This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2021.3072642, IEEE Transactions on Knowledge and Data Engineering

Missing Value Imputation for Multi-view Urban Statistical Data via Spatial Correlation Learning

Yongshun Gong, *Member, IEEE*, Zhibin Li, Jian Zhang*, *Senior Member, IEEE*, Wei Liu, *Senior Member, IEEE*, Yilong Yin*, and Yu Zheng, *Fellow, IEEE*

Abstract—As a developing trend of urbanization, massive amounts of urban statistical data with multiple views (e.g., views of Population and Economy) are increasingly collected and benefited to diverse domains, including transportation service, regional analysis, etc. Unfortunately, these statistical data that are divided into fine-grained regions usually suffer from missing value problem during the acquisition and storage processes. It is mianly caused by some inevitable circumstances, e.g., the document defacement, statistical difficulty in remote districts, and inaccurate information cleaning, etc. Those missing entries which make valuable information invisible may distort the further urban analysis. To improve the quality of missing data imputation, we propose an improved spatial multi-kernel learning method to guide the imputation process incorporating with the adaptive-weight non-negative matrix factorization strategy. Our model takes into account the regional latent similarities and the real geographical positions as well as the correlations among various views that are able to complete missing values precisely. We conduct intensive experiments to evaluate our method and compare with other state-of-the-art approaches on real-world datasets. All the empirical results show that the proposed model outperforms all the other state-of-the-art methods. Additionally, our model represents a strong generalization ability across multiple cities.

Index Terms—Missing data imputation, Spatial data, Statistic data, Multi-view

1 Introduction

Urbanization's rapid progress has currently modernized many people's lives. The generated statistic data play an irreplaceable role in a large number of city development and social services, e.g., regional planning, urban computing, failure detection, and transportation management [1]–[5]. These statistical data record various types of information that usually contain multi-fold views (e.g., views of Family, Income, Population and Business). Such statistics reveal the growth gaps among different administrative regions from various perspectives. Fig. 1 gives an example of the regional statistics. The Business area r_2 contains four views. Among them, the economy view records the key economic indicators for fine-grained regions[1], such as employee statistics and the number of industries; and the family view consists of detailed family information of all living sizes.

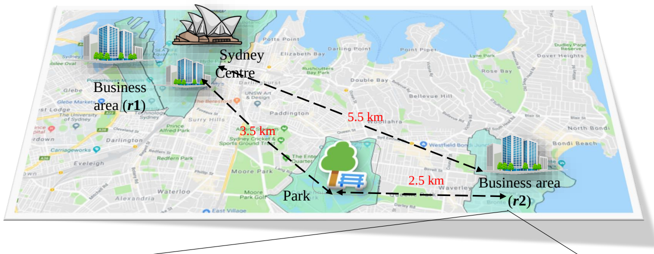
The statistic data provide key statistics to governments, business and the community on social science, for the benefit of many aspects of human life[2]. However, in some places, statistical data are hard to be entirely acquired due to document defacement, error recordings, and statistician misplay, resulting in data-missing and sparsity problem. Such missing values hide useful information which may cause distorted results for further analysis. For example, in the point of interest recommendation problem [6], economy and population are two of the most important influence factors of determining area functions. A business area should have a high-quality economy and a larger population. If the economy attributes were missing, this area would be grouped into the residential district because the solely consideration of population. To the best of our knowledge, to present, it is an under-studied field concerning on this specific problem. Yet it is a real-world demand from the national government. Accordingly, an effective missing-data imputation method for urban statistical data should be devised, which is important for the reliable urban computing and government services.

In this paper, we explore the missing-data imputation problem in urban statistical datasets collected from the Australian Bureau of Statistics (ABS) and the New Zealand Stats (NZS)[3]. Such a missing value imputation task for the multi-view urban statistical data is much more difficult than completing missing values for other datasets as this type of problem has some unique challenges:

Spatial Correlation Mining. The statistical data focusing on fine-grained regions may change over locations significantly and non-linearly. Even though the First Law of Geography emphasizes that everything is related to everything else, but near things are more related than distant ones [7], the potential similarities need to be considered when analyzing the spatially related data [8], [9]. Therefore, to properly recover the missing information of statistical data, we need to consider the regional similarities. As illustrated in Fig. 1,

Yongshun Gong and Yilong Yin are with the School of Software, Shandong University, Jinan, China. (e-mail: {ysgong,ylyin}@sdu.edu.cn).
Zhibin Li is now with the Commonwealth Scientific and Industrial Research Organisation, Australia. (e-mail: Zhibin.Li@csiro.au).
Jian Zhang and Wei Liu are with the Faculty of Engineering and Information Technology University of Technology Sydney, NSW 2007, Australia. (e-mail: {Jian.Zhang,Wei.Liu}@uts.edu.au).
Yu Zheng is a Vice President of JD.COM and the Chief Data Scientist of JD Digits. Beijing, China (e-mail: msyuzheng@outlook.com)
**Corresponding authors.*

1. The fine-grained area partition is based on the Australian Bureau of Statistics (ABS) standard. https://www.abs.gov.au/
2. https://www.abs.gov.au/about?OpenDocument&ref=topBar
3. https://www.stats.govt.nz/



Economy View	Total number of businesses	Total number of businesses entries	Total number of businesses exits	Value of total building	...
	11,843	2,027	1,496	615	...
Family View	Total households	Total families	Average Family Size	Separate house	...
	7417	6074	3	3203	...
Income View	Mean Employee income	Mean Investment income	Mean Superannuation and annuity income	Median Employee income	...
	56923	7808	26200	53489	
Population View	Person Total	Working Age Population %	Persons/km2	Australian citizen %	...
	132,733	67.8	3909.1	81.8	...

Fig. 1: Urban statistical data illustration and the regional similarity. *First, the tabular data give an example of the property of statistics; second, to properly recover the missing information of statistical data, we need to consider the regional similarities, as represented in the map figure.*

the properties of business area b_1 are similar to the 'Sydney center' because they are neighboring each other. However, although the business area r_2 is closer to the park area in terms of the physical distance, the statistical data of r_2 are more analogous to the 'Sydney center' than the 'Park' because they have a similar functional property (business centre).

Multi-view Problem. The complicated underlying interactions suggest that simply recovering the missing information without considering the correlations among attributes and multi-modes will end up with a poor performance [10]. For example, the economy view has strong correlations with the income and population views, so that a high-quality economy in a region usually goes along with a better income and a larger population; and a low-level economy in a region has a high probability of being connected with a lower income and a smaller population. In this case, only considering per view separately does not utilize the relationships between attributes and multi-modes. Therefore, how to integrate the whole views into a unified model is a principal challenge need to be solved.

Missing Temporal Information. In this specific problem, almost all the missing statistic values in the current year were also missing in the past years, which may be caused by the region restriction and complicated human-made errors. Based on the real phenomenon, the temporal dimension is unavailable. Besides, this violates the basic assumption of matrix completion [11] that the unobserved entries are sampled uniformly at random. Thus matrix completion-based approaches may not work in this case. Note that, this challenge is not solely a matter for statistical data but appears in other fine-grained spatio-temporal data mining problems [12], [13], where the temporal information in some fine-grained areas cannot be collected at any time.

In the current literature, the existing missing data imputation approaches, while yielding good estimations for missing values in the single view, are poor in terms of completing missing data on multiviews, especially when associated with the spatial characteristic. For example, numerous approaches can be applied in urban statistical data, e.g., mean-filling (MF), k-nearest-neighbor (KNN) filling [14], and collaborative filtering based methods [15]. Most of them, however, have been proposed to focus on the single view problem. Besides, although several spatiotemporal methods can infer the missing information based on the knowledge from both spatial and temporal domains [8], [16], [17], they do not perform well when the missing temporal information challenge appears. To handle all challenges well, we devise a model via the spatial correlation learning. In detail, the method integrates a spatial multi-kernel clustering method and an adaptive-weight non-negative matrix factorization[4] (NMF) method for solving the multi-view spatially related tasks. We summarize the main contributions and innovations of this paper as follows:

- To address the multi-view problem with spatial characteristics, we design a Spatially related Multi-Kernel K-Means (S-MKKM) approach to identify the underlying relationships among multi-fold views and capture the regional similarities.
- We propose an adaptive-weight non-negative matrix factorization method to leverage the information learned above to tackle the multi-view missing data imputation problem. Besides, the proposed method also takes guidance from the single-view and the real geographic information with *KNN* strategy into consideration.
- A multi-view missing data imputation method for urban statistical data via spatial correlation learning named SMV-NMF is proposed. SMV-NMF does not rely on the temporal information but achieves a great performance by only using the spatial information.
- We perform a collection of experiments on six real-world datasets to prove the effectiveness of our method compared with other state-of-the-art models. All the evaluation results show that the proposed method SMV-NMF yields the best performance. Furthermore, SMV-NMF shows the strong generalization ability that can transfer the constructed model from one urban dataset to another well.

This paper is an extended version of [18], the novelty and improvements of this paper are: we leverage an improved adaptive-weight matrix factorization strategy to control the knowledge extraction, provide more detailed theoretical analyses and conduct intensive experiments for the comprehensively evaluation from many perspectives. The remainder of this paper is organized as follows: Section 2 includes a literature review. Section 3 formally defines the problem and shows the preliminary methods. Our model is proposed in Section 4. All experimental results are shown in Section 5. Finally, conclusions are drawn in Section 6.

2 Related Work

In this section, we review the current studies on the spatio-temporal missing data imputation. Then provide the multi-view model discussions.

4. We add the non-negativity constraint since almost all the urban statistic data are non-negative.

2.1 Missing Data Imputation for Spatio-temporal Data

Missing data imputation is a significant task for data analysis [19]. It aims at filling out the data with estimation value. In the spatially related problem, neighborhood and collaborative filtering [8], [20] based methods are two kinds of dominant approaches in missing data filling. Although some classical methods (e.g., zero-filling, mean value filling, regression models) can be applied to the spatial missing data imputation, they have disadvantages in nature, i.e, they are not designed for this spatial problem. [21] used the inverse distance weighting (IDW) method to interpolate the spatial rainfall distribution. [22] utilized the spatial information as inputs in a residual kriging method to estimate the average monthly temperature. Unlike the spatial model, some successful spatio-temporal models were proposed for use with time stream data [8], [9], [16], [17]. For example, [8] develops a spatio-temporal multi-view approach (ST-MVL) to collectively complete missing values in a collection of geosensory time series data. It considers that 1) the temporal correlations between readings at various time spans in the same series and 2) the spatial correlations between different time series. However, they focused on filling missing entries by considering both spatial and temporal properties, and would not perform well on the static spatial data without the temporal information. Furthermore, these discussed methods leveraged the spatial guidance but did not consider the problem on multi-view datasets.

Here we also discuss other spatio-temporal missing data imputation methods used in the real-time applications. For a large city network, we are faced with a truth that data is not everywhere, especially in the real-time system, there may be not enough time to collect complete data.

A naive way is to average the values near missing data. Research by [23] used the simple average method to impute the missing data. As they said, this was mainly because the missing ratios for the selected sensors are sufficiently low. However, when faced with large-scale traffic network, the number of missing data is probably huge, and the average method cannot be adopted.

Many traffic prediction methods incorporate missing data completion into prediction steps. [24] dealt with missing data by 'expanded Bayesian network'. They made use of the causal relations in traffic network, and constructed the network by replacing the missing data with its causal variables. The main shortage for this method is that once the structures and parameters of the Bayesian network is trained, the relative position and time for missing data is also fixed. This is usually unreal, because data is likely to be missed at different time and sites. In other words, one model can only handle with one case of missing data. If we are facing a real traffic network, it is impossible for us to enumerate every condition and train for each condition a model.

[25] proposed a data completion method by matrix factorization. The Traffic data were structured as matrix with each entry X_{ij} denotes traffic speed between node i and node j. Based on the non-negative matrix tri-factorization framework, they got the latent attribute matrix of nodes and the attribute interaction matrix. By minimising the known error together with constraint using Laplacian matrix [26], missing data completion was accomplished by reconstructing data matrix with factor matrices.

Tensor decomposition [27] was used by [28] to complete missing computer network traffic data. They used a weighted optimization version of CP decomposition to impute the missing data. [29] improve this method through Tucker decomposition. They got a comparatively accurate result even when the missing ratio of data was quite high (up to 75%). These methods organised data as a three-way tensor, with day mode, hour mode and interval mode. [30] used this method to floating car data and get a better coverage of traffic state. In this paper, data is organised as a three-way tensor of link mode, interval mode and day mode. Similar method was used in research of [31]. They treated data to be predicted as missing data, and trained the decomposition model with historical data as rough prediction. There are two main problems of these tensor-based methods. First, they can only deal with one road or several road segments at a time, which is not enough for a citywide traffic network. Second, they did not define a rule for choosing the ranks for tensor decompositions, but the rank is one of the most crucial parameters for tensor decompositions.

2.2 Multi-view Learning

We discuss the multi-view studies because the missing-data usually contain the multiply views. Multi-view learning methods involved the diversity of different views that can jointly optimize functions based on various feature subsets [32], [33]. [34] proposed a matrix co-factorization based method (MVL-IV) to embed different views into a shared subspace, such that the incomplete views can be estimated by the information on observed views. To connect multiple views, MVL-IV assumes that different views have distinct 'feature' matrices (i.e., $\{H_i\}_{i=1}^{d}$), but correspond to the same coefficient matrix (i.e., W). However, it does not exploit the spatial correlations and may suffer from the imbalance problem, i.e., if there is a substantial missing ratio gap between views, the coefficient matrix W is mostly learned from the dense view. The proposed method has addressed this weakness by introducing guidance matrices. Another widely used strategy for solving the multi-view problem is tensor factorization [35], [36], but this restricts a regular tensor that requires the number of dimensions per view to be the same. Moreover, multiple kernel learning with incomplete views [37], [38] only focuses on completing missing kernels instead of filling missing values. To the best of our knowledge, none of the above studies considered both spatial and multi-view problems. Hence, this paper proposes an effective missing value imputation model for multi-view urban statistical data.

3 Problem Description and Preliminaries

Before clarifying our model, we firstly introduce some basic notations, operations and algorithms used in this paper. The main symbols used in this paper are summarized in Table 1.

3.1 Problem Description

Focusing on the multi-view missing-value problem, each set of urban statistical data includes multi-fold views, e.g.,

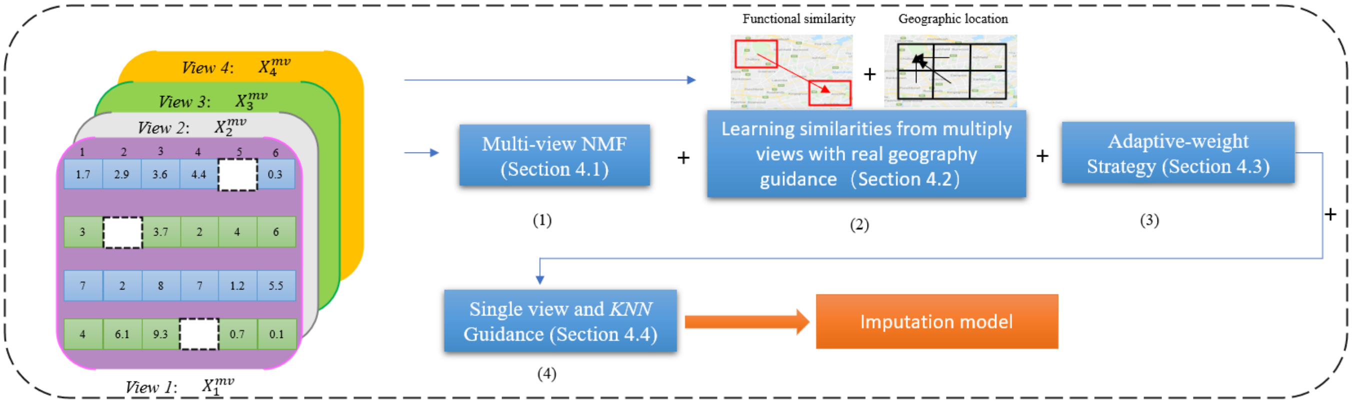


Fig. 2: The flowchart of our proposed method. In the learning process, given a set of multi-view urban statistical data, we first use the muti-view NMF as a basic imputation method, then we propose a multi-view spatial similarity guidance with adaptive-weight strategy to build guidance X_p^{mv}. Next, SMV-NMF also considers single-view and real geographic locations. Finally, the target missing value can be inferred by Algorithm 1.

TABLE 1: Symbol description.

Symbols	Descriptions
$X = [X_1 X_2 ... X_d]$	original data matrix contains d views
$W; H_p$	latent space matrices, p indicates the p-th view
$Y_p; \bar{Y_p}$	indication matrices for all complete entries and missing entries of p-th view
$k; l$	the number of dimensions of latent space; and the number of clusters
$n; d; m_p$	the number of regions; and the number of views; the dimension of attributes in the p-th view
$Z_p; Z_p^{'}$	weight matrices of the p-th view
L	graph Laplacian matrix
V	the clustering matrix
$X^{mv}; X^{sv}; X^{knn}$	three guidance matrices
$K_{\boldsymbol{\beta}}; \boldsymbol{\beta}$	the kernel matrix; the coefficients of kernels
$\lambda_1; \lambda_2; \lambda_3; \alpha$	regularization parameters

Income, Population, Economy views, etc. As shown in Fig. 3, given the incomplete statistic dataset with n regions (r_1,...,r_n) and d views, where the dimension of attributes in the p-th view is m_p ($1 \leq p \leq d$), this paper aims to find the interactions among views and fill missing entries precisely.

3.2 Non-negative Matrix Factorization (NMF)

In the missing data imputation problem, the non-negative matrix factorization method decomposes the original matrix $X \in \mathbb{R}_+^{n\times m}$ into two matrices $W \in \mathbb{R}_+^{k\times n}$ and $H \in \mathbb{R}_+^{k\times m}$, where W and H represent the latent spaces. In our problem, W indicates the latent features of the regions; H indicates the latent features of the data view. Each column in these matrices represents k attributes of corresponding regions and statistical fields. The interaction between these attributes determines the statistical value between the regions and statistic fields. Therefore, the basic missing data imputation model based on NMF can be described as the following optimization objective:

$$\min_{W\geq 0, H\geq 0} ||Y \odot (X - W^\top H)||_F^2 \tag{1}$$

where $W^\top$ is the transposed matrix of W; Y is an indicator matrix whose entry (i, j) is one if $X(i, j)$ has been recorded (for observed values) and zero otherwise (for missing values); $||.||_F$ is the Frobenius norm of matrix; and $\odot$ is the Hadamard product (entry-wise product) operator.

3.3 Multiple Kernel K-means (MKKM)

The original data X is combined by d data views ($X = [X_1 ... X_d]$) as shown in Fig. 3. Let $\{\mathbf{x}_i\}_{i=1}^n$ be a collection of n samples (region), $\mathbf{x}_i$ represents the statistical features of the i-th region, and $\phi_p(\cdot)$ be the p-th view mapping that maps $\mathbf{x}$ onto the p-th reproducing kernel Hilbert space. In this case, each sample has multiple feature representations defined by a group of feature mappings $\phi_{\boldsymbol{\beta}}(\mathbf{x}_i) = [\beta_1\phi_1(\mathbf{x}_i)^\top, \cdots, \beta_d\phi_d(\mathbf{x}_i)^\top]^\top$, where $\boldsymbol{\beta} = [\beta_1, \cdots, \beta_d]^\top$ consists of the coefficients of the d base kernels. These coefficients will be optimized during learning. Based on the definition of $\phi_{\boldsymbol{\beta}}(\mathbf{x})$, a kernel function can be expressed as $\boldsymbol{\kappa}_{\boldsymbol{\beta}}(\mathbf{x}_i, \mathbf{x}_j) = \phi_{\boldsymbol{\beta}}(\mathbf{x}_i)^\top \phi_{\boldsymbol{\beta}}(\mathbf{x}_j) = \sum_{p=1}^d \beta_p^2 \kappa_p(\mathbf{x}_i, \mathbf{x}_j)$. And a kernel matrix $K_{\boldsymbol{\beta}}$ is then calculated by applying the kernel function $\boldsymbol{\kappa}_{\boldsymbol{\beta}}(\cdot,\cdot)$ to $\{\mathbf{x}_i\}_{i=1}^n$. Based on the kernel matrix $K_{\boldsymbol{\beta}}$, the objective of MKKM can be written as:

$$\begin{aligned} &\min_{V,\beta} \mathrm{Tr}(K_\beta(\mathbf{I}_n - VV^\top)) \\ &s.t.\ V \in \mathbb{R}^{n\times l}, V^\top V = \mathbf{I}_l, \beta^\top \mathbf{1}_d = 1, \beta_p \geq 0, \forall p, \end{aligned} \tag{2}$$

where V is the clustering matrix; $\mathbf{1}_d \in \mathbb{R}^d$ is a column vector with all 1 elements; $\mathbf{I}_n$ and $\mathbf{I}_l$ are identity matrices with size n and l; l is the number of clusters.

4 The Proposed Method

In this section, we propose our missing data imputation model SMV-NMF. We will describe how to address the multi-view problem, and how to capture the spatial correlation via the multiple kernel learning. Fig. 2 shows the flowchart of our model.

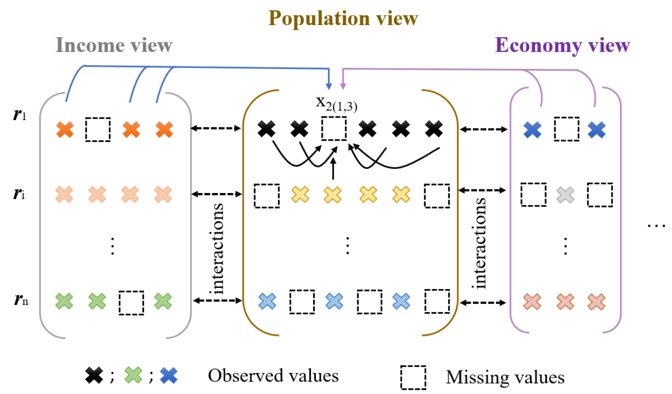


Fig. 3: Problem description. *For an urban dataset containing n regions (r_1,...,r_n), our method aims to impute the missing values with a high accuracy. For example, the imputation process for $x_{2(1,3)}$ will both consider the internal (feature similarities) and external knowledge (view interactions) to fill the target entry.*

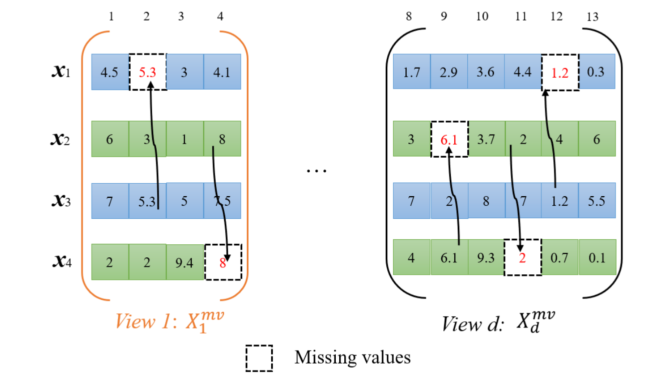


Fig. 4: An example of generating X_1^{mv} and X_d^{mv}. *Assume that regions $\mathbf{x}_1$ and $\mathbf{x}_3$ are falling into one cluster with the blue background, and $\mathbf{x}_2$ and $\mathbf{x}_4$ belong to another cluster with green background. $\mathbf{x}_2$ and $\mathbf{x}_3$ are the centroid regions of two clusters, respectively. For a missing entry x_{12}, its corresponding value x_{32} is used as an imputation guide. Moreover, if the value in centroid region is missing, then a greedy strategy is implemented to find the nearest observed value (use x_{49} to fill x_{29}).*

4.1 Multi-view NMF

The multi-view NMF is an effective method to handle the multi-view problem. In our problem, we aim to learn a latent subspace $W \in \mathbb{R}_+^{n\times k}$ by multiple views $\{X_1 ... X_d\}$ through the multi-view generation matrices $H_p \in \mathbb{R}_+^{k\times m_p}$. In this case, the basic missing data imputation model can be described as the following optimization objective:

$$\min_{W\geq 0, H_p \geq 0} \mathcal{N} = \sum_{p=1}^{d} ||Y_p \odot (X_p - WH_p)||_F^2, \tag{3}$$

where Y_p are indicator matrices for the p-th view whose entry $Y_p(i,j)$ is one if $X_p(i,j)$ has been recorded (for observed values).

We utilize the multi-view NMF method to find the potential connections among views. One of the advantages of non-negative constraint is the reasonable assumptions of latent characters and interpretability of the results [39]–[41]. Furthermore, due to the fact of urban statistical data, the missing values must be non-negative, thus W and H should be constrained into non-negative field.

4.2 Multi-view Spatial Similarity Guidance

In general, the multi-view matrix factorization based methods usually suffer from the imbalance problem as discussed in Section 2.2. To address this problem, we propose the similarity guidance X_p^{mv} for the p-th view X_p in this paper. To extract associations among different views of spatially related data, we devise a method to capture regional similarities via the spatial multiple kernel learning, named S-MKKM. The basic idea is that the development of a city gradually fosters different functional groups, such as educational and business districts, where the regions belonging to the same group would have strong connections with each other [3]. S-MKKM leverages the multi-kernel k-means (MKKM) clustering algorithm combined with a graph Laplacian dynamics strategy (an effective smoothing approach for finding spatial structure similarity [25], [41]–[43]) to cluster regions into the functional groups. Specifically, we construct a graph Laplacian matrix L, defined as $L = D - M$, where M is a graph proximity matrix that is constructed from the regional physical topology, i.e., $M_{(i,j)}$ = 1 if and only if the region $\mathbf{x}_i$ is contiguous to $\mathbf{x}_j$, and D is a diagonal matrix $D_{(i,i)} = \sum_j (M_{(i,j)})$. With this constraint, the S-MKKM model can be formulated as follows:

$$\begin{aligned} &\min_{V, \boldsymbol{\beta}} \ \mathrm{Tr}(K_\beta(\mathbf{I}_n - VV^\top)) + \alpha \mathrm{Tr}(V^\top L V) \\ &s.t.\ V \in \mathbb{R}^{n\times l}, V^\top V = \mathbf{I}_l, \boldsymbol{\beta}^\top \mathbf{1}_d = 1, \beta_p \geq 0, \forall p, \end{aligned} \tag{4}$$

where α is the regularization parameter; V is the consensus clustering matrix.

To get the complete kernels, we initially impute the missing data for each view by a simple method, such as KNN or MF (the effects of different initializations are illustrated in Section 5.7). Therefore, this problem can be solved by alternately updating V and $\boldsymbol{\beta}$ [38]:

i) Optimizing $\boldsymbol{V}$ with fixed $\boldsymbol{\beta}$. With the kernel coefficients $\boldsymbol{\beta}$ fixed, $\boldsymbol{V}$ can be obtained by the following strategy:

$$V \leftarrow choose\ the\ l\ smallest\ eigenvectors\ of\ (-K_{\boldsymbol{\beta}} + \alpha L) \tag{5}$$

ii) Optimizing $\boldsymbol{\beta}$ with fixed V. With V fixed, $\boldsymbol{\beta}$ can be optimized via solving the quadratic programming with linear constraints:

$$\beta_p = \frac{tr(K_p(\boldsymbol{I} - VV^\top))^{-1}}{\sum_{p=1}^{d}(tr(K_p(\boldsymbol{I} - VV^\top))^{-1})} \tag{6}$$

The objective of the S-MKKM method is to discover the regions with similar properties and build the guidance matrices X_p^{mv}. We employ the clustering result to build a matrix X_p^{mv}, that is after having gotten V, X_p^{mv} can be built. Fig. 4 illustrates an example of this process. The construction process of X_p^{mv} is that **i)** for the unknown entry x_{ij}, and the region $\mathbf{x}_i \in c$-th cluster, we use its corresponding value $x_{c(i),j}$ from the centroid region to impute x_{ij}; **ii)** if the corresponding value of centroid region is also missed, a

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2021.3072642, IEEE Transactions on Knowledge and Data Engineering

greedy strategy will be used to find the nearest observed value for imputation.

4.3 Adaptive-Weight NMF

To learn the knowledge from X_p^{mv} more reliably, we propose an adaptive weighting strategy in the NMF imputation process. The adaptive-weight matrix of the p-th view is denoted as $Z_p \in \mathbb{R}_+^{n\times m_p}$, which is built by an exponential function as shown in Equation (7) and (8).

$$z_{p(i)} = e^{-Dist(\mathbf{v}_i, \mathbf{v}_{c(i)})}, \tag{7}$$

$$Z_p = z_p \mathbf{1}_{m_p}^{\top}, \tag{8}$$

where $Dist(\cdot,\cdot)$ is the Euclidean distance calculating from the geo-location ($\mathbf{v}_i$) and its corresponding centroid region ($\mathbf{v}_{c(i)}$), here we use the latent embedding $\mathbf{v}_i$ to represent the geo-location of region i, and $\mathbf{v}_{c(i)}$ represents the centroid of the c-th cluster which contains region $\mathbf{v}_i$; $z_p \in \mathbb{R}_+^n$ is a column vector and $\mathbf{1}_{m_p}$ is all-ones vector with size m_p. It is not a straight way for imputation, but the adaptive-weight matrix Z_p controls how much information can be extracted. Z_p adjusts the penalty of each estimated entry. As emphasised in the First Law of Geography [7], the near things have more spatial correlations than distant things. If the distance between $\mathbf{x}_i$ and $\mathbf{x}_{c(i)}$ is small, we want a high penalty to guide the imputation process.

To this stage, the adaptive-weight NMF can be described as the following optimization function:

$$\min_{W\geq 0, H_p\geq 0} \mathcal{M} = \sum_{p=1}^{d} ||\bar{Y_p} \odot Z_p \odot (X_p^{mv} - WH_p)||_F^2, \tag{9}$$

where $\bar{Y_p} = \mathbf{1} - Y_p$, $\mathbf{1}$ is an all one matrix that has the same size as Y_p; X_p^{mv} is a homomorphic matrix of X_p.

We consider the adaptive-weight NMF jointly, then we can get:

$$\min_{W\geq 0, H_p\geq 0} \mathcal{J}_1 = \mathcal{N} + \lambda_1 \mathcal{M}, \tag{10}$$

where λ_1 is the regularization parameter to control the learning rate of X_p^{mv}.

4.4 Improved by Single-view and KNN Guidances

S-MKKM aims to find the regional groups by considering multiple views simultaneously. However, it is obvious that each view has its characteristics, and the relationships between regions in one specific view are also critical for imputing missing entries. To consider the above knowledge, we apply the spatially related kernel k-means (S-KKM) to capture the similarities among regions of each view. It is essentially analogous to the learning process of S-MKKM as discussed in Section 4.2, but considering each view, respectively. For one view X_p, the S-KKM model is expressed as follows:

$$\begin{aligned} \min_{V_p} \; & \mathrm{Tr}(K_p(\mathbf{I}_n - V_p V_p^{\top})) + \alpha \mathrm{Tr}(V_p^{\top} L V_p) \\ & s.t.\ V_p \in \mathbb{R}^{n\times l}, V_p^{\top} V_p = \mathbf{I}_l, \end{aligned} \tag{11}$$

where K_p is one separate kernel and V_p represents the p-th clustering matrix based on X_p.

In fact, to reduce the complexity of our model, we assume that the physical location affects the clustering performance with the same degree and the number of clusters is the same as that in S-MKKM, i.e., l and α are the same as used in Equation (4). The reason behind this assumption is that most cities have the same functional regions, such as the residential region and business region. Thus, it is reasonable that we choose the same α and l in this practical task. Besides, α and l are very stable due to the intrinsic property of the urban statistical data, and we fixed them in the experiments. The single view guidance matrix X_p^{sv} and adaptive-weight matrix $Z_p^{'}$ can be constructed by the same strategy of building X_p^{mv} and Z_p.

Furthermore, for each region, its k-nearest spatial neighbors imply rich information that should be considered in our model. Even though the regional physical topology is already involved in multi-view and single-view learning processes, the KNN is a more flexible method. After structuring X_p^{knn} which is an imputed matrix with the average value of k-nearest neighbors, our optimization functions of single view and KNN are expressed as:

$$\min_{W\geq 0, H_p\geq 0} \mathcal{S} = \sum_{p=1}^{d} ||\bar{Y_p} \odot Z_p^{'} \odot (X_p^{sv} - WH_p)||_F^2, \tag{12}$$

$$\min_{W\geq 0, H_p\geq 0} \mathcal{K} = \sum_{p=1}^{d} ||\bar{Y_p} \odot (X_p^{knn} - WH_p)||_F^2, \tag{13}$$

Taking all the above techniques into consideration, our final jointly loss function is shown as follows:

$$\min_{W\geq 0, H_p\geq 0} \mathcal{J} = \mathcal{N} + \lambda_1 \mathcal{M} + \lambda_2 \mathcal{S} + \lambda_3 \mathcal{K}, \tag{14}$$

where λ_2 and λ_3 are the regularization parameters to control the learning rate of X_p^{sv} and X_p^{knn}, respectively.

After solving Equation 14, the learned matrices W and H_p can be used to do the missing data imputation. The filled data are estimated by:

$$\hat{X_p} = Y_p \odot X_p + \bar{Y_p} \odot (WH_p) \tag{15}$$

4.5 Learning Process

As Equation (14) is a non-convex problem, we use the multiplicative update strategy [44] to discover the local optimization. Additionally, to generate the complete kernels, we need to initialize the missing values in data matrices $\{X_1 ... X_d\}$, shown in the Section 5.7. The update rules for W and H_p are presented in Equation (16) - (17).

Theorem 1. *$\mathcal{J}$ is non-increasing under the following update rules in Equation 16-17 by optimizing W and H_p alternatively:*

$$W = W \odot \frac{\sum_{p=1}^{d} (Y_p \odot X_p + \bar{Y_p} \odot (\lambda_1 Z_p \odot X_p^{mv} + \lambda_2 Z_p^{'} \odot X_p^{sv} + \lambda_3 X_p^{knn})) H_p^{\top}}{\sum_{p=1}^{d} ((Y_p + \bar{Y_p} \odot (\lambda_1 Z_p + \lambda_2 Z_p^{'} + \lambda_3 \mathbf{1})) \odot (W^{\top} H_p) H_p^{\top})} \tag{16}$$

$$H_p = H_p \odot \frac{W(Y_p \odot X_p + \bar{Y_p} \odot (\lambda_1 Z_p \odot X_p^{mv} + \lambda_2 Z_p^{'} \odot X_p^{sv} + \lambda_3 X_p^{knn}))}{W(Y_p + \bar{Y_p} \odot (\lambda_1 Z_p + \lambda_2 Z_p^{'} + \lambda_3 \mathbf{1})) \odot (W^\top H_p)} \tag{17}$$

The prove of **Theorem 1** is given in Section 4.6.1. The above two multiplicative update rules guarantee to be non-negative if the initialization is positive. Without this constraint, the matrices W and H_p could be negative, thus the imputation results could be negative too, which is a contradiction to the facts. We now derive the update rule of W as an example, other variables can be solved with a similar process. The objective of $\mathcal{J}$ could be rewritten as follows:

$\mathcal{J} = L_0 + L_1 + L_2 + L_3$, where:

$$\begin{aligned} L_0 &= \sum_{p=1}^{d} ||Y_p \odot (X_p - WH_p)||_F^2, \\ L_1 &= \lambda_1 \sum_{p=1}^{d} ||\bar{Y_p} \odot Z_p \odot (X_p^{mv} - WH_p)||_F^2, \\ L_2 &= \lambda_2 \sum_{p=1}^{d} ||\bar{Y_p} \odot Z_p^{'} \odot (X_p^{sv} - WH_p)||_F^2, \\ L_3 &= \lambda_3 \sum_{p=1}^{d} ||\bar{Y_p} \odot (X_p^{knn} - WH_p)||_F^2 \end{aligned} \tag{18}$$

We provide the derivative of L_0 respect to W as an example, the other components can be derived in the same way. L_0 could also be rewritten as follows:

$$L_0 = \langle Y_p \odot (X_p - WH_p), Y_p \odot (X_p - WH_p) \rangle \tag{19}$$

where $\langle , \rangle$ presents the inner product of matrix. Then:

$$\begin{aligned} dL_0(W) &= -2 \sum_{p=1}^{d} \langle dWH_p, Y_p \odot (X_p - WH_p) \rangle \\ &= -2 \sum_{p=1}^{d} \left\langle dW, Y_p \odot (X_p - WH_p) H_p^\top \right\rangle \\ \Rightarrow \frac{\partial L_0}{\partial W} &= -2 \sum_{p=1}^{d} Y_p \odot (X_p - WH_p) H_p^\top \end{aligned} \tag{20}$$

Analogously, we can get:

$$\frac{\partial L_1}{\partial W} = -2\lambda_1 \sum_{p=1}^{d} \bar{Y_p} \odot Z_p \odot (X_p^{mv} - WH_p) H_p^\top \tag{21}$$

$$\frac{\partial L_2}{\partial W} = -2\lambda_2 \sum_{p=1}^{d} \bar{Y_p} \odot Z_p^{'} \odot (X_p^{sv} - WH_p) H_p^\top \tag{22}$$

$$\frac{\partial L_3}{\partial W} = -2\lambda_3 \sum_{p=1}^{d} \bar{Y_p} \odot (X_p^{knn} - WH_p) H_p^\top \tag{23}$$

As discussed in [44], the traditional gradient descent method is expressed as: $W_t = W_t - \gamma g(W_t) = W_t - \gamma(P_{item} + N_{item})$, where P_{item} and N_{item} denote all positive and negative items in $g(W_t)$, respectively (e.g., $P_{item} = \sum_{p=1}^{d} ((Y_p + \bar{Y_p} \odot (\lambda_1 Z_p + \lambda_2 Z_p^{'} + \lambda_3 \mathbf{1})) \odot (W^\top H_p) H_p^\top)$). We can set the step γ to:

$$\gamma = \frac{W_t}{P_{item}} \tag{24}$$

then, we got the update rule of W as shown in Equation 16.

Algorithm 1 summarizes our learning and estimation process of SMV-NMF.

Algorithm 1: SMV-NMF

Input: original data $\{X_p\}$; graph Laplacian matrix L.
Output: complete data $\{\hat{X_p}\}$.
Impute X_p by KNN for an initialization.
Initialize W and H_p by decomposing X_p.
Construct X_p^{mv}, X_p^{sv} and X_p^{knn} by S-MKKM, S-KKM, and KNN respectively.
for *Epoch = 1 to T* **do**
    if $|\mathcal{J}_t - \mathcal{J}_{t+1}| \,/\, \mathcal{J}_t \geq \varepsilon$ **then**
        update W **By** Equation (16)
        update H **By** Equation (17)
    else
        Break
Return $\hat{X}_p$ **By** Equation (15).

4.6 Time complexity and convergence

We discuss the time complexity and convergence of SMV-NMF here. The time complexity of guidance matrices X_p^{mv} and X_p^{sv} is mainly affected by MKKM. Even though MKKM has a high computational complexity ($O(n^3)$), it is not involved in updating loop of variables (W and H_p). Equation (16) and Equation (17) present that the time complexity of our final function is governed by matrix multiplication operations in each iteration. Therefore, the time complexity per iteration is dominated by $O(nk^2)$. Due to the pursuing of pinpoint accuracy, we sacrifice efficiency to some degree in this real-world problem. In terms of convergence, we give the strict convergence proof of W because other variables can similarly be proofed.

4.6.1 *Proof of Theorem 1*

To prove **Theorem 1**, we need to find an auxiliary function for SMV-NMF objective function as expressed in Equation (14).

Definition 1. *$G(h, h^{'})$ is an auxiliary function for our final function $\mathcal{J}(h)$ if the following conditions are satisfied:*

$$G(h', h) \geq \mathcal{J}(h) \quad \text{and} \quad G(h, h) = \mathcal{J}(h). \tag{25}$$

Lemma 1 *If G is an auxiliary function, then $\mathcal{J}$ is non-increasing under the update:*

$$h^{t+1} = \arg\min_h G\left(h, h^t\right), \tag{26}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2021.3072642, IEEE Transactions on Knowledge and Data Engineering

consequently, we have:

$$\mathcal{J}\left(h^{t+1}\right) \leq G\left(h^{t+1}, h^{t}\right) \leq G\left(h^{t}, h^{t}\right) = \mathcal{J}\left(h^{t}\right). \quad (27)$$

The proof of Lemma 1 is given by [44]. Lemma 1 illustrates that $\mathcal{J}\left(h^{t+1}\right) \leq \mathcal{J}(h^t)$ when exits $\mathcal{Q}\left(h, h^t\right)$.

Lemma 2. *If $K(h^t)$ is a diagonal matrix under the following definition,*

$$K(h^t) = diag(W diag(v) W^T h./h), \quad (28)$$

where v is a column vector of $V = Y_p + \overline{Y}_p \odot \left(\lambda_1 Z_p + \lambda_2 Z_p' + \lambda_3 \mathbf{1}\right)$ then,

$$\begin{aligned} G\left(h, h^t\right) =& J\left(h^t\right) + \left(h - h^t\right)^T \nabla \mathcal{J}\left(h^t\right) \\ &+ \frac{1}{2}\left(h - h^t\right)^T K\left(h^t\right)\left(h - h^t\right), \end{aligned} \quad (29)$$

is an auxiliary function for $\mathcal{J}(h)$.

Proof: Since $G(h,h) = \mathcal{J}(h)$ is obvious, we need only show that $G(h,h^t) \geq \mathcal{J}(h)$. To do this, we compare

$$\begin{aligned} \mathcal{J}(h) =& \mathcal{J}\left(h^t\right) + \left(h - h^t\right)^T \nabla \mathcal{J}\left(h^t\right) \\ &+ \frac{1}{2}\left(h - h^t\right)^T \left(W diag(v) W^T\right)\left(h - h^t\right) \end{aligned} \quad (30)$$

with Equation (29) to find that $G(h,h^t) \geq \mathcal{J}(h)$ is equivalent to

$$0 \leq \left(h - h^t\right)^T \left[K\left(h^t\right) - W diag(v) W^T\right]\left(h - h^t\right) \quad (31)$$

The next step is to prove $\left[K\left(h^t\right) - W\,\mathrm{diag}(v) W^T\right]$ is positive semi-definite. Let $Q = W diag(v) W^T$, then $\left[K\left(h^t\right) - W\,\mathrm{diag}(v) W^T\right]$ can be expressed as $[diag(Qh./h) - Q]$. As the **Lemma 1** provided in [45], if Q is a symmetric non-negative matrix and h be a positive vector, then the matrix $\hat{Q} = diag(Qh./h) - Q \succeq 0$.

Replacing $G(h,h^t)$ in Equation (26) by Equation (29) results in the update rule:

$$h^{t+1} = h^t - K\left(h^t\right)^{-1} \nabla \mathcal{J}\left(h^t\right) \quad (32)$$

Since Equation (29) is an auxiliary function, $\mathcal{J}$ is nonincreasing under this update rule, according to **Lemma 1**. Writing the components of this equation explicitly, we obtain

$$h_a^{t+1} = h_a^t \frac{(Wx)_a}{(W(v \odot W^T h))_a} \quad (33)$$

where x is the column vector of $X = \left(Y \odot X + \overline{Y} \odot \left(\lambda_1 Z \odot X^{mv} + \lambda_2 Z' \odot X^{sv} + \lambda_3 X^{knn}\right)\right)$.

By reversing the roles of W and H in **Lemma 1** and **Lemma 2**, $\mathcal{J}$ can similarly be shown to be nonincreasing under the update rules for W. □

5 EXPERIMENTS

In this chapter, we have conducted comprehensive experiments to demonstrate the effectiveness of our method. The source code has been released at https://github.com/SMV-NMF.

5.1 Datasets

We use eight real-world urban statistical datasets collected from the Australian Bureau of Statistics 2017 (ABS) and the New Zealand Stats 2018 (NZS), i.e., **Sydney**, **Melbourne**, **Brisbane**, **Perth**, **SYD-large** and **MEL-large** are collected from the ABS, and **Auckland** and **Northland** are collected from NZS. **-large** datasets contains much more fine-grained regions. In ABS datasets, each one contains four views, i.e., Economy, Family, Income, and Population; the size (number of fine-grained areas) of the first six datasets are 174, 284, 220, 130, 2230, 1985 respectively, and the numbers of attributes of the four views are 43, 44, 50, 97. The last two datasets are provided by NZS, which includes 563 areas with eleven views, and the average number of the view's dimension is six. Notably, due to the fact that the NZS geography map is inaccessible to us, we tested our method without any geography guidance on these datasets. And for the ABS data, the designation of regions is based on the Statistical Geography Standard[5] for the best practical value. The scales of different views are normalized into the same range [0,10] so that we can evaluate the results together. Besides, to guarantee the diversity of testing, for each missing ratio, we randomly select the test columns and repeat the experiment 20 times and report average results.

5.2 Baselines & Measures

5.2.1 Baselines

We compare the proposed method SMV-NMF with the following 13 baselines. All parameters of the proposed method and baselines are optimized by the grid search method.

sKNN: A classical method that uses the average values of its k nearest spatial neighbors as an estimate (k=5).

MKKMIK$^{\mathbf{a}}$: A MKKM based method to handle the incomplete views [38]. We modified it to adapt to the spatially related data, then interpolated a missing value by its k nearest spatial neighbors (k=5);

MKKMIK$^{\mathbf{b}}$: Similar to MKKMIK$^{\mathbf{a}}$ but utilize the mean value of each cluster to fill the missing data.

NMF: Fill the missing data by NMF.

IDW: A global spatial learning method compared in many works [16], [21].

UCF: The Local spatial learning method based on collaborative filtering [8], [20].

IDW+UCF: Fill missing entries by the average result of IDW and UCF.

MVL-IV: A state-of-the-art multi-view learning method based on matrix co-factorization, which learns a same coefficient matrix to connect multiple views [34].

ST-MVL: A state-of-the-art method to impute spatio-temporal missing data [8]. We only use its spatial part due to the problem of missing temporal information.

SMV-MF: We remove the non-negativity constraint in SMV-NMF to test the effects of this constraint.

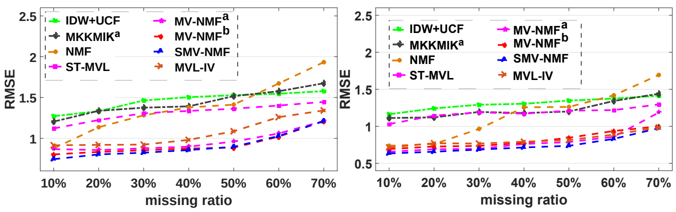
MV-NMF$^{\mathbf{a}}$: Remove the graph Laplacian dynamics strategy in SMV-NMF when building the X_p^{mv} and X_p^{sv};

MV-NMF$^{\mathbf{b}}$: Remove the KNN guidance in SMV-NMF.

MV-NMF$^{\mathbf{c}}$: Remove all the geography guidance that can be used in the NZS datasets.

5. https://www.abs.gov.au/geography

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2021.3072642, IEEE Transactions on Knowledge and Data Engineering



(a) Test on Sydney (b) Test on Melbourne

Fig. 5: Average RMSE with the variation of missing ratios.

Measures. We utilized the most widely used evaluation metrics in this chapter, namely Mean Relative Error (MRE) and Root Mean Square Error (RMSE).

$$MRE = \frac{\sum_{i=1}^{Q} |u_i - \hat{u}_i|}{\sum_{i=1}^{Q} u_i}, \quad RMSE = \sqrt{\frac{\sum_{i=1}^{Q} (u_i - \hat{u}_i)^2}{Q}},$$

where $\hat{u}_i$ is a prediction for missing value, and u_i is the ground truth; Q is the number of prediction values.

5.3 Performance Results and Analysis

The first set of experiments is designed to assess performance on each dataset. We pick up 1/3 of statistical fields (properties) in each urban dataset randomly as the validation set, and the other set as the test set. In the test set, we randomly select missing ratios from 10% to 70% to evaluate the imputation accuracy. As mentioned in Section 5.1, NZS data do not provide geography knowledge that some methods and baselines cannot be used.

Because of the page width limitation, we partition our main results into two tables, i.e., Table 2 and Table 3. These two tables report the average errors of all missing ratios across different compared algorithms. In this test, it is apparent that the series of proposed approaches (SMV-MF, MV-NMF^a, MV-NMF^b, MV-NMF^c, SMV-NMF) achieve the better results on eight real-world datasets. The SMV-NMF method which corporates both geography and latent spatial guidances that performs almost most through all experiments. MVL-IV yields better results than ST-MVL, MKKMIK$^{\mathbf{a}}$, IDW+UCF, and NMF becuase it considers the multi-view problem. Compared with the best baseline MVL-IV, our method learned the similarities from latent spaces instead of clustering in the original data; second, we have added prior knowledge (adaptive weights) to each missing entries and observed entries. Such prior knowledge also considered the spatial correlation which is significant for the spatially correlated data [2], [21]. Although ST-MVL is a great method of filling spatio-temporal missing data, it would not perform well when the missing temporal information challenge appears.

To represent the error change with the varying missing ratios, we draw the top eight methods with different missing ratios on the Sydney and Melbourne datasets, which is shown in Fig. 5. The results drawn in Fig. 5 demonstrate the results of Table 2-3 and the discussions aforementioned. For the methods NMF, MV-NMF^a, they are sensitive to the missing ratios, which could get good results under the lower level missing ratios, but performs worse when the missing ratio increases. To evaluate the improvement of our model, we also report the T-test results (p-value 1) compared with the best baseline MVL-IV, the result is significant at $p < 0.05$. Our methods, (SMV-MF, MV-NMF^a, MV-NMF^b, SMV-NMF) have significant improvements compared with current baselines.

Ablation Study. To analyses the contribution of each component of the final method SMV-NMF, we analyze the ablation study here. All the results are shown in Tables 2 and 3. Table 4 illustrates the different strategies used in proposed models. n-constraint represents the non-negativity constraint; Laplacian means whether the model considers the graph Laplacian dynamics when building X_p^{mv} and X_p^{sv}; knn indicates the KNN guidance is used or not.

As the results shown in Tables 2 and 3, we can see that without the non-negativity constraint, SMV-MF performs worse than SMV-NMF, which demonstrates the effectiveness of this constraint. Two models (MV-NMF^a and MV-NMF^b) will perform worse when casting off the graph Laplacian dynamics strategy. If we only consider the non-negativity constraint but without any geography guidance, our model, MV-NMF^c cannot achieve a good result. But it is still better than other baselines. We have provided the T-test (p-value 2) between the best model MV-NMF and the second best MV-NMF^b to present the improvements, the result is significant at $p < 0.05$.

Overall, SMV-NMF outperforms the other baselines because it integrates both multi-view and spatial problems to address the specified missing data imputation task. MV-NMF^a MV-NMF^b and MV-NMF^c remove a part of the spatial guidance which results in slightly worse performances than SMV-NMF.

5.4 Experiments on Generalization Ability

In this section, we try to explore the generalization ability of our method. The test process is that the dataset Sydney is chosen as the validation set and two urban datasets (Melbourne and Brisbane) are chosenas the test sets. Fig. 6 reports the performances among eight outstanding approaches. We clearly see that the SMV-NMF achieves the best performance.

Our method represents strong generalization ability which can transfer the constructed model from one urban dataset to another. This is because there are high correlations among cities. For example, the number of functional regions of each city is mostly the same, resulting in the same amount of clusters. The gap between SMV-NMF and MVL-IV narrows as the missing ratio increases, but the former is more robust than the latter because SMV-NMF achieves the best results across all missing ratios. Table 5 reveals the average errors using two evaluation metrics. The generality test demonstrates that our model SMV-NMF is a universal model that performs well crossing different urban statistical datasets.

5.5 View Correlation Analysis

To evaluate the correlations between views, we represent the view weight changes with varying missing ratios. Fig. 7 shows the results on the Sydney dataset. As we can see that the view of Economy occupies the highest priority in

Methods	Sydney		Melbourne		Brisbane		Perth	
	MRE	**RMSE**	**MRE**	**RMSE**	**MRE**	**RMSE**	**MRE**	**RMSE**
sKNN	0.332 ± 0.011	1.530 ± 0.097	0.310 ± 0.014	1.372 ± 0.079	0.355 ± 0.006	1.518 ± 0.065	0.381 ± 0.010	1.575 ± 0.083
MKKMIK^b	0.329 ± 0.015	1.550 ± 0.089	0.346 ± 0.011	1.462 ± 0.076	0.377 ± 0.009	1.593 ± 0.070	0.398 ± 0.018	1.699 ± 0.064
IDW	0.332 ± 0.009	1.518 ± 0.077	0.318 ± 0.010	1.318 ± 0.089	0.351 ± 0.008	1.466 ± 0.090	0.372 ± 0.009	1.557 ± 0.081
UCF	0.356 ± 0.008	1.663 ± 0.083	0.338 ± 0.009	1.463 ± 0.079	0.362 ± 0.007	1.592 ± 0.072	0.375 ± 0.008	1.655 ± 0.080
IDW+UCF	0.330 ± 0.007	1.460 ± 0.056	0.314 ± 0.006	1.304 ± 0.064	0.340 ± 0.008	1.397 ± 0.070	0.361 ± 0.007	1.495 ± 0.069
MKKMIK^a	0.308 ± 0.011	1.439± 0.105	0.288 ± 0.009	1.226 ± 0.092	0.316 ± 0.010	1.347± 0.097	0.354 ± 0.008	1.506 ± 0.083
NMF	0.221 ± 0.014	1.384 ± 0.111	0.199 ± 0.009	1.155 ± 0.093	0.225 ± 0.010	1.304 ± 0.108	0.248 ± 0.013	1.288 ± 0.127
ST-MVL	0.294 ± 0.007	1.313 ± 0.069	0.283 ± 0.007	1.179 ± 0.077	0.311 ± 0.006	1.295 ± 0.067	0.332 ± 0.006	1.394 ± 0.077
MVL-IV	0.198 ± 0.005	1.063 ± 0.032	0.174 ± 0.004	0.818 ± 0.029	0.197 ± 0.004	0.969 ± 0.041	0.225 ± 0.005	1.067 ± 0.044
SMV-MF	0.191 ± 0.004	0.960 ± 0.031	0.181 ± 0.004	0.800 ± 0.027	0.183 ± 0.005	0.854 ± 0.021	0.219 ± 0.004	1.002 ± 0.029
MV-NMF^a	0.185 ± 0.005	0.925 ± 0.020	0.177 ± 0.004	0.815 ± 0.022	0.163 ± 0.003	0.757 ± 0.015	0.223 ± 0.004	0.976 ± 0.022
MV-NMF^b	0.180 ± 0.003	0.927 ± 0.013	0.173 ± 0.002	0.804 ± 0.015	0.164 ± 0.003	0.770 ± 0.014	0.217 ± 0.004	0.972 ± 0.021
MV-NMF^c	0.195 ± 0.005	0.957 ± 0.033	0.182 ± 0.004	0.853 ± 0.030	0.190 ± 0.005	0.860 ± 0.035	0.228 ± 0.005	1.064 ± 0.042
MV-NMF	**0.175 ± 0.002**	**0.901 ± 0.012**	**0.168 ± 0.002**	**0.747 ± 0.011**	**0.157 ± 0.002**	**0.705 ± 0.009**	**0.208 ± 0.003**	**0.933 ± 0.014**
p-value 1	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
p-value 2	$= 0.14$	< 0.01	$= 0.02$	< 0.01	< 0.01	< 0.01	< 0.02	< 0.03

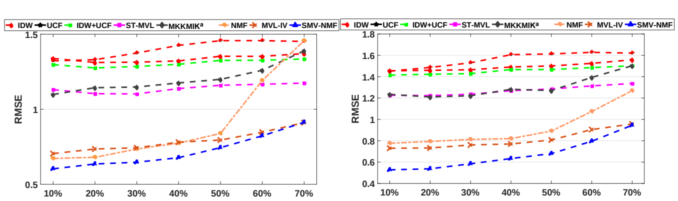
TABLE 2: The average MRE, RMSE, std and p-value on real-world urban statistical datasets (Part I). Best results are bold.

Methods	SYD-large		MEL-large		Auckland		Northland	
	MRE	**RMSE**	**MRE**	**RMSE**	**MRE**	**RMSE**	**MRE**	**RMSE**
sKNN	0.297 ± 0.010	1.256 ± 0.088	0.310 ± 0.012	1.313 ± 0.070	–	–	–	–
MKKMIK^b	0.348 ± 0.013	1.611 ± 0.094	0.306 ± 0.011	1.456 ± 0.068	0.527 ± 0.027	2.139 ± 0.156	0.570 ± 0.034	2.575 ± 0.201
IDW	0.327 ± 0.008	1.499 ± 0.065	0.308 ± 0.009	1.258 ± 0.072	–	–	–	–
UCF	0.332 ± 0.009	1.423 ± 0.068	0.332 ± 0.009	1.509 ± 0.070	–	–	–	–
IDW+UCF	0.304 ± 0.006	1.223 ± 0.047	0.297 ± 0.007	1.211 ± 0.055	–	–	–	–
MKKMIK^a	0.291 ± 0.012	1.288 ± 0.080	0.301 ± 0.014	1.230 ± 0.101	0.496 ± 0.022	1.957 ± 0.123	0.512 ± 0.028	2.200 ± 0.104
NMF	0.238 ± 0.009	1.199 ± 0.092	0.203 ± 0.011	1.066 ± 0.073	0.406 ± 0.017	1.580 ± 0.092	0.444 ± 0.016	1.787 ± 0.099
ST-MVL	0.294 ± 0.006	1.077 ± 0.077	0.282 ± 0.008	1.145 ± 0.071	–	–	–	–
MVL-IV	0.179 ± 0.004	0.895 ± 0.059	0.184 ± 0.003	0.922 ± 0.046	0.327 ± 0.012	1.226 ± 0.066	**0.322 ± 0.010**	1.467 ± 0.039
SMV-MF	0.177 ± 0.003	0.835 ± 0.024	0.192 ± 0.004	0.901 ± 0.031	–	–	–	–
MV-NMF^a	0.171 ± 0.005	0.822 ± 0.017	0.185 ± 0.003	0.861 ± 0.020	–	–	–	–
MV-NMF^b	0.168 ± 0.002	0.804± 0.014	0.176 ± 0.004	0.812 ± 0.017	–	–	–	–
MV-NMF^c	0.182 ± 0.005	0.960 ± 0.034	0.193 ± 0.004	0.915 ± 0.035	**0.317 ± 0.009**	**1.208 ± 0.061**	0.341 ± 0.009	**1.405 ± 0.032**
SMV-NMF	**0.166 ± 0.001**	**0.774 ± 0.007**	**0.169 ± 0.002**	**0.791 ± 0.008**	–	–	–	–
p-value 1	< 0.01	< 0.01	< 0.01	< 0.01	$=0.09$	$= 0.35$	< 0.01	< 0.01
p-value 2	$= 0.34$	< 0.05	< 0.04	< 0.03	–	–	–	–

TABLE 3: The average MRE, RMSE, std and p-value on real-world urban statistical datasets (Part II). Best results are bold.

TABLE 4: Ablation Studies. The strategies used in different models.

n-constraint	Laplacian	knn	Method
✓		✓	MV-NMF^a
✓	✓		MV-NMF^b
✓			MV-NMF^c
	✓	✓	SMV-MF
✓	✓	✓	SMV-NMF



(a) Test dataset Melbourne. (b) Test dataset Brisbane.

Fig. 6: The average RMSE in generalization ability tests.

the similarity learning process when the missing ratio is below (or equal to) 50%. When data is relatively adequate,

Methods	Dataset Melbourne		Dataset Brisbane	
	MRE	**RMSE**	**MRE**	**RMSE**
UCF	0.331 ± 0.009	1.405 ± 0.096	0.368 ± 0.008	1.560 ± 0.089
IDW	0.333 ± 0.010	1.339 ± 0.070	0.369 ± 0.006	1.496 ± 0.071
IDW+UCF	0.320 ± 0.007	1.306 ± 0.047	0.351 ± 0.005	1.455 ± 0.072
MKKMIK^a	0.282 ± 0.010	1.201 ± 0.103	0.313 ± 0.009	1.302 ± 0.086
ST-MVL	0.279 ± 0.006	1.139 ± 0.077	0.312 ± 0.006	1.269 ± 0.038
NMF	0.183 ± 0.017	0.955 ± 0.126	0.199 ± 0.015	0.989 ± 0.093
MVL-IV	0.152 ± 0.003	0.787 ± 0.022	0.163 ± 0.004	0.808 ± 0.029
SMV-NMF	**0.148 ± 0.002**	**0.720 ± 0.006**	**0.149 ± 0.001**	**0.671 ± 0.010**

TABLE 5: Generalizability test. We report the average MRE and RMSE of all missing ratios and best results are bold.

the attributes of economy are the most important factor that influences the imputation method. With increasing missing ratio, the weight of Economy view decreases significantly because the data from other views are more considered to utilize the observed data comprehensively.

5.6 The Sensitivity of Parameters

This section evaluates the performances of SMV-NMF by varying the critical parameters (k, λ_1, λ_2, and λ_3). We here

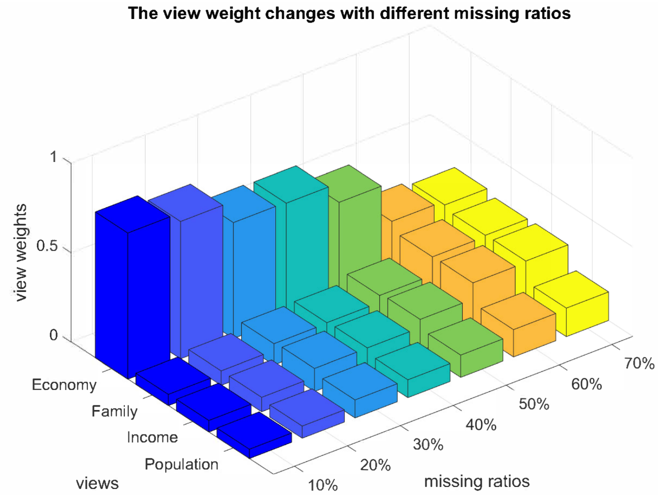
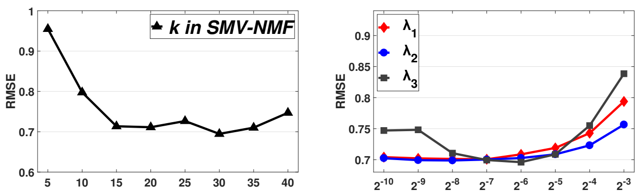


Fig. 7: View correlation analysis on the Sydney dataset.



(a) Factor k.

(b) Factor λ_1&λ_2&λ_3.

Fig. 8: Effect of Parameters.

show the experimental results for the Sydney validation dataset. We discuss them separately but pick them up by the grid search method because four parameters have high dimensional correlations that are hard to visualize. Our illustration approach that discusses parameters separately has been widely used in many other research papers [25], [43].

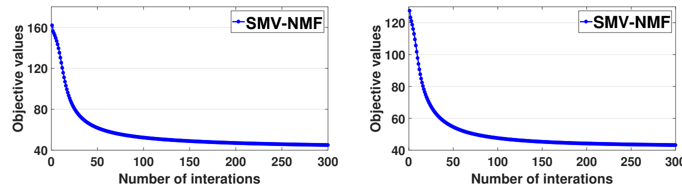
Fig. 8 (a) shows the different performances with a varying setting for k. When we increase k from 5 to 15, the results improve significantly. However, the performance tends to stay stable at $15 \leq k \leq 35$. In particular, SMV-NMF achieves the best result when $k = 30$, while it can get good performance if the k is set between 15 and 35. This indicates that a low-rank latent space representation can already capture the attributes of the urban statistical data.

Fig. 8 (b) reveals the effect of varying λ_1, λ_2, and λ_3. These three parameters determine the strength of the three guidance matrices X^{mv}, X^{sv}, and X^{knn}, respectively. λ_1=2^{-7}, λ_2=2^{-8} and λ_3=2^{-6} yield the best results for SMV-NMF. We observe that the performance is stable when these three parameters are ranged between 2^{-8} and 2^{-6}.

In summary, both parameters used in this chapter bring benefits to the improvement of our models. Furthermore, our model is stable and easy fine-tuning because it is insensitive to these parameters.

5.7 Initialization and Convergence

To get the complete kernels, we first impute the missing data for each view by an efficient method, such as KNN and MF.



(a) Conduct on the Melbourne dataset.

(b) Conduct on the Perth dataset .

Fig. 9: Convergence rate.

TABLE 6: Effects of different initialization methods.

	Zero-init	Random-init	Mean-init	KNN-init
RMSE	1.173	1.134	0.949	**0.908**

The effects of different initializations are reported in Table 6. Based on the results, we easily find that the initialization method KNN could achieve the great performance for SMV-NMF. Accordingly, we choose the KNN method for a good balance between time-consuming and accuracy.

Figs. 9 (a) and (b) show the convergence trends of iterative model SMV-NMF on both the Melbourne and Perth datasets. It illustrates that our algorithm can converge into a local solution in terms of the objective value in a small number of iterations.

6 Conclusion

Due to some inevitable issues, urban statistical data usually suffer from the missing data problem. To overcome it, we propose a missing data imputation model for multi-view urban statistical data via the spatial correlation learning, which called SMV-NMF in this paper. To handle the multi-view problem, we develop an improved spatial multi-kernel method to guide the imputation process based on the NMF strategy. Furthermore, the spatial correlations among different regions are taken into consideration from two aspects. First, the latent similarities are discovered by S-MKKN and S-KKM based on the idea of finding functional regions, and secondly, KNN is used for capturing the information of real geographical positions. We conduct intensive experiments on eight real-world datasets to compare the performance of our model and other state-of-the-art approaches. The results not only show that our approach outperforms all other methods, but also represent strong generalizabilities crossing different urban datasets.

Acknowledgment

The authors would like to thank the Australian Bureau of Statistics and the New Zealand Stats to provide data for the research exploration. And we also thank all editors and anonymous reviewers for their constructive comments that help to improve the quality of this paper.

References

[1] B. Murgante and M. Danese, "Urban versus rural: the decrease of agricultural areas and the development of urban zones analyzed with spatial statistics," *International Journal of Agricultural and Environmental Information Systems*, vol. 2, no. 2, pp. 16–28, 2011.

[2] A. Gomez-Lievano, H. Youn, and L. M. Bettencourt, "The statistics of urban scaling and their connection to zipf's law," *PloS one*, vol. 7, no. 7, p. e40393, 2012.

[3] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing: concepts, methodologies, and applications," *ACM Transactions on Intelligent Systems and Technology*, vol. 5, no. 3, p. 38, 2014.

[4] Z. Li, J. Zhang, Q. Wu, Y. Gong, J. Yi, and C. Kirsch, "Sample adaptive multiple kernel learning for failure prediction of railway points," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 2848–2856.

[5] Y. Gong, Z. Li, J. Zhang, W. Liu, and J. Yi, "Potential passenger flow prediction: A novel study for urban transportation development." in *AAAI*, 2020, pp. 4020–4027.

[6] B. Liu, Y. Fu, Z. Yao, and H. Xiong, "Learning geographical preferences for point-of-interest recommendation," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2013, pp. 1043–1051.

[7] W. R. Tobler, "A computer movie simulating urban growth in the detroit region," *Economic geography*, vol. 46, no. sup1, pp. 234–240, 1970.

[8] X. Yi, Y. Zheng, J. Zhang, and T. Li, "St-mvl: filling missing values in geo-sensory time series data." in *Proceedings of 25th International Joint Conference on Artificial Intelligence*, 2016, pp. 2704–2710.

[9] G. Atluri, A. Karpatne, and V. Kumar, "Spatio-temporal data mining: A survey of problems and methods," *ACM Computing Surveys (CSUR)*, vol. 51, no. 4, pp. 1–41, 2018.

[10] C. Zhu, L. Cao, Q. Liu, J. Yin, and V. Kumar, "Heterogeneous metric learning of categorical data with hierarchical couplings," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 7, pp. 1254–1267, 2018.

[11] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, vol. 9, no. 6, p. 717, 2009.

[12] Y. Liang, K. Ouyang, L. Jing, S. Ruan, Y. Liu, J. Zhang, D. S. Rosenblum, and Y. Zheng, "Urbanfm: Inferring fine-grained urban flows," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 3132–3142.

[13] K. Ouyang, Y. Liang, Y. Liu, Z. Tong, S. Ruan, D. Rosenblum, and Y. Zheng, "Fine-grained urban flow inference," *IEEE Transactions on Knowledge and Data Engineering*, 2020.

[14] L. Pan and J. Li, "K-nearest neighbor based missing data estimation algorithm in wireless sensor networks," *Wireless Sensor Network*, vol. 2, no. 2, p. 115, 2010.

[15] M. Ranjbar, P. Moradi, M. Azami, and M. Jalili, "An imputation-based matrix factorization method for improving accuracy of collaborative filtering systems," *Engineering Applications of Artificial Intelligence*, vol. 46, pp. 58–66, 2015.

[16] S. Cheng and F. Lu, "A two-step method for missing spatio-temporal data reconstruction," *ISPRS International Journal of Geo-Information*, vol. 6, no. 7, p. 187, 2017.

[17] J. Zhou and Z. Huang, "Recover missing sensor data with iterative imputing network," in *Workshops at the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[18] Y. Gong, Z. Li, J. Zhang, W. Liu, B. Chen, and X. Dong, "A spatial missing value imputation method for multi-view urban statistical data." in *Proceedings of 29th International Joint Conference on Artificial Intelligence*, 2020, pp. 1310–1316.

[19] S. Van Buuren, *Flexible imputation of missing data*. CRC press, 2018.

[20] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Advances in artificial intelligence*, vol. 2009, 2009.

[21] F.-W. Chen and C.-W. Liu, "Estimation of the spatial rainfall distribution using inverse distance weighting (idw) in the middle of taiwan," *Paddy and Water Environment*, vol. 10, no. 3, pp. 209–222, 2012.

[22] T. Wu and Y. Li, "Spatial interpolation of temperature in the united states using residual kriging," *Applied Geography*, vol. 44, pp. 112–120, 2013.

[23] L. Li, X. Su, Y. Wang, Y. Lin, Z. Li, and Y. Li, "Robust causal dependence mining in big data network and its application to traffic flow predictions," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 292–307, 2015.

[24] S. Sun, C. Zhang, and G. Yu, "A bayesian network approach to traffic flow forecasting," *IEEE Transactions on intelligent transportation systems*, vol. 7, no. 1, pp. 124–132, 2006.

[25] D. Deng, C. Shahabi, U. Demiryurek, L. Zhu, R. Yu, and Y. Liu, "Latent space model for road networks to predict time-varying traffic," in *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2016, pp. 1525–1534.

[26] X. He and P. Niyogi, "Locality preserving projections," in *Advances in neural information processing systems*, 2004, pp. 153–160.

[27] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.

[28] E. Acar, D. M. Dunlavy, T. G. Kolda, and M. Mørup, "Scalable tensor factorizations for incomplete data," *Chemometrics and Intelligent Laboratory Systems*, vol. 106, no. 1, pp. 41–56, 2011.

[29] H. Tan, G. Feng, J. Feng, W. Wang, Y.-J. Zhang, and F. Li, "A tensor-based method for missing traffic data completion," *Transportation Research Part C: Emerging Technologies*, vol. 28, pp. 15–27, 2013.

[30] B. Ran, L. Song, J. Zhang, Y. Cheng, and H. Tan, "Using tensor completion method to achieving better coverage of traffic state estimation from sparse floating car data," *PloS one*, vol. 11, no. 7, p. e0157420, 2016.

[31] H. Tan, Y. Wu, G. Feng, W. Wang, and B. Ran, "A new traffic prediction method based on dynamic tensor completion," *Procedia-Social and Behavioral Sciences*, vol. 96, pp. 2431–2442, 2013.

[32] A. P. Singh and G. J. Gordon, "Relational learning via collective matrix factorization," in *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2008, pp. 650–658.

[33] Y. Li, M. Yang, and Z. M. Zhang, "A survey of multi-view representation learning," *IEEE Transactions on Knowledge and Data Engineering*, 2018.

[34] C. Xu, D. Tao, and C. Xu, "Multi-view learning with incomplete views," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5812–5825, 2015.

[35] S. Rendle, L. Balby Marinho, A. Nanopoulos, and L. Schmidt-Thieme, "Learning optimal ranking with tensor factorization for tag recommendation," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2009, pp. 727–736.

[36] L. Xiong, X. Chen, T.-K. Huang, J. Schneider, and J. G. Carbonell, "Temporal collaborative filtering with bayesian probabilistic tensor factorization," in *Proceedings of the 2010 SIAM International Conference on Data Mining*. SIAM, 2010, pp. 211–222.

[37] A. Trivedi, P. Rai, H. Daumé III, and S. L. DuVall, "Multiview clustering with incomplete views," in *The workshop pf 24th Conference on Neural Information Processing Systems*, 2010, pp. 1–7.

[38] X. Liu, M. Li, L. Wang, Y. Dou, J. Yin, and E. Zhu, "Multiple kernel k-means with incomplete kernels." in *Proceedings of 31th Conference on AAAI*, 2017, pp. 2259–2265.

[39] X. Li, G. Cui, and Y. Dong, "Graph regularized non-negative low-rank matrix factorization for image clustering," *IEEE transactions on cybernetics*, vol. 47, no. 11, pp. 3840–3853, 2016.

[40] X. Luo, H. Wu, H. Yuan, and M. Zhou, "Temporal pattern-aware qos prediction via biased non-negative latent factorization of tensors," *IEEE transactions on cybernetics*, vol. 50, no. 5, pp. 1798–1809, 2019.

[41] Y. Gong, Z. Li, J. Zhang, W. Liu, and Y. Zheng, "Online spatio-temporal crowd flow distribution prediction for complex metro system," *IEEE Transactions on Knowledge and Data Engineering*, 2020.

[42] R. Lambiotte, J.-C. Delvenne, and M. Barahona, "Laplacian dynamics and multiscale modular structure in networks," *arXiv preprint arXiv:0812.1770*, 2008.

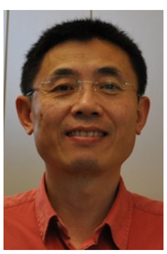
[43] Y. Gong, Z. Li, J. Zhang, W. Liu, Y. Zheng, and C. Kirsch, "Network-wide crowd flow prediction of sydney trains via customized online non-negative matrix factorization," in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 2018, pp. 1243–1252.

[44] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in neural information processing systems*, 2001, pp. 556–562.

[45] M. D. Gupta and J. Xiao, "Non-negative matrix factorization as a feature selection tool for maximum margin classifiers," in *CVPR 2011*. IEEE, 2011, pp. 2841–2848.

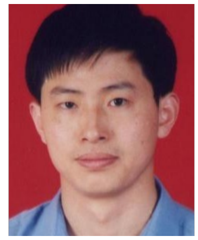
Yongshun Gong (SM'19-M'21) is an Associate Professor at the School of Software, Shandong University, China. He received his Ph.D. degree from University of Technology Sydney in 2021. His principal research interest covers the data science and machine learning, in particular, the following areas: spatio-temporal data mining; traffic prediction; recommender system and sequential pattern mining. He has published above 20 papers in top journals and refereed conference proceedings, including the IEEE Transactions on Knowledge and Data Engineering, IEEE Transactions on Neural Networks and Learning Systems, IEEE Transactions on Cybernetics, IEEE Transactions on Multimedia, Pattern Recognition, NeurIPS, KDD, CIKM, AAAI and IJCAI.

Zhibin Li obtained the Ph.D. degree from University of Technology Sydney in 2021. He is now a postdoctoral research fellow of the Commonwealth Scientific and Industrial Research Organisation, Australia. His principal research interest includes traffic analysis, factorization models, and theoretical analysis of machine learning. He has published 9 papers in top journals and refereed conference proceedings including TKDE, KDD2019 and NeurIPS2020.



Jian Zhang (SM'04) is currently an Associate Professor with the Global Big Data Technologies Centre, School of Electrical and Data Engineering, Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney. He received the PhD in electrical engineering from the University of New South Wales (UNSW), Sydney, Australia, in 1999. He is the author or co-author of more than 180 paper publications, book chapters, and seven issued US and China patents. His current research interests include social multimedia signal processing, large scale image and video content analytics, retrieval and mining, 3D based computer vision and intelligent video surveillance systems. Dr. Zhang was the General Co-Chair of the International Conference on Multimedia and Expo in 2012 and Technical Program Co-Chair of IEEE Visual Communications and Image Processing 2014. Currently, he is an Associated Editors for the IEEE TRANSACTIONS ON MULTIMEDIA and was an Associate Editor for the IEEE TRANSCTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (2006 – 2015).

Wei Liu (M'15-SM'20) is an Associate Professor and the Director of Robust Machine Learning Lab at the School of Computer Science, Faculty of Engineering and Information Technology, the University of Technology Sydney (UTS). Before joining UTS, he was a Research Fellow at the University of Melbourne and then a Machine Learning Researcher at NICTA. He obtained his PhD from the University of Sydney in 2011. He works in the areas of machine learning and data mining and has published more than 80 papers in the research topics of tensor factorization, adversarial learning, multimodal machine learning, graph mining, causal inference, and anomaly detection. He has won three best paper awards.



Yilong Yin received the Ph.D. degree from Jilin University, Changchun, China, in 2000. From 2000 to 2002, he was a Postdoctoral Fellow with the Department of Electronic Science and Engineering, Nanjing University, Nanjing, China. He is the Director of the Machine Learning and Applications Group and a Professor with Shandong University, Jinan, China. His research interests include machine learning, data mining, computational medicine, and biometrics.

Dr. Yu Zheng is the Vice President of JD.COM and the Chief Data Scientist at JD Digits, passionate about using big data and AI technology to tackle urban challenges. Before Joining JD.COM, he was a senior research manager at Microsoft Research. Zheng is also a Chair Professor at Shanghai Jiao Tong University. He currently serves as the Editor-in-Chief of ACM Transactions on Intelligent Systems and Technology and has served as chair on over 10 prestigious international conferences, e.g. as the program co-chair of ICDE 2014 (Industrial Track), CIKM 2017 (Industrial Track) and IJCAI 2019 (industrial track). He is also a keynote speaker of AAAI 2019, KDD 2019 Plenary Keynote Panel and IJCAI 2019 Industrial Days. His monograph, entitled Urban Computing, has been used as the first text book in this field. In 2013, he was named one of the Top Innovators under 35 by MIT Technology Review (TR35) and featured by Time Magazine for his research on urban computing. In 2017, Zheng is honored as an ACM Distinguished Scientist.