Dissecting Urban Noises from Heterogeneous Geo-Social Media and Sensor Data

Hsun-Ping Hsieh
National Taiwan University
Taipei, Taiwan
sandoh714@gmail.com

Tzu-Chi Yen
Sensoro, Inc.
Beijing, China
junipertcy@gmail.com

Rui Yan
Baidu, Inc.
Beijing, China
yanrui02@baidu.com

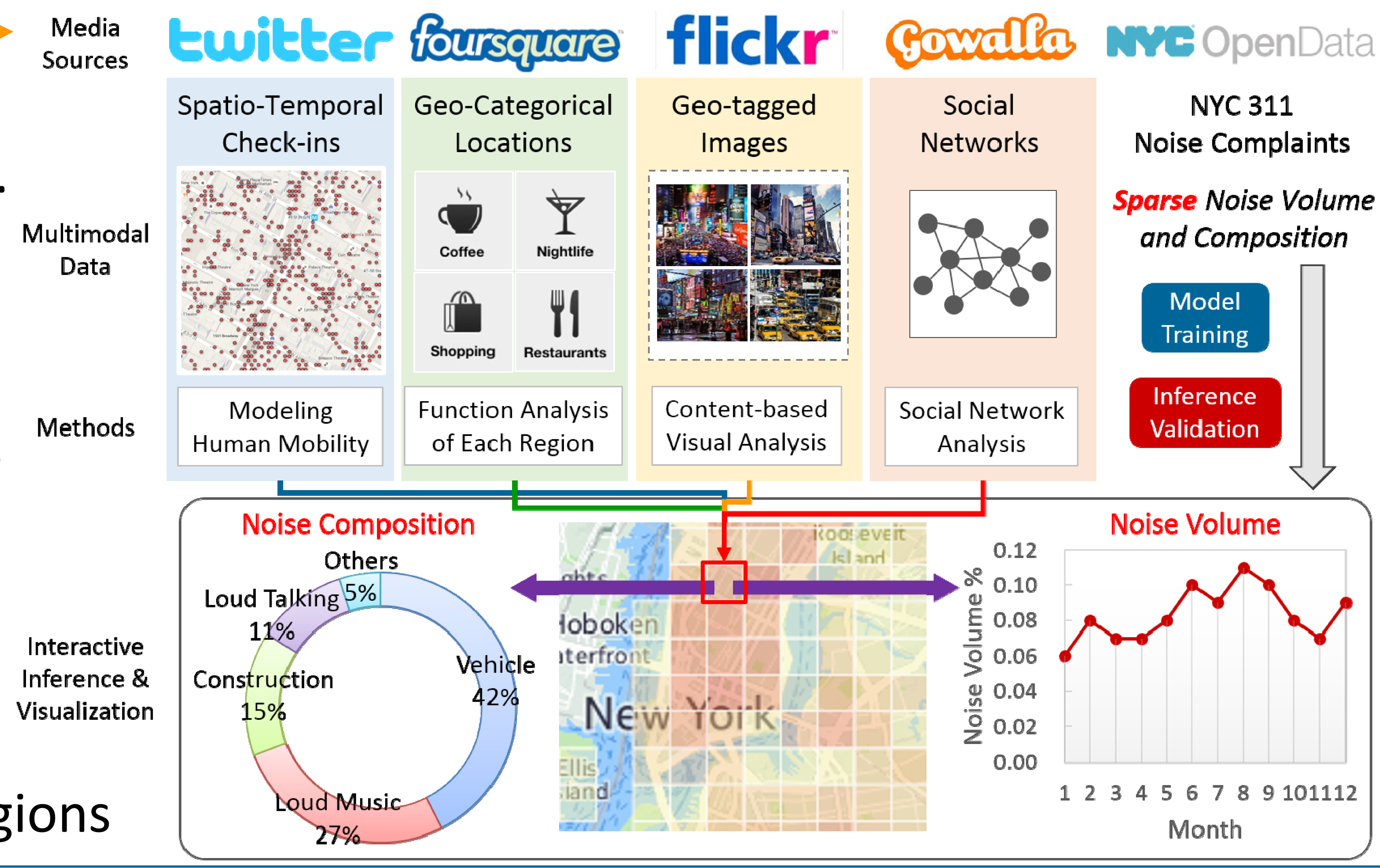
Cheng-Te Li
Academia Sincia
Taipei, Taiwan
ctli@citi.sinica.edu.tw

Motivation

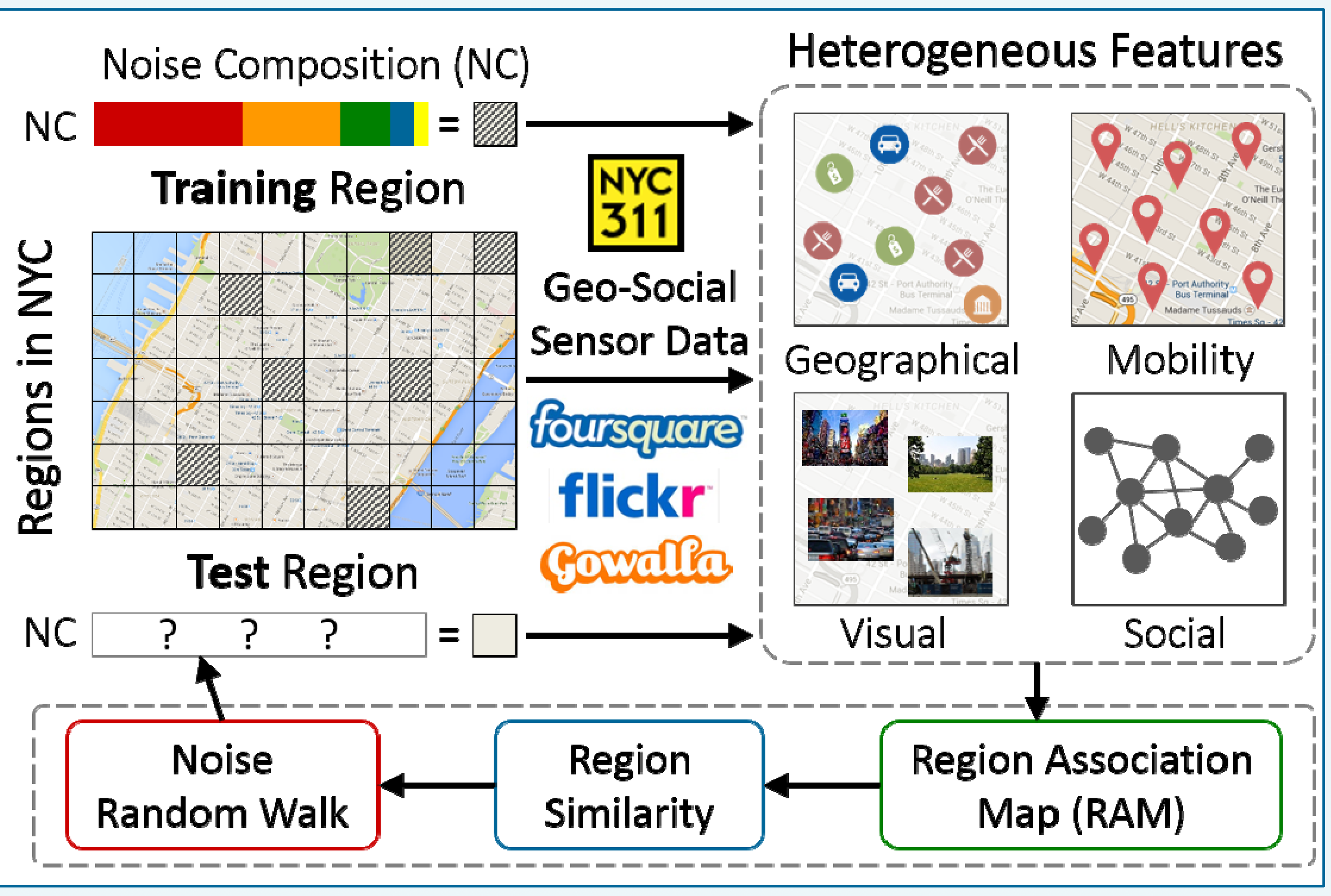
- **Geo-social media** can be considered as a kind of sensors that monitor urban human activities.
- Urban noise pollution damages the mental health (e.g. work efficiency and sleep quality).
- We need to first understand the elements and causes that produce the noise. E.g. @TimeSquare *"in evening rush hour, 60% noise come from vehicle traffic, 30% from loud music of stores, and 10% from people' talking"*

Problem Statement

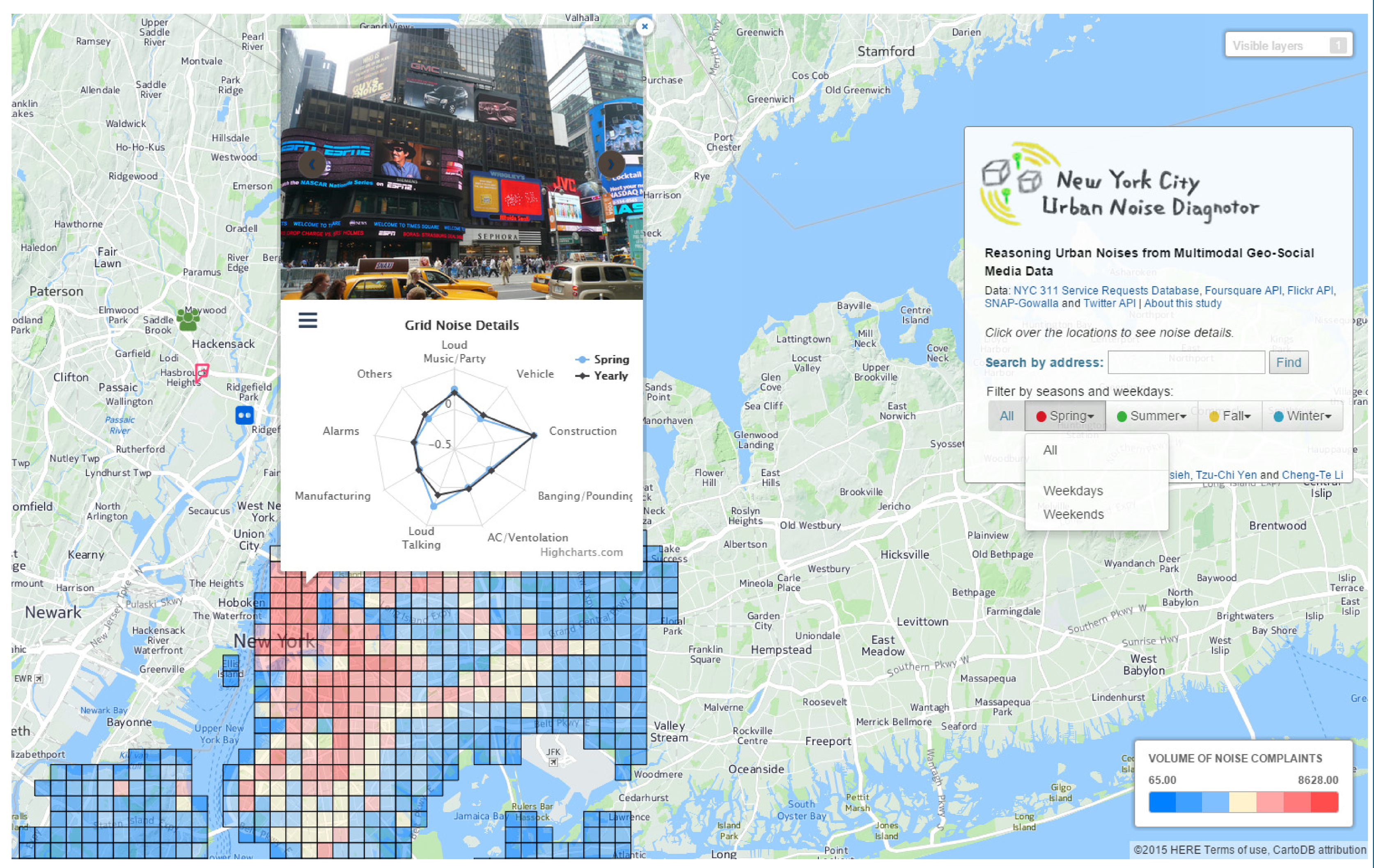
- Given a few NYC regions whose noise info are known, infer the noise composition of other regions



Approach Overview



System Demonstration

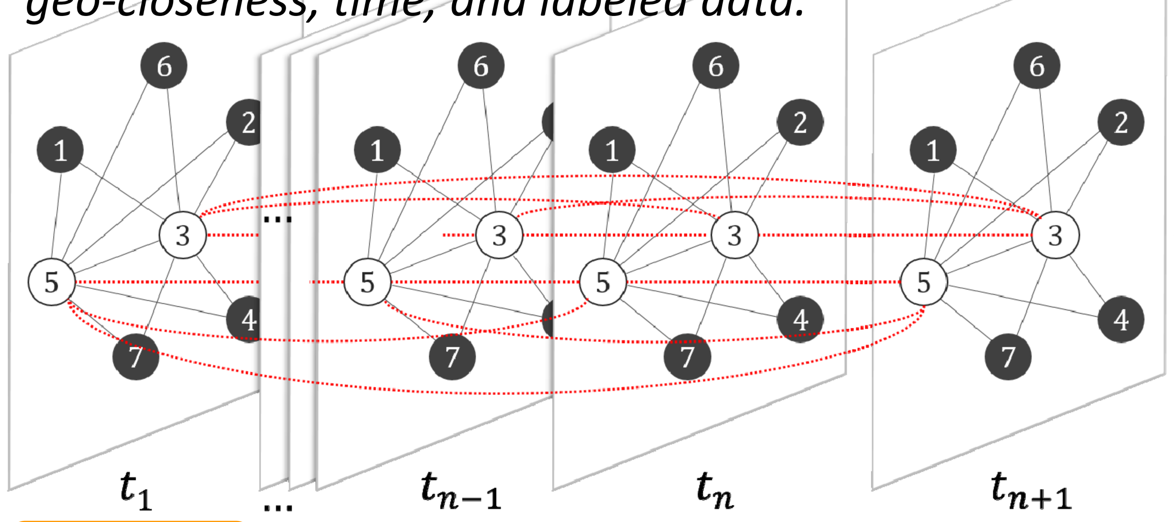


Step 1 Extracting Geo-Social Features

Geographical	Mobility	Visual	Social
Region Density, Venue Closeness, Category Entropy, Competitiveness	Region Popularity, Movement Density, Incoming Flow	Color Histogram, Color Moments, Gabor Textures, SIFT features	PageRank, SimRank, Spatial Centralities, #Communities, CC, density

Step 2 Region Association Map

A region is a node. Edges are constructed based on geo-closeness, time, and labeled data.



Step 3 Region Similarity

$$EdgeWeight(u,v) = RegSim(u,v) = \exp\left(-\sum_{k=1}^{m} \pi_k \times \|\mathbf{f}_k(u) - \mathbf{f}_k(v)\|\right)$$

totally k features

Feature Weights Learned from validation set!

Feature Vector of feature k for node v

Step 4 Noise Random Walk in RAM

$$n_c(v) \leftarrow \sum_{u \in \Gamma(v)} w_{v,u} \times n_c(u)$$

Edge weight

Neighbor nodes of v

- Iteratively update each unlabeled node based on the noise composition of its neighbors

Experimental Results

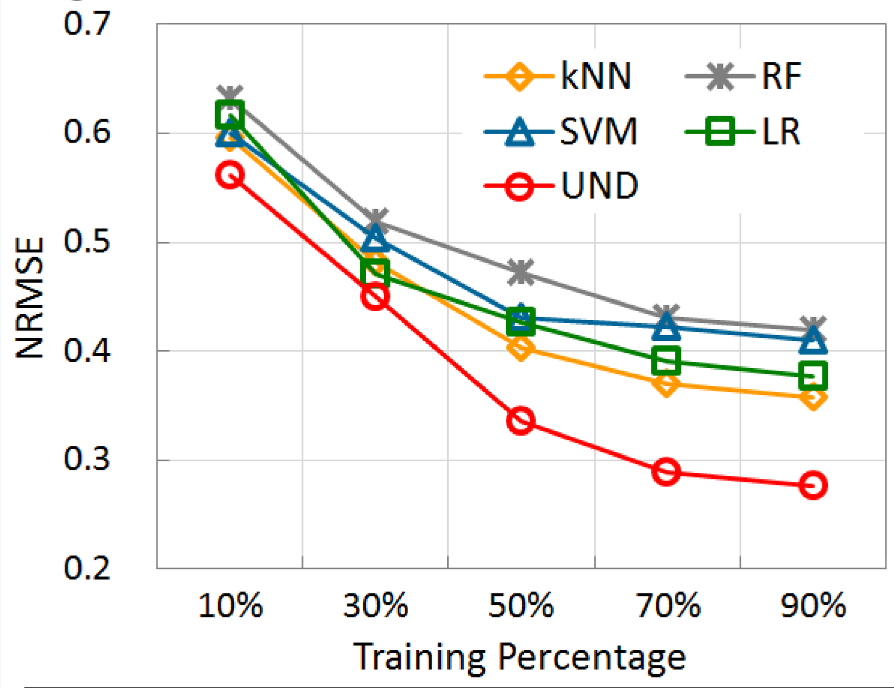
560 regions in NYC

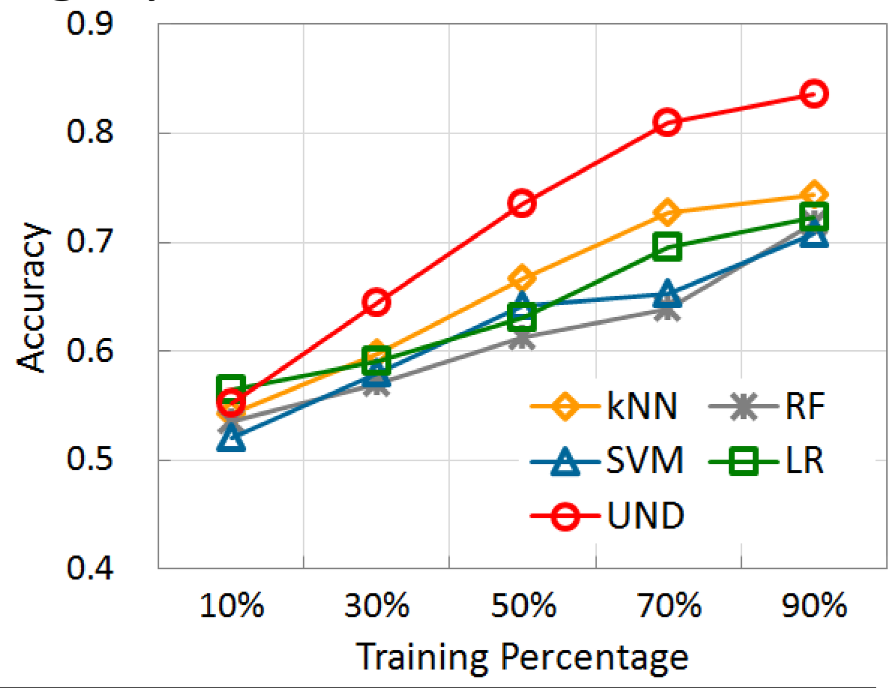
- NYC 311 Noise Compliant Data – 14 noise categories *e.g. Loud Music/Party, Construction, Loud Talking, Vehicle*
- 47,581 venues, 196,591 users, 6,442,890 check-ins

① Measure the error between the ground-truth and the inferred noise values

5% data used for tuning parameters

② Whether the main noise category is identified





	Days		Hour Intervals					
	Weekday	Weekend	00-03	04-07	08-11	12-15	16-19	20-23
RMSE	0.118	0.159	0.167	0.061	0.081	0.075	0.106	0.219
Accuracy	0.909	0.858	0.853	0.923	0.904	0.911	0.895	0.836