Joint Representation Learning for Multi-Modal Transportation Recommendation

Hao Liu¹, Ting Li², Renjun Hu³, Yanjie Fu⁴, Jingjing Gu⁵, Hui Xiong¹∗
¹The Business Intelligence Lab, Baidu Research
²National University of Defense Technology, Changsha, China
³SKLSDE Lab, Beihang University, Beijing, China
⁴Missouri University of Science and Technology, Missouri, USA
⁵Nanjing University of Aeronautics and Astronautics, Nanjing, China

Present by: Dr. Hao Liu
Emerging user requirements

High route planning decision cost across multiple transportation modes

- Increasing activity radius
- Complex travel context
- Diversified transportation choices

Personalized and context-aware intelligent route planning
Multi-Modal Transportation Recommendation
Related Work

- Liu et al.\cite{1} discussed generating multi-modal shortest routes based on heterogeneous transportation network.
- MPR\cite{2} and TPMFP\cite{3} mines the most popular routes and the most frequent paths from massive trajectories on the road network, respectively.
- Rogers et al.\cite{4} considers personal preference to improve route recommendations quality.

- Metapath2vec\cite{5} studies network embedding in heterogeneous networks.
- Yao et al.\cite{6} and Wang et al.\cite{7} leverages network embedding for region function profiling.
- Feng et al.\cite{8} and Zhao et al.\cite{9} applies network embedding on POI recommendations.
Trans2vec: Multi-Modal Transportation Recommendation Architecture

Multi-modal data

OD profiling

POI KG

User profiling

Multi-modal transportation graph construction

User

Modes

OD

Real time ETA

Station service

Trans2vec

User profile

Context sensing

Joint representation learning

Online recommendations
Multi-Modal Transportation Graph Construction

- A multi-modal transportation graph is a heterogeneous undirected weighted graph \( G = (V, E) \), where \( V = \cup OD \cup M \) is a set of heterogeneous nodes, and \( E = E_{\text{um}} \cup E_{\text{odm}} \cup E_{\text{uu}} \cup E_{\text{odod}} \) is a set of heterogeneous edges including user-mode edges \( E_{\text{um}} \), OD-mode edges \( E_{\text{odm}} \), user-user edges \( E_{\text{uu}} \) and OD-OD edges \( E_{\text{odod}} \).
The Base Model

• Analogize travel events to sentences and random walks, in order to learn low-dimensional representations of users, OD pairs, and transport modes.

User-mode-OD embedding:

\[ O_0 = \sum_{(u,m) \in \mathcal{E}_{um}} \log \sigma(x_u^\top \cdot x_m) + \sum_{(od,m) \in \mathcal{E}_{odm}} \log \sigma(x_{od}^\top \cdot x_m) \]

Embedding with Negative sampling:

\[ O_0 = \sum_{(u,m) \in \mathcal{E}_{um}} \left( \log \sigma(x_u^\top \cdot x_m) + \log \sigma(-x_u^\top \cdot x_{m'}) \right) + \sum_{(od,m) \in \mathcal{E}_{odm}} \left( \log \sigma(x_{od}^\top \cdot x_m) + \log \sigma(-x_{od}^\top \cdot x_{m'}) \right) \]
Anchor Embedding

Problem

- there are only several (e.g., 5 in our case) transport mode nodes whereas there are a large number of user nodes and OD nodes.

Solution

- each node is assigned a discriminative embedding that reflects its inherent context information.
Modeling User Relevance

• The choice of transport mode is highly influenced by the characteristics of users
  • e.g., age, sex, martial
• User-user relevance:
  \[ \text{rel}(u, u') = \sum_i w_i I(A(u)_i, A(u')_i) / \sum_i w_i \]
• User constraints:
  \[ O_1 = -\frac{1}{2} \sum_{(u, u') \in E_{uu}} (x_u^T \cdot x_{u'} - \text{rel}(u, u'))^2. \]

Beyond travel preference: fined-grained user profile at Baidu
Modeling OD Relevance

• Distance and travel purpose (e.g., home-work, home-commercial) are among the most influential factors for choosing transport modes

• OD relevene: 
  \[ \text{od} = d_{od} \oplus p_o \oplus p_d \]
  \[ \text{rel}(od, od') = \exp\{-||w \odot (od - od')||\} \]

• OD constraints:
  \[ O_2 = -\frac{1}{2} \sum_{(od, od') \in E_{odod}} (x_{od}^T \cdot x_{od'} - \text{rel}(od, od'))^2 \]

OD heat map
Joint Learning & Online Recommendations

- Overall objective:
  \[ O = O_0 + \beta_1 O_1 + \beta_2 O_2 \]

- The score of each mode is computed by:
  \[ f(u, od, m) = \gamma x_u^T \cdot x_m + (1 - \gamma) x_{od}^T \cdot x_m \]

**Algorithm 1: Joint learning algorithm of Trans2Vec**

**Input:** A multi-modal transportation graph G, number d of dimensions, number K, learning rate α, parameters β1 and β2;

**Output:** \( x_u/x_{od}/x_m \) for \( ulodm \in U/O/D/M \);

1. Initialize entries of \( x_u/x_{od}/x_m \) with
   uniform \([-\frac{1}{2d}, \frac{1}{2d}]\);
2. Compute user and OD relevance with Eqs. (4) & (7);
3. \( \text{iter} \leftarrow 1 \);
4. repeat
   5. \( \text{foreach} (u, u') \in E_{uu} \text{ do} \)
      6. \( x_u \leftarrow x_u - \frac{\alpha}{\text{iter}} (x_u^T \cdot x_{u'} - \text{rel}(u, u')) x_{u'}; \)
      7. \( x_{u'} \leftarrow x_{u'} - \frac{\alpha}{\text{iter}} (x_{u'}^T \cdot x_u - \text{rel}(u, u')) x_u; \)
   8. \( \text{foreach} (od, od') \in E_{odod} \text{ do} \)
      9. \( x_{od} \leftarrow \)
         10. \( x_{od} - \frac{\alpha}{\text{iter}} (x_{od}^T \cdot x_{od'} - \text{rel}(od, od')) x_{od'}; \)
     11. \( x_{od'} \leftarrow \)
         12. \( x_{od'} - \frac{\alpha}{\text{iter}} (x_{od'}^T \cdot x_{od} - \text{rel}(od, od')) x_{od}; \)
   13. \( \text{foreach} (u, m) \in E_{um} \text{ do} \)
       14. Sample a transport mode \( m' \sim U; \)
       15. \( x_u \leftarrow x_u - \frac{\alpha}{\text{iter}} (\sigma(x_u^T \cdot x_m) - 1) x_m - \frac{\alpha}{\text{iter}} \sigma(x_u^T \cdot x_{m'}) x_{m'}; \)
   16. \( \text{foreach} (od, m) \in E_{odm} \text{ do} \)
       17. Sample a transport mode \( m' \sim U; \)
       18. \( x_{od} \leftarrow x_{od} - \frac{\alpha}{\text{iter}} (\sigma(x_{od}^T \cdot x_m) - 1) x_m - \frac{\alpha}{\text{iter}} \sigma(x_{od}^T \cdot x_{m'}) x_{m'}; \)
   19. \( \text{iter} \leftarrow \text{iter} + 1; \)
5. until converge;
6. return \( x_u/x_{od}/x_m \) for \( ulodm \in U/O/D/M \);
Experiments – Objectives & Data Sets

Objectives

- The overall performance of Trans2Vec
- The parameter sensitivity
- The transport mode relevance
- The robustness of Trans2Vec

Data sets

- BEIJING and SHANGHAI
- Produced based on the map queries and user feedbacks on the Baidu Map,
- Time window April 1, 2018 - August 20, 2018.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
<th>BEIJING</th>
<th>SHANGHAI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>Q</td>
<td>$</td>
<td># of travel events</td>
</tr>
<tr>
<td>$</td>
<td>U</td>
<td>$</td>
<td># of users</td>
</tr>
<tr>
<td>$</td>
<td>OD</td>
<td>$</td>
<td># of ODs</td>
</tr>
<tr>
<td>$</td>
<td>M</td>
<td>$</td>
<td># of modes</td>
</tr>
</tbody>
</table>

Table 1. Data Statistics
Experiments – Overall Results

Evaluation metrics

- NDCG@5,
- The weighted precision (PREC)
- Recall (REC)
- F1

Table 2. Overall performance

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>BEIJING</th>
<th></th>
<th></th>
<th></th>
<th>SHANGHAI</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NDCG@5</td>
<td>PREC</td>
<td>REC</td>
<td>F1</td>
<td>NDCG@5</td>
<td>PREC</td>
<td>REC</td>
<td>F1</td>
</tr>
<tr>
<td>LR</td>
<td>0.804</td>
<td>0.704</td>
<td>0.589</td>
<td>0.633</td>
<td>0.848</td>
<td>0.682</td>
<td>0.657</td>
<td>0.658</td>
</tr>
<tr>
<td>LTR</td>
<td>0.824</td>
<td>0.667</td>
<td>0.662</td>
<td>0.664</td>
<td>0.830</td>
<td>0.671</td>
<td>0.666</td>
<td>0.668</td>
</tr>
<tr>
<td>PTE</td>
<td>0.770</td>
<td>0.493</td>
<td>0.518</td>
<td>0.499</td>
<td>0.807</td>
<td>0.564</td>
<td>0.610</td>
<td>0.585</td>
</tr>
<tr>
<td>Metapath2Vec</td>
<td>0.731</td>
<td>0.718</td>
<td>0.439</td>
<td>0.515</td>
<td>0.736</td>
<td>0.728</td>
<td>0.451</td>
<td>0.528</td>
</tr>
<tr>
<td>BTrans2Vec</td>
<td>0.876</td>
<td>0.682</td>
<td>0.736</td>
<td>0.704</td>
<td>0.878</td>
<td>0.695</td>
<td>0.754</td>
<td>0.718</td>
</tr>
<tr>
<td>Trans2Vec</td>
<td><strong>0.893</strong></td>
<td>0.700</td>
<td><strong>0.770</strong></td>
<td><strong>0.711</strong></td>
<td><strong>0.891</strong></td>
<td>0.708</td>
<td><strong>0.778</strong></td>
<td><strong>0.719</strong></td>
</tr>
</tbody>
</table>

Baselines

- Logistic regression
- L2R\(^{[10]}\)
- PTE\(^{[11]}\)
- Metapath2Vec\(^{[5]}\)
Experiments – Parameter Sensitivity

Effect of $d$ on BEIJING

Effect of $k$ on BEIJING

Effect of $\beta_1$ on BEIJING

Effect of $\beta_2$ on BEIJING
Experiments – Robustness Check

- We test the performance on four subgroups of users (resp. OD pairs)
  - Group users (resp. OD pairs) by K-means
- The performance is stable in different groups of users and OD pairs.

![Graph showing performance metrics for different groups](image1)

![Graph showing performance metrics for different OD pairs](image2)
Multi-Modal Transportation Recommendation on Baidu Map

Faster than bus & drive

20%

Cheaper than taxi

50%
References


Thanks !
Q & A