Learning Urban Community Structures: A Collective Embedding Perspective with Periodic Spatial-temporal Mobility Graphs

Pengyang Wang, Yanjie Fu, Jiawei Zhang, Xiaolin Li, Dan Lin
Outline

- **Background and Motivation**
  - Definition and Problem Statement
  - Methodology
  - Application
  - Evaluation
  - Conclusion
Background and Motivation

- Urban life is getting more diverse and vibrant
Why we study urban communities?

- Spatial Imbalance

---vibrancy differences between communities
Challenges & Insights

- **Challenge I – Graph construction**
  How to unify and represent the POIs and human periodic mobility records as a set of mobility graphs?

- **Insight I**
  a set of periodic spatial-temporal mobility graphs
Challenges & Insights

- **Challenge II – Collective embedding**
  How to collectively learn the embeddings of POIs from multiple periodic mobility graphs?

- **Insight II**
  Collective deep auto-encoder
Challenge III - Embedding aggregation

How to align and aggregate POI embeddings for community structure representation learning?

Insight III

unsupervised graph-based weighting method
Outline

- Background and Motivation
- **Definition and Problem Statement**
- Methodology
- Application
- Evaluation
- Conclusion
Definition I

- **Urban communities**

  - residential complex
  - neighborhood area

  ![Map with a circle radius of 1km indicating urban communities and residential complex areas]
Definition II

- **Mobility Graph**
Definition III

- **Periodic Mobility Graphs**
Problem Statement

- **Given**
  - Residential communities (locations, POIs)
  - Human mobility (e.g., taxi GPS traces)

- **Objective**
  - Learning representations about static spatial configurations
  - Learning representations about dynamic human mobility connectivity of POIs in the community

- **Core tasks**
  - Construction of the periodic mobility graph set for a community
  - Collectively embedding
  - Aggregating and aligning POI embedding into community embedding.
Framework Overview

- Human Mobility
- POI

Periodic Mobility Graph Construction

Community Embedding

POI Embedding Alignment and Aggregation

Collective POI Embedding

POI Embeddings
Outline

- Background and Motivation
- Problem Statement
- Methodology
- Application
- Evaluation
- Conclusion
Methodology

- Periodic Mobility Graph Construction
- Collective POI Embedding
- Aligning and Aggregating POI Embeddings to Community Embeddings
Periodic Mobility Graph Construction

Propagate visit probability

\[ P(x) = \frac{\beta_1}{\beta_2} \cdot x \cdot \exp\left(1 - \frac{x}{\beta_2}\right), \]
\[ \beta_1 = \max_x P(x) \quad \text{and} \quad \beta_2 = \arg \max_x P(x) \]

the closer, the more likely to visit?

Probability distribution w.r.t \( \beta_1 = 0.8, \beta_2 = 100 \).
Collective POI Embedding
Collective POI Embedding

Encoder
\[
\begin{align*}
\hat{y}^{(k),1}_{i,t} &= \sigma(\hat{W}^{(k),1}_{i,t} p^{(k)}_{i,t} + b^{(k),1}_{i,t}), \forall t \in \{1, 2, \ldots, 7\}, \\
y^{(k),r}_{i,t} &= \sigma(\hat{W}^{(k),r}_{i,t} p^{(k)}_{i,t} + b^{(k),r}_{i,t}), \forall r \in \{2, 3, \ldots, o\}, \\
y^{(k),o+1}_{i,t} &= \sigma(\sum_t \hat{W}^{(k),o+1}_{t} y^{(k),o}_{i,t} + b^{(k),o+1}_t), \\
z^{(k)}_i &= \sigma(\hat{W}^{(k),o+2}_{i} y^{(k),o+1}_i + b^{(k),o+2}),
\end{align*}
\]

Decoder
\[
\begin{align*}
\hat{y}^{(k),o+1}_{i} &= \sigma(\hat{W}^{(k),o+2}_{i} z^{(k)}_i + b^{(k),o+2}), \\
\hat{y}^{(k),o}_{i,t} &= \sigma(\hat{W}^{(k),o+1}_{i} \hat{y}^{(k),o}_{i,t} + b^{(k),o+1}_t), \\
\hat{y}^{(k),r-1}_{i,t} &= \sigma(\hat{W}^{(k),r}_{i,t} \hat{y}^{(k),r}_{i,t} + b^{(k),r}_{i,t}), \forall r \in \{2, 3, \ldots, o\}, \\
\hat{p}^{(k)}_{i,t} &= \sigma(\hat{W}^{(k),1}_{i,t} \hat{y}^{(k),1}_{i,t} + b^{(k),1}_{i,t}),
\end{align*}
\]

Loss Function:
\[
L^{(k)} = \sum_{t \in \{1, 2, \ldots, 7\}} \sum_i \| (p^{(k)}_{i,t} - \hat{p}^{(k)}_{i,t}) \odot v^{(k)}_{i,t} \|^2_2
\]
- **Graph based weighting method**

\[
sim_{i,j} = \frac{\sum_l \tilde{G}^{(k)}[i, l] \times \tilde{G}^{(k)}[j, l]}{\sqrt{\sum_l \tilde{G}^{(k)}[i, l]^2} \times \sqrt{\sum_l \tilde{G}^{(k)}[j, l]^2}}
\]
Graph based weighting method

- **Weight Calculation**

\[
\omega_l^{(k)} = \frac{\sum_{i \in c_k} \sum_{j \in c_k} sim_{i,j} \times |\tilde{G}^{(k)}[i, l] - \tilde{G}^{(k)}[j, l]|}{M}
\]

if the l-th dimension of the latent feature makes more sense, when POI \(p_i\) and \(p_j\) are very similar, the difference of \(p_i\) and \(p_j\) on the l-th dimension \(|\tilde{G}^{(k)}[i, l] - \tilde{G}^{(k)}[j, l]|\) should be very small. Therefore, if the l-th dimension of the latent feature does not make much sense, \(|g[i, l] - g[j, l]|\) will increase; if \(p_i\) and \(p_j\) are very similar, \(Sim_{i,j}\) will further penalize \(|g[i, l] - g[j, l]|\)

\[
\hat{G}^{(k)}[s, l] = \sum_{p_i \in \Phi_s} \tilde{G}^{(k)}[i, l] \times \omega_l^{(k)}
\]
Outline

- Background and Motivation
- Definition and Problem Statement
- Methodology
- Application
- Evaluation
- Conclusion
Predicting Willing to Pay (WTP)

\[ r = \frac{P_f - P_i}{P_i} \]

- Final Price
- Initial Price
Spotting vibrant urban communities

\[ u_k = \frac{2 \times \text{freq}^{(k)} \times \text{div}^{(k)}}{\text{freq}^{(k)} \times \text{div}^{(k)}} \]

- **Density of Consumer Activities**
- **Urban Vibrancy Value**
- **Diversity of Consumer Activities**
Outline

- Background and Motivation
- Definition and Problem Statement
- Methodology
- Application
- **Evaluation**
- Conclusion and Future Work
## Data Description

### From Beijing City

<table>
<thead>
<tr>
<th>Data Sources</th>
<th>Properties</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taxi Traces</td>
<td>Number of taxis</td>
<td>13,597</td>
</tr>
<tr>
<td></td>
<td>Effective days</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>Number of trips</td>
<td>8,202,012</td>
</tr>
<tr>
<td></td>
<td>Number of GPS points</td>
<td>111,602</td>
</tr>
<tr>
<td></td>
<td>Total distance (km)</td>
<td>61,269,029</td>
</tr>
<tr>
<td>Residential Communities</td>
<td>Number of residential communities</td>
<td>2,990</td>
</tr>
<tr>
<td></td>
<td>Latitude and Longitude</td>
<td>04/2011 - 09/2012</td>
</tr>
<tr>
<td></td>
<td>Time period of transactions</td>
<td></td>
</tr>
<tr>
<td>POIs</td>
<td>Number of POIs</td>
<td>328668</td>
</tr>
<tr>
<td></td>
<td>Number of POI categories</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Latitude and Longitude</td>
<td></td>
</tr>
<tr>
<td>Check-Ins</td>
<td>Number of check-in events</td>
<td>2,762,128</td>
</tr>
<tr>
<td></td>
<td>Number of POI categories</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Time Period</td>
<td>01/2012-12/2012</td>
</tr>
</tbody>
</table>
Baselines

- Explicit Features (EF): (i) POI numbers per category; (ii) Average commute distance; (iii) Average commute speed; (iv) Average commute time; (v) Number of mobilities; (vi) Average distance between POIs.
- Latent Features (LF): Specifically, the latent features are learned from the proposed collective embedding method.
- The combination of EF and LF (ELF).
- Variation of step1 (V-1): using distance-based matching of the records.
- Variation of step2 (V-2): computing the POI embedding as an average of the embeddings.
- Variation of step3 (V-3): averaging over the POI embeddings.

Evaluation Metric

- Root-Mean-Square Error (RMSE)
The Application of WTP Prediction

- **Results**

<table>
<thead>
<tr>
<th>Feature set</th>
<th>ELF</th>
<th>LF</th>
<th>EF</th>
<th>V-1</th>
<th>V-2</th>
<th>V-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>0.0036</td>
<td>0.0057</td>
<td>0.0422</td>
<td>0.0273</td>
<td>0.0350</td>
<td>0.0193</td>
</tr>
</tbody>
</table>
Spotting vibrant urban communities

Baselines

Learning to Rank

(1) MART: it is a boosted tree model, specifically, a linear combination of the outputs of a set of regression trees.

(2) RankBoost (RB): it is a boosted pairwise ranking method, which trains multiple weak rankers and combines their outputs as final ranking.

(3) LambdaMART (LM): it is the boosted tree version of LambdaRank.

(4) ListNet (LN): It is a listwise ranking model with permutation top-k ranking likelihood as objective function.

(5) RankNet (RN): it uses a neural network to model the underlying probabilistic cost function.

Feature Set

(1) Explicit Features

(2) Latent features

(3) Explicit&Latent features
Evaluation

- **Evaluation Metrics**
  - **Root-Mean-Square Error (RMSE)**
  - **Normalized Discounted Cumulative Gain (NDCG@N)**
    - Evaluate the ranking performance at TopN
  - **Kendall’s Tau Coefficient (Tau)**
    - Measure the overall ranking accuracy.
  - **F-measure@N**
    - “high-vibrancy” and the rating > 3
    - “low-vibrancy” and the rating < 3
    - Measure the ranking precision and recall @ TopN
Overall performance

- NDCG
- F-measure
- Tau

Graphs showing performance metrics for different models and conditions.
Comparison with Representation Learning Algorithms

![Comparison with Representation Learning Algorithms](image-url)

The diagrams above compare the performance of different models using NDCG (Normalized Discounted Cumulative Gain) at various cutoff points (@5, @10, @15, @20). The models compared are:

- **Our Model**
- **RBM**
- **NMF**
- **Skip-gram**

The y-axis represents NDCG, which is a measure of the effectiveness of the ranking. Each bar represents the performance of the models at different cutoff points, showing how well the models perform in ranking items.

**Legend:**
- Red: Our Model
- Turquoise: RBM
- Green: NMF
- Purple: Skip-gram
Investigation of Community Structure Properties

- Community Connectivities.
Investigation of Community Structure Properties

- The Learned Representation of the Community Structure

Visualization of the learned structure representations of two similar communities
Outline

- Background and Motivation
- Definition and Problem Statement
- Methodology
- Application
- Evaluation
- Conclusion
Conclusion

- We formulate the problem as a learning task over multiple mobility graphs of POIs and propose a novel collective embedding framework.

- We started with a probabilistic propagation method to unify and represent static POIs and dynamic human mobility records as periodic spatial-temporal mobility graphs.

- We then developed a collective embedding method to learn the embeddings of POIs from the obtained mobility graphs.

- Based on the POIs embeddings, we further proposed an unsupervised graph based weighted aggregation method to identify community embeddings.

- The method is effective.
Thanks!

Questions?