ARTICLE IN PRESS

Computers, Environment and Urban Systems xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/ceus

Using machine learning and small area estimation to predict building-level municipal solid waste generation in cities

Constantine E. Kontokosta[a,b,*], Boyeong Hong[a,b], Nicholas E. Johnson[b,c], Daniel Starobin[d]

[a] Department of Civil and Urban Engineering, New York University, United States
[b] Center for Urban Science and Progress, New York University, United States
[c] University of Warwick, United Kingdom
[d] New York City Department of Sanitation, United States

ARTICLE INFO

Keywords:
Urban waste management
Municipal waste
Machine learning
Data analytics
GIS

ABSTRACT

Municipal solid waste management represents an increasingly significant environmental, fiscal, and social challenge for cities. Understanding patterns of municipal waste generation behavior at the household and building scales is a critical component of efficient collection routing and the design of incentives to encourage recycling and composting. However, high spatial resolution estimates of building refuse and recycling have been constrained by the lack of granular data for individual properties. This paper presents a new analytical approach, which combines machine learning and small area estimation techniques, to predict weekly and daily waste generation at the building scale. Using daily collection data from 609 New York City Department of Sanitation (DSNY) sub-sections over ten years, together with detailed data on individual building attributes, neighborhood socioeconomic characteristics, weather, and selected route-level collection data, we apply gradient boosting regression trees and neural network models to estimate daily and weekly refuse and recycling tonnages for each of the more than 750,000 residential properties in the City. Following cross-validation and a two-stage spatial validation, our results indicate that our method is capable of predicting building-level waste generation with a high degree of accuracy. Our methodology has the potential to support collection truck route optimization based on expected building-level waste generation rates, and to facilitate new equitable solid waste management policies to shift behavior and divert waste from landfills based on benchmarking and peer performance comparisons.

1. Introduction

Waste management is an increasingly complex quality-of-life issue for cities around the world, especially given the rapid growth of urban populations over the past two decades (World Health Organization Centre for Heath Development, 2010; Leao, Bishop, & Evans, 2004). Proper waste management is essential in order to provide sustainable, livable cities, as the collection and removal of waste impacts carbon emissions, traffic congestion, and air quality, as well as requiring significant operating expenditures (Adeyemi, Olorunfemi, & Adewoye, 2001; Esin & Cosgun, 2007). To improve waste management services and reduce the amount of waste sent to landfills, local governments are developing new methods to create efficient waste management systems and increase diversion rates through recycling and composting programs (MacDonald, 1996; Wang, Richardson, & Roddick, 1996; Bhargava & Tettelbach, 1997; Guerrero, Maas, & Hogland, 2013; Pappu, Saxena & Asolekar, 2007; Tam & Tam, 2006). New data streams, and the application of machine learning statistical methods, enable data-driven approaches to persistent problems in urban environmental management. Such data have proven to be a great resource for waste management planning, but they are typically collected at too coarse of an aggregation to fully optimize collection routing, and provide the empirical basis for policies that can shift, or "nudge", behavior through incentives or regulations based on performance metrics. The need for high resolution and targeted municipal solid waste management policy is crucial to minimize the future negative environmental impacts of urban waste.

Previous work has aimed to improve municipal waste management by using systems dynamics or data-driven modeling techniques to predict waste generation and identify factors that explain waste and recycling behavior. In particular, studies using temporal models with lagged waste generation data have performed well for prediction and forecasting, in large part due to the time series auto-correlation observable in waste generation rates at the regional or local level. Missing

* Corresponding author at: Department of Civil and Urban Engineering, New York University, United States.
E-mail address: ckontokosta@nyu.edu (C.E. Kontokosta).

https://doi.org/10.1016/j.compenvurbsys.2018.03.004
Received 27 November 2017; Received in revised form 6 March 2018; Accepted 7 March 2018
0198-9715/ © 2018 Elsevier Ltd. All rights reserved.

Please cite this article as: Kontokosta, C., Computers, Environment and Urban Systems (2018), https://doi.org/10.1016/j.compenvurbsys.2018.03.004

from the literature, however, are attempts to predict waste generation for individual buildings in a large-scale municipality. Part of the challenge emerges from data constraints, as few sanitation agencies collect and make available granular waste collection data. Furthermore, small-area estimation problems can confound attempts to accurately downscale predictions from the city or district to individual buildings.

Given this context, this research attempts to inform municipal waste management operations by developing high spatial and temporal resolution estimates of waste and recycling generation rates for individual residential buildings. Using an extensive and granular waste collection dataset from the New York City Department of Sanitation (DSNY), coupled with detailed land use, demographic, socioeconomic, and weather variables, we are able to overcome previous data limitations to build a socio-spatial machine learning model to predict building-specific waste generation rates based on derived building population estimates for more than 750,000 residential properties in New York City. We validate our prediction using a sample of individual truck route collection data for specific days and locations that are representative of the City's land use types and densities. The results of the model and validation indicate that our method performs well at estimating building-level waste generation. Our findings also provide a comprehensive understanding of the socioeconomic and land use drivers of municipal waste generation in New York City and can enable more efficient, and equitable, waste management practices, particularly through route optimization and peer comparison benchmarking programs.

We begin by presenting a literature review of previous studies that attempt to predict waste generation using machine learning techniques, as well as applications of small unit estimation to determine building and household population size. Section 3 includes a description of our data and machine learning methods. Results are presented in Section 4, followed by a discussion of the findings and their implications for urban waste management and data-driven environmental policy.

2. Background

Numerous studies have used available waste data to identify significant factors that influence refuse and recycling rates, and to develop statistical models to forecast waste generation. Previous work has demonstrated that a broad range of factors drive waste generation depending on the particular study area, although most focus on macro-scale analyses of regions or metropolitan areas. Keser, Duzgun, and Aksoy (2012), for example, found that regional characteristics such as the unemployment rate, the asphalt-to-paved road ratio, temperature, higher education ratio, and agricultural production values have a significant influence on waste generation in Turkey. In Xiamen, China, Zhang et al. (2015) identified population, land use, and building coverage as important factors driving waste generation. Oribe-Garcia, Kamara-Esteban, Martin, Macarulla-Arenaza, and Alonso-Vicario (2015b) support and extend these studies by suggesting that urban morphology, tourism activity, educational level, economic status, and resources of the population impact aggregate residential waste generation. Denafas et al. (2014) estimate seasonal variations of waste generation in Eastern European cities by using time series forecasting models. Their results suggest that geographical latitude is a major factor in the seasonal pattern of waste generation across cities primarily due to differences in local weather.

A variety of models have been developed to predict waste generation, typically at the national or city scale. Karadimas and Loumos (2008) used the ant colony system algorithm to predict waste generation based on spatially-dependent characteristics such as the location of waste bins, the road network topology, and population density. Other researchers have built predictive models to forecast waste generation using temporal features such as seasonality and historical trends. For instance, Rimaityte, Ruzgas, Denafas, Racys, and Martuzevicius (2012) found that an Autoregressive and Integrated Moving Average (ARIMA) model combined with seasonal exponential smoothing is an effective method to predict weekly waste generation. Modern machine learning methods, such as neural networks, have also been used in temporal models to predict waste generation. Zade and Noori (2008) for example, used a feed-forward artificial neural network (ANN) to forecast weekly waste generation in the tourist city of Mashhad, Iran. They found that ANNs perform well when predicting waste generation at low spatial resolutions, and introducing time lag features into an ANN model can address serial correlation in the time series data. Antanasijevic, Pocajt, Popovic, Redzic, and Ristic (2013) also used ANNs to predict solid waste generation at the national level in Bulgaria and Serbia.

Combinations of spatial and temporal models have also been used to improve prediction results. For instance, Dyson and Chang (2005) developed a systems dynamics model using population, median household income, household size, and historical waste generation data to predict waste generation in San Antonio, Texas. In a particularly relevant study, Johnson et al. (2017) developed a spatiotemporal model using a gradient boosting regression tree algorithm and features such as weather, urban morphology, and socioeconomic and demographic information to predict weekly waste generation for 232 administrative sections (geographic divisions) in New York City. The results demonstrate high predictive accuracy for waste generation using historical waste generation rates.

These studies, however, are often limited by spatially-aggregated data that do not allow for analysis or prediction of waste generation patterns across small areal units. Data are typically aggregated into larger geographical units due to privacy and confidentiality issues, or the lack of data collection at the truck, household, or building scale. Previous attempts at building or household population estimation tend to rely on surveys (Ojeda-Benítez, Armijo-de Vega, & Marquez-Montenegro, 2008; Thanh, Matsui, & Fujiwara, 2010) or on down-sampling from predictions at larger geographies, such as the city or province (Oribe-Garcia, Kamara-Esteban, Martin, Macarulla-Arenaza, & Alonso-Vicario, 2015a; Purcell & Magette, 2009). These studies are either difficult to generalize, given their small samples sizes and limited historical data, or do not account for significant variations in the occupancy of specific buildings that can be obscured by relying solely on population surveys, such as census data. This is a significant gap in the literature, as building-level waste generation data can provide important insights for collection route optimization, household waste benchmarking programs, and more equitable waste reduction incentive and behavior change initiatives built on information transparency.

3. Materials and methods

This study aims to predict weekly and daily municipal waste generation from residential properties at the building level using a data mining and machine learning approach. We first develop a predictive model by comparing the performance of gradient boosting regression tree (GBRT) and Neural Network (NN) machine learning algorithms to estimate weekly waste generation for each of the 609 *DSNY sub-sections*, which are collection areas (also known as "frequencies") within the 232 DSNY sections. We then estimate individual building populations for all residential properties in NYC by implementing small area estimation methods that combine census population data with specific building characteristics including type, size, and density. Weekly generation at the building-level is then calculated by multiplying the predicted per capita weekly waste generation for each DSNY sub-section with the estimated building population of a given building located within that DSNY sub-section. This approach accounts for inter-sub-section variations in waste generation behavior, driven by such variables as demographics or socioeconomic characteristics, and building-specific factors, such as the number of residential units and their relative size. Fig. 1 presents a flow chart of our methodology.

Download English Version:

https://daneshyari.com/en/article/6921845

Download Persian Version:

https://daneshyari.com/article/6921845

Daneshyari.com