YIELD STRENGTH MODELLING OF FORMED MATERIAL USING EVOLUTIONARY COMPUTATIONAL METHOD

L. Gusel, R. Rudolf

University of Maribor, Faculty of Mechanical Engineering, Smetanova 17, SI – 2000 Maribor, Slovenia, leog@uni-mb.si

Abstract

In this paper we propose an evolutionary computation approach for the modelling of yield strength in formed material. One of the most general evolutionary computation methods is genetic programming, which was used in our research. Genetic programming is an automated method for creating a working computer program from a problem's high-level statement. Genetic programming does this by genetically breeding a population of computer programs using the principles of Darwinian's natural selection and biologically inspired operations. During our research, material was cold formed by drawing using different process parameters and then determining yield strengths (dependent variable) of the specimens. On the basis of a training data set, various different genetic models for yield strength distribution were developed during simulated evolution. The accuracies of the best models were proved by a testing data set and comparing between the genetic and regression models. The research showed that very accurate genetic models can be developed by the proposed approach.

Key words: metal forming, yield strength, genetic programming, modelling

1. Introduction

Many modelling methods for predicting dependent output variables have been developed to reduce the costs of the experiments and computer computations. In most conventional deterministic modelling methods, such as regression analysis, a prediction model is determined in advance. Traditional methods often employed to solve real complex problems tend to inhibit elaborate explorations of the search space. They can be expensive and often result in sub-optimal solutions. Because of the pre-specified size and shape of the model, the latter is often incapable of capturing complex relationships between influencing parameters. It is very important that the independent input variables influence on the dependent output variables and, consequently, on the product quality has been already examined in the early stages of a metal forming process.

Evolutionary computation (*EC*) is generating considerable interest for solving real engineering problems. They are proving robust in delivering global optimal solutions and helping to resolve those limitations encountered in traditional methods. *EC* harnesses the power of natural selection to turn computers into optimization tools. This is very applicable to different problems in the manufacturing industry [1,2]. One of most important EC methods is genetic programming (GP) which is, similarly to a genetic algorithm, an evolutionary computation method for imitating biological evolution of living organisms. Several researches have been carried out using a neural network or genetic algorithms for modelling, thus forming process parameters [2-6], but only a few dealing with much more general genetic programming method [7-9]. In the GP method, the structure subject for adaptation is the population of hierarchically-organized computer programs. The GP method is most often used for complex system modelling, but it can also be effectively used for the modelling of a relatively simple system, such as the systems described in our paper.

This paper describes an evolutionary computation method approach for the modelling of yield strength. Experimental data obtained during the cold drawing processes under different conditions serves as an environment which, during simulated evolution, models for the yield strength have to be adapted to. Different values for effective strains and coefficients of friction were used as independent input variables (parameters), while yield strength was a dependent output variable. Then, GP was used for the evolutionary development of the models for yield strength prediction, on the basis of a training data set. No assumptions about the model's form and size were determined in advance, but were left to the evolutionary process. Finally, the prediction accuracy of the model was proved using the testing data set.

2. Method used

Genetic programming is probably the most general approach from among evolutionary computation methods in which the structures subject to adaptation are those hierarchically organized computer programs whose size and form dynamically change during simulated evolution.

The space for solutions in the GP method is the huge space of all possible computer programs consisting of components describing the problem area studied. The aim of GP is to find out the computer program that best solves the problem. Possible solutions in GP are all those possible computer programs that can be composed in a recursive manner from a set of function genes *F* and a set of terminal genes *T*. Function genes *F* are arithmetical functions, Boolean functions, relation functions, etc., while terminal genes are numerical constants, logical constants, variables, etc. [1]

The initial population is obtained by the creation of random computer programs consisting of available function genes from set *F* and available terminal genes from set *T*. Each program represents a random point in the searching space. The creation of an initial population is a blind random search for solutions within the huge space of possible solutions. The next step is the calculation of individual's adaptation to the environment (i.e., calculation of fitness for each computer program). Fitness is a guideline for modifying those structures undergoing adaptation. Computer programs change in GP, in particular during genetic operations regarding reproduction and

crossover. The reproduction operation gives a higher probability of selection to more successful organisms. They are copied unchanged into the next generation. The crossover operation ensures the exchange of genetic material between computer programs. The mutation operation increases the genetic diversity of a population. After finishing the first cycle, which includes creation of the initial population, calculation of fitness for each individual of the population, and genetic modification of the contents of the computer programs and formation of a new population, an iterative repetition of fitness calculation and genetic modification follows. After a certain number of generations the computer programs are usually much better adapted to the environment. The definition of the environment depends on the problem dealt with. The evolution is terminated when the termination criterion is fulfilled. This can be a prescribed number of generations or sufficient quality of the solution. Since evolution is a non-deterministic process, it does not end with a successful solution after each run (i.e., civilization). In order to obtain a successful solution, the problem must be processed over several independent runs. The number of runs required for the satisfactory solution depends on the difficulty of the problem.

3. Experimental work

The aim of the experimental work was to determine the influence of the effective strain ε_e and coefficient of friction μ during cold drawing on the change of yield strength of cold drawn copper alloy CuCrZr. This is a special copper alloy with 0,71%Cr, 0,05% Zr. It has high electrical and thermal conductivity, with excellent mechanical and physical properties at elevated temperatures.

Copper alloy rods were deformed by cold drawing under different conditions. The drawing speed was 20 m/min and the angle of drawing die was $\delta = 28°$. Copper alloy rods were drawn from an initial diameter of $+ D=20$ mm to six different diameters (i.e. six different effective strains). Three different lubricants with different coefficients of friction ($\mu=0,07$, $\mu=0,11$ and $\mu=0,16$) were used for the drawing process. In order to evaluate the yield strength, standard specimens for tensile tests were prepared from locations in the middles of the drawn rods. In this way we obtained 18 different experimental specimens. The yield strengths of all specimens were determined by providing three tensile tests for each specimen in order to provide reliable results. The results (average values) for yield strength are presented in Table 1. Experimental data serve as an environment which, during simulated evolution, models for impact toughness have to adapt.

Table 1. Experimental results

No.	Effective strain, ε_e	Coeficent of friction, μ	Yield strength, $Rp_{0,2}$ [N/mm^2]
initial spec.	/	/	409
1	0.10	0.07	498
2	0.21	0.07	513
3	0.32	0.07	523

No.	Effective strain, εe	Coeficent of friction, μ	Yield strength, $Rp_{0,2}$ [N/mm^2]
4	0.44	0.07	529
5	0.57	0.07	532
6	0.71	0.07	533
7	0.10	0.11	500
8	0.32	0.11	528
9	0.71	0.11	537
10	0.10	0.16	502
11	0.44	0.16	536
12	0.71	0.16	544
13	0.21	0.11	515
14	0.44	0.11	532
15	0.57	0.11	535
16	0.21	0.16	520
17	0.32	0.16	529
18	0.57	0.16	540

4. Yield strenght modelling by genetic programming

In the *GP* method the initial random population *P(t)* consists of randomly generated organisms which are, in fact, mathematical models. The variable *t* represents the generation time. Each organism in the initial population consists of the available function genes *F* and terminal genes *T*. Terminal genes are in fact independent variables: strain and coefficient of friction. Random floating-point numbers within the range [-10, 10] are added to the set of terminals to increase the genetic diversities of the organisms. Function genes *F* are basic arithmetical, exponential and cosine functions.

4.1. Evolutionary parameters

The absolute deviation *R* (*i, t*) of individual model *i* (organism) in generation time *t* for the *GP* approach, was introduced as the standard raw fitness measurement [1]:

$$R(i,t) = \sum_{j=1}^{n} |E(j) - P(i,j)|, \tag{1}$$

where: *E(j)* is the experimental value for measurement *j*, *P(i, j)* is the predicted value returned by the individual model *i* for measurement *j*, and *n* is the maximum number of measurements. The aim of the optimisation task is to find those models that equation (1) would give as having as low an absolute deviation as possible. However, because it is unnecessary that the smallest values of the above equation also means the

smallest percentage deviation of this model, the average absolute percentage deviation of all measurements for individual model *i* was defined as [1]:

$$\Delta(i) = \frac{R(i,t)}{|E(j)|\,n} \cdot 100\%$$

(2)

Equation (2) was not used as the fitness measurement for evaluating population, but only to find the best organism in population after completing the run.

In the *GP* method*,* reproduction, crossover, and mutation operations were used for altering the population *P(t)*. Evaluation and altering of the population *P(t)* were repeated until termination condition had been fulfilled. The termination condition was the prescribed maximum number of generation to be run. Reproduction, crossover, and mutation were used as genetic operations. For example, Fig. 1 shows the operation of the crossover. Two randomly selected parts of two parental organisms (in boldface) are interchanged. Thus two offspring are created [8]

$$\frac{\boldsymbol{r+z}}{5.6\,\mu} + r\mu \qquad\qquad \frac{(1-z\mu)}{5.6\,\mu} + r\mu$$

Parent 1   Crossover →   Child 1

$$r\,(\boldsymbol{1-z\,\mu}) \qquad\qquad\qquad r\,(r+z)$$

Parent 2   Child 2

Fig. 1. Crossover of two mathematical expressions

The evolutionary processes were controlled by the following evolutionary parameters: population size 1000, maximum number of generations to be run 50, probability of reproduction 0.15, probability of crossover 0.7, probability of mutation 0.15, maximum depth for initial random organisms 6, maximum depth of mutation 6, and maximum permissible depth of organisms after crossover 12.

4.2. Realisation of the evolutionary process

The modelling of yield strength was carried out by the special *GP* system (computer program), which comprises 49 modules, and was programmed in our laboratory. The GP system ensures repeated development of the individual civilization if necessary. This is very useful when it is necessary to repeat evolution of the civilization with a greater number of generations or when evolution is interrupted for any reason. Each individual GP run started with the training phase by the training data set shown in Table 1 (No.1 to No.12). The testing data set (Table 1: No.13 to No. 18) was not included within the training range. Each run lasted up to generation 30 when it was temporarily interrupted. If an average percentage deviation $\Delta(i)$ of at least one prediction model (organism) in the population was smaller than 5%, the evolution of the population continued up to generation 50, otherwise it was terminated. After each training phase, the accuracy of predicting the best models was tested using the testing data set. More a 500 independent runs were executed for modelling the yield strength.

The GP models in our research were originally developed as prefix LISP expression [10], and then converted into an infix notation.

5. Genetic models-Results and discussion

GP modelling was executed by two different genes function sets $F = (+,-,^*, /)$ and $F = (+,-,^*, /, \text{ZEXP})$. The best (the most accurate) model obtained with genes function set $F = (+,-,^*, /)$ is quite complicated and is written is LISP as:

(- (+ (+ (* (- 0.330041 ε) (% ε μ)) (+ (* (% (% μ 9.1439) (% ε -8.45165)) 6.49328) (* (% ε μ) (- (+ (* (- μ ε) (- ε ε)) (+ (* 6.70192 μ) (% μ 9.1439))) (% 8.87929 (- μ (- ε - 6.00992)))))))) 8.80115) (* (- (+ (- μ 7.33881) (* -1.68909 ε)) (- (* -7.79857 4.25639) (- (+ (% (+ (- ε 7.33881) (* 6.49328 ε)) (+ (* 7.39886 ε) (- 1.79688 μ))) (+ (% ε -8.45165) (+ (* 6.70192 μ) (% μ 9.1439)))) μ))) (+ (- -9.42629 ε) (- -8.5559 μ))))

The same model written as mathematical expression:

$$8.801 + \frac{0.33\varepsilon - \varepsilon^2}{\mu} - \frac{6\mu}{\varepsilon} + 6.811\varepsilon - \frac{8.879\varepsilon}{\mu^2 - \varepsilon\mu - 6\mu} + \frac{1}{1.796 + 7.398\varepsilon - \mu} \cdot$$

$$\left(\left(17.982 + \varepsilon + \mu\right)\left(-13.372\varepsilon^2 - 6.811(\mu - 1.597)(3.596 + \mu) + 195.543\varepsilon + 52.203\varepsilon\mu\right)\right) \quad (5.1)$$

The model (5.1) was generated in generation No. 47 and has the average percentage deviation of the training data Δ (i) = 0.18 % and that of the testing data $\Delta(i)$ = 0.,185%. Percentage deviation is in fact the percentage error between a single experimental value and the value predicted by the genetic model. Slightly better accuracy (Δ (i) = 0.175 %, and that of the testing data $\Delta(i)$ = 0.18%) of the GP model was obtained when the genes function set which includes the exponent function was used: $F = (+,-,^*, /, \text{ZEXP})$:

(- (- (* (+ (* 8.218 μ) (+ -6.93491 7.17839)) (+ (% (ZEXP μ) μ) (- 2.8788 -9.30543))) (+ (% (ZEXP 5.93875) (* (+ 3.20968 2.16393) (ZEXP ε))) (% (* (- μ -4.38819) (* 8.218 ε)) (% (ZEXP (+ ε 7.35216)) (ZEXP (ZEXP ε))))))) (- (+ (+ (- (% (ZEXP (ZEXP ε)) ε) (% (* (+ 3.20968 2.16393) (ZEXP ε)) (ZEXP ε))) μ) (% (* (+ ε 8.218) (% (* 8.218 μ) (* ε (% 6.56898 (ZEXP μ)))))) (- (+ (* 8.218 μ) (+ -6.93491 7.17839)) (* ε -5.45287)))) (+ (ZEXP (- (- 2.8788 -9.30543) (% 6.56898 1.10215))) (% (* (+ (* 8.218 μ) (+ -6.93491 7.17839)) (+ (% (ZEXP μ) μ) (- 2.8788 -9.30543))) (* (+ (* 8.218 μ) (+ -6.93491 7.17839)) (+ (% (ZEXP ε) μ) (- 2.8788 -9.30543)))))))

The upper GP model can be written as a mathematical expression:

$$513.098 - 70.615e^{-\varepsilon} + 0.023\frac{e^{e^{\varepsilon}}}{\varepsilon} + e^{-\varepsilon e^{\varepsilon}}\varepsilon +$$

$$e^{\varepsilon}\left(8.218 + \frac{0.243}{\mu} - \frac{10.281\mu + 1.251\varepsilon}{\varepsilon(0.234 + 5.452\varepsilon + 8.218\mu)} + \frac{1}{e^{\varepsilon} + 12.184\mu}\right) +$$

$$\mu\left(99.13 + \frac{12.184}{e^{\varepsilon} + 12.184\mu}\right)$$

(5.2)

The most accurate simple model with model depth 5, containing 7 function genes has average deviation $\Delta(i) = 1.67\%$ (testing data $\Delta(i) = 2.2\%$) is:

$$(+ (\% -1.85689\ \varepsilon)\ (* (-\varepsilon -9.72416)\ (- (+ \mu\ \varepsilon)\ (* 5.39679 -8.66756))))$$

or

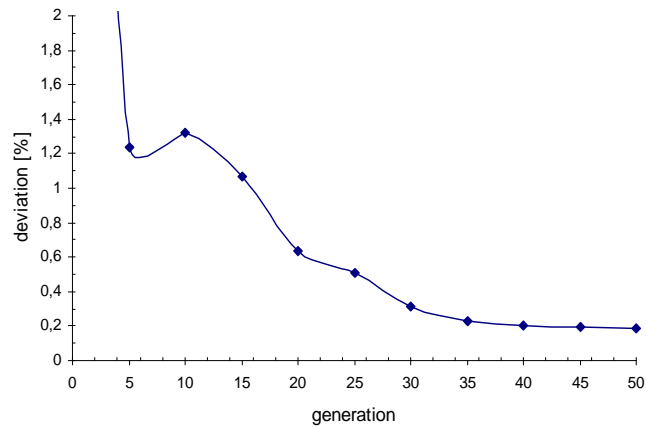$$46.76\ \mu - 454.79 - 1.85/\varepsilon \qquad\qquad (5.3)$$

Fig. 2. Percentage deviation curve between the best models regarding individual generation and experimental results (F = +, -, *, /, ZEXP)

Figure 2 shows the percentage deviation curve (Δ_i) between the best model regarding individual generation and experimental results when using the set of function genes F = {+, -, *, /, ZEXP}. It is obvious that in early generations the best models are not as precise as the models generated in late generations. The relatively slow improvement of the best models in later generations (after generation 35) is due to the unification trends of the population leading to the shortage of new genetic ideas.

Figure 3 shows the depth curve of the best models (generated with function genes F = {+, -, *, /}) in each generation. In generation 0 created randomly, the best models have a depth of 5. Then, from generations 2 to 9 the depth of the best models increases and reaches the value of 10. In the next generations (from 10 to 30) the depths

of models jumped from 10 to 8 and back to 10. Finally, after generation 30, the depth is constant and reaches a maximum depth of 10. The higher number of the model depth usually means higher complexity of the genetic model.
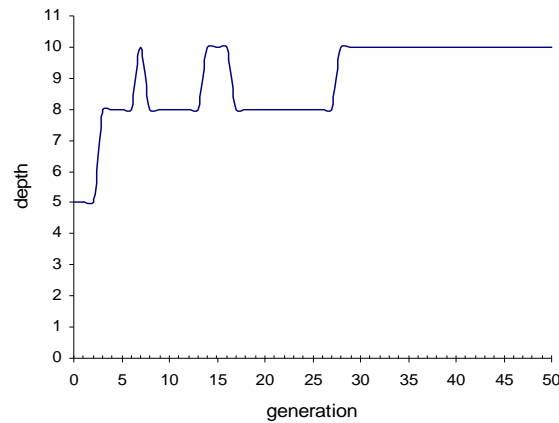

Fig. 3. The depth of the best models regarding individual generation

6. Modelling results obtained by regression analysis

A mathematical model for regression method was chosen according to [11]:

$$y(x) = b_0 + \sum_{i=1}^{n} b_i x_i + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} b_{ij} x_{ij} + \sum_{i=1}^{n}\sum_{j=i}^{n} b_{ij} x_{ij}^2$$

(6.1)

Where y is dependent variable, x_i, x_{ij} are independent variables, while b_0, b_i, b_{ij} are coefficients to be determined by using regression analysis. In our case, the dependent variable was yield strength ($Rp_{0,2}$), while effective strain ε_e and coefficient of friction μ were independent variables. Coefficients b_0, b_i and b_{ij} were determined by using the regression analysis computer program. By inserting the computed values of coefficients into the equation (6.1), the regression model for impact toughness can be presented as:

$$R_{p0,2} = 408.837 + 213.845\varepsilon + 104.061\mu - 176.166\varepsilon^2 - 386.503\mu^2 + 139.688\varepsilon\mu \quad (6.2)$$

Equation (6.2) represents a mathematical model of effective strain's influence and the coefficient of friction on yield strength for chosen material within experimental area. It has the average percentage deviation of the training data set $\Delta(i) = 0.28$ % and that of the testing data set $\Delta(i) = 0.32$ %.

When regression models are compared to genetic programming models, the first important difference is the complexities of the genetic models. Due to its evolutionary concept, genetic programming models are complex, with lots of genes, and the forms of

these models can be confusing. But the form of a GP model (5.3) is very simple. Of course, when it comes to the accuracies of different models, GP models show much greater accuracy then regression models.

7. Conclusion

The genetic development of models took place on the basis of experimental data. The experimental data in this research were in fact the environment to which the population of models had to be adapted as much as possible. The models presented are a result of the self-organization and stochastic processes taking place during simulated evolution, and not of human intelligence. The accuracies of the models developed during the training phase were also confirmed using testing data not included within the training range. Only two genetically developed models out of many successful solutions are presented here.

The accuracies of solutions obtained by *GP* depend on applied evolutionary parameters and also on the number of measurements and the accuracy of measurement. In general, more measurements supply more information to evolution which improves the structures of models. At the same time, the greater number of measurements causes time-consuming computer processing and the execution of experiments is very expensive and requires much time.

Because of the high precision regarding the models developed by the *GP* approach, with the proposed concept, the excessive number of experiments/simulations can be avoided, which leads to the reduction of the product development costs. The research showed that simple, and in the same time, very precise models are often hard to reach. This is due to the fact that evolution is a stochastic process, therefore, rationality in the development of the models is rare. However, in many metal-forming processes the accuracy of prediction is of vital importance, not the model complexity.

References

[1]   J.R. Koza, Genetic programming, The MIT Press, Massachusetts, 1992.

[2]   B. Tanguy, J. Besson, R. Piques, A. Pineau, Eng. Fract. Mech. 72 (2005) 49-72.

[3]   S. Sette, L. Boullart, Eng. Appl. Artif. Intell. 14 (2001) 727–736.

[4]   N.N. Zadeh, A. Darvizeh, A. Jamali, A. Moeini, J. Mater. Process. Technol. 164-165 (2005) 1561-1571.

[5]   H. Pierreval, C. Caux, J.L. Paris, F. Viguier, Computers & Industrial Engineering 44 (2003) 339-364.

[6]   V. Odugava, A. Tiwari, R. Roy, Applied Soft Computing 5 (2005) 281–299.

[7]   L. Gusel, M. Brezocnik, Comp. Mat. Sc. 37 (2006) 476–482.

[8]   M. Brezocnik, M. Kovacic, L. Gusel, Mater. Manuf. Processes 20 (2005) 497–508.

[9]   Y.S.Chang, S.P. Kwang, B.Y. Kim, Future Generation Computer Systems 21 (2005) 1117-1123.

[10]   T.M. Mitchell, Machine learning, McGraw-Hill, New York, 1997.

[11]   D.C. Montgomery, G.C. Runger, N.F. Hubele, Engineering statistics; Second Edition, John Wiley & Sons, New York, 2001.