

Machine learning methods for classifying normal vs. tumorous tissue with spectral data

Félix F. González-Navarro^{1,2} and Lluís A. Belanche-Muñoz²

¹ Instituto de Ingeniería, Universidad Autónoma de Baja California
Mexicali, México

² Dept. de Llenguatges i Sistemes Informàtics
Universitat Politècnica de Catalunya, Barcelona, Spain.
{fgonzalez,belanche}@lsi.upc.edu

Abstract. Machine learning is a powerful paradigm within which to analyze ¹H-MRS spectral data for the automated classification of tumor pathologies aimed to facilitate clinical diagnosis. The high dimensionality of the involved data sets makes the discover of computational models a challenging task. In this study we apply a feature selection algorithm in order to reduce the complexity of the problem. The obtained experimental results yield a remarkable classification performance of the final induced models, both in terms of prediction accuracy and number of involved spectral frequencies. A dimensionality reduction technique that preserves the class discrimination capabilities is used for the visualization of the final selected frequencies, thus enhancing their interpretability.

Keywords: Brain tumor classification, Feature Selection, Visualization

1 Introduction

Proton Magnetic Resonance Spectroscopy (¹H-MRS) is a powerful technique that helps to observe metabolic processes in living tissue [1]. Although these metabolic functions are not fully understood, it is possible to employ *machine learning* (ML) techniques on this kind of data for the diagnosis and grading of adult brain tumors [2]. Several recent examples in the literature use ML techniques for distinguishing between different brain tumor pathologies (e.g. [3], [4]). Due to the high dimensionality (nearly 390 spectral points in this study), many efforts have used dimensionality reduction methods (mainly feature extraction) to lower the complexity of the problem. Methods such as Principal Components Analysis (PCA) have recognized drawbacks that limit their applicability in this problem: the result is difficult to interpret in original terms and PCA it is not designed to preserve class separability. In the present study we are interested in performing a feature selection study in two types of ¹H-MRS spectral data by constructing an *ad hoc* combination of both. We use a simple filter technique for feature selection as a fast method to generate relevant subsets of spectral frequencies, complemented by the use of *bootstrapping* techniques. In this paper, we report experimental results that support the practical advantage of combining robust feature selection and classification techniques in this application field.

2 An entropic filtering algorithm

Mutual Information (MI) measures the mutual dependence of two random variables. It has been used with success as a criterion for feature selection in machine learning tasks. In this work we use this concept embedded in a fast algorithm that computes MI between a set of variables and the class variable by generating first a “super-feature”, obtained considering the concatenation of each combination of possible values of its forming features. In symbols, let $X = \{X_1, \dots, X_n\}$ be the original feature set and consider a subset $\tau = \{\tau_1, \dots, \tau_k\}$. A single feature \mathcal{V}_τ can be obtained uniquely, whose possible values are the concatenations of all possible values of the features in τ . The conditional entropy between \mathcal{V}_τ and the class feature Y is then:

$$H(Y|\tau_1, \dots, \tau_k) = H(Y|\mathcal{V}_\tau) = - \sum_{v \in \mathcal{V}_\tau} \sum_{y \in Y} p(v, y) \log \frac{p(v, y)}{p(y)}. \quad (1)$$

Proceeding in this way, the MI can be determined as a simple bivariate case: $I(\mathcal{V}_\tau; Y) = H(Y) - H(Y|\mathcal{V}_\tau)$. An *index of relevance* of the feature $X_i \in X$ to a class Y with respect to a subset $\tau \subset X$, inspired on [5], is given by:

$$R(X_i; Y|\tau) = \frac{I(X_i; Y|\mathcal{V}_\tau)}{H(Y|\mathcal{V}_\tau)} = \frac{H(Y|\mathcal{V}_\tau) - H(Y|X_i; \mathcal{V}_\tau)}{H(Y|\mathcal{V}_\tau)}. \quad (2)$$

This index of relevance of a feature subset is to be maximized (it has a maximum value of 1). This measure is used in this study to evaluate subsets of spectral frequencies, embedded into a filter forward-selection strategy, conforming the *Entropic Filtering Algorithm* (EFA). A detailed description and a fast implementation of the whole algorithm can be found in [reference removed due to blind evaluation]

3 Experimental setup

The *echo time* is an influential parameter in ^1H -MRS spectra acquisition. In short-echo time (SET) spectra (typically 20-40 ms) some metabolites are better evaluated (e.g. lipids, myo-inositol, glutamine and glutamate). However, there may be numerous overlapping resonances (e.g. glutamate/glutamine at 2.2 ppm and NAA at 2.01 ppm) which make the spectra difficult to interpret [6]. A long-echo time or LET (270-288 ms) yields less metabolites but also less baseline distortion, resulting in a more readable spectrum. There are a few studies comparing the classification potential of the two types of spectra (see e.g. [6], [7]). These works seem to give a slight advantage to using SET information or else suggest a combination of both types of spectra, which is the possibility explored in this work. The analyzed ^1H -MRS data sets are detailed as follows:

- 195 single voxel *long-echo time LET* spectra acquired *in vivo* from brain tumor patients, including: meningiomas (55 cases), glioblastomas (78), metastases (31), astrocytomas Grade II (20), oligoastrocytomas Grade II (6) and oligodendrogliomas Grade II (5);

- 217 single voxel *short-echo time SET* spectra: meningiomas (58 cases), metastases (38), glioblastomas (86), astrocytomas Grade II (22), oligoastrocytomas Grade II (6) and oligodendrogliomas Grade II (7).
- 195 single voxel *long/short-echo time LSET* spectra, obtained by merging the 195 common observations of the two previous data sets.

The merged LSET spectra were bundled into four pathology groupings (super-classes): *Normal tissue* (normal brain tissue white matter or brain abscesses), *High-grade malignant tumors* (metastases and glioblastomas), *Low-grade malignant gliomas* (astrocytomas, oligodendrogliomas and oligoastrocytomas) and *Others* (the rest of tumors). The spectra consist of 390 frequency intensity values, from 4.21 ppm down to 0.51 ppm. Three different experiments were designed in order to test the capability of the proposed techniques in distinguishing normal vs. tumorous tissue: *1-Normal tissue vs. high-grade malignant tumors* –named EXP1–, *2-Normal tissue vs. low-grade malignant tumors* –named EXP2– and *3-Normal tissue vs. Others* –named EXP3– generating three data sets (LSET1, LSET2 and LSET3), one for each experiment.

To obtain more reliable relevant features we advocate for the use of *bootstrap* resampling in the feature selection process. This technique is here used to yield mean performance estimates and their variability, and thus a more reliable measure of predictive ability; these methods are also well-suited when sample size is small or the distribution of the statistic is unknown. The three data sets LSET1 to LSET3 were used to generate 1,000 bootstrap samples for each experiment.

First, the EFA is applied to the bootstrap samples to obtain a Best Spectral Subset for each experiment (named BSS1 to BSS3). Notice that each relevance value calculated in the EFA is the *average* across the bootstrap samples, to guide and stabilize the algorithm. Second, the classifier development stage is conducted using the original datasets (LSET1 to LSET3) with the obtained BSSs. Five classifiers were evaluated by means of 10 times 10-fold Cross Validation (10x10cv): a *nearest-neighbor* (kNN) with parameter k (number of neighbors), the *Naïve Bayes classifier* (NBC), a *Linear Discriminant classifier* (LDC), *Support Vector Machine* with *linear kernel* (lSVM) and parameter C (regularization constant) and *Support Vector Machine* with *radial kernel* (rSVM) and parameters C and σ^2 (amount of smoothing in the kernel). The parameters are optimized via a grid search process using 10x10cv.

4 Experimental results

The obtained BSSs are detailed in Table 2 and plot in Fig. 1 against average spectra per super-class. All three (BSS1, BSS2 and BSS3) deliver maximum relevance ($R = 1$) with as little as 25, 7 and 16 spectral frequencies, respectively. It is observed that LET frequencies are preferred over SET frequencies; selected spectral points are specially located in the Choline/Creatine (3.32 ppm to 3.01 ppm), Glutamate/Glutamine (2.71 ppm to 2.41 ppm), N-acetylaspartate (2.10 ppm to 1.80 ppm) and Lipids (around 1.19 ppm) peaks.

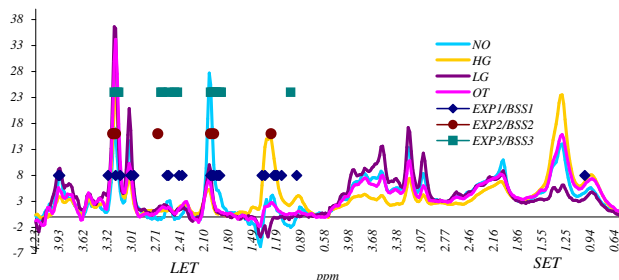


Fig. 1. Spectral subsets found by the EFA that reach maximum relevance ($R = 1$) as positioned in the whole spectrum. The resonance frequency position of each peak on the plot is dependent on the chemical environment of the nucleus and is usually expressed as parts per million or ppm.

Experiment	NBC	kNN	LDC	ISVM	rSVM
EXP1	92.81±0.005	95.32±0.002	94.47±0.04	96.45±0.004	96.62±0.004
EXP2	95.05±0.007	98.19±0.000	95.22±0.006	98.00±0.003	98.37±0.002
EXP3	92.68±0.006	95.33±0.005	91.11±0.006	94.21±0.007	94.22±0.005

Table 1. 10x10cv accuracies (in %) and standard errors per experiment and classifier using the selected subsets of spectral points BSS1 to BSS3 –see Fig. 1. Best results are marked in bold.

Experiment	Best Spectral Subset BSS
EXP1	L2.04, L2.01, L1.99, L1.19, L2.43, L1.95, L0.94, L1.93, L1.21, L1.91, L2.56, L1.13, L1.23, L3.03, L2.58, L1.34, L1.38, L3.17, L3.95, L3.22, L3.93, L2.39, L3.32, S1.02, L3.00
EXP2	L3.22, L3.24, L2.04, L1.99, L2.69, L3.26, L1.27
EXP3	L3.19, L3.20, L2.04, L2.01, L1.99, L3.24, L1.97, L2.44, L2.63, L2.56, L2.65, L1.93, L1.91, L1.89, L2.48, L1.02

Table 2. Spectral subsets found by the EFA that reach maximum relevance ($R = 1$). The prefix (L or S) indicates LET or SET origin.

Classification stage results in Table 4 show remarkable values, specially when using the rSVM. Confidence intervals (CI) for these results (at the 95% level) are: (95.91-97.73) for the rSVM (EXP1), (98.01-98.73) for the rSVM (EXP2) and (94.31, 96.35) for kNN (EXP3). Note that the CIs are about 1% to 2% wide, a relatively low value, signaling a consistently stable generalization ability. A Wilcoxon signed rank test is carried out as a way to assess significant differences between the best result and all other classifiers, for each experiment –Table 4.

Moreover, in order to ascertain which super-classes are the most difficult to separate, the full confusion matrices of the three experiments are shown in Table 3. It can be seen how average prediction is very accurate. However, there remain some data observations (one, or two at the most) that are wrongly predicted as non-tumorous when in fact they are, which is serious. These observations should receive special attention by an oncologist to check whether they are correct.

In a medical context, data visualization in a low-dimensional representation space may become extremely important, helping radiologists to gain insights into what undoubtedly is a complex domain. We use in this work a method based on the decomposition of the data scatter matrix, with the property of maximizing

True Class	EXP1 (rSVM)		True Class	EXP2 (kNN)		True Class	EXP3 (rSVM)	
	NO	HG		NO	LG		NO	OT
NO	6.37±0.18	1.30±0.18	NO	10.03±0.10	0.03±0.08	NO	7.07±0.19	0.60±0.17
HG	0.02±0.07	36.13±0.09	LG	0.00±0.00	7.67±0.09	OT	0.50±0.20	15.50±0.20

Table 3. 10x10cv average confusion matrices (using the rSVM on EXP1 and EXP2 and kNN on EXP3). These results correspond to the highest accuracy values (Table 4) in each experiment. True class falls vertically. Super-classes are as follows: NO is Normal tissue, HG are High-grade tumors, LG are Low-grade tumors and OT are Others.

Test	$p < 0.05$	Test	$p < 0.05$	Test	$p < 0.05$
rSVM vs. NBC	0.002	rSVM vs. NBC	0.004	kNN vs. NBC	0.004
rSVM vs. kNN	0.020	rSVM vs. kNN	1.000	kNN vs. LDC	0.002
rSVM vs. LDC	0.002	rSVM vs. LDC	0.002	kNN vs. ISVM	0.234
rSVM vs. ISVM	0.438	rSVM vs. ISVM	0.049	kNN vs. rSVM	0.012

Table 4. Wilcoxon signed rank test p -values in each experiment. Values lower than 0.05 signal significant differences at the 95% level.

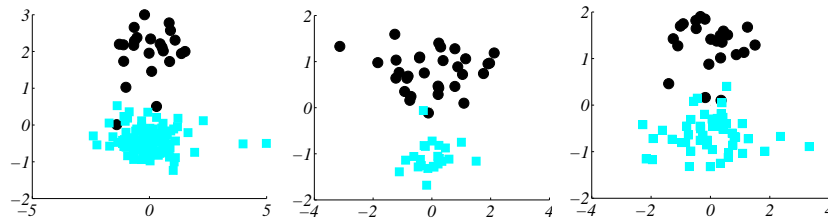


Fig. 2. Projection of the data sets (using the *selected BSS*) onto the first two eigenvectors of the scatter matrices as coordinate system. Left: EXP1, center: EXP2, right: EXP3. Circles represent normal tissue *gliomas*; filled squares represent tumorous samples (left: high-grade *malignant tumors*, center: low-grade *malignant tumors* and right: other tumors).

the separation between the projections of compact groups of data (tumor classes, in this work). This is a recently introduced method that leads onto the definition of low-dimensional projective spaces with good separation between classes, even when the data covariance matrix is singular; further details about this method can be found in [8]. The visualizations of our best results (again those boldfaced in Table 4) show a clear separation between super-classes, being this latter result highly coincident with the reported 10x10cv accuracy values (95-98%) reached by the developed classifiers, despite the differences in dimension –Fig. 2, a fact that adds support to the consistency and clinical value of the results.

5 Conclusions and future work

¹H-MRS is yet to become a standard method for day-to-day clinical diagnosis of brain tumors, despite being a non-invasive technique. In this study, we report experimental results that support the practical advantage of combining robust feature selection and classification techniques in this application field. An attractive accurate classification is obtained with parsimonious and interpretable subsets of spectral frequencies. A dimensionality reduction technique that provides a data projection –while preserving the class discrimination achieved by a

classifier– is also used in our study. The feature filter algorithm has advantage in being simple to implement and requiring no parameter tuning. The reported results also signal a few troublesome data observations that shall deserve clinical attention. Future research will extend the use of the proposed methodology to the analysis of other brain tumor classification problems involving different pathologies and pathology groupings.

Acknowledgements

Authors gratefully acknowledge the former INTERPRET partners (INTERPRET, EU-IST-1999-10310) and, from 1st January 2003, Generalitat de Catalunya (grants CIRIT SGR2001-194, XT 2002-48 and XT 2004-51); data providers: Dr. C. Majós (IDI), Dr. A. Moreno-Torres (CDP), Dr. F.A. Howe and Prof. J. Griffiths (SGUL), Prof. A. Heerschap (RU), Dr. W. Gajewicz (MUL) and Dr. J. Calvar (FLENI); data curators: Dr. A.P. Candiota, Ms. T. Delgado, Ms. J. Martín, Mr. I. Olier, Mr. A. Pérez and Prof. Carles Arús (all from GABRMN-UAB). Authors also acknowledge funding for CICYT TIN2006-08114 and SAF2005-03650 projects; the Mexican CONACyT and Baja California University and thank the anonymous reviewers for their helpful suggestions.

References

- [1] N. Sibtain. The clinical value of proton magnetic resonance spectroscopy in adult brain tumours. *Clinical Radiology*, 62:109–119, 2007.
- [2] A. Tate and et. al. Development of a decision support system for diagnosis and grading of brain tumours using in vivo magnetic resonance single voxel spectra. *NMR in Biomedicine*, 19:411–434, 2006.
- [3] C. Ladroue. *Pattern Recognition Techniques for the Study of Magnetic Resonance Spectra of Brain Tumours*. PhD thesis, St. George’s Hospital Medical School, 2003.
- [4] A. Devos. *Quantification and classification of MRS data and applications to brain tumour recognition*. PhD thesis, Katholieke Univ. Leuven, 2005.
- [5] H. Wang. *Towards a unified framework of relevance*. PhD thesis, Univ. of Ulster, 1996.
- [6] C. Majos and et. al. Brain tumor classification by proton mr spectroscopy: Comparison of diagnostic accuracy at short and long te. *American Journal of Neuroradiology*, 25:1696–1704, December 2004.
- [7] J. Garcia and et. al. On the use of long te and short te sv mr. spectroscopy to improve the automatic brain. tumor diagnosis. Technical report, In <ftp://ftp.esat.kuleuven.ac.be/pub/SISTA/ida/reports/07-55.pdf>, 2007.
- [8] P. Lisboa, I. Ellis, A. Green, F. Ambrogi, and M. Dias. Cluster based visualisation with scatter matrices. *Pattern Recognition Letters*, 29(13):1814–1823, 2008.