Metric Learning on Healthcare Data with Incomplete Modalities

Qiuling Suo[1] , **Weida Zhong**[1] , **Fenglong Ma**[1] , **Ye Yuan**[2] , **Jing Gao**[1] **and Aidong Zhang**[3]
[1]Department of Computer Science and Engineering, SUNY at Buffalo, NY, USA
[2]College of Information and Communication Engineering, Beijing University of Technology, China
[3]Department of Computer Science, University of Virginia, VA, USA
{qiulings, weidazho, fenglong}@buffalo.edu, yuanye91@emails.bjut.edu.cn, aidong@virginia.edu

Abstract

Utilizing multiple modalities to learn a good distance metric is of vital importance for various clinical applications. However, it is common that modalities are incomplete for some patients due to various technical and practical reasons in healthcare datasets. Existing metric learning methods cannot directly learn the distance metric on such data with missing modalities. Nevertheless, the incomplete data contains valuable information to characterize patient similarity and modality relationships, and they should not be ignored during the learning process. To tackle the aforementioned challenges, we propose a metric learning framework to perform missing modality completion and multi-modal metric learning simultaneously. Employing the generative adversarial networks, we incorporate both complete and incomplete data to learn the mapping relationship between modalities. After completing the missing modalities, we use the nonlinear representations extracted by the discriminator to learn the distance metric among patients. Through jointly training the adversarial generation part and metric learning, the similarity among patients can be learned on data with missing modalities. Experimental results show that the proposed framework learns more accurate distance metric on real-world healthcare datasets with incomplete modalities, comparing with the state-of-the-art approaches. Meanwhile, the quality of the generated modalities can be preserved.

1 Introduction

With the advances in data collection techniques, large amounts of healthcare data collected from multiple sources are becoming available. Such multi-source data can provide complementary information that can reveal the fundamental characteristics of patients. For example, in the study of Alzheimer's disease, different types of measurements such as magnetic resonance imaging (MRI), positron emission tomography (PET) and cerebrospinal fluid (CSF) are used to examine morphological changes, metabolic changes and cerebrospinal pathology associated with the disease respectively. Extracting valuable information from such multi-source (a.k.a multi-modal) data may effectively improve clinical decision support.

For many clinical applications such as personalized medicine, trajectory analysis and cohort study, it is crucial to learn a proper distance function or similarity measurement metric from multi-modal data. However, existing metric learning models for measuring similarity among patients [Zhan *et al.*, 2016; Huai *et al.*, 2018] only focus on single-modal data, instead of multi-modal data. In other applications, such as image retrieval and face recognition, multi-modal metric learning methods [Xie and Xing, 2013] have been developed through linear or non-linear integration of features from multiple modalities.

However, one common problem that hampers the use of multi-modal metric learning approaches for patient similarity analysis is the issue of missing data. In the healthcare area, each examination generates a modality of data. Due to the potential risks in certain examinations such as PET scans, or invasive procedures such as CSF biomarkers, patients may not be recommended to take all examinations in disease diagnosis. Also, the dropout of patients during the study and data privacy policies can cause the missing modality issue. In the multi-modal datasets, data can be missing in one or multiple modalities, i.e., for a patient, certain data source is either available or missing. Due to the unique challenge of healthcare data, existing unimodal or multi-modal metric learning methods cannot directly measure the distance between two patients with different available modalities.

A simple approach for solving the missing data problem is to remove all samples with missing values, but this dramatically reduces the amount of samples and results in a severe loss of valuable information. In fact, the large amount of data with incomplete modalities provides more information, which is important to characterize the similarity of samples and the relationship between modalities. This motivates us to complete the missing modalities for complementary study, which not only enables patient similarity learning, but also provides potential patterns of the missing sources that can be used for further clinical analysis. Traditional missing data imputation techniques, such as mean and matrix completion are not suitable for large-scale high-dimensional datasets.

To tackle the aforementioned problems, we propose

a Metric Learning with Incomplete Modalities (MeLIM) method that jointly infers the missing modalities and similarity information. The proposed model contains a modality completion part and a metric learning part. In the model, we employ generative adversarial network (GAN) to capture the relationship between the existing modality and the missing one. The missing modality is completed by mapping the existing modality through a generator. A discriminator is used to distinguish the true data and the generated one. Meanwhile, the large amount of incomplete data is incorporated into the GAN framework during the training process. In this way, the relationship between modalities is captured by the generative network and the missing modality can be imputed. We then connect the imputation part with metric learning by incorporating an auxiliary task in the discriminator to make use of the extracted non-linear multi-modal representations. In the proposed model, data of high quality and discriminativeness can be generated, which helps to better measure the similarity among samples. Our main contributions can be summarized as follows:

- We propose a new framework of patient similarity learning on multi-modal healthcare data with incomplete modalities. It imputes the missing modality and learns the sample representation jointly, without a need for a separate imputation step.
- Our proposed method can utilize both complete and incomplete data in the training process. The complete data provides supervised information, and the large amount of incomplete data provides more information of modality characteristics and relationships.
- Comparing with the state-of-the-art approaches, the proposed method not only learns a more accurate distance metric, but also preserves the quality of the generated data, which is validated on a real-world dataset.

2 Related Work

The goal of metric learning is to learn a distance metric so that similar samples are grouped together and dissimilar samples are separated. Metric learning has prompted the development of patient similarity analysis [Zhan *et al.*, 2016; Huai *et al.*, 2018; Ni *et al.*, 2017; Suo *et al.*, 2018], which is a key task in clinical decision support applications.

When it comes to integrating information from multiple sources, multi-modal learning approaches [Xie and Xing, 2013; Zhang *et al.*, 2017; Hu *et al.*, 2014; Zhang *et al.*, 2018; Yuan *et al.*, 2018] cannot be easily applied in healthcare domain because of the missing modality problem. To tackle the missing modality problem, [Li *et al.*, 2018] partitions the data into multiple complete subgroups; [Yang *et al.*, 2018] integrates the consistency of modalities. For high dimensional data such as bio-images, deep learning based approaches are adopted. [Li *et al.*, 2014] learns PET from MRI images by minimizing the pixel difference between predicted images and true ones. However, the loss function may lead to a blurry problem. To generate high quality data, models based on GAN [Goodfellow *et al.*, 2014] are developed to learn the mapping between modalities. GAN models have achieved great success in image-to-image translation [Isola *et al.*, 2017; Gan *et al.*, 2017; Zhu *et al.*, 2017], attribute-to-image generation [Du *et al.*, 2018], etc. The basic GAN framework consists of a generator network and a discriminator network. The generator takes a known distribution and generates new data, while the discriminator is used to distinguish the generated samples from real data distribution. The generator and discriminator play a minimax game.

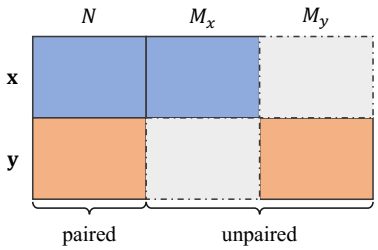


Figure 1: Illustration of multi-modal data with missing modalities.

Recently, several works [Wang *et al.*, 2018; Cai *et al.*, 2018; Shang *et al.*, 2017] have emerged utilizing the generative ability of GAN models to impute missing data. These works mainly focus on data generation using GAN, while few of them explore the downstream tasks, such as metric learning. We combine the data generation process and metric learning to resolve the missing modality issue in patient similarity analysis. In terms of cross modal generation and classification, the proposed model is also related to domain adaptation works [Hoffman *et al.*, 2017; Russo *et al.*, 2018]. However, different from these methods that evaluate only on the target domain, the proposed model is a multi-modal learning, i.e., the prediction utilizes the complementary information of all domains. Moreover, they minimize the difference between domains via an unsupervised mapping, while we learn the domain relationship using both paired and unpaired modalities.

3 Methodology

3.1 Problem Formulation

Performing patient similarity analysis relies on learning a proper distance metric among patients. In metric learning, a linear or non-linear transformation function $f(\cdot)$ maps the input data into a new space. The metric in the transformed space measures the sample distances for a considered task. Without loss of generality, the distance metric between two samples $\mathbf{p}_i$ and $\mathbf{p}_j$ can be written as:

$$d^2(\mathbf{p}_i, \mathbf{p}_j) = \| f(\mathbf{p}_i) - f(\mathbf{p}_j) \|^2 . \quad (1)$$

For each pair of samples $\mathbf{p}_i$ and $\mathbf{p}_j$, a pairwise label g_{ij} denotes whether these two samples are similar or not. If $\mathbf{p}_i$ and $\mathbf{p}_j$ are similar (i.e. they belong to the same group), then g_{ij} is set to 1, otherwise -1. The distance constraints can be constructed via various types of loss functions. In this paper, we minimize the pairwise hinge loss:

$$\mathcal{L}_m = \sum_{\mathcal{T}} \left[1 - g_{ij}(\gamma - d^2(\mathbf{p}_i, \mathbf{p}_j))\right]_+ , \quad (2)$$

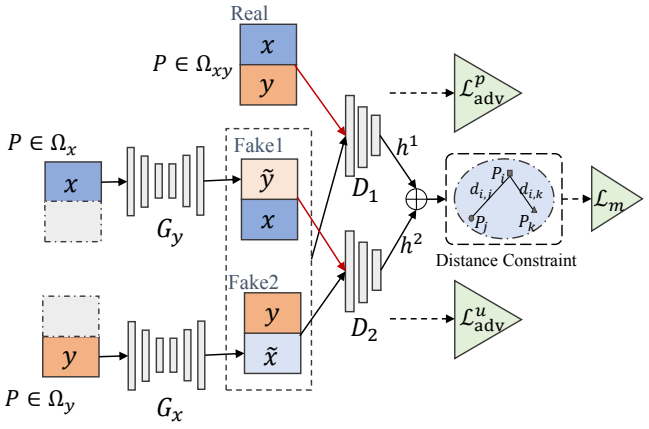


Figure 2: Overall framework.

where $[\cdot]_+ = \max(\cdot, 0)$, γ is the unit margin, and $\mathcal{T} = \{(\mathbf{p}_i, \mathbf{p}_j, g_{ij})\}$ is the set of annotated positive and negative sample pairs. Each patient $\mathbf{p}_k$ ideally has two modalities $\mathbf{x}$ and $\mathbf{y}$, and such patients are considered as complete/paired samples. We assume that there are N complete samples, denoted as $\Omega_{xy} = \{(\mathbf{p}_k^x, \mathbf{p}_k^y)\}_{k=1}^N$. However, modality missing is a common issue in healthcare area. There are two cases in our setting: $\mathbf{x}$ or $\mathbf{y}$ modality missing. The sample set without $\mathbf{y}$ modality is denoted as $\Omega_x = \{\mathbf{p}_k^x\}_{k=1}^{M_x}$, and $\Omega_y = \{\mathbf{p}_k^y\}_{k=1}^{M_y}$ is the set without $\mathbf{x}$ modality. The total number of samples is $N+M_x+M_y$. Figure 1 illustrates the three types of data. In Eq. (2), $\mathbf{p}_i$ can be $(\mathbf{p}_i^x, \mathbf{p}_i^y)$, $\mathbf{p}_i^x$ or $\mathbf{p}_i^y$, and so as $\mathbf{p}_j$.

Due to the issue of incomplete modalities, it is hard to directly apply traditional metric learning to learn similarities among patients. To tackle this issue, we need to design an effective model to automatically infer an appropriate mapping from the *observed modality* to the *missing one*. Thus, the proposed model consists of two main parts: missing modality generation and metric learning, and connects them in an end-to-end way, which is illustrated in Figure 2. We first employ GAN to generate the missing modality based on the observed modality, and then feed the latent representation which contains the multi-modal information into a metric learning layer to learn the distance metric. The two parts are optimized simultaneously in the framework. For simplicity, we use x for $\mathbf{p}^x$ and y for $\mathbf{p}^y$ when it is unambiguous.

3.2 Modality Generation

Intrinsically, multiple modalities share consistency with each other and can provide complementary information together. Through learning the hidden relationship between modalities, the missing modality can be reconstructed according to the observed one. To achieve this goal, we develop the following framework for modality completion.

The proposed generative network includes two generators $G_y : \mathbf{x} \to \mathbf{y}$ and $G_x : \mathbf{y} \to \mathbf{x}$. The goal is to train the generator networks to infer the missing modality from the observed one. From a probabilistic perspective, suppose that x is drawn from the distribution $p_x(x)$, and y is drawn from $p_y(y)$. The generator G_y characterizes the conditional distribution $p_y(y|x)$, and G_x defines the conditional distribution in the other direction $p_x(x|y)$. In the generation process, a sample x is drawn from the data belonging to domain $\mathbf{x}$, and then the generator G_y produces a fake sample $\tilde{y} = G_y(x)$ in domain $\mathbf{y}$ following $p_y(y|x)$. Hence, the fake sample pair $(x, \tilde{y})$ follows the joint distribution $p_{G_y}(x, y) = p_y(y|x)p_x(x)$. Similarly, a fake pair $(\tilde{x}, y)$ can be obtained from G_x, i.e. $\tilde{x} = G_x(y)$, following $p_{G_x}(x, y) = p_x(x|y)p_y(y)$.

Loss of Paired Data

For the paired training data, both of the two modalities $\mathbf{x}$ and $\mathbf{y}$ are available. We use a discriminator D_1 to distinguish whether a sample (x, y) is real or fake. Since the incompleteness exists in both modalities in our problem, this requires the proposed model to be able to generate data from two directions. It leads to the failure of the unidirectional conditional generation process proposed in [Cai *et al.*, 2018]. Therefore, we develop the following objective function to enable the conditional generation in two directions:

$$\begin{aligned}\min_{G_x,G_y}\max_{D_1}\mathcal{L}^p_{adv} = &\mathbb{E}_{x\sim p_x, y\sim p_y}\log(D_1(x,y)) \\ &+\mathbb{E}_{x\sim p_x}\log(1-D_1(x,\tilde{y})) \\ &+\mathbb{E}_{y\sim p_y}\log(1-D_1(\tilde{x},y)).\end{aligned} \tag{3}$$

The adversarial loss is optimized based on a minimax game. We treat D_1 as a binary classification network. The true data sample (x, y) is set with label 1, and the predicted samples $(x, \tilde{y})$ and $(\tilde{x}, y)$ are given label 0. We minimize the binary cross-entropy loss $\mathcal{L}_{ce}(\hat{c}, c) = -(c\log(\hat{c}) + (1-c)\log(1-\hat{c}))$ to train the classifier. Therefore, the loss function of D_1 can be expressed as:

$$\begin{aligned}\mathcal{L}_{D_1} = &\mathcal{L}_{ce}(D_1(x,y),1) + \mathcal{L}_{ce}(D_1(x,\tilde{y}),0) \\ &+\mathcal{L}_{ce}(D_1(\tilde{x},y),0).\end{aligned} \tag{4}$$

To minimize Euclidean distance between the predicted and true data, the mean squared error loss $\mathcal{L}_{mse}$ is calculated as:

$$\mathcal{L}_{mse}(x, y, \tilde{x}, \tilde{y}) = \|x - \tilde{x}\|_2^2 + \|y - \tilde{y}\|_2^2. \tag{5}$$

Loss of Unpaired Data

In real world applications, it is usually hard to obtain complete data. The small amount of available complete data may not provide enough information to characterize the dataset. The incomplete data contain partial extra information and are expected to improve the performance of downstream tasks. Therefore, how to incorporate the incomplete data into training is the issue that we aim to resolve. For unpaired data, we lack supervision in the form of paired modalities. To make use of the unpaired data, motivated by [Dumoulin *et al.*, 2016; Donahue *et al.*, 2016], we use the following equation to enable joint distribution matching:

$$\begin{aligned}\min_{G_x,G_y}\max_{D_2}\mathcal{L}^u_{adv} = &\mathbb{E}_{y\sim p_y}\log D_2(\tilde{x},y) \\ &+\mathbb{E}_{x\sim p_x}\log(1-D_2(x,\tilde{y})),\end{aligned} \tag{6}$$

where the discriminator D_2 is used to distinguish whether a fake modality pair is from $p(x, \tilde{y})$ or $p(\tilde{x}, y)$. Similarly to D_1, the loss function of D_2 is:

$$\mathcal{L}_{D_2} = \mathcal{L}_{ce}(D_2(\tilde{x},y),1) + \mathcal{L}_{ce}(D_2(x,\tilde{y}),0). \tag{7}$$

Loss of Generation

The generators G_x and G_y learn the mapping between two domains, and are optimized to make the imputed data hard to be distinguished from the real data. Following the optimization strategy in [Goodfellow *et al.*, 2014], we train the generators with the objective function:

$$\begin{aligned}\mathcal{L}_{G_x,G_y} &= \mathcal{L}_{ce}(D_1(x,\tilde{y}),1) + \mathcal{L}_{ce}(D_1(\tilde{x},y),1) \\ &+ \mathcal{L}_{ce}(D_2(\tilde{x},y),0) + \mathcal{L}_{ce}(D_2(x,\tilde{y}),1) \\ &+ \mathcal{L}_{mse}(x,y,\tilde{x},\tilde{y}).\end{aligned} \tag{8}$$

The generators and discriminators are all deep neural networks. In particular, the generators G_x and G_y follow a U-net encoder-decode network [Ronneberger *et al.*, 2015] structure with skip connections between layers. The discriminators D_1 and D_2 extract high-level representations h^1 and h^2 respectively from their input sample pairs. The representation vectors h^1 and h^2 are not only used for discriminating the sample pairs, but also for performing the metric learning task.

3.3 Metric Learning Layer

Through the generators G_x and G_y in Section 3.2, we can predict the missing modalities and obtain three types of data (x,y), $(\tilde{x},y)$ and $(x,\tilde{y})$. We then perform the multi-modal metric learning on the whole training data. We make use of the latent representations of the discriminators to perform metric learning, by training the distance metric as an auxiliary task [Odena *et al.*, 2017] in the discriminator. In this way, the adversarial network and metric learning layer are optimized in an end-to-end way. The similarity information learned by metric learning is expected to improve the discriminability of the generated domains, i.e., a generated sample should be more similar to the samples in the same group and dissimilar to the samples from another group. Meanwhile, the high quality and discriminability of the generated data in return can improve metric learning performance.

During the training process, a sample $\mathbf{p}_i \in \Omega_x$ with only modality x_i is fed into G_y to generate the missing modality $\tilde{y}_i$, and then a complementary sample $\mathbf{p}_i = (x_i, \tilde{y}_i)$ can be derived. The sample $(x_i, \tilde{y}_i)$ is fed into discriminators D_1 and D_2 for adversarial training as described in Section 3.2. Meanwhile, latent representations h_i^1 and h_i^2 are extracted by D_1 and D_2, respectively. The vectors h_i^1 and h_i^2 are non-linear abstract representations which capture the characteristics of $(x_i, \tilde{y}_i)$. We then transform h_i^1 and h_i^2 to obtain a latent vector h_i, i.e., $h_i = f(h_i^1 \oplus h_i^2)$, where $\oplus$ is the concatenation operator and $f(\cdot)$ is a fully-connected layer. Here h_i can be considered as the representation in the transformed space for $\mathbf{p}_i$ in Eq. (1). Similarly, the vector representations for (x_j, y_j) and $(\tilde{x}_k, y_k)$ are obtained from the latent vectors learned from D_1 and D_2. Note that (x_j, y_j) does not contribute to the adversarial loss of D_2. After obtaining the representations for a batch of samples, including $\{(x_i, \tilde{y}_i)\}$, $\{(x_j, y_j)\}$, and $\{(\tilde{x}_k, y_k)\}$, we calculate the metric learning loss in Eq. (2) for sample pairs in the mini-batch.

3.4 The Learning Framework

In the proposed method, the generators are trained to produce fake data which can fool the discriminators. The discriminators produce not only the probability distribution of real/fake modality pairs, but also the distance distribution of sample pairs. Metric loss $\mathcal{L}_m$ is involved to optimize the parameters of discriminators and generators. Generators are trained to minimize $\alpha\mathcal{L}_m + \mathcal{L}_{adv}^u + \mathcal{L}_{adv}^p$, and discriminators are trained to minimize $\alpha\mathcal{L}_m - \mathcal{L}_{adv}^u - \mathcal{L}_{adv}^p$, where α is a trade-off parameter. In the training process, we first randomly choose a batch of paired and unpaired samples, and then use generators to predict the corresponding fake data. Generator and discriminator are updated iteratively by fixing one and updating the other. When it comes to testing, we use the generators to obtain the imputed data, and the metric learning network to obtain the distance metric.

4 Experiments

4.1 Dataset

The Alzheimer's Disease Neuroimaging Initiative (ADNI) [1] project aims to track the progression of Alzheimer's disease using biomarkers and clinical measures. There are 840 patients in three cohorts: 231 cognitively normal (CN), 410 mild cognitive impairment (MCI), and 199 Alzheimer's disease (AD) patients. In this work, we use the available modalities in the database: MRI images and PET images, and generate the missing images from each other. We first preprocess the MRI and PET images for each patient. The T1-weighted MRI images are skull-stripped and intensity inhomogeneity corrected. After that, each MRI is segmented into gray matter, white matter and cerebrospinal fluid, and further spatially normalized into a template space using SPM[2] and CAT12[3] softwares. In the experiments, we use the gray matter tissue density maps as the MRI modality. The PET modality is coregistered, spatially normalized and rigidly aligned to the MRI modality. The MRI gray matter images and PET images are further smoothed using a unit standard deviation Gaussian kernel. To reduce the computational cost, we downsample the images to 32×32 2D slices as the inputs.

4.2 Experimental Setup

Baseline Approaches

We compare the proposed metric learning framework on the incomplete dataset with the state-of-the-art approaches. For the baseline approaches, we follow a two-step strategy: first impute the missing modalities using data completion methods, and then perform multi-modal metric learning. The approaches for imputing missing modalities include: Multimodal Autoencoder (MultiAE), Pix2pix [Isola *et al.*, 2017] and CycleGAN [Zhu *et al.*, 2017]. The approaches for multimodal metric learning include: LM3L [Hu *et al.*, 2014], FISH-MML [Zhang *et al.*, 2018] and HM3L [Zhang *et al.*, 2017]. Among the imputation approaches, MultiAE and Pix2pix are trained on paired data, and CycleGAN is trained on unpaired data. All the methods have the same generator and discriminator network structure as the proposed method, but different loss functions. For the metric learning approaches, since they cannot handle incomplete examples, we

[1] https://adni.loni.usc.edu/

[2] https://www.fil.ion.ucl.ac.uk/spm/

[3] http://www.neuro.uni-jena.de/cat/

Datasets	Methods	5% paired				20% paired			
		Classification		Clustering		Classification		Clustering	
		Accuracy	F1	Purity	RI	Accuracy	F1	Purity	RI
CN-AD	LM3L	.507(.019)	.425(.030)	.550(.032)	.506(.006)	.525(.024)	.396(.081)	.540(.030)	.504(.007)
	FISH-MML	.663(.019)	.559(.030)	.514(.000)	.500(.000)	.665(.011)	.596(.016)	.517(.006)	.500(.001)
	HM3L	.577(.017)	.447(.030)	.523(.006)	.501(.001)	.635(.004)	.480(.014)	.514(.001)	.500(.000)
	MultiAE	.698(.017)	.667(.015)	.692(.023)	.575(.018)	.727(.013)	.701(.014)	.722(.013)	.599(.012)
	Pix2pix	.728(.011)	.701(.015)	.723(.011)	.600(.010)	.734(.010)	.714(.012)	.729(.008)	.605(.007)
	CycleGAN	.719(.002)	.696(.005)	.705(.019)	.584(.015)	.732(.013)	.710(.020)	.726(.013)	.602(.012)
	MeLIM	**.756(.004)**	**.737(.009)**	**.755(.006)**	**.630(.006)**	**.783(.005)**	**.759(.005)**	**.782(.005)**	**.659(.006)**
MCI-AD	LM3L	.499(.037)	.558(.023)	.557(.038)	.508(.012)	.519(.023)	.583(.033)	.531(.007)	.500(.001)
	FISH-MML	.583(.015)	.663(.011)	.555(.033)	.507(.009)	.576(.023)	.645(.017)	.540(.024)	.502(.006)
	HM3L	.547(.026)	.640(.022)	.584(.002)	.513(.001)	.583(.007)	.666(.004)	.575(.020)	.511(.006)
	MultiAE	.645(.035)	.660(.035)	.649(.025)	.545(.015)	.672(.012)	.688(.011)	.665(.009)	.554(.006)
	Pix2pix	.672(.012)	.688(.011)	.665(.009)	.554(.006)	.709(.010)	.723(.009)	.703(.012)	.582(.010)
	CycleGAN	.700(.008)	.714(.007)	.692(.020)	.574(.014)	.703(.007)	.720(.010)	.695(.013)	.576(.010)
	MeLIM	**.708(.006)**	**.732(.007)**	**.706(.006)**	**.584(.005)**	**.731(.005)**	**.761(.004)**	**.725(.007)**	**.601(.006)**

Table 1: Performance evaluation of learned distance metrics on two datasets.

first fill in the missing modalities with the imputation method that gives the best performance among baselines. We also conduct experiments by first completing the data using different modality completion methods, and then training a deep metric learning network on each imputed dataset to optimize the loss in Eq. (2). The metric learning network has the same structure as the discriminator in the proposed framework.

Implementation Details and Measurement

We randomly divide the patient set into training, validation and testing sets in a 0.75:0.05:0.2 ratio. Adam optimizer is used for models with deep architectures. We set the learning rate and the network structures the same as [Cai *et al.*, 2018] but in a 2D fashion. We evaluate the learned distance metric in two tasks: disease prediction and patient clustering. For the disease prediction task, k-nearest neighbor (KNN) classification is performed on the learned distance metric. For each testing patient, we assign the predicted label with the most common class label among the top-k neighbors from the training set. We set $k = 3$ in the KNN classifier, and use accuracy and F1 as measures for the comparison. For the patient clustering task, we perform k-means algorithm with $k = 2$ on the testing examples based on the learned distance metric. The quality of clustering is reported in terms of purity and Rand index (RI). We also evaluate the performance of data generation part, by employing structural similarity (SSIM) [Cai *et al.*, 2018] to quantitatively measure the structural difference between predicted images and the corresponding true ones. The SSIM values are in the range of [0,1], and the higher the better.

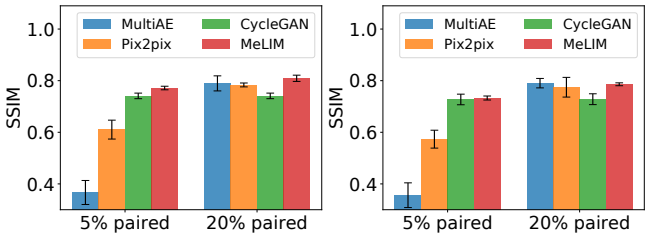
4.3 Experimental Results

We evaluate the performance on two tasks separately: distinguishing CN and AD patients (CN-AD), and MCI and AD patients (MCI-AD). Each task can be viewed as a dataset. For the training set, we use a small ratio of paired data and the available unpaired data. For the testing set, we keep half paired data and half unpaired data. We repeat all the approaches five times on the two datasets.

Patient Similarity Analysis

There are two main ways for similarity comparison of incomplete data. One way is to first complete the missing modalities and then conduct metric learning. The other is to integrate missing modality imputation and metric learning into an end-to-end learning framework. For the baseline algorithms, we follow the two-step procedure. For the proposed method, image generation and metric learning are trained jointly.

In Table 1, we present the average and standard deviation values of different methods by varying the ratio of paired data in the training set (i.e. 5% paired and 20% paired). In the table, we can see that the proposed method MeLIM outperforms baselines in terms of both disease classification and patient clustering tasks. Since we measure the distance metric on the testing set with both complete and imputed samples, we expect that the imputed data can be both realistic and distinguishable. In general, if the imputed data are in high quality, the performance of the downstream task (i.e. metric learning in our problem) will be better. MultiAE and Pix2pix can utilize only paired data samples. The number of available data is so small that they cannot learn a good mapping. Since MultiAE optimizes the content loss $\mathcal{L}_{mse}$, it suffers from the so-called "blurry" problem. CycleGAN is an unsupervised image-to-image translation method, and it suffers from the lack of supervision. These data completion methods optimize the data generation process, but can not guarantee a good representation for learning the similarity among patients. MeLIM optimizes modality generation and metric learning simultaneously, and thus obtains the best results. The traditional multi-modal metric learning models, LM3L, FISH-MML and HM3L learn linear transformations to map the original inputs to new spaces, which is not suitable for high-dimensional complex dataset. Therefore, they cannot obtain satisfactory results for lack of the ability to extract precise representations. MeLIM utilizes the latent space learned by the discriminator which has a deep architecture, and thus high level features which better characterize the inputs can be extracted. For each method, the performance on



(a) CN-AD dataset (b) MCI-AD dataset

Figure 3: SSIM values between generated images and corresponding true images on two datasets.

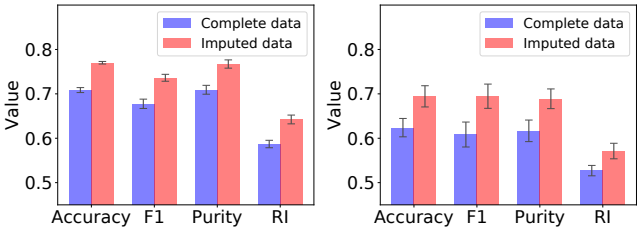
the CN-AD dataset is generally better than that on the MCI-AD dataset. Since MCI is the mild stage of AD, it is easier to distinguish AD patients from CN groups than from MCI groups. We also observe that using more modality complete samples during training leads to better performance. This demonstrates that supervised information helps the data generation process, and the higher quality of generated data can help to improve downstream tasks.

Data Quality Comparison

In this subsection, we show that the proposed framework not only learns accurate distance metric, but also preserves the quality of generated images. We calculate the SSIM values between the predicted images and the true images. Figure 3 shows the results using different modality completion methods. In the figure, we provide the average SSIM values of MRI and PET generations. It can be seen that MultiAE and Pix2pix are not able to generate high quality data when there is not enough paired training data. CycleGAN is trained on the whole datasets without paired information, so that its results do not change with the ratio of paired data. Since the proposed method takes into consideration both the paired and unpaired data, it generates high quality images that are more similar to the ground truth. By comparing the results on the data with different ratios of complete samples, we can see that using 20% paired samples is generally better than using 5% paired samples, especially for MultiAE and Pix2pix. This is due to the fact that more complete samples can provide more supervision to the data generation process. In the diagnosis of AD, doctors focus more on some critical regions and less on the whole images. Therefore, in this work, we mainly focus on the patient similarity analysis using the generated images, and do not emphasize too much on the overall image quality.

Complete vs. Imputed Data

To validate the importance of modality imputation, we conduct the following experiments by comparing the metric learning performance on complete and imputed dataset. We train a deep metric learning model using only the complete/paired data in the training set, and report model performance on the testing set of complete data. For comparison, we train MeLIM using the whole training set (including both paired and unpaired data), and report the performance on the same testing set. The comparison results between the above two strategies can be seen in Figure 4. We conduct experiments on the two datasets with 5% paired training samples. We observe that by incorporating the imputed data during training, metric learning can achieve much better results. As the majority of patients have missing modalities, the information in the complete dataset may not be sufficient, so that metric learning model cannot be well optimized using only complete data. Our proposed method enables metric learning model to incorporate the incomplete data into the training process. The imputed data provides useful information to improve the performance of metric learning.



(a) CN-AD dataset (b) MCI-AD dataset

Figure 4: Performance comparison between complete and imputed data on two datasets.

5 Conclusion and Discussion

The problem of incomplete modalities hampers the development of metric learning approaches on healthcare domain. Moreover, the incomplete data contains extra information which should not be abandoned. To tackle this issue, we propose a new framework to perform metric learning on healthcare data with incomplete modalities. In the proposed method, modality imputation and metric learning are conducted simultaneously in an end-to-end way. In the modality generation part, we incorporate both complete and incomplete data in the learning process, making use of the complementary information contained in the two types of data. Meanwhile, the non-linear high-level representations of both complete and imputed samples are extracted by the discriminators, and then fed into a metric learning layer as an auxiliary task. Experiments show that the proposed model learns an accurate distance metric and also generates high quality data. The proposed method can be generalized for other types of data, by adjusting the network structures of generators and discriminators. The future research may address the scenario in which the datasets contain more than two modalities, but there exists incomplete or missing data in some modalities.

Acknowledgments

The authors would like to thank the anonymous reviewers for their valuable comments. This work was supported in part by the US National Science Foundation under grants NSF-IIS 1553411. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[Cai *et al.*, 2018] Lei Cai, Zhengyang Wang, Hongyang Gao, Dinggang Shen, and Shuiwang Ji. Deep adversarial learning for multi-modality missing data completion. In *Proceedings of SIGKDD*, pages 1158–1166, 2018.

[Donahue *et al.*, 2016] Jeff Donahue, Philipp Krahenbuhl, and Trevor Darrell. Adversarial feature learning. *arXiv preprint arXiv:1605.09782*, 2016.

[Du *et al.*, 2018] Changying Du, Changde Du, Xingyu Xie, Chen Zhang, and Hao Wang. Multi-view adversarially learned inference for cross-domain joint distribution matching. In *Proceedings of SIGKDD*, pages 1348–1357, 2018.

[Dumoulin *et al.*, 2016] Vincent Dumoulin, Ishmael Belghazi, Ben Poole, Olivier Mastropietro, Alex Lamb, Martin Arjovsky, and Aaron Courville. Adversarially learned inference. *arXiv preprint arXiv:1606.00704*, 2016.

[Gan *et al.*, 2017] Zhe Gan, Liqun Chen, Weiyao Wang, Yuchen Pu, Yizhe Zhang, Hao Liu, Chunyuan Li, and Lawrence Carin. Triangle generative adversarial networks. In *Proceedings of NeurIPS*, pages 5247–5256, 2017.

[Goodfellow *et al.*, 2014] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Proceedings of NeurIPS*, pages 2672–2680, 2014.

[Hoffman *et al.*, 2017] Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei A Efros, and Trevor Darrell. Cycada: Cycle-consistent adversarial domain adaptation. *arXiv preprint arXiv:1711.03213*, 2017.

[Hu *et al.*, 2014] Junlin Hu, Jiwen Lu, Junsong Yuan, and Yap-Peng Tan. Large margin multi-metric learning for face and kinship verification in the wild. In *Proceedings of ACCV*, pages 252–267, 2014.

[Huai *et al.*, 2018] Mengdi Huai, Chenglin Miao, Qiuling Suo, Yaliang Li, Jing Gao, and Aidong Zhang. c. In *Proceedings of SDM*, pages 270–278, 2018.

[Isola *et al.*, 2017] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. pages 1125–1134, 2017.

[Li *et al.*, 2014] Rongjian Li, Wenlu Zhang, Heung-Il Suk, Li Wang, Jiang Li, Dinggang Shen, and Shuiwang Ji. Deep learning based imaging data completion for improved brain disease diagnosis. In *Proceedings of MICCAI*, pages 305–312, 2014.

[Li *et al.*, 2018] Yan Li, Tao Yang, Jiayu Zhou, and Jieping Ye. Multi-task learning based survival analysis for predicting alzheimer's disease progression with multi-source block-wise missing data. In *SDM*, pages 288–296, 2018.

[Ni *et al.*, 2017] Jiazhi Ni, Jie Liu, Chenxin Zhang, Dan Ye, and Zhirou Ma. Fine-grained patient similarity measuring using deep metric learning. In *Proceedings of CIKM*, pages 1189–1198, 2017.

[Odena *et al.*, 2017] Augustus Odena, Christopher Olah, and Jonathon Shlens. Conditional image synthesis with auxiliary classifier gans. In *Proceedings of ICML*, pages 2642–2651, 2017.

[Ronneberger *et al.*, 2015] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Proceedings of MICCAI*, pages 234–241, 2015.

[Russo *et al.*, 2018] Paolo Russo, Fabio M Carlucci, Tatiana Tommasi, and Barbara Caputo. From source to target and back: symmetric bi-directional adaptive gan. In *Proceedings of CVPR*, pages 8099–8108, 2018.

[Shang *et al.*, 2017] Chao Shang, Aaron Palmer, Jiangwen Sun, Ko-Shin Chen, Jin Lu, and Jinbo Bi. Vigan: Missing view imputation with generative adversarial networks. In *Proceedings of BigData*, pages 766–775, 2017.

[Suo *et al.*, 2018] Qiuling Suo, Weida Zhong, Fenglong Ma, Yuan Ye, Mengdi Huai, and Aidong Zhang. Multi-task sparse metric learning for monitoring patient similarity progression. In *Proceedings of ICDM*, pages 477–486, 2018.

[Wang *et al.*, 2018] Qianqian Wang, Zhengming Ding, Zhiqiang Tao, Quanxue Gao, and Yun Fu. Partial multi-view clustering via consistent gan. In *Proceedings of ICDM*, pages 1290–1295, 2018.

[Xie and Xing, 2013] Pengtao Xie and Eric Xing. Multi-modal distance metric learning. In *Proceedings of IJCAI*, 2013.

[Yang *et al.*, 2018] Yang Yang, De-Chuan Zhan, Xiang-Rong Sheng, and Yuan Jiang. Semi-supervised multi-modal learning with incomplete modalities. In *Proceedings of IJCAI*, pages 2998–3004, 2018.

[Yuan *et al.*, 2018] Ye Yuan, Xun Guangxu, Fenglong Ma, Yaqing Wang, Nan Du, Kebin Jia, Lu Su, and Aidong Zhang. Muvan: A multi-view attention network for multivariate temporal data. In *Proceedings of ICDM*, pages 717–726, 2018.

[Zhan *et al.*, 2016] Mengting Zhan, Shilei Cao, Buyue Qian, Shiyu Chang, and Jishang Wei. Low-rank sparse feature selection for patient similarity learning. In *Proceedings of ICDM*, pages 1335–1340, 2016.

[Zhang *et al.*, 2017] Heng Zhang, Vishal M Patel, and Rama Chellappa. Hierarchical multimodal metric learning for multimodal classification. In *Proceedings of CVPR*, pages 3057–3065, 2017.

[Zhang *et al.*, 2018] Changqing Zhang, Yeqinq Liu, Yue Liu, Qinghua Hu, Xinwang Liu, and Pengfei Zhu. Fish-mml: Fisher-hsic multi-view metric learning. In *Proceedings of IJCAI*, pages 3054–3060, 2018.

[Zhu *et al.*, 2017] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. pages 2223–2232, 2017.