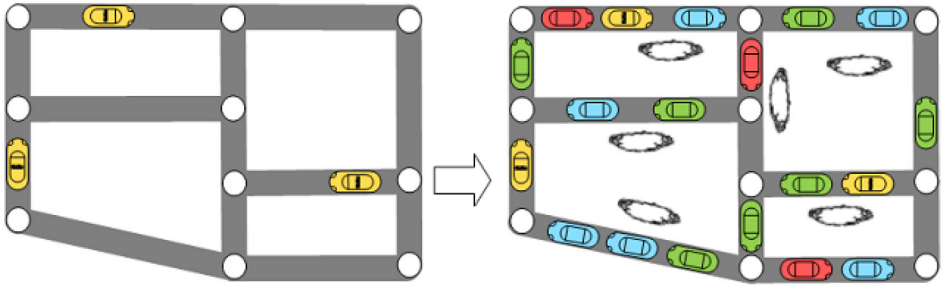
Inferring Gas Consumption and Pollution Emission of Vehicles throughout a City

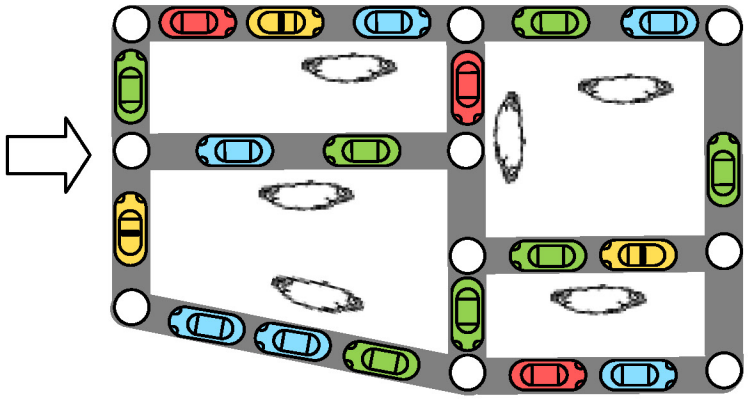
Dr. **Yu Zheng** yuzheng@microsoft.com
Lead Researcher, Microsoft Research
Chair Professor at Shanghai Jiao Tong University



Questions

How many liters of gas have been consumed by the vehicles, in the entire city, in the past one hour?

What is the volume of CO that has been generated accordingly?

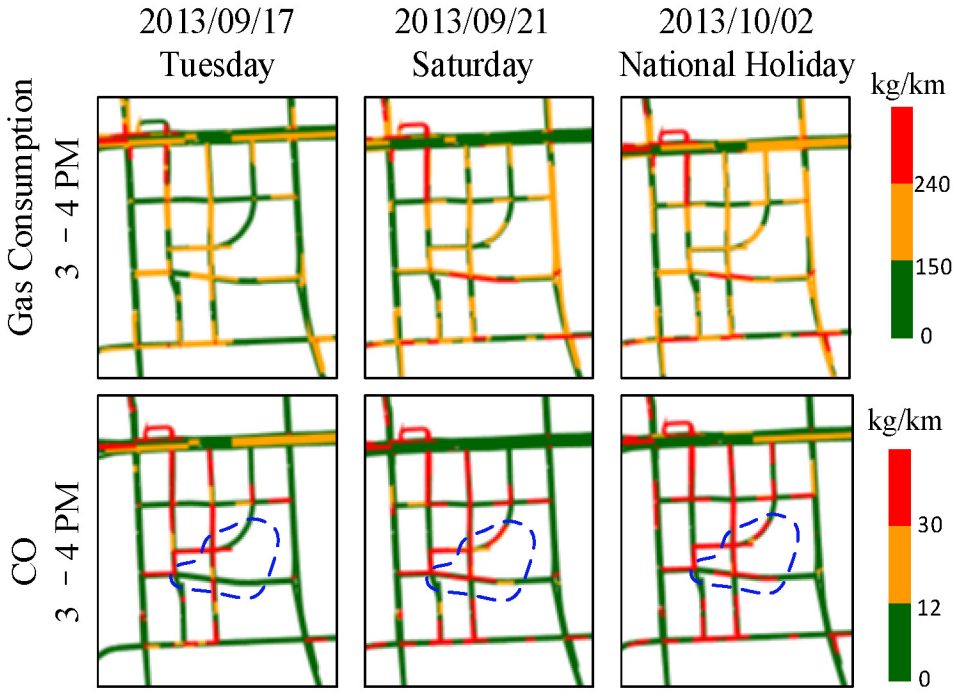


Download data here

Jingbo Shang, **Yu Zheng,** et al. Inferring Gas Consumption and Pollution Emission of Vehicles throughout a City. **KDD 2014**.

Goals

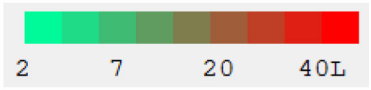
- Estimate the gas consumption and vehicle emissions
  - on arbitrary road segment
  - at any time intervals
  - using GPS trajectories of a sample of vehicles



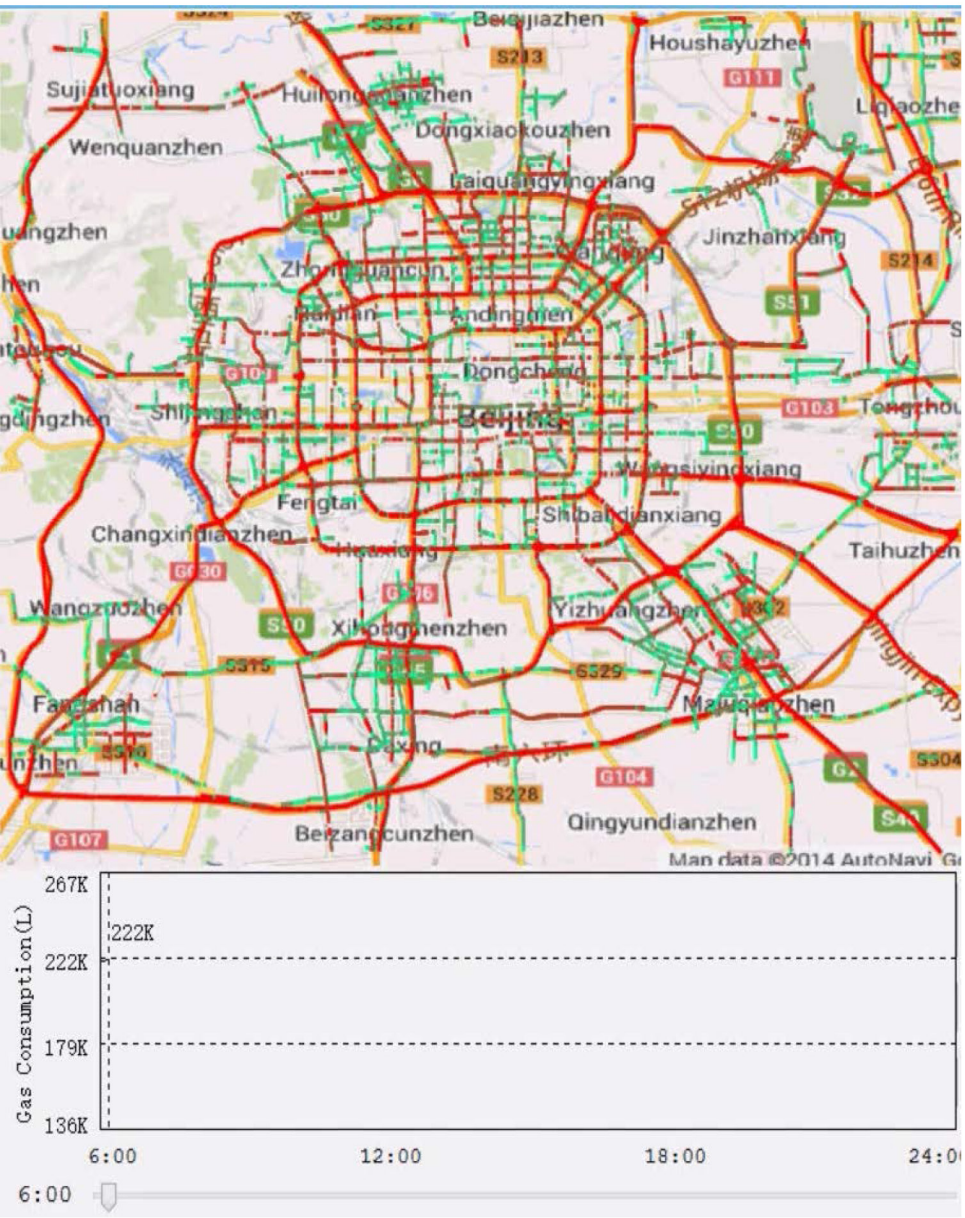
Jingbo Shang, **Yu Zheng,** et al. Inferring Gas Consumption and Pollution Emission of Vehicles throughout a City. **KDD 2014**.

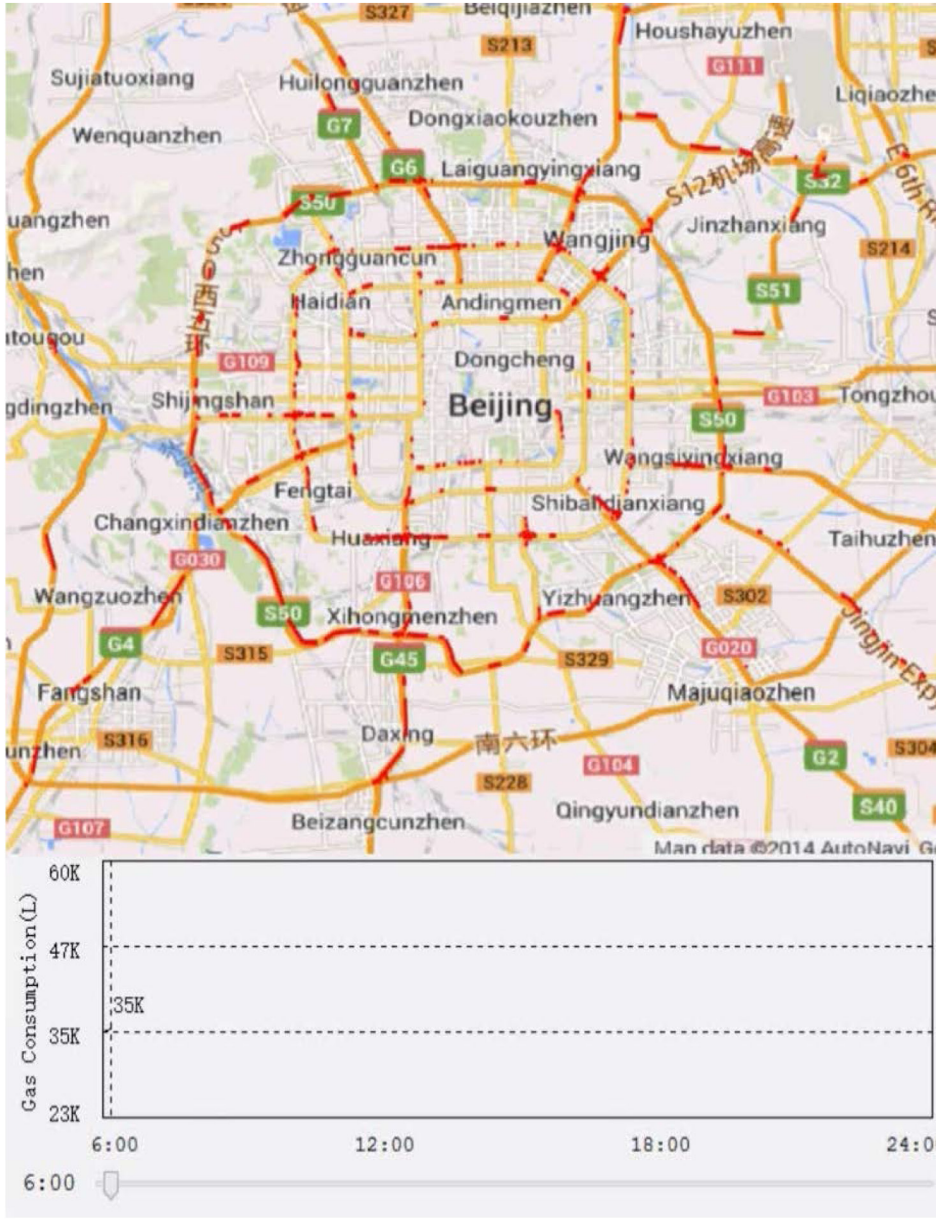
Real-Time Gas Consumption

Citywide



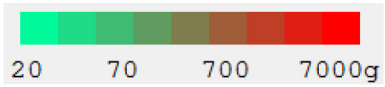
Top 3%



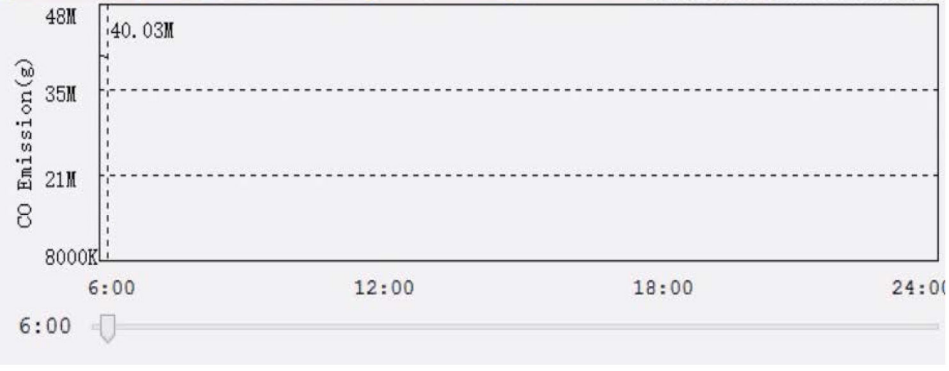


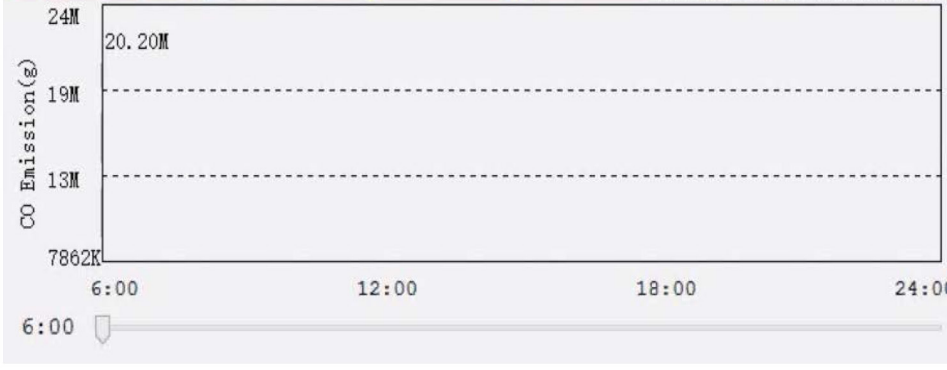
Real-Time CO Emission

Citywide



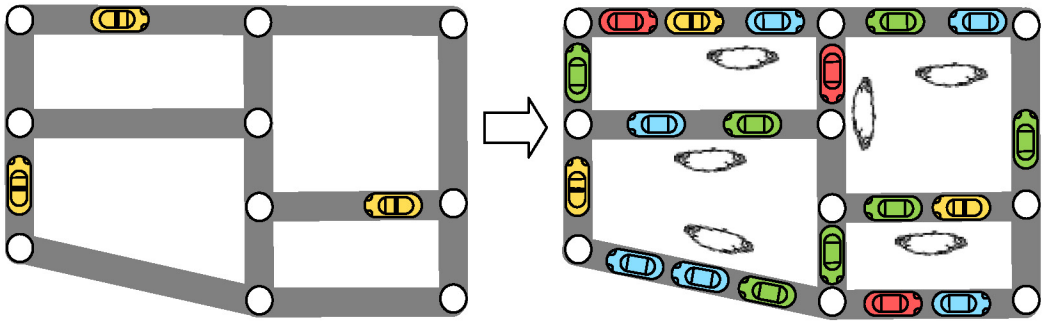
Top 4%



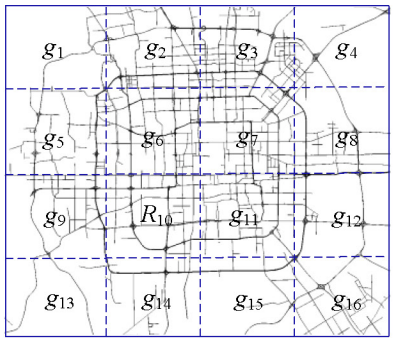


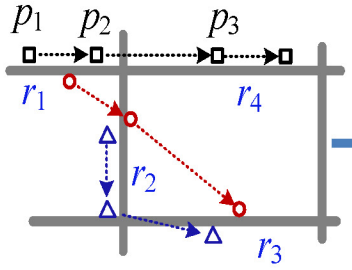
Our Approach

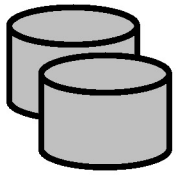
- Using the GPS trajectories of a sample of vehicles
  - Travel Speed Estimate (**TSE**) on each road segment
  - Traffic Volume Inference (**TVI**) for each road segment
  - Calculate the gas consumption and emission of vehicles

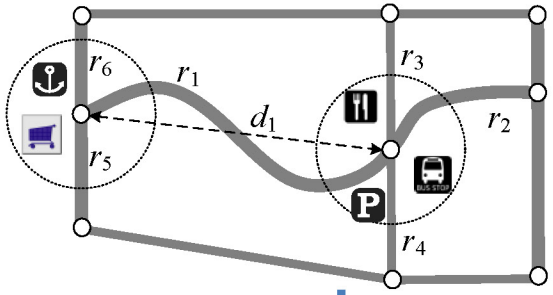


Travel Speed Estimation (TSE)









$$Y = \begin{array}{c} \\ t_i \\ t_{i+1} \\ \vdots \\ t_j \end{array} \begin{array}{c} \begin{array}{cccc} g_1 & g_2 & \cdots & g_{16} \end{array} \\ \left[\begin{array}{c|c} M'_G & M_G \end{array}\right] \end{array} \quad X = \begin{array}{c} \\ t_i \\ t_{i+1} \\ \vdots \\ t_j \end{array} \begin{array}{c} \begin{array}{cccc} r_1 & r_2 & \cdots & r_n \end{array} \\ \left[\begin{array}{c|c} M'_r & M_r \end{array}\right] \end{array} \quad Z = \begin{array}{c} \\ r_1 \\ r_2 \\ \vdots \\ r_n \end{array} \begin{array}{c} \begin{array}{ccc} f_1 & f_2 & \cdots \end{array} \\ \left[\begin{array}{c|c|c} f_r & f_p & f_g \end{array}\right] \end{array}$$

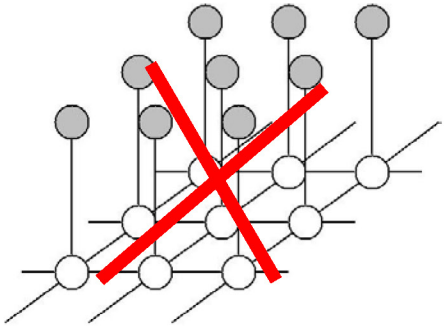
$$Y \approx T \times (G;G)^T;\ X \approx T \times (R;R)^T;\ Z \approx R \times F^T$$

- X: denotes fine-grained traffic conditions
- Y: denotes coarse-grained traffic conditions
- Z: geographical contexts of road segments

$$L(T,R,G,F) = \frac{1}{2}||Y - T(G;G)^T||^2 + \frac{\lambda_1}{2}||X - T(R;R)^T||^2 + \frac{\lambda_2}{2}||Z - RF^T||^2 + \frac{\lambda_3}{2}(||T||^2 + ||R||^2 + ||G||^2 + ||F||^2),$$

Traffic Volume Inference (TVI)

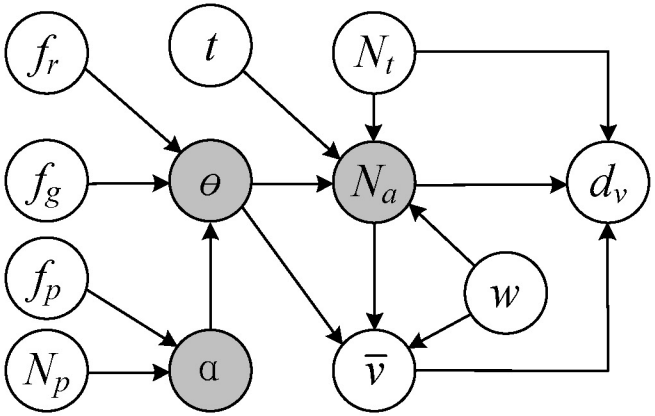
- Objective
  - Given the travel speed of a road segment and other observations
  - Infer the volume of traffic on the road segment
- Difficulties
  - Biased distribution between taxis and other vehicles
  - Traffic volume depends on
    - the current travel speeds and density of vehicles
    - the length, shape, and capacity of a road, weather conditions
  - Not enough training data (complex models do not work)



Traffic Volume Inference (TVI)

Unsupervised learning approach

- N_a: the number of vehicles per minute per lane
  - The volume of observed sample vehicles N_t
  - Weather conditions w
  - Time of day t
- $\bar{v}$ the average travel speed
- dv The variance of the speed
- The type of the road θ
  - Road network features f_r
  - Global position feature f_g
- surrounding POIs α
  - Distribution of POIs f_p
  - total number of POIs N_p



Gray node: hidden variable
White nodes: observations

Energy and Emission Calculation

Adopt environmental theories: European-3 standards

Average vehicle capacity: 1.4-2.0L

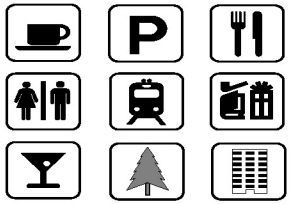
$$EF = (a + cv + ev^2)/(1 + bv + dv^2).$$

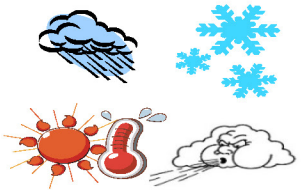
	a	b	c	d	e
CO	71.7	35.4	11.4	-0.248	0
Hydrocarbon	5.57×10^{-2}	3.65×10^{-2}	-1.1×10^{-3}	-1.88×10^{-4}	1.25×10^{-5}
Nox	9.29×10^{-2}	-1.22×10^{-2}	-1.49×10^{-3}	3.97×10^{-5}	6.53×10^{-6}
Fuel Consumption	217	9.6×10^{-2}	0.253	-4.21×10^{-4}	9.65×10^{-3}

$$E = EF \times r.N_a \times r.n \times r.len$$

Experiments

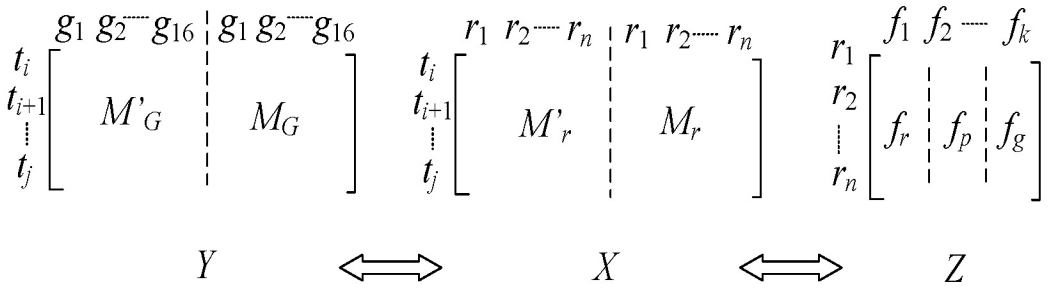
- *Road networks*:
  - 148,110 nodes and 196,307 edges
  - total length of road segments: 21,895km
- *POIs*:
  - 273,165 POIs (195 tier two categories)
  - choose the top 10 most frequent categories nearby road segments
- *GPS Trajectories*:
  - generated by 33,000 Beijing taxis over 47 days
  - 673,469,757 GPS points
  - total length is over 26,218,407km
  - average sampling rate is 96 seconds per point
- *Weather data*
  - Sunny, windy, rainy, overcast…





Experiments

- Evaluation on TSE



Methods	RMSE of $\overline{v}$	RMSE of dv	Time (sec)
MF(M'_r)	2.172	1.833	2.2
MF($M'_r + Z$)	1.939	1.385	18.2
MF($M'_r + M'_g + Z$)	1.908	1.314	20.2
TSE	**1.369**	**1.035**	**22.2**
Kriging	2.340	1.300	1,000
KNN	3.360	1.590	0.14

RMSE	level 0-1	level 2	level 3	level >=4
Speed	2.575	1.189	0.977	1.612
Variance	1.526	1.128	0.628	1.234

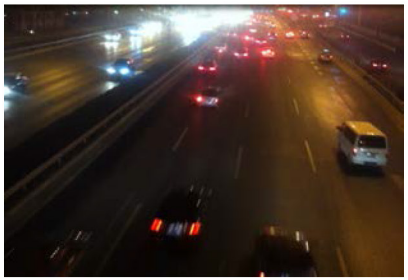
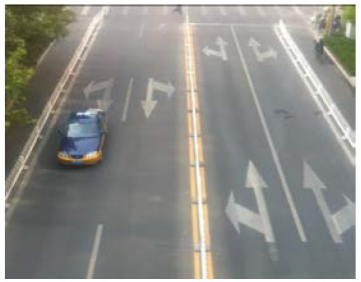
Experiments

- Evaluation on TVI

[Download data here]

Methods	MAE	MRE	Inference time (us/road)
TVI	3.01	29%	7.27
TVI w/o dv	3.19	31%	7.18
TVI w/o w	3.15	29%	7.10
LR	3.06	27%	0.15
FD-SC	3.9	42%	0.13

	level 0-1	level 2	Weekday	Weekend
MAE	5.55	2.23	2.97	3.28

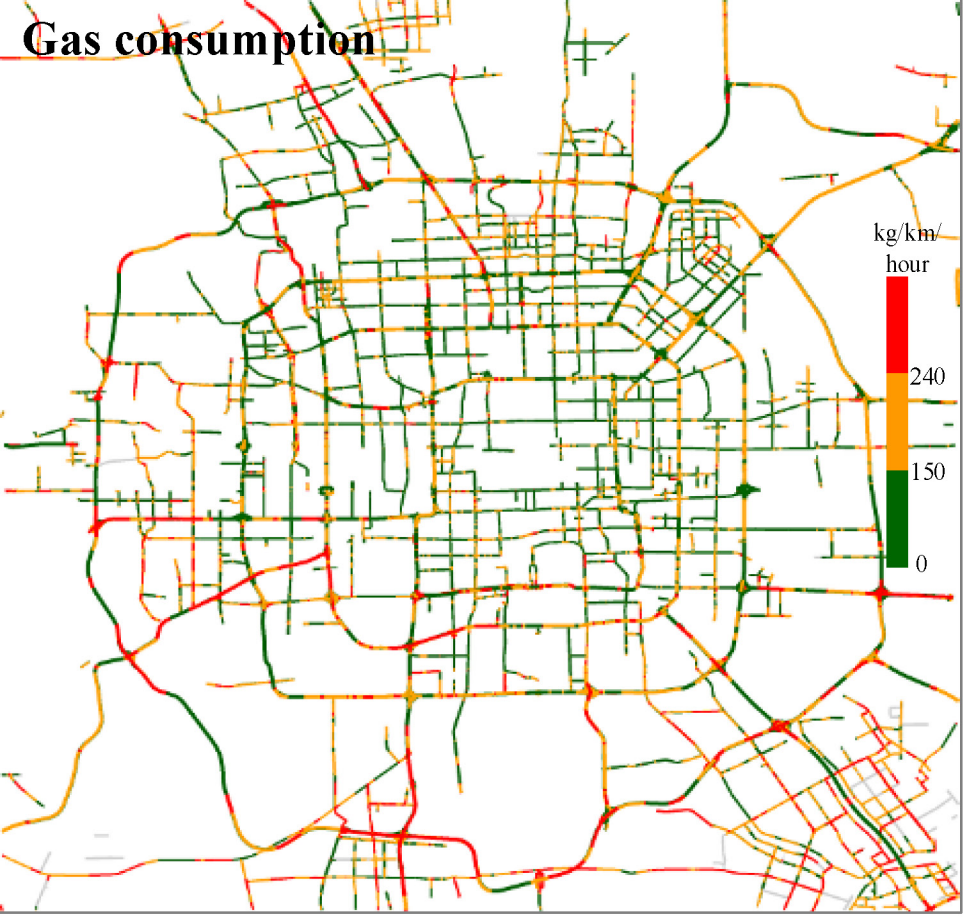




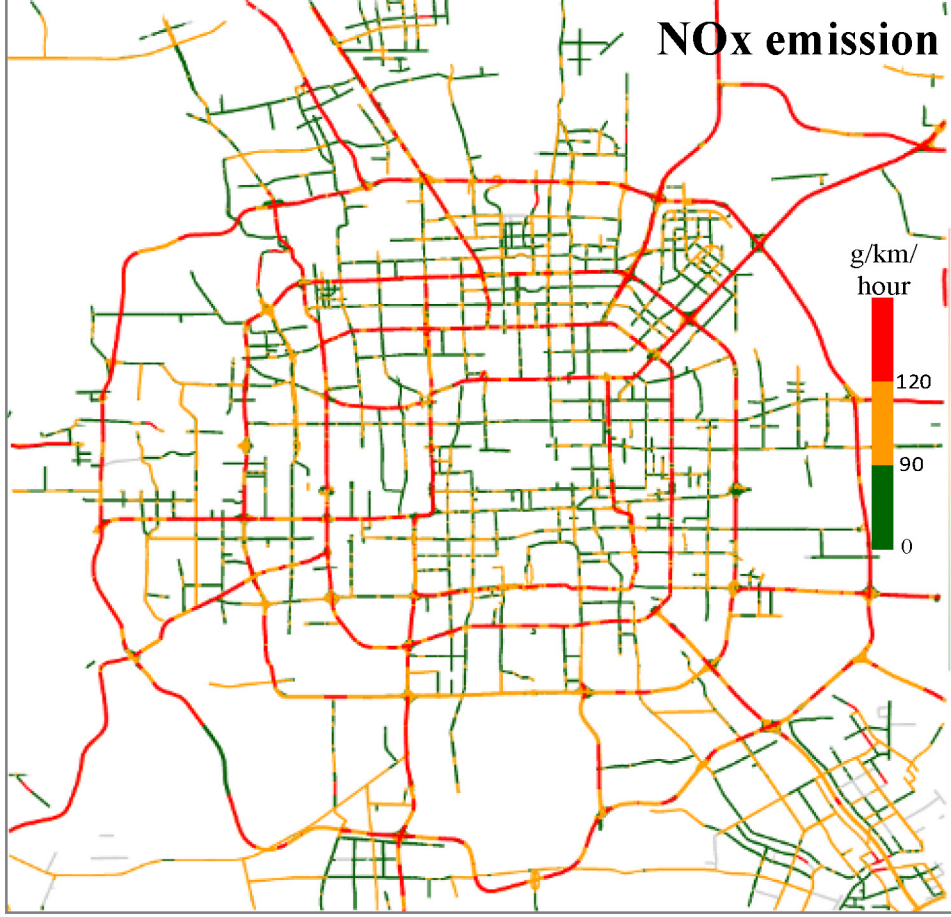
Time	7:00 ~ 10:00			10:00~16:00			16:00~20:00			after 20:00			total
Level	0,1	2	3	0,1	2	3	0,1	2	3	0,1	2	3	
Holiday	0	0	0	6	14	4	6	8	1	4	6	0	49
Workday	7	28	8	29	74	9	28	92	7	6	17	4	309
Total	43			136			142			37			358

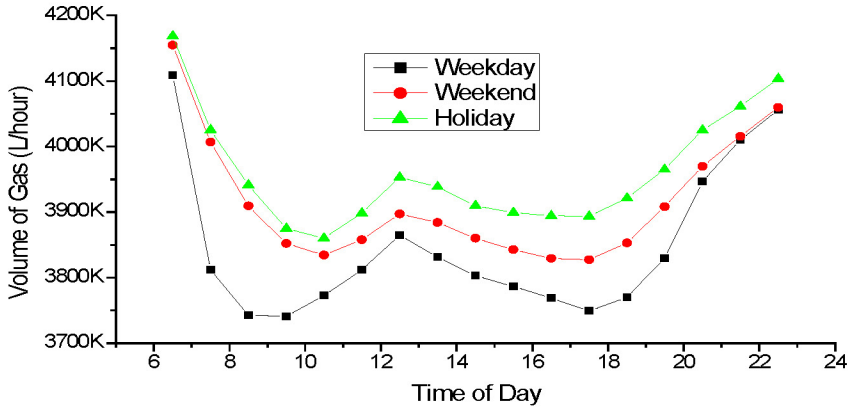
Experiments

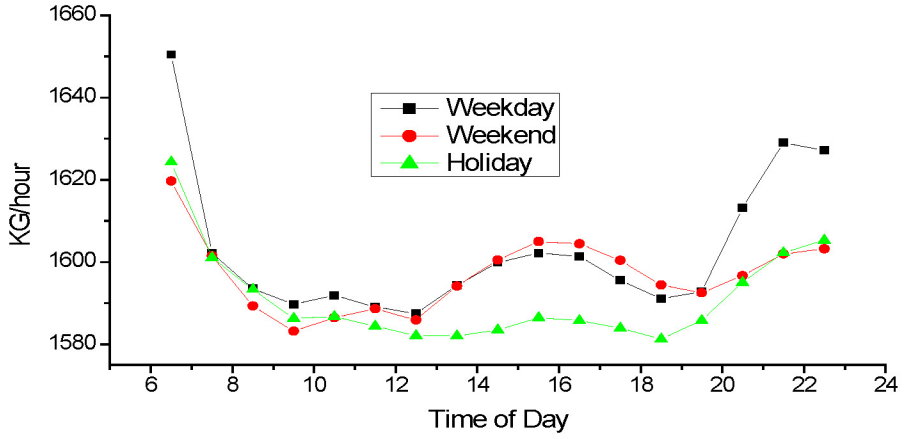
- Efficiency
  - Tested on a single machine
  - About 5 minutes to process 10-minute data

Online components	Time	Offline components	Time
Map-matching	4.94min	Geo-feature extraction	149s
TSE	22.2s	Historical pattern extraction	240s
TVI (inference)	0.84s	TVI learning	89s
Total	**5.32min**	Total	478s


Gas consumption
kg/km/
hour
240
150
0


NOx emission
g/km/
hour
120
90
0


Volume of Gas (L/hour)
4200K
4100K
4000K
3900K
3800K
3700K
6
8
10
12
14
16
18
20
22
24
Time of Day
Weekday
Weekend
Holiday


KG/hour
1660
1640
1620
1600
1580
6
8
10
12
14
16
18
20
22
24
Time of Day
Weekday
Weekend
Holiday

Conclusion

- Infer the citywide gas consumption and vehicle emission
  - Using a sample of vehicles
  - 5 minutes on a single machine
  - With an accuracy of 70%+
- Deal with data sparsity and data bias

Download data here

Urban Computing

Search for “Urban Computing”

Thanks!

Yu Zheng

yuzheng@microsoft.com

Homepage