ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

A Kernel Based Multilinear SVD Approach for Multiple Sclerosis Profiles Classification

Berardino Barile[1,2], Pooya Ashtari[1,2], Francoise Durand-Dubief[1,4], Frederik Maes[2], Dominique Sappey-Marinier[1,3] and Sabine Van Huffel [2]

1-CREATIS (UMR 5220 CNRS & U1294 INSERM),
Universite Claude Bernard Lyon 1, France
2-KU Leuven, Department of Electrical Engineering (ESAT),
Division Stadius, Leuven, Belgium
3-CERMEP Imagerie du Vivant, Bron, France
4-Hospices Civils de Lyon, Service de Neurologie A, France

Abstract. In machine learning, kernel data analysis represents a new approach to the study of neurological diseases such as Multiple Sclerosis (MS). In this work, a kernelization technique was combined with a tensor factorization method based on Multilinear Singular Value Decomposition (MLSVD) for MS profile classification. Our simple, yet effective, approach generates a meaningful feature embedding of multi-view data, allowing good classification performance. The results presented in this work define an interesting approach, given that only the anatomical T1-weighted image was used, which represents the most important modality in clinical applications.

1 Introduction

Multiple Sclerosis (MS) is a chronic disease of the central nervous system and its evolution is not uniform among patients. Thus, the neurologist's challenge is to classify MS patients into four clinical phenotypes depending on the patterns of cognitive or physical impairment. In about 85% of the cases, disease onset is characterized by a first Clinically Isolated Syndrome (CIS) and/or a Relapsing-Remitting course (RRMS) followed by a Secondary-Progressive course (SPMS), while the remaining 15% of MS patients evolve directly into a Primary-Progressive course (PPMS) [1].

In this work, a factorization approach based on MultiLinear Singular Value Decomposition (MLSVD) [2], was used for the discrimination of MS patients between early stages of the disease, such as CIS and RRMS, and progressive courses (PPMS and SPMS). The multi-view information, obtained from the cortical Grey Matter (GM) surface reconstruction output of FreeSurfer [7], was exploited to enhance classification performance based on three morphological features, namely thickness, curvature and area. To our knowledge, this is the first attempt to perform classification of MS patients using a factorization method applied to kernel multi-view data. Moreover, only the anatomical MR image modality (T1-weighted) was used for the task, avoiding the burden of acquiring Diffusion Tensor Imaging (DTI) and/or functional Magnetic Resonance Imaging (MRI) data, which require long and expensive procedures. Furthermore, in order to demonstrate the generalization capacity of the proposed approach, the

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

properties of MLSVD were exploited, generating a projection matrix. Such a matrix was used to project new incoming patients in the same embedding space obtained from our labeled training dataset, and subsequently classified using Random Forest (RF) [11].

2 Data Acquisition and Pre-processing

In this study, ninety MS patients divided in four clinical profiles (12 CIS, 30 RRMS, 28 SPMS, 20 PPMS) were used. Each patient underwent several consecutive MR examinations, six on average, resulting in one scan per visit, using a 1.5T Siemens Sonata system (Siemens Medical Solution, Erlangen, Germany) with 8-channel head-coil. For a thorough description of the dataset we refer to [14]. The preprocessing of the anatomical images, was performed using Freesurfer v6.0.0 image analysis suite and three GM features, namely thickness, curvature and area were calculated. In Figure 1, an intuitive pipeline for the kernel data generation is proposed. The GM parcellation task was performed using the Desikan atlas [8], yielding 68 different brain regions. For each region, a list of values $v \in \mathbb{R}_{+}^{(d\times 1)}$, with d representing its total number (usually in the order of thousands), were calculated for each specific GM region and are called vertices. These vertex values represent small triangles on the flattened surface, following the approach proposed by [4], from which morphological features were calculated by means of the Freesurfer toolkit. For more information we refer to [5] and [6]. It is important to notice that, for each scan, more than 180.000 features were generated, which represents a problem for standard Machine Learning (ML) models due to the curse of dimensionality [13].

3 Kernelization, Tensorization and Normalization

In order to reduce the dimensionality of the problem, a three-steps approach was implemented. First, we perform kernelization of the high-dimensional dataset by means of Minkowski distance formulation. The kernelization step maps the high-dimensional data in a possibly higher-dimensional one in which relationships are assumed to be linear. Second, the obtained kernel dataset was restructured into a third order tensor which is assumed to be low-rank. Third the obtained tensor was decomposed using a tensor-based higher-order SVD (where the learnt components are not necessarily "sources" but correspond to bases) for the extraction of feature embedding used as input to the RF classifier.

3.1 Kernelization

Let's define with $K^f(x_i, x_j)$ the kernel function for the morphological feature f. $K^f(x_i, x_j)$ represents a scalar value defining (dis-)similarity between region i and region j where $i \neq j \ \forall i, j \in \{1, 2, ..., q\}$ such that:

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

$$K^f(x_i, x_j) = \left(\sum_{i,j;\forall i \neq j}^{q} |x_i^f - x_j^f|^\gamma \right)^{\frac{1}{\gamma}} \tag{1}$$

where x_i^f and x_j^f represent the set of vertex values for morphological feature f pertaining to brain region i and j respectively, and $\gamma = 1$ defines the norm. This distance measure is also called Manhattan or taxicab distance. Thus the kernel matrix $X^f \in \mathbb{R}_+^{(q \times q)}$ can be obtained by arranging each pairwise comparison $K^f(x_i, x_j)$ in a squared matrix, where i and j represent rows and columns respectively. The process was performed independently for all morphological features f, finally obtaining three real positive kernel matrices $X_t, X_c, X_a \in \mathbb{R}_+^{(q \times q)}$ (i.e., thickness, curvature and area).

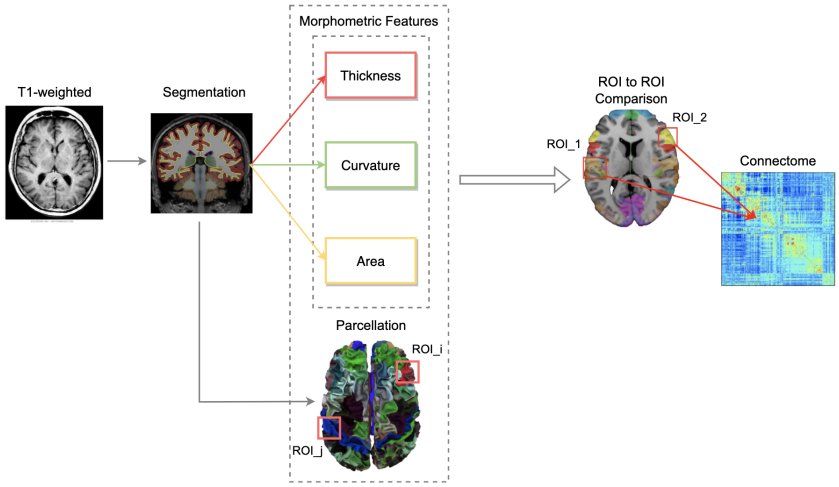

Fig. 1: Intuitive pipeline of the GM connectome data generation.

3.2 Tensorization and Normalization

The three morphometric kernels were stacked along the first dimension such that $X_{tca} \in \mathbb{R}_+^{(z \times q)}$ and $z = 3q$. Thus, X_{tca} represents the vertically stacked adjacency matrices of a patient scan. A third-order tensor can be obtained by concatenating all the stacked kernel matrices along the third mode, obtaining a tensor $\mathcal{X}_{tca} \in \mathbb{R}_+^{(z \times q \times p)}$, where p defines the number of patient scans. Finally, normalization was performed along the third mode employing a min-max operation such that each value (i.e. dissimilarity) of the matrix is standardized in the range [0 - 1].

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

4 Tensor-based multi-view factorization

Blind Source Separation (BSS) represents a general and widely used approach to discover hidden relations among different modes of high-dimensional data. In this work, a factorization approach, based on MLSVD, was employed for the extraction of feature embedding from a third-order tensor, and used for MS patient profiling. Such a tensor can be decomposed in the product of three nonzero factors, representing latent components (i.e., factor loadings), each of which describing a specific mode of the tensor. In this work, we refer to multilinear rank $L = [r_1, r_2, r_3] \in \mathbb{N}^3$, where possibly $r_1 \neq r_2 \neq r_3$, as a triplet of values such that r_n describes the mode-n rank of the mode-n unfolding of the original third-order tensor $\mathcal{X}$, which we denote as $X_{(n)}$.

4.0.1 Multilinear Singular Value Decomposition

The general method is based on the Singular Value Decomposition (SVD) obtained from a higher-order tensor. Mathematically, we can define the MLSVD decomposition as follows:

$$\mathcal{X} \approx \mathcal{S} \bullet_1 U^{(1)} \bullet_2 U^{(2)} \bullet_3 U^{(3)} \tag{2}$$

where $U^{(1)} \in \mathbb{R}^{(z \times r_1)}$, $U^{(2)} \in \mathbb{R}^{(q \times r_2)}$ and $U^{(3)} \in \mathbb{R}^{(p \times r_3)}$ respectively form orthonormal bases for the first, second and third mode of the tensor $\mathcal{X}$. In other words, the matrices $U^{(1)}$, $U^{(2)}$, $U^{(3)}$ represent respectively the first r_1,r_2 and r_3 left singular vectors of the SVD of the mode-1, mode-2 and mode-3 unfolding of $\mathcal{X}$. The operator "$\bullet_n$" represents a mode-n tensor-to-matrix product. The core tensor $\mathcal{S} \in \mathbb{R}^{(r_1 \times r_2 \times r_3)}$ explains the interaction between the different modes and has all-orthogonality properties meaning that all slices of $\mathcal{S}$ are mutually orthogonal. Moreover, this structured representation ensures that variation in each mode is captured independently. For a more in-depth description of the model, we refer to [2].

4.0.2 Feature Extraction and Classification

In this work, we exploited the properties of the MLSVD factorization model to generate feature embedding able to classify MS patients. Practically speaking, from Eq. 2 it is possible to rearrange the terms on the right-hand-side such that:

$$\vec{x}_{tca} \approx u^{(3)} S_{(3)} \left(U^{(1)} \otimes U^{(2)} \right)^T \tag{3}$$

where $S_{(3)} \in \mathbb{R}_+^{(r_3 \times r_1 r_2)}$ represents the mode-3 unfolding of the core tensor $\mathcal{S}$, $\otimes$ defines kronecker product and $u^{(3)} \in \mathbb{R}_+^{(1 \times r_3)}$ represents the row vector of the mode-3 matrix $U^{(3)}$ associated to the multi-view matrix X_{tca} of a specific patient scan. Also, let's define with $\vec{x}_{tca} \in \mathbb{R}_+^{(1 \times zq)}$ the vectorization of the matrix X_{tca}. Consequently, $\vec{X}_{tca} \in \mathbb{R}_+^{(p \times zq)}$ represents the vectorization matrix describing all patients. The embedding space can be obtained as follows:

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

$$X_{embed} = \vec{X}_{tca} \left(U^{(1)} \otimes U^{(2)} \right) S_{(3)}^{T} \tag{4}$$

where $X_{embed} \in \mathbb{R}^{(p \times r_3)}$ represents the embedded matrix used for classification, while $S_{(3)}^{T}$ represents the transposed matrix of the mode-3 core tensor unfolding. RF was used as classifier and trained on the embedding representation of the multi-view data. The number of trees was set to 300 for computational reasons.

5 Results and Discussion

The experiment was performed employing a stratified 10-fold Cross Validation (CV) strategy. Each patient (including all consecutive scans to guarantee independency of the test set) was randomly assigned to only one of the 10 folds, thereby preserving the percentage of samples of each class. Also, at each fold iteration, factorization was applied only on the training data (9 folds) while the test data (left-out fold) were hold out. The projection matrix $\left(U^{(1)} \otimes U^{(2)} \right) S_{(3)}^{T}$ was used to project each left-out multi-view matrix (i.e. test set) in the desired embedding space.

Mode-3 Rank	**F1**	**Precision**	**Recall**	**AUC**
50	0.680 (0.118)	0.693 (0.117)	0.687 (0.118)	0.699 (0.132)
100	0.679 (0.124)	0.688 (0.122)	0.688 (0.124)	0.721 (0.154)
150	0.681 (0.128)	0.696 (0.127)	0.689 (0.128)	0.713 (0.162)
200	0.713 (0.131)	0.728 (0.127)	0.718 (0.129)	0.741 (0.118)
250	0.672 (0.116)	0.686 (0.115)	0.679 (0.114)	0.751 (0.131)
300	0.677 (0.127)	0.693 (0.121)	0.684 (0.124)	0.748 (0.151)

Table 1: Classification performance between early MS stages (CIS+RRMS) and progressive MS courses (PPMS+SPMS). Standard deviations were reported in parentheses.

Table 1 reports the classification results obtained when comparing early MS stages (CIS and RRMS) with progressive MS courses. All possible mode-3 multilinear rank combinations between 50 and 300 at steps of 50 were considered. Mode-1 and mode-2 multilinear ranks were constrained to their maximal value equal to $q = 68$, representing the size of the kernel data, following the approach proposed in [9]. Notice that the results remain stable at different mode-3 multilinear rank values, which defines the dimensionality of the embedding space to which we project new incoming patients. Specifically, for almost all mode-3 multilinear ranks, the Area Under the Operating Curve (AUC) scored higher than 0.7, which represents a good result considering the low amount of data and that only the T1-weighted image was used. For completeness, the F1 score as well as the Precision and Recall scores were also reported. In this case, lower values were obtained due to the thresholding binarization used to discretize probabilities, imposed at a standard value of 0.5. Conversely, the AUC score represents

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

a threshold free classification metric and it might represent a more accurate measure of classification performance.

6 Conclusion

In this work, a new pipeline was implemented for the classification of MS patients. Interesting results were obtained considering that only the T1-weighted image modality was used. To our knowledge, this is the first attempt to apply tensor factorization for MS patients profiling. We demonstrated that the proposed approach provides stable results for different mode-3 multilinear rank values, which represents the size of the embedding space used for classification. In future works, we aim to improve the results by further optimizing the factorization approach, used for dimensionality reduction, including relevant constraints and automated rank-selection. Also, our method will be applied on a much larger dataset (Observatoire Francais de la Sclerose en Plaques - OFSEP), demonstrating the generalization performances of the proposed pipeline.

References

[1] Dobson R., Giovannoni G., Multiple sclerosis: a review, 26(1):27-40, 2019

[2] Lathauwer L.D., Moor B.D., Vandewalle J., A multilinear singular value decomposition, SIAM J. Matrix Anal. Appl, 21(4):1253-1278, 2000

[3] S. Basu, A. Banerjee, R. Mooney. Semi-supervised Clustering by Seeding. Proceedings of the 19th International Conference on Machine Learning (ICML-2002), pp. 19-26, Sydney, Australia, July 2002

[4] E.L. Schwartz, B. Merker Computer-Aided Neuroanatomy: Differential Geometry of Cortical Surfaces and an Optimal Flattening Algorithm. IEEE Comp. Graph. Appl. 6:36-44, 1986

[5] A. Dale, B. Fischl, M.I. Sereno, Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. NeuroImage, 9(2):179-194, 1999

[6] B. Fischl, M.I. Sereno, A. Dale. Cortical Surface-Based Analysis: II: Inflation, Flattening, and a Surface-Based Coordinate System. NeuroImage, 9(2):195-207, 1999

[7] M. Reuter and N.J. Schmansky and H.D. Rosas. Within-subject template estimation for unbiased longitudinal image analysis. Neuroimage, 61:1402-1418,2012

[8] M. Rosenke, K.S. Weiner, M. Barnett et al. A cross-validated cytoarchitectonic atlas of the human ventral visual stream. NeuroImage, 2017

[9] Bousse M., Goovaerts G., Vervliet N. et al. Irregular heartbeat classification using kronecker product equations. International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 92:438-441, 2017

[10] Rabanser S., Shchur O., Gunnemann S. Introduction to tensor decompositions and their applications in machine learning. arXiv preprint, 1711.10781, 2017

[11] L. Braiman, Random forests, Machine Learning 45, 5-32 (2001).

[12] V. Satopaa, J. Albrecht, D. Irwin, Finding a "Kneedle" in a Haystack: Detecting Knee Points in System Behavior. International Conference on Distributed Computing Systems Workshops. 166-171, 2011

[13] R. Bellman, Dynamic programming. Science. 153(3731): 34-37, 1966

[14] G. Kocevar, C. Stamile, S. Hannoun et al. Graph Theory-Based Brain Connectivity for Automatic Classification of Multiple Sclerosis Clinical Courses. Frontiers in Neuroscience. 10, 2016