Available online at www.sciencedirect.com
SCIENCE DIRECT®
Chemometrics and Intelligent Laboratory Systems 80 (2006) 109 – 119
Chemometrics and intelligent laboratory systems
www.elsevier.com/locate/chemolab

A comparison of different discriminant analysis techniques in a steel industry welding process

José M. Prats-Montalbán [a,*], Alberto Ferrer [a], J.L. Malo [b], J. Gorbeña [b]

[a] Department of Applied Statistics, Operations Research and Quality, Technical, University of Valencia. Cno. De Vera s/n, Edificio I3, 46022, Valencia, Spain
[b] Regional Development Unit, Fundación LABEIN, Parque Tecnológico de Bizkaia, Edificio 700, 48160, Derio, Spain

Received 22 November 2004; received in revised form 3 August 2005; accepted 19 August 2005
Available online 3 October 2005

Abstract

The present work compares several statistical discriminant analysis techniques applied to a steel industry welding process. Data from 85 variables collected from 18,605 links, classified as *Good* (18,381), *Defective* (195) or *Bad* (29) from laboratory analysis, are available. Process engineers want to find out which variables explain the main differences between the three defined types, so they can implement effective action to reduce the percentage of *Defective* and *Bad* links. The approaches used are SIMCA, Global PCA, PLS-DA and Fisher's Linear Discriminant Analysis (LDA). The dataset comprises two kinds of variables, one for the chemical properties of the links, and the other related to the welding process. All the above approaches basically lead to the same results and match the ones derived from the more traditional Fisher s Linear Discriminant Analysis (LDA) technique. The pros and cons of the approaches used are discussed.

Keywords: PCA; SIMCA; PLS-Discriminant Analysis; Linear Discriminant Analysis; Welding process

1. Introduction

Production of oil rig chains constitutes a delicate series system manufacturing process, made up of as many links as necessary in order to make the work possible at different level (or depth) stages.

The main risk derived from this product is the high cost associated to the breaking of just one of the chain links, not only in terms of the chain itself, but also those associated to the oil extraction process, and the recovery of the chain. This risk, as the chain is a so-called series system, is related to the breaking of just one of the links, making their manufacturing and quality control essential, in order to guarantee their reliability.

The links manufacturing process is as follows: when one link has to be produced, the steel cylinder that constitutes the original material passes through an oven to reach an adequate temperature that permits it to get blended.

Once the cylinder is completely blended to the link form and entwined with the previous one so as to build the chain, the welding process is executed on both left and right extremes of the link, for a period of 90 s approximately. During this process (split up in several sub-phases) different variables derived from their time evolution and from some events occurring during the welding phase are calculated. Examples of these variables are the average values of electrical voltages and intensities, time of change from one sub-phase to another, temperatures, average value of the approximation velocity between the clamps that press the extremes of the links, pressures at a certain instant during the process, distances between the clamps at each initial time of each subphase, etc.; which are computed to summarize the information provided by those time dependent variables.

Afterwards, the link, which is now actually part of the chain, leaves the welding machine, leaving the next one to be welded, and cools by air convection to the ambient temperature. Then, the link passes through an echo sounder test for detecting internal fissures produced during the welding process that reduce its strength. Finally, links are classified as *Good*, *Defective* or *Bad*, according to the results of the echo sounder test.

The *Good* links remain in the chain, whilst the *Defective* ones are removed and processed again so that they receive a

* Corresponding author. Departamento de Estadística e I.O., Aplicadas y Calidad. Universidad Politécnica de Valencia. Cno. De Vera s/n, Edificio, I3, 46022, Valencia, Spain.
E-mail address: jopramon@eio.upv.es (J.M. Prats-Montalbán).

0169-7439/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2005.08.005

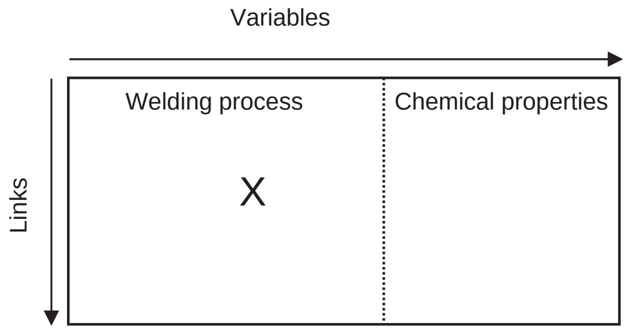

Fig. 1. Data structure **X**. First mode is formed by the links, while the second mode is joined by the chemical and welding process variables.

thermic treatment in order to give them *Good* properties. On the other hand, *Bad* links are completely removed from the chain and also from the process. All these activities lead to logistical problems and increase the manufacturing cost.

The fact that the quality of the chain is obviously related to the quality of the links, and the latter to the existence of internal fissures (which are produced during the welding process), is the reason why the plant engineers have considered that this is the critical phase of the whole manufacturing cycle.

Add to the welding process variables, information on the chemical composition of the links (percentages of Carbon, Copper, Aluminium, Manganese, etc.) were also available. In this way, the whole data set was made up with the 66 variables derived from their time evolution and from some events occurring during the welding phase, together with another 19 variables associated to the chemical composition of the links. As data from 18,605 links are available, the data structure becomes a 2-way array **X** (18605, 85), as shown in Fig. 1.

Data structure, with a multitude of variables that are usually correlated, makes the projection methods a good choice for analyzing these types of arrays, when trying to classify the different types of links as *Good*, *Defective* or *Bad*. Furthermore, some of these methods can let one know which variables govern the welding process, and which are responsible for the different behaviour of the different types during this process.

The present work analyses this data structure by using different approaches: Global PCA [1], SIMCA (Soft Independent Modelling of Class Analogy) [2], PLS-DA [3] and Fisher's LDA [4], introduced in Section 2. Section 3 presents a discussion on the different approaches studied and the conclusions of the paper.

2. Welding process analysis

2.1. Pre-treatment of data

As is well known, pre-treatment of data is a critical task when trying to efficiently analyse any kind of data. There are some works which carry on with this problem, concluding that, in order to get good results, data pre-processing is sometimes more important that the model used [5].

As the goal of this work was to establish differences with respect to the main behaviour of the links, and different kinds of variables were present in the study, a classical pretreatment of data was carried out, centring across the links mode, and scaling along the variables mode.

On the other hand, due to the high number of available links (18,381 *Good*, 195 *Defective* and 29 *Bad* links), training and test sets were derived for each type. This way, 9191 *Good*, 130 *Defective* and 20 *Bad* links were used as training data to build the models. Test set was made up of the rest of the links.

As commented by the referees, both variables and observations are disproportionately distributed in the training data set.

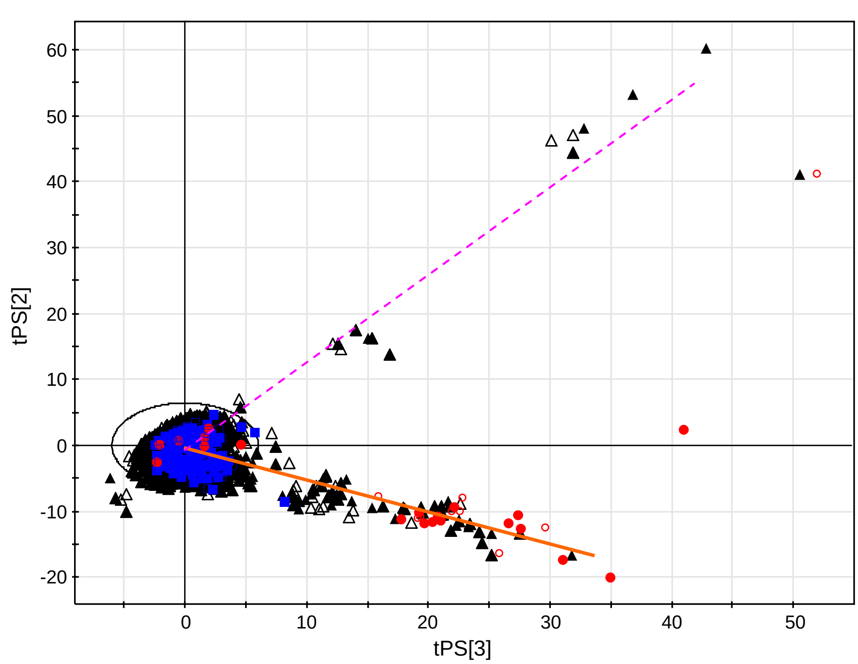

Fig. 2. Score Plot of PCs 2 and 3 from Global PCA. Solid icons correspond to training set, while opened icons correspond to test set. Triangles are relates to *Good* links, squares to *Defective* links, and circles to *Bad* links.