LCARS: A Location-Content-Aware Recommender System

Hongzhi Yin†, Yizhou Sun‡, Bin Cui†
Zhiting Hu†, Ling Chen

†Peking University  ‡Northeastern University
University of Technology, Sydney
Outline

- Introduction
  - Background
  - Challenges

- Our Solution – LCARS
  - Offline Modeling - LCA-LDA
  - Online Recommendation – TA algorithm

- Experiments
  - Experimental Setup
  - Experimental Results

- Conclusions
Outline

■ Introduction
  □ Background
  □ Challenges

■ Our Solution – LCARS
  □ Offline Modeling - LCA-LDA
  □ Online Recommendation – TA algorithm

■ Experiments
  □ Experimental Setup
  □ Experimental Results

■ Conclusions
Background

Location-based Social Networks (LBSNs)

- Loopt
- Foursquare
- Facebook Places

- Users share photos, comments or check-ins associated with a location
- Expanded rapidly, e.g., Foursquare gets over *3 million* check-ins every day
Background

- Event-based Social Networks, e.g. Meetup.com (EBSNs)

- **Let's Go Retro: 80s Theme Dance**
  - @ Culture Club Dance Club
  - Fri 04/20 10:00 PM

- **Spring Is Here!!! Sangria & Co**
  - Fri 04/20 10:00 PM

- **Battle of the Sexes Multi Group**
  - @ THE WORLD BAR
  - Fri 04/20 10:00 PM

- **Battle of the Sexes Multi Group**
  - Fri 04/20 10:00 PM
Problem Definition

We aim to mine useful knowledge from the user activity history data in LBSNs and EBSNs to answer two typical questions in our daily life:

- If we want to visit venues in a city such as Beijing, where should we go?
- If we want to attend local events such as dramas or exhibitions in a city, which events should we attend?

- **Spatial Item**: venue or event associated with location
- **Problem**: given a querying user $u$ with a querying city $l_u$, find $k$ interesting spatial items within $l_u$, that match the preference of $u$. 
Challenge(1/4)

- **Spatial Item Recommendations in LBSN and EBSN**
- **Existing Solutions**
  - Based on item/user collaborative filtering
  - Similar users gives the similar ratings to similar items

### Challenge (2/4)

- **User-item rating/visiting matrix**

<table>
<thead>
<tr>
<th></th>
<th>$V_1$</th>
<th>$V_2$</th>
<th>$V_3$</th>
<th>...</th>
<th>...</th>
<th>...</th>
<th>$V_{m-2}$</th>
<th>$V_{m-1}$</th>
<th>$V_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>User</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$U_0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$U_i$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$U_j$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$U_n$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Noulas, S. Scellato, C Mascolo and M Pontil “An Empirical Study of Geographic User Activity Patterns in Foursquare” (ICWSM 2011)*
### User-item rating/visiting matrix

*Millions of spatial items around the world*

<table>
<thead>
<tr>
<th></th>
<th>$V_1$</th>
<th>$V_2$</th>
<th>$V_3$</th>
<th>...</th>
<th>...</th>
<th>...</th>
<th>$V_{m-2}$</th>
<th>$V_{m-1}$</th>
<th>$V_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>User</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$U_0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$U_i$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$U_j$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$U_n$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A user visits ~100 spatial items

---

Noulas, S. Scellato, C Mascolo and M Pontil “An Empirical Study of Geographic User Activity Patterns in Foursquare” (ICWSM 2011)
**Challenge (2/4)**

- **User-item rating/visiting matrix**

  **Millions of spatial items around the world**

Los Angeles

<table>
<thead>
<tr>
<th></th>
<th>$V_1$</th>
<th>$V_2$</th>
<th>$V_3$</th>
<th>...</th>
<th>...</th>
<th>...</th>
<th>$V_{m-2}$</th>
<th>$V_{m-1}$</th>
<th>$V_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>User</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$U_0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$U_i$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$U_j$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$U_n$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A user visits ~100 spatial items

---

Noulas, S. Scellato, C. Mascolo and M. Pontil “An Empirical Study of Geographic User Activity Patterns in Foursquare” (ICWSM 2011)
### Challenge (2/4)

**User-item rating/visiting matrix**

Millions of spatial items around the world

<table>
<thead>
<tr>
<th></th>
<th>V₁</th>
<th>V₂</th>
<th>V₃</th>
<th>…</th>
<th>…</th>
<th>…</th>
<th>Vₘ₋₂</th>
<th>Vₘ₋₁</th>
<th>Vₘ</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>User</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U₀</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>…</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uᵢ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uⱼ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>…</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uₙ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A user visits ~100 spatial items

User activity histories are locally clustered

---

Noulas, S. Scellato, C Mascolo and M Pontil “An Empirical Study of Geographic User Activity Patterns in Foursquare” (ICWSM 2011)
**User-item rating/visiting matrix**

*Millions* of spatial items around the world

(a) New York users in Los Angeles  
(b) New York users in New York City.

Noulas, S. Scellato, C Mascolo and M Pontil “An Empirical Study of Geographic User Activity Patterns in Foursquare” (ICWSM 2011)
User activity histories are **locally clustered**

When $U_5$ travels to Los Angeles that is new to him
- **User-based CF?** Users similar to $U_5$ rarely rated items in LA.
- **Item-based CF?** Rating patterns of Items in LA are not similar to that of items in NYC.
- Data sparsity
- User’s activities are very limited in distant locations
  - Things can get worse in totally **NEW Areas**
    - *(Where you need recommendations the most)*
  - New City Problem
Outline

- Introduction
  - Background
  - Challenges

- Our Solution – LCARS
  - Offline Modeling - LCA-LDA
  - Online Recommendation – TA algorithm

- Experiments
  - Experimental Setup
  - Experimental Results

- Conclusions
Our Main Ideas (1/3)

For spatial item recommendation, we consider
• the querying user’s interest;
• the local preference of the querying city,
  • the local word-of-mouth opinion for a spatial item in the querying city.

1. User Personal Interests/Preferences

2. Local Preference

Recommender System
Our Main Ideas (2/3)

Main idea #1:
Identify user interest using semantic information from the user activity history

Main idea #2:
Discover local preference in a specific querying city

Main idea #3:
Combine user interest & local preference for recommendation in a unified way

User Personal Interests/Preferences

Local Preference in a querying city
Our Main Ideas (3/3)

Content Words of Items
Such as tags and category (e.g., movie, shopping, nightlife)

The users in one side and the items in the other side can be linked together by the item contents.
The model learns:

- **Topic**: Each topic $z$ in our work has two topic models $\phi_z$ and $\phi'_z$. The former is a probability distribution over items (item ID) and the latter is a probability distribution over content words.

- **User Interest**: The intrinsic **interest of user $u$** is represented by $\theta_u$, a probability distribution over topics.

- **Local Preference**: The **local preference in region $l$** is represented by $\theta_l$, a probability distribution over topics.
We use LCA-LDA model to simulate the process of user decision-making for visiting behaviors.
Outline

■ Introduction
  - Background
  - Challenges

■ Our Solution – LCARS
  - Offline Modeling - LCA-LDA
  - Online Recommendation – TA algorithm

■ Experiments
  - Experimental Setup
  - Experimental Results

■ Conclusions
Online Recommendation

- The **offline model** learned the parameters:
  - topics $\phi_z$ and $\phi'_z$
  - user interest $\theta_u$
  - local preference $\theta_l$
  - mixing weights $\lambda_u$

- For a query $(u, l_u)$, the **online recommendation part** computes a ranking score for each spatial item $v$ within querying region $l_u$, and then returns top-$k$ ranked spatial items as the recommendations.

$$P(v|\theta_u, \theta'_l, \phi, \phi') = \lambda_u P(v|\theta_u, \phi, \phi') + (1-\lambda_u) P(v|\theta'_l, \phi, \phi')$$
A brute-force alg. computes ranking scores for all items within the querying region \( l_u \), which is computationally expensive and too slow for online recommendation, esp. when there are millions of items.

We extend the **Threshold Algorithm (TA)**

- A greedy (but exact) algorithm
- capable of correctly finding top-k results by examining the minimum number of spatial items.
- The alg. maintains a threshold value \( T_a \) during examining items, representing the maximum possible ranking score that can be achieved by remaining unexamined items.
- Hence, if the smallest ranking score of the top-k examined items is no less than the threshold score, the algorithm can terminate immediately.
Outline

- Introduction
  - Background
  - Challenges

- Our Solution – LCARS
  - Offline Modeling - LCA-LDA
  - Online Recommendation – TA algorithm

- Experiments
  - Experimental Setup
  - Experimental Results

- Conclusions
### Experimental Data Sets

#### Data Sets

- **DoubanEvent**: DoubanEvent is China’s largest event-based social networking site where users can publish and participate in social events. This data set consists of 100,000 users, 300,000 events and 3,500,000 check-ins.

- **Foursquare**: This dataset contains 11,326 users, 182,968 venues and 1,385,223 check-ins.

#### User and Event Distributions over Cities in DoubanEvent

![User Distribution](image1.png)

- **Beijing**: 38.04%
- **Shanghai**: 31.72%
- **Guangzhou**: 19.45%
- **Shenzhen**: 6.61%
- **Others**: 4.18%

![Event Distribution](image2.png)

- **Beijing**: 48.44%
- **Shanghai**: 24.00%
- **Guangzhou**: 17.26%
- **Shenzhen**: 4.43%
- **Others**: 6.00%
Two real settings to evaluate the recommendation effectiveness:

- Querying cities are new cities to querying users;
- Querying cities are home cities to querying users;

Baseline:

- USG, User-based CF, Item-based CF, LDA
- Location-Aware LDA (LA-LDA): One component of LCA-LDA
- Content-Aware LDA (CA-LDA): Another component of LCA-LDA
Experimental Results

- Recommendation Effectiveness

(a) Users Traveling in New Cities
(b) Users Traveling in Home Cities

Figure 3: Top-$k$ Performance on DoubanEvent
Experimental Results

- Recommendation Effectiveness

Figure 4: Top-\(k\) Performance on Foursquare
Experimental Results

- Efficiency of online recommendation, querying cities are Beijing and Shanghai

![Graphs showing processing times in Beijing and Shanghai](image)

(a) Processing Time in Beijing
72,000 items

(b) Processing Time in Shanghai
51,784 items
Outline

■ Introduction
  ❑ Background
  ❑ Challenges

■ Our Solution – LCARS
  ❑ Offline Modeling - LCA-LDA
  ❑ Online Recommendation – TA algorithm

■ Experiments
  ❑ Experimental Setup
  ❑ Experimental Results

■ Conclusions
Conclusion

Spatial item Recommendations

- Data sparsity is a big challenge in recommendation systems
- New city problem amplify the data sparsity challenge
- Mobile scenario requires the recommender system to generate real-time response to the user query.

Our Solution - LCARS

- Exploit the Local Preference of the querying city to alleviate the data sparsity. Local word-of-mouth is a valuable resource for making a recommendation.
- Take advantage of Content Information of items to overcome the sparsity. The contents build a bridge between users and items from disjoint regions.
- Extend the Threshold-based algorithm (TA) to produce fast online recommendations

Result

- LCARS is more effective and more efficient
Thanks