Computational Intelligence Techniques Applied to Magnetic Resonance Spectroscopy Data of Human Brain Cancers

Barton, Alan; Valdés, Julio

National Research Council Canada
Conseil national de recherches Canada

Institute for Information Technology
Institut de technologie de l'information

NRC · CNRC

Computational Intelligence Techniques Applied to Magnetic Resonance Spectroscopy Data of Human Brain Cancers *

Barton, A., Valdés, J.
2008

* published in the Conference Proceedings by Springer-Verlag in the Lecture Notes on Artificial Intelligence (LNAI 5306). pp. 485-494. 2008. NRC 50395.

Canada

Computational Intelligence Techniques Applied to Magnetic Resonance Spectroscopy Data of Human Brain Cancers

Alan J. Barton[1] and Julio J. Valdes[1]

National Research Council Canada, Institute for Information Technology,
M50, 1200 Montreal Rd., Ottawa, ON K1A 0R6 ,
alan.barton@nrc-cnrc.gc.ca, julio.valdes@nrc-cnrc.gc.ca,
WWW home page: http://iit-iti.nrc-cnrc.gc.ca

Abstract. Computational intelligence techniques were applied to human brain cancer magnetic resonance spectral data. In particular, two approaches, Rough Sets and a Genetic Programming-based Neural Network were investigated and then confirmed via a systematic Individual Dichotomization algorithm. Good preliminary results were obtained with 100% training and 100% testing accuracy that differentiate normal versus malignant samples.

1 Introduction

Magnetic resonance spectroscopy (MRS) and magnetic resonance imaging (MRI) are two non-invasive and harmless clinical techniques that can provide useful biochemical information about a region of interest in the body. They can be particularly helpful when the organ under investigation is difficult or dangerous to reach (e.g. the brain) where direct inspection and surgery should be avoided as much as possible.

Both techniques are based on magnetic resonance (MR), which is related to the physical property called quantum spin. The MRI technique reveals water concentration levels and is used in routine examinations by clinicians; whereas the MRS technique is not used as frequently as MRI (despite its great potential). MRS information consists of a signal, possibly noisy, composed of peaks whose location and height correspond to different metabolites and their relative concentrations. Reading the most frequent chemical in an MR spectrum is relatively straightforward, but the complete interpretation of a spectrum or the comparison between two spectra usually requires an expert [14]. This reliance on specialized expertise may be one of the reasons why it has been more difficult to introduce MRS into routine medical practice.

An international project, INTERPRET http://azizu.uab.es/INTERPRET, gathered the efforts of 5 centers across Europe with the long term goal of generalizing the use of MRS. During this project, a large database of 1HMR spectra was built in order to develop an automatic MRS-based system to aid clinicians to diagnose brain tumors. Each spectrum in the database was acquired according to a pre-defined protocol and formally validated by clinicians and pathologists [9].

This paper has a preliminary character and will focus on the study of the tumor vs normal differentiation (i.e. $\{G1, G2, G3\} vs \{normal\}$), with 204 and 15 cases respectively. Future studies will cover the distinction between the different types of tumors.

2 Rough Sets

The Rough Set Theory [17], [16] bears on the assumption that in order to define a set, some knowledge about the elements is needed. This is in contrast to the classical approach where a set is uniquely defined by its elements. In the Rough Set Theory, some elements may be indiscernible from the point of view of the available information and it turns out that vagueness and uncertainty are strongly related to indiscernibility.

Reducts and Minimum Reducts Let $O = \{o_1, o_2, \cdots, o_m\}$ be a set of m objects and $A = \{a_1, a_2, \cdots, a_N\}$ a set of N attributes. Let d be a special attribute called the decision attribute. O is consistent if $\forall k, n, \forall i \in [1, N], a_i(o_k) = a_i(o_n) \rightarrow d(o_k) = d(o_n)$. A reduct is a subset $R \subseteq A$ so that $\forall k, n, \forall a \in R, a(o_k) = a(o_n) \rightarrow d(o_k) = d(o_n)$. Minimal reducts are those for which no proper subset is a reduct and are extremely important, as decision rules can be constructed from them [3]. However, the problem of reduct computation is NP-hard, and several heuristics have been proposed [21].

Reduct Computation Genetic algorithms are the most popular representative of the evolutionary computation family of algorithms [5], [1].They have been used as an approach to reduct computation by [20], which proposed several methods based on the notion of a distinction table; which is a $(m^2 - m)/2 \times (N + 1)$ matrix B where columns i are attributes (the last one is the decision attribute d) and the rows are pairs of objects k, n. For every row $i \in [1, N]$ and every $k, n \in [1, m]$ the values of B are constructed as follows: $B[(k, n), i] = 1$ if $a_i(o_k) \neq a_i(o_n)$ and 0 otherwise. For the last row $B[(k, n), N + 1] = 1$ if $d(o_k) = d(o_n)$ and 0 otherwise. In terms of B, a reduct is a subset of columns R with the property [20] $\forall k, n, \exists i \in R, (B[(k, n), i] = 1) \vee (B[(k, n), N + 1] = 1)$. In its simplest representation, a GA with binary chromosomes of length N encodes subsets of attributes (the indices of the chromosomes for which the value is 1). The evolution is guided by a fitness function given by: $F(r) = ((N - L_r)/N) + C_r/K$, where r is a chromosome, L_r is the cardinality of the set of attributes (given by the number of 1s in the chromosome, C_r is the number of object pairs (with different values of the decision attribute) which are discerned by the attributes in R. $K = (m(m - 1))/2$ is the number of object pairs.

3 Genetic Programming

Analytic functions are among the most important building blocks for modeling, and are a classical way of expressing knowledge and have a long history of usage in science. From a data mining perspective, direct discovery of general analytic functions poses enormous challenges because of the (in principle) infinite size of the search space. Within computational intelligence, genetic programming techniques aim at evolving computer programs, which ultimately are functions. Genetic Programming (GP) introduced in [10] and further elaborated in [11], [12] and [13], is an extension of the Genetic Algorithm. The algorithm starts with a set of randomly created computer programs and this initial population goes through a domain-independent breeding process over a series of generations. It employs the Darwinian principle of survival of the fittest with

operations similar to those occurring naturally, like sexual recombination of entities (crossover), occasional mutation, duplication and gene deletion.

3.1 Gene Expression Programming

There are many approaches to GP leading to a plethora of variants (and implementations). A discussion about their relative merits, drawbacks and properties is beyond the scope of this paper. One of these GP techniques is the so-called Gene Expression Programming (GEP) [7], [8]. GEP individuals are nonlinear entities of different sizes and shapes (expression trees) encoded as strings of fixed length. For the interplay of the GEP chromosomes and the expression trees (ET), GEP uses an unambiguous translation system to transfer the language of chromosomes into the language of expression trees and vise versa. The structural organization of GEP chromosomes allows a functional genotype/phenotype relationship, as any modification made in the genome always results in a syntactically correct ET or program. The set of genetic operators applied to GEP chromosomes always produces valid ETs.

3.2 Neural Networks Constructed via Genetic Programming (NN-GP)

A general extension to GEP for vector valued functions was previously introduced [19], whereby GEP individuals consist of multiple chromosomes. Such an extension was the starting point for the construction of a technique to evolve explicit neural networks. Figure 1 shows an example of an explicit neural network consisting of $(n+m+c)$ neurons and (3) layers (other topologies are also possible), where each neuron is a chromosome in an individual. For this example, n neurons in the input layer are determined by the number of variables in the input data set; m neurons in the hidden layer determine the dimension of the non-linear space to be constructed (in this paper, $m = 1$); and c determines the number of classes that need to be discriminated. In general, c neurons in the output layer may be used, but other approaches exist. For example, this paper uses $c = 2$ and uses 1 output neuron in order to construct explicit classifiers. Future studies will investigate these issues more deeply, for example, when determining class discrimination between $c > 2$ classes.

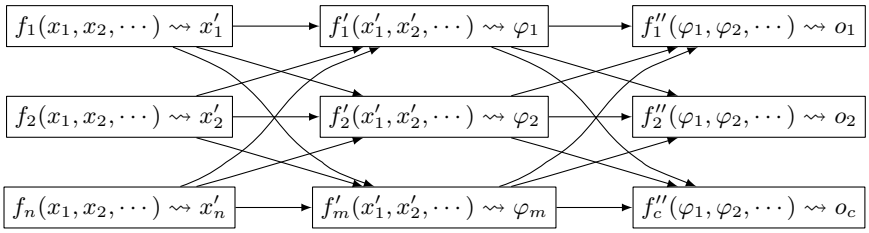

Fig. 1: Neural network representation of one specific topology containing (3) layers and $(n+m+c)$ neurons. Each box is a neuron in the network where all activity occurs (e.g. activation, aggregation, etc). Weights are learned within the neuron by NN-GP.

4 Individual Dichotomization

This is a simple screening algorithm used with the purpose of finding individual attributes that are relevant from the point of view of their ability to differentiate the classes (in a binary problem), when their values are dichotomized. The inputs for the algorithm are: *i)* the values of a given attribute A for all the objects , *ii)* the classes C_1, C_2 associated with each sample (Cancer vs Normal in this case), and *iii)* a probability threshold p_T. The algorithm proceeds as follows: *(1)* construction of the set of distinct values of A (call it Δ). If O is the set of objects and $A(o)$ is the value of the attribute for any object $o \in O$, $\Delta = \{\delta_1, \delta_2, \cdots, \delta_k\}, (k \in [1, card(O)])$ with the following properties: ($\forall \delta_i, \delta_j \in \Delta$, $\delta_i \neq \delta_j$), ($\forall o \in O, \exists \delta \in \Delta$ s.t. $A(o) = \delta$) and ($\forall \delta \in \Delta, \exists o \in O$ s.t. $A(o) = \delta$). *(2)* sort Δ in increasing order. (3) construct the set $\hat{\Delta}$ composed by the mean of all consecutive values of Δ. That is, for every pair $\delta_i, \delta_{i+1} \in \Delta$ compute $(\hat{\delta}_i = (\delta_i + \delta_{i+1})/2$. Clearly, $\hat{\Delta}$ has one element less than Δ. *(4)* use each $(\hat{\delta}_i \in \hat{\Delta}$ as a binary threshold for the values of attribute A. This divides the set of objects into two disjointed classes A_1, A_2 . *(5)* compute the contingency table of A_1, A_2 vs C_1, C_2 *(6)* on the table, compute the conditional probabilities $p_1 = p(C_1/A_1)$, $p_2 = p(C_1/A_2)$and retain $p_{max} = \mathtt{max}(p_1, p_2)$. *(7)* if $p_{max} \geq p_T$ select the attribute as relevant, and discard it otherwise. The process is repeated for all attributes and the resulting set of selected attributes gives an indication on how many of them contain a differentiation power equal or better than the pre-set probability threshold p_T . Specifically, if $p_T = 1$ the algorithm will give a set of attributes such that each of them (*individually*) will perfectly differentiate the classes $\{C_1, C_2\}$.

5 Experimental settings

The height and shape of each resonance in the MR spectrum is determined by several parameters related to the way in which signal produced by the exited proton spin decays by a relaxation process. One of them, called the echo time (TE) is very important. The longer the TE, the more the signal has attenuated before acquisition. Hence, a short echo time spectrum (TE $\leq$ 50ms) has larger peaks than a long echo time spectrum (TE $\geq$ 130ms). A short echo time spectrum also contains more peaks, as resonances with a small relaxation value or complex coupling pattern, like mI (myo-Inositol), Glu (glutamate) and Gln (glutamine) are less pronounced at longer echo times. At short echo time signals, macromolecules are prominent; originating from proteins and membrane components. They have very broad peaks with a large contribution to an underlying and partially unknown baseline [14], [6]. The data used in this study consist of 219 long-echo MR spectra (echo time TE $\geq$ 130ms). The data acquisition protocol and the signal processing procedure is described in [18]. Each spectrum covers a range between [4.23 .. 0.45] parts per million (ppm) along the x-axis, where 200 equally spaced samples were taken. The available validated set represents different types of tumors and normal cases grouped into 4 main classes: G1: astrocytome, oligoastrocytome and oligodendrogliome, G2: glioblastome and metastasis and G3: meningiomes. This paper has a preliminary character and so will focus on the study of the tumor vs normal differentiation (i.e. $\{G1, G2, G3\} vs \{normal\}$), with 204 and 15 cases respectively. In order

Table 1: Experimental settings for the two series of experiments involving NN-GP.

	Series 1 (240)	Series 2 (2250)
GEP Max. Num. Generations	50	same
GEP Population Size	5, 10, 15	10
GEP Num. Elite Individuals	1	same
GEP Inversion Rate	0.1	same
GEP Mutation Rate	0.044	same
GEP IS Transposition Rate	0.1	same
GEP RIS Transposition Rate	0.1	same
GEP One Point Recomb. Rate	0.3	same
GEP Two Point Recomb. Rate	0.3	same
GEP Gene Recombination Rate	0.1	same
GEP Gene Transposition Rate	0.1	same
GEP Num. Genes Per Chromosome	1	same
GEP Gene Headsize	2	same
GEP Gene Linking Function	Addition	same
GEP Num. Real Constants Per Gene	2, 4, 8, 200	1, 2, 3, 4, 5
GEP Constants Limits	$[-100.0, 100.0]$	same
GEP Seeds	5 unique seeds	Series 1 and 45 more
GEP Species RNC Mutation Rate	0.01	same
GEP Species DC Mutation Rate	0.044	same
GEP Species DC Inversion Rate	0.1	same
GEP Species DC IS Transposition Rate	0.1	same
GEP Functions For All Symbol Sets	Addition, Subtraction, Multiplication	
GEP Number of Symbol Sets	Determined by NN topology: 3 (one/layer)	
GEP Number of Chromosomes	Determined by NN topology: 202	
Neural Network (NN) Topology	200 Input Nodes, 1 Hidden, 1 Output	
NN Input Layer Constant Weights	1, 200	1, 100, 200
NN Input Layer Terminal Weights	1	same
NN Hidden Layer Constant Weights	1, 200	1, 100, 200
NN Hidden Layer Terminal Weights	1	same
NN Output Layer Constant Weights	1	same
NN Output Layer Terminal Weights	1	same

to simplify the application of some procedures, in particular genetic programming, the dataset (219 individuals and 200 predictive variables) was linearly re-scaled from its original range $[-44.850571, 56.267685]$ to the $[1, 100]$ range. The purpose was to work with strictly positive values and since the target range is almost the same as the original (99 vs 101.118256), the re-scaling operation is essentially a shifting. The re-scaled data was divided into a training and a test set using random stratified sampling so that class proportions were preserved. The training set contained 80% of the data (175 objects) and the test set the remaining 20% (44 objects). The NN-GP approach was investigated within a series of two experiments (See Table nn-gep-experimental-settings). The first series of 240 attempted to broadly sweep the parameter space; with the second series of 2250 being used to more closely investigate the parameter space around the good solution obtained within the first series.

6 Results

Results from Rough Sets and NN-GP are reported, along with validation via the individual dichotomization approach.

Rough sets Results Rough sets analysis was conducted as follows: *i)* the training set was discretized according to the global method described in [2], [4], *ii)* reducts (see Section 2) were computed using exhaustive and genetic algorithms [2], [20], *iii)* classification rules were generated from the reducts, *iv)* the test set was discretized using the same cuts produced by the discretization of the training set, and finally, *v)* the set was classified using the rules obtained for the training set. Remarkably, both reduct computation algorithms found a single reduct on the training set. Moreover, it was a simple reduct composed of a singleton attribute ($\{V270\}$). Accordingly, both sets of classification rules consist of the common single rule:

$$\text{IF } V_{270} \begin{cases} \geq 69.374496 \Rightarrow C_1 \text{ (i.e. Normal)} \\ < 69.374496 \Rightarrow C_2 \text{ (i.e. Diseased)} \end{cases}$$

which classifies the training set with 100% accuracy. When applied to the test set, it turned out that it also classifies with 100% accuracy. This is very interesting, as it shows that a single attribute (V_{270}) (out of the original 200) is capable of discriminating the spectra from normal cases from those of the malignant class. It corresponds to a concentration of approx. 1.969 ppm.

NN-GP Results Two series of experiments, one of size 240, and the other size 2250 led to 26 explicit neural networks that, when interpreted as classifiers, had 100% training and 100% testing error; a very interesting preliminary result. In order to study the properties of these high performing solutions, the space constructed from the mapping function associated with each of the 26 networks is summarized in Fig.2. It can be seen that all 26 spaces (horizontal lines in Fig.2) perfectly separate the 2 classes and that the 26 solutions can be divided into 4 equivalence classes based on constructed space magnitude: *i)* extra large magnitude $[-150000, 200000]$ (1 solution), *ii)* large magnitude $[-4000, 6000]$ (14 solutions), *iii)* medium magnitude $[-1000, 2000]$ (2 solutions), and *iv)* small magnitude $[-200, 100]$ (9 solutions); with the small magnitude solutions lying closest to the magnitude of the training and testing data. The 26 spaces shown in Fig.2 may also be analyzed in terms of their associated mapping functions. In particular, the 26 equations contain only 50 of the 200 attributes present within the input data; with 43 attributes occurring in exactly one equation, 3 attributes occurring in two equations and 2 attributes occurring in exactly three equations. The two most frequent attributes are V_{270} occurring in exactly eleven equations and V_{271} occurring most frequently, and in sixteen equations. In addition, it is observed that V_{271} was more frequently used than V_{270} within good solution networks and that it was not discovered by the Rough Sets approaches that were investigated, which only discovered attribute V_{270}. Of the 26 good solution results (100% training and 100% testing accuracy), 3 are now highlighted that show use of the 2 most frequent variables (as both independent and joint usage) in the

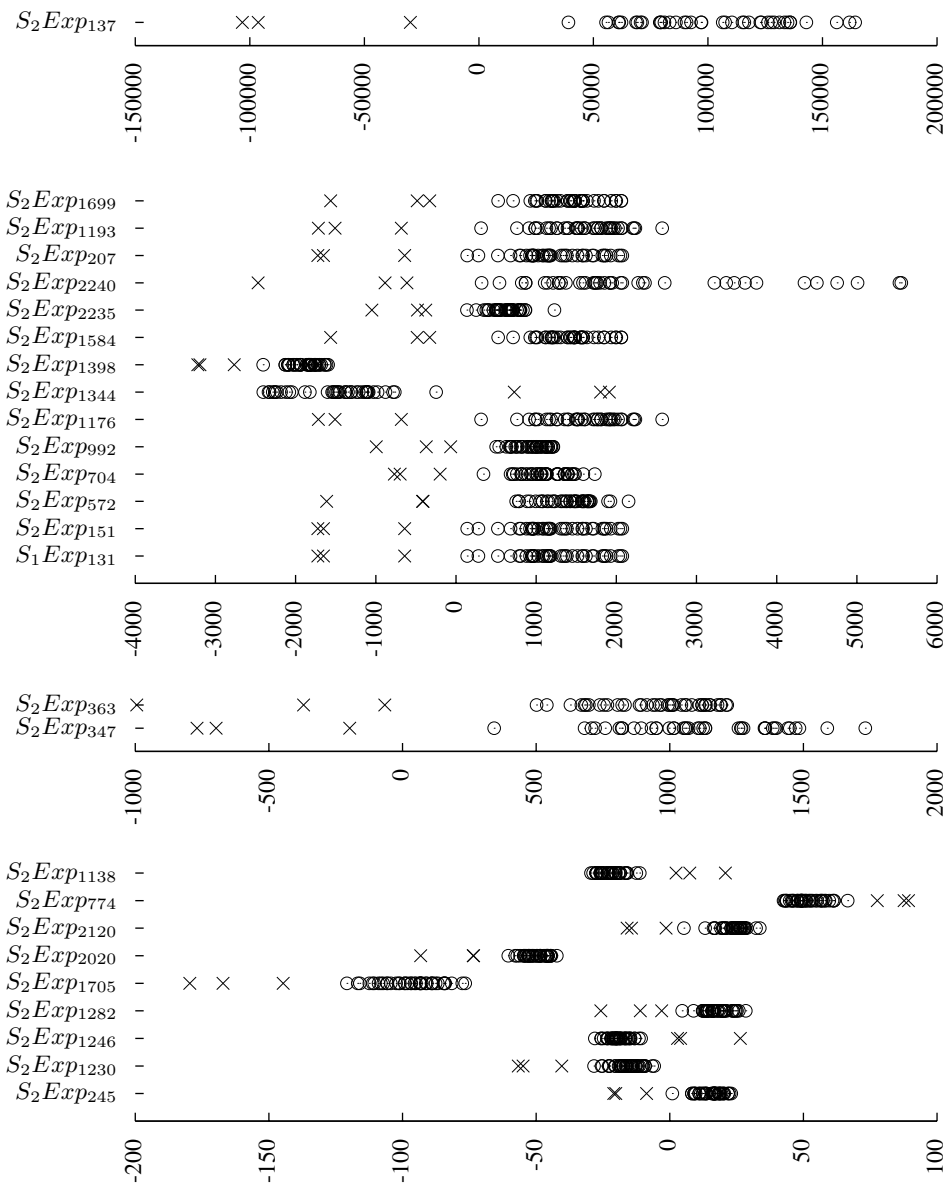

Fig. 2: Best 26 mapped 1D spaces (varying orders of magnitude) from nonlinear discriminant analysis of neural network (NN-GP) solutions having 200 input variables. All 26 spaces have an associated classifier (not shown) with 0.00 training and validation error. X = healthy class. O = diseased patient samples.

mapping and classifier results. It can be observed from Fig.2, that the mapping results may be converted into the good classifiers through rescaling (and possibly reflection about a point) of the constructed spaces. An example NDA and classifier result involving V_{270} was discovered in experiment S_2 Exp_{207} and resulted in the construction of a 200D to 1D mapping function $\varphi_1(\cdot) = 66.86 - V_{270}$ and the following classifier (with 100% training and testing accuracy):

$$\text{IF } (66.86 - V_{270})^3 \begin{cases} < 0.5 \Rightarrow C_1 \text{ (i.e. Normal)} \\ = 0.5 \Rightarrow \text{Undecidable} \\ > 0.5 \Rightarrow C_2 \text{ (i.e. Diseased)} \end{cases}$$

An example NDA and classifier result involving V_{271} was discovered in experiment S_2 Exp_{347} and resulted in the construction of a 200D to 1D mapping function $\varphi_1(\cdot) = V_{271} - V_{234} - 27.69$ and the following classifier (with 100% train/test accuracy):

$$\text{IF } -28.75(V_{271} - V_{234} - 27.69) - 50.78 \begin{cases} < 0.5 \Rightarrow C_1 \text{ (i.e. Normal)} \\ = 0.5 \Rightarrow \text{Undecidable} \\ > 0.5 \Rightarrow C_2 \text{ (i.e. Diseased)} \end{cases}$$

An example NDA and classifier result involving both V_{270} and V_{271} was discovered in experiment S_2 Exp_{1699} and resulted in the construction of a 200D to 1D mapping function $\varphi_1(\cdot) = V_{331} - V_{295} - V_{271} - V_{270} - V_{195} + V_{179}$ and the following classifier (with 100% train/test accuracy):

$$\text{IF } V_{331} - V_{295} - V_{271} - V_{270} - V_{195} + V_{179} + 137.40 \begin{cases} < 0.5 \Rightarrow C_1 \text{ (i.e. Normal)} \\ = 0.5 \Rightarrow \text{Undecidable} \\ > 0.5 \Rightarrow C_2 \text{ (i.e. Diseased)} \end{cases}$$

Individual Dichotomization Results A systematic exploration of each single attribute in the training set was made with the individual dichotomization algorithm (see Section 4). The probability threshold was set to 1 ($p_T = 1$) in order to find the highest conditional probabilities of the classes given the attribute dichotomization. It was found that $P(class = normal/(V_{270} \geq 69.375)) = 1$ and that $P(class = normal/(V_{271} \geq 68.257)) = 1$. When these probabilities are computed on the test set using the same conditionals, the result was the same, showing that both V_{270} and V_{271} (spectral peaks at 1.969 and 1.95 ppm respectively), can individually discriminate the normal from the malignant cases, thus confirming the results found with rough sets and especially with the NN-GP network. Rough sets found V_{270} but not V_{271}, whereas NN-GP detected V_{270} and V_{271} as the two most important attributes, confirmed by individual dichotomization.

7 Conclusions

Computational intelligence techniques were applied to brain cancer data. Good preliminary results were obtained with 100% training and testing accuracy that differentiate normal versus malignant samples. Two out of 200 attributes were found to be most

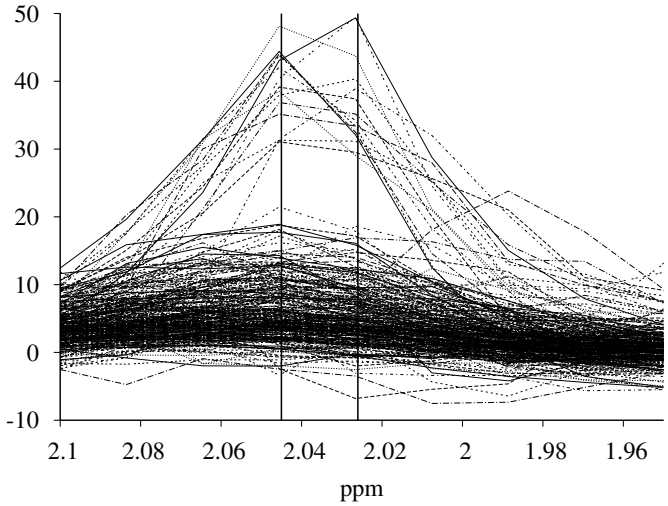

Fig. 3: All 285 MR spectra. 2 out of 200 variables may be used (independently or jointly) for discrimination. Larger values ($[31.075169..48.118134]$ for V_{270} and $[29.067427..49.497776]$ for V_{271}) are normal samples.

important. Rough Sets found one; whereas the NN-GP experiments found both. The results were confirmed via a systematic algorithm, which disregards attribute interactions; something that cannot (in general) be assumed *a priori*. The NN-GP approach, which, although more complex, did not miss a relevant attribute as did the Rough Sets approach. Future studies will focus on differentiation of the different cancers.

8 Acknowledgments

This paper is a cooperation between the SOCO Group of the Polytechnic University of Catalonia (UPC, Spain) and the Knowledge Discovery Group (National Research Council Canada). The authors thank Alfredo Vellido (UPC, Spain) for his support of the research presented in this paper. Authors gratefully acknowledge the former INTERPRET (EU-IST-1999-10310) European project partners. Data providers: Dr. C. Majós (IDI), Dr. À. 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 and Mr. A. Pérez (all from GABRMN-UAB). Prof. Carles Arús, GABRMN group leader.

References

1. T. Bäck, D.B.Fogel, and Z. Michalewicz. *Evolutionary Computation 1 and 2*. Institute of Physics Publishing, Bristol and Philadelphia, 2000.
2. J. G. Bazan, H. S. Nguyen, S. H. Nguyen, P. Synak, and J. Wróblewski. *Rough set algorithms in classification problem.* In: L. Polkowski, S. Tsumoto and T.Y. Lin (eds.): Rough Set Methods and Applications. Physica-Verlag, Heidelberg, New York 2000, pp. 49–88.

3. J. G. Bazan, A. Skowron, and P. Synak. *Dynamic reducts as a tool for extracting laws from decision tables*. In Proc. of the Symp. on Methodologies for Intelligent Systems, volume 869 of Lecture Notes in Artificial Intelligence LNAI, pages 346–355. Springer-Verlag, 1994.
4. J. G. Bazan, M. S. Szczuka, and J. Wróblewski. *A new version of rough set exploration system.* In Third. Int. Conf. on Rough Sets and Current Trends in Computing (RSCTC), v. 2475 of LNCS, pages 397–404. Springer-Verlag, 2002.
5. T. Bäck, D. B. Fogel, and Z. Michalewicz. *Handbook of Evolutionary Computation*. Institute of Physics Publishing and Oxford Univ. Press, New York, Oxford, 1997.
6. A. Devos. *Quantification and classification of magnetic resonance spectroscopy data and applications to brain tumour recognition*. Technical Report U.D.C. 616.831-073, Katholieke Universiteit Leuven. Dept. of Electronics, 2005.
7. C. Ferreira. *Gene expression programming: A new adaptive algorithm for problem solving*. Journal of Complex Systems, 13, 2001.
8. C. Ferreira. *Gene Expression Programming: Mathematical Modeling by an Artificial Intelligence*. Springer Verlag, 2006.
9. M. Juliá-Sapé, D. Acosta, M. Mier, C. Arús, and D. Watson. *A multi-centre, web-accessible and quality control-checked database of in vivo mr spectra of brain tumour patients*. Magn Reson Mater Phy, 19, 2006.
10. J. Koza. *Hierarchical genetic algorithms operating on populations of computer programs*. In Proc. of the 11th International Joint Conf. on Artificial Intelligence. San Mateo, CA, 1989.
11. J. Koza. *Genetic programming: On the programming of computers by means of natural selection.* MIT Press, 1992.
12. J. Koza. *Genetic programming II: Automatic discovery of reusable programs*. MIT Press, 1994.
13. J. R. Koza, F. H. Bennett III, D. Andre and M. A. Keane. *Genetic Programming III: Darwinian Invention and Problem Solving*. Morgan Kaufmann, 1999.
14. C. L. C. Ladroue. *Pattern Recognition Techniques for the Study of Magnetic Resonance Spectra of Brain Tumours*. PhD thesis, St George's Hospital Medical School. London, 2004.
15. A. Øhrn. *Discernibility and rough sets in medicine: Tools and applications*. Technical Report NTNU report 1999:133, Norwegian University of Science and Technology, Department of Computer and Information Science, 1999.
16. Z. Pawlak. *Rough Sets, Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers, New York, 1991.
17. Z. Pawlaw. *Rough Sets*. International Journal of Information and Computer Sciences, 11:341–356, 1982.
18. A. R. Tate, J. Underwood, D. M. Acosta, M. Juliá-Sapé, C. Majo, A. Moreno-Torres, F. A. Howe, M. van der Graaf, V. Lefournier, M. M. Murphy, A. Loosemore, C. Ladroue, P. Wesseling, J. L. Bosson, M. E. C. n. as, A. W. Simonetti, W. Gajewicz, J. Calvar, A. Capdevila, P. R. Wilkins, B. A. Bell, C. Rémy, A. Heerschap, D. Watson, J. R. Griffiths1, and C. Aruś. *Development of a decision support system for diagnosis and grading of brain tumours using in vivo magnetic resonance single voxel spectra*. NMR Biomed, 19, 2006.
19. J.J.Valdés, R. Orchard, and A.J. Barton. *Exploring Medical Data using Visual Spaces with Genetic Programming and Implicit Functional Mappings*. GECCO Workshop on Medical Applications of Genetic and Evolutionary Computation. The Genetic and Evolutionary Computation Conference. London, England. 2007.
20. J. Wróblewski. *Finding minimal reducts using genetic algorithm.* In Proc. of the Second Annual Join Conference on Information Sciences, 1995.
21. J. Wróblewski. *Ensembles of classifiers based on approximate reducts*. Fundamenta Informaticae, 47, 2001.