Hidden Markov models for tool wear monitoring during turning operations

by

G van den Berg and P S Heyns
Dynamic Systems Group
Department of Mechanical and Aeronautical Engineering
University of Pretoria
Pretoria

Abstract

To date the use of neural networks has been the most popular approach to monitoring tool condition ~~monitoring. Recently,~~but hidden Markov models have recently been proposed for this ~~application.~~purpose. Hidden Markov models have excellent capabilities for capturing the temporal and spatial characteristics of signals. This article presents a method of using hidden Markov models and dimensional reduction to classify tool wear during a turning operation. ~~A turning~~An experiment on turning was conducted where a strain signal was measured on the tool holder, using two strain gauges from a strain gauge rosette. The two gauges were oriented perpendicularly and were connected in a ~~half~~ half-bridge configuration. The rosette was ~~located in such a manner~~positioned so that it measured the feed force during cutting. A set of signal features was extracted from the sensor signal. These comprised statistical time domain features and spectral band features. The most consistent signal features were identified using a correlation coefficient approach. The signal dimensionality was then reduced to one with the help of a principal component analysis. From the data twoTwo classes were identified from the data and a hidden Markov model was trained for each class. The ~~optimal state~~optimal-state topology was found by using an exhaustive method. The hidden Markov models were tested and trained from different training sequences. During the testing the hidden

Markov model trained on ~~sharp tool~~sharp-tool data achieved ~~of~~a more than 90~~-~~% correct classification. The hidden Markov model trained on data from a worn tool also achieved a greater than 90% correct classification. ~~This is compared~~These results compare favorably to a Bayesian classifier that achieved a 96.4% correct classification of sharp tools but only a 73.7% correct classification of worn tools.

1. Introduction

Tool condition monitoring (TCM) strives to achieve a generic system that can be applied to machining processes to monitor the condition of the machining tools. By far the most research has gone into indirect methods ~~that measure~~for measuring process parameters that are in turn related to tool ~~wear. This is~~wear, because indirect methods employ parameters that are easier and cheaper to measure.

Indirect process parameters such as vibration or strain on the tool holder, or spindle motor current are unfortunately also typically non-linear and chaotic, according to Byrne et al. [1]. When signals of the process are compressed into characteristic features, these features usually show only slight changes with the progression of~~tool wear. Performing~~as tool wear progresses. The classification or evaluation of tool wear ~~with~~from these features is therefore complex and computationally ~~expensive and calls for~~expensive, and requires the use of advanced methods. This is why artificial neural networks (ANNs) have become the standard approach for signal classification in TCM. ANNs have the ability to extract the underlying knowledge and relations from training ~~samples. Complex partial patterns can be dealt with very succesfully.~~samples, and can deal very successfully with complex partial patterns. Unfortunately, the technique is not always well adapted to dealing with temporal variations ~~of~~in the training samples.

Hidden Markov models (HMMs) are commonly used in speech recognition systems, but have only recently been advocated for ~~classification of~~classifying the faults in mechanical systems. They are advocated because ~~of their ability to handle~~they can easily handle the spatial and temporal variations in patterns~~ very well~~. They also have a ~~very well developed~~well-developed mathematical structure and are easy to train. HMMs work well

with mechanical systems because they allow for a modular construction of classification systems. This means that improvements ~~on a system can easily be made.~~can readily be made to a system. Since the tutorial article by Rabiner [2], HMMs have become more approachable and~~has seen~~ a number of applications to mechanical systems have been seen.

Kwon and Kim [3] produced an accident classification system ~~that used a self organising~~using a self-organizing map (SOM) to fuse ~~twenty two~~22 input variables into a single feature ~~that is~~which was classified by a set of HMMs. Each HMM corresponded to ~~an~~a particular accident type that ~~was~~had to be identified. Atlas et al. [4] devised a "widely applicable and reconfigurable, real time transient classifications system" that can could predict failures in a manufacturing environment. It was found that signals from longer time periods produced more consistent results. Bunks et al. [5] used a continuous HMM with 8-dimensional state distributions~~,~~ to classify helicopter gearbox faults, using eight accelerometers. The HMMs were trained and tested at different torque levels with seeded defects. HMMs ~~are then~~were shown to have great potential but ~~are~~had not been applied to a real model. Ertunc et al. [6] created a ~~drill wear~~system for monitoring ~~system~~drill wear, using HMMs applied to a bar graph method and a multiple model method. Both methods ~~are~~were found to have potential for predicting the status of wear ~~of~~on a given tool. Wang et al. [7] approached tool wear monitoring through a discrete wavelet transformation (DWT). An HMM was trained on a codebook created from the DWT. This was shown to be effective for a binary ~~tool state assignment.~~assignment of tool state. Lee et al. [8] produced a rig on which rotor faults could be simulated using HMMs. ~~Both continuous~~Continuous and discrete HMMs were trained and tested on autospectrum data from accelerometers on the rig. ~~Discrete~~The discrete HMMs were found to produce more robust recognition results. Ge et al. [9] used HMMs and auto-regressive models (AR) to diagnose faults in a stamping process.

This study ~~will use~~used multiple signal features from a sensor signal recorded from a 90° strain gauge rosette, to perform a binary~~sharp/worn~~ classification between sharp/worn on a turning tool. Much research has gone into the use of ANN technologies in TCM systems and subsequently, ~~much~~a great deal of research has gone into the use of multiple

sensors and sensor fusion. In order ~~that~~to make good use~~is to be made~~ of the research that has been done with ANN technologies, HMM methods ~~should be able to handle~~ought to be capable of handling multiple features in a similar way. In this study an open source HMM toolbox for Matlab was used for the classification [10]. A dimensional reduction scheme was ~~In~~introduced in order to implement a multiple feature classification on this ~~toolbox, adimensional reduction scheme was introduced.~~
toolbox.

2. Hidden Markov models

The HMM consists of a number of states, say N, ~~which~~where each in turn contains the description ~~for a state process.~~of a process state. At any time the HMM can occupy only one state. For a ~~first-order~~first-order HMM the current state ~~is only dependent~~depends only on the previous state. This dependency is~~a~~ probabilistic~~one~~ and is described within a state transition matrix A, for which the elements are given by:

$$a_{ij} = P(q_{t+1} = S_j \mid q_t = S_i) \quad 1 \le i, j \le N \qquad (1)$$

In this equation a_{ij} is the probability that the HMM, currently in state i will transit to state j for the next time step. The state S_i occupied by the HMM at time instant t is denoted as q_t . The idea of state transition probabilities can be ~~graphically shown as in figure~~displayed graphically, as shown in Figure 1. This shows a ~~first-order~~first-order HMM with three states and the transition probabilities drawn in. At each time step the HMM emits an observation from the state it is currently occupying. In the case of a discrete HMM this observation will be a symbol from an alphabet V of size M . The state process B then describes the probability that the HMM, currently in state S_i will emit the k -th symbol from the alphabet. This is given in ~~equation~~Equation 2.

$$b_i(k) = P(v_k \mid q_t = S_i) \quad at\ time\ t \quad 1 \le k \le M \qquad (2)$$

(Figure 1: A ~~three state~~three-state hidden Markov model.)

The last parameter needed to describe the HMM~~,~~ is the initial state distribution:

(3)

An HMM can therefore be fully described by the parameters A, B and π or $\lambda = (A, B, \pi)$.

3. Hidden Markov ~~Models~~models and classification

From the above it can be seen that the HMM is a doubly stochastic process ~~that~~which can be trained by adjusting the model parameters to mimic the statistical characteristics of a certain type of signal. ~~The HMM can then be used to produce~~Then the HMM can be used for producing signals with the same statistical parameters, or alternatively, given a signal and an HMM, one can calculate the probability that the HMM will produce the exact signal. The latter is how HMMs are used for classification.

Consider a system that can occupy a number of operational states. If signals from this system are recorded, one can train an HMM for each operational state of the system. Any future signal from the system can be classified by associating it with an HMM. This is done by calculating the probability ~~for~~that each HMM ~~to~~will produce the unclassified signal. This probability is called the forward probability and is calculated with a recursive algorithm called the forward procedure. The signal can then be classified into the category of the HMM with the highest forward probability. The diagram in Figure 2 shows~~a schematic diagram of~~ how this is done.

(Figure 2: A classification procedure using HMMs.)

The forward probabilities can shrink to very small numbers and will almost certainly cause underflow problems on ~~present computers.~~the computers currently available. To counteract this, logarithmic probabilities are introduced. Whereas normal probabilities range from 0 to 1, logarithmic probabilities range from -∞ to 0. A probability of 0.5

therefore corresponds to a logarithmic probability of -0.301. This ensures that ~~very~~extremely small numbers can be dealt with conveniently.

4. Experimental setup

A cutting experiment was conducted where the strain signals from a tool holder were recorded using a strain gauge rosette with ~~2~~two strain gauges. ~~The two strain~~These two gauges were ~~located~~positioned on the side of the tool holder and connected in a ~~half bridge~~half-bridge configuration. The strain signals from the two gauges in the rosette were thus fused into a single signal. The gauge elements were oriented perpendicularly with one element oriented along the length of the tool holder. Figure 3 ~~shows a schematic~~contains a diagram of the tool holder as well as the approximate location of the strain gauges.

~~(figure 3: A schematic~~(Figure 3: A diagram of the tool holder showing the approximate location of the strain gauges)

The rosette was positioned to capture the feed force and the radial force.~~It has been observed by~~ Dimla [11] observes that the feed and radial forces may be influenced more by tool wear ~~then~~than by the main cutting force itself. Strain gauges are ~~relatively cheap~~fairly inexpensive compared ~~to~~with other measurement techniques but~~their preparation is time consuming and keeping them protected~~it is imperative to protect them from the cutting area ~~is imperative.~~and preparing the strain gauges is time-consuming. The side of the tool holder is therefore a~~very~~ sensible location.

Table 1 defines the equipment ~~that was used. In figure 4 a schematic setup~~used in the experiment. Figure 4 contains a diagram of the whole measurement setup~~ is shown~~.

Table 1: List~~ing~~ of all the equipment

Item	**Make**
Lathe	Tortona SAG14

Tool holder	Mitsubishi S160 SCI PRO9
Tool insert	Mitsubishi US7020 MV
Strain gauge	HBM 1.5/120XY91 (rosette)
Strain gauge amplifier	Clip AE101
A/D conversion	National Instruments PCI-6024E
Signal processing	Matlab Data Acquisition toolbox
Workpiece material	EN19

~~(figure 4: A schematic~~(Figure 4: A diagram of the setup of the system)

The signal from the strain gauge amplifier was fed through an analogue anti-alias filter. A 4th order Chebyshev filter was used for anti-alias protection. The filter had a –3 dB roll-off at 4,350 Hz. The signals were recorded on a personal computer with a National Instruments analogue to a digital conversion card using the Matlab Data Acquisition Toolbox [12].

The experiment was conducted for one set of machining parameters only. Keeping the cutting speed constant during the experiment turned out to be a ~~problem. This is because of the nature of the~~problem because in this particular type of cutting process~~where~~ the diameter of the workpiece is reduced after each pass. ~~Because~~Since the rotational speed can only assume a fixed value, this caused each subsequent cut to have a lower cutting speed. It was decided ~~to keep~~that the cutting speed would be kept constant within a certain ~~tolerable~~tolerance band. This band was chosen to be ±5% around 130 m/min. ~~The~~However, the feed rate and depth of cut could ~~however be kept constant. This was selected to be 0.25 mm/rev and 0.5 mm respectively.~~
be kept constant: the feed rate was selected as 0.25 mm/rev and the depth of cut as 0.5 mm.

The experiment was conducted using ~~EN19 roundbar~~an EN19 round-bar with a diameter of around 77 mm after the “crust” ~~was~~had been removed. The workpieces were received from the vendor in the so-called “T-condition”. To keep within the ~~tolerable~~tolerance band for the cutting parameters, a total of 10 cuts with a length of 170 mm each were

made on each workpiece.

The cutting speed ~~turned out to be~~was rather slow for the combination of material and cutting ~~insert combination. This was proved~~insert. This was proven by the long continuous chip that was formed during cutting. ~~During inspection of~~When the insert was examined under an optical microscope, only nose wear could be detected, even after ~~thirty~~30 minutes of cutting time. Nose wear is ~~characteristic~~typical of low cutting speeds.

During the experiment the tool insert was worn down from 0 mm to 0.1 mm. This is a third of the usual~~allowable~~ 0.3 mm allowable for flank ~~wear. This makes~~wear, making the recognition harder since there will less separation between the features. This may however ~~proof the ability of the system~~prove the system's ability to perform robust classification on features with little separation. Unfortunately, the experiment could not be ~~continued for more that~~conducted for more than one tool, which limited the amount of good data that could be used for training and testing.

5. Signal processing

5.1 Feature extraction

A typical cutting signal is shown in ~~figure~~Figure 5. The signal features a drift~~that is~~ caused by a rise in temperature~~increase~~ as the cutting progresses. The part indicated in the figure was extracted and detrended to leave a signal with no transients and a zero mean. Finally the signal was segmented into ~~segments of 2048~~2,048 samples. The features were calculated ~~From~~from each of these segments~~ the features are calculated~~.

(Figure 5: A typical cutting signal.)

Two sets of features were ~~calculated;~~calculated: statistical time domain features and frequency domain features. Statistical features are ~~very easy and very fast~~quick and easy to calculate, making them ideal for on-line applications. Previous studies by Silva et al. [13] and Scheffer and Heyns [14] have found these features to be indicative ~~to~~of tool

wear. The following features were considered:

Variance, which is the second statistical moment of the data. After detrending, the mean of the data is 0 which makes the variance of the data equal to the square of the root mean squares (RMS) of the data. The RMS is an indicator of the energy content of a signal. As tool wear progresses, more energy is needed to drag the tool insert through the workpiece. It follows~~to reason~~ that the RMS (or variance in this case) should increase too. The variance of the signal $x(t)$ over period T ~~is~~was calculated using~~:~~

$$\sigma^2 = \frac{1}{T}\int x(t)^2 dt \tag{4}$$

In ~~equation~~Equation 4 ? is the standard deviation.

Skewness is the third statistical moment and describes the distribution of the data in terms of symmetry or lack ~~thereof:~~of symmetry:

$$S = \frac{1}{\sigma^3 T}\int x(t)^3 dt \tag{5}$$

Kurtosis is the fourth statistical moment and is ~~very popular in bearing condition monitoring. The kurtosis~~remarkably popular in monitoring the condition of bearings. Kurtosis is a measure of the relative peakedness of the distribution, and the pulses in the time-domain waveform of the signal. ~~The kurtosis~~Kurtosis is also a measure of how close the distribution is to a Gaussian ~~distribution. It thus carries~~distribution and therefore it conveys valuable information for condition monitoring. ~~The kurtosis~~Kurtosis is calculated using~~:~~

$$K = \frac{1}{\sigma^4 T}\int x(t)^3 dt \tag{6}$$

Crest factor is another feature~~that is~~ widely used in bearing condition monitoring and is a measure of the impulsiveness of a vibration signal. A truly random signal has a crest factor generally less than 3. The crest factor is calculated using~~:~~

$$CF = \frac{X_{\max}}{X_{rms}} \tag{7}$$

Entropy is a measure of the uncertainty or disorder of a given signal. One can intuitively see that a signal with a higher energy content, as in the case of a worn tool, will display more disorder. The entropy measure used was Shannon entropy which is often used in

wavelet analysis. Shannon entropy is calculated using ~~equation~~Equation 8:

$$E = -\sum_{i=1}^{N-1} x_{i^2} \log(x_{i^2}) \tag{8}$$

In ~~equation~~Equation 8, x_i is the value of x at time $t=t_i$. N is the number of samples for which the feature is calculated ~~for~~. This is the same as the time interval for the statistical features.

Dynamism is a measure of the rate of change ~~of~~in a quantity. This feature also captures ~~dynamic behaviour~~the dynamic behavior of a signal in a way similar~~way~~ to that for the crest factor. ~~Dynamism was used~~Ajmera et al. [15] used dynamism for speech and music ~~segmentation by Ajmera et al. [15]. Dynamism~~segmentation. In the present study, dynamism is calculated ~~with:~~using:

$$D = \sum_{i=1}^{N} [x_i - x_{i-1}]^2 \tag{9}$$

The second set of features ~~are~~is that of the frequency domain. These ~~will be~~are levels of <<please check if levels is the right word here>> energy within certain frequency bands. This is calculated ~~using:~~as follows:

$$E = \int_{fl}^{fh} S_x df \tag{10}$$

S_x is the ~~one sided~~one-sided power spectral ~~density and,~~density. and f_l and f_h are the lower and higher limits for the frequency band. ~~For the selection of frequency domain features a~~A number of frequency bands were identified for selecting the features of the frequency domain. These are bands where the spectral energy increases (or decreases) as tool life progresses.

Twelve features were extracted from the data. Figure 6 shows how these features evolved as functions of the sample number. This is also a measure of time since each data point is the result of the features for 2,048 samples, which is in turn equal to ~~around~~approximately 0.1 second. The frequency domain features are shown in ~~figure~~Figure 7. In both the figures the most desirable features are ~~the ones~~those that show

a consistent trend throughout the range. This ~~can~~may be a negative or a positive trend, meaning that the feature value might decrease or increase as tool wear progresses. Furthermore, some features may ~~show~~display a tendency to become more sensitive to certain stages of the wearing process.

(Figure 6: The time domain features)
(Figure 7: The frequency domain features)

The most salient of the features were selected using the method suggested by Scheffer and Heyns [14]. Features are selected according to ~~its~~their correlation coefficient with a theoretical tool weartool-wear function. The theoretical tool weartool-wear function was ~~taken to be~~arbitrarily chosen as a straight line with a gradient of 40°.~~This gradient was chosen arbitrarily.~~ The correlation coefficient ~~shows~~indicates how consistent the trends in the features are with ~~that of the tool wear~~those of the tool-wear function. The correlation coefficient was calculated using~~:~~

$$corr(X,Y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2\right]^{\frac{1}{2}}} \qquad (11)$$

If X and Y were the two signals ~~that were to be compared~~to be compared, then X was taken to be the feature and Y the ~~tool wear function~~tool-wear function; $\bar{x}$ and $\bar{y}$ denote the mean values of the variables. Table 2 shows the features as well as their correlation coefficients.

Table 2: The calculated features and their corresponding correlation coefficients

Feature	**Correlation coefficient**
Entropy	-0.650
Crest factor	-0.596
Kurtosis	-0.057

Energy in band: 3110-3505 Hz	0.037
Skewness	0.085
Energy in band 166-175 Hz	0.379
Dynamism	0.404
Energy in band 185-214 Hz	0.445
Energy in band 126-244 Hz	0.516
Energy in band 3600-3800 Hz	0.540
Energy in band 30-85 Hz	0.601
Variance	0.704

Based on these correlation coefficients, all the features except the kurtosis, skewness and the ~~3110–3505~~3,110–3,505 Hz energy band were chosen as features for the classification process. This reduced the feature set to ~~9.~~nine.

5.2 Dimensional reduction

The theory ~~on HMM can easily be~~of HMM can be easily extended to more than one dimension as is the case with the work of Bunks et al. [5], buthaving high dimensionality in a ~~dataset, requires that more data be collected in order to get~~data set requires the collection of more data so that good estimates of the distributions can be obtained. For cases where ~~a high~~large quantities of data ~~is~~are readily available, this would be the preferable route. This might not always be the case, though, and the normal HMM toolbox can be used in these situations. A reduction in dimensionality is ~~In order for~~needed so that this HMM toolbox can handle the information from a number of different ~~features to be handled by this HMM toolbox areduction in dimensionality is~~ features. In this study, all the features were used in the dimensional reduction, ~~needed. In this study all the features except:~~except for Energy in band: 3110-3505 Hz, Skewness, and Energy in band 166-175 Hz. The correlation coefficients of these features were much smaller than those of the others. The data was subjected to a principal component decomposition. The data was then projected onto the first principal component to reduce the dimensionality from 9 dimensions to 1 dimension.

Principal component analysis (PCA) is a standard statistical function that utilizes singular

value decomposition to calculate the principal components of a data matrix. Table 3 shows the variance represented by each principal component. This table indicates that the first principal component comprises 73.5% percent of the total variance in the features. The principal components were not calculated from all the data available. One-third of the components was randomly selected from the available data and on this data sample, the first principal component was calculated.

Table 3: The principal components and percentage of the total variance they represent.

Principal component no.	Percentage variance
1	73.57
2	10.53
3	7.17
4	3.81
5	2.45
6	1.07
7	0.84
8	0.50
9	0.02

The data archive was segmented into three parts for training the HMM. The first segment was used for training and testing the HMM for sharp/new tools. Likewise the last part was used for training and testing the HMM for worn tools. Random samples were drawn from these two segments to make up a feature space that could be used for determining the principal components. Figure 8 shows the dimensionally reduced feature. This feature has a correlation coefficient of -0.695 with the theoretical tool wear function and is therefore comparable to the best of the calculated features. An important issue to address here is the fact that not all the features are sensitive to tool condition during the entire life of the tool. This means that certain features are sensitive to changes in tool condition in the early stages of wear, whereas others are more sensitive to tool condition during the later stages of tool life. This is why the use of only a single feature might not perform as well as a combined set of features.

The type of noise shown in Figure 8 illustrates the need for complex classification schemes which are not highly sensitive to workpiece-related issues, such as the clamping conditions of the workpiece and/or the tool insert, metallurgical composition, etc. This technique also demonstrates how features from different sensor signals can be fused to form a robust feature that can be used for classification.

(Figure 8: The dimensionally reduced feature)

5.3 Discretization

A discrete HMM was used for the classification procedure. A decision was taken to discretize the signal into 150 levels, as there is still a great deal of detail left in the signal at such a level. Discretization may serve as a form of noise filtering. With fewer levels of discretization the data might be less noisy. The simplest HMM model, namely the discrete HMM, was then used for the classification. This technique also serves as a method of data compression.

5.4 Observation sequences

Hidden Markov models do not work with single samples at a time, as is the case with neural networks. Instead an observation sequence is treated by the HMM classification. The observation sequence in this case consisted of 16 consecutive samples from the dimensionally reduced feature vector. This constituted about 2 seconds of cutting time during the experiment. These observation sequences convey more information about the characteristics of the cutting signal. Figure 9 shows a few observation sequences for the different classes, and there is clearly a good separation. There is also separation within the classes, which can be seen as outlier sequences in the figure. Owing to the nature of the data, this may be a result of clamping conditions on the workpiece as well as the specific composition of the workpiece itself. Fortunately, these factors work constructively with the deterioration of the tool insert.

(Figure 9: Some observation sequences.)

6. Classification results

6.1 Classification strategy

A multi-model approach as described by Ertunc et al. (6) was used for the classification procedure. This means that an HMM was trained for each system condition that would be identified. Two-thirds of the data for each class was used for training the models and the remaining one-third was used for testing. The data sets for training and testing were randomly selected from the data set for each class.

6.2 Training

The two HMMs were trained for 10 iterations after the random initialization of the HMM parameters. The training sets for each HMM consisted of 444 observation sequences. After training the sharp-tool HMM, the batch logarithmic probability for the training set was around –18,000. The batch logarithmic probability is the sum of all the forward probabilities for an HMM for a data set. This translates into an average of -40 for any one set. This number represents the expected range for the forward probability of any signal that should fall into this category. For the worn-tool HMM the batch probability was –22,000. The average forward probability for a signal from this class was -50.

6.3 Classification

The HMM was then tested with a data set that contained the data from the classes not selected for training. The forward probabilities for each HMM were calculated for each observation sequence. The observation sequence was classified into the sharp category if the sharp-tool HMM had a higher forward probability than the worn-tool HMM, and vice versa.

6.4 State topology

As with neural networks, the optimal state topology (or number of neurons in the hidden layer for neural networks) of an HMM cannot be determined analytically, so it was determined exhaustively. The whole training and classification procedure was repeated a number of times, with the number of states in each model increasing at each step. This procedure is shown in Figure 10. Each point in this figure is the average result of 5

iterations. The state topology was chosen as a trade-off between complexity and performance. After a few simulations, the optimal topologies were chosen as

2 states for the HMM to be trained on worn tool data;

7 states for the HMM to be trained on sharp tool data.

This figure now represents the kind of classification performances that can be achieved by an HMM as a function of the number of states contained in the HMM. The classification performance can be further explored by investigating the performance with these parameters in place.

(Figure 10: The number of states versus the classification performance.)

6.5 Performance

The data set used for the performance evaluation consisted of two parts: the first half of the data set for testing contained data belonging to the "sharp" category and the second half the data from the "worn" category. The classification results are shown in the top part of Figure 11. One can see that somewhere in the middle, the forward probabilities of the two models cross each other.

(Figure 11: (top) Classification probabilities. (bottom) Classes awarded)

If this graph is conditioned so that 1 is awarded if the forward probability of the sharp-tool HMM is higher than that of the worn-tool HMM and if it is lower, a zero is awarded, then the bottom half of Figure 11 is obtained. This gives a clearer picture of the classification performance.

If more observation sequences are randomly selected from the two data classes and the process of training and testing is repeated with the selected state topologies, the average result of the performance can be calculated. In addition one can also see the behavior of the classification scheme, as shown in Figure 12. The performance is rather erratic and this may be attributed to either the quality of the data set or the limited amount of data.

(Figure 12: The performance behavior of the HMM classification.)

The average classification results after 20 iterations are as follows:
91.5 % correct classifications of sharp tools
94.5 % correct classifications of worn tools.
It is suspected that with more data the erratic behavior shown in Figure 13 would smooth out and classification might improve.

6.6 Reduced data set

In order to test the hypotheses that the classification will be improved with more training data, the converse of this statement was proved. The classification procedure described above was repeated with a reduced data set that contained only two-thirds of the original data. In this data set, the tool wear progressed only two-thirds of the way, compared with the original data set.

Once again the optimal state topology was determined, as shown in Figure 13. This figure indicates there was far less of a trend between the number of states in the HMM and the performance of the HMM.

(Figure 13: The number of states versus the classification performance.)

The optimal number of states was as follows:

2 states for the worn-tool HMM
8 states for the sharp-tool HMM.

Even with the optimal number of states, the HMM classification fared more poorly than the HMM trained with the full set of data. The behavior of the model was more erratic and spread over a wider range, indicating that the amount of data is a factor in the classification performance of the HMM. After a number of simulations, the average performance of the HMMs trained on the reduced data set was

65% correct classifications for worn tools;
92% correct classifications for sharp tools.

5.1 Bayesian classification

The classification performance with the HMM was also checked against the performance of another classification scheme. The simple-to-use and well-known Bayesian classifier was chosen for this comparison. This classifier is based on a maximum likelihood method.

In this method, a probability density function (PDF) is created for each class. This PDF is also created with a training data set assembled in the same manner as for the previous method. Classification is then done with a testing data set (which is also partitioned in the same manner as before). If $P_a(x)$ is the PDF of one class and $P_b(x)$ the PDF of the second class, then classification can be done in a manner similar to that used for HMMs.

If $P_a(x)$ and $P_b(x)$ are calculated for any x of the testing data set, then x can be classified into the class of the set with the highest probability. Figure 14 shows histograms of the two classes and their corresponding Gaussian approximations that were used for this classification. The Gaussian approximations to the histograms can be used as a PDF that describes a class. This is the area of maximum likelihood for each class in the discretized space. In Figure 14 the tops of the bins of the histograms were connected to form a curve and the whole curve was normalized so that its area would be equal to unity. The decision boundary for the maximum likelihood is also shown in Figure 14. When the training, testing and classifying of randomly chosen data sets is repeated a few times in the same manner as before, Figure 15 is obtained. This figure shows the performance behavior of the Bayesian classifier. This classifier has a more stable performance. The average classification performance for the Bayesian classifier after 20 iteration was
73,7% correct classifications of worn tools;
96.8% correct classifications of sharp tools.

In the case of a sharp tool this is better than the classification performance of the HMMs trained on the full data set, but in the case of worn tools, this is much worse than the performance of the HMM trained on the full data set. The classification performance of

the HMM that had been trained on the reduced data set was comparable to that of the Bayesian classifier.

(Figure 14: The histograms describing the different classes.)
(Figure 15: The performance behavior of the Bayesian classifier.)

The poor classification results for the HMM using the reduced data set and the Bayesian classifier are due to the fact that a clear separation of the classes was not achieved. It is obvious that there is a large degree of overlap between the classes shown in the histograms in Figure 14. The performance of the Bayesian classifier also depends on the approximations of the class distributions. In this case, the assumption made was that the classes could be described using a Gaussian distribution. Figure 14 also indicates that a better type of distribution would increase the performance of the Bayesian classifier.

The power of the observation sequences that the HMMs use for classification is also shown in this comparison. Because more than one sample is used simultaneously in the observation sequence, information about the relation of one sample to its consecutive partners is also captured and used in the classification.

7. Conclusion

In this study, hidden Markov models were applied to the problem of tool wear classification in turning operations. A strain signal was recorded from strain gauges that measured the feed force during cutting. Signal features correlating to tool wear were extracted and combined by applying a principal component analysis and using the first principal component. This method was therefore used as a technique for dimensional reduction. Two classes were identified from the machining data.

A discrete, ergodic HMM with a codebook of size 150 was trained for each class. Training and testing was done with random observation sequences from the data sets for each class. The optimal state topology was determined by means of an exhaustive method. The average result of this method was compared to a Bayesian classifier. The

Bayesian classifier performed better with the classification of sharp tools but performed disappointingly badly with the classification of worn tools. The reason for this is the overlapping of the histograms (and consequently the theoretical probability density functions) of the two classes. The HMMs do not have this problem and their successful classification is due to the fact that the HMMs use observation sequences instead of single values. This fact illustrates the strength of an HMM's ability to capture the temporal as well as spatial characteristics of a signal.

This study has also shown how features from different sensors may be fused to form a robust recognition system using HMMs. The feature extraction and selection process from neural network research may be applied to HMMs using the suggested method in this article. [9]

References

[1] G. Byrne, D. Dornfeld, I. Inasaki, G. Ketteler, W. Konig & R. Teti, Tool condition monitoring - the status of research and industrial application. *Annals of the CIRP* 44, 2001, 541-567.

[2] L.R. Rabiner A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77(2), 1989, 257-286.

[3] K.C. Kwon & J.H. Kim, Accident identification in nuclear power plants using hidden Markov models. *Engineering Applications of Artificial Intelligence* 12, 1999, 491-501.

[4] L. Atlas, M. Ostendorf & G.D. Bernard, Hidden Markov models for monitoring machining tool wear. *IEEE* 2000, 3887-3890.
(download at:
http://cisi.berkely.edu/~dpwe/research/etc/icassp2000/pdf/4250_08.pdf)

[5] C. Bunks, D. McCarthy & T. Al-Ani, Condition-based maintenance of machines using hidden Markov models. *Mechanical Systems and Signal Processing*, 14 (4), 2000, 597-612.

[6] H.M. Ertunc, K.A. Loparo & H. Ocak Tool wear condition monitoring in drilling operations using hidden Markov models (HMMs). *International Journal of Machine Tools and Manufacture*, 41 (9) , 2001, 1363-1384.

[7] L. Wang, M.G. Mehrabi & E. Kannatey-Asibu, Hidden Markov model based tool wear monitoring and turning. *Journal of Manufacturing Science and Engineering* 124, August 2002, 651-658.

[8] J.M. Lee, S.J. Kim, Y. Hwang & C.S. Song, Pattern recognition of mechanical fault signal using hidden Markov model. <<check if model or models in source cited>> *International Congress on Sound and Vibration*, Stockholm Sweden, July 2003, 4725-4730.

[9] M. Ge, R. Du & Y. Xu, Hidden Markov model-based fault diagnosis for stamping operations. *Mechanical Systems and Signal Processing*, 19(2), 2003, 391-408.

[10] K. Murphy, Hidden Markov Model Toolbox, 1998, http://www.ai.mit.edu/murphyk/Software/hmm.html

[11] D.E. Dimla, Sensor signals for tool-wear monitoring in metal cutting operations – a review of methods. *International Journal of Machine Tools and Manufacture* 40, 2000, 1073-1098.

[12] The Mathworks, Data Acquisition Toolbox for use with MATLAB, 2001.

[13] R.G. Silva, R.L. Reuben, K.J. Baker & S.J. Wilcox, Tool wear monitoring of turning operations by neural network and expert system classification of a feature set generated from multiple sensors. *Mechanical Systems and Signal Processing*, 12(2), 1998, 319-332.

[14] C. Scheffer & P.S. Heyns. Wear monitoring in turning operations using vibration and strain measurements. *Mechanical Systems and Signal Processing*, 15(6), 2001, 1185-1202.

[15] J. Ajmera, I McCowan & H. Bourlard. Speech/music segmentation using entropy and dynamism features in an HMM classification framework. *Speech Communication*, 40(3), 2002, 351-363.