Neurocomputing 131 (2014) 15–22

Contents lists available at ScienceDirect

Neurocomputing

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

Classification in high-dimensional spectral data: Accuracy vs. interpretability vs. model size

Andreas Backhaus *, Udo Seiffert

Fraunhofer Institute for Factory Operation and Automation IFF, Sandtorstr. 22, 39106 Magdeburg, Germany

ARTICLE INFO

Article history:
Received 30 October 2012
Received in revised form
4 August 2013
Accepted 23 September 2013
Available online 23 November 2013

Keywords:
Learning Vector Quantization (LVQ)
Radial Basis Function (RBF) networks
Support Vector Machines (SVM)
Supervised Neural Gas (SNG)
Hyperspectral imaging
Raman spectroscopy

ABSTRACT

Against the background of classification in data mining tasks typically various aspects of *accuracy*, and often also of *model size* are considered so far. The aspect of *interpretability* is just beginning to gain general attention. This paper evaluates all three of these aspects within the context of several computational intelligence based paradigms for high-dimensional spectral classification of data acquired by hyperspectral imaging and Raman spectroscopy. It is focused on state-of-the-art paradigms of a number of different concepts, such as prototype based, kernel based, and support vector based approaches. Since the application point of view is emphasized, three real-world datasets are the basis of the presented study.

1. Introduction

Hyperspectral imaging as recent extension to traditional non-invasive spectroscopic analysis techniques (e.g. NIR spectroscopy) has paved the way to obtain the biochemical constitution of inspected solid materials with the additional advantage of a two-dimensional spatial resolution [1,2]. For the examination of liquid samples, Raman spectroscopy has been shown to be a viable tool to gather information without extensive sample preparation [3].

Often the direct relationship between spectral information and biochemical target values or material category is not known in a closed mathematical form. In this case a machine learning approach is used to acquire an analysis model from reference data, a paradigm often referred to as 'soft-sensor'. Sensor data analysis becomes a pattern recognition task. Regarding pattern recognition and data mining in the acquired spectral data, computational intelligence based methods are still providing powerful tools to cope with this kind of high-dimensional and complex data (see Fig. 1).

From the computational intelligence point of view the recent developments in hyperspectral camera technology with increasingly high resolution in both the spectral and spatial domain have led to high-dimensional input spaces and a large number of training vectors. Both aspects even more motivate and demand computational intelligence based algorithms.

Besides unsupervised visualization and clustering typically used to get a (first) graphical representation of the acquired spectral data, classification and multivariate regression is often required by the underlying application. Here, corresponding labeled data is necessary. Since suitable wet lab analyses to provide continuously valued reference data are typically expensive, frequently categorical labels are provided. This leads to a classification task. Industrial applications in product quality control and sorting also demand on-line classification at a low systems cost.

Therefore this classification task has in general three, sometimes conflicting, objectives to address. The first objective is a classification model of high accuracy. The second objective is an as small as possible classification model for quick calculation. A third objective is the restriction to necessary information/features of the examined objects for the classification task at hand. In spectral data processing this means the restriction to necessary spectral bands. This not only speeds up calculation but also leads to less expensive spectral sensor systems. Therefore classification models need to offer a certain degree of interpretability. Relevance profiles for example can indicate the importance of the used input variables, in this case the acquired spectral bands. Additionally, classification models should require small or no expert interference in order to tune model parameters which could lead to biased, non-optimal decisions by the user.

Keeping these requirements in mind, a number of computational intelligence paradigms appear to be particularly suitable. Among them are prototype-based neural networks, such as the Generalized Learning Vector Quantization (GLVQ) family [4],

* Corresponding author.

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.09.048

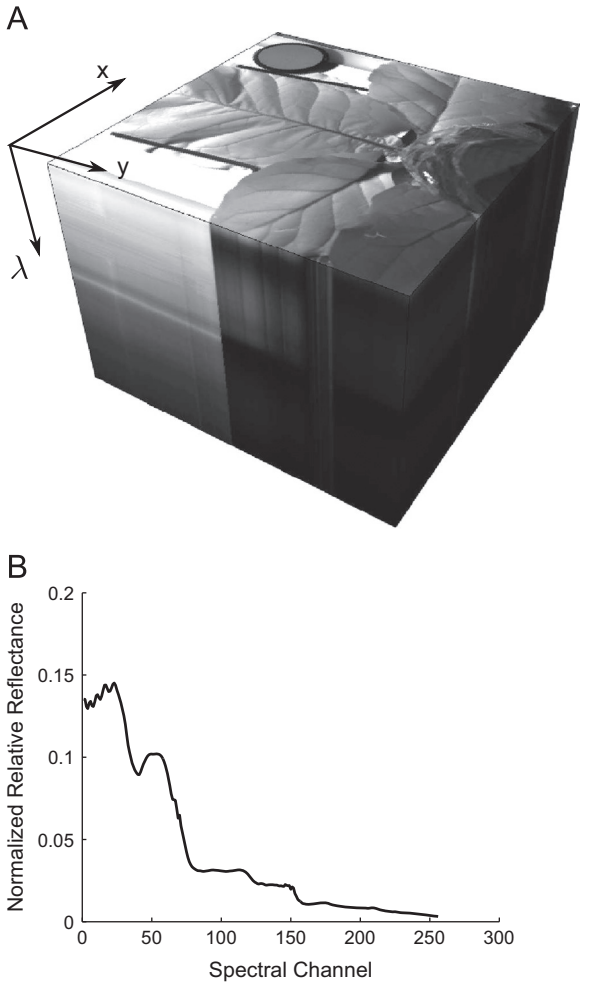

Fig. 1. Example of hyperspectral imaging cube: (A) Reflectance properties of objects can be recorded with spatial resolution; (B) each spatial pixel contains a spectral signature that contains information about the chemical composition of the observed material.

Supervised Neural Gas [5], and RBF (Radial Basis Function) networks [6] as well as Multilayer Perceptron (MLP) networks [7,8] and Support Vector Machines (SVM) [9]. These five approaches span the scope of the presented paper. The qualification of these different approaches regarding classification data from the hyperspectral imaging domain as well from Raman spectra data in terms of several theoretical considerations as well as practical aspects is evaluated. In order to derive practically relevant information from this study, several real-world datasets are used.

2. Material and methods

2.1. Green coffee spectra

Quality control of coffee products, from basic green coffee to the finished roasted coffee, by hyperspectral imaging offers the means for a non-invasive, on-line and automated screening method to control large product quantities [10,11]. For example green coffee has to be inspected for the Robusta or Arabica varieties since Arabica based coffee is sold at a different price than Robusta based coffee. The spatial resolution of hyperspectral imaging makes it the ideal tool for loose material sorting especially in the case where information from color, shape, or texture is not sufficient for differentiation.

For the hyperspectral image acquisition coffee beans of four different green coffee varieties, two varieties of Arabica and two varieties of Robusta and a standard optical PTFE (polytetrafluoroethylene) calibration pad were positioned on a translation table, one class at a time. Hyperspectral images were recorded using a HySpex SWIR-320m-e line camera (Norsk Elektro Optikk A/S). Spectra are from the short-wave infra-red range (SWIR) between 970 nm and 2500 nm at 6 nm resolution yielding a 256-dimensional spectral vector per pixel. The camera line has a spatial resolution of 320 px and can be recorded with a maximum frame rate of 100 fps. Radiometric calibration was performed using the vendors software package. Spectra are normalized to a vector length of one.

Coffee beans were segmented from background via Neural Gas (NG) clustering [12]. We used five prototype vectors, each representing a cluster with a receptive field determined by the smallest Euclidean distance from data sample $\mathbf{v}$ to the prototypes. The prototype spectra $\mathbf{w}$ are randomly initialized and updated by minimizing the following energy function [13]:

$$E(\mathbf{V},\mathbf{W}) = \frac{1}{C(\gamma,K_c)} \sum_{\mathbf{v}\in\mathbf{V}} \sum_{\mathbf{w}\in\mathbf{W}} h_\gamma(r,\mathbf{v},\mathbf{W}) d(\mathbf{v},\mathbf{w}), \tag{1}$$

with d being the Euclidean distance and where

$$h_\gamma(r,\mathbf{v},\mathbf{W}_c) = \exp\left(-\frac{k_r(\mathbf{v},\mathbf{W}_c)}{\gamma}\right) \tag{2}$$

denotes the degree of neighborhood cooperation. The function $k_r(\mathbf{v},\mathbf{W})$ gives the number of prototypes that have equal or smaller distance to the input spectra than prototype $\mathbf{w}_r$, and $C(\gamma,K_{c_v})$ is a normalization constant depending on the neighborhood range γ and cardinality K of $\mathbf{W}$. Minimization was achieved with the freely available Matlab 'minFunc' optimization toolbox[1] using the nonlinear conjugate gradient approach with automatic step size. The cluster representing coffee was chosen through manual inspection and all spectra in this cluster formed the respective coffee class. Fig. 2 depicts the clustering/segmentation process.

The dataset contains the four green coffee varieties forming a 4-class problem with 2000 spectra per class. Fig. 3A shows average spectra for the four green coffee classes.

2.2. Scotch whisky spectra

The automated, on-line assessment of high-priced liquor products is essential for the standardization and quality monitoring in liquor production as well as potential fraud detection. An ideal sensor should be compact for mobile applications and requires no special sample preparation while measuring sample quality instantaneously. In [3] an optofluidic chip was presented that uses Raman spectroscopy to acquire a Raman spectrum of the fluid sample.

The procedure to acquire the Raman spectra from Whisky samples is shown in detail in [3]. In Raman spectroscopy a sample is illuminated with a laser beam. The laser light interacts with molecular vibrations, phonons or other excitations in the system, resulting in the energy of the laser photons being shifted up or down. The shift in energy gives information about the vibrational modes in the system. Raman spectroscopy is commonly used in chemistry, since vibrational information is specific to the chemical bonds and symmetry of molecules. Therefore, it provides a fingerprint by which molecules can be identified.

Whisky samples of 20 µl were directly loaded into the microfluidic chip without any preparation. After Raman acquisition, any remaining liquid at the sample inlet was wiped off and 40 µl of

[1] http://www.di.ens.fr/~mschmidt/Software/minFunc.html