Tool Wear and Breakage Detection Using a Process Model
Y. Koren (1), A. G. Ulsoy, and K. Dansi

On-line sensing of tool wear and breakage in machining has been a long standing goal of the manufacturing community. Wear and breakage detection systems are typically based on force, acoustic emission, current or temperature measurement. They are important for reliable unmanned operation, and also for implementation of an adaptive control optimization system. This paper proposes a model-based approach to on-line tool wear and breakage detection under varying cutting conditions based on force measurement. The proposed model is used together with on-line parameter estimation to track flank wear due to tool wear and tool breakage. The proposed method is illustrated with a simulation study, the results of which confirm the feasibility of the model-based approach.

1. INTRODUCTION

A successful tool wear/breakage sensor has been a long standing goal in the manufacturing community [1-8]. In addition, a wear sensor has always been regarded as a necessary step toward the implementation of an adaptive control optimization system [6,8,10]. Wear and breakage detection systems are typically based on acoustic emission, motor current, temperature measurement, or cutting force.

Acoustic emission [11-16] and vibration [17] measurements are receiving a great deal of attention due to the non-obtrusive nature of the transducers, but improvement in sensitivity and linearity problems. Temperature measurement is particularly important for wear [20-22], however, a practical on-line transducer has yet to be developed [7]. With all these methods the tool wear or breakage are measured indirectly through another system variable. Direct methods based on optical or radiometric techniques have been demonstrated, but not yet proven to be practical [23,24].

Among the indirect methods those based on force and torque sensing [25-28] are perhaps the most developed. Such commercial units have begun to appear on the market in recent years, and have been used with some success in certain limited production situations. With this recent introduction of the first commercial tool wear/breakage sensors it is worthwhile to evaluate these sensors and to reassess their economic advantages [29-30]. For this purpose the machining operation can be divided into roughing and finishing cuts. In roughing operations one can consider two distinct types of production:

1. short production cycles (e.g., t < 5 min)
2. long production cycles (e.g., t > 10 min)

The first situation is frequently encountered in high volume production, while the second situation occurs in the production of large complex workpieces, or in the machining of hard materials such as titanium.

The importance of wear sensing, breakage detection, and adaptive control is summarised according to the production mode in Table 1. As can be seen tool breakage detection is economically important in all machining operations except finishing operations. Tool wear sensing is important as a predictor of tool failure due to excessive wear. In long production cycle manufacturing operations, tool wear detection becomes important for surface roughness control. Tool breakage sensing is also important in finishing cuts due to the effect on product quality.

Commercially available systems for wear and breakage detection are typically based on force or power limits. When the measured force falls outside these predetermined fixed limits the tool is assumed to have failed due to excessive wear or breakage. The disadvantage of the fixed force limit method is that the cutting conditions must remain nearly identical throughout the whole cutting operation, and therefore this is applicable only in very simple cases.

To extend the force limit approach to various cutting conditions, force signature methods typically use "learning" or "averaging" strategies. These methods, however, have the following disadvantages:

1. many parts must be cut to allow the system to learn the force signature
2. the method is not effective in long production cycle situations
3. the method is not applicable in adaptive control (AC) because AC cutting conditions are not repetitive from one part to another.

Table 1 also summarises the importance of various types of adaptive control systems. Adaptive control is not particularly significant in high volume/short production cycle manufacturing. Geometric adaptive control (GAC) is important in finishing operations [31-34]. In manufacturing operations with a long production cycle, both adaptive control optimization (ACO) and adaptive control with constraint (ACC) type systems are important.

The tool wear sensor is also a necessary component in ACO systems for turning and milling [8-10, 35]. There is a fundamental difference, however, between using a tool wear monitoring system (which operates in open-loop) and applying the same sensor in a closed-loop adaptive control system. ACO systems (e.g., the mid 1960's Bendix system [3]) are based on maintaining the optimal cutting conditions by incrementing the feed and/or cutting speed in small steps. However, in practice an indirect measurement of the tool wear is used, and feeding back the sensor signals in order to close the ACO loop is not a straightforward task.

For example, assume that the tool wear is monitored through measurements of the cutting force. Typically, increasing wear increases the magnitude of the output signal of this sensor. An incremental increase in the feed (as may be done automatically by the ACO system) will have two effects:

1. Since the cutting force is directly dependent on the feed, an increase in the feed causes a consequent increase in the cutting force. This is the direct effect.
2. An increase in the feed shortens the tool life, namely increasing the tool wear rate. As a consequence the wear increases and the force again becomes larger. This is the indirect effect.

An ACO system should use only the latter effect. However, since both mechanisms affect the cutting force similarly, the problem of isolating the second effect from the first is not trivial. One might say that the direct effect happens almost immediately (theoretically after one revolution of the spindle), while the other mechanism affects the force after "some time". However, since the ACO convergence strategy is based upon incremental feed variations, and the signal-to-noise ratio in force measurements is low, it is difficult in practice to isolate the two phenomena by using simple electronics. A solution to this problem might be obtained by programming a mathematical model of the cutting process and updating it in real time in order to obtain an accurate estimate of the two effects. The separation of the two effects may have other significant outcomes: it permits the separate detection of tool breakage and excessive tool wear. Tool breakage is detected by a sudden change in the force due to the first (direct) effect, whereas wear can be sensed by changes in the force due to the second (indirect) effect. By contrast to the fixed limits and the force signature approaches, this proposed method does not require the cutting of many parts for the learning mode, and is also effective in long production cycles.

Annals of the CIRP, Vol. 35/1/1986 283
2. THE FORCE EQUATION

Before proceeding to the model-based approach of this paper, it is worthwhile to further clarify the problem introduced at the end of the previous section by using a simplified mathematical analysis. The relationship between each component of the cutting force $F$ and the flank wear $W$ is approximated by [36]

$$ F = F_0 + a(W) $$

(1)

where $F_0$ is the initial cutting force (with a sharp tool), $a$ is the depth of cut, and $W$ is a constant for a certain tool and workpiece material; and a fixed set of cutting conditions (feed $f$ and speed $v$). Equation (1) has been experimentally verified (see Fig. 1) for cases in which flank wear has a dominant effect on tool life. The initial cutting force is given by

$$ F_0 = aaf $$

(2)

where $a$ and $f$ are constants depending on the tool and workpiece material; typically 0.6 $f$ ≤ 1.

If the wear is assumed to be in its linear progression zone (see Fig. 5) it obeys the equation

$$ W = W_0 + b(t) $$

(3)

where $W_0$ is the corresponding wear at which the tool is replaced, and $b(t)$ is a slowly varying function used to account for the non-constant slope of the flank wear curve. Combining Eq. (1) and (3) yields the following force equation:

$$ F = F_0 + b(t) $$

(4)

An acceptable model for $T$ is the extended Taylor's tool life equation

$$ C_TD_F = 1 $$

(5)

substituting $T$ from Eq. (6) into Eq. (5) yields

$$ F_1 = T(1) + b $$

(6)

We now make the assumption that tool wear and tool life are essentially independent. Thus, the wear sensors could be used to determine the wear rate, and the tool life equation could be used to predict the time until replacement is necessary.

3. ESTIMATION OF PROCESS PARAMETERS

The estimation of the unknown coefficients $a$, $f$, and the unknown exponents $k$ and $m$ in Eqs. (2), (4) and (7) can be accomplished by on-line (recursive) parameter estimation methods. In this paper the standard recursive least squares (RLS) algorithm is used [49, 50],

$$ \theta(k+1) = \theta(k) + \frac{P(k)}{\lambda(k) + |\theta(k)|} (y(k) - \phi(k)\theta(k)) $$

(8)

and,

$$ P(k+1) = \frac{1}{\lambda(k) + |\theta(k)|} P(k) $$

(9)

where $y(k)$ is the value of the measured variable at time $t = k$ for $k = 0, 1, 2, 3, ...$. $\theta(k)$ is a vector of measured (or known) variables, $\phi(k)$ provides exponential data weighting, and $|\theta(k)|$ is a vector of parameter estimates. The $P(k)$ is the matrix of estimation gains. The above algorithm recursively updates the estimated parameter vector $\theta(k)$ for any process whose equations can be written in the form.

$$ y(k) = \phi(k)^T \theta(k) $$

(10)

Thus, the process model must be written in a form that is linear in the unknown parameters, which are the elements of the vector $\theta(k)$.

For the process model presented in the previous section, we first consider the estimation of $a$ and $f$ using Eq. (2). When the tool is sharp, the term $F_0$ in Eq. (4) is approximately zero, and one can write

$$ F = F_0 = aaf $$

This can be rewritten in the form of Eq. (10) by taking natural logarithms of both sides,

$$ \ln F = \ln a + \ln f + (1)\ln f $$

(11)

or,

$$ \ln F - \ln a = (1)\ln f $$

(12)

So, initially the estimation algorithm in Eqs. (8) and (9) is used with,

$$ \theta(k) = 1\ln F(k) - \ln a(k) $$

(13)

$$ \phi(k)^T = [1 \ln F(k)] $$

(14)

to estimate $a$ and $f$. Thus, assuming that $a$ and $f$ remain constant during the cutting operation, one can calculate

$$ F_1(k) = aaf $$

(15)

Next, use the same RLS algorithm with Eq. (4) to estimate $f$, $a$, and $m$ in Eq. (7). For example, if we assume that $Y(t) = 1$, and $a$, $m$, and $n$ are all unknown and constant, taking natural logarithms of both sides of the equation gives,

$$ \ln F(k) - \ln F_0(k) = \ln a + \ln f + \ln m $$

(16)

$$ \phi(k)^T = [1 \ln F(k) \ln F_0(k)] $$

(17)

to estimate $m$, $n$, and $f$. Then, assuming that $a$ and $f$ remain constant during the cutting operation, one can calculate

$$ F_2(k) = aaf $$

(18)

Finally, use the same RLS algorithm with Eq. (5) to estimate $a$, $f$, and $b$. The above algorithm recursively updates the estimated parameter vector $\theta(k)$ for any process whose equations can be written in the form.

$$ y(k) = \phi(k)^T \theta(k) $$

(19)

Thus, the process model must be written in a form that is linear in the unknown parameters, which are the elements of the vector $\theta(k)$.
to estimate $b$, $m$, and $n$.

The RLS estimation algorithm requires that one select initial values for $\hat{b}(k)$. These selections can either be made from typical values of $b$, $m$, and $n$ as published in the literature [36, 37], or assigned arbitrarily. Initial values of the gain matrix $P(k)$ are usually selected to be of the form

$$P(0) = \alpha I$$

where $I$ is the identity matrix, and the scalar constant $\alpha > 0$ is chosen by trial and error through simulation studies. The RLS algorithm is typically not too sensitive to the choice of $\hat{b}(0)$ and $P(0)$.

The choice of the weighting factor $\alpha(k)$ in Eqs. (6) and (9) is more significant. The standard RLS algorithm is obtained when $\alpha(k) = 1$. When it is necessary to track parameters with a high degree of accuracy, a value of $0 < \alpha(k) < 1$ exponentially weights the data to ensure that the "alertness" of the algorithm [49, 50]. In the simulation results presented in the next section a value of $\alpha(k) = 1$ was used. In those simulation studies scaling and factorization of the $(k)$ matrix into upper triangular and diagonal factors was also used. Such measures are often required in practice to eliminate number overflows and to simplify the calculations. For some processes a model as in Eq. (10) with constant coefficients it can be proven that the RLS algorithm ensures convergence of the solution error $\varepsilon(k)$ to zero. To ensure that the parameter estimates $\hat{b}(k)$ converge to reasonable values there is an additional requirement on the richness (frequency content) of the process inputs. For further discussion on estimation algorithms the interested reader is referred to [49, 50].

4. SIMULATION RESULTS

In this section we present simulation results to illustrate the model-based approach to tool wear and breakage detection as presented in the previous sections. The basic scheme is illustrated in Figure 3, where the "process model" block is based on the model presented in Eqs. (4), (7), and (8). The "adaptation algorithm" is based on the RLS algorithm presented in Eqs. (8) and (9). For the purposes of this simulation study, the "process" is represented by a model described previously in [48]. That model is more detailed than the simple process model presented here and accounts for crater wear as well as flank wear due to both thermally and mechanically activated mechanisms. Results presented in [48] show that the model gives good results for force, temperature, flank wear, and crater wear. The model estimation process as used in the simulation is summarised in the Appendix.

Under these cutting conditions flank wear dominates, and the effect of crater wear is negligibly small.

To provide sufficient input richness for parameter convergence using the RLS algorithm, the feed $f(t)$ is varied as shown in Fig. 4, by $\pm 0.025$ mm/rev about the nominal value of $f_{平均}$. This process is assumed to be relatively constant wear rate for $3 < t < 10$ min. For $t > 10$ min the wear rate again increases and the useful life of the tool is $T = 11$ min. The simulated force shows the effects of the variations in feed, and also an increase in level due to increasing flank wear.

The RLS algorithm in Eqs. (8) and (9) is used to estimate $\alpha$ and $\beta$ using Eq. (12). This estimation is carried out during the first few sampling periods and leads to constant values of $\alpha = 3000$ and $\beta = 0.67$. Assuming that $\beta$ is constant, $f_0$ is calculated and used to estimate the $f_0$ term in Eq. (7). The estimate of $f_0$ is shown in Fig. 7. The value $v = 300$ m/min is constant, and $\alpha = 2.5$ is constant, and $\beta = 1$ is assumed to be known, so we use the following form of Eq. (7): $F(t) \approx \frac{v \alpha^2}{2} f_0^2$. The value $f_0 = 1$ has been suggested in this form. We have found that $f_0 = 1$ gives good results in the estimation. As shown in Figs. 8 and 9, when the wear is in the constant wear rate region the estimated values are $\alpha = 80$ and $\beta = 0.1$.

These lead to good agreement between the actual $F(t)$ and the estimated $F(t)$ (i.e., $F(t)$), as shown in Figure 10. For the constant wear rate region. When the tool wear rate increases, the error $e = F(t) - \hat{F}(t)$ becomes large, and serves as a good indicator of tool failure. Thus, the simulation results shown in Figs. 4-10 illustrate the potential usefulness of the model-based approach for predicting the effects of feed and flank wear on force and for predicting tool failure due to excessive flank wear.

5. SUMMARY AND CONCLUSIONS

This paper has presented a model-based approach to on-line tool wear and breakage detection in metal cutting. Such a model-based approach is considered important for machining under variable cutting conditions, and for use with adaptive control systems that automatically adapt feedrates. The basic approach has been developed, and illustrated with a simple simulation example.

The simulation results confirm the feasibility of the proposed model-based approach, and indicate the need for further research to obtain experimental confirmation. Further research may also be desirable on process modeling, estimation algorithms, and on-line training of the model-based approach by using artificial intelligence methods.

6. REFERENCES


285


APPENDIX

The "cutting process" in Fig. 3 is simulated using the following model from (48).

\[
W_e = \frac{(W_e)^2}{(W_e^2 + \alpha)^2} \exp(-k_2/(273 + \nu))
\]

\[
k_3 = k_2 \exp(-k_2/(273 + \nu))
\]

\[
k_3 = k_2 \exp(-k_2/(273 + \nu))
\]

where \(W_e\) is the mechanically activated component of flank wear, \(k_2\) is the thermally activated component of flank wear, \(\nu\) is the temperature at the tool flank, \(v\) is the cutting speed, \(f\) is the feed, \(a\) is the depth of cut, and \(C\) is the cutting force. The parameters \(k_1, k_2, k_3\) are selected to be typical of turning steel with a carbide tool [48]. The values used in the simulation study are.

- \(a = 2.5\) mm
- \(K_1 = 1960\)
- \(r = 0.35\) mm/sec
- \(K_2 = 0.57\)
- \(v = 300\) m/min
- \(K_3 = 86\)
- \(a_0 = 10^7\)
- \(K_4 = 0.1\)
- \(k_5 = 500\)
- \(K_6 = 500\)
- \(k_7 = 5.2 \times 10^{-5}\)
- \(K_8 = 0.4\)
- \(k_9 = 15\)
- \(K_9 = 0.6\)
- \(k_{10} = 8000\)
- \(n_1 = 1.45\)
- \(k_{11} = 72\)
- \(n_2 = 0.76\)
- \(k_{12} = 2500\)
- \(n_3 = 0.05\) min.

Table 1. Summary of Comments on the Importance of Wear, Breakage, and Adaptive Control

<table>
<thead>
<tr>
<th>Type of Operation</th>
<th>Wear</th>
<th>Breakage</th>
<th>Adaptive Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short Production</td>
<td>Important</td>
<td>Important</td>
<td>Important</td>
</tr>
<tr>
<td>Cancellation Volume</td>
<td>Condition</td>
<td>Volume</td>
<td>Condition</td>
</tr>
<tr>
<td>Long Production</td>
<td>Important</td>
<td>Important</td>
<td>Important</td>
</tr>
<tr>
<td>Finishing</td>
<td>Important</td>
<td>Important</td>
<td>Important</td>
</tr>
</tbody>
</table>
Fig. 9 Estimated p versus Time for the Simulation Example

Fig. 10 Actual and Estimated \( F_t \) (i.e., \( \hat{F}_t \)) Versus Time for the Simulation Example