Neurocomputing 72 (2009) 3085–3097

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Outlier exploration and diagnostic classification of a multi-centre ^{1}H-MRS brain tumour database

Alfredo Vellido [a,*], Enrique Romero [a], Félix F. González-Navarro [a], Lluís A. Belanche-Muñoz [a], Margarida Julià-Sapé [b,c], Carles Arús [c,b]

[a] Dept. de Llenguatges i Sistemes Informàtics, Universitat Politècnica de Catalunya, C./ Jordi Girona, 1-3, 08034 Barcelona, Spain
[b] Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Cerdanyola del Vallès, Spain
[c] Grup d'Aplicacions Biomèdiques de la RMN (GABRMN), Departament de Bioquímica i Biología Molecular (BBM), Unitat de Biociències Universitat Autònoma de Barcelona (UAB), Cerdanyola del Vallès, Spain

ARTICLE INFO

Article history:
Received 30 June 2008
Received in revised form
10 March 2009
Accepted 22 March 2009
Communicated by T. Heskes
Available online 9 April 2009

Keywords:
Proton magnetic resonance spectroscopy
Brain tumours
Outlier detection
Nonlinear dimensionality reduction
Feature selection
Medical decision support systems

ABSTRACT

Non-invasive techniques such as magnetic resonance spectroscopy (MRS) are often required for assisting the diagnosis of tumours. Radiologists are not always accustomed to make sense of the biochemical information provided by MRS and they may benefit from computer-based support in their decision making. The high dimensionality of the MR spectra obscures atypical aspects of the data that may jeopardize their classification. In this study, we describe a method to overcome this problem that combines nonlinear dimensionality reduction, outlier detection, and expert opinion. MR spectra subsequently undergo a feature selection process followed by classification. The impact of outlier removal on classification performance is assessed.

1. Introduction

Decision making in oncology is a sensitive matter, and even more so in the specific area of brain tumour oncologic diagnosis, for which the direct and indirect costs—both human and financial—of misdiagnosis are very high. In this area, in which most diagnostic techniques must be non-invasive, clinicians should benefit from the use of an at least partially automated computer-based medical decision support system (DSS).

AIDTumour (artificial intelligence decision tools for tumour diagnosis [1]) is a research project for the design and implementation of a medical DSS to assist experts in the diagnosis of human brain tumours on the basis of biological signal data obtained by magnetic resonance spectroscopy (MRS). This is a technique that can shed light on cases that remain ambiguous after clinical investigation. The MRS data used in AIDTumour and analysed in this paper belong to a complex multi-centre set containing cases of several brain tumour pathologies [16]. These data have undergone a rigorous pre-processing quality control that validates them from the viewpoint of the radiologists. Nevertheless, and for their use in an automated computer-based DSS, the various origins of these spectra and the complexity of their pre-processing make further data exploration advisable.

It might be problematic to include some of the spectra in an automated DSS without further ado for three different reasons. Firstly, some may contain measurement or acquisition artefacts that, even if not completely precluding diagnosis by visual inspection, might induce errors in computer-based diagnosis: these are what we call here *artefact-related outliers*. Secondly, atypical cases that do not contain artefacts but are nevertheless unrepresentative of the main distributions of the whole dataset: herein, these will be referred to as *distinct outliers* [33]. Thirdly, some cases with a clear biopsy-based diagnosis (tumour type attribution) may yield spectra that are quantitatively similar to those of other tumour types, misleading a computer-based classification system. Even if representative of the data as a whole, they are still unrepresentative of their own tumour type: these we will call *class outliers*. Note that these three outlier typologies are not always mutually exclusive.

Machine learning (ML) and related methods can play a useful role [35] in dealing with the uncertainty introduced by the presence of outliers in a diagnostic setting. Here, we show the effectiveness of a method to identify and characterize potentially conflicting MRS data that combines techniques of nonlinear

* Corresponding author.
E-mail address: avellido@lsi.upc.edu (A. Vellido).

0925-2312/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2009.03.010

dimensionality reduction (DR), exploratory visualization, and outlier detection, with expert knowledge. The introduction of the latter is paramount, as it will help to skim those cases truly conflictive out of those shortlisted by blind quantitative criteria. Dimensionality reduction is not trivial in this setting, as the available MRS data are scarce and high dimensional. Sammon's mapping [29] is used to this end. Generative topographic mapping (GTM [3]), a manifold learning model, is used to quantify the atypicality of spectra [33,34].

Overall, the aforementioned method is conceived as a preliminary step to data classification in the DSS, in which specific cases are tagged with information regarding their possible atypicality and its characteristics. The fact that the MRS data analysed in this study are scarce and of high dimensionality makes their computer-based automated classification a difficult undertaking. Most importantly, this high dimensionality also precludes the straightforward interpretation of the obtained results, limiting their usability in a practical medical setting. Consequently, dimensionality reduction, in the form of either feature selection or feature extraction, would help to reduce the complexity of the problem at hand. Feature extraction, though, may not comply with the interpretability requirement. The expert radiologists who are meant to be assisted by the medical DSS are not usually trained to make sense of new features extracted from the MRS frequencies. Instead, they often have knowledge of specific MRS frequencies related to metabolites of known significance for tumour type discrimination. Note also that one goal of exploratory studies of this kind is to understand where the variables selected by the model fit in relation to prior knowledge from the medical domain [23]. This may limit the practical applicability of methods such as PCA or ICA as used, for instance, in [19,13,24,37] for assisting brain tumour diagnosis. As an example, in analysing these type of data, ICA will often yield components that "would correspond with identifying the independent degrees of freedom in MRS, not with individual metabolites, but with characteristic tissue generators" [13].

An entropic filtering algorithm (EFA) is used in this study for feature selection as a fast method to generate a relevant subset of MR spectral frequencies. Bootstrap resampling techniques are used to obtain mean performance estimates and their variability. The main goal is obtaining simple models (in terms of low numbers of hopefully interpretable MR spectral frequencies) that generalize well. Outliers might still unduly bias the automated classification process in the DSS, even if for different reasons. We hypothesized that, by removing the cases labelled as outliers, classification accuracy would improve and feature selection would experiment significant variations. The experimental results reported in this paper provide partial support for the first hypothesis but not for the second.

The remaining of the paper is structured as follows. First, the ^{1}H-MRS dataset available for experimentation is briefly described. This is followed, in Section 3, by a description of the different analytical methods. Experimental results are presented in Section 4. The paper closes with a section summarizing our conclusions.

2. ^{1}H-MRS brain tumour data

The echo time is an influential parameter in ^{1}H-MRS data acquisition. In short-echo time (SET) spectra (typically acquired at 20–40 ms) some metabolites are better resolved (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 [26]. The use of a long-echo time (LET) yields less clearly resolved metabolites but also less baseline distortion, resulting in a more readable spectrum. There are a few studies comparing the classification potential of these two types of spectra (see, e.g. [26]). In this study, we focus on LET data.

The analysed data correspond to 195 LET single voxel ^{1}H-MR spectra acquired in vivo from brain tumour patients. They include 55 meningiomas (*mm*), 78 glioblastomas (*gl*), 31 metastases (*me*), 20 astrocytomas grade II (*a2*), 6 oligoastrocytomas grade II (*oa*), and 5 oligodendrogliomas grade II (*od*). Following a common procedure [26,24], the clinically-relevant regions of the spectra were sampled to obtain 195 frequency intensity values (measured in parts per million (ppm), an adimensional unit of relative frequency position in the data vector), from 4.25 ppm down to 0.56 ppm. These frequencies become data attributes in the reported experiments and, as a result, the analysed data consist of 195 cases and 195 attributes.

These data are extracted from a database resulting from the *international network for pattern recognition of tumours using magnetic resonance* (INTERPRET) European research project [16]. The criteria for the selection of cases to be included in the original complete database (in which there are more tumour types than the ones analysed in this study as well as cases corresponding to normal tissue and abscesses) were: (a) that the case had a single voxel SET, 1.5 T spectrum acquired from a nodular region of the tumour; (b) that the voxel was located in the same region as where subsequent biopsy was obtained; (c) that the short-echo spectrum had not been discarded because of acquisition artefacts or other reasons and (d) that a histopathological diagnosis was agreed among a committee of neuropathologists. In those cases in which the spectra were obtained from normal volunteers without the pathology, or corresponded to abscesses or clinically proven metastases, biopsy was not required. For further details on data acquisition and processing, and on database characteristics, see, for instance, [15,16].

Class labelling was performed according to the World Health Organization (WHO) system for diagnosing brain tumours by histopathological analysis of a biopsy sample. For the analyses reported in this study, a subset of spectra from the database were bundled into three groups, namely: G1: *low-grade gliomas* (*a2*, *oa* and *od*); G2: *high-grade malignant tumours* (*me* and *gl*); and G3: *meningiomas* (*mm*). This type of grouping is justified [31] by the well-known difficulty in distinguishing between metastases and glioblastomas, due to their similar spectral pattern produced by the highly necrotic nature of these tumours.

3. Methods

3.1. Outlier characterization

3.1.1. MRS data dimensionality reduction and visualization through Sammon's mapping

There are several decisions involved in the choice of a dimensionality reduction method. To name just a few [22]: hard vs. soft DR; generative vs. non-generative methods; implicit vs. explicit mappings; or linear vs. nonlinear DR. For this study, a nonlinear DR method was preferred in principle (instead of a linear alternative such as PCA or classical Multi-Dimensional Scaling, for instance), as there existed no *a priori* reason to assume only linear dependencies. Given that DR in this study does not aim at providing generalization, an explicit mapping procedure was also preferred. A typical *desiderata* for the visual representation of data and knowledge can be formulated in terms of maximizing structure preservation and, therefore, a method with "in-built" preservation of inter-point distances was also preferred. The nonlinear Sammon's mapping method [29] fits all those