Inferring User Context From Spatio-Temporal Pattern Mining for Mobile Application Services

2012 IEEE/WIC/ACM WI-IAT Special Session 1

SkillUpJapan Corporation - R&D Department

Daniel Pereira  dpereira@skillupjapan.co.jp
Luis Loyola    lloyola@skillupjapan.co.jp
Summary

• Introduction
• Purpose
• Analysis
• Proposed approach
• Results
• Future work & conclusion
Introduction

• Our company provides several services to our users throughout several platforms:
  • Video and advertisement platforms
  • Recommender engine
  • Mobile applications
  • Social medical website
  • Electronic commerce

• We launched some mobile applications that make use of our advertisement platform.

• However, user information is difficult to gather.
Introduction

• Our company provides several services to our users throughout several platforms:
  • Video and advertisement platforms
  • Recommender engine
  • Mobile applications
  • Social medical website
  • Electronic commerce

• We launched some mobile applications that make use of our advertisement platform

• However, mobile user information is difficult to gather
Purpose

• On the other hand, advertisers want to target their users as accurately as possible

• Most of current online advertisement research still focuses on desktop users

• Advertisers want to know more about user habits
  
  • However, advertisements are integrated into a mobile applications by means of an external SDKs

• External parties cannot ask user information
Purpose

• So far, a user is but a random number with a history of items combined with latitudes and longitudes

• Recently, Location-Based Social Networks like Foursquare, Facebook or Google Places are increasingly becoming rich with location metadata

• **Goal**: Integrate user geographic data with location metadata into the current advertisement platform and recommender engines
Clustering analysis

- There are several ways to cluster
  - Hierarchical or partitional
  - Exclusive, overlapping or fuzzy
  - Complete or partial

- As well as several algorithms
  - Centroid-based (based on center of mass)
  - Density-based (how close information is)
  - Distribution-based (statistical distribution)
  - Connectivity-based (hierarchy)

- Some techniques are a better fit for some tasks than others
Clustering analysis

But ... which one to use?!
Clustering analysis

• We found three main aspects that narrows down our choices:

1. We do not know, beforehand, how many location clusters a user has
   • Work and home can be assumed, but a mobile device provides many more choices, depending on user activities, family and circle of friends

2. We do not need to cluster every point a user has visited, just the ones the user has visited more often
   • This does not tell us directly if a place is relevant or not, but we assume so after many visits from a user

3. User geographic patterns differ from typical clustering patterns
Clustering analysis

• Hierarchical clustering
  • Too sensitive to noise and outliers
  • Does not handle outliers very well

• Fuzzy clustering
  • Computation is slower, even with low dimensionality
  • Need to specify the number of clusters (K)
Clustering analysis

- **K-means (K=2)**
  - Considers every point
  - Need to specify K
  - Non deterministic

- **Orthogonal regression**
  - Too sensitive
  - Better fit for evenly dispersed data

- **Density clustering**
  - Considers outliers
  - No need to specify K
  - Supports any shape
Clustering analysis

• Density clustering
  • Considers outliers
  • No need to specify K
  • Supports any shape
Clustering analysis

- Density clustering
  - Considers outliers
  - No need to specify K
  - Supports any shape

- We chose DBScan [1]
  - Its parameters are further analyzed in the paper

Proposed platform

1. Application requests banner

2. Server logs request and gets list of recommended banners

3. Server sends a banner

4. User logs are parsed

5. User data is available to analysis

6. Location is analyzed and context inferred with LSN

7. User recommendation is updated
Proposed approach

• Parse logs from the advertisement platform
  • This platform can be used in any support, however mobile devices provide us richer geographic information

• Unsupervised learning
  • Clustering seems a good candidate

• Extract context information from Location Social Networks (LSNs)
  • Foursquare, Facebook, Google Places
Context analysis

- We aim to use LSNs, rich on location content provided by the users

Sample Facebook result:

```
"data": [
    {
        "name": "昭和大学病院",
        "category": "Local business",
        "location": {
            "street": "旗の台1-5-8",
            "city": "Shinagawa-ku",
            "state": "Tokyo",
            "country": "Japan",
            "zip": "142-8666",
            "latitude": 35.60806622729,
            "longitude": 139.70385268404
        },
        "id": "145842375451592"
    }
],
```

Sample Foursquare result:

```
"venues": [
    {
        "id":"4b5dcfaaf964a520046d29e3",
        "name":"Coca-Cola",
        "location":{
            "address":"Avenida Kennedy 5757",
            "crossStreet":"Manquehue",
            "lat":-33.39930950236016,
            "lng":-70.57355403900146,
            "distance":51,
            "city":"Las Condes",
            "state":"Santiago",
            "country":"Chile"
        },
        "categories":[
            {
                "id":"4bf58dd8d48988d124941735",
                "name":"Office",
                "pluralName":"Offices",
                "shortName":"Corporate \ Office",
                "icon":{
                    "prefix":"https://foursquare.com/img\categories\building\default_",
                    "sizes":[32,44,64,88,256],
                    "name":"
                },
                "primary":true
            }
        ],
        "verified":false,
        "stats":{
            "checkinsCount":2123,
            "usersCount":124,
            "tipCount":4
        },
        "hereNow":{"count":0}
    }
],
```

Sample Google result:

```
"results": [  
    {
        "geometry" : {  
            "location" : {  
                "lat" : -33.400320,
                "lng" : -70.5763180
            }
        },
        "icon" : "http://maps.gstatic.com/...png",
        "id" : "86aa312400ac897443dc52ac1fc185e85",
        "name" : "Parque Arauco",
        "rating" : 4.30,
        "types" : [ "establishment" ],
        "vicinity" : "Presidente Kennedy 5413, Las Condes"
    }
],
```
Context analysis

• Some networks, like Foursquare, provide a hierarchy of venues

• This hierarchy greatly simplifies the effort needed to categorize places

  • Users pro-actively update the locations with new data

  • A hierarchy means that by simply analyzing the upper levels we can get a general idea of the location

Sample of Foursquare’s hierarchy of venues
## Results

- **Context analysis for week day (Foursquare)**

<table>
<thead>
<tr>
<th>Time</th>
<th>Clusters</th>
<th>Contexts</th>
</tr>
</thead>
<tbody>
<tr>
<td>6am-8am</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>8am-10am</td>
<td>1</td>
<td>Residence:6 Travel:1</td>
</tr>
<tr>
<td>10am-12p</td>
<td>1</td>
<td>Professional:4 Shops:3</td>
</tr>
<tr>
<td>12pm-2pm</td>
<td>3</td>
<td>Professional:6, Shops:4, Travel:3, Residence:3, Food:2</td>
</tr>
<tr>
<td>2pm-4pm</td>
<td>2</td>
<td>Shops:8, Food:7, Professional:5</td>
</tr>
<tr>
<td>4pm-6pm</td>
<td>1</td>
<td>Professional:6 Shops:4</td>
</tr>
<tr>
<td>6pm-8pm</td>
<td>1</td>
<td>Professional:4 Shops:3</td>
</tr>
<tr>
<td>8pm-10pm</td>
<td>3</td>
<td>Food:4, Shops:4, Residence:3 Outdoors:2</td>
</tr>
<tr>
<td>10pm-12a</td>
<td>1</td>
<td>Residence:2 Outdoors:2</td>
</tr>
<tr>
<td>12am-2am</td>
<td>0</td>
<td>-</td>
</tr>
</tbody>
</table>
Results

- Context analysis for week day (Foursquare)

<table>
<thead>
<tr>
<th>Time</th>
<th>Clusters</th>
<th>Contexts</th>
</tr>
</thead>
<tbody>
<tr>
<td>6am-8am</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>8am-10am</td>
<td>1</td>
<td>Residence:6, Travel:1</td>
</tr>
<tr>
<td>10am-12p</td>
<td>1</td>
<td>Professional:4, Shops:3</td>
</tr>
<tr>
<td>12pm-2pm</td>
<td>3</td>
<td>Professional:6, Shops:4, Travel:3, Residence:3, Food:2</td>
</tr>
<tr>
<td>2pm-4pm</td>
<td>2</td>
<td>Shops:8, Food:7, Professional:5</td>
</tr>
<tr>
<td>4pm-6pm</td>
<td>1</td>
<td>Professional:6, Shops:4</td>
</tr>
<tr>
<td>6pm-8pm</td>
<td>1</td>
<td>Professional:4, Shops:3</td>
</tr>
<tr>
<td>8pm-10pm</td>
<td>3</td>
<td>Food:4, Shops:4, Residence:3, Outdoors:2</td>
</tr>
<tr>
<td>10pm-12a</td>
<td>1</td>
<td>Residence:2, Outdoors:2</td>
</tr>
<tr>
<td>12am-2am</td>
<td>0</td>
<td>-</td>
</tr>
</tbody>
</table>
## Results

- **Context analysis for weekend (Foursquare)**

<table>
<thead>
<tr>
<th>Time</th>
<th>Clusters</th>
<th>Contexts</th>
</tr>
</thead>
<tbody>
<tr>
<td>10am-12pm</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>12pm-2pm</td>
<td>3</td>
<td>Parks &amp; Outdoors:3, Residence:9</td>
</tr>
<tr>
<td>2pm-4pm</td>
<td>1</td>
<td>Parks &amp; Outdoors:6, Shops:3</td>
</tr>
<tr>
<td>4pm-6pm</td>
<td>2</td>
<td>Arts &amp; Entertainment:2, Residence:20</td>
</tr>
<tr>
<td>6pm-8pm</td>
<td>1</td>
<td>Parks &amp; Outdoors:1, Shops:3</td>
</tr>
<tr>
<td>8pm-10pm</td>
<td>0</td>
<td>-</td>
</tr>
</tbody>
</table>
Results

• Foursquare is highly categorized, with a very well structured tree of venues

• Facebook provides similar capabilities, however the information input by users is less accurate and categories are not as rich

• Twitter is also an interesting approach, however the information is uncategorized so we’d need to use other techniques to analyze the text

• Google Places suffers from the same problem as Twitter
Conclusion

- This work can readily be fed in existing recommendation systems
- Easily integrated in SDKs for mobile devices
  - No need to bother the user for information
  - Can be used in many situations where we have no total control over the mobile applications
- When used in advertisement platforms, it can provide advertisers with precise ways to target its campaigns
- Some results can already be achieved
Future work

• Integrate this work into current recommendation systems
• Further evaluate/validate the clusters
  • Evaluate the minimum amount of data versus clustering quality
• Further search for other ways to get location metadata
• Analyze time divisions in order to be more adaptive to the user behavior
Questions

Thank you for your attention!