Non-negative Matrix Factorization as a pre-processing tool for travelers temporal profiles clustering

Léna CAREL[1,2], Pierre ALQUIER[1]

[1]CREST, ENSAE, Université Paris Saclay
3 avenue Pierre Larousse, 92245 Malakoff CEDEX - France
[2]TRANSDEV Group
32 boulevard Gallieni, 92130 Issy-les-Moulineaux - France

Introduction

In recent years, more and more travel networks use smart card automated fare collection systems. The main purpose of these systems is to collect the fare revenues. However, they also allow to collect a large amount of information on onboard transactions that can be used for various objectives: to analyse nowadays cities through global urban problematics [1], to study the variability of the travels from a spatial and temporal perspective [2], to help transit planners [3] or to analyze the travel habits of smart card holders as in [4].

More precisely, in [4], the authors propose a mixture of k multinomial distributions as a model for the travelers temporal profiles. They then estimate the parameters of the models, and assign the travelers to clusters, using the Expectation-Maximization (EM) algorithm.

Methods comparaison

Use of EM as a clustering tool [4]

In [4] the authors assume that there is a given number of clusters of users, and in each cluster, the profiles are independently generated from a common multinomial distribution. Thus, the distribution on all profiles is a mixture of multinomial distributions.

$$\mathbf{u}|z \sim \mathcal{M}(D, \beta_z)$$

With $\mathbf{u}$ the temporal profile of an user, z the cluster $\mathbf{u}$ comes from, D the number of displacements recorded in $\mathbf{u}$ and β the cluster profiles.

Traveler profiles → EM → Clustering profiles

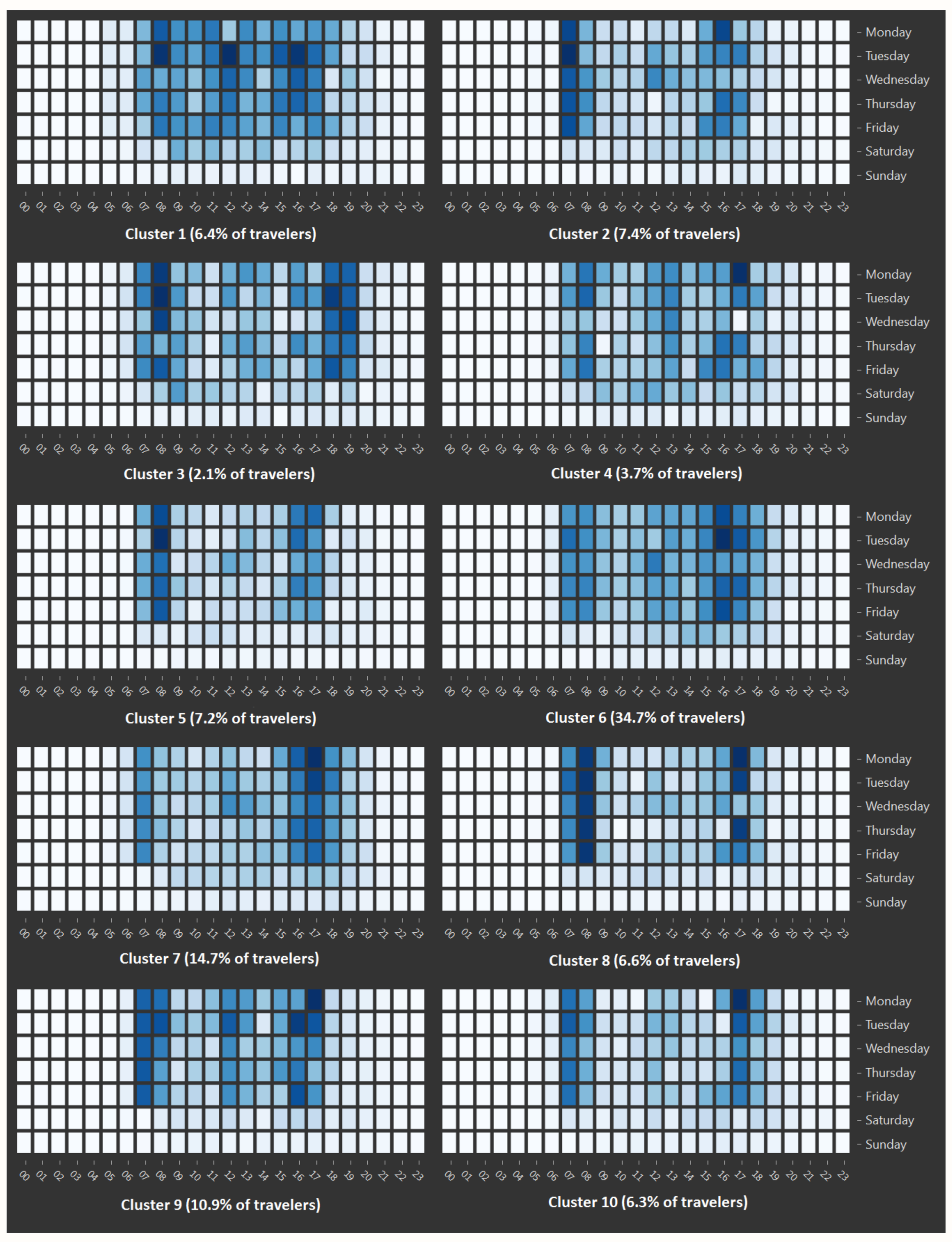

Figure 1: Clusters obtained by EM-algorithm

Use of NMF as a pre-processing tool

Consider the matrix V that contains the temporal profiles V_i of all users as rows.

$$V \approx WH$$

With $V \in \mathbb{R}^{n\times m}$, $W \in \mathbb{R}^{n\times K}$ and $H \in \mathbb{R}^{K\times m}$. H is a dictionnary of profiles and W contains the travelers profiles in reducted space, as a positive linear combination of these "words".

Traveler profiles → NMF → Traveler profiles in reducted space → k-means → Clustering profiles

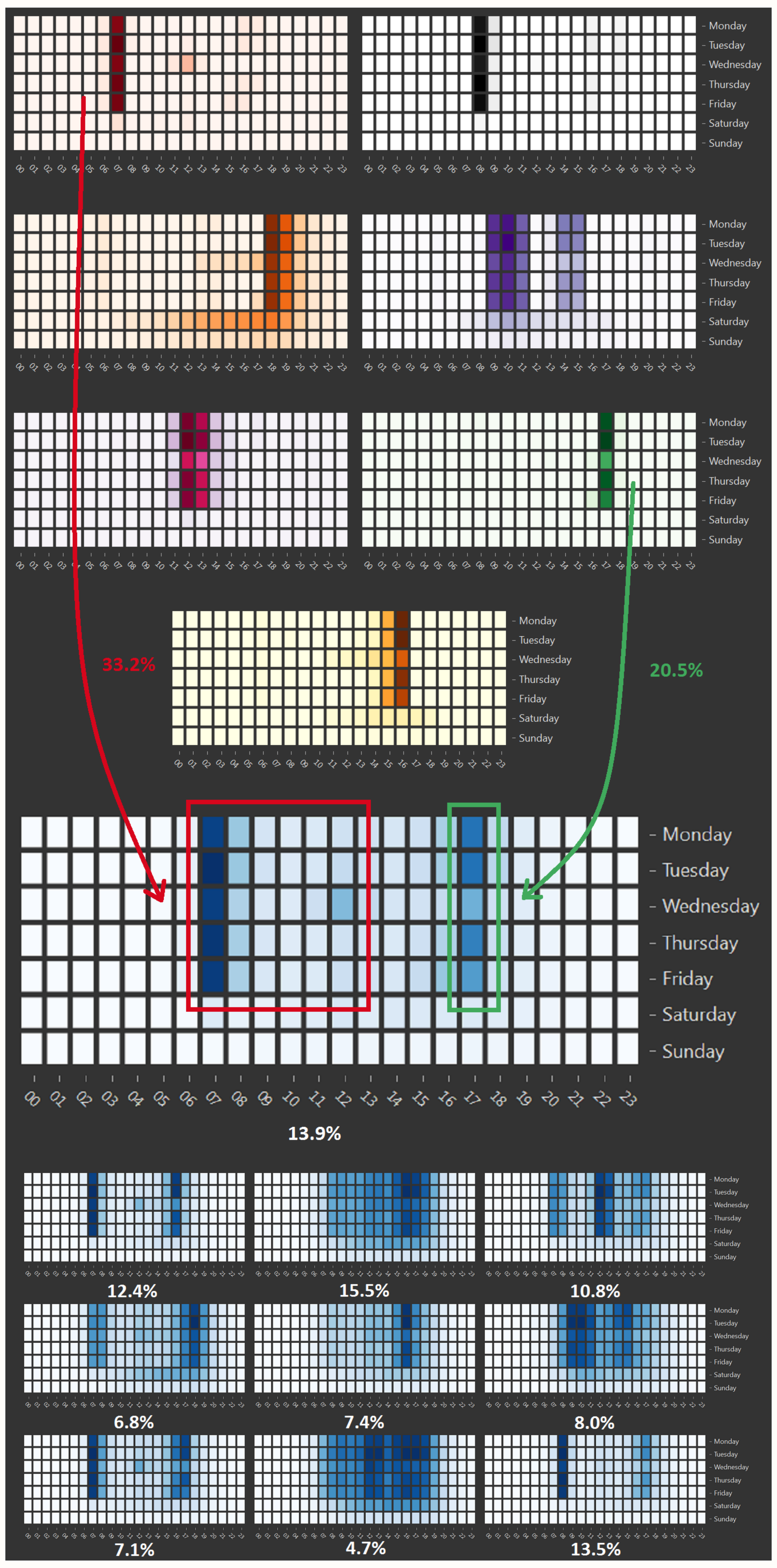

Figure 2: Up : "words" of the dictionnary obtained by NMF; Middle : Decompozition in "words" of one of the clusters obtained by k-means; Down : The other clusters obtained by k-means on the reducted space

Table 1: Repartition of individuals between the clusters obtained by EM-algorithm and the clusters obtained by k-means on the reducted space.

		NMF + k-means									
		1	2	3	4	5	6	7	8	9	10
EM-algorithm	1	4%	6%	24%	13%	3%	7%	22%	3%	5%	12%
	2	31%	22%	11%	5%	4%	5%	8%	8%	2%	5%
	3	4%	12%	36%	5%	17%	2%	6%	1%	1%	14%
	4	6%	11%	28%	14%	6%	5%	8%	3%	4%	16%
	5	14%	9%	12%	6%	2%	4%	6%	6%	2%	40%
	6	15%	8%	13%	12%	6%	11%	9%	8%	7%	11%
	7	9%	14%	20%	6%	15%	10%	5%	10%	7%	4%
	8	4%	26%	11%	5%	4%	2%	5%	7%	1%	35%
	9	14%	16%	12%	22%	2%	5%	7%	8%	2%	11%
	10	3%	33%	15%	13%	12%	2%	2%	7%	1%	12%

By observing the Table 1, we note that the clusters obtained with our method of NMF as a pre-processing tool do not correspond to clusters obtained by EM-algorithm. Indeed, the clusters obtained by EM are distributed between all the clusters obtained by NMF with k-means.

Perspectives

The results shown here are preliminary results. Future work will include discussion on the choice of the size of the dictionary K and the number of clusters k. A BIC criterion can be considered as in [10]. More importantly, we use a two-step procedure: the dictionary W is chosen in order to minimize a square error (unrelated to clustering), we then perform clustering on the matrix H. An important objective would be to define a one-step procedure that would directly estimate a dictionary W optimizing a criterion related to the clustering objective.

References

[1] Y. Zheng, L.Capra, O. Wolfson, H. Yang, Urban computing: concepts, methodologies, and applications, *ACM TIST*

[2] C. Morency, M. Trepanier, B. Agard, Measuring transit use variability with smart-card data, *Transport Policy*

[3] M.-P. Pelletier, M. Trépanier, C. Morency, Smart card data in public transit planning: a review, *CIRRELT*

[4] M. K. El Mahrsi, E. Côme, J. Baro, L. Oukhellou, Understanding passenger patterns in public transit through smart card and socioeconomic data, *UrbComp'*

[5] D. D. Lee, H. S. Seung, Learning the parts of objects by non-negative matrix factorization, *Nature*

[6] W. Xu, X. Liu, Y. Gong, Document clustering based on non-negative matrix factorization, *ACM SIGIR*

[7] F. Shahnaz, M. W. Berry, V. P. Pauca, R. J. Plemmons, Document clustering using nonnegative matrix factorization, *Information Processing & Management*

[8] C. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), *Springer, New York*

[9] J. Kim, H. Park, Sparse nonnegative matrix factorization for clustering

[10] C. Bouveyron, E. Côme, J. Jacques and others, The discriminative functional mixture model for a comparative analysis of bike sharing systems, *The Annals of Applied Statistics*

ESANN, 26 - 28 April 2017, Bruges, Belgium