Towards Space and Time Coupled Social Media Analysis

Chao Zhang, Quan Yuan, Shi Zhi, Sha Li, Jiawei Han

University of Illinois at Urbana-Champaign Facebook Inc.
Explicit Space and Time Information in Social Media

Daily Stats:

- 10 M+ geo-tagged tweets
- 20 M+ geo-tagged Instagram posts
- 8 M+ Foursquare checkins

Time

Location
Implicit Space and Time Information in Social Media

Location Entity

Time Expression
The Rise of Social Sensing

The confluence of people, devices, and environments
Example 1: Hurricane Harvey

“Social media becomes a savior in hurricane Harvey relief.”
— NBC News

21.2 Million tweets in a five-day span
Example 2: Urban Exploration

People leave behind them billions of traces of their visited places.

Travelers see and understand their destinations before they arrive.
Example 3: Smart Transportation

People as traffic sensors:

- Accidents
- Jams
- Hazards
- Construction
An Ecosystem of People and Data

Activities

Data Production

Devices

Data

Applications

Knowledge

Models

Data Consumption

Data Production
General Challenges

How to combine multiple factors for joint analysis?

How to address data sparsity in the multi-dimensional space?

How to design online and scalable methods?
What Will Be Covered in This Tutorial?

1. Spatiotemporal activity modeling
   ○ How to find the typical activities in different regions and periods?

2. Spatiotemporal event detection and forecasting
   ○ How to detect and forecast unusual spatiotemporal events?

3. Spatiotemporal mobility modeling
   ○ How to model human movement regularities from semantic trajectories?

4. Location recommendation and prediction
   ○ How to improve location recommendation and prediction systems?
Outline

Introduction

Part 1: Spatiotemporal Activity Modeling

Part 2: Spatiotemporal Event Detection

Part 3: Spatiotemporal Mobility Modeling

Part 4: Location Recommendation and Prediction

Summary & Research Frontiers
Part I: Spatiotemporal Activity Modeling
Problem Definition

**Input:** a collection of geo-tagged social media records

- Each record has: a location, a timestamp, a text message

**Task:** predict people’s typical activities in different regions and periods

- Multiple schemas, e.g.
  - Region + time -> keywords
  - Region + keywords -> time
  - Time + keyword -> region

“What are the fun things to do around the Hilton Hotel?”

“Where should I go to hang out with my friends at 9pm?”
Representative Approaches

**Similarity-Based Methods**
- Li et. al. WWW 2015

**Probabilistic Graphical Modeling Methods**
- Sizov et. al. WSDM 2010

**Representation Learning Methods**
- Zhang et. al. WWW 2017
Representative Approaches

**Similarity-Based Methods**
- Li et. al. WWW 2015

**Probabilistic Graphical Modeling Methods**
- Sizov et. al. WSDM 2010

**Representation Learning Methods**
- Zhang et. al. WWW 2017
Semantic Annotation of Mobility Data Using Social Media [Wu et. al. WWW 2015]

Input: GPS location history, a collection of spatiotemporal social media

Goal: annotate GPS location history with social media keywords to reflect user activities

Why using social media?
- “Static annotation” vs “dynamic annotation”
- Including up-to-date event information.

<table>
<thead>
<tr>
<th>Record ID</th>
<th>Time</th>
<th>Longitude</th>
<th>Latitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>r1</td>
<td>2013-1-20</td>
<td>40.75051</td>
<td>-73.99349</td>
</tr>
<tr>
<td>r2</td>
<td>2013-2-10</td>
<td>40.68312</td>
<td>-73.97597</td>
</tr>
<tr>
<td>r3</td>
<td>2013-2-19</td>
<td>40.75051</td>
<td>-73.993499</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Record ID</th>
<th>Annotations</th>
</tr>
</thead>
<tbody>
<tr>
<td>r1</td>
<td>madison, garden, rangers, penguins</td>
</tr>
<tr>
<td>r2</td>
<td>nets, barclays, center, nba</td>
</tr>
<tr>
<td>r3</td>
<td>rangers, madison, square, montreal, canadiens</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>Longitude</th>
<th>Latitude</th>
<th>Tweets</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013-1-20</td>
<td>40.75051</td>
<td>-73.99349</td>
<td>LETS GO RANGERS</td>
</tr>
<tr>
<td>2013-1-20</td>
<td>40.61219</td>
<td>-74.15814</td>
<td>I’m at Buffalo Wild Wings</td>
</tr>
<tr>
<td>2013-1-20</td>
<td>40.75050</td>
<td>-73.99350</td>
<td>I’m @ Madison Square Garden for Pittsburgh Penguins vs New York Rangers</td>
</tr>
</tbody>
</table>
Three Annotation Strategies

Frequency-Based Methods
- Use TF-IDF weighting to select representative keywords

Gaussian Model
- Model the distribution of a keyword as a mixture of Gaussians

Kernel Density Estimation
- Estimate the kernel densities of keywords based on observed samples
Kernel Density Estimation: More Details

**Assumption:** the semantics of a location record $r_i$ can be inferred from the documents posted at nearby locations within a short time.

**Time sensitivity:** define a time window and collect documents $D_i$ that fall into it.

**Relevance:** Rank words based on the scores estimated by Kernel Density Estimation (KDE)

\[
s^K_D E_w(r_i; L_i(w), H) = \frac{1}{|L_i(w)|} \sum_{loc^D_j \in L_i(w)} K_H(loc^U_i - loc^D_j)
\]

\[
K_H(x) = |H|^{-1/2} K(H^{-1/2}(x))
\]

Locations of all the occurrences of word $w$  \hspace{1cm} Kernel function  \hspace{1cm} Geo-coordinates of record $r_i$  \hspace{1cm} Bandwidth
Datasets:

- **Context documents**: geo-tagged tweets from NYC, Chicago and LA
- **Users’ location histories**: GPS coordinates and timestamps of tweets
- **Groundtruth**: manually judge whether the extracted keywords reflect user intention

Experiments
Representative Approaches

Similarity-Based Methods
- Li et. al. WWW 2015

Probabilistic Graphical Modeling Methods
- Sizov et. al. WSDM 2010

Representation Learning Methods
- Zhang et. al. WWW 2017
GeoFolk: Latent Spatial Semantics in Web 2.0 Social Media [Sizov et. al. WSDM 2010, TIST 2012]

Goal: use topic modeling to uncover the latent activities from social media.

Each latent topic is **multidimensional** (location, time, text):
- Single dimension: insufficient for reliable content disambiguation
- Combination: better content characterization

Applications
- Topic discovery for regions
- Recommender system
- Content categorization
Examples

Geographical Topic Discovery and Comparison, Yin et al, WWW 2011
GeoFolk: Geographical Topic Modeling

**Input:** each document is associated with tags, GPS coordinates (and time).
Experiments

Dataset: 28,770 Flickr images with tags and coordinates

Task 1: Classification and clustering

<table>
<thead>
<tr>
<th>Model</th>
<th>avg(accuracy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GeoFolk</td>
<td>0.421</td>
</tr>
<tr>
<td>LDA</td>
<td>0.374</td>
</tr>
<tr>
<td>Tags</td>
<td>0.282</td>
</tr>
<tr>
<td>Coordinates</td>
<td>0.187</td>
</tr>
</tbody>
</table>

classification

<table>
<thead>
<tr>
<th>Model</th>
<th>avg(accuracy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GeoFolk</td>
<td>0.328</td>
</tr>
<tr>
<td>LDA</td>
<td>0.255</td>
</tr>
<tr>
<td>Tags</td>
<td>0.117</td>
</tr>
<tr>
<td>Coordinates</td>
<td>0.102</td>
</tr>
</tbody>
</table>

clustering

Task 2: Tag recommendation

<table>
<thead>
<tr>
<th>Model</th>
<th>MRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>GeoFolk</td>
<td>0.212</td>
</tr>
<tr>
<td>LDA</td>
<td>0.119</td>
</tr>
<tr>
<td>Tags</td>
<td>0.073</td>
</tr>
<tr>
<td>Coordinates</td>
<td>0.027</td>
</tr>
</tbody>
</table>
Representative Approaches

Similarity-Based Methods
- Li et. al. WWW 2015

Probabilistic Graphical Modeling Methods
- Sizov et. al. WSDM 2010

Representation Learning Methods
- Zhang et. al. WWW 2017
Regions, Periods, Activities: Uncovering Urban Dynamics via Cross-Modal Representation Learning [Zhang et. al. WWW 2017]

Idea: map geographical regions, temporal periods, and textual keywords into a latent semantic space to preserve their correlations.
Regions, Periods, Activities: Uncovering Urban Dynamics via Cross-Modal Representation Learning [Zhang et. al. WWW 2017]

Detect regions and periods where people’s activities burst

Map regions, periods, and keywords into the same space
Hotspot Detection: A Mode-Seeking Procedure

A spatial (temporal) hotspot is a density maxima in the 2D (1D) space.

We design a fast mode seeking procedure to find the hotspots.

Benefits:

- Fast
- No distribution assumptions

Kernel density estimation:

\[ f(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \]
Cross-Modal Embedding: Designing Philosophy

Map regions, periods, and keywords into the same space:
- **Region**: a spatial hotspot
- **Period**: a temporal hotspot

Aim to preserve two types of correlations:
1. **Co-occurrence**: two units are correlated if there co-occur frequently
2. **Neighborhood**: two regions (periods) are correlated if they are adjacent
Cross-Modal Embedding: Two Strategies

**Reconstruction-based embedding**

1. Consider each record as a relation
2. Mark off one unit \( i \) and try to predict it from the observed units

\[
p(i|r_{-i}) = \frac{\exp(s(i, r_{-i}))}{\sum_{j \in X} \exp(s(j, r_{-i}))}
\]

Overall objective:

\[
O = - \sum_{r \in C} \sum_{i \in r} \log p(i|r_{-i})
\]

**Graph-based embedding**

1. Use a graph to encode the correlations between regions, periods, and activities
2. Learn graph node embeddings to preserve the graph structure

\[
O = O_{WW} + O_{LL} + O_{TT} + O_{WL} + O_{WT} + O_{LT}
\]

\[
O_{XY} = \sum_{i \in X} d_i KL(p'(\cdot|i)\|p(\cdot|i)) + \sum_{j \in Y} d_j KL(p'(\cdot|j)\|p(\cdot|j))
\]
Embedding Visualization

Visualizing the feature vectors generated by LGTA and CrossMap for three activity categories: “Food” (cyan), “Travel & Transport” (blue), and “Residence” (orange).
Quantitative Evaluation: Attribute Recovery

Mark off one attribute (location, time, or text) and predict it based on the observed ones.

Mean reciprocal ranks:

<table>
<thead>
<tr>
<th>Method</th>
<th>Text</th>
<th>4SQ</th>
<th>Location</th>
<th>Tweet</th>
<th>4SQ</th>
<th>Time</th>
<th>4SQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>LGTA</td>
<td>0.376</td>
<td>0.6107</td>
<td>0.3792</td>
<td>0.6083</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MGMTM</td>
<td>0.3874</td>
<td>0.5974</td>
<td>0.4474</td>
<td>0.5753</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TF-IDF</td>
<td>0.62</td>
<td>0.8505</td>
<td>0.4298</td>
<td>0.7097</td>
<td>0.3197</td>
<td>0.3431</td>
<td>-</td>
</tr>
<tr>
<td>SVD</td>
<td>0.4475</td>
<td>0.7137</td>
<td>0.3953</td>
<td>0.646</td>
<td>0.3256</td>
<td>0.3187</td>
<td>-</td>
</tr>
<tr>
<td>Tensor</td>
<td>0.4382</td>
<td>0.6826</td>
<td>0.3871</td>
<td>0.6251</td>
<td>0.3179</td>
<td>0.2983</td>
<td>-</td>
</tr>
<tr>
<td>RECON</td>
<td>0.6877</td>
<td>0.9219</td>
<td>0.6526</td>
<td>0.9044</td>
<td>0.3582</td>
<td>0.3612</td>
<td>-</td>
</tr>
<tr>
<td>GRAPH</td>
<td>0.7011</td>
<td>0.9449</td>
<td>0.6758</td>
<td>0.9168</td>
<td>0.3895</td>
<td>0.3716</td>
<td>-</td>
</tr>
</tbody>
</table>
Application: Activity Classification

The embeddings can be used as feature vectors for downstream applications.

Example: Foursquare checkins belong to nine categories. We predict the category based on the embeddings.
Reference

Annotate POIs with categories
- On the semantic annotation of places in location-based social networks. Ye et al. KDD 2011
- Placer: semantic place labels from diary data. Krumm et al. UbiComp 2013

Annotate Regions with functions
- Inferring urban land use using large-scale social media check-in data. Zhang et al. Networks and Spatial Economics 2014
- Geographical topic discovery and comparison, Yin et al. WWW 2011
- Discovering regions of different functions in a city using human mobility and pois. Yuan et al. KDD 2012
- Latent geospatial semantics of social media. Sizov et al. TIST 2012

Annotate Visits with Semantics
- A conceptual view on trajectories. Spaccapietra et al. DKE 2008
- Semantic trajectories: Mobility data computation and annotation. Yan et al. TIST 2013
- Semantic annotation of mobility data using social media. Wu et al. WWW 2015

Multimodal Representation Learning
Part II: Spatiotemporal Event Detection
What is a Spatiotemporal Event?

An unusual activity bursted in a local area and a specific duration while impacting a considerable number of people.
Prior Art: Global Event Detection

Feature-based methods:

- Text Corpus
  - News [Fung 2005]
  - Social Media [He 2007, Weng 2011, Li 2012]

- Bursty Features
  - Wavelet Transform [Weng 2011]
  - Gaussian Mixture [He 2007]
  - Binomial Distribution [Fung 2005, Li 2012]

- Clustering
  - Co-occurrence [Fung 2005, He 2007]
  - Temporal Distribution [Weng 2012]
  - Mixed Similarity [Li 2012]
Prior Art: Global Event Detection

Document-based methods:

Text Corpus
- News  [Allan 1998]
- Social Media  [Aggarwal 2012]
- Diplomatic Record  [Chaney 2016]

Document Cluster
- Topic Modeling  [Chaney 2016]
- TF-IDF Similarity  [Allan 1998]
- Content & Structure  [Aggarwal 2012]

Filtering
- Business Score  [Chaney 2016]
- Novelty Discovery  [Allan 1998, Aggarwal 2012]
How About Using Them to Detect Spatiotemporal Events?

Existing global detection tools

Detect global burstiness

Cluster similar text

Batch Detection

Spatiotemporal Events

Locally Bursty

Text is Short & Messy

Online Data
Representative Approaches

**Feature-Based Detection**: First detect bursty keywords/phrases from the input, then group relevant features into events.
- E.g., Chen et. al. CIKM 2009, Abdelhaq et. al. PVLDB 2013

**Document-Based Detection**: Consider each document (e.g., tweet, check-in) as a basic unit and detect bursty document clusters as events.
- E.g., Zhang et. al. SIGIR 2016, Zhang et. al. KDD 2017

**Spatiotemporal Event Forecasting**: Predict whether a specific type of spatiotemporal event will occur in the future
- E.g., Zhao et. al. KDD 2016
Representative Approaches

**Feature-Based Detection**: First detect bursty keywords/phrases from the input, then group relevant features into events.
- E.g., Chen et. al. CIKM 2009, Abdelhaq et. al. PVLDB 2013

**Document-Based Detection**: Consider each document (e.g., tweet, check-in) as a basic unit and detect bursty document clusters as events.
- E.g., Zhang et. al. SIGIR 2016, Zhang et. al. KDD 2017

**Spatiotemporal Event Forecasting**: Predict whether a specific type of spatiotemporal event will occur in the future.
- E.g., Zhao et. al. KDD 2016
Event Detection from Flickr Data through Wavelet-based Spatial Analysis [Chen et. al. CIKM 2009]

A feature-based approach:

1. Use Wavelet analysis to find spatiotemporally localized Flickr tags
2. Cluster event-related tags into events based on both semantic and spatiotemporal similarities.
Detecting Event-related Tags

- For each tag, map every occurrence into a point in the 3D (lat, lng, time) space
- User Wavelet analysis handle 3D signals and find event-related (bursty) tags

(a) usage occurrences in the original 3D space

(b) surface plot in the original 3D space
Event Generation

Cluster bursty tags into spatiotemporal events:

\[
S(q_i, q_j) = \frac{\text{SemSim}(q_i, q_j)}{1 + \text{SpaDist}(q_i, q_j)}
\]

\[
\text{SemSim}(q_i, q_j) = \frac{|P(q_i) \cap P(q_j)|}{\min\{|P(q_i)|, |P(q_j)|\}}
\]

\[
\text{SpaDist}(q_i, q_j) = KL^N(m_i, sd_i; m_j, sd_j) = \\
\frac{1}{2} \left( \log\left( \frac{sd^2_j}{sd^2_i} \right) + \frac{sd^2_i}{sd^2_j} + \frac{(m_i - m_j)^2}{sd^2_j} \right) - 1
\]
## Example Bursty Tags and Events

<table>
<thead>
<tr>
<th>No.</th>
<th>Event Tags</th>
<th>Time</th>
<th>Location ((l_a, \ l_o))</th>
<th>Event Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(E_1)</td>
<td>partnershipwalk akf agakhanfoundation</td>
<td>10/29/2006, 11/10/2007</td>
<td>(29.719322, -95.37212)</td>
<td>Partnership Walk is an initiative of Aga Khan Foundation USA to raise funds and awareness to help communities in Africa and Asia. It is held annually at Atlanta, Chicago, Dallas, Houston, Los Angeles.</td>
</tr>
<tr>
<td>(E_3)</td>
<td>crosswalkamerica crosswalk scottgriessel creatista griessel</td>
<td>07/02/2006, 08/01/2006, 08/20/2006, 09/01/2006, 07/02/2007, 08/06/2007, 08/23/2007, 09/01/2007</td>
<td>(33.99294, -110.07808)</td>
<td>Crosswalk is a journey made by a couple of progressive Christians who trekked across the country from April to September. Griessel is the photographer of this walk.</td>
</tr>
<tr>
<td>(E_4)</td>
<td>f1 formulaone unitedstatesgrandprix</td>
<td>07/02/2006, 06/17/2007</td>
<td>(39.693844, -86.23974)</td>
<td>The United States Grand Prix was a Formula One race held on July 2, 2006, and June 15-17, 2007, at the Indianapolis Motor Speedway.</td>
</tr>
<tr>
<td>(E_5)</td>
<td>asl northpark deaf gpccd</td>
<td>04/22/2006, 04/14/2007</td>
<td>(34.239143, -116.894745)</td>
<td>The annual ASL fundraising picnic party at Pittsburgh North Park hosted by GPCCD in April.</td>
</tr>
<tr>
<td>(E_6)</td>
<td>beachjam amusementrides moreyssiers wildwoodbeachjam amusement beachcycling</td>
<td>05/20/2006, 05/20/2007</td>
<td>(38.987007, -74.81043)</td>
<td>The Beach Jam is an annual camping event on the Wildwood, NJ, beach at Morey’s Piers that includes amusement rides. There is a 3-day Spring Beach Jam before Memorial Day.</td>
</tr>
</tbody>
</table>
EvenTweet: Online Localized Event Detection from Twitter [Abdelhaq et. al. PVLDB 2013]

Online detection:
- Partition time into intervals
- Trigger the detector when the current bin is saturated

Feature-based detection:
- Each keyword is a feature
- Detect bursty features and cluster them into local events

Figure 1: System overview of EvenTweet
The Online Detection Process

1. Select temporally bursty keywords from the current query bin by comparing against previous bins.
2. Select spatially localized keywords by computing keyword entropies.
3. Cluster the localized keywords into events based on spatial distributions.
4. Compute the burstiness score of the clusters and rank them.
A local event is represented as a collection of bursty and localized keywords

Examples [Abdelhaq et. al. PVLDB 2013]
Representative Approaches

**Feature-Based Detection:** First detect bursty keywords/phrases from the input, then group relevant features into events.

- E.g., Chen et. al. CIKM 2009, Abdelhaq et. al. PVLDB 2013

**Document-Based Detection:** Consider each document (e.g., tweet, check-in) as a basic unit and detect bursty document clusters as events.

- E.g., Zhang et. al. SIGIR 2016, Zhang et. al. KDD 2017

**Spatiotemporal Event Forecasting:** Predict whether a specific type of spatiotemporal event will occur in the future

- E.g., Zhao et. al. KDD 2016
GeoBurst: Real-Time Local Event Detection in Geo-Tagged Tweet Streams [Zhang et. al. SIGIR 2016]

A document-based approach:

A local event is a geo-topic cluster that is spatiotemporally bursty.
GeoBurst: A Two-Step Detector

1. Candidate generation: detect all the geo-topic clusters in the query window
2. Candidate ranking: select top-K candidates by spatiotemporal burstiness
Candidate Generation & Ranking

**Candidate generation**
- The event occurring spot acts as a **pivot** that produces relevant tweets around it.
- Define geo-topic authorities and perform authority ascent to find pivot tweets.

**Candidate filtering**
Summarize typical activities in different regions to rank the candidates.
Empirical Performance

Effectiveness comparison:

(a) Precision comparison (NY).
(b) Precision comparison (LA).

Efficiency:

(a) Running time (NY).
(b) Running time (LA).
(a) Time v.s. # update (NY).
(a) Time v.s. # tweet (NY).
TrioVecEvent: Embedding-Based Online Local Event Detection in Geo-Tagged Tweet Streams [Zhang et. al. KDD 2017]

Motivation:

1. How can we better capture the semantics of short text during candidate generation?
   - Previous methods are mostly based on bag-of-words representations

2. How to better identify true events from the candidate pool?
   - Previous methods manually design burstiness scoring functions and select top-k bursty clusters
An Overview of TrioVecEvent

Cross-Modal Embedding
Embed location, time, and text into the same low-dimensional space.

Candidate Generation
Cluster the tweets into geo-topic clusters as candidate events.

Classification
Extract features for the candidates and identify true events.
Cross-Modal Embedding

Correlated units tend to be close in the embedding space.

Preserve:

1. Intra-type similarity
2. Inter-type similarity
Learning Cross-Modal Embeddings

Objective of preserving the information of corpus C

$$J_C = - \sum_{d \in C} \sum_{i \in d} \log p(i|d_{-i})$$

Probability of observing unit i given the others in document d

$$p(i|d_{-i}) = \frac{\exp(s(i, d_{-i}))}{\sum_{j \in X} \exp(s(j, d_{-i}))}$$

Similarity between unit i and the others.

$$s(i, d_{-i}) = v_i^T \sum_{j \in d_{-i}} v_j / |d_{-i}|$$
Find Geo-Topic Clusters: A Bayesian Mixture Model

Each tweet consists of: 1) a location; and 2) a text embedding.
Find Geo-Topic Clusters: A Bayesian Mixture Model

Each tweet consists of: 1) a location; and 2) a text embedding.

\[ \pi \sim \text{Dirichlet}(\cdot | \alpha) \]
\[ \{\eta_k, \sigma_k\} \sim \text{NIW}(\cdot | \eta_0, \lambda_0, S_0, \nu_0) \]
\[ \{\mu_k, \kappa_k\} \sim \Phi(\cdot | m_0, R_0, c) \]
\[ z_d \sim \text{Categorical}(\cdot | \pi) \]
\[ l_d \sim N(\cdot | \eta_{zd}, \sigma_{zd}) \]
\[ x_d \sim \text{vMF}(\cdot | \mu_{zd}, \kappa_{zd}) \]
Inferring Cluster Membership

MCMC Inference:

\[
p(z_d = k | x_d, l_d, \Theta) \propto p(z_d = k | Z^{-d}, \Theta) \cdot p(x_d | z_d = k, \Theta) \cdot p(l_d | z_d = k, \Theta)
\]

\[(n^k, -d + \alpha) \quad \text{“Rich-Get-Richer”}
\]

Semantic Coherence

\[
\frac{C_D(\kappa_k)C_D(\|\kappa_k(R_0m_0 + x^{k,-d})\|_2)}{C_D(\|\kappa_k(R_0m_0 + x^{k,-d} + x_d)\|_2)}
\]

Geographical Proximity

\[
\frac{\lambda^{k,-d}(v^{k,-d} - 1)|S^k \cap L^{-d}|v^{k,-d}/2}{2(\lambda^{k,-d} + 1)|S^k \cup \{l_d\}|(v^{k,-d} + 1)/2}
\]
Candidate Filtering: Classification

What geo-topic clusters are true events?

Should be coherent, bursty, and locally unusual.

Coherency Features:
spatial, temporal, and semantic concentrations

Unusual Features:
spatial and temporal unusualness

Burstiness Feature:
number of tweets
Event in LA: a protest rally at the LAPD headquarters

- Standing for justice! @ LAPD Headquarters http://t.co/YxNUAlOQcE
- At the LAPD protest downtown #EzellFord #MikeBrown http://t.co/kWphv6dXO
- Hands Up. Don't Shoot. @ Los Angeles City Hall
- Black, Brown, poor white, ALL oppressed people unite. #ftp #lapd #ferguson #lapd #mikebrown #ezellford http://t.co/szf3mJRJwV
- Finished marching now gathered back at LAPD police as organizers speak some truth #EzellFord #MikeBrown #ferguson http://t.co/M33m9IMOzC

Event in NYC: Hoboken Fall Arts & Music Festival

- Hoboken Fall Arts & Music Festival with bae @alli_holmes93 @ Washington St. Hoboken
- On Washington Street. (at Hoboken Music And Arts Festival) https://t.co/YbLSdZhLZV
- Sweeeet. Bonavita Guitars, at the Hoboken festival. http://t.co/2Cw1Qz4UGo
- I'm at Hoboken Music And Arts Festival in Hoboken, NJ https://t.co/i4bSM3mrjb
- It's a festy music day.
# Quantitative Evaluation

<table>
<thead>
<tr>
<th>Method</th>
<th>LA</th>
<th>NY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>R</td>
</tr>
<tr>
<td>EventTweet</td>
<td>0.132</td>
<td>0.212</td>
</tr>
<tr>
<td>GeoBurst</td>
<td>0.282</td>
<td>0.451</td>
</tr>
<tr>
<td>GeoBurst+</td>
<td>0.368</td>
<td>0.483</td>
</tr>
<tr>
<td>TrioVecEvent</td>
<td><strong>0.804</strong></td>
<td><strong>0.612</strong></td>
</tr>
</tbody>
</table>

EvenTweet [PVLDB’13]: bursty feature selection, top-k selection

GeoBurst [SIGIR’16]: document-based detection, top-k selection

GeoBurst+ [TIST’17]: supervised detection
Efficiency of TrioVecEvent

(a) Geo-topic clustering convergence.

(b) Summarization throughput.

(c) Online clustering time.

(d) Candidate filtering time.
An Overview of Representative Approaches

**Feature-Based Detection**: First detect bursty keywords/phrases from the input, then group relevant features into events.
- E.g., Chen et. al. CIKM 2009, Abdelhaq et. al. PVLDB 2013

**Document-Based Detection**: Consider each document (e.g., tweet, check-in) as a basic unit and detect bursty document clusters as events.
- E.g., Zhang et. al. SIGIR 2016, Zhang et. al. KDD 2017

**Spatiotemporal Event Forecasting**: Predict whether a specific type of spatiotemporal event will occur in the future
- E.g., Zhao et. al. KDD 2016
Hierarchical Incomplete Multi-Source Feature Learning for Spatiotemporal Event Forecasting [Zhao et. al. KDD 2016]

**Task:** use multiple data sources to predict whether certain event types will occur in the future.

Why multiple data sources?
- Spatiotemporal events are often influenced by different aspects of the society.
- Different data sources complement each other.
- One single source cannot cover all aspects of an event.

---

Slide from Liang Zhao: [link](http://people.cs.vt.edu/liangz8/materials/papers/HIML_slides.pdf)
Challenges for Event Forecasting

Hierarchical topology
- E.g., country-level, state-level, city-level
- Higher-level features can influence lower-level ones

Interactive missing values
- Different data sources, different spans
- Need to consider the interactions among different sources.

Feature sparsity
- Only a small set of features are useful
- Need to use geo-hierarchy to select useful features
Hierarchical Incomplete Multi-Source Feature Learning

Given the multi-source data for a location $l$ at time $t$, predict whether the event will happen at time $\tau$

$$f : \{X_{t,l_1}, \cdots, X_{t,l_N}\} \rightarrow Y_{\tau,l}$$

- Each location has features at multiple levels $l=(l_1, l_2, \ldots, l_N)$ E.g., (San Francisco, CA, USA)

Variables are dependent on the variables in their parent level

(level - 1) $Y_{\tau,l} = \alpha_0 + \sum_{i=1}^{\mid F_1 \mid} \alpha_i^T \cdot [X_{t,l_1}]_i + \varepsilon$

(level - 2) $\alpha_i = \beta_{i,0} + \sum_{j=1}^{\mid F_2 \mid} \beta_{i,j}^T \cdot [X_{t,l_2}]_j + \varepsilon_i$

(level - 3) $\beta_{i,j} = W_{i,j,0} + \sum_{k=1}^{\mid F_3 \mid} W_{i,j,k}^T \cdot [X_{t,l_3}]_k + \varepsilon_{i,j}$

Tensor form:

$$Y_{\tau,l} = W \odot Z_{t,l} + \varepsilon$$
Hierarchical Incomplete Multi-Source Feature Learning
## Experiments

Datasets: 10 datasets for civil unrest (CU) and influenza (FLU)

<table>
<thead>
<tr>
<th>Method</th>
<th>Argentina</th>
<th>Brazil</th>
<th>Chile</th>
<th>Colombia</th>
<th>El Salvador</th>
<th>Mexico</th>
<th>Paraguay</th>
<th>Uruguay</th>
<th>Venezuela</th>
</tr>
</thead>
<tbody>
<tr>
<td>LASSO</td>
<td>0.5267</td>
<td>0.7476</td>
<td>0.5624</td>
<td>0.8032</td>
<td>0.3148</td>
<td>0.7823</td>
<td>0.5572</td>
<td>0.4693</td>
<td>0.8073</td>
</tr>
<tr>
<td>LASSO-INT</td>
<td>0.5268</td>
<td>0.7191</td>
<td>0.5935</td>
<td>0.7861</td>
<td>0.5269</td>
<td>0.777</td>
<td>0.4887</td>
<td>0.5069</td>
<td>0.7543</td>
</tr>
<tr>
<td>iMSF</td>
<td>0.4795</td>
<td>0.4611</td>
<td>0.5033</td>
<td>0.7213</td>
<td>0.0</td>
<td>0.5569</td>
<td>0.4486</td>
<td>0.4904</td>
<td>0.5</td>
</tr>
<tr>
<td>MTL</td>
<td>0.3885</td>
<td>0.5017</td>
<td>0.5011</td>
<td>0.4334</td>
<td>0.3452</td>
<td>0.4674</td>
<td>0.4313</td>
<td>0.3507</td>
<td>0.5501</td>
</tr>
<tr>
<td>Baseline</td>
<td>0.5065</td>
<td>0.7317</td>
<td>0.6148</td>
<td>0.8084</td>
<td><strong>0.777</strong></td>
<td><strong>0.8037</strong></td>
<td>0.7339</td>
<td>0.7264</td>
<td><strong>0.7846</strong></td>
</tr>
<tr>
<td>HIML</td>
<td><strong>0.5873</strong></td>
<td><strong>0.8353</strong></td>
<td>0.5705</td>
<td><strong>0.8169</strong></td>
<td>0.7191</td>
<td>0.7973</td>
<td><strong>0.7478</strong></td>
<td><strong>0.8537</strong></td>
<td>0.7488</td>
</tr>
</tbody>
</table>

### CU forecasting performance AUC

<table>
<thead>
<tr>
<th>Method</th>
<th>21%</th>
<th>30%</th>
<th>50%</th>
<th>70%</th>
<th>runtime (second)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LASSO</td>
<td>0.9180</td>
<td>0.9056</td>
<td>0.9036</td>
<td>0.8753</td>
<td>493.92</td>
</tr>
<tr>
<td>LASSO-INT</td>
<td>0.9142</td>
<td>0.9027</td>
<td>0.9073</td>
<td>0.8403</td>
<td>508.49</td>
</tr>
<tr>
<td>iMSF</td>
<td>0.8949</td>
<td>0.8899</td>
<td>0.8930</td>
<td>0.8628</td>
<td>88.90</td>
</tr>
<tr>
<td>MTL</td>
<td>0.6129</td>
<td>0.5303</td>
<td>0.6253</td>
<td>0.5568</td>
<td>223.78</td>
</tr>
<tr>
<td>Baseline</td>
<td>0.9044</td>
<td>0.9045</td>
<td>0.8562</td>
<td>0.4359</td>
<td><strong>31.97</strong></td>
</tr>
<tr>
<td>HIML</td>
<td><strong>0.9372</strong></td>
<td><strong>0.9368</strong></td>
<td><strong>0.9364</strong></td>
<td><strong>0.9357</strong></td>
<td>851.83</td>
</tr>
</tbody>
</table>

### FLI forecasting performance AUC
Reference

Spatiotemporal Event Detection

- Event detection from flickr data through wavelet-based spatial analysis. Chen et al. CIKM 2009
- A probabilistic model for spatio-temporal signal extraction from social media. GIS 2013
- Identifying local events via space-time signals in twitter feeds. Krumm et al. GIS 2015
- Eventweet: Online localized event detection from twitter. Adbelhaq et al. PVLDB 2013
- TrioVecEvent: Embedding-Based Online Local Event Detection in Geo-Tagged Tweet Streams. Zhang et al. 2017
- Earthquake shakes twitter users: real-time event detection by social sensors. Sakaki et al. WWW 2010
- Crowd sensing of traffic anomalies based on human mobility and social media. Pan et al. GIS 2013
- Extracting city traffic events from social streams. Anantharam et al. TIST 2015

Spatiotemporal Event Forecasting

- “beating the news” with embers: forecasting civil unrest using open source indicators. Ramakrishnan et al. KDD 2014
- Spatiotemporal event forecasting in social media. Zhao et al. SDM 2015
- Combining heterogeneous data sources for civil unrest forecasting. Korkmaz et al. ASONAM 2015
- Spatiotemporal model fusion: multiscale modelling of civil unrest. Hoegh et al. J. the Royal Statistical
Part III: Spatiotemporal Mobility Modeling
Problem Description

Semantic trajectory: each GPS record also has semantic information (e.g., place category, text message)

Task: given a collection of semantic trajectories, how to model the movement regularities of the populace?

- Mobility modeling can occur at either user level or crowd level.
Representative Approaches

**Pattern-based approaches:** mining pre-defined mobility patterns from semantic trajectories

- Sequential pattern mining (e.g., Zhang et al. PVLDB 2014)

**Model-based approaches:** build statistical model to describe human movements

- User-level mobility models (e.g., Yuan et al. KDD 2013)
- Hidden markov models (e.g., Zhang et al. KDD 2016)
- Recurrent neural network models (e.g., AAAI 2016, Yao et al. CIKM 2017)
Splitter: Mining Frequent Sequential Patterns from Semantic Trajectories [Zhang et al. PVLDB 2014]

Each record in the trajectory has category information (e.g., office, hotel, gym).

Frequent sequential movement pattern: a movement sequence that frequently appear in the input trajectories.
How to Define Sequential Movement Patterns?

We cannot consider each location as an independent item because of spatial continuity.

Similar places should be grouped while respecting:

- Semantic consistency
- Spatial compactness
- Temporal continuity

Example pattern: $G_1 \rightarrow G_2 \rightarrow G_3$:

Problem Description

Input: a collection of $N$ semantic trajectories, a support threshold $K$

Output: the sequential movement patterns that appear in no less than $K$ trajectories
Splitter: A Top-down Mining Approach

How do we group similar places to form sequential movement patterns?
- There is an exponential number of possibilities, impossible to enumerate every option!

Key idea of Splitter:
- First mine category-level frequent transitions (coarse pattern)
- Then break each coarse pattern into spatially compact ones (fine-grained pattern)
Splitter: A Top-down Mining Approach

How do we group similar places to form sequential movement patterns?

- There is an exponential number of possibilities, impossible to enumerate every option!

Key idea of Splitter:

- First mine semantics-level frequent transitions (coarse pattern)
- Then break each coarse pattern into spatially compact ones (fine-grained pattern)
Splitter: A Top-down Mining Approach

**Coarse Pattern Mining:** (1) Group the places with the same category; (2) Apply time-constrained sequential pattern mining

<table>
<thead>
<tr>
<th>Object</th>
<th>Semantic Trajectory</th>
</tr>
</thead>
<tbody>
<tr>
<td>$o_1$</td>
<td>$&lt;(p_2, 0), (p_1, 10), (p_7, 30), (p_9, 40)&gt;$</td>
</tr>
<tr>
<td>$o_2$</td>
<td>$&lt;(p_7, 0), (p_1, 30), (p_2, 360), (p_7, 400), (p_{10}, 420)&gt;$</td>
</tr>
<tr>
<td>$o_3$</td>
<td>$&lt;(p_3, 0), (p_6, 30)&gt;$</td>
</tr>
<tr>
<td>$o_4$</td>
<td>$&lt;(p_2, 0), (p_1, 120), (p_6, 140), (p_8, 150), (p_{11}, 180)&gt;$</td>
</tr>
<tr>
<td>$o_5$</td>
<td>$&lt;(p_{12}, 50), (p_8, 80), (p_{11}, 120), (p_4, 210)&gt;$</td>
</tr>
</tbody>
</table>

**Fine-Grained Pattern Mining:** (1) regard each transition as a high-D point; (2) perform iterative clustering in the high-D space to find patterns.

E.g., regard every length-2 transition as a 4D point: $p_1 \rightarrow p_5$, $p_2 \rightarrow p_5$, $p_4 \rightarrow p_5$, $p_4 \rightarrow p_6$, $p_3 \rightarrow p_6$. 

![Diagram](image)
Example Patterns

### Coarse Patterns

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Sup</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shop → Food</td>
<td>1819</td>
</tr>
<tr>
<td>Food → Shop</td>
<td>1464</td>
</tr>
<tr>
<td>Professional → Nightlife Spot</td>
<td>1121</td>
</tr>
<tr>
<td>Outdoor → Food</td>
<td>947</td>
</tr>
<tr>
<td>Residence → College &amp; University</td>
<td>647</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Sup</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shop → Food → Shop</td>
<td>262</td>
</tr>
<tr>
<td>Professional → Food → Nightlife Spot</td>
<td>240</td>
</tr>
<tr>
<td>Entertainment → Food → Shop</td>
<td>178</td>
</tr>
<tr>
<td>Transportation → Shop → Shop</td>
<td>174</td>
</tr>
<tr>
<td>Residence → Outdoor → Food</td>
<td>163</td>
</tr>
</tbody>
</table>

### Fine-grained Patterns

![Graph showing patterns between shops and restaurants with support values]
Experiments

Pattern Coverage

(a) Coverage w.r.t. \( \sigma \).
(b) Pattern number w.r.t. \( \sigma \).

Efficiency

(a) Running time w.r.t. \( \sigma \).
(b) Running time w.r.t. \( \Delta t \).

**Input:** a collection of checkins of different users
- Each post contains a user ID, a timestamp, a venue and a text message
- Each venue is associated with a venue ID and geo-coordinates

**Output:** multi-dimensional user-level mobility models
- who visits which place at what time for what activity

**Previous studies:** consider at most three factors out of the four
- Where What: geographical topic modeling
- Where When What: geographical event detection
- Who Where When: spatiotemporal mobility behavior modeling for users
- Who Where What: user-level geographical topic profiling
Overview

User u’s mobility centers at several personal geographical regions r (home, work, …)
The region r where a user u stays is influenced by day s
- E.g., weekday: work region; weekend: shopping region
Visiting time is determined by region r and day s
- E.g., visiting shopping region at weekday evening & weekend afternoon
Overview

User $u$’s topic interests is influenced by $u$’s topic preference and region $r$

- E.g., $u$: “reading” and “shopping”. $u$@Times Square: “shopping”

User $u$ chooses a POI $l$ based on either topic $z$ or region $r$

- Nearby POI within $r$ that meets the topic requirement $z$ (e.g., meal)
- Different users make different trade-offs between $z$ and $r$
Overview

User $u$ chooses a set of words $w$ based on either topic $z$ or region $r$.
- Different user makes different trade-offs between $z$ and $r$
- E.g., user $u$ is shopping at home region: “grocery”, “family”
Experiments

Datasets
- 89,007 world-wide tweets (WW)
- 171,768 microblogs in USA (USA)

Venue prediction for tweet
- Rank venues by $P(l|u,s,t,w)$

Visitor prediction
- Rank users by $P(u|s,t,l)$

Venue prediction for user
- Rank venues by $P(l|u,s,t)$

<table>
<thead>
<tr>
<th>Acc</th>
<th>WW</th>
<th>USA</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMM</td>
<td>0.4163</td>
<td>0.4021</td>
</tr>
<tr>
<td>$W^4$</td>
<td>0.5063</td>
<td>0.5863</td>
</tr>
<tr>
<td>$EW^4$</td>
<td>0.5351</td>
<td>0.7679</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Acc</th>
<th>WW</th>
<th>USA</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMM</td>
<td>0.0423</td>
<td>0.1102</td>
</tr>
<tr>
<td>$W^4$</td>
<td>0.0776</td>
<td>0.2953</td>
</tr>
<tr>
<td>$EW^4$</td>
<td>0.1423</td>
<td>0.5054</td>
</tr>
</tbody>
</table>
GMove: Group-Level Mobility Modeling Using Geo-Tagged Social Media [Zhang et. al. KDD 2016]

**Input:** the semantic trajectories for a collection of users

**Goal:** (1) What are the intrinsic states underlying people’s movements? (2) How do people move sequentially between those latent states?
Dilemma in Learning Mobility Models

Individual-level modeling: learn a model for each individual user

Each user has a limited number of records, we suffer from data scarcity!

Global-level modeling:

Different users have different moving behaviors, we suffer from data inconsistency!
Gmove: Group-Level Mobility Modeling

Idea: divide the users into coherent groups, and learn one model for each group.

- Reduce data sparsity by aggregating the movements of multiple users.
- Ensure data consistency as the users in the same group have similar movement regularity.

<table>
<thead>
<tr>
<th></th>
<th>Data Sparsity</th>
<th>Data Consistency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual-level</td>
<td>X</td>
<td>O</td>
</tr>
<tr>
<td>Global-level</td>
<td>O</td>
<td>X</td>
</tr>
<tr>
<td>Group-level</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>
HMM Ensemble Learner

User grouping and mobility modeling mutually enhance each other:

- Better user grouping leads to more consistent training data
- Better mobility modeling helps infer the user membership more accurately

An Iterative Process:

For each user $u$, compute the posterior probability that $u$ belongs to group $g$:

$$p(g|u; \mathcal{H}^{\text{new}}) \propto p(g)p(u|g; \mathcal{H}^{\text{new}})$$
Quantitative Evaluation: Next Location Prediction

(a) LA

(b) NY
Predicting the Next Location: A Recurrent Model with Spatial and Temporal Contexts \[\text{[Liu et al. AAAI 2016]}\]

Model the sequential aspect of user mobility

- Recent visits are more influential

Some mobility behaviors are periodic

Users tend to visit nearby places

How to model the spatial, periodic and sequential data?

Spatial Temporal Recurrent Neural Networks
Recurrent Neural Networks

Perform the same task for every element of a sequence-->unfold

RNNs have a “memory” which captures information about what has been calculated so far

Variants: Long Short Term Memory (LSTM), Gated Recurrent Units (GRU)
Spatial Temporal Recurrent Neural Networks

**Basic RNN**: User’s status at $t$ is influenced by
- The characteristics of the visited location
- The status at $t-1$

$$h_{t_k}^u = f \left( M q_{t_k}^u + Ch_{t_k-1}^u \right)$$

**RNN With Temporal Context**: the status at $t$ is further influenced by
- The locations visited within a period in the past
- The gap time $t-t_i$ of these visits to $t$
Spatial Temporal Recurrent Neural Networks

RNN With Spatial Temporal Context: the importance of a previous visit is also dependent on the distance to current location

\[ h_{t,q_t}^u = f \left( \sum_{q_{t_i}^u \in Q^u, t - \omega < t_i < t} S_{q_i^u - q_{t_i}^u} T_{t-t_i} q_{t_i}^u + C_{h_{t-\omega, q_{t-\omega}}^u} \right) \]

Prediction: how likely user u visits location v?

\[ o_{u,t,v} = (h_{t,q_v}^u + p_u)^T q_v \]

- u’s current status
- u’s static preference
- v’s characteristics

Location prediction on Gowalla dataset

<table>
<thead>
<tr>
<th>Method</th>
<th>recall@1</th>
<th>recall@5</th>
<th>recall@10</th>
<th>F1-score@1</th>
<th>F1-score@5</th>
<th>F1-score@10</th>
<th>MAP</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOP</td>
<td>0.0052</td>
<td>0.0292</td>
<td>0.0585</td>
<td>0.0052</td>
<td>0.0097</td>
<td>0.0106</td>
<td>0.0372</td>
<td>0.6685</td>
</tr>
<tr>
<td>MF</td>
<td>0.0100</td>
<td>0.0538</td>
<td>0.1146</td>
<td>0.0100</td>
<td>0.0179</td>
<td>0.0208</td>
<td>0.0527</td>
<td>0.7056</td>
</tr>
<tr>
<td>MC</td>
<td>0.0091</td>
<td>0.0543</td>
<td>0.1015</td>
<td>0.0091</td>
<td>0.0181</td>
<td>0.0184</td>
<td>0.0510</td>
<td>0.7029</td>
</tr>
<tr>
<td>Gowalla</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TF</td>
<td>0.0116</td>
<td>0.0588</td>
<td>0.1120</td>
<td>0.0116</td>
<td>0.0196</td>
<td>0.0204</td>
<td>0.0551</td>
<td>0.7097</td>
</tr>
<tr>
<td>PFMC</td>
<td>0.0159</td>
<td>0.0792</td>
<td>0.1535</td>
<td>0.0159</td>
<td>0.0264</td>
<td>0.0279</td>
<td>0.0671</td>
<td>0.7363</td>
</tr>
<tr>
<td>PFMC-LR</td>
<td>0.0186</td>
<td>0.0940</td>
<td>0.1823</td>
<td>0.0186</td>
<td>0.0313</td>
<td>0.0331</td>
<td>0.0763</td>
<td>0.7580</td>
</tr>
<tr>
<td>PRME</td>
<td>0.0203</td>
<td>0.0990</td>
<td>0.1896</td>
<td>0.0203</td>
<td>0.0330</td>
<td>0.0344</td>
<td>0.0847</td>
<td>0.7695</td>
</tr>
<tr>
<td>RNN</td>
<td>0.0257</td>
<td>0.1349</td>
<td>0.2286</td>
<td>0.0257</td>
<td>0.0450</td>
<td>0.0416</td>
<td>0.0921</td>
<td>0.7875</td>
</tr>
<tr>
<td>ST-RNN</td>
<td><strong>0.0304</strong></td>
<td><strong>0.1524</strong></td>
<td><strong>0.2714</strong></td>
<td><strong>0.0304</strong></td>
<td><strong>0.0508</strong></td>
<td><strong>0.0493</strong></td>
<td><strong>0.1038</strong></td>
<td><strong>0.8115</strong></td>
</tr>
</tbody>
</table>
SERM: A Recurrent Model for Next Location Prediction in Semantic Trajectories

Modeling users’ interests.

- Input data: semantic enriched trajectories
- A record has: user, time, coordinates and text

To model a user’s next step mobility, we need to consider

- Current location and time. *Location and time embeddings*
- What is the user doing? Activity semantics. *Content embeddings*
- The user’s static preference over locations. *User embedding*
SERM: A Recurrent Model for Next Location Prediction in Semantic Trajectories

Factors to consider

- Current location and time. \textit{Location and time embeddings}: \( e_{lk}, e_{tk} \)
- What is the user doing? Activity semantics. \textit{Content embeddings}: \( e_{ck} \)
- The user’s static preference over locations. \textit{User embedding}: \( e_{uk} \)
Part IV: Location Recommendation & Prediction
Location Recommendation

**Input:** a collection of users’ check-in records

- Each record has: user, a visiting timestamp, a visited point-of-interest (POI) and its geolocation

**Task:** recommending unvisited POIs to users

- Help user explore new places
- Help business owners attract more customers

Time is an important context information

- Recommend brunch restaurants in the morning
- Recommend theaters in the evening
Representative Approaches

User-Based Collaborative Filtering

- Yuan et. al. SIGIR 2013

Matrix Factorization Model

- Gao et. al. RecSys 2013

Probabilistic Models

- Yin et. al. CIKM 2015
Time-aware Point-of-interest Recommendation
[Yuan et. al. SIGIR 2013]

Goal: to recommend POIs for a target user to visit at a specific timeslot in a day:

- Split a day into 24 slots based on hour

Observations: user mobility is influenced by

- Geographical factor: users tend to visit their nearby places
- Temporal factor: users tend to visit different places at different time (morning: libraries; night: pubs)

1. Modeling geographical factor: power-law willingness function

- The willingness of a user to visit $dis$ km far away POI:
  $$w_i(dis) = a \cdot dis^k$$
- The probability a user $u$ visit a POI $l$ at time $t$ is proportional to
  - The product of the willingness of visiting $l$ from each visited POI of $u$
  - The visiting popularity of $l$ at time $t$
Geographical and Temporal Factors

2. Modeling temporal factor: time-aware collaborative filtering

- Consider time dimension when estimating similarity between two users $u$ and $v$ at time $t$

\[
    w_{u,v} = \frac{\sum_{i,j} c_{u,i} c_{v,j}}{\sqrt{\sum_i c_{u,i}^2} \sqrt{\sum_j c_{v,j}^2}} \quad \rightarrow \quad w_{u,v}^{(t)} = \frac{\sum_{i=1}^T \sum_{t=1}^T c_{u,i,t} c_{v,i,t}}{\sqrt{\sum_{i=1}^T \sum_{t=1}^T c_{u,i,t}^2} \sqrt{\sum_{i=1}^T \sum_{t=1}^T c_{v,i,t}^2}}
\]

- Users’ check-in behaviors at different timeslots are correlated
  - Users tend to visit the same POIs at 12pm and 6pm (for meals)
  - Estimate the cosine similarity $\rho_{t,t'}$ between each pair of timeslots $t$ and $t'$, and use it to smooth check-in records
  - Estimate the temporal similarities using the smoothed records, and then rank POIs based on collaborative filtering

Data Sparsity!!!

<table>
<thead>
<tr>
<th>$c_{u,i}$</th>
<th>$l_1$</th>
<th>$l_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_1$</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$t_2$</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$c_{v,i}$</th>
<th>$l_1$</th>
<th>$l_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_1$</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$t_2$</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Similarity: 0?
Performance

Linearly combine the geographical and temporal recommendation scores of each POI as its final score

Experiments

- Datasets: Foursquare and Gowalla check-ins
- Metrics: Precision@N and Recall@N
- Methods: user-based CF(U), U + time-unaware spatial model (U+SB), our time-aware CF and spatial model (UTE+SE)

Exploiting time information can significantly improve the recommendation accuracy
Exploring Temporal Effects for Location Recommendation on Location-Based Social Networks [Gao et. al. RecSys 2013]

Goal: to recommend POIs for a target user to visit:
  - Split a day into 24 slots based on hour

Observations:
  - Human movement exhibits significant daily pattern
    - A user regularly arrives to office at 9am, and goes to a restaurant at 12pm
  - Temporal properties of a user’s daily check-in preferences
    - Non-uniformness: visiting distinct POIs at different hours of a day
    - Consecutiveness: visiting similar POIs in consecutive hours
Modeling Temporal Properties

Basic matrix factorization: 
\[
\min_{U_{ij}, L_{ij} \geq 0} \sum_{i} \sum_{j} Y_{ij}(C_{ij} - U_{ij}L_{ij}^T)^2
\]

- \(m\) users, \(n\) POIs. user-location check-in count matrix \(C \in \mathbb{R}^{m \times n}\), user preference matrix \(U \in \mathbb{R}^{m \times d}\), POI characteristic matrix \(L \in \mathbb{R}^{n \times d}\), indicator matrix \(Y \in \mathbb{R}^{m \times n}\)

- Non-uniformness: visiting distinct POIs at different hours of a day
  - Learn time dependent user preference matrices \(U_t\) at each time \(t\)
- Consecutiveness: a user’s characteristics at consecutive hours are similar.
  - Use cosine coefficient \(\psi_i(t, t-1)\) of user \(u_i\)'s check-in preference at time \(t\) and \(t-1\) to regularize the \(U_t\) and \(U_{t-1}\)
- Jointly optimize and learn \(U_t\) and \(L\) w.r.t. non-uniformness and consecutiveness
Recommendation and Performance

Recommendation: given a target user $u_i$ and a candidate POI $l_j$

1. Compute the time dependent recommendation scores for all timeslots
2. Aggregate these scores by some strategies (sum, mean, etc.) as the final score

Performance on Foursquare check-ins

Time effects are helpful for POI recommendation
Joint Modeling of User Check-in Behaviors for Point-of-Interest Recommendation [Yin et. al. CIKM 2015]

**Goal**: to suggest home-town or out-of-town POIs to users

Factors to be exploited

- Content effect: POI description, e.g., “vegetarian restaurant”
- Temporal effect: user activity exhibits temporal cyclic patterns w.r.t. hour and day
- Geographical influence: people tend to check in around several centers, e.g., “office”
- Word-of-mouth effect: people tend to visit region-level popular POIs
Generative model

Each user has a distribution over topics (e.g., “dinning”)

We learn temporal topics based on the descriptions (w) of POIs and visiting time (t)

- POIs visited at the same time with similar descriptions tend to belong to the same topic

Each region $r$ has a distribution over POIs ($v$) and their geo coordinates ($l_v$)

Parameters are estimated by Gibbs Sampling.
Recommendation and Performance

Given a target user, sort candidate POIs $v$ according to its generative probability

$$P(v, l_v, W_v|u_q, t_q, l_q, \hat{\Psi})$$

Performance

- In-town/out-of-town recommendation: distance(user’s current loc, user’s home) > 100km

Top-k Performance on Foursquare

(a) Out-of-town Recommendation  (b) Home-town Recommendation

Top-k Performance on Twitter

(a) Out-of-town Recommendation  (b) Home-town Recommendation
Location Prediction

**Input**: a collection of users’ trajectories

- Each visit has: user, a visiting timestamp, the geo-coordinates of the visited place

<table>
<thead>
<tr>
<th>User</th>
<th>Time</th>
<th>Coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>7909162</td>
<td>2015 06 03 15:44:24</td>
<td>58.34985748, 11.93956238</td>
</tr>
<tr>
<td>7909162</td>
<td>2015 06 03 17:56:07</td>
<td>58.36145307, 11.91453369</td>
</tr>
<tr>
<td>3818771</td>
<td>2015 06 03 18:11:49</td>
<td>43.17244007, -79.03869084</td>
</tr>
<tr>
<td>7111652</td>
<td>2015 06 03 20:16:33</td>
<td>35.96498954, -83.91925795</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

**Task**: to predict the location that a target user is going to visit next or visit at a target time

**Applications**:

- Logistic planning
- Advertisement targeting
- Pick-up location prediction for carpool
Representative Approaches

Frequent Pattern-based Models
- Monreale et. al. KDD 2009

Hidden Markov Models
- Ye et. al. SDM 2013

Probabilistic Models
- Yuan et. al. WSDM 2017

Supervised Ranking Models
- Noulas et. al. ICDM 2012
WhereNext: a Location Predictor on Trajectory Pattern Mining [Monreale et. al. KDD 2009]

Goal: given the trajectories of users, and the prefix of the trajectory of a target user, predict the next location the user is going to visit.

Basic assumption: people often follow the crowd (common paths)

- Model typical behaviors → predict future movement

Typical behaviors: T-patterns

- A sequence of regions and commuting time

\[ \text{RailwayStation}^{15\text{min}} \rightarrow \text{CastleSquare}^{50\text{min}} \rightarrow \text{Museum} \]
Prediction Pipeline

Extract T-patterns from trajectories

Build T-pattern Tree
  ● decision tree

Given a target trajectory, find the best path on the tree that matches the trajectory
  ● Calculate the matching score for each path

Select the children of the best node that produces the prediction as the next possible locations
What’s Your Next Move: User Activity Prediction in Location-based Social Networks [Ye et. al. SDM 2013]

Goal: to predict the next venue on LSBNs data

- Each record is labeled with semantic information (name and category of the venue)

The prediction space is huge: millions of venues

2-step method

- Predict the category of user activity at the next step (e.g., entertainment:cinema)
- Predicting a venue given the estimated category distributions
Mixed HMM

Basic HMM

- Observations: check-in categories

Mixed HMM with Temporal and Spatial Covariates

- Checkin behaviors are influenced by time and location
  - Outdoor \(\rightarrow\) food; @midnight: Outdoor \(\rightarrow\) nightlife
  - Current pos: a shopping mall \(\rightarrow\) next activity: shopping
- Checkin categories are determined by states and time
  - Time: day of a week + hour of a day

Given the most likely category, rank locations of the category by some ranking schemes (e.g., check-in count, user count, etc.)
Performance

Dataset: 13 million Gowalla check-ins, each check-in has user_id, lat, lng, time, name and category of the location

Category prediction

![Graph showing prediction accuracy across different categories and number of clusters]

Location prediction

<table>
<thead>
<tr>
<th>Method</th>
<th>Top-1</th>
<th>Top-2</th>
<th>Top-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMM [5] with two states</td>
<td>16.82</td>
<td>26.91</td>
<td>34.31</td>
</tr>
<tr>
<td>Max check-in by user</td>
<td><strong>45.63</strong></td>
<td><strong>58.80</strong></td>
<td><strong>62.81</strong></td>
</tr>
</tbody>
</table>
Goal: predict the geo-coordinates of a user at a specified time

Input: GPS trajectories, each record consists of <user, lat/lng, time>

Basic assumption: a user tends to periodically visit a set of regions

- Visiting a coffeehouse every morning
- Visiting a supermarket every Friday evening

Predict locations using the user’s periodic regions

- The regions in which a user shows periodic visiting behavior
- Consists of geo-regions and visiting periods
- Neither is known
Basic Idea

Clustering records by exploiting spatial and temporal information jointly

- Close in distance
- The gap time between two consecutive records should approximately be a multiple of the period

Generative model: for each record

- Sample a region \( r \)
- Sample geo coordinates from the spatial Gaussian distribution of \( r \)
- Sample a gap visiting time from the temporal Gaussian distribution of \( r \)
Prediction and Performance

Given a target user $u$ and target time $t$, we find the region with the greatest visiting probability, and return its center as the predicted location.

Datasets: Gowalla check-ins (Gowalla) of 550 users and Tweets (Twitter) of 550 users

Evaluation metric

- Error distance: Euclidean distance between the true and predicted locations
- Macro (averaged by users); Micro (averaged by test instances)
Mining User Mobility Features for Next Place Prediction in Location-based Services [Noulas et. al. ICDM 2012]

Goal: given a collection of check-ins, a target user u and his/her current check in \(<l', t'>\), predict the venue that u is going to visit next.

- Check-in: \(<\text{venue, time}>\)

Model the prediction problem as a ranking task: compute score for each venue Features

- User Mobility Features
  - Historical Visits: the count of visits of u at a venue
  - Categorical Preferences: the count of visits of u at venues of a category
  - Social Filtering: how many times u’s friends visit a venue
Mining User Mobility Features for Next Place Prediction in Location-based Services [Noulas et. al. ICDM 2012]

Features

- User Mobility Features
- Global Mobility Features
  - Popularity: count of check-ins at a venue
  - Geographic Distance: how far a venue is to u’s all visited venues
  - Activity Transitions: sequence of category transition between consecutive check-ins
- Temporal Features
  - The numbers of check-ins made at a venue (category) within a hour / day

Supervised classifiers:
- linear regression
- M5 model trees
Part V: Research Frontiers and Summary
Integrating Heterogeneous Modalities and Sources

A trend in the confluence of multiple modalities (text, image, space, time, user) and different sources (surveillance camera, traffic sensor, social media, etc)

How to combine different modalities and sources for ambient intelligence?
Online and Efficient Learning

In many scenarios, the data continuously arrive in large volumes.

How do we design online and efficient learning algorithms?

- Semantic drift, high throughput
Representation Learning for Social Sensing Data

Unsupervised representation learning

- How to capture the correlations among different factors
- The learned representations are useful for downstream applications

Supervised representation learning

- Optimize representations for the task at hand
- Often need sufficient data
Learning with Data and Label Sparsity

Data sparsity:

- How to model human activities from small data?
- Can we incorporate external knowledge effectively?

Label sparsity:

- How to build models if the labels are scarce?
- Unsupervised learning, semi-supervised learning, transfer learning
Summary

Space and time coupled social media analysis can be a game changer for many applications.

- Smart city, traffic scheduling, disaster control, mobile healthcare

Many different methodologies have been proposed:

- Topic models, cross-modal embedding, sequential models

Important research challenges remain to be solved:

- Online learning, modality fusion, data scarcity
Thanks!