
Global Report

가정용 운송에너지 사용량 예측을 위한 기계학습 접근법

1. 개요
2. 원시 데이터 및 사전 처리
3. 분석방법 및 분석모델의 사양
4. 기본 통계 결과
5. 결론

1. 개요

지난 세기 동안 전 세계가 급속히 도시화가 되면서 도시교통(특히 여객운송) 부문이 약 45%를 차지한 것으로 나타났다. 거대 에너지사용량으로 인해 전 세계 온실가스 배출량 1~2위를 기록하고 있는 미국의 경우, 에너지 사용 및 온실가스 배출량은 1차 에너지의 37%를 운송부문에서 사용하고 있으며, 산업 부문은 35%, 주거부문은 16%, 상업부문은 12%로 나타났다. 특히 주거 및 운송부문을 합친 에너지는 43.2 quadrillion Btu(quads)를 차지하고 있다. 이처럼 주거 및 운송부문에서 1차 에너지사용 점유율 대비 효율성을 보면, 운송부문보다는 각 가정에서의 에너지소비를 줄일 수 있는 가능성이 크다고 볼 수 있다. 건설/환경산업 분야에서 자원 소비량 대비 환경영향 결과를 예측하는 것은 도시환경의 지속가능성을 높이기 위한 정책입안자에게 직면한 핵심과제이다. 그간 가정용 건물의 에너지소비 예측기술에 대한 모델링 관련 연구는 많지만, 가정용 운송에너지 소비에 초점을 맞추는 연구는 거의 없는 상황이다 [1][2][3]. 연구자들은 오랫동안 회귀 기반 분석(regression-based analyses)방법을 적용하여 도시민의 여행패턴과 경험적 관계를 탐구해왔다. 이러한 분석방법은 주로 에너지사용량 예측보다는 여행패턴에 초점을 맞추고 있다[4][5]. 2009년 Pan 등은 분산된 데이터를 기반으로 회귀 모델을 개발하여 여행패턴과 이동거리 사이의 관계를 통해 도시민의 특성과 자가 차량 소유의 관계를 연구한바 있다[6].

분산된 데이터를 기반으로 회귀 모델링을 통한 대부분의 ML(machine learning : 기계학습)은 에너지 예측 방법에서 운송에너지 사용보다는 건물을 대상으로 하고 있다. 가정용 운송에너지 예측을 위한 ML 방법에 대한 문헌적 연구는 아직 초기단계이다. 이 연구는 이러한 고급 통계 모델이 개발환경 영향을 정량화하기 위한 시스템 기반 프레임워크인 IUMAT(Integrated Urban Metabolism Analysis Tool : 통합 도시 대사 분석도구)에서 운용되는 TM(Transportation Module)의 모델링 엔진으로 어떻게 수행되는지를 보여준다. 즉, IUMAT-TM은 물과 토지 사용모듈이 에너지와 연결되어 있음을 알 수 있다[7]. 이러한 모듈은 건물 및 근린특성에 따라 리소스 수요 프로파일을 생성하고 도시 대사 분석 플랫폼 내에서 관련 환경영향을 계산하기 위한 상황식 구조를 갖고 있다.

이 연구에서는 IUMAT-TM에서 운송에너지 소비를 예측할 수 있는 다른 통계적 방법에 대해 설명하고 대체 접근방식과 비교하여 교통에너지 소비가 도시민의 특성에 미치는 영향을 탐구한다.

2. 원시 데이터 및 사전 처리

2.1 데이터 준비

DVRPC(Delaware Valley Regional Planning Commission : 델라웨어 밸리 지역계획위원회)의 오픈 소스인 HTS(Household Travel Survey : 가정여행 설문 조사)를 이용하여 가정용 운송에너지 소비 예측 모델을 개발하였다. DVRPC는 델라웨어 밸리 지역(Pennsylvania counties of Bucks, Chester, Delaware, Montgomery, and Philadelphia; New Jersey counties of Burlington, Camden, Gloucester and Mercer)에 거주하는 20,216명의 주민을 대상으로 2015년 한 해 동안 10개 카운티에서 14,903대의 자동차를 이용하여 48,646건의 여행을 한 9,235세대의 인구통계 및 여행정보를 2012HTS에 수록하였다. 각 가구구성원의 여행모드(걷기/자전거 타기/차량 운전/버스), 여행목적, 목적지 및 여행시간 등에 대한 정보와 연령/성별/차량가용성/고용상태와 같은 인구통계학적 특성이 기록되어 있다. 아울러 DVRPC의 구역 경계 데이터 파일(고용밀도 및 가구, 학교 및 버스정류장 등)를 이용하여 주거민의 특성을 파악하였다. 단, HL(Household Level)의 이동거리와 같은 운송정보 계산에 직접 사용되는 변수는 제거하였다. 그 결과, 19개의 연속 및 12개의 범주형 변수(인구통계, 여행정보 및 영역특성 등)가 분석에 포함되었다. DVRPC 지역의 인구통계학적 특성(인구통계, 여행정보 및 지역특성 등)을 <표 1>에 나타낸다.

<표 1> DVRPC 지역의 인구통계학적 특성

구분	설명	최소값	최대값	평균값	왜곡
하우스 홀드	총 가구 수	0	2764	978	0.55
자동차	사용가능한 총 차량 수	0	9237	2163	1.22
HHS_0-35K	0~35K 사이의 소득(1999년)을 가진 가구	0	1614	359	1.3
HHS_OVER_35K	35K 이상의 소득(1999년)을 가진 가구	0	1695	619	0.58
학교 수	지역 내 K-12 학교 수	0	7	0.93	1.29
대학 수	대학 또는 대학 수	0	3	0.48	5.63
자전거	TAZ 링크의 평균 자전거가치 점수	0	2	0.79	0.16
공원	TAZ 내 공원의 주차장 수	0	2882	12.3	83.7
철도	TAZ와 교차하는 총 철도역 수	0	10	0.61	3.49
버스정류장	TAZ와 교차하는 총 교차버스정류장 수	0	501	32.8	2.42
총 고용	고용수	0	1414.9	1040	3.10

※ TAZ : Traffic Analysis Zones

* 자료 : Shideh Shams Amiri, Nariman Mostafavi, Earl Rusty Lee, Simi Hoque, “Machine Learning Approaches for predicting Household Transportation Energy Use“, City and Environment Interactions, 21 August 2020, p.7.

2.2 운송에너지 사용량 추정

가정용 운송에너지 사용량은 각 가정의 여행패턴과 관련된 여행모드, 연료유형 및 여행거리 등에 따라 다르게 나타날 수 있다. 교통용 운송에너지는 2010년 Jiang이 제안한 방법을 이용하여 DVRPC의 일일 여행패턴을 기반으로 계산하였다. TAZ(Traffic Analysis Zones) 수준의 운송에너지 소비량 측정수치의 계산식을 식(1-3)에 나타낸다. 연비 및 에너지강도 요인은 2013년 DoT(Department of Transportation : 미국 교통부) 자료를 참고하였다.

$$E_i^T = \sum_m E_i^m \dots\dots\dots (식 1)$$

$$E_i^m = \sum_j \sum_k (TF_j^m \cdot i, j, k \cdot \frac{TD^m_{i, j, k}}{TO^m_{i, j, k}}) \cdot EI^m \dots\dots\dots (식 2)$$

$$EI^m = FU^m \cdot EC^m \dots\dots\dots (식 3)$$

$i-f^h$: 가정용 운송에너지 사용량

$j-f^h$: 각 가정의 구성원

k : 여행목적

m : 여행모드

E_i^T : 가정용 일일 운송에너지 총 소비량 (kWh/HH)

$TF^m_{i, j, k}$: 가정 구성원(J)의 일별 여행 빈도 (Trips/Day)

$TD^m_{i, j, k}$: 가정 구성원(J)의 평균 여행거리 (Mile/Trip)

$TO^m_{i, j, k}$: 가정 구성원(J)의 목적별 여행 점유율

EI^m : 모드(m[kWh/마일])에 대한 에너지강도 계수 (kWh/mile)

FU^m : 모드(m[kWh/마일])에 대한 연비 계수 (L/mile or KWH/mile)

EC^m : 모드(m[kWh/마일])에 대한 에너지 함량 계수 (kWh/L)

3. 분석방법 및 분석모델의 사양

3.1. 선형 회귀 모델

일반화된 예측모델에 대한 직관적인 방법으로 선형회귀 방법이 많이 이용되고 있다. 응답변수를 입력 파라미터의 선형 조합으로 표시하고, 오류손실을 최소화하기 위한 간단한 함수가 포함된다. 일반적으로 최소 정사각형 방식은 여러 회귀 모델에서 초기 추정 목적으로 사용된다. 회귀 계수를 찾은 후 선형 회귀 방정식은 하나 이상의 독립적인 예측변수의 선형함수로서 연속 응답변수의 값을 예측할 수 있다. 회귀 모델은 모델 파라미터의 구현 및 해석이 용이하여 많이 이용되고 있다.

3.2 정규화

LSA(Lasso Regression Analysis : Lasso 회귀 분석)을 이용한 간단한 선형 회귀 정규화를 적용하여 선형 모델의 오버 피팅(over fitting)을 줄이고 예측오류를 개선할 수 있다. 선형 모델에 그림을 너무 많이 추가하면 분산계수가 증가할 수 있어 LSA의 예상계수 크기를 제한하기 위해 페널티 기간을 추가하고 계수 수가 줄어든 간단한 모델을 선택한다. 이러한 페널티 기반 변수선택 방법을 이용하면 자연순서(2013년 Ren 등이 제안)를 선택하는 동시에 계수행렬의 하위구조를 결정할 수 있다. LSA(Lasso 회귀 분석)은 (식 4)와 같이 계산할 수 있다:

$$\min_{\beta} \left\{ \frac{1}{N} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1 \right\} \dots\dots\dots (식 4)$$

(식 4)에 페널티 기간을 추가하여 LSA를 확장하여 예측변수를 개선할 수도 있는데, 이에 대한 상관관계는 (식 5)와 같이 계산할 수 있다:

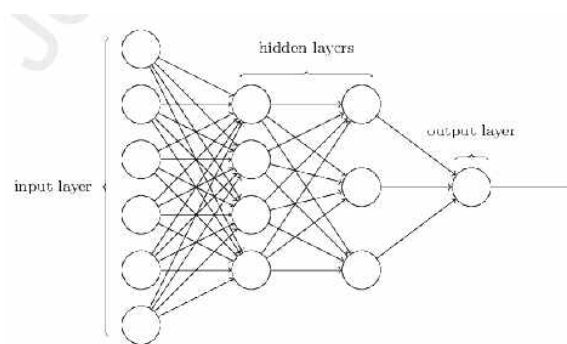
$$\beta \equiv \operatorname{argmin}_{\beta} (y - X\beta)^2 + \lambda_2 \beta^2 + \lambda_1 \beta_1) \dots\dots\dots (식 5)$$

3.3. 신경망 모델

NN(Neural Network : 신경망) 기술은 복잡한 비선형 패턴을 캡처할 수 있는 가장 널리 사용되는 ML 알고리즘 중 하나이다. NN 모델은 원래 생물학적 뇌의 신경생리학을 모방하여 개발되었다. 프레임워크의 훈련과정은 표준 ANN(Artificial Neural Network : 인공 신경망)에는 입력, 출력 및 숨겨진 레이어에 노드가 포함되어 있는 인간의 뇌와 유사한 구조를 갖는 층(layer)과 뉴런(neurons : 신경)으로 되어 있다. 인공 뉴런(노드)은 학습데이터에 맞게 모델이 학습하는 동안 조정되는 가중치가 있다. 이는 (식 6)과 계산할 수 있다:

$$p_j(t) = \sum_i o_i(t) w_{ij} + w_{0j} \dots\dots\dots (식 6)$$

여기서 $p_j(t)$ 는 입력뉴런 j , $O_i(t)$ 는 이전 뉴런작업에서 나온 출력값 이고, w_{ij} 는 연결의 중량이며, w_{0j} 는 바이어스 값이다. 전파함수는 일반적으로 비선형 활성화 함수가 출력되기 전에 숨겨진 레이어에 적용되는 것과 유사하다. 숨겨진 레이어의 수는 시스템의 복잡성과 특성에 따라 다르게 나타날 수 있다. 시간 파라미터, 임계값 및 두 개의 활성화 및 출력함수에 따라 뉴런의 입출력 값이 결정된다. 출력뉴런은 활성화 기능이 없는 단일 뉴런으로 되어 있다. 두 개의 숨겨진 레이어가 있는 비교적 간단한 구조를 갖는 NN(Neural Network) 모델의 회로구조를 [그림 1]에 나타낸다.

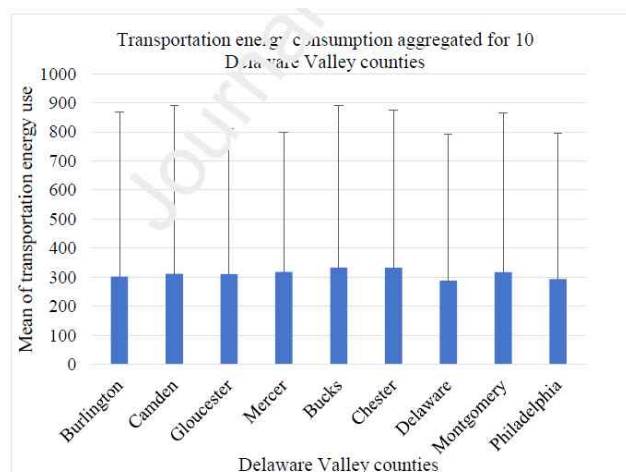


- * 자료 : Shideh Shams Amiri, Nariman Mostafavi, Earl Rusty Lee, Simi Hoque, “Machine Learning Approaches for predicting Household Transportation Energy Use“, City and Environment Interactions, 21 August 2020, p.13.

[그림 1] NN(Neural Network) 모델의 회로구조

4. 기본 통계 결과

가정용 운송에너지 사용량에 대한 기본통계에 대한 비교결과를 [그림 2]에 나타낸다.



- * 자료 : Shideh Shams Amiri, Nariman Mostafavi, Earl Rusty Lee, Simi Hoque, “Machine Learning Approaches for predicting Household Transportation Energy Use“, City and Environment Interactions, 21 August 2020, p.16.

[그림 2] 가정용 운송에너지 사용량에 대한 기본통계 비교

이는 델라웨어 밸리 카운티에 대한 교통에너지 소비량에 대한 집계결과이며, 모든 델라웨어 카운티는 에너지 사용량 면에서 넓은 범위와 높은 표준편차를 가지고 있음을 알 수 있다.

5. 결론

가정용 에너지 소비는 도시설계 기획자, 건물 및 운송 엔지니어의 의사결정에 매우 중요한 역할을 한다. 특히 회귀모델, SVM 및 ANN 기술은 건물수준에서 가정용 에너지 소비량을 추정하는 데 매우 중요한 인자이다. 아울러 학습모델 등 ML방법은 최근 에너지 연구 분야에서 고도의 빠르고 정확한 옵션을 제공하고 있다.

이 연구에서는 탄력적인 순 정규화 분석, 결정 트리 및 무작위 신경망을 이용하여 가정용 운송에너지 사용량(소비)을 예측하기 위한 모델링 기술을 제시하였다. 이 목적은 이러한 모델이 도시의 지속가능성 평가를 위한 시스템 기반 계산 플랫폼(IUMAT)에서 작동하는 교통모듈에 얼마나 효과적으로 적용할 수 있는지를 평가하는 것이다. 이 연구결과는 국내 각 가정용 운송에너지 관련 정책에 벤치마킹 자료로 활용할 수 있을 것이다.

[참고문헌]

- [1] Næss, Petter. "Residential location, travel, and energy use in the Hangzhou Metropolitan Area." *Journal of transport and land use* 3, no. 3 (2010): 27–59.
- [2] Yu, Feifei. "Simulation of household in-home and transportation energy use: an integrated behavioral model for estimating energy consumption at the neighborhood scale." PhD diss., Massachusetts Institute of Technology, 2013.
- [3] Pede, Timothy. "Exploring relationships between building and transportation energy use of residents in US metropolitan regions." (2014).
- [4] Boarnet, Marlon, and Randall Crane. "The influence of land use on travel behavior: specification and estimation strategies." *Transportation Research Part A: Policy and Practice* 35, no. 9 (2001): 823–845.
- [5] Newbold, K.B., Scott, D.M., Spinney, J.E., Kanaroglou, P. and Páez, A., 2005. Travel behavior within Canada's older population: a cohort analysis. *Journal of Transport Geography*, 13(4), pp.340–351.
- [6] Pan, Haixiao, Qing Shen, and Ming Zhang. "Influence of urban form on travel behaviour in four neighbourhoods of Shanghai." *Urban studies* 46, no. 2 (2009): 275–294.
- [7] Mostafavi, Nariman, Mohamad Farzinmoghdam, and Simi Hoque. "Urban residential energy consumption modeling in the Integrated Urban Metabolism Analysis Tool (IUMAT)." *Building and Environment* 114 (2017): 429–444.

<원문제목> Machine Learning Approaches for predicting Household Transportation Energy Use <원문출처> City and Environment Interactions
--