ESANN 2010 proceedings, European Symposium on Artificial Neural Networks - Computational Intelligence and Machine Learning. Bruges (Belgium), 28-30 April 2010, d-side publi., ISBN 2-930307-10-2.

Spectral Prototype Extraction for dimensionality reduction in brain tumour diagnosis

Sandra Ortega-Martorell[1,2], Iván Olier[3], Alfredo Vellido[4], Margarida Julià-Sapé[2,1] and Carles Arús[1,2] *

1- Departament de Bioquímica i Biología Molecular - Unitat de Biociències - UAB
08193, Cerdanyola del Vallès (Barcelona) - Spain

2- Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales
y Nanomedicina (CIBER-BBN)
08193, Cerdanyola del Vallès (Barcelona) - Spain

3- Institut de Neurociències - UAB
Edifici M, Despatx M3/206, 08193, Cerdanyola del Vallès (Barcelona) - Spain

4- Dept. de Llenguatges i Sistemes Informàtics - UPC
Edifici Omega, Campus Nord, 08034, Barcelona - Spain

Abstract. Diagnosis in neuro-oncology can be assisted by non-invasive data acquisition techniques such as Magnetic Resonance Spectroscopy (MRS). From the viewpoint of computer-based brain tumour classification, the high dimensionality of MRS poses a difficulty, and the use of dimensionality reduction (DR) techniques is advisable. Despite some important limitations, Principal Component Analysis (PCA) is commonly used for DR in MRS data analysis. Here, we define a novel DR technique, namely Spectral Prototype Extraction, based on a manifold-constrained Hidden Markov Model (HMM). Its formulation within a variational Bayesian framework imbues it with regularization properties that minimize the negative effect of the presence of noise in the data. Its use for MRS pre-processing is illustrated in a difficult brain tumour classification problem.

1 Introduction

Brain tumour diagnosis is a sensitive and complex task. Due to the anatomical constraints of these pathologies, experts' decision making is strongly supported by information acquired through non-invasive measurement methods. Some of the most commonly used techniques for this task are image-based, such as Magnetic Resonance Imaging (MRI). Other Magnetic Resonance (MR) techniques, such as single-voxel proton MRS (SV-^{1}H-MRS), provide metabolic information about the tissues investigated. However, one of the main difficulties for the analysis of MRS data is their high dimensionality, given that spectral frequencies are

*This work was funded by CIBER-BBN (an initiative of the Spanish ISCiii); MICINN projects PHENOIMA: SAF2008-03323 and AIDTUMOUR: TIN2009-13895-C02-01. Partial funding also contributed by the European Commission: eTUMOUR (FP6-2002-LIFESCIHEALTH 503094) and HealthAgents (IST2004-27214). The authors thank former INTERPRET partners: C. Majós (IDI-Bellvitge), A. Moreno-Torres (CDP), J. Griffiths and F. Howe (SGUL), A. Heerschap (RU), W. Gajewicz (MUL) and J. Calvar (FLENI) for access to data.

taken to be data variables. This is further complicated by the usually small number of cases available in medical MRS databases of brain tumours.

Feature Selection (FS) and Feature Extraction (FE) for DR are often performed in MRS datasets prior to diagnostic classification [1, 2]. PCA is, by far, the FE technique most commonly used in MRS data analysis. Unfortunately, PCA has some important limitations in this scenario: First, the DR results are bound to be affected by the presence of uniformative noise in the data, which is commonplace in MRS. Secondly, the resulting components must be selected according to some *ad hoc* thresholding on the basis of variance retention. Thirdly, each component is a linear combination of all the spectral frequencies; this seriously limits the interpretability of the results, which is paramount in brain tumour diagnosis. Finally, PCA (and, by that matter, other FE techniques used for the same purpose, such as Independent Component Analysis: ICA) completely bypasses the fact that MRS data do not comply with the independent and identically-distributed (i.i.d.) condition.

In this brief paper, we define a novel FE technique, namely Spectral Prototype Extraction (SPE), that overcomes all of the aforementioned limitations of PCA and similar techniques. It is based on a manifold-constrained HMM, suitable for non-i.i.d. data, and its formulation within a variational Bayesian framework imbues it with regularization properties that minimize the negative effect of the presence of noise in the data. This model, Variational Bayesian Generative Topographic Mapping Through Time (VB-GTM-TT: [3]) segments the MRS in an interpretable way. Its use for FE in MRS is illustrated in a difficult brain tumour classification problem: that of discriminating between glioblastomas and metastases, two types of agressive brain tumours.

2 Spectral Prototype Extraction using VB-GTM-TT

Manifold learning techniques are meant to model usually complex and high-dimensional multivariate data through simpler low-dimensional, manifold-based representations. When defined within the Statistical Machine Learning framework, they can be made to rely in sound principles, while embodying attractive properties such as adaptive parameter optimization and modularity.

Generative Topographic Mapping Through Time, or GTM-TT [4], is one such technique, defined as a constrained HMM and capable of providing simultaneous clustering and visualization of multivariate non-i.i.d. data such as time series and spectra. This model was recently assessed in some detail in [5].

2.1 Variational Bayesian GTM-TT

The presence of uninformative noise in a dataset and the associated potential problem of data overfitting can seriously hamper the modeling of non-i.i.d. data. In its basic formulation, GTM-TT is prone to overfitting unless active regularization methods are applied. The reformulation of this model within a Variational Bayesian framework confers it with regularization capabilities in a natural way, avoiding unnecessary approximations. The resulting Variational Bayesian

GTM-TT (VB-GTM-TT) has been shown to deal effectively with the problem of overfitting caused by model learning in the presence of noise [3].

Avoiding a direct Maximum Likelihood approach -that might require the use of approximations in algorithms such as Expectation-Maximization (EM) for adaptive parameter estimation- variational inference allows the definition of a lower bound for the marginal log-likelihood of the model, defined as

$$\ln p\left(\mathbf{X}\right) = \ln \int \sum_{\text{all } \mathbf{Z}} p\left(\mathbf{Z}, \mathbf{X}|\mathbf{\Theta}\right) p\left(\mathbf{\Theta}\right) d\mathbf{\Theta} \tag{1}$$

where $\mathbf{X}$ are the MRS data; $\mathbf{Z}$ are the hidden states defined by the model; and $\mathbf{\Theta}$ are the model parameters, including a matrix with the centroids or prototypes embedded in the model manifold $\mathbf{Y}$, initial state probabilities $\boldsymbol{\pi}$, and transition probabilities $\mathbf{A}$. These parameters depend, in turn, on a set of hyperparameters $\boldsymbol{\nu}$, $\boldsymbol{\lambda}$, ϵ, α, d_β, s_β. The complete model is graphically illustrated by Fig. 1. Details on the calculations involved are beyond the scope of this paper and can be found elsewhere [3].

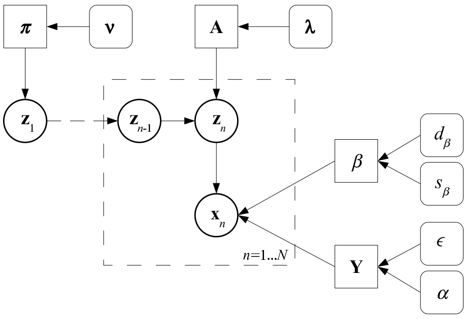

Fig. 1: Graphical representation of the Bayesian GTM-TT model. Variables are noted by circles, parameters, by squares, and hyperparameters, by rounded squares.

2.2 Spectral Prototype Extraction

The VB-GTM-TT provides a hidden space representation of the data by assigning each point (in the case of MRS data, each spectral frequency) to the hidden state bearing maximum responsibility for the generation of that point. Hidden states are arranged in a regular and topology-preserving 2-D grid for data visualization. This mapping assignment (equivalent to a cluster-membership assignment) is carried out according to a mode-projection that takes the form $h_n^{mode} = \underset{k}{\texttt{argmax}} \langle z_{k,n} \rangle$, where the variational parameter $\langle z_{k,n} \rangle$ is calculated as part of the model estimation of its adaptive parameters and represents the probability for each hidden state of being the generator of each data point. Each spectral frequency is therefore assigned to what we call here a spectral prototype (SP). Importantly, previous research [3] has shown that VB-GTM-TT, unlike unregularized counterparts, models data using a very limited number of non-empty hidden states (SPs). This behaviour eliminates the need to select a number of extracted features.

Another consequence of this assignment procedure is that each of the MRS frequencies will be assigned solely to one SP. Moreover, given the HMM-based nature of the model, each SP is likely to consist of complete intervals of frequencies or collections of these intervals. All these should make the interpretation of the SPE process easier than PCA and similar methods.

3 MRS of human brain tumours

The analyzed data were extracted from a multi-center, international database [6] resulting from the INTERPRET European research project [7] and processed according to [1].

The echo time (TE) is an influential parameter in MRS data acquisition. In this study, we analyze data acquired at both short (SET) and long (LET) TE, and their combination. The data sets analyzed consist of SV-^{1}H-MRS spectra acquired at 1.5T from brain tumour patients: 124 SET, including 86 glioblastomas (gl) and 38 metastases (me); 109 LET, including 78 gl and 31 me; and 109 items built by combination (through concatenation)[8] of the spectra measured at LET and SET for the same patients. 195 frequency intensity values measured in parts per million (ppm) were used from each spectrum, in the [4.24-0.50]ppm interval. These frequencies become the data features in all cases.

4 Experiments

The reported experiments aim, first, to assess the interpretably of the extracted SP and, second, to compare PCA and SPE as FE techniques for the preprocessing of the available MRS prior to classification. The application of SPE was followed by a sequential forward (greedy stepwise) process of selection of the most relevant SP. A subsequent classification step, using Linear Discriminant Analysis (LDA), made use of these selections. These processes were implemented in SpectraClassifier [9]. Classifier results were validated through bootstrap with 1,000 repetitions, and averaged accuracy (AA) and standard deviation (SD) values as well as the Area Under the ROC Curve (AUC) were obtained.

4.1 Results and discussion

Due to space limitations, illustrative examples of the SPE results are shown, and their interpretation in metabolic terms is discussed, for LET (Fig. 2) and SET (Fig. 3), but not for LET+SET.

A sample of comparative classification results of the application of PCA and SPE together with LDA is compiled in Table 1. Both models yield similar classification performances for LET, SET and for the combination of LET+SET. Statistically significant differences (according to a t-Student test with $p \leq 0.05$ significance threshold) were found in favour of SPE with LET, and in favour of PCA with SET. The combination of LET+SET did not produce statistically

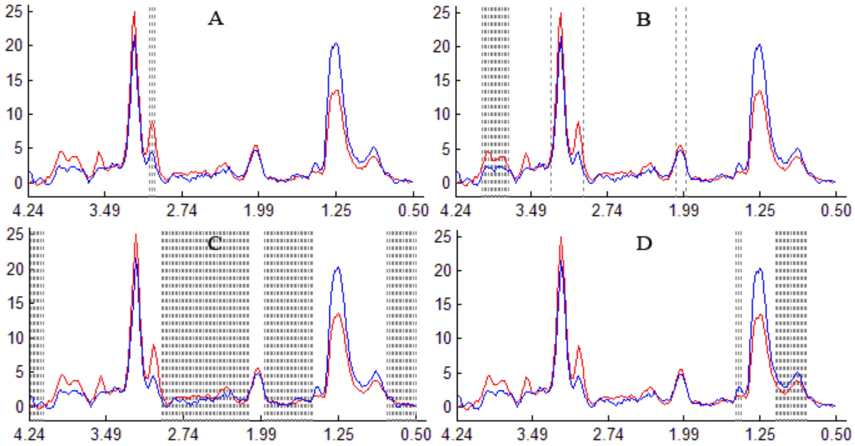

Fig. 2: Four of the 20 SP describing LET. In this figure, red is the mean spectrum of glioblastomas and blue the mean spectrum of metastases. **A)** The SP consists of 3 frequencies corresponding to Creatine (from 3.05 to 3.01ppm). **B)** The SP consists of 18 frequencies corresponding to posible metabolite contributions that resonate in the 3.95 to 3.72ppm range, namely Alanine/Glx/Phosphocreatine and Creatine; other frequencies found were Taurine/Myo-inositol/Scyllo-inositol (3.30ppm), Phosphocreatine and Creatine/Glutathione (2.97ppm), Mobile Lipids (2.07 and 1.97ppm). **C)** The 92 frequencies of this SP correspond mainly to baseline noise and/or minoritary components. **D)** The SP consists of 20 frequencies corresponding to Alanine (from 1.48 to 1.44ppm) and Mobile Lipids (from 1.09 to 0.79ppm).

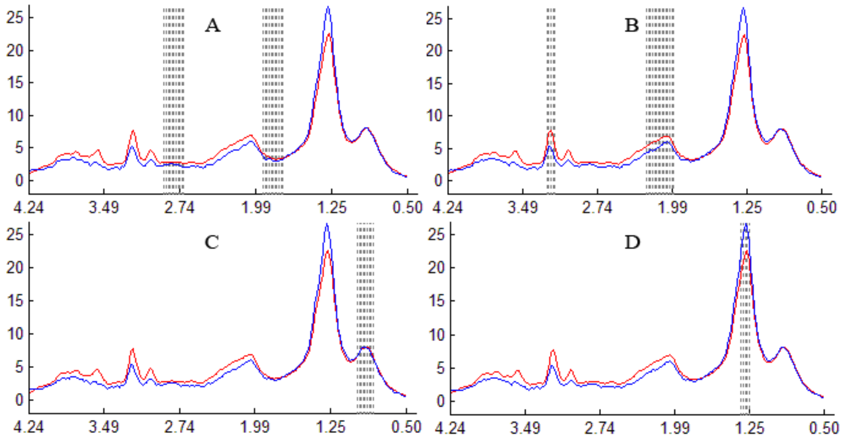

Fig. 3: Four of the 22 SP describing SET. Again, the mean spectrum of glioblastomas and metastases are displayed, in turn, in red and blue. **A)** The SP consists of 22 frequencies corresponding to Glutathione and Mobile Lipids (ML), from 2.90 to 2.70ppm, and from 1.92 to 1.73ppm, respectively. **B)** The SP consists of 20 frequencies corresponding to total Choline (from 3.24 to 3.16ppm), and ML/Glx/Macromolecules (from 2.24 to 1.97ppm). **C)** The SP consists of 9 frequencies corresponding to ML centered at 0.9ppm (from 0.98 to 0.82ppm). **D)** The SP contains 5 frequencies corresponding to ML (from 1.32 to 1.25ppm).

significant differences. Furthermore, both approaches perform better with LET than with SET. The combination of both echo times does not improve the performance significantly for the discrimination challenge investigated. In conclusion, SPE has been shown to be competitive as FE method previous to classification, while improving on the interpretability of PCA. Future research should compare the reported classification results obtained with LDA with those obtained with alternative nonlinear classifiers.

	PCA				**SPE**			
	PC	AA±SD	AA±SD per class	AUC	SP	AA±SD	AA±SD per class	AUC
LET	17	76.85±4.08	gl: 76.61±4.81	0.812	18	78.03±3.98	gl: 74.39±4.92	0.847
			me: 77.52± 7.37				me: 87.18±5.99	
SET	17	70.05±4.11	gl: 69.57±4.97	0.769	22	67.85±4.32	gl: 67.63±5.07	0.771
			me: 71.17±7.59				me: 68.24±7.55	
LET+ SET	20	77.05±3.98	gl: 73.05±4.92	0.846	19	77.10±4.13	gl: 71.88±5.26	0.828
			me: 87.02±5.85				me: 90.18±5.27	

Table 1: Classification results for PCA-based LDA and SPE-based LDA. The results chosen for comparison are the best ones obtained for each method.

References

[1] A. Vellido, E. Romero, F.F. González-Navarro, Ll. Belanche-Muñoz, M. Julià-Sapé and Arús, C. Outlier exploration and diagnostic classification of a multi-centre ^{1}H-MRS brain tumour database. *Neurocomputing*, 72(13-15):3085-3097, Elsevier, 2009.

[2] L. Lukas, et al., Brain tumor classification based on long echo proton MRS signals. Artificial Intelligence in Medicine, 31:73-89, Elsevier, 2004.

[3] I. Olier and A. Vellido, A variational formulation for GTM Through Time. In proceedings of the *International Joint Conference on Neural Networks* (IJCNN 2008), pages 517-522, Hong Kong, 2008.

[4] C.M. Bishop, G. Hinton and I. Strachan, GTM Through Time. In proceedings of the IEE Fifth *International Conference on Artificial Neural Networks*, pages 111-116, Cambridge, U.K., 1997.

[5] I. Olier and A. Vellido, Advances in clustering and visualization of time series using GTM Through Time. *Neural Networks*, 21(7):904-913, Elsevier, 2008.

[6] M. Julià-Sapé, D. Acosta, M. Mier, C. Arús, D. Watson and The INTERPRET Consortium: A multi-centre, web-accessible and quality control checked database of in vivo MR spectra of brain tumour patients. Magnetic Resonance Materials Physics, Biology and Medicine MAGMA, 19:22-33, Springer, 2006.

[7] INTERPRET project (International Network for Pattern Recognition of Tumours using Magnetic Resonance. IST-1999-10310). Project URL: http://azizu.uab.es/INTERPRET.

[8] J.M. García-Gómez, S. Tortajada, C. Vidal, M. Julià-Sapé, J. Luts, A. Moreno-Torres, S. Van Huffel, C. Arús, M. Robles: The effect of combining two echo times in automatic brain tumor classification by MRS. NMR in Biomedicine 2008 Nov, 21(10):1112-1125.

[9] S. Ortega-Martorell, I. Olier, M. Julià-Sapé and C. Arús, TumourClassifier, a Java tool for fast development and implementation of MRS-based classifiers. In proceedings of the *17th Scientific Meeting & Exhibition of the International Society for Magnetic Resonance in Medicine* (ISMRM), page 3477, Honolulu, U.S.A., 2009.