Computers & Geosciences 37 (2011) 1318–1323

Contents lists available at ScienceDirect

Computers & Geosciences

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

Function identification for the intrinsic strength and elastic properties of granitic rocks via genetic programming (GP)

Murat Karakus*

School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide, SA 5005, Australia

ARTICLE INFO

Article history:
Received 11 May 2010
Received in revised form
4 September 2010
Accepted 8 September 2010
Available online 23 November 2010

Keywords:
Genetic programming (GP)
Symbolic regression (SR)
Elasticity modulus
Compressive strength
Tensile strength
Granitic rocks

ABSTRACT

Symbolic Regression (SR) analysis, employing a genetic programming (GP) approach, was used to analyse laboratory strength and elasticity modulus data for some granitic rocks from selected regions in Turkey. Total porosity (n), sonic velocity (vp), point load index (Is) and Schmidt Hammer values (SH) for test specimens were used to develop relations between these index tests and uniaxial compressive strength (σ_c), tensile strength (σ_t) and elasticity modulus (E). Three GP models were developed. Each GP model was run more than 50 times to optimise the GP functions. Results from the GP functions were compared with the measured data set and it was found that simple functions may not be adequate in explaining strength relations with index properties. The results also indicated that GP is a potential tool for identifying the key and optimal variables (terminals) for building functions for predicting the elasticity modulus and the strength of granitic rocks.

1. Introduction

Strength, which is the stress at which a material ruptures or fails, is an essential parameter in the design and construction of underground structures in rock masses. Uniaxial compressive, tensile, shear and flexural strengths are commonly used parameters in rock engineering design. In addition, modulus of elasticity is directly related to the strength of rocks. There have been many attempts to understand the relation between strength and other properties of rocks such as Point Load Index, Schmidt Hammer Rebound Values, Ultrasonic Velocity and moisture content (Karakus et al., 2005; Karakus and Tutmez, 2006; Katz et al., 2000; Yavuz et al., 2006). As an example of soft computing applications, Grima and Babuška (1999) developed a fuzzy model to estimate unconfined compressive strength (UCS) of rocks and they concluded that fuzzy modelling has the potential of solving multi-variable engineering geological systems. Some of the recent research on estimating UCS using Multiple Regression (MR), artificial neural network (ANN) and ANFIS models was carried out by Yilmaz and Yüksek (2008, 2009). A higher prediction performance of ANFIS over MR and ANN models was reported by Yilmaz and Yüksek (2009). Sonmez et al. (2004) applied Fuzzy Logic in indirect determination of compressive strength of Ankara Agglomerates and they concluded that the developed model using fuzzy logic provided high performance prediction capacity for Ankara Agglomerates. Majdi and Beiki (2010) used Genetic Algorithms (GA) in design and optimising the Back Propagation Neural Network (BPNN) structure and applied the GA–ANN to predict the modulus of deformation of rock masses. In general, the models developed are successful in predicting the mechanical properties of rocks with index properties. However, there have been few attempts found in the literature on the identification of the parameters and functions which play a vital role in defining strength properties of rocks with regard to the aforementioned rock properties.

Symbolic regression (SR) defines a process of finding a function that fits a given finite data set as well as providing a method of function identification (Koza, 1992); in other words, a measured data set is fitted to an appropriate mathematical formula. Determination or identification of key variables and variable combinations, providing comprehension of developed models, are among the benefits of symbolic regression analysis. In this study, SR analysis was conducted using a GP approach. GP is well suited to geotechnical problems, as GP attempts to find the key variables for a problem and establishes mathematical expressions to explain the relation between the variables. Therefore, use of GP for optimum solution and function identification of geomechanical problems is increasing recently. For example, Johari et al. (2006) have successfully applied GP for the prediction of the soil–water characteristic curve. Baykasoğlu et al. (2008) applied multi-expression programming (MEP), gene expression programming (GEP) and linear genetic programming (LGP) to estimate compressive and tensile strength of limestone for the first time, with good predictions. Javadi et al. (2006) introduced a new technique based on genetic

* Tel.: +61 8 8303 6471; fax: +61 8 8303 4359.
E-mail address: mkarakus@civeng.adelaide.edu.au

0098-3004/$ - see front matter
doi:10.1016/j.cageo.2010.09.002

programming (GP), for the determination of liquefaction-induced lateral spreading. They concluded that the GP model was able to define the relations between liquefaction-induced lateral displacements with high accuracy. Subsidence due to underground mining operations was investigated for the first time by Shuhua et al. (2006) using a GP approach. Shuhua et al. (2006) pointed out that GP has the ability to solve complex problems and also stressed the importance of training data and experience in choosing the main factors for a model. When dealing with multi-variable inputs, genetic programming requires a large amount of time to solve the problem (Shuhua et al., 2006). A similar study was carried out by Lia et al. (2007) using a fuzzy genetic programming method (FGPM) for solving ground movement patterns due to underground mining, and very good agreement between predicted and measured data was achieved. Another interesting Genetic Programming application was conducted by Cabalar and Cevik (2009) for the prediction of peak ground acceleration (PGA) using strong-ground-motion data from Turkey. In this research, they demonstrated a high correlation between PGA and predictions.

In the current study, the GP approach is described in Section 2. Selection of independent variables, functions and fitness rule are included in Section 3. In Section 4, goodness of the fit for predicting modulus of elasticity, compressive strength and tensile strength of granitic rocks are summarised and outcomes of the GP analysis are presented in three categories. Firstly, relations between index properties of granitic rocks and the modulus of elasticity are explained and GP functions for this model presented. Secondly, a GP model for predicting uniaxial compressive strength of granitic rocks is included. Thirdly, a GP model for estimating tensile strength of granitic rocks is given in Section 4.

2. Overview of genetic programming

In this section, the GP paradigm will be discussed and the essentials of GP will be highlighted briefly. Further concepts and terminology behind GP can be found from the inventor of this paradigm (Koza, 1992). However, it is advised that the genetic algorithm concept developed in 1975 by Holland (1992) and work from his student, Goldberg (1989), can also be visited for further insight.

2.1. Basic concepts of GP

Genetic programming is an extension of the conventional genetic algorithm (GA), generating novel solutions to complex problems, developed by Koza (1992). Unlike GA which uses a string of numbers to represent the solution, the GP automatically creates several computer programs (CP) with a parse tree structure to solve the problem at hand. Solving any problem with GP is merely searching for the best individual CP among several possible randomly generated computer programs. The generated CP is based on the Darwinian concept of survival and reproduction of the fittest, as well as appropriate mating of CPs. The problem will be solved using the Darwinian genetic operators such as reproduction, crossover and mutation. The population in GP is initialised with randomly generated CPs. These programs are composed of functions and terminals appropriate to the characteristics of the problem. If the functions and terminals selected are not appropriate for the problem, the desired solution cannot be achieved. Therefore, it is also important to have a deep understanding of the problem. A basic flow chart of the genetic programming paradigm is given in Fig. 1.

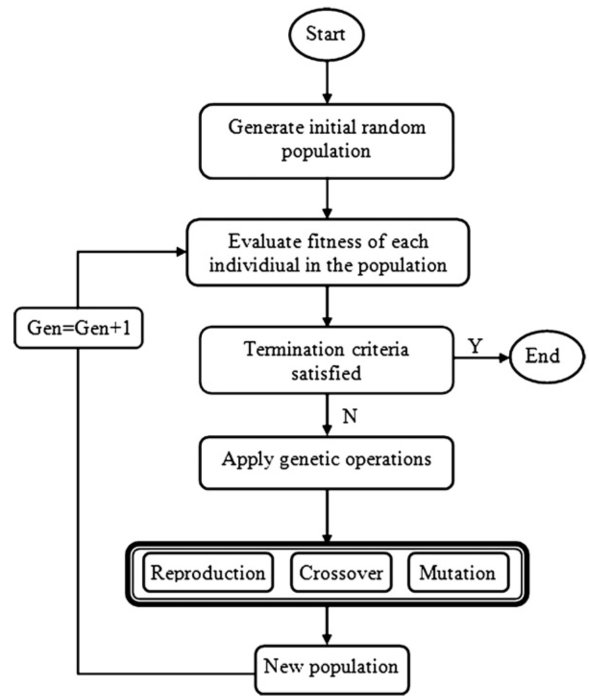

Fig. 1. A general flow chart for GP.

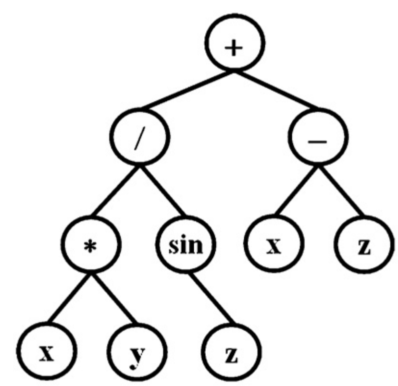

Fig. 2. A typical parse tree structure of the function $xy/(\sin(z))+(x-z)$.

As stated earlier, CP is composed of functions and terminals. The functions can be standard arithmetic operations, e.g. $f=\{/, \times, +, -, \sin, \cos, \log 2, power, \ldots\}$ and/or any mathematical functions, logical functions, as well as user-defined operators. Depending on the nature of the problem investigated, the computer program might be Boolean-valued, integer-valued, real-valued, complex-valued, vector-valued, symbolic-valued or multiple-valued (Koza, 1992). A typical program, representing the expression is shown in Fig. 2. In this example, the function set (F) is composed of multiplication, division, addition, subtraction and the sine function, $F=\{\times, /, +, -, \sin\}$. The terminal set ($T$) is composed of $N=3$ variable as $T=\{x, y, z\}$. The functions and terminals must fulfil two important parameters in order to solve the problem with an appropriate representation (Koza, 1992). These parameters are *closure property* and *sufficiency property*. The *closure property* includes protection of the function set and the terminal set against all possible argument values, e.g. protection of negative square root. *Sufficiency property* is the selection of the appropriate functions and terminals to the solve problem at hand.

2.2. Genetic operations

Genetic operations used in GP are composed of *reproduction*, *crossover* and *mutation*. *Reproduction* involves selecting, in proportion to fitness, a computer program from the current population of