HHS Public Access

Author manuscript

Deep Learn Med Image Anal Multimodal Learn Clin Decis Support (2017). Author manuscript; available in PMC 2017 November 02.

Published in final edited form as:

Deep Learn Med Image Anal Multimodal Learn Clin Decis Support (2017). 2017 September ; 10553: 160–168. doi:10.1007/978-3-319-67558-9_19.

Multi-stage Diagnosis of Alzheimer's Disease with Incomplete Multimodal Data via Multi-task Deep Learning

Kim-Han Thung, Pew-Thian Yap, and Dinggang Shen

Department of Radiology and BRIC, University of North Carolina, Chapel Hill, USA

Abstract

Utilization of biomedical data from multiple modalities improves the diagnostic accuracy of neurodegenerative diseases. However, multi-modality data are often incomplete because not all data can be collected for every individual. When using such incomplete data for diagnosis, current approaches for addressing the problem of missing data, such as imputation, matrix completion and multi-task learning, implicitly assume linear data-to-label relationship, therefore limiting their performances. We thus propose multi-task deep learning for incomplete data, where prediction tasks that are associated with different modality combinations are learnt jointly to improve the performance of each task. Specifically, we devise a multi-input multi-output deep learning framework, and train our deep network subnet-wise, partially updating its weights based on the availability of modality data. The experimental results using the ADNI dataset show that our method outperforms the state-of-the-art methods.

1 Introduction

Neuroimaging technologies such as Magnetic Resonance Imaging (MRI) and Positron Emission Topography (PET) afford different views of the human brain, both in terms of structure and function. Utilizing neuroimaging data jointly with other non-imaging data can significantly improve diagnostic accuracy [13]. However, multimodal or longitudinal datasets are quite often incomplete because not all data can be collected for every individual. We will show in this work how the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, which is incomplete, can be used effectively for diagnosis.

Though incomplete data can still be used by discarding incomplete samples, this will reduce the sample size. Doing so with the ADNI data will cause a loss of about 50% of the data. To use the data without discarding samples, one can impute the missing values using methods like k -nearest neighbor, expectation maximization, low-rank matrix completion (LRMC), etc. [2,11]. However, these imputation methods only work well if data are missing in random. This is seldom the case in clinical study design, e.g., data in ADNI are missing in blocks (e.g., an entire modality is missing). To tackle this problem, Goldberg et al. [4] propose to simultaneously impute the missing data and the diagnostic labels using low rank

matrix completion. This approach is improved in [7] by selecting discriminative features and samples before matrix completion.

Another approach to tackle this problem is by using multi-task learning [16]. Specifically, this method first divides the incomplete dataset into several subsets of complete data with different combinations of modalities. Then, a classification model is trained for each subset, with all models trained jointly across subsets. This avoids the need to perform imputation and allows utilization of all available data. However, like the matrix completion method, this method still assumes *linear* relationship between the predictors (multimodal data) and the targets (class labels), therefore imposing a strong assumption on their relationships [13].

Deep learning can learn complex nonlinear mappings, but most of the deep learning models assume the data are complete. Though some models are applicable to incomplete data, almost all of them are based on data synthesis/imputation [5,6], which may not work well for large amount of missing data [7]. In this work, we show how one can use deep learning on incomplete multi-modal data via multi-task learning, without involving data synthesis. Similar to [16], we associate a learning task with each data subset and learn these tasks jointly. More specifically, we design a multi-input multi-output deep learning framework, i.e., one input for one modality and one output for one learning task, and train the network iteratively and partially, based on data availability. The proposed method allows multi-class classification, in contrast to most current methods that focus on binary classification [1,7–10,15,16].

2 Method

2.1 Multi-task Learning

Figure 1 shows a typical multi-task learning (MTL) framework for dealing with classification problems involving incomplete multimodal data [16]. Without loss of generality, here we assume that there are two modalities, denoted by M1 and M2, which could be incomplete (i.e., some subjects do not have data from M1 or M2). Due to the missing data, we cannot build a prediction model directly. However, by grouping the data into subsets of complete data with different combinations of modalities, we can define a classification task for each of these subsets (right in Fig. 2). Yuan et al. [16] then learned these tasks jointly, by constraining all the tasks to select common features for each modality. However, there are several limitations in Yuan's method. First, it implicitly assumes linear data-to-label relationship. Second, it did not utilize the maximum number of samples for each task. For example, for Task 1 where the relationship between modality M1 and target Y is to be learned, only a portion of M1 samples is used (left in Fig. 1). Third, it uses original features of all the modalities for classification, ignoring heterogeneity among modalities. Fourth, its current formulation only supports binary classification. Many of these weaknesses can be readily overcome by using deep learning.

2.2 Multi-task Deep Learning for Incomplete Multimodal Data

Figure 2 shows an overview of the proposed Multi-Task Deep Learning (MTDL) framework for two incomplete modalities. This is a multi-input multi-output deep learning framework,

as we have two types of input (i.e., one for each modality), and three types of output (i.e., one for each classification task). The architecture consists of mainly three layers, i.e., input layer, hidden layers, and output layer. Each blue rectangular box denotes a set of neurons and the arrows denote the connections between the neurons. Note that there are two sets of neurons in the input layer, each of which receives input data from one modality. In addition, there are three sets of neurons in the output layer, which correspond to three different classification tasks, i.e., one for each combination of modalities. Each classification task is learned by using the maximum number of available samples, as shown on the right side of Fig. 2. There are two types of hidden layers in our framework, i.e., “modality-specific layers” and “task-specific layers”, focusing on learning the latent data representation for specific modality and also learning the data-to-label relationship for specific classification task, respectively. Details on these two layers are described next.

Modality-Specific Layers—Heterogeneity is inevitable in multimodal data. In our case, we use MRI and PET data for information on structure and functional. The number of samples for each modality is different. PET data have far less number of samples compared with MRI data. Thus, to use the maximum number of samples to effectively learn the latent representation, i.e., high-level features for each modality, we adopt modality-specific hidden layers in our framework. The number of neurons and the number of layers can be different for different modalities.

Task-Specific Layers—A task is a classification problem associated with each modality combination. Using the example in Fig. 2, we have a total of three tasks, i.e., (Task 1) classification using only M1 data, (Task 2) classification using only M2 data, and (Task 3) classification using M1+M2 data. In “task-specific layers”, we first concatenate the high-level features learned from each modality (via “modality-specific layers”) based on the combination mentioned above for all the three classification tasks (Fig. 2). Then we learn the data-to-label mapping for each task in this layer.

Iterative Subnet Training—Training the network in Fig. 2 is not straightforward, as the number of samples differs across modalities. To maximally utilize all samples during training, we train the MTDL network in Fig. 2 iteratively and subnet-wise, as shown in Fig. 3. Again, using the example of Fig. 2, we first divide the data into three subsets, i.e., M1, M2, and M1+M2, each consists of maximum number of samples for each data combination of modalities. Then, in each training batch, only a certain part of the network (i.e., subnet) is updated, using the corresponding subset of data.

Implementation—We used Keras [3] to implement our framework. To regularize the network weights, we use $l_{2,1}$ -norm plus l_2 -norm for the first hidden layer, and use l_2 -norm for other hidden layers. We use $l_{2,1}$ -norm to regularize the network weights as it is useful for feature selection by imposing row sparsity on the network weight matrix. We use softmax activation function for the output layer, and use categorical crossentropy as the loss function for the output of each task. We also divided our data into ten parts according to [14], and then use eight parts for training, one part for validation, and the remaining one part for testing. For each subnet, we train it using all the available training samples that contain the

needed modality combination for 200 epochs, and stop the training if the validation loss stops improving for 20 epochs. We then update the network using the network weights that is obtained when the validation loss is minimum, for each subnet training. We iteratively trained all the subnets for several iterations (Fig. 3) until convergence condition (e.g., no improvement of loss minimization) or maximum number of iterations was met.

3 Materials, Preprocessing and Feature Extraction

In this study, we used MRI and PET data from the ADNI dataset¹ (MRI data are complete, while PET data are incomplete, at baseline) for multi-stage AD diagnosis, i.e., to differentiate between Normal Control (NC), Mild Impairment Cognitive (MCI - prodromal stage of AD) and Alzheimer's Disease (AD). In addition, to improve the diagnostic performance, we also included demographic information (i.e., age, gender, education) and genetic information (i.e., Apoe4). Moreover, we also used samples from all the time points for each training subject to significantly increase the number of samples for training. We summarize the demographic information of the subjects used in this study in Table 1. From the Table 1, we can see that 805 samples (here we define a sample as the data collected from one subject at a certain time point), which corresponded to 805 subjects used in this study, were collected at the baseline. After included the samples from all the time points for these 805 subjects, we ended up with a total of 4830 samples in this study.

We extracted the region-of-interest (ROI)-based features from the MRI and PET images by using the following processing pipeline. Each MR image was AC-PC aligned, N3 intensity inhomogeneity corrected, skull stripped, tissue segmented, and registered to a template with 93 ROIs [12]. We then used the normalized Gray matter (GM) volumes from the 93 ROIs as MRI features. We also linearly aligned each PET image to its corresponding MRI image, and used the mean intensity values of each ROI as a PET feature.

4 Results and Discussions

We compared our proposed method with a Baseline method (i.e., sparse multi-label feature selection followed by a multi-class support vector classifier) that uses only the complete portion of the data, and two other popular methods that use incomplete multimodal data, i.e., low-rank matrix completion with sparse feature selection (LRMC) [7], and incomplete multi-source joint feature learning (iMSF) [16]. We used the standardized cross-validation indices in [14] to perform 10-fold cross validation experiment for all the comparison methods, and reported the mean accuracy as the evaluation metric. The accuracy of the classification is defined as the total number of correctly classified samples divided by the total number of samples.

We performed two experiments. The first experiment (i.e., Exp.1) used incomplete multimodal data for multi-stage AD diagnosis via the proposed multitask deep learning method. The second experiment (i.e., Exp.2) used extra demographic information (i.e., age,

¹<http://adni.loni.ucla.edu>.

gender, education and Apoe4). We appended this extra information to the MRI and PET data to increase their feature dimension to 97 ($93 + 4$) and use the same framework to solve the problem. Figure 4 shows the results of these two experiments using the proposed framework. The two bar groups on the left are for the first experiment, and the other two on the right are for the second experiment. There are 4 types of results for each group of bars, namely “MRI”, “MRI+PET”, “Overall-bl” and “Overall”, which respectively denote the results for prediction task that uses only MRI data, prediction task that uses MRI and PET data, overall classification result (for all the prediction tasks) using all the baseline testing samples, and overall classification result (for all the prediction tasks) using all the testing samples at different time points. As can be seen, “MRI+PET” in general has higher classification accuracy than others, as it uses more information. To show that all the tasks in our framework are complementary to each other, and training them jointly and iteratively will improve each individual task, we also present the results of the first and last iterations, respectively denoted by “Initial” and “Final” in Fig. 4. As can be seen from Fig. 4, the final network in general has higher classification accuracy than the initial network, e.g., for the first experiment, the initial classification accuracy using MRI data is 55.5%, but after several iterations, it is improved to 58.0%. Similarly for the second experiment, the classification accuracy using MRI+PET data has improved from 62.4% to 65.8%. Also note that, the multistage classification results reported here is lower than the binary classification results reported in the literature [7], as multi-stage classification is much more challenging as it needs to differentiate the subtle differences among NC, MCI and early AD samples simultaneously.

In addition, we also compared our results with the state-of-the-art methods in Table 2, i.e., iMSF [16] and LRMC [7]. We modified the iMSF code to one-vs-others setting for multi-class classification. The baseline method can only work with complete data and therefore used a smaller sample size of data (marked with ‘*’). Its results are not comparable with those that use all the samples (i.e., column 3, 4 and 5 of Table 2). For comparison, we also included our results for these subset of samples in the last column of Table 2, and marked them with ‘*’ as well. From the Table, it can be seen that our methods is better than the compared methods that using either incomplete or complete data. It is interesting that the iMSF algorithm, which also uses multi-task learning strategy, did not perform well for multi-stage classification. This is probably due to its use of linear regression model, which is unable to learn the complex data-to-label relationship in this multi-stage AD study. LRMC is the best performing method after MTDL. It is noteworthy that MTDL can be easily extended to use more complicated neuroimaging data, e.g., the raw images, by simply replacing the fully connected modality-specific layers with the convolutional layers.

5 Conclusion

We have demonstrated how a deep neural network can be applied on incomplete multimodal data by using a multi-task learning strategy, so that no data synthesis is needed. To this end, we have devised a multi-input multi-output deep neural network, which can be trained using an iterative subnet training strategy. The experiment results show that different classification tasks can benefit from each other during training. In addition, the results also show that the

proposed method (MTDL) outperforms other compared methods, such as LRMC and iMSF that were designed for incomplete multimodal data. Moreover, our framework can be easily extended to use more complicated imaging data by replacing the full-connectivity network in the modality-specific layers with the convolutional layers.

Acknowledgments

This work was supported in part by NIH grants NS093842, EB006733, EB008374, EB009634, EB022880, AG041721, MH100217, and AG042599.

References

1. Adeli-Mosabbe E, et al. Robust feature-sample linear discriminant analysis for brain disorders diagnosis. *Advances in Neural Information Processing Systems*. 2015:658–666.
2. Candès EJ, et al. Exact matrix completion via convex optimization. *Found Comput Math*. 2009; 9(6):717–772.
3. Chollet, F. Keras. 2015. <https://github.com/fchollet/keras>
4. Goldberg A, et al. Transduction with matrix completion: three birds with one stone. *Adv Neural Inf Process Syst*. 2010; 23:757–765.
5. Li F, et al. A robust deep model for improved classification of AD/MCI patients. *IEEE J Biomed Health Inf*. 2015; 19(5):1610–1616.
6. Liu S, et al. Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. *IEEE Trans Biomed Eng*. 2015; 62(4):1132–1140. [PubMed: 25423647]
7. Thung KH, et al. Neurodegenerative disease diagnosis using incomplete multi-modality data via matrix shrinkage and completion. *Neuroimage*. 2014; 91:386–400. [PubMed: 24480301]
8. Thung KH, et al. Identification of progressive mild cognitive impairment patients using incomplete longitudinal MRI scans. *Brain Struct Funct*. 2016; 221(8):3979–3995. [PubMed: 26603378]
9. Thung, KH., Yap, PT., Adeli-M, E., Shen, D. Joint diagnosis and conversion time prediction of progressive mild cognitive impairment (pMCI) using low-rank subspace clustering and matrix completion. In: Navab, N.Hornegger, J.Wells, WM., Frangi, AF., editors. *MICCAI 2015 LNCS*. Vol. 9351. Springer; Cham: 2015. p. 527-534.
10. Thung, KH., Adeli, E., Yap, PT., Shen, D. Stability-weighted matrix completion of incomplete multi-modal data for disease diagnosis. In: Ourselin, S.Joskowicz, L.Sabuncu, MR.Unal, G., Wells, W., editors. *MICCAI 2016 LNCS*. Vol. 9901. Springer; Cham: 2016. p. 88-96.
11. Troyanskaya O, et al. Missing value estimation methods for DNA microarrays. *Bioinformatics*. 2001; 17(6):520–525. [PubMed: 11395428]
12. Wang, Y., Nie, J., Yap, PT., Shi, F., Guo, L., Shen, D. Robust deformable-surface-based skull-stripping for large-scale studies. In: Fichtinger, G.Martel, A., Peters, T., editors. *MICCAI 2011 LNCS*. Vol. 6893. Springer; Heidelberg: 2011. p. 635-642.
13. Weiner MW, et al. The Alzheimer's disease neuroimaging initiative: a review of papers published since its inception. *Alzheimer's Dementia*. 2013; 9(5):e111–e194.
14. Wyman BT, et al. Standardization of analysis sets for reporting results from ADNI MRI data. *Alzheimer's Dementia*. 2013; 9(3):332–337.
15. Yu G, et al. Multi-task linear programming discriminant analysis for the identification of progressive MCI individuals. *PLoS ONE*. 2014; 9(5):e96458. [PubMed: 24820966]
16. Yuan L, et al. Multi-source feature learning for joint analysis of incomplete multiple heterogeneous neuroimaging data. *NeuroImage*. 2012; 61(3):622–632. [PubMed: 22498655]

Fig. 1. Left: Original classification problem with incomplete multimodal data. Right: Multi-task learning for incomplete multimodal data. (X: input data; Y: target data; M1, M2: multimodal data; White: Missing data; Grey: Available data)

Fig. 2.
The proposed multi-task deep learning (MTDL) via multi-input multi-output neural network.

Fig. 3.

Iterative subnet training. Subnets (i.e., partial network) in (a), (b) and (c) are corresponding to prediction tasks of Task 1, Task 2, and Task 3, respectively. During the training of each subnet, only the network weights within the red blocks are updated, while the remaining network weights (in gray blocks) are frozen. The network in Fig. 2 is learned iteratively in the sequence of $\{(a) \rightarrow (b) \rightarrow (c) \rightarrow (a) \rightarrow \dots\}$ until convergence.

Fig. 4.

Accuracy of multi-stage AD diagnosis using MRI and PET data for the proposed multi-task deep learning method. Exp.1 and Exp.2 denote the experiments that ignored and used the demographic information, respectively. Error bars: standard deviations.

Table 1

Demographic information of the subjects involved in this study. (Edu.: Education)

	# Subjects	Gender (M/F)	Age (years)	Edu. (years)
NC	226	118/108	75.8 ± 5.0	16.0 ± 2.9
MCI	393	253/140	74.9 ± 7.3	15.6 ± 3.0
AD	186	99/87	75.3 ± 7.6	14.7 ± 3.1
Total	805	470/335	-	-

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Comparison with the baseline, iMSF [16] and LRMC [7].

Data	Baseline	iMSF	LRMC	MTDL
MRI	55.1	54.4	56.9	58.0
MRI+PET	56.9 [*]	54.0	58.5	61.1 65.6 [*]
MRI+PET+DEM	62.0 [*]	58.3	61.9	63.6 65.8 [*]

Bold: Best results,
^{*} Complete data results (less number of samples)