A Symbolic Regression Approach for Modeling the Temperature of Metal Cutting Tool

Hossam Faris*

Abstract

Determining the temperature of the cutting tool is important due to its influence on the quality and cost of the production process. However, the process of modeling the temperature of the tool is a complex process; this due to the nonlinear relationship between the system variables. In this paper, a symbolic regression approach via genetic programming (GP) is used to model a cutting machine temperature and compared to other approaches which based on estimating the parameters of the nonlinear regressive curve of the cutting tool. The developed GP model shows an promising results compared to models developed based on parameter estimation such as Least Squares regression (LS) , Genetic Algorithms (GA) and Particle Swarm Optimization (PSO).

1 Introduction

Cutting tool is one of the most important machines in manufacturing and industries. The quality of the tool is vital and it depends on number of factors such as its temperature, cutting speed, feed, depth of cut and cutting tool wear. However, the process of modeling the temperature of the tool is hard because the relation between the large number of experiment data is very complex. Therefore, several researchers proposed applying different approaches based on empirical models for modeling this complex problem. Among all modeling approaches applied in engineering applications, heuristic approaches such as genetic algorithms, genetic programming, neural networks and swarm intelligence proved their competence in terms of accuracy, efficiency and performance [1, 2].

The problem of modeling the cutting machine presented in this paper was tackled previously twice. The first study was conducted by Jianin and others in [3]. Authors applied both LS and GA approaches and compared the results. Least squares is a standard common approach aims to minimize the sum of the squares of the errors (e.g; the difference between an observed value and the actual value provided by the model) made in the results of every single equation of overdetermined system. However, one of the problems of this method is that sometimes it fails in reaching the global minimum when the objective function is not differential or linear in model parameters.

To overcome the problems of the LS and enhance the performance of the model, the application of GA was proposed also in [3]. GA refers to a model introduced and investigated by John Holland (1975). GAs are heuristic search algorithms premised on the evolutionary ideas of natural selection and genetic [4]. The basic concept of GA is designed to simulate the processes in natural system necessary for evolution, specifically for those that follow the principle of survival of the fittest. In [3], results showed that GA enhanced the modeling quality of the cutting tool.

*H. Faris is Assistant Professor at Business Information Technology Department, The University of Jordan, Amman, Jordan. (hossam.faris@ju.edu.jo)

Later in [5], authors proposed the PSO algorithm for estimating the parameters of the empirical model of the same cutting tool. PSO is a population based computation technique introduced first by Kennedy and Eberthart in [6]. PSO is inspired by animal social behavior such as bird flock and fish school. A population of candidate solutions to the problem under consideration is used to probe the search space when improved positions of the solutions are discovered these positions will then guide the movements of the swarm. The advantage of PSO is that it can deal with large search space with few or no assumptions about the problem being optimized is required. PSO showed an improvement in the precision of the temperature nonlinear model compared with the LS and GA methods.

In this paper, we will investigate the application of Symbolic Regression via Genetic Programming in optimizing the parameters of an empirical model for the temperature of a cutting tool. Symbolic regression is a powerful modeling technique introduced by John Koza [7–10]. Symbolic regression develops models when the underlying physical relationships between input and output data can't be abstracted. Moreover, it provides a direct insight into the underlying process structures, as well as making accurate numeric predictions compared to the empirical values.

In order to justify the evaluation of the GP models developed in this work, Performance of the GP models will be compared with the three previously investigated approaches, Least Squares regression (LS), Genetic Algorithms (GA) and Particle Swarm Optimization (PSO).

This paper is structured as follows, the case study and the empirical data are presented in section 2. Genetic Programming is described in section 3. In section 4, the performance criteria used in order to compare the Cuckoo search with other techniques are listed. Finally, the results of performance and comparison are discussed in section 5.

2 Empirical Model and Experiment Data

The case study of this paper is the P05 horny alloy steel cutting machine. The machine is composed of a 38CrNi3Mo workpiece metal, tall stock and cutting tool (P05) as shown in Figure 1. The most commonly applied empirical model to calculate the temperature T_r for the P05 horny alloy steel cutting system is shown in equation 1.

$$T_r = k\,\alpha_p^x\,f^y\,v^z \tag{1}$$

where;
T_r is the output of temperature model,
k is a coefficient depending on the machined material,
α_p is the depth of cut (mm),
f is the cutting feed rate (mm/rev),
v is the cutting speed (m/min),
x, y, z are coefficients dependent on the type material of the cutting tool.

Based on nine experiments data measured and collected in [3] and shown in Table 1, previous studies aimed to estimate and optimize the coefficients x, y, z values using different approaches.

Authors in [3] used two approaches to estimate values of the coefficient of the empirical model. The first approach was the Least Square method and the resulted model is shown in equation 2. The second approach is GA with model equation 3. GA model showed better performance. Later on, authors in [5] used PSO approach to develop the model represented in equation 4. In this paper, LS,

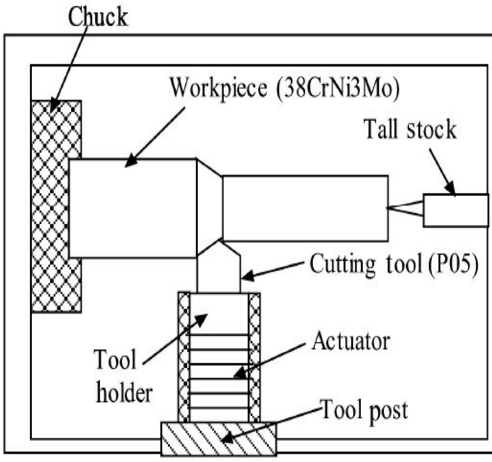

Figure 1. Metal cutting machine [5].

Table 1. Experimental data

No.	α_p (mm)	f (mm/ver)	v (m/min)	$T(i)$ (°C)
1	1	0.1	60	740.24
2	1.41	0.2	60	793.8
3	2	0.4	60	854.8
4	1.41	0.1	94.8	819.42
5	2	0.2	94.8	870
6	1	0.4	94.8	904.8
7	2	0.1	150	879.6
8	1	0.2	150	911.1
9	1.41	0.4	150	974.1

GA and PSO models will be evaluated using different measurements. The results of the evaluation will be compared with those obtained by the GP model.

$$T_{LS} = 467\,\alpha_p^{0.033} f^{0.083}\, v^{0.16} \tag{2}$$

$$T_{GA} = 471\,\alpha_p^{0.0301} f^{0.083}\, v^{0.159} \tag{3}$$

$$T_{PSO} = 470.3\,\alpha_p^{0.0324} f^{0.0828}\, v^{0.159} \tag{4}$$

3 Genetic Programming

Genetic Programming (GP) is an evolutionary algorithm for automatically solving problems. GP is inspired by biological evolution Darwin's theories [7, 8]. GP has been applied successfully to a large number of complex problems like feature selection and classification [11], industrial processes and robots [12, 13], engineering optimization [14] and modeling [15]. The evolutionary process of

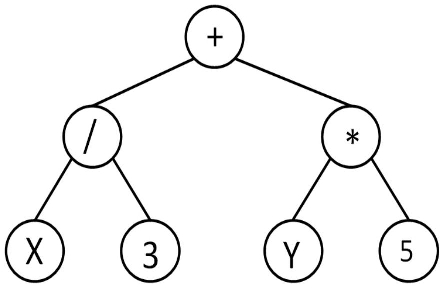

Figure 2. Tree structure model based GP.

the GP starts by generating some initial individuals. Each individual is a hierarchical computer program and all the individuals form one population. Each GP program is tree structured which can be shown as a graphical representations of so-called S-expressions of the programming language LISP [16]. Figure 2 shows an example of a very simple genetic program represented as a tree. After generating the initial population, the fitness of each individual in this population is computed. While stopping criteria is not met, the following steps are performed:

- Selection: some individuals are selected for reproduction using some selection mechanisms. The common Tournament selector is chosen for the implementation.
- Perform reproduction operators which include:
  - Crossover: this operator is applied on selected individuals (parents) in order to create new ones (children) by swapping subtrees between the parents randomly. This operation is illustrated in 4.
  - Mutation operates on one individual by replacing a subtree below a random chosen point by a randomly generated subtree, Figure 5.
  - Elitism: few individuals are copied to the next generation without any modification. The probability of crossover and mutation are selected based on the application but typically the probability of mutation is much smaller that crossover.
- Evaluate fitness of all individuals in the new generation.
- Termination condition: the evolutionary cycle of GP algorithm goes until either the optimal solution is found or the maximum number of generations is reached.

By this process the individual programs evolves and have better fitness values by time. The whole process can be summarized in the flow chart shown Figure 3.

In the case of the cutting machine introduced in this paper, GP will try to find the the best model that fits the nine experimental data points. This process is involves finding both the functional structure and the numeric coefficients for the model. In general, J.Koza identifies symbolic regression which is also called (function identification) as ”finding a mathematical expression, in symbolic form, that provides a good, best, or perfect fit between a given finite sampling of values of the independent variables and the associated values of the dependent variables.”.

The input of the GP generated model or mathematical expression is a set of independent variables while the output is one dependent variable. While the ultimate goal is to minimize the error of the model with respect to the given experimental data points [16].

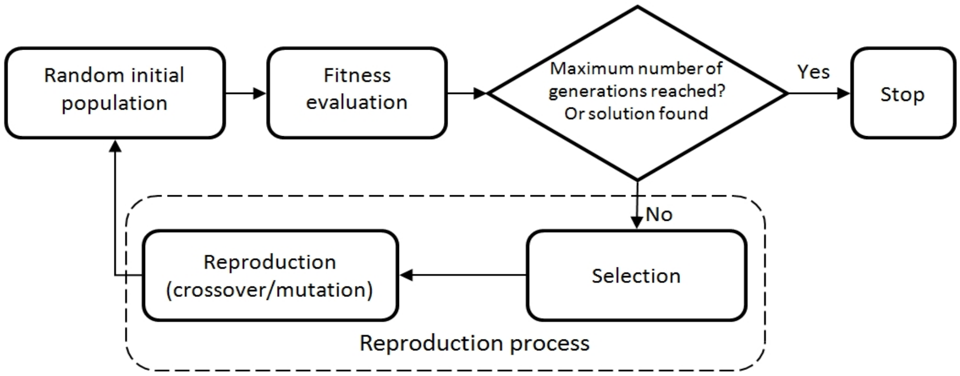

Figure 3. Flow chart of genetic programming.

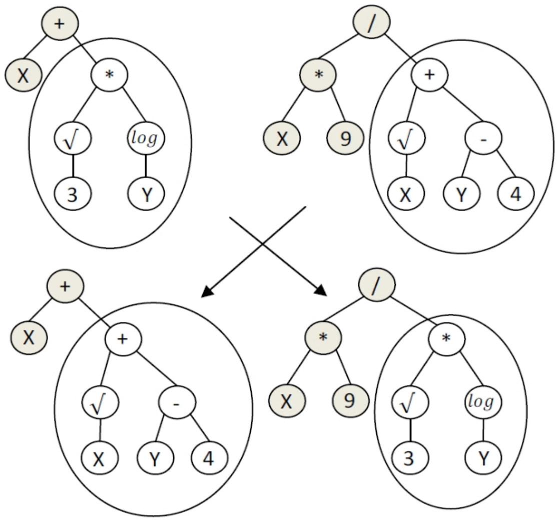

Figure 4. Example of GP crossover operation.

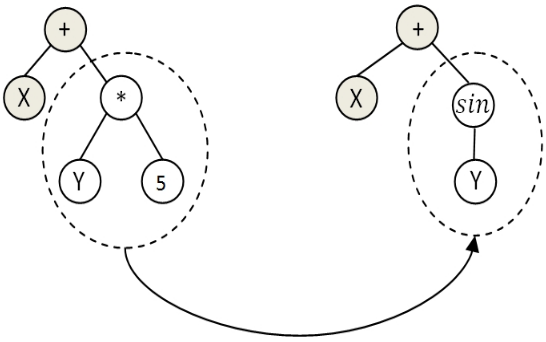

Figure 5. Example of GP mutation operation.

4 Performance Measurements

In order to check the performance of the developed GP models, the Variance-Accounted-For (VAF), the Mean Squares Error (MSE), the Euclidian distance (ED) and The Manhattan distance (MD) were measured. These performance criteria are assessed to measure how close the measured values to the values developed using the genetic programming approach. VAF, MSE, ED and MD are computed as:

$$VAF = \left[1 - \frac{var(y - \hat{y})}{var(y)})\right] \times 100\% \tag{5}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y - \hat{y})^2 \tag{6}$$

$$ED = \sqrt{\left(\sum_{i-1}^{n}(y_i - \hat{y}_i\right)^2} \tag{7}$$

$$MD = \frac{1}{N}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{8}$$

where y is real actual value, $\hat{y}$ it the estimated target value. n is the total number of measurements.

5 Experiments and Results

HueristicLab framework[1] was used to apply GP in the experiments designed in this research [17]. The data set in Table 1 was loaded into HeuristicLab framework then a symbolic regression via GP was applied with parameters set shown in Table 2. The cross validation was tuned to 50% for training and 50% for testing. After a cycle of 100 generations, GP converged to the best model shown in Figure 6. The best GP individual obtained was able to model the P05 steel cutting machine with a VAF value of 99.58%, 13.961 for MSE, 11.209 for ED and 3.1756 for MD. Actual and GP Estimated Temperature response is shown in Figure 7.

The temperature values obtained using GP, LS, GA and PSO are shown in Table 3. A simple comparison between the four approaches is summarized in Table 4 and Figure 8. The four approaches are compared using different evaluation criteria (VAF, MSE, ED and MD). It can be noticed that the symbolic regression model developed by GP outperformed all other techniques, therefore it can replace the common empirical models since GP model has higher performance results.

[1]HeuristicLab is a framework for heuristic and evolutionary algorithms that is developed by members of the Heuristic and Evolutionary Algorithms Laboratory (HEAL), http://dev.heuristiclab.com

Table 2. GP parameters

Parameter	Value
Mutation probability	15%
Population size	1000
Maximum generations	100
Selection mechanism	Tournament selector
Elites	1
Operators	{+,-,*,/,power,log,root}

$$T = \left(\log\left(\left(c_0 \cdot \alpha + \left(\left((c_1 \cdot \mathrm{f})^2 + c_2 \cdot \mathrm{v}\right) + \log(c_3 \cdot \mathrm{f})\right) \cdot \sqrt{c_4 \cdot \mathrm{f}}\right)\right) \cdot c_5 + c_6\right)$$

$$c_0 = 15997,\ c_1 = 12149,\ c_2 = 0.34404,\ c_3 = 0.56633,\ c_4 = 0.6993,\ c_5 = 151.78,\ c_6 = 461.23$$

Figure 6. Best GP generated model

Table 3. Temperature values obtained by different modeling approaches.

	T	PSO	GA	LS	GP
1	740.24	745.25	746.01	742.71	740.66
2	793.8	798.11	798.41	795.67	797.77
3	854.8	854.88	854.63	852.56	855.77
4	819.42	810.44	810.63	808.22	812.60
5	870	868.09	867.72	866.02	868.19
6	904.8	898.96	900.13	896.56	901.51
7	879.6	881.71	881.21	879.88	882.91
8	911.1	913.06	914.12	910.91	916.72
9	974.10	977.82	978.32	975.86	971.73

Table 4. Evaluation results of GP,LS,GA and PSO

	PSO	GA	LS	GP
VAF	99.522	99.514	99.52	99.674
MSE	20.463	20.929	25.229	13.961
ED	13.571	13.725	15.069	11.209
MD	3.7683	3.9048	3.5815	3.1756

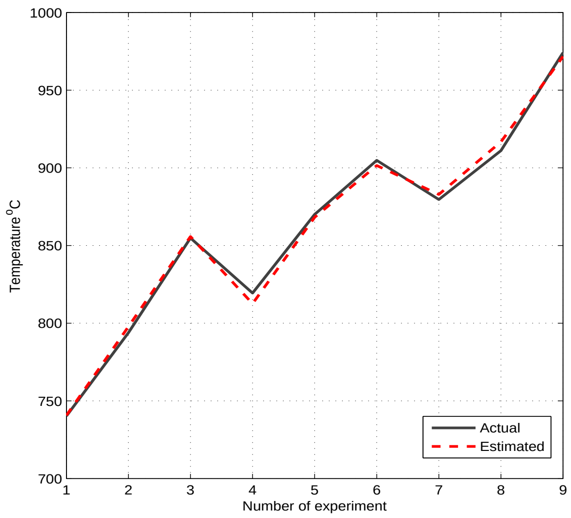

Figure 7. Actual and GP Estimated Temperature response.

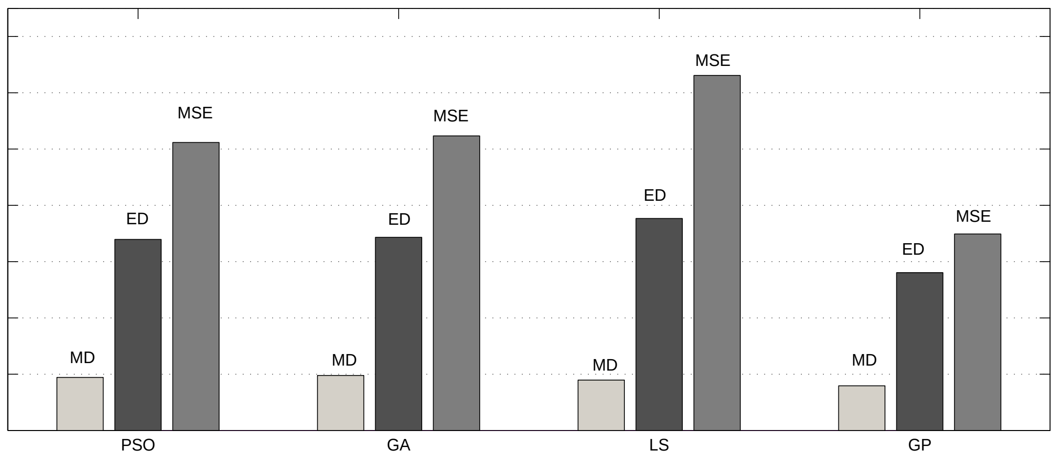

Figure 8. Evaluation results of GP, LS, GA and PSO.

6 Conclusions

In this paper, a symbolic regression model was developed using genetic programming for the temperature of a horny alloy steel cutting machine. Performance of GP model was compared to other approaches based on estimating the parameters of common empirical models. These approaches include; least squared estimation, genetic algorithm and particle swarm intelligence. GP showed superior performance in modeling the cutting tool.

References

[1] H. Faris and A. Sheta, "Identification of the tennessee eastman chemical process reactor using genetic programming," *International Journal of Advanced Science and Technology*, vol. 50, pp. 121–140, Jan. 2013.

[2] H. Faris, A. Sheta, and E. Oznergiz, "Modeling hot rolling manufacturing process using soft computing techniques," *International Journal of Computer Integrated Manufacturing*, 2013.

[3] Z. Jianin, L. Yutong, and Y. Qiun, "Ga algorithm for cutting experiment data drawing," in *Journal of Southwest Petroleum Institute*, pp. 62–63, 1998.

[4] T. Back, D. B. Fogel, and Z. Michalewicz, eds., *Handbook of Evolutionary Computation*. Bristol, UK, UK: IOP Publishing Ltd., 1st ed., 1997.

[5] L. Yi-jian, Z. Jian-ming, and W. Shu-qing, "Parameter estimation of cutting tool temperature nonlinear model using pso algorithm," in *Journal of Zhejiang University SCIENCE*, pp. 1026–1029, Zhejiang University Press, 2005.

[6] J. Kennedy and R. C. Eberhart, "Particle swarm optimization," in *Proceedings of the IEEE International Conference on Neural Networks*, pp. 1942–1948, Piscataway, NJ, USA, 1995.

[7] J. Koza, "Evolving a computer program to generate random numbers using the genetic programming paradigm," in *Proceedings of the Fourth International Conference on Genetic Algorithms*, Morgan Kaufmann, La Jolla,CA, 1991.

[8] J. R. Koza, *Genetic Programming*. MIT Press, 1982.

[9] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. The MIT Press, 1992.

[10] J. R. Koza, *Genetic Programming II*. MIT Press, 1994.

[11] F. A. Alsulaiman, N. Sakr, J. J. Valdés, A. El Saddik, and N. D. Georganas, "Feature selection and classification in genetic programming: application to haptic-based biometric data," in *Proceedings of the Second IEEE international conference on Computational intelligence for security and defense applications*, CISDA'09, (Piscataway, NJ, USA), pp. 339–345, IEEE Press, 2009.

[12] J. U. Dolinsky, I. D. Jenkinson, and G. J. Colquhoun, "Application of genetic programming to the calibration of industrial robots," *Comput. Ind.*, vol. 58, pp. 255–264, Apr. 2007.

[13] A. Sheta and H. Faris, "Improving production quality of a hot rolling industrial process via genetic programming model," *International Journal of Computer Applications in Technology, special issue on Computational Optimisation and Engineering Applications*, 2013, Accepted.

[14] G. Wang, *Application of genetic programming and artificial neural networks to improve engineering optimization*. PhD thesis, Beaumont, TX, USA, 1999. AAI9938783.

[15] L. Gusel and M. Brezocnik, "Application of genetic programming for modelling of material characteristics," *Expert Syst. Appl.*, vol. 38, pp. 15014–15019, Nov. 2011.

[16] M. Affenzeller, S. Winkler, S. Wagner, and A. Beham, *Genetic Algorithms and Genetic programming- Modern Concepts and Practical Applications*. CRC Press, 2009.

[17] S. Winkler, M. Affenzeller, and S. Wagner, "New methods for the identification of nonlinear model structures based upon genetic programming techniques," in *Proceedings of the 15th International Conference on Systems Science*, pp. 386–393, 2004.