Computers & Geosciences 39 (2012) 64–76

Contents lists available at ScienceDirect

Computers & Geosciences

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

Support vector regression to predict porosity and permeability: Effect of sample size

A.F. Al-Anazi, I.D. Gates *

Department of Chemical and Petroleum Engineering, Schulich School of Engineering, University of Calgary, 2500 University Drive NW, Calgary, Alberta, Canada T2N 1N4

ARTICLE INFO

Article history:
Received 3 December 2010
Received in revised form
21 May 2011
Accepted 21 June 2011
Available online 2 July 2011

Keywords:
Porosity prediction
Permeability prediction
Small sample size
Loss function
Support vector machines
Artificial neural networks
Log data
Core data

ABSTRACT

Porosity and permeability are key petrophysical parameters obtained from laboratory core analysis. Cores, obtained from drilled wells, are often few in number for most oil and gas fields. Porosity and permeability correlations based on conventional techniques such as linear regression or neural networks trained with core and geophysical logs suffer poor generalization to wells with only geophysical logs. The generalization problem of correlation models often becomes pronounced when the training sample size is small. This is attributed to the underlying assumption that conventional techniques employing the empirical risk minimization (ERM) inductive principle converge asymptotically to the true risk values as the number of samples increases. In small sample size estimation problems, the available training samples must span the complexity of the parameter space so that the model is able both to match the available training samples reasonably well and to generalize to new data. This is achieved using the structural risk minimization (SRM) inductive principle by matching the capability of the model to the available training data. One method that uses SRM is support vector regression (SVR) network. In this research, the capability of SVR to predict porosity and permeability in a heterogeneous sandstone reservoir under the effect of small sample size is evaluated. Particularly, the impact of Vapnik's ε-insensitivity loss function and least-modulus loss function on generalization performance was empirically investigated. The results are compared to the multilayer perception (MLP) neural network, a widely used regression method, which operates under the ERM principle. The mean square error and correlation coefficients were used to measure the quality of predictions. The results demonstrate that SVR yields consistently better predictions of the porosity and permeability with small sample size than the MLP method. Also, the performance of SVR depends on both kernel function type and loss functions used.

1. Introduction

One of the most important tasks in modeling geoscience data is development of robust and accurate correlation models calibrated to small sample size problems. For permeability estimation, correlation models between porosity and permeability are often built based on the relationship between geophysical logs and core-measured porosity and permeability. In typical practice, core plugs are extracted from a few key wells during drilling whereas geophysical logs are run for all wells in the oil/gas field. The limited number of core plug data poses a challenging problem to existing empirical techniques that employ the empirical risk minimization (ERM) principle such as linear regression and neural networks. Statistical learning theory (SLT) shows that these techniques can be safely used as a measure of the true risk when the sample size is sufficiently large. SLT calls for introducing a structure to match the complexity of the predictive learning technique to the available training data. The structure is composed of elements of increasing complexity that need to be chosen to imitate the response of the learning problem using a limited number of data (Cherkassky and Mulier, 2007).

Artificial neural networks (ANNs) is a universal approximator that is capable of approximating any nonlinear function to any degree of accuracy provided that there are a sufficient number of neurons in the network (Hornik et al., 1989). The structure implemented by ANN may be captured by the number and size of the hidden layers, which are controlled explicitly by the user. This structure may also lead to an overfitting problem during learning, particularly in the presence of a small sample size, which potentially yields a poor generalization model. Although ANN has shown some successful applications to porosity and permeability (Helle and Ursin, 2001; Huang et al., 2001), the underlying learning algorithm has been developed for learning problems of large sample sizes. Hence, for a given small sample size, extensive experiments with several different learning techniques are required to devise an accurate ANN-based regression model (Kaviani et al., 2008).

* Corresponding author. Tel.: +1 403 220 5752; fax: +1 403 284 4852.
E-mail address: ian.gates@ucalgary.ca (I.D. Gates).

0098-3004/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cageo.2011.06.011

Nomenclature

b	bias constant
MLP	multilayer perceptron neural networks
c	regularization parameter
DT	sonic porosity log
f	an unknown function
GR	gamma ray
h	Vapnik-Chervonenkis dimension
ILD	deep inductive laterolog
K	kernel function/permeability
L	Lagrangian equation for a dual programming problem or loss function
NPHI	neutron porosity log
RBF	radial basis function
RHOB	bulk density log
R_{emp}	empirical risk
R	structural risk
SVR	support vector regression
SVM	support vector machines
x	input variable
y	output variable
$\hat{y}$	estimated output value

Greek symbols

α,α^*	Lagrangian multiplier to be determined
ε	error accuracy
η,η^*	Lagrangian multipliers
κ, ϑ	sigmoid function parameters
ξ,ξ^*	slack variables
w	weight vector

Subscripts and superscripts

k,l	indices
N	number of samples
n	input space dimension

Recently, support vector machines (SVMs) have been gaining popularity in regression and classification due to their excellent generalization performance. The SVM approach has been successfully applied to several different applications such as face recognition, object detection, handwriting recognition, text detection, speech recognition and prediction, lithology identification, and porosity and permeability determination from log data (Li et al., 2000; Lu et al., 2001; Choisy and Belaid, 2001; Gao et al., 2001; Kim et al., 2001; Ma et al., 2001; Van Gestel et al., 2001; Al-Anazi and Gates, 2010a, 2010b, 2010c, 2010d). The SVM formulation is based on the structural risk minimization (SRM) inductive principle where the empirical risk minimization (ERM) inductive principle and the Vapnik and Chervonenkis (VC) confidence interval are simultaneously minimized (Vapnik and Chervonenkis, 1974; Vapnik, 1982, 1995). The SRM principle introduces a structure where each element of the structure is indexed by a measure of complexity defined by the margin size between two classes in a classification learning problem and by an insensitivity zone size in a regression problem (Cherkassky and Mulier, 2007). The SVM optimization formulation implicitly matches a suitable structure of certain complexity to the available small size sample. This type of structure is controlled independently of the dimension of the problem, which is an advantage over classical learning techniques. In regression applications, the empirical error (the training error) is minimized by Vapnik's ε-insensitivity loss function rather than the quadratic error and absolute-value loss functions used in neural networks and classical regression methods. To generalize to nonlinear regression, kernel functions are used to project the input space into a feature space where a linear or nearly linear regression hypersurface results. A regularization term is used to determine a trade-off between the training error and the VC confidence term. The learning problem is formulated as a constrained convex optimization problem whose solution is used to construct the mapping function between the empirical input and the output data (Kecman, 2005).

Previously, our research demonstrated the generalization capability of SVM in lithology classification and porosity and permeability predictions with sensitivity analysis of kernel function types and SVM regularization parameters. In this research, however, the sensitivity of the SVM-based prediction of porosity and permeability to sample size and empirical loss functions are examined and compared to a multilayer perceptron network. The empirical evaluation of the generalization performance under small sample setting is conducted for two loss functions: first, the ε-insensitivity loss function, and second, the least-modulus (or absolute value) loss function.

2. Background

2.1. Multilayer perceptron neural network model

ANN has been frequently used as an intelligent regression technique in petrophysical properties estimation (Rogers et al., 1995; Huang et al., 1996, 2001; Fung et al., 1997; Helle and Ursin, 2001; Helle and Bhatt, 2002). Backpropagation multilayer perceptron neural networks are considered to be universal approximators: it has been mathematically proven that a network with a hidden layer of an arbitrary large number of nonlinear neurons can approximate any continuous nonlinear function over a compact subset to any desirable accuracy (Hornik et al., 1989). In our study, a backpropagation conjugate gradient learning algorithm is used to train the multilayer perception (MLP) network by minimizing a squared residual cost function. During training, input patterns are propagated forward through hidden layers toward the output while the output error signal is backpropagated toward the input layer to adjust the weights of the hidden and output layers in order to approximate the target hypersurface. One well-known problem with such a training algorithm is that it can get trapped in local minima because the algorithm performance is sensitive to the selection of the starting weight values (Hastie et al., 2001). To overcome this issue, the initial range of initial weight values is chosen by the Nguyen–Widrow algorithm (Nguyen and Widrow, 1990). The conjugate gradient algorithm is used to optimize values of the weights. Optimization is done several times with different starting values of the weight values (chosen randomly) to improve the chances of converging to the global solution. The MLP network can also overfit the training data leading to poor generalization to new data. In this study, a cross-validation technique was used to terminate training to select the best model (Sherrod, 2009).