Joint Modeling of User Check-in Behaviors for Point-of-Interest Recommendation

Hongzhi Yin\textsuperscript{1}, Xiaofang Zhou\textsuperscript{1}
Yingxia Shao\textsuperscript{2}, Hao Wang\textsuperscript{3}, Shazia Sadiq\textsuperscript{1}

\textsuperscript{1}The University of Queensland, \textsuperscript{2}Peking University
\textsuperscript{3}Institute of Software, Chinese Academy of Sciences
Outlines

• Introduction
• JIM Model
• Experiment
• Conclusion
• Q&A
Introduction
Motivation

- Location-based social network (LBSN) is popular
  - 65% of Americans access social networks on mobile devices
  - 48% of smartphone owners use LBSNs
  - Location-based marketing is expected to be a $39.87 billion business worldwide by 2016.

Refer to: http://www.marketsandmarkets.com/PressReleases/location-based-service.asp
Motivation

- Users can post their physical locations or geo-tag information via “check-in”, and share their visiting experiences with their friends.

POI Contents

A check-in record consists of a five-tuple (user, POI-ID, check-in time, geo-location, check-in contents)
Motivation

• Recommendation is important
  – Too many choices!!
  – “best match” potentially billions of locations to billions of users globally.

**Problem 1. (POI Recommendation)** Given a check-in activity dataset $D$ and a querying user $u_q$ with his/her current location $l_q$ and time $t_q$ (that is, the query is $q = (u_q, t_q, l_q)$), our goal is to recommend a list of POIs that $u_q$ would be interested in. Given a distance threshold $d$, the problem becomes an **out-of-town recommendation** if the distance between the target user’s current location and his/her home location (that is, $|l_q - l_u|$) is greater than $d$. Otherwise, the problem is a **home-town recommendation**.
Challenges

• **Sparsity problem**
  – Billions of POIs around the world.
  – A user visits ~100 POIs.
  – The densities of user-POI matrix are 0.55% and 0.02% for Foursquare and Twitter datasets used in our experiments

• **Travel locality**\(^1\)
  – 45% of users travel 10 miles or less, 75% travel 50 miles or less, and more than 97% users travel 100 miles or less
  – According to our investigations, the check-in records generated by users in their non-home cities are very scarce and only take up 0.47% of the check-in records left in their home cities, which aggravates the data sparsity problem with POI recommendation for out-of-town users.

\(^1\) Levandoski *et al.* Lars: A Location-Aware Recommender System. In ICDE, 2012
Example of Out-of-Town Recommendation
Travel Locality: When U3 travels to Los Angeles that is new to her since she has no check-in history there, how can we recommend POIs to her? In other words, how to link the users in one side to the items in the other side?

Both Graph-based methods and Collaborative Filtering methods would fail in this scenario.
Performance of Mainstream RS

On the Foursquare Dataset

Performance of Traditional user-based CF

Performance of Matrix Factorization
Our Solution - JIM
How Decision Process be Influenced

• **Geographical Influence:** People tend to explore POIs near their current positions or the ones that they have visited before. So, POIs visited by users often form spatial clusters, i.e., people tend to check in around several centers (e.g., “home” and “office”)

• **Temporal Effect:** Users’ activity contents exhibit strong temporal cyclic patterns. For example, a user is more likely to go to a restaurant rather than a night-life spot at lunch time.

• **Content Effect:** A user only prefer a small number of categories of POIs. The check-in activities of users exhibit a strong semantic regularity, e.g., the entropy of categories of POIs an individual user checks-in is very small.
How Decision Process be Influenced

• **Word-of-Mouth Effect**: The region-level popularity of POIs also affects user visiting behaviors. In fact, the probability of a user visiting a POI is largely affected by the local word-of-mouth about the POI, especially when users travel in unfamiliar regions.
  
  – Local attractions. For example, a professor never goes gambling when he lives in Beijing. But when he travels in Las Vegas, he is most likely to visit casinos.

• Need a model that jointly encodes the personal interests, geographical influence, temporal effect and word-of-mouth effect in a unified way to alleviate the issue of data sparsity for effective POI recommendation.
JIM Model

• As a POI has both semantic and geographical attributes, JIM consists of two components:
  – Time-aware User Interest Component (TIC) to exploit content effect and temporal effect to model personal interests
  – Popularity-aware User Mobility Component (PMC) to exploit geographical influence and word-of-mouth effect to account for users’ spatial mobility

<table>
<thead>
<tr>
<th>ID-Name: Darling Harbour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location: Longitude: 151.200, Latitude: -33.877</td>
</tr>
<tr>
<td>Content: plaza, scenic views, harbour, bar, cafés, entertainment sunsets, harbors, museums, leisure, marina</td>
</tr>
</tbody>
</table>
JIM-TIC

- TIC adopts latent topics to characterize users’ interests, i.e., we infer individual user’s interest distribution over a set of topics according to the contents (e.g., tags and categories) of her visited POIs.

The contents play the role of medium which can transfer users’ interests inferred at their home regions to out-of-town regions.
As users’ interests and activities are influenced by time, so the learnt topics should be not only semantic-coherent, but also time-aware.

Each topic in our TIC is not only associated with a distribution over words, but also with a distribution over time. This design is also helpful to improve the quality of topics, since the check-in time provides important clues about the contents of POIs.

Topic: Restaurant & Eating

Topic: Night Life Spot & Entertainment
JIM-PMC

- The spatial clustering phenomenon indicates that users are most likely to visit a small number of POIs, and these POIs are usually limited to some geographical regions.
- The geographical space is divided into $R$ regions. We assume a Gaussian distribution for each region $r$, and the location for POI $v$ is characterized by: $l_v \sim \mathcal{N}(\mu_r, \Sigma_r)$
- We estimate the probability that the user visits each region according to her historical visited POIs or her current location.
- Then, we infer the probability of the POI generated from each region, according to both the public’s preferences at that region and the distance between the POI and the centre of that region.
Generative Process of JIM Model

\[
\begin{align*}
\text{for each region } r & \text{ do} \\
& \quad \text{Draw } \theta_r \sim \text{Dirichlet}(\cdot | \tau) \\
\end{align*}
\]

\[
\begin{align*}
\text{for each user } u & \text{ do} \\
& \quad \text{Draw } \theta_u \sim \text{Dirichlet}(\cdot | \alpha) \\
& \quad \text{Draw } \phi_u \sim \text{Dirichlet}(\cdot | \gamma) \\
\end{align*}
\]

\[
\begin{align*}
\text{for each check-in } (u, v, l_v, W_v, t) & \in D_u \text{ do} \\
& \quad \text{Draw a topic index } z \sim \text{Multi}(\theta_u) \\
& \quad \text{Draw a time } t \sim \text{Beta}(\psi_z, 1, \psi_z, 2) \\
& \quad \text{for each token } w \in W_v \text{ do} \\
& \quad & \quad \text{Draw } w \sim \text{Multi}(\phi_z) \\
& \quad \text{end} \\
& \quad \text{Draw a region index } r \sim \text{Multi}(\theta_u) \\
& \quad \text{Draw a POI index } v \sim \text{Multi}(\phi_r) \\
& \quad \text{Draw a location } l_v \sim \mathcal{N}(\mu_v, \Sigma_v) \\
& \text{end}
\end{align*}
\]
Recommendation Using JIM

- After we estimate the model parameters using the collapsed Gibbs sampling, given a query user $u_q$ with the query time $t_q$ and location $l_q$, i.e., $q = (u_q, t_q, l_q)$, we compute the probability of each unvisited POI $v$ w.r.t. the query $q$, and then select the top-$k$ ones with highest visiting probabilities as recommendations.
Experiment
Data sets

• Foursquare
  – Publicly available
  – Contain 483,813 check-in records of 4163 users who live in the California, USA

• Twitter
  – Publicly available
  – Contain 1,434,668 check-ins of 114,058 users who live across whole USA

• Distributions

![Foursquare](image1) ![Twitter](image2)
Comparative Approaches

- **State-of-Art POI Recommendation Methods**
  - LCA-LDA[^1]
    - Consider both personal interests and local preferences of the target city
  - TCAF[^2]
    - Time-Aware collaborative filtering method
  - UPS-CF[^3]
    - A mixture of user-based collaborative filtering and social-based collaborative filtering
  - SVDFeature[^4]
    - A feature-based matrix factorization model

Comparative Approaches

• Variant versions of JIM
  – JIM-S1: Without consideration of temporal effect
  – JIM-S2: Without consideration of content effect
  – JIM-S3: Without consideration of word-of-mouth effect
  – JIM-S4: Without consideration of geographical influence

<table>
<thead>
<tr>
<th>Methods</th>
<th>Features</th>
<th>Spatial</th>
<th>Temporal</th>
<th>Social or Crowd</th>
<th>Textual</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVDFeature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LCA-LDA</td>
<td></td>
<td>✔️</td>
<td>✔️</td>
<td></td>
<td>✔️</td>
</tr>
<tr>
<td>TACF</td>
<td></td>
<td>✔️</td>
<td></td>
<td></td>
<td>✔️</td>
</tr>
<tr>
<td>UPS-CF</td>
<td></td>
<td>✔️</td>
<td></td>
<td></td>
<td>✔️</td>
</tr>
<tr>
<td>JIM</td>
<td></td>
<td>✔️</td>
<td>✔️</td>
<td></td>
<td>✔️</td>
</tr>
<tr>
<td>JIM-S1</td>
<td></td>
<td>✔️</td>
<td></td>
<td></td>
<td>✔️</td>
</tr>
<tr>
<td>JIM-S2</td>
<td></td>
<td>✔️</td>
<td>✔️</td>
<td></td>
<td>✔️</td>
</tr>
<tr>
<td>JIM-S3</td>
<td></td>
<td>✔️</td>
<td></td>
<td></td>
<td>✔️</td>
</tr>
<tr>
<td>JIM-S4</td>
<td></td>
<td>✔️</td>
<td>✔️</td>
<td></td>
<td>✔️</td>
</tr>
</tbody>
</table>

Features of Different Methods
The recommendation accuracies of all methods are significantly higher in the home-town scenario than in out-of-town scenario;

The same methods have different performances in these two recommendation settings, indicating that the two recommendation scenarios are intrinsically different, and should separately evaluated.
Impact of Different Factors

According to the importance of the four factors, they can be ranked:
1) Content Effect > Word-of-Mouth > Temporal Influence > Geographical Influence for the out-of-town scenario;
2) Temporal Influence > Geographical Influence > Word-of-Mouth > Content for the home-town scenario.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Out-of-Town Scenario</th>
<th>Home-Town Scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ac@1</td>
<td>Ac@10</td>
</tr>
<tr>
<td>JIM-S1</td>
<td>0.052</td>
<td>0.108</td>
</tr>
<tr>
<td>JIM-S2</td>
<td>0.045</td>
<td>0.096</td>
</tr>
<tr>
<td>JIM-S3</td>
<td>0.049</td>
<td>0.101</td>
</tr>
<tr>
<td>JIM-S4</td>
<td>0.056</td>
<td>0.113</td>
</tr>
<tr>
<td>JIM</td>
<td>0.062</td>
<td>0.121</td>
</tr>
</tbody>
</table>

Table 4: Recommendation Accuracy on Foursquare Dataset.

Table 5: Recommendation Accuracy on Twitter Dataset.

Obviously, the content information plays a dominant role in overcoming the issue of data sparsity of out-of-town recommendation scenario, while the temporal effect is most important to improve home-town recommendation.
Impact of Different Factors

• The importance of the same factor w.r.t. the two recommendation scenarios is different.
• This is because the two recommendation scenarios have different characteristics:
  – Most of users have enough check-in records in their home towns while few check-in activities are left in out-of-town regions;
  – The limitation of travel distance in the out-of-town scenario does not matter as much as that in home town;
  – Users’ daily routines may change when they travel out of town.
Conclusions

- To alleviate data sparsity and travel locality, we design a joint probabilistic generative model JIM to model users’ check-in behaviours
  - Time-aware User Interest Component (TIC): temporal effect and content effect
  - Popularity-aware User Mobility Component (PMC): geographical influence and word-of-mouth effect
- Study the importance of each factor in improving both home-town and out-of-town recommendation on two real datasets, and find
  - Content information and the wisdom of crowd play a dominant role in out-of-town recommendation scenario
  - Temporal influence and geographical influence are most important to improve home-town recommendation
- Most recommendation methods display different performance in these two recommendation scenarios, showing the two recommendation scenarios are intrinsically different, and should separately evaluated.
Thank you!

Q&A