Published in final edited form as:

Med Image Comput Comput Assist Interv. 2018 September ; 11072: 446–454. doi:

10.1007/978-3-030-00931-1_51

Temporal Correlation Structure Learning for MCI Conversion Prediction

Xiaoqian Wang¹, Weidong Cai², Dinggang Shen³, and Heng Huang¹

¹Department of Electrical and Computer Engineering, University of Pittsburgh, Pittsburgh, USA

²School of Information Technologies, University of Sydney, Sydney, Australia

³Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, USA

Abstract

In Alzheimer's research, Mild Cognitive Impairment (MCI) is an important intermediate stage between normal aging and Alzheimer's. How to distinguish MCI samples that finally convert to AD from those do not is an essential problem in the prevention and diagnosis of Alzheimer's. Traditional methods use various classification models to distinguish MCI converters from non-converters, while the performance is usually limited by the small number of available data. Moreover, previous methods only use the data at baseline time for training but ignore the longitudinal information at other time points along the disease progression. To tackle with these problems, we propose a novel deep learning framework that uncovers the temporal correlation structure between adjacent time points in the disease progression. We also construct a generative framework to learn the inherent data distribution so as to produce more reliable data to strengthen the training process. Extensive experiments on the ADNI cohort validate the superiority of our model.

Keywords

Deep learning; Temporal correlation structure; MCI conversion prediction; Alzheimer's disease

1 Introduction

Alzheimer's disease (AD) is a complex chronic progressive neurodegenerative disease that gradually affects human memory, judgment, and behavior. As an important intermediate stage between normal aging and AD, MCI possesses an increased risk of transiting to AD. That being the case, how to recognize the MCI samples with high potential of switching to AD prior to dementia becomes an essential problem in Alzheimer's prophylaxis and early treatment.

Neuroimaging provides an effective tool to characterize the structure and functionality of nervous system, thus has greatly contributed to Alzheimer's study [22]. Extensive work has been proposed to predict MCI conversion using neuroimaging data [6,15]. Previous methods usually formulate MCI conversion prediction as a binary classification (distinguishing MCI converters from non-converters) [15] or multi-class classification problem (when considering other classes such as AD or health control (HC)) [6], where the methods take

the neuroimaging data at baseline time as the input and classify if the MCI samples will convert to AD in years.

Despite the prosperity and progress achieved in MCI conversion prediction, there are still several problems existing in previous methods. (1) Although we expect the model to be capable of forecasting the MCI conversion years before the change of disease status, the training process should not be limited to just baseline data. In the longitudinal study of AD, usually the data at several time points along the disease progression is available, such as baseline, month 6, month 12, *etc.* However, previous methods only consider the baseline data in the training process, thus ignore the temporal correlation structure among other time points.

(2) The labeling process for Alzheimer's is time-consuming and expensive, so the MCI conversion prediction suffers greatly from limited training data.

To deal with these problems, we propose a novel model for MCI conversion prediction. Firstly, we study the temporal correlation structure among the longitudinal data in Alzheimer's progression. Since AD is a chronically progressive disorder and the neuroimaging features are correlated [11], it can be helpful to analyze the temporal correlation between neuroimaging data in the disease progression as in other nervous system diseases [3]. We construct a regression model to discover such temporal correlation structure between adjacent time points. Our model incorporates the data at all time points along the disease progression and uncovers the variation trend that benefits MCI conversion prediction.

Secondly, we construct a classification model to predict the disease status at each time point. Different from previous classification models that use the baseline data to forecast the progression trend in two or three years, our classification model focuses on adjacent time points. Compared with previous models that require a highly distinguishable conversion pattern appears several years before dementia, our model predicts the progression trend for consecutive time points, thus is more accurate and reliable.

Thirdly, we construct a generative model based on generative adversarial network (GAN) to produce more auxiliary data to improve the training of regression and classification model. GAN model is proposed in [5], which uses the adversarial mechanism to learn the inherent data distribution and generate realistic data. We use the generative model to learn the joint distribution of neuroimaging data at consecutive time points, such that more reliable training data can be obtained to improve the prediction of MCI conversion.

2 Temporal Correlation Structure Learning Model

2.1 Problem Definition

In MCI conversion prediction, for a certain sample and a time point t , we use $\mathbf{x}_t \in \mathbb{R}^p$ to denote the neuroimaging data at time t while $\mathbf{x}_{t+1} \in \mathbb{R}^p$ for the next time point, where p is the number of imaging markers. $\mathbf{y}_t \in \mathbb{R}$ is the label showing the disease status at time t and t

+ 1. Here we define three different classes for $\mathbf{y}_t$: $\mathbf{y}_t = 1$ means the sample is AD at both time t and $t + 1$; $\mathbf{y}_t = 2$ shows MCI at time t while AD at time $t + 1$; while $\mathbf{y}_t = 3$ indicates that the sample is MCI at both time t and $t + 1$. In the prediction, given the baseline data of an MCI sample, the goal is to predict whether the MCI sample will finally convert to AD or not.

2.2 Revisit GAN Model

GAN model is proposed in [5], which plays an adversarial game between the generator G and discriminator D . The generator G takes a random variable $\mathbf{z}$ as the input and outputs the generated data to approximate the inherent data distribution. The discriminator D is proposed to distinguish the data $\mathbf{x}$ from the real distribution and the data produced from the generator. Whereas the generator G is optimized to generate data as realistic as possible to fool the discriminator. The objective function of the GAN model has the following form.

$$\min_G \max_D \mathbb{E}_{\mathbf{x} \sim p(\mathbf{x})} [\log(D(\mathbf{x}))] + \mathbb{E}_{\mathbf{z} \sim p(\mathbf{z})} [\log(1 - D(\mathbf{z}))],$$

where $p(\mathbf{z})$ denotes the distribution of the random variable and $p(\mathbf{x})$ represents the distribution of real data. The min-max game played between G and D improves the learning of both the generator and discriminator, such that the model can learn the inherent data distribution and generate realistic data.

2.3 Illustration of Our Model

Inspired by [2], we propose to approximate the joint distribution of neuroimaging data at consecutive time points and data label $([\mathbf{x}_t, \mathbf{x}_{t+1}], \mathbf{y}_t) \sim p(\mathbf{x}, \mathbf{y})$ by considering the following:

$$\min_{G_t, G_{t+1}} \max_D \mathbb{E}_{([\mathbf{x}_t, \mathbf{x}_{t+1}], \mathbf{y}_t) \sim p(\mathbf{x}, \mathbf{y})} [\log(D([\mathbf{x}_t, \mathbf{x}_{t+1}], \mathbf{y}_t))] + \mathbb{E}_{\mathbf{z} \sim p(\mathbf{z}), \mathbf{y} \sim p(\mathbf{y})} [\log(1 - D([G_t(\mathbf{z}, \mathbf{y}), G_{t+1}(\mathbf{z}, \mathbf{y})], \mathbf{y}))],$$

where the generators take a random variable $\mathbf{z}$ and a pseudo label $\mathbf{y}$ as the input and output a data pair $([G_t(\mathbf{z}, \mathbf{y}), G_{t+1}(\mathbf{z}, \mathbf{y})], \mathbf{y})$ that is as realistic as possible. Still, the discriminator is optimized to distinguish real from fake data. The construction of such generative model approximates the inherent joint distribution of neuroimaging data at adjacent time points and label, which generates more reliable samples for the training process.

To uncover the temporal correlation structure among the neuroimaging data between consecutive time points, we construct a regression network R to predict $\mathbf{x}_{t+1}$ from $\mathbf{x}_t$, such that progression trend among neuroimaging data along the disease progression can be learned. The network R takes data from both real distribution and the generators as the input and optimize the following:

$$\min_R \mathbb{E}_{(\mathbf{x}_t, \mathbf{x}_{t+1}, \mathbf{y}_t) \sim p(\mathbf{x}, \mathbf{y})} [\|\mathbf{x}_{t+1} - R(\mathbf{x}_t)\|_1] \quad (1)$$

$$+ \lambda_{reg} \mathbb{E}_{\mathbf{z} \sim p(\mathbf{z}), \mathbf{y} \sim p(\mathbf{y})} [\|G_{t+1}(\mathbf{z}, \mathbf{y}) - R(G_t(\mathbf{z}, \mathbf{y}))\|_1],$$

where the hyper-parameter λ_{reg} balances the importance of real and generated data. We consider ℓ_1 -norm loss to make the model R more robust to outliers.

In addition, we construct a classification structure C to predict the label $\mathbf{y}_t$ given data $\mathbf{x}_t$. The optimization of C is based on the following:

$$\min_C - \mathbb{E}_{(\mathbf{x}_t, \mathbf{x}_{t+1}, \mathbf{y}_t) \sim p(\mathbf{x}, \mathbf{y})} [\mathbf{y}_t \log(C(\mathbf{x}_t))] \quad (2)$$

$$- \lambda_{cly} \mathbb{E}_{\mathbf{z} \sim p(\mathbf{z}), \mathbf{y} \sim p(\mathbf{y})} [\mathbf{y} \log(C(G_t(\mathbf{z})))],$$

where λ_{cly} is a hyper-parameter to balance the role of real and generated data.

Given a set of real data $\{(\mathbf{x}_t^i, \mathbf{x}_{t+1}^i, \mathbf{y}_t^i)\}_{i=1}^n$, the above three loss terms can be approximated by the following empirical loss:

$$\mathcal{L}_{adv} = \frac{1}{n} \sum_{i=1}^n \log(D(\mathbf{x}_t^i, \mathbf{x}_{t+1}^i, \mathbf{y}_t^i)) + \sum_{j=1}^{n_z} \log(D(G_t(\mathbf{z}^j, \mathbf{y}^j), G_{t+1}(\mathbf{z}^j, \mathbf{y}^j))),$$

$$\mathcal{L}_{reg} = \frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_{t+1}^i - R(\mathbf{x}_t^i)\|_1 + \lambda_{reg} \sum_{j=1}^{n_z} \|G_{t+1}(\mathbf{z}^j, \mathbf{y}^j) - R(G_t(\mathbf{z}^j, \mathbf{y}^j))\|_1,$$

$$\mathcal{L}_{cly} = -\frac{1}{n} \sum_{i=1}^n \mathbf{y}_t^i \log(C(\mathbf{x}_t^i)) - \lambda_{cly} \sum_{j=1}^{n_z} \mathbf{y}^j \log(C(G_t(\mathbf{z}^j, \mathbf{y}^j))).$$

For a clear illustration, we plot a figure in Fig. 1 to show the structure of our Temporal-GAN model (temporal correlation structure learning for MCI conversion prediction with GAN). The implement details of the networks can be found in the experimental setting section. The optimization of our model is based on a variant of mini-batch stochastic gradient descent method.

3 Experimental Results

3.1 Experimental Setting

To evaluate our Temporal-GAN model, we compare with the following methods: **SVM-Linear** (support vector machine with linear kernel), which has been widely applied in MCI conversion prediction [6,15]; **SVM-RBF** (SVM with RBF kernel), as employed in [10,21]; and **SVM-Polynomial** (SVM with polynomial kernel) as used in [10]. Also, to validate the improvement by learning the temporal correlation structure, we compare with the **Neural Network** with exactly the same structure in our classification network (network C in Fig. 1)

that only uses baseline data. Besides, we compare with the case where we do not use the GAN model to generate more auxiliary samples, *i.e.*, only using network C and R in Fig. 1, which we call **Temporal-Deep**.

The classification accuracy is used as the evaluation metric. We divide the data into three sets: training data for training the models, validation data for tuning hyper-parameters, and testing data for reporting the results. We tune the hyper-parameter C of SVM-linear, SVM-RBF and SVM-Polynomial methods in the range of $\{10^{-3}, 10^{-2}, \dots, 10^3\}$. We compare the methods when using different portion of testing samples and report the average performance in five repetitions of random data division.

In our Temporal-GAN model, we use the fully connected neural network structure for all the networks G , D , R and C , where each hidden layer contains 100 hidden units. The implementation detail is as follows: the number of hidden layers in structure G , D , R and C is 3, 1, 3, 2 respectively. We use leaky rectified linear unit (LReLU) [12] with leakiness ratio 0.2 as the activation function of all layers except the last layer and consider weight normalization [14] for layer normalization. Also, we utilize the dropout mechanism in the regression structure R with the dropout rate of 0.1. The weight parameters of all layers are initialized using the Xavier approach [4]. We use the ADAM algorithm [9] to update the weight parameters with the hyper-parameters of ADAM algorithm set as default. Both values of λ_{reg} in Eq. (1) and λ_{cl} in Eq. (2) are set as 0.01.

3.2 Data Description

All data were downloaded from the ADNI database (adni.loni.usc.edu). Each MRI T1-weighted image was first anterior commissure (AC) posterior commissure (PC) corrected using MIPAV2, intensity inhomogeneity corrected using the N3 algorithm [17], skull stripped [20] with manual editing, and cerebellum-removed [19]. We then used FAST [23] in the FSL package3 to segment the image into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF), and used HAMMER [16] to register the images to a common space. GM volumes obtained from 93 ROIs defined in [8], normalized by the total intracranial volume, were extracted as features. Out of the 93 ROIs, 24 disease-related ROIs were involved in the MCI prediction [18]. This experiment includes data from six different time points: baseline (BL), month 6 (M6), month 12 (M12), month 18 (M18), month 24 (M24) and month 36 (M36). All 216 samples with no missing MRI features at BL and M36 time are used by all the comparing methods, where there are 101 MCI converters (MCI at BL time while AD at M36) as well as 115 non-converters (MCI at both BL and M36). Since our Temporal-GAN model can use data at time points other than BL and M36, we include a total of 1419 data pairs with no missing neuroimaging measurement for training the classification, regression and generative model in our Temporal-GAN model. All neuroimaging features in the data are normalized to zero mean and unit variance.

3.3 MCI Conversion Prediction

We summarize the MCI conversion classification results in Table 1. The goal of the experiment is to accurately distinguish converter subjects from non-converters among the MCI samples at baseline time. From the comparison we notice that Temporal-GAN

outperforms all other methods under all settings, which confirms the effectiveness of our model. Compared with SVM-Linear, SVM-RBF, SVM-Polynomial and Neural Network, the Temporal-GAN and Temporal-Deep model illustrates apparent superiority, which validates that the temporal correlation structure learned in our model substantially improves the prediction of MCI conversion. The training process of our model takes advantage of all the available data along the progression of the disease, which provides more beneficial information for the prediction of MCI conversion. By comparing Temporal-GAN and Temporal-Deep, we can notice that Temporal-GAN always performs better than Temporal-Deep, which indicates that the generative structure in Temporal-GAN could provide reliable auxiliary samples to strengthen the training of regression R and classification C model, thus improves the prediction of MCI conversion.

3.4 Visualization of the Imaging Markers

We use feature weight visualization in Fig. 2 to validate if our Temporal-GAN can detect disease-related features when using all 93 ROIs in the MCI conversion prediction. We adopt the Layer-wise Relevance Propagation (LRP) [1] method to calculate the importance of neuroimaging features in the testing data. We can notice that our Temporal-GAN model selects several important features from all 93 ROIs. For example, our method identifies fornix as a significant feature in distinguishing MCI non-converters. The fornix is an integral white matter bundle that locates inside the medial diencephalon. [13] reveals the vital role of white matter in Alzheimer's, such that the degradation of fornix indicates essential predictive power in MCI conversion. Moreover, cingulate region has been found by our model to be related with MCI converters. Previous study [7] finds significantly decreased Regional cerebral blood flow (rCBF) measurement in the left posterior cingulate cortex in MCI converters, which serves as an important signal in forecasting the MCI conversion. The replication of these findings proves the validity of our model.

4 Conclusion

In this paper, we proposed a novel Temporal-GAN model for MCI conversion prediction. Our model considered the data at all time points along the disease progression and uncovered the temporal correlation structure among the neuroimaging data at adjacent time points. We also constructed a generative model to produce more reliable data to strengthen the training process. Our model illustrated superiority in the experiments on the ADNI data.

Acknowledgments

H. Huang—This work was partially supported by U.S. NIH R01 AG049371, NSF IIS 1302675, IIS 1344152, DBI 1356628, IIS 1619308, IIS 1633753.

References

1. Bach S, Binder A, Montavon G, Klauschen F, Müller KR, Samek W: On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS One* 10(7), e0130140 (2015) [PubMed: 26161953]
2. Chongxuan L, Xu T, Zhu J, Zhang B: Triple generative adversarial nets. In: *Advances in Neural Information Processing Systems*, pp. 4091–4101 (2017)

3. Fiorini S, Verri A, Barla A, Tacchino A, Bricchetto G: Temporal prediction of multiple sclerosis evolution from patient-centered outcomes. In: Machine Learning for Healthcare Conference, pp. 112–125 (2017)
4. Glorot X, Bengio Y: Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, pp. 249–256 (2010)
5. Goodfellow I, et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2672–2680 (2014)
6. Hu K, Wang Y, Chen K, Hou L, Zhang X: Multi-scale features extraction from baseline structure MRI for MCI patient classification and AD early diagnosis. *Neurocomputing* 175, 132–145 (2016)
7. Huang C, Wahlund LO, Svensson L, Winblad B, Julin P: Cingulate cortex hypoperfusion predicts Alzheimer's disease in mild cognitive impairment. *BMC Neurol.* 2(1), 9 (2002) [PubMed: 12227833]
8. Kabani NJ: 3D anatomical atlas of the human brain. *Neuroimage* 7, P-0717 (1998)
9. Kingma D, Ba J: Adamml: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
10. Lemos L: Discriminating Alzheimer's disease from mild cognitive impairment using neuropsychological data. *Age (M ± SD)* 70(8.4), 73 (2012)
11. Liu S, et al.: Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. *IEEE Trans. Biomed. Eng* 62(4), 1132–1140 (2015) [PubMed: 25423647]
12. Maas AL, Hannun AY, Ng AY: Rectifier nonlinearities improve neural network acoustic models. In: International Conference on Machine Learning (ICML), vol. 30 (2013)
13. Nowrangi MA, Rosenberg PB: The fornix in mild cognitive impairment and alzheimers disease. *Front. Aging Neurosci* 7, 1 (2015) [PubMed: 25653617]
14. Salimans T, Kingma DP: Weight normalization: a simple reparameterization to accelerate training of deep neural networks. In: Advances in Neural Information Processing Systems (NIPS), pp. 901–909 (2016)
15. Schmitter D: An evaluation of volume-based morphometry for prediction of mild cognitive impairment and Alzheimer's disease. *NeuroImage: Clin.* 7, 7–17 (2015) [PubMed: 25429357]
16. Shen D, Davatzikos C: Hammer: hierarchical attribute matching mechanism for elastic registration. *IEEE Trans. Med. Imaging* 21(11), 1421–1439 (2002) [PubMed: 12575879]
17. Sled JG, Zijdenbos AP, Evans AC: A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans. Med. Imaging* 17(1), 87–97 (1998) [PubMed: 9617910]
18. Wang H, et al.: Identifying quantitative trait loci via group-sparse multitask regression and feature selection: an imaging genetics study of the ADNI cohort. *Bioinformatics* 28(2), 229–237 (2011) [PubMed: 22155867]
19. Wang Y, et al.: Knowledge-guided robust MRI brain extraction for diverse large-scale neuroimaging studies on humans and non-human primates. *PloS One* 9(1), e77810 (2014) [PubMed: 24489639]
20. Wang Y, Nie J, Yap P-T, Shi F, Guo L, Shen D: Robust deformable-surface-based skull-stripping for large-scale studies In: Fichtinger G, Martel A, Peters T (eds.) MICCAI 2011. LNCS, vol. 6893, pp. 635–642. Springer, Heidelberg (2011). 10.1007/978-3-642-23626_678
21. Wei R, Li C, Fogelson N, Li L: Prediction of conversion from mild cognitive impairment to Alzheimer's disease using MRI and structural network features. *Front. Aging Neurosci.* 8, 76 (2016) [PubMed: 27148045]
22. Weiner MW, et al.: The Alzheimer's disease neuroimaging initiative: a review of papers published since its inception. *Alzheimer's Dement.* 9(5), e111–e194 (2013) [PubMed: 23932184]
23. Zhang Y, Brady M, Smith S: Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imaging* 20(1), 45–57 (2001) [PubMed: 11293691]

Fig. 1.

Illustration of our Temporal-GAN model. $\mathbf{x}_t$ and $\mathbf{x}_{t+1}$ are the neuroimaging data at two adjacent time points and y_t is the label ($y_t = 1$ if both $\mathbf{x}_t$ and $\mathbf{x}_{t+1}$ are at AD status; $y_t = 2$ if $\mathbf{x}_t$ is MCI while $\mathbf{x}_{t+1}$ is AD; $y_t = 3$ if both $\mathbf{x}_t$ and $\mathbf{x}_{t+1}$ are MCI.). The regression network R predicts $\mathbf{x}_{t+1}$ from $\mathbf{x}_t$ so as to uncover the temporal correlation between adjacent time points. The classification network C predicts the label y_t from $\mathbf{x}_t$. We also construct a generative model with generator G and discriminator D to approximate the joint distribution underlying data pair $([\mathbf{x}_t, \mathbf{x}_{t+1}], y_t)$ to generate more reliable data for training R and C . In the prediction process, the neuroimaging data $\mathbf{x}_0$ at baseline time for MCI samples is given, and we use R and C to predict whether the MCI sample will convert to AD at time T .

Fig. 2.

Visualization figure showing the feature weights from our Temporal-GAN model. The upper figure shows features on the left hemisphere while the lower corresponds to the right hemisphere.

Table 1.

MCI conversion prediction with different portion of testing data.

Methods	10%	20%	50%
SVM-Linear	0.6273 ± 0.0668	0.6558 ± 0.0648	0.6093 ± 0.0484
SVM-RBF	0.5818 ± 0.0881	0.5767 ± 0.0770	0.5852 ± 0.0381
SVM-Polynomial	0.6545 ± 0.0617	0.5953 ± 0.1253	0.5611 ± 0.0429
Neural Network	0.3727 ± 0.0340	0.4233 ± 0.0426	0.4685 ± 0.0324
Temporal-Deep	0.7455 ± 0.0464	0.7209 ± 0.0441	0.6741 ± 0.0451
Temporal-GAN	0.7818 ± 0.0445	0.7488 ± 0.0342	0.7000 ± 0.0570