Chemometrics and Intelligent Laboratory Systems 107 (2011) 318–332

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Genetic algorithm for informative basis function selection from the wavelet packet decomposition with application to corrosion identification using acoustic emission

G. Van Dijck *, M.M. Van Hulle

Computational Neuroscience Research Group, Katholieke Universiteit Leuven, Herestraat 49, B-3000 Leuven, Belgium

A R T I C L E I N F O

Article history:
Received 6 August 2010
Received in revised form 19 April 2011
Accepted 2 May 2011
Available online 11 May 2011

Keywords:
Acoustic emission
Corrosion monitoring
Feature subset selection
Genetic algorithm
Mutual information
Wavelet Packet Transform

A B S T R A C T

Chemical process installations are exposed to aggressive chemicals and conditions leading to corrosion. The damage from corrosion can lead to an unexpected plant shutdown and to the exposure of people and the environment to chemicals. Due to changes within and on the surface of materials subjected to corrosion, energy is released in the form of acoustic waves. This acoustic activity can be captured and used for corrosion monitoring in chemical process installations. Wavelet packet coefficients extracted from the acoustic activity have been considered to determine whether corrosion occurs, and to identify the type of corrosion process, at least for the most important corrosion processes in the chemical process industry. Feature subset selection is then applied to these wavelet coefficients to achieve a much higher accuracy in the identification of different corrosion processes than when no feature subset selection is applied to the acoustic waves. However, due to the statistical dependencies that potentially exist between the wavelet coefficients, the latter should not be selected independently from each other. Local discriminant basis selection algorithms do not take the statistical dependencies between wavelet coefficients into account. In this paper, we have used several mutual information-based approaches that take these dependencies into account and compared them to the wavelet-specific local discriminant basis selection algorithm. Furthermore, a hybrid filter–wrapper genetic algorithm, which uses a relevance–redundancy approach as a local search procedure, was designed. The highest classification accuracies are obtained with the hybrid filter–wrapper genetic algorithm, for all classifiers used in this paper. Furthermore, the proposed algorithm easily outperformed one of the most commonly used classifiers in chemometrics: partial least squares discriminant analysis (PLS-DA). A naïve Bayes classifier that uses the features selected by the hybrid filter–wrapper genetic algorithm was able to identify the absence of corrosion, uniform corrosion, pitting and stress corrosion cracking, with an accuracy of up to 87.20%.

1. Introduction

1.1. The continued need for corrosion monitoring

Corrosion destroys each year a large part of the world's economy. The global cost of corrosion, consisting of direct and indirect costs, is estimated at 3.8% of the GWP (gross world product) [1]. This global cost equals $1930 billion in US dollars for the year 2004. For the United States, the total cost of corrosion is estimated at $504 billion ($304 billion direct costs + $200 billion indirect costs) per year in 2004 [1,2], which represents about 4.7% of its GDP (gross domestic product). The direct costs are the costs incurred by the owners or operators. These costs consist of the following [1,2]: the use of more expensive or additional materials to prevent corrosion, the labor and equipment for corrosion management, the loss of revenue due to disruption of the supply of the product, the loss of reliability, and the loss of capital due to corrosive deterioration. The indirect costs are incurred by the user of the products or the society. In the chemical, petrochemical and pharmaceutical sectors [2], which are the targeted sectors here, the direct costs are extrapolated to $1.9 billion for 2004 [1,2]. About 60% of the mechanical failures in the chemical process industry are due to corrosion [3]. A large part, 25 to 40%, of the direct and indirect costs can be saved by the use of corrosion monitoring and control systems [4]. Corrosion detection provides feedback to operators about the state of the plant so that they can participate in managing the high corrosion costs [4]. Direct costs that can be avoided by the use of monitoring systems are due to the increased reliability of the plant, avoidance of the disruption of the supply of products, decreased loss of capital and avoidance of lawsuits against companies (e.g., due to pollution caused by leaks of the installations), among other factors. Indirect costs can be equally important as these costs have an impact on the society and environment. In some sectors, damage due to corrosion can be tolerable, but in the chemical, petrochemical and nuclear sectors, corrosion damage can be catastrophic, even resulting in the loss of lives and environmental damage.

* Corresponding author. Tel.: +32 16330428; fax: +32 16345960.
E-mail address: gert.vandijck@med.kuleuven.be (G. Van Dijck).

0169-7439/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2011.05.001

Regular practice in the chemical process industry consists of periodic inspections of the plant, e.g., every 3 months, every 6 months or every year [5]. A recurring problem with such periodic inspections is that one can overlook the active damage that occurs in the plant; furthermore, immediately after inspection, the damage can continue to grow until the next periodic inspection is scheduled. Clearly, such situations should be avoided. A solution is offered by continuous monitoring using corrosion monitoring systems. Different techniques are available for corrosion detection and monitoring in the chemical process industry [5,6]. In this research, we identify the most important types of corrosion in the chemical process industry using the acoustic emission signals that are emitted during the corrosion process. Chemical reactions, as occurring during corrosion, emit acoustic activity [7,8] as well as the microscopic damage and fracture processes that occur during corrosion [9]. The acoustic emission technique has the advantage that it is low cost and allows for a continuous, on-line monitoring so that the damage can be detected as soon as it occurs [6].

1.2. Wavelet packet feature extraction and selection from acoustic emission

Although future successes in corrosion prevention still depend on selecting and developing more corrosion resistant materials, it is expected that the main progress in corrosion prevention will be achieved with better information-processing strategies and the development of more efficient monitoring tools that support corrosion control programs [10]. Feature extraction, feature subset selection, and classifier choice and design are all information-processing strategies that should be explored in the design of better corrosion monitoring systems.

Features to characterize the acoustic emission activity have often been obtained in the time-amplitude domain [5,8,11], the frequency domain [5,8,12], or the time–frequency domain using the Continuous Wavelet Transform (CWT) [13,14], the Discrete Wavelet Transform (DWT) [14] or the Wavelet Packet Transform (WPT) [15]. The process of constructing informative features that discern between different classes is often not trivial, but some generic approaches are available [16]. One generic approach is to consider basis functions that can be used to extract features. A library of basis functions can be obtained from the Wavelet Packet Transform [17–19]. Moreover, Wavelet Packet Decompositions are more flexible than the Discrete Wavelet Transform (DWT) and the Fourier Transform (FT) [19].

One of the challenges that arises after the use of the Wavelet Packet Transform is the selection of a basis that is optimal in some sense, or the selection of a few coefficients for signal compression or pattern recognition purposes [18,20–25]. The current paper contributes to the selection of the most informative basis functions, from a library of wavelet packets, to distinguish between different classes of corrosion, using information theory. We use mutual information [26] to guide the search for informative basis functions by taking into account the statistical dependencies between the wavelet coefficients. Mutual information is intensively used in chemometrics as a feature selection criterion [27–31]. It is a filter-based variable selection technique, meaning that it does not take the interaction with the final machine learning algorithm used for pattern prediction into account [16,32]. This may lead to an inferior performance compared to wrapper [32] approaches. However, the latter often come with an increased computational cost. A wrapper-based feature subset selection approach will become computationally expensive when thousands of features are obtained, which is typically the case after wavelet packet coefficients have been extracted from acoustic emission signals. A solution exists in combining the mutual information-based approach with a wrapper search, leading to a so-called hybrid filter–wrapper approach [33,34]. In this article, we will follow this hybrid filter–wrapper strategy by performing the expensive local search in a genetic algorithm [35] with a simple, mutual information-based filter. Our approach is generically applicable to classification problems only requiring a set of training signals with corresponding class labels. The approach proposed here is called GIBFS (Genetic Informative Basis Function Selection). It is a genetic algorithm driven approach using mutual information as a local search procedure applied to a dictionary of basis functions for feature extraction.

2. Materials and methods

2.1. Experimental set-up

This section briefly sketches the experimental set-up for recording the acoustic emission signals. The experimental set-up is shown on the left side of Fig. 1.

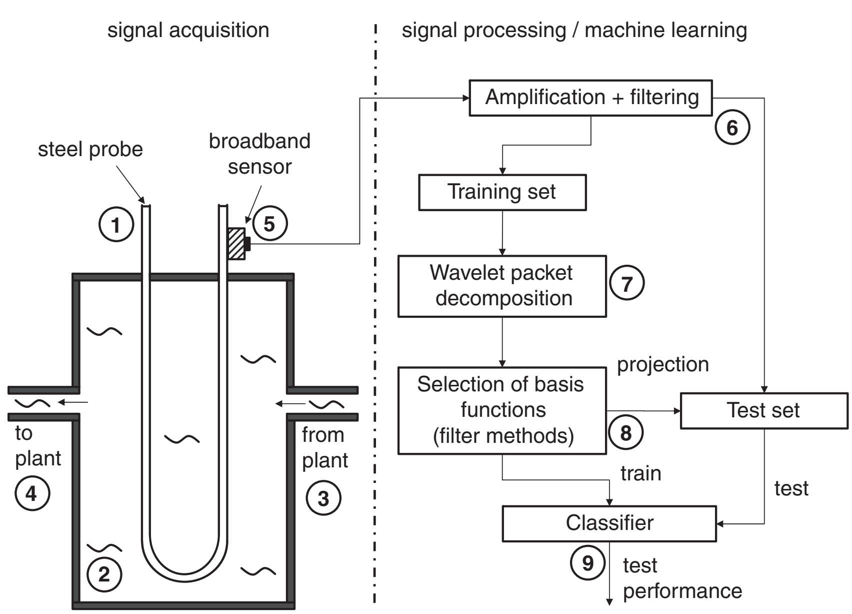

Fig. 1. Experimental set-up of the probe in a by-pass of the installation on the left and the signal processing and machine learning part on the right.