Smart Card in Public Transportation: Designing a Analysis System at the Human Scale

E. Tonnelier, N. Baskiotis, V. Guigue and P. Gallinari

LIP6 - UPMC - Sorbonne Universités

November 3rd, 2016
IEEE 19th International Conference on Intelligent Transportation Systems
Data sources
- City / Smart City: cellphones, GPS, smart street furnitures
  ...
- Explosion of available data, rich literature over the last decade

Development policies
- Trip prediction
- Users representations

[Black et al., 2002, Golias, 2002]
[Ceapa et al., 2012, Louail et al., 2014]
[Song et al., 2010, Foell et al., 2013]
[Poussevin et al., 2014]
**URBAN MOBILITY – User-centered study**

- Temporal patterns, habits
  - at the individual scale
  - for a standard week day

Illustrations from CityVille
Contributions and challenges

- Logs = entries of **10k users** during **13 weeks**
CONTRIBUTIONS AND CHALLENGES

- Logs = entries of 10k users during 13 weeks

- Characterize noisy users
  - Aggregation / Clustering
  - Habits modeling of week days:

⇒ Variance overestimate
CONTRIBUTIONS AND CHALLENGES

- Logs = entries of **10k users** during **13 weeks**

- Characterize **noisy users**
  - Aggregation / Clustering
  - Habits modeling of week days:

  8am Week days  7pm Week days  11pm Thursdays

⇒ Variance overestimate

New hypothesis:

**Habits are shared...**

**but with individual schedule**

Vincent Guigue
Matrix factorization

User decomposition = Habit extraction

Data matrix (logs) = Code matrix × Dictionary

- \( i \)-th row

Goal: optimizing both code & dictionary

- Variations SVD
  - Non-negative matrix factorization
  - Sparseness

[Golub and Van Loan, 1996]
[Lee and Seung, 2000]
[Hoyer, 2002]
NMF ON AN EXAMPLE

5 random users

- 24h = 96 intervals of 15min

Dictionary: (most used atoms)

Code:
NMF ON AN EXAMPLE

5 random users
- 24h = 96 intervals of 15min

Dictionary: (most used atoms)

Code:
NMF ON AN EXAMPLE

5 random users
- 24h = 96 intervals of 15min

Focusing on user #1

Dictionary: (most used atoms)

Code:

Vincent Guigue

Smart Card Analysis at the Human Scale
5 random users

- 24h = 96 intervals of 15min

Focusing on user #1
NMF on an Example

5 random users

- 24h = 96 intervals of 15min

Focusing on user #1

Dictionary: (most used atoms)

Code:
NMF on an example

5 random users
○ 24h = 96 intervals of 15min
Focusing on user #1

Dictionary: (most used atoms)

Code:
NMF ON AN EXAMPLE

5 random users
- 24h = 96 intervals of 15min
Focusing on user #1

Dictionary: (most used atoms)

Code:
NMF on an example

5 random users
  ○ 24h = 96 intervals of 15min
Focusing on user #1

Dictionary: (most used atoms)

Code:
NMF on an Example

5 random users
- 24h = 96 intervals of 15min
Focusing on user #1

Dictionary: (most used atoms)

Code:

Vincent Guigue
**NMF: ANALYSIS**

- **Number of atoms in the dictionary (rank constraint)**
  - 5 atoms
  - 5 best atoms among 10
  - 5 best atoms among 40

- More atoms = **finer reconstruction**...
  - reconstruction of the noise...
  - + meaningless atoms

- Less atoms = **no longer local event modeling** (variance overestimate)
  - Parameters are wasted **modeling translated events**
  - **Evaluation ?**

Vincent Guigue

Smart Card Analysis at the Human Scale
NMF: ANALYSIS

- Number of atoms in the dictionary (rank constraint)
  - More atoms = finer reconstruction...
  - + reconstruction of the noise...
  - + meaningless atoms
  - Less atoms = no longer local event modeling (variance overestimate)

Reconstruction (poor dictionary)  Reconstruction (standard dictionary)  Reconstruction (rich dictionary)

- Parameters are wasted modeling translated events
- Evaluation?
CONTRIBUTION: TS-NMF

Idea:

- Keeping the NMF framework
- Defining **compact atoms** which shapes are learned on all users (=NMF) which can be positioned for each user

\[
\text{Signal} \quad \uparrow \quad \text{Shift} \quad \downarrow \text{Signal}
\]

\[
\text{Atom}
\]
CONTRIBUTION: TS-NMF

Idea:

- Keeping the NMF framework
- Defining **compact atoms**
  which shapes are learned on all users (=NMF)
  which can be positioned for each user

\[
u = \sum_z \tau_{u,z}(w_{u,z}d_z) = w_{u,z}d_z(t + \phi_{u,z})\]

Data matrix (logs) \hspace{2cm} \phi \text{ matrix} \hspace{2cm} \text{Code matrix} \hspace{2cm} \text{Dictionary}

\[
\begin{array}{c|c|c|c}
\hline
\text{i-th row} & \text{\phi matrix} & \text{Code matrix} & \text{Dictionary} \\
\hline
\end{array}
\]
Algorithm 1: TS NMF learning algorithm

**Data:** $X \in \mathbb{R}_+^{U \times T}$, $Z$, $\text{max}_{\text{iter}}$, $\alpha_\Phi$

1. $\text{max}_{\text{iter}}$: to reach convergence
2. $\alpha_\Phi$: window of search in the atom shift procedure

**Result:** Optimized matrix $\Phi$, $D$ and $W$

2. $D, W, \Phi = \text{init}(X, Z)$

3. for $\text{it} \in 0...\text{max}_{\text{iter}}$ do
4.     for $u \in \text{range}(0, U)$ do
5.         $x_u = X[u, .]$
6.     atoms = descendingEntropy($D$)

   return atoms indexes sort in descending order

3. $D$ and $W$ randomly initialized, $\Phi$ regularly scattered along time band
**Algorithm 2:** TS NMF learning algorithm

Data:

\[ X \in \mathbb{R}^{U \times T} +, \ Z, \ \max_{\text{iter}}, \ \alpha, \ \Phi \]

\( \alpha, \ \Phi \): window of search in the atom shift procedure

Result:

Optimized matrix \( \Phi, D \) and \( W \)

\[ \text{init}(X, Z) \]

\( D \) and \( W \) randomly initialized, \( \Phi \) regularly scattered along time band

3. \( \text{for } \text{it } \in 0...\max_{\text{iter}} \text{ do} \)

4. \( \text{for } u \in \text{range}(0, U) \text{ do} \)

5. \( x_u = X[u, .] \)

6. \( \text{atoms } = \text{descendingEntropy}(D) \)

   return atoms indexes sort in descending order

7. \( \text{for } a \in \text{atoms do} \)

8. \( \Phi_{u,a} = \text{minimizeLocalCost}(x_u, D_a) \)

   finding optimal time-shift \( t \) in a window of size \( \alpha, \Phi \)

9. \( W_{u,a} = \text{update}_W(x_u, W_{u,.}, D, \Phi_{u,a}) \)

   Simple gradient descent

10. \( x_u = x_u - f(D_a, \Phi_{u,a})W_{u,a} \)

   Matching pursuit like update

11. \( D = \text{update}_D(W, D, \Phi) \)

12. \( D = \text{centerAtoms}(D) \)

Centering procedure to make atoms comparable
TS-NMF: OUTPUTS

Reconstructed users:

Dictionary:

Code:

$\phi$: 

Vincent Guigue
TS-NMF: OUTPUTS

Reconstruction process:
Atom selection ...

Dictionary:

Code:

φ:

Vincent Guigue
TS-NMF: OUTPUTS

Reconstruction process:
+ shift

Dictionary:

Code:

ϕ:
TS-NMF: OUTPUTS

Reconstruction process:

Dictionary:

Code:

φ:
TS-NMF: OUTPUTS

Reconstruction process:

Dictionary:

Code:

\( \phi \):
TS-NMF: OUTPUTS

Reconstruction process:

Dictionary:

Code:

Vincent Guigue
TS-NMF: OUTPUTS

Reconstruction process:

Dictionary:

Code:

φ:
Evaluation

- Impossible on training data:
  - more degrees of freedom ⇒ less reconstruction error

- 9 weeks for learning, 4 weeks for testing
  - Random initialization + non convex optimization ⇒ averaging performance on 5 runs
  - Reconstruction of unseen data (=predictive skills)
  - Necessary but not sufficient

⇒ meaning/interpretation of the model is required
  - link with the number of parameters
**Baselines & Dimensionality**

- **10k users**
- **480 time intervals (3 minutes)**

**# parameters**

- **General model** = 1-mean model  
  \[ \Rightarrow 0 \]  

- **k-means**, \( k = 16 \):  
  - 16 prototypes \( \in \mathbb{R}^{480} \) + 10k assignments  
  \[ \Rightarrow 17,680 \]  

- **NMF**, \( Z = 16 \):  
  - 16 prototypes \( \in \mathbb{R}^{480} \) + 10k \times 16 weights  
  \[ \Rightarrow 167,680 \]  

- **GMM** = 3 Gaussian atoms \( (\mu, \sigma_1), (\mu, \sigma_2), (\mu, \sigma_3) \) centered on each of the 480 time interval & weighted  
  \[ \Rightarrow 14,400,003 \]  

- **TSNMF** = 16 atoms of size 60, weighted & shifted  
  \[ \Rightarrow 320,960 \]
## COMPARISON

### 2 metrics

**MSE**  Mean Squared Error (between real & estimated pdf)

**ML**  Likelihood of the logs according to the model

<table>
<thead>
<tr>
<th>Model</th>
<th># param.</th>
<th>MSE -train- (mean (std))</th>
<th>MSE -test- (mean (std))</th>
</tr>
</thead>
<tbody>
<tr>
<td>General model</td>
<td>0</td>
<td>0.033 (0)</td>
<td>0.040 (0)</td>
</tr>
<tr>
<td>KMeans (16 clusters)</td>
<td>17,680</td>
<td>0.027 (6.3e-6)</td>
<td>0.038 (1.5e-5)</td>
</tr>
<tr>
<td>NMF</td>
<td>167,680</td>
<td>0.024 (7.7e-5)</td>
<td>0.036 (6.7e-5)</td>
</tr>
<tr>
<td>GMM</td>
<td>14,400,003</td>
<td>0.023 (0)</td>
<td>0.050 (0)</td>
</tr>
<tr>
<td>TS-NMF</td>
<td>320,960</td>
<td>0.016 (5.8e-4)</td>
<td>0.042 (8.9e-4)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th># param.</th>
<th>ML -train- (mean (std))</th>
<th>ML -test- (mean (std))</th>
</tr>
</thead>
<tbody>
<tr>
<td>General</td>
<td>0</td>
<td>0.0038 (0)</td>
<td>0.0036 (0)</td>
</tr>
<tr>
<td>KMeans (16 clusters)</td>
<td>17,680</td>
<td>0.010 (6.3e-6)</td>
<td>0.008 (5.7e-6)</td>
</tr>
<tr>
<td>NMF</td>
<td>167,680</td>
<td>0.013 (8.3e-5)</td>
<td>0.009 (3.8e-5)</td>
</tr>
<tr>
<td>GMM</td>
<td>14,400,003</td>
<td>0.027 (0)</td>
<td>0.018 (0)</td>
</tr>
<tr>
<td>TS-NMF</td>
<td>320,960</td>
<td>0.026 (9.3e-4)</td>
<td>0.016 (4.8e-4)</td>
</tr>
</tbody>
</table>
Qualitative analysis

Shapes of the atoms

- +/- variance
- Different shapes

Atoms Positions (distrib. over the population)

- Most atoms correspond to a defined period of the day

Vincent Guigue

Smart Card Analysis at the Human Scale
Introduction

According to the shapes of the atoms

[Poussevin et al., 2014]
According to the **time positions of the atoms** (morning = departure to work)
Characterizing both habits and their schedules

- ... at the individual scale
- ⇒ valuable information on users
- Costly, but scalable for a transportation system

Perspective

- Work on the cost function...
- ... to discover less compact atoms (more meaningful)
BIBLIOGRAPHY


