, 시군구 단위로 표출) · 다양한 통행 정보를 표출 (총 통행량 표출 → 총 통행량, 일평균 통행량, 면적대비 통행량 표출)

내부 통행 지역 분석		내부 통행량이 많은 지역(전체 통행량 대비 내부 통행량 비율)을 확인할 수 있는 기능	- 신규 기능 개발
통행 시간 거리 분석		지역별 접근성(평균 통행시간, 평균 통행거리)을 확인할 수 있는 기능	- 신규 기능 개발

- 2017년도의 통행정보를 기준으로 분석 서비스 제공

- 2017년도에 생성된 모바일 통신 빅데이터로부터 집계된 통행정보(통행량)를 2017년 기준 교통폴리곤(2016년 기준 행정구역)을 통해 표출
- 사용자가 시스템에 구축된 기반 데이터를 이중 데이터와 연계하여 분석하거나 다른 방식으로 결과를 표출하는데 활용할 수 있도록 데이터 다운로드 서비스 제공

2. 2020년 개발 방향 및 범위

가. 분석 서비스 고도화

1) 이용자 특성에 맞춰 분석 서비스 재구조화

- 일반사용자를 위한 분석서비스와 고급사용자를 위한 분석서비스를 별도로 개발
 - 기 개발된 분석서비스에는 전문적인 용어가 많고, 조건 입력 등 사용방법이 어려워 일반 국민(교통 비전문가)이 이용하기 쉽지 않았음
 - 일반 국민도 쉽고 빠르게 통행 정보를 확인할 수 있도록 일반사용자용 “Light version”의 분석서비스를 추가로 개발하고, 기존에 제공했던 분석서비스는 “Expert version”으로 명명하고 전문가에게 친숙한 UI/UX로 변경
 - Light version은 출근, 관광 등 일반 국민에게 친숙한 주제로 분석 기능을 개발

<표 4- 2> Light version과 Expert version 구분

구분	Light version	Expert version
이용 대상	교통 비전문가, 일반 국민	학계·산업계 등 교통관련 전문가
특징	<ul style="list-style-type: none"> - 특정 목적을 위한 간편 분석기능으로 구성됨 - 간단한 마우스 조작을 통한 쉬운 분석 결과 도출 - 목적에 특화된 DB를 구성하여 가볍고 빠른 분석 기능 제공 	<ul style="list-style-type: none"> - 전체 데이터를 활용할 수 있는 다양한 상세 분석 옵션 제공 (특정 위치, 특정 시간, 특정 기간, 특정 통행자 조건 설정 가능) - 분석 사용자가 원하는 방식으로 시각화 할 수 있는 다양한 시각화 옵션 제공 - 데이터를 통한 분석이 필요한 사용자를 위한 다양한 데이터 다운로드 기능 제공

2) 신규 분석 기능 개발

- 심층 분석을 가능케 하는 분석 기능 탑재
 - 2개 이상의 분석 수행 결과를 주제도로 동시에 확인할 수 있는 “정보 연계 분석 기능” 개발
 - 통행변화를 시계열로 확인할 수 있는 “통행변화 시뮬레이션 분석 기능” 개발
 - 집계값이나 추정값이 아닌 1시간 단위의 실제값을 제공하는 모바일 통신 통행 데이터 특성을 적극 활용하여, 시간 흐름에 따른 통행 변화를 확인할 수 있는 분석 기능을 개발
- View-T Light의 분석서비스 중 일부를 Expert에 개발

3) 이용자 편의성을 고려한 분석서비스 개선

- View-T 및 모바일 통신 데이터 기반 통행 분석 대한 이해를 돕기 위한 portal 사이트 제작
 - 통행 분석 서비스를 제공하기 위해 구축되어 있는 데이터, 각종 분석지표, 분석도구에 대한 설명과 함께 분석 활용 사례를 소개
 - 사용자가 바로 분석서비스(Light version, Expert version)를 이용해볼 수 있도록 접근 루트 생성

- 직관적인 분석 기능사용을 위한 UI/UX 개선
 - 다수의 분석 조건을 선택하는데 발생하는 불편함을 줄일 수 있도록 UI/UX를 개선
- 분석 기능 튜토리얼 제공
 - 사용자의 이해도를 높일 수 있도록 조건 설정/분석 과정/분석 결과를 동영상 또는 그림으로 제작
 - 기존에는 사용자가 사용방법을 찾아서 확인해야 하는 번거로움이 존재

<표 4-3> 분석 기능 사용방법 안내 방식 변경

변경 전 (기존 방식)	변경 후
<ul style="list-style-type: none"> - PDF 파일로 된 매뉴얼 제공 - 사용자가 분석기능 동작시(마우스오버) 사용 순서, 설명 표출 	<ul style="list-style-type: none"> - 동영상, 그림으로 제작된 튜토리얼 제공

나. 기반 DB 현행화 및 속도 개선

1) 최신 통행 DB, 분석 맵 업데이트

- 사용자가 최신의 통행 데이터로 분석서비스를 이용할 수 있도록 2019년 과업에서 구축한 2018년 기준 통행 DB와 분석 맵을 시스템에 탑재

<표 4- 4> 기반 DB 업데이트

구분		기존 구축 DB (2019년 기준)	신규 구축 DB (2020년 기준)
통행 DB	시간적 범위	2017년 1월 ~ 2017년 12월	2018년 1월 ~ 2018년 12월
	공간적 범위	전국	좌동
	내용적 범위	통행량	통행량, 평균통행시간
분석 맵	교통폴리곤	- 2016년 집계구와 2017년 기지국 정보를 활용하여 구축한 영역 (2018년 개발)	- 2018년 (6월 기준) 집계구와 2018년 기지국 정보를 활용하여 구축한 영역 (2019년 개발)
	행정구역	- 2016년 기준 읍면동, 시군구, 시도 영역	- 2018년 기준 읍면동, 시군구, 시도 영역

2) 분석·표출 속도 향상을 위한 DB 재구성

- 분석 수행시 조회하는 데이터 건수를 최소화할 수 있도록 데이터를 그룹화
 - 교통폴리곤별 통행정보 집계 후 인덱스 생성
 - 교통폴리곤 단위로 1차 집계된 통행정보를 다시 행정구역(읍면동, 시군구, 시도) 단위로 재집계
 - 사용자가 자주 검색하는 분석 조건(예: 오전첨두, 주말 등)을 기준으로 통행 정보를 집계

다. 다운로드 정책 개선

- 데이터 제공처(통신업체), 데이터 수요처(국민, 민간기업 등)의 요구를 반영하여 다운로드 서비스를 “공공용”, “인가용”, “주문형” 세 가지로 구분
 - 기존에는 통신업체에서 판매하고 있는 데이터와 View-T를 통해 제공하는 데이터의 시·공간적 위계가 유사하여, 사용자에게 데이터 제공시 통신업체와 별도의 협의가 필요하다는 것과, 협의가 완료되는 시점까지 사용자(신청자)는 무한정 대기해야 한다는 문제가 존재

<표 4- 5> 2020년 기준 다운로드 정책

구분	공공용	인가용	주문형
목적	사용자(일반 국민) 대기시간 최소화	데이터 제공처의 데이터 판매사업과의 상충 문제 해소	사용자(민간기업, 지자체)의 분석 목적 반영
데이터 형태	시/도 단위의 통행 정보	시군구 이하 단위의 통행 정보	시군구 이하 단위의 통행 정보 (이종 데이터 연계 고려)
절차	데이터 신청서 작성(기본 정보만 작성) → 데이터 다운로드	데이터 신청서 작성(상세 정보까지 작성) → 데이터 제공처(통신업체) 승인 → 데이터 다운로드	데이터 신청서 작성(상세 정보까지 작성) → 데이터 제공처(통신업체) 승인 → 데이터 가공(KOTI) → 데이터 다운로드

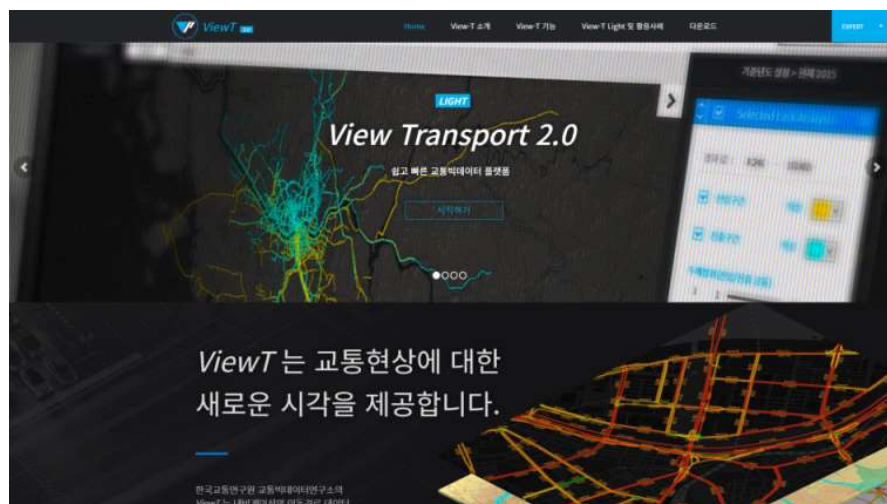
제2절 분석 서비스 고도화

1. View-T Portal

- View-T Portal 사이트는 View-T를 처음 접한 사용자를 위한 서비스로, View-T의 구성요소에 대한 설명 및 활용사례를 제공하여 시스템 전반의 이해를 도움
 - “View-T 소개, View-T 기능, View-T Light 및 활용사례, 다운로드” 네 가지 메뉴로 구성

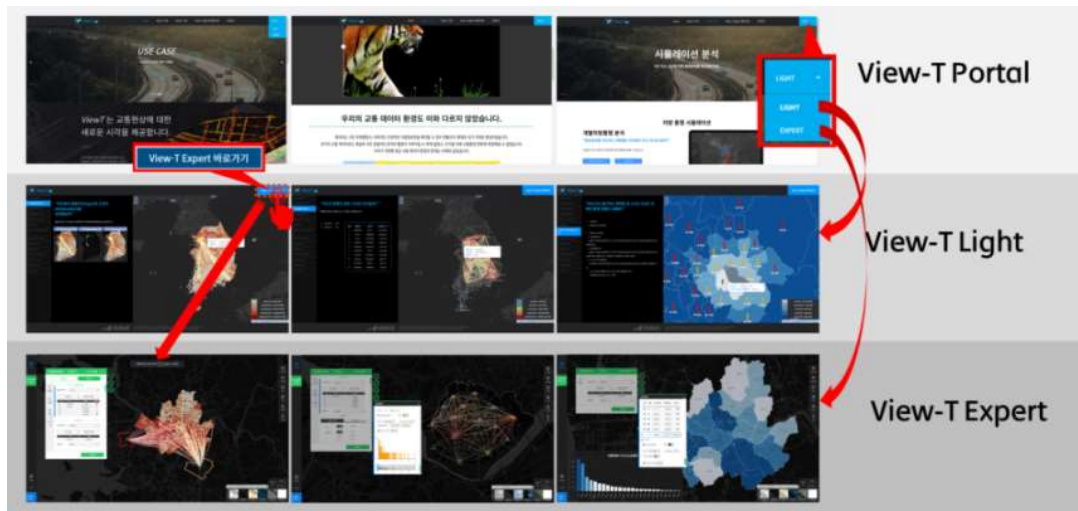
<표 4- 6> View-T Portal 사이트 구성

View-T 소개	View-T 기능	View-T Light 및 활용사례	다운로드
- View-T 서비스의 배경 및 방향성 소개 (구축 배경, 구축 연혁, 구축데이터, 구축 철학, 비전 및 로드맵에 대한 설명으로 구성)	- 각종 분석지표, 분석도구, 시뮬레이션 기능, 분석 리포트, 부가기능과 사용법을 소개 (기능 설명과 함께 해당 기능의 튜토리얼을 제공하여 해당 기능의 사용법을 쉽게 숙지할 수 있도록 함)	View-T를 통한 다양한 분석 사례를 소개	



<그림 4- 1> View-T Portal 메인 화면

- 서브시스템인 View-T Light와 View-T Expert의 진입 통로의 역할을 수행
 - 직접 진입 통로: 우측 상단 드롭다운 메뉴 또는 팝업창을 통해 진입



<그림 4- 2> View-T Light와 View-T Expert 직접 진입 방법

- 연계 진입 통로: 기본 메뉴에서 제공하는 바로가기 버튼을 통해 연계 진입
 - View-T Light 진입 통로: “View-T Light 및 활용사례” 메뉴에서 간편 분석 기능으로 연계 진입
 - View-T Expert 진입 통로: “View-T 기능” 메뉴에서 상세 분석기능으로 진입하거나, View-T Light에 진입 후 우측 상단의 “View-T Expert로 바로가기”버튼을 클릭하여 진입



<그림 4- 3> View-T Light와 View-T Expert 연계 진입 방법

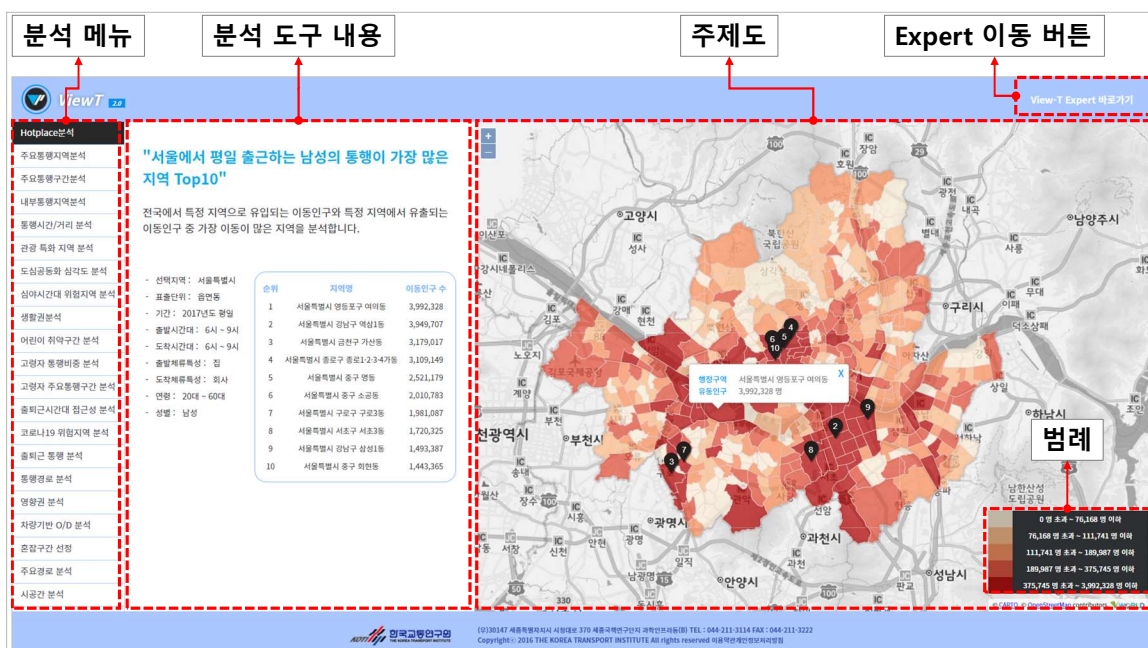
2. View-T Light

가. 개요

- View-T Light는 좀 더 간단하게, 빠르게, 이해하기 쉽게 View-T의 분석 도구들을 사용하자는 의도에서 개발된 분석서비스로, 기존 View-T 주요 분석 도구들의 목적에 최적화 된 간소화 버전을 제공하여 사용자가 몇 번의 클릭만으로 의도한 분석결과가 시각화 된 모습을 볼 수 있어 기능성을 강조하고, 사용자가 View-T 전반에 대해 직관적으로 이해할 수 있도록 도움
- 기존 View-T는 방대한 교통 데이터 위에서 복잡한 UI, 상세한 분석 조건 설정을 통하여 사용자가 원하는 차량 및 통신 이동 통행을 분석 및 시각화하고 있었으나,

이는 View-T에 대한 이해도가 낮은 사용자에게는 시스템 전반의 기능을 이해하기 어렵고, 친숙도가 떨어지는 애로사항이 있었음

- View-T Light는 각 주요 분석 도구 메뉴, 각 도구마다의 간단한 설명과 기본 분석 설정 값, 주제도 및 범례로 구성
 - View-T Light에서 사용할 수 있는 기능은 기존 View-T의 분석 도구 중 차량과 통신 각 파트별 주요 분석 도구들만을 제공
 - 메뉴에서 각 분석 도구를 선택하면 이미 설정되어 있는 기본 설정 값과 해당 설정 값에 최적화 된 정제 데이터를 이용하여 우측 주제도 상에 분석결과를 표출
 - 메뉴와 주제도 사이에는 해당 분석 도구의 대표적인 캐치프레이즈, 분석 내용에 대한 간단한 설명과 함께 분석에 사용된 기본 설정 값을 표출
 - 분석 도구에 따라 이미지, 테이블 차트, 그래프 등을 함께 시각화하여 제공

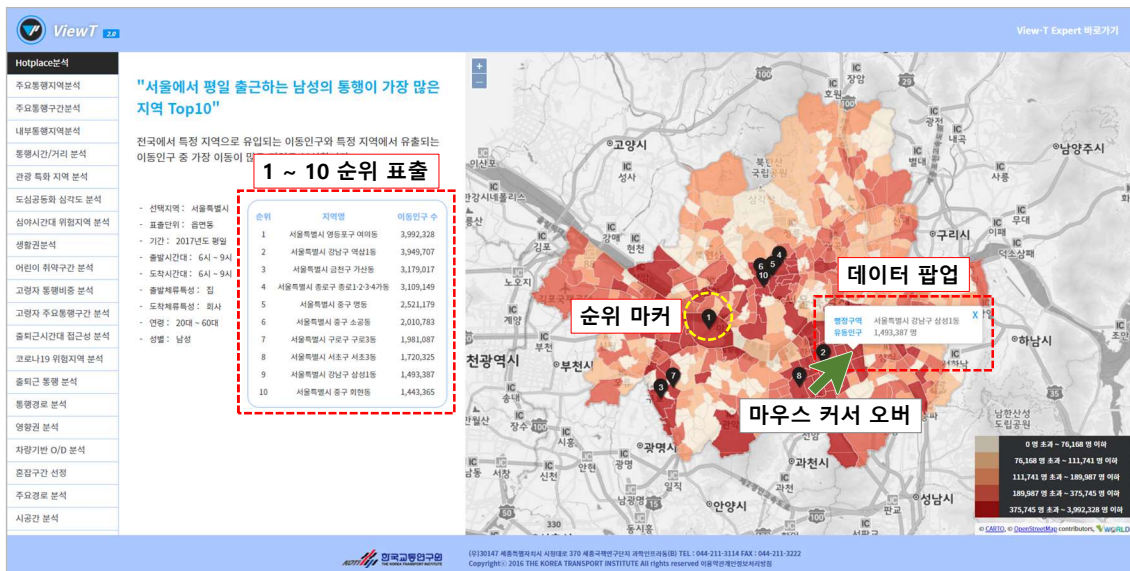


<그림 4- 4> View-T Light 화면 구성

나. 기존 분석 기능의 View-T Light 버전 개발

1) Hotplace 분석

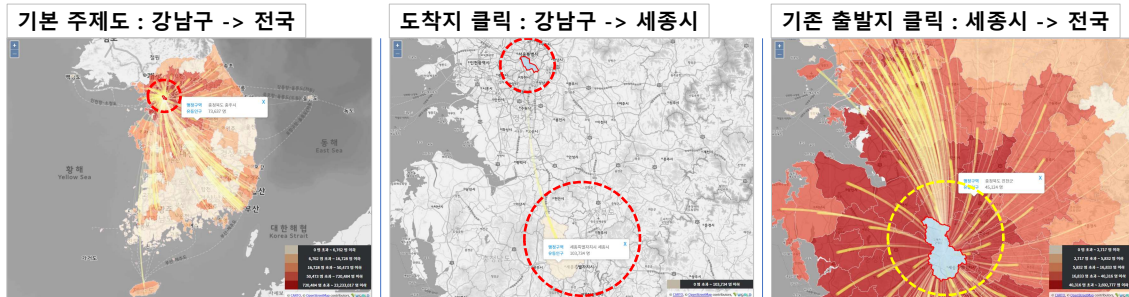
- 전국에서 특정 지역으로 유입되는 이동인구와 특정 지역에서 유출되는 이동인구 중 가장 이동이 많은 지역을 간단하게 분석
 - 서울에서 평일 출근하는 남성의 통행이 가장 많은 지역 Top10을 표출
 - 우측에는 주제도를 표현하고, 좌측에는 우측 주제도의 조건 및 내용을 설명
 - 우측 하단에는 범례를 표시
 - 주제도상에 마우스 오버 이벤트로 해당 지역의 행정구역명과 유동인구 수를 표현



<그림 4- 5> View-T Light Hotplace 분석 화면

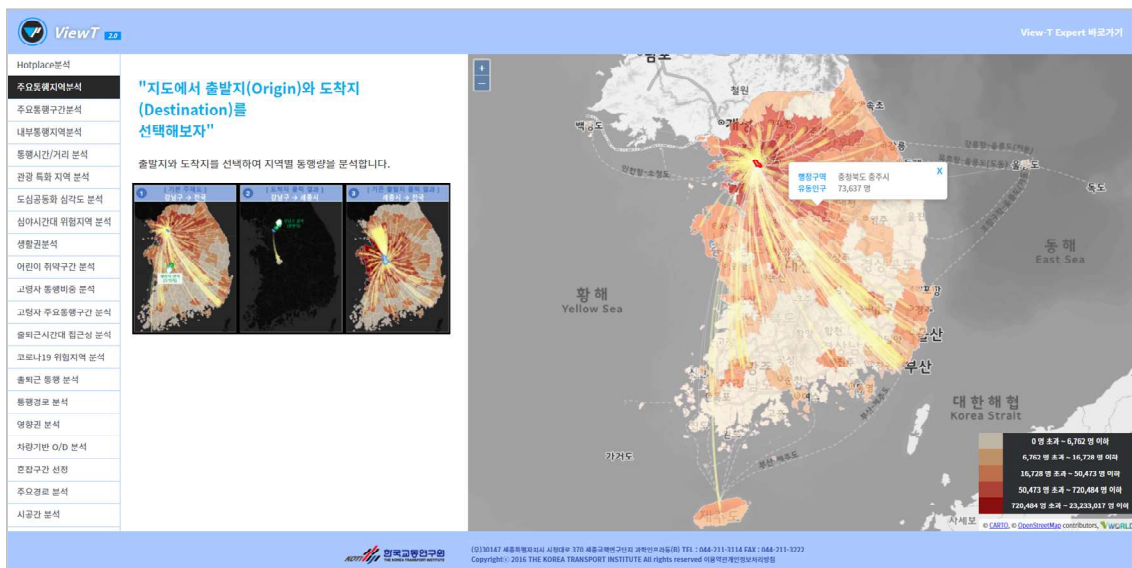
2) 주요통행지역 분석

- 특정 출발지에서 특정 도착지의 지역별 통행량을 간단하게 분석
 - 마우스 클릭만으로 출발지와 도착지를 선택하여 유동인구를 파악할 수 있는 기능
 - 출발지를 제외한 다른 시/군/구를 선택하면 기존 선택 시/군/구에서 새로 선택한 시군구로의 유동인구를 표출
 - 선택된 2개의 시/군/구 중 하나를 클릭하면 해당 지역이 선택 해제되며 나머지 선택 지역에서 전국으로의 유동인구를 표출



<그림 4- 6> View-T Light 주요통행지역 분석 출/도착지 변경 화면

- 기본 주제도 출력은 강남구를 출발지로 하는 전국 시군구로의 유동인구를 표출
- 우측 하단에는 범례를 표시



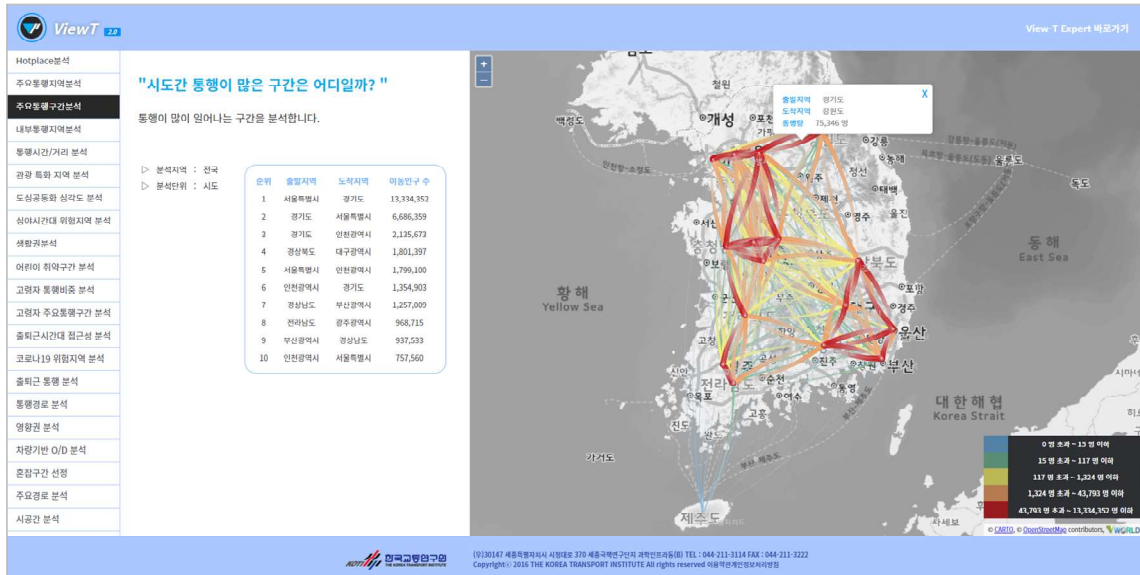
<그림 4- 7> View-T Light 주요통행지역 분석 화면

3) 주요통행구간 분석

- 통행이 많이 일어나는 구간을 간단하게 분석
 - 시도간 통행이 많이 일어나는 구간을 분석한 내용을 시각화 함
 - 기본 주제도 출력은 전국의 시도단위에서 각각 시도에서의 출발 및 도착의 이동인구수를 희망선도로 표현함
 - 주제도상에 마우스 오버 이벤트로 해당 희망선도의 출발지역, 도착지역, 통행자 수

를 표현

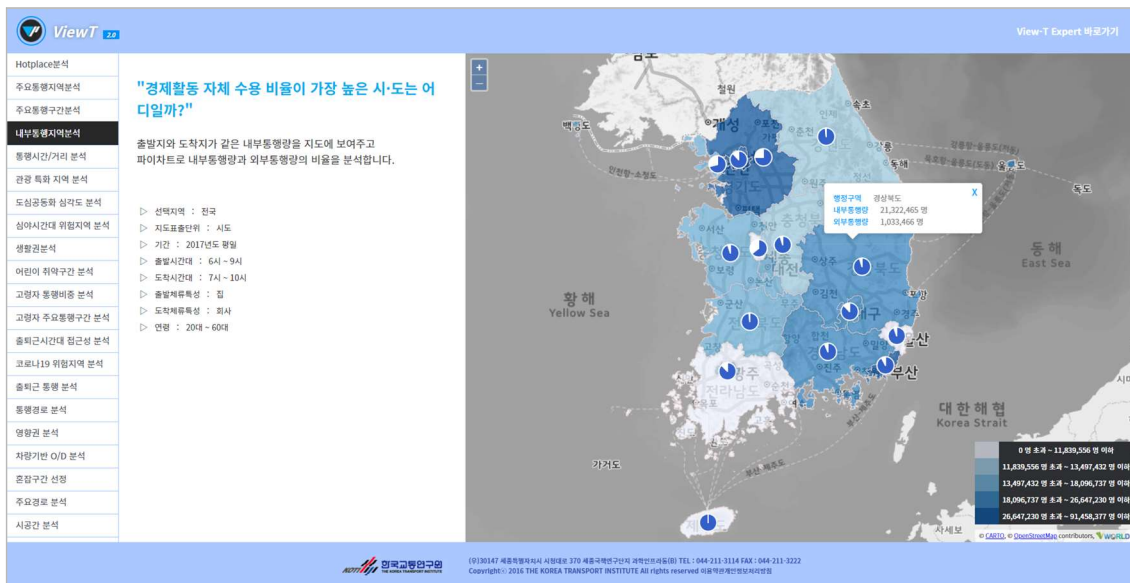
- 좌측에는 우측 주제도의 조건 및 상위 10개 희망선도의 내용을 표현
- 우측 하단에는 범례를 표시



<그림 4- 8> View-T Light 주요통행구간 분석 화면

4) 내부통행지역 분석

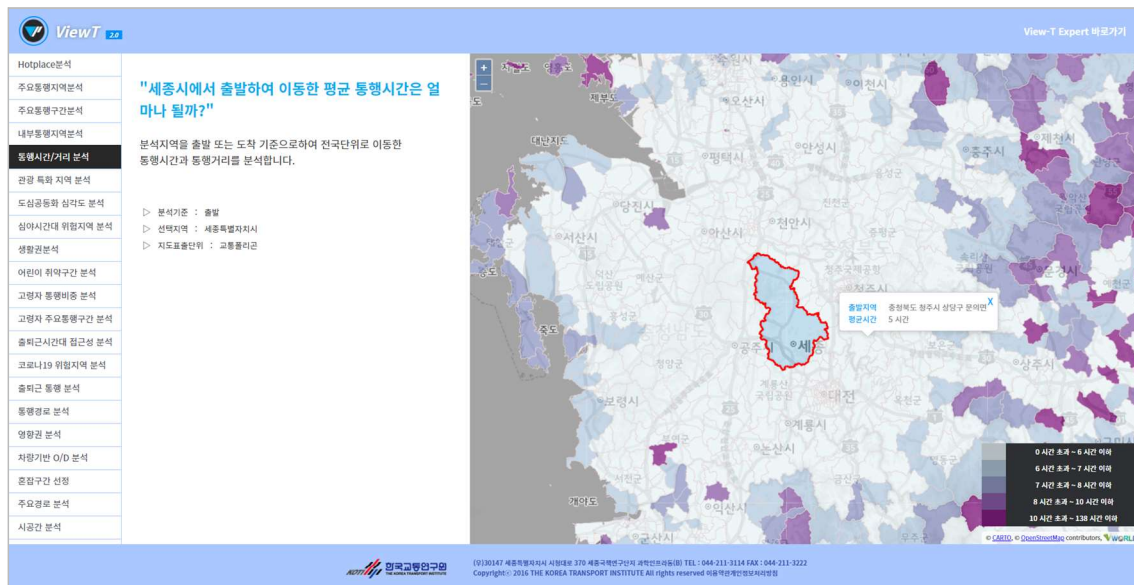
- 출발지와 도착지가 같은 내부통행량을 주제도로 보여주고, 파이차트를 이용하여 내부 통행량과 외부통행량의 비율을 간단하게 분석
 - 경제활동 자체 수용 비율이 가장 높은 시/도를 표현하기 위한 기능
 - 기본 주제도 출력은 출발지와 도착지가 같은 행정구역인 내부통행량을 주제도 상에 표현하고 외부통행량과 내부통행량의 비율을 파이 차트로 표시함
 - 주제도상에 마우스 오버 시 해당 지역의 행정구역명, 내부통행량, 외부통행량의 수를 표현
 - 좌측에는 우측 주제도의 조건을 표출, 우측 하단에는 범례를 표출



<그림 4- 9> View-T Light 내부통행지역 분석 화면

5) 통행시간/거리 분석

- 특정 분석지역을 출발 또는 도착 기준으로 하여 전국단위로 이동한 통행시간과 통행 거리를 간단하게 분석
 - 경제활동 자체 수용 비율이 가장 높은 시/도를 표현하기 위한 기능
 - 기본 주제도 출력은 세종시에서 출발한 통행량의 평균 통행시간과 거리를 표현함
 - 주제도상에 마우스 오버 이벤트로 해당 지역의 행정구역명과, 평균 통행시간을 표현
 - 좌측에는 우측 주제도의 조건을 표출, 우측 하단에는 범례를 표출



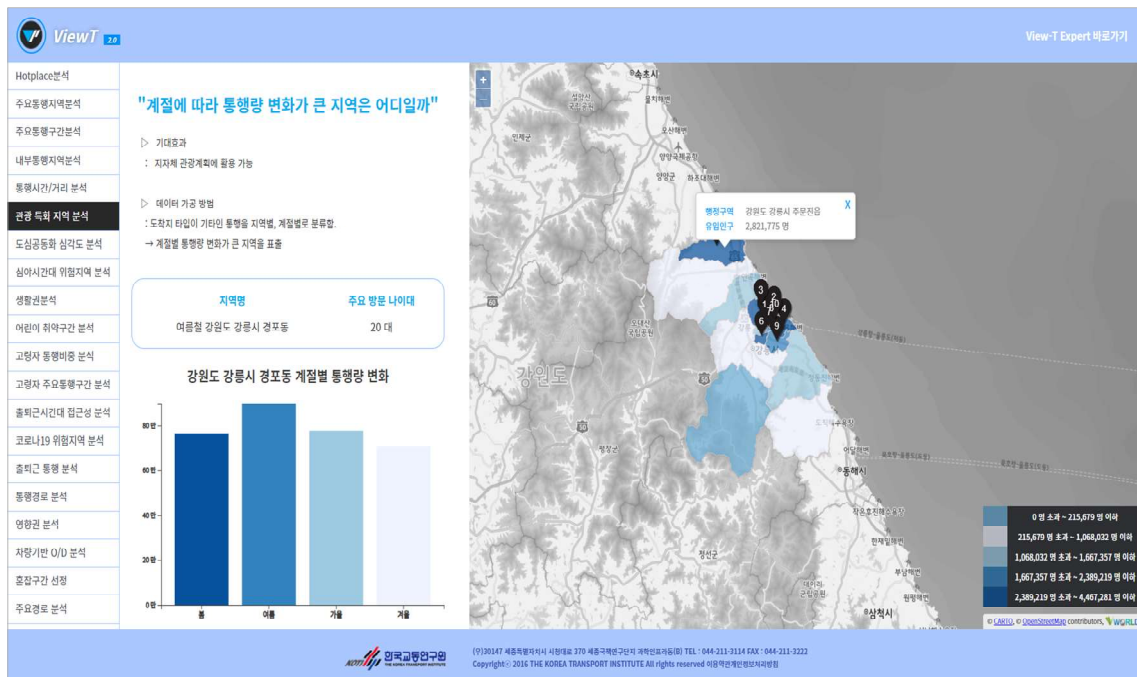
<그림 4- 10> View-T Light 통행시간/거리 분석 화면

다. 목적별 View-T Light 신규 기능 개발

- 기 구축된 통행량 데이터 또는 필요 시 추가 데이터를 사용하여 통신 데이터의 특성이 잘 나타나도록 목적에 알맞은 콘텐츠를 구성하여 사용자에게 다양한 분석의 결과를 제공

1) 관광 특화 지역 분석

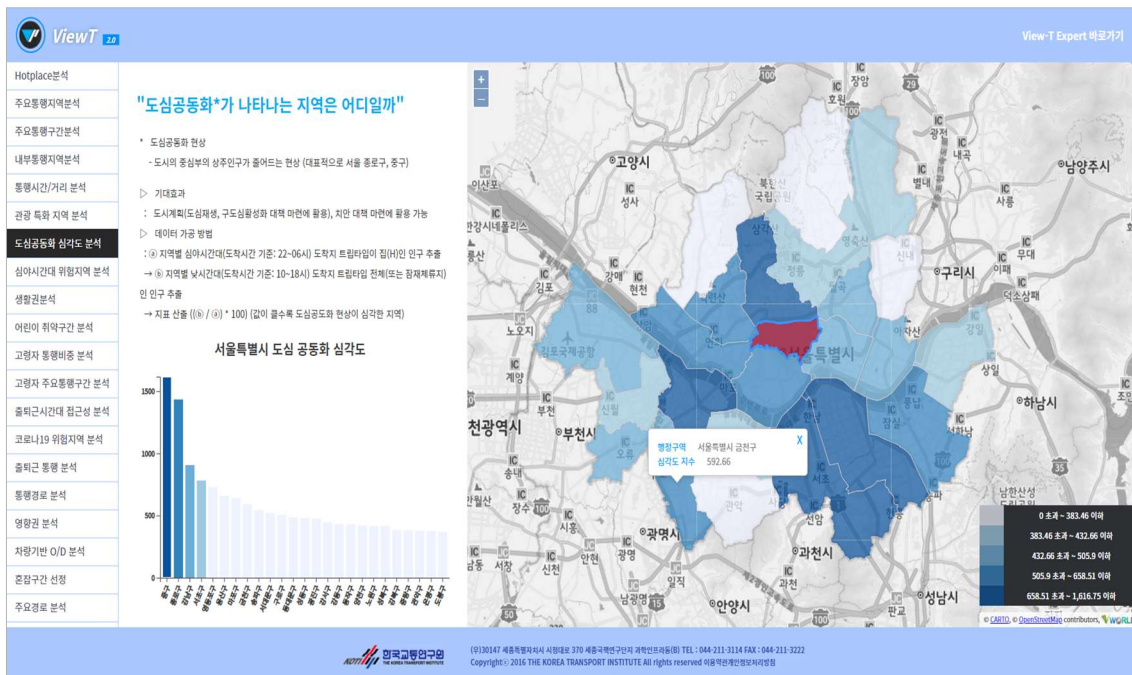
- 계절에 따라 통행량 변화 큰 지역을 확인할 수 있는 기능
- 데이터 가공 방법 : 도착지 타입이 비정기적 체류지인 통행을 지역별, 계절별로 분류
- 관광 특화 지역 분석 UI
 - 기본 주제도는 강원도 강릉시의 유입량 지표를 색상으로 표출
 - 유입량지표가 높은 상위 10개 지역의 순위를 표시
 - 주제도상에 마우스 오버 이벤트로 해당 지역의 유입인구수를 표현
 - 좌측에는 우측 주제도의 조건과 지역명, 주요방문 나이대, 계절별 통행량 차트를 표출
 - 우측 하단에는 범례를 표시



<그림 4- 11> 관광 특화 지역 분석 화면

2) 도심공동화 심각도 분석

- 도심공동화가 나타나는 지역을 분석하기 위한 기능
 - 데이터 가공 방법
 - . ㉠ 지역별 심야시간대(도착시간 기준: 22~06시) 도착지 트립타입이 집(H)인 인구 추출
 - . ㉡ 지역별 낮시간대(도착시간 기준: 10~18시) 도착지 트립타입 전체(또는 잠재 체류지)인 인구 추출
 - . 도심공동화 지표 산출 ($((㉡ / ㉠) * 100)$)
 - 도심공동화 심각도 분석 UI
 - 기본 주제도는 서울시의 도심공동화 지표를 색상으로 표출
 - 주제도상에 마우스 오버 이벤트로 해당 지역의 도심공동화 심각도 지수를 표현
 - 좌측에는 도심공동화 심각도에 대한 차트를 표출
 - 우측 하단에는 범례를 표시



<그림 4- 12> 도심공동화 심각도 분석 화면

3) 생활권분석

○ 주요활동 권역을 표현하기 위한 기능

– 데이터 가공 방법

- 기종점 권역의 중심을 기준으로 SDE분석
- SDE 분석을 통해 그려진 타원의 중첩되는 비율을 측정
- 중첩되는 비율이 50%를 넘는다면 생활권역으로 포함

– SDE(Standard Deviation Ellipse)

- 공간적으로 속성값을 지닌 점들의 공간분포 형상을 분석하는 공간통계분석기법
- 기종점 Point를 기준으로 거리 및 표준편차 등을 활용하여 해당 점들을 포괄하는 타원을 생성하는 기법
- 표준편차 거리계산(\bar{X}, \bar{Y} 는 각각 x 의 평균과 y 의 평균)

$$SDE_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n}} \quad SDE_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{Y})^2}{n}}$$

- 타원체의 기울어진 각도를 계산

$$\tan \theta = \frac{A+B}{C}$$

$$A = \left(\sum_{i=1}^n x_i^2 - \sum_{i=1}^n y_i^2 \right)$$

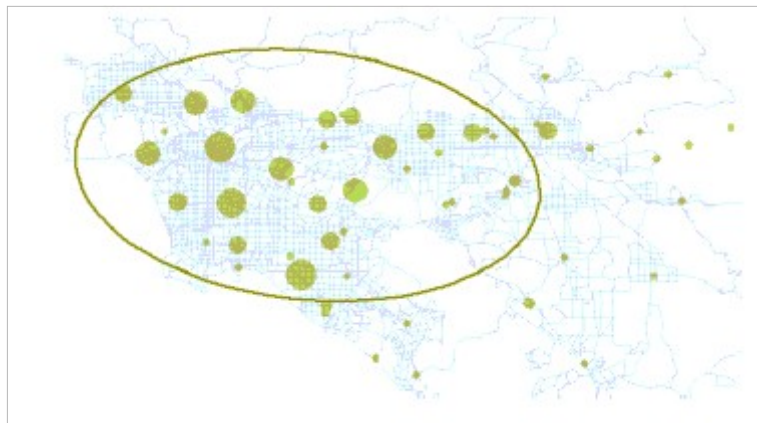
$$B = \sqrt{\left(\sum_{i=1}^n x_i^2 - \sum_{i=1}^n y_i^2 \right)^2 + 4 \left(\sum_{i=1}^n x_i y_i \right)^2}$$

$$C = 2 \sum_{i=1}^n \bar{x}_i \bar{y}_i$$

$$x'_i = x_i - \bar{x} \quad , y'_i = y_i - \bar{y}$$

- . 타원체의 주축과 보조 축을 계산해 타원체를 그림
- . $\sigma(x)$: 타원체의 장축 표준편차거리
- . $\sigma(y)$: 타원체의 단축 표준편차거리

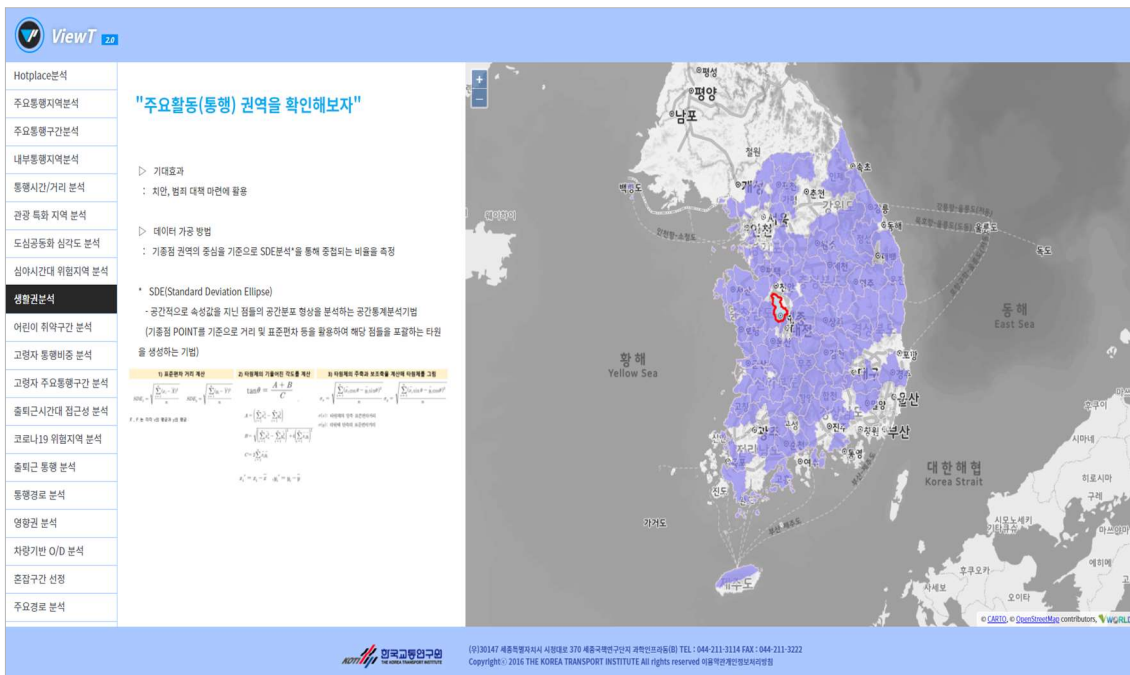
$$\sigma_x = \sqrt{\frac{\sum_{i=1}^m (x_i \cos \theta - \bar{y}_i \sin \theta)^2}{n}} \quad \sigma_y = \sqrt{\frac{\sum_{i=1}^m (\bar{x}_i \sin \theta - y_i \cos \theta)^2}{n}}$$



<그림 4- 13> SDE로 그려진 타원체 예시

○ 생활권분석 기능의 UI

- 지도의 특정 행정구역을 선택하면 사용자가 선택한 행정구역의 생활권은 어디인지 주제도로 표출



<그림 4- 14> 생활권분석 화면

4) 어린이 취약구간 분석

- 어린이 교통사고가 발생할 가능성이 높은 지역을 표현하기 위한 기능
 - 주요 통학구간 산정
 - . ㉠ 모바일 데이터를 활용하여 통학시간대 어린이의 통행구간 표출
 - . ㉡ 학교(어린이집, 유치원, 초중고교) 위치를 기준으로 반경 300m 버퍼를 형성
 - . ㉠과 ㉡를 중첩하여 주요통학구간 추정
 - 어린이 교통사고 가능성 산출
 - . ㉢ 내비게이션 데이터를 활용하여 통학시간대 교통량이 많은 구간 표출
 - . 주요통학구간과 ㉢의 중첩도를 바탕으로 어린이 교통사고 가능성 산출
 - 어린이 교통사고 위험도 산출
 - . ㉣ 주요통학구간 내 교통사고 데이터 표출
 - . 주요통학구간과 ㉣의 중첩도를 바탕으로 어린이 교통사고 위험도 산출
- 어린이 취약구간 분석기능의 UI



<그림 4- 15> 어린이 취약구간 분석 화면

5) 고령자 통행비중 분석

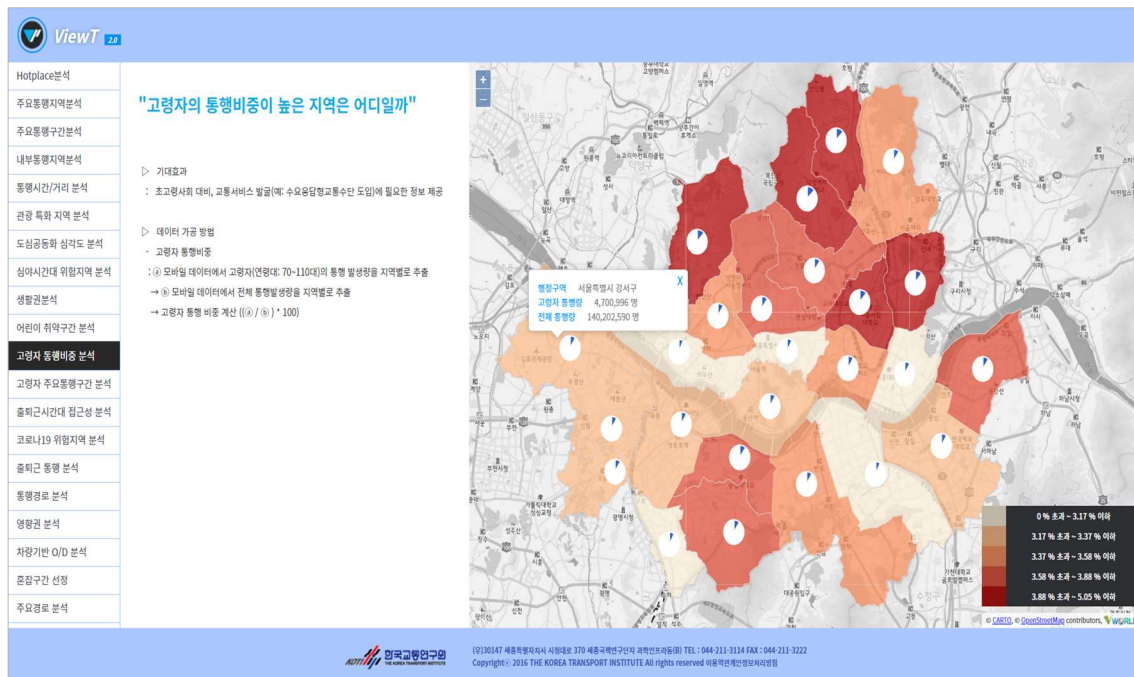
○ 고령자의 통행비중을 분석하기 위한 기능

- 모바일 데이터에서 고령자(연령대:70~110대)의 통행 발생량을 지역별로 추출
- 데이터 가공 방법

- . ㉠ 모바일 데이터에서 고령자(연령대: 70~110대)의 통행 발생량을 지역별로 추출
- . ㉡ 모바일 데이터에서 전체 통행발생량을 지역별로 추출
- . 고령자 통행 비중 계산 ((㉠ / ㉡) * 100)

○ 고령자 통행비중 분석 UI

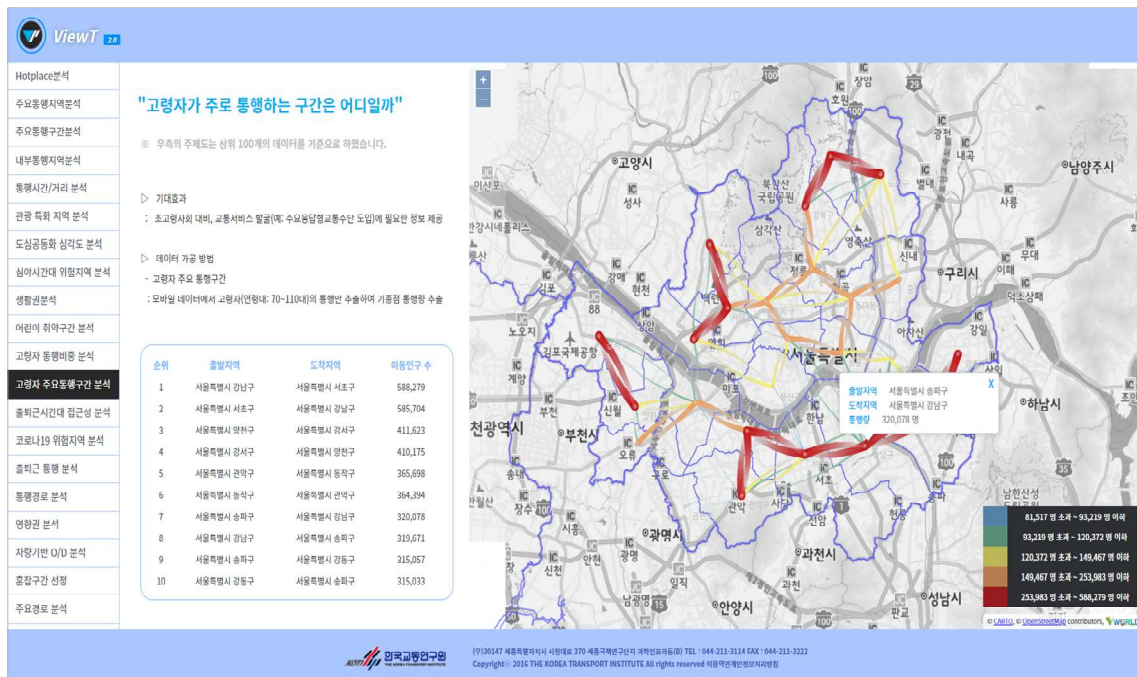
- 고령자 통행 비중 계산하여 주제도 상에 표시
- 주제도상의 고령자 비중을 파이차트로 표시
- 지도의 시도 단위의 행정구역을 클릭하면 사용자가 선택한 시도의 고령자 통행비중 에 대한 결과를 표출해주며, 마우스 오버 시 고령자 통행량과 전체 통행량의 속성정보를 표출
- 우측 하단에는 범례를 표시



<그림 4- 16> 고령자 통행비중 분석 화면

6) 고령자 주요통행구간 분석

- 고령자의 주요통행구간을 분석하기 위한 기능
 - 데이터 가공 방법
 - . 모바일 데이터에서 고령자(연령대: 70~110대)의 통행만 추출하여 기종점 통행량 추출
- 고령자 주요통행구간 분석 UI
 - 희망선도로 고령자 주요 통행구간을 표출함
 - 지도의 시도 단위의 행정구역을 클릭하면 사용자가 선택한 시도의 고령자 주요통행구간에 대한 결과를 표출해주며, 마우스 오버 시 출발지역과 도착지역, 통행량의 속성정보를 표출
 - 주제도상에 마우스 오버 이벤트로 해당 지역의 고령자 통행량을 표현
 - 좌측에는 주요통행구간 상위 10개 순위를 표시
 - 우측 하단에는 범례를 표시



<그림 4- 17> 고령자 주요통행구간 분석 화면

7) 출퇴근시간대 접근성 분석

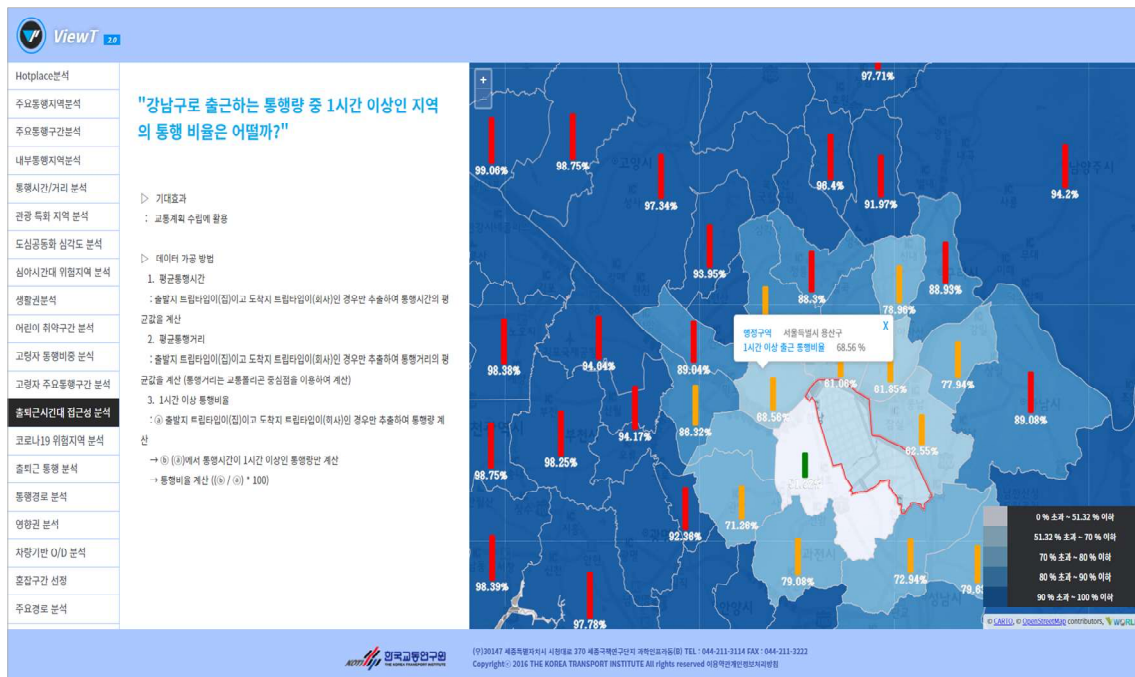
- 출퇴근시간대에 목적지까지 접근성을 확인할 수 있는 기능

- 데이터 가공 방법

- 출발지 트립타입이(집)이고 도착지 트립타입이(회사)인 경우만 추출하여 통행시간의 평균값을 계산하여 평균 통행시간을 산출
- 출발지 트립타입이(집)이고 도착지 트립타입이(회사)인 경우만 추출하여 통행거리의 평균값을 계산하여 평균통행거리 산출 (통행거리는 교통폴리곤 중심점을 이용하여 계산)
- ㉠ 출발지 트립타입이(집)이고 도착지 트립타입이(회사)인 경우만 추출하여 통행량 계산
- ㉡ (㉠)에서 통행시간이 1시간 이상인 통행량만 계산
- 통행비율 계산 ((㉡ / ㉠) * 100)

- 출퇴근시간대 접근성 분석 UI

- 1시간 이상 출근 통행비율의 결과 값을 주제도에 표출
- 마우스 오버 시, 행정구역 명칭과 1시간 이상 출근 통행비율의 속성정보 표출
- 우측 하단에는 범례를 표시

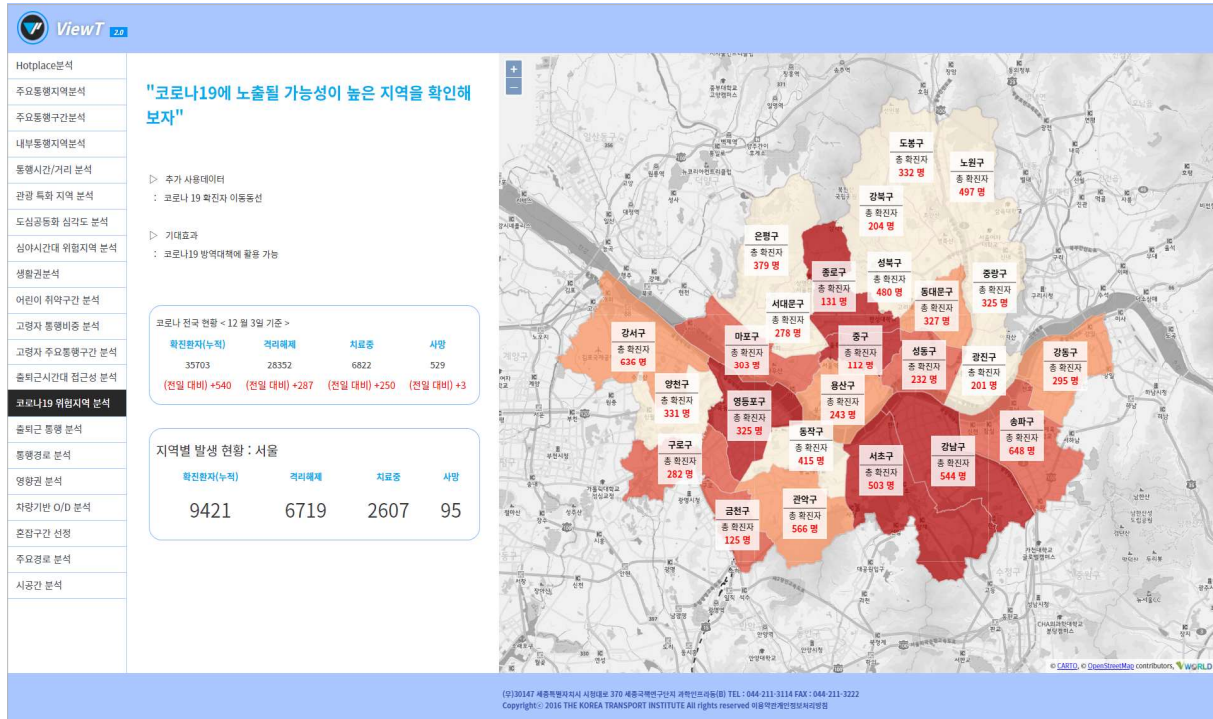


<그림 4- 18> 출퇴근시간대 접근성 분석 화면

8) 코로나19 위험지역 분석

- 코로나19에 노출될 가능성이 높은 지역을 확인할 수 있는 기능
 - 필요데이터 : 시간대별 통행발생량, 코로나 19 확진자 이동 동선
 - 주제도 표출을 위한 데이터 가공 방법
 - . 모바일 데이터 기반 시간대별 통행발생량을 지역별로 추출
 - . 코로나 현황 데이터를 실시간으로 획득 할 수 있는 모듈 개발
 - . 통행 발생량이 높은 지역에서 확진자수 발생량이 많은 지역을 위험지역으로 판단
- 코로나19 위험지역 분석 UI
 - 기본 주제도는 전국의 코로나 위험도를 지역별로 표현
 - 팝업 클릭시 시도 및 시군구 단위로 확대 및 축소 가능
 - 기본 주제도는 지역별 통행량을 기준으로 핫플레이스 표출
 - 지역별(시도,시군구) 코로나 누적 확진자수에 대한 값을 팝업 형태로 표출
 - 전국 코로나 현황(누적 확진자수, 격리해제, 치료중, 사망)에 대한 데이터 테이블 표출

– 시도별 코로나 현황(누적 확진자수, 격리해제, 치료중, 사망)에 대한 데이터 테이블 표출



<그림 4- 19> 코로나19 위험지역 분석 화면

3. View-T Expert

가. 사용자 편의성을 위한 UI/UX 개선

- 심층적인 분석보다 간단한 분석을 선호하는 사용자를 위해 조건 설정, 범례 설정을 간소화함
- “키워드 분석” 옵션을 조건 설정 부분에 추가하여, 사용자가 상세하게 조건을 입력하지 않아도 분석을 수행할 수 있도록 함
- 사용자가 자주 사용하는 분석 조건을 기본 옵션으로 두어, 사용자의 조건 선택을 최소화

개선

상세분석

키워드분석

<그림 4-20> 분석 조건값 UI 변경

<표 4-7> 주제별 키워드 사전 조건

통행목적	출발 조건		도착 조건		기타 조건	비고
	시간대	트립타입	시간대	트립타입		
출근	06~09	집 (H)	07~10	회사(C)	20~60대	필수
등교	06~09	집 (H)	07~10	학교(S)	10~20대	필수
퇴근	18~21	회사(C)	19~22	집 (H)	20~60대	필수
하교	15~21	학교(S)	16~21	집 (H)	10~20대	필수
귀가	18~20	정기적 잠재체류지(R),	19~23	집 (H)	-	필수
		비정기적 잠재체류지(X)				
여가	-	집 (H),	-	정기적 잠재체류지(R), 비정기적잠재체류지(X)	-	필수
		회사(C),				
		학교(S),				
		정기적 잠재체류지(R), 비정기적 잠재체류지(X)				
사적업무 후 복귀	-	정기적 잠재체류지(R),	-	회사(C), 학교(S)	10~60대	보류
		비정기적 잠재체류지(X)				

* ① 도착지 타입이 '-'인 경우, ② 회사(C)→학교(S) 또는 학교(S)→회사(C) 타입인 경우, ③출발일자와 도착일자가 다른 경우는 제외

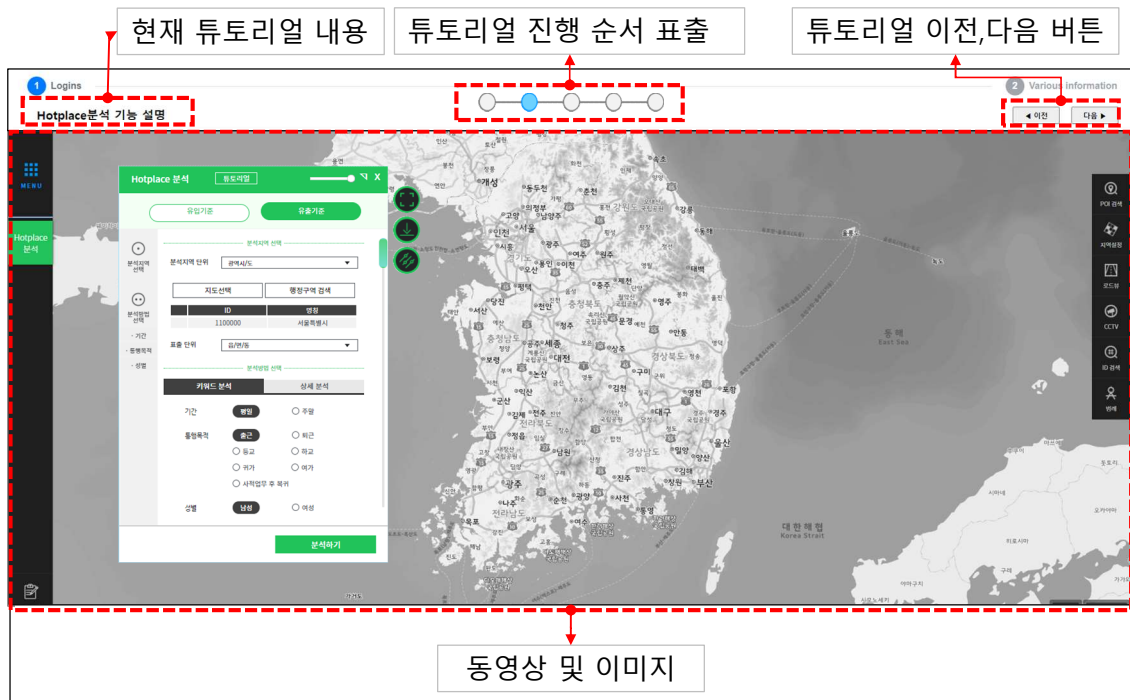
- 필수입력 조건이 아닌 범례 및 표출정의를 별도의 아이콘으로 배치하여 레이아웃의 여유 공간을 확보하고, 요소 전체 UI를 개선해 사용자가 정보를 식별하는데 수월할 수 있도록 개선



<그림 4- 21> 범례설정 UI 변경

나. 분석 기능 튜토리얼 제작

- 통신 주요분석기능의 기능 및 분석 결과 설명, 분석 도구의 사용 방법을 제공
 - 분석도구 사용이 익숙하지 않은 사용자에게 동영상 또는 그림을 통하여 튜토리얼을 제공함으로써 사용 편의를 높이고, 조건 설정 등 분석 과정 및 분석결과에 대한 사용자의 이해도를 높일 수 있도록 함
 - 이전, 다음 버튼을 통하여 사용자가 원하는 튜토리얼을 진행할 수 있도록 구현
 - 상단 중앙에 튜토리얼 진행 순서를 표시하는 상태바 표출
 - 좌측 상단에 현재 진행 중인 튜토리얼의 기능명 표출
 - 중앙에 해당 분석 기능의 사용방법의 이해를 도울 수 있는 동영상 혹은 이미지를 표출



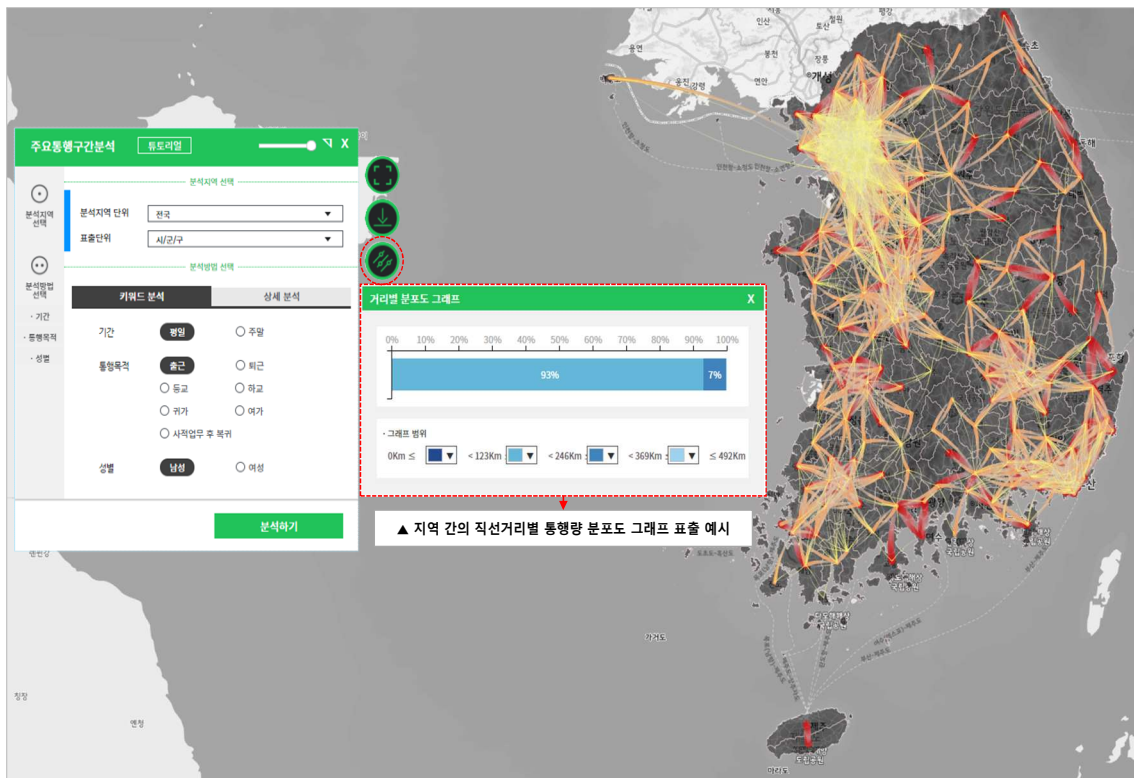
<그림 4- 22> 분석기능 튜토리얼 UI

다. 기존 분석기능 (주요통행구간분석) 개선

- 단거리 구간 통행과 장거리 구간 통행의 분석이 가능하도록 개선
 - 전국의 시도, 시군구 단위 간 구간 분석이 가능하도록 개선
 - 단거리 구간 통행과 장거리 구간 통행 비교를 위한 그래프 표출

<표 4- 8> 주요통행구간 분석 기능의 연도별 표출형태 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 단위선택



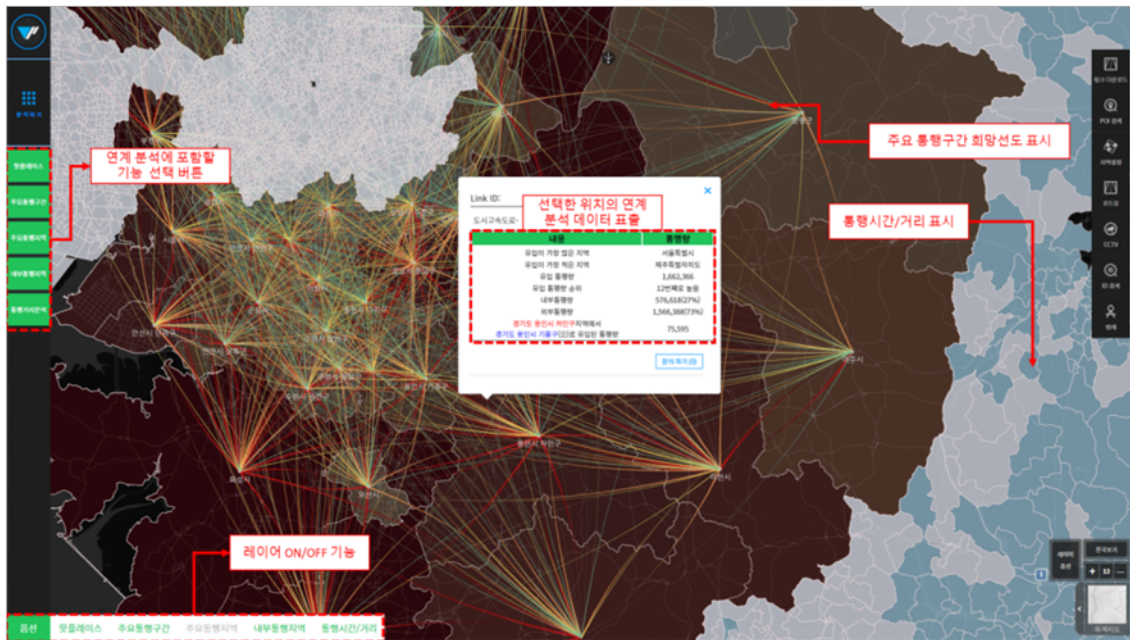
<그림 4- 23> 주요통행구간 분석의 전국분석 결과 및 통행량 분포도 그래프

라. 신규 분석 기능 개발

1) 심층 분석을 가능케 하는 분석 기능 개발

① 정보 연계 분석 기능

- 여러 분석 기능을 선택하고 그에 따른 결과를 동시에 주제도로 표출하여 확인 할 수 있는 기능
 - － 핫플레이스, 주요통행구간, 주요통행지역, 내부통행지역, 통행거리분석을 연계분석 가능함
- 정보 연계 분석 기능 UI
 - － 각 분석 기능을 선택 한 후 분석하기 버튼을 클릭하여 연계분석을 실행함
 - － 주제도상 클릭하여 연계분석된 데이터를 표출함
 - － 하단 옵션 버튼을 통하여 연계분석 기능들의 레이어를 끄거나 켜서 개별 주제도를 볼 수 있음
 - － 희망선도를 통해서 주요 통행구간을 인지 할 수 있음
 - － 희망선도 혹은 지역을 선택하여 해당 위치의 데이터 개괄을 볼 수 있음



<그림 4-24> 정보 연계 분석 기능 화면

② 통행변화 시뮬레이션 기능

- 통행 변화를 시계열로 확인할 수 있는 기능
- 통행변화 시뮬레이션 분석 기능 UI
 - 분석유형(유입, 유출 및 O/D기준), 분석지역 및 분석기간을 설정해 사용자 의도에 맞는 공간데이터를 시각화하여 통행변화의 흐름분석이 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 4- 25> 유입 기준 폴리곤 형태의 통행변화 시뮬레이션

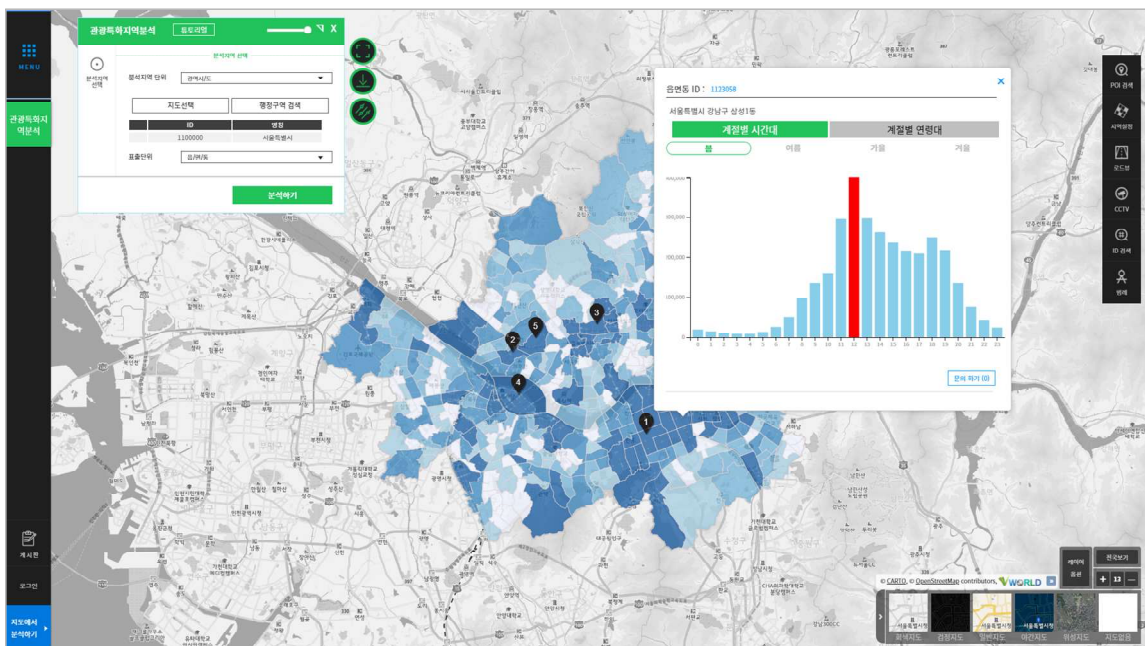


<그림 4- 26> 지역 내 기준 폴리곤 및 희망선 형태의 통행변화 시뮬레이션

2) View-T Light의 분석서비스 중 일부를 Expert에 개발

① 관광 특화지역 분석 기능

- 계절에 따라 통행량 변화가 큰 지역을 확인 할 수 있는 기능
 - 도착지 트립타입이 비정기적 잠재체류지인 통행을 지역별, 계절별로 분류 표시
- 관광 특화지역 분석 기능 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



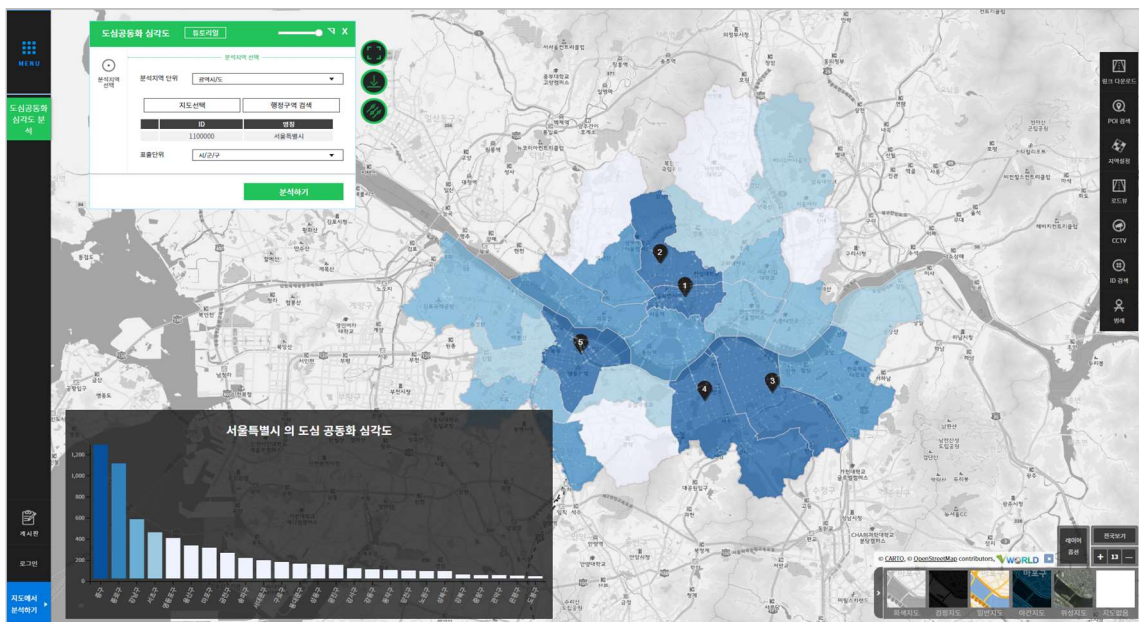
<그림 4-27> 관광 특화지역 분석 기능 화면

② 도심공동화 심각도 분석 기능

- 도시 중심부의 상주인구가 줄어드는 현상을 확인할 수 있는 기능
 - 심야 시간대 집으로의 통행량과 낮 시간대 통행량의 비율을 통한 상주인구 측정

○ 도심공동화 심각도 분석 기능 UI

- 분석지역 단위를 광역시/도, 시/군/구에서 선택
- 분석하기 원하는 지역을 지도선택 혹은 행정구역 검색을 통하여 선택
- 표출단위는 분석지역 단위의 하위 지역으로 고정됨
- 상위5개 지역을 주제도 상에 표시
- 해당 지역의 공동화 심각도를 순위 그래프로 표시



<그림 4- 28> 도심공동화 심각도 분석화면

제3절 기반 DB 현행화 및 속도 개선

1. 모바일 통신 빅데이터 기반 통행정보 DB

가, 개요

1) 구축 대상 (가공 전 데이터 형태)

- 2017년 데이터 수는 1년 기준 약 145억 건이 수집되었으며, 데이터 용량은 약 700GB임. 2018년 데이터는 2017년 대비 전체적으로 약 40% 감소하여 2018년 데이터 건 수는 약 86억 건 수집되었으며, 데이터 용량은 약 430GB임

<표 4- 9> 2017년, 2018년 통신 데이터 현황

구분	2017년 통신 데이터	2018년 통신 데이터
수집 지역	전국	
수집 기간	2017년도 1년 (365일)	2018년도 1년 (365일)
데이터 수	약 145억 건	약 86억 건
데이터 크기	약 700GB	약 430GB
데이터 유형	일자별 csv 파일 (년월일(8자리)).csv	

<표 4- 10> 모바일 통신 빅데이터 기반 통행정보 DB 컬럼명 및 데이터 타입

NO	필드명	설명	데이터 타입	비고
1	o_polygon	출발지 (교통폴리곤 코드)	STRING	시도코드2자리 + 일련번호 4자리
2	o_base_ymd	출발일자	STRING	연월일 (YYYYMMDD)
3	o_timezn_cd	출발시간대	STRING	0~23시
4	o_trip_type	출발체류특성	STRING	4가지 (H: 집, C: 회사, S: 학교, R: 기타)
5	d_polygon	도착지 (교통폴리곤 코드)	STRING	시도코드2자리 + 일련번호 4자리

6	d_base_ymd	도착일자	STRING	연월일 (YYYYMMDD)
7	d_timezn_cd	도착시간대	STRING	0~23시
8	d_trip_type	도착체류특성	STRING	5가지 (H: 집, C: 회사, S: 학교, R: 기타, -:이동정보 없음)
9	age_itg_cd	연령대	INT	0~110세까지 10세 단위로 구성 (예: 10은 10~19세를 의미)
10	sex_trip_itg_cd	성별	STRING	2가지 (M: 남성, F: 여성)
11	total	통행량	INT	1~2통행 미만인 경우 3통행으로 표기
12	trv_time	평균통행시간	FLOAT	-

2) DB 구축 방안

- 검색·표출 시간을 단축할 수 있도록 집계 형태의 DB를 별도로 구축
 - 통행 위치 정보(교통폴리곤 단위, 행정구역 단위)와 사용자가 자주 검색하는 분석 조건을 기준으로 통행 정보를 집계
 - 가공 전 데이터는 약 145억 건으로 별도의 가공 없이 View-T의 서비스를 하게 될 경우 1시간 이상의 처리 시간이 소요되기 때문

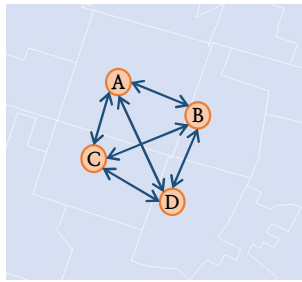
3) DB 구축 절차

- (1단계) 데이터 전처리 및 검증
 - 교통폴리곤이 속하는 행정구역 정보를 매칭하여 출발지·도착지 위치 정보를 추가 생성
 - 3통행 미만 보정계수를 적용하여 과다하게 집계되던 통행량을 보정
 - 원천데이터에서 기종점 통행량 정보로 가공할 때 개인정보보호법에 따라 3통행 미만인 경우에는 모두 3통행으로 처리하여 실제 통행량보다 높게 집계되는 문제 발생
 - 검색 쿼리와 DB 적재 용량을 확보하기 위해 문자형 데이터 형식을 숫자형 데이터 형식으로 변환
 - 분석 서비스에 활용하기 위해 출발 교통폴리곤과 도착 교통폴리곤 간의 직선거리를 생성

데이터 형식 변환 예시

기존	변환후	비고
H	11	집
C	12	회사
S	13	학교
R	14	종교집회장소
X	15	기타 잠재체류지
F	1	여자
M	2	남자

거리 데이터 산출 예시



출발 폴리곤	도착 폴리곤	거리(m)
A	B	100
A	C	150
A	D	200
B	C	200
B	D	120
C	D	130

<그림 4- 29> 통신데이터의 형식변환 및 거리산출

- JAVA 기반의 검증시스템을 개발하여 정보가 누락된 데이터에 대해 오류 데이터를 생성
- o (2단계) 교통폴리곤별 통행정보 집계 후 인덱스 생성
 - 출발 교통폴리곤ID, 도착 교통폴리곤ID, 체류특성, 연령대, 성별 필드를 키(Key)로 구성하고 키 값이 동일한 데이터들을 그룹화

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

<그림 4- 30> 교통폴리곤 기준 통행량 데이터 구조

- o (3단계) 행정구역(읍면동, 시군구, 시도) 단위로 통행정보 재집계
 - 1단계에서 매칭한 행정구역 정보를 활용하여 출발 행정구역 기준 집계 데이터, 도착 행정구역 기준 집계 데이터, 출발과 도착이 모두 행정구역 단위인 집계 데이터를 구축

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

<그림 4- 31> 행정구역 단위의 출발/도착 기준 통행량 DB 구조

- (4단계) 사용자가 자주 검색하는 분석 조건(예: 오전첨두, 주말 등)을 기준으로 통행 정보를 집계

분류	옵션	조건		
분석 기간	주중	월, 화, 수, 목, 금		
	주말	토, 일		
통행 목적	출근	집 - 회사	06시~09시 - 07시~10시	20대 ~ 60대
	등교	집 - 학교	06시~09시 - 07시~10시	10대 ~ 20대
	퇴근	회사 - 집	18시~21시 - 19시~22시	20대 ~ 60대
	하교	학교 - 집	15시~21시 - 16시~21시	10대 ~ 20대
	귀가	정기적, 비정기적 잠재체류지 - 집	18시~20시 - 19시~23시	전체 연령대
	여가	집, 회사, 학교, 정기적, 비정기적 잠재체류지 - 정기적, 비정기적 잠재체류지	전체 시간대	전체 연령대
	사적업무 후 복귀	정기적, 비정기적 잠재체류지 - 회사, 학교	전체 시간대	10대 ~ 60대
	* ① 도착지 타입이 '-'인 경우, ② 회사(C)→학교(S) 또는 학교(S)→회사(C) 타입인 경우, ③출발일자과 도착일자가 다른 경우는 제외			
성별	남자	-		
	여자	-		

<그림 4- 32> 통행목적별 집계 통행량 데이터 조건

나. DB 구축 결과

1) 교통폴리곤 기준 집계 통행량 DB

- 교통폴리곤 단위별 출발/도착에 대한 통행량을 집계하여 저장해 놓은 데이터
 - _id 컬럼은 출발/도착 교통폴리곤 ID, 체류특성, 연령대, 성별의 값을 구성된 기준키이며 대용량 데이터의 건수를 줄이는 데 사용됨
 - CG 컬럼은 _id 컬럼의 값과 동일한 데이터의 출발 시간대와 도착 시간대의 값으로

그룹을 구성하여 모아 놓았으며 시간대별 통행량을 그룹화 하는 데 사용됨

- K 컬럼은 _id 컬럼, CG 컬럼의 값과 동일한 데이터의 기준일과 통행량의 값을 그룹으로 구성하여 모아 놓았으며 기준일에 해당되는 통행량을 구하는 데 사용됨

<표 4-11> 교통폴리곤 기준 통행량 DB 테이블 정의서

No	Column	Type	구성	비고
1	_id	Object	출발 교통폴리곤 ID	6자리
			도착 교통폴리곤 ID	6자리
			출발 체류특성	2016년 : 1 ~ 3
			도착 체류특성	2017년 : 11 ~ 16
			연령대	2018년 : 1 ~ 4
			성별	0 ~ 110 (10단위)
2	CG	Object Array	출발 시간대	1, 2
			도착 시간대	hh
			K Column	-
3	K	Object Array	기준일	2017년, 2018년 출발일자, 도착일자
			통행량	-
4	A1	Integer	출발 읍면동 ID	7자리
5	A2	Integer	출발 시군구 ID	7자리
6	A3	Integer	출발 시도 ID	7자리
7	E1	Integer	도착 읍면동 ID	7자리
8	E2	Integer	도착 시군구 ID	7자리
9	E3	Integer	도착 시도 ID	7자리

<표 4-12> 교통폴리곤 기준 집계 통행량 DB 구축 효과 (도큐먼트 수 변화)

구분	가공 전	가공 후	감소율
교통폴리곤 기준 집계	약 145억 건	약 9억 건	약 94%

2) 행정구역 단위의 통행량 DB

- 행정구역 단위별 출발/도착에 대한 통행량을 집계하여 저장해 놓은 데이터
- json 데이터 형식이며 다양한 출발 행정구역, 도착 행정구역에 대한 집계를 사용하는 서비스에서 사용됨

<표 4-13> 행정구역 단위의 출발/도착 기준 통행량 DB 테이블 정의서

No	Column	Type	구성		비고
			출발기준	도착기준	
1	_id	Object	출발 교통폴리곤 ID	출발 행정구역 ID	-
			도착 행정구역 ID	도착 교통폴리곤 ID	-
			출발 체류특성		2016년 : 1 ~ 3
			도착 체류특성		2017년 : 11 ~ 16 2018년 : 1 ~ 4
			연령대		0 ~ 110 (10단위)
			성별		1, 2
2	CG	Object Array	출발 시간대		hh
			도착 시간대		hh
			K Column		-
3	K	Object Array	기준일		2017년, 2018년 출발일자, 도착일자
			통행량		-
4	A1	Integer	출발 읍면동 ID		행정구역 단위 별 출발/도착 기준에 따라 포함 여부 변경
5	A2	Integer	출발 시군구 ID		
6	A3	Integer	출발 시도 ID		
7	E1	Integer	도착 읍면동 ID		
8	E2	Integer	도착 시군구 ID		
9	E3	Integer	도착 시도 ID		

<표 4- 14> 행정구역 기준 집계 통행량 DB 구축 효과

구분	가공 전	가공 후	감소율
읍면동 기준 집계	약 9억 건	약 3억 건	약 30%
시군구 기준 집계	약 9억 건	약 1,300만 건	약 98%
시도 기준 집계	약 9억 건	약 12만 건	약 99%

② 출발과 도착 모두 행정구역 단위인 통행량 DB

- 시도/시군구/읍면동별 출발과 도착이 모두 행정구역 단위인 통행량을 집계하여 저장해 놓은 데이터
- json 데이터 형식이며 서비스 중 선택 지역 단위와 분석 지역 단위가 같은 내부통행지역분석 및 지역간 OD 통행량 데이터 다운로드에서 사용됨

<표 4- 15> 출발과 도착 모두 행정구역 단위인 통행량 DB 테이블 정의서

No	Column	Type	구성	비고
1	_id	Object	출발 행정구역 ID	시도/시군구/읍면동
			도착 행정구역 ID	시도/시군구/읍면동
			출발 체류특성	2016년 : 1 ~ 3
			도착 체류특성	2017년 : 11 ~ 16
				2018년 : 1 ~ 4
			연령대	0 ~ 110 (10단위)
2	CG	Object Array	성별	1, 2
			출발 시간대	hh
			도착 시간대	hh
3	K	Object Array	K Column	-
			기준일	2017년, 2018년
			통행량	출발일자, 도착일자
5	A2	Integer	출발 시군구 ID	행정구역 단위에 따라 포함 여부 변경
6	A3	Integer	출발 시도 ID	
8	E2	Integer	도착 시군구 ID	
9	E3	Integer	도착 시도 ID	

<표 4-16> 출발과 도착 모두 행정구역 단위인 집계 통행량 DB 구축 효과

구분	가공 전	가공 후	감소율
읍면동 기준 집계	약 9억 건	약 3억 건	약 30%
시군구 기준 집계	약 9억 건	약 1,300만 건	약 98%
시도 기준 집계	약 9억 건	약 12만 건	약 99%

3) 분석특성별(통행목적별) 집계 통행량 DB

- 교통폴리곤 기준 통행량 DB를 통행목적 조건에 부합하는 데이터를 압축(경량화)한 데이터
- 교통폴리곤 기준 통행량 DB에서 키로 사용되는 _id 컬럼과 출발/도착 행정구역 ID 컬럼은 동일하며 Object Array 컬럼에 해당하는 출발/도착 시간대와 출발/도착 일자가 압축된 형태로 구성됨
- 세밀하게 조건을 설정하는 심층 분석의 경우와 다르게 사용자가 많은 조건을 설정하지 않고 자주 분석하는 조건을 선택하여 분석하는 일반 분석에서 활용

<표 4-17> 통행목적별 집계 통행량 데이터 DB 테이블 정의서

Column	Type	구성	비고
_id	Object	출발 읍면동	통행목적 1 : 출근 2 : 등교 3 : 퇴근 4 : 하교 5 : 귀가 6 : 여가 7 : 사적업무 후 복귀
		도착 읍면동	
		성별	
		통행목적	
WD	int	평일 집계 통행량	-
WE	int	주말 집계 통행량	-
A2	int	출발 시군구	-
E2	int	도착 시군구	-
A3	int	출발 시도	-
E3	int	도착 시도	-
S	double	출발 읍면동 - 도착 읍면동 거리	-

<표 4- 18> 분석특성별(통행목적별) 집계 통행량 DB 구축 효과

구분	가공 전	가공 후	감소율
통행목적 기준 집계	약 9억 건	약 1.7억 건	약 79%

다. DB 검증

1) 교통폴리곤 기준 통행량 DB 검증

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

<표 4- 19> 교통폴리곤 기준 통행량 DB 검증 항목 예시

검 증 항 목	비 고
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	구축되어진 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	

❖ 교통폴리곤 기준 통행량 DB 검증 쿼리 예시

```

1 db.TMOBILE_2017.aggregate( [
2 {
3   $match:{
4     $and : [
5       {"A3": {"$gt" : 29000000 }},
6       {"_id.D": 11 },
7       {"_id.H": 13 },
8       {"_id.I": {"$in" : [20,30,40,50,60] }},
9       {"$expr": {"$ne" : ["$_id.A", "$_id.E"] }},
10    ]
11  },
12 },
13 {
14   $unwind: "$CG"
15 },
16 {
17   $unwind: "$CG.K"
18 },
19 {
20   $match:{
21     $and : [
22       {"CG.C": {"$in" : [6,7,8,9] }},
23       {"CG.G": {"$in" : [7,8,9,10] }},
24       {"$expr": {"$eq" : ["$CG.K.SD", "$CG.K.ED"] }},
25     ]
26   },
27 },
28 {
29   $group : {
30     "_id" : "$E3",
31     "SUM" : {"$sum": "$CG.K.V"}
32   },
33 },
34 { $limit : 100}
35 ],{allowDiskUse : true})

```

❖ 쿼리 결과 예시

```

1 /* 1 */
2 {
3   "_id" : 1100000,
4   "SUM" : 25832885
5 },
6
7 /* 2 */
8 {
9   "_id" : 3700000,
10  "SUM" : 19676116
11 },
12
13 /* 3 */
14 {
15   "_id" : 3300000,
16   "SUM" : 12848713
17 },
18
19 /* 4 */
20 {
21   "_id" : 3100000,
22   "SUM" : 86253471
23 },
24
25 /* 5 */

```

<그림 4- 33> 교통폴리곤 기준 통행량 DB 검증 쿼리와 결과 예시(시도별 통행량 확인)

교통폴리곤 단위의 통행량 DB 검증					
출/도착	행정구역 단위	행정구역 ID	통신데이터 전처리 DB	교통폴리곤 단위의 통행량 DB	통행량 차
출발	교통폴리곤	110006	178,364	178,364	0
	교통폴리곤	110054	1,387,766	1,387,766	0
	교통폴리곤	250279	2,325,815	2,325,815	0
	교통폴리곤	310527	1,470,119	1,470,119	0
	교통폴리곤	311498	279,547	279,547	0
	시도	3100000	3,309,923,708	3,309,923,708	0
	시도	2900000	78,763,333	78,763,333	0
도착	교통폴리곤	220500	896,918	896,918	0
	교통폴리곤	330013	145,228	145,228	0
	교통폴리곤	380962	216,664	216,664	0
	교통폴리곤	381014	348,031	348,031	0
	교통폴리곤	390324	960,493	960,493	0
	시도	1100000	3,224,194,723	3,224,194,723	0
	시도	2500000	506,762,799	506,762,799	0

<그림 4- 34> 통신데이터 전처리DB와 교통폴리곤 단위 집계 통행량 DB 비교검증 결과

2) 행정구역 단위 통행량 DB 검증

- 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 4-20> 행정구역 단위 통행량 합계 비교 검증 항목

검 증 항 목	비 고
행정구역이 매칭된 통신 데이터의 합계와 시도 단위 통행량 합계	출발/도착 각각의 값으로 비교 검증
행정구역이 매칭된 통신 데이터의 합계와 시군구 단위 통행량 합계	
행정구역이 매칭된 통신 데이터의 합계와 읍면동 단위 통행량 합계	

읍면동 단위의 출발 기준 통행량 DB 검증					
출/도착	행정구역 단위	행정구역 ID	교통폴리곤 단위의 통행량 DB	읍면동 단위의 출발 기준 통행량 DB	통행량 차
출발	읍면동	1101056	3,530,070	3,530,070	0
	읍면동	3109158	5,243,365	5,243,365	0
	시군구	1107000	90,828,601	90,828,601	0
	시군구	3113000	148,604,202	148,604,202	0
	시도	1100000	3,226,649,521	3,226,649,521	0
	시도	3100000	3,309,923,708	3,309,923,708	0
도착	교통폴리곤	320006	103,813	103,813	0
	교통폴리곤	330277	897,396	897,396	0
	읍면동	3232011	5,746,120	5,746,120	0
	읍면동	3338031	643,991	643,991	0
	시군구	3233000	8,659,057	8,659,057	0
	시군구	3304200	67,805,452	67,805,452	0
	시도	3200000	511,558,046	511,558,046	0
	시도	3300000	478,665,927	478,665,927	0

<그림 4-35> 교통폴리곤 단위 집계 통행량 DB와 읍면동 단위 집계 통행량 DB 비교검증 결과

시군구 단위의 출발 기준 통행량 DB 검증					
출/도착	행정구역 단위	행정구역 ID	읍면동 단위의 출발 기준 통행량 DB	시군구 단위의 출발 기준 통행량 DB	통행량 차
출발	시군구	1101000	125,665,140	125,665,140	0
	시군구	3105000	223,565,282	223,565,282	0
	시도	1100000	3,226,649,521	3,226,649,521	0
	시도	3100000	3,309,923,708	3,309,923,708	0
도착	교통폴리곤	210076	112,401	112,401	0
	교통폴리곤	341254	206,280	206,280	0
	읍면동	2113051	8,981,015	8,981,015	0
	읍면동	3435035	331,064	331,064	0
	시군구	2106000	75,557,531	75,557,531	0
	시군구	3433000	21,162,535	21,162,535	0
	시도	2100000	1,047,125,776	1,047,125,776	0
	시도	3400000	618,207,761	618,207,761	0

<그림 4- 36> 읍면동 단위 집계 통행량 DB와 시군구 단위 집계 통행량 DB 비교검증 결과

시도 단위 통행량 DB(출/도착이 시도 단위) 검증					
유형	행정구역 단위	행정구역 ID	시도 단위의 출발 기준 통행량 DB	출발/도착이 모두 시도 단위인 통행량 DB	통행량 차
출발	시도	3100000	3,309,923,708	3,309,923,708	0
	시도	3900000	276,386,268	276,386,268	0
도착	시도	1100000	3,224,194,723	3,224,194,723	0
	시도	2600000	349,364,773	349,364,773	0
출발 체류특성	집	3100000	981,980,740	981,980,740	0
	집 이외 주체류지	3100000	141,782,187	141,782,187	0
	회사	3100000	359,385,379	359,385,379	0
	학교	3100000	70,579,025	70,579,025	0
	종교집회장소	3100000	336,461,388	336,461,388	0
	기타	3100000	1,419,734,989	1,419,734,989	0

<그림 4- 37> 시도 단위 출발 기준 집계 통행량 DB와 출발/도착이 모두 시도 단위인 집계 통행량 DB 비교검증 결과

1 db.TMOBILE_POLY_UMD_2017_umdupdate.aggregate([

2 {

3 \$match:{

4 \$and : [

5 {"A3": 3900000},

6 {"_id.D": {"\$in" : [15,16]} },

7 {"_id.H": 11 },

8 {"\$expr": {"\$ne" : ["\$A1", "\$_id.E1"] } },

9]

10 }

11 },

12 {

13 \$unwind: "\$CG"

14 },

15 {

16 \$unwind: "\$CG.K"

17 },

18 {

19 \$match:{

20 \$and : [

21 {"CG.C" : {"\$in" : [18,19,20]} },

22 {"CG.G" : {"\$in" : [19,20,21,22,23]} },

23 {"\$expr": {"\$eq" : ["\$CG.K.SD", "\$CG.K.ED"] } },

24]

25 }

26 },

27 {

28 \$group : {

29 "_id" : {A1 : "\$A1", "I" : "\$_id.I"},

30 SUM : {\$sum:"\$CG.K.V"}

31 }

32 },

33 {\$limit : 10}

34],{allowDiskUse : true})

TMOBILE_POLY_UMD_2017_umdupdate 11.460 s 10 Docs

Key	Value	Type
(1) { A1 : 3901057, I : 20 }	{ SUM : 9870 }	Document
id	{ A1 : 3901057, I : 20 }	Object
A1	3901057	Int32
I	20	Int32
SUM	9870	Int32
(2) { A1 : 3902053, I : 20 }	{ SUM : 12244 }	Document
(3) { A1 : 3902053, I : 50 }	{ SUM : 22482 }	Document
(4) { A1 : 3901067, I : 80 }	{ SUM : 563 }	Document

<그림 4- 38> 행정구역 단위 집계 통행량 DB 검증 쿼리와 결과 예시

3) 분석특성별(통행목적별) 통행량 DB 검증

- 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 동일할 때 까지 진행
- 통행목적별 집계 통행량 DB 검증 항목은 아래와 같음

<표 4-21> 통행목적별 집계 통행량 합계 비교 검증 항목

비교 대상	비교 조건	비고
교통폴리곤 기준 통행량 데이터	- 주중/주말 조건 통행량 합계 - 통행목적 조건 통행량 합계 성별 통행량 합계	출발/도착 각각의 값으로 비교 검증
행정구역 단위 출발/도착 통행량 데이터		

○ 통행목적별 집계 통행량과 교통폴리곤 기준 통행량 DB 검증 수행 결과

교통폴리곤 단위로 집계한 통행량 DB							
구분	주말-출근	주말-등교	주말-퇴근	주말-하교	주말-귀가	주말-여가	주말-사적업무후복귀
1100000	8,959,808	1,203,952	5,953,004	1,236,161	33,090,751	368,999,783	16,060,904
2100000	3,516,447	459,781	1,937,408	446,397	10,638,770	128,724,018	5,922,820
2200000	2,995,862	382,936	1,464,160	346,306	8,068,478	97,415,404	4,678,168
2300000	2,833,069	331,225	1,370,768	306,616	7,282,260	88,005,438	4,003,205
2400000	1,635,160	226,061	838,248	225,849	4,753,812	59,699,463	2,747,174

통행목적별 집계 통행량 DB							
구분	주말-출근	주말-등교	주말-퇴근	주말-하교	주말-귀가	주말-여가	주말-사적업무후복귀
1100000	8,959,808	1,203,952	5,953,004	1,236,161	33,090,751	368,999,783	16,060,904
2100000	3,516,447	459,781	1,937,408	446,397	10,638,770	128,724,018	5,922,820
2200000	2,995,862	382,936	1,464,160	346,306	8,068,478	97,415,404	4,678,168
2300000	2,833,069	331,225	1,370,768	306,616	7,282,260	88,005,438	4,003,205
2400000	1,635,160	226,061	838,248	225,849	4,753,812	59,699,463	2,747,174

집계 통행량 차이							
구분	주말-출근	주말-등교	주말-퇴근	주말-하교	주말-귀가	주말-여가	주말-사적업무후복귀
1100000	0	0	0	0	0	0	0
2100000	0	0	0	0	0	0	0
2200000	0	0	0	0	0	0	0
2300000	0	0	0	0	0	0	0
2400000	0	0	0	0	0	0	0


검증 결과 집계 통행량 차이가 '0'이면 검증 완료

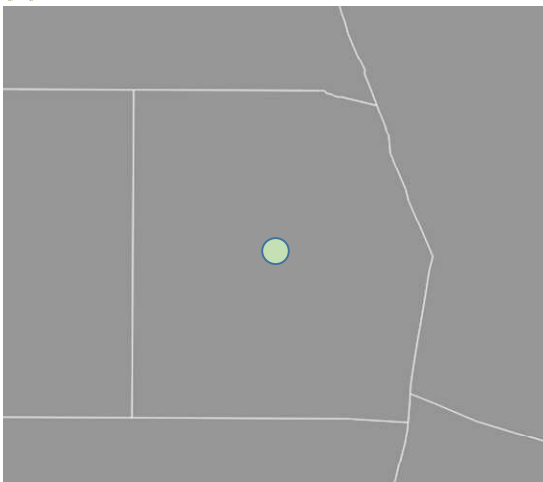
<그림 4-39> 교통폴리곤 단위 집계 통행량 DB와 통행목적별 집계 통행량 DB 비교검증 결과(2017년 기준)

2. 모바일 분석 맵 DB

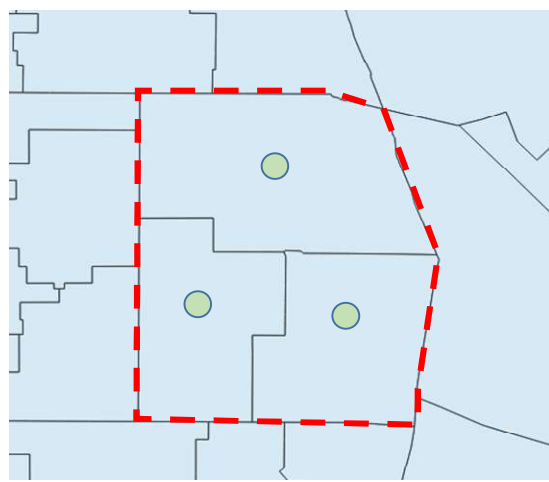
가. 개요

- 교통폴리곤, 읍면동, 시군구, 시도 총 4가지 형태의 맵을 구축
 - 교통폴리곤(2018년 기준): 기지국 수신범위(셀 반경)와 집계구를 매칭시켜 모바일 통신 빅데이터 기반 통행 정보(통행량)를 집계하는데 용이한 형태로 구축된 맵 (읍면동의 1/3~1/4 크기로 분할되어 있는 영역)
 - 읍면동, 시군구, 시도(2017년 기준): 시도 17개, 시군구 250개, 읍면동 3,500개로 구성

 읍면동과 교통폴리곤 형상 비교 예시 : 대전시 둔산동



<읍면동 행정구역 형상정보>



<교통폴리곤 형상정보>

<그림 4- 40> 교통폴리곤의 형상정보

<표 4- 22> 교통폴리곤 레이어 필드

컬럼명	타입	내용	비고
sido	TEXT	시도 ID	2자리의 시도 ID
emd	TEXT	읍면동 ID	8자리의 읍면동 ID
polycode	TEXT	교통폴리곤 ID	6자리의 교통폴리곤 ID
area	DOUBLE	면적	교통폴리곤의 면적

나. 구축절차

○ 데이터 가공 및 검수

- 데이터 수급 이후 해당 데이터에 대한 검수 작업을 진행함

. 형상 검수 해당 데이터에 대한 홀 폴리곤 및 이상 형상 또는 좌표계 등의 오류에 대한 검수 작업을 진행

. 속성 데이터의 행정구역 정보와 형상과의 매칭 데이터가 정확히 매칭 되어 있는지를 공간 조인 방식으로 검수

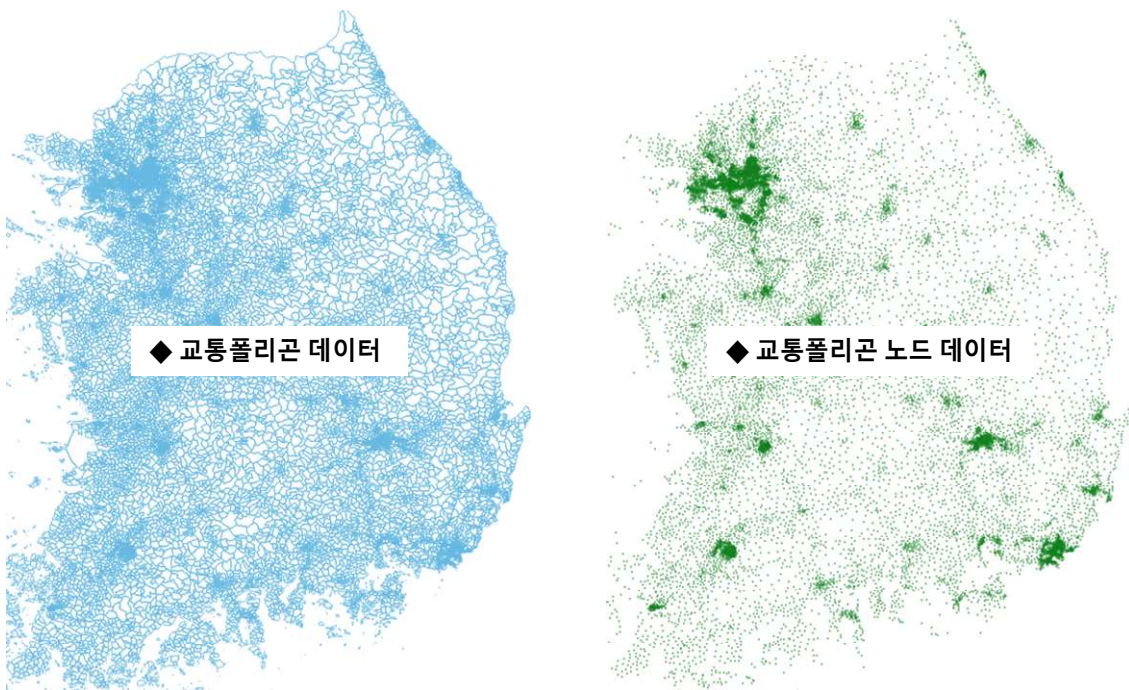
○ 교통폴리곤의 대표 노드 데이터 생성

- 각 데이터 표출을 위한 교통 폴리곤의 대표 포인트(노드)를 구축함

- 대표 포인트는 상용 GIS 툴을 이용하여 해당 폴리곤의 중심점 좌표를 생성 함

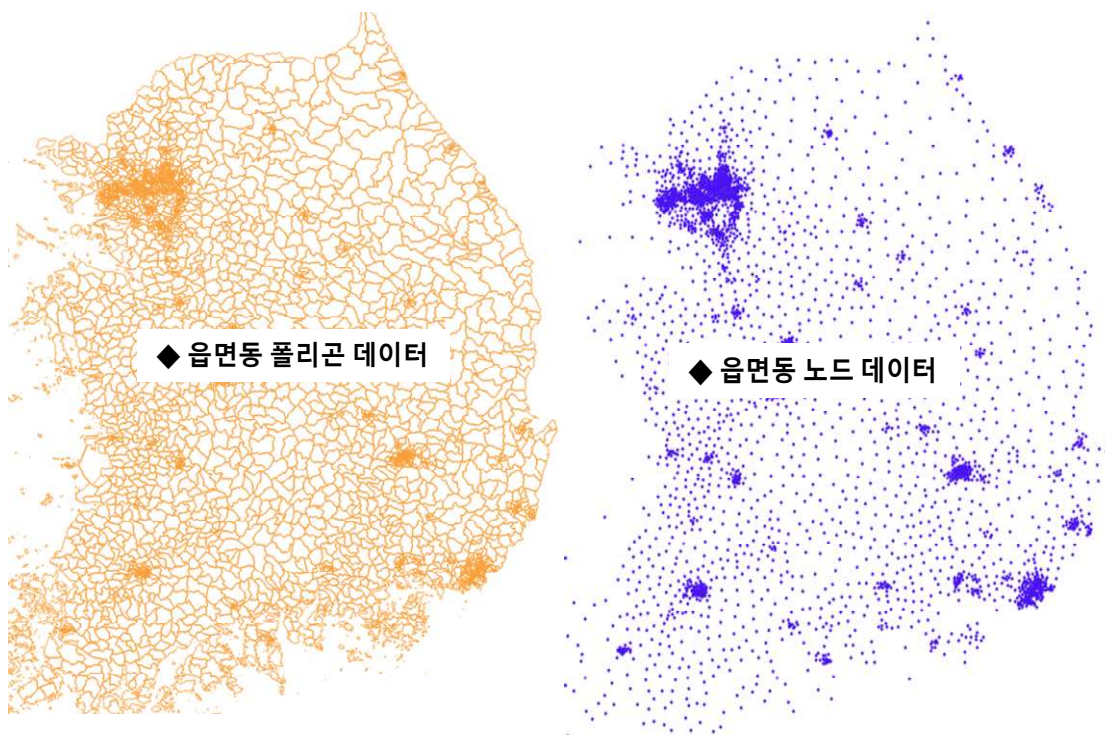
- GIS 툴을 이용하여 해당 폴리곤의 중심점 좌표를 생성할 경우 일부 형상이 복잡한 교통폴리곤의 경우 해당 폴리곤 외부에 대표 노드가 생성됨. 이러한 데이터들은 수작업으로 해당 노드들의 위치를 보정하여 차후 사용 할 수 있도록 구축함

- 분석 맵 노드 이외에 각 행정구역에 해당하는 노드도 같은 방식으로 대표 노드를 생성함



<그림 4- 41> 교통폴리곤 형상 데이터



- GIS 서비스를 위한 행정구역 정보 및 가공
 - 교통 폴리곤 구축 시 활용된 집계구 경계 기준연도와 동일한 행정구역을 수급 후 진행
 - GeoServer의 VectorTile, WMS 서비스 형태로 시도, 시군구, 읍면동 행정구역 폴리곤 가공
- 행정구역 노드 레이어 구축
 - 각 기반정보 데이터 표출 기능에서 사용되는 행정구역 폴리곤의 대표 포인트(노드)를 구축
 - 교통폴리곤과 같이 노드는 상용 GIS 툴 또는 GeoSever가 지원하는 기능을 활용하여 구축
 - 형상이 복잡한 행정구역의 경우, 범위를 벗어난 노드가 발생하지 않도록 데이터를 검수/수정



<그림 4- 42> 읍면동 형상 데이터

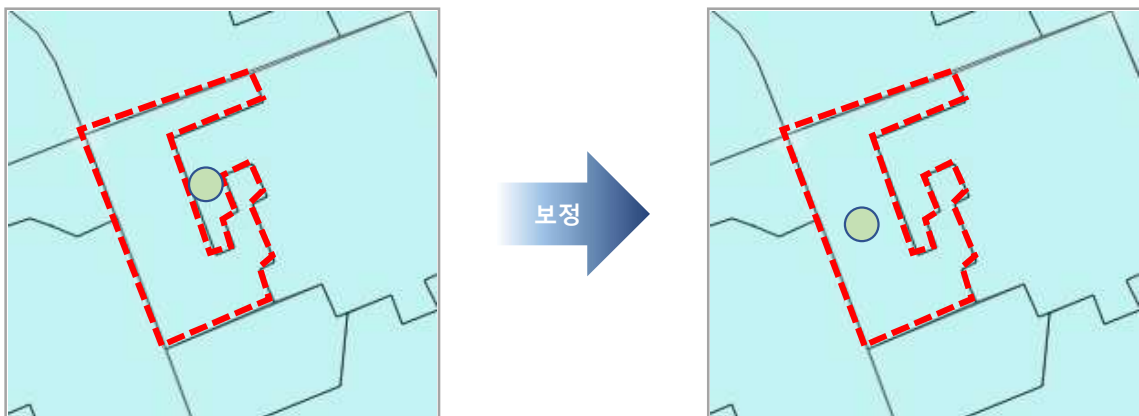
- VectorTile, WMS의 지도 일반화
 - GeoServer의 VectorTile, WMS 서비스는 지도의 뷰 레벨에 따라 최적화된 형상의 이미지를 제공하기 위한 지도 일반화 작업을 진행

<표 4- 23> VectorTile, WMS의 지도 일반화

구분	일반화 적용 전	일반화 적용 후
분석구 (Polygon)		
비고	일반화 적용 전에는 곡선부분 등에서 절점이 중복되거나 가까이 위치하여 데이터의 크기 증가	데이터의 형상은 최대한 유지하면서 절점의 개수를 줄여 데이터 크기 감소 및 그리기 속도 향상

○ 교통폴리곤/행정구역 대표 노드 위치 보정

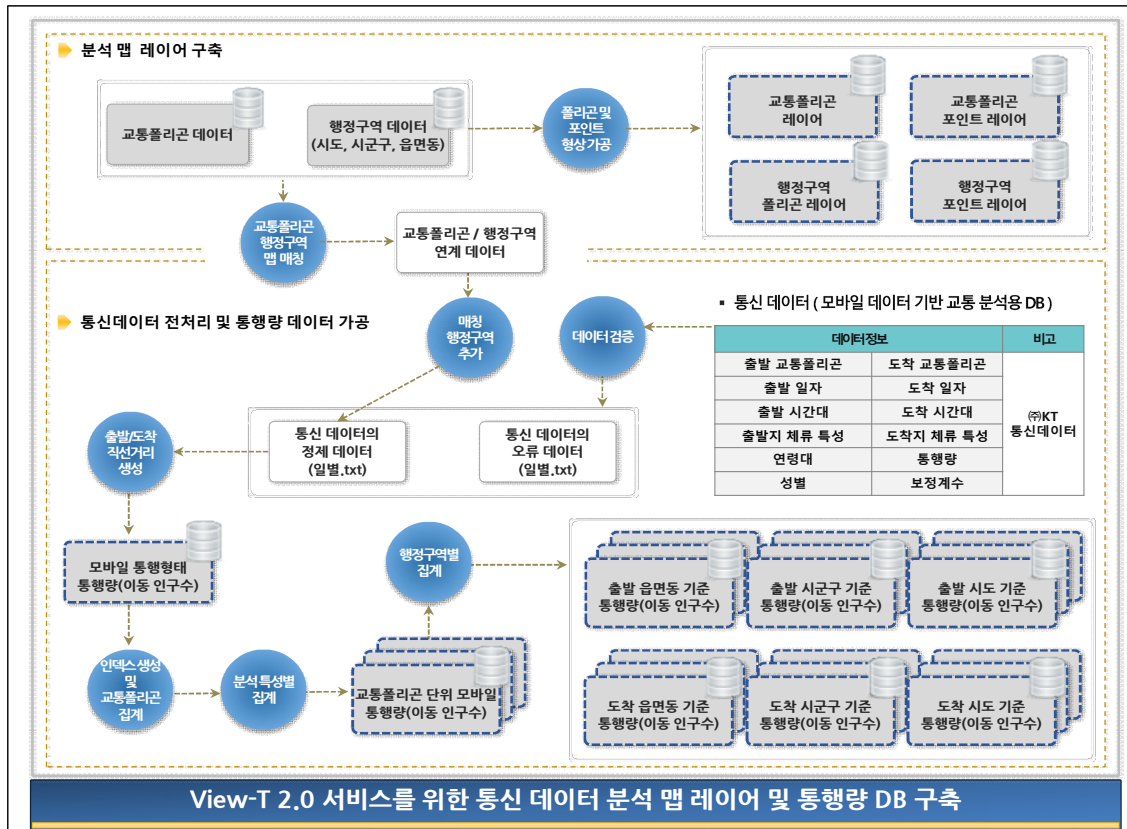
- 교통폴리곤과 행정구역의 대표 노드 구축 시 복잡한 형상은 수작업으로 노드의 위치 보정 후 구축을 진행



<그림 4- 43> 노드 위치 보정 작업 예시

3. 기반 DB 구축 과정

- 앞서 설명한 모바일 통신 빅데이터 기반 통행정보 DB와 모바일 분석 맵 DB를 구축하기 위한 전반적인 과정은 다음과 같음



<그림 4- 44> 기반 DB 구축 과정

- 데이터 구축은 크게 모바일 통신 빅데이터 기반 통행정보 DB(통신 데이터 전처리와 통신 통행량 데이터가공)과 모바일 분석 맵 DB(분석 맵 레이어 구축)로 크게 구분함
- 모바일 데이터 기반의 교통 분석용 데이터를 그대로 서비스 시 많은 데이터를 가져오는 데에 발생하는 처리 속도 이슈로 인하여 데이터 그룹화를 통해 조회하는 데이터 건수를 감소시켜 속도를 개선함
 - 통신 데이터 전처리는 교통 폴리곤 데이터와 행정구역 데이터를 이용해 교통 폴리곤과 행정구역의 연계 데이터를 생성하는 작업과 통신 데이터의 검증, 통신 데이터의 교통 폴리곤 코드와 읍면동, 시군구, 시도 매칭 및 출발 교통 폴리곤과 도착 교통 폴리곤의 대표 노드 간 직선거리 코드를 추가하는 작업을 진행함

- 통신 통행량 데이터가공은 행정구역이 매칭된 통신 데이터를 이용해 인덱스 생성 및 교통 폴리곤별 그룹화 작업을 통해 데이터 건수를 줄여 교통 폴리곤 기준 통행량 구축과 교통 폴리곤 기준 통행량을 이용한 각 행정구역 단위의 출발/도착 기준 집계 통행량 데이터를 구축하는 작업을 진행함
- 서비스 만족도를 높이기 위하여 검색, 표출 속도를 개선하기 위한 작업으로 교통 폴리곤과 행정구역 단위 출발/도착 기준 통행량 데이터를 통행목적별로 재집계하여 데이터 건수를 간추려 구축함
- 사용자가 자주 검색하는 분석 조건(예: 오전/오후첨두, 평일/주말, 출근/퇴근 등)을 기준으로 통행 정보를 집계함
- o 분석 맵 레이어 구축은 교통 폴리곤 데이터와 행정구역 데이터를 이용해 교통 폴리곤과 행정구역의 폴리곤 형상 레이어를 경량화하는 작업과 중심점을 생성하여 구축하는 포인트 레이어 구축 작업을 진행함

제4절 데이터 다운로드 서비스 개선

- 데이터 제공처(통신업체), 데이터 수요처(국민, 민간기업 등)의 요구를 반영하여 다운로드 서비스를 “공공용”, “인가용”, “주문형” 세 가지로 구분
 - 기존에는 통신업체에서 판매하고 있는 데이터와 View-T를 통해 제공하는 데이터의 시·공간적 위계가 유사하여, 사용자에게 데이터 제공시 통신업체와 별도의 협의가 필요하다는 것과, 협의가 완료되는 시점까지 사용자(신청자)는 무한정 대기해야 한다는 문제가 존재

1. 공공용 다운로드 서비스

- 시/도 단위의 통행 정보가 담긴 데이터를 다운로드 받아볼 수 있는 서비스
- 신청 즉시 다운로드 가능
 - 데이터 신청서 작성(기본 정보만 작성) → 데이터 다운로드

<표 4-24> 공공용 다운로드 서비스에서 제공되는 데이터 형태

시도 (17개)	월 (12개)	요일 (2개)	성 (2개)	연령 (4개)	통행목적 (3개)
서울 부산 대구 인천 광주 대전 울산 세종 경기 강원 충북 충남 전북 전남 경북 경남 제주	1월 2월 3월 4월 5월 6월 7월 8월 9월 10월 11월 12월	평일 주말	남성 여성	20세 미만 20세 이상 40세 미만 40세 이상 60세 미만 60세 이상	통근 통학 기타

2. 인가용 다운로드 서비스

- 시군구 단위의 통행 정보가 담긴 데이터를 다운로드 받아볼 수 있는 서비스
- 데이터 제공처(통신업체)의 승인 후에 데이터 다운로드 가능
 - 데이터 신청서 작성(상세 정보까지 작성) → 데이터 제공처(통신업체) 승인 → 데이터 다운로드

3. 주문형 다운로드 서비스

- 사용자가 원하는 형태의 통행 정보가 담긴 데이터를 다운받아 볼 수 있는 서비스
 - 예시) 모바일 통신 빅데이터 기반 통행정보와 이종 데이터의 정보를 연계
- 데이터 제공처(통신업체)의 승인 후, 한국교통연구원 데이터 가공 담당자와 협의 후 데이터 다운로드 가능
 - 데이터 신청서 작성(상세 정보까지 작성) → 데이터 제공처(통신업체) 승인 → 데이터 가공(KOTI) → 데이터 다운로드

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 '19년에 개발한 모바일 가공 알고리즘을 보완함
 - 기존의 이동속도 및 패턴기반 기지국 신호 보정 알고리즘을 보완하여, 데이터의 이상치를 효과적으로 보정하기 위한 알고리즘을 제시함
 - 신규 알고리즘은 합리적 수준의 시·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 점단위의 위치 추정 정보로 변환하는 방법을 제안하여 모바일통신 데이터의 공간적 해상도를 개선하였으며, 향후 이동경로 및 수단 추정을 위한 개발 알고리즘의 적용방안을 검토함
- 2019년 1월 1일부터 12월 31일까지 생성된 로그 기록에 새로 개발한 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
 - 모바일통신 빅데이터기반으로 구축된 통행DB의 시간대별, 요일별, 성별, 연령별 통행량 및 통행시간을 기존 국가교통DB사업의 '가구통행실태조사자료 (2016년)'와 통계청의 '인구총조사자료 (2015년)'비교하여 검증함
- 모바일통신 빅데이터 기반 통행 분석 서비스의 성능 및 사용성을 개선하기 위하여 DB를 재구조화하고 이용자 맞춤형 'Light version'을 추가 개발함
 - 일반 국민도 쉽고 빠르게 통행 정보를 확인할 수 있도록 일반사용자용 'Light version'의 분석서비스를 추가로 개발하고, 기존에 제공했던 분석서비스는 'Expert version'으로 명명하고 전문가에게 친숙한 UI/UX로 변경함
 - 또한, 2개 이상의 분석 수행 결과를 주제도로 동시에 확인할 수 있는 '정보 연계 분석 기능' 개발하고, 통행변화를 시계열로 확인할 수 있는 '통행변화 시뮬레이션 분석 기능' 개발함

제2절 차년도 수행계획

- 모바일통신 빅데이터 가공 로직 개발
 - 기 구축된 DB (전국, 2019년도) 검증
 - 국가교통DB 등 기존 조사기반 통계를 이용한 검증
 - GPS 정보 등을 활용한 신뢰성 검증 (신규)
 - 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
 - 모바일통신 빅데이터 기반 추정 가능 체류지(거주지, 근무지, 여가 등) 선정
 - 거주지 및 근무지 추정 로직 개선
 - 통행경로 및 수단 구분 알고리즘 개발 (신규)
 - 개별통행에 대한 이동경로 추정 및 보정 로직 개발 (신규)
 - 대중교통(철도, 버스) / 승용차 / 기타 이용자 구분 로직 개발 (신규)
- 모바일통신 빅데이터기반 교통지표 개발
 - 교통지표 개발
 - 모바일통신 빅데이터 DB 기반 산출 가능한 통행지표 개발 (통행량, 평균통행시간, 평균통행거리 등)
 - 이종 데이터와 모바일통신 빅데이터 DB를 연계·가공하여 산출 가능한 각종 지표 개발 (교통안전지수, 도시공동화지수, 교통서비스지수, 경제활동지수 등) (신규)
 - 교통빅데이터 플랫폼 분석기능 개발 지원
 - 교통빅데이터 플랫폼을 위한 개별통행 DB 재구조화 지원
 - 교통지표 제공 및 분석 서비스 개발 지원