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

Emeric Tonnelier[1], Nicolas Baskiotis[1], Vincent Guigue[1] and Patrick Gallinari[1]

***Abstract*— In the 20th century, most mobility studies were based on costly surveys with few samples; nowadays, the data from static and mobile sensors allow to track the habits of a massive number of citizens. However, the counterpart of sensors data is that they generally provide noisy and partial signals lacking semantic information: the purpose of each human activity captured by the sensor is unknown. Extracting this latent semantic information from raw sensors data is a challenging and crucial task. In this paper, a novel algorithm based on non negative matrix factorization (NMF) is proposed in order to extract precise and meaningful user temporal profiles from logs of smart card data in a transportation system. The proposed NMF based algorithm allows a natural and informative clustering of the profiles which can lead to semantic information on the mobility of the users. The approach is compared to 4 others algorithms and focuses on the human scale, indeed, individual profiles differ quite substantially from group profiles. Experiments are conducted on a 3 months dataset supplied by the STIF, the Parisian public transport authority.**

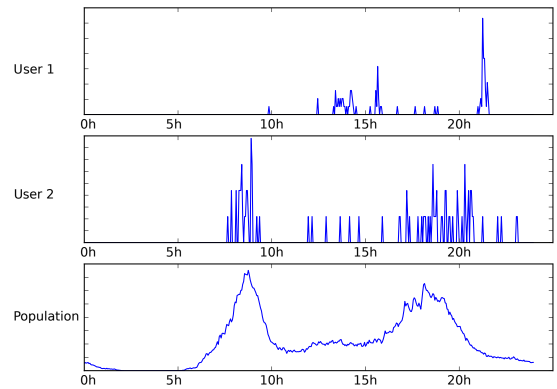

Fig. 1. Probability density functions of check-in for 2 users and the whole population. At the population scale, main habits appear clearly; At the individual level, peaks and discrepancies are different: more precision is required.

I. Introduction

Over the last decade, sensors data have been widely used in mobility studies. The growing availability of data, covering more and more urban spaces, led to major results (regarding the regularity of human mobility for instance [1], [2]). Other researches characterized the dynamics of the urban space: inference of Origin-Destination matrices for bike sharing networks [3], for road networks [4], or congestion detection [5]. All those studies aggregate individual information to model the average behavior of the population. However, zooming in on the behavior of a particular citizen with respect to the general model generally leads to generic information: the lack of semantics, the high level of noise in raw data and the excessive smoothing associated to the global averaging delete most fine level information. Thus, it is difficult to build a relevant analysis at both the individual and system levels. In this paper, this issue is tackled with an original approach consisting in extracting patterns at the population level for robustness (namely finding patterns or habits shared by large groups of users) and positioning those patterns at the individual scale to preserve a fine modeling.

We focus in this paper on data provided by smart card data of the French Parisian public transportation system (STIF). Each log corresponds to an entry in the transport network: it is made of a user id, a time stamp and a location information[1].

Smart card data have been widely used to characterize mobility according to temporal and spatial axes of analysis [6]. However, most of scientific literature is dedicated to a global and aggregated level (e.g. inference of origin-destination matrices [7], prediction of congestion [8]). Fewer studies focus on the user level given that smart card data is very noisy (due to the variability of human habits), incomplete (entrance only, data missing due to sensor failures) and lacks of semantics: labels like *going to work*, *family outing*, *seeing friends* are not available. On top of that, data generally concern a short period of time for privacy reasons (3 months in the presented use-case): only few samples are available for each user which penalizes statistical methods. Thus, most user level models tackle simple prediction tasks (for instance, next stop or next validation problem [9]).

The section II presents the formalization of the task and 4 usual models for the temporal profiling of users based on their smart card logs in the French Parisian public transportation network. The probability density functions of the entries of each user are considered and estimated using clustering or Gaussian mixture models. In

[1]Sorbonne Universités, UPMC Univ Paris 06, CNRS, LIP6 UMR 7606, 4 place Jussieu 75005 Paris. `firstname.name@lip6.fr`

[1]Exit information is not available, as in many smart card systems.

section III, a novel approach is proposed which consists in combining a pattern extraction approach at the group level and a personalized time shifted reconstruction to fit accurately habits of a given user. In section IV, the proposed approach is compared with the usual approaches and qualitative and quantitative analysis are proposed to assess its relevance. Finally, section V presents related works.

II. Baselines & formalism for user profiling

This section describes first the considered assumptions and the proposed formalism for the profiling task, then the investigated modelings and their specificities.

A. Data representation

The raw dataset of logs is made of triplets containing an anonymous user id, a time stamp and a location of an entry in the network : (id, t, ℓ). In this paper, only the temporal information is considered for the profiling, all spatial variables are aggregated as in [10]. In order to produce a first representation of the user temporal activity, the logs of each user are aggregated in a daily base of 24 hours; this time window is next discretized in 480 regular intervals of 3 minutes. Thus, each user u is represented by a vector $\mathbf{x}^{(u)} \in \mathbb{R}^T$, where $T = 480$ and $x_t^{(u)}$ corresponds to the percentage of his validations occurring in the time interval t. As a consequence, $\mathbf{x}^{(u)}$ sums to one and can be seen as discrete probabilistic density function (pdf) of the user's check-ins: Fig 1 illustrates the pdf of two users and the global activity of the network. Concatenating all user vectors leads to the matrix $X \in \mathbb{R}^{U \times T}$, where U denotes the number of users.

Figure 1 shows that an user pdf is very noisy and hardly exploitable at the user level. Dictionary learning algorithms are the most common approaches to tackle the decomposition of a collection of signals according to shared robust patterns under the constraint of a minimal reconstruction error. We propose in the following to compare different dictionary learning algorithms, where patterns are learned at different human scales (population, groups, individual person) in order to measure their ability to rebuild accurately individual users.

B. Evaluation

The experimental section IV will provide both qualitative and quantitative analysis. An important point concerns the evaluation of the considered approaches. The provided dataset includes 13 weeks of logs: 9 weeks will be used to train the various models and 4 weeks of virgin data will be used to evaluate the reconstruction ability for each user. Two quantitative indicators will be investigated: the mean squared error (MSE) and the mean likelihood (ML). The reconstruction ability of the studied models will be evaluated according to the MSE for each user, then according to the averaged MSE over the whole dataset. As explained previously we will distinguish a training error on the data used to learn the model and a test error, on the remaining 4 weeks.

The MSE based metrics are robust but hardly understandable. That is why a probabilistic measure will be also considered. In order to deal with outliers and unpredictable logs that hinder even robust likelihood measures[2], an unusual metric will be used: the mean likelihood (ML), by computing the average likelihood of the test set logs. This indicates how the modelings are able to predict future logs. As all outputs are normalized pdf, all measures can be compared.

C. Models & human scales

The first proposed baseline is the overall pdf estimated using all the users (Fig 1). In such naive approach, no particular pattern is extracted and the modeling is obviously too rough to capture different kinds of users.

A second widely used approach is the k-means clustering algorithm [11]. Given k -the number of clusters- and X as inputs, k profiles are extracted corresponding to homogeneous user groups. Each profile can be seen as a 24h-pattern, but users are represented by only one of them: the representation is not able to mix different habits during the day, it will poorly represent for instance users with same habits in the morning and different ones in the evening.

Non-negative Matrix Factorization techniques (NMF) received a lot of attention as a source decomposition algorithm [12]. A dedicated version has been implemented within the context of smart card analysis to extract localized pattern and to map population dataset on new axes [13]. Similarly to k-means, patterns are extracted at the user group level. The details are discussed in section III.

In order to focus on the individual user scale, a Gaussian Mixture Model is finally considered. First, a very large Gaussian dictionary combining 4 standard deviations (2, 12, 22 and 32 minutes) and 480 means (one per discreet interval) is built. Every user is decomposed in this base in a sparse manner using the $\mathcal{L}_1$ regularized LASSO algorithm [14]. In order to tackle efficiently the mapping step in a very noisy context, the Least Angle Regression solver is used and the number of Gaussian atom is limited to 5 per user. This approach performs actually no dictionary learning and extract predefined Gaussian behaviors: robustness is improved but no analysis is possible on the shape of the atoms. The next section presents a novel strategy able to perform relevant dictionary learning at the individual user scale.

III. Time Shift-NMF

First, this section presents matrix factorization approaches dedicated to the user profiling task; then it describes the proposed novel Time-Shift NMF algorithm.

[2]For a given user, when one check-in in the test dataset occurs during a time period with no training check-ins, the likelihood measure will be null.

A. *Non negative matrix factorization*

Matrix factorization approaches are widely used in unsupervised learning contexts when the objective is to decompose objects of interest into small atoms with the constraint that each atom is used for the reconstruction of the largest amount of objects [12]. This constraint guarantees a regularization of the retrieved atoms and ensures the generalization capability of the model. More specifically, non negative matrix factorization (NMF) introduces an additional constraint: objects are reconstructed as an additive composition of (positive) atoms, each atom describing a part of the object.

Formally, given a non negative representation matrix $X \in \mathbb{R}_+^{U\times T}$ of U objects along T dimensions (each value in matrix being positive), the goal of NMF is to find two non negative matrix W and D such that $X \approx \hat{X} = WD$ with $W \in \mathbb{R}_+^{U\times R}$ and $D \in \mathbb{R}_+^{R\times T}$. Each row $D_{j,.}$ of D corresponds to an atom and the value $w_{i,j}$ of the matrix W denotes the weight of the atom j for the object i. The i-th object, represented by the row $X_{i,.}$, is thus approximated by : $\hat{X}_{i,.} = W_{i,.}D = \sum_j w_{i,j} D_{j,.}$. The number of atoms R allows to control the compactness and the expressiveness of the representation: fewer atoms leads to kind of k-means algorithm while a larger number of atoms creates a potential risk of over-fitting, atoms becoming too specific and explaining few situations. Adding sparsity constraints to control over-fitting is classically done by penalizing $\|W\|$. Finally, additional constraints can be easily considered during the learning step, for instances to insure the normalization of the reconstructed profiles.

B. *Time Shift Nonnegative Matrix Factorization (TS-NMF)*

The main idea of the algorithm is to consider that the patterns in check-in distributions are shared among a category of users for a given activity, regardless of its precise time location : an user can go to work at 7:00 and an other at 7:30 depending on the trips duration, but the two pdf of their check-ins for this activity will have the same shape and variability if both users have the same kind of work. Thus, atoms definition becomes time independent: relevant shapes are learned without the noise associated to the slight time-shift between users.

The dictionary matrix $D \in \mathbb{R}_+^{R\times S}$, with $S << T$ corresponds to the time range encoded by atoms: S is assumed to be the maximal time interval where all the check-ins linked to the same activity can occur for a particular user. Let denote $\Phi \in \mathbb{R}_+^{U\times R}$ the phase matrix corresponding to the time position of each atom for each person and $f : \mathbb{R}_+^{R\times S} \times \mathbb{R}_+^{U\times R} \rightarrow \mathbb{R}_+^{U\times R\times T}$ the operator such that $f(D,\Phi)$ is a 3D-tensor containing each R shifted atoms in the original T time space for U users. To ensure that each atom represents a probabilistic distribution, each row of D is constrained to have unit norm. As in the usual NMF, the reconstruction of each user is done by weighting and summing up the atoms according to the weight matrix W. The new regularized loss is given by : $L(D,W,\Phi) = ||X - Wf(D,\Phi)||_F||$ so that $\forall i, \ \sum_j w_{ij} = 1$

Algorithm 1: TS NMF learning algorithm

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

$D, W, \Phi = init(X, R)$
> 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
  - **for** $a \in atoms$ **do**
    - $\Phi_{u,a} = minimizeLocalCost_t(\mathbf{x}_u, D_a)$
      > finding optimal time-shift t in a window of size α_Φ
    - $W_{u,a} = update\_W(\mathbf{x}_u, W_{u,.}, D, \Phi_{u,a})$
      > Simple gradient descent
    - $\mathbf{x}_u = \mathbf{x}_u - f(D_a, \Phi_{u,a})W_{u,a}$
      > Matching pursuit like update
- $D = update\_D(W, D, \Phi)$
- $D = centerAtoms(D)$
  > Centering procedure to make atoms comparable

Algorithm 1 describes the learning algorithm used to infer W, Φ and D: it is directly inspired from the matching pursuit algorithm [15]. A gradient descent algorithm is used to learn W and D matrices with respect to the loss function[3]. As the operator f is highly non-linear, a greedy optimization procedure is used to update the Φ matrix: for each user and each atom (in a descending entropy order to place meaningful atoms first), the algorithm looks at the best position close to the actual one which minimizes the local cost. Once an atom is placed, user data explained by this atom are removed from the user representation in order to focus on other parts of the signal.

IV. Experiments

A. *Data description and experimental setup*

The dataset used in this section is provided by the "Syndicat des Transports d'Île-de-France" (STIF). It

[3] Parameters like the number of iterations max_{iter} are chosen to achieve convergence

represents over 1 million logs of 10,000 randomly chosen users having a smart card (Navigo Card) over the 308 metro stations in Paris, France. Users have between 1 and 500 logs over 13 weeks. Data are processed according to section II-A leading to a matrix $X \in \mathbb{R}^{U \times T}$ of user check-in pdf estimates with $U = 10,000$ rows and $T = 480$ columns (time intervals of 3 minutes). Only the first 9 weeks of the data are considered in order to construct the training matrix X, the 4 remaining weeks are used for testing purpose.

In order to assess the performances of our approaches, we evaluate the results of the TS-NMF algorithm and the 4 baselines according to Mean Squared Error (MSE) and Mean Likelihood (ML) metrics presented in section II-B. Good performances correspond to a minimal MSE and a maximal ML. The k-means algorithm and the NMF model are set to use 16 clusters; the Gaussian Mixture Model (GMM) is detailed in section II-C. For the TS-NMF algorithm, 16 atoms are used, each of them describing a time interval of 3 hours ($S = 60$)[4].

B. Experimental Results

Table I presents the results in terms of MSE and ML measures for the training and the test sets. As expected, the general model is the worst : the mean likelihood is slightly better than 0.002, the ML score of a uniform random model[5]. The k-means results are better in terms of MSE but especially in terms of ML. More surprisingly, the NMF results are very close to those of k-means showing the incapacity of this model to fit the data accurately at the user level despite the high number of involved parameters. Table I shows that the two best models for our task are the GMM and the TS-NMF: ML scores of both models are very close; in terms of MSE, TS-NMF shows a better capacity to represent the training dataset with a very low MSE, but both approaches seem to over-fit reaching a degraded MSE on the test set. Finally, the low variance of the results computed in 5 independent runs guarantees for all models a very high stability.

Looking at the number of parameters of each model (2^{nd} column of the table I), NMF uses 10 times more parameters than k-means to achieve same results, which confirms the inability of NMF to fit the data and the human scale. TS-NMF uses only around two times more parameters than usual NMF but 50 times less than GMM for comparable results. The high number of parameters for the GMM model is due to the number of Gaussian functions contained into the dictionary and can be computationally problematic for large datasets. Although it can explain well observed data, it over-fits and obtains a highest testing cost. In summary, the proposed TS-NMF shows a good trade-off between performances and number of parameters.

Given the user temporal profile extraction task, qualitative analysis is an important part of the models' evaluations. Indeed, our main goal is to extract meaningful atoms in order to exploit them. Figure 2 presents the atoms extracted by classical NMF: atoms are highly correlated as most of them explain the same period of the day (morning and evening). The failure of the model to fit data is thus explained by the large human variability; the algorithm tends to focus on the few most energetic time intervals of the day and describes large groups of users at once, overestimating the individual variance of persons. Figure 3 presents the top 10 used atoms for the GMM models: they are all quite similar, with the same (low) variance and a slight time shift. It illustrates the difficulty of any semantic labeling based on this representation: similar behaviors are scattered among multiple atoms, hindering clustering and interpretability. The good news resides in the quantitative results: we verify that working at the individual scale leads to sharper pdf modeling. Figure 4 shows atoms obtained with TS-NMF[6]. The flexibility of our approach is fully exploited : each atom presents a different check-in distribution, from very punctual distribution (1^{st} atom) to distribution with a high late variance like the last one. Some TS-NMF atoms uses the entire 3 hours while other uses only a smaller part. Various atoms entropies discriminate diffuse activities from punctual ones.

Finally, to illustrate the interest of the phase matrix Φ in TS-NMF, figure 5 shows the histogram of the use of each atom with respect to the time in the day. Some atoms are localized in the morning (first ones), other in the afternoon (last ones), while last ones are used during a large time window. This fact confirms the capacity of our model to extract features characterizing latent activities. We are also able to describe a user according to the timing of his activities. For instance, it is possible to group users according to both mean and variance in work departure time.

Figure 6 shows for a random user the raw check-in density estimation (top) and the reconstruction inferred by the TS-NMF model, each atom being represented by a different color.

V. Related Work

Various approaches have been proposed for the extraction of mobility patterns. Clustering techniques similar to k-means have been used to characterize trips pattern [16] or temporality of the trips [17] but as shown in section IV these clustering approaches are not well suited for the reconstruction of each individual user profile. Probabilistic models have been used to capture habits of networks users [3], [10] but they focus essentially on the characterization at the system level. [13] proposes

[4] running time: Kmeans, NMF: 5 minutes, Gaussian Mixture Model: 10 minutes, TS-NMF: 4 hours.

[5] The uniform model has a probability of $1/480 = 0.002$ at each time interval and thus an average likelihood of 0.002.

[6] The time scale is noted same in this figure, as these atoms are defined for an interval of 3 hours.

Model	# param.	MSE -train- (mean (std))	MSE -test- (mean (std))
General model	0	0.033 (0)	0.040 (0)
KMeans (16 clusters)	15 339	0.027 (6.3e-6)	0.038 (1.5e-5)
GMM	14 628 690	0.023 (0)	0.050 (0)
NMF	130 224	0.024 (7.7e-5)	**0.036 (6.7e-5)**
TS-NMF	252 768	**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)	15 339	0.010 (6.3e-6)	0.008 (5.7e-6)
GMM	14 628 690	**0.027 (0)**	**0.018 (0)**
NMF	130 224	0.013 (8.3e-5)	0.009 (3.8e-5)
TS-NMF	252 768	**0.026 (9.3e-4)**	0.016 (4.8e-4)

TABLE I

Models optimized parameter number, reconstruction cost and test cost (MSE (less is better) and Likelihood (high is better)) mean and variance over 5 runs on the learn/test dataset

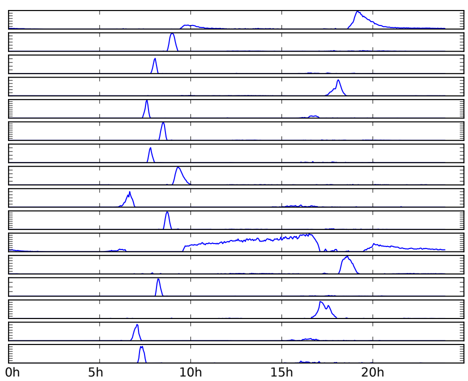

Fig. 2. NMF extracted atoms.

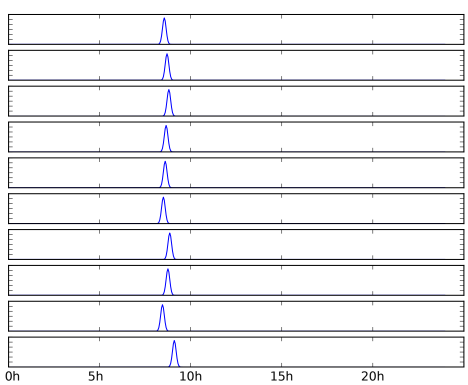

Fig. 3. Gaussian Mixture Model top 10 atoms.

a probabilistic NMF to mine transportation network logs but the modeling of check-in events with fixed distributions leads to a less flexible representation and prevents to discover unexpected behaviors. Other NMF approaches have been used in mobility studies for traffic modeling [18] and urban transportation network [19] but once more at the network level. Other latent models focus on the semantic extraction task - labeling trips, urban space, user activity or latent models [20], [21] - but they generally lack of reconstruction ability which hinders the analysis at an aggregated level or at the network level.

VI. Conclusion

In this paper, we proposed a dictionary learning algorithm based on matrix factorization for temporal profiling of transportation network users. The proposed approach has better modeling properties than state-of-the-art algorithms. The extracted atoms can be used for a wide range of applications : usual statistical analysis of the system, of the users habits, but also for latent semantic labeling of the trips or meaningful clustering based on the shape of the atoms.

Moreover, considering the compactness of the proposed user representation with respect to the whole user log (with marginal loss of information) leads to a powerful compression tool. Incorporating in a same manner spatial information - which are not taken into account in this paper - can lead to an efficient compact and complete representation of users network activity, helpful for compression purposes but also for semantic indexing of the users and the network.

Acknowledgment: This work was partially founded by the STIF. The authors gratefully acknowledge the support of the STIF staff regarding the exploitation and the analysis of the dataset.

References

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

[2] Sui Tao, David Rohde, and Jonathan Corcoran. Examining the spatial–temporal dynamics of bus passenger travel behaviour using smart card data and the flow-comap. 41:21–36, 2014.

[3] Andry Randriamanamihaga, Etienne Come, Latifa Oukhellou, and Gérard Govaert. Clustering the velib' origin-destinations flows by means of poisson mixture models. pages 273–278.

[4] Ennio Cascetta and Sang Nguyen. A unified framework for estimating or updating origin/destination matrices from traffic counts. *Transportation Research Part B: Methodological*, 22(6):437–455, 1988.

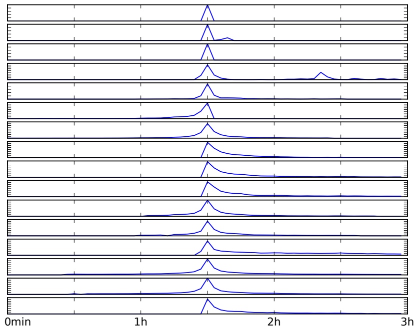

Fig. 4. TS-NMF Atoms - activity pattern shared by all users defined over 3 hours. Each is then shift for each user day aggregation representation.

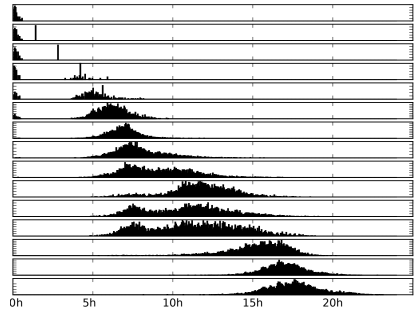

Fig. 5. Time shift distribution for each TS-NMF activity pattern.

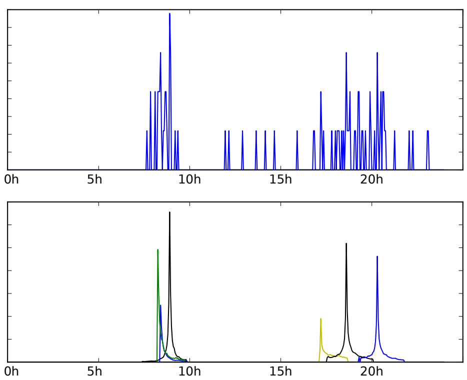

Fig. 6. Time Shift NMF reconstruction (bottom) and original data (top).

[5] Chengkun Liu, Kun Qin, and Chaogui Kang. Exploring time-dependent traffic congestion patterns from taxi trajectory data. In *2015 2nd IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services (ICSDM)*, pages 39–44.

[6] Marie-Pier Pelletier, Martin Trépanier, and Catherine Morency. Smart card data use in public transit: A literature review. 19(4):557–568.

[7] Marcela A Munizaga and Carolina Palma. Estimation of a disaggregate multimodal public transport origin–destination matrix from passive smartcard data from santiago, chile. *Transportation Research Part C: Emerging Technologies*, 24:9–18, 2012.

[8] Irina Ceapa, Chris Smith, and Licia Capra. Avoiding the crowds: understanding tube station congestion patterns from trip data. pages 134–141. ACM, 2012.

[9] Stefan Foell, Gerd Kortuem, Reza Rawassizadeh, Santi Phithakkitnukoon, Marco Veloso, and Carlos Bento. 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*, pages 1239–1248. ACM, 2013.

[10] A. S. Briand, E. Côme, M. K. E. Mahrsi, and L. Oukhellou. A mixture model clustering approach for temporal passenger pattern characterization in public transport. In *IEEE International Conference on Data Science and Advanced Analytics (DSAA), 2015. 36678 2015*, pages 1–10.

[11] Richard O Duda, Peter E Hart, and David G Stork. *Pattern classification.* John Wiley & Sons, 2012.

[12] Andrzej Cichocki, Rafal Zdunek, Anh Huy Phan, and Shun-ichi Amari. *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation.* John Wiley & Sons, 2009.

[13] Mickaël Poussevin, Emeric Tonnelier, Nicolas Baskiotis, Vincent Guigue, and Patrick Gallinari. Mining ticketing logs for usage characterization with nonnegative matrix factorization. In Martin Atzmueller, Alvin Chin, Frederik Janssen, Immanuel Schweizer, and Christoph Trattner, editors, *Big Data Analytics in the Social and Ubiquitous Context*, pages 147–164. Springer International Publishing, 2014.

[14] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[15] Stéphane G Mallat and Zhifeng Zhang. Matching pursuits with time-frequency dictionaries. *Signal Processing, IEEE Transactions on*, 41(12):3397–3415, 1993.

[16] Xiaolei Ma, Yao-Jan Wu, Yinhai Wang, Feng Chen, and Jianfeng Liu. Mining smart card data for transit riders' travel patterns. *Transportation Research Part C: Emerging Technologies*, 36:1 – 12, 2013.

[17] Juanjuan Zhao, Chen Tian, Fan Zhang, Chengzhong Xu, and Shengzhong Feng. Understanding temporal and spatial travel patterns of individual passengers by mining smart card data. In *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 2991–2997.

[18] Chengbin Peng, Xiaogang Jin, Ka-Chun Wong, Meixia Shi, and Pietro Liò. Collective human mobility pattern from taxi trips in urban area. 7(4):e34487, 2012.

[19] Yufei Han and Fabien Moutarde. Analysis of large-scale traffic dynamics in an urban transportation network using non-negative tensor factorization. *Int. J. Intelligent Transportation Systems Research*, 14(1):36–49, 2016.

[20] Sébastien Gambs, Marc-Olivier Killijian, and Miguel Núñez del Prado Cortez. Next place prediction using mobility markov chains. In *Proceedings of the First Workshop on Measurement, Privacy, and Mobility*, page 3. ACM.

[21] Ji Yuan, Xudong Liu, Richong Zhang, Hailong Sun, Xiaohui Guo, and Yanghao Wang. Discovering semantic mobility pattern from check-in data. In *Web Information Systems Engineering–WISE 2014*, pages 464–479. Springer, 2014.