전국 도로 교통량 조사, AI가 대신한다:
인공지능과 모빌리티 빅데이터를 활용한 교통량 추정의 진화

이승봉 교통빅데이터연구본부 부연구위원
이채영 교통빅데이터연구본부 연구원 · **전준수** 교통빅데이터연구본부 전문연구원

공공 교통량 관측 지점은 전체 도로망의 약 2%에 불과하여 전국 교통현상 설명에 구조적 한계가 있으나, 차량 GPS 기반 모빌리티 빅데이터와의 융합을 통해 전국 미관측 도로구간의 교통량 추정이 가능해지고 있다. 본 연구는 인공지능 기반 교통량 추정 모형을 단계적으로 고도화하여 전국 단위 교통량 DB의 신뢰도를 실증적으로 개선하고, 교통정책·SOC 투자·탄소배출 분석 등 정책 활용 기반을 확장하였다.

01 개요

◦ 도로 교통량은 교통혼잡 관리, 도로투자 타당성, 교통안전 등 국가 교통정책의 핵심 근거이자 도로망의 소통·지체·혼잡 수준을 판단하고 정책 효과를 검증하는 기본 지표임

- 정책 집행과 투자 의사결정의 신뢰도는 교통량 데이터의 정확도에 크게 좌우됨

◦ 그러나, 국가에서 조사하고 있는 교통량 조사 지점은 전체 도로망 대비 매우 제한적(약 2% 수준[2])으로, 전국 단위 교통망 성능을 파악하기 어려움

- 루프·영상검지기 등 기존 센서 기반 수집 방식은 특정 지점에 고정되어 있어 공간적 대표성 부족, 설치·유지관리 비용 부담, 포장 변형 시 성능저하 등의 문제가 제기됨
- 또한, 기관별 조사방식·장비가 상이하여 시·공간적 일관성이 낮고, 미관측구간이 광범위하게 발생하여 정책평가 과정에서 대표성·정확성·일관성 제약이 지속적으로 지적됨

2) 2023년 교통량 조사지점 11,378 지점(양방향)/2023년 12월 기준 국가표준노드링크수 548,614 개 = 2.07%

표 1 통행수단 분포 (단위:개)

구분		조사 지점수	방향별 관측교통량 DB 구축 지점 수		
			상행	하행	양방향
건설기술연구원	**소계**	**3,893**	**3,893**	**3,893**	**7,786**
	상시 조사지점	601	601	601	1,202
	수시 조사지점	3,292	3,292	3,292	6,584
한국도로공사		427	213	213	426
서울특별시		139	139	139	278
7대 특·광역시	**소계**	**713**	**1,066**	**1,074**	**2,140**
	인천광역시	195	246	249	495
	대전광역시	40	37	37	74
	대구광역시	124	244	244	488
	광주광역시	33	61	61	122
	부산광역시	135	237	237	474
	울산광역시	38	49	49	98
	세종특별자치시	48	192	197	389
합계		**5,262**	**6,519**	**4,870**	**11,389**

◦ 데이터 수집 환경은 고정 센서에서 이동 센서 기반으로 빠르게 전환되고 있음

- 차량 GPS, 스마트폰, DSRC 등 이동 프로브 데이터의 급격한 확산으로, 개별차량의 실제 주행 경로와 전국 커버리지 확보가 가능해졌으며, 데이터 양·질이 기존 대비 비약적으로 개선되었며, 실시간성, 전국 단위 연속성, 대규모 표본 확보가 가능한 환경으로 변화

◦ 국가통합교통체계효율화법(제12조·제14조) 및 시행령 제8조에 따라, 국가교통데이터베이스(KTDB) 구축·운영에는 전국 도로망을 대표할 신뢰성 높은 교통량 DB 구축이 필수임

- 교통망 성능평가 지표는 교통혼잡, 온실가스, 물류 효율성, 안전 등 주요 정책의 근거이므로, 전국 단위에서 대표성과 정확성을 갖춘 교통량 DB 구축이 요구됨

02
교통량 추정 모형의 진화

1) 교통량 추정 모형 개요

◦ 전국 도로 교통량 DB 구축의 핵심은 2% 수준의 관측교통량과 전국 단위 커버리지를 가진 차량 GPS 데이터를 융합하여, 나머지 98% 미관측 지점을 신뢰성 있게 추정하는 데 있음

- 국가 교통량 조사는 전체 도로망의 약 2%만 관측되나, 차량 GPS · 스마트폰 · 내비게이션 등 이동 프로브 데이터는 전국적 연속성을 갖춰 이러한 한계를 보완하는 핵심 자원임

◦ 이에 따라 KTDB 사업은 관측교통량과 차량GPS 기반 주행 패턴을 결합해 미관측 구간을 추정하는 체계를 구축하고, 교통량 추정 모형을 단계적으로 고도화하고 있음

- Model① KNN(K-nearest neighbors)과 지역가중선형회귀(Locally weighted linear regression) 알고리즘이 결합된 방법론[2]으로 미관측 링크 구간에 대한 교통량을 추정함
- Model② 랜덤포레스트(Random Forest) 기반 AI 회귀모형을 도입해 입력변수를 다변화하고 비선형 관계를 반영하여 일반화 성능을 강화
- Model③ 기존 Model②에 다년도 누적 학습데이터 + 네트워크 연결성* 변수를 추가하여 시공간적 연속성과 추정 정밀도를 높인 고도화 모델

* 개별차량 경로 데이터를 이용하여 특정 도로구간을 통과하는 차량의 연결성이 높은 관측지점 정보와 프로브 통행량을 생성하여 입력데이터로 활용

2) 모델별 교통량 추정 모형의 특징

◦ Model① (KNN+LWLR 기반 초기 모형)

- KNN은 미관측 구간과 통행 특성이 유사한 관측지점을 선택하는 군집화 역할 수행하고, LWLR은 선택된 관측지점 정보를 기반으로 국부 가중 회귀식을 적용해 교통량 추정

◦ Model② (AI 모형: 랜덤포레스트)

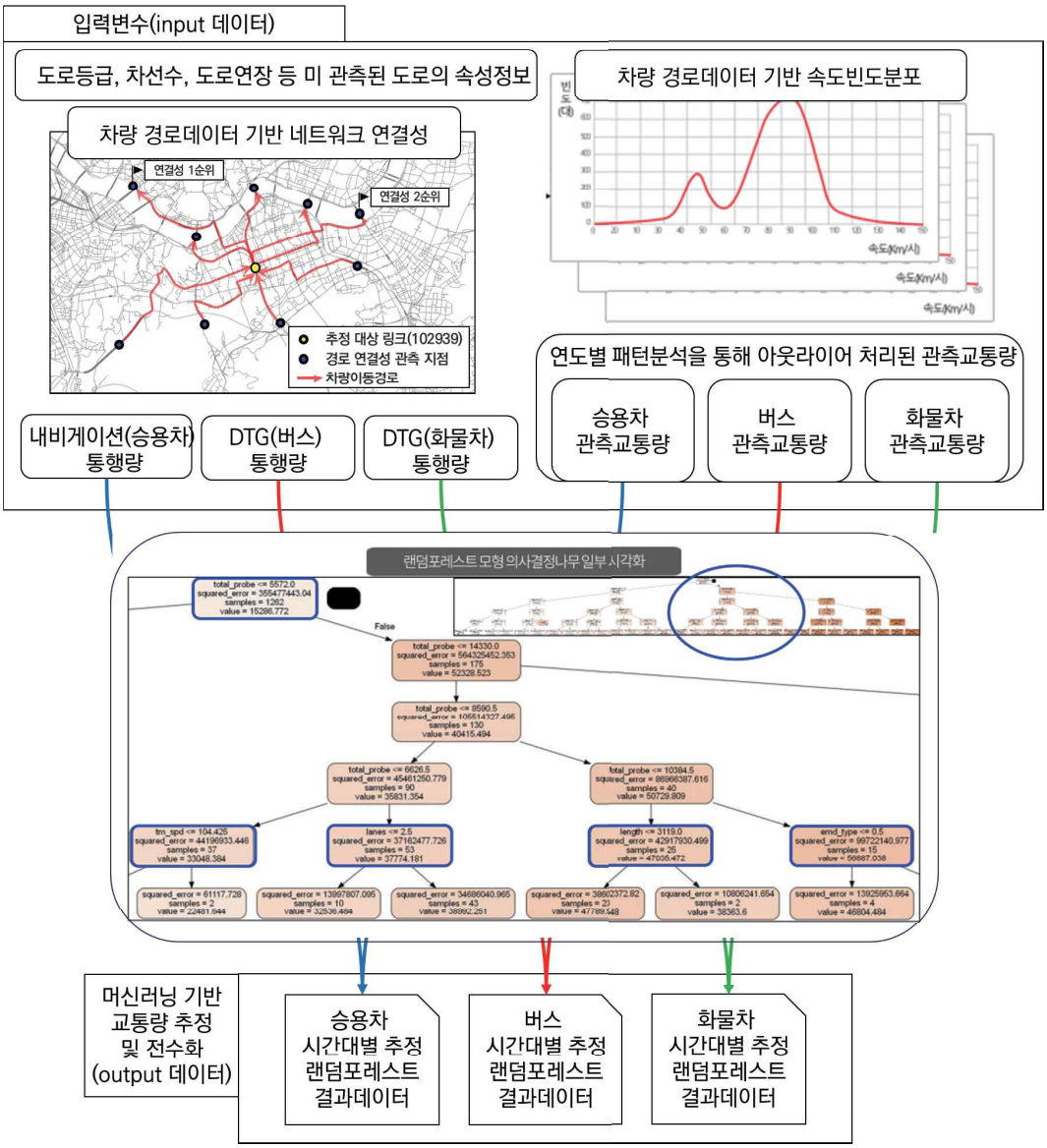
- 다수의 의사결정나무를 앙상블해 예측하는 랜덤포레스트를 적용하여 비선형 관계를 효과적으로 반영하고, 무작위 데이터 샘플링과 변수 선택을 통해 과대적합을 줄여, 변동성이 큰 도로에서도 일별·시간대별 교통량을 안정적으로 추정

◦ Model③ (누적 학습 + 네트워크 연결성 변수 기반 고도화 모형)

- 기존 Model②에 다년도 누적 데이터와 네트워크 연결성 변수를 함께 활용하여 도로망의 구조적·시공간적 특성을 정밀하게 반영하고 모형의 일반화 성능과 추정 정확성 개선

2) 천승훈 외(2016), 『첨단교통자료를 활용한 교통지표DB구축(1) - 전국도로망 기초교통DB 구축 및 활용방안』, 한국교통연구원, pp.106-111

그림 1 | Model③의 교통량 추정 프로세스



자료 : 한국교통연구원(2024), "2024년 국가교통조사. 및 분석 – 차량 GPS 빅데이터 구축" 한국교통연구원, pp. 61

03 교통량 추정 모형별 성능 비교 결과

1) 분석 개요

◦ 교통량 추정 모형의 단계별 고도화에 따른 성능 개선 효과를 신뢰도 분석을 통해 검증함

- 분석에 활용된 핵심 데이터는 차량 GPS 기반 개별 차량 경로 데이터와 관측교통량 자료로 모든 모형에서 동일함
- 다만, 모형별 적용 시점 차이에 따라 기준연도와 샘플 구조에는 차이가 존재하나, 동일한 검증 기준을 적용하여 성능을 비교함
- 또한, Model①에서는 내비게이션 데이터만을 활용하고, Model ②, ③에서는 내비게이션 데이터는 승용차, DTG 데이터는 버스·화물 교통량 추정에 활용함

◦ 모델별로 적용된 시점의 차이로 수집 데이터 및 통행량의 차이는 존재할 수 있음

방법론	추정 모형	수집 데이터	기준연도	학습데이터	입력변수
Model①	(비모수 기법) KNN+LWLR	내비게이션, 관측교통량	2019년	단일연도 (2019년)	링크별 프로브 통행량, 인접구간 관측교통량 등
Model②	(인공지능) 랜덤포레스트	내비게이션, DTG, 관측교통량	2021년	단일연도 (2021년)	링크별 차종별 프로브 통행량, 도로 정보(차선수, 도로등급 등), 평균속도, 관측교통량 등
Model③	(인공지능) 랜덤포레스트	내비게이션, DTG, 관측교통량	2023년	누적연도 (2020-2023년)	링크별 차종별 프로브 통행량, 도로 정보(차선수, 도로등급 등), 평균속도, 관측교통량 등 네트워크 연결성 변수

2) 신뢰도 분석결과

◦ (분석조건) 신뢰도 검증은 관측교통량이 24시간 모두 확보된 지점을 대상으로 수행함

- Model①은 학습과 검증 지점 별도 구분없이, 전체 자료를 이용하여 성능을 평가, Model②, ③은 학습데이터와 검증데이터를 8:2 비율로 분리하여 모형구축 및 검증 수행
- 모형별 신뢰도 검증 외에도 교통량 수준별 검증을 통해 모형의 활용 가능성 검토

◦ (평가기준) 교통량 추정 결과의 신뢰도 평가는 평균절대비율오차(MAPE, %) 지표를 활용함

- MAPE는 관측교통량 대비 추정교통량의 상대적 오차 수준을 정량적으로 평가하는 지표로, 모형 간 성능 비교에 주로 활용됨

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

여기서, q_i = 지점 i의 관측교통량(대/일)

$\hat{q}$ = 지점 i의 추정교통량(대/일)

n = 전체 지점 개수(개)

① Model별 분석 결과

◦ 관측교통량과 추정교통량 간 신뢰도 분석 결과, 교통량 추정모형의 고도화에 따라 평균 오차율(MAPE)이 단계적으로 크게 개선된 것으로 나타남

- Model①의 평균 오차율은 33.4% 수준으로, 고속도로 등급을 제외하고는 미관측 구간 추정에 한계가 있었으나, 인공Model②, ③의 신뢰도는 크게 개선됨
- Model③은 다년도 누적 학습과 네트워크 연결성 변수 반영을 통해 평균 오차율 8.3%까지 개선됨에 따라 교통량 추정의 정밀도와 전국 단위 활용 가능성이 실증적으로 확보됨

② 도로등급별 분석 결과

◦ Model③은 기존에 신뢰도가 낮았던 교통량 낮은 구간에서 뚜렷한 개선 효과를 보임

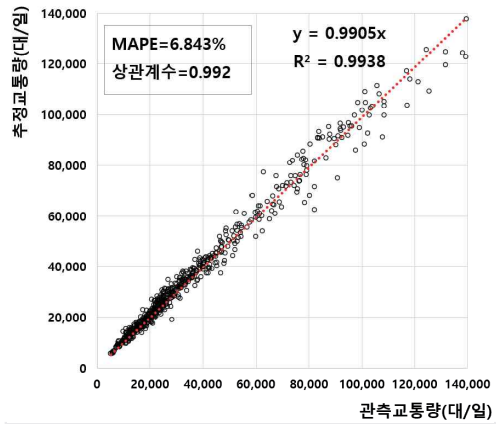
- 특별·광역시도는 오차율 7.1%, 지방도는 8.4%, 시군도는 9.2%로 기존 Model① 대비 각각 약 41%p, 53%p 이상 오차가 감소하여 안정적인 추정 수준을 확보함
- 도시고속국도 역시 약 6% 수준의 오차율을 보이며, 대도시 고속 소통축에서도 실무 적용이 가능한 수준으로 평가됨
- 국가지원지방도는 57.1% → 11.3%로 -45.9%, 일반국도는 27.1% → 9.9%로 –17.1% 개선되어 간선도로 전반에서 일관된 성능 향상이 확인됨

◦ 이러한 결과는 차량 GPS 기반 교통량 추정모형의 고도화를 통해 전 도로등급을 포괄하는 전국 단위 교통량 DB로서의 활용 가능한 신뢰도 수준을 확보하였음을 의미함

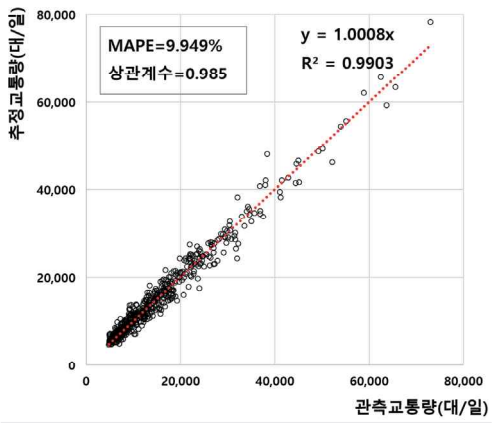
표 3 도로등급별 교통량 추정 신뢰도 분석 결과 (단위: %, %p)

도로등급	오차율(MAPE)			초기 모델과 비교 (Model③-Model①)
	Model①	Model②	Model③	
고속국도	14.0	5.9	6.9	-7.1
도시고속국도	26.3	3.2	6.3	-20.0
일반국도	27.1	11.4	9.9	-17.1
특별광역시도	21.4	9..	7.1	-14.3
국가지원지방도	57.1	13.5	11.3	-45.9
지방도	49.2	15.3	8.4	-40.7
시군도	62.6	23.2	9.2	-53.3
평균	**33.4**	**11.8**	**8.3**	**-25.2**

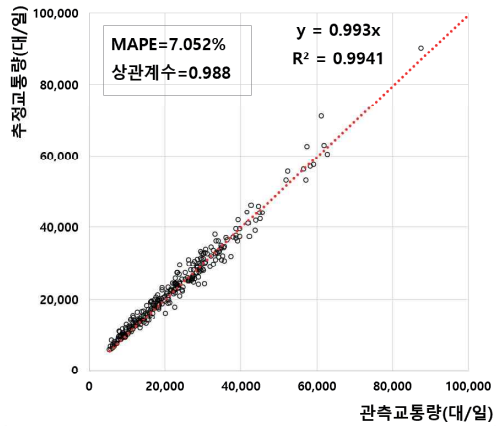
그림 2 Model③의 도로등급별 교통량 추정 검증 결과



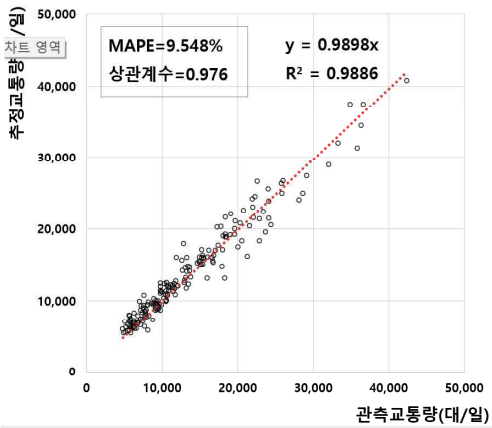
고속국도+도시고속국도



일반국도



특별광역시도



국가지원지방도+지방도

04 결론

1) 주요결과 요약

◦ 공공 교통량 조사자료의 공간적 한계와 데이터 수집 환경 변화에 대응하여, 모빌리티 빅데이터 기반 교통량 추정 모형을 단계적으로 고도화하여 구축함

- 기존 고정식 관측 중심 체계에서 차량 GPS 기반 이동형 데이터 중심 체계로 전환

◦ 차량 GPS 기반 개별차량 경로데이터와 관측교통량 자료를 활용하여 전국 미관측 도로구간의 교통량을 추정하는 체계를 구축함

◦ 다양한 방법론 및 입력변수 적용을 통해 교통량 추정모형의 신뢰도를 실증적으로 개선함

- 네트워크 연결성 변수를 적용한 인공지능 기반 추정모형을 통해 평균 오차율이 Model① 약 34% → Model③ 약 8% 수준으로 약 26%p 이상 개선

◦ 이를 통해 고속국도부터 시군도까지 전 도로등급을 포괄하는 전국 단위 교통량 DB로서의 활용 가능한 신뢰도 수준을 달성함

- 특히, 기존 관측 사각지대였던 국가지원지방도·시군도 등 교통량 수준이 낮은 도로구간의 활용성이 크게 향상됨

2) 정책적 시사점

◦ 현재 공공에서 조사·운영 중인 교통량 관측 지점은 전체 도로망의 2% 수준으로 제한적이어서, 기존 관측자료만으로는 전국 단위 교통현상을 충분히 설명하는 데 한계가 있음

- 이에 따라 차량 GPS 기반 교통량 추정기술과의 결합을 통해 미관측 구간을 포함한 전국 단위 교통량 DB 구축으로 공간적 공백을 해소할 필요가 있음

◦ GPS 데이터는 표본 자료인 반면, 관측교통량은 전수자료로 추정모형 학습·검증의 핵심 기준이므로, 교통량 조사 자료의 표준화와 통합 관리체계 구축이 교통량 DB 신뢰도의 출발점임

- 조사 지점별 장비·관측 방식 차이에 따른 자료 불일치를 해소하기 위해 품질 기준 정립과 정합성 관리가 필수적임
- 특히 시군도 등 저교통량 도로의 관측지점 비중이 낮으므로, 교통량 수준을 반영한 전략적 표본 설계를 통해 저·고 교통량 구간의 추정 정확도를 균형 있게 확보할 필요가 있음

◦ 현재 교통량 조사는 특정 1주일 중심으로 이루어지고 있어, 연평균 평일(AADT) 기준 통계에 한정되는 한계가 존재함

- 향후 정책 활용도를 높이기 위해서는 주말·계절 변동성을 반영할 수 있도록 상시 교통량 조사 비중을 단계적으로 확대할 필요가 있음
- 지자체 스마트교차로 교통량 데이터는 보조 관측수단으로 활용 가치가 높으나, 정식 통계 활용을 위해서는 표준화된 품질관리 기준 정립이 병행되어야 함

◦ 교통량은 도로 SOC 투자평가, 혼잡 개선 효과, 탄소배출량 산정 등 국가 교통정책의 핵심 정량지표로 활용되므로, 교통량 데이터의 품질 수준이 곧 정책 신뢰도를 좌우함

- 모빌리티 빅데이터 수집 확대 및 인공지능 기반 교통량 추정모형의 정확도 고도화를 통해 승인통계 수준의 교통량 통계 생산 체계로 발전시킬 필요가 있음

◦ 차량 GPS 기반 교통량 DB는 교통혼잡비용, 차량주행거리, 수송부문 탄소배출량, SOC 투자효과 분석 등 다양한 정책 분야에서 활용 가능한 국가 핵심 인프라 데이터로 확장 가능

- e-나라지표, 도로망 성능평가, 탄소공간지도 등과의 연계를 통해 혼잡 완화, 탄소중립, 투자 효율성 제고를 뒷받침하는 범정부 정책 활용 기반으로 지속 확장될 것으로 기대됨