Multi-Tensor Completion with Common Structures

[1,2]Chao Li, [2]Qibin Zhao, [2]Junhua Li, [2,3]Andrzej Cichocki and [1]Lili Guo
[1]Harbin Engineering University, Harbin, Heilongjiang, China
[2]RIKEN Brain Science Institute, Wako, Saitama, Japan
[3]Systems Research Institute, PAS, Warsaw, Poland
{chao.li, junhua.li}@riken.jp, {qbzhao, cia}@brain.riken.jp, guolili@hrbeu.edu.cn

Abstract

In multi-data learning, it is usually assumed that common latent factors exist among multi-datasets, but it may lead to deteriorated performance when datasets are heterogeneous and unbalanced. In this paper, we propose a novel common structure for multi-data learning. Instead of common latent factors, we assume that datasets share Common Adjacency Graph (CAG) structure, which is more robust to heterogeneity and unbalance of datasets. Furthermore, we utilize CAG structure to develop a new method for multi-tensor completion, which exploits the common structure in datasets to improve the completion performance. Numerical results demonstrate that the proposed method not only outperforms state-of-the-art methods for video in-painting, but also can recover missing data well even in cases that conventional methods are not applicable.

Introduction

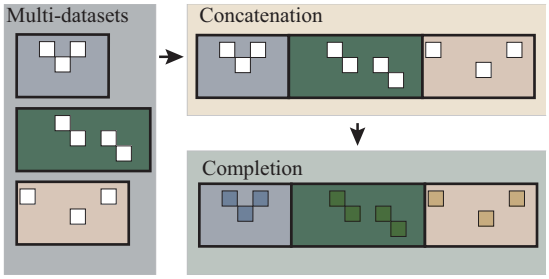
In multi-data learning, a general assumption is that datasets contain common latent factors, and such an assumption has been successfully applied in many fields, e.g., Collective Matrix Factorization (Singh and Gordon 2008), Multi Tensor Decomposition (Takeuchi et al. 2013; Acar, Kolda, and Dunlavy 2011), and Group Analysis (Calhoun, Liu, and Adali 2009; Karhunen, Hao, and Ylipaavalniemi 2013; Ylmaz, Cemgil, and Simsekli 2011). For example, in the matrix case, a set of matrices $\mathbf{X}_k,\ k = 1, \ldots, K$ can be decomposed as (Zhou, Cichocki, and Xie 2012)

$$\mathbf{X}_k = \mathbf{U}^C \mathbf{V}_k^{\mathrm{T}} + \mathbf{N}_k, \qquad k = 1, \ldots, K, \tag{1}$$

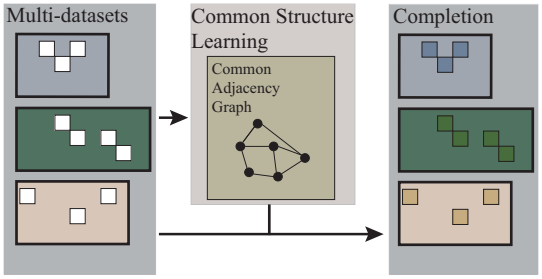
where $\mathbf{U}^C$ denotes common factors shared among $\mathbf{X}_k$, and $\mathbf{N}_k$ denotes residuals such as distinctive factors or noise. To solve this problem, "concatenation" usually plays a key role since (1) can be represented by

$$\begin{aligned} \bar{\mathbf{X}} &= [\mathbf{X}_1 \ldots \mathbf{X}_K] \\ &= \mathbf{U}^C \left[\mathbf{V}_1 \ldots \mathbf{V}_K\right]^{\mathrm{T}} + [\mathbf{N}_1 \ldots \mathbf{N}_K] \\ &= \mathbf{U}^C \bar{\mathbf{V}}^{\mathrm{T}} + \bar{\mathbf{N}}, \end{aligned} \tag{2}$$

where $\bar{\mathbf{X}}, \bar{\mathbf{V}}, \bar{\mathbf{N}}$ denote concatenated matrices of $\mathbf{X}_k, \mathbf{V}_k, \mathbf{N}_k$, respectively. From (2) we can estimate common factors $\mathbf{U}^C$ by factorizing the concatenated matrix $\bar{\mathbf{X}}$ directly. However, in the problem of data completion, the model given by (2) does not work well when different datasets are heterogeneous or unbalanced, even though they indeed mutually share common information. Such situation frequently occurs in multi-view/media data in-painting, link prediction for social networks and recommendation systems, due to that datasets might be collected from different devices or domains.



(a) Conventional model



(b) The proposed model

Figure 1: Comparison between conventional model and the proposed model for multi-data completion where white squares represent missing data, and filled squares represent estimated data.

In the problem of recommendation system, a well-known principle is "Similar users select similar items.". Thus it is straightforward to infer that utilizing similarity structure to describe common information is better than common latent factors in multi-data completion. Therefore, in this paper, we develop a novel framework to exploit common information

among datasets. Instead of common latent factors, we assume that datasets share Common Adjacency Graph (CAG) structure (see Fig. 1). In addition, we propose a new method for multi-tensor completion by taking the common structure into account.

Basic model

Common adjacency graph

For clarity, we first introduce the new model in the matrix case, and then apply it for tensor data in the next subsection. Suppose that we have K matrices $\mathbf{X}_k \in \mathbb{R}^{I \times N_k}$, $k = 1, \ldots, K$, and let the first mode be the common mode, i.e. rows of matrices $\mathbf{X}_k$ have common structure information, and each row of datasets is defined as an *entity*. Let G denote a graph in which nodes represent entities and edges represent relationships between every pair of entities in datasets (see Fig. 2). Furthermore, a weight is assigned to each edge to quantify how 'close' of entities. We call such a graph as Common Adjacency Graph (CAG) since it can reflect the common neighbour relationships among entities of datasets.

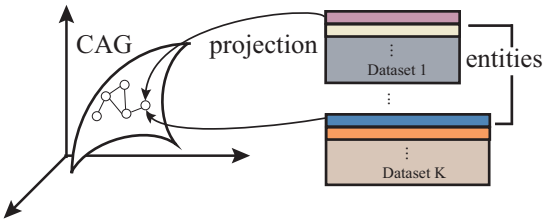


Figure 2: Schematic diagram illustrating how matrices have CAG structure.

More specifically, we use an adjacency matrix $\mathbf{W} \in \mathbb{R}^{I \times I}$ to describe weights of edges in the graph G. The i, jth element of $\mathbf{W}$, denoted by $\mathbf{W}(i, j)$, presents the similarity between ith and jth entities. In the simplest case, we can specify $\mathbf{W}$ by binary values according to a priori knowledge, that is

$$\mathbf{W}_p(i,j) = \begin{cases} 1 & \text{if node } i \text{ and } j \text{ are connected} \\ 0 & \text{otherwise} \end{cases}. \quad (3)$$

An alike manner was proposed by (Chen, Hsu, and Liao 2013), in which smoothness was imposed for single image in-painting by a so-called differential matrix. However, such a priori knowledge cannot be obtained in advance for most applications. In order to learn CAG structure from data directly, one straightforward data-driven measure can be defined by

$$\mathbf{W}_c(i,j) = \frac{1}{K}\sum_{k=1}^{K} \frac{\langle \mathbf{X}_k(i,:) - \mu_{k,i}, \mathbf{X}_k(j,:) - \mu_{k,j} \rangle}{\|\mathbf{X}_k(i,:) - \mu_{k,i}\|_2 \|\mathbf{X}_k(j,:) - \mu_{k,j}\|_2}, \quad i,j = 1, \ldots I, \quad (4)$$

where $\langle \cdot, \cdot \rangle$ denotes inner product of two vectors, $\|\cdot\|_2$ denotes Euclidean norm of a vector, $\mathbf{X}_k(i,:)$, $k = 1, \ldots, K$, $i = 1, \ldots, I$ represents the ith row of matrix $\mathbf{X}_k$, and $\mu_{k,j}$ is given by $\mu_{k,j} = \sum_{l=1}^{N_k} \mathbf{X}_k(j,l)/N_k$. It can be found from (4) that $\mathbf{W}$ is actually an average correlation matrix among $\mathbf{X}_k \in \mathbb{R}^{I \times N_k}$, $k = 1, \ldots, K$, and its element $\mathbf{W}(i,j) \in [-1, 1]$ describes similarity between two entities. If we restrict the measure to be non-negative, an alternative measure is defined by

$$\mathbf{W}_g(i,j) = \exp\left(-\frac{\sum_{k=1}^{K} \|\mathbf{X}_k(i,:) - \mathbf{X}_k(j,:)\|_2^2 / N_k}{K\sigma}\right), \quad i,j = 1, \ldots, I, \quad (5)$$

where σ denotes a tuning parameter to control the degree of attenuation. To avoid scale diversity of datasets, we can also use a normalized version of (5) in which we force entities to be zero mean and unit variance. Note that $\mathbf{W}(i,j) \in (0, 1]$ in (5) can be considered as a non-linear version of (4), and the values are kept being always positive.

Formulas (3)-(5) represent three possible measures to describe relationships of entities. According to specific applications, we can easily extend any existing similarity measure to the model. In this paper, we only use the normalized version of (5) in experiments since it generally provides good performance in physical data like video in-painting. For other kinds of data, e.g., ratings and links, extensions of Jaccard similarity and cosine similarity (Shepitsen et al. 2008) are expected to achieve satisfactory results.

Multi-tensor completion with common structures

In the problem of tensor completion, low rank approximation is usually applied to estimate missing values. Many methods minimize the nuclear norm of a tensor since nuclear norm is well known as the tightest convex envelope of rank of a matrix (Fazel 2002), and has been directly extended into tensor by (Gandy, Recht, and Yamada 2011). In the case of multi-tensor completion, we also follow this idea, but it should be noted that only minimizing the nuclear norm of each individual tensor may break the structure of Common Adjacency Graph. Thus, in the following, we will introduce a novel multi-tensor completion method, in which not only low rank approximation but also CAG structure is imposed to estimate missing values in multi-tensor.

Suppose we have K tensors $\underline{\mathbf{X}}_k$, $k = 1, \ldots, K$ with missing values that are indexed by binary tensors $\underline{\mathbf{B}}_k$ (0-unobserved, 1-observed) whose size is same to $\underline{\mathbf{X}}_k$. Furthermore, suppose that there is one common mode for all tensors, and datasets have CAG structure along the common mode. Additionally we suppose the dimension of the common mode is equal to M, and let a vector $\mathbf{c} \in \mathbb{R}^{K \times 1}$ describe the order of the common mode in each tensor. Then the fully reconstructed tensor $\underline{\mathbf{Y}}_k$, $k = 1, \ldots, K$ can be ob-

tained by minimizing the following cost function

$$\begin{aligned}\min\ & f\left(\{\underline{\mathbf{Y}}_k|\forall k\},\mathbf{W}\right)= \\ & \sum_{k,i,j\neq i}\mathbf{W}\left(i,j\right)\|\mathbf{Y}_k^{(\mathbf{c}(k))}\left(i,:\right)-\mathbf{Y}_k^{(\mathbf{c}(k))}\left(j,:\right)\|_2^2 \\ & +\sum_k\lambda_k\|\underline{\mathbf{Y}}_k\|_* \qquad (6)\\ s.t.\quad & P_{\underline{\mathbf{B}}_k}\left(\underline{\mathbf{Y}}_k\right)=P_{\underline{\mathbf{B}}_k}\left(\underline{\mathbf{X}}_k\right)\\ & \mathbf{W}\text{ satisfies one of (3)-(5) over }\underline{\mathbf{Y}}_k,\ \forall k\end{aligned}$$

where $\mathbf{c}\left(k\right)$ denotes the kth element of vector $\mathbf{c}$, $\mathbf{Y}_k^{(\mathbf{c}(k))}$ denotes unfolding of tensor $\underline{\mathbf{Y}}_k$ along the $\mathbf{c}\left(k\right)$th order (Cichocki et al. 2009), and $\mathbf{W}\in\mathbb{R}^{M\times M}$. $\|\underline{\mathbf{Y}}_k\|_*$ denotes nuclear norm of $\underline{\mathbf{Y}}_k$, which is defined by $\|\underline{\mathbf{Y}}_k\|_*=\sum_{q,i}\alpha_q\cdot\sigma_i\left(\mathbf{Y}_k^{(q)}\right)/\sum_j\alpha_j$, where $\sigma_i\left(\mathbf{Y}_k^{(q)}\right)$ denotes the ith singular value of $\mathbf{Y}_k^{(q)}$ and $\alpha_q,q=1,\ldots,Q_k$ represents any positive number that was set as 1 for simplicity. $P_{\underline{\mathbf{B}}_k}\left(\cdot\right)$ indicates that we choose observed elements of which the corresponding elements in $\underline{\mathbf{B}}_k$ are equal to 1, and $\lambda_k>0$ are tuning parameters. From (6), we can find that matrix $\mathbf{W}$ builds a relationship among all datasets. In other words, it means that estimated tensors $\underline{\mathbf{Y}}_k$ have common neighbour (i.e., CAG) structure along the common mode. Meanwhile, we also minimize nuclear norm of individual tensors as penalty terms to achieve low rank approximation of observations. It is worth noting that, if let $\mathbf{W}$ equal a identity matrix, (6) degenerates to estimating the low rank approximation of each dataset individually. It means, in such case no common information is utilized, but in practice there would be more or less common information which can be exploited to achieve better performance for multi-data analysis. If there are more than one common mode, (6) can be straightforwardly extended by imposing additional CAG structures.

Algorithm

Note that it is difficult to optimize (6) directly due to the existence of tensor nuclear norm and the additional constraints. In order to search the optimal solution of (6) efficiently, we relax the cost function by introducing auxiliary matrices $\mathbf{Z}_{k,q_k}\forall k,q_k$ which can be seen as an estimation for the unfolding $\mathbf{Y}_k^{(q_k)}$ of tensor $\underline{\mathbf{Y}}_k$, and the relaxed cost function can be easily solved by Proximal Gradient Descent (PGD) and Alternating Update (AU) methods. Let $\underline{\mathbf{X}}_k,\ k=1,\ldots,K$ be K Q_kth order tensors, then a relaxation of (6) can be formulated as

$$\begin{aligned}\min\ & f\left(\{\underline{\mathbf{Y}}_k|\forall k\},\{\mathbf{Z}_{k,q_k}|\forall k,q_k\},\mathbf{W}\right)\\ & =\sum_{k,i,j\neq i}\mathbf{W}\left(i,j\right)\|\mathbf{Z}_{k,\mathbf{c}(k)}\left(i,:\right)-\mathbf{Z}_{k,\mathbf{c}(k)}\left(j,:\right)\|_2^2 \\ & +\sum_{k,q_k}\lambda_{k,q_k}\|\mathbf{Z}_{k,q_k}\|_* \qquad (7)\\ s.t.\quad & P_{\underline{\mathbf{B}}_k}\left(\underline{\mathbf{Y}}_k\right)=P_{\underline{\mathbf{B}}_k}\left(\underline{\mathbf{X}}_k\right),\ \mathbf{Y}_k^{(q_k)}=\mathbf{Z}_{k,q_k}\\ & \mathbf{W}\text{ satisfies one of (3)-(5) over }\underline{\mathbf{Y}}_k,\ \forall k,\ q_k\end{aligned}$$

where $\mathbf{Z}_{k,q_k}$ has the same size to the corresponding unfolding matrix $\mathbf{Y}_k^{(q_k)}$. To search the stationary point of (7), we update $\mathbf{Z}_{k,q_k}$, $\underline{\mathbf{Y}}_k$, $\forall k,q_k$ and $\mathbf{W}$ sequentially. For each matrix and tensor, we update $\mathbf{Z}_{k,q_k}$, $\forall k,q_k$ by PGD methods which has been discussed by (Cai, Candès, and Shen 2010; Toh and Yun 2010), and update $\underline{\mathbf{Y}}_k$ and $\mathbf{W}$ by standard gradient methods. Pseudo-code of the algorithm is shown in Alg.1.

Algorithm 1 Multi-tensor completion with common structures

Input: Observation $\underline{\mathbf{X}}_k$ and index tensors $\underline{\mathbf{B}}_k$, the order of common mode $\mathbf{c}$ for $\underline{\mathbf{X}}_k$ and tuning parameter $\lambda_{k,q_k}>0,\ k=1,\ldots,K,\ q_k=1,\ldots,Q_k$.

Output: Estimated tensors $\underline{\mathbf{Y}}_k$.

1: Initialize $\underline{\mathbf{Y}}_k^0$ by $P_{\underline{\mathbf{B}}_k}\left(\underline{\mathbf{Y}}_k^0\right)=P_{\underline{\mathbf{B}}_k}\left(\underline{\mathbf{X}}_k\right)$, and fill unobserved elements by zero, compute $\mathbf{W}^0$ over $\underline{\mathbf{Y}}_k^0$, and let iteration counter $t=0$.
2: **repeat**
3: $t=t+1$.
4: $\mathbf{Z}_{k,\mathbf{c}(k)}^{t-1}\leftarrow\mathbf{R}\mathbf{Y}_k^{(\mathbf{c}(k)),t-1},\ \forall k$, where entries of $\mathbf{R}$ are given by
$$\mathbf{R}\left(i,i\right)=0,\ i=1,\ldots,M,$$
$$\mathbf{R}\left(i,j\right)=\mathbf{W}^{t-1}\left(i,j\right)/\sum_{m\neq i}\mathbf{W}^{t-1}\left(i,m\right),\ i\neq j$$
5: $\mathbf{Z}_{k,q_k}^{t-1}\leftarrow\mathbf{Y}_k^{q_k,t-1},\forall k,q_k\neq\mathbf{c}\left(k\right)$
6: $\mathbf{Z}_{k,q_k}^{t}\leftarrow D\left(\mathbf{Z}_{k,q_k}^{t-1}|\ \lambda_{k,q_k}/2\right)_+,\ \forall k,q_k$
7: $\underline{\mathbf{Y}}_k^t\leftarrow\sum_{q_k}ten\left(\mathbf{Z}_{k,q_k}^t\right)/Q_k,\ \forall k$
8: $P_{\underline{\mathbf{B}}_k}\left(\underline{\mathbf{Y}}_k^t\right)\leftarrow P_{\underline{\mathbf{B}}_k}\left(\underline{\mathbf{X}}_k\right),\ \forall k$
9: Compute $\mathbf{W}^t$ over $\underline{\mathbf{Y}}_k^t,\ \forall k$.
10: **until** convergence

In Alg.1, steps 4-6 are performed according to PGD, where step 4 is derived by computing partial derivative of (7) w.r.t. $\mathbf{Z}_{k,\mathbf{c}(k)}^{t-1}$ under the constraint $\mathbf{Y}_k^{(q_k)}=\mathbf{Z}_{k,q_k},\ \forall k,q_k$. Matrix $\mathbf{R}$ denotes a intermediate matrix in order to compact expression, and $D\left(\cdot|\cdot\right)_+$ denotes soft-thresholding the singular values of the matrix with a scalar to achieve a low rank approximation. If $\mathbf{X}=\mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}}$ is the Singular Value Decomposition (SVD) of $\mathbf{X}$, then $D\left(\mathbf{X}|\lambda\right)_+=\mathbf{U}\bar{\mathbf{\Sigma}}\mathbf{V}^{\mathrm{T}}$, where any element $\bar{\mathbf{\Sigma}}\left(i,j\right)$ of $\bar{\mathbf{\Sigma}}$ satisfies

$$\bar{\mathbf{\Sigma}}\left(i,j\right)=\begin{cases}\mathbf{\Sigma}\left(i,j\right)-\lambda & \mathbf{\Sigma}\left(i,j\right)>\lambda\\ 0 & \text{otherwise}\end{cases}. \qquad (8)$$

Note that $ten\left(\cdot\right)$ denotes folding a matrix into a tensor which can be considered as an reverse procedure to tensor unfolding. It should be noted from Alg.1 that auxiliary variables $\mathbf{Z}_{k,q_k}$ can provide closed-form solutions for each iteration, additionally reduce mutual influence among unfolding matrices of tensors when computing tensor nuclear norm.

Related works

The problem of a single tensor completion has been extensively discussed in (Liu et al. 2013; Signoretto et al. 2014; Xu et al. 2013; Zhao, Zhang, and Cichocki 2014; Acar et al. 2011), where (Zhao, Zhang, and Cichocki 2014; Acar et al. 2011) are based on tensor decomposition to achieve a low rank approximation for completion, while (Liu et al. 2013; Signoretto et al. 2014; Xu et al. 2013) minimize nuclear norm of a tensor directly. However, all these methods only focus on one single tensor (e.g., images and videos), thus they cannot sufficiently exploit common information in the case of multi-data task. In the case of multiple datasets, (Singh and Gordon 2008; Bouchard, Yin, and Guo 2013; Klami, Bouchard, and Tripathi 2013) discuss multi-matrix completion in recommendation system, and (London et al. 2013; Nickel 2013; Nickel and Tresp 2013) developed methods for multi-relational learning. However, all these methods hold a basic assumption that common latent factors are contained by datasets. It has been discussed previously that such an assumption is often not sufficient in many applications. In machine learning, (Niyogi 2004) and later (Shu, Ma, and Latecki 2014; He, Cai, and Niyogi 2005) discussed methods to find a locality preserving subspace for feature extraction and classification/clustering. Inspiring by these works, we find that CAG is an inherent structure to share common information in multiple datasets.

Experimental results

Multi-view video in-painting with one frame lost

In computer vision and graphics, many methods were developed for image and video in-painting (Liu et al. 2013; Xu et al. 2013; Zhao, Zhang, and Cichocki 2014). In our paper, we extend in-painting problem to multi-view videos. Furthermore, we intend to reconstruct severely damaged videos in which not only pixels but a whole frame is lost. In this case, numerical experiments show that conventional methods can not satisfactorily recover the lost frame, but our method can estimate not only missing pixels but also the lost frame efficiently by utilizing the CAG structure among datasets.

Specifically, in Experiment 1, we utilized multi-view videos (Height×Width×Channel×Time) of human action (body movement by a lady) recorded by 3 cameras simultaneously[1]. Since the three videos were recorded at the same time from different views, it is straightforward to infer that datasets are mutually related along the time mode (i.e., common mode). Furthermore, in order to validate our method in the case of unbalanced datasets, the three videos were resized. This unbalanced situation is always happened in real applications due to videos are possibly recorded by different devices. For performance evaluation, we independently run the experiment 20 times, and randomly removed some pixels from the three videos with specific missing percentage, and randomly removed one frame from the first video in each run. Tab.1 shows Related Square Error ($RSE = \sum_k \|\underline{\mathbf{X}}_k - \underline{\mathbf{Y}}_k\|^2 / \sum_k \|\underline{\mathbf{X}}_k\|^2$, where $\underline{\mathbf{X}}_k$, $k = 1, \ldots, K$ are original tensors and $\underline{\mathbf{Y}}_k$ are the corresponding estimations.) of missing values and lost frame estimation in different missing percentage.

Table 1: RSE of estimation of multi-view videos in-painting in different missing percentages

Methods	90%		80%		50%	
	Miss	Lost	Miss	Lost	Miss	Lost
Proposed	**0.25**	**0.25**	**0.18**	**0.19**	**0.12**	**0.13**
MTC	0.29	0.77	0.25	0.79	0.14	0.91
FaLRTC	0.27	N/A	0.20	N/A	0.12	N/A
HardC	0.31	N/A	0.22	N/A	0.13	N/A
TMac	0.31	N/A	0.29	N/A	0.28	N/A
CPWOPT	0.27	N/A	0.24	N/A	0.21	N/A

For comparison, FaLRTC (Liu et al. 2013), HardC (Signoretto et al. 2014), TMac (Xu et al. 2013), and CPWOPT (Acar et al. 2011) was individually implemented on every video since all these methods only focused on one single tensor. Additionally, MTC (Li, Guo, and Cichocki 2014) represents a "common factor" based method which was implemented on the three videos, and the key point of MTC is to concatenate all datasets along the common mode. It is shown from Tab. 1 that the proposed method outperforms other methods. For lost frame estimation, only MTC and the proposed method work, but the performance of MTC for lost frame estimation is seriously poor. Fig. 3 shows prediction of two frames chosen from the first video, where the first row represents a frame with missing pixels and the second row represents the lost frame. Fig. 4 shows comparison of average runtime in different missing percentage. It is shown from Fig. 3, 4 that the proposed method can recover the lost frames well, and its runtime is comparable to other methods.

Video in-painting with audio information

In multi-media analysis, not only videos but also audio are usually recorded into a multi-media file. Thus one question arises whether we can in-paint a video in serious damaged cases by using auxiliary information from audio or not? The answer is obvious, but the biggest challenge is that video and audio data are heterogeneous and extremely unbalanced. In Experiment 2, we demonstrate that the proposed method can recover videos with numerous frames lost (randomly or consecutively) by exploiting common structures of audio data.

In Experiment 2, a video we used was that a train passed by a camera with sound of whistles and noise generated by wheels, etc.[2]. We extracted video and audio dataset from the original file individually, and organize data into a tensor and matrix respectively. For the video, the tensor was organized as $\underline{\mathbf{V}} \in \mathbb{R}^{61\times 88\times 3\times 227}$ (Height×Width×Channel×Time). For audio, we only choose one channel and cut it up into 227 segments which correspond to frames of the video. As

[1] http://media.au.tsinghua.edu.cn/index.jsp

[2] http://v.youku.com/v_show/id_XNDI3NTAxMjIw.html

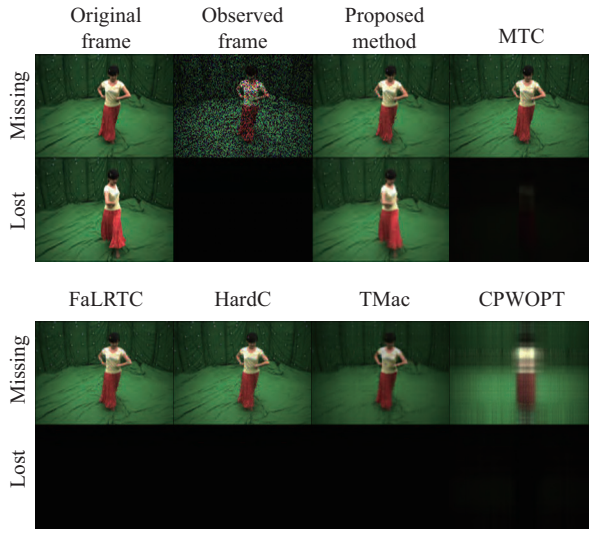


Figure 3: Performance comparisons among methods when missing percentage of pixel was equal to 50%.

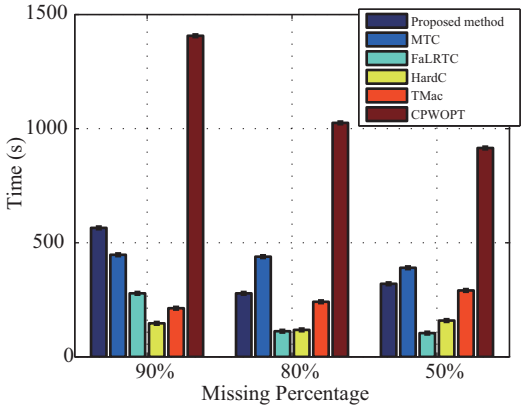


Figure 4: Comparison of average run time for various methods.

preprocessing, we applied Fourier transform on each segment with Hamming window to obtain spectral amplitude, and removed high frequency components since important audio information only focused on low frequency. Finally, we obtained the audio data as a matrix $\mathbf{A} \in \mathbb{R}^{227\times 701}$.

The content of the whole video can be mainly divided into four stages. In the first stage, camera was recording the background, the train didn't pass by the camera yet, and there is only noise in audio; In the second stage, a yellow railway engine passed by the camera with whistle; In the third stage, a series of red carriages passed by the camera with the sound of clicks of the train; The final stage was also the background like the first stage but the corresponding sound was still click sounds since the train was not far away from the camera. Fig. 5 shows typical frames and their corresponding audio segments at each stage.

Since the video and audio were recorded at the same time,

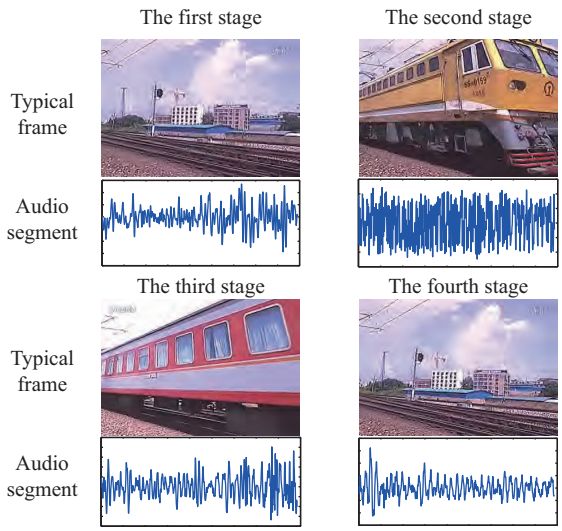
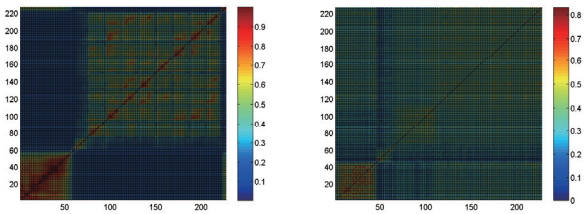


Figure 5: Typical frames and corresponding audio segments for each stage of the data in Experiment 2.

it is easy to understand that they share common information along the time mode. Fig. 6 shows adjacency matrices of unfolding of $\underline{\mathbf{V}}$ and $\mathbf{A}$ along the time mode by using (5), where warm colors represent two entities are similar while cool colors represent they are different from each other. It can be easily found from Fig. 6 that both the two adjacency matrices have some common structures to reflect different stages of the video, which indicates that our assumption is reasonable for video in-painting via audio information.



(a) video data (b) audio data

Figure 6: Adjacency matrix of video and audio data for Experiment 2.

In the experiment, we first randomly removed 50% pixels of each frame from $\underline{\mathbf{V}}$, and randomly removed eight frames from four stages of $\underline{\mathbf{V}}$. We implemented the proposed method to recover the damaged video tensor with audio matrix $\mathbf{A}$. Fig. 7 shows estimations of lost frames by the proposed method.

It is evident from Fig. 7 that the proposed method can provide rough but reasonable estimations for lost frames by using CAG structure shared by video and audio data. Some details (e.g., the gate of the carriage and the 227th frame) cannot be exactly recovered. It is because video data has totally lost this information and common information from $\mathbf{A}$

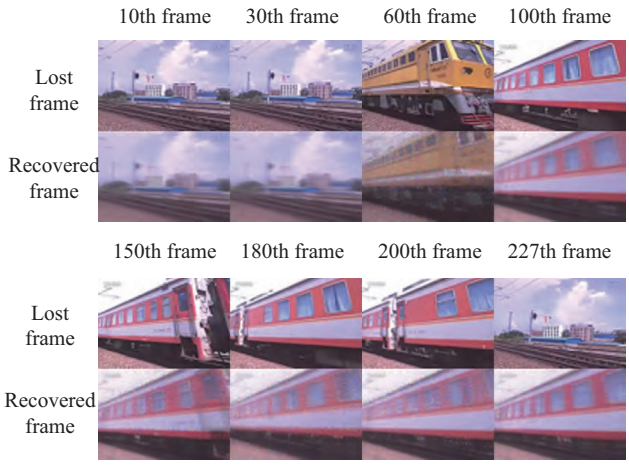
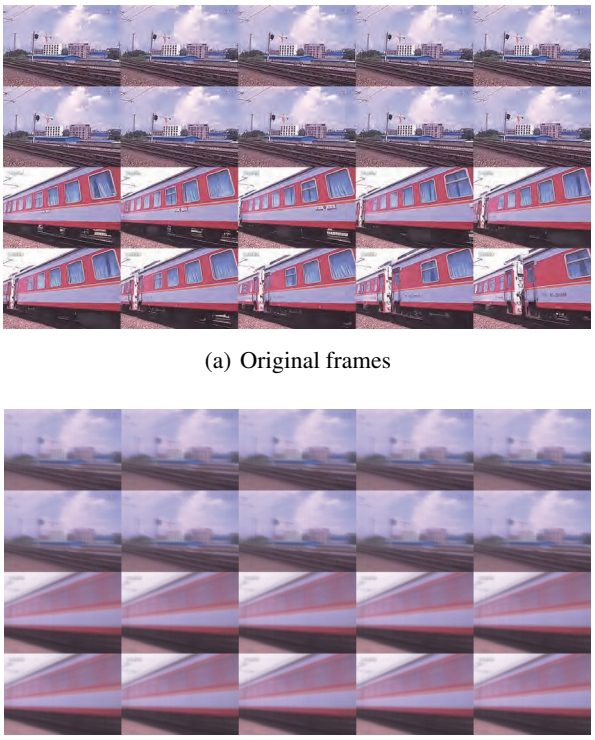


Figure 7: Recovery of lost frames by using audio information.

cannot exactly reflect details of $\underline{\mathbf{V}}$. For example, the 227th frame represents that the train has passed by the camera, but the sound of clicks of the train was still recorded by audio since the train was still near to the camera. Therefore our method can only infer from the audio that the train was still passing by the camera during the 227th frame. However, it is worth noting that our method sufficiently exploits the relationship between video and audio data to provide a fairish estimation, while conventional methods entirely fail in this case, which has been shown in Experiment 1.

One would argue that, since lost frames can be estimated by imposing a priori knowledge about correlation along time domain during completion, what is the main advantage for the proposed method? It should be noted that the proposed method does *NOT* impose any priori knowledge about video in-painting, while we only exploit some useful information from multi-observations. We believe that the proposed method can work for many different applications, e.g. Recommendation System and Link Prediction, as long as multi-datasets share common information.

Compared to the previous experiments in which lost frames were discretely chosen, subsequently, we removed blocks of frames from $\underline{\mathbf{V}}$, and applied the proposed method for in-painting. Similarly, we first randomly removed 50% pixels of each frames from $\underline{\mathbf{V}}$, but next removed consecutive 10 frames (it will last 1 sec in the video for each block) from the first and third stages of $\underline{\mathbf{V}}$, respectively. Fig. 8 shows two blocks of lost consecutive frames and their estimations by the proposed method. It is shown in Fig. 8 that the proposed method can estimate lost consecutive frames reliably. Since details of lost frames of $\underline{\mathbf{V}}$ cannot be exactly reflected from audio data $\mathbf{A}$, the estimation for lost consecutive frames is not very high-quality. However, the result shown in Fig. 8 indicates that common information contained by audio can be exploited for video in-painting problem, even if video and audio data is heterogeneous and extremely unbalanced.



(a) Original frames

(b) Estimations

Figure 8: Recovery of consecutive lost frames by using audio information.

Conclusion

In this paper, we proposed a novel framework to exploit common structure existing in heterogeneous datasets for tensor completion, and such a new idea breaks the conventional assumption that multi-datasets should contain common latent factors. We developed a new structure called Common Adjacency Graph to describe the shared information existing in multiple datasets, regardless of whether datasets are heterogeneous or not. By using Common Adjacency Graph, a novel multi-tensor completion method was introduced, and results of numerical experiments show that the proposed method outperforms conventional tensor completion methods, and can successfully reconstruct videos in the case that conventional methods entirely fail.

Acknowledgements

The authors would like to thank the reviewers for their positive feedback and constructive comments. This work was supported by the National Natural Science Foundation of China (Grant No. 61271263, 61401115, 61202155).

References

Acar, E.; Dunlavy, D. M.; Kolda, T. G.; and Mrup, M. 2011. Scalable tensor factorizations for incomplete data. *Chemometrics and Intelligent Laboratory Systems* 106(1):41–56.

Acar, E.; Kolda, T. G.; and Dunlavy, D. M. 2011. All-at-once optimization for coupled matrix and tensor factorizations. *arXiv preprint arXiv:1105.3422*.

Bouchard, G.; Yin, D.; and Guo, S. 2013. Convex collective matrix factorization. *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics* 144–152.

Cai, J.-F.; Candès, E. J.; and Shen, Z. 2010. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization* 20(4):1956–1982.

Calhoun, V. D.; Liu, J.; and Adali, T. 2009. A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. *Neuroimage* 45(1):S163–S172.

Chen, Y.; Hsu, C.; and Liao, H. 2013. Simultaneous tensor decomposition and completion using factor priors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 36(3):577–591.

Cichocki, A.; Zdunek, R.; Phan, A. H.; and Amari, S.-i. 2009. *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*. John Wiley and Sons.

Fazel, M. 2002. *Matrix rank minimization with applications*. Ph.D. Dissertation, PhD thesis, Stanford University.

Gandy, S.; Recht, B.; and Yamada, I. 2011. Tensor completion and low-n-rank tensor recovery via convex optimization. *Inverse Problems* 27(2):025010.

He, X.; Cai, D.; and Niyogi, P. 2005. Tensor subspace analysis. In *Advances in neural information processing systems*, 499–506.

Karhunen, J.; Hao, T.; and Ylipaavalniemi, J. 2013. Finding dependent and independent components from related data sets: A generalized canonical correlation analysis based method. *Neurocomputing* 113(0):153–167.

Klami, A.; Bouchard, G.; and Tripathi, A. 2013. Group-sparse embeddings in collective matrix factorization. *arXiv preprint arXiv:1312.5921*.

Li, C.; Guo, L.; and Cichocki, A. 2014. Multi-tensor completion for estimating missing values in video data. *arXiv preprint arXiv:1409.0347*.

Liu, J.; Musialski, P.; Wonka, P.; and Ye, J. 2013. Tensor completion for estimating missing values in visual data. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 35(1):208–220.

London, B.; Rekatsinas, T.; Huang, B.; and Getoor, L. 2013. Multi-relational learning using weighted tensor decomposition with modular loss. *arXiv preprint arXiv:1303.1733*.

Nickel, M., and Tresp, V. 2013. Logistic tensor factorization for multi-relational data. *arXiv preprint arXiv:1306.2084*.

Nickel, M. 2013. *Tensor factorization for relational learning*. Ph.D. Dissertation, Ludwig Maximilian University.

Niyogi, X. 2004. Locality preserving projections. In *Neural information processing systems*, volume 16, 153–161.

Shepitsen, A.; Gemmell, J.; Mobasher, B.; and Burke, R. 2008. Personalized recommendation in social tagging systems using hierarchical clustering. In *Proceedings of the 2008 ACM conference on Recommender systems*, 259–266. ACM.

Shu, L.; Ma, T.; and Latecki, L. J. 2014. Locality preserving projection for domain adaptation with multi-objective learning. *AAAI*.

Signoretto, M.; Dinh, Q. T.; De Lathauwer, L.; and Suykens, J. A. 2014. Learning with tensors: A framework based on convex optimization and spectral regularization. *Machine Learning* 94(3):303–351.

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

Takeuchi, K.; Tomioka, R.; Ishiguro, K.; Kimura, A.; and Sawada, H. 2013. Non-negative multiple tensor factorization. In *Proc. IEEE International Conference on Data Mining (ICDM13)*.

Toh, K.-C., and Yun, S. 2010. An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems. *Pacific Journal of Optimization* 6(615-640):15.

Xu, Y.; Hao, R.; Yin, W.; and Su, Z. 2013. Parallel matrix factorization for low-rank tensor completion. *arXiv preprint arXiv:1312.1254*.

Ylmaz, Y. K.; Cemgil, A. T.; and Simsekli, U. 2011. Generalised coupled tensor factorisation. In *Advances in Neural Information Processing Systems*, 2151–2159.

Zhao, Q.; Zhang, L.; and Cichocki, A. 2014. Bayesian CP factorization of incomplete tensors with automatic rank determination. *arXiv preprint arXiv:1401.6497*.

Zhou, G.; Cichocki, A.; and Xie, S. 2012. Common and individual features analysis: Beyond canonical correlation analysis. *arXiv preprint arXiv:1212.3913*.