Pergamon International Journal of Machine Tools & Manufacture 41 (2001) 1363–1384

INTERNATIONAL JOURNAL OF MACHINE TOOLS & MANUFACTURE DESIGN, RESEARCH AND APPLICATION

Tool wear condition monitoring in drilling operations using hidden Markov models (HMMs)

Huseyin M. Ertunc, Kenneth A. Loparo *, Hasan Ocak

Electrical Engineering and Computer Science, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106-7070, USA

Received 10 September 1999; accepted 1 October 2000

Abstract

Monitoring of tool wear condition for drilling is a very important economical consideration in automated manufacturing. Two techniques are proposed in this paper for the on-line identification of tool wear based on the measurement of cutting forces and power signals. These techniques use *hidden Markov models* (*HMMs*), commonly used in speech recognition. In the first method, bargraph monitoring of the HMM probabilities is used to track the progress of tool wear during the drilling operation. In the second method, sensor signals that correspond to various types of wear status, e.g., sharp, workable and dull, are classified using a multiple modeling method. Experimental results demonstrate the effectiveness of the proposed methods. Although this work focuses on on-line tool wear condition monitoring for drilling operations, the HMM monitoring techniques introduced in this paper can be applied to other cutting processes.

Keywords: Drilling; Tool wear; Condition monitoring

1. Introduction

Machining operations such as turning, milling, drilling and grinding are material removal processes that have been widely used in manufacturing since the industrial revolution. Of these processes, drilling is one of the most common machine tool operations in manufacturing. The drilling operation is also frequently a preliminary step for many machining processes like boring, tapping and reaming. It is estimated that approximately 250 million twist drills are used annually by US industry alone [1]. Hence, a considerable amount of money is spent on drilling tools each

* Corresponding author. Tel.: +1-216-368-4088; fax: +1-216-368-2668.
E-mail address: kal4@po.cwru.edu (K.A. Loparo).

0890-6955/01/$ - see front matter
PII: S0890-6955(00)00112-7

year. For instance, according to the US Department of Commerce, approximately $1.62 billion was spent in the production of drill bits in the US in 1991 [2]. It was also estimated that drilling accounts for nearly 40% of all metal removal operations in the aerospace industry [3]. Even for a small jet fighter, more than 245,000 holes need to be drilled [4].

Whereas manufacturing technology has moved to the stage of automation in recent years, there are still two unsolved problems in drilling, like in many other cutting processes: *tool wear* and *tool breakage*. Due to its complexity, the dynamics of tool wear are not completely understood. After a certain limit, tool wear can cause catastrophic failure of the tool that can cause considerable damage to the workpiece and even to the machine tool set-up. Even if the price of a drill is relatively low, failure can cause an incomparably higher production cost. The economic aspect of downtime (production loss during tool replacement) because of tool failure is very important in an unmanned factory.

Therefore, monitoring systems are necessary to detect the progress of tool wear using sensors during the cutting operations, so that worn tools can be replaced in time. For this reason, sensor-based *tool condition monitoring* (*TCM*) systems have become the topic of considerable research in machining processes. TCM techniques can be broadly classified as *direct methods* and *indirect methods* based on the type of sensor measurements; or *off-line monitoring* and *on-line monitoring* with regard to the implementation of the monitoring technique. Direct methods have the advantage of continuous measurement of actual tool wear, while a parameter correlated to tool wear is measured in an indirect method. On the other hand, on-line monitoring techniques are used during the machining process and off-line techniques require the machine to be stopped. Direct measurements are very difficult due to the continuous contact between the tool and the workpiece while the machine is running. By using different sensors, information regarding the machining process can be obtained without interrupting the process. Therefore, direct (indirect) methods generally tend to be off-line (on-line) monitoring.

In the drilling operation, a direct measurement of tool wear such as the worn area on the tool cutting lip (edge) can be obtained after drilling a certain number of holes by either removing the drill from the machine or installing a measuring device on the machine. For example, by analyzing the image of the tool surface, one can characterize tool wear and assess the condition of the tool — because tool wear is a deformation of the tool surface and the intensity of the light reflected from the worn sections of the drill surface changes as the drill tool wears. Even though direct methods are likely to be more accurate, both implementations — i.e., removing the drill or installing a measurement device — are generally not suitable in an automated machining process. This is because machine interruption might hurt finished part quality and the measuring devices must be located in close proximity to where tool wear and breakage take place. Because the drill is engaged in the workpiece during drilling, it is not possible to monitor the condition of the drill either visually or optically by means of a vision system. Therefore, indirect methods based on sensor signals for on-line tool condition monitoring are necessary for use in high-volume manufacturing. However, the determination of a reliable criterion that is proportional to tool wear under all cutting conditions is the main problem for both indirect and direct methods.

Many models have been developed and different approaches to tool wear monitoring have been tried over the years such as neural networks [5,6], model-based techniques [7], dynamic force modeling [8,9], statistical approach [10], control system design [11–13] and abductive networks [14]. For a detailed review of these techniques, refer to Ref. [15].

The objective of this paper is to propose a new, on-line, data-trainable tool condition monitoring system for drilling operations using hidden Markov models (HMMs). Based on force (thrust and torque) and power (spindle and servo power) signals collected during the drilling operation, HMMs are used to evaluate the health of the tool. The effectiveness of the proposed monitoring technique is presented from drilling experiments using a split-point twist drill and a steel workpiece.

2. Tool wear in drilling operation

2.1. Wear mechanism

Tool wear is a result of physical (mechanical) and chemical interactions between the cutting tool and workpiece, that remove small parts of material from the cutting tool. The basic wear mechanisms can be classified as follows: abrasion, adhesion, diffusion and fatigue. Even though abrasive and adhesive wear are the most commonly encountered mechanisms, the wear in cutting tools cannot be described *completely* by these mechanisms *alone*. Tool wear is a highly complicated process associated with many variables such as contact stress, nature and composition of the workpiece and cutting tool, temperature on the cutting edge, and cutting conditions.

2.2. Wear in drilling

Tool wear in drilling is a progressive and comparatively slow phenomenon whereas tool failure and cutting edge breakage are usually catastrophic and sudden without any warning. Even though a drill tool begins to wear as soon as it is placed into operation, the wear occurs at an accelerated rate once a drill becomes dull.

As illustrated in Fig. 1, drill wear states can be classified into different wear stages as a function of tool life [6]:

1. initial wear;

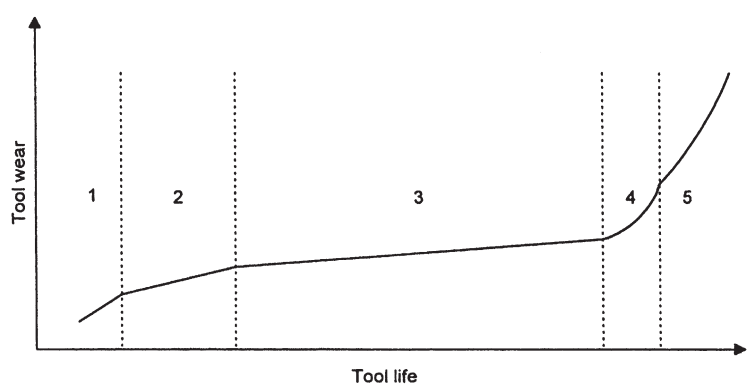

Fig. 1. Tool wear evolution.

2. slight wear (regular stage of wear);
3. moderate wear (micro breakage stage of wear);
4. severe wear (fast wear stage); and
5. worn-out (or tool breakage).

2.2.1. Wear patterns

Since the geometry of a drill bit and the cutting conditions vary along the cutting lips from the margin to the chisel edge, diversified wear patterns can be seen at the drill bit. In general, the distribution of heat generated during cutting, gradients of pressure, friction, and the stress distribution at the tool–workpiece interface influence wear patterns.

Kanai and Kanda [16] classified the various types of drill wear for large drills (of diameter $\geq$6 mm); these are illustrated in Fig. 2 in addition to chipping at the cutting lips (P_T and P_M). Many researchers have selected outer corner wear as the predominant type of wear in drilling. Because the highest cutting speed appears at the outer corner of the cutting lips, maximum wear typically occurs at these corners when compared with other parts of the drill.

3. Theoretical background for HMMs

Hidden Markov models (HMMs) were introduced by Baum and colleagues at the end of the 1960s. Then, Baker and Jelinek implemented this theory for speech processing applications in the 1970s [17]. Currently, HMM is a state-of-the-art technique for speech recognition because of its elegant mathematical structure and the availability of computer implementation of these models. Recently, HMM applications have been spreading steadily to other engineering fields that include communications systems [18], target tracking [19], machine tool monitoring [20], fault detection and diagnosis [21,22], robotics [23,24] and character recognition [25].

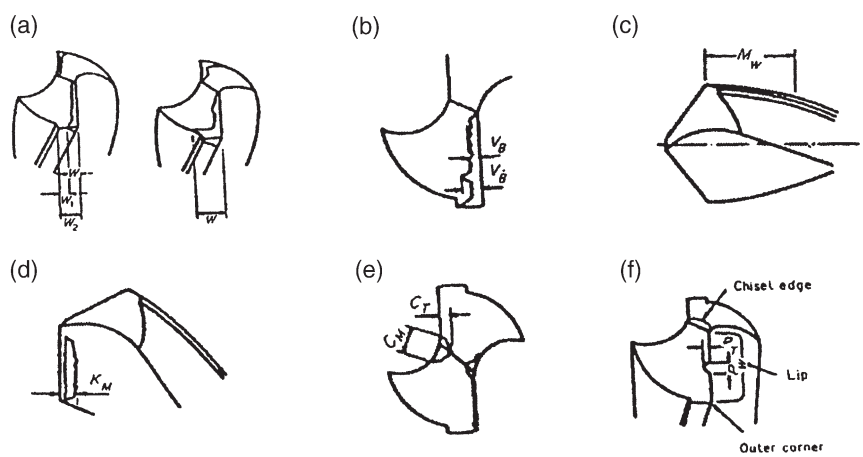

Fig. 2. Different types of wear. (a) Outer corner wear; (b) Flank wear; (c) Margin wear; (d) Crater wear; (e) Chisel edge wear and (f) Chipping at lip.

Hidden Markov models are an extension of Markov chains. A Markov chain is a random process of discrete-valued variables involving a number of states linked by a number of possible transitions. At different times, the system is in one of these states; each transition between the states has an associated probability, and each state has an associated observation output (symbol). The state transition probabilities depend only on the current state, not on past states. Unlike Markov chains, hidden Markov models are *doubly* stochastic processes; i.e., not only is the transition from one state to another state stochastic, but the output symbol generated at each state is also stochastic. Thus the model can only be observed through another set of stochastic processes. The actual sequence of states is not directly observable but is "*hidden*" from the observer. This is why these models are called *hidden Markov models*.

3.1. Elements of an HMM

The following elements, using the same notation as given in Ref. [17], completely characterize an HMM:

1. **N**, the number of states of the model. The set of individual states is denoted as $S=\{S_1, S_2, \ldots, S_N\}$.
2. **T**, the number of observations. A typical observation sequence is denoted as $O=\{O_1, O_2, \ldots, O_T\}$.
3. $\mathbf{Q}=\{q_t\}$, the set of states. q_t denotes the current state such that $q_t \in S$ and $t=1, 2, \ldots, T$.
4. **M**, the number of observation symbols in the alphabet. If the observation space is "*continuous*", then **M** is defined to be infinite.
5. $\mathbf{V}=\{v_1, v_2, \ldots, v_M\}$, the set of observation symbols. The observation symbols per state matches the physical output of the model.
6. $\mathbf{A}=\{a_{ij}\}$, the set of state transition probabilities of the underlying Markov process.
7. $\mathbf{B}=\{b_j(k)\}$, probability distribution of observation symbols for each state.
8. The initial state distribution, $\boldsymbol{\pi}=\{\pi_i\}$, where π_i is the probability of being in state S_i at the beginning of the observation sequence.

If the observations are continuous signals (or vectors), a continuous probability distribution function in the form of a finite mixture is assigned to each state instead of a set of discrete probabilities. In this case, the parameters of the continuous probability distribution function are often approximated by a weighted sum of M Gaussian distributions η,

$$b_j(O_t)=\sum_{m=1}^{M} c_{jm}\eta(\mu_{jm}, U_{jm}, O_t), \quad 1\leq j\leq N \tag{1}$$

where c_{jm} is a weighting (mixture) coefficient, μ_{jm} is the mean vector, U_{jm} is the covariance matrix and M is the number of mixture components.

Note that $\eta(\mu, U, O)$ is a multivariate Gaussian probability density function with mean vector μ and covariance matrix U, that is

$$\eta(\mu, U, O)=\frac{1}{\sqrt{(2\pi)^n|U|}}\exp\left(-\frac{1}{2}(O-\mu)'U^{-1}(O-\mu)\right), \tag{2}$$

where n is the dimensionality of O.

The c_{jm} must also satisfy stochastic constraints, i.e.,

$$c_{jm} \geq 0, \quad 1 \leq j \leq N, \quad 1 \leq m \leq M \tag{3}$$

and

$$\sum_{m=1}^{M} c_{jm} = 1, \quad 1 \leq j \leq N. \tag{4}$$

Thus, the *compact representation* for an HMM with a discrete output probability distribution is given by

$$\lambda = (A, B, \pi) \tag{5}$$

and the model with a continuous output distribution is given by

$$\lambda = (A, c_{jm}, \mu_{jm}, U_{jm}, \pi). \tag{6}$$

To illustrate the ideas, consider a three-state HMM shown in Fig. 3. As seen from the figure, an HMM is defined by a set of transition probabilities and each state has an associated output with a probability distribution b_i defined over a set of output symbols, say A, B, C, D, E and F. In fact, an HMM is a finite-state machine which changes state once every time unit and, at each time t, an output symbol is chosen according to the probability distribution function corresponding to the current state. Additionally, if the discrete number of output symbols is replaced by a continuous-valued vector, the output process is defined by a continuous distribution function over the possible values of the random output vector.

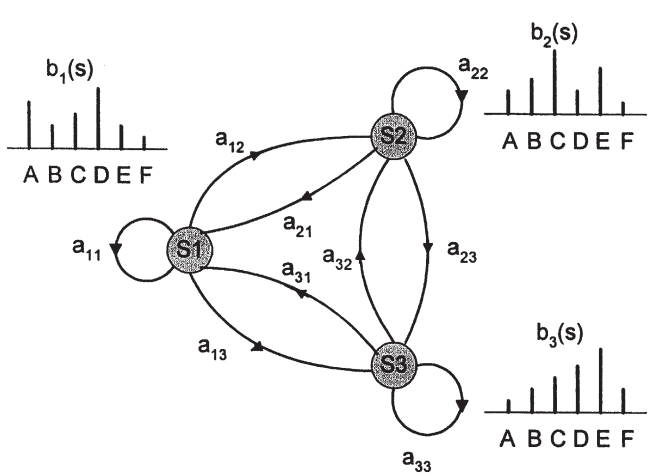

Fig. 3. A three-state hidden Markov model.

3.1.1. Basic assumptions in the theory of HMMs

The following are the basic assumptions in the theory of HMMs.

1. *The Markov assumption*: Similar to Markov chains, the probability of being in a state at any given instant of time depends only on the state at the previous instant of time, i.e.,

$$a_{ij}=P[q_t=S_j|q_{t-1}=S_i]. \tag{7}$$

2. *The stationary assumption*: The state transition probabilities are independent of the actual time at which the transition occurs and can be specified for any t_1 and t_2, i.e.,

$$P[q_{t_1}=S_j|q_{t_1-1}=S_i]=P[q_{t_2}=S_j|q_{t_2-1}=S_i]. \tag{8}$$

3. *The output independence assumption*: The current output (observation) depends only on the current state and it is statistically independent of the previous outputs (observations), i.e.,

$$P[O|q_1, q_2, \ldots, q_T, \lambda]=\prod_{t=1}^{T} P[O_t|q_t, \lambda]. \tag{9}$$

3.2. Three basic problems of HMMs

3.2.1. Evaluation problem

This problem is to calculate the probability of the observation sequences given the model, i.e., $P(O|\lambda)$. This probability can be calculated by using simple probabilistic arguments and is given by

$$P(O|\lambda)=\sum_{q_1, q_2, \ldots, q_T} \pi_{q_1} b_{q_1}(O_1) a_{q_1 q_2} b_{q_2}(O_2) \ldots a_{q_{T-1} q_T}(O). \tag{10}$$

The number of operations for this calculation is of the order of $2T\cdot N^T$. This calculation has exponential complexity with respect to T and is intractable if the number of observations in the sequence, T, is too large. Therefore, a more efficient procedure known in the literature as the *forward–backward procedure* is used to evaluate the required probability, $P(O|\lambda)$. For more details, please refer to Ref. [17].

The calculation of both the forward and backward variables requires computations of the order of $N^2\cdot T$; considerably less complex than the direct calculation. The probability of the observation sequence in terms of these variables is given by

$$P(O|\lambda)=\sum_{i=1}^{N} P(O, q_t=i|\lambda)=\sum_{i=1}^{N} \alpha_t(i)\beta_t(i), \tag{11}$$

where $\alpha_t(i)$ and $\beta_t(i)$ are the forward and backward variables, respectively.

3.2.2. Decoding problem

The decoding problem is to find the most likely state sequence associated with a given sequence of observations, $O=O_1, O_2, \ldots, O_T$ and a model, $\lambda=(A, B, \pi)$ for the discrete observation case or $\lambda=(A, C, \mu, U, \pi)$ for the continuous observation case. A dynamic programming type procedure, the so-called Viterbi algorithm, is used to solve this problem [26]. The details of this algorithm are beyond the scope of this paper.

3.2.3. Training problem

The training problem is to adjust the HMM parameters ($\{a_{ij}, b_j(k), \pi_i\}$ or $\{a_{ij}, c_{jm}, \mu_{jk}, U_{jk}, \pi_i\}$) to maximize the probability of the observation sequence, $P(O|\lambda)$, given the model, $\lambda=(A, B, \pi)$. Iterative learning procdures such as the Baum–Welch method or a gradient-based method can be used to solve this problem. For the Baum–Welch algorithm, it is necessary to make an initial guess of a suitable set of HMM parameters first, then re-estimation formulae are used to find more accurate parameters in the sense of maximum likelihood. They are

$$\bar{\pi}_i=\gamma_1(i), \quad 1\le i\le N \tag{12a}$$

$$\bar{a}_{ij}=\frac{\sum_{t=1}^{T-1}\xi_t(i,j)}{\sum_{t=1}^{T-1}\gamma_t(i)}, \quad 1\le i\le N,\ 1\le j\le N \tag{12b}$$

$$\bar{b}_j(k)=\frac{\sum_{t=1}^{T}\gamma_t(j)}{\sum_{t=1}^{T}\gamma_t(j)}, \quad 1\le j\le N,\ 1\le k\le M \tag{12c}$$

where $\xi_t(i, j)$, *the state transition probability*, is the probability of being in state S_i at time t and state S_j at time $t+1$, given the model and observation sequence. $\xi_t(i, j)$ can be expressed in terms of the forward and backward variables by

$$\xi_t(i,j)=\frac{\alpha_t(i)a_{ij}\beta_{t+1}(j)b_j(O_{t+1})}{\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_t(i)a_{ij}\beta_{t+1}(j)b_j(O_{t+1})}. \tag{13}$$

The other variable in the equations is $\gamma_t(i)$, *the state probability*, which is the a posteriori probability of being in state S_i at time t, given the observation sequence and the model. $\gamma_t(i)$ can also be expressed in terms of the forward and backward variables, i.e.,

$$\gamma_t(i)=\frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^{N}\alpha_t(i)\beta_t(i)}. \tag{14}$$

The re-estimated model, $\bar{\lambda}=(\bar{A}, \bar{B}, \bar{\pi})$ or $\bar{\lambda}=(\bar{A}, \bar{C}, \bar{\mu}, \bar{U}, \bar{\pi})$, satisfies the property that $P(O|\bar{\lambda})>P(O|\lambda)$, where $\bar{\lambda}$ are the new model parameters and λ are the current model parameters. Starting from a given set of model parameters, the procedure of updating the HMM parameters is repeated until the estimates do not change significantly. The procedure may need to be run several times from different initial guesses of the parameters to obtain an appropriate HMM model, because the Baum–Welch algorithm only guarantees finding local maximum optima.

3.3. Types of HMM

Three types of HMM are considered:

1. ergodic model (fully connected model);
2. left-to-right model (Bakis model); and
3. two parallel left-to-right model.

The *ergodic model* has the property that a transition from any state to any other state can be made in a finite number of steps. Consequently, all entries of the transition matrix (a_{ij}) are positive (see the first model in Fig. 4).

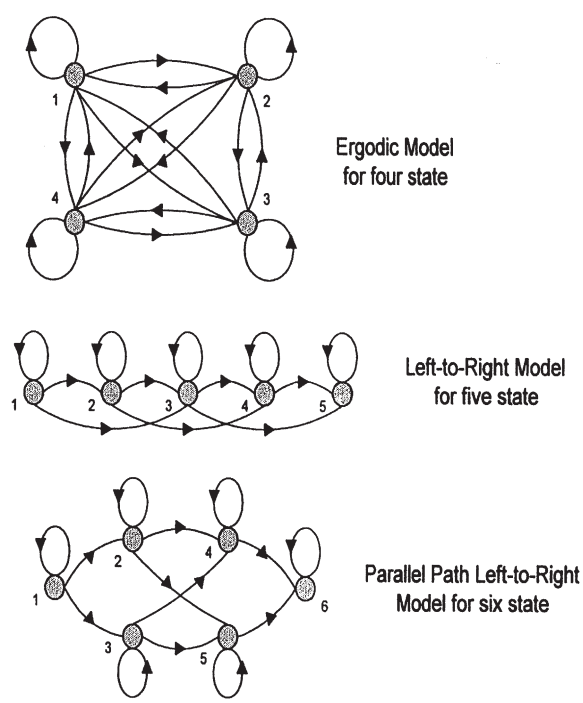

Fig. 4. Types of HMM.

The *left-to-right model* has the property that the next state index is always greater than or equal to the current state index as time increases. The states then proceed from left to right as seen in the second model in Fig. 4. The state transition matrix is triangular and the entries, a_{ij}, satisfy the following constraints:

1. $a_{ij}=0, \quad j<i$
2. $a_{ij}=0, \quad j>i+\Delta$
3. $\pi_i=\begin{Bmatrix}0, & i\neq 1\\ 1, & i=1\end{Bmatrix}$
4. $a_{NN}=1$ and $a_{Ni}=0, \quad i<N.$

The first property implies that the transitions are unidirectional, i.e., no transitions are allowed to states whose index is lower than the current state. The state transition matrix, $A=[a_{ij}]$, is upper triangular and the state evolution is such that each state can only have a transition to itself or to a subset of neighboring states determined by Δ as stated in the second property. According to the third property, the initial state is always the first state. The last property comes from the definition of the left-to-right model. Note that the final state is a trapping state.

The *two parallel left-to-right models* have all the a_{ij} constraints of the left-to-right model, except there are two parallel paths as seen in the third model in Fig. 4.

3.4. Scaling

The computation of probabilities for the forward and backward variables [given by α_t and β_t in Eq. (11), respectively] involves taking the product of a large number of probabilities as time increases. The *P*-values can become extremely small quickly, and this can lead to underflow in computer implementations. Hence, in order to avoid numerical problems, the variables in the forward–backward procedure need to be scaled using

$$c_t=\frac{1}{\sum_{i=1}^{N}\alpha_t(i)}, \tag{15}$$

which is independent of the states but depends on t. Then, scaled forward variables, $\hat{\alpha}_t(i)$, become

$$\hat{\alpha}_t(i)=c_t\alpha_t(i)=\frac{\alpha_t(i)}{\sum_{i=1}^{N}\alpha_t(i)}. \tag{16}$$

The same scaling factor can be used for backward variables, i.e.,

$$\hat{\beta}_t(i)=c_t\beta_t(i). \tag{17}$$

Note that the scaling factors do not affect the re-estimation formulas, since they appear in both

the numerator and the denominator. Even though the total effect of the scaling is removed at the end of the computation, the procedure for computing $P(O|\lambda)$ is affected by the scaling. Recall that

$$P(O|\lambda)=\sum_{i=1}^{N}\alpha_T(i). \tag{18}$$

Then, from the definition of the scaling factor, this expression can be written as

$$\prod_{t=1}^{T} c_t \sum_{i=1}^{N}\alpha_T(i)=\prod_{t=1}^{T} c_t \cdot P(O|\lambda)=1. \tag{19}$$

So,

$$P(O|\lambda)=\frac{1}{\prod_{t=1}^{T} c_t}. \tag{20}$$

Finally,

$$\log[P(O|\lambda)]=-\sum_{t=1}^{T}\log c_t. \tag{21}$$

This formula is very useful, because even if the probability of an observation sequence for a given HMM is a very small number, the *log* of P can still be in the range of the machine. The other issues associated with the implementation of HMMs are the following: dealing with multiple observation sequences, initializing parameter estimates, missing data and the choice of the model size and type; refer to Ref. [17] for more details.

4. Experimental set-up

The experimental set-up used in this study is illustrated in Fig. 5. Drilling life tests were conducted on an MAHO 700S computer numerical controlled (CNC) five-axis machining center, which has movement in three perpendicular axes and a rotary/tilt table. A special CNC program was run to track force and power signals as a function of wear. The thrust force (F_z) and torque (T) were measured by a Kistler two-component dynamometer (Kistler model 9271A). The dynamometer consists of a two-component force transducer fitted under high preload between a base plate and a top plate. The force transducer contains two sets of quartz disks arranged in a circular configuration. One set is sensitive to pressure in the **z** direction and yields a charge proportional to F_z (thrust), while the other responds to shear stress and gives a charge proportional to T (torque). The resulting charge signals are converted into output voltages proportional to the forces sustained and then these voltage signals are amplified by two charge amplifiers. The power cell, a motor load sensor, was used to measure servo power and spindle power in terms of horsepower (HP) or Watts.

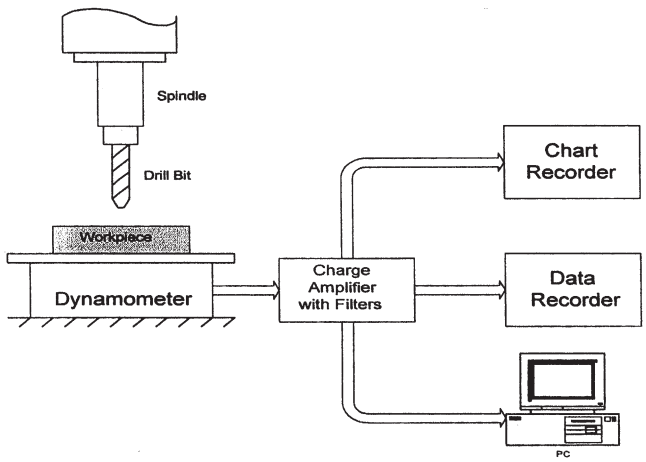

Fig. 5. Schematic diagram of the experimental set-up and data-acquisition system.

Force and power signals were recorded simultaneously to a tape by a Sony PC216Ax data recorder. Four of its 16 analog input channels were used to collect the data by sampling the signals at 12 kHz. As seen from Fig. 5, collected signals could be monitored in real time either on the computer screen or on the chart recorder. The data recorded were loaded into a Pentium II computer via PCScan MKII software for data analysis after the test. The data were resampled at 1 kHz and saved to the hard disk of the computer.

Holes were drilled on a steel workpiece material using an M7 cobalt split-point twist drill bit, which has a diameter of 5/16 in. and a 135° point angle. The holes were drilled to a depth of 1/2 in. The feed rate was kept constant at 11 in/min and different spindle speeds in the range of 1825 to 2500 rev/min were used to obtain different tool wear rates. The cutting conditions (speed, feed rate, depth of hole) remained constant throughout the tests.

All tests were performed without using any cutting fluid, i.e., under dry cutting conditions. The workpiece material was fairly homogeneous and its geometrical features were not considered.

5. Drill wear monitoring methods

As pointed out previously, cutting forces (both thrust and torque) are very useful for drill wear monitoring [27] because these forces generally increase as tool wear increases. Thus, within the tool wear region, cutting forces provide a good assessment of the tool conditions. If the cutting tool cannot withstand the increased cutting forces, catastrophic tool failure becomes inevitable. Consequently, tool life, which is a direct function of tool wear, is best determined by monitoring both torque and thrust force. Generally, thrust and torque magnitudes are 50% larger when machining the last hole than when machining the first hole [28]. Moreover, serious tool wear can cause 50% or 100% increase in the amplitude of the force signals [27]. Fig. 6 illustrates the latter case, showing the force signals for a new drill and a dull drill when drilling holes in a steel workpiece.

As a drill wears, Fig. 6 also shows that not only the amplitude but also the signature of the signals changes according to the type of drill wear. There is a close relationship between the

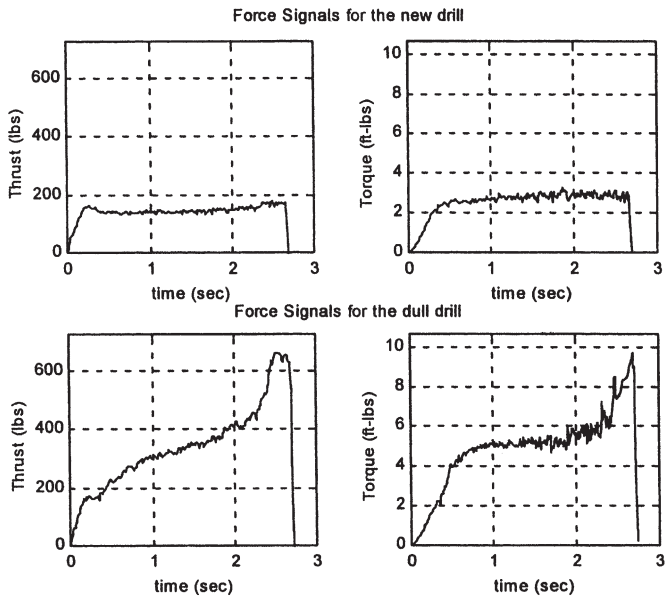

Fig. 6. The force signals for the new and dull drill.

dynamic components of the force signals and tool wear. Hong et al. [29] classified the dynamic properties of different kinds of tool wear and failure wear types for thrust, torque and vibration signals. For instance, corner wear causes some thrust fluctuation and lip height difference results in periodicity in the torque signals.

Because force signal measurements using a dynamometer may not be available in most machining applications, a cost-effective method for tool condition monitoring is necessary. In general, the process of tool wear impairs the sharpness of the tool's cutting edge, increases the friction between the tool and workpiece, and also increases the power consumption. There are commercially available power cells that can measure spindle or servo power delivered to the motor by sensing the instantaneous power change into a machine or process. These devices are relatively inexpensive and are easily installed in the machine without any intrusive procedures.

Hidden Markov models can be considered as a method of modeling pattern classes that consist of time-series data and then comparing patterns to recognize or classify new data. This idea can be applied to tool wear condition monitoring. Here, the force and power signals are the time series used to build the HMM and to monitor the machining process. If a model is developed for a sharp drill, then force or power signals obtained throughout the drilling process could be used to monitor the dullness of the drill by computing the probabilities generated by the HMM. Another way of classifying the data signals is to develop multiple models for various states of drill wear such as *sharp*, *workable* and *dull*, or a similar set of labels. In this paper, two methods are proposed using HMMs for tool wear condition monitoring in drilling operations:

1. the bargraph method; and
2. the multiple modeling method.

For additional details, refer to [15].

5.1. The bargraph method

In this method, after drilling a hole in a workpiece material with a sharp drill, the time-series data are used to develop HMMs by training all types of data signal including thrust, torque, spindle power and servo power, separately. Developing an HMM requires determining the HMM parameters: A, C_{jm}, μ_{jm}, U_{jm} and π. Once these parameters are determined, time-series data obtained from the drilling process can be used to evaluate the health of the tool. During the drilling process, measured data signals are processed continuously using the corresponding HMM; e.g., measured torque signals are processed in the HMM developed for torque and so on. Each model generates a probability that quantifies the similarity between the current measurement signal and the signal used for building the model, or the signals obtained from a sharp drill. The probabilities generated by the model are represented in a logarithmic scale because the values of these probabilities can be extremely small, for example of the order of 10^{-50} for the processing of thrust data near to a drill breakage condition. In an alternative representation, the probabilities generated from each of the HMMs corresponding to a sharp drill are normalized to *one*, thus the probabilities take values between 0 and 1 as shown in Fig. 7.

The more the drill wears, the more the data signals (force and power signals) differ from those signals obtained for a sharp drill, in terms of both magnitude and shape. When these data are processed through the corresponding HMM, the probabilities begin to decline indicating the wearing process of the drill. If the normalized probabilities decrease below a certain threshold, it can be concluded that the drill bit has accumulated excessive wear and should be replaced before breakage occurs.

5.2. The multiple modeling method

This method is based on classifying the data signals using an HMM to predict the wear status of the drill. The data signals measured from the sensors during drilling are dependent upon machining variables such as spindle speed, feed rate, type of workpiece material, depth of cut, etc. However,

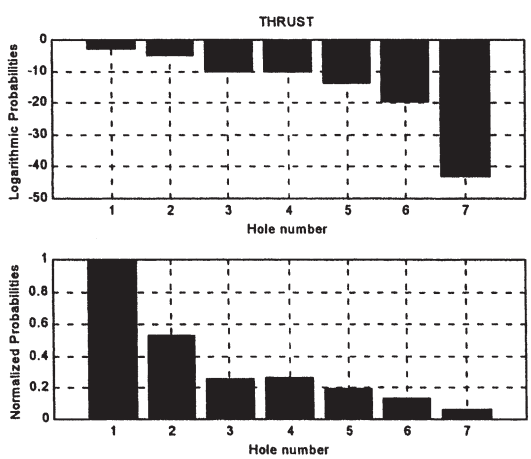

Fig. 7. Application of the bargraph method for a sample drill.

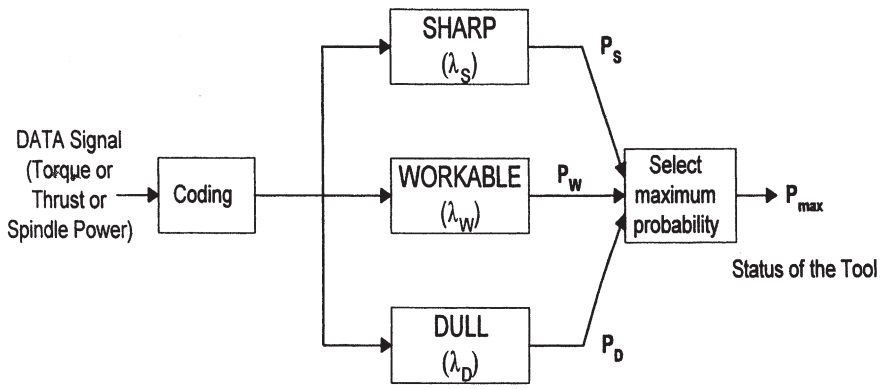

Fig. 8. Block diagram for the multiple modeling method.

under the same operating conditions, the data signals (thrust, torque, spindle power and servo power) are expected to be quite similar with respect to shape and magnitude for identical drill types.

In this approach, different models such as *sharp*, *workable* and *dull* are developed for the distinct cases of a test drill. As mentioned earlier, wearing is a progressive process that has different stages such as initial, slight, moderate, severe and worn-out. Once these stages have been determined in a test drill, hidden Markov models are developed for the various stages of wear; e.g., *workable*, which corresponds to initial–slight–moderate wear, and *dull*, which corresponds to severe and worn-out stages of wear.

These models can be used to monitor and track the wear status of any identical tool under the same machining conditions. The data signals are processed in the associated models. The HMM that gives the maximum probability indicates the wear status of the tool, i.e., whether the drill is still sharp, or workable, or dull, as illustrated in Fig. 8.

6. Experimental results and discussion

Several sets of data were collected by drilling holes in a steel workpiece under different experimental conditions using a split-point angle twist drill. The results of the proposed methods will be presented for three sample data sets running under the same experimental conditions, which are summarized in Table 1.

Table 1
Experimental conditions

Test no.	Spindle speed (rev/min)	Feed rate (in/min)	Depth of hole (in)	Workpiece	Drilled holes
1	1825	11	1/2	Steel	7
2	1825	11	1/2	Steel	69
3	1825	11	1/2	Steel	93

Each experiment started with a new/sharp drill and drills were allowed to wear until final breakage occurred by drilling many holes into two separate workpieces: one of the workpieces was attached to a dynamometer, the other one was not. The thrust and torque signals were collected only when holes were drilled in the workpiece mounted on a dyno table. Unlike the force signals, power signals were available for measurement when drilling into both materials.

6.1. Results of the bargraph method

6.1.1. Thrust and torque

Fig. 9 shows the results of the bargraph method using force signals. The run numbers correspond to the number of holes drilled into the steel workpiece attached to the dynamometer. In the first test, the drill was intentionally dulled from the corners before drilling the second hole and third holes in order to accelerate the tool wear rate. Once the drill reached an excessive wear stage, the rest of the holes were drilled without interrupting the process until the drill broke at the seventh hole. The tool wear occurred naturally in the second test and the run numbers corre-

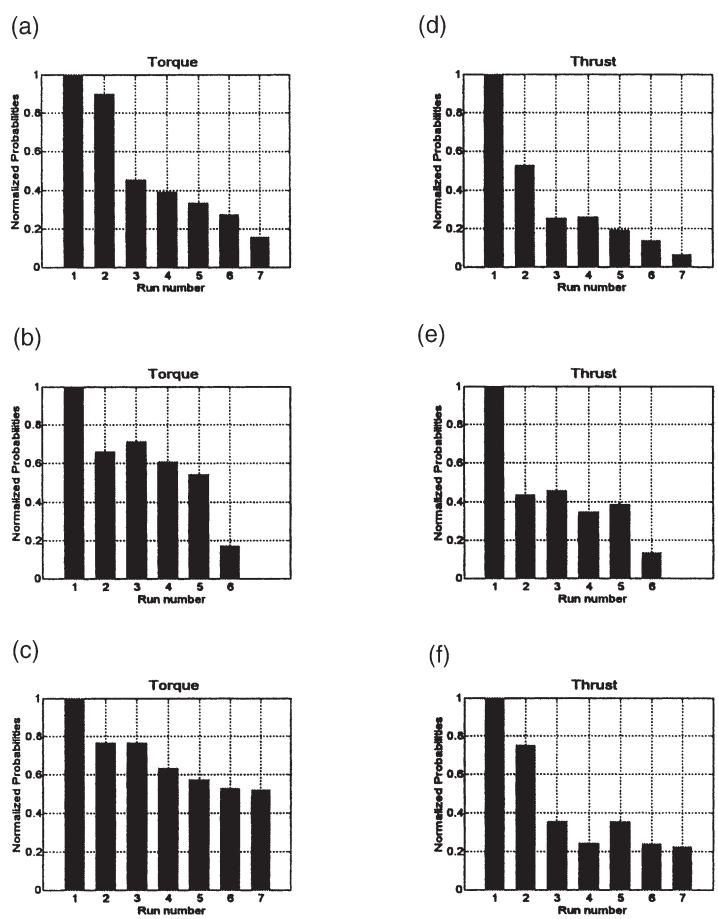

Fig. 9. The results of the bargraph method for force signals. (a and d) Test 1; (b and e) Test 2 and (c and f) Test 3.

spond to drilling of holes 1, 14, 28, 41, 55 and 68, respectively. The force signals were obtained when drilling holes in the workpiece located on the dyno table. After drilling a single hole, the tool holder was shifted to another similar workpiece, a series of holes was drilled where only spindle power signals were collected, and then the tool holder was shifted back to the dyno table. The last measured force signals correspond to the drilling of the hole which precedes tool breakage. The condition of the tool can be clearly monitored from the normalized probabilities generated by the HMM as seen in Fig. 9(a), (b), (d) and (e), which show the bargraph results using torque and thrust for Tests 1 and 2.

The run numbers for Test 3 correspond to the drilling of holes 1, 15, 28, 42, 55, 69, and 82, respectively. Note that, even though the drill was destroyed at the 93rd hole, the last sample of torque and thrust data was at the 82nd hole and the drill was still in a workable condition. Therefore, the last normalized probabilities were still high compared with the ones in the previous tests.

6.1.2. Spindle power

Before applying the bargraph method using HMMs, spindle power signals need to be preprocessed in order to eliminate the oscillation associated with the spindle power signals in the experiments. Thus, the spindle power signals were first filtered using a zero-phase FIR filter and then the signals were evaluated using the HMM to produce normalized probabilities that track the tool wear in the drill, refer to Fig. 10. Despite the filtering, the fluctuation of the bars is the result of natural oscillation of the spindle power signals. Recall that in Test 3 force signal data were not available for drilling holes when the tool was near to failure. But, using the spindle power signal, it can be observed from Fig. 10(c) that the probabilities begin to decrease abruptly as an incipient indication of imminent tool breakage.

Because servo power signals do not change significantly throughout the drilling process, the bargraph method is not applicable.

6.2. Results of the multiple modeling method

6.2.1. Torque and thrust

Three models (sharp, workable and dull) were developed based on the force signals obtained in Test 1. Tables 2 and 3 list the probabilities generated by the torque and thrust signals, respectively, in Test 2. Here, bold numbers in the tables indicate the maximum probabilities that determine the status of the tool (the wear condition of the tool) as illustrated in Fig. 8. As seen from both of these tables, the sharp and dull status of the tool is determined exactly from the first and last force signals, respectively. Even though the HMMs for torque and thrust provide different decisions in the second and third runs, the tool is still categorized as being usable: either as sharp or workable.

For the same models, Tables 4 and 5 list the multiple modeling results of torque and thrust in Test 3, respectively. Although the drill broke at the 93rd hole in this test, the last data were collected at the 82nd hole where the tool was still in a workable condition. As will be explained in the next section, spindle power signals are necessary to determine the status of the tool for this test.

Note that the probabilities given in the tables are logarithmic probabilities because the numerical

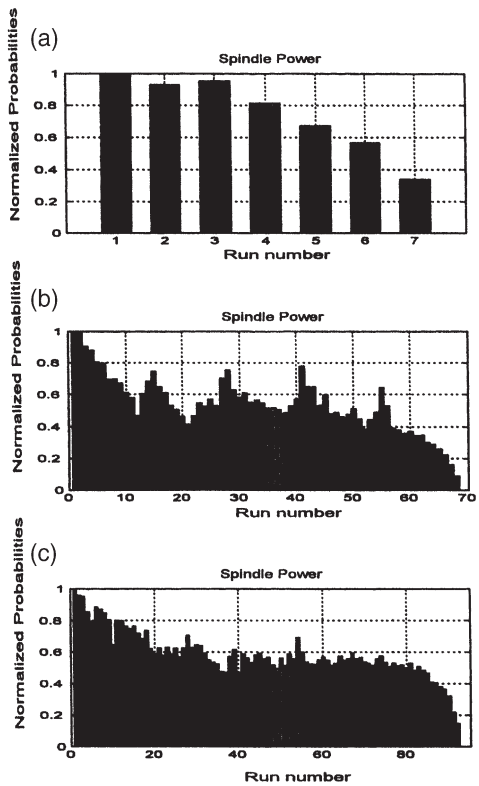

Fig. 10. The results of the bargraph method for the spindle power signals. (a) Test 1; (b) Test 2 and (c) Test 3.

Table 2
The multiple modeling method results of torque signals for Test 2

Run no.	Sharp	Workable	Dull	Status
1	**−5.08**	−5.72	−7.50	Sharp
2	**−5.29**	−5.38	−6.60	Sharp
3	**−5.26**	−5.39	−6.64	Sharp
4	−5.36	**−5.33**	−6.43	Workable
5	−5.42	**−5.33**	−6.38	Workable
6	−6.99	−5.74	**−5.59**	Dull

values of the actual probabilities generated by the models are very small; e.g., $10^{-5.08} \simeq 0.0000083176$ for the sharp case in Table 2. In general, as the tool wears, the probabilities generated by the λ_S (HMM for the *sharp* case) decrease and the probabilities generated by the λ_D (HMM for the *dull* case) increase. The multiple modeling method using the thrust signal seems to provide a better estimate about the tool wear status than the one which uses the torque signals. The second run and third run data signals were collected after drilling 13 holes and 27 holes, respectively, and the tool was already worn to some extent when drilling these particular holes.

Table 3
The multiple modeling method results of thrust for Test 2

Run no.	Sharp	Workable	Dull	Status
1	**−2.68**	−2.71	−4.11	Sharp
2	−3.70	**−2.71**	−3.23	Workable
3	−3.70	**−2.69**	−3.20	Workable
4	−4.08	**−2.84**	−3.07	Workable
5	−3.93	**−2.79**	−3.12	Workable
6	−6.66	−4.30	**−2.90**	Dull

Table 4
The multiple modeling method results of torque for Test 3

Run no.	Sharp	Workable	Dull	Status
1	**−5.11**	−5.72	−7.18	Sharp
2	**−5.24**	−5.41	−6.60	Sharp
3	**−5.25**	−5.38	−6.55	Sharp
4	−5.35	**−5.32**	−6.39	Workable
5	−5.40	**−5.36**	−6.44	Workable
6	−5.46	**−5.29**	−6.26	Workable
7	−5.50	**−5.30**	−6.25	Workable

Table 5
The multiple modeling method results of thrust for Test 3

Run no.	Sharp	Workable	Dull	Status
1	**−2.55**	−3.00	−5.29	Sharp
2	−2.69	**−2.68**	−4.40	Workable
3	−3.12	**−2.62**	−3.70	Workable
4	−3.51	**−2.68**	−3.32	Workable
5	−3.17	**−2.58**	−3.54	Workable
6	−3.58	**−2.70**	−3.27	Workable
7	−3.66	**−2.73**	−3.25	Workable

However, the tool was still workable for the rest of the holes. Thus, the estimate of tool status for the second and third run appears more reasonable in the case when the thrust signal is used.

6.2.2. Spindle power

Similar to the force signals, the models associated with the different wear conditions of the tool are developed for the spindle power signals obtained in Test 1. Figs. 11 and 12 show the results of the multiple modeling method using spindle power signals for each hole drilling in Test 2 and Test 3, respectively. Because the workpiece material was steel in these tests, the drill

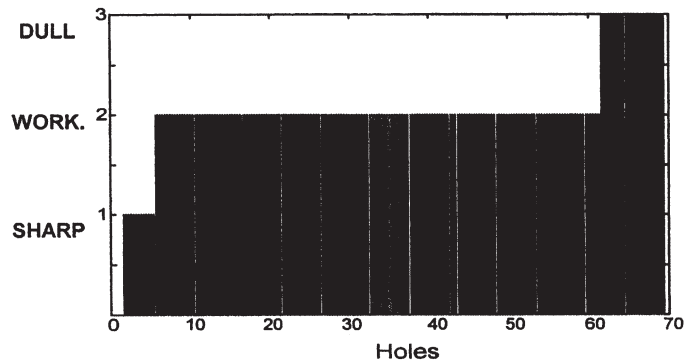

Fig. 11. The multiple modeling method results of spindle power for Test 2.

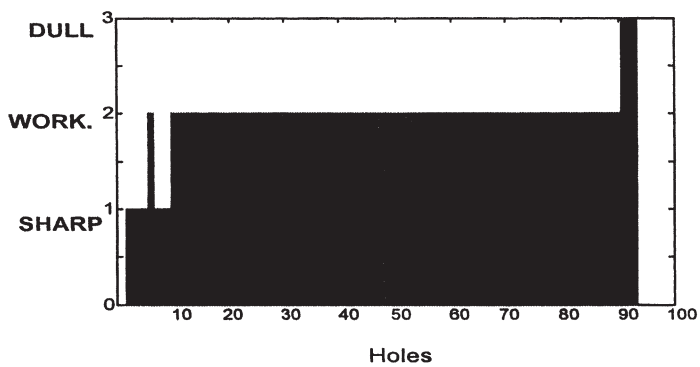

Fig. 12. The multiple modeling method results of spindle power signals for Test 3.

became worn rapidly after a couple of holes and it kept the workable status until very close to the end of its life. As seen from the figures, the multiple modeling method predicts efficiently the dull state before tool breakage.

It is important to develop models using suitable signals, especially for the dull case model. As long as the tool is workable, it does not matter whether the tool status given by the model is sharp or workable. The problem is how to select the signals to develop a workable model because every type of tool wear affects the torque and thrust signals differently. To overcome this issue, HMM parameters are generated for different tool wear types and the average of these parameters can be taken to develop a final HMM for the *workable* case. Because only corner wear is considered in this investigation, the *workable* HMM was developed by training the data based on acceptable corner wear for the drilling operation.

7. Conclusions

It must be emphasized that the ultimate purpose of tool wear condition monitoring is to determine the exact time when a tool should be changed before breakage or more serious wear conditions that can adversely affect the machining process occur — independent of the type of wear.

Because both torque and thrust forces are very sensitive to tool wear and cutting powers are functions of force signals, the *bargraph method* or the *multiple modeling method* is potentially feasible for an on-line tool condition monitoring (TCM) system in a drilling process. On the other hand, due to the lack of force measurement systems, which require a dynamometer, it is necessary to develop a monitoring system based on power signal measurements. In order to apply the methods proposed in this paper, the power signals need to be preprocessed in order to eliminate fluctuations which can impair the effectiveness of HMM application.

The experimental results show that the progress of tool wear can be monitored by using the bargraph method for torque, thrust and spindle power signals; or the status of the tool can be predicted by the multiple modeling method for torque and thrust signals, or spindle power after filtering. The probabilities for the thrust signal generated in the bargraph method and the multiple modeling method have a wider range between the sharp and dull status of tool. Therefore, it can be concluded that thrust signals are better indicators of tool wear than torque signals in this particular application.

The force and power signals are measured signals which indirectly track tool wear. These signals are affected by the cutting conditions, workpiece material and type of tool. For example, increasing the feed rate during the drilling process leads to a proportional increase in the measured signals. This situation could be confused with increased forces due to tool wear in the system.

Finally, the methods proposed in this paper can also be used for monitoring other machining operations such as turning, milling and grinding as long as there are readily available data signals that are sensitive to tool wear.

Acknowledgements

The authors would like to thank John Gorse and Karl Ziegler from Eaton MTC for providing technical support and access to their laboratoy for the experimental part of our work.

References

[1] J. Pletting, Superior Tooling Builds Outstanding Workmanship, Pletting and Associates, 1999, www.pletting.com.
[2] US Industrial Outlook 1992: Business Forecasts for 350 Industries, US Department of Commerce, International Trade Administration, 1992.
[3] K. Subramanian, N.H. Cook, Sensing of drill wear and prediction of drill life, ASME, Journal of Engineering for Industry 99 (1977) 295–301.
[4] S.Y. Hong, Knowledge-based diagnosis of drill conditions, Journal of Intelligent Manufacturing 4 (1993) 233–241.
[5] E. Govekar, I. Grabec, Self-organizing neural network application to drill wear classification, Journal of Engineering for Industry 116 (May) (1994) 233–238.
[6] T.I. Liu, K.S. Anantharaman, Intelligent classification and measurement of drill wear, Journal of Engineering for Industry 116 (1994) 392–397.
[7] R. Isermann, M. Ayoubi, H. Konrad, T. Reib, Model Based Detection of Tool Wear and Breakage for Machine Tools, in: Proceedings of the IEEE International Conference Syst. Man Cybern., IEEE, Piscataway, NJ, vol. 3, 1993, pp. 72–77.
[8] K. Danai, A.G. Ulsoy, Dynamic modeling of cutting forces in turning, milling and drilling, from experimental data, Control Methods for Manufacturing Processes 9 (1988) 27–34.

[9] R.J. Furness, C.L. Wu, A.G. Ulsoy, Dynamic modeling of the thrust force and torque for drilling, ACC/WA11, pp. 384–390, 1992.
[10] H. Pan, Y. Chen, E. Orady, Monitoring methods of tool wear in a drilling process, in: 2nd Industrial Engineering Research Conference Proceedings, 1999, pp. 380–383.
[11] S. Kawaji, Y. Suenaga, T. Maeda, N. Mmatsunaga, T. Sasaoka, Control of cutting torque in the drilling process using disturbance observer, in: American Control Conference, Seattle, WA, June, 1995, pp. 723–728.
[12] R.J. Furness, A.G. Ulsoy, C.L. Wu, Feed, speed, and torque controllers for drilling, Journal of Engineering for Industry 118 (1996) 2–9.
[13] J. Kim, S. Lee, Y. Park, Development of a drilling process with torque stabilization, Journal of Manufacturing Systems 13 (6) (1999) 435–441.
[14] B.Y. Lee, H.S. Liu, Y.S. Tarng, Modeling and optimization of drilling process, Journal of Material Processing Technology 74 (1998) 149–157.
[15] H.M. Ertunc, Techniques for tool wear condition monitoring in drilling operations, Case Western Reserve University, 1999.
[16] M. Kanai, K. Inata, S. Fujii, Y. Kanda, Statistical characteristics of drill wear and drill life for the standardized performance tests, CIRP Annual 27 (1) (1988) 61–66.
[17] L.R. Rabiner, A tutorial on Hidden Markov models and selected applications in speech recognition, Proceedings of IEEE 77 (2) (1989) 257–286.
[18] R.J. Elliot, L. Aggoun, J.B. Moore, Hidden Markov Models, Estimation and Control, Springer, Berlin, 1995.
[19] Y. BarShalom, X.R. Li, Estimation and Tracking: Principles, Techniques, and Software, Artech House, Boston, MA, 1993.
[20] L.M.D. Owsley, L.E. Atlas, G.D. Bernard, Self-organizing feature maps and Hidden Markov models for machine-tool monitoring, IEEE Transactions on Signal Processing 45 (11) (1997) 2787–2798.
[21] P. Smyth, Hidden Markov models for fault detection in dynamic systems, Pattern Recognition 27 (1) (1994) 149–164.
[22] Y. Zhang, X.R. Li, K. Zhou, A fault detection and diagnosis approach based on Hidden Markov chain model, in: Proceedings of the American Control Conference, Philadelphia, June, 1998, pp. 2012–2016.
[23] B. Hannaford, P. Lee, Hidden Markov model analysis of force/torque information in telemanipulation, The International Journal of Robotics Research 10 (5) (1991) 528–539.
[24] G.E. Hovland, B.J. McCarragher, Hidden Markov models as a processor monitor in robotic assembly, The International Journal of Robotics Research 17 (2) (1998) 153–168.
[25] H.J. Kim, S.K. Kim, K.H. Kim, J.K. Lee, An HMM-based character recognition network using level building, Pattern Recognition 30 (3) (1997) 491–502.
[26] A.J. Viterbi, Error bounds for convolutional codes and an asymptotically optimal decoding algorithm, IEEE Transaction on Information Theory 13 (1967) 260–269.
[27] G.S. Li, W.S. Lau, Y.Z. Zhang, In-process drill wear and breakage monitoring for a machining centre based on cutting force parameters, International Journal of Machine Tools Manufacturing 32 (6) (1992) 855–867.
[28] S.A. Jalali, W.J. Kolarik, Tool life and machinability modes for drilling steels, International Journal of Machine Tools Manufacturing 31 (3) (1991) 273–282.
[29] S.Y. Hong, J. Ni, S.M. Wu, Analysis of drill failure modes by multi-sensors on a robotic end effector, ASME in: Proceedings of 1992 Japan/USA Symposium on Flexible Automation, San Francisco, CA, vol. 2, 1992, pp. 947-956.