Collective and Semantic Exploration of Human Mobility Data
—Modeling, Representation, and Applications

Yanjie Fu
Outline

- **Background and Motivation**
  - Modeling Spatiotemporal Dynamics
  - Collective Representation Learning
  - Applications
  - Conclusion and Future Work
Pervasive Sensing for Human Movements

IoT, GPS, wireless sensors, mobile Apps
Human Mobility Data

- **Human mobility data** are people’s movement trajectories which can be the traces of:
  - **devices**: phones, WIFIs, network stations, RFID
  - **vehicles**: bikes, taxicabs, buses, subways, light-rails
  - **location based services**: geo-tweets (Facebook, Twitter), geo-tagged photos (Flickr), check-ins (Foursquare, Yelp)

Represent the **spatial**, **temporal**, **social**, and **semantic** contexts of **dynamic human behaviors** within and across regions.
Important Applications of Human Mobility

- Understand Human Movement Patterns
- Automated User Profiling
- Intelligent Transportation Systems
- Smart and Connected Communities
- Smart Health Care
- City Governance and Emergency Management
Unprecedented and Unique Complexity

- **Spatio-temporal-textual**
- **Networked**
  - Collectively-related
- **Heterogeneous**
  - Multi-source
  - Multi-domain
  - Multi-format
- **Semantically-rich**
  - Trip purposes
  - User profiles
  - Outlier events/incidents
  - Spatial configuration and urban functions of regions
The Overview of The Talk

Collective and Semantic Exploration

Modeling spatial diffusion and temporal dynamics

Modeling

Representation

Collective representation learning of urban regions with multi-source data

Applications

Smart transfer and driving behavior analysis
Outline

- Background and Motivation
- **Modeling Spatiotemporal Dynamics**
- Collective Representation Learning
- Applications
- Conclusion and Future Work
Learn the patterns of spatiotemporal arrival matrix, and forecast 3W (where, when, what) of future arrivals.
If two regions are similar in urban functions, they show similar arrival patterns.
Linking Arrivals, Regions and Purposes

1 Urban functions of regions

2 Trip purposes

3 Arrival rates

4 Synchronization
Linking Arrivals, Regions and Purposes

1 → 2: The urban functions of origin and destination regions show trip purposes

1 Urban functions of regions

2 Trip purpose

3 Arrival rates

4 Synchronization

Time Square

Synchronization effect

Hotel
Empire State
Statue of Liberty
Central Park

Home
Newport
Brooklyn
Apartment

Office
Scenic Spots
Restaurant
Bldg.

Relax
Work
Travel
...
2 → 3: Different trip purposes have different arrival rates in different time slots.
Linking Arrivals, Regions and Purposes

1. Urban functions of regions

2. Trip purpose

3. Arrival rates

4. Synchronization

3→4: If two regions share similar urban functions, they share similar arrival rate patterns.
Framework

Modeling the arrivals of a single region for single trip purpose

Modeling the arrivals of a single region for multiple trip purposes

Modeling the arrivals of multiple regions for multiple trip purposes

Incorporating human mobility synchronization effects

Incorporating the modeling of origin and destination regions

Integrating human knowledge
Each trajectory is a five-element arrival event is: $E_n = \{g_n, z_n, t_n, w_n^d, w_n^o\}$

- $g_n$: the trip purpose of the $n$-th arrival
- $t_n$: the timestamp of the $n$-th arrival
- $w_n^d$: POIs of destination region
- $w_n^o$: POIs of origin region

For each region, we organize trajectories as a sequence of arrivals: $E = \{E_1, E_2, ..., E_N\}$

Benefits: support multi-source mobility data, e.g., trajectories, check-ins
Modeling mobility arrivals as a stochastic point process

- Hawkes Process: \( \lambda(t) = \mu + \int_{-\infty}^{t} g(t - s) dN(s) \)

Self-exciting for multi-peak gradually-excited human activities

- The to-work arrivals at 9am are self-excited by the increasingly intensive to-work arrivals at 8am
Modeling Arrivals of Single Region for Multiple Trip Purposes (1)

Mobility arrivals in the i-th region:

$$\lambda_i = \lambda_{i,\text{eat}}(t) + \lambda_{i,\text{work}}(t) + \lambda_{i,\text{relax}}(t) + \cdots$$
Mixture Hawkes processes with respect to different trip purposes

- \( \lambda_{i,m}(t) = \mu_{i,m} + \int_{-\infty}^{t} g(t - s) dN(s) = \mu_i * \gamma_m + \int_{-\infty}^{t} g(t - s) dN(s) \)

- \( i \): the i-th region
- \( m \): the m-th trip purpose
- \( \mu_{i,m} \): the base rate that region i get visited with trip purpose m
- \( \mu_i \): the base visit rate of region i
- \( \gamma_m \): the base visit rate of trip purpose m
- \( g(t - s) \): memory decay function

Decouple the base rates of location and trip purpose to reduce the number of parameters
Synchronization Effect Across Regions

- **Region synchronization graph**
  - Road networks as graph
  - Regions as nodes in the graph
  - Synchronization rate between two regions as the edge weight between two nodes

If Region(i) and Region(j) are both office areas, and many to-work arrivals are observed in Region(j), then it is likely to observe many to-work arrivals in Region(i)
Modeling Synchronization Effect Across Regions in Mixture Hawkes Processes

- Integrating the synchronization effects across regions into mixture Hawkes processes
  \[ \lambda_{i,m}(t) = \mu_i \cdot \gamma_m + \sum_{j=1}^{I} a_{ji}^{m} \int_{-\infty}^{t} g(t - s)dN(s) \]
  - Base arrival rate
  - Sync effect when \( j \neq i \) (region-region peer dependency)
  - Self-exciting effect when \( j = i \) (past-current temporal dependency)

- Synchronization (Mutual-exciting)
  - The arrivals are not just self-excited by previous arrivals within a region, but also excited by the arrivals of peer regions
  - Example: The to-work arrivals of the i-th region at 9am are excited by the to-work arrivals of the j-th similar region at 9am
The urban functions of origin and destination regions can jointly show trip purposes.
Analogies between region modeling and textual mining

<table>
<thead>
<tr>
<th>Region-Building</th>
<th>Document-Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region</td>
<td>Document</td>
</tr>
<tr>
<td>Building category</td>
<td>Word</td>
</tr>
<tr>
<td>Urban function</td>
<td>Topic</td>
</tr>
</tbody>
</table>
Probabilistic generative model of buildings in origin and destination regions

- Draw a trip purpose for each trip
- Draw buildings of origin region from the trip purpose
- Draw buildings of destination region from the trip purpose

- Generate a purpose $m \sim Multi(\pi)$
- Generate the POI Topic for the origin $z_o \sim Multi(\Phi_{mz})$
  - For each POI $w^o$ in the origin neighborhood
  - Generate the POI $w^o \sim Multi(\beta_{zw})$
- Generate the POI Topic for the origin $z_d \sim Multi(\Phi_{mz})$
  - For each POI $w^d$ in the origin neighborhood
  - Generate the POI $w^d \sim Multi(\beta_{zw})$
Solving the Co-optimization (1)

Modeling origin and destination regions

Urban function topic

Origin

Destination

Modeling latent trip purposes

Trip Purpose

Modeling arrival sequences

Sync-aware Mixture Hawkes

Observed Events

Event Rate \( \lambda(t) \)

Expected Event Volume \( \xi(t) \)

Synchronization Effects \( s(t) \)
Solving the Co-optimization (2)

1. Training Data

\[(G, z, t, W) = \{(G_n, z_n, t_n, W_n)\}\] with \(t_0 = 0\) and \(t_N = T\)

2. Likelihood Function

\[L(G, t, W) = \prod_{n=1}^{N} p(G_n)p(W^o_n, W^d_n|G_n)p(t_n|G_n)\]

\[\mathcal{L}(t, W) \equiv \log \left( \int \left( \prod_{n=1}^{N} p(G_n)p(W^o_n, W^d_n|G_n)p(t_n|G_n) \right) d\{(G, z)\} \right)\]

\[\geq \int_{\{(G, z)\}} \log \left( \frac{L(G, z, t, W)}{dq((G, z))} \right) dq((G, z))\]

\[= E_q[\mathcal{L}(G, z, t, W)] + \mathcal{E}[q] \equiv \mathcal{L},\]

3. A Lower Bound

4. Parameter Update Rules

\[\xi^o_{m,r} \propto \prod_{c=1}^{C} (\beta_{rc})^c W^o_{nc}\]

\[\xi^d_{m,r} \propto \prod_{c=1}^{C} (\beta_{rc})^c W^d_{nc}\]

\[\beta_{rc} \propto \sum_n \sum_m \phi_{nm} (\xi^o_{m,e} W^o_{m,e} + \xi^d_{m,e} W^d_{m,e}).\]

\[\xi^o_{m,r} \propto \sum_n \sum_m \phi_{nm} \eta^m_{nn} \delta_{in} \delta_{inj}.\]

\[\xi^d_{m,r} \propto \sum_n \sum_m \phi_{nm} \eta^m_{nn} \delta_{in} \delta_{inj}.\]

\[\phi_{nm} \propto \sum_{n=1}^{N} \phi_{nm},\]

\[\beta_{rc} \propto \sum_{n=1}^{N} \phi_{nm} \eta^m_{nn} \delta_{in} \delta_{inj}.\]

\[\mu_{i} \propto \frac{\sum_{m=1}^{M} \phi_{nm} \eta^m_{nn} \delta_{in} \delta_{inj}}{\sum_{m=1}^{M} \gamma_{m} T},\]

\[\alpha_{ij} = \frac{\sum_{n=1}^{N} \sum_{l=1}^{N-1} \phi_{nl} \phi_{lm} \eta_{ln} \delta_{il} \delta_{inj}}{\sum_{n=1}^{N} \Theta(T - t_n) \phi_{nm} \delta_{in}^j},\]
Study of Forecasting Next Arrivals

Benchmark time intervals of every two arrival events

Predicted time intervals of every two arrival events
Study of Trip Purpose Clustering

- Experiments on synthetic data: validate the identified trip purposes
- Synthetic data generation: Ogata’s modified thinning algorithm for sampling arrival sequences
- Task: Clustering the trajectories based on the inferred trip purposes
- Baseline methods: MHP, LDA, K-means
- Metrics: purity, F1-Measure, Rand Statistics
### Study of Trip Purpose Interpretation

#### Data
- Taxi trips of NYC: 7-day taxi trips, millions of GPS trajectories, 152 valid regions
- Point of Interests data of NYC

#### Identified trip purposes

<table>
<thead>
<tr>
<th>Trip Purpose</th>
<th>prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>nightlife</td>
<td>prob.</td>
</tr>
<tr>
<td>dining</td>
<td>prob.</td>
</tr>
<tr>
<td>work</td>
<td>prob.</td>
</tr>
<tr>
<td>shopping</td>
<td>prob.</td>
</tr>
<tr>
<td>schooling</td>
<td>prob.</td>
</tr>
<tr>
<td>sightseeing</td>
<td>prob.</td>
</tr>
<tr>
<td>home</td>
<td>prob.</td>
</tr>
</tbody>
</table>

### Table

<table>
<thead>
<tr>
<th>TOPIC 1</th>
<th>prob.</th>
<th>TOPIC 2</th>
<th>prob.</th>
<th>TOPIC 3</th>
<th>prob.</th>
<th>TOPIC 4</th>
<th>prob.</th>
<th>TOPIC 5</th>
<th>prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bar</td>
<td>0.1884</td>
<td>Chinese Rest.</td>
<td>0.1286</td>
<td>Bar</td>
<td>0.0933</td>
<td>Office</td>
<td>0.3331</td>
<td>Clothing Store</td>
<td>0.0995</td>
</tr>
<tr>
<td>Home</td>
<td>0.0953</td>
<td>Italian Rest.</td>
<td>0.0913</td>
<td>Italian Rest.</td>
<td>0.0565</td>
<td>General Entertain</td>
<td>0.1035</td>
<td>Cafe</td>
<td>0.0693</td>
</tr>
<tr>
<td>Nightclub</td>
<td>0.0571</td>
<td>Asian Rest.</td>
<td>0.0541</td>
<td>American Rest.</td>
<td>0.0442</td>
<td>Hotel</td>
<td>0.1023</td>
<td>Office</td>
<td>0.0574</td>
</tr>
<tr>
<td>Event Space</td>
<td>0.0495</td>
<td>Tea Room</td>
<td>0.0481</td>
<td>Wine Bar</td>
<td>0.0373</td>
<td>Building</td>
<td>0.0869</td>
<td>Coffee Shop</td>
<td>0.0535</td>
</tr>
<tr>
<td>Cocktail Bar</td>
<td>0.0495</td>
<td>Bar</td>
<td>0.0472</td>
<td>Sushi Rest.</td>
<td>0.0319</td>
<td>Event Space</td>
<td>0.0593</td>
<td>Cosmetics Shop</td>
<td>0.0419</td>
</tr>
<tr>
<td>Lounge</td>
<td>0.0495</td>
<td>Spa or Massage Parlor</td>
<td>0.0416</td>
<td>Mexican Rest.</td>
<td>0.0306</td>
<td>Sandwich Place</td>
<td>0.0376</td>
<td>General Entertain</td>
<td>0.0408</td>
</tr>
<tr>
<td>Speakeasy</td>
<td>0.0471</td>
<td>Salon or Barbershop</td>
<td>0.0403</td>
<td>Lounge</td>
<td>0.0297</td>
<td>Hotel Bar</td>
<td>0.0342</td>
<td>French Rest.</td>
<td>0.0406</td>
</tr>
<tr>
<td>Breakfast Spot</td>
<td>0.0382</td>
<td>Vietnamese Rest.</td>
<td>0.039</td>
<td>Pizza Place</td>
<td>0.0278</td>
<td>Lounge</td>
<td>0.0342</td>
<td>High Tech Outlet</td>
<td>0.0388</td>
</tr>
<tr>
<td>French Rest.</td>
<td>0.0334</td>
<td>Art Gallery</td>
<td>0.0342</td>
<td>Coffee Shop</td>
<td>0.0256</td>
<td>Other Outdoors</td>
<td>0.0298</td>
<td>Salon or Barbershop</td>
<td>0.0368</td>
</tr>
<tr>
<td>Boat or Ferry</td>
<td>0.0316</td>
<td>Cocktail Bar</td>
<td>0.0316</td>
<td>Salon or Barbershop</td>
<td>0.0256</td>
<td>Performing Arts Venue</td>
<td>0.0289</td>
<td>Miscellaneous Shop</td>
<td>0.0331</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TOPIC 6</th>
<th>prob.</th>
<th>TOPIC 7</th>
<th>prob.</th>
<th>TOPIC 8</th>
<th>prob.</th>
<th>TOPIC 9</th>
<th>prob.</th>
<th>TOPIC 10</th>
<th>prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>College Acad.</td>
<td>0.0808</td>
<td>Park</td>
<td>0.1343</td>
<td>Art Gallery</td>
<td>0.2773</td>
<td>American Rest.</td>
<td>0.1023</td>
<td>Home</td>
<td>0.2005</td>
</tr>
<tr>
<td>Food Truck</td>
<td>0.0756</td>
<td>Other Outdoors</td>
<td>0.1</td>
<td>Park</td>
<td>0.1021</td>
<td>Deli or Bodega</td>
<td>0.0619</td>
<td>Building</td>
<td>0.0591</td>
</tr>
<tr>
<td>University</td>
<td>0.0653</td>
<td>Scenic Lookout</td>
<td>0.0767</td>
<td>Other Outdoors</td>
<td>0.0892</td>
<td>Office</td>
<td>0.0569</td>
<td>Deli or Bodega</td>
<td>0.0471</td>
</tr>
<tr>
<td>College Library</td>
<td>0.0639</td>
<td>General Travel</td>
<td>0.0753</td>
<td>Cafe</td>
<td>0.0555</td>
<td>Pizza Place</td>
<td>0.0464</td>
<td>Pizza Place</td>
<td>0.0442</td>
</tr>
<tr>
<td>General College/University</td>
<td>0.0573</td>
<td>Building</td>
<td>0.074</td>
<td>Playground</td>
<td>0.049</td>
<td>Bar</td>
<td>0.0448</td>
<td>Laundromat or Dry Cleaner</td>
<td>0.0342</td>
</tr>
<tr>
<td>College Dorm</td>
<td>0.0565</td>
<td>Airport</td>
<td>0.074</td>
<td>Automotive Shop</td>
<td>0.0386</td>
<td>Food Truck</td>
<td>0.0434</td>
<td>Coffee Shop</td>
<td>0.0317</td>
</tr>
<tr>
<td>College Classroom</td>
<td>0.0499</td>
<td>Harbor or Marina</td>
<td>0.0616</td>
<td>Event Space</td>
<td>0.033</td>
<td>Sandwich Place</td>
<td>0.0392</td>
<td>Drugstore or Pharmacy</td>
<td>0.0291</td>
</tr>
<tr>
<td>Plaza</td>
<td>0.0485</td>
<td>Taxi</td>
<td>0.0534</td>
<td>Strip Club</td>
<td>0.0265</td>
<td>Coffee Shop</td>
<td>0.0346</td>
<td>Chinese Rest.</td>
<td>0.0256</td>
</tr>
<tr>
<td>Park</td>
<td>0.0382</td>
<td>Government Building</td>
<td>0.048</td>
<td>Sculpture Garden</td>
<td>0.0241</td>
<td>Burger Joint</td>
<td>0.0326</td>
<td>Mexican Rest.</td>
<td>0.0236</td>
</tr>
<tr>
<td>College Classroom</td>
<td>0.0374</td>
<td>Seafood Rest.</td>
<td>0.0343</td>
<td>Plaza</td>
<td>0.0233</td>
<td>Cafe</td>
<td>0.0307</td>
<td>Apartment Building</td>
<td>0.0206</td>
</tr>
</tbody>
</table>
Study of Synchronization Effect

\[ \alpha_{AB} = 8.27066832 \]

\[ \alpha_{AC} = 0.00711464 \]

A and B have a higher synchronization rate
A and C have a lower synchronization rate
Study of Synchronization Effect

The POI and arrival distributions of A, B, C are consistent with the pairwise sync rates of A, B, C.
Summary

- **Task**
  - Modeling spatial diffusion and temporal dynamics of human mobility data

- **Property (provide in-depth understanding)**
  - Identify the synchronization property of human mobility

- **Modeling (make it predictable and traceable)**
  - Model human mobility as stochastic point processes
  - Develop a synchronization-aware mixture Hawkes model to jointly capture synchronization effects, mobility arrivals, urban regions, and trip purposes
  - Unify mobility arrival forecasting and trajectory semantic annotation
Outline

- Background and Motivation
- Modeling Spatiotemporal Dynamics
- **Collective Representation Learning**
- Applications
- Conclusion and Future Work
Spatial Representation Learning

Given: urban regions, single-source human mobility data
Objective: learn the vector representations of regions in a latent feature space
Constraints: similar regions share similar representations

$f(\text{Spatial Objects (e.g., Regions)} , \text{Single-source Human Mobility}) = \text{Vector Representations}$
Collective Representation Learning with Multi-source Mobility Data

Spatial Objects (e.g., Regions) \[ f(, , ) \]

Multi-Source Human Mobility Data

Vector Representations

- Given: urban regions, multi-source human mobility data
- Objective: learn the vector representations of regions in a latent feature space
- Constraints
  - Similar regions share similar representations
  - Integrate the mutual validation of multi-source human mobility patterns
Why Collective Representation Learning?

- Automated representation learning from widely-available data without domain experts
  - Non-automated: Find domain experts, design variables, and extract vector representations

- Automated fusion of multi-source unbalanced data
  - Non-automated: Design features, select features, weigh features, weighted combination of features

- Enable the availability of existing algorithms
  - Enable classification, ranking, clustering, outlier detection for spatial contexts
The Patterns of Three Mobility Events

- **Checkin mobility pattern**
  - \(<\text{day, hour, location category}>\) of a checkin event

- **Taxi mobility pattern**
  - \(<\text{day, hour, leaving or arriving}>\) of a taxi pickup or delivery event

- **Bus mobility pattern**
  - \(<\text{day, hour, leaving or arriving}>\) of a bus pickup or delivery event
If the representations of two regions are similar, the mobility patterns are similar.
If the representations of two regions are similar, the mobility patterns in different time slots are similar.
If the representations of two regions are similar, the mobility patterns in different time slots and of different sources are similar.
A Probabilistic Hierarchical Model for Collective Representation Learning

Observed Mobility Data
A Probabilistic Hierarchical Model for Collective Representation Learning

Generative Structure
A Probabilistic Hierarchical Model for Collective Representation Learning

Parameters
A Probabilistic Hierarchical Model for Collective Representation Learning

Region Representation
A Probabilistic Hierarchical Model for Collective Representation Learning

(M regions for N time periods on K hidden status with C/T/B mobility)

- The region $m$ is represented by a latent probabilistic vector $\eta_m$
- The hidden status $f$ of the region $m$ changes over time
- In a period, a region shows checkin ($C$), taxi ($T$), and bus ($B$) clusters of mobility patterns that reflect the hidden status $f$
- A cluster of mobility patterns = a document
- A mobility event= a word
- Model doc-word with topic modeling
Solving the Optimization Problem

Collapsed Gibbs Sampling to Solve Probabilistic Hierarchical Model

For the i-th taxi pattern $t_{m,n,i} \in t_{m,n}$, the conditional posterior for its latent taxi topic is computed by

$$P(u_{m,n,i} = r | D, \Upsilon - u_{m,n,i}) = \frac{P^{-(m,n,i)} T_{r,t_{m,n,i}} + \omega_{t_{m,n,i}}}{\sum_{l=1}^{R} P_{l}^{-(m,n,i)} T_{r,l} + \omega}.$$

For the i-th bus pattern $b_{m,n,i} \in b_{m,n}$, the conditional posterior for its latent bus topic is computed by

$$P(v_{m,n,i} = w | D, \Upsilon - v_{m,n,i}) = \frac{P^{-(m,n,i)} \sum_{b=1}^{B} B_{w,b}^{-1} + \zeta_{w}}{\sum_{b=1}^{B} B_{w,b}^{-1} + \zeta_{w}}.$$

For the n-th mobility segment in estate m, the conditional posterior probability for its latent function assignment $f$ is computed by

$$P(f_{m,n} = k | D, \Upsilon - f_{m,n}) = \frac{\prod_{l=1}^{Q} \Gamma(Z_{k,z} + \mu_{z}) \Gamma(Y_{k,m}) \prod_{l=1}^{R} \Gamma(U_{k,u} + \nu_{u}) \Gamma(S_{k,c} + \kappa_{c}) \prod_{l=1}^{W} \Gamma(V_{k,v} + \zeta_{v}) \Gamma(S_{k,v} + \kappa_{v})}{\prod_{l=1}^{Q} \Gamma(Z_{k,m,n} + \mu_{z}) \prod_{l=1}^{R} \Gamma(U_{k,m,n} + \nu_{u}) \prod_{l=1}^{W} \Gamma(V_{k,m,n} + \zeta_{v})}.$$
Study of Restaurant Popularity Prediction

Accuracy comparison of human-defined explicit features and machine-learned latent representations over different predictive models

Square Error

Empirical variables extracted by human
Latent features learned by our model
Apply K-Means to cluster regions into 4 categories using the learned representations.

- Transportation
- Office mixed with scenic spots
Apply K-Means to cluster regions into 5 categories using the learned representations

Office mixed with scenic spots
Transportation
Annotating Regions of Urban Functions

Apply K-Means to cluster regions into 8 categories using the learned representations

Office mixed with scenic spots

Transportation
Outline

- Background and Motivation
- Modeling Spatiotemporal Dynamics
- Collective Representation Learning
- **Applications**
- Conclusion and Future Work
SmartTransfer: Modeling the Spatiotemporal Dynamics of Passenger Transfers for Crowdedness-aware Route Recommendations
Route and Transfer Recommendations in Public Transportation Systems

Fastest routes != best routes

Spatial distribution of transfer passenger flow

Cumulative distribution of transfer passenger flow for top 256 subway stations

Root Cause: Spatial-temporal unbalance of traffic demand and transportation capacity supply
Crowdedness-aware Route Recommendations

- Feature extraction of subway stations
- Predict the transfer demands of subway stations with spatial-temporal multi-task learning
- Given origin and destination, generate candidate routes from subway networks and bus networks
- Recommend routes based on potential time cost and crowdedness
You Are How You Drive: Peer and Temporal-Aware Representation Learning for Driving Behavior Analysis
Beyond Accidents: Vehicles as Weapons

Charlottesville, Virginia

A car plows into counterprotesters marching against white

**Date of attack:** August 12, 2017

**Number of casualties:** A 32-year-old woman was killed

Nice, France

French citizens in mourning over Nice attack 02:19

**Date of attack:** July 14, 2016

**Number of casualties:** Eighty-four people were killed and more than 200 wounded.

What can we do to protect human-transportation systems from vehicle-ramming attacks?
Toward Machine-Learning Based Driving Behavior Analysis

Driving Behavior Analysis

Emergency Alarm Systems

Turn Left

Turn Right

Accelerate

<\(t_1, \text{lat}_1, \text{lon}_1\)>

<\(t_2, \text{lat}_2, \text{lon}_2\)>

<\(t_3, \text{lat}_3, \text{lon}_3\)>

<\(t_4, \text{lat}_4, \text{lon}_4\)>

<\(t_5, \text{lat}_5, \text{lon}_5\)>

Emergency Beep Alerts:

- Idle
- Over Speed
- Rapid Acceleration
- Hard Braking
1. Learn driving behavior profiles from driving state transition graphs with spatiotemporal representation learning
2. Exploit driving behavior profiles to automatically score driving performances and detect risky areas
More Applications

Energy Consumption
Gas Refilling Event Detection and Gas Station Site Selection (DASFAA16)

Community Planning
Residential Community Analysis for Affordable Housing (KDD14, KDD15, TKDD)

Mitigate Traffic Congestion
Bike Station Site Selection and Rebalancing (ICDM15)

User Modeling
Point-Of-Interests Recommender Systems (KDD13, SDM14, ICDM16)
Outline

- Background and Motivation
- Modeling Spatiotemporal Dynamics
- Collective Representation Learning
- Applications
- Conclusion and Future Work
Conclusion Remarks

- **Data Environments**
  - Human mobility data

- **Data Science Foundations**
  - Modeling spatial diffusion and temporal dynamics as mixture stochastic point processes integrated with human knowledge
    - Generalized for ecommerce click rate data, online hospital comment data, network intrusion data, malware/disease infection data, paypal e-payment data
    - Spatiotemporal forecasting of 3W (when, where, what)
  - Collective representation learning with multi-source data
    - Generalized for automated heterogeneous data fusion and automated representation learning
    - Spatiotemporal embedding + semantic labeling

- **Data Science Applications**
  - Smart transfer systems
  - Driving behavior analysis