MULTIPLE-GRAPH RECURRENT GRAPH CONVOLUTIONAL NEURAL NETWORK ARCHITECTURES FOR PREDICTING DISEASE OUTCOMES

*Juliette Valenchon and Mark Coates**

Dept. of Electrical and Computer Engineering, McGill University, Montréal, Québec, Canada

ABSTRACT

Improving disease outcome prediction can greatly aid in the strategic deployment of secondary prevention approaches. We develop a method to predict the evolution of diseases by taking into account personal attributes of the subjects and their relationships with medical examination results. Our approach builds upon a recent formulation of this problem as a graph-based geometric matrix completion task. The primary innovation is the introduction of multiple graphs, each relying on a different combination of subject attributes. Via statistical significance tests, we determine the relevant graph(s) for each medically-derived feature. We then employ a multiple-graph recurrent graph convolutional neural network architecture to predict the disease outcomes. We demonstrate the efficacy of the technique by addressing the task of predicting the development of Alzheimer's disease for patients exhibiting mild cognitive impairment, showing that the incorporation of multiple graphs improves predictive capability.

***Index Terms*—** Graph Convolutional Neural Network, Geometric Matrix Completion, Graph signal processing, Disease outcome prediction

1. INTRODUCTION

Extracting the most information from medical datasets can greatly aid in the strategic deployment of secondary prevention approaches. Machine learning algorithms can potentially discover patterns that are not obvious to a doctor. The prediction accuracy can be improved by using as much information as possible, including, for example, the age and sex of a subject. These types of subject attributes can impact both the medically-derived features and the disease outcome that is the prediction target. For example, women are more likely to develop Alzheimer's disease (AD) than men [1, 2]. The MRI-derived brain volumes of cortical subregions are potential predictors, and larger values are observed for men [3, 4].

Recently, prediction techniques have been developed based on graph convolutional neural networks (CNNs) and graph-based geometric matrix completion [5, 6]. These methods connect subjects by constructing a single graph based on attributes such as age and sex. Graph-based learning approaches such as those developed in [7–9] are then employed to process the medical features for each subject and perform the prediction. The geometric matrix completion approach also eliminates the need for imputation of missing features.

In general, it is not the case that every feature is dependent on each of the attributes used to construct the graph. For example, for Alzheimer's disease prediction, intracranial volume is dependent on sex but does not vary significantly with age (see Figure 1). The major innovation in this paper is the use of multiple attribute graphs for graph CNN matrix completion. Via a general linear model and statistical significance tests, we identify an appropriate association of specific features to each graph. Our approach is the first to employ multiple feature-specific adjacency matrices for learning using graph convolutional neural networks.

In the following subsection we discuss related work. Section 2 provides a more formal statement of the problem. Section 3 provides a brief overview of graph-based matrix completion and graphical convolutional neural networks. Section 4 describes our approach and algorithm, and Section 5 presents the results of the application of our approach to the prediction of Alzheimer's disease development.

1.1. Related work

Only recently have graph-based learning methods started to appear for disease outcome prediction. Previously, state-of-the-art approaches employed more traditional classification approaches including random forests and support vector machines [10, 11], or convolutional neural networks [12]. Parisot et al. were the first to propose a graph-based learning algorithm for disease outcome prediction [5]. They employed a graph convolutional neural network.

In [6], Vivar et al. improved the performance by using graph-based geometric matrix completion, employing the recurrent graph convolutional neural network architecture de-

*Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf. In particular, this work uses the TADPOLE data sets https://tadpole.grand-challenge.org constructed by the EuroPOND consortium http://europond.eu funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 666992

veloped in [9]. This built on the work of Thung et al. [13], who used the low-rank matrix completion approach developed in [14] for jointly performing imputation of missing values and transductive classification.

In both [5] and [6], a single graph is used for all features. In contrast, we develop an architecture that processes multiple graphs; our algorithm associates different features to different graphs by fitting a general linear model (GLM) and assessing the significance of each regression coefficient. Since it builds on the algorithm in [9], our work is related to graph-based matrix completion techniques [9,15,16] and graph convolutional neural networks [7,8]. Most of these algorithms employ a single graph. Kipf et al. discuss the possibility of using multiple graphs in [8]; Such et al. and Monti et al. explicitly use multiple graphs in [9,17]. Although multiple graphs are employed, each graph is used to process all features at each node. As a result, the graph neural network must learn an embedding from a higher dimensional feature space using many variables that are unlikely to be related to the graph.

2. PROBLEM FORMULATION

We consider the following prediction task for disease outcomes. Let $X \in \mathbb{R}^{m \times n}$ be the feature matrix, m being the number of subjects and n the number of features. The features are assumed to be derived from medical examinations. X may have missing values. Let $Y \in \{0,1\}^{m \times 1}$ be a vector denoting the disease outcomes for the m subjects. Some of these are unknown and these are the focus of the prediction task.

Let $\mathcal{G}_i = \{V_i, E_i, A_i\}$ be a graph on the subjects with edges derived by a similarity metric from a subject attribute s_i. The attribute can be categorical, or real- or integer-valued. V_i denotes the vertices, $V_i = \{1, ..., m\}$, E_i the edges, $E_i \subseteq V_i \times V_i$, and $A_i \in \{0,1\}^{m \times m}$ the adjacency matrix. We assume that there are P such graphs derived from different combinations of subject attributes and thus capturing different relationships between subjects. Taking into account the features X and the relationships formed by the similarities of the attributes s_i and captured by G_i, $i = 1, \ldots, P$, our task is to predict the unknown disease outcomes in Y and impute the missing values in the matrix X.

3. BACKGROUND MATERIAL

The goal of matrix completion is to recover the missing values of a matrix, usually by making structural assumptions such as low rank. Geometric matrix completion involves constraining the space of solutions to be smooth with respect to some geometric structure (such as a graph on the rows of X). We have an initial matrix M with some known values and use $||.||_F$ to denote the Frobenius norm. Let Ω_a be the indicator matrix of the known feature values and $*$ an element-wise multiplication. We denote by $||X||_{\mathcal{G}} = \mathrm{trace}(X^T \Delta X)$ the Dirichlet norm with respect to a graph $\mathcal{G}$ with adjacency matrix A, degree matrix D and graph Laplacian $\Delta = I - D^{-1/2} A D^{-1/2}$, for identity matrix I. The geometrix matrix completion task can then be expressed as:

$$\min_X ||X||^2_{\mathcal{G}} + \frac{\mu}{2} ||\Omega_a * (X - M)||^2_F, \tag{1}$$

where μ controls the balance between fidelity to known values and smoothness with respect to the graph.

In [9], Monti et al. proposed a method that combines a graph convolutional neural network and a recurrent neural network to construct a graph diffusion process to identify a solution to (1). Graph convolutional neural networks (GCNNs) are a generalization of CNNs to data on graphs [7, 8]. Collecting the feature vectors from all nodes in the graph as the rows of a matrix X, the layers of a GCNN [7,8] are of the form:

$$H^{(1)} = \sigma(\widetilde{A}_{\mathcal{G}} X W^{(0)}) \tag{2}$$

$$H^{(l+1)} = \sigma(\widetilde{A}_{\mathcal{G}} H^{(l)} W^{(l)}) \tag{3}$$

Here $W^{(l)}$ are the weights of the neural network at layer l, $H^{(l)}$ are the output features from layer $l-1$, and σ is a nonlinear activation function. The matrix $\widetilde{A}_{\mathcal{G}}$ is an operator derived from the observed graph and determines how the output features are mixed across the graph at each layer. In [8], $\widetilde{A} = D^{-1/2}(I + A)D^{-1/2}$; in [7], $\widetilde{A} \approx [T_0(\Delta) \ldots T_{K-1}(\Delta)]$ is a learnable operator constructed from a Chebyshev expansion of a K-th order polynomial of the graph Laplacian and here $\widetilde{A}_{\mathcal{G}} H^{(l)} W^{(l)} = (\widetilde{A}^T H^{(l)})^T W^{(l)} \approx \sum_k^{K-1} T_k(\Delta) H^{(l)} W_k^{(l)}$ where $W^{(l)} = [W_0^{(l)} \ldots W_{K-1}^{(l)}]^T$ are the weights of the neural network at layer l.

The matrix completion procedure in [9] involves initialization of a matrix X_0 and then iterative training of (i) a graph CNN to perform an embedding of X_t; and (ii) a recurrent neural network that processes the embedding to calculate an update $\delta\widetilde{X}$ to obtain $X_t = X_{t-1} + \delta X_{t-1}$. The parameters of the GCNN and the recurrent NN are trained to minimize a loss function of the form (1).

4. METHODOLOGY

In the task of disease outcome prediction, most of the datasets have missing values and inaccurate measurements. Formulating the task as matrix completion as in [14] allows us to jointly perform transductive classification and imputation of missing values. To do this, we form a matrix $Z = [X, Y]$ and apply a matrix completion algorithm.

We commonly have knowledge of attributes that can be used to identify relationships or similarities between subjects. Attributes such as age and sex often impact the probability of a disease outcome and the likelihood of a feature derived from a medical examination. In trying to recover a matrix with missing values and unknown disease outcomes, it is reasonable to assume that there is smoothness with respect to a

graph that connects individuals who share similar attributes (close in age, same sex). Once such a graph has been constructed, geometric graph completion can be performed; Vivar et al. [6] use the algorithm from [9] to do so.

The problem with the approach outlined above is that in general it is not the case that every medically-derived feature is dependent on all of the attributes used to construct the graph. For example, for prediction of progression to Alzheimer's disease, Vivar et al. construct a weighted adjacency matrix that includes an edge between people of the same sex and those of similar age. Many of the features in the matrix have no dependency on age; requiring such features to be smooth with respect to age imposes an undesirable penalty in the optimization and results in incorrect information diffusion throughout the graph. With regard to imputation, if one is estimating a missing value that is sex-dependent, but not age-dependent, it is better to use all of the values from subjects with the same sex and not bias the imputation by processing values from the other sex.

In our proposed approach we construct multiple graphs based on the available attributes and associate a feature with one or more of these graphs by fitting a general linear model (GLM) with the attributes as the independent variables and the features as the dependent variables. We then assess the significance of the regression coefficients. The features with a statistically significant non-zero value for attribute a_i are included in a subset Z_i of Z that is associated with each graph $\mathcal{G}_i$. The GLM is fit using ordinary least squares and we assess significance of coefficients using multiple ANOVA and post-hoc t-tests. In this procedure, controlling the Type I error is not as significant a concern as is usually the case in regression procedures. Erroneous association of a feature with a specific attribute graph leads to an additional smoothness penalty that should not be included, but in most cases this has a minor effect on the overall inference procedure. Improvement in prediction outcomes is achieved by ensuring that the majority of features with no dependence on an attribute are excluded from the subset.

4.1. Multiple-Graph Recurrent Graph Convolutional Neural Network (MG-RGCNN)

We develop an architecture based on the Recurrent Graph Convolutional Neural Network (RGCNN) from [9]. We adapt it to take into account the multiple graphs and the prediction task. The GCNN layer as described in Section 3 computes features from the initial matrix Z using multiple graph convolutional neural networks based on the graphs $\mathcal{G}_i$. The P different GCNN outputs are concatenated and provided to the recurrent neural network. The algorithm is described in Algorithm 1. The parameters of the multiple GCNNs and the RNN are trained using a loss function that has a Dirichlet norm penalty for each graph. It would also be possible to consider a weighted sum of Dirichlet norms, where each weight is dependent on the number of features associated with the subset. As we are addressing a classification problem in addition to imputation of missing entries, we also add a binary cross-entropy term $l_{\Omega_b}(Z, M) = -(y \log(p) + (1-y)\log(1-p))$, where Ω_b is the indicator matrix of known outcomes, p is the classification output vector and y the label vector. We also add a l2 regularization term for the overfitting on the q weight matrices $(W_1, \ldots, W_q)$ used in the architecture. $\mu_i (i \in [1, P])$, μ and γ_{l2} are parameters controlling the balance between the different loss terms.

$$l(\theta) = \sum_{i=1}^{P} \frac{\mu_i}{2} ||Z_i||^2_{\mathcal{G}_i} + \frac{1}{2} ||\Omega_a * (Z - M)||^2_F + \mu l_{\Omega_b}(Z, M) + \gamma_{l2} \sum_{i=1}^{q} W_i \quad (4)$$

Algorithm 1 RGCNN with P graphs (MG-RGCNN)

procedure MG-RGCNN($M = [X, Y], \{A_i\}, \{Z_i\}$)
  Initialization : $Z = [X, Y_{train}]$, weights (Xavier [18]) and biases (zero)
  for k in number iterations **do**
    for i in $1, ..., P$ **do**
      $v_i = GCNN(Z_i, A_i)$
    $v_{tot} = \text{concatenation}([v_i], i = 1, \ldots, P)$
    $\delta Z = LSTM(v_{tot})$
    $Z = Z + \delta Z$
  Compute loss function (4) and update weights by backpropagation

5. RESULTS

We apply the proposed MG-RGCNN to the TADPOLE dataset [19], a dataset for the prediction of conversion from Mild Cognitive Impairment (MCI) to Alzheimer's Disease (AD). MCI is an intermediate stage between the normal stage and dementia. Patients with MCI are in a stage where the disease could evolve to AD (MCI converters, MCIc) or not (MCI non converters, MCInc). Early prediction of the conversion can aid in the strategic deployment of secondary prevention approaches. In forming predictions, we use the baseline data acquired from the first examination of a subject. We include subjects that were diagnosed as MCI in their baseline scan and that have converted to AD 48 months later (MCIc) or that have remained stable for the course of the study (MCInc). We excluded features where more than 50% of the values were missing. After the preprocessing steps, we have 779 subjects, 296 MCIc and 483 MCInc, and 563 features. We added the label (disease outcome = MCIc or MCInc) column as the last column of the matrix.

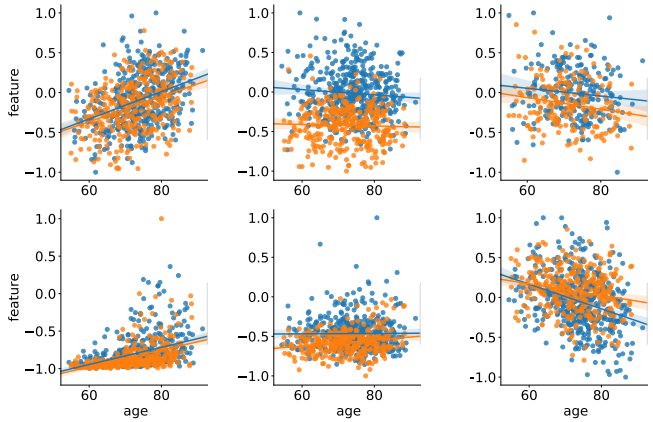
We built 4 graphs: $\mathcal{G}_a$ for age-related features, $\mathcal{G}_s$ for sex-related features, $\mathcal{G}_{a\&s}$ for age and sex-related features and $\mathcal{G}_{no}$ for features that are neither related to age or sex. At the node of each graph, we have the values of the features that are related to this subject's characteristic. These four graphs lead

to four adjacency matrices A_a, A_s, $A_{a\&s}$ and A_{no}, A_{no} being the identity. A_a is constructed by including an edge between subject r and s if $|\text{age}(s) - \text{age}(r)| < 2$. A_s includes an edge if $\text{sex}(s) = \text{sex}(r)$. $A_{a\&s}$ adds an edge when both conditions are satisfied. The loss function reads as Eq. (5) where μ_a, μ_s, $\mu_{a\&s}$, μ_{no}, μ and γ_{l2} are parameters to control the trade-off between the different loss terms.

$$l(\theta) = \frac{\mu_a}{2}||Z_a||^2_{\mathcal{G}_a} + \frac{\mu_s}{2}||Z_s||^2_{\mathcal{G}_s} + \frac{\mu_{a\&s}}{2}||Z_{a\&s}||^2_{\mathcal{G}_{a\&s}} + \frac{\mu_{no}}{2}||Z_{no}||^2_{\mathcal{G}_{no}} + ||\Omega_a * (Z - M)||^2_F + \mu l_{\Omega_b}(Z, M) + \gamma_{l2} \sum_{i=1}^{q} W_i \quad (5)$$

5.1. Graph construction

AD and MRI-derived brain volume features are known to be related to age and sex [1–4], so these attributes are used to construct the graphs, as in [5, 6]. We conducted the GLM analysis using the three variables age, sex, and age&sex, employing a significance threshold for the p-values of 0.05. As expected, different features have different relationships with age and sex, as illustrated by the examples in Fig. 1. The analysis leads to 452 age-related features, 188 sex-related features, 123 age&sex related features, and 89 features with no relationship to age or sex.



Age-related features | **Sex**-related features. | **Age & Sex** features.

Fig. 1. Age and sex (Men and Women) dependencies. The age-related features are the left caudal anterior cingulate cortical thickness and the hypointensities volume; the sex-related features are intracranial volume and the left caudate volume; the age & sex-related features are the raw volume value for the right pars orbitalis and the cortical thickness average of the left pars orbitalis.

5.2. MG-RGCNN performance

Features were normalized to range between -1 and 1 and missing values were initialized as 0. For the included subjects, 21% of feature values are missing. We used the GCNN from [7]. We optimized the AUC on the validation set to find the hyperparameters: hidden units=51, learning rate=8×10^{-4}, μ_a=84, μ_s=100, $\mu_{a\&s}$=29, μ_{no}=82, μ=84 and γ_{l2}=1. We took a split of 0.6/0.2/0.2 for the training/validation/testing set. Table. 1 depicts the mean and the variance of the Area Under the receiver operating characteristic Curves (AUCs) on the test set when executing 100 different training/validation/test partitions. We evaluated the statistical significance of the difference between the AUC for the MG-RGCNN and the sRGCNN [6] by performing a Wilcoxon signed-rank test. We obtained a p-value of 4.1×10^{-3}, indicating that the MG-RGCNN AUC scores were statistically significantly higher than the sRGCNN scores.

Method	**AUC** (mean ± std)
Linear SVM	0.6896 ± 0.0270
Multi-Layer Perceptron	0.7360 ± 0.0372
Random Forest	0.7705 ± 0.0318
Parisot et al [5]	0.7671 ± 0.0355
Vivar et al [6]	0.7191 ± 0.0556
MG-RGCNN	0.7394 ± 0.0435

Table 1. Performances for prediction from MCI-to-AD.

6. CONCLUSION

We introduce a multiple-graph architecture based on a graph-based geometric matrix completion method to predict disease outcomes for datasets with missing values. We use a statistical significance test to determine the subsets of the features that are relevant to each of the graphs. This leads to an improvement of 2% on the mean AUC compared to [6]. The MG-RGCNN algorithm helps in performing better classification as it takes into account more accurately the feature dependencies with age and sex and allows to better recover the missing values. However, it is being outperformed by random forest and Parisot et al. [5], architectures where the missing values are imputed by a mean of the known values for this feature. This could be due to the fact that the dataset is not large and there are only 21% of missing values which is not enough to interfere with the classification results.

In future work, we will test the algorithm when more data is missing to see the importance of the matrix completion task on the classification result. Moreover, instead of focusing on classification and providing an output that only indicates if the subject will progress to the disease or not, we will incorporate a calibration mechanism so that the provided value represents the probability of conversion. We will also apply the proposed multiple graph algorithm to other disease outcome datasets and explore methods for automatically choosing the attributes used to construct graphs instead of using ad hoc rules.

7. REFERENCES

[1] C.M. Mazure and J. Swendsen, "Sex differences in Alzheimer's disease and other dementias," *The Lancet Neurology*, vol. 15, no. 5, pp. 451–452, 2016.

[2] J. Viña and A. Lloret, "Why women have more Alzheimer's disease than men: gender and mitochondrial toxicity of amyloid-β peptide," *J. Alzheimer's disease*, vol. 20, no. s2, pp. S527–S533, 2010.

[3] S.J. Ritchie et al., "Sex differences in the adult human brain: Evidence from 5216 UK Biobank participants," *Cerebral Cortex*, vol. 28, no. 8, pp. 2959–2975, 2018.

[4] A.N.V. Ruigrok et al, "A meta-analysis of sex differences in human brain structure," *Neuroscience & Biobehavioral Rev.*, vol. 39, pp. 34–50, 2014.

[5] S. Parisot et al., "Spectral graph convolutions for population-based disease prediction," in *Proc. Int. Conf. Medical Image Computing and Comput-Assisted Intervention*, 2017, pp. 177–185.

[6] G. Vivar, A. Zwergal, N. Navab, and S.-A. Ahmadi, "Multi-modal disease classification in incomplete datasets using geometric matrix completion," in *Proc. Int. Workshop Graphs Biomedical Image Analysis (GRAIL)*, Granada, Spain, Sept. 2018, pp. 24–31.

[7] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. Advances Neural Inform. Process. Syst.*, 2016.

[8] T. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. Int. Conf. Learning Representations*, Toulon, France, 2017.

[9] F. Monti, M. Bronstein, and X. Bresson, "Geometric matrix completion with recurrent multi-graph neural networks," in *Proc. Advances Neural Inform. Process. Syst.*, 2017, pp. 3697–3707.

[10] E. Moradi, A. Pepe, C. Gaser, H. Huttunen, and J. Tohka, "Machine learning framework for early MRI-based Alzheimers conversion prediction in MCI subjects," *NeuroImage*, vol. 104, pp. 398–412, 2015.

[11] J.E. Arco, J. Ramírez, C.G. Puntonet, J.M. Górriz, and M. Ruz, "Improving short-term prediction from MCI to AD by applying searchlight analysis," in *Proc. IEEE. Int. Symp. Biomedical Imaging*, 2016, pp. 10–13.

[12] H. Choi and Kyong H. Jin, "Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging," *Behavioural Brain Res.*, vol. 344, pp. 103–109, 2017.

[13] K.H. Thung, E. Adeli, P.T. Yap, and D. Shen, "Stability-weighted matrix completion of incomplete multi-modal data for disease diagnosis," in *Proc. Int. Conf. Medical Image Computing and Comput-Assisted Intervention*, 2016, pp. 88–96.

[14] A. Goldberg, B. Recht, J. Xu, R. Nowak, and X. Zhu, "Transduction with matrix completion: Three birds with one stone," in *Proc. Adv. Neural Information Proc. Systems*, 2010, pp. 757–765.

[15] V. Kalofolias, X. Bresson, M. Bronstein, and P. Vandergheynst, "Matrix completion on graphs," *arXiv preprint arXiv:1408.1717*, 2014.

[16] R. van den Berg, T. Kipf, and M. Welling, "Graph convolutional matrix completion," *arXiv preprint arXiv:1706.02263*, 2017.

[17] F. Such et al., "Robust spatial filtering with graph convolutional neural networks," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 6, pp. 884–896, Sept 2017.

[18] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. artificial intell. and statist.*, 2010, pp. 249–256.

[19] R. V. Marinescu et al., "Tadpole challenge: Prediction of longitudinal evolution in Alzheimer's disease," *arXiv preprint arXiv:1805.03909*, 2018.