Comparing a Neural-Fuzzy Scheme with a Probabilistic Neural Network for Applications to Monitoring and Diagnostics in Manufacturing Systems

Kai Goebel, Bill Wood, Alice Agogino and Punit Jain

University of California at Berkeley

Abstract

The success of unattended manufacturing depends largely on control mechanisms that monitor the machining state and take actions to rectify unsatisfactory performance. Direct sensing methods like quality inspection lack on-line capability, whereas indirect methods using sensors can be thwarted by noise and changes in operating conditions. While knowledge about these changes exists, it does not generally correspond with an available sensor. Two different techniques are applied to the problem of integrating data from multiple sensors in the manufacturing environment: one featuring the integration of fuzzy logic and neural networks, and one using a probabilistic neural network. These techniques are applied to monitor and diagnose tool wear in unattended milling machines - an application with implications toward extension to other manufacturing machines.

Data from spindle motor current, acoustic emission, and vibration gathered in experiments on a Matsuura machining center are used as input to the two systems. In the case of the fuzzy-neural system, clusters for tool wear are established using the dendrogram method, then membership functions for these clusters are learned by a neural network. These clusters can be interpreted as fuzzy rules which are then applied to tool wear diagnosis using other principles of fuzzy logic. For the probabilistic neural network system, a network with fixed size is used for clustering of data and estimating the probability density function using a self-organizing probabilistic neural network (SOPNN).

Both systems show promising results with regard to tool wear. The advantage of the fuzzy neural-fuzzy system is that its classification seems to exhibit high reliability due to its redundant structure and efficiency of the preclustering. The advantage of the probabilistic network, on the other hand, is that it allows the use of rigorous probabilistic analysis, supports Bayesian network models and provides a means for the continuous updating of the density functions. The neural-probabilistic system has been tested successfully on data from an industrial power generation plant for application to sensor validation.

Introduction

The need of manufacturers to produce inexpensive, quality products has resulted in increasing demand for unattended and/or automated manufacturing systems. One problem in automating machining is how to deal with common malfunctions and disturbances such as tool wear, chatter, and tool breakage. Tool wear is a process which is very difficult to deal with for a variety of reasons. It is not a linear process: a tool wears fast initially, then at a moderate rate for a longer period of time, and finally at an accelerated rate until total failure. To complicate things, tool life is not constant under the seemingly same operating conditions. Many factors affect the operating life up to the wear limit: slight variations in the material of the workpiece, the degree of inclusions in the workpiece and slight temperature changes are but a few. To avoid costly damage due to tool wear or breakage, manufacturers use conservative operating procedures to prevent these malfunctions (Rangwala and Dornfeld, 1989). However, these result in less efficient and more costly production because of premature tool replacement and excessive machine downtime.

To increase operating efficiency, manufacturers can consider the use of sensors to diagnose tool status and control the system on-line. Since each sensor alone cannot reliably render the state of a tool in changing cutting conditions, integrating the information of various sensors becomes the major challenge. By using partly redundant information this sensor fusion can provide data for decision making about the process that will yield accurate diagnostic predictions and early warning of incipient failures. Early research focused on extracting relevant features from sensor data and inferring the tool status; others (Agogino, 1988, 1990) proposed the use of expert systems and sensor fusion using probabilistic influence diagrams. However, these approaches suffer from a high sensitivity to changing cutting conditions and varying sensor integrity and precision.

This paper summarizes two approaches to the problems outlined above: (1) a hybrid fuzzy-neural system and (2) a system using a probabilistic neural network which can be integrated into a Bayesian network or Influence Diagram. Preliminary results are compared and key questions that will drive future research discussed.

Neural-Fuzzy Reasoning

A neural-fuzzy approach is considered because it combines the advantages of both fuzzy logic and neural networks. Fuzzy logic can be useful in describing systems which are difficult to model and measure precisely, such as manufacturing processes. Reasoning in fuzzy logic is performed with IF-THEN relations and fuzzy membership functions. Initial intuition or experience and later tuning is necessary in most cases to achieve reasonable results. This becomes increasingly difficult with more sensors because higher dimensional membership functions are not easy to identify. Neural networks are used to learn the fuzzy relations by a simple mapping from clustered input to membership value, requiring no knowledge about the fuzzy function itself.

The procedure is divided in several steps (Takagi, 1991). First, the number of rules is determined by dividing the training data into different groups using a standard clustering method. The repeated use of a dendrogram finds classification-distinct clusters utilizing the centroid method, thus determining the number of rules. Next, the system is trained with a perfect fit membership value ($\mu = 1$) for the cluster it belongs to and no fit ($\mu = 0$) otherwise. This part resembles the IF portion of the fuzzy rule and can be expressed as:

$$NN_{mem}(x_1, x_2, x_3, \dots\ x_k) \text{ is } A^i \text{ is } \mu_i$$

where:

$NN_{mem}(x_1, x_2, x_3, \dots\ x_k)$ is the neural net into which x_1 - x_k are fed

A^i is a fuzzy number relating to the cluster i that x_1 - x_k belong to

μ_i is the membership value that is learned by this step

x_1 - x_k are the k different data gathered from the sensors at one time instance

Note that A^i is a fuzzy number with an unknown meaning. It can be thought of as a combination of quantities of the inputs, for example: acoustic emission at the table LARGE and vibration at the spindle SMALL and spindle motor current MEDIUM ...

Lastly, the THEN part of the rules is determined and the amount of diagnosing value for each rule trained. This is done by introducing as many additional neural nets as there are rules. Hereby, partly redundant information is combined it into a single justified value which has the effect of smoothing the output. Spikes which appear in a pure neural network are avoided. The categorization is therefore also much more consistent. The inputs to these nets are the sensor data; the output is the diagnosis for a rule. This relation can then be expressed as:

$$y_i = NN_i(x_1, x_2, x_3, \dots\ x_k)$$

where:

y_i is the diagnosis value for a rule

$NN_i(x_1, x_2, x_3, \dots\ x_k)$ is the neural net for rule i

The overall diagnostic output is obtained by taking the weighted average of membership values with the diagnostic value of each rule. The relation for the defuzzification for the diagnosis can be expressed as follows:

$$\beta^* = \frac{\sum y^i \mu^i}{\sum \mu^i}$$

where:

β^* = overall control value, in this case for tool wear

μ^i = membership value of rule i

This architecture is expressed in Figure 1.

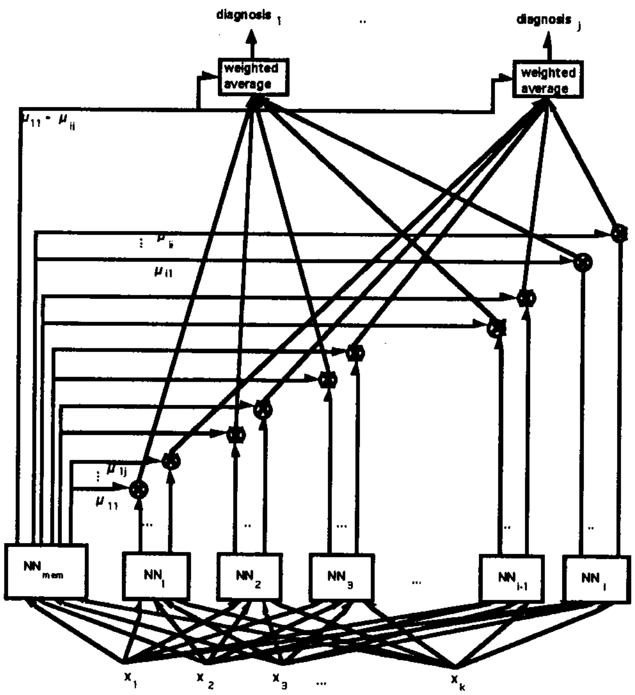

Figure 1: Architecture of the neural-fuzzy system

SOPNN (Self-Organizing Probabilistic Neural Network)

For comparison, a probability-based approach has been applied to the same data set using an algorithm designed by Tseng [1991]. The particular algorithm used is a Self-Organizing Probabilistic Neural Network, or SOPNN. This is based in the work of Specht [1988], but differs in that K-means clustering method is used to create a finite set of distribution centroids producing a fixed size network. The probability density function (pdf) of the system is then modeled by a weighted sum of non-covariant multivariate Gaussian probability distributions about the cluster centroids with a common variance used to smooth the distribution between clusters:

$$P(\mathbf{x}) = \frac{1}{2\pi^{d/2}\sigma^{d} n}\sum_{k} n_k e^{-\left(\frac{(\mathbf{x}-\mathbf{m}_k)(\mathbf{x}-\mathbf{m}_k)^T}{2\sigma^2}\right)}$$

where:

- $\mathbf{n}$ is the number of training points
- $\mathbf{k}$ is the number of clusters
- $\mathbf{m_k}$ is the centroid of cluster k
- $\mathbf{n_k}$ is the number of points in cluster k
- $\mathbf{d}$ is the dimension of the data space
- σ is the smoothing variance

This method produces a joint distribution which can be decomposed into conditional distributions used for detection of both process failure and sensor failure. Kim et al. [1992] have described a methodology for validation of sensor input data useful in industrial power generation applications. This method requires some preprocessing of data to generate features which are then modeled with a joint pdf. Knowledge of sensor failure modes is coupled to conditional pdf's generated from this joint pdf to separate data deviations generated by process faults from those caused by sensor error. Explicitly reasoning about the sensor is an essential part of applying knowledge based techniques to on-line systems and one which has been aided through the use of pdf's for process modeling. The SOPNN method should prove valuable for the integration of reasoning about sensor integrity in the manufacturing environment.

Figure 2 is the neural representation of the SOPNN system. Each set of four circled network subnodes represents a cluster of data. The subnodes are used to separate the input and output spaces to create the proper conditional pdf's used in diagnosis. Node inputs are derived from the quadratic term in the above definition of the

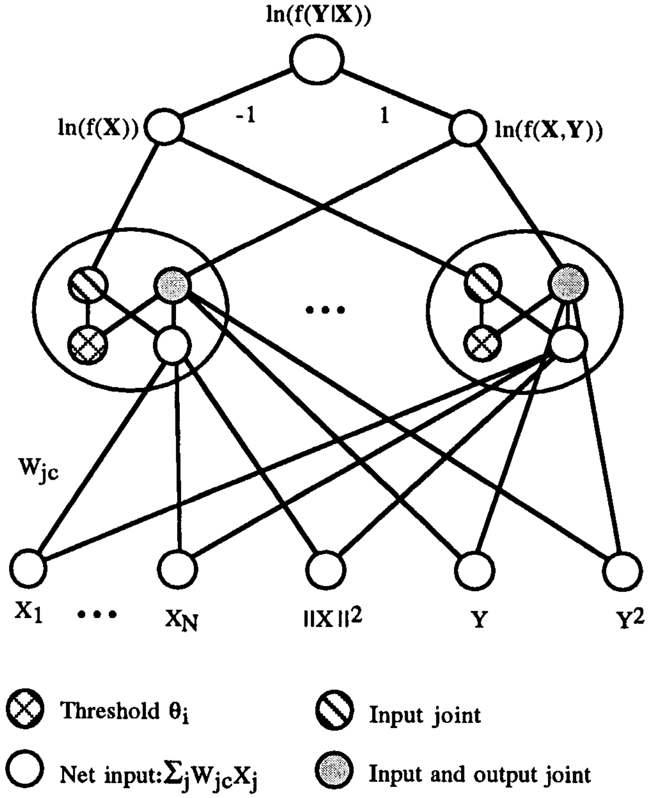

Figure 2: SOPNN probabilistic-neural system architecture

multivariate Gaussian pdf. The input data vector, **X**, is compared to each cluster centroid. Two thresholds are applied to this comparison: θ_i, the norms of the data clusters; and the norm of the input vector, $\|\mathbf{X}\|^2$. To form the joint distribution, these data sets are augmented by the one dimensional output variable, **Y**, and its appropriate thresholds. The output of the network is properly scaled as a post process to produce **P(Y|X)**.

Experimental Set-up

A Matsuura machining center was instrumented with five sensors measuring spindle motor current and vibration and acoustic emission at both the spindle and the table, sampled at 250 Hz. Two features were extracted from each data stream - the mean value and standard deviation within a window of fifty sample points. With this instrumentation, a series of machining operations was performed to establish training data to diagnose tool wear. From each of these operations, training sample sets of one hundred data points were taken at five stages of tool wear. Duplicate test runs using different tool inserts and workpieces were used to create data for testing the results of training.

Results

Fuzzy-Neural. The clustering algorithm of the neural-fuzzy system identified 7 different clusters according to their corresponding diagnosis values. The neural net for the membership function (NN_{mem}) was trained, using a 10-10-7 net. NN_1 - NN_7 used a 10-10-5 net each for 10,000 epochs. 100 data points is only a fraction of the data available (< 1%). Testing was done on a different set of data also from 5 different time periods in the tool life. The categorization of the diagnosis of the test run is displayed in Figure 3, the respective membership values are shown in Figure 4.

The system architecture divides the data properly into the different diagnosis categories fresh tool, slightly worn tool, half worn tool, considerably worn tool and tool at wear limit. Snapshots of the tool wear are shown after 3 min., 12 min., 42 min., 49 min., and 67 min., respectively. The degree of membership that is associated with the categories is shown in the diagrams right below the diagrams for the categories. The output node with the maximum membership value was selected for categorization of the tool wear class The membership is particularly crisp for a fresh and a worn tool. In between categorization is adequate. For some of the values the system alternates between the categories half worn tool and strong tool wear because this time period is at an in between stage. The trend can be read off correctly.

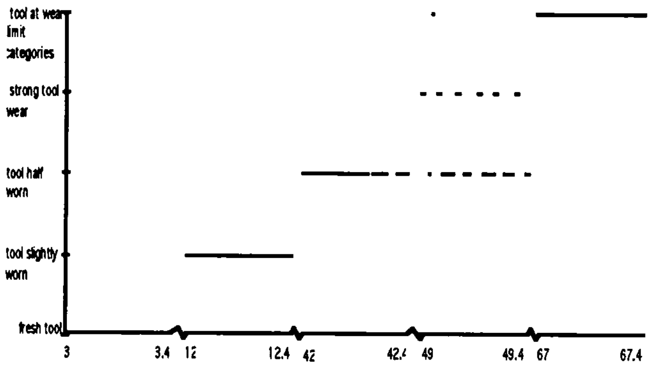

Figure 3: Categorization of the neural-fuzzy system

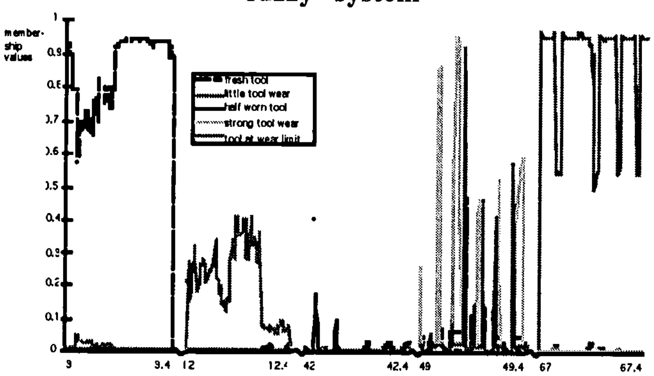

Figure 4: Membership values associated with the categories

Based on the test data obtained so far the hybrid fuzzy-neural system performs very well. It categorizes the diagnosis properly and delivers higher membership values for the categories that yield maximum membership values. It renders low membership values for the other categories.

Probabilistic-Neural: Clustering within the SOPNN was done using two slightly different algorithms. The first used random data presentation to seed the K-mean clusters with the first **k** samples after which data is parsed into the appropriate cluster by a simple distance measuring metric. A second clustering method was used in an attempt to reduce the number of clusters needed to provide adequate diagnostic performance. Because the data had been divided into five approximately discrete classification sample sets, each diagnostic class was clustered independently. This is similar in philosophy to the dendrogram method used in the Fuzzy-Neural scheme which also separates clusters by discrete classification; the original SOPNN algorithm was designed for continuous data only and so was somewhat handicapped in comparison.

Diagnosis within this probabilistic framework can be done using one of several methods. For example, typical pattern recognition methods can be used for discrete classification: the conditional pdf for each class cluster set can be integrated to get a class probability. For simplicity in a multimodal environment, selecting the diagnosis with maximum probability might be most appropriate. For this application, an expected value method captures more of the flavor of the system operation - tool wear is not by nature discrete, so its diagnosis need not be discrete.

Training the system according to the original SOPNN clustering algorithm (treating decision class as a continuous variable) resulted in performance that varies strongly with the number of clusters. Figure 5 shows the results of testing the SOPNN system on data gathered from instances of machine operation nominally the same as the training data but removed two weeks from component calibration and using different workpiece. In Case 1, 250 clusters were used to generate the probability distribution; performance is reasonable given the different operating conditions. Case 2 is the result of reducing the number of clusters to 100; Case 3, 50. Diagnostic performance shows an increase in error resulting from the lower number of clusters used,

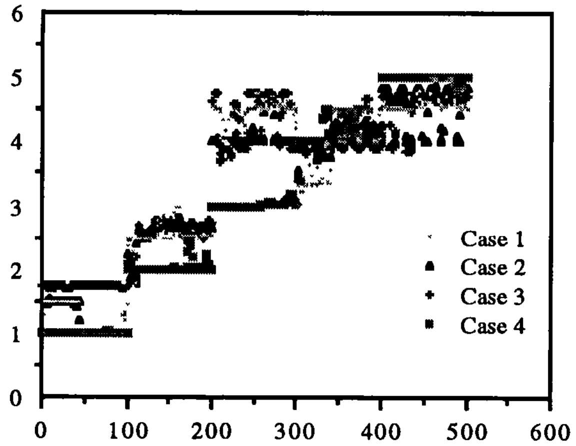

Figure 5: Diagnosis of tool wear using continuous classification clusters.

especially at the extremes of the diagnosis range. The effect is that of averaging training instances, projecting all clusters toward the overall centroid. Case 4 is the performance of the system when tested on data taken from the same operating conditions as the training data - performance here is quite good.

The second clustering method, forcing diagnosis clusters to be centered on discrete classifications, performed flawlessly on data taken from similar operation to the training set, shown as Case 1 of Figure 6. Generalizing the diagnosis to the same tested conditions of Cases 1 - 3 of Figure 5, the system performs poorly. The number of clusters for this training method is generally a factor of ten fewer than that used in the continuous case: Case 2 uses 10 clusters; Case 3, 50. Performance near the extremes of the range is greatly improved since cluster centroids are constrained to be there, but premature diagnosis of a worn tool largely obscures

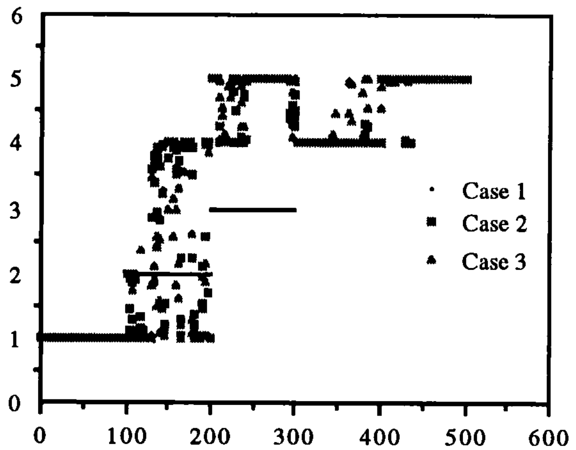

Figure 6: Diagnosis of tool wear using discretized classification clusters.

any gains made.

In both of the above tests, diagnostic error is unacceptably high, rendering the performance differences due to network size relatively insignificant. The SOPNN system using discrete classification was applied to separate training and testing data sets made by combining samples from the two operating conditions. In this case, diagnostic performance was again perfect, showing that the methodology is extensible to more difficult classification tasks if adequately trained.

Discussion

Perhaps the most highly touted aspect of neural systems is their ability to generalize from training instances to instances outside of their 'experience'. In our comparison, neither method significantly outperforms the other when confronted with data taken from operating conditions outside of direct experience. For the SOPNN system without discrete wear clustering, some generalization is already being done in the sense that cluster centroids exist where no samples have been taken; this seems to improve performance when the system is applied to new instances, but does not perform flawlessly on its own training data. Using discrete classification clustering produces flawless performance over the training data and like circumstances, but seems somewhat brittle when applied in new situations. Further investigation into incorporating fuzzy membership function into continuous wear values should be done. The SOPNN system is already capable of acting on continuous values.

One fundamental issue in both systems is the selection of features. Here, we chose mean and standard deviation of the sensor readings as features. Others might be more efficient. A related issue is the level of appropriate preprocessing of the data. Data that are not preprocessed (and that were stored for that purpose) should be used to check whether satisfying results can be obtained. Different signal conditioning techniques should also be pursued to see whether they might render satisfactory results. The Fast-Fourier Transform gives relevant features that might improve the diagnostic capabilities, as indicated by past work (Rangwala, 1988; Agogino, 1988). In the same spirit, other techniques from speech recognition, such as Cepstrum, could be used to identify and expand the feature set.

Future work in the Neural-Fuzzy system is to investigate the role of overlearning and overfitting as well as the size of the neural net to determine to which extent they influence the result (Takagi, 1991). Alternate neural net types should be tried. Recurrent nets are one possibility which might recognize the trend of sensor readings with regard to tool wear. One interesting application beyond tool wear is predicting the spontaneous occurrence of chatter which occurs when the geometry of the piece is not straight and the feed rate cannot easily be determined. This event might be more easily recognized by this type of neural net than the standard type. Also interesting might be the use of radial basis functions as activation functions. They have an inherent fuzzy behavior which might make them in particular useful for this type of problem.

Further investigations should look at the problem of noise in learning. A backward elimination can be used that considers only four out of five sensor readings and checks the error after a given number of epochs. This can be done in turn for all sensors and the error compared. Sensor inputs that just contribute noise to the results can thus be eliminated. Multisensor systems are highly sensitive to the reliability of the sensor values upon which they are based. Thus the integrity of the sensors defines the ability of these systems to monitor operations and diagnose failure. We see this as a motivation to extend this research into on-line sensor validation building on previous work in Bayesian influence diagrams [Tseng, 1991] and fuzzy influence diagrams [Jain, 1990].

Algorithm efficiency is a prominent aspect of any discussion of real-time or near real-time methodologies. In our comparison, there really is no contest. The fuzzy-neural scheme is perhaps an order of magnitude faster for diagnosis. This efficiency may come at the expense of fault tolerance for the probabilistic system can be easily extended to diagnose and correct for sensor faults, something which is a key research direction to pursue toward applying the fuzzy-neural methodology. Overall, both methods provide good results when applied in situations close to their 'experience'. Training and testing over a wider array of operating situations is a crucial step in the assessment of the ultimate usefulness of either scheme for manufacturing applications.

Acknowledgements

The authors thank Prof. D. Dornfeld and Prof. Tomizuka for providing access to and expertise on the machining and test equipment. Special thanks go to Dr. Takagi for his valuable advice. Thanks also to Andrzej Sokolowski and Piotr Dworak for helping with the experiments. The authors gratefully acknowledge the support of Project MICRO (contract #93-004), General Electric, Ricoh, Autodesk and Neuron Data.

References

Agogino, A., and S. Srinivas "Multiple Sensor Expert System for Diagnostic Reasoning, Monitoring and Control of Mechanical Systems," *Mechanical Systems and Signal Processing*, Vol. 2, No. 2, 1988.

Agogino, A. M., and K. Ramamurthi "Real Time Influence Diagrams for Monitoring and Controlling Mechanical Systems," *Influence Diagrams, Belief Nets and Decision Analysis*, Ed.: Oliver, R. M., and J.O. Smith. J.R. Wiley & Sons, Ltd., 1990.

Dornfeld, D.A. "Acoustic Emission Process Monitoring for Untended Manufacturing", *Proc. Japan-USA Symposium on Flexible Automation*, JAACE, Osaka, Japan, 1986.

Jain, P. and A.M. Agogino, "Stochastic Sensitivity Analysis using Fuzzy Influence Diagrams," in *Uncertainty in Artificial Intelligence 4*, eds., R.D. Shachter, T.S. Levitt, L.N. Kanal, J.F. Lemmer; Elsevier Science Publishers V.V., North-Holland Press, 1990, pp. 79-92.

Kim, Y.J., W.H. Wood, and A.M. Agogino, "Signal Validation for Expert System Development," *Proceedings of the 2nd International Forum on Expert Systems and Computer Simulations in Energy Engineering*, March 17-20, 1992 (Erlangen, Germany), pp. 9-5-1 to 9-5-6.

Rangwala, S.S. "Machining Process Characterization and Intelligent Tool Condition Monitoring Using Acoustic Emission Signal Emission Signal Analysis", PhD thesis, UC Berkeley, 1988.

Takagi, H., and I. Hayashi "NN - Driven Fuzzy Reasoning", *International Journal of Approximate Reasoning 1991*, 5: Elsevier Science Publishing Co.

Specht, D.F., "Probabilistic Neural Networks for Classification, Mapping, or Associative Memory," IEEE International Conf. on Neural Networks, vI, san Diego, CA, July 24-27, 1988, pp. 525-532.

Tseng. M.-L., "Integrating Neural Networks and Influence Diagrams for multiple sensor diagnostic systems," PhD thesis, Department of Mechanical Engineering, University of California at Berkeley, 1991.