Available online at www.sciencedirect.com
SCIENCE DIRECT®
International Journal of Heat and Mass Transfer 49 (2006) 4352–4359
International Journal of HEAT and MASS TRANSFER
www.elsevier.com/locate/ijhmt

Heat transfer correlations by symbolic regression

Weihua Cai [a], Arturo Pacheco-Vega [b,*], Mihir Sen [a], K.T. Yang [a]

[a] Department of Aerospace and Mechanical Engineering, University of Notre Dame, Notre Dame, IN 46556, USA
[b] CIEP-Facultad de Ciencias Químicas, Universidad Autónoma de San Luis Potosí, San Luis Potosí, SLP 78210, México

Received 1 June 2005; received in revised form 22 April 2006
Available online 7 July 2006

Abstract

We describe a methodology that uses symbolic regression to extract correlations from heat transfer measurements by searching for both the form of the correlation equation and the constants in it that enable the closest fit to experimental data. For this purpose we use genetic programming modified by a penalty procedure to prevent large correlation functions. The advantage of using this technique is that no initial assumption on the form of the correlation is needed. The procedure is tested using two sets of published experimental data, one for a compact heat exchanger and the other for liquid flow in a circular pipe. In both situations, predictive errors from correlations found from symbolic regression are smaller than their published counterparts. A parametric analysis of the penalty function is also carried out.

Keywords: Heat transfer; Correlations; Symbolic regression; Genetic programming; Heat exchanger

1. Introduction

In the design, selection and control of thermal components for industrial and commercial applications, it is necessary to predict their performance under specific conditions of operation. Though in theory this calculation can be carried out from first principles by formulating the governing equations, complexities arising from factors like turbulence, temperature dependence of properties, and the geometry makes it difficult to achieve in practice. As a result, most calculations are based on experimental data. The information is compressed in the form of correlations from which the heat transfer coefficient can be obtained. Most commonly, correlations are developed in terms of dimensionless groups like the Nusselt, Reynolds and Prandtl numbers; sometimes for greater generality geometrical factors are also included. Assuming a functional relationship between the groups with a certain number of free constants, a regression analysis to minimize the error between predicted and experimental values is carried out to determine the appropriate values of the constants.

A disadvantage of this procedure is that predictive errors in the heat transfer rate are normally larger than the experimental uncertainties from which the correlation was generated. Assumptions such as using average transfer coefficients or constant property values [1] and the fact that the error minimization function may have more than one local minimum [2,3] are among the reasons for this loss of accuracy. Another source of error is the specific form of the correlation function assumed for the regression analysis. The functional form is selected on the basis of simplicity, compactness and common usage [4], but cannot be completely justified from first principles. There is usually not much physics behind the choice of the form. Although power laws are commonly used in heat transfer studies, a variety of other forms have also been used [5], though it is not obvious how the form should be chosen. As an example, for heat exchangers Pacheco-Vega et al. [6] have shown that *different* functional forms may predict performance with more or less similar accuracy. It would thus be

* Corresponding author. Tel.: +52 444 826 2440; fax: +52 444 826 2372.
E-mail address: apacheco@uaslp.mx (A. Pacheco-Vega).

0017-9310/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijheatmasstransfer.2006.04.029

Nomenclature

A_r	area ratio	Pr	Prandtl number
a_1, a_2	penalty parameters	p_c	probability of crossover
C	set of constants	p_m	probability of mutation
D	inner diameter of pipe	Q_f, Q_j, Q_{Nu}	penalized fitness function
F	set of operators	Re	Reynolds number
f	correlation function	S_j, S_{Nu}	variance of error
F_f, F_j, F_{Nu}	fitness function	**T**	terminal set
g	penalty function	x_j	variable
G	generation number		
G_{max}	maximum number of generations	*Greek symbol*	
j	Colburn j-factor	μ	dynamic viscosity of fluid
L	length of pipe		
L_f, L_j, L_{Nu}	size of correlation	*Subscripts and superscripts*	
M	population size	e	experimental value
N	number of experimental data sets	p	predicted value
N_v	number of variables	t	target value
Nu	Nusselt number	w	wall

advantageous to have an algorithmic way to determine the best correlation that fits experimental data without the need to assume its functional form.

The genetic algorithm (GA) [7,8] is an optimization technique based on stochastic, evolutionary principles that is used to find global extrema of a given function. Genetic programming (GP) [9] is a symbolic regression extension that works with a set of possible functions to find the best for a given set of data. Applications of GP to thermal engineering are scarce: the correlations obtained by Lee et al. [10] for critical heat flux for water flow in vertical round pipes and Pacheco-Vega et al. [11] for artificial heat-exchanger data are among the very few.

The aim of the present study is to describe a methodology based on GP to develop heat transfer correlations that can be used to predict the performance of thermal components. Since compact forms of the correlations are to be preferred, the standard procedure will be modified by a penalty function that weights against complicated forms. The procedure is described first. Then, two sets of published experimental data, one corresponding to heat transfer in compact heat exchangers and the other to heating and cooling of liquids in pipes, are used to demonstrate the capability of GP to find accurate correlations. The effect of the parameters of the penalty function on the results is also analyzed.

2. Genetic programming

2.1. Description

GP is a soft computing search technique in which computer codes, representing functions as parse trees, evolve as the search proceeds. The objective is to extremize a certain quantity called the *fitness function*. Developed originally to automatically generate computer programs, it has been used in a variety of applications, e.g., finance [12], electronic design [13], signal processing [14], and system identification [15], among others. GP is discussed in detail in the monograph by Koza [9].

Compared to the GA [7,8], in GP functions take the place of numbers in an attempt to find the best solution to a particular problem by genetically recombining a population of individuals that portray candidate solutions. This is achieved by using tree-structured representations of functions; an example of the function $5x\cos(5x+1)$ is shown in Fig. 1(a). Branch nodes may be operators with one or two arguments (such as sin, cos, exp, log, +, −, ∗, /, ∧), or may be Boolean (such as AND, OR, NOT) or conditional (IF–THEN–ELSE, etc.) operators. Leaf or terminal nodes, on the other hand, are the variables ($x_j, j = 1,\ldots,N_v$) in a particular problem, or constants to be

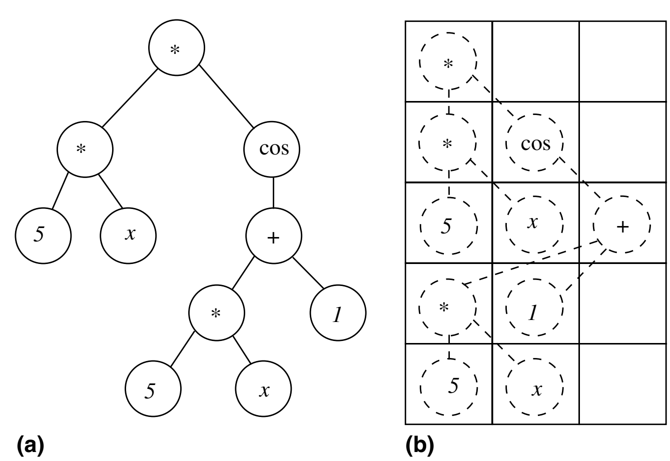

Fig. 1. Representation of function $5x\cos(5x+1)$ as (a) parse tree, and (b) array.