Computing confidence and prediction intervals of industrial equipment degradation by bootstrapped support vector regression

Isis Didier Lins\textsuperscript{a,b,*}, Enrique López Droguett\textsuperscript{c}, Márcio das Chagas Moura\textsuperscript{a,b}, Enrico Zio\textsuperscript{d,e}, Carlos Magno Jacinto\textsuperscript{f}

\textsuperscript{a} Center for Risk Analysis and Environmental Modeling, Federal University of Pernambuco, Recife, Brazil
\textsuperscript{b} Department of Production Engineering, Federal University of Pernambuco, Recife, Brazil
\textsuperscript{c} Department for Risk and Reliability, Mechanical Engineering Department, University of Maryland, College Park, USA
\textsuperscript{d} Department of Energy, Polytechnic of Milan, Milan, Italy
\textsuperscript{e} European Foundation for New Energy, Electricité de France, Ecole Centrale Paris – Supelec, France
\textsuperscript{f} CENPES, PETROBRAS, Rio de Janeiro, Brazil

A R T I C L E   I N F O

Article history:
Received 4 June 2013
Received in revised form 26 December 2014
Accepted 15 January 2015
Available online 23 January 2015

Keywords:
Degradation
Prognostics
Uncertainty
Confidence and prediction intervals
Support vector regression
Bootstrap
Scale growth rate

A B S T R A C T

Data-driven learning methods for predicting the evolution of the degradation processes affecting equipment are becoming increasingly attractive in reliability and prognostics applications. Among these, we consider here Support Vector Regression (SVR), which has provided promising results in various applications. Nevertheless, the predictions provided by SVR are point estimates whereas in order to take better informed decisions, an uncertainty assessment should be also carried out. For this, we apply bootstrap to SVR so as to obtain confidence and prediction intervals, without having to make any assumption about probability distributions and with good performance even when only a small data set is available. The bootstrapped SVR is first verified on Monte Carlo experiments and then is applied to a real case study concerning the prediction of degradation of a component from the offshore oil industry. The results obtained indicate that the bootstrapped SVR is a promising tool for providing reliable point and interval estimates, which can inform maintenance-related decisions on degrading components.

\textcopyright{} 2015 Elsevier Ltd. All rights reserved.

1. Introduction

The prediction of equipment reliability measures is of great importance in many industrial sectors, due to the fact that equipment failures may lead to accidents and/or plant unavailabilities which can translate into costs and production losses \cite{1}. For example, in the context of the oil industry, scale deposition due to salt accumulation may prevent equipment from properly actuating; this may cause the interruption of oil production\cite{2}. Kinetic formulas have been introduced to predict scale formation by correlation of the relevant explanatory variables \cite{4-6}. These deterministic mathematical formulae correlate the scale output variable to the relevant explanatory input variables. Establishing and validating such formula requires a large number of reliable data and is not easy in practice. Lifetime distributions and stochastic processes are also used to model the failure behavior of equipment \cite{7}, but generally require making simplifying assumptions rarely met in practice.

Another alternative is to use data-driven learning methods for predicting the evolution of the degradation process affecting the equipment. These methods do not require specific knowledge on the functional relationship between the influential factors (the explanatory variables) and the degradation variable of interest. Among these methods, Support Vector Regression (SVR) has provided promising results in reliability \cite{8,9}, economic \cite{10,11}.
environmental [12,13], electrical [14,15] applications, among others. The SVR learning (or training) phase involves the resolution of a convex quadratic optimization problem, for which the Karush–Kuhn–Tucker (KKT) first order conditions are necessary and sufficient for a global optimum [16]. Indeed, this is an advantage of SVR over other learning techniques, such as artificial neural networks (ANNs), which may be trapped into local optima [17]. After the training step, in correspondence of a new observation symbol \( \mathbf{x}_+ \), the estimate \( \hat{y}_+ \) of the true mean response \( \mu_+ \) can be obtained via the adjusted regression function (i.e. the estimator) [17,18].

In many applications, particularly reliability and failure prediction problems, it is important to account for the variability of the estimator and/or assess the uncertainty on the prediction of the response variable \( Y_+ \). This means that besides point estimates, confidence intervals for \( \mu_+ \) and/or prediction intervals for \( Y_+ \) need to be calculated [18,19].

Given that SVR does not require any hypothesis about the distribution of the error term, the central limit theorem enables the approximation of confidence and prediction intervals when large data sets are available [20]. On the other hand, for small numbers of data points (e.g. 100 or less), the intervals based on bootstrap [21,22] tend to be more accurate, given that they do not rely on asymptotic results but on the construction of the limit distribution from the available data [20].

The main idea of bootstrap methods is to estimate probability distributions for statistics of interest obtained from the available data. For regression, bootstrap can be based on: (i) pairs, when the observations from the available data set are sampled with replacement; (ii) on residuals, when the residuals associated with a regression model adjusted over the original sample are sampled with replacement. They are widely used in (generalized) linear, non-linear and nonparametric regression [23]. For example, in linear regression, Cribari-Neto [24] and Cribari–Neto and Lima [25] use bootstrapped hypothesis testing and intervals tailored to account for heteroskedasticity with an estimator of the covariance matrix that considers the effects of leverage points in the design matrix. In nonparametric regression, Zio [26], Cadini et al. [27], Secchi et al. [28] and Zio et al. [29] analyze by bootstrap the uncertainty of ANN predictions of nuclear process parameters. In the context of support vector machines, bootstrap approaches have been mainly applied to classification problems [30,31]. Specifically for SVR, Lin and Weng [32] and Yang and Ong [33] have proposed probabilistic outputs, but assuming a probability distribution for the response variable.

De Brabanter et al. [20] develops approximate pointwise and simultaneous intervals for Least Squares Support Vector Machines (LS-SVM). The authors then compare the approximate intervals with intervals obtained via bootstrap based on residuals (for both homoskedastic and heteroskedastic errors). In the present work, differently from [20], we also consider bootstrap based on pairs, and the SVR procedure put forward involves Vapnik’s \( \epsilon \)-insensitivity loss function [34] instead of quadratic errors as in LS-SVM.

Lins et al. [35] present a comprehensive approach for variable selection, parameter tuning and uncertainty analysis based on the coupling of Particle Swarm Optimization (PSO), bootstrap methods and SVR. They present an application example for the prediction of Times Between Failures (TBFs) of onshore oil wells located in the Northeast of Brazil. However, in [35], the authors do not validate the combination of bootstrap with SVR for uncertainty characterization and quantification.
دریافت فوری متن کامل مقاله

امکان دانلود نسخه تمام متن مقالات انگلیسی
امکان دانلود نسخه ترجمه شده مقالات
پذیرش سفارش ترجمه تخصصی
امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
امکان دانلود رایگان ۲ صفحه اول هر مقاله
امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
دانلود فوری مقاله پس از پرداخت آنلاین
پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات