DEPARTMENT OF THE NAVY
NAVAL UNDERSEA WARFARE CENTER DIVISION
1176 HOWELL STREET
NEWPORT RI 02841-1708

IN REPLY REFER TO:

Attorney Docket No. 79833
Date: 20 October 2004

The below identified patent application is available for licensing. Requests for information should be addressed to:

PATENT COUNSEL
NAVAL UNDERSEA WARFARE CENTER
1176 HOWELL ST.
CODE 00OC, BLDG. 112T
NEWPORT, RI 02841

Serial Number	10/857,371
Filing Date	26 May 04
Inventor	Christopher M. DeAngelis

If you have any questions please contact James M. Kasischke, Deputy Counsel, at 401-832-4736.

DISTRIBUTION STATEMENT A
Approved for Public Release
Distribution Unlimited

Attorney Docket No. 79833
Customer No. 23523

PROBABILITY DISTRIBUTION CLASSIFICATION PROCESSOR

TO WHOM IT MAY CONCERN:

BE IT KNOWN THAT CHRISTOPHER M. DeANGELIS AND ROBERT C. HIGGINS, employees of the United States Government, citizens of the United States of America, residents respectively of Cranston, County of Providence, State of Rhode Island and Tiverton, County of Newport, State of Rhode Island, have invented certain new and useful improvements entitled as set forth above of which the following is a specification:

JEAN-PAUL A. NASSER, ESQ.
Reg. No. 53372
Naval Undersea Warfare Center
Division Newport
Newport, RI 02841-1708
TEL: 401-832-4736
FAX: 401-832-1231

Attorney Docket No. 79833

PROBABILITY DISTRIBUTION CLASSIFICATION PROCESSOR

STATEMENT OF GOVERNMENT INTEREST

The invention described herein may be manufactured and used by or for the Government of the United States of America for governmental purposes without the payment of any royalties thereon or therefore.

CROSS REFERENCE TO OTHER RELATED APPLICATIONS

Not applicable.

BACKGROUND OF THE INVENTION

(1) Field of the Invention

The present invention relates generally to classifying probability distributions and, more particularly, to a neural network system and method for processing random data with an unknown probability distribution function to thereby classify the best represented probability distribution thereof.

(2) Description of the Prior Art

The basic prior art technique for classifying probability distributions involves the use of the histogram. The purpose of a histogram is to graphically summarize the distribution of a data set. The histogram graphically attempts to show the following: the center (i.e., the location) of the data, the spread (i.e., the scale) of the data, the skewness of the data, the presence of outliers, and the presence of multiple modes in the data. One disadvantage of the histogram technique is that refinements must often be made in the data interval sizes to classify the distribution, which often involves a trial and error approach. Another disadvantage is the time-consuming requirement for visual examination of the features of the histogram to provide indications of the proper distributional model for the data so that a hypothesis can be formed as to the type of distributional model. Mathematical means of confidence such as the probability plot or a goodness-of-fit test can then be used to verify the hypothesized distributional model.

After obtaining a hypothesis utilizing the histogram, the probability plot is a graphical technique for assessing whether or not a data set follows a given distribution such as the normal or Weibull. The data are plotted against a theoretical distribution in such a way that the points should form approximately a straight line. Departures from this straight line indicate departures from the specified distribution. The correlation coefficient associated with the linear fit to the data in the probability plot is a measure of the goodness of the fit. Estimates of the location and scale parameters of the distribution are given by the intercept and slope. Probability plots can be generated for several competing distributions to see which provides the best fit, and the probability plot generating the highest correlation coefficient is the best choice since it generates the straightest probability plot.

Another means for verifying the hypothesis for the probability distribution is the Anderson-Darling goodness-of-fit test, which tests if a sample of data came from a population with a specific distribution. It is a modification of the Kolmogorov-Smirnov (K-S) test and gives more weight to the tails than does the K-S test. The K-S test is distribution-free in the sense that the critical values do not depend on the specific distribution being tested. The Anderson-Darling test makes use of the specific distribution in calculating critical values. This has the advantage of allowing a more sensitive test and the disadvantage that critical values must be calculated for each distribution.

In the field of statistics, due to the difficulty as explained above, for determining the probability distribution of random data, assumptions are usually made about the probability characteristics of a sample of random data because the probability distribution is otherwise unknown. The probability distribution function, if known, would provide information about the frequency of occurrence of each data element in a sample of random (non-deterministic) data that consists of several data elements. Statistics like the mean and standard deviation of the data are estimated based on assumptions about the shape of the probability distribution function that characterizes random data. Therefore, the calculated statistics associated with a Binomial Distribution will be different from a Poisson or a Gaussian Distribution, or any other type of distribution. When modeling parameters (i.e. phase, amplitude, and frequency) associated with certain types of acoustic interference like reverberation, multipath or noise, the models may typically assume a Gaussian or some other distribution to represent these acoustic phenomena.

The following U.S. Patents describe various prior art systems that may be at least related to the problems discussed above, but which do not provide suitable solutions.

U.S. Patent Application No. 2001/0031064 A1, published October 18, 2001, to Donescu et al., discloses a method of inserting a watermarking signal in a set of coefficients representing a digital image in which at least one subset of coefficients is modified by the watermarking signal. For each representative coefficient to be modified, a neighborhood of the representative coefficient to be modified is determined. A neighborhood in a dictionary of neighborhoods is selected according to a predetermined criterion of similarity with the neighborhood for the representative coefficient under consideration. The representative coefficient is modified as a function of the watermarking signal.

U.S. Patent No. 6,324,532, issued November 27, 2001, to Spence et al., discloses a signal processing apparatus and concomitant method for learning and integrating features from multiple resolutions for detecting and/or classifying objects. The signal processing apparatus comprises a hierarchical pyramid of neural networks (HPNN) having a "fine-to-coarse" structure or a combination of the "fine-to-coarse" and the "coarse-to-fine" structures.

U.S. Patent No. 6,278,970, issued August 21, 2001, to Benjamin P. Milner, discloses how to calculate the log frame energy value of each of a pre-determined number n of frames of an input speech signal and apply a matrix transform to the n log frame energy values to form a temporal matrix representing the input speech signal. The matrix transform may be a discrete cosine transform.

U.S. Patent No. 6,006,186, issued December 21, 1999, to Chen et al., discloses a method and an apparatus for a parameter sharing speech recognition system. Speech signals are received into a processor of a speech recognition system. The speech signals are processed using a speech recognition system hosting a shared hidden Markov model (HMM) produced by generating a number of phoneme models, some of which are shared. The phoneme models are generated by retaining as a separate phoneme model any triphone model having a number of trained frames available that exceeds a prespecified threshold. A shared phoneme model is generated to represent each of the groups of triphone phoneme models for which the number of trained frames having a common biphone exceed the prespecified threshold. A shared phoneme model is generated to represent each of the groups of triphone phoneme models for which the number of trained frames having an equivalent effect on a phonemic context exceed the prespecified threshold. A shared phoneme model is generated to represent each of the groups of triphone phoneme models having the same center context. The generated phoneme models are trained, and shared phoneme model states are generated that are shared among the phoneme models. Shared probability distribution functions are generated that are shared among the phoneme model states. Shared probability sub-distribution functions are generated that are shared among the phoneme model probability distribution functions. The shared phoneme model hierarchy is reevaluated for further sharing in response to the shared probability sub-distribution functions. Signals representative of the received speech signals are generated.

U.S. Patent No. 5,924,066, issued July 13, 1999, to Amlan Kundu, discloses a system and method for classifying a speech signal within a likely speech signal class of a plurality of speech signal classes. Stochastic models include a plurality of states having state transitions and output probabilities to generate state sequences, which model evolutionary characteristics and durational variability of a speech signal. The method includes extracting a frame sequence, and determining a state sequence for each stochastic model with each state sequence having full state segmentation. Representative frames are determined to provide speech signal time normalization. A likely speech signal class is determined from a neural network having a plurality of inputs receiving the representative frames and a plurality of outputs corresponding to the plurality of speech signal classes. An output signal is generated based on the likely stochastic model.

U.S. Patent No. 6,336,109, issued January 1, 2002, to Gary Howard, discloses a method of processing data relating to a plurality of examples using a data classifier arranged to classify input data into one of a number of classes, and a rule inducer, comprising the steps of: (i) inputting a series of inputs to the data classifier so as to obtain a series of corresponding outputs; (ii) inputting the series of outputs and at least some of the series of inputs to the rule inducer so as to obtain a series of rules which describe relationships between the series of inputs to the data classifier and the series of corresponding outputs from the data classifier.

U.S. Patent No. 6,314,399, issued November 6, 2001, to Deligne et al., discloses an apparatus that generates a statistical class sequence model called A class bi-multigram model from input training strings of discrete-valued units, where bigram dependencies are assumed between adjacent variable length sequences of maximum length N units, and where class labels are assigned to the sequences. The number of times all sequences of units occur are counted, as well as the number of times all pairs of sequences of units co-occur in the input training strings. An initial bigram probability distribution of all the pairs of sequences is computed as the number of times the two sequences co-occur, divided by the number of times the first sequence occurs in the input training string. Then, the input sequences are classified into a pre-specified desired number of classes. Further, an estimate of the bigram probability distribution of the sequences is calculated by using an EM algorithm to maximize the likelihood of the input training string computed with the input probability distributions. The above processes are then iteratively performed to generate statistical class sequence model.

U.S. Patent No. 6,239,740, issued May 29, 2001, to Collins et al., discloses an efficient algorithm for evaluating the (weighted bipartite graph of) associations between two sets of data with Gaussian error, e.g., between a set of measured state vectors and a set of estimated state vectors. First a general method is developed for determining, from the covariance matrix, minimal d-dimensional error ellipsoids for the state vectors, which always overlap when a gating criterion is satisfied. Circumscribing boxes, or d-ranges, for the data ellipsoids are then found and whenever they overlap the association probability is computed. For efficiently determining the intersections of the d-ranges a multidimensional search tree method is used to reduce the overall scaling of the evaluation of associations. Very few associations that lie outside the predetermined error threshold or gate are evaluated. Empirical testing for variously distributed data in both three and eight dimensions indicate that the scaling is significantly reduced from N^2, where N is the size of the data set. Computational loads for many large scale (N>10-100) data association tasks may therefore be significantly reduced by this or related methods.

U.S. Patent No. 6,131,089, issued October 10, 2000, to Campbell et al., discloses classifiers and a comparator that perform an identification method to identify a class as one of a predetermined set of classes. The identification method is based on determining the observation costs associated with the unidentified class. The identification method includes combining models representing the predetermined set of classes and the unidentified vectors representing the class. The predetermined class associated with the largest observation cost is identified as the class. Additionally, a unique, low-complexity training method includes creating the models, which represent the predetermined set of classes.

U.S. Patent No. 6,041,299, issued March 21, 2000, to Schuster et al., discloses an apparatus for calculating a posteriori probabilities of phoneme symbols and a speech recognition apparatus using the apparatus for calculating a posteriori probabilities of phoneme symbols. A feature extracting section extracts speech feature parameters from a speech signal of an uttered speech sentence composed of an inputted character series, and a calculating section calculates the a posteriori probability of a phoneme symbol of the speech signal, by using a bidirectional recurrent neural network. The bidirectional recurrent neural network includes an input layer for receiving the speech feature parameters extracted by the feature extracting means and a plurality of hypothetical phoneme symbol series signals, an intermediate layer of at least one layer having a plurality of units, and an output layer for outputting the a posteriori probability of each phoneme symbol. The input layer includes a first input neuron group having a plurality of units, for receiving a plurality of speech feature parameters and a plurality of phoneme symbol series signals, a forward module, and a backward module.

U.S. Patent No. 5,999,893, issued December 7, 1999, to Lynch, Jr. et al., discloses a classification system that uses sensors to obtain information from which features, which characterized a source or object to classified, can be extracted. The features are extracted from the information and compiled into a feature vector, which is then quantized to one of M discrete symbols. After N feature vectors have been quantized, a test vector having components which are defined by the number of occurrences of each of the M symbols in N the quantized vectors is built. The system combines the test vector with training data to simultaneously estimate symbol probabilities for each class and classify the test vector using a decision rule that depends only on the training and test data. The system classifies the test vector using either a Combined Bayes test or a Combined Generalized likelihood ratio test.

The above disclosed prior art does not provide an automatic system and method which can be easily implemented with hardware or on a computer to classify random data into any type of the probability distribution and which does not involve visual examination, trial and error approaches, iterative techniques, measures of confidence, and the like. It is the inventors' belief that in the design of many signal processors, knowledge about the probability and statistical characteristics of the noise that contaminates a signal input may improve the capability of the processor. Accordingly, exact knowledge of these phenomena would provide more accurate results. Moreover, it is the inventors' belief that an automatic mechanism to conveniently process a set of data samples and determine the associated probability distribution would therefore likely be invaluable to researchers. As discussed above, the prior art does not provide such a method. The need for solutions to the above-described problems has been long felt but the solutions have not been forthcoming. Consequently, those skilled in the art will appreciate the present invention that addresses the above and other problems.

SUMMARY OF THE INVENTION

It is a general purpose and object of the present invention to provide an improved system and method for determining a probability distribution for random data.

It is another general purpose and object of the present invention to provide a system and method to which the random data can be input for automatically determining which of several different types of probability distributions best describe the random data.

It is another objection of the present invention to utilize a plurality of trained neural networks to sample the random data for determining the probability distribution thereof.

These and other objects, features, and advantages of the present invention will become apparent from the drawings, the descriptions given herein, and the appended claims. However, it will be understood that above listed objects and advantages of the invention are intended only as an aid in understanding aspects of the invention, are not intended to limit the invention in any way, and do not form a comprehensive list of objects, features, and advantages.

According, the present invention provides a system for determining a probability distribution of random data wherein the system may comprise one or more elements such as, for example, one or more first neural networks each trained to recognize a first probability distribution of the random data, one or more second neural networks each trained to recognize a second probability distribution of the random data, and a decision logic module for comparing outputs for the one or more first neural networks and the second neural networks to thereby select the first probability distribution or the second probability distribution as best describing the probability distribution of the random data. The decision logic module is preferably operable for selecting the probability distribution of the random data based on a set of rules related to specified ranges of values of the outputs for first neural networks and the second neural networks.

The system may further comprise a window generator for providing a selected size sample of the random data to the one or more first neural networks and the one or more second neural networks. In one preferred embodiment, the system comprises a plurality of the first neural networks and a plurality of the second neural networks wherein the window generator is operable for providing variable size samples to each of the plurality of first neural networks and the plurality of second neural networks.

Other elements of the system may further comprise a parameter estimator operable for determining one or more statistical parameters of the selected size sample. The parameter estimator preferably produces an output receivable by the decision logic module for determining whether to continue processing the random data or to stop processing the random data.

Preferably, a normalizer is utilized for normalizing the random data to fall within a selected range of values.

In operation, the present invention provides a method for automatically determining a probability distribution for random data which may comprise one or more steps such as, for instance, providing different size samples of the random data to a first plurality of neural networks, providing the different size samples of the random data to a second plurality of neural networks, and comparing outputs of the first plurality of neural networks with the outputs of the second plurality of neural networks for determining the probability distribution of the random data.

The method further comprises providing a set of logic rules for determining the probability distribution of the random data based on ranges of values of the outputs of the first plurality of neural networks and the outputs of the second plurality of neural networks.

The method may further comprise training the first plurality of neural networks to produce a selected output value in response to random data corresponding to a first probability distribution, and training the second plurality of neural networks to produce a selected output value in response to random data corresponding to a second probability distribution.

The method may preferably further comprise determining statistical parameters of the different size samples of the random data.

In a preferred embodiment, the method comprises utilizing a sliding window of random data for the different size data samples whereby the sliding window of random data changes after each use by adding a selected number of new data samples and subtracting a corresponding number of previously processed data samples.

BRIEF DESCRIPTION OF THE DRAWINGS

A more complete understanding of the invention and many of the attendant advantages thereto will be readily appreciated as the same becomes better understood by reference to the following detailed description when considered in conjunction with the accompanying drawings, wherein like reference numerals refer to like parts and wherein:

FIG. 1A is a schematic of one possible neural network configuration for a component of the system in accord with the present invention;

FIG. 1B is a schematic of one neuron in a layer of neurons of a neural network such as the neural network of FIG. 1A;

FIG. 2 is a block diagram for one possible probability distribution classification processor in accord with the present invention for two probability distributions;

FIG. 3 is a graph showing possible outputs for the six artificial neural networks (ANN) processors of the probability distribution classification processor of FIG. 2;

FIG. 4 is a graph showing outputs of the parameter estimator for the probability distribution classification processor of FIG. 2; and

FIG. 5 is a block diagram for the rule-based decision aide of the probability distribution classification processor of FIG. 2.

DESCRIPTION OF THE PREFERRED EMBODIMENT

Referring now to the drawings and, more particularly, to FIG. 2, there is shown system 10 which is a block diagram for a preferred embodiment of a probability distribution classification processor in accord with the present invention. The present invention therefore provides system and methods for classifying the form of the probability distribution that describes a sample of random data elements.

To simplify the description and operation of this processor, the design of system 10 presented here will have the capability to classify only two different probability distributions. However, it will be understood that the addition of capability to classify data into other probability distributions will be substantially the same as discussed herein below for two frequently occurring probability distributions. Thus, it will be understood that the processor capability can be readily expanded to automatically classify the random data into any number of different types of possible probability distributions. In one embodiment, the capability for considering different probability distributions could be switchable or selectable so that consideration of any of numerous different probability distributions could be easily switched on as desired for testing by the user.

In the embodiment of the invention disclosed in FIG. 2, two different probability distributions are considered by system 10. The first probability distribution is the well-known Uniform distribution that represents data where each random element (number) has the same probability of occurrence. The second probability distribution to be considered is the well-known Gaussian distribution that represents data where each element has a different probability of occurrence with the larger positive and negative random elements (numbers) occurring less frequently than the smaller ones. When the mean is zero, the Gaussian distribution is referred to as a Normal distribution.

FIG. 1A and 1B show artificial neural network (ANN) elements that are trained to recognize a particular type of probability distribution, such as the Uniform distribution or the Gaussian distribution. Neural network 12 may comprise several layers of interconnected neurons wherein each neuron, such as neuron 14, may preferably be connected to each neuron of the previous layer of neurons and to each neuron of the subsequent layer of neurons. It will be understood that neural network 12 may be much larger than shown in FIG. 1A and/or comprise more layers and/or more neurons per layer. For instance, to process 100 data inputs simultaneously, the neural network may comprise a first layer of 100 neurons. Each neuron in each layer, such as neuron 14 of FIG. 1B, may simply weight and sum inputs 16 and offset by a bias 17 with adder 18. The result of adder 18 is passed through activation function 20.

Activation function 20, which may be a Sigmoid function, is used to limit the permissible amplitude range of output 22. For instance, the output may be limited to between zero to one or from minus one to plus one. Neuron output 22 is passed on to each neuron in the next layer of neurons and so on until the last layer produces output 24 for the neural network, such as the respective outputs of neural networks 26, 28, 30, 32, 34, and 36 of system 10 shown in FIG. 2.

Training the neural network(s) involves adaptively and systematically selecting the weights and bias of each neuron within the network to minimize the error between the actual output and the desired one. For example, if it desired to have a neural network produce a one when the input random data is uniformly distributed and a zero otherwise, the internal weights and bias of each neuron are established by an adaptive iterative process to make this happen. After the network has been trained, the output will provide a measure of the uniformity of the distribution of the random data. A detailed description of a neural network may be found in the literature (e.g., *Neural Networks* by Simon Haykin, 1994). The specific type of neural network that is used in this application may preferably be a back propagation model. In this type of model, the correct weights and bias of each are established by modifying the last layer of neuron parameters first in an attempt to produce the desired output, then if necessary modifying the previous layer, and so forth.

FIG. 2 is therefore a block diagram of one embodiment of system 10 for a Probability Distribution Classification Processor in accord with the present invention. Input data 38 preferably consists of or comprises a finite set of random numbers. As a first step in a preferred embodiment of the invention, the entire set of random numbers is preferably normalized to a new set of data between -1 and +1 in normalizer 40. Normalization simplifies the process of training the neural networks. The Gaussian distribution artificial neural network (ANN) classifier can thereby be trained using a finite number of statistics. For example, training may be performed using Gaussian data representing standard deviations from 0.1 to 1 in increments of 0.1 and means from -1 to +1 in increments of 0.1. Each Gaussian and Uniform neural network classifier will be trained using a different sized sliding window of normalized data as indicated at 42. Therefore, system 10 will be comprised of several different neural networks, such as networks 26, 28, 30, 32, 34, and 36 corresponding to the number of different sliding windows of data samples for each type of distribution to be classified.

After normalizing, the numbers are processed by artificial neural networks (ANN) 26, 28, 30, 32, 34, and 36. In this example, the intervals or windows comprise 100, 500, and 1000 data samples. However, other interval sizes could also be utilized and/or additional interval sizes could be utilized. Preferably at least two different interval or window sizes are utilized for comparison purposes as discussed hereinafter. Sliding window function 42 may be used to select the data samples for the respective networks. For instance, using a 100 sample sliding window, the first 100 data values (samples 1 through 100) are selected and processed, then the second 100 values (samples 2 through 101) are selected and processed; this process continues until all the data has been processed. The 500 and 1000 sample windows of data are processed in the same manner with all sliding window operations (100, 500, and 1000) performed concurrently.

In FIG. 2, three ANN Gaussian Distribution Classification processors 26, 28, and 30 as well as three ANN Uniform Distribution Classification processors 32, 34, and 36 are shown. The three ANN Gaussian processors will be capable of processing data windows of 100, 500, and 1000 samples of the normalized random data respectively. The three ANN Uniform processors will have a corresponding capability. As discussed earlier, a 100 sample sliding window function 42 will initially select data samples 1 through 100 for processing. Then, data samples 2 through 101 are selected. This process continues incrementally until all the data has been processed. Each ANN Uniform and each ANN Gaussian processor will be trained to produce an output between 0 and 1 for each sliding window increment of data, where the output of the ANN is proportional to the degree in which the data matches the respective distribution. Each output will be stored within Rule-Based Decision Aide 44 until all the data has been processed by the neural networks. It should be noted that any number of ANNs can be employed at any window size to enhance the data classification process. Figure 3 shows one set of possible ANN 26, 28, 30, 32, 34, and 36 outputs with respect to each increment of the sliding window.

Parameter Estimator 46 shown in FIG. 2 will also provide information to Rule-Based Decision Aide 44. Relevant statistics like the mean and standard deviation and/or other statistics are preferably measured and stored within Rule-Based Decision Aide 44 as each sliding window of data is processed by the Parameter Estimator 46. FIG. 4 shows the type of information that the Parameter Estimator provides such as mean and standard deviation for each increment of the sliding window. In this case, it will be noticed that curve 54 for the mean 100 window samples, curve 56 for the mean 500 window samples, and curve 58 for the mean 1000 window samples tend to gravitate roughly about the same approximate values. The standard deviations 60, 62, and 64 of the three different window sample sizes are somewhat more variable. This information is fed to Rule-Based Decision Aid 44 to be acted upon based on the rules provided therein and to be available if required to further describe the selected probability distribution.

Rule-Base Decision Aide 44 uses the inputs from the Parameter Estimator 46 and Neural Networks 26-36 to decide what type of distribution best describes the data samples. FIG. 5 provides a block diagram of the components of Rule-Based Decision Aide 44. As neural networks 26-36 and Parameter Estimator 46 process the data, the outputs of these functions are stored in Database 48. Then Data Evaluator 50 processes and examines these outputs (like the ones seen in Figures 3 and 4) to determine if the progressively larger sliding windows have produced less variability in the ANN 26-36 and Parameter Estimator 46 outputs. If the standard deviation and/or mean of the outputs calculated by the Data Evaluator 50 do not significantly change as the sliding window size increases, the decision is made to STOP processing and any one of the sample neural network outputs may be utilized. If the standard deviation does decrease with window size, the outputs corresponding to the 1000 sample sliding window are averaged to get the most representative single values for ANN Uniform output 36, ANN Gaussian output 30. These values along with the normalized mean and standard deviation (Parameter Estimator 46) outputs, are passed along to Decision Logic component 52 in FIG. 5. This component classifies the Probability Distribution based on a pre-determined logic scheme.

For example, in one logic scheme, if the ANN Uniform mean output is between 0.0 and 0.5, and the ANN Gaussian mean output is between 0.5 and 1.0, then Rule-Based Decision Aide 44 decides the data fits most closely to a Gaussian probability distribution. On the other hand, if the mean ANN Uniform output is between 0.5 and 1.0, and the mean ANN Gaussian output is between 0.0 and 0.5, then Rule-Based Decision Aide 44 decides the data fits most closely to a Uniform probability distribution. However, if the mean ANN Uniform output and the mean ANN Gaussian output are both within a selected range around 0.5, then Rule-Based Decision Aide 44 decides that more data is needed and the data processing continues using larger windows as discussed above. Other logic schemes can be utilized which will relate to the value for which each group of neural networks have been trained to output for the probability distribution the respective group of neural networks is trained to recognize.

As mentioned earlier, each of the six neural networks (ANN) are trained in advance, each with a different sliding window size. Each ANN Uniform network 32-36 is trained on data that is known in advance to be uniformly distributed data. This data can be generated using one of many commercially available uniform random number generators. The ANN Uniform adaptive elements 32-36 will be trained to converge to values that produce an output close to 1 when the input data is Uniformly distributed and close to 0 if it is not. Hence, the output of each Uniform neural network 32-36 when processing non-training data will be a value between 0 and 1 that will indicate how close the data matches uniformly distributed data. Likewise, the ANN Gaussian networks 26-28 are trained on data that is known in advance to be Gaussian distributed data. This data can be generated using one of many commercially available Gaussian random number generators. This training process is more complicated because the ANN Gaussian networks are trained to recognize Gaussian data regardless of the mean or Standard deviation. Training will consist of using data characterized by different means or standard deviations. The range of standard deviations will be 0 to 1 in increments of 0.1; the range of means will be -1 to +1 in increments of 0.1. The output will be 1 if the input data is Gaussian.

The advantage that this disclosure provides is the increased statistical knowledge about a data set with probability characteristics that are unknown. This disclosure enables one to process the data and classify the probability distribution. The process is performed automatically, is not an iterative process, and does not involve visual examination or the implementation of mathematical measures of confidence. Furthermore, the disclosure may be extended to include any other distribution.

Many additional changes in the details, materials, steps and arrangement of parts, herein described and illustrated to explain the nature of the invention, may be made by those skilled in the art within the principle and scope of the invention. It is therefore understood that within the scope of the appended claims, the invention may be practiced otherwise than as specifically described.

Attorney Docket No. 79833

PROBABILITY DISTRIBUTION CLASSIFICATION PROCESSOR

ABSTRACT OF THE DISCLOSURE

A system and method are disclosed which preferably comprises several groups of artificial neural networks (ANNs) for classifying the probability distribution of random data. Each group of artificial neural networks is preferably trained to produce a selected output in response to data having a particular probability distribution. Each of the group of artificial neural networks preferably analyzes a different sample size of data. A parameter estimator module calculates statistical parameters for the different size data samples. The outputs of the several groups of artificial networks and of the parameter estimator are analyzed by a rule based decision logic module which then selects the type of probability distribution that best describes the random data based on rules that correspond to ranges of values of the outputs of the artificial neural networks.

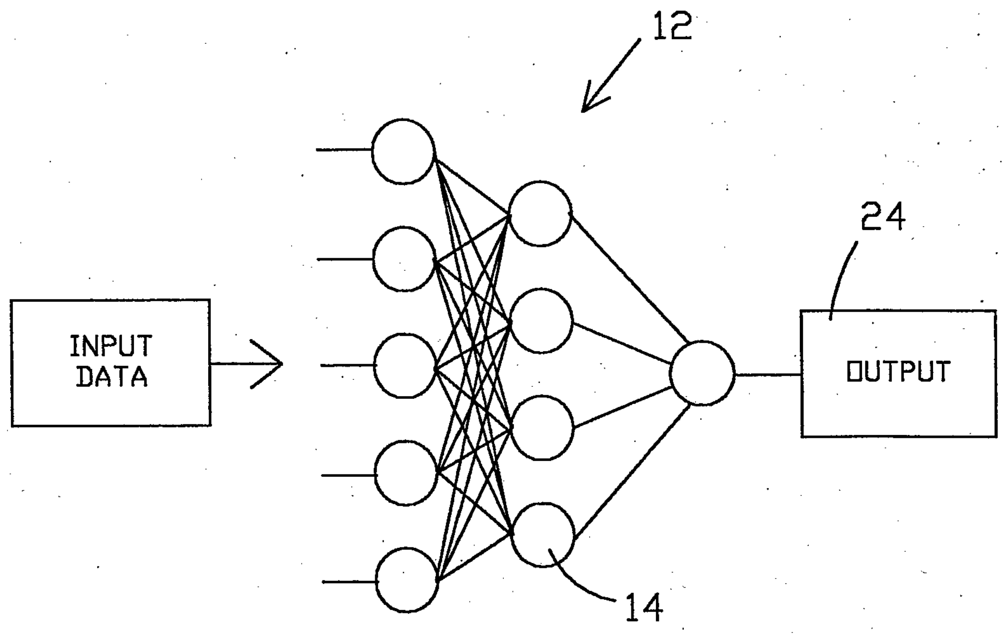
12
24
INPUT DATA
OUTPUT
14

FIG. 1A

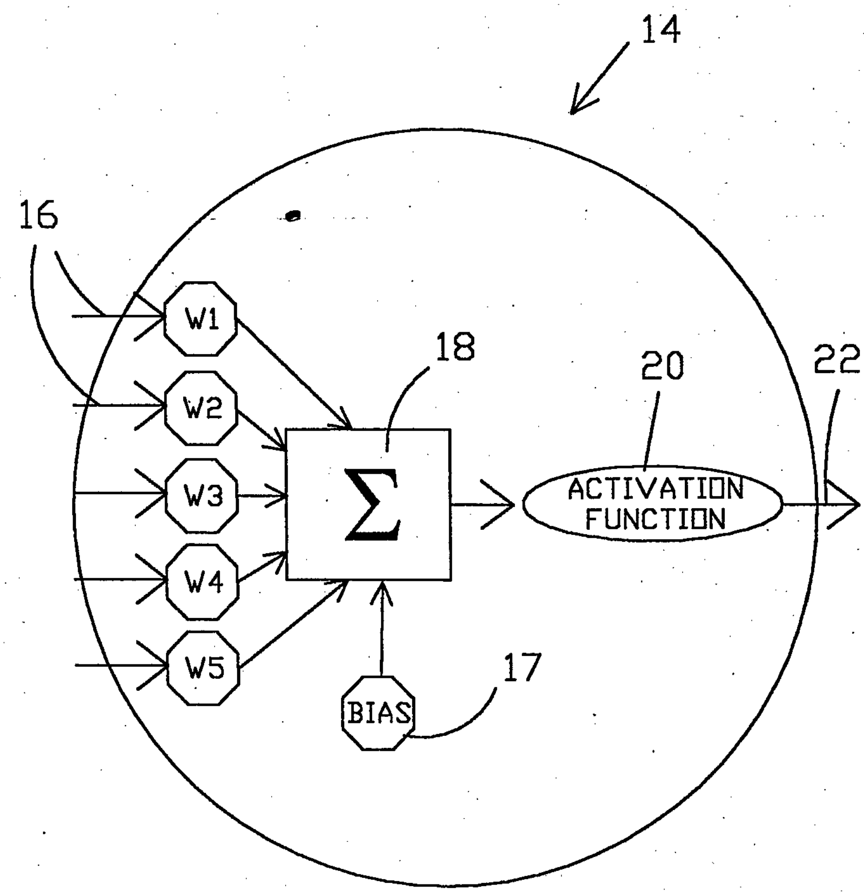
14
16
W1
W2
W3
W4
W5
18
Σ
20
ACTIVATION FUNCTION
22
BIAS
17

FIG. 1B

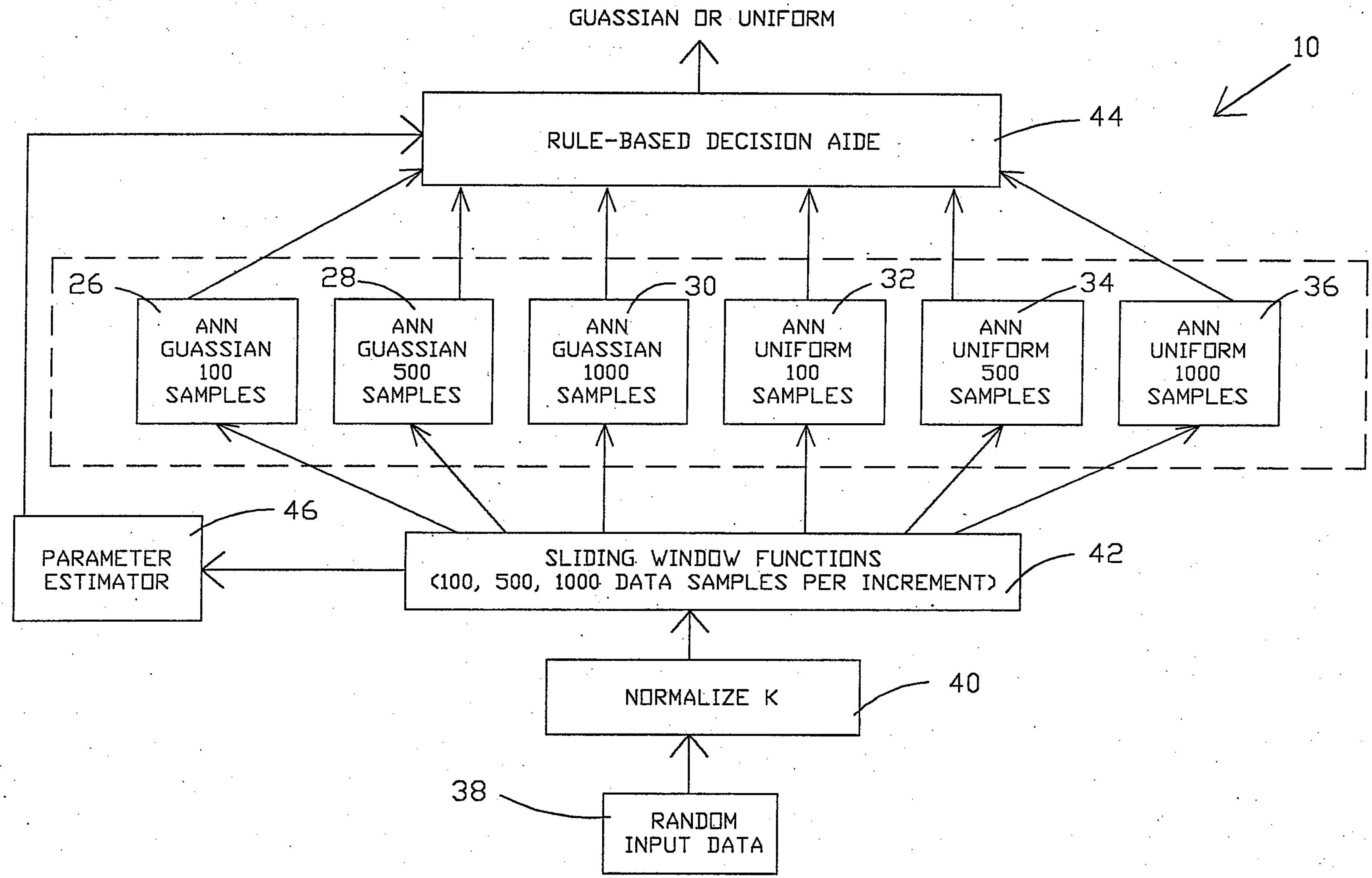
GUASSIAN OR UNIFORM
10
RULE-BASED DECISION AIDE
44
26
28
30
32
34
36
ANN GUASSIAN 100 SAMPLES
ANN GUASSIAN 500 SAMPLES
ANN GUASSIAN 1000 SAMPLES
ANN UNIFORM 100 SAMPLES
ANN UNIFORM 500 SAMPLES
ANN UNIFORM 1000 SAMPLES
46
PARAMETER ESTIMATOR
SLIDING WINDOW FUNCTIONS
(100, 500, 1000 DATA SAMPLES PER INCREMENT)
42
NORMALIZE K
40
38
RANDOM INPUT DATA

FIG. 2

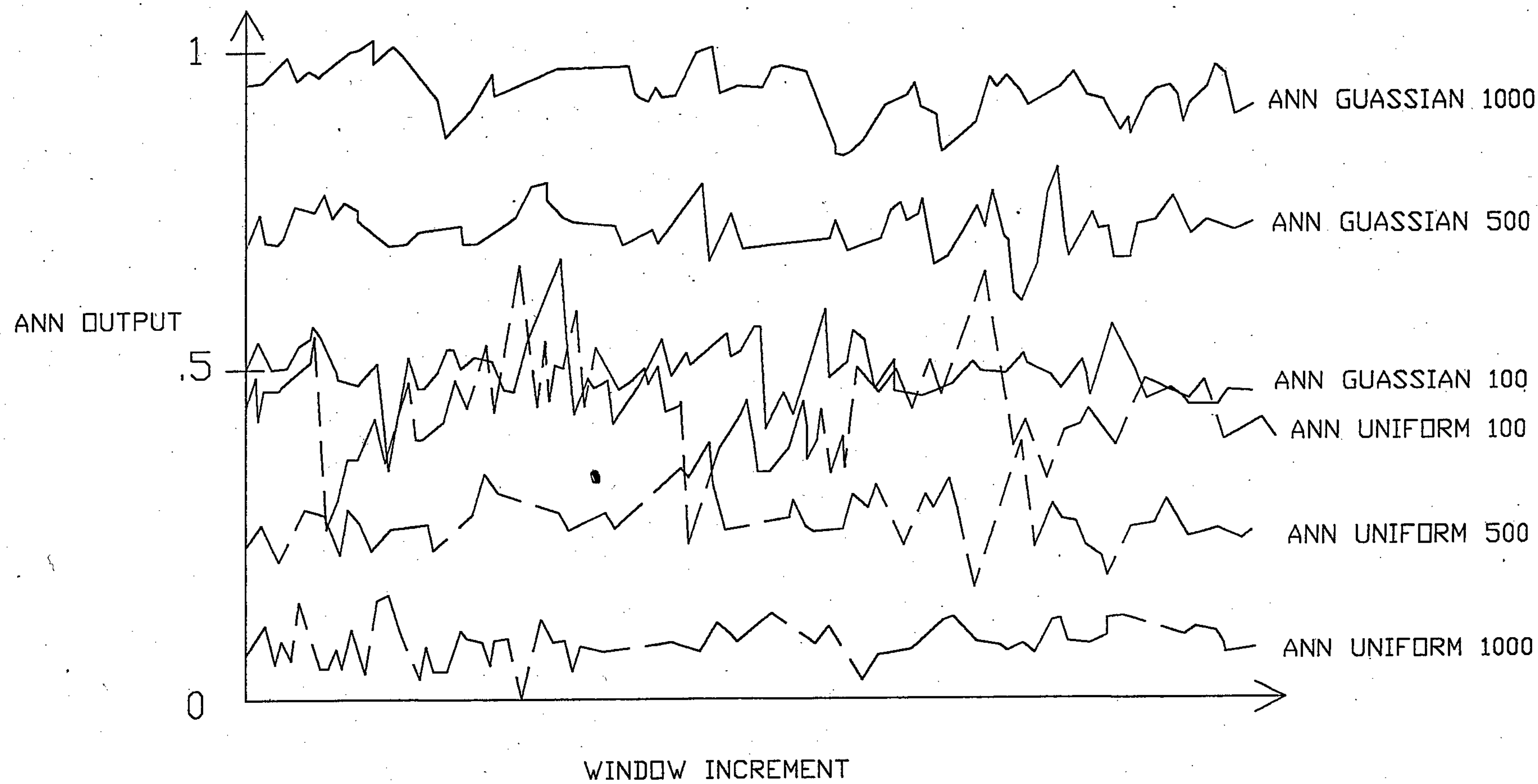
1
.5
0
ANN OUTPUT
WINDOW INCREMENT
ANN GUASSIAN 1000
ANN GUASSIAN 500
ANN GUASSIAN 100
ANN UNIFORM 100
ANN UNIFORM 500
ANN UNIFORM 1000

FIG. 3

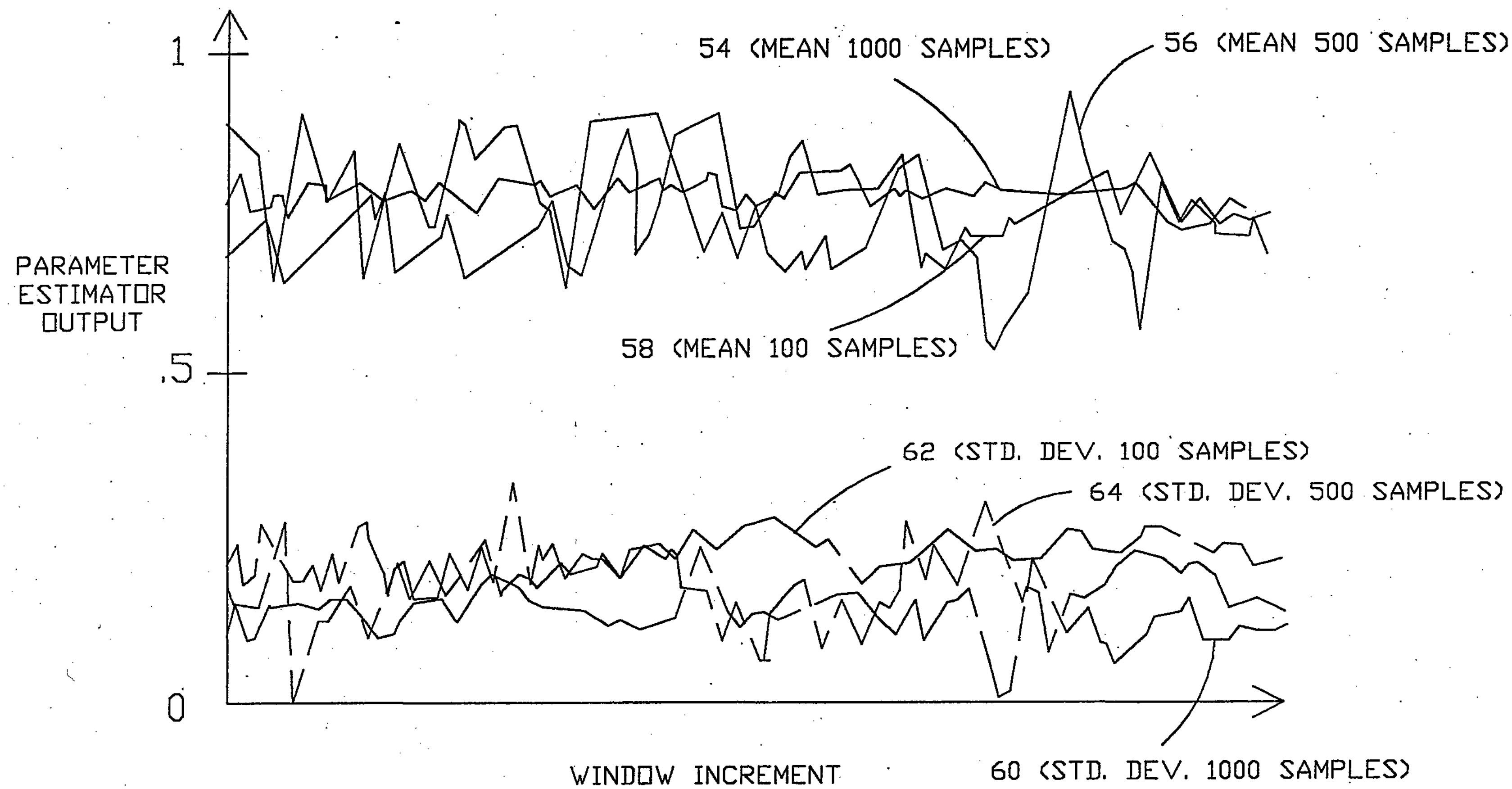
1
.5
0
PARAMETER ESTIMATOR OUTPUT
54 (MEAN 1000 SAMPLES)
56 (MEAN 500 SAMPLES)
58 (MEAN 100 SAMPLES)
62 (STD. DEV. 100 SAMPLES)
64 (STD. DEV. 500 SAMPLES)
60 (STD. DEV. 1000 SAMPLES)
WINDOW INCREMENT

FIG. 4

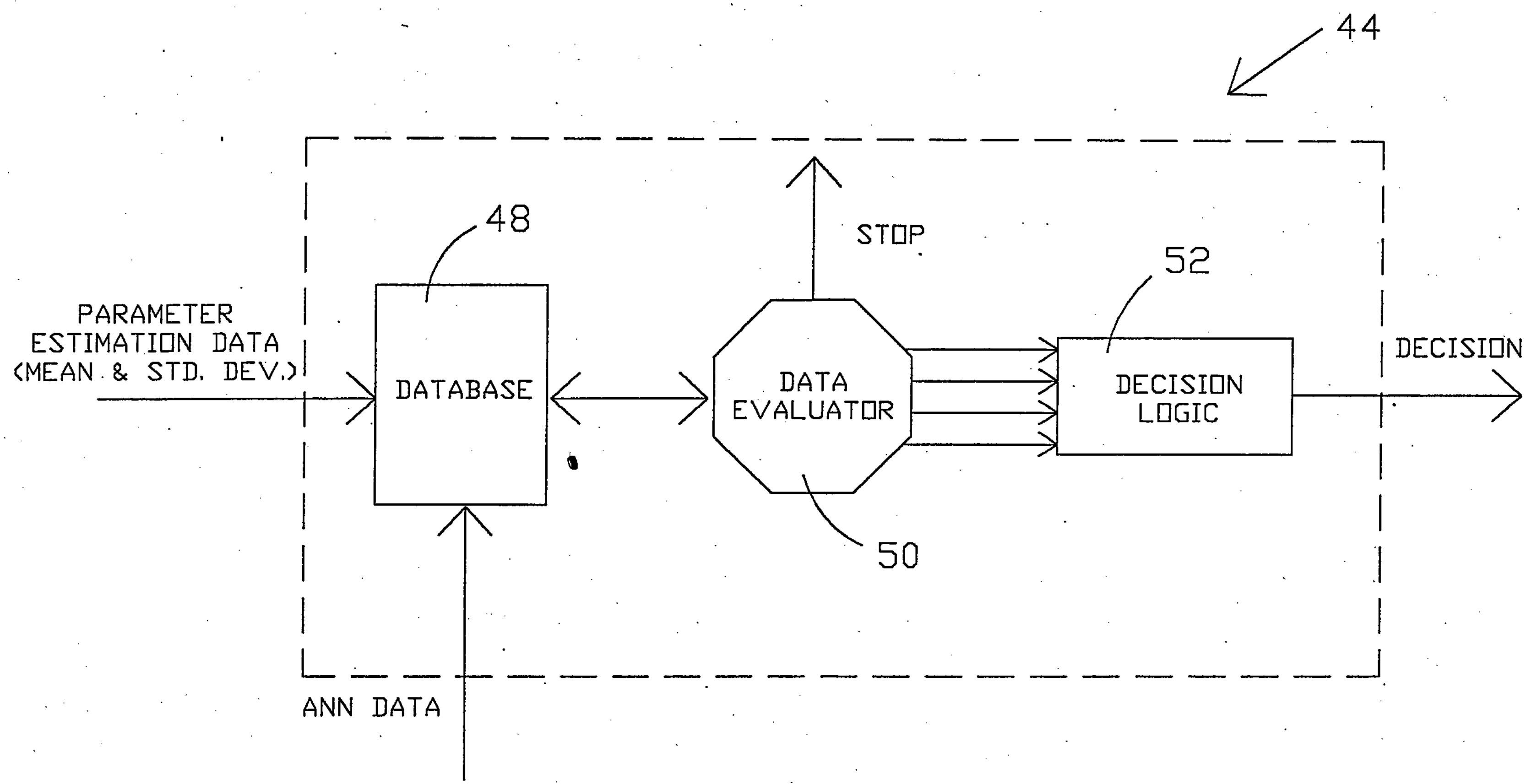
44
48
PARAMETER
ESTIMATION DATA
(MEAN & STD. DEV.)
DATABASE
ANN DATA
STOP
DATA
EVALUATOR
50
52
DECISION
LOGIC
DECISION

FIG. 5