International Journal of Pressure Vessels and Piping 86 (2009) 403–411

Contents lists available at ScienceDirect

International Journal of Pressure Vessels and Piping

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

Risk assessment for pipelines with active defects based on artificial intelligence methods

Calin I. Anghel*

Department of Chemical Engineering, Faculty of Chemistry and Chemical Engineering, University "Babes-Bolyai", Cluj-Napoca, Romania

ARTICLE INFO

Article history:
Received 24 April 2007
Received in revised form
16 October 2008
Accepted 22 January 2009

Keywords:
Pressure piping
Safety index
Limit state function
Support vector machine
Minimax approach
Binary classification

ABSTRACT

The paper provides another insight into the pipeline risk assessment for in-service pressure piping containing defects. Beside of the traditional analytical approximation methods or sampling-based methods safety index and failure probability of pressure piping containing defects will be obtained based on a novel type of support vector machine developed in a *minimax* manner. The safety index or failure probability is carried out based on a binary classification approach. The procedure named *classification reliability procedure*, involving a link between artificial intelligence and reliability methods was developed as a user-friendly computer program in MATLAB language. To reveal the capacity of the proposed procedure two comparative numerical examples replicating a previous related work and predicting the failure probabilities of pressured pipeline with defects were presented.

1. Introduction

Failure probability and implicit risk assessment for pipelines subjected to different degradation mechanism and systematic inspection program have been widely studied in the last decades [1–8]. In technical domains risk assessment based on safety index and failure probability is still of greater importance in engineering activities. Generally the probabilistic approaches were obtained based on well-known principles of structural reliability. Simulation techniques alike the *response surface*, *directional simulation or asymptotic techniques*, analytical approximation known by their acronyms *FORM/AFORM/FOSM* and *SORM*, sampling-based methods (*Monte Carlo Simulation, Latin Hypercube Sampling*, etc.) or analyses based on the commercial finite element code were the most common used techniques. To enhance performances and to increase the computational efficiency in the last decades innovative ideas in the field of artificial intelligence were integrated in many engineering and scientific disciplines. The ability of these methods to investigate new phenomenon cases where the information cannot be easily accessed theoretically or explicitly relational physics it is well known. Excepting few items [9–12] such modern procedures have been rarely used in reliability analyses. Based on general principles of structural reliability methods [15] especially on *FORM/AFORM/FOSM* principles the main goal of this paper is to introduce a novel procedure implemented on support vector machine in a *minimax* manner, able to predict the failure probability and implicit the risk for pipelines with active defects. The proposed procedure named *classification reliability procedure* estimates the location of the most probable point (*MPP*) and calculates the safety index. The idea is that the safety index reported to an equivalent hyper plane (equivalent linear form of the limit state function into a feature space) would be achieved into a binary *minimax* classification manner. The procedure is closed to solve reliability problems involving intricate structural systems, with explicit or implicit highly non-linear performance functions with a large number of random variables. The main parts of this paper are: (a) introductions, (b) theoretical background for safety index assessment based on support vector machine in a *minimax* approach, (c) numerical examples implemented in the *MATLAB* package and (d) conclusions. The two-implemented examples were directed to prove the validity of the procedure into failure probabilities assessment. These examples do not represent "case study". The results were compared to those obtained by the traditional consecrated techniques. The former replicates a previous related study [3] predicting pipeline failure probability for pipelines with active corrosion defects. The later, presents a case study on the failure probability for in-service pressure piping, with active defects subjected to erosive and corrosive slurries, working as

* Tel.: +40 264 593833; fax: +40 264 590818.
E-mail address: canghel@chem.ubbcluj.ro

0308-0161/$ – see front matter
doi:10.1016/j.ijpvp.2009.01.009

sludge pipelines into a technological process of gold leaching. The comparative results point out a reasonable agreement and the opportunity of such assessments based on artificial intelligence in risk analysis. Today we are unaware of other similar approaches based on support vector machine in a *minimax* manner and related to structural reliability analysis. By proper developments the procedure may be applied in many other scientific or engineering domains.

2. Assessment of safety index based on support vector machine developed in a *minimax* manner

Only main principles and basic theoretical framework related to safety index assessments in *minimax* classification procedure will be presented. More details will be found elsewhere [16–22]. Usually in codes of practice [1–3,13–15] one or combination of more limit state functions represents a common expression of a limit state. In classical reliability techniques the evaluation of the limit state functions (*LSF*) is based on terms of the reduced variables (standard normal, independent and uncorrelated variates $\mathbf{u}$) and all approximations are made in a reduced space U^d. The replacements are grounded on well-known isoprobabilistic transformation [15], thus $LSF = LSF\,(x_i) \twoheadrightarrow LSF\,(\mathbf{u}_i) = 0$. Thus the failure domain F characterized by the existence of a limit state function whose non-positive values define the non-reliability domain was expressed as:

$$F = \left\{\mathbf{u}\in U^d | LSF(\mathbf{u}) < 0\right\} \text{when } LSF(\mathbf{u})\in C2 \tag{1}$$

The safety domain S characterized by the existence of limit state function whose positive values define the reliability domain may be expressed as:

$$S = \left\{\mathbf{u}\in U^d | LSF(\mathbf{u}) > 0\right\} \text{when } LSF(\mathbf{u})\in C1 \tag{2}$$

Basically limit state functions (*LSF*) represent a failure limit hyper surface that is a boundary between the safe and failure regions. The safety index is reported to the point that lies closest to the origin of the system in transformed reduced space $U^{\boldsymbol{d}}$ and belongs to the limit state hyper surface. This point is often referred as the most probable point (*MPP*). According to *FORM/AFORM* methods, still importance in engineering domains, there is a direct relationship between the safety index and the first order approximation for the probability of failure:

$$P_f = \text{Prob}\left\{\mathbf{u}\in U^d | LSF(\mathbf{u}) < 0\right\} = \Phi(-\beta) \tag{3}$$

where $\Phi(\ldots)$ represents the one-dimensional standard Gauss (normal) cumulative density function and β is the safety (reliability) index. Basically previous relationship eqn. (3) is only approximate, but in the unique case of a linear limit state function of Gauss distributed random variables the relationship is exact. Well known by their simplicity *FORM/AFORM* has some drawbacks: (a) need statistical characterisation of parameters, (b) variable stability, (c) give possible local extremes, and (d) give inaccurate results when the failure surface is highly non-linear, (e) usually give approximate results. To overcome certain previous mentioned drawbacks the paper proposes a procedure based on support vector machine in a *minimax* approach. Based on its nature [16–19,22] important incentives for the support vector machine in *minimax* approach are: (a) it does not assume normality of the data and makes no prior assumptions concerning the data distribution, *as it is a non-parametric classifier*, (b) provides better accuracy even with a small number of training samples and is fast and simple in implementation, (c) avoids specific problems such as over-fitting and local minima, (d) has a relative explicitly nature, (e) avoids drawbacks of simple assessment of the same covariance for each classes and thus defining the margin in a local way, (f) provides an explicit direct upper bound on the probability of misclassification of new data, without making any specific distribution assumptions and (g) defines the margin in a global way and obtains explicit decision boundaries based on a global information of available data. The basic idea of our procedure is to establish the safety index and the failure probability reported to an *equivalent linear form of the limit state function (equivalent hyper plane) in a binary minimax classification manner.*

2.1. *Basic principles of support vector machine in a* minimax *approach*

Basically our work started from support vector machine [16,17] that is a primarily two-class optimised classifier. In support vector machine the optimisation criterion used is the width of the margin between the classes. Unlike support vector machine, for which the closest points to the decision boundaries are most important, a *minimax* approach [18,19] minimizes the maximum of distances to the "means" of classes. As was earlier stated [18,19] into a *minimax* approach for a binary classification problem, with $\mathbf{u_1}$ and $\mathbf{u_2}$ denoting random vectors data from each of two classes as $\mathbf{u}_1\in$ Class 1 and $\mathbf{u}_2\in$ Class 2, a hyper plane that separates the two classes of points,

$$H(\mathbf{w},b) = \left\{\mathbf{u}\middle|\mathbf{w}^T\cdot\mathbf{u} = b\right\} \text{where } \mathbf{w}\in\mathbf{R}^d/\{0\} \text{and } b\in\mathbf{R} \tag{4}$$

with maximal probability in respect to all distributions having mentioned means $\bar{\mathbf{u}}_1, \bar{\mathbf{u}}_2$ and covariance matrices $\sum\mathbf{u}_1, \sum\mathbf{u}_2$, may be determined by:

$$\max_{\alpha,\mathbf{w}\neq 0,b}\ s.t.\begin{cases}\inf\limits_{\mathbf{u}_1(\bar{\mathbf{u}}_1,\sum\mathbf{u}_1)} & \text{Probability}\left\{\mathbf{w}^T\,\mathbf{u_1} \geq b\right\} \geq \alpha \\ \inf\limits_{\mathbf{u}_2(\bar{\mathbf{u}}_2,\sum\mathbf{u}_2)} & \text{Probability}\left\{\mathbf{w}^T\,\mathbf{u_2} \leq b\right\} \geq \alpha\end{cases} \tag{5}$$

The term α represents lower bound of the correct classification of future data, $\mathbf{w} \in R^d/\{0\}$ is the outward normal vector of the hyper plane, b offset of the hyper plane from origin, $\boldsymbol{u}$ generic random vector of the original input space and R, R^d real and Euclidian d-dimensional input space. In a *minimax* approach the classifier must to minimize the misclassification probability $(1-\alpha)$ by an optimal separating hyper plane, named *minimax probabilistic decision hyper plane*.

In many real analyses, problems–patterns are scattered in high-dimensional (often) non-linear subspaces; thus it must be handled by the non-linear classification problems. To overcome this the *kernel trick* is used to map the input data points into a high-dimensional feature space. To simplify kernels can be viewed as inner products into mapped space named features space or some Hilbert space (an inner product in feature space has an equivalent kernel in input space). The principle behind kernel trick is to map original d-dimensional input processed points $\boldsymbol{u} \in \mathbf{u}\in\mathbf{U}(\mathbf{u}_1,\ldots,\mathbf{u_d})\subset R^d$ from real input space into a high f-dimensional feature space R^f by the eigenvalues and eigenfunctions of the reproducing kernel. The mapping function, Φ (usually unknown), transforms input vectors (data points) of a dimension d to some other higher dimensional $R^{\boldsymbol{f}}$ space: $\phi : R^d \mapsto R^f$. Into this named feature space a linear classifier–surface between the classes corresponds to a non-linear decision into original input space (Fig. 1). To carry out this a kernel function (a symmetric real valued function of two variables), $K(u_i,u_j) = \langle\phi(u_i)\times\phi(u_j)\rangle$ satisfying Mercer's condition must be defined [16–19]. It must be mentioned that *minimax* classification is similar to maximum margin classification with respect to the mean of the classes, where a factor that