PERSONALIZED POI RECOMMENDATION ON LOCATION-BASED SOCIAL NETWORKS

PhD Dissertation Defense
by
Huiji Gao

http://www.public.asu.edu/~hgao16

Committee Members:
Dr. Huan Liu, Chair
Dr. Guoliang Xue
Dr. Jieping Ye
Dr. James Caverlee

November 17, 2014
Outline

- Introduction
- Challenges and Contributions
- Personalized POI Recommendation
- Conclusion and Future Work
Introduction

What is POI Recommendation?

Personalized POI Recommendation on Location-Based Social Networks

Why Personalized?

Why on LBSNs?
What is POI Recommendation?

- If this is your first time visiting **Phoenix**, where should you go?

- Among hundreds of restaurants in **San Francisco Bay Area**, which one should you go to for dinner?
Choice Paralysis

- More choices than ever before – it could cause more problems

- Choice Paralysis

  So many choices! Noooooooooo!

- Recommendation is helpful
  - Help users filter uninteresting POIs
  - Reduce time in decision making
POI Recommendation

- A POI (Point of Interest) is a geographical point with specific functions (e.g., hotel, restaurant, museum, store) that a user may find useful or interesting.

- POI Recommendation (Location Recommendation)
  Recommend POIs (locations) to a user to fulfill his requirements and satisfy his interests
Introduction

What is POI Recommendation?

Personalized POI Recommendation on Location-Based Social Networks

Why Personalized?

Why on LBSNs?
Why Personalized?

- Examples of POI Recommender Systems
Among hundreds of thousands of restaurants in San Francisco Bay Area, which one should I go for dinner?

Yelp

Yelp can suggest some restaurants based on their ratings and your current locations automatically.

- Based on POI Popularity
- Ignore Personal Interests
Why Personalized?

- Examples of POI Recommender Systems

If this is the first time visiting **Phoenix**, where should I go?

**Foursquare**

Foursquare can suggest some places to visit and as some useful tips base on your locations

- Recently released the App with personalization (in Aug 2014*)

*http://searchengineland.com/new-foursquare-app-tips-tastes-deliver-big-personalization-199279
Introduction

What is POI Recommendation?

Personalized POI Recommendation on Location-Based Social Networks

Why Personalized?

Why on LBSNs?
What are Location-Based Social Networks?

Location-based social networks are social networks in which GPS features of mobile devices are used to locate people (and you) and that let you broadcast your location and other content from your mobile device.

Why on Location-Based Social Networks?

- Bridging the Gap between Real World and Online Social Networks

Real World

CHECK IN

Online Social Networks
A Check-in Example on LBSNs

Who: Felix

Where: Shanghai Flavor Shop, Sunnyvale, CA

When: August 1, 2013

What: Best pan-fried pork bun and Shanghai wonton on the west coast!!
How Popular are Location-Based Social Networks?

- 26% of Americans access social networks on mobile devices*
- 18% of smartphone owners use location-based social services*
- Location-based marketing is anticipated to be a $1.8 billion business worldwide by 2015**

**P. Finocchiaro. Mobile advertisers forecast to spend $1.8 billion on location-based campaigns in 2015. 2010.
Personalized POI Recommendation on LBSNs

Problem Statement

Given a user \( u \), a set of POIs (locations) he has checked-in, recommend him some POIs for his future visits based on the LBSN context related to him.
“Existing Location” vs. “New Location”

- Power-Law Distribution on Check-in Frequency

New Location

Cold-Start Check-in Behavior

Existing Location

Repetitive Check-in Behavior
Outline

- Introduction
- Challenges and Contributions
- Personalized POI Recommendation
- Conclusion and Future Work
W⁴: Information Layout on LBSNs

What
- POI-Associated Content (e.g., Tips, POI descriptions)

Who
- Friendships

Where
- Check-in POIs

When
- Time Stamps
Challenges

- Geo-Social Correlations
  - Social
  - Geo
  - Temporal
  - Content

- Geo-Temporal Patterns
- Geo-Content Indications
Contributions

- Investigate Geo-Social correlations to estimate user similarity in terms of various correlation strength and measures, and propose personalized Geo-Social POI recommender systems.

- Study the temporal patterns of user check-in behavior to solve the temporal sparseness problem on LBSNs, propose a personalized geo-temporal POI recommendation framework.

- Identify the challenges of analyzing Geo-Content information on LBSNs, propose to leverage content information for personalized Geo-Content POI recommendation.
Outline

Introduction

Challenges and Contributions

Personalized POI Recommendation

Conclusion and Future Work
Personalized POI Recommender System

- Personalized Geo-Social POI Recommendation
  - Geo-Social Correlations

- Personalized Geo-Temporal POI Recommendation
  - Temporal Cyclic Patterns of Check-ins

- Personalized Geo-Content POI Recommendation
  - User Interests, POI Properties, Sentiment Indications
Geo-Social Correlations

- Geographical Properties of Social Connections

Geographical Distance

\[ \downarrow \quad \uparrow \]

Social Connections

![Graph showing the relationship between geographical distance and the probability of friendship.](image)
Geo-Social Correlations

- Friends with long distance share a small number of commonly visited locations
- Non-friends with short distance share a large number of commonly visited locations
- Users are segmented into four geo-social circles

<table>
<thead>
<tr>
<th></th>
<th>Geo-Social Circles</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( \overline{D} )</td>
<td>( S_{FD} ): Local Friends</td>
<td>( S_{FD} ): Local Non-friends</td>
</tr>
<tr>
<td>( D )</td>
<td>( S_{FD} ): Distant Friends</td>
<td>( S_{FD} ): Distant Non-friends</td>
</tr>
</tbody>
</table>
gSCorr: Geo-Social POI Recommendation Framework

- POI recommendation based on geo-social circles

\[ P_u^t(l) = \Phi_1 P_u^t(l \mid S_{FD}) + \Phi_2 P_u^t(l \mid S_{FD}) + \Phi_3 P_u^t(l \mid S_{FD}) + \Phi_4 P_u^t(l \mid S_{FD}). \]
Personalized POI Recommender System

- Personalized Geo-Social POI Recommendation
  - Geo-Social Correlations

- Personalized Geo-Temporal POI Recommendation
  - Temporal Cyclic Patterns of Check-ins
  - Temporal Chronological Patterns of Check-ins

- Personalized Geo-Content POI Recommendation
  - User Interests, POI Properties, Sentiment Indications
Geo-Temporal Patterns

- **Temporal Cyclic Patterns of Geographical Check-ins**
  - go to restaurant around 11:30 am daily
  - watch movies in a theater on every Friday
  - buy the same camera on 3:00 pm daily?
  - watch “Batman” on every Monday?
Temporal Cyclic Patterns

- **Temporal Cyclic (Periodic Patterns)**
  - One user’s daily check-in activity w.r.t. his top 5 frequently visited locations

- **Temporal Non-uniformness**
  A user presents different check-in preferences at different hours of the day

- **Temporal Consecutiveness**
  A user presents similar check-in preferences at nearby hours of the day
POI Recommendation with NMF

Basic POI Recommendation without Temporal Effects

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

<table>
<thead>
<tr>
<th>( Y \in R^{m \times n} )</th>
<th>Check-in indicator</th>
<th>( C \in R^{m \times n} )</th>
<th>User-Location matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>( U \in R^{m \times d} )</td>
<td>Low-rank representation of user check-in preference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( L \in R^{n \times d} )</td>
<td>Low-rank representation of location preference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( Y_{i,j} = 1 )</td>
<td>User i has checked-in at location j</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Modeling Temporal Non-uniformness

- A user presents different check-in preferences at different hours of a day

$$\min_{U_i \geq 0, L_j \geq 0} \sum_{(i,j) \in \Omega} (C_{i,j} - U_i L_j^T)^2$$

$$\min_{U_i \geq 0, L_j \geq 0} \sum_{t=1}^{24} \sum_{(i,j) \in \Omega} Y_{i,j}^t (C_{i,j}^t - U_i^t L_j^T)^2$$
Modeling Temporal Consecutiveness

- A user presents similar check-in preferences at nearby hour of the day

\[
\min_{U \geq 0} \sum_{t=1}^{T} \sum_{i=1}^{m} \psi_i(t, t-1) \left\| U_t(i,:) - U_{t-1}(i,:) \right\|_F^2
\]

\[
\psi_i(t, t-1) = \frac{C_t(i,:) \cdot C_{t-1}(i,:)}{\sqrt{\sum_j C_t^2(i,:) \sqrt{\sum_j C_{t-1}^2(i,:)}}}
\]
Framework of POI Recommendation with Temporal Effects

\[ C \rightarrow C_1 \rightarrow U_1 \rightarrow L \rightarrow U_T \rightarrow \tilde{C}_1 \]

\[ C_2 \rightarrow U_2 \rightarrow \cdots \rightarrow \tilde{C}_2 \]

Unobserved Check-ins

\[ T=24 \]

\[ \tilde{C} \]

Approximated Check-in Preference

Data Mining and Machine Learning Lab
Personalized POI Recommender System

- Personalized Geo-Social POI Recommendation
  - Geo-Social Correlations

- Personalized Geo-Temporal POI Recommendation
  - Temporal Cyclic Patterns of Check-ins

- Personalized Geo-Content POI Recommendation
  - User Interests, POI Properties, Sentiment Indications
Geo-Content Indications

- **POI-Associated Content** (POI-Based Descriptive Tags)
  - e.g., Tags for “Shima Japanese Restaurant”
    {sashimi, fresh fish, Japanese cuisine, sushi}
  - POI Property
  - Users who visit this location may have interests on these facets

- **User-Generated Content** (User Check-in Tips)
  - e.g., {The Spicy beef tendon, cumin lamb, and mapo tofu are all great.}

  - User interested topics
    Spicy, tofu, beef tendon, …
  - Sentiment Indication
    all great {Positive}
Geo-Content Indications

Content Information on LBSNs

- POI-Associated Content
  - POI Property
- User-Generated Content
  - User Interests
  - Sentiment Indication

Check-in Actions

Key factors of Check-in Actions?

- What is the POI about? POI Property
- Am I interested? User Interests
- How good is the POI? Sentiment Indication
Methodology

A Basic POI Recommendation Model

$$\min_{U_i \geq 0, H \geq 0, L_j \geq 0} \sum_i \sum_j W_{i,j} (C_{i,j} - U_i H L_j^T)^2$$

<table>
<thead>
<tr>
<th>Input</th>
<th>$W \in R^{m \times n}$</th>
<th>Check-in indicator</th>
<th>$C \in R^{m \times n}$</th>
<th>User-POI matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$W_{i,j} = 1$</td>
<td>User i has checked-in at POI j</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Output</td>
<td>$U \in R^{m \times k}$</td>
<td>User interests in latent dimension</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$L \in R^{n \times k}$</td>
<td>POI property in latent dimension</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$H \in R^{k \times k}$</td>
<td>Data-Dependent Information</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Modeling Sentiment Indication

\[
\min_{U_i \geq 0, H \geq 0, L_j \geq 0} \sum_{i}^{m} \sum_{j}^{n} W_{i,j} (C_{i,j} - U_i H L_j^T)^2
\]

- **Determine the Importance of a POI**
  - Check-in Weighting Matrix \( W \)
  - Sentiment Indication Matrix \( S \in R^{m \times n}, S_{i,j} \in [-1,1] \)

- **Sentiment-Enhanced Weighting Scheme**

\[
\hat{W} = f(W, S)
\]
Modeling Sentiment Indication

\[
\min_{U_i \geq 0, H \geq 0, L_j \geq 0} \sum_i \sum_j W_{i,j} (C_{i,j} - U_i H L_j^T)^2
\]

- Sentiment-Enhanced Weighting Scheme

\[
\hat{W} = f(W, S)
\]

- Sentiment Consistency
- Sentiment Scaling
- Non-Negativity

\[
\hat{W} = W + \eta S, \eta \in [0,1]
\]

For any observed Check-ins

\[
W_{i,j} = 1, S_{i,j} \in [-1,1], \hat{W}_{i,j} \geq 0
\]
Modeling User Interested Topics

\[
\min_{U_{i} \geq 0, G_{j} \geq 0} \sum_{i} \sum_{j} (A_{i,j} - U_{i}G_{j})^{2}
\]

<table>
<thead>
<tr>
<th>$A \in R^{m \times d}$</th>
<th>Observed User-Tag Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>$G \in R^{k \times d}$</td>
<td>Tag-Topic Matrix in Latent Dimension</td>
</tr>
</tbody>
</table>
Modeling POI-Associated Content

\[
\min_{V_i \geq 0, G_j \geq 0} \sum_{i}^{n} \sum_{j}^{d} (B_{i,j} - V_i \tilde{G}_j)^2
\]

<table>
<thead>
<tr>
<th>$B \in \mathbb{R}^{n \times d}$</th>
<th>Observed POI-Tag Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tilde{G} \in \mathbb{R}^{k \times d}$</td>
<td>Tag-Topic Matrix in Latent Dimension</td>
</tr>
</tbody>
</table>
Modeling POI-Associated Content

\[ \min \left\| G - \hat{G} \right\|_1 \]

Latent Topics

Check-in Actions
Methodology

Objective Function

\[
\min W \circ \left\| C - UHL^T \right\|_F^2 + \lambda_1 \left\| A - UG \right\|_F^2 + \lambda_2 \left\| B - LG \right\|_F^2 \\
+ \delta \left\| G - \tilde{G} \right\|_1 + \alpha \left( \left\| U \right\|_F^2 + \left\| H \right\|_F^2 + \left\| L \right\|_F^2 + \left\| G \right\|_F^2 + \left\| \tilde{G} \right\|_F^2 \right)
\]

<table>
<thead>
<tr>
<th>Input</th>
<th>( W \in R^{m \times n} )</th>
<th>Check-in indicator</th>
<th>( C \in R^{m \times n} )</th>
<th>User-Location matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A \in R^{m \times d} )</td>
<td>User-Tag Matrix</td>
<td>( B \in R^{n \times d} )</td>
<td>Location-Tag Matrix</td>
<td></td>
</tr>
<tr>
<td>( S \in R^{m \times n} )</td>
<td>Sentiment Indication</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Output</th>
<th>( U \in R^{m \times k} )</th>
<th>User Interests</th>
<th>( L \in R^{n \times k} )</th>
<th>Location Property</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H \in R^{k \times k} )</td>
<td>Data-Dependent Information</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( G \in R^{d \times k} )</td>
<td>Tag-Topic Matrix</td>
<td>( \tilde{G} \in R^{d \times k} )</td>
<td>Tag-Topic Matrix</td>
<td></td>
</tr>
</tbody>
</table>
Methodology

Objective Function

\[
\begin{align*}
\min W \circ \|C - UHL^T\|_F^2 + \lambda_1 \|A - UG\|_F^2 + \lambda_2 \|B - L\hat{G}\|_F^2 \\
+ \delta \|G - \hat{G}\|_1 + \alpha (\|U\|_F^2 + \|H\|_F^2 + \|L\|_F^2 + \|G\|_F^2 + \|\hat{G}\|_F^2 )
\end{align*}
\]

\[D = G - \hat{G}\]

\[
\begin{align*}
\min W \circ \|C - UHL^T\|_F^2 + \lambda_1 \|A - UG\|_F^2 + \lambda_2 \|B - L(G - D)\|_F^2 \\
+ \delta \|D\|_1 + \alpha (\|U\|_F^2 + \|H\|_F^2 + \|L\|_F^2 + \|G\|_F^2 )
\end{align*}
\]
Datasets

Foursquare Dataset: (Food Category)

<table>
<thead>
<tr>
<th></th>
<th>CA</th>
<th>NY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Users</td>
<td>4,287</td>
<td>6,043</td>
</tr>
<tr>
<td>Number of Check-ins</td>
<td>134,556</td>
<td>207,591</td>
</tr>
<tr>
<td>Number of POIs</td>
<td>5,878</td>
<td>5,937</td>
</tr>
<tr>
<td>Number of Tips</td>
<td>19,741</td>
<td>40,539</td>
</tr>
<tr>
<td>Number of Comments</td>
<td>56,718</td>
<td>78,290</td>
</tr>
<tr>
<td>Check-in Duration</td>
<td>May, 2008-Sep, 2013</td>
<td></td>
</tr>
</tbody>
</table>
Experimental Setup

- **Sentiment Indication Matrix S**
  - word-matching scheme
  - **MPQA** (2,718 positive and 4,902 negative words)
  - normalized to \([-1, 1]\)

- **Training/Testing Data:**
  For each individual, mark off 80% of all locations chronologically that he has checked-in for training, the rest 20% are used as testing.

- **Evaluation Metrics:** Precision@N, Recall@N
- Comparing with State-of-the-art Approaches
- Investigating the Effect of Different Content Information
- Variation of Parameters
Experiment

- Comparing with State-of-the-art Approaches
  - User-Based Collaborative Filtering (UCF)
  - Probabilistic Matrix Factorization (PMF)
  - Non-negative Matrix Factorization (NMF)
  - Spatial Topic Location Recommender (STLR)
  - Sentiment-Enhanced Location Recommender (SELR)
  - CAPRF (Our Method)

### Table 4: Performance Comparison (CA)

<table>
<thead>
<tr>
<th>Methods</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P@5</td>
<td>P@10</td>
</tr>
<tr>
<td>UCF</td>
<td>0.0083</td>
<td>0.0077</td>
</tr>
<tr>
<td>PMF</td>
<td>0.0114</td>
<td>0.0104</td>
</tr>
<tr>
<td>NMF</td>
<td>0.0126</td>
<td>0.0111</td>
</tr>
<tr>
<td>STLR</td>
<td>0.0173</td>
<td>0.0150</td>
</tr>
<tr>
<td>SELR</td>
<td>0.0134</td>
<td>0.0121</td>
</tr>
<tr>
<td>CAPRF</td>
<td>0.0186</td>
<td>0.0169</td>
</tr>
</tbody>
</table>

### Table 5: Performance Comparison (NY)

<table>
<thead>
<tr>
<th>Methods</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P@5</td>
<td>P@10</td>
</tr>
<tr>
<td>UCF</td>
<td>0.0052</td>
<td>0.0047</td>
</tr>
<tr>
<td>PMF</td>
<td>0.0109</td>
<td>0.0099</td>
</tr>
<tr>
<td>NMF</td>
<td>0.0114</td>
<td>0.0102</td>
</tr>
<tr>
<td>STLR</td>
<td>0.0138</td>
<td>0.0125</td>
</tr>
<tr>
<td>SELR</td>
<td>0.0124</td>
<td>0.0113</td>
</tr>
<tr>
<td>CAPRF</td>
<td>0.0158</td>
<td>0.0143</td>
</tr>
</tbody>
</table>
# Experiment

- Investigating Different Content Information

## Table 6: Recommendation Effect of Different Types of Content Information

<table>
<thead>
<tr>
<th>Information</th>
<th>SI</th>
<th>UIC</th>
<th>PPC</th>
<th>CA P@5</th>
<th>R@5</th>
<th>P@10</th>
<th>R@10</th>
<th>NY P@5</th>
<th>R@5</th>
<th>P@10</th>
<th>R@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>NONE</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>0.0126</td>
<td>0.0177</td>
<td>0.0111</td>
<td>0.0310</td>
<td>0.0110</td>
<td>0.0120</td>
<td>0.0102</td>
<td>0.0223</td>
</tr>
<tr>
<td>SI</td>
<td>√</td>
<td>×</td>
<td>×</td>
<td>0.0130</td>
<td>0.0182</td>
<td>0.0115</td>
<td>0.0323</td>
<td>0.0117</td>
<td>0.0127</td>
<td>0.0108</td>
<td>0.0234</td>
</tr>
<tr>
<td>UIC</td>
<td>×</td>
<td>√</td>
<td>×</td>
<td>0.0164</td>
<td>0.0230</td>
<td>0.0148</td>
<td>0.0416</td>
<td>0.0138</td>
<td>0.0151</td>
<td>0.0128</td>
<td>0.0279</td>
</tr>
<tr>
<td>PPC</td>
<td>×</td>
<td>×</td>
<td>√</td>
<td>0.0154</td>
<td>0.0217</td>
<td>0.0138</td>
<td>0.0387</td>
<td>0.0130</td>
<td>0.0141</td>
<td>0.0120</td>
<td>0.0261</td>
</tr>
<tr>
<td>UIC+SI</td>
<td>√</td>
<td>√</td>
<td>×</td>
<td>0.0172</td>
<td>0.0242</td>
<td>0.0154</td>
<td>0.0433</td>
<td>0.0140</td>
<td>0.0152</td>
<td>0.0130</td>
<td>0.0283</td>
</tr>
<tr>
<td>PPC+SI</td>
<td>√</td>
<td>×</td>
<td>√</td>
<td>0.0157</td>
<td>0.0221</td>
<td>0.0141</td>
<td>0.0396</td>
<td>0.0133</td>
<td>0.0144</td>
<td>0.0123</td>
<td>0.0267</td>
</tr>
<tr>
<td>UIC+PPC</td>
<td>×</td>
<td>√</td>
<td>√</td>
<td>0.0178</td>
<td>0.0250</td>
<td>0.0162</td>
<td>0.0456</td>
<td>0.0149</td>
<td>0.0163</td>
<td>0.0138</td>
<td>0.0301</td>
</tr>
<tr>
<td>SI+UIC+PPC</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>0.0186</td>
<td>0.0261</td>
<td>0.0169</td>
<td>0.0474</td>
<td>0.0154</td>
<td>0.0168</td>
<td>0.0143</td>
<td>0.0311</td>
</tr>
</tbody>
</table>
Experiment

(a) Sentiment Indications- $\eta$

(b) User-Interest Content- $\lambda_1$

(c) POI-Property Content- $\lambda_2$

(d) Semantic Overlapping- $\delta$
Conclusion and Future Work

- Temporal-based Content Analysis
- Tensor-Based Methods
- Relationships Among Multiple Information
- Location-Based Mobile Applications
Achievements

- 10+ First Author Papers
- PERSONALIZED RECOMMENDATION ON LOCATION-BASED SOCIAL NETWORKS
- BOOK
- 591 Citations, h-index: 14
- Third Prize in Nokia Mobile Challenge
- Tutorial
- Encyclopedia Entry
- 20+ co-author papers
Acknowledgments

- Committees
  - Professors Huan Liu, Guoliang Xue, and Jieping Ye, James Caverlee

- Data Mining and Machine Learning Lab (DMML) @ ASU

http://dmml.asu.edu/

- Office of Naval Research (ONR)
- Everyone attending my defense