Sensor Based Condition Monitoring Feature Selection Using a Self-Organizing Map

Rui G. Silva and Steven J. Wilcox

Abstract— **This paper presents a new approach to sensor based condition monitoring feature selection using a self-organizing map. Self-Organizing Maps perform classification in a non-supervised fashion performing vector quantization and therefore place similar vectors close together in the two dimensional output space. The unsupervised process leads to the self organization of modeling with no previous knowledge of what is being modeled and therefore it does not model a predetermined environment. Taking the above into account feature selection was performed by analyzing the contributions of different sensor based features towards tool wear classification. It was found that some of the features, not previously evaluated and justified, have a strong contribution towards tool wear classification.**

Index Terms— **Self-Organizing Map, Condition Monitoring, Tool Wear, Feature Selection.**

I. Introduction

The late 1990s and early 2000s have witnessed a change from the old practice of changing tools automatically, to the feasibility of instituting tool change procedures based on monitoring the amount of wear on the cutting tool-edges through the implementation of adaptive tool inspection mechanisms. Thus, an appropriate and timely decision for tool change is significantly required in the machining system. The traditional ability of the operator to determine the condition of the tool based on his/her experience and senses, i.e. vision and hearing, is now the expected role of the monitoring system. One important strategy to support this goal is sensor-based, real-time control of key characteristics of both machines and products, throughout the manufacturing process.

Several factors have impeded advances in the development of TCMSs including inappropriate choice of sensor signals and their utilization. The random behavior can be attributed to the large-scale variation and non-homogeneities that exist in the workpiece. Typically, most metal cutting processes can be classified as having one or more of the following characteristics, Warneche et al. [1]: Complex to chaotic behavior due to non-homogeneities in workpiece material, sensitivity of the process parameters to cutting conditions, and a non-linear relationship of the process parameters to tool wear.

Wear monitoring has been performed using many different sensing techniques. These techniques include; temperature, motor current, acoustic emission (AE), audible emissions, vibration and force, [2]. Some of these have been successfully applied under laboratory conditions although industrial applications have been rather unsuccessful. Clearly, the quality of the sensor information is adequate to make judgments of the state of wear in idealized conditions but much work has to be performed in information processing and decision making in order to correctly classify the tool wear state from the available sensors. It is therefore the aim of this work to integrate some of the above mentioned sensors to extract the largest possible amount of information from the cutting process and provide an indication of the wear level.

Previous work on the relationship between audible emissions and tool wear has established that audible emissions are capable of indicating the extent of the cutting edge wear, Weller et al.[3]. McNulty et al.[4] have also highlighted the use of noise spectra for tool life evaluation applied to several cutting processes and have found significant changes in certain frequency bands that appear to be characteristic of wear in certain cutting processes. Lee [5] found that, during turning the machine noise exhibited a wear related change of sound pressure level (SPL) at certain frequencies (4 - 6 kHz) for several materials. A drop in the SPL before the tertiary zone (third and last stage of wear) was suggested as an end of tool-life predictor. Experiments carried out by Ya et al. [6] using two different types of turning tool showed that both the tool angle and the cutting speed exerted no great influence on the average cutting noise.

Vibration has also been used to recognize the wear state of a tool whilst turning [7] and the main advantage of this method is its ease of application. Taking into consideration previous research (e.g. Jiang et al. [8]) vibration has been chosen in this work as a secondary source of information because of the correlation between machine tool vibration and tool wear that have been demonstrated successfully in the

Manuscript received March 9, 2008. This work was supported in part by FCT – Fundação para a Ciência e a Tecnologia, Portugal. Project Reference POCTI/EME/59491/2004.

R. G. Silva is with Universidade Lusíada de Vila Nova de Famalicão, Largo Tinoco de Sousa, 4760-108 VNF, Portugal (phone: 351 252 309 200; fax: 351 252 376 363; e-mail: rsilva@fam.ulusiada.pt).

S. J.. Wilcox is with the University of Glamorgan, Pontypridd, Wales CF37 1DL, UK (e-mail: sjwilcox@glam.ac.uk).

laboratory. The vibrations arising from the shearing action of the tool is transmitted to the base of the machine where they are transduced by the accelerometer.

A mechanistic model derived from first principles is theoretically the most accurate model that can be developed for any system. Unfortunately, the resources required to develop such a model for even the simplest of systems tends to prohibit their use. Therefore, forecasting in complex systems that are poorly understood, noisy and often non-linear can be practically impossible when based on the traditional model predictive algorithms, Parlos et al. [9]. Consequently engineers tend to rely on system identification techniques to establish process models. As with linear models, Artificial Neural Networks (ANN) provide a description of the relationship between cause and effect variables. The benefit of ANNs over linear models is that they are capable of modeling non-linear relationships. In fact studies have shown them to be capable of modeling any non-linear function to arbitrary accuracy Cybenko [10] and Hornik et al. [11]. Also, artificial neural networks have found increasing favor in manufacturing systems research because of their ability to perform robustly in noisy environments, Balazinski et al. [12]. Abstraction of hardly accessible knowledge and generalization from distorted sensor signals are some of the most attractive features of neural networks when applied to sensor fusion and classification in tool wear monitoring. Nevertheless, although working in certain conditions, most of the previous applications of neural networks have some limitations, as reported by Lennox et al. [13] in a study of the application of artificial neural networks in the area of process monitoring and control.

This article is subdivided into four main sections: an introduction to condition monitoring and its current state of the art; an introduction to spiking neuron networks and their feasibility for condition monitoring; and, finally, experimental work and simulation results.

II. Preliminary Experimental Work

Based on the above considerations experimental background work was conducted on the turning process to collect tool wear data. In this work a set of tool wear cutting data was acquired by machining a block of mild steel under realistic production conditions that consisted of a cutting speed of 350 m/min, a feed rate of 0.25 rev/min and a depth of cut of 1 mm, with a coated cemented carbide tip. The set of sensors used were: an accelerometer for measuring vertical vibration, a microphone for recording sound emission, a strain gauged tool holder for force measurement and a meter for the spindle current of the CNC machine. The turning operation was carried out on an MT 50 CNC Slant Bed Turning Centre. The analogue signals were sampled at 20 kHz with tool wear and sensor data being acquired at intervals of 2 min, taking into account an expected tool life, for each insert, with a typical value of 15 min. Sample data were recorded for 6 inserts. The length of each sample was 512 points, and these were acquired approximately in the middle of the bar.

Each 512 point record was processed to generate the features used in the classification stage. A total of 12 features were extracted from the sound and vibration data: absolute deviation, average, kurtosis, skewness and the energy in the frequency bands (2.2-2.4 and 4.4-4.6 kHz) obtained from the spectra. Two additional features were presented from the means of the feed and tangential forces. Results have shown that tool wear classification is difficult in the presence of such noisy data and it is therefore required that classification is made by a method that can resolve the complex interrelation between features to produce a robust wear classification. Also the use of multiple sensors should prove to be of great value towards tool wear evaluation since the noisy character of each sensor alone would lead to certain failure of the monitoring system [14].

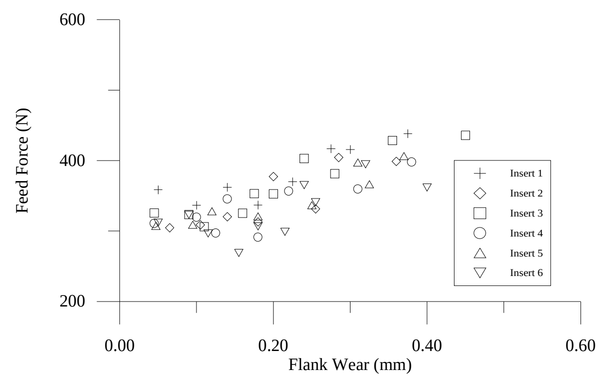


Figure 1 - Feed Force versus tool wear

As can be observed, e.g. Figure 1, both tangential and feed forces show an increase with tool wear which is consistent between tools.

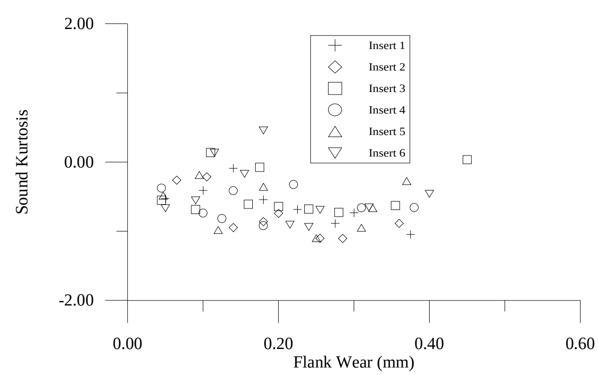


Figure 2 - Kurtosis of Sound versus tool wear

From Figure 3 it can be seen that the power spectrum of data obtained from the microphone data varied consistently with the wear level. The intervals which show this relation for the microphone, are: [3.5;5.5] kHz, [6.2;7.5] kHz and for the accelerometer, are: [3.6;5.2] kHz, [6.2;7.2] kHz. The results obtained from the statistical and frequency parameters, as well as forces and spindle current, are somewhat difficult to interpret considering them one at a time as some appear to correlate, whilst others appear to hold no correlation with tool wear. This can be overcome by taking into account the neural networks' ability to extract information from apparently scattered information.

The remaining features (absolute deviation, mean, kurtosis and skewness of both sound and vibration) exhibited little correlation with flank wear, e.g. Figure 2, data points

ISBN:978-988-17012-3-7

appearing to be randomly distributed through the entire space. Although the statistical parameters did not present any obvious relation to tool wear evolution, it is not possible at this stage to judge their importance for tool wear monitoring due to the complexity of the process. The remainder of this paper, however, shows that some of these data can still be used in monitoring the cutting process. Despite this it was decided to use them is training the NN's as there may have been features corresponding to tool wear that a simple regression analysis would not show.

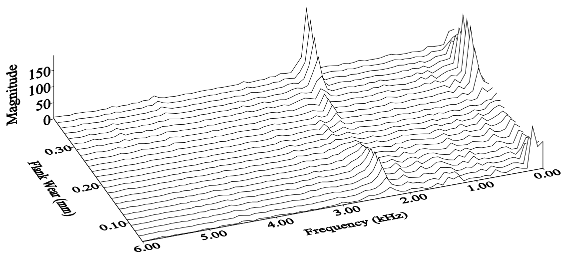


Figure 3 Waterfall plot of frequency spectrum of sound emissions (Insert 1)

In the present study 14 inputs (average, absolute deviation, skewness and kurtosis for sound and vibration; tangential and feed force; 2 spectrum bands from both the sound and vibration) have been used although speed would not be compromised by a higher number of inputs. The number of frequency bands was selected in order not to compromise data reliability due to misuse of certain frequencies subject to changes which are not due to tool wear (e.g. machine environment). Therefore, two bands were selected from the power spectrum for the audible emissions and two for the machine vibration.

III. The Self-Organizing Map (SOM)

Unlike other neural network approaches, the SOM network performs unsupervised training; that is, during the learning process the processing units in the network adjust their weights primarily based on the lateral feedback connections. The more common approach to neural networks required supervised training of the network (i.e., the network is fed with a set of training cases and the generated output is compared with the known correct output). Deviations from the correct output result in adjustment of the processing units' weights. On the other hand, unsupervised learning does not require the knowledge of target values. The nodes in the network converge to form clusters to represent groups of entities with similar properties. The number and composition of clusters can be visually determined based on the output distribution generated by the training process.

Cluster analysis is a technique for grouping subjects into clusters of similar elements. In cluster analysis, we try to identify similar elements by their attributes. We form groups, or clusters, that are homogeneous and different from other groups. SOM networks combine competitive learning with dimensionality reduction by smoothing the clusters with respect to an a priori grid and provide a powerful tool for data visualization.

The success of applying neural networks to a problem depends upon the type of problem domain and the representativeness of the data sets that are used to train the neural network. The SOM was coded based on the theory developed by Kohonen [15]. The present network consists of two layers of neurons. The first is the sensory or input layer, consisting in this case of 15 neurons, one for each feature obtained from the sensors plus one that provides the bias. The computation is carried out in the second layer, called the map, that also acts as the output layer and this was 10×10 neurons. The learning procedure consists of two stages. In the first, the map unfolds until a global ordering of the neurons is reached. Every neuron tunes to a pattern or class of patterns, and neighbor neurons tune to similar inputs. In the second stage, the statistical distribution of the synaptic weights approaches that of the input variables. A set of experiments was carried out using the SOM. The results achieved demonstrate the ability of this Neural Network to classify sets of data into.

The SOM network typically has two layers of nodes, the input layer and the Kohonen layer. The input layer is fully connected to a two-dimensional Kohonen layer. During the training process, input data are fed to the network through the processing elements (nodes) in the input layer. An input pattern is denoted by a vector of order equal to the number of features. As the training process proceeds, the nodes adjust their weight values according to the topological relations in the input data. The node with the minimum distance is the winner and adjusts its weights to be closer to the value of the input pattern.

Taking into account the previous considerations and combining it with learning paradigms of unsupervised artificial neural networks, all weights are updated according to,

$$\Delta w_{ij} = \eta . \lambda_{ij} \left(f_i - w_{ij} \right) \qquad (1)$$

Where λ_{ij} is the neighborhood function, W_{ij} the weight to output neuron ij and η the learning rate.

In order to extend this formulation of competitive learning to a realization of self-organization it is required the formulation of a neighborhood function that describes the inhibitory behavior of nearby neurons on the output map. This function allows neurons which are topologically close together initially to have strong excitatory lateral connections whereas remote neurons have strong inhibitory connections. This means that the winner neuron drives the neurons in the neighborhood thus increasing the values they encode. The response of remote neurons is inhibited by the lateral connections. The following neighborhood function was used for the above purpose:

$$\lambda_{ij} = e^{-\frac{\left| v_{winner} - v_{ij} \right|}{d_{\max}}} \qquad (2)$$

where $|v_{winner}-v_{ij}|$ describes the Euclidean distance between the winner neuron, v_{winner}, and some other neuron, v_{ij}, in the neighborhood, and d_{max} the maximum distance between neurons in the topological output map.

The implementation consists of three major components; input vector normalization, training, and test data

ISBN:978-988-17012-3-7

interpretation. Upon training, the weights start to stabilize until there is no significant change in their value. Interpretation of the output results is achieved by visual analyzes of the output map classification results.

In real-time, the only available information concerning a configuration's success will reside in its training performance. The ideal policy will recommend employing a neural network exhibiting "good" sample set classification. The testing to be performed will assess the validity of such a policy for competitive learning, i.e. it will observe its generalization ability. In addition, testing will identify the configurations which typically yield good results, and mark them as good candidates for the application. Two policies exist for training pertaining to weight update. In this work the policy dictating that weights freeze after "sufficient" training is followed because this provides better control over test classification.

Self organization of topologically close neurons is realized taking into account that initial neurons that are topologically close together have strong excitatory lateral connections whereas remote neurons have strong inhibitory connections.

IV. Simulation and Result Analysis

Simulation was performed with an artificial neuron network algorithm, similar to the above description, using 15 input neurons (one for each feature extracted from experimental data plus a bias neuron) and a 10 by 10 grid of neurons. Training was performed on experimental data from 6 cutting inserts representing several wear stages in a total of 33 feature vector sets. The learning procedure consists of two stages. In the first, the map unfolds until a global ordering of the neurons is reached. Every neuron tunes to a pattern or class of patterns, and neighbor neurons tune to similar inputs. In the second stage, the statistical distribution of the synaptic weights approaches that of the input variables.

When the number of clusters desired is different from the number of nodes on the SOM output map, additional steps are required to analyze and group the points on the output map into the desired number of clusters. Currently, this process is done manually and is usually assessed by visual inspection. Sometimes it is hard to visually group the output from SOM especially when the map is highly populated. Hence, a more scienti6c approach that can help the user to group the output from SOM network based on certain objective criterion is needed. On the other hand, assessing the performance of the map is not always a straightforward task and usually takes into account one's ability to visually evaluate it's performance. To overcome this limitation, we have employed a variogram analysis to introduce a more scientific approach to SOM map distribution evaluation.

The variogram is a quantitative descriptive statistic that can be graphically represented in a manner which characterizes the spatial continuity (i.e. roughness) of a data set. It is not surprising that the common descriptive statistics and the histograms fail to identify, let alone quantify, the textural difference between two data sets. Common descriptive statistics and histograms do not incorporate the spatial locations of data into their defining computations.

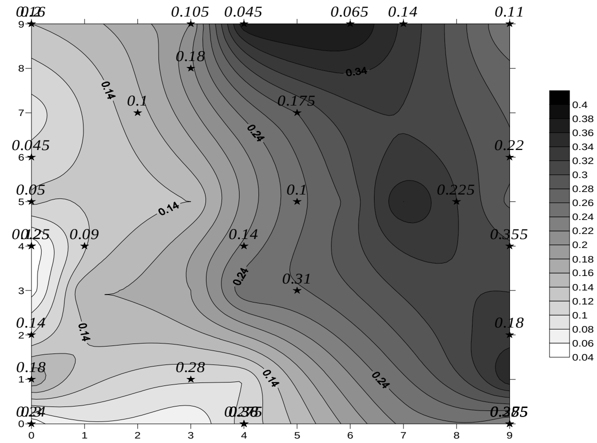
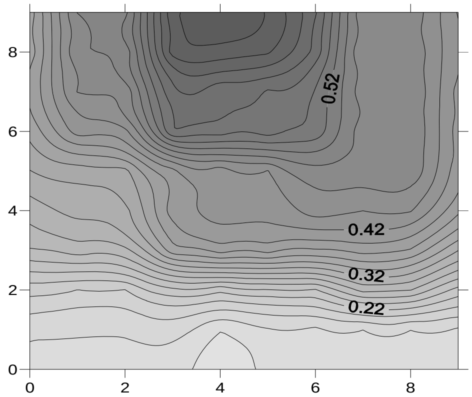


Figure 4 – Contour map of tool wear state classification after 400 epochs trainings with full feature vector

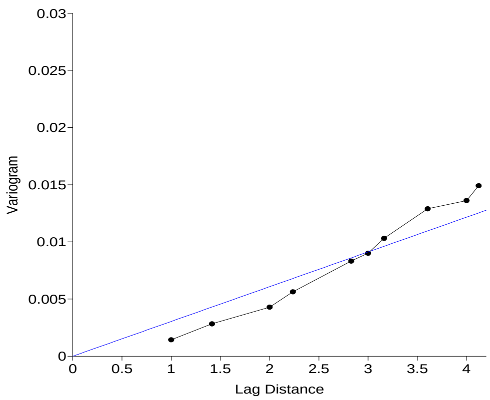
Figure 4 presents the classification results performed using the self-organizing map upon training with the 33 feature vectors and results presented in a 10 x 10 output grid. From the contour maps it can be seen that several distinct areas were created, the shaded areas correspond to the area allocated for the stage of tool wear. This network shows good performance although the interpretation of results is rather difficult, that is it does not provide a direct measure of wear. Visual inspection of the contour map shows the darker areas of the map represent clustering of worn data sets and it is shown that classification of the different tool wear states represented by a star(*) are placed in contiguous areas and therefore it is shown that clustering occurs. A further analysis of the map leads us to conclude that the roughness of the map, presented by the sharpness of the contour lines, might be due to randomness introduced by information predominantly related to the cutting noise. Since the self-organizing algorithm performs classification without any previous knowledge of what is being classified it is probable that interference might be caused by weak features that are mainly machine related.

This study is targeted to the evaluation of feature strength and will be conducted based on the analysis of the weight connection resultant from the use of the full feature vector presented earlier. Given that the winner, or most representative, neuron is achieved through the sum of weighted contributions of each of the features we intend to show, by analyzing each set of weights associated with each feature, that feature selection can be performed in a systematic fashion.

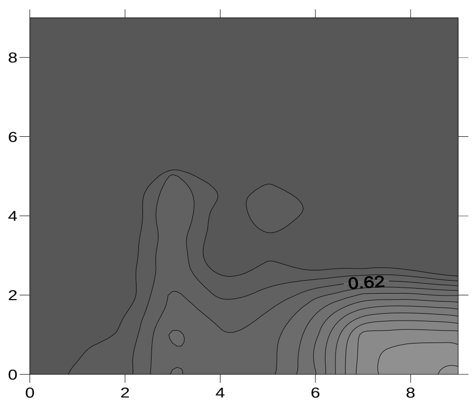
ISBN:978-988-17012-3-7



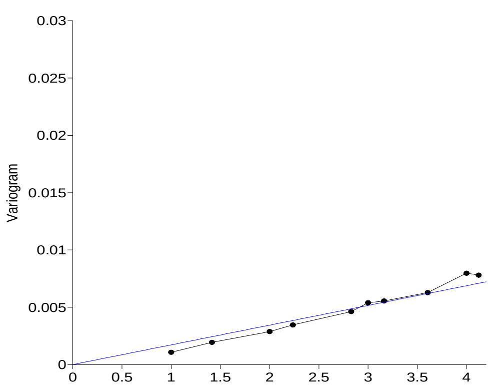
Feed Force – Feature Map (a)



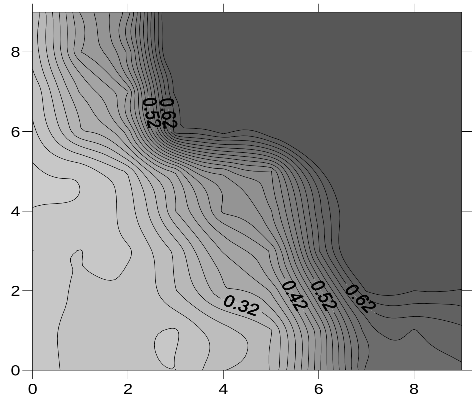
Feed Force – Feature Map Variogram (b)



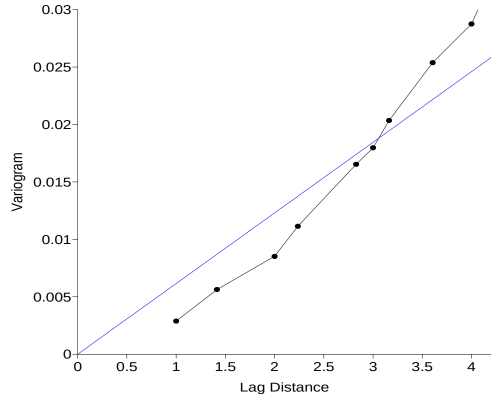
Sound Kurtosis – Feature Map (c)



Sound Kurtosis – Feature Map Variogram (d)



Vibration Frequency Band 1 – Feature Map (e)



Vibration Frequency Band 1 – Feature Map Variogram (f)

Figure 5 – Contour maps of feature strength for different features and the corresponding feature map variogram

Figure 5 presents the mapping of individual weights contribution considering isolated features. Since the self-organizing map performs a sort of vector quantization the weight value indicates the amount of contribution individual features have in the final value and therefore being constant or small little influence have on the final results. Having this in consideration we have used a variogram analysis to, in a more scientific approach, evaluate each feature weight related contribution towards the final classification. Figure 5 shows the plot for some of feature related weights and the corresponding variogram analysis and corresponding linear fit. The slope of this linear fit gives us an indication of variation of weights throughout the map and therefore the indication of fitness or adequacy. It is shown that the variation in weights for kurtosis, Figure 5(d), has a smaller slope than the other two presented in Figure 5 and is related to the fact that little change occurs in the weights, which in result represents that similar contributions of this feature are give to the full map. The data sets are significantly different in ways that are not captured by the common descriptive statistics and histograms.

Table 1 presents the resulting linear fit slope for all the features that resulted from the variogram analysis. It can be seen that some of the features present a higher slope value which correspond to the stronger features. As expected the stronger features are related to the frequency change at the given frequency bands and also at naked eye show a strong correlation with the tool ear evolution.

Table 1 - Linear fit slope for all features

Feature	**Linear Fit's Slope**
Feed Force	0.00304
Tangential Force	0.00308
Sound Average	*0.00281**
Sound Standard Deviation	*0.00172**
Skewness Sound	0.00353
Sound Kurtosis	0.00348
Sound Frequency Band 1	0.00304
Sound Frequency Band 2	0.00322
Vibration Average	*0.00264**
Vibration Standard Deviation	*0.00179**
Vibration Skewness	0.00395
Vibration Kurtosis	0.00407
Vibration Frequency Band 1	0.00615
Vibration Frequency Band 2	0.00651

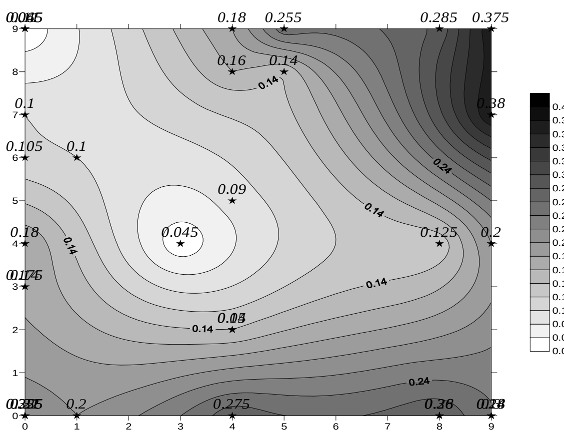


Figure 6 – Contour map of tool wear state classification after 400 epochs trainings with reduced feature vector

In Figure 6 we present the results of the self-organizing map training using a reduced feature vector. The selection of features was performed taking into account the previous analysis and consisted in eliminating the features with the smaller variogram slope value, identified in Table 1 with a star (*). This map shows a smother displacement of classification feature vectors acknowledging the fact that some of the features are weak contributions to tool wear classification. The output map shows that clustering is performed in a smother fashion and self-organization takes place.

V. Conclusion

This paper described the implementation of a prototype decision support system for tool wear monitoring feature selection based on the self-organizing map. It was shown that the modeling technique proposed is highly effective for the classification of wear levels of tool inserts using apparently weak features.

The results obtained from the statistical and frequency parameters, as well as forces, are somewhat difficult to interpret considering them one at a time as some appear to correlate, whilst others appear to hold no correlation with tool wear. This can be overcome by taking into account the neural networks' ability to extract information from apparently scattered information. The use of a Self Organizing Map (SOM) structure has shown that classification was performed quite efficiently although the interpretation of results was not that easy, due to the complexity of the output structure.

The results show that the self-organizing map neural network is a powerful tool for feature selection and validation as it performs vector quantization and hence feature contribution towards final classification can be analyzed in a straightforward manner. Tests presented show a case study where this has been applied with success.

This work has illustrated the potential of Neural Networks when applied to tool wear monitoring. Further, it has enhanced the potential of neural networks, and in particular the self-organizing map, to perform tasks other than classification providing a insight view of feature value and potential towards data modeling.

References

[1] A. Warnecke, A. Jenewein & A. Reinfelder, Tool monitoring based on process identification, monitoring and control for manufacturing processes, Winter Annual Meeting of the ASME, Dallas, TX, USA, 1990, 43–55.

[2] E. Dimla Snr, Sensor signals for tool-wear monitoring in metal cutting operations—a review of methods. International Journal of Machine Tools & Manufacture, 40, 2000, 1073–1098.

[3] E. J. Weller, H.M. Scrier, and B. Weichbrodt, What Sound Can Be Expected from a Worn Tool?. ASME Paper 68-wa/prod, 1969, 1-10.

[4] G.J. McNulty, and N. Popplewell, Health Monitoring of Cutting Tools Through Noise Spectra. 1st Joint Polytechnic Symposium on Manufacturing Engineering June 14&15, 1977, E41-6.

[5] L.C. Lee, A Study of Noise Emission for Tool Failure Prediction. International Journal Tool Design Research 26, 1986, 205-215.

[6] W. Ya, K. Shiqiu, Y. Shuzi, Z. Qilin, X. Shanxiang and W. Yaozu, An Experimental Study of Cutting Noise Dynamics. Machinery Dynamics and Element Vibrations 36, 1991, 313-318.

[7] D. Dimla, and P. Lister, On-line metal cutting tool condition monitoring. I: force and vibration analyses. International Journal of Machine Tools & Manufacture 40, 2000, 739–768.

[8] C.Y Jiang, Y.Z. Zhang and H.J. Xu, In-Process Monitoring of Tool Wear Stage by the Frequency Band-Energy Method. Annals CIRP 36(1), 1987, 45-48.

[9] A. Parlos, O. Rais and A. Atiya, Multi-Step-Ahead Prediction Using Recurrent Neural Networks, Neural Networks, 13, 2000, 765-786.

[10] G. Cybenko, Approximation by superposition of a sigmoidal function, Mathematics of Control, Signals and Systems. 2, 1989, 303-314.

[11] K. Hornik, M. Stinchcombe, and H. White, Multilayer feedforward networks are universal approximators, Neural Networks, 2, 1989, 359-366.

[12] M. Balazinski, E. Czogala, K. Jemielniak and J. Leski, Tool condition monitoring using artificial intelligence methods, Engineering Applications of Artificial Intelligence, 15, 2002, 73-80.

[13] B. Lennox, G. Montague, A. Frith, C. Gent and V. Bevan, Industrial application of neural networks – an investigation, Journal of Process Control, 11, 2001, 497-507.

[14] D. Dimla, P. Lister and N. Leighton, Neural network solutions to the tool condition monitoring problem in metal cutting – a critical review of methods, International Journal of Machine Tools and Manufacture, 37, 1997, 1219-1241.

[15] T. Kohonen, The self organizing map. Proceedings of IEEE 78, 1990, 1464-1480.