Reconstruction of Physical Correlations Using Symbolic Regression

Klaus Weinert and Marc Stautner

Dept. of Machining Technology
University of Dortmund
D-44227 Dortmund, Germany

July 24, 2001

Abstract

Modeling the particle flow mechanisms in orthogonal cutting of turn-
ing processes is a vital task in mechanical engineering. This paper presents
a new approach that differs from techniques like finite element analyzes
(FEA) or molecular dynamics (MD). Using symbolic regression, a genetic
programming system evolves mathematical formulae that describe the tra-
jectories of single particles of steel recorded during the turning process by
a high-speed camera.

1 INTRODUCTION

Modelling the chip-building process in metal cutting has been in the centre of
interest for a long time. The chip geometry, the material movement and the
thermal processes which take place at the centre of cutting (the so called "con-
tact zone") are decisive for manufacturing quality, reduction of machining times
and tool wear. Most existing approaches incorporate FE- or MD-methods as
well as analytical techniques based on cutting force models. Some of them can
be found in [Ina97, SB93, SITU94, XBZ98]. The ideas presented in this pa-
per differ from those. No external knowledge is provided besides the images of
the process taken by a high-speed camera. The principle of the system works
similar to a human observer [War74] but of course, a computer can deal much
better with the large amount of information provided by the process images.
In order to "simulate" a human observer, as he is searching for the process-
describing formulae, Genetic Programming is utilized. This method has proven
to be helpful to search for solutions within a search space, especially when deal-
ing with finding formulae matching some given discrete requirements, a task
which is known as "symbolic regression". The evolved formulae are expected
to reveal geometrical aspects of the cutting process. They will be used as a
basis for advanced simulation tools, as well as for analytical investigations in
the field of metal cutting. Before they are applicable, they have to be validated
through correspondence with existing formulae or with observed properties of
the cutting process. Furthermore, the system may produce some unexpected

results, which do not correspond to any known properties of the process. If such phenomena occur, new knowledge about the process has been generated. This knowledge may be helpful to support the development of new simulation tools, which are able to increase the productivity of the process as well as its reliability.

The task of describing the particle flow mechanisms in orthogonal cutting leads to the problem of describing the particle trajectories of the crystalline particles in the metallic workpiece. This implies the problem of finding the mathematical functions which underly these trajectories.

These functions may be illustrated by relations of this type:

$$\vec{f}(t) = \left(\begin{array}{c} f_x(t) \\ f_y(t) \end{array} \right) \quad resp. \quad f(x,y) = 0$$

To succeed in this objective, some problems have to be solved. The first task is to obtain the trajectories of the particles in the crystalline structure of the metal workpiece. Until now the studies are based on movies which were recorded by Warnecke [War74] in the 1970's. A concept for the experimental setup, shown in figure 1, has been developed:
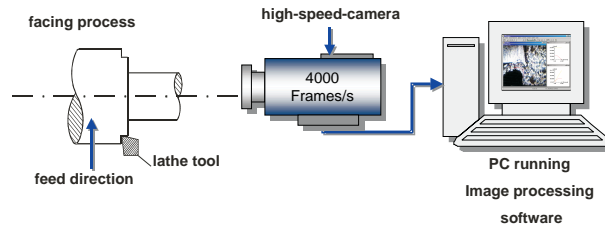

Figure 1: Experimental Setup of Camera and Turning Machine

A "Boehringer M-670" turning lathe will be used. The cutting process will be filmed by a high speed camera "Weinberger SpeedCam+", which allows to take 4000 images per second. The size of the display window is 0.44 mm by 0.33 mm. In Figure 2 the trajectory of a single particle has been sketched on to the frame of the digitized film.

In order to extract the particle trajectories, the programm "WINAnalyze" developed by Mikromak will be used. This program generates ASCII trace files that allow to analyze the trajectories of the particles in the GP-kernel. The positions of the particles are interpreted as values of a parametric function which is generated using symbolic regression and genetic programming.

2   MAIN ALGORITHM

The evolution starts with a randomly generated set of functions (initial population). The genetic operations are applied to this first generation of functions
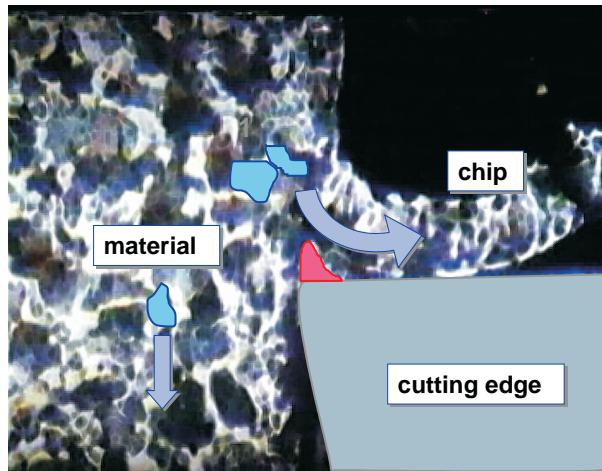
Figure 2: Single Frame Taken from the Film Sequence

and a succeeding generation is produced. The fitness of an individual is measured directly by evaluating the evolved formulae. Technically the formulae are compiled as a C-program and linked dynamically to the main GP-program, so evaluation can take place in a fast and generic way. It is also possible to utilize an interpreter, what may be more flexible in some cases (e.g. for debugging tasks), but slows down evaluation. The basic structure of the program follows the classic GP-scheme [Koz92]. The data structure and the genetic operators will be described next.

2.1 DATA STRUCTURES

In analogy to the infix notation of mathematical functions a tree based representation forms the genotype. A sample individual is shown in Figure 3.

The primary goal of getting the mathematical representations of two-dimensional curves leads to the phenotypic representation of an individual as graph of a parametric function.

This difference between genotypic and phenotypic representation and the need for genetic operators in which small changes in the genotype shall result in small changes in the phenotype yields two problems. First a set of genetic operators which allows to do these small changes on the genotype must be used. Second a fitness function which correctly evaluates the fitness of one individual has to be defined.

2.2 FITNESS FUNCTION

The naive way of applying a fitness function is to compare the function values of the individual with corresponding points of the trajectory. If all points of the trajectory lie on the function plot this individual will represent a perfect solution to the problem. The fitness values will be represented by a weighted
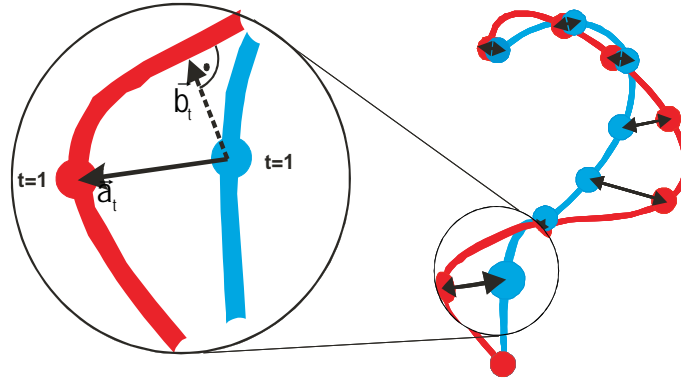
3

Fx(t)

Fy(t)

$$f_x(t) = sin(t * cos(3 + t))$$
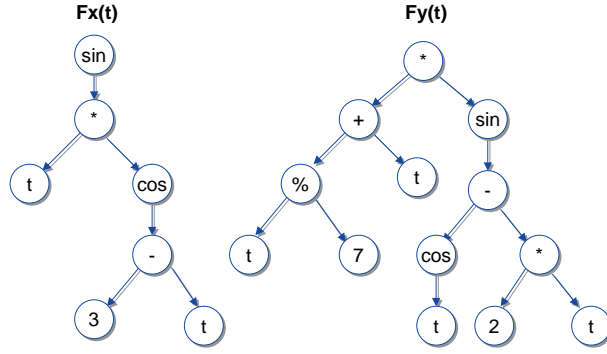
$$f_y(t) = t/7 + t * sin(cos(t) - 2 * t)$$

Figure 3: Sample structure of a single individual

point to value distance scheme. See Figure 4. In order to improve the efficiency of the naive approach, some variations have been realized.

$$f_{indiv.} = \sum_{t=0}^{t_{max}} (|\vec{a_t}|)^2$$

Figure 4: The Point to Value Distance Scheme

The evaluation of the distance of the sampling points to the evaluated function has been simplified to point-to-point distances instead of the mathematically correct orthogonal distances. This simplification is sufficient if the density of the sampling points is high enough.

Due to the fact that many solutions of the GP-algorithm show shapes that are similar to the correct solution but differ only in orientation and size, a de-

terministic step is inserted between mutation and the fitness calculation of the individual.
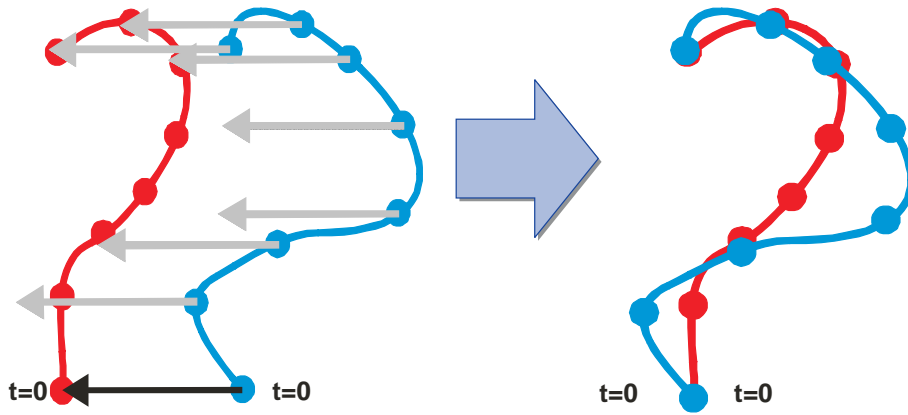

Figure 5: Translation of an Individual

The translation vector $\vec{a}$ is determined by setting the starting point of the individual at the position of the first point of the trajectory. See Figure 5. The translation is performed by adding $\vec{a}$ to each point of the genetically generated function.

In a second step the individual will be scaled to the size of the trajectory curve. The scaling values are determined by setting the maxima of the individual to the same value as the maxima of the trajectory points. See Figure 6.
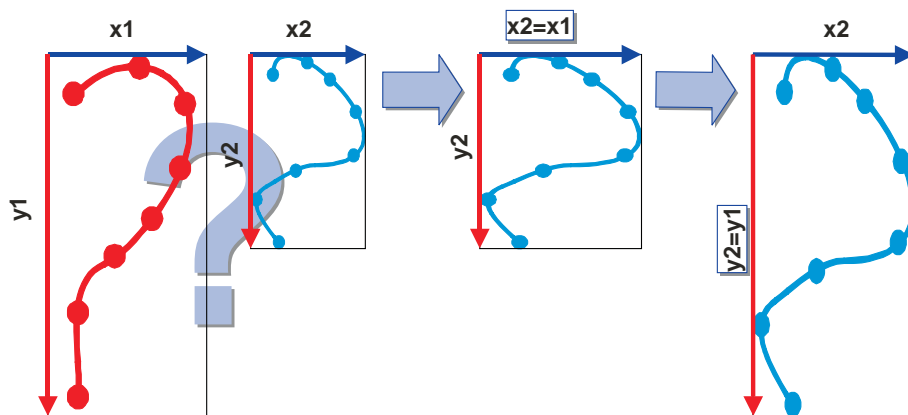

Figure 6: Scaling of an Individual

2.3 GENETIC OPERATORS

The basic algorithmic structure of the GP algorithm follows the description of Koza [Koz92]. Due to the fact that the algorithm has to evolve parametric functions, two symbolic representations have to be generated in parallel.
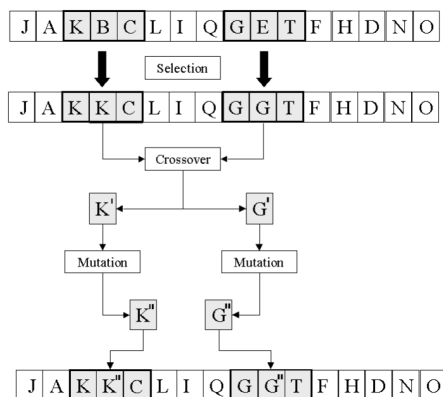

Figure 7: Sample of the Variation Scheme

Figure 7 illustrates the variation scheme of the algorithm consisting of recombination (crossover) and mutation. A tournament selection operator defines two groups (dark gray and light gray) of individuals that compete with each other. A letter is assigned to each individual. Every survivor of a tournament replaces the inferior individual. A survivor of a tournament is the individual which has a higher fitness value compared to its competitor. In Figure 7 **K** and **G** are the winners that replace **B** and **E**. The survivors undergo recombination realized by crossover. Each resulting individual **K´** and **G´** is varied by mutation yielding **K´´** and **G´´**.

The probability to choose a constant during mutation, the maximum size of an individual or the probability to change a function or a terminal and the number of changes is defined in a parameter file.

3 RESULTS

3.1 TEST FUNCTION

In the first example the following simple function had to be reconstructed:

$$\vec{f}(t) = \left(\begin{array}{c} f_x(t) \\ f_y(t) \end{array} \right) = \left(\begin{array}{c} \frac{cos(t)}{t+t^2} \\ \frac{sin(t)}{6t} \end{array} \right)$$

In this case a known function was reconstructed in order to illustrate the power of the symbolic regression algorithm. The shape of the graph of $\vec{f}(t)$ resembles a

6

chip. For simulation purpose this function was sampled yielding discrete points which represent points which could have been sampled from a film sequence. The size of the population was 200 and 400 generations were evolved. After this
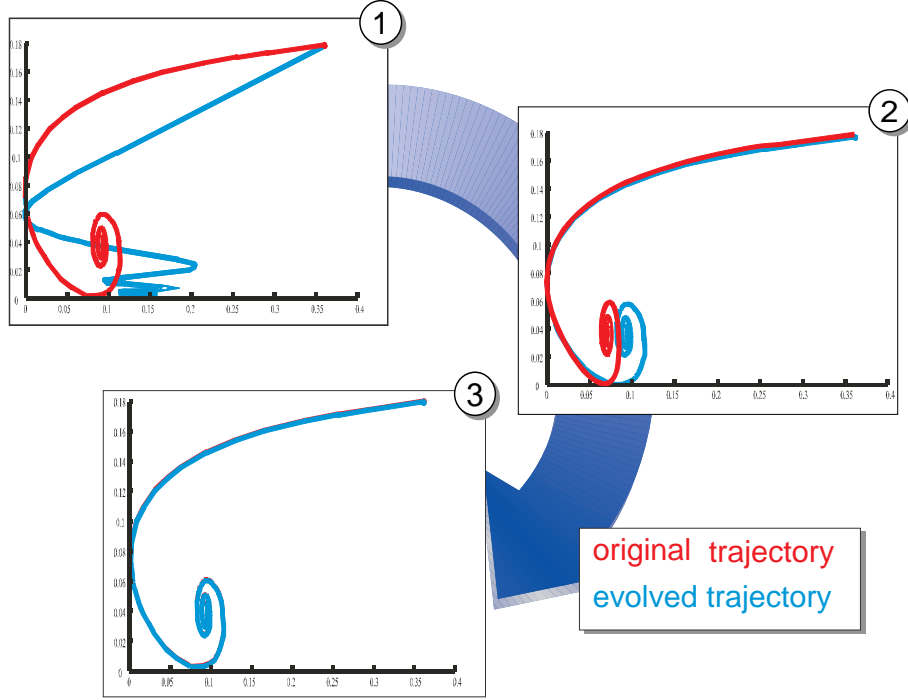

Figure 8: Approximation of the Test Function, after 200, 300 and 400 Generations

test run the evolved formula was,

$$f_x(t) = \frac{cos(t)}{t * (t + 1.02618) * cos\left(\frac{1.272908745}{(t+1.21945)*(t+3.81393)}\right)}$$

$$f_y(t) = \frac{sin(t)}{t}$$

After rescaling with the factor $1\frac{2}{3}$ the target function has been approximated with a negligible error. See Figure 8.

3.2 REAL_WORLD DATA

Next the algorithm is used on the data extracted from the film data of the turning process. The size of the filmed window is 0.44 mm by 0.33 mm, so the trajectory which is to be reconstructed stretches from 0.01 mm to 0.08 mm and from 0.06 mm to 0.16 mm.

Terminal Set	t , $K \in \{ 1.0; 5.0\}$
Func. Set	+ - * / sin() cos()
Selection	tournament
# individuals	250
# programs	2 programs per individual

Table 1: Parameter Table for real world test data

Table 1 shows the parameters for this test run. The result after 500 generations is shown in Figure 9.
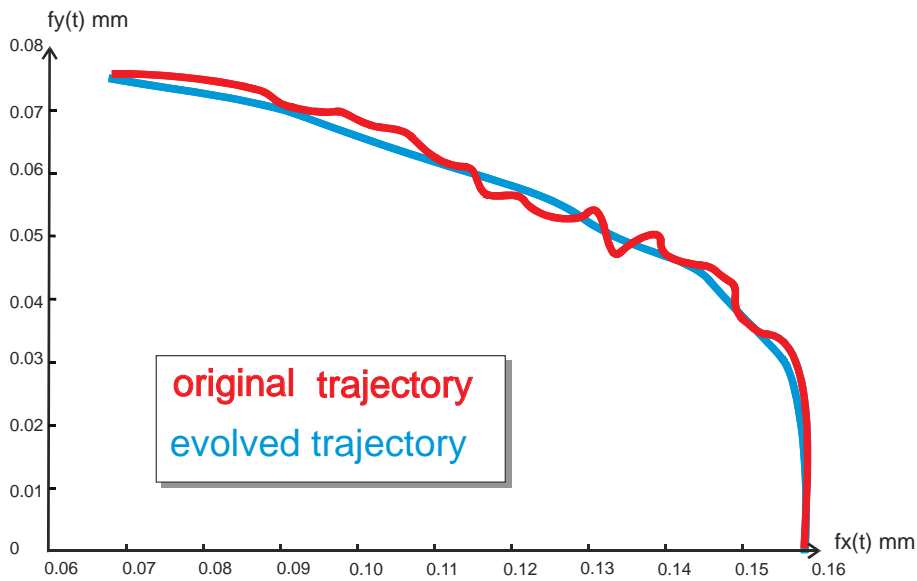

Figure 9: Approximation of real-world data after 500 Generations

The figure illustrates the tendency of the algorithm to equalize the trajectories. The approximations can be improved locally by increasing the number of sampling points. This way the reconstructions are forced to follow also small bendings of the particle flow.

4   CONCLUSION

This paper presents a method to reconstruct and analyze unknown mathematical correlations. In this case one particle trajectory of material movement in the turning process has been reconstructed.

This reconstruction uses symbolic regression and genetic programming. First tests show the general applicability of this approach. Complex two dimensional functions can be reconstructed by assembling terminal functions.

In order to develop a physical model of the turning process, there are additional considerations to be done. For example adding parameters, like the cutting speed or the tool angle to the reconstruction process. Furthermore, restrictions had to be developed to constrain the length of the reconstructed individuals.

The proposed tracking of particles by GP is the first step towards an overall system, which will include another CI-method, namely a CA (Cellular Automaton). This automaton will be provided with the rules evolved by the GP part of the system. The authors intend to use the modeling properties of CAs (simple local interactions can model complex global behavior) together with the GP-evolution of formulae to develop a complete model of the geometric properties of the chip-building process in orthogonal cutting which can be a basis for systems as described in the introduction section.

Acknowledgments

References

[Ina97]   T . Inamura. Brittle/ductile phenomena observed in computer simulations of machining defect-free monocrystalline silicon. *Annals of the CIRP*, 46:31–34, 1997.

[Koz92]   J. R Koza. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, Cambridge, MA, 1992.

[SB93]   H. Schulz and K. Bimschas. Optimisation of precision machining by simulation of the cutting process. *Annals of the CIRP*, 42:55–58, 1993.

[SITU94] S. Shimada, N. Ikawa, H. Tanaka, and J. Uchikoshi. Structure of micromachined surface simulated by molecular dynamics analysis. *Annals of the CIRP*, 43:51–54, 1994.

[War74]   G. Warnecke. *Spanbildung bei metallischen Werkstoffen*. Technischer Verlag Resch, Munich, 1974.

[XBZ98]   J. Q. Xie, A. E. Bayoumi, and H. M. Zbib. Fea modeling and simulation of shear localized chip formation in metal cutting. *Intl. J. of Machine Tools and Manufacture.*, 38:1057–1087, 1998.