Computational Materials Science 99 (2015) 316–325

Contents lists available at ScienceDirect

Computational Materials Science

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

Comprehensive mechanical property classification of vapor-grown carbon nanofiber/vinyl ester nanocomposites using support vector machines

O. Abuomar [a,*], S. Nouranian [a,1], R. King [a,b], T.M. Ricks [c], T.E. Lacy [c]

[a] Center for Advanced Vehicular Systems (CAVS), Mississippi State, MS 39762, USA
[b] Department of Electrical and Computer Engineering, Mississippi State University, Mississippi State, MS 39762, USA
[c] Department of Aerospace Engineering, Mississippi State University, MS 39762, USA

ARTICLE INFO

Article history:
Received 30 September 2014
Received in revised form 17 December 2014
Accepted 23 December 2014
Available online 23 January 2015

Keywords:
Support vector machines
Data mining
Knowledge discovery
Vapor-grown carbon nanofiber/vinyl ester
Confusion matrix
Cross validation

ABSTRACT

In the context of data mining and knowledge discovery, a large dataset of vapor-grown carbon nanofiber (VGCNF)/vinyl ester (VE) nanocomposites was thoroughly analyzed and classified using support vector machines (SVMs) into ten classes of desired mechanical properties. These classes are high true ultimate strength, high true yield strength, high engineering elastic modulus, high engineering ultimate strength, high flexural modulus, high flexural strength, high impact strength, high storage modulus, high loss modulus, and high tan delta. Resubstitution and 3-folds cross validation techniques were applied and different sets of confusion matrices were used to compare and analyze the classifier's resulting classification performance. The designed SVMs model is resourceful for materials scientists and engineers, because it can be used to qualitatively assess different nanocomposite mechanical responses associated with different combinations of the formulation, processing, and environmental conditions. In addition, the lead time required to develop VGCNF/VE nanocomposites for particular engineering application will be significantly reduced using the designed SVMs classifier. This work specifically present a framework for a fast and reliable classification of a large material dataset with respect to desired mechanical properties, and can be used for all materials within the context of materials science and engineering.

1. Introduction

The support vector machines (SVMs) methodology [1] is considered a widely used technique by the artificial intelligence community. It can be employed to design classifiers using datasets of different sizes and dimensions and from different knowledge domains. SVMs can be used for both supervised and unsupervised learning methodologies [1]. Supervised learning can be implemented using a relatively small number of data vectors (points). However, some prior knowledge of the problem is needed to assist the SVMs model in generalizing to unknown data vectors and thus predicting the correct quantity. Unsupervised learning typically requires a large number of data vectors within a particular dataset to adequately discover the relationship between the dataset's dimensions and to model the problem appropriately without over-training (over-fitting) the model [1]. The development of SVMs involves theory first, then implementation and experiments take place whereas other classifiers, like ANN follow a heuristic path, with applications and extensive experimentation preceding theory [1]. A significant advantage of SVMs is that other classifiers like ANN can suffer from multiple local minima whereas the solution to an SVM is global and unique. Two more advantages of SVMs are that that have a simple geometric interpretation and give a sparse solution. Unlike other techniques, the computational complexity of SVMs does not depend on the dimensionality of the input space. In addition, ANNs for example use empirical risk minimization, whereas SVMs use structural risk minimization [1]. The reason that SVMs often outperform other classifiers in practice, especially ANNs, is that they are less prone to overfitting [1].

SVMs classifiers generally perform poorly on highly unbalanced datasets because they are designed to generalize from sample data and output the simplest hypothesis that best fits the data, based on the principle of Occam's razor. This principle is embedded in the inductive bias of many machine learning algorithms including decision trees, which favor shorter trees over longer ones [2]. With

* Corresponding author at: Center for Advanced Vehicular Systems (CAVS), 200 Research Blvd., Mississippi State University, Mississippi State, MS 39762, USA. Tel.: +1 662 325 5576.
E-mail address: abuomar@cavs.msstate.edu (O. Abuomar).
[1] Present address: Department of Chemical Engineering, The University of Mississippi, University, MS 38677, USA.

http://dx.doi.org/10.1016/j.commatsci.2014.12.029
0927-0256/

unbalanced data, the simplest hypothesis is often the one that classifies almost all instances as negative or all instances as positive. In addition, highly unbalanced datasets will have a negative effect on the designed classifier by making it too sensitive to noise and more prone to learn an erroneous hypothesis [2].

These problems are encountered on highly unbalanced datasets. According of what has mentioned in literature, an imbalance of 100 samples of one class to 1 sample of another class exists in fraud detection domains, even approaching 100,000 samples of one class to 1 sample of another class in other applications [2].

SVMs can also classify linearly and nonlinearly separable data into two or more classes [1]. SVMs have recently been employed as an application of data mining and knowledge discovery techniques in the context of materials science and engineering to facilitate the discovery of new knowledge [3–6]. For example, materials scientists can use SVMs to interpret experimental data. This activity can not only accelerate research, but also aid in the development of new materials with desired mechanical properties. In short, data mining approaches are being fueled by dynamic growth in the information technology sector and is driving the interest in SVMs, machine learning, information retrieval, and other knowledge representation in different engineering disciplines [7]. Abuomar et al. [8] applied an artificial neural network (ANN) technique to a dataset associated with the viscoelastic response of a vapor-grown carbon nanofiber (VGCNF)/vinyl ester (VE) nanocomposite material system. The ANN was trained using the resubstitution method and the 3-fold cross validation (CV) technique to predict the responses (*i.e.* storage modulus, loss modulus, and tan delta) with the minimal mean square error [8]. Nunes et al. [9] evaluated the efficiency and accuracy of artificial intelligence techniques to classify ultrasound signals, raw data and feature selection methods, background echo and backscattered signals acquired at frequencies of 4 and 5 MHz to characterize the microstructural kinetics of phase transformations on a Nb-base alloy, thermally aged at 650 and 950 °C for 10, 100 and 200 h. Papa et al. [10] implemented SVMs, Bayesian and Optimum-Path Forest (OPF) based classifiers, and also the Otsu's method for automatic characterization of particles in metallographic images. De Albuquerque et al. [11] presented an ANN model to automatically segment and quantify material phases from SEM metallographic images and then the results were compared to a commercial software used for quantifying material phases from metallographic images. The results of the new ANN model were precise, reliable and more accurate and faster than the commercial software [11]. In addition, De Albuquerque et al. [12,13] presented a comparative analysis between backpropagation multilayer perceptron and self-organizing maps (SOMs) topologies applied to segment microstructures from metallographic images as well as they applied an ANN computational solution to segment and quantify the constituents of metallic materials from images. As another application of radiographic images segmentation task, an ANN model was employed to evaluate the delamination in laminate plates due to drilling operation [14]. Roberts et al. [15] presented a model that classifies different materials based on their microstructure. Based on microstructural characteristics such as Haralick variables [16], the Euler parameter [15], and the fractal dimension [15], the designed SVMs classifier identifies the appropriate class of given material sample [15]. Swaddiwudhipong et al. [17] utilized and implemented least squares support vector machines (LS-SVMs) [18]. Four LS-SVMs models that simulate the relationship between the elasto-plastic material properties and indentation load–displacement characteristics were designed; it was determined that the LS-SVMs technique was robust in determining the power hardening parameters given the fact that no iterative approaches were used [18]. Hu et al. [19] used knowledge discovery to resolve the problem of materials science image data sharing. Different annotations for non-structured materials science data were developed that utilize a complete ontology-based approach with the aid of semantic web technologies [19]. Sabin et al. [20] used a Gaussian process framework as a statistical technique to predict the output (*i.e.* the mean logarithm of grain size (D)) based on a probability distribution function over the training dataset. This framework was trained based on the available input variables (*i.e.* Strain, temperature (°C), and annealing time (s)) and tested to make the corresponding predictions and estimations.

In this work, a specific class of advanced engineering materials was studied, *i.e.*, polymer nanocomposites [21]. These materials have multifunctional properties and are extensively being used for fuel cell, aerospace, automotive, catalysis, biomedical, and other engineering applications. For example, nano-enhanced polymer composites meet the requirements of improved stiffness properties and energy absorption characteristics in automotive structural applications [22]. They have been the subject of extensive research recently [23,24]. Abuomar et al. [25] applied data mining and knowledge discovery techniques in order to analyze a thermosetting viscoelastic VGCNF/VE nanocomposite material system [26–29]. These techniques included SOMs, which are sometimes referred to as Kohonen maps [30,31] and fuzzy C-means (FCM) clustering [32,33]. The SOMs were used to determine the VGCNF/VE nanocomposite systems that produce the same storage and loss modulus for the lowest cost [25]. The FCM algorithm has also been used to discover some of the mechanical and physical patterns in the VGCNF/VE nanocomposite behavior after using the principal component analysis (PCA) technique to reduce the number of dimensions in the original dataset [34].

The current knowledge of the influence of formulation, processing, and environmental factors on the mechanical behavior of VGCNF/VE nanocomposites has been expanded in this study. This was accomplished by including a wider range of measured mechanical properties, *i.e.*, viscoelastic [26], compressive and tensile [35], flexural [36], and impact strengths [29]. Abuomar et al. [37] implemented this idea initially for a smaller dataset, where SVMs technique was used to analyze and classify a VGCNF/VE dataset, including viscoelastic data, compressive and tensile property data, and flexural property data into three classes of desired mechanical properties, i.e., high storage modulus, high true ultimate strength, and high flexural modulus. This new study, however, provides a more general and comprehensive insight into the mechanical behavior of VGCNF/VE nanocomposites for data mining purposes by including the VGCNF/VE impact strengths data as well as classifying and analyzing ten desired mechanical properties instead of three. The application of data mining and knowledge discovery techniques to a comprehensive dataset of mechanical responses of polymer nanocomposites is unprecedented and novel. The SVMs technique is used in this work to separate the new VGCNF/VE nanocomposite test data into ten different desired mechanical property classes. Thus, an unknown VGCNF/VE sample whose configuration is not represented by the current dataset can be easily identified, analyzed, and classified into its corresponding VGCNF/VE mechanical class without the need to conduct expensive and time-consuming experiments. Materials scientists and engineers can use the results of this study as a guideline to efficiently design or optimize a material system for a certain engineering application. The lead time required to develop a new material system for a specific engineering application can be significantly reduced using this study's fast and reliable qualitative assessment.

2. Materials and methods

The majority of data samples used in this work were generated using various statistics-based designed experiments, utilizing a general mixed level full factorial and central composite designs