Extracting Commuting Patterns in Railway Networks through Matrix Decompositions

Shashank Jere*, Justin Dauwels*, Muhammad Tayyab Asif*, Nikola Mitrovic*, Andrzej Cichocki† and Patrick Jaillet‡§
*School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
†Laboratory for Advanced Brain Signal Processing, RIKEN, Brain Science Institute, Japan
‡Laboratory for Information and Decision Systems, MIT, Cambridge, MA, USA
§Center for Future Urban Mobility, Singapore-MIT Alliance for Research and Technology, Singapore

***Abstract*—With the rise in the population of the world's cities, understanding the dynamics of commuters' transportation patterns has become crucial in the planning and management of urban facilities and services. In this study, we analyze how commuter patterns change during different time instances such as between weekdays and weekends. To this end, we propose two data mining techniques, namely Common Orthogonal Basis Extraction (COBE), and Joint and Individual Variation Explained (JIVE) for Integrated Analysis of Multiple Data Types and apply them to smart card data available for passengers in Singapore. We also discuss the issues of model selection and interpretability of these methods. The joint and individual patterns can help transportation companies optimize their resources in light of changes in commuter mobility behavior.**

I. Introduction

Due to the advancements in sensor technologies and rapid expansion of transportation infrastructure in modern cities such as Singapore, there have been rapid developments in the field of Data-Driven Intelligent Transportation Systems (D^2ITS) [1]. Our aim is to extract dominant commuter mobility patterns and track changes in these patterns during different time periods. With the aid of such analysis, pubic transportation companies can optimize their resources in a better manner by tracking changes in mobility patterns during different time instances (e.g., morning and evening rush hours, or between weekdays and weekends).

Although it is possible to apply different techniques to obtain travel patterns based on any criterion, our focus shall be to analyze the variations between the travel patterns observed on weekdays and on weekends. In particular, we are interested in knowing the "joint" patterns, which are observed in both cases, and "individual" patterns, which can only be seen either during a weekday or a weekend, and are thus, unique. To extract these common and individual patterns, we propose two techniques, Joint and Individual Variation Explained (JIVE) [2] and Common Orthogonal Basis Extraction (COBE) [3] and apply them to smart card (EZ-Link) data obtained from the passengers using the Mass Rapid Transit (MRT) services in Singapore. The data was provided by Singapore's Land Transport Authority (LTA). We have focused on the formulation of a model showing the macro-level travel patterns of passengers using the MRT in Singapore using the EZ-Link smart card data.

A significant number of studies on the topic of origin destination (OD) matrices for traffic networks have been performed in the past. While some have formulated a unified framework for estimating or updating OD matrices based on traffic counts for a road network [4], others have focused on doing the same for congested road networks [5]. The topic of inferring passenger travel patterns from smart card data of rail networks in urban regions has also been studied in some detail in the past. One particular example is the case study in Shenzhen, China, in which Liu et al. [6] analyzed individual and collective passenger mobility patterns. However, their analysis relied heavily on empirical observations and could not make a strong distinction between common and individual spatial or temporal patterns. Li et al. [7] introduced the idea of clustering nearby stations and proposed models for identifying morning, afternoon and evening peak patterns. This study focused on using smart card data records resolved in both space and time to study the collective spatial and temporal mobility patterns. It also touched on studying travel patterns at the individual passenger level and their regularity. However, as in [6], Li et al. [7] did not make a clear distinction between patterns that are "common" and "individual". In some studies, clustering algorithms were applied to obtain spatio-temporal patterns [8]. The focus, here, however, was mainly on temporal patterns, and consequently spatial patterns received little attention, which are of particular importance. In conclusion, none of the above studies have shed light on joint and individual spatial patterns across different days of the week, or even different times of the day. We address this issue, by developing common and individual low-dimensional models for urban rail road networks by applying JIVE and COBE on smart card data.

Both JIVE and COBE function in similar ways on the raw data but with subtle differences, to generate joint (common) and individual (unique) features, depending on the type of data sets being dealt with. In our application, we apply these methods to raw passenger travel data to arrive at spatial passenger travel patterns between different days of the week. We also look at issues concerning interpretability of the results obtained, in addition to factoring in model complexity by using the Bayesian Information Criterion (BIC) as a model selection criterion.

The remainder of the paper is structured as follows. In section II, we introduce the data set used. In section III, we briefly explain the mathematical models of JIVE and COBE and how they are applied to our case. In section IV, we look at methods to optimize the techniques used in terms of model accuracy and model complexity. Finally, in section VI, we summarize our contributions, and suggest topics for future work.

II. Data Set

According to LTA, a journey is defined as a set of rides on bus and train from the origin to the destination. It may involve one or more rides. Rides are considered to belong to one journey when they fulfil the transfer condition for the Distance-based journey fare. In the data provided, one record is one ride. We represent the smart card (EZ-Link) data as origin-destination (OD) pair matrices whose rows correspond to origin MRT stations, while their columns correspond to destination MRT stations. We assigned a unique index number (starting from 1) to each MRT station in the network, which serves as an identifier for that station in the OD pair matrices. Every element of an OD pair matrix shows the number of passengers travelling from the origin MRT station corresponding to the specific row to the destination MRT station corresponding to that particular column. At the time the data was acquired, which is April 2011, there were 107 functional MRT stations, and thus, we constructed 107 x 107 sized square matrices for two days, Monday (11^{th} April 2011) representing weekdays, and Sunday (17^{th} April 2011) representing weekends. We chose these two days to look at specific differences between passenger traffic on weekdays and weekends.

III. Patterns in Origin-Destination (OD) pair data

Many fields of scientific research today deal with high-dimensional data, one of them being the field of transportation. These data tend to include multiple high-dimensional data sets measured for a common set of objects. In our case, this translates to the OD pair matrices containing passenger travel counts for 107 MRT stations during two separate days. Given the relatively periodic and regular nature of transportation patterns in large metropolitan cities [9], it is reasonable to expect shared patterns between multiple data sets, in particular, the travel patterns between different days of the week, or between different time slots of the day, and so forth. We shall refer to such shared patterns as joint structure. It is also expected that a particular subset of data would have certain variation unrelated to other data sets. We shall refer to such individual variation as individual structure. We will now briefly explain the models of JIVE and COBE in the following subsections.

A. Joint and Individual Variation Explained (JIVE)

JIVE separates joint and individual passenger travel patterns between data sets by decomposing a data set as a sum of three terms as follows:

(a) A low-rank approximation capturing “joint” passenger travel patterns common to all the days.
(b) A low-rank structure capturing “individual” passenger travel patterns unique to a particular day.
(c) Residual noise that falls under neither joint or individual patterns.

The individual structure can identify potentially useful information that exists in one set of variables, but not others. Accounting for the individual structures also implies more accurate estimation of the joint structure. JIVE assumes the data sets to be arranged in the form of multiple matrices, $\mathbf{X_1}$, $\mathbf{X_2}$,$\cdots$,$\mathbf{X_k}$, where $k \geq 2$. All matrices have n columns each, corresponding to a common set of n objects. The i^{th} matrix has p_i rows, with each row corresponding to a variable. Applying this definition to our case, we can have $\mathbf{X_1}, \mathbf{X_2}, \cdots, \mathbf{X_k}$ represent the OD pair matrices for different days of the week. In this paper, we limit our analysis to only two days of the week; Monday (weekday) and Sunday (weekend). Therefore, the data sets in concern are only $\mathbf{X_1}$ and $\mathbf{X_2}$. As OD pair matrices are square, we have $p_i = n$ in our case. The two matrices may be combined into a single matrix $\mathbf{X}$ of size $p \times n$ as

$$\mathbf{X} = \begin{bmatrix} \mathbf{X_1} \\ \mathbf{X_2} \end{bmatrix}. \tag{1}$$

Joint structure is represented by a single $p \times n$ matrix $\mathbf{J}$ of rank $r < \min\{rank(\mathbf{X_1}), rank(\mathbf{X_2})\}$ defined as

$$\mathbf{J} = \begin{bmatrix} \mathbf{J_1} \\ \mathbf{J_2} \end{bmatrix}. \tag{2}$$

Individual structure for each $\mathbf{X_i}$ is represented by a $p_i \times n$ matrix of rank $r_i < rank(\mathbf{X_i})$. Let $\mathbf{A_i}$ be the sub-matrix of the individual structure matrix $\mathbf{A}$, representing the individual structure of $\mathbf{X_i}$, such that

$$\mathbf{A} = \begin{bmatrix} \mathbf{A_1} \\ \mathbf{A_2} \end{bmatrix}, \tag{3}$$

and let $\mathbf{J_i}$ be the sub-matrix of the joint structure matrix $\mathbf{J}$ that is associated with $\mathbf{X_i}$. The JIVE model for two data sets can then be described as

$$\begin{aligned} \mathbf{X_1} &= \mathbf{J_1} + \mathbf{A_1} + \varepsilon_\mathbf{1}, \\ \mathbf{X_2} &= \mathbf{J_2} + \mathbf{A_2} + \varepsilon_\mathbf{2}, \end{aligned} \tag{4}$$

for the two data matrices in our case. Here $\varepsilon_\mathbf{i}$ are error matrices with independent entries, i.e., $E(\varepsilon_\mathbf{i}) = \mathbf{0}_{p_i \times n}$. The orthogonality constraint $\mathbf{J}\mathbf{A}_i^T = \mathbf{0}_{p \times p_i}$ ensures that patterns responsible for joint structure between data sets are unrelated to patterns responsible for individual structure [2].

Various rank selection criteria are discussed in [2] and [10]. JIVE estimates joint and individual structures by minimizing the Frobenius Norm of the Residual Structure matrix $\mathbf{R}$

$$\mathbf{R} = \begin{bmatrix} \mathbf{R_1} \\ \mathbf{R_2} \end{bmatrix} = \begin{bmatrix} \mathbf{X_1} - \mathbf{J_1} - \mathbf{A_1} \\ \mathbf{X_2} - \mathbf{J_2} - \mathbf{A_2} \end{bmatrix}, \tag{5}$$

by iteratively applying Alternating Least Squares (ALS) [2]. The Frobenius norm of the matrix $\mathbf{R}$ (r_{ij}) is defined as

$$||\mathbf{R}||_F = \sqrt{\sum_{i,j} r_{ij}^2}.$$

B. Common Orthogonal Basis Extraction (COBE)

JIVE requires intermediate matrices to be stored during every iteration of the process. This involves extra cost in terms of memory. This problem can be tackled with another approach termed as Common Orthogonal Basis Extraction (COBE).

The raw data of interest is contained in matrices $\mathbf{X_n} \in \mathbb{R}^{I\times M_n}$; $n \in \{1,2\}$, for the two OD pair matrices, with $\mathbf{X_1}$ containing the raw OD pair data for Monday (11$^{\text{th}}$ April 2011), representative of weekdays and $\mathbf{X_2}$ containing that for Sunday (17$^{\text{th}}$ April 2011), representative of weekends. COBE aims to solve the following factorization problem for each matrix

$$\min_{\mathbf{P_n},\ \mathbf{Q_n}} \quad ||\mathbf{X_n} - \mathbf{P_n}\mathbf{Q_n}^T||_F;\ n \in \{1,2\}.$$

The columns of matrix $\mathbf{P_n} \in \mathbb{R}^{I\times R_n}$ contains the latent variables in $\mathbf{X_n}$, while matrix $\mathbf{Q_n} \in \mathbb{R}^{M_n\times R_n}$ is the corresponding co-efficient matrix; R_n represents the number of latent components. Although COBE imposes the restriction $R_n < I$, it is more than likely to be obeyed in our case since we are computing low-rank approximations of the actual data. The probability of this being true is further increased when we consider that the latent rank is often significantly lower than the dimensionality of the actual data [11].

COBE considers the data sets $\mathbf{X_1}$ and $\mathbf{X_2}$ to be naturally linked sharing some common components. For this purpose, $\mathbf{P_n}$ is partitioned as [3]

$$\mathbf{P_n} = \left[\bar{\mathbf{P}}\ \breve{\mathbf{P}}_\mathbf{n}\right], \tag{6}$$

with $\bar{\mathbf{P}} \in \mathbb{R}^{I\times c}$ and $\breve{\mathbf{P}}_\mathbf{n} \in \mathbb{R}^{I\times(R_n-c)}$. The variable c is a positive integer defined as

$$c \le \min\{R_1, R_2\},$$

where c can be described as the number of common components shared by matrices $\mathbf{X_1}$ and $\mathbf{X_2}$. With the above partitioning of $\mathbf{P_n}$ and a compatible partitioning of $\mathbf{Q_n^T}$, each data matrix can be factorized as

$$\begin{aligned} \mathbf{X_n} &= \left[\bar{\mathbf{P}}\ \breve{\mathbf{P}}_\mathbf{n}\right]\begin{bmatrix}\bar{\mathbf{Q}}_\mathbf{n}^\mathbf{T}\\ \breve{\mathbf{Q}}_\mathbf{n}^\mathbf{T}\end{bmatrix}, \\ &= \bar{\mathbf{P}}\bar{\mathbf{Q}}_\mathbf{n}^\mathbf{T} + \breve{\mathbf{P}}_\mathbf{n}\breve{\mathbf{Q}}_\mathbf{n}^\mathbf{T}, \\ &= \mathbf{J_n} + \mathbf{A_n}, \end{aligned} \tag{7}$$

with $n \in \{1,2\}$. Each matrix $\mathbf{X_n}$ is represented by two parts: the common space $\mathbf{J_n}$ spanned by the common components (columns of $\bar{\mathbf{P}}$), and the individual space $\mathbf{A_n}$ spanned by the individual components (columns of $\breve{\mathbf{P}}_\mathbf{n}$). The objective of COBE is therefore to seek $\bar{\mathbf{P}}$ and $\breve{\mathbf{P}}_\mathbf{n}$ without the knowledge of $\mathbf{Q_n}$ and possibly the number c. The optimization problem earlier defined now becomes [3]

$$\begin{aligned} \min_{\mathbf{P_n},\ \mathbf{Q_n}} \quad & \sum_{n\in\{1,2\}} ||\mathbf{X_n} - \bar{\mathbf{P}}\bar{\mathbf{Q}}_\mathbf{n}^\mathbf{T} - \breve{\mathbf{P}}_\mathbf{n}\breve{\mathbf{Q}}_\mathbf{n}^\mathbf{T}||_F^2, \\ \text{subject to:} \quad & \bar{\mathbf{P}}^\mathbf{T}\bar{\mathbf{P}} = \mathbf{I_c}, \\ & \breve{\mathbf{P}}_\mathbf{n}^\mathbf{T}\breve{\mathbf{P}}_\mathbf{n} = \mathbf{I}_{R_n\times c}, \\ & \bar{\mathbf{P}}^\mathbf{T}\breve{\mathbf{P}}_\mathbf{n} = \mathbf{0},\ \text{(Orthogonality Constraint)}. \end{aligned} \tag{8}$$

The above orthogonality constraint is similar to that in JIVE. The restriction that $rank(\bar{\mathbf{P}}) + rank(\breve{\mathbf{P}}_\mathbf{n}) = R_n$ implicitly guarantees that we are seeking common components [3]. In order to facilitate easier comparison with JIVE, we use the variant "COBEc" of the COBE algorithm, which assumes that the number of common components c is known beforehand. This allows a certain control over the model, similar to the control provided by rank r of the joint structure in JIVE.

Additionally, it can be shown [3] that

$$\bar{\mathbf{Q}}_\mathbf{n} = \mathbf{X_n^T}\bar{\mathbf{P}}. \tag{9}$$

With $\bar{\mathbf{P}}$ and thus the common space $\bar{\mathbf{J}}_\mathbf{n}$ computed by COBEc, the individual space $\mathbf{A_n}$ can be calculated from a rank k truncated-SVD (tSVD) of the residual structure $\mathbf{X_n} - \mathbf{J_n}$, where $k < R_n - rank(\mathbf{J_n})$.

C. Equivalence between JIVE and COBEc

We can draw a certain degree of similarity between the parameters used in JIVE and those used in COBEc. While the free parameters in JIVE are the rank r of joint structure $\mathbf{J}$ and ranks r_1 and r_2 of individual structures $\mathbf{A_1}$ and $\mathbf{A_2}$ respectively, those in COBEc are the number of common components c and the rank k of the truncated-SVD (tSVD) used in computing the respective individual spaces. We shall use this equivalence while implementing the two algorithms and comparing the results.

TABLE I: Equivalence between JIVE and COBEc

Model: $\mathbf{X_n}$=$\mathbf{J_n}$+$\mathbf{A_n}$+$\varepsilon_\mathbf{n}$			
JIVE		**COBEc**	
Nomenclature	**Notation**	**Nomenclature**	**Notation**
Joint structure	$\mathbf{J}$	Common space	$\mathbf{J}$
Ind. structure	$\mathbf{A}$	Ind. space	$\mathbf{A}$
Joint structure rank	r	Common components	c
Ind. structure ranks	r_1, r_2	t-SVD rank	k
$\mathbf{J}$ and $\mathbf{A}$ computed simultaneously		$\mathbf{J}$ computed first followed by each $\mathbf{A_i}$	

IV. Model Selection

Both the algorithms JIVE and COBEc aim to minimize the sum of squared error (SSE), which is $||\mathbf{X} - \mathbf{J} - \mathbf{A}||_F^2$ in both the models. However, they do not account for the complexity of the model formed as a result of the choice of the free parameters mentioned in the previous section. In this section, we use Bayesian Information Criterion (BIC) to account for model complexity and strike a balance between model accuaracy and complexity for both JIVE and COBEc. We shall go on to select optimal values of free parameters (r, r_1, r_2) in the JIVE model and (c, k) in the COBEc model, based on the results of BIC.

A. Factoring in Model Complexity-JIVE

The JIVE algorithm has a pre-requisite of specifying the appropriate values of the joint structure rank r and the individual structures ranks r_1 and r_2. Each combination of these ranks is essentially a model to be chosen or discarded. BIC calculates a score for each model, and states that the optimum model can be identified as the one having the lowest score [12]. We compute the BIC score for a model according to

[13], under the assumption that model errors are independent and follow a normal distribution. The definition of the BIC score for the JIVE model is given as

$$\text{BIC}_{\text{JIVE}} = n \ln\left(\frac{\text{SSE}}{n - p_{\text{JIVE}} - 1}\right) + p_{\text{JIVE}} \ln(n). \quad (10)$$

Let us define $m = 107$, which is the number of MRT stations in the network. Then, n, which is the total number of data points considered, can be defined as

$$n = 2m^2 = 2(107^2) = 22,898.$$

It should be noted that n is the total number of elements in both $\mathbf{X_1}$ and $\mathbf{X_2}$. The Sum of Squared Error (SSE) is defined as

$$\text{SSE} = ||\mathbf{X} - \mathbf{J} - \mathbf{A}||_F^2, \quad (11)$$

where $\mathbf{X}$, $\mathbf{J}$ and $\mathbf{A}$ are defined in (1), (2) and (3) respectively. We define p_{JIVE}, which is a measure of the number of free parameters (model complexity), as

$$p_{\text{JIVE}} = 2mr + mr_1 + mr_2 = m(2r + r_1 + r_2).$$

Although r_1 and r_2 may have different positive integer values generally, we cosider $r_1 = r_2$ in order to limit the number of free parameters to two, i.e., r and $r_i = (r_1 = r_2)$.

We are interested in finding low-rank approximations for the available OD pair data. Therefore, we shall restrict the joint structure rank r as well as the individual ranks r_i from taking values above a certain threshold. Considering that the data matrices $\mathbf{X_i}$ are of size 107 $\times$ 107, we shall place the following restriction on r and r_i

$$\begin{aligned} r &\leq 40, \\ r_i &\leq 40, \;\; i = \{1, 2\}, \end{aligned} \quad (12)$$

so that the joint structure and individual structure ranks are limited to less than half the full-rank value. BIC scores are then computed for all possible models formed with the above limits, i.e., for $1 \leq r \leq 40$ and $1 \leq r_i \leq 40$. We plot the scores as a 3-dimensional surface plot in Fig. 1.

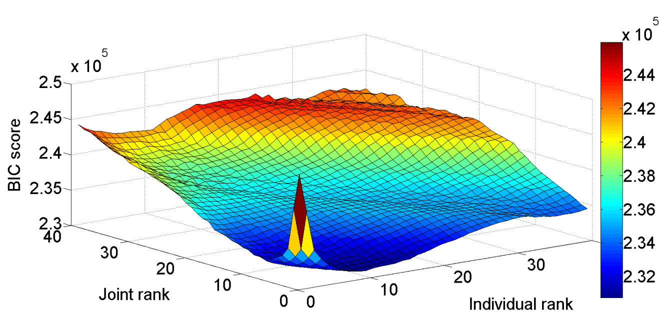

Fig. 1: Surface plot of BIC scores for JIVE model with rank restrictions.

The model with the lowest BIC score is considered the optimal model out of all the available models [12]. We found that the lowest BIC score occurs at rank values $r = 1$, $r_1 = r_2 = 10$. Consequently, we implement JIVE on the raw data contained in matrices $\mathbf{X_1}$ and $\mathbf{X_2}$ with these rank values.

B. *Factoring in Model Complexity-COBEc*

The two free parameters used in the COBEc model are c and k. We use BIC to facilitate the selection of optimal values for these parameters, as we did for JIVE. The BIC score definition for COBEc is given as

$$\text{BIC}_{\text{COBEc}} = n \ln\left(\frac{\text{SSE}}{n - p_{\text{COBEc}} - 1}\right) + p_{\text{COBEc}} \ln(n). \quad (13)$$

The Sum of Squared Error (SSE) for the COBEc model is defined in the same manner as the JIVE model, as in (11). The variable p_{COBEc} is the measure of the number of free parameters and is defined as

$$p_{\text{COBEc}} = 2mc + 2mk = 2m(c + k).$$

Similar to the restrictions on the values of the ranks in the JIVE model, we place the following restrictions on the values of c and k, since we are interested in the low-rank approximations of the OD pair data

$$\begin{aligned} c &\leq 40, \\ k &\leq 40. \end{aligned} \quad (14)$$

We calculate BIC scores for $1 \leq c \leq 40$ and $1 \leq k \leq 40$. The surface plot of the BIC scores is shown in Fig. 2.

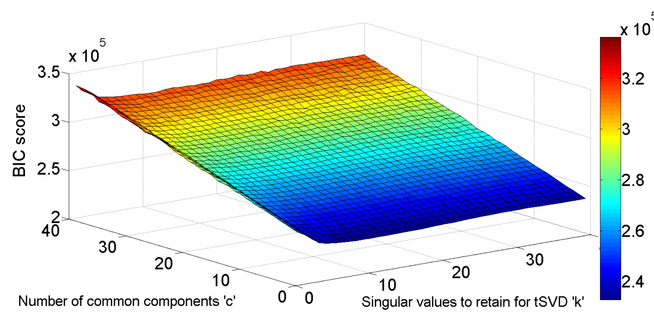

Fig. 2: Surface plot of BIC scores for COBEc model with parameter restrictions.

As in the JIVE model, we search for parameter values yielding the lowest BIC score [12]. We found that $c = 1$ and $k = 11$ produces the lowest BIC score with the above restrictions holding. Therefore, we implemented COBEc with these parameter values,

However, we observed that the COBEc implementation results in $\mathbf{J}$ containing a large number of negative values, which pose interpretability issues. This issue was encountered in JIVE as well, but the negative values observed were much fewer and their magnitudes were much smaller. In order to tackle this problem in the COBEc case, we applied Non-Negative Matrix Factorization (NMF) instead of the normal Singular Value Decomposition (SVD) approach used in COBEc. This modified algorithm (Algorithm 1) is briefly presented below.

Here, $\mathbf{W_{c_i}}$ $(n \times c)$ and $\mathbf{H_{c_i}}$ $(c \times n)$ are low-rank non-negative matrices factored out of the original data matrices $\mathbf{X_i}$ $(n \times n)$ using a suitable Non-Negative Matrix Factorization algorithm [14], [15]. For a general matrix $\mathbf{Y}$ of size $n \times m$, the function

Algorithm 1 Modified COBEc algorithm with NMF

1. $[\mathbf{W_{c_1}}, \mathbf{H_{c_1}}] = nnmf(\mathbf{X_1}, c)$
2. $[\mathbf{W_{c_2}}, \mathbf{H_{c_2}}] = nnmf(\mathbf{X_2}, c)$
3. $\mathbf{J_1} = \mathbf{W_{c_1}} \mathbf{H_{c_1}}$
4. $\mathbf{J_2} = \mathbf{W_{c_2}} \mathbf{H_{c_2}}$
5. $\mathbf{A_1} = tSVD(|\mathbf{X_1} - \mathbf{J_1}|, k)$
6. $\mathbf{A_2} = tSVD(|\mathbf{X_2} - \mathbf{J_2}|, k)$

$nnmf(\mathbf{Y}, c)$ factors $\mathbf{Y}$ into non-negative factors $\mathbf{W}$ $(n \times c)$ and $\mathbf{H}$ $(c \times m)$. In our implementation, we use the alternating least-squares algorithm for implementing Non-Negative Matrix Factorization. The function $tSVD(\mathbf{Z}, p)$ returns the closest rank p approximation to $\mathbf{Z}$ using the p largest singular values and associated singular vectors of matrix $\mathbf{Z}$. Using the above approach, the severity of the problem of interpretability of negative values in $\mathbf{J}$ and $\mathbf{A}$ in the COBEc model is significantly reduced.

V. Results and Discussion

In this section, we analyze the passenger travel patterns obtained from the results of implementation of both JIVE and COBEc. We analyze passenger traffic patterns on Monday using the data provided by LTA for 11th April 2011 as a representative for weekday travel patterns, and on Sunday using the data for 17th April 2011 as a representative for weekend travel patterns. In particular, we are interested in knowing the travel patterns that occur during both the days and are hence "joint" or "common", and the travel patterns that occur during one of the days but not the other, and are thus "individual" or "unique" to that day.

We implement JIVE on the raw OD pair data contained in matrices $\mathbf{X_1}$ and $\mathbf{X_2}$ using $r = 1$ and $r_1 = r_2 = 10$. For visualization, we set a threshold value of 1000 for the number of passengers travelling from any origin station to any destination station. In other words, we are interested in knowing which routes have more than 1000 passengers travelling along them via MRT over the course of the day. Depending on whether such a large value is observed in $\mathbf{J}$ or $\mathbf{A}$, the pattern is accordingly classified as a "joint" (occurring in both $\mathbf{J_1}$ and $\mathbf{J_2}$) pattern or an "individual" (occurring in one of $\mathbf{A_1}$ or $\mathbf{A_2}$) pattern.

As an example, we analyze the travel patterns for three major stations, namely, Boon Lay, Tampines, and Harbour Front. Considering these stations as "sources" or origins of high volume passenger traffic, we find the important routes by examining those elements in $\mathbf{J}$ and $\mathbf{A}$ with values of more than 1000 in the row corresponding to that particular source station. In this way, uni-directional passenger travel patterns for an entire day can be obtained. This results in travel patterns maps, as shown in Figs. 3, 4, 5 and 6. The blue arrows in Fig. 3 and Fig. 5 represent uni-directional travel patterns common to both weekdays and weekends. On the other hand, in Fig. 4 and Fig. 6, the blue arrows represent uni-directional travel patterns unique to weekdays, while the yellow arrows represent travel patterns unique to weekends. In all figures, the red dots represent the three chosen major traffic sources, while the black dots represent the major passenger destinations or "sinks".

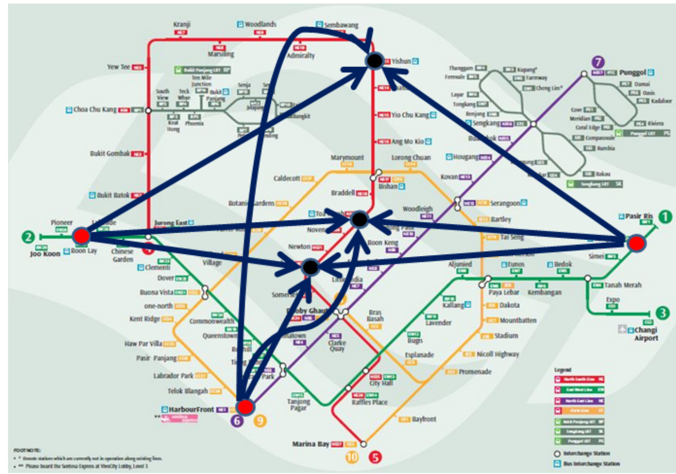

Fig. 3: Passenger travel patterns common to both weekdays and weekends (blue arrows): JIVE results.

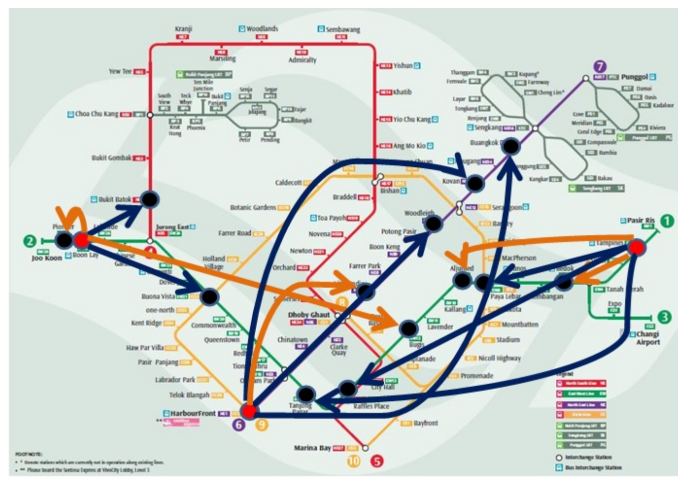

Fig. 4: Passenger travel patterns unique to weekdays (blue arrows) and to weekends (yellow arrows): JIVE results.

A quick comparison between the individual day-wise patterns indicated by JIVE and those indicated by COBEc with NMF show a close resemblance. The differences between them are the omission of some patterns in COBEc while a change in the "sink" of some, although a few patterns may be added or lost. This is an assuring fact since it implies that the patterns obtained are stable and do not change substantially with a change in the technique or method used.

The patterns have some intuitive sense as well. For instance, the centrally located stations such as Toa Payoh and Orchard were found to be popular destinations on both weekdays and weekends. Travel patterns unique to Monday (weekdays) indicate a high volume of traffic (>1000) to industrial and commercial hubs such as Bukit Batok and Raffles Place,

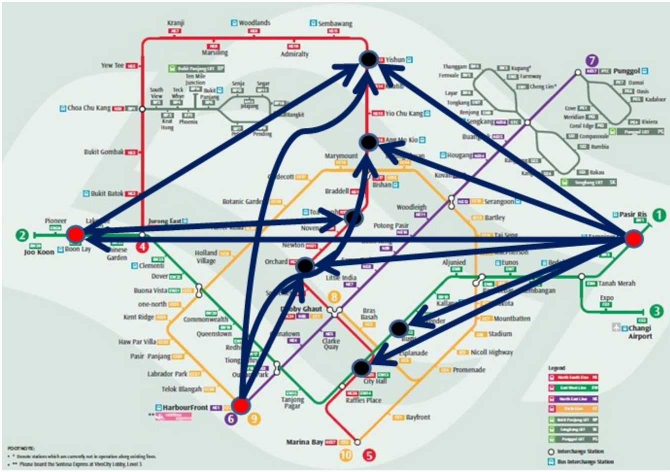

Fig. 5: Passenger travel patterns common to both weekdays and weekends (blue arrows): COBEc results.

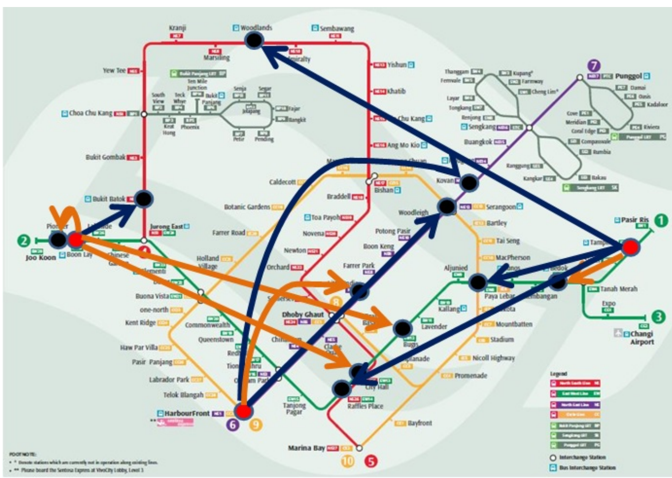

Fig. 6: Passenger travel patterns unique to weekdays (blue arrows) and to weekends (yellow arrows): COBEc results.

typical of a working day routine. Travel patterns unique to Sunday (weekends) indicate a high volume of traffic (>1000) to shopping and leisure destinations such as Bugis and City Hall, expected of a holiday routine. Yishun was observed as a heavy traffic destination on both the days. Thus, the travel patterns deciphered follow basic intuition, and can be qualitatively justified in that sense.

VI. Conclusion and Future Work

Although JIVE and COBEc have similarities, they have certain differences in their methodologies. Therefore, a reasonable degree of similarity between the travel patterns obtained from JIVE and COBEc was expected. This assumption was found to be true when the patterns returned by JIVE and COBEc were found to be similar.

In summary, both JIVE and COBEc provide useful tools for quick discovery of travel patterns hidden inside high-dimensional data. However, the interpretability issues posed in COBEc and to some degree in JIVE due to negative values in output matrices need to be effectively addressed. It is possible to extend this analysis to other temporal patterns, i.e., travel patterns specific to a certain time period of the day, e.g. morning and evening rush hours or the afternoon period, etc. Factoring in time of travel will also aid in identifying such cases where more than required transport resources are being put to use during a particular time and instead transfer them to a time slot where greater resources are needed.

Acknowledgment

The research described in this paper was funded in whole or in part by the Singapore National Research Foundation (NRF) through the Singapore-MIT Alliance for Research and Technology (SMART) Centre for Future Urban Mobility (FM).

References

[1] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation systems: A survey," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 12, no. 4, pp. 1624–1639, 2011.

[2] E. F. Lock, K. A. Hoadley, J. Marron, and A. B. Nobel, "Joint and individual variation explained (jive) for integrated analysis of multiple data types," *The annals of applied statistics*, vol. 7, no. 1, pp. 523–542, 2013.

[3] G. Zhou, A. Cichocki, and S. Xie, "Common and individual features analysis: beyond canonical correlation analysis," *arXiv preprint arXiv:1212.3913*, 2012.

[4] E. Cascetta and S. Nguyen, "A unified framework for estimating or updating origin/destination matrices from traffic counts," *Transportation Research Part B: Methodological*, vol. 22, no. 6, pp. 437–455, 1988.

[5] H. Yang, T. Sasaki, Y. Iida, and Y. Asakura, "Estimation of origin-destination matrices from link traffic counts on congested networks," *Transportation Research Part B: Methodological*, vol. 26, no. 6, pp. 417–434, 1992.

[6] L. Liu, A. Hou, A. Biderman, C. Ratti, and J. Chen, "Understanding individual and collective mobility patterns from smart card records: A case study in shenzhen," in *Intelligent Transportation Systems, 2009. ITSC'09. 12th International IEEE Conference on.* IEEE, 2009, pp. 1–6.

[7] M. Li, B. Du, and J. Huang, "Travel patterns analysis of urban residents using automated fare collection system," in *ITS Telecommunications (ITST), 2012 12th International Conference on.* IEEE, 2012, pp. 442–446.

[8] C. Morency, M. Trépanier, and B. Agard, "Analysing the variability of transit users behaviour with smart card data," in *Intelligent Transportation Systems Conference, 2006. ITSC'06. IEEE.* IEEE, 2006, pp. 44–49.

[9] G. W. Hilton, "Rail transit and the pattern of modern cities: the california case," *Traffic Quarterly*, vol. 21, no. 3, 1967.

[10] P. R. Peres-Neto, D. A. Jackson, and K. M. Somers, "How many principal components? stopping rules for determining the number of non-trivial axes revisited," *Computational Statistics & Data Analysis*, vol. 49, no. 4, pp. 974–997, 2005.

[11] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[12] E. Wit, E. v. d. Heuvel, and J.-W. Romeijn, "all models are wrong...: An introduction to model uncertainty," *Statistica Neerlandica*, vol. 66, no. 3, pp. 217–236, 2012.

[13] M. B. Priestley, *Spectral analysis and Time Series.* London: Academic Press, 1981.

[14] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in neural information processing systems*, 2000, pp. 556–562.

[15] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons, "Algorithms and applications for approximate nonnegative matrix factorization," *Computational Statistics & Data Analysis*, vol. 52, no. 1, pp. 155–173, 2007.