ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017, i6doc.com publ., ISBN 978-287587039-1.
Available from http://www.i6doc.com/en/.

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

Léna CAREL[1,2] and 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

**Abstract**. We propose to use non-negative matrix factorization (NMF) to build a dictionary of travelers temporal profiles. Clustering based on decomposition in this dictionary rather than on the full profiles (as in previous works) lead to more interpretable clusters.

## 1 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. Although the results they obtained allow to identify relevant users profiles, some clusters are not easily interpretable. To overcome this issue, we propose to reduce the dimension of the profiles by non-negative matrix factorization (NMF). NMF was introduced by [5] and leads in many high-dimensional applications to the definition of a sparse and easily intepretable dictionary: [5] provided examples in image analysis, [6, 7] in text document clustering, among others. Here, NMF provides a dictionary of temporal profiles, and a projection of each profile in the span of this dictionary. Any clustering method can then be used in this smaller space (we use $k$-means in this paper). This leads to easily interpretable clusters.

## 2 The data

We study here validations made during the month of September 2014 on the network of Rouen metropolis. Ticketing data are the information obtained at each transaction made by a smart card on a validator system. For privacy reasons it is not possible to connect each validation to the user that made it.

ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017, i6doc.com publ., ISBN 978-287587039-1. Available from http://www.i6doc.com/en/.

The feature that permits us to realize our study and create temporal profiles is a card number which is encrypted, and re-initialized every three months. It is thus impossible to follow the long-term behaviour of a user. This is the reason why we focus on a one month period in a first time. This period (September) have been chosen because it has no vacation or bank holidays. We use the same method as [4] in order to keep only the regular smart card holders: to be a regular card holder, the traveler must have used his card for at least ten days during the studied period and must have made his first boarding after 4am each day at the same station 50% of the time. The data are then agregated so that for each traveler, for each day of the week (Monday to Sunday) and each hour (00 to 23) we have the mean of the number of validation during the studied period.

## 3 Results obtained by EM

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. They used an EM-algorithm to estimate the parameters of each multinomial and the probabilities for any users to belong to each cluster (we refer the reader to [4] for more details on this model, and to Chapter 9 in [8] for an introduction to the EM algorithm). We use the same methodology on our dataset. The results for 10 clusters are shown in Figure 1 (we tested other numbers of cluster but do not show the results here for the sake of shortness). Two comments are in order: first, while some profiles are easily interpretable, it is not so easy to give an interpretation to Cluster 1 when compared to Cluster 6 and Cluster 9. Moreover, the clusters are really unbalanced: almost 35% of the travelers are in Cluster 6 while Cluster 4 contains only 3.7% of the travelers.

## 4 Results obtained by NMF

Consider the matrix $V$ that contains the temporal profiles $V_i$ of all users as rows. The principle of Nonnegative Matrix Factorization is to factorize the matrix $V \in \mathbb{R}^{n\times m}$ into two matrices $W \in \mathbb{R}^{n\times K}$ and $H \in \mathbb{R}^{K\times m}$ such that $V \approx WH$. When $K \ll n, m$ the number of entries in $V$ is much bigger than the ones in $W$ and $H$: $Kn + Km \ll nm$. So each profile $V_i$ is approximated by $W_{i,1}H_1 + \ldots W_{i,K}H_K$ and so the $H_j$'s can be interpreted as a dictionary of profiles. Moreover, the nonnegativity constraint sets some $W_{i,j} = 0$ and so each profile is approximated as a small number of elements in the dictionary.

It is important to find a $K$ small enough to ensure that $Kn + Km \ll nm$, but large enough so that $WH$ remains an acceptable approximation of $V$. Let $D(V|WH)$ denote a generic function measuring the distance between $V$ and its approximation $WH$. To summarise, the aim of the NMF is to solve the following problem :

$$\min D(V|WH), \text{ subject to } W \geq 0, H \geq 0 \tag{1}$$

ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017, i6doc.com publ., ISBN 978-287587039-1.
Available from http://www.i6doc.com/en/.

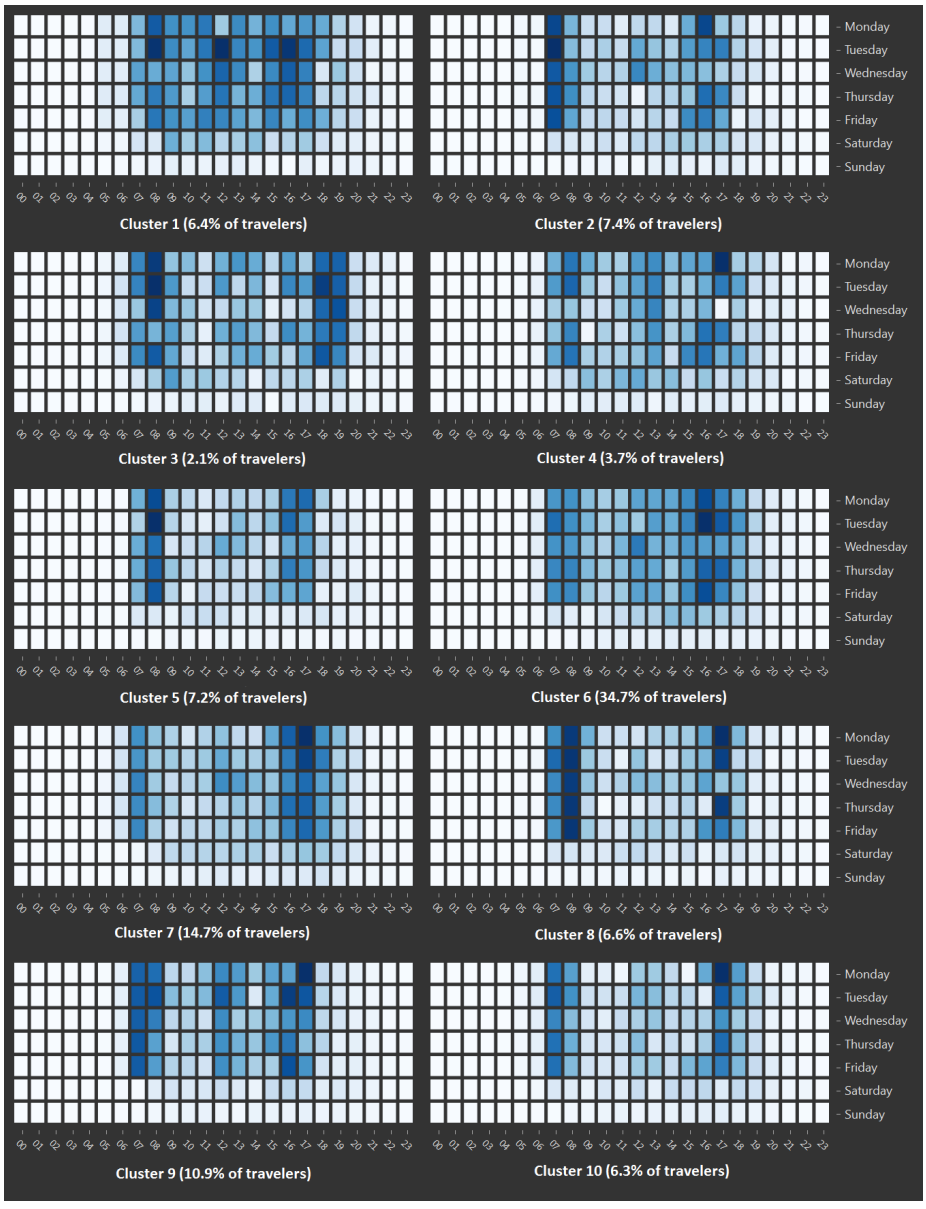


Fig. 1: Clusters obtained by EM-algorithm

ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017, i6doc.com publ., ISBN 978-287587039-1. Available from http://www.i6doc.com/en/.

where the inequalities are interpreted elementwise. Several algorithms are known to compute $W$ and $H$. These methods are discussed and compared in [9].

We chose $D(V|WH) = \|V - WH\|_F^2$ in (1). We tested different algorithms recommended in [9]: the multiplicative algorithm and the projected gradient method. The results being similar on our dataset we only present the results obtained by the multiplicative algorithm. For the sake of completeness, we remind that the algorithm is an iteration of the following udpates:

$$W_{i,a} \leftarrow W_{i,a} \frac{(VH^T)_{i,a}}{(WHH^T)_{i,a}} \text{ and } H_{a,\mu} \leftarrow H_{a,\mu} \frac{(W^TV)_{a,\mu}}{(W^TWH)_{a,\mu}} \ \forall i, \mu, a.$$

Here again we tested several dimensions $K$, and then used the $k$-means algorithm on the matrix $V$ to get our clusters. The dictionary and the clusters centers are shown in Figure 2 for $K = 7$, which were particularly easy to analyse. Indeed, as it can be seen on the Figure 2 the first word corresponds to the first hour of the morning peak (7 a.m.). The second word corresponds to the second hour of morning peak (8 a.m.), the third to the last two hours of afternoon peak (6-7 p.m.) and Saturdays afternoons, the fourth to the off-peak periods, mostly in the morning (9-11 a.m. and 2-3 p.m.), the fifth to the midday hours (12 and 1 p.m.), the sixth to the first hour of afternoon peak (5 p.m.) and the seventh to the off-peak period in the afternoon (3 to 4 p.m.).

In the middle of the Figure 2, the first cluster obtained with a k-means method applied to our reducted space is mostly a combination of the first and the sixth words that respecively explain 33.2% and 20.5% of the cluster. The rest of the cluster is explained by all the others words in negligible proportions. Each cluster is similarly a linear combination of the seven "words".

We can note that the clusters of travelers temporal profiles are more easily interpretable than the ones obtained by EM-algorithm. Clusters 1, 2, 5, 8 and 10 represent groups of people traveling during the peaks during the week and sometimes in Wednesdays noons. Cluster 4 represents travelers who use the public transportation during the peaks but also during the midday hours. Cluster 6 represents people travelling almost only during the afternoon (it may be people who use an other modal transport in the morning). Cluster 7 gathers users who only travel in off-peak. Finally, Clusters 3 and 9 contain travelers who have diffuse habits of travel during the day. Also note that the groups are more balanced as largest cluster contains only 15.5% of the users.

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.

## 5 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

ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017, i6doc.com publ., ISBN 978-287587039-1.
Available from http://www.i6doc.com/en/.

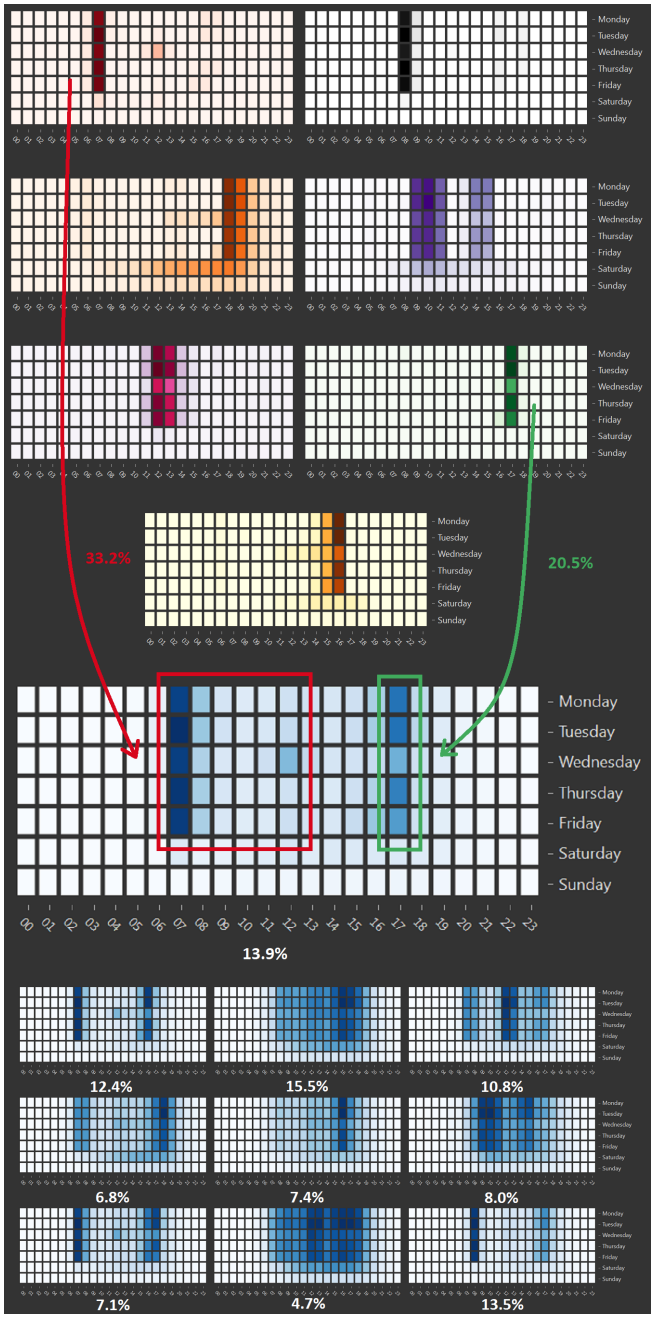


Fig. 2: Up : "words" of the dictionnary obtained by NMF; Middle : Decompozition in "words" of one of the clusters obtained by k-means; Down : The other clusters obtained by k-means on the reducted space

ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 26-28 April 2017, i6doc.com publ., ISBN 978-287587039-1. Available from http://www.i6doc.com/en/.

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% |

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] Yu Zheng, Licia Capra, Ouri Wolfson, and Hai Yang. Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(3):38, 2014.

[2] Catherine Morency, Martin Trepanier, and Bruno Agard. Measuring transit use variability with smart-card data. *Transport Policy*, 14(3):193–203, 2007.

[3] Marie-Pier Pelletier, Martin Trépanier, and Catherine Morency. *Smart card data in public transit planning: a review*. CIRRELT, 2009.

[4] Mohamed K El Mahrsi, Etienne Côme, Johanna Baro, and Latifa Oukhellou. Understanding passenger patterns in public transit through smart card and socioeconomic data. 2014.

[5] Daniel D Lee and H Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.

[6] Wei Xu, Xin Liu, and Yihong Gong. Document clustering based on non-negative matrix factorization. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 267–273. ACM, 2003.

[7] Farial Shahnaz, Michael W Berry, V Paul Pauca, and Robert J Plemmons. Document clustering using nonnegative matrix factorization. *Information Processing & Management*, 42(2):373–386, 2006.

[8] C Bishop. Pattern recognition and machine learning (information science and statistics), 1st edn. 2006. corr. 2nd printing edn, 2007.

[9] Jingu Kim and Haesun Park. Sparse nonnegative matrix factorization for clustering. 2008.

[10] C. Bouveyron, E. Côme, and J. Jacques. The discriminative functional mixture model for a comparative analysis of bike sharing systems. *The Annals of Applied Statistics*, 9(4):1726–1760, 2015.