MONOPOLY: Learning to Price Public Facilities for Revaluing Private Properties with Large-Scale Urban Data

Miao Fan, Jizhou Huang, An Zhuo, Ying Li, Ping Li, Haifeng Wang

{fanmiao, huangjizhou01, zhuoan, liying, liping11, wanghaifeng}@baidu.com

Motivation

- M1: Have you ever wondered about the true value of your **private property**?
- M2: How much do these public facilities contribute to the price of your house?
- M3: If you are a tycoon, how much will you pay to acquire these **public facilities**?

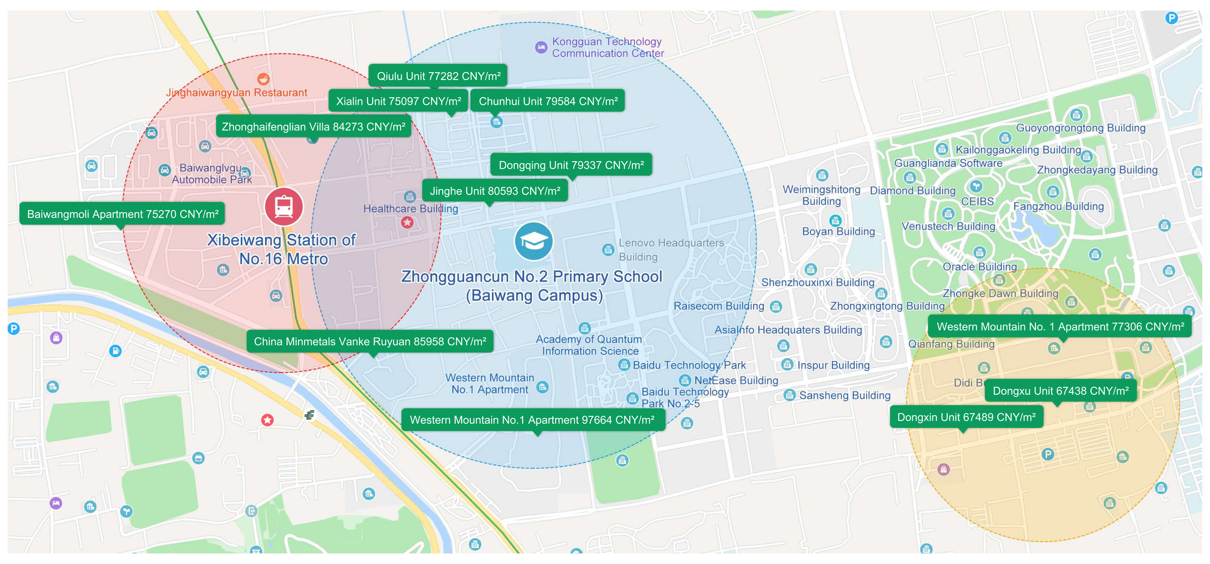

Figure 1: A screenshot of the housing prices of the Haidian District in Beijing. We can see that there are three kinds of dash circles colored by red, blue, and orange. Red circle: the average price of the properties near the *Xibeiwang Station of No.16 Metro* is $79,772\ \mathrm{CNY/m^2}$. Blue circle: the average price of the properties near the *Zhongguancun No. 2 Primary School (Baiwang Campus)* is $82,216\ \mathrm{CNY/m^2}$. Orange circle: however, the average price of the properties without those public facilities nearby is $70,744\ \mathrm{CNY/m^2}$, noticeably lower than the two other areas.

Idea

Generally speaking, the "Monopoly" project aims to assign virtual prices to public facilities based on the values of existing private properties, and in turn, the virtual prices of public facilities can help estimate the worth of a newly-established realty.

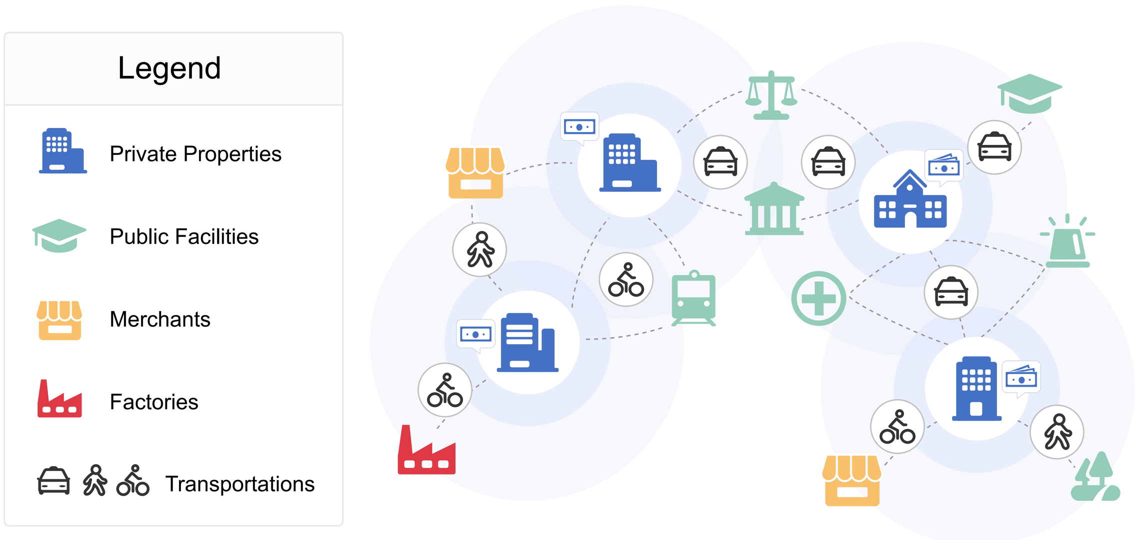

Figure 2: An illustration of the organization of urban data employed by the "Monopoly" project. To be specific, our approach regards many points of interest (POIs) as nodes in an undirected weighted graph based on their geographic information. Then we formulate the factors, including the variables that indicate the values of surrounding public facilities, to parallelly regress to the housing prices we know. As a result, the estimated values of both public facilities and private properties can be updated iteratively until convergence.

Significance

Besides deploying the "Monopoly" project in our web mapping service (i.e., Baidu Maps), we plan to directly establish an independent platform where multiple business intelligent agents could automatically give more suggestions on investments, urban planning, and taxation powered by our large-scale urban data.

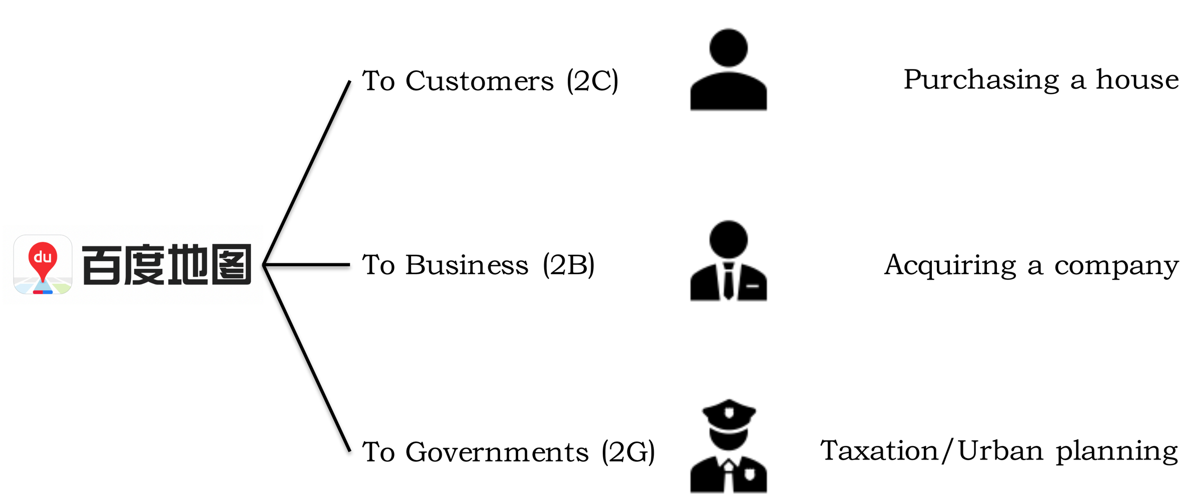

Figure 3: Creative applications serving customers (2C), business (2B), and governments (2G).

Solution

Model Overall, the formulation we propose to assess the value of a private property is:

$$\hat{h} = S(\mathbf{x}; \Theta) \times (\mathbf{w} \cdot F(\mathbf{D}; \Phi)). \quad (1)$$

Suppose that we have n instances of private properties in a training set Δ, the learning objective can be defined as:

$$\text{minimize} \quad \mathcal{L} = \sum_{i=1}^{n} \left(h^{(i)} - \hat{h}^{(i)}\right)^2, \quad (2)$$

where i is the index of the i-th instance in the training set, and $h^{(i)}$ stands for the ground-truth price of that instance.

Algorithm

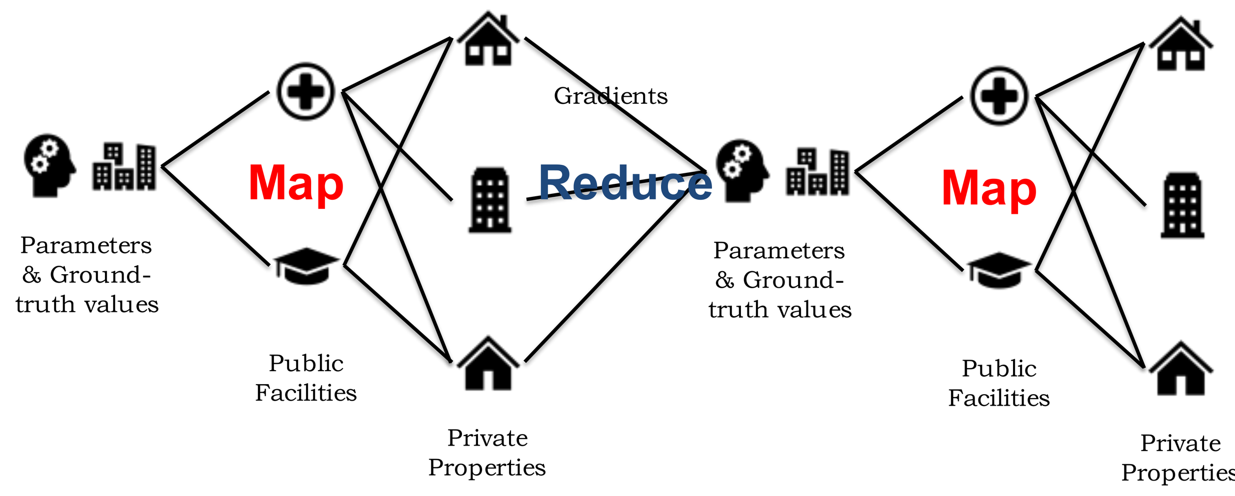

Figure 4: The distributed learning algorithm of "Monopoly" under the framework of MapReduce .

Assessment

Real-world Datasets

City	#(Res. Blocks)	#(Pub. Facilities)	#(Other Res. Blocks) per Res. Block	#(Pub. Facilities) per Res. Block
Beijing	7,573	843,426	1.5 (within 0.5 km) | 5.9 (within 1.0 km)	115.6 (within 0.5 km) | 296.3 (within 1.0 km)
Shanghai	11,604	970,566	3.6 (within 0.5 km) | 12.3 (within 1.0 km)	147.4 (within 0.5 km) | 313.9 (within 1.0 km)
Guangzhou	6,508	810,465	4.1 (within 0.5 km) | 14.4 (within 1.0 km)	213.0 (within 0.5 km) | 364.7 (within 1.0 km)
Shenzhen	3,849	724,320	2.9 (within 0.5 km) | 10.4 (within 1.0 km)	197.0 (within 0.5 km) | 371.9 (within 1.0 km)
AVERAGE	7,384	837,194	3.0 (within 0.5 km) | 10.8 (within 1.0 km)	168.3 (within 0.5 km) | 336.7 (within 1.0 km)
TOTAL	29,534	3,348,777	-	-

Table 1: The statistics of the urban data we used to conduct experiments. The datasets are collected from four metropolises (i.e., Beijing, Shanghai, Guangzhou, and Shenzhen) in China. For the other columns, #(Res. Blocks): the number of residential blocks; #(Pub. Facilities): the number of public facilities; #(Other Res. Blocks) per Res. Block: the average number of surrounding residential blocks per residential block; #(Pub. Facilities) per Res. Block: the average number of public facilities per residential block.

Experimental Results

Method	Mean Absolute Error (MAE)	Root Mean Squared Error (RMSE)	R^2 Score (R^2)
Avg. Prices (citywide)	$18,631 \pm 4,517\ \mathrm{CNY/m^2}$	$25,151 \pm 6,260\ \mathrm{CNY/m^2}$	0.0001 ± 0.0001
Macro Avg. Prices (within 0.5 km)	$15,110 \pm 3,928\ \mathrm{CNY/m^2}$	$21,593 \pm 5,826\ \mathrm{CNY/m^2}$	0.2668 ± 0.1580
Macro Avg. Prices (within 1.0 km)	$14,963 \pm 3,771\ \mathrm{CNY/m^2}$	$20,745 \pm 5,492\ \mathrm{CNY/m^2}$	0.3251 ± 0.1349
Micro Avg. Prices (within 0.5 km)	$15,089 \pm 3,889\ \mathrm{CNY/m^2}$	$21,589 \pm 5,793\ \mathrm{CNY/m^2}$	0.2690 ± 0.1572
Micro Avg. Prices (within 1.0 km)	$14,172 \pm 3,650\ \mathrm{CNY/m^2}$	$20,575 \pm 5,508\ \mathrm{CNY/m^2}$	0.3361 ± 0.1342
Linear Regression (within 0.5 km)	$12,764 \pm 3,250\ \mathrm{CNY/m^2}$	$18,897 \pm 5,086\ \mathrm{CNY/m^2}$	0.4400 ± 0.0414
Linear Regression (within 1.0 km)	$10,717 \pm 2,150\ \mathrm{CNY/m^2}$	$16,303 \pm 4,027\ \mathrm{CNY/m^2}$	0.5769 ± 0.0512
Boosting Trees (within 0.5 km)	$11,515 \pm 2,847\ \mathrm{CNY/m^2}$	$18,650 \pm 4,617\ \mathrm{CNY/m^2}$	0.5082 ± 0.0681
Boosting Trees (within 1.0 km)	$10,594 \pm 2,085\ \mathrm{CNY/m^2}$	$16,125 \pm 4,018\ \mathrm{CNY/m^2}$	0.5846 ± 0.0670
DNN (within 0.5 km)	$11,280 \pm 2,425\ \mathrm{CNY/m^2}$	$17,144 \pm 4,359\ \mathrm{CNY/m^2}$	0.5351 ± 0.0484
DNN (within 1.0 km)	$9,947 \pm 1,845\ \mathrm{CNY/m^2}$	$15,440 \pm 3,580\ \mathrm{CNY/m^2}$	0.6161 ± 0.0566
Monopoly (within 0.5 km)	$9,531 \pm 1,745\ \mathrm{CNY/m^2}$	$15,352 \pm 3,764\ \mathrm{CNY/m^2}$	0.6192 ± 0.0746
Monopoly (within 1.0 km)	$9,544 \pm 1,728\ \mathrm{CNY/m^2}$	$15,234 \pm 3,640\ \mathrm{CNY/m^2}$	0.6231 ± 0.0773

Table 2: The comparison results between our method on "Monopoly" and other baselines evaluated by the metrics of MAE, RMSE, and R^2 score. All the numbers (i.e., mean $\pm$ standard deviation) in this table are obtained by averaging the results of the four cities in Table 1.

Discoveries

Type of Public Facility	Average Virtual Price			
	Beijing	Shanghai	Guangzhou	Shenzhen
Governmental Agency	$(66,125) + 6,082\ \mathrm{CNY/m^2}$	$(55,670) + 3,833\ \mathrm{CNY/m^2}$	$(31,209) + 1,328\ \mathrm{CNY/m^2}$	$(59,478) + 183\ \mathrm{CNY/m^2}$
Educational Institution	$(66,125) + 4,441\ \mathrm{CNY/m^2}$	$(55,670) + 4,379\ \mathrm{CNY/m^2}$	$(31,209) + 2,501\ \mathrm{CNY/m^2}$	$(59,478) + 3,099\ \mathrm{CNY/m^2}$
Financial Institution	$(66,125) + 5,238\ \mathrm{CNY/m^2}$	$(55,670) + 4,918\ \mathrm{CNY/m^2}$	$(31,209) + 3,616\ \mathrm{CNY/m^2}$	$(59,478) + 1,603\ \mathrm{CNY/m^2}$
Recreational Facility	$(66,125) + 2,361\ \mathrm{CNY/m^2}$	$(55,670) + 3,222\ \mathrm{CNY/m^2}$	$(31,209) + 554\ \mathrm{CNY/m^2}$	$(59,478) + 2,863\ \mathrm{CNY/m^2}$
Medical Treatment	$(66,125) + 4,225\ \mathrm{CNY/m^2}$	$(55,670) + 3,440\ \mathrm{CNY/m^2}$	$(31,209) + 867\ \mathrm{CNY/m^2}$	$(59,478) + 112\ \mathrm{CNY/m^2}$
Commercial Office	$(66,125) + 1,313\ \mathrm{CNY/m^2}$	$(55,670) + 1,161\ \mathrm{CNY/m^2}$	$(31,209) - 146\ \mathrm{CNY/m^2}$	$(59,478) - 636\ \mathrm{CNY/m^2}$
Transportation	$(66,125) + 4,393\ \mathrm{CNY/m^2}$	$(55,670) + 2,750\ \mathrm{CNY/m^2}$	$(31,209) + 2,007\ \mathrm{CNY/m^2}$	$(59,478) + 3,278\ \mathrm{CNY/m^2}$
Scenic Spot	$(66,125) + 6,425\ \mathrm{CNY/m^2}$	$(55,670) + 5,415\ \mathrm{CNY/m^2}$	$(31,209) + 1,055\ \mathrm{CNY/m^2}$	$(59,478) + 3,855\ \mathrm{CNY/m^2}$
Wasteyard	$(66,125) - 7,647\ \mathrm{CNY/m^2}$	$(55,670) - 6,221\ \mathrm{CNY/m^2}$	$(31,209) - 1,873\ \mathrm{CNY/m^2}$	$(59,478) - 1,043\ \mathrm{CNY/m^2}$
Cemetery	$(66,125) - 4,129\ \mathrm{CNY/m^2}$	$(55,670) - 8,276\ \mathrm{CNY/m^2}$	$(31,209) - 569\ \mathrm{CNY/m^2}$	$(59,478) - 2,607\ \mathrm{CNY/m^2}$

Table 3: The virtual prices of the public facilities acquired by "Monopoly" in the four metropolises of China.

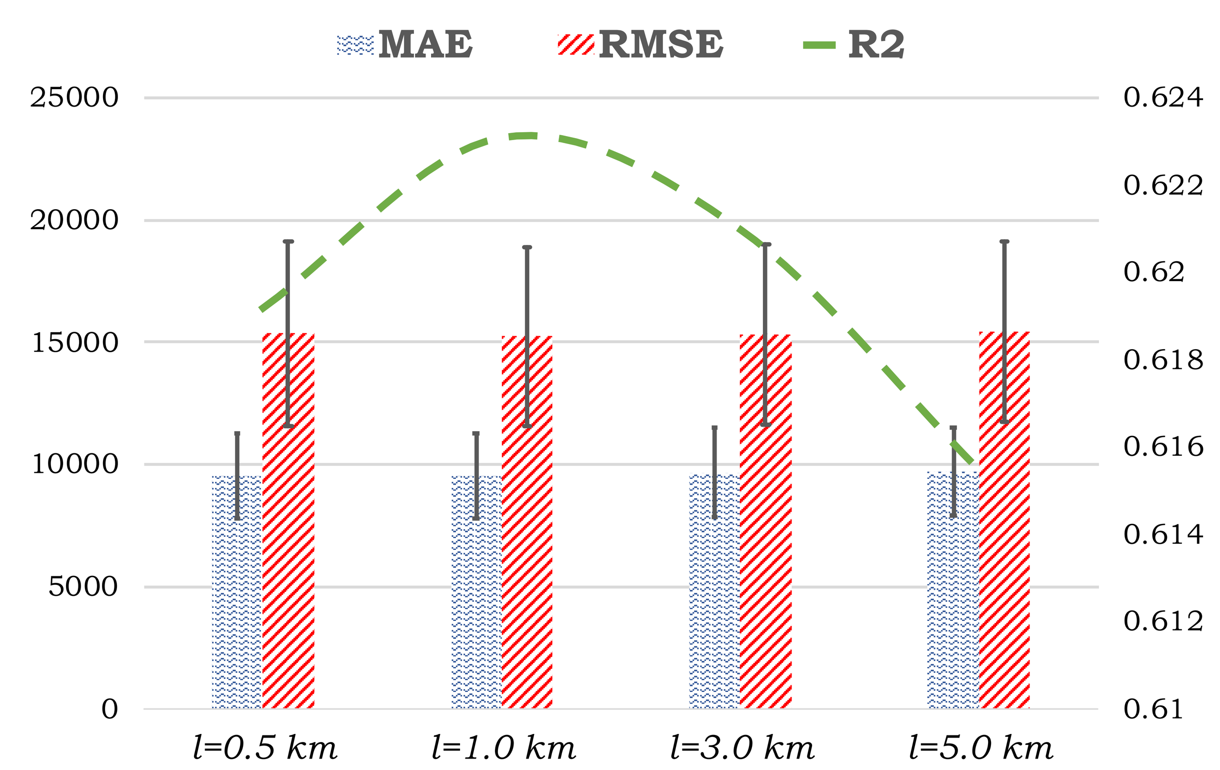

Figure 5: An illustration on the average performance (i.e., mean $\pm$ standard deviation) of "Monopoly" measured by MAE and RMSE, along with different values (i.e., 0.5 km, 1.0 km, 3.0 km, and 5.0 km) of influencing radius.

Downloads

CONTACT US (pwd: w166) PAPER CODES

The 28th ACM International Conference on Information and Knowledge Management (CIKM'19), November 3–7, 2019, Beijing, China