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 3^{rd}, 2016

IEEE 19th International Conference on Intelligent Transportation Systems

Urban mobility – many issues & sensors

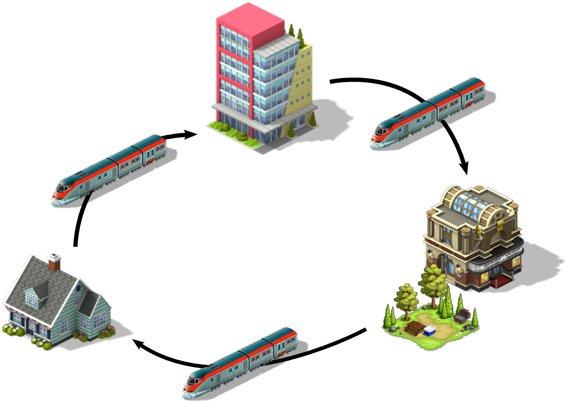
- Data sources
  - City / Smart City: cellphones, GPS, smart street furnitures ...
  - Explosion of available data, rich literature over the last decade

- Development policies [Black et al., 2002, Golias, 2002]
- Global view on traffic [Ceapa et al., 2012, Louail et al., 2014]
- Regularity of users
  - Trip prediction [Song et al., 2010, Foell et al., 2013]
  - Users representations [Poussevin et al., 2014]

Urban mobility – *User-centered study*

- Temporal patterns, habits

⇒ at the individual scale
⇒ for a standard week day



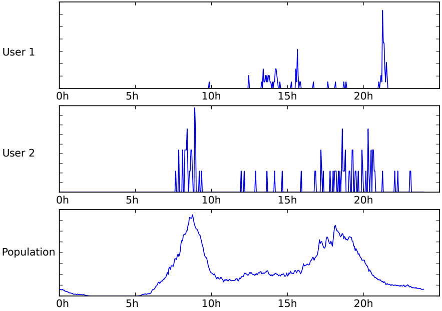
Illustrations from CityVille

Contributions and challenges

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

CONTRIBUTIONS AND CHALLENGES

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



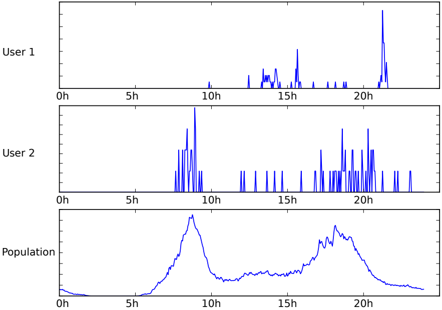
- Characterize **noisy** users
  - Aggregation / Clustering
  - Habits modeling of week days:

8am Week days | 7pm Week days | 11pm Thursdays

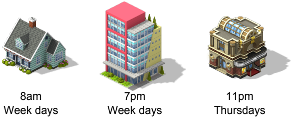
⇒ Variance overestimate

Contributions and challenges

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



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



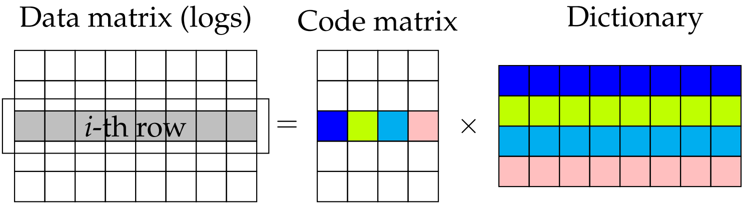
$\Rightarrow$ Variance overestimate

New hypothesis:

Habits are shared...
but with individual schedule

Matrix factorization

User decomposition = Habit extraction



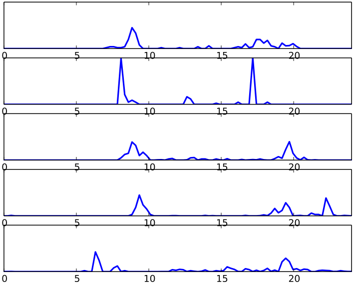
Goal: optimizing both **code & dictionary**

- Variations SVD [Golub and Van Loan, 1996]
  - Non-negative matrix factorization [Lee and Seung, 2000]
  - Sparseness [Hoyer, 2002]

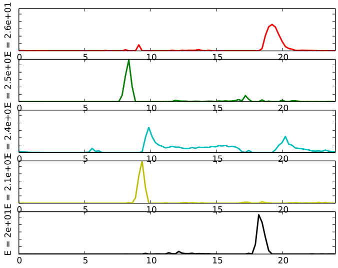
NMF ON AN EXAMPLE

5 random users

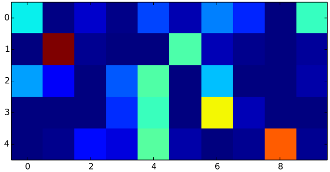
- 24h = 96 intervals of 15min



Dictionary: (most used atoms)



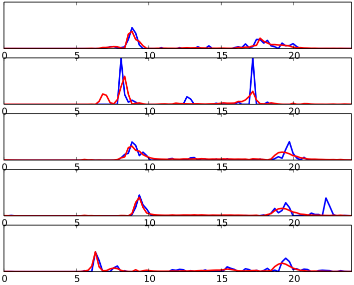
Code:



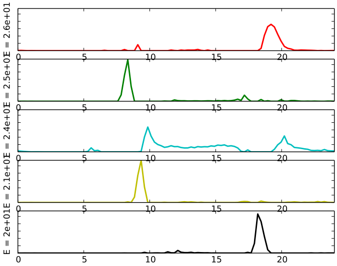
NMF on an example

5 random users

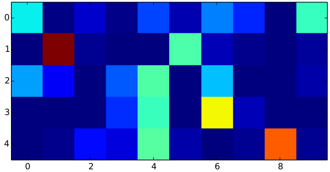
- 24h = 96 intervals of 15min



Dictionary: (most used atoms)



Code:

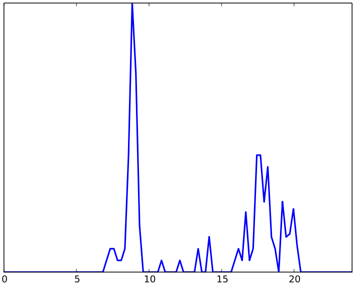


NMF on an example

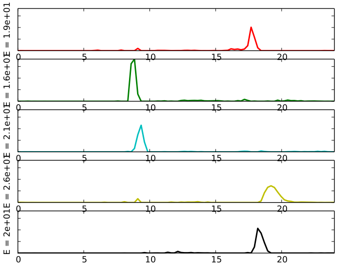
5 random users

- 24h = 96 intervals of 15min

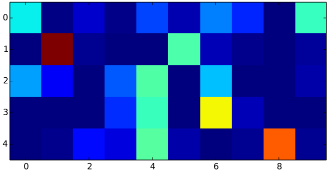
Focusing on user #1



Dictionary: (most used atoms)



Code:

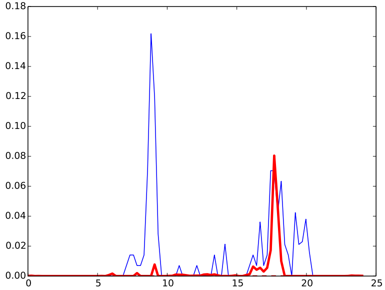


NMF on an example

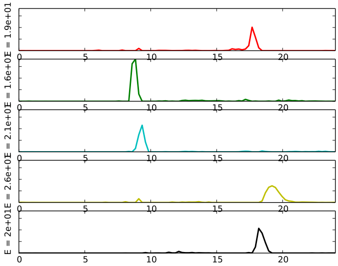
5 random users

- 24h = 96 intervals of 15min

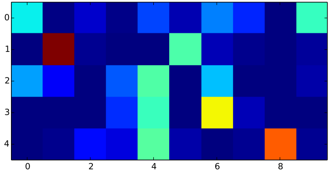
Focusing on user #1



Dictionary: (most used atoms)



Code:

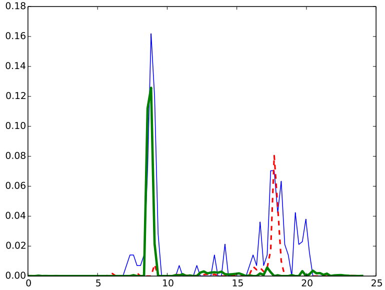


NMF on an example

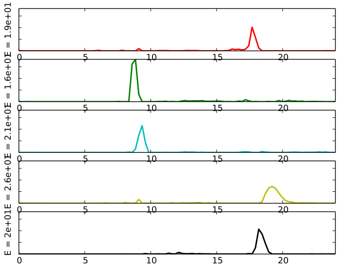
5 random users

- 24h = 96 intervals of 15min

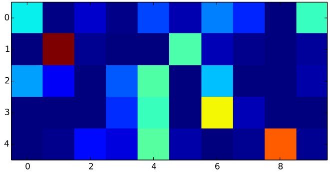
Focusing on user #1



Dictionary: (most used atoms)



Code:

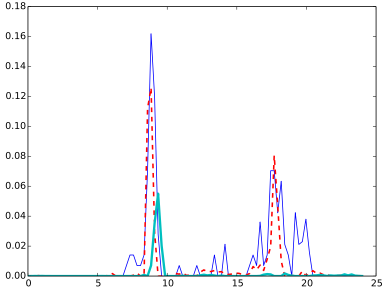


NMF on an example

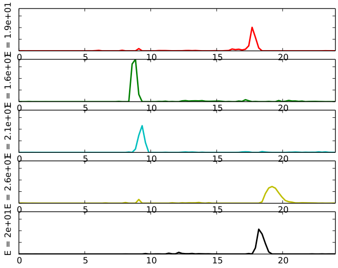
5 random users

- 24h = 96 intervals of 15min

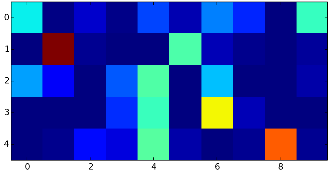
Focusing on user #1



Dictionary: (most used atoms)



Code:

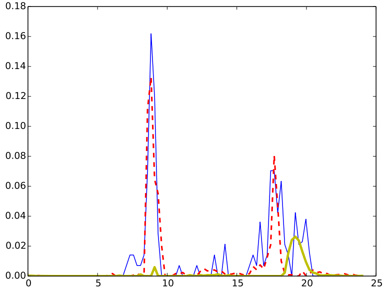


NMF on an example

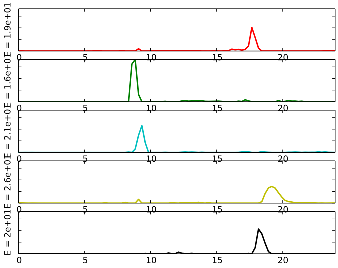
5 random users

- 24h = 96 intervals of 15min

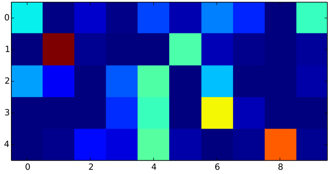
Focusing on user #1



Dictionary: (most used atoms)



Code:

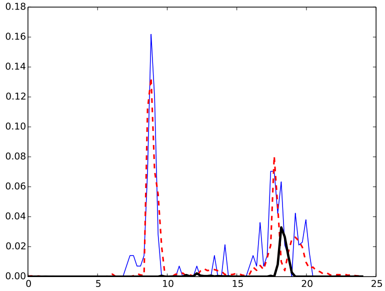


NMF on an example

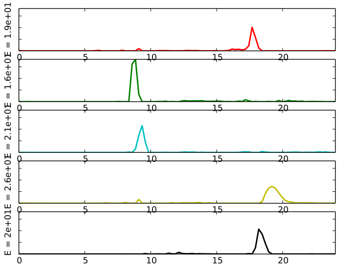
5 random users

- 24h = 96 intervals of 15min

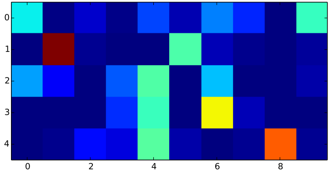
Focusing on user #1



Dictionary: (most used atoms)



Code:

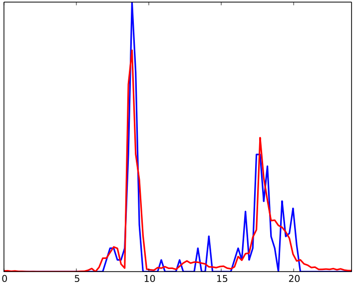


NMF on an example

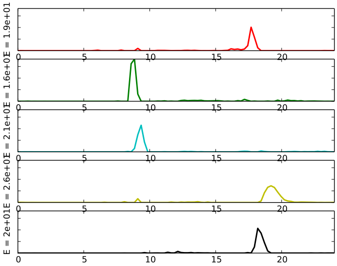
5 random users

- 24h = 96 intervals of 15min

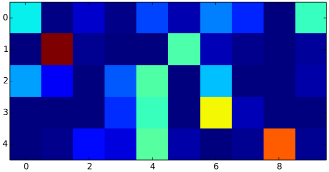
Focusing on user #1



Dictionary: (most used atoms)



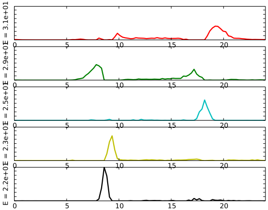
Code:



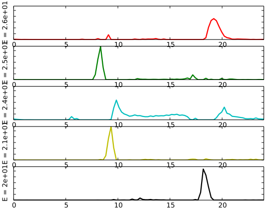
NMF: analysis

- Number of atoms in the dictionary (rank constraint)

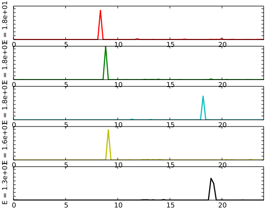
5 atoms



5 best atoms among 10



5 best atoms among 40

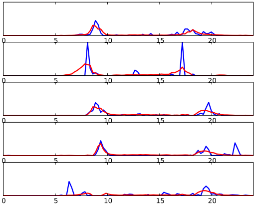


  - More atoms = finer reconstruction...
    +reconstruction of the noise...
    + meaningless atoms
  - Less atoms = no longer local event modeling (variance overestimate)
- Parameters are wasted modeling translated events
- Evaluation ?

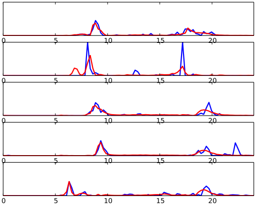
NMF: analysis

- 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)

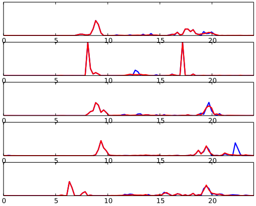
Reconstruction (poor dictionary)



Reconstruction (standard dictionary)



Reconstruction (rich dictionary)

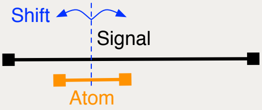


- Parameters are wasted modeling translated events
- Evaluation ?

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**

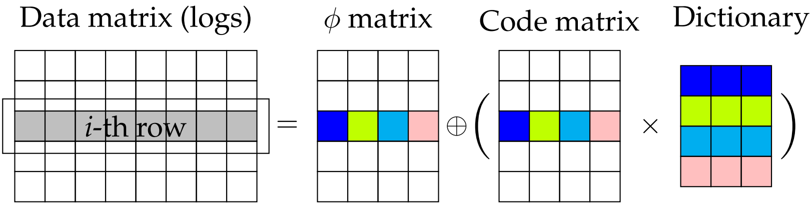


Contribution: TS-NMF

Idea:

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

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



Algorithm

Algorithm 1: TS NMF learning algorithm

Data: $X \in \mathbb{R}_+^{U \times T}$, Z, max_{iter}, α_Φ

max_{iter}: to reach convergence
α_Φ: window of search in the atom shift procedure

Result: Optimized matrix Φ, D and W

$D, W, \Phi = init(X, Z)$

D and W randomly initialized, Φ regularly scattered along time band

for $it \in 0...max_{iter}$ **do**
    for $u \in range(0, U)$ **do**
        $\mathbf{x}_u = X[u, .]$
        $atoms = descendingEntropy(D)$

return atoms indexes sort in descending order

Algorithm

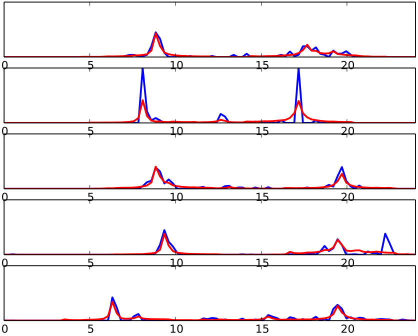
$D, W, \Phi = init(X, Z)$

D and W randomly initialized, Φ regularly scattered along time band

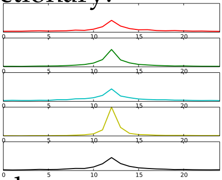
```
for it ∈ 0...max_iter do
    for u ∈ range(0, U) do
        x_u = X[u, .]
        atoms = descendingEntropy(D)
            [return atoms indexes sort in descending order]
        for a ∈ atoms do
            Φ_{u,a} = minimizeLocalCost_t(x_u, D_a)
                [finding optimal time-shift t in a window of size α_Φ]
            W_{u,a} = update_W(x_u, W_{u,.}, D, Φ_{u,a})
                [Simple gradient descent]
            x_u = x_u − f(D_a, Φ_{u,a}) W_{u,a}
                [Matching pursuit like update]
    D = update_D(W, D, Φ)
    D = centerAtoms(D)
        [Centering procedure to make atoms comparable]
```

TS-NMF: outputs

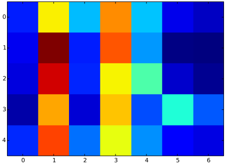
Reconstructed users:



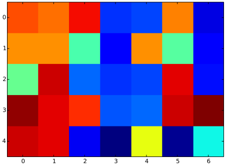
Dictionary:



Code:



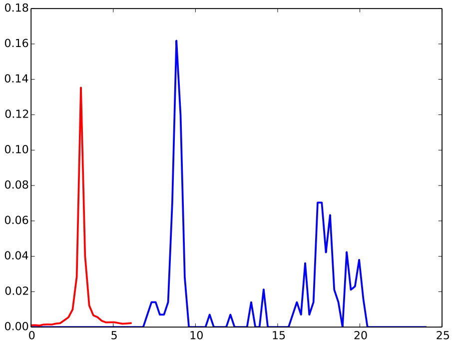
ϕ:



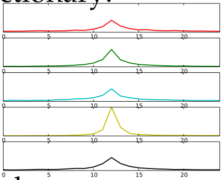
TS-NMF: OUTPUTS

Reconstruction process:

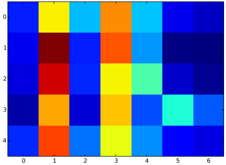
Atom selection ...



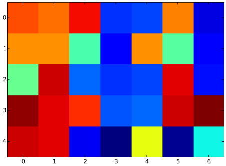
Dictionary:



Code:



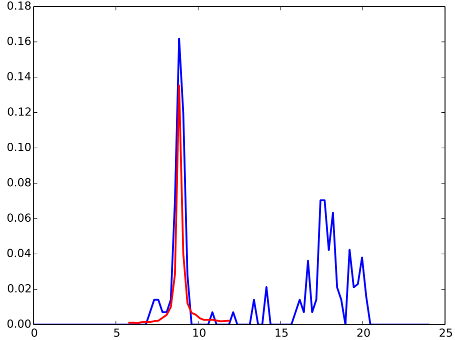
ϕ:



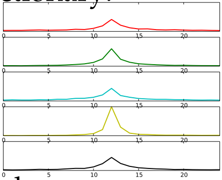
TS-NMF: outputs

Reconstruction process:

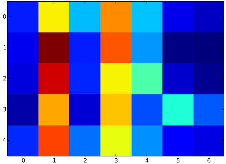
+ shift



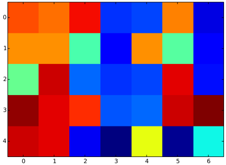
Dictionary:



Code:

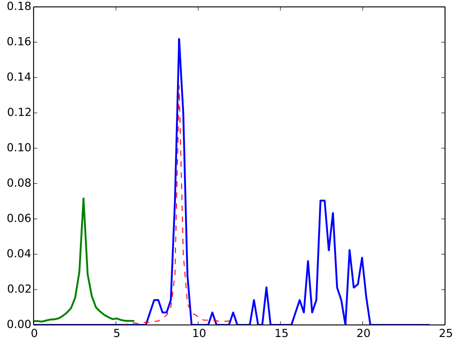


ϕ:

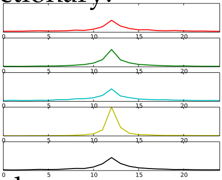


TS-NMF: outputs

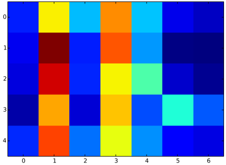
Reconstruction process:



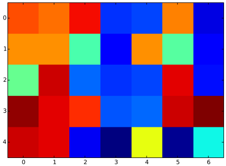
Dictionary:



Code:

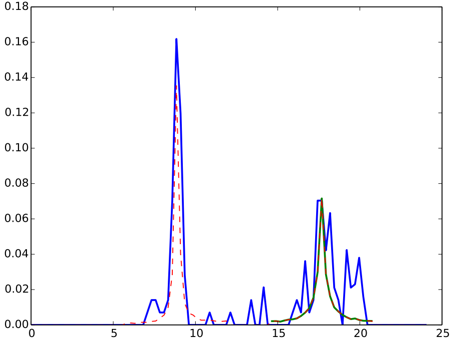


ϕ:

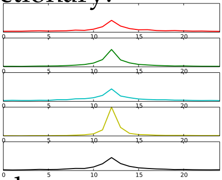


TS-NMF: outputs

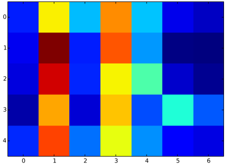
Reconstruction process:



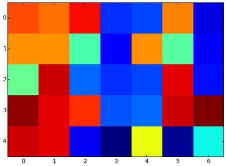
Dictionary:



Code:

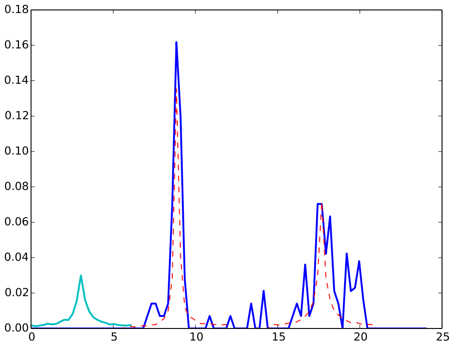


ϕ:

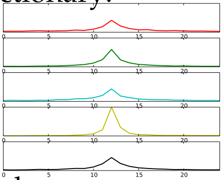


TS-NMF: outputs

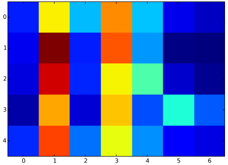
Reconstruction process:



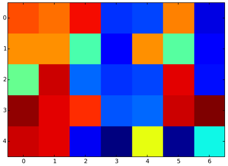
Dictionary:



Code:

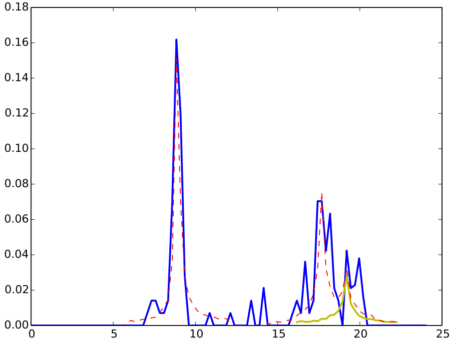


ϕ:

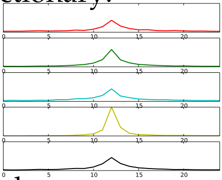


TS-NMF: outputs

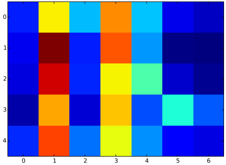
Reconstruction process:



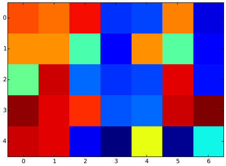
Dictionary:



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 (μ, σ_1), (μ, σ_2), (μ, σ_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

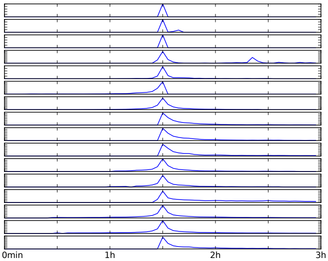
2 metrics

- MSE Mean Squared Error (between real & estimated pdf)
- ML Likelihood of the logs according to the model

Model	# param.	MSE -train- (mean (std))	MSE -test- (mean (std))
General model	0	0.033 (0)	0.040 (0)
KMeans (16 clusters)	17,680	0.027 (6.3e-6)	0.038 (1.5e-5)
NMF	167,680	0.024 (7.7e-5)	**0.036 (6.7e-5)**
GMM	14,400,003	0.023 (0)	0.050 (0)
TS-NMF	320,960	**0.016 (5.8e-4)**	0.042 (8.9e-4)
Model	**# param.**	**ML -train- (mean (std))**	**ML -test- (mean (std))**
General	0	0.0038 (0)	0.0036(0)
KMeans (16 clusters)	17,680	0.010 (6.3e-6)	0.008 (5.7e-6)
NMF	167,680	0.013 (8.3e-5)	0.009 (3.8e-5)
GMM	14,400,003	**0.027 (0)**	**0.018 (0)**
TS-NMF	320,960	**0.026 (9.3e-4)**	0.016 (4.8e-4)

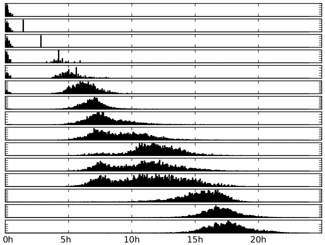
Qualitative analysis

Shapes of the atoms



- +/- variance
- Different shapes

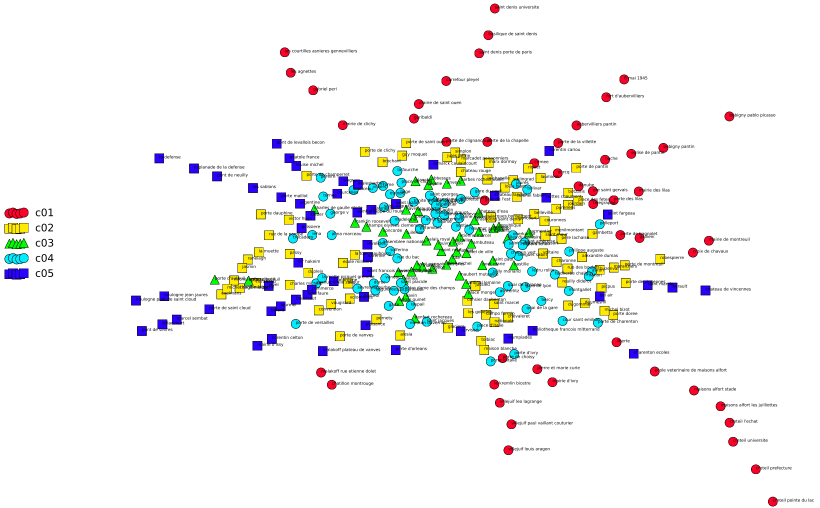
Atoms Positions (distrib. over the population)



- Most atoms correspond to a defined period of the day

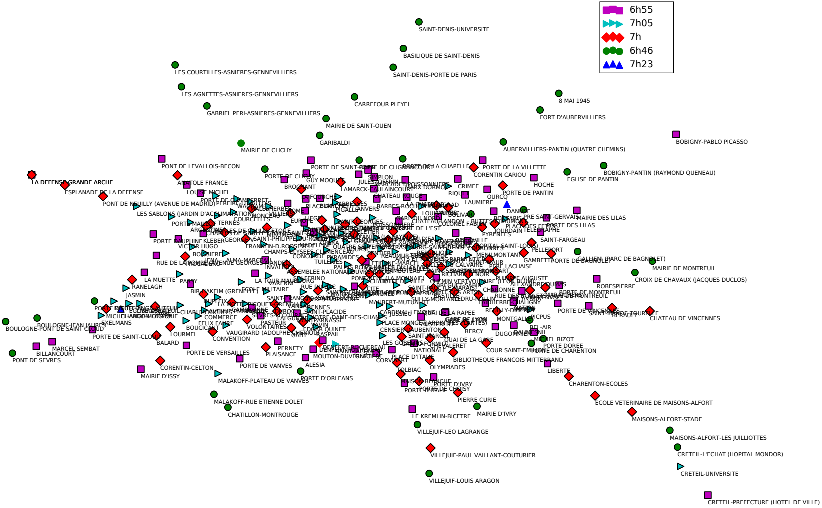
User mapping on Paris map

According to the **shapes of the atoms** [Poussevin et al., 2014]



User mapping on Paris map

According to the **time positions of the atoms** (morning = departure to work)



Conclusion

Characterizing both **habits** and **their schedules**

- ... at the individual scale
- $\Rightarrow$ 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

Black, J. A., Paez, A., and Suthanaya, P. A. (2002).
Sustainable urban transportation: performance indicators and some analytical approaches.
Journal of urban planning and development, 128(4):184–209.

Ceapa, I., Smith, C., and Capra, L. (2012).
Avoiding the crowds: understanding tube station congestion patterns from trip data.
In *ACM SIGKDD 2012*, pages 134–141. ACM.

Foell, S., Kortuem, G., Rawassizadeh, R., Phithakkitnukoon, S., Veloso, M., and Bento, C. (2013).
Mining temporal patterns of transport behaviour for predicting future transport usage.
In *UbiComp 13*, pages 1239–1248. ACM.

Golias, J. C. (2002).
Analysis of traffic corridor impacts from the introduction of the new athens metro system.
J. Transp. Geogr., 10(2):91–97.

Golub, G. H. and Van Loan, C. F. (1996).
Matrix Computations (3rd Ed.).
Johns Hopkins University Press.

Hoyer, P. O. (2002).
Non-negative sparse coding.
In *W. on Neural Networks for Signal Processing*, pages 557–565. IEEE.

Lee, D. D. and Seung, H. S. (2000).
Algorithms for non-negative matrix factorization.
In *NIPS*, pages 556–562.

Louail, T., Lenormand, M., Cantú, O. G., Picornell, M., Herranz, R., Frias-Martinez, E., Ramasco, J. J., and Barthelemy, M. (2014).
From mobile phone data to the spatial structure of cities.
Nature.

Poussevin, M., Tonnelier, E., Baskiotis, N., Guigue, V., and Gallinari, P. (2014).
Mining ticketing logs for usage characterization with nonnegative matrix factorization.
In *International Workshop on Modeling Social Media*, pages 147–164. Springer.

Song, C., Qu, Z., Blumm, N., and Barabási, A.-L. (2010).
Limits of predictability in human mobility.
Science, 327(5968):1018–1021.