# Diagnosing New York City’s Noises with Ubiquitous Data

# Data definition

### Road network:

##### A set of road segment

### POI

##### Having a name, address, coordinates, category, and other attributes

### User Check-in

##### A time stamp and a geospatial coordinate, usually associated with a POI category

### Noise Complaint

##### A timestamp ns.time

##### A location ns.l (denoted by a (latitude, longitude) and a street address)

##### A complaint category ns.c

# Related work

## ||Urban noise sensing||

### Using wireless sensor networks to monitor environmental noise pollution in urban areas

##### Very expensive, in terms of money and human resources

###### 08-Data collection in wireless sensor networks for noise pollution monitoring

###### 08-First experiences using wireless sensor networks for noise pollution monitoring

###### 98-Wireless sensor networks for environmental noise monitoring

### Leverage crowdsourcing

##### People share information using a mobile device

10-A survey of mobile phone sensing

12-Towards crowd-aware sensing platform for metropolitan environments

13-UbiComp-Crowd++: unsupervised speaker count with smartphones

##### Leverages noise measurements shared by mobile phone users to paint a noise map

09-NoiseTube: Measuring and mapping noise pollution with mobile phones

9th-Participatory noise mapping

10-Crowdsourcing of pollution data using smartphones

##### The noise map generated through this approach is usually very sparse

Recover from incomplete and random samples based on compressive sensing

10- Ear-phone: an end-to-end participatory urban noise mapping system

### Paper contribution

##### Beyond raw sensor data, the 311 data we use does not indicate the noise levels but also people’s reaction and tolerance to noise

##### Explores the distribution of noises over different categories. The information can inform decision on tackling noise pollution

##### When dealing with the data sparsity problem, we incorporate other data sources

## ||Noise Understanding||

### Propose a statistical framework for a noise event recognition, including noises from cars, trucks and airplanes

###### 97-Environmental sound recognition: a statistical approach

### Proposed a hidden Markov model (HMM)-based classifier to recognize five noise events (car, truck, moped, aircraft and train), which is claimed to be better than human listeners

###### 98-Automatic classification of environmental noise events by hidden Markov models

### Use a HMM-based strategy capable of classifying ten environments

###### 03-Environmental noise classification for context-aware applications

### SoundSense combines supervised and unsupervised learning techniques to classify not only general sound types but also novel sound events

###### 09- SoundSense: scalable sound sensing for people-centric applications on mobile phones

### Paper contribution

##### The proportion of 311 complaints of different noise categories in a location could well describe the composition of the noises in the location

##### Applying such classification methods to a major city like NYC needs a huge volume of training data that covers different locations and time spans

## ||Tensor decomposition for urban computing||

### Papers

###### 13-UbiComp-Fine-grained preference-aware location search leveraging crowdsourced digital footprints from LBSNs

###### 14-KDD-Travel Time Estimation of a Path using Sparse Trajectories

###### 10-AAAI- Collaborative Filtering Meets Mobile Recommendation: A User-centered Approach

###### 12-JAI-Towards mobile intelligence: Learning from GPS history data for collaborative recommendation

###### 13-UbiComp-Sensing the pulse of urban refueling behavior

# Framework

## 

# Data analysis

### 311 Data about noises

##### Categories

##### Difference between weekdays and weekends

##### Relation between complaints and real noises

##### 

14-MST-TR- Methods for sensing urban noises

##### 

### User check-in data

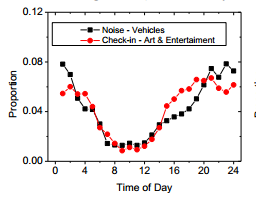
##### Data sources

Gowalla 127558 (4/24/2009~10/13/2013)

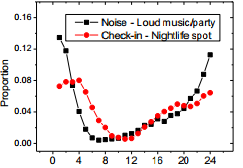
Foursqurare 173275 (5/5/2008~7/23/2011)

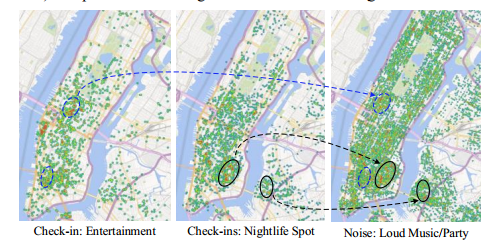
##### Strong correlation

Between Art & Entertainment check-ins and complaints about vehicles



Between nightlife spot check-ins and complaints about Loud music/Party





### Road Network and POIs

##### Examples

If a region has many bars, the amount of loud music and talking tend to be high

The structure of a road network in a region also has a strong correlation with the region’s traffic patterns

##### 

# Noise inference

### Map Segmentation

##### Partition NYC’s map into disjoint regions by major roads (s.lev<5)

##### 

###### 12-MSR-TR- Segmentation of urban areas using road networks

##### Other methods: uniform-grid, ZIP code

### Tensor Construction

Region dimension:

Time span dimension:

Category dimension:

An entry: An entry stores the total number of 311 complaints of category 𝑐𝑗 in region 𝑟𝑖 and time slot 𝑡 𝑘 over the given period of time (normalized to [0,1])

##### A common approach to fill the missing entries of tensor

Minimize

A core tensor

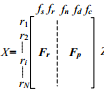
Three matrices:, is very small

Recovery

09- Tensor decompositions and applications

### Context-aware matrix decomposition

##### The geographical feature

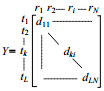


POI features: total number of POIs, density of POIs and distribution of POI over 15 categories

Road network features : the number of intersections and the total length of road segment in different levels

Matrix incorporates the similarity between two regions in terms of their geographic features   
(Assumption: similar geographic features could have a similar noise situation)

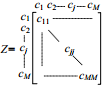
##### Human mobility features

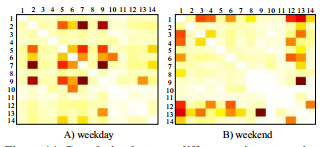


Entry : the number of check-ins generated in region , time slot

Matrix reveals the correlation between different time slots in terms of the distribution of check-ins over different regions  
(Time slots sharing a similar check-in pattern could have similar noise situation)

##### The correlation between different noise categories



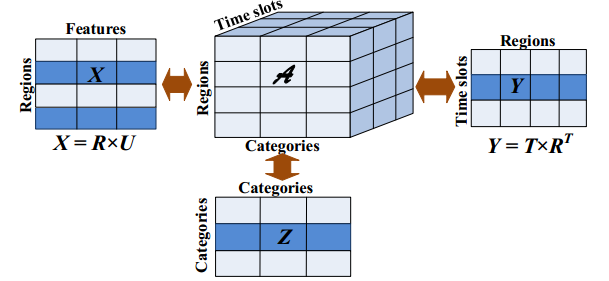


For a complaint record (of the 𝑖th category), we count the complaints of other categories within a circle distance to 𝑛𝑠

is the collection of 311 data

Though tensor 𝒜 can capture the correlation between different noise categories to some extent, matrix 𝑍 can further intensify the correlation

##### Decomposition



Minimize

A core tensor

Three matrices:, is very small

avoid over-fitting

Interpretation of : guarantees two noise categories with a higher similarity should also have a closer distance between the vectors

Recovery

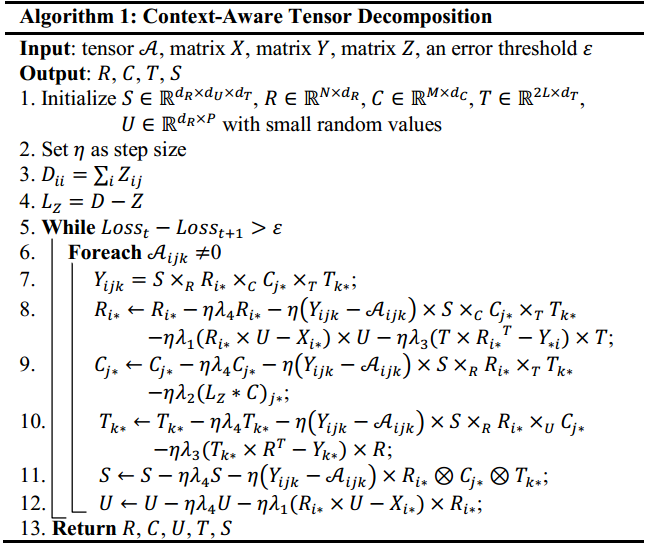
Explanation

and share matrix

and share matrix and

influences matrix

Algorithm



10--Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering   
(element-wise optimization algorithm)