Mining ticketing logs for usage characterization with nonnegative matrix factorization

Mickaël Poussevin, Emeric Tonnelier, Nicolas Baskiotis, Vincent Guigue, and Patrick Gallinari

Sorbonne-Universités, UPMC-Paris 6
Laboratoire d'Informatique de Paris 6, UMR 7606, CNRS

Abstract. Understanding urban mobility is a fundamental question for institutional organizations (transport authorities, city halls) and it involves many different fields like social sciences, urbanism or geography. With the increasing number of probes tracking human locations, like RFID pass for urban transportation, road sensors, CCTV systems or cell phones, mobility data are exponentially growing. Mining the activity logs in order to model and characterize efficiently our mobility patterns is a challenging task involving large scale noisy datasets.
In this article, we present a robust approach to characterize activity patterns from the activity logs of a urban transportation network. Our study focuses on the Paris subway network. Our dataset includes more than 80 millions travels made by 600k users. The proposed approach is based on a multi-scale representation of the user activities, extracted by a nonnegative matrix factorization algorithm (NMF). NMF is used to learn dictionaries of usages that can be exploited in order to characterize user mobility and station patterns. The relevance of the extracted dictionaries is then assessed by using them to cluster users and stations. This analysis shows that public transportation usage patterns are tightly linked to sociological patterns. We compare our approach with a k-means baseline that does not take into account user information and demonstrate the interest of characterizing user profiles to obtain better representations of stations.

1 Introduction

The literature on urban mobility is vast and diverse but until recently, it has mainly focused on explanatory statistics of global behaviors. With the development of tracking techniques such as mobile phone networks, the last decade has seen a multiplication of quantitative statistical studies. For example, frequency scales of travels are analyzed in [3] and multiple studies have shown that it is possible to predict most daily travels [19]. For public transportations, some early studies focused on opinion pool to analyze the modification of user behaviors after the creation of new lines [6]. In this domain, quantitative data are recent and linked to the adoption of smart cards to authenticate users in most cities, like e.g. in London, Lisbon or Paris. Up to now, they have been exploited for

problems like bottleneck detection [4] or frequent pattern prediction [5], which turns out to determine the time and location of the next trip for a given user.

However, no study has focused on mining the temporal and spatial profiles of individual users. Analysis have been performed on Shanghai taxis [16] and Parisian public bicycle sharing system [17] for example, but they have been mainly focused on extracting global statistics and none of them has analyzed individual user traces across multiple travels during multiple days. Discovering latent usage of public transportations is a crucial need for institutional authorities: lot of efforts are spent on conducting ground surveys to achieve a partial understanding of the habits of their clients. Such knowledge is important for pricing policies, load management and planning. We propose to exploit ticketing data to extract regularities in usage patterns in order to identify the hidden activities that causes each individual event (i.e. to explain all logs in the data).

Our analysis focuses on data provided by the STIF (*Syndicat des Transports en le de France*, Paris area transport authority) and contains around 80 millions log entries overs 91 days, with an explicit identifier for both the station (300 locations) and the user (600k ids). The noise inherent to the individual activity traces and the size of the logs explain why ticketing data has hardly been used. We propose here an experimental study to show the potential of machine learning to exploit transport data. We introduce a user centered multi-scale representation of the data and we use a constrained nonnegative matrix factorization (NMF) to extract latent activity patterns. Based on this extracted representation, we build station profiles that we cluster and analyze the obtained segmentation. We demonstrate the interest of our approach with respect to a k-means baseline : we show that an efficient profiling requires a user modeling step, even if finally we focus on station representation. We also focus on data reconstruction and abnormality detection: our model is able to output log predictions. Using a symmetrised Kullback-Leibler divergence between ground truth and predictions reveals some abnormalities in the log flow associated to each station.

The paper is organized as follows. We briefly review literature on related work in section 2. To face the challenge of data size and sparsity, we propose in section 3 a model to aggregate ticketing logs per user and station on three frequency and two temporal scales. In section 4 we present the nonnegative matrix factorization model used to extract the usages and latent activities from user events. Analysis of the results on our dataset is presented in section 5. As an application case, in section 6 we illustrate how to use this model for clustering stations and extracting correlations between temporal habit and sociological realities.

2 Related work

We present below a synthesis on the related work both on urban mobility understanding and on nonnegative matrix factorization algorithms.

2.1 Urban mobility

The problem of understating urban mobility has been studied at different levels. [2] studied city planning policies in order to promote the use of performance indicators for sustainable public transports. Many studies [3, 8] have focused on private vehicle travels, using mobile phone networks to track a population and to characterize the time scales of these travels. In [19], the authors showed that most private vehicle travels are predictable and in [22] they also linked travel behaviors and social network behaviors. The recent study [15] also uses mobile phone networks to extract hot spots from week day travels in the 31 biggest Spanish cities while [10] focused on car traffic analysis and abnormality detection. [14] used 6 months of GPS data coming from the 33000 taxis in Beijing, covering an impressive 800 million kilometers, to analyze the causes of possible abnormalities. Also with GPS data but on 2000 Shanghai taxis, [16] used nonnegative matrix factorization to characterize the behavior of taxi drivers. Finally, several recent works have focused on new ways to track users. For example [13] describe the use of Bluetooth scanners to track pedestrians visiting Duisburg zoo. Similarly in Paris, [17] mined spatio-temporal clusters using Paris Velib' data (city shared bicycle service), but without having access to the user identification, they could not track individual users.

The creation of smart cards to authenticate users [6] allows for wider scale and finer analysis. One important target has been the identification of bottlenecks in the network. For instance [4] studied spatio-temporal distribution of users in the subway of London. In the same way, [5] focused on itinerary prediction, during week days for buses, so as to inform users in case of problems in the network, mining a dataset of 24 million travels of 800000 users over 61 days.

In contrast with previous work focused retrieving global traffic information, the analysis presented here focuses on discovering a collection of trip habits that can be used to describe individual users. As in [16], our model relies on a modified nonnegative matrix factorization, described in section 4, to extract behavioral atoms. We also exploit the user identification to characterize travels through the notion of periodicity.

2.2 Nonnegative matrix factorization (NMF)

Matrix factorization approaches have long been used in data mining and are well described in [7]. Nonnegative matrix factorizations [1] extract constructive representations over a set of extracted basic components out of nonnegative data. Each basic component is named *atom* and the extracted component collection is named *dictionary*. The dictionary is learned such that each data can be approximate by the positive weighted combination of atoms of the dictionary. A such weighted vector over the atoms is named *code*. Matrix factorization has shown interesting performances as a feature extractor when the nonnegativity is a sensible part of the data either to obtain a composition percentage over a dictionary like in face detection [24] or when a negative coding does not make sense like in topic extractions from document application [18].

Matrix factorizations solve two problems at once: learning the dictionary and the associated code. Constraints are generally added during the learning algorithm as regularization terms. Nonnegativity [12] and sparseness [11] are two common constraints. As no subtraction is allowed with the nonnegativity constraint, Nonnegative Matrix Factorization (NMF) is more likely to obtain part-based representation with atoms being parts of the initial signals and the code being the proportion of each atom in the signal. Sparseness forces the model to reconstruct initial data using only a few atoms.

While a completely different field, our task might be linked to the separation of musical sources [21] or note identification in a music track [20]. The log of a user corresponds to frequent and multiple activities, it can be seen as a stream of pulses generated by multiple sources [9]. However the variability of the time of authentication, which is our reference event, hinders the approaches based frequency representation used in most signal applications. In [23], nonnegative matrix factorization has been adapted to time event detection to identify *time-shift* invariant patterns. It cannot be used as such to detect temporal behaviors as they are typically time dependent: our idea is to characterize activities based on event occurrence.

3 Subway data analysis and modeling

In this study, we aim at discovering patterns in the subway usage of any user in order to characterize each log by a corresponding activity: an event which occurs every working day around 7 a.m. may be categorized as going to work, an event which occurs some Friday night as nightlife, etc. Tagging authentication events with a social activity allows us to characterize both the user (workplace, home, friends' home, recreational habits) and the station at the same time. We propose a model extracting meaningful activity temporal patterns and allowing a categorization of the subway traffic according to different usages. In this section, we first describe the data and then our modeling.

3.1 Ticketing logs

We use a dataset collected by the *Syndicat des Transports en Île-de-France* (STIF), the transport organization authority in charge of Paris public transport. More than seven millions of Paris transport users have subscribed to a pass managed by the STIF. The ticketing logs record every authentication of any pass with its location and precise time. This dataset provides an accurate real-time picture of the use of a public transportation system. The analysis is challenging for two reasons: the size of the data (5GB/month) coupled to the sparsity of individual user data hinders the mining of frequent behavioral patterns; secondly, the dataset contains only a subset of the urban network activity. The STIF estimates that 20% to 30% of logs are missing, either due to the malfunctioning of a turnstile or to the user voluntary not authenticating itself.

Moreover, the logs correspond to a check-in action, as pass checkpoints are positioned at entry points in metro stations and buses, and not at exit points. Thus the user's itineraries are not explicitly present in the dataset, only a (large) part of their check-ins are recorded.

3.2 Modeling

In the following, u will refer to a user, s to a station and t to a time. An authentication at time t is thus a triplet $\ell = (u, s, t)$. The mobility of a user is partially described by the log of its authentications, $L = \{\ell = (u, s, t)\}$. Each authentication is related to a latent activity of the user. We propose in this section a latent model using a multi-scale aggregation of events by day and by week.

A first difficulty for identifying latent activities lies in the wide frequency scale of the events. Non-frequent events are hidden by frequent ones, any approximation will catch frequent day-to-day activities and miss the less frequent patterns. We propose to dispatch user activities in three frequency bands: high frequency for events occurring more than twice a week, medium frequency for events occurring at least once each 10 days and less than twice a week, and low frequency for unusual events. We use the spatial information to filter events in the frequency bands: for a given user, each event is associated to a frequency band based on the number of times the event station appears in the user log.

We make the assumption that within a frequency band activities are more characterized by their occurrence time than by their location. On the one hand, this assumption correctly interprets regular activities occurring at the same place (like leaving for work, going back home); on the other hand, it allows us to capture recreational activities that occur in wide areas (like going out to restaurant, visiting friends).

Within a frequency band, an event is represented by the couple (u, t). The user behavior can be modeled by the probability function of the authentication time for this user in the three frequency bands: $p^b(t|u)$, $b \in \{low, medium, high\}$. Our goal is to discover a set of activities $\mathcal{A}$ over which the user log can be decomposed as: $p^b(t|u) = \sum_{a \in \mathcal{A}} p(t|a) * p(a|u)$. Our approach relies on a multi-scale representation of authentication events, by day and by week. We chose to characterize an activity, that we denote a, by the probability function of the ticketing event during the day $f_{d,a}(t) = p(t|a)$ and during the week $f_{w,a}(t) = p(t|a)$ with t respectively a day time and a week time variable. As we are more interested on the discovery of widespread usages, our objective is to infer a small set of activities $\mathcal{A}$ sensible for all users.

This formalism supposes a weekly pattern and does not allow to model explicitly the localization information. However, the station information can still be decoded from the individual user data knowing the activity decomposition for a user.

3.3 Notations and data representation

We first filter authentication triplets (u, s, t) in three frequencies bands: *low* for couples (u, s) occurring only a few times, *high* for frequent couples (u, s) and *medium* for everything in between. On figure 1 is represented the authentication set of a single user with stations sorted by frequency, from low to high. The two most frequent stations for this user are likely to be its home and workplace. For all frequency bands f in $\{low, medium, high\}$, we build a multi-scale vector representation of a user as the concatenation of the two probability functions of the authentication time during the day and the week. Both are empirically estimated using a time step of 15 minutes for the day and 2 hours for the week. Thus each user is represented by a n dimensional vector with $60/15 * 24 = 96$ dimensions for the day and $24/2*7 = 84$ for the week. We denote m the number of users. As a result, data are represented by a set of 3 matrices: $\{X^{(b)} \in \mathbb{R}^{m \times n} | \forall b \in \{low, medium, high\}\}$.

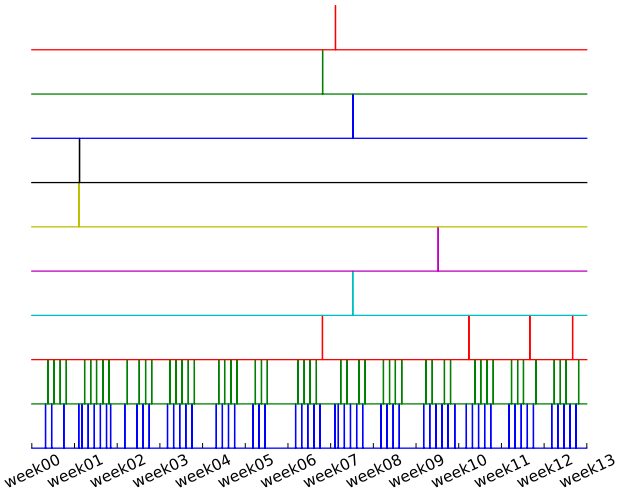


Fig. 1: A user of the Parisian network authenticated at 10 stations over 91 days. Stations are ordered by decreasing frequency from bottom to top. The most frequent stations are relative to the residence and the workplace. Less frequent ones are likely to correspond to recreational activities.

Figure 2 shows a subset of user profiles aggregated without the frequency filtering. The data is noisy reflecting users' individual variance. Still this representation extracts peaks of activities. As argued above, without the frequency filtering most of the density is used by recurrent commuting patterns. User profiles after frequency filtering are represented in figure 3, where *low*, *medium* and

high are respectively the bottom, middle and top. Typically, commuting patterns are present in the high frequency band and week-end and evening events are present in the other two bands. Some users have no event in a particular band, which is a strong characterization of their behaviors.

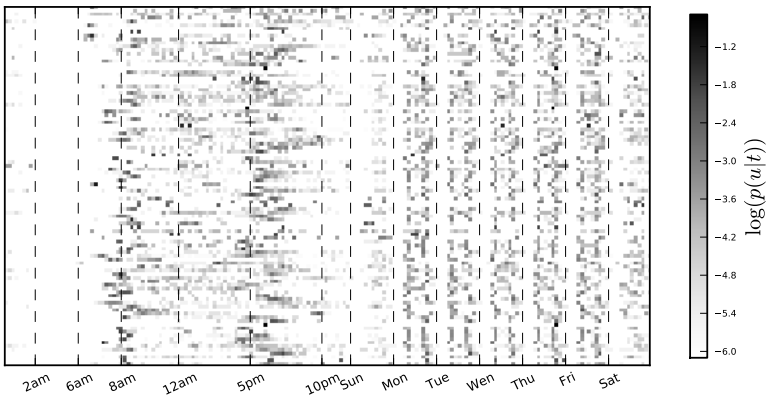


Fig. 2: Aggregated user profiles over time ordered by time of their highest peak. Concatenation of the day and week scales.

4 Learning usage atoms with NMF

The previously extracted matrices from authentication data represent the averaged daily and weekly user profiles. We aim at extracting the latent behavioral patterns which compose these profiles. Our goal is to achieve a granular identification of behavioral patterns, allowing us to characterize heavy and light traffic periods, during evening, nights and week-end. Thus the challenge here is to extract generic patterns reflecting the most common patterns like commuting but also less frequent ones happening during evenings and week-ends without over-fitting and learning patterns specific to only of small set of users.

We propose to use a nonnegative matrix factorization algorithm to extract the pattern dictionary with a mono-modal constraint on dictionary rows. As nonnegative matrix factorization decomposes each sample as a positive linear combination of atoms, it is well-suited to our problem: we seek to approximate a user by a set of complementary behaviors.

Formally, the goal is to approximate each matrix X by the product of two positive matrices, D the pattern dictionary and A the code matrix: a row of D, named atom, represents the time profile $p(t|a)$ of a particular activity a and a row of A contains the coordinates of a user in this usage space, that is the repartition $p(a|u)$, for the user u, of the activities in $\mathcal{A}$. These coordinates can be viewed as the usage probabilities for the user.

We want to take into account the following constraints:

- normalization: each dictionary atom is the concatenation of the daily and the weekly information representing the probability density function of the ticketing event; thus both parts of the dictionary atom has to sum up to 1;
- mono-modal atoms: an atom is supposed to explain a unique activity in the day, localized at a precise time window;
- sparsity of the reconstruction: a user has naturally few activities, thus only a small set of dictionary atoms should have a non-zero weight in each row of A.

The dictionaries for the three frequency bands are distinct: they can be learned separately. For each frequency band, the problem can be formalized as the minimization of a loss function $\mathcal{L}$ under the constraint $C(D)$ on the dictionary D:

$$\mathcal{L}(X, A, D) = \frac{1}{m}\|X - A.D\|^2 + \lambda|A| \tag{1}$$

$$C(D) : D \geq 0, \forall i, \sum_{j<t_{day}} D_{ij} = 1, \sum_{t_{day}\leq j<t_{week}} D_{ij} = 1 \tag{2}$$

The NMF optimization is done using a projected gradient (with projection ϕ on the constraints) for the dictionary and multiplicative update rules for the codes (see algorithm 1).

$D, A \leftarrow$ rand ;
while *not converged* **do**
  $D = D - \mu(A^T(X - AD)$;
  $D = \phi(D)$;
  $A = A \odot \frac{XD^T}{\lambda + ADD^T}$;
end

Algorithm 1: NMF update rules: projected gradient on D and multiplicative update rules on A

To force the mono-modal property of dictionary atoms, an additional projection is used every 100 iterations, by applying a Gaussian filter to each atom in order to capture only the highest peak of the daily part:

$$\forall i, \quad d_{i,day} \leftarrow d_{i,day} \odot \exp(-\frac{(t_{day} - t_{peak})^2}{2\sigma^2}) \tag{3}$$

where $\odot$ is the element-wise multiplication, t_{day} ranges over the day dimensions of the atom, t_{peak} is the time corresponding to the peak day time in the atom and σ a fixed parameter defining the granularity of the time window and $d_{i,day}$ is the part of i-th row of the dictionary D accounting for daily patterns.

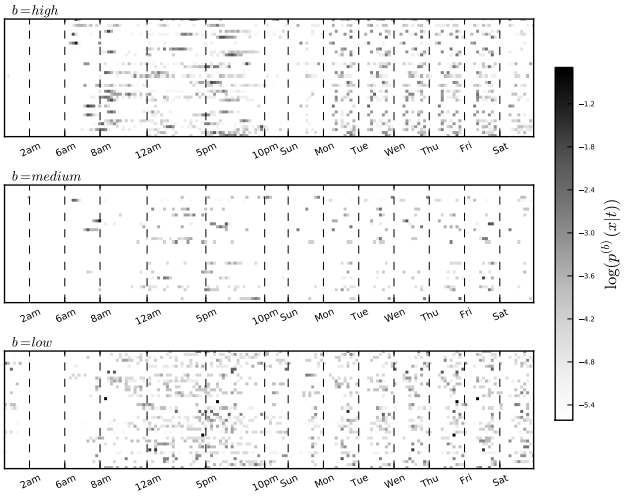


Fig. 3: Aggregated user profiles overtime with frequency filtering. Events with high, medium and low frequencies are in top, middle and bottom charts respectively. Concatenation of the day and week scales.

5 Analysis of extracted representation

We focus our analysis on the subway traffic, excluding buses, trains and trams. As we are looking for mobility usages. Users were filtered to retain only those with enough travels to have a significant activity and subscribing a monthly pass. Our dataset is composed of around 80 million user authentications at one of the 300 stations over 91 days for the subway network of the city of Paris, concerning around 600k unique users.

The $k = 100$ atoms were extracted with 180 features of $m = 600000$ users, using a 8 cores (3.07 GHz) and 16 GB of RAM PC. It takes approximately 10 hours to complete the 1k iterations of the nonnegative matrix factorization algorithm.

Extracted atoms are represented in figure 4 where they are sorted by their occurrence of highest peak. In the high frequency band, where commuting patterns are the most frequent ones, the mono-modality allows the discovery of joint *leaving-for-work* and *going-back-home* patterns. The extraction is mainly focused on working days since only a small part of the density occurs during week-ends. With most atoms occurring in the morning the high frequency band has a fine granularity on the *leaving-for-work* usage. In the medium frequency band more atoms are devoted to lunch activities and they typically are active during working days too. Still some atoms, bottom ones, are catching evening and week-end activities. In the low frequency band, density of the week is more

evenly distributed through the seven days. A vast majority of the atoms are representing evening and nightly activities. Note that the working hours of the subway are 5:30 am to 2 am, with a slight variance over days, which explains why there is no activity corresponding to the 2 am to 5:30 am period.

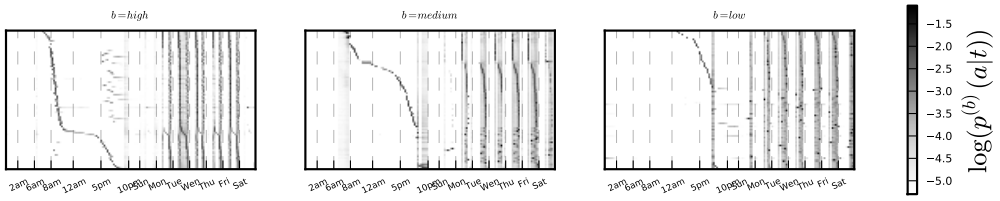


Fig. 4: Activity atoms extracted from the profiles per frequency band. Concatenation of day and week scales. Blue values are zeros, red values are peaks.

Quality measures of the nonnegative matrix factorizations are presented in table 1. The data sparseness is the average percentage of nonzero entries per user. The dictionary sparseness is the average percentage of nonzero entries per dictionary atom. The code sparseness is the average percentage of nonzero entries per user code (a row of α). The important activities row counts the average number of activities responsible of 90% of a user's profile. It is interesting to see that the factorization on the *low* frequency band scatters the density over more activities than the *medium* and *high* band. It is coherent with the fact that infrequent validations correspond to more diverse usages (visiting friends, hiking in the city, going to a restaurant,...). This value has to be put in perspective with the bar plots and histograms of figure 5. The bar plots of the top row of figure 5 represents, per frequency band, the percentage of users that have a non-zero weight attributed to each dictionary atom. The shape of the curves indicates that all dictionary atoms are evenly used to describe users. This is the sign that the model did not over-fit the dataset and it is one of the property we were looking for: the extracted behavior pattern, the dictionary atoms, are not to specific to one particular user. The bottom row histograms count the number of dictionary atoms per user. The first noticeable effect of the frequency filter is that in the medium and high frequency bands a great proportion of users are characterized by a lack of events. It also confirms the good sparsity of the extracted representation: the NMF uses few dictionary atoms to reconstruct each user profile. NMF appears as a robust solution to extract latent activity patterns from noisy temporal data.

6 From users to stations

Using NMF, each user is now represented as a vector of usage weights. In this section we analyze the distribution of the individual usage patterns according

NMF	*low*	*medium*	*high*
Data sparsity (std)	16.98 (0.38)	7.97 (0.27)	16.88 (0.37)
Dictionary sparsity (std)	51.15 (0.50)	52.41 (0.50)	51.93 (0.50)
Code sparsity (std)	9.23 (0.29)	4.33 (0.20)	5.46 (0.23)
Important activities (std)	4.38 (0.20)	1.94 (0.14)	1.68 (0.13)

Table 1: NMF sparsity measures on the *low*, *medium* and *high* frequency bands

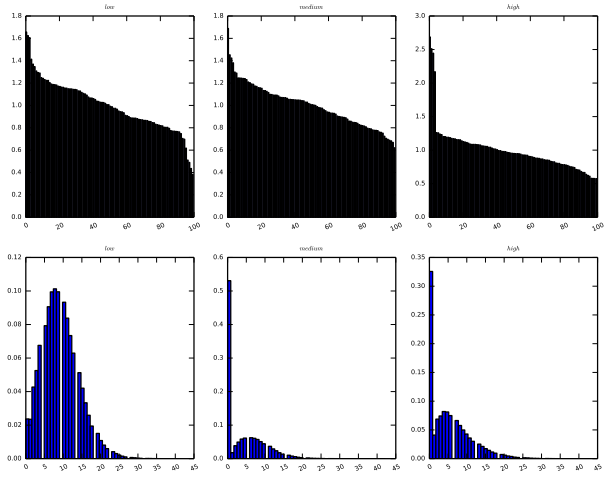


Fig. 5: Top row represents usage of each dictionary atom over the set of users per frequency band. Bottom row contains the histogram of the number of dictionary atoms having non-zero weights per user per frequency band.

to the metro station geographical position. Formally, we want to estimate the probability function $p(a|s)$ of activities over subway stations. We will use the following decomposition: $p(a|s) = \sum_u p(a|u)p(u|s)$. To simplify the estimation of $p(u|s)$, we consider that it is uniform over the stations visited by the user in the frequency band of the activity.

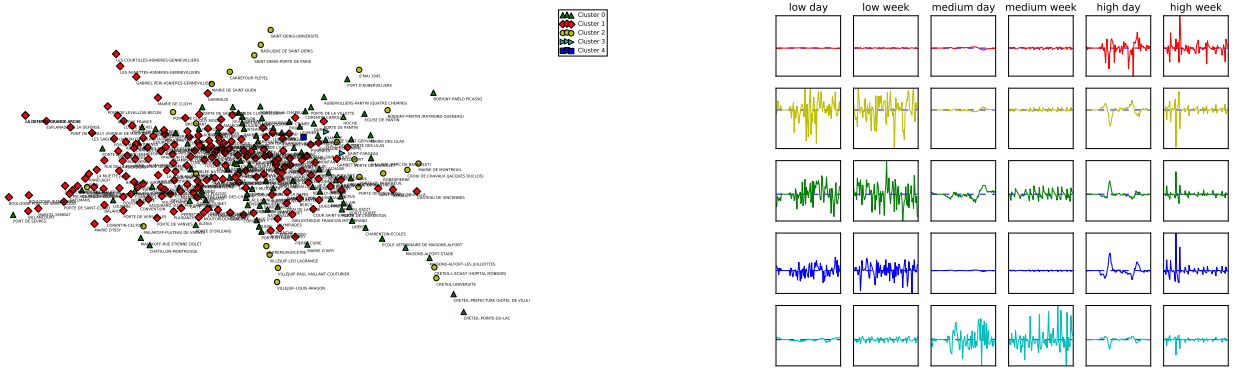
We build the representation of each station for each frequency band as the sum of all the activities of all the users authentication at this station weighted by the overall usage in this frequency band. This gives us a set of three representation matrices for the station, one per frequency band, a row of which representing a particular station. As we extract 100 atoms per frequency band and there are around 300 stations in our problem, these matrices have small dimensions.

We use a multi-scale clustering algorithm, similar to what [25], to cluster together stations into coherent groups of behaviors. Since the matrices are small (300×100), the clustering is fast (a couple of seconds) on a typical computer. It is stable with respect to the initialization. For simplicity, we present the result using 5 clusters to capture macroscopic groups of behaviors in the network. In order to evaluate our approach, we compare NMF-clustering results with a basic k-means learnt over raw station logs (without taking into account user profiles).

Figure 6 illustrates the result of the k-means procedure as well as the associated prototypes. Results are very hard to interpret: classically, a main cluster corresponds to standard behavior while smaller classes model what appear as random variations around the first prototype.

Figure 7 represents the NMF clustered map of the subway stations in Paris where each station is colored (and shaped) according to its cluster. Some geographical patterns clearly emerge. There are two inner clusters in the center, one belt-like cluster around this center and the separation of the western and eastern sub-urban regions around Paris. It is interesting to compare this clustering to the sociological geography of the city. First the touristic center of Paris with *Champs Élysées* and *Concorde*, the Louvre Museum, the Garnier Opera, *Notre Dame* and the *Sacré Cœur* are in the same cluster. Second, the belt-shaped cluster, here in yellow squares, corresponds to the limits between Paris and the surrounding cities. The city limits are marked by *Porte* (gates in French) as a reminder of the gates piercing the old fortified walls of the medieval Paris. And last but not least, the clustering opposes the posh western sub-urban regions of Paris to the relatively poorer eastern sub-urban regions: the distinction, based on temporal patterns, is interesting as the users might have the same patterns since they are at the same distance from the city center. So the distinction goes beyond the simple geographic explanation and touches a sociological repartition of work hours.

The NMF has provided centroids composed of one pattern per frequency band. For any given cluster, the high frequency band comes from the usage pattern of people that frequently check-in in one of the stations meaning they either live or work there. The low frequency band corresponds to people that hardly use the station (less than once every ten days) and is basically composed



(a) Map of Paris subway stations with colors coming from the basic k-means clustering, without taking into account user profiles.

(b) Prototypes associated to each cluster from the k-means procedure. As all prototypes were very close, we choose to plot the differences between the prototype and the global mean.

Fig. 6: k-means negative result: no relevant behavior emerge from this basic procedure.

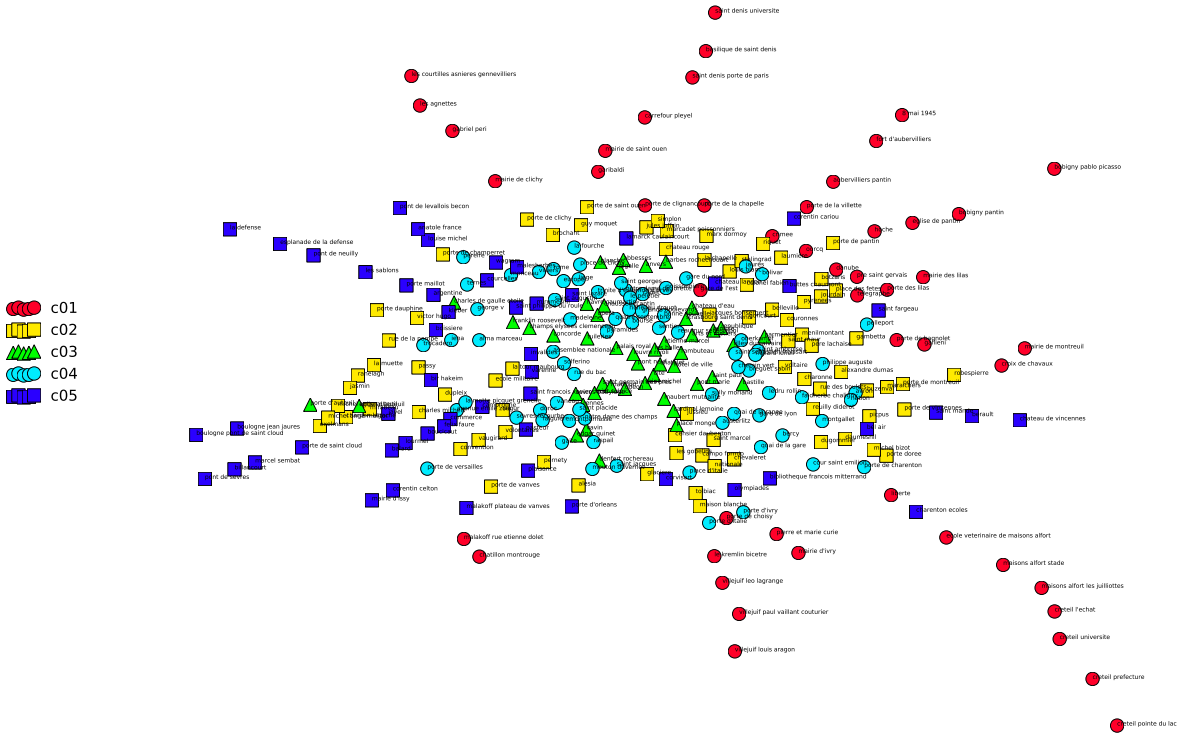


Fig. 7: Map of Paris subway stations with colors coming from our NMF clustering.

of evening events, meals or sleepovers. The medium band corresponds to everything in between. The difference between the patterns of the centroids and the average network load are represented in figure 8, with each centroid being colored as the corresponding cluster in figure 7. A positive spike means more users authenticating themselves at any station of the cluster than in average over the network. A negative spike is the opposite: a deficit of users compared to average load.

As only the check-in authentications are available to us, the high frequency band corresponds to the habit of the people living or working near the stations of the cluster while the medium and low frequency bands correspond to people check-in to travel elsewhere, meaning that they came to the stations of the cluster for some periodical activity like pubs, restaurants, theaters and so on and are now leaving.

The last line is the belt cluster that corresponds to the average behavior of standard Paris dwellers which is coherent with the fact that stations composing this cluster are at the limit of the city. The first and third lines are close one to another. The former corresponds to the touristic center of Paris which is sensible as it is characterized by a lack of check-ins in the morning for no working class is living there. Once again the temporal pattern is linked to a sociological repartition. The latter contains the big clusters corresponding to train stations. People working in sub-urban regions where the subway network is not present typically take the train to one hub and then finish their journey with the subway. This explains the main difference in the high frequency band between the two clusters: the peak around nine in the morning. It is also noticeable that these two clusters are the only ones having people departing after infrequent activities, as can be seen on the low frequency band. The second and fourth lines correspond respectively to the western and eastern sub-urban regions around Paris. As said earlier they are mainly residential areas with the western part being wealthier than the eastern one. This is confirmed by the night part of the medium and low frequency bands: in contrast to the center clusters, few people are leaving these clusters after some episodic activities. The phenomenon of a peak in the medium band correspond to the activity of leaving for work from a station that is not the main home of a user and is typical of sleepovers. Finally the commuting patterns of the two clusters is different. The model is able to extract fine grained representations and distinguish the clusters as they do not commute at the same time. Stations in poorer regions see a peak early in the morning followed by a deficit of users at the same time as the affluence peak in stations of posh neighborhoods.

7 Abnormality detection

The NMF decomposition allows to reconstruct the temporal log flow of a station. We exploit this ability to build the standard weekly profile associated to every stations and we measure the distance between reconstructed data and ground truth using a symmetrised Kullback-Leibler divergence (KL). In order to improve

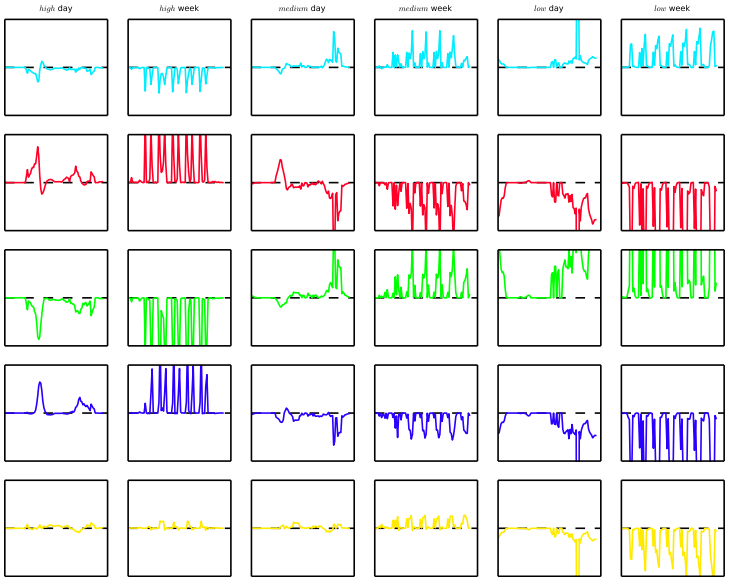


Fig. 8: Difference to the average behavior per multi-instance centroids. Juxtaposition of the day and week scales for the high, medium and low frequency bands respectively.

robustness, we aggregate the results respectively over days and weeks. Figure 9 illustrates some promising results: whereas raw log flows seem roughly standard, KL measure enables us to point out clearly local abnormality at different scales.

8 Conclusion

In this work, we have proposed a new approach to urban mobility analysis introducing a machine learned based modeling that fully exploits the data available since the introduction of RFID cards in public transportation. We gave proposed a multi-scale modeling of event logs of users in order to retrieve latent activities of users. A nonnegative matrix factorization algorithm with sparsity, mono-modality and normalization constraints is used to build the set of dictionary atoms representing these activities. We have analyzed and exploited the extracted representations of the users to build stations profiles and to cluster them. We showed the interest of our formulation with respect to simpler a k-means that does not take into account user modeling. Finally we demonstrate our ability to reconstruct a temporal log distribution at the station scale as well as to detect abnormalities in the log flow.

The study of these clusters has revealed that temporal patterns are able to capture fine grained representations of behaviors from roughly aggregated noisy ticketing logs. We used here the extracted latent activities in a qualitative study. From a machine learning point of view, the extracted dictionary atoms contains

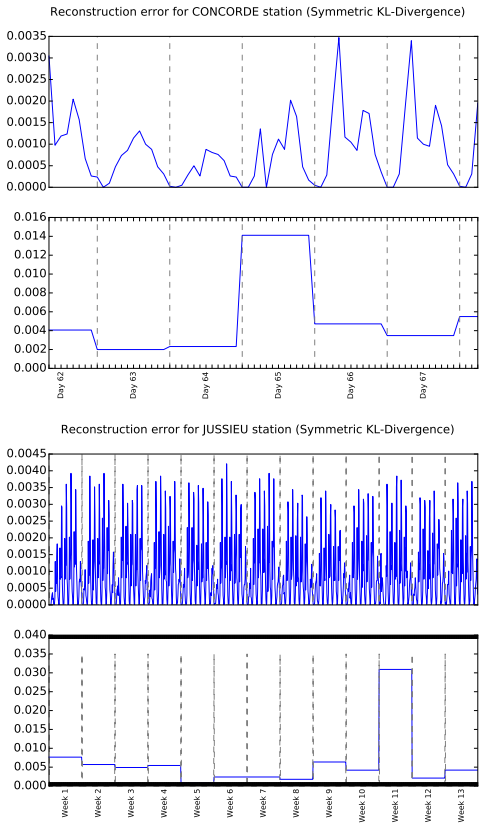


Fig. 9: Ground truth profiles and symmetrised KL divergences for 2 stations (Concorde and Jussieu) respectively aggregated by day and week.

meaningful high-level information that can be exploited further to jointly characterize users and locations.

One main perspective of this work concerns the design of an embedded model able to model all frequency scale at the same time. Such a model will not be easy to develop; indeed, preliminary works show that cascade approches mainly focus on frequent events without describing rare phenomena.

References

1. Berry, M.W., Browne, M., Langville, A.N., Pauca, V.P., Plemmons, R.J.: Algorithms and applications for approximate nonnegative matrix factorization. Computational Statistics & Data Analysis (2007)
2. Black, J.A., Paez, A., Suthanaya, P.A.: Sustainable urban transportation: performance indicators and some analytical approaches. Journal of urban planning and development (2002)

3. Brockmann, D., Hufnagel, L., Geisel, T.: The scaling laws of human travel. Nature (2006)
4. Ceapa, I., Smith, C., Capra, L.: Avoiding the crowds: understanding tube station congestion patterns from trip data. In: ACM SIGKDD (2012)
5. Foell, S., Kortuem, G., Rawassizadeh, R., Phithakkitnukoon, S., Veloso, M., Bento, C.: Mining temporal patterns of transport behaviour for predicting future transport usage. In: Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication (2013)
6. Golias, J.C.: Analysis of traffic corridor impacts from the introduction of the new athens metro system. Journal of Transport Geography (2002)
7. Golub, G.H., Van Loan, C.F.: Matrix Computations (3rd Ed.). Johns Hopkins University Press (1996)
8. Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.L.: Understanding individual human mobility patterns. Nature (2008)
9. Hegde, C., Baraniuk, R.G.: Sampling and recovery of pulse streams. IEEE T. on Signal Processing (2011)
10. Herring, R.J.: Real-Time Traffic Modeling and Estimation with Streaming Probe Data using Machine Learning. Ph.D. thesis, University of California, Berkeley (2010)
11. Hoyer, P.O.: Non-negative matrix factorization with sparseness constraints. JMLR (2004)
12. Lee, D.D., Seung, H.S.: Algorithms for non-negative matrix factorization. In: NIPS (2000)
13. Liebig, T., Xu, Z., May, M., Wrobel, S.: Pedestrian quantity estimation with trajectory patterns. In: Machine Learning and Knowledge Discovery in Databases (2012)
14. Liu, W., Zheng, Y., Chawla, S., Yuan, J., Xing, X.: Discovering spatio-temporal causal interactions in traffic data streams. In: ACM SIGKDD (2011)
15. Louail, T., Lenormand, M., Cantú, O.G., Picornell, M., Herranz, R., Frias-Martinez, E., Ramasco, J.J., Barthelemy, M.: From mobile phone data to the spatial structure of cities. arXiv preprint arXiv:1401.4540 (2014)
16. Peng, C., Jin, X., Wong, K.C., Shi, M., Liò, P.: Collective human mobility pattern from taxi trips in urban area. PLoS ONE (2012)
17. Randriamanamihaga, A., Côme, E., Oukhellou, L., Govaert, G.: Clustering the vélib origin-destinations flows by means of poisson mixture models. In: ESANN 2013 (2013)
18. Shahnaz, F., Berry, M.W., Pauca, V., Plemmons, R.J.: Document clustering using nonnegative matrix factorization. Information Processing & Management (2006)
19. Song, C., Qu, Z., Blumm, N., Barabási, A.L.: Limits of predictability in human mobility. Science (2010)
20. Vincent, E., Bertin, N., Badeau, R.: Harmonic and inharmonic nonnegative matrix factorization for polyphonic pitch transcription. In: IEEE ICASSP (2008)
21. Wang, B., Plumbley, M.D.: Musical audio stream separation by non-negative matrix factorization. In: Proc. DMRN summer conf (2005)
22. Wang, D., Pedreschi, D., Song, C., Giannotti, F., Barabási, A.L.: Human mobility, social ties, and link prediction. In: ACM SIGKDD (2011)
23. Wang, F., Lee, N., Hu, J., Sun, J., Ebadollahi, S.: Towards heterogeneous temporal clinical event pattern discovery: a convolutional approach. In: ACM SIGKDD (2012)

24. Zafeiriou, S., Tefas, A., Buciu, I., Pitas, I.: Exploiting discriminant information in nonnegative matrix factorization with application to frontal face verification. IEEE TNN (2006)
25. Zhang, M.L., Zhou, Z.H.: Multi-instance clustering with applications to multi-instance prediction. Applied Intelligence (2009)