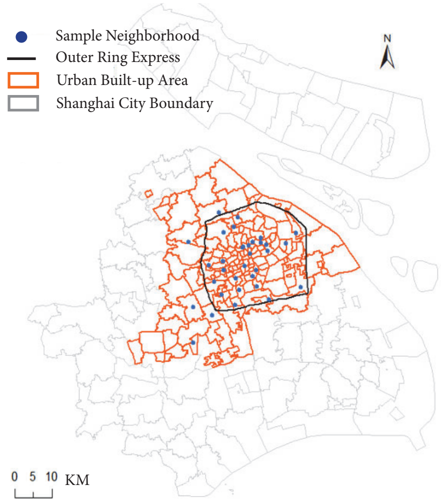
Effects of Home and Workplace Built Environments on Open Space Usage A Case Study of Shanghai

Urban parks benefits people's physical and psychological health. Promotion of the park usage will lead to a healthier city. Among different factors related to the park usage, built environment was proved relatively important. It's essential to understand to what extent does different built environment elements exert effects on the park usage and what are the actual relationships. Therefore, this study conducted a questionnaire among Shanghai citizens and use Gradient Boosting Decision Tree to figure out the relationship between residential, work-place environment and the park usage of citizens, as well as the relative importance of each predictors. The result shows that the importance of work-place and residential built environment are equal, which is much larger than social-economic and preference predictors. Moreover, all the built environment predictors have non-linear relationship with the park usage.

Data Source

The data used in the study were obtained from the 2018 "Daily Activity and Health Status of Shanghai Residents" questionnaire. The survey used a stratified, size-proportional sampling method to survey 1052 residents in 33 neighborhoods in the built-up area of Shanghai(as the figure on the right), of which 1005 were valid samples.



Response Variable

Based on the question "How often do you go to the park/open space from your home", assign a value of 0 to "less than once a week" and a value of 1 to "at least once a week".

Predictors	
Home Built Environment	
Home Population Density	Population density within 1km around home (thousand people/km2)
Home Mixed Land Use	Mixed Land Use Index within 1km around home
Home Road Density	Road density within 1km around home (km/km2)
Home Open Space Density	Open space density within 1km around home (number of open spaces/km2)
Home Transit Density	Transit density within 1km around home (number of bus/metro stations/km2)
Workplace Built Environment	
Workplace Population Density	Population density within 1km around workplace (thousand people/km2)
Workplace Mixed Land Use	Mixed Land Use Index within 1km around workplace
Workplace Road Density	Road density within 1km around workplace (km/km2)
Workplace Open Space Density	Open space density within 1km around workplace (number of open spaces/km2)
Workplace Transit Density	Transit density within 1km around workplace (number of bus/metro stations/km2)
Commute Distance	Distance from home to workplace (km)
Socio-economic variables	
Age	Age of the interviewee
Hukou	Have Hukou of Shanghai or not (no=0, yes=1)
Education	Years of education
Sex	Female=0, male=1
Household Income	Annual household income(1= below 50,000 2 = 50,000-100,000 3=110,000-200,000 4=210,000-300,000 5=310,000 to 400,000 6=410,000-500,000 7=510,000-600,000 8=over 600,000 yuan)
Household Size	How many people in this household
Children	How many children in this household
Car ownership	Have a car or not (no=0, yes=1)
Healthy preference variables	
Sitting hours at home	How many hours does this interviewee sit constantly at home
Sitting hours at work	How many hours does this interviewee sit constantly at workplace
Exercise preference	Does this interviewee like exercise (0=Don't like at all 1=Not so like 2=Neutral 3=Relatively like 4=Very like)

Algorithm: Gradient Boosting Decision Tree

Since the response variable only have 0 and 1, y is the true value and $f(x)$ is the predicted value, the loss function is:

$$L(y, f(x)) = -2(yf(x) - \log(1 + \exp(f(x))))$$

$$f_0(x) = argmin_c \sum_{i=1}^{N} L(y_i, c), \quad \text{so,} \quad f_0(x) = c = \log\left(\frac{y}{1-y}\right)$$

Generate m=1, 2…, M trees, and in each tree iterate i=1, 2…, N (Sample size), fit the tree with training data and calculate the residual by the following gradient formula:

$$r_{mi} = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]$$

Use r_{mi} to fit another new tree and minimize the Loss function:

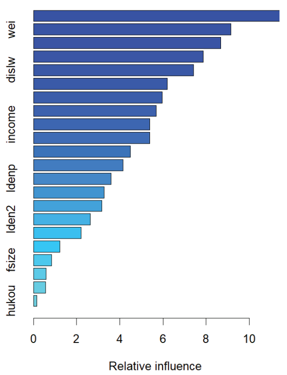
$$c_{mj} = argmin_c \sum_{x_i \in R_{mp}} L(y_i, f_{m-1}(x_i) + c)$$

The new model $f_m(x) = f_{m-1}(x) + \sum_{p=1}^{J} c_{mp}\, I(I = 1, x \in R_{mp}; I = 0, x \notin R_{mp})$

In case of the overfitting, we add a shrinkage parameter ζ, so the new model is $f_m(x) = f_{m-1}(x) + \zeta \sum_{p=1}^{J} c_{mp}\, I(0 < \zeta \leq 1)$, and the final tree can be interpreted as below:

$$f_M(x) = \sum_{m=1}^{M} \sum_{p=1}^{P} \zeta c_{mp}\, I(I = 1, x \in R_{mp}; I = 0, x \notin R_{mp})$$

Relative Influence of Predictors



Predictors	Relative Importance	Rank
Home Built Environment	**28.78%**	
Home Population Density	3.59%	13
Home Mixed Land Use	7.87%	4
Home Road Density	11.42%	1
Home Open Space Density	2.63%	16
Home Transit Density	3.27%	14
Workplace Built Environment	**28.88%**	
Workplace Population Density	5.38%	10
Workplace Mixed Land Use	9.16%	2
Workplace Road Density	5.96%	7
Workplace Open Space Density	2.20%	17
Workplace Transit Density	6.19%	6
Commute Distance	**7.41%**	5
Socio-economic variables	**18.58%**	
Age	5.69%	8
Hukou	0.16%	22
Education	0.56%	21
Sex	0.57%	20
Household Income	5.38%	9
Household Size	0.85%	19
Children	1.22%	18
Car ownership	4.16%	12
Healthy preference variables	**16.34%**	
Sitting hours at home	3.17%	15
Sitting hours at work	4.49%	11
Exercise preference	8.68%	3

Non-linear Relations