Chemometrics and Intelligent Laboratory Systems 159 (2016) 89–96

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Rigorous and compliant approaches to one-class classification

Oxana Ye. Rodionova[a], Paolo Oliveri[b,*], Alexey L. Pomerantsev[a]

[a] N.N. Semenov Institute of Chemical Physics RAS, Kosygin 4, 119991 Moscow, Russia
[b] Department of Pharmacy, University of Genoa, Viale Cembrano, 4, I-16148 Genoa, Italy

ARTICLE INFO

Keywords:
One-class classification
Class modelling
Discriminant analysis
Sensitivity
Specificity

ABSTRACT

A wide number of real problems requiring qualitative answers should be addressed by one-class classification (OCC), as in the case of authentication studies, verification of particular claims and quality control. The key feature of OCC is that models are developed using only samples from the target class, so that a representative sampling is not strictly required for non-target classes. On the contrary, in the discriminant analysis (DA) approach, all of the classes considered (at least two) have a non-negligible influence in the definition of the delimiter. It follows that faults in the definition of the classes involved and in representative sampling for each of them may determine a bias in the classification rules. A key aspect in one-class classification concerns model optimisation. When the optimal modelling conditions are searched by considering parameters such as type II error or specificity ('compliant' approach), information from the non-target class is being used and may therefore determine a bias in the model. In order to build pure class models ('rigorous' approach), only information from the target class should be regarded: in other words, optimisation should be performed only considering type I error, or sensitivity. In the present study, 'compliant' and 'rigorous' approaches are critically compared on real case studies, by applying two novel modelling techniques: partial least squares density modelling (PLS-DM) and data driven soft independent modelling of class analogy (DD-SIMCA).

1. Introduction

One-class classification (OCC) [1,2] consists in making a description of a target class of objects and in detecting whether a new object resembles this class or not. The term class modelling is often used for denoting OCC methods [3]. In some sense, this approach is opposite to the discrimination problem that is to allocate a new object to one of distinct and exhaustive classes [4]. The critical difference between OCC and discriminant analysis (DA) is that the OCC model is developed using target class samples only.

The work of Harold Hotelling on multivariate quality control (1947) can be considered as the first example of multivariate one-class classification in chemistry [5]. The unequal class models (UNEQ) method was developed by Derde and Massart (1986) as an evolution of these concepts [6]. In fact, such a method – closely related to quadratic discriminant analysis (QDA) – is based on the hypothesis of a multivariate normal distribution in the class to be modelled and defines the width of the class space based on Hotelling's T^2 statistics, at a selected confidence level.

The first method specifically developed for one-class classification in chemometrics was soft independent modelling of class analogy (SIMCA), by Svante Wold [7,8]. This method performs PCA on the samples of the class to be modelled – the SIMCA model being defined as the range of sample scores on the significant PCs. A critical distance, at a given confidence level, is obtained by application of the Fisher F statistics to residuals of each training sample to the model, and is used to define the boundaries of the SIMCA class space around the model.

OCC modelling is a rather new strategy in comparison with DA. The classical OCC version does not utilise any information about non-target (extraneous) classes, even when the data regarding such extraneous classes is available. We call such an approach a 'rigorous' one. Contributing to the OCC technique elaboration, we consider the outcomes that can be yielded in case the rigorous concept is violated. The most common violation – which we call a 'compliant' approach – makes use of some relevant non-target information that can influence the results of the OCC modelling.

The main objective of the present study is the comparison between the outcomes of 'rigorous' and 'compliant' approaches. For this purposes, two different OCC methods, namely, partial least squares density modelling (PLS-DM) [9], and data-driven soft independent modelling of class analogy (DD-SIMCA) [10] are employed. Method descriptions are presented in Sections 3.1 and 3.2. An additional goal is to compare these techniques using two real world examples.

* Corresponding author.
E-mail address: oliveri@difar.unige.it (P. Oliveri).

http://dx.doi.org/10.1016/j.chemolab.2016.10.002
Received 1 September 2016; Received in revised form 2 October 2016; Accepted 4 October 2016
Available online 14 October 2016
0169-7439/

2. Theory

2.1. Figures of merit

Performances of one-class classifiers are usually reported using two parameters: sensitivity and specificity. Sensitivity is the fraction of samples of the target class which are correctly recognised as consistent with the model. It can also be defined as the rate of true positives and, therefore, it is complementary to type I error (*i.e.*, the false negative rate). Specificity is the fraction of samples extraneous to the target class which are correctly recognised as inconsistent with the model, corresponding to the rate of true negatives. This parameter is therefore complementary to type II error (*i.e.*, the false positive rate). Efficiency of one-class classifiers is usually defined as the geometric mean of sensitivity and specificity [11].

When sensitivity and specificity are considered, it is very important to realise on which sample subset each parameter was calculated. First of all, let us consider the type I error, α. At the stage of model building, some OCC methods enable to set a prior value of α and to use it for establishing the corresponding threshold. After that, it is possible to calculate sensitivity for the training set, referred to as SENS_T in the following sections. This value is a classification analogue of the root mean square error of calibration (RMSEC) for calibration problems. It is important to verify that SENS_T is in agreement with the a-priori α value. Varying the α value, it is possible to control the risk of wrong rejections of target objects. The value of sensitivity that characterises the quality of predictions should be calculated as the fraction of samples from the target test set which are correctly recognised (SENS_P). In this case, SENS_P is the classification analogue of the root mean square error of prediction (RMSEP). We can also calculate sensitivity using the cross-validation approach. The corresponding value is referred to as SENS_V. It is worth mentioning that calculation of sensitivity as the fraction of all samples form the target class (training plus test samples) can provide misleading results, especially in case the test set is rather small.

For OCC models, specificity is calculated only in the presence of non-target objects. This figure of merit is obtained empirically or, for some OCC method, it can be calculated theoretically [12] as the type II error, β. In case non-target objects are organised in several extraneous classes, specificity should be calculated for each extraneous class separately; otherwise, the reported value of specificity would not reflect the true relationships between the target and alternative classes. Such an example is presented below in Section 4. At the same time, total specificity can be reported, if the customer is not interested in the details.

It should be mentioned that both sensitivity and specificity depend on the selected value of type I error, α. The first parameter has a direct relation: SENS=(1−α). Conversely, dependence of specificity is more complex and will be considered in Section 4.

2.2. 'Rigorous' vs. 'Compliant' approach

We distinguish two approaches when building OCC models. We call the first one 'rigorous' OCC. This means that a model is developed based merely on the target training dataset, and optimal conditions are obtained employing the type I error, α (the rate of wrong rejections of the target samples), or sensitivity, computed as (1-α). Depending on the method, this α value may be estimated a-priori, and/or calculated a-posteriori. This evaluation is made using the target samples only. Considering that sensitivity is an experimental estimate of type I error, α, of a given model, outcomes whose sensitivity is closest to (1-α) should be considered as optimal, when optimising a model in a 'rigorous' way.

The second approach is called here 'compliant' OCC. Such a very common modelling strategy utilises additional information regarding non-target samples when, except for data from the target class, one/several datasets from extraneous classes are available. For each alternative class, the type II error, β (the rate of wrong acceptances of objects from the alternative class), or the corresponding specificity, computed as (1-β), is estimated. In this case, model optimisation is performed with respect to the estimates of both α and β and the OCC model that has maximum efficiency is selected.

2.3. Data driven soft independent modelling of class analogy (DD-SIMCA)

The DD-SIMCA method develops a decision rule that delineates the objects from the target class by exploring the corresponding data matrix. The procedure consists of two steps. The first step is the application of principal component analysis (PCA) [13], establishing a model using training samples from the target class. The $(I{\times}J)$ data matrix $\mathbf{X}$ (duly pre-processed, *e.g.* centred) is decomposed by:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^t + \mathbf{E} \tag{1}$$

where $\mathbf{T}=\{t_{ia}\}$ is the $(I{\times}A)$ scores matrix; $\mathbf{P}=\{p_{ja}\}$ is the $(J{\times}A)$ loadings matrix; $\mathbf{E}=\{e_{ij}\}$ is the $(I{\times}J)$ matrix of residuals; and A is the number of principal components (PC). Matrix $\mathbf{T}^t\mathbf{T}=\mathbf{\Lambda}=\mathrm{diag}(\lambda_{1,\dots},\lambda_A)$ is a diagonal matrix with elements $\lambda_a = \sum_{i=1}^{I} t_{ia}^2$, which are the eigenvalues of matrix $\mathbf{X}^t\mathbf{X}$ ranked in descending order.

In the second step, we employ the PCA results when calculating two relevant distances for each object i=1,..., I of the training set. They are the score distance (SD), h_i, and the orthogonal distance (OD), v_i:

$$h_i = \mathbf{t}_i^t(\mathbf{T}^t\mathbf{T})^{-1}\mathbf{t}_i = \sum_{a=1}^{A}\frac{t_{ia}^2}{\lambda_a}, \qquad v_i = \sum_{j=1}^{J} e_{ij}^2 \tag{2}$$

SD represents the position of a sample within the score space, while OD characterises a sample distance to the score space.

In a previous study [14], it was shown that distributions of both distances are well approximated by the scaled chi-squared distribution:

$$N_h\frac{h}{h_0} \propto \chi^2(N_h) \quad N_v\frac{v}{v_0} \propto \chi^2(N_v) \tag{3}$$

where v_0 and h_0 are the scaling factors, N_h and N_v are the numbers of the degrees of freedom (DoF). These parameters are considered unknown and estimated using a data-driven method explained in ref. [10].

Statistics c, called the *total distance*:

$$c = N_h\frac{h}{h_0} + N_v\frac{v}{v_0} \propto \chi^2(N_h + N_v). \tag{4}$$

is used to generate the decision rules. Any decision rule (*i.e.*, an acceptance area) is determined by an inequality:

$$c \le c_{crit} \tag{5}$$

The first decision rule is developed for a given type I error, α:

$$c_{\mathrm{crit}} = \chi^{-2}(1-\alpha,\ N_h + N_v) \tag{6}$$

where χ^{-2} is the quantile of the chi-squared distribution with N_v+N_h DoF. To calculate the type II error β, we should assume that an alternative class is available:

$$\beta = \Pr\left\{\chi'^2(N_h + N_v,\ s) < \frac{c_{\mathrm{crit}}}{c'_0}\right\}, \tag{7}$$

where c_{crit} is defined in Eq. (5) and χ'^2 is the non-central chi-squared distribution. Parameters c'_0 and s are found by the method explained in ref. [12].

Using this approach, every sample and the acceptance areas can be plotted in the coordinates of h/h_0 against v/v_0. Fig. 2 illustrates an example of this distance plot. Applying theory (Eqs. (6) and (7)) in practice, we can yield two acceptance areas. First, we can develop a 'rigorous' decision rule defined by a given α, and then calculate a subsequent β error. On the other hand, we can employ a 'compliant'