Mechanical Systems and Signal Processing 37 (2013) 76–91

Contents lists available at SciVerse ScienceDirect

Mechanical Systems and Signal Processing

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

A methodology for exploiting the tolerance for imprecision in genetic fuzzy systems and its application to characterization of rotor blade leading edge materials

Luciano Sánchez [a], Inés Couso [b,*], Ana M. Palacios [a], José L. Palacios [c]

[a] Departamento de Informática, Universidad de Oviedo, 33071 Gijón, Asturias, Spain
[b] Departamento de Estadística, Universidad de Oviedo, 33071 Gijón, Asturias, Spain
[c] Department of Aerospace Engineering, The Pennsylvania State University, University Park, PA 16802, USA

ARTICLE INFO

Article history:
Received 17 December 2011
Received in revised form
13 February 2012
Accepted 18 February 2012
Available online 16 March 2012

Keywords:
Genetic fuzzy systems
Fuzzy rule-based classifiers
Vague data
Mutual information
Ice-phobic materials
Shear adhesion strength

ABSTRACT

A methodology for obtaining fuzzy rule-based models from uncertain data is proposed. The granularity of the linguistic discretization is decided with the help of a new estimation of the mutual information between ill-known random variables, and a combination of boosting and genetic algorithms is used for discovering new rules. This methodology has been applied to predict whether the coating of an helicopter rotor blade is adequate, considering the shear adhesion strength of ice to different materials. The discovered knowledge is intended to increase the level of post-processing interpretation accuracy of experimental data obtained during the evaluation of ice-phobic materials for rotorcraft applications.

1. Introduction

Expert knowledge elicitation from experimental data is a valuable tool in the engineering design process, that aids to make agile decisions in the absence of a causal model or extensive prototype testing. Often, expert knowledge is best organized as a set of linguistic rules, describing the circumstances under which the behavior of an element is admissible or unsuitable for a given application. The benefits of computer-produced knowledge depend on the accuracy of the conclusions that might be drawn from it, and also on it being described at a level that is understandable to the design engineer [4].

It is generally agreed that, given the right amount of quality data, computer algorithms are capable of obtaining accurate and informative rule bases, however in practice this is not consistently so. Comprehensive experimental descriptions can be time consuming and expensive, and therefore there is a need for exploiting the information in low quality data, including scarce, incomplete and/or imprecise sources of information [12]. Numerous studies have been published regarding the representation of uncertain empirical information, and the difference between different types of uncertainties [5,16–18,21,23,24]. According to [16], there are two main categories of uncertainty: stochastic uncertainty, that arises from random variability related to natural processes such as the heterogeneity of population or the fluctuations

* Corresponding author. Tel.: +34 98 51 81906.
E-mail addresses: luciano@uniovi.es (L. Sánchez), couso@uniovi.es (I. Couso), palaciosana@uniovi.es (A.M. Palacios), jlp324@psu.edu (J.L. Palacios).

0888-3270/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ymssp.2012.02.009

Table 1
SAT for aluminum with a temperature of -11 °C, according to different authors.

Author	Ice	Test	SAT-kPa
Loughborough and Hass [30]	Freezer ice	Pull	558
Stallabrass and Price [50]	Impact ice	Rotating instrumented beam	97
Itagaki [28]	Impact ice	Rotating rotor	27–157
Scavuzzo and Chu [47]	Impact ice	Shear window	90–290
Reich [40]	Freezer ice	Pull	896

of a quantity with time, and epistemic uncertainty, that arises from the incomplete or imprecise nature of available information. Stochastic uncertainty can be modeled with classical probability theory, however there are different theories for handling incomplete and imprecise information [15], which often appears in engineering problems.

The purpose of this study is therefore to define a methodology for obtaining expert knowledge, comprising fuzzy classification rules, from ill-known data. This methodology will be demonstrated on a real-world problem: discovering a set of linguistic rules describing the ice accretion strength of different materials used in helicopter rotor blades, for different ambient conditions.

This problem is out of the ordinary because neither the ambient conditions nor the experimental parameters of the essays can be precisely determined on all occasions; they may change during the realization of an essay, and sometimes they cannot be directly measured. For that matter, the properties of some materials cannot be reliably estimated. This last fact will be illustrated with the help of an example: In Table 1, some estimations of the shear adhesion strength (SAT) for Aluminum at -11 °C, found in the literature of the field, were collected. These values are much different among themselves, arguably as consequence of the impossibility of accurately determining the value of some ambient or experimental parameters.

Following with the example, imagine that a SAT of 100 kPa was measured in an experiment for which the initial temperature was -14 °C and the final temperature was -10 °C. In this case, it is not correct to write "SAT=100 kPa at a temperature of -12 °C" neither it is to state "SAT=100 kPa for temperatures between -14 and -10 °C". Our knowledge is restricted to the fact "SAT=100 kPa at an unknown temperature between -14 and -10 °C". Clearly, linguistic modeling techniques are well suited for expressing this kind of information.

In view of the above, the first part of this paper contains a description of the proposed modeling methodology and their assumptions, and the second part of this study describes a practical application of this methodology to rotor blade characterization. In Section 2 the use of fuzzy sets for describing the uncertainty in the data is explained. In Section 3 it will be explained how ill-known data is discretized into linguistic values. The selection of an informative discretization is addressed in Section 4, where a mutual information measure for fuzzy discretized data is proposed. In Section 5 two algorithms for finding fuzzy classification rules from imprecise data are described, and in Section 6 the demonstration problem is detailed. The paper is finished with some concluding remarks, in Section 7.

2. Fuzzy models of uncertainty in the data

The use of stochastic techniques for describing the numerical uncertainty in experimental data is prevalent among researchers and practitioners. However there may be cases where there are better alternatives. For example, in presence of coarse digital measurements (lack of significant digits), censored data or missing values, a probabilistic model is too restrictive. Interval-valued descriptions or other characterizations of the uncertainty, based on families of probability distributions, are to be preferred [31].

The use of a possibility distribution for describing partial ignorance about a value falls in the second of these groups. Fuzzy membership functions can be derived from possibility distributions, and interval-valued descriptions are particular cases of this representation. Moreover, the same description can also be used for summarizing conflicting data, as happens for instance when a set of measurements of the same physical magnitude is produced by different sensors. Here is a case in point: suppose that these conflicting measurements are

$$X = \{2, 1, 3, 3, 2, 2, 4\}. \tag{1}$$

Their average is 2.429. Nevertheless, using this value for describing the unknown value of the physical magnitude discards information that might be relevant: there are some items as low as 1, and others as high as 4. To gain additional insight about the importance of the dispersion of the values, it can be assumed that the set of items X is a sample of a larger population, whose mean is unknown. Confidence intervals for the value of this mean can be computed for different signification levels, and the knowledge about the unknown magnitude described by a list of nested confidence intervals. Following with the same example, if a Gaussian population is assumed, these confidence intervals are

$$\tilde{X}_\alpha = 2.429 \pm 0.9759 \cdot qt_6\left(1-\frac{\alpha}{2}\right), \tag{2}$$

where 0.9759 is the standard deviation and qt_6 is the quantile function for the t distribution. According to [7,8,46], this set of intervals contains the same information about the unknown variable that a possibility distribution defined by a fuzzy