Learning and Combining Image Similarities for Neonatal Brain Population Studies

Veronika A. Zimmer[1], Ben Glocker[3], Paul Aljabar[4], Serena J. Counsell[4], Mary A. Rutherford[4], A. David Edwards[4], Jo V. Hajnal[4], Miguel Ángel González Ballester[1,2], Daniel Rueckert[3], and Gemma Piella[1]

[1] SIMBioSys Group, Universitat Pompeu Fabra, Barcelona, Spain
[2] ICREA, Barcelona, Spain
[3] Biomedical Image Analysis Group, Imperial College London, United Kingdom
[4] Division of Imaging Sciences & Biomedical Engineering, King's College London, UK

Abstract. The characterization of neurodevelopment is challenging due to the complex structural changes of the brain in early childhood. To analyze the changes in a population across time and to relate them with clinical information, manifold learning techniques can be applied. The neighborhood definition used for constructing manifold representations of the population is crucial for preserving the similarity structure in the embedding and highly application dependent. It has been shown that the combination of several notions of similarity and features can improve the new representation. However, how to combine and weight different similarites and features is non-trivial. In this work, we propose to learn the neighborhood structure and similarity measure used for manifold learning through Neighborhood Approximation Forests (NAFs). The recently proposed NAFs learn a neighborhood structure in a dataset based on a user-defined distance. A characterization of image similarity using NAFs enables us to construct manifold representations based on a previously defined criterion to improve predictions regarding structural and clinical information. In particular, NAFs can be used naturally to combine the affinities learned from multiple distances in a joint manifold towards a more meaningful representation and an improved characterization of the resulting embedding. We demonstrate the utility of NAFs in manifold learning on a population of preterm and in term neonates for classification regarding structural volume and clinical information.

1 Introduction

During early childhood, the brain undergoes complex structural changes, which makes it challenging to characterize normal and abnormal brain development. There is a need for identifying brain imaging biomarkers to improve the diagnostic and therapy. To analyze the changes and differences in a population and to relate them with clinical information, manifold learning techniques can be applied. Classical manifold learning techniques use the neighborhood of images, defined, e.g., by the L2-distance of intensities, to construct manifold representations of the population. The neighborhood definition is crucial for the quality of

the resulting representation and highly application dependent. There has been much interest in identifying and combining additional information in the manifold learning step to improve the resulting embeddings. The manifold structure of brain images has been estimated in [1] based on pairwise non-rigid transformations, whereas in [2] similarities were derived from overlaps of their structural segmentations. In [3], shape and appearance information was combined in a joint embedding for an improved characterization of brain development and in [4] clinical information was incorporated into the embedding construction. However, in general it is not clear how to combine and weight multiple features.

For deriving manifold representations from imaging information, typically the whole image is used (e.g., the L2-distance between all the intensities). But it may not be known in advance which voxels or features are important for a given classification task and manifold learning does not provide insight into this question. The extraction of relevant features requires prior knowledge to the underlying data which might not be available for all applications.

Random forests, on the other hand, have shown to be a powerful approach to feature selection and classification. In [5], they were applied to manifold learning by deriving the pairwise similarity measures from random forest classifiers for different modalities. Additionally, the most important features for the classification problem could be extracted. Recently, the neighborhood approximation forests (NAFs) [6] have been proposed which learn the neighborhood structure in a dataset and the most discriminative features based on a user-defined distance.

In this work, we learn the neighborhood structure used for manifold learning through NAFs. The ability of NAFs to learn on arbitrary distances enables us to construct manifold representations for specific high-level information. NAFs generate affinities based on co-occurrence in leaf nodes and give a natural way to combine the affinities learned from multiple distances. We consider the problem of classification of the data samples regarding structural and clinical information. We train the NAFs on a population of preterm and in term neonatal MR images based on the differences in structural volume (cerebellum and left lateral ventricle volume) and clinical information (gestational age (GA) at scan, birth weight in kilograms (kg) and whether oxygen was supplied after birth). Using the obtained affinity matrices, we construct manifold representations and show their improved performance for classification regarding the learned information. In particular, we show that joint embeddings, obtained by combining affinity matrices from different NAFs, are able to encode simultaneously different information.

2 Methods

Neighborhood Approximation Forests A NAF learns in a supervised manner a neighborhood structure of a given dataset induced by an arbitrary notion of similarity between images. In the training step, the algorithm uses features based on appearance to cluster the images according to the distance function. For testing, the learned features are used to predict the closest neighbors in the training database of a test image. Given a population $\mathcal{I}$ of images, a subset

$\mathbf{I} = \{I_p\}_{p=1}^{P} \in \mathcal{I}$ is used for training and each I_p is represented by a high-dimensional intensity-based feature vector $\mathbf{f}(I_p) \in \mathbb{R}^Q$. The population $\mathcal{I}$ is equipped with a user-defined distance function $\rho : \mathcal{I} \times \mathcal{I} \to \mathbb{R}$ which allows the definition of pairwise distances $\rho(I_m, I_n)$ between the training images.

Training phase: In the training phase N individual trees are constructed. For each tree T, a random subset of features $\mathbf{f}_T \subset \mathbf{f}$ is selected with $\mathbf{f}_T \in \mathbb{R}^q$, $q < Q$. At each node of tree T, the algorithm divides the data samples present in the current node into two sets. This branching of the set of images $\mathbf{I}_s$ present in node s is based on a binary test: for $I_n \in \mathbf{I}_s$, $I_n \in \mathbf{I}_{s_R}$ if $\mathbf{f}_T^m(I_n) > \tau$ and $I_n \in \mathbf{I}_{s_L}$ if $\mathbf{f}_T^m(I_n) \leq \tau$. Here, $\tau \in \mathbb{R}$, s_R and s_L are the children nodes of s and $\mathbf{f}_T^m$ is the mth component of $\mathbf{f}_T$. For each node in each tree, t_s is optimized with respect to the parameters m and τ such that the data samples are clustered according to the distance function ρ in the most compact way.

Testing phase: Given a test set $\hat{\mathbf{I}} = \{\hat{I}_r\}_{r=1}^{R} \in \mathcal{I}$, a test image $\hat{I}_r$ is passed down each tree in the forest. At a node, the binary test with the parameters learned in the training phase is applied. According to this test, the image is sent to the left or to the right child of the current node. This is repeated until the image arrives at a leaf node. If the leaf node contains the training image I_p, their affinity a_{pr} is increased by one. This procedure yields an affinity matrix $A = \{a_{pr}\}_{\substack{p=1,\ldots,P \\ r=1\ldots,R}}$ between the samples of the training and testing set.

Feature Selection: During the training phase of NAFs, the parameters m and τ of the binary test t_s in node s are optimized to obtain an optimal partitioning of the training data samples. The parameter m denotes the component of the feature vector $\mathbf{f}_T^m$ which is tested at the current node. There exist several ways of determining the importance of individual features for the growing of the decision trees. In [6], a feature is considered as important, if it is selected in the first three levels of the trees. A more sophisticated approach was used in [5], where the decrease in the Gini impurity criterion was measured for the individual features in each node. In this work, we adopt the former and simpler approach. The frequency of the selected features in the first three levels of the trees of the forest is recorded, and the values are normalized by the number of nodes in the tree level.

NAFs for Manifold Learning The NAFs F_ρ are trained using the training set $\mathbf{I}$. For each distance function ρ, a pairwise affinity matrix $A_\rho \in \mathbb{R}^{P \times P} = \{a_\rho^{(ij)}\}_{i,j=1,\ldots,P}$ is computed, where $a_\rho^{(ij)}$ reports, how often image $I_j \in \mathbf{I}$ and $I_i \in \mathbf{I}$ finish in the same node. The corresponding distance matrix D_ρ is constructed as $D_\rho = \{d_\rho^{(ij)}\}_{i,j=1,\ldots,P}$ with $d_\rho^{(ij)} = 1 - a_\rho^{(ij)}/N$, where N is the number of trees in the forest. The matrix D_ρ can now be interpreted as pairwise distances of the image set $\mathbf{I}$ and can be used for constructing a manifold representation of the training set. We employ Isomap [7], a global approach, for learning the manifold which we found to give better embeddings than local approaches, such as Laplacian Eigenmaps. The combination of different affinity matrices A_{ρ_k}, $k = 1, \ldots, K$ learned with NAFs based on user-defined distances ρ_k, to create a joint embedding can be done by linear combination. That is, if the NAFs

F_{ρ_k} contain the same number N of trees, the affinity matrices are additively combined by $A_{\rho_1,\ldots,\rho_K} = \frac{1}{K}\sum_{k=1}^{K} A_{\rho_k}$ and the components of the joint distance matrix $D_{\rho_1,\ldots,\rho_K}$ are $d^{(ij)}_{\rho_1,\ldots,\rho_K} = 1 - a^{(ij)}_{\rho_1,\ldots,\rho_K}/N$.

3 Data and Results

Data We tested the proposed approach on a population of 343 neonatal brain T2 weighted MR images, both preterm and in term subjects, with an age range of $26.71 - 49.86$ GA at scan. For all subjects, automatic segmentation into 87 regions were available [8]. For a subset of 314 subjects, the weight at birth in kg is known and for a subset of 212 subjects it is known whether oxygen was supplied right after birth. In the experiments, we used this information to construct manifold representations of the populations.

All images were skull-stripped using BET [9],corrected for bias using N4 [10] and intensity normalized. To account for the size differences in the population, all subjects were affine aligned to an atlas template of 37 GA [11]. A non-rigid alignment was further applied with a large control point spacing to preserve detailed differences in the images. The aligned images in the atlas space were of size $117 \times 159 \times 126$ with an isotropic voxel size of 0.86 mm. The images were smoothed using a Gaussian filter with physical size of 4.3 mm in each dimension.

NAF Construction The feature vector for each image was composed of the intensities of randomly chosen voxels inside the brain mask of the atlas template. We chose a feature vector of length $Q = 100{,}000$. We trained NAFs for five different definitions of the distance function ρ: (i) F_{GA}, ρ is age difference; (ii) F_{BW}, ρ is difference in weight at birth; (iii) F_{LVV}, ρ is difference in left lateral ventricle volume; (iv) F_{CV}, ρ is difference in cerebellum volume. In addition, we trained a NAF on (v) F_{O2} using instead of a distance function the labels to train the trees. The parameters of the NAFs in the training phase were determinted empirically. We used 500 trees in each forest, in each tree $q = \mathrm{round}(\sqrt{Q}) = 316$ features are evaluated, the maximum tree depth was 12 and the minimum sample size at a leaf was 5.

Manifold Learning and Regression Isomap was applied to the affinity matrices obtained by the NAFs as described in Section 2. We kept the first 20 dimensions as new embedding coordinates. To evaluate the quality of the resulting embeddings, we predicted clinical and structural information for the test images and compared it to the real values. For each embedding, we predicted the following: (a) GA: GA at scan, (b) BW: birth weight in kg, (c) O2: O_2 supply after birth, (d) LVV: volume of the left ventricle, (e) CV: volume of the cerebellum. The prediction $\tilde{v}_i(I)$ of the real value $v_i(I)$, $i \in \{\mathrm{GA}, \mathrm{BW}, \mathrm{O2}, \mathrm{LVV}, \mathrm{CV}\}$, for a test image I is obtained using the weighted mean:

$$\tilde{v}_i(I) = \frac{\sum_{I_n \in \mathbf{N}_\rho^k} \mathbf{w}(I, I_n) v_i(I_n)}{\sum_{I_n \in \mathbf{N}_\rho^k} \mathbf{w}(I, I_n)},$$

where $\mathbf{N}_\rho^k$ is the set of k nearest neighbors of I in the training set and $\mathbf{w}(I, \cdot)$ are the affinities of I with respect to the training images.

Experiments and Results We trained the forests F_{GA}, F_{LVV}, F_{CV}, F_{BW} and F_{O2} with two different configurations. For the first option, we trained the NAFs on the whole populations such that we get pairwise affinity matrices A_i and corresponding distance matrices D_i, $i \in \{\mathrm{GA}, \mathrm{BW}, \mathrm{O2}, \mathrm{LVV}, \mathrm{CV}\}$, for the whole populations. Isomap was applied to each D_i in order to get a manifold representation of the dataset. For the evaluation of the performance in classification, we excluded the ground truth of the current test image in the regression step. For the second option, we performed leave-ten-out cross validations to estimate the performance of out-of-sample predictions. Multiple forests were constructed for each F_i, excluding ten samples each time, which were used for testing.

The regression results are shown in Table 1, where for each embedding the correlation between predicted and real value is presented. The columns correspond to the information we want to predict and the rows to the embeddings used for prediction. It can be seen that the quality of the predictions differ significantly between the different embeddings. All embeddings yield high correlation values for the prediction of the GA at scan. This is due to the fact, that the appearance of the images differ strongly between the age groups. Even the simple L2-distance between images is able to capture those differences. This is not the case for the predictions of the other clinical and structural information. All embeddings lead to poor correlation values for the prediction of the birth weight or the O_2 supply, except for the embeddings based on F_{BW} and F_{O2}, respectively. This indicates that it is challenging to estimate neighborhoods which capture all the structural and functional changes in the brain. By constructing specified neighborhoods using various similarity definition, we are able to better classify and categorize the population according to previously defined criteria. When using leave-ten-out cross validation, the correlation values decrease in most of the cases. This expected effect was particularly marked for the embedding based on F_{O2}.

Table 1. Correlation between real and predicted values by the embedding based on normal L2-distance and NAFs as explained in Section 3.

	Whole population					Leave-ten-out				
	GA	BW	O_2	LVV	CV	GA	BW	O_2	LVV	CV
L2	0.93	0.51	-0.03	0.65	0.92	0.87	0.42	0.11	0.43	0.85
F_{GA}	**0.99**	0.44	0.09	0.43	0.94	**0.96**	0.42	0.02	0.52	0.92
F_{BW}	0.93	**0.96**	0.23	0.50	0.89	0.92	**0.73**	0.23	0.52	0.87
F_{O2}	0.91	0.49	**0.81**	0.56	0.86	0.87	0.48	**0.26**	0.61	0.79
F_{LVV}	0.93	0.30	0.12	**0.95**	0.92	0.92	0.34	-0.01	**0.89**	0.90
F_{CV}	0.96	0.55	0.16	0.49	**0.99**	0.95	0.50	-0.03	0.50	**0.94**

As an example, two embeddings, obtained with the simple L2-distance and with F_{BW}, are visualized through their first two embedding coordinates in Fig. 1. The color coding is according to the weight at birth in kg. It can be clearly seen that the embedding based on the L2-distance is not able to separate the data samples according to birth weight, whereas the embedding based on F_{BW} provides a good separation.

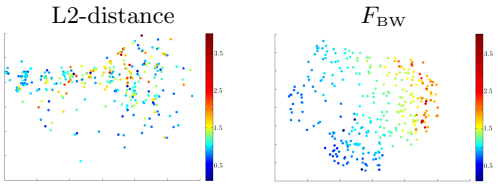

Fig. 1. Scatter plots of the first two embedding coordinates. Embeddings obtained with the L2-distance (left) and F_{BW} (right) as similarity measure. The color coding corresponds to the weight at birth in kg.

The embeddings based on NAFs are specialized embeddings, meaning that they are constructed to estimate the neighborhood according to one specific criteria. The downside of this approach can be seen, e.g., for the prediction of the left lateral ventricle volume. The embeddings based on the L2-distance, F_{GA}, F_{BW} and F_{O2} obtain rather low correlation values in the range of $0.4 - 0.65$ for predicting the left lateral ventricle volume. However, the joint embedding, which combines the affinity matrices of the NAFs trained on the ventricle volume and the birth weight, as explained in Section 2, is able to predict both information more accurately. This is shown in Table 2, where the correlation of the real and predicted values are shown for three joint embeddings based on (i) $F_{\mathrm{BW,LVV}}$: combination of F_{BW} and F_{LVV}, (ii) $F_{\mathrm{BW,CV}}$: combination of F_{BW} and F_{CV} and (iii) $F_{\mathrm{LVV,CV}}$: combination of F_{LVV} and F_{CV}.

Table 2. Correlation between real value of clinical and structural information and predicted values by the joint embedding as explained in Section 3.

	Whole population					Leave-ten-out				
	GA	BW	O_2	LVV	CV	GA	BW	O_2	LVV	CV
$F_{\mathrm{BW,LVV}}$	0.95	**0.90**	0.14	**0.95**	0.92	0.95	**0.67**	0.05	**0.88**	0.91
$F_{\mathrm{BW,CV}}$	0.97	**0.89**	0.05	0.55	**0.98**	0.97	**0.66**	0.13	0.55	**0.95**
$F_{\mathrm{LVV,CV}}$	0.95	0.42	0.20	**0.96**	**0.93**	0.94	0.37	-0.10	**0.86**	**0.93**

Figure 2 plots the regression results using the joint embedding based on $F_{\mathrm{BW\text{-}LVV}}$ and the individual embeddings of F_{BW} and F_{LVV} for predicting the weight at birth and the left lateral ventricle volume. The differences are clearly seen. Neither F_{LVV} is able to predict the birth weight, nor F_{BW} to predict the volume of the left ventricle. The correlation between predicted and real values are poor and the root mean square error (RMS) is high. The joint embedding based on both NAFs, however, yields correlation values of 0.90 for the birth weight and 0.95 for the lateral ventricle volume and smaller RMS.

During training, NAF selects the most discriminative features according to the selected distance function. In Fig. 3, the features selected in the first three levels of the trees are shown for F_{LVV} and F_{GA}. As expected, the most discrimi-

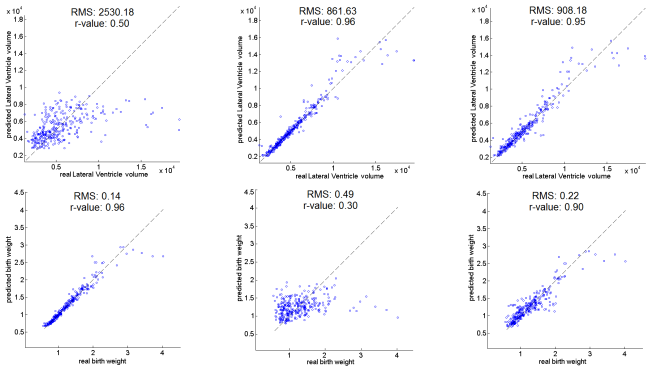

Fig. 2. Scatter plots for left lateral ventricle volume (top row) and birth weight prediction (bottom row) based on F_{BW} (left), F_{LVV} (middle) and joint $F_{\mathrm{BW,LVV}}$ (right).

native features in forest F_{LVV} are in the left lateral ventricle. The most discriminative features in forest F_{GA} are found in the deep gray matter and part of the cortex. In this regions the appearance in MR differ strongly between younger (26-28 GA) and older neonates (37-42 GA).

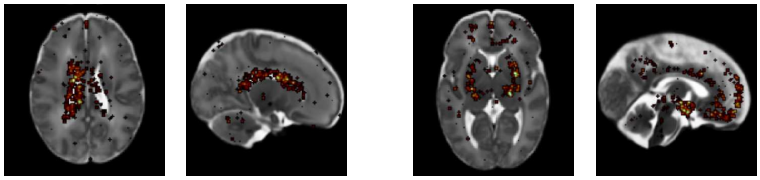

Fig. 3. Features selected in the first three levels of the trees for distance based on left ventricle volume (left: axial and sagittal) and GA at scan (right: axial and sagittal).

4 Conclusions

We have proposed a framework for manifold learning, where the pairwise similarities are learned and combined through NAFs. We used the resulting embeddings to perform classification regarding structural and clinical information. One key motivation of using NAFs is that they provide a natural way for the combination of similarities learned from multiple distances in the manifold learning step. In addition, the NAFs approximate neighborhoods based on arbitrary distances and select automatically the features which are most discriminative for the given distance function which make a priori feature extraction not necessary.

The method was applied to a population of preterm and in term neonatal MR images. We trained the NAFs on appearance features of the images based on differences in cerebellum and left lateral ventricle volume, GA at scan, birth weight and oxygen supply. The resulting embeddings were specific to the criterion their neighborhoods were trained on (structural volumes, birth weight, etc.)

and showed an accurate classification performance regarding this criterion while embeddings based on classical similarity measures fail. In particular, we showed how the combination of pairwise affinity matrices based on different NAFs can improve the overall performance of the joint embedding.

Encoding simultaneously different information (clinical and image-based) in the embeddings may help in studying abnormal brain development which is characterized by the change in multiple biomarkers.

Acknowledgements The authors would like to thank E. Konukoglu for providing the implementation of the NAFs (publicly available at http://www.nmr.mgh.harvard.edu/~enderk/software.html).

V. A. Zimmer is supported by the grant FI-DGR 2013 (2013 FI B00159) from the Generalitat de Catalunya. This research was partially funded by the Spanish Ministry of Economy and Competitiveness (TIN2012-35874).

References

1. S. Gerber, T. Tasdizen, P. Fletcher, S. Joshi, R. Whitaker: Manifold modeling for brain population analysis. Med. Imag. Anal. 14(5), 643–53 (2010)
2. P. Aljabar, D. Rueckert, W. Crum: Automated morphological analysis of magnetic resonance brain imaging using spectral analysis. NeuroImage 43(2), 225–235 (2008)
3. P. Aljabar, R. Wolz, L. Srinivasan, S.J. Counsell, M.A. Rutherford, A.D. Edwards, J.V. Hajnal, D. Rueckert: A Combined Manifold Learning Analysis of Shape and Appearance to Characterize Neonatal Brain Development. IEEE Trans. Med. Imag 30(12), 2072–86 (2011)
4. R. Wolz, P. Aljabar, J.V. Hajnal, J. Lötjönen, D. Rueckert: Nonlinear dimensionality reduction combining MR imaging with nonimaging information. Med. Imag. Anal. 16(4), 819–830 (2012)
5. K.R. Gray, P. Aljabar, R.A. Heckemann, A. Hammers, D. Rueckert for the Alzheimer's Disease Neuroimaging Initiative: Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. NeuroImage 65, 167–175 (2013)
6. E. Konukoglu, B. Glocker, D. Zikic, A. Criminisi: Neighborhood approximation using randomized forests. Med. Imag. Anal. 17, 790–804 (2013)
7. J.B. Tenenbaum, V. de Silva, J.C. Langford: A Global Geometric Framework for Nonlinear Dimensionality Reduction. Science 290 (5500), 2319–2323 (2000)
8. A. Makropoulos, I.S. Gousias, C. Ledig, P. Aljabar, A. Serag, J. Hajnal, A.D. Edwards, S. Counsell, D. Rueckert: Automatic whole brain MRI segmentation of the developing neonatal brain. IEEE Trans. Med. Imag. 33(9), 1818–1831 (2014)
9. S.M. Smith: Fast robust automated brain extraction. Human Brain Mapping 17(3), 143–155 (2002)
10. N.J. Tustison, B.B. Avants, P.A. Cook, Y. Zheng, A. Egan, P.A. Yushkevich, and J.C. Gee: N4ITK: Improved N3 Bias Correction. IEEE Trans. Med. Imag. 29(6), 1310–1320 (2010)
11. A. Serag, P. Aljabar, G. Ball, S.J. Counsell, J.P. Boardman, M.A. Rutherford, A.D. Edwards, J.V. Hajnal, D. Rueckert: Construction of a consistent high-definition spatio-temporal atlas of the developing brain using adaptive kernel regression. NeuroImage 59(3), 2255–65 (2012)