Fuel Processing Technology 129 (2015) 120–129

Contents lists available at ScienceDirect

Fuel Processing Technology

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

Proximate analysis based prediction of gross calorific value of coals: A comparison of support vector machine, alternating conditional expectation and artificial neural network

Qihong Feng [a], Jiyuan Zhang [a,*], Xianmin Zhang [a], Shengming Wen [b]

[a] School of Petroleum Engineering, China University of Petroleum (East China), Qingdao, China
[b] PetroChina Coalbed Methane Company Limited, Beijing, China

ARTICLE INFO

Article history:
Received 17 April 2014
Received in revised form 1 September 2014
Accepted 1 September 2014
Available online 19 September 2014

Keywords:
Gross calorific value
Alternating conditional expectation
Artificial neural network
Support vector machine
Proximate analysis

ABSTRACT

The gross calorific value (GCV) of coal is important in both the direct use and conversion into other fuel forms of coals. The measurement of GCV usually requires sophisticated bomb calorimetric experimental apparatus and expertise, whereas proximate analysis is much cheaper, easier and faster to conduct. This paper presents the application of three regression models, i.e., support vector machine (SVM), alternating conditional expectation (ACE) and back propagation neural network (BPNN) to predict the GCV of coals based on proximate analysis information. Analytical data of 76 Chinese coal samples, with a large variation in rank were acquired and used as input into these models. The modeling results show that: 1) all three methods are generally capable of tracking the variation trend of GCV with the proximate analysis parameters; 2) SVM performs the best in terms of generalization capability among the models investigated; 3) BPNN has the potential to outperform SVM in the training stage and ACE in both training and testing stages; however, its prediction accuracy is dramatically affected by the model parameters including hidden neuron number, learning rate and initial weights; 4) ACE performs slightly better with respect to the generalization capability than does BPNN, on an averaged scale.

1. Introduction

Unconventional fuel resources, such as coalbed methane [1–3] and shale gas [4,5] are increasingly adding to the world energy supply. Coal as a conventional fossil fuel, however, still plays a vital role in nowadays industrial fields like electricity generation, cement making and conversion to coke for the smelting of iron ore [6]. The calorific value of coal is of great importance in both its direct use and the conversion to other useful forms of fuel [7]. The calorific value is usually expressed as gross calorific value (GCV) or higher heating value (HHV). Estimating the GCV from the elemental composition of fuel is one of the basic steps in performance modeling and calculations on thermal systems [8].

To date, a variety of multivariate parametric regression (MPR) models [7,9–20] have been proposed for predicting the GCV of coal based on proximate and/or ultimate analysis information. These models, according to the input parameters, can be classified into three groups: proximate analysis based [9–16], ultimate analysis based [15,17,18], and hybrid analyses based [7,14,19,20] models (Table 1). The proximate analysis based models incorporate one or several compositions, i.e., fixed carbon (FC), ash (A), moisture (M), and volatile matter (VM) measured from proximate analysis whereas the ultimate analysis based models use the element(s) (i.e., C, H, O, N, and S) as input. The hybrid analysis based models use a combination of both analysis information. A majority of these studies found that GCV models incorporating ultimate analysis information generally have higher prediction accuracy than those incorporating only proximate analysis parameters. Essentially, ultimate analysis based models are superior to proximate analysis based ones in terms of accuracy. Ultimate analysis, however, is usually more expensive and tedious than proximate analysis and even bomb calorimeter experimental method. By contrast, proximate analysis is often easier, cheaper and faster to carry out; therefore, developing its correlation with GCV might be of more significance and should provide influential contribution to this area [9]. A disadvantage of applying MPR is that it is subject to a priori assumption of function form before regression. The function forms may lead to inaccurate or even absurd results [21,22] if assumed improperly.

In addition to MPR, the artificial neural networks (ANNs) [23–30] were also applied for predicting GCV based on proximate and/or ultimate analysis information [14,16,18,23]. Among the ANNs used in this area, the back propagation neural networks (BPNNs) are most widely utilized. Mesroghli et al. [14] demonstrate that ANNs are not better or much different from MPR in terms of error assessment whereas other researchers [16,18] show that the ANN models have higher accuracy than that by the MPR method. Despite the relatively strong power in exploring the correlation between dependent and independent variables, ANNs often suffer from a deficiency that their performance is

* Corresponding author. Tel.: +86 150 64247179.
E-mail address: zhjy221@126.com (J. Zhang).

http://dx.doi.org/10.1016/j.fuproc.2014.09.001
0378-3820/

Table 1
Summary of the MPR models for prediction of GCV of coals in previous literatures.

Model classification	Ref.	GCV (MJ/kg)	Basis	Country	R^2
Proximate analysis based					
	[9]	$0.836\ M^{-8.155}A^{-3.559}VM^{0.35}FC^{0.626}$	a.r.	Turkey	0.97
		$0.561\ M^{-6.137}VM^{0.381}FC^{0.666}$			0.97
		$33.078–0.72\ M + 0.012\ M^2 - 1.163\ M^3 - 0.324A^2$			0.97
	[10]	$4.183 \times 10^{-3} \times (82FC + kVM)$	daf	n.g	n.g
	[11]	$4.183 \times 10^{-3} \times (8000 + VM \times (70 - 1.65 \times VM))$	daf	n.g	n.g
	[12]	$4.184 \times 10^{-3} \times (9170 - 16VM - 60\ M(1 - 0.001\ M)$	n.g.	India	n.g
	[13]	$-0.03A - 0.11\ M + 0.33VM + 0.35FC$	a.r.	India	0.98
	[14]	$37.777 - 0.647\ M - 0.387A - 0.089VM$	a.r.	US	0.97
	[15]	$35.391 - 0.47\ M - 0.364A - 0.028\ V$	n.g.	Afghan	0.998
	[16]	$-3.57 + 0.31VM + 0.34FC$	a.r.	Slovenia	0.971
Ultimate analysis based					
	[15]	$-0.408 + 1.243H + 0.348C - 0.1\ N - 0.111O + 0.112S$	n.g.	Afghan	0.998
	[17]	$0.3278C + 1.419H + 0.09257S - 0.1379O + 0.637$	dmmf	n.g	n.g
	[18]	$64.62 - 0.262C - 0.579O - 0.46S$	daf	US	0.69
Hybrid analysis based					
	[7]	$10^{-3} \times (198.11C + 620.31H + 80.93S + 44.95A - 5159)$	dry	US	n.g.
	[14][a]	$-26.29 + 0.275A + 0.605C + 1.352H + 0.840\ N + 0.321S$	a.r.	US	0.99
		$6.971 + 0.269C + 0.195\ N - 0.061A - 0.251O_{ex} + 1.08H_{ex} - 0.21\ M$	a.r.	US	0.995
	[19]	$4.183 \times 10^{-3} \times (8781 + 19VM - 144O)$	n.g.	n.g	n.g
	[20][b]	$4.183 \times 10^{-3} \times (144.54C + 610.2H + 40.5S - (65.88 - 30.96O/(100 - A)))$	ua.	ua.	ua.

Note: a.r. = as received; daf = dry-ash-free; dmmf = dry mineral matter free; n.g. = not given.
[a] The subscript ex represents excluding hydrogen.

significantly influenced by a set of model parameters such as initial weights, hidden neuron number, and learning rate [30,31]. The reproducibility as well of results from ANN may be questioned partly due to the random initialization of the networks [32].

Support vector machine (SVM) has been demonstrated to be a powerful regression tool and applied in numerous prediction problems in varying fields [31,34,35]. The unique merit of using the structural risk minimization principle has made it especially effective in dealing with small samples than ANNs [31]. Another advantage of SVM over ANNs is that fewer parameters need to be optimized.

Alternating conditional expectation (ACE) is a nonparametric method for multivariate nonparametric regression that was first proposed by Breiman and Friedman [36], and later refined by Xue et al. [37]. Similar to SVM and ANN, a most distinguishing advantage of ACE over parametric regression is that it is totally data driven and does not require a prior assumption of functional forms [32,38,39]. The effect of each input on the response can be uncovered once optimal transforms are estimated. Unlike BPNN or SVM, ACE does not require, and therefore, is not influenced by model parameters, which eliminates the complicated work of parameter optimization. To date, neither SVM nor ACE has been used to predict GCV of coals. To bridge this gap, this paper intends to investigate and compare the prediction performance of three correlation tools, i.e., SVM, ACE and BPNN.

2. Basics of the regression models

2.1. SVM

SVM regression models are supervised machine learning models with associated learning algorithms. Given an n-dimension input vector $x = \{x_1, x_2, \ldots, x_n\}$ and its corresponding target vector $y = \{y_1, y_2, \ldots, y_n\}$, SVM aims to find a function $f(x)$ that has at most ε deviation from the actually obtained targets y_i for all training data and, at the same time, is as smooth as possible [40]. To realize this goal, SVM maps the input data in true space into a higher dimensional feature space via a nonlinear mapping function and linear regression is conducted in this space [41]. Thus, the problem of estimating the nonlinear regression in input space is converted into approximation of linear regression in the mapped higher feature space. SVM can be mathematically expressed as

$$f(x) = w\phi(x) + b \tag{1}$$

where $\phi(x)$ is a mapping function; w and b are a weight vector and bias value respectively which can be estimated by minimizing a regularized risk function (see Appendix A). The minimization problem is further converted to a dual optimization problem through the minimization of the Lagrange function, which involves the use of a selected kernel function.

2.2. ACE

The basic idea behind ACE is to estimate the transformations of a dependent and a set of independent variables that produce the maximum linear effect between the transformed independent and dependent variables [37]. Given the same variables as defined in Section 2.1, the ACE regression model can be written as:

$$\psi(y) = a + \sum_{i=1}^{n} \varphi_i(x_i) + \delta \tag{2}$$

where φ_i and ψ are the transform functions of the independent variable x_i and dependent variable y, respectively. To obtain the solutions of optimal transforms, a series of single function minimizations are conducted. Two basic mathematical operations, i.e., conditional expectations and iterative minimization are used during the minimization process [38]. The iteration process terminates when the error variance is satisfied, resulting in the optimal transforms referred to as $\varphi_i^*(x_i)$ and $\theta^*(y)$. It is noted that these functions appear as a series of point pairs $[\psi^*(y_i) - y_i]$ or $[\varphi_i^*(x_i) - x_i]$ in the computing algorithm rather than being explicitly available [42]. More detail concerning the mathematical formulations of ACE is presented in Appendix B.

2.3. BPNN

BPNNs are computational models consisting of a group of highly interconnected nodes (neurons). The most basic and commonly used