Self-organising Maps Use for Intelligent Data Analysis

Doug Myers, Kok Wai Wong** and Chun Che Fung**

** School of Electrical and Computer Engineering*
Curtin University of Technology
Bentley, Western Australia 6102
Phone (618) 9266 7912
Fax (618) 9266 2584
Email: rmyersdg@cc.curtin.edu.au

*** School of Information Technology*
Murdoch University
Murdoch, Western Australia 6150
Phone: (618) 9360 2918
Fax: (618) 9360 2941
Email:kwong@murdoch.edu.au

Abstract: *A neural-network-based data-analysis model for the prediction and classification of field data has many attractions. However, there are problems in ensuring the generalisation capability of the data analysis model, in measuring the similarity between the original training data and the new unknown data, and in processing large data volumes. This paper proposes the use of self-organising maps (SOMs) to overcome these difficulties and illustrates the utility of the approach though applications in the agricultural, resource exploration and mineral processing areas. In most SOM applications, its self-organising and clustering capabilities have always been the focus. In this paper, SOM is used as enhancement approach that can be incorporated within another intelligent data analysis approach.*

1. INTRODUCTION

Self-Organising Maps (SOMs) [8,10] have been recognised as an important tool for information processing and data analysis. Most SOM applications focus on their self-organising and clustering capabilities as SOM has the ability to organise the input vectors in an unsupervised learning mode.

Intelligent data analysis, in the form of neural networks, is critical to an increasing number of application areas, but there remain problems to resolve. Many of these would not exist if the information to be analysed could be clustered before processing proper. It is suggested that SOMs offer that capability and so allow inferential data analysis procedures to be enhanced.

Data analysis is usually performed on a sample set of observations taken from some population [11]. In most practical situations, a sample is all that is available and it may provide incomplete information on the population. The objective of data analysis is nonetheless to extract maximum information from that sample, to exhibit reasonable interpolation skill and provide some indication where that has been used for extrapolation purposes.

In this paper, SOM is used to examine three aspects of this data analysis problem. First, the problem of ensuring the generalisation capability of the data analysis model is investigated. SOM data splitting validation is proposed to solve this. After the interpretation model is established, SOM is then used to provide measurement of the similarity between the training data and the new unknown data. However, in cases where the available data is large, it is always safer to assume that the underlying function the interpretation model needs to learn is difficult to realise. SOM can then be used in establishing modular models for overcoming this problem.

The proposed SOM solutions to the intelligent data analysis problems have each been successfully applied to problems relating to local industrial problems in Western Australia. The proposed approach has been applied in the area of agriculture, resource exploration and mineral processing activities. In particular, the SOM approach is used in classifying Australian wheat varieties [1, 12], to aid a Backpropagation Neural Network (BPNN) in providing better and more accurate well log analysis [14] and in assisting a BPNN in providing reliable hydrocyclone data analysis [2].

2. SELF-ORGANISING MAPS

SOM mimics aspects of brain behaviour and has a close relationship to brain maps [8, 10]. Its main feature is the ability to visualise high dimensional input spaces onto a smaller dimensional display, usually two-dimensional. For the discussion and applications in this paper, only a two-dimensional array is of interest.

Consider some input data space $\Re^n$ to be mapped by the SOM onto a two-dimensional array with I nodes. For each of the I nodes, there is an associated parametric reference vector $m_I=[\mu_{I1},\mu_{I2},\ldots,\mu_{In}]^T \in \Re^n$, where μ_{ij} is the connection weight between node I and input j. The input data space $\Re^n$ consists of input vectors $X=[x_1,x_2,..,x_n]^T$. Then any $X \in \Re^n$ can be visualised as being connected to all nodes in parallel via scalar weights μ_{ij}. The aim of learning is to map

all n input vectors X_n onto m_I by adjusting weights μ_{ij} such that the SOM gives the best match response locations.

SOM can also be said to be a nonlinear projection of the probability density function $p(X)$ of the high dimensional input vector space onto the (two-dimensional) display map. Normally, to find the best matching node I, a given input vector X is compared to all reference vectors m_I by searching the smallest Euclidean distances $\| X - m_I \|$, signified by some parameter c. Therefore,

$$c = \arg\min_i \{\| X - m_i \|\} \quad (1)$$

or $$\| X - m_c \| \cong \min_i \{\| X - m_i \|\} \quad (2)$$

During the learning process, the node that best matches the input vector X is allowed to learn, but in addition those nodes close to that node within a given Euclidean distance are also allowed to learn. The learning process is expressed as:

$$m_i(t+1) = m_i(t) + h_{ci}(t)[X(t) - m_i(t)] \quad (3)$$

where t is discrete time coordinate
and $h_{ci}(t)$ is the neighbourhood function

After the learning process has converged, the map will display the probability density function $p(X)$ that best describes the entire input vector space. On completion, an average quantisation error of this map can be generated to indicate how well the map has matched the entire input vector set X_n. The average quantisation error is defined as:

$$E = \int \| X - m_c \|^2 \, p(X) dX \quad (4)$$

Beside the average quantisation error, an individual quantisation error may also be used to measure how well any input vector matches the closest node I,. This is similar to equation (2).

3. SOM DATA SPLITTING

Split-sample validation is the most commonly used method for estimating the generalisation capability of a BPNN using the early-stopping approach [17]. Here, a set of validation data that is not used in the training process is used to calculate the validation error. The stopping point in this method is suggested to be the point where the validation error starts to rise. This point also indicates that the generalisation ability starts to degrade. When training starts, the errors for both data sets will normally reduce. After much training iteration, the validation error normally starts to rise although the training error may continue to fall. The BPNN training process can be stopped at this point as further training will result in overfitting. A typical plot of the training and validation errors is shown in Figure 1.

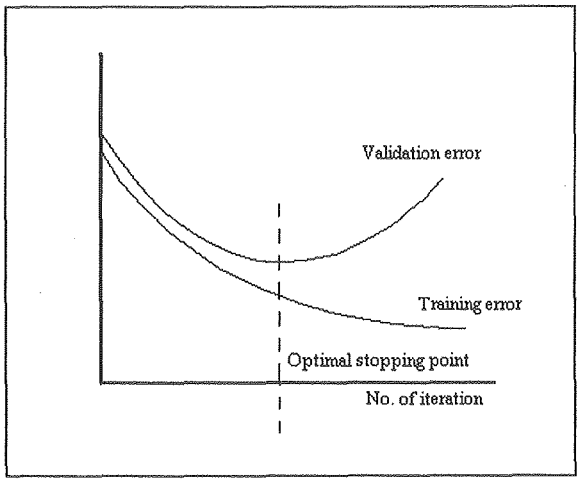

Figure 1: Typical plot of the training and validation errors

As the generalisation ability of the BPNN is highly dependent on the validation data set, the splitting method used is important. However, there are no rules to suggest the best method. Nevertheless, the validation data set should demonstrate two characteristics: (1) the validation set should be statistically close to the training set, and (2) the validation error should indicate the generalisation ability of the final BPNN and it should be possible to use it as the stopping criteria for the training process.

If U is the universal sample space of all the cases of data to be processed by the network, then the training set, TR should be contained in or equal to set U:

$$TR \subseteq U \quad (5)$$

If TR follows the condition in equation (5), the validation set, VA, and testing set, TE, should a proper subset to the training set. That is:

$$VA \subset TR \quad (6)$$

$$TE \subset TR \quad (7)$$

with the condition:

$$VA \cap TE = \varnothing$$

However, if the traditional random approach of data splitting is used, this may result in a worst case situation as illustrated by the following equations.

$$TR \subset U \quad (8)$$

$$VA \subset U \quad (9)$$

$$TE \subset U \quad (10)$$

with the following conditions:

$TR \cap VA \cap TE = \varnothing$,
$VA \not\subset TR$, and
$TE \not\subset TR$.

In this case, the statistical characteristics of the three data sets are all mutually exclusive, the training set does not cover all the sample space, and the validation and testing sets will not be able to give a fair indication of the generalisation ability of the network.

In the proposed SOM data-splitting technique [6, 18], the available data are first classified into different clusters using unsupervised learning. If U is classified into C_1 to C_n clusters, then U can be written as:

$$U = \{ C_1, C_2, C_3, ..., C_n \} \quad (11)$$

If the training data set is selected from each one of the n clusters and the rest are left for testing and validation, then the conditions on equation (6) and (7) are satisfied. In this case, the training set will cover all the desired underlying cases. The validation set and testing set are subsets from the clusters from which the training set is selected.

From the above, an important and crucial task is splitting the available data into training and validation sets. The training set will give information on what the BPNN should learn and the validation set acts as a teacher to guide the BPNN such that it will learn the correct function. As the BPNN is based on a training set to obtain the underlying knowledge, therefore it should contain more data than the validation set.

When obtaining the training set, there will be some environmental factors that affect the measurements. As a result, it is not possible to have an exact function that describes the relationship between X and Y. However, a probabilistic relationship governed by a joint probability law $P(\nu)$ can be used to describe the relative frequency of occurrence of vector pair (X_n, Y_n) for an n training set. The joint probability function $P(\nu)$ can be further separated into an environmental probability function $P(\mu)$ and a conditional probability function $P(\gamma)$. Thus the probability function may be expressed as:

$$P(\nu) = P(\mu)P(\gamma) \quad (12)$$

The environmental probability function $P(\mu)$ describes the occurrence of the input X. The conditional probability function $P(\gamma)$ describes the occurrence of the output Y based on the given input X.

A vector pair (X, Y) is considered as noise if X does not follow the environmental probability function $P(\mu)$, or the output Y based on the given X does not follow the conditional probability function $P(\gamma)$.

The rule for splitting the available data into training and validation sets is that the training set should be statistically similar to the whole sample space. The validation set should also be statistically similar to the training set as it has to act as a teacher. This rule suggests deploying the SOM algorithm of the last section. SOM can be used as a nonlinear probability density function projection on the two-dimensional map. Therefore, in each node I the probability density function of the input vectors being mapped onto it should have a similar probability density function. This also implies that the input vectors that are mapped onto the same node should have similar relative occurrences as denoted by $P(X)$. This $P(X)$ is similar to the environmental probability function $P(\mu)$ in equation (12). From the analysis of equation 12, the role of training the BPNN can be said to be a search for the conditional probability law $P(\gamma)$. The formulation of the $P(X)$ here has to be extended. Instead of mapping just the input vector X, both the input vector X and target vector Y are used in the learning of the SOM. A joint probability between X and Y is assumed and is denoted as $P(X,Y)$. It can be further expressed as:

$$P(X,Y) = P(X \mid Y)P(Y) = P(Y \mid X)P(X) \quad (13)$$

As equation (13) is similar to equation (12), it implies that the joint probability function density of a SOM is directly related to the joint probability function. With this, it can also be realised that the joint vectors of X and Y falling in the same node should have very similar statistical characteristics.

The methodology for satisfying the splitting data rule has been formulated. The n available data sets that consist of X input vectors and Y output vectors are first used to train the SOM. After the map has been trained and individual quantisation errors have been generated, selection can be made. A data set is selected as validation data if it has a small quantisation error as compared to the other data sets in the same node. This will ensure that the validation set is a sub-set of the training set. However, for cases where there is only one data set in that node, it will be left in the training set. This is to ensure that the training set can cover the whole sample space of the available data and to ensure that the training set is always larger than the validation set. After all the available data has been split into training and validation sets, the BPNN can start to learn and the process is stopped by using the early stopping validation technique.

4. STATISTICAL COMPARISON

The issue of evaluating an indication of the confidence level for the predicted properties in unknown cases is considered. The objective is to provide an indication of the usability of the trained interpretation model when it is used for any new data that may be statistically different from the training data. In cases where the indication shows that the unknown new cases are very different from the trained cases, the predicted results cannot be totally trusted. This will be useful in providing a confidence indication to the analyst.

To perform the confidence level indication, a SOM is used to classify the training data to a pre-defined two-dimensional map. At the completion of this unsupervised learning stage, an average quantisation error is generated that gives a measure of the fitness of the training data in the resultant clusters. Any subsequent unknown input data to be applied to the prediction model are now mapped onto the trained SOM. An average quantisation error is generated that measures the statistical similarity between the unknown data set and the trained map. Comparing the average quantisation errors of the training data set and the unknown data set indicates the similarity. These values suggest to the users how similar or different are the trained and predicted data sets. It provides the user an assurance of the predicted output from the BPNN interpretation model.

5. SOM FOR MODULAR NEURAL NETWORKS

When there is a large volume of available training data, the Modular Neural Network (MNN) is proposed for analysis. The MNN is based on the Self-organising Map (SOM) [8, 10], Learning Vector Quantisation (LVQ) [9] and BPNN [15]. However, an MNN [7] can only be used when the available training data is large. As compared to the usual BPNN approach with its single network, the MNN employs a number of sub-networks. SOM and LVQ are used to classify the raw data. Several BPNNs corresponding to the number of classes obtained from the SOM are then trained for the purpose of prediction. Since the number of data to be handled by each sub-network is relatively small, the training time is significantly shortened. As the data that falls into the same sub-network will have similar characteristics, this effectively reduces the complexity of the function that the ANN needs to learn. Figure 2 shows the block diagram of the MNN.

An MNN is arranged into two major sections. The first focuses on classification. The second covers the prediction results of the MNN.

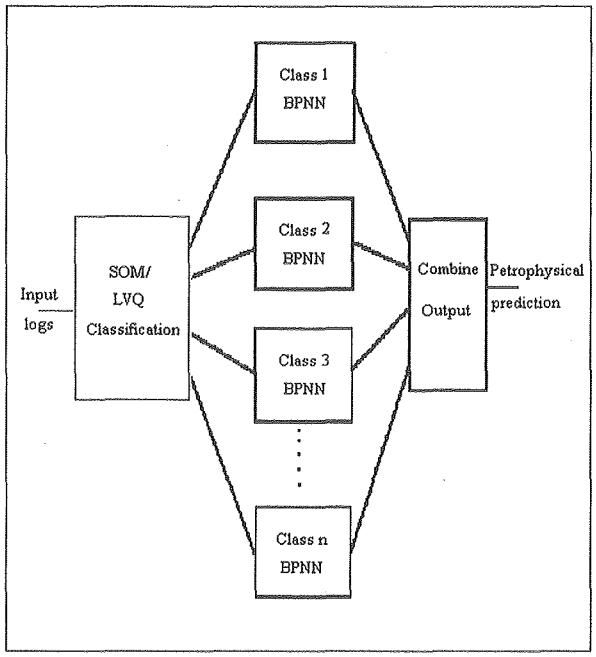

Figure 2: Block diagram of Modular Neural Network.

An ANN is capable of learning any non-linear function from the available training data. However, if the available training data is large and complex like that of Figure 3, the underlying function may be too complex for a single ANN to cope with. This may be overcome by modularising the task as shown in Figure 4. If the data can be first classified before the ANN learning process, then the functions handled by each sub-section of a modular structure will be very much simpler compared to the whole training data set. Consequently, the function should be able to learn in a shorter time and better prediction results obtained.

There are several ways of performing classification. However, a technique that can be done automatically and transparently to a human analyst is most desirable. SOM is selected as the best classification approach in designing this MNN as it uses unsupervised learning. It has the ability to learn and organise information without being given correct outputs for its inputs. A SOM network consists of two layers of nodes. Each output node is computed with the dot product of its weight vector and the input vector. The result will reflect the similarity between the two vectors. At the end of the training, the SOM will make use of its learning ability to arrange the available training data into a different cluster. After the SOM classification of the training data, supervised learning in the form of LVQ is employed to fine-tune the classification process such that it could be used for any unknown input data. LVQ is closely related to SOM, but uses the given classification information to define the class regions

in the input space. In this case, SOM and LVQ will learn from the data and perform their own classification process. This removes the need for any human intervention in sorting data.

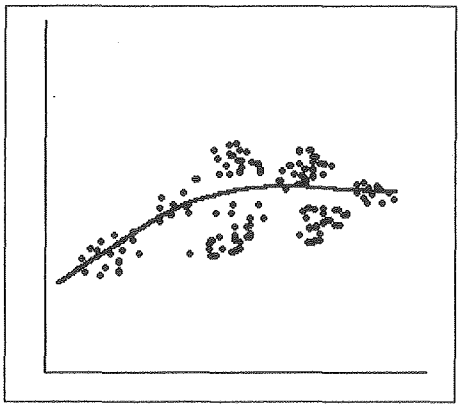

Figure 3: Function handle by one BPNN

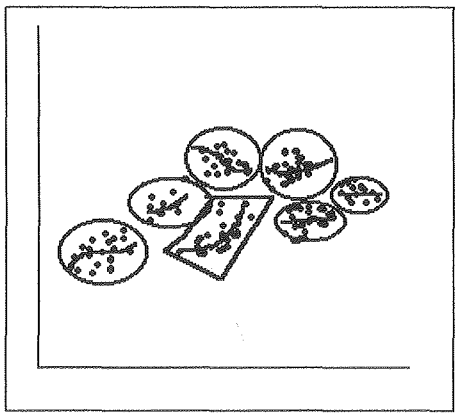

Figure 4: Functions handle by MNN

Since it is known that a relationship exists between the input vectors and the characteristics within the output data, an approach to determine that has been formulated. SOM is first applied to classify the input and output data. The classes obtained are then used to label the input vectors. The input vectors coupled with the output class labels are then applied to the LVQ algorithm. A number of BPNN networks corresponding to the number of classes obtained from the SOM are trained. After the classification process, the data fed into the different BPNN has similar characteristics. In this way, training of the BPNN is expected to take shorter time.

The process is summarised in the following steps:

Step 1: Normalise the input and output data.
Step 2: Determine the number of classes required and apply the SOM algorithm to the input and output vectors.
Step 3: Label the input vectors according to output classifications from Step 2.
Step 4: Apply the LVQ algorithm to the normalised inputs and establish the network.
Step 5: Prepare to train a few BPNNs, each corresponding to each class from step 2.
Step 6: Train each BPNN using the SOM data splitting validation approach mentioned in the previous section.

Once the network is trained, new input data can be classified by applying the normalised data to the network.

6. APPLICATIONS

The following application examples are used to illustrate the usefulness of the proposed SOM data analysis approach. The authors have investigated these problems over the past few years. SOM has enhanced the performance of the data analysis model and improved the results in these cases. Further details and descriptions of the problems can be found in the referred papers.

6.1 In Agriculture

For a variety of reasons, Australian wheat varieties derive from four base varieties introduced in the 19th century. Cross breeding programs have lead to about 180 varieties currently being registered, but due to these origins they are genetically very similar. While the plants themselves can be physically quite different, the grains are almost identical. Even for experienced agronomists it is impossible to tell the variety of wheat from the kernel alone.

There are circumstances, though, where it is important to be able to determine variety from a kernel. The most pressing is in quality control.

Once wheats were sold in broad classifications such as bread, biscuit, industrial and feed. However, the trend in world markets is for customers to issue quite detailed specifications on what they require and seek a supplier who will meet those. In particular, each type of pasta and noodle tends to have its own specification.

With increasing affluence, the demand for noodles is growing rapidly in Asia. This is a market of considerable interest to Australian grain growers, in part because many Australian wheats are very well suited to noodle production. No one variety of wheat, though, can meet a noodle specification, but an admixture of several can. Thus the marketing practice is to call in the quantities of each variety needed to meet the specification of an order, mix and then ship them. This raises a quality control problem, namely how to ensure that the grain supplied is of the correct variety and that the admixture is correctly formed.
There are two solutions to this problem, each with some drawbacks [1, 5, 12]. One is to use chemical

analysis. That is relatively slow and expensive, but it is precise and can easily test bulk volumes. The other is to employ an automated system that uses statistical pattern recognition techniques to identify variety through shape analysis.

To undertake such automated analysis, the grain kernels need to be presented individually to a camera in some common orientation. In the case of wheat, this is usually with the crease down. The kernel is imaged and a computer processes the image to extract the contour. Then various shape parameters are derived from it and used to arrive at a decision on the variety or some other aspect of the grain.

Automated methods of grain recognition have been investigated in most of the world's leading agricultural nations. They have been applied not only to wheat, but also to rice, corn and barley, plus some oilseeds. Except in Australia, though, the visual differences between kernels of different varieties are quite apparent. Thus application of these automated systems is a convenience. Such systems have the advantage they may be deployed anywhere, but as they treat only one kernel at a time, a statistically uncertain result may occur if time is limited.

The similarity of grain physical characteristics makes automated recognition of Australian wheats very difficult. SOM has been used to perform the classification task [5]. The input vectors are sample sets of shape features. Given a shape contour, moments may be computed and the principal and minor axes plus the centroid determined. These form a reference set for the shape measurements. A variety of shape features can be extracted, but experience has shown rays emanating from the centroid at regular angles, or aspect ratio measurements – the ratio of the width of the grain at set points along the principal axis to the length of that axis – adequately capture the information required.

After the network has been trained with a selective set of data, the output node which gives the highest response to a specific class of wheat variety within the input training data sets is labelled to that class of wheat. When an output node gives the same response to two or more wheat varieties, its neighbouring nodes are taken into account and the majority rule is applied to determine the labelling for the node.

The test outcomes indicate that SOM is able to classify up to two or three wheat varieties with a maximum accuracy of 96.5% and 88% respectively when the task is identifying variety in a group of four. However, when the test set increases to six, problems are encountered and there may be no classification. If the set exceeds about 10, then accuracy falls to around 40% [16] SOM's do not perform as well for this task as some other methods [13]. In spite of that, they have some attractions. When they converge they do so quickly, they are a learning network and they are easy to implement. This would suggest that a combination of SOM and other methods would be of some benefit in this application. This has yet to be investigated.

6.2 In Resource Exploration

Developing a petroleum reservoir demands a huge capital investment. The exploration process therefore has to be managed and controlled carefully. The initial phase normally involves a number of boreholes being drilled at different locations around the region believed to hold the reservoir. Well logging instruments are then lowered into each borehole to collect data typically at every 150mm or so of depth. These data are known in the industry as *well log data*. The next stage involves an intense process of analysing the available well log data in order to evaluate the reservoir's potential.

In order to obtain an accurate picture of the physical characteristics of the well, actual rock samples from various depths are retrieved using a coring barrel. These samples are then sent to a laboratory and they are examined using various physical and chemical processes. Data obtained from this phase are known as *core data* in the analysis process.

Two key issues in the reservoir evaluation of petroleum exploration using well log data are the characterisation of formation and the prediction of petrophysical properties such as porosity, permeability and volume of clay [14]. While a set of core data gives an accurate picture of the petrophysical properties at specific depths, it is a lengthy process and great expense is incurred in obtaining it. Hence only limited core data are available at selected wells and depths. The objective of well log data analysis is to therefore establish an accurate interpretation model for the prediction of the petrophysical properties for uncored depths and boreholes around that region. An accurate prediction is essential to the ultimate determination of the economic viability of the exploration and the production capacity of the particular well or region.

However, the establishment of an accurate well log interpretation model is not an easy task due to the complexity of different factors that influence the log responses. This demands a high level of human expertise, experience and knowledge.

A large number of techniques have been introduced over the past 50 years with an intention to establish an adequate interpretation model. The way that well log interpretation is carried out has also changed due to developmenst in logging tools. The analysis process has also undergone substantial changes due to the

development and understanding of the physics of porous media and the rapid development of computer technology. Nevertheless, the derivation of a well log interpretation model normally falls into one of two main approaches: empirical and statistical.

For the empirical approach, the unique geophysical characteristic of each region prevents a single formula from being universally applicable. In addition, as the number of parameters that the mathematical functions can handle is limited, it is also difficult to establish an accurate model. Statistical techniques lack universal capabilities and their successful application is an inverse function of the problem complexity. When the problem becomes too complex, the assumptions are more difficult to estimate correctly. Statistical techniques also limit the number of well log data that can be handled at the same time. With the increasing number of instruments and log data, it becomes difficult to apply the traditional statistical and graphical methods.

BPNN has been the emerging technology in this field. A BPNN is suited to this application as it resembles the characteristics of regression analysis in statistical approaches. However, it performs analysis in a fundamentally different way from the traditional empirical and statistical approaches. BPNNs can be used to address most of the mentioned factors that could possibly affect the accuracy of the model. A BPNN does not require a prior assumption of the functional form of the dependency. It also offers a numerical model free of estimators and dynamic systems. In addition, BPNNs are able to model complex nonlinear processes with acceptable accuracy and have the ability to reject noise.

The raw application of a BPNN may not provide reliable well log analysis. The three problems raised in the beginning of this paper are the major concerns for the application of BPNN techniques in this field. However, with the application of SOMs for data analysis in the manner outlined, these concerns can be eliminated. All the three proposed SOM approaches have been incorporated into the BPNN data analysis model, and research results indicate this leads to an increase in the reliability of the prediction [7, 3].

6.3 In Mineral Operations

Hydrocyclones find extensive application in the mineral process industry where they are used for the classification and separations of solids suspended in fluids [2]. They are manufactured in different shapes and sizes to suit specific purposes. Hydrocyclones normally have no moving parts. The feed slurry containing all sizes of particles enters the hydrocyclone. Inside, due to centrifugal force experienced by the slurry, the heavier particles will be separated from the lighter.

After the particles suspended in the fluid are classified, they are discharged either from the vortex finder as overflow or from the spigot opening as underflow. Due to the complexity of the separation mechanism in the hydrocyclone, the interpretation of the physical behaviour and forces acting on the particles is not clear.

The performance of a hydrocyclone is normally described by a parameter known as d50. This parameter determines the classification efficiency. It represents a particular particle size reporting 50% to the overflow and 50% to the underflow streams. The separation efficiency of hydrocyclones depends on the dimensions of the hydrocyclone and the operational parameters. Examples of the operational parameters are flowrates and densities of slurries. D50 is not a monitored parameter, but determined from separation curves known as tromp curves. They are used to provide the relationship between the weight fraction of each particle size in the overflow and underflow streams.

In practical applications, the d50 curve is corrected by assuming that a fraction of the heavier particles is entering the overflow stream. This is equivalent to the fraction of water in the underflow. This correction of d50 is designated as d50c. The correct estimation of d50c is important since it is directly related to the efficiency of operations. Under normal industrial applications of hydrocyclones, any deviation from a desired d50c value cannot be restored without changing the operation conditions or/and the geometry of the hydrocyclone. Also, sensing the changes in d50c is a difficult task. It requires external interference by taking appropriate samples from the overflow and underflow streams. At the same time, lengthy size distribution analyses of these samples needs to be conducted.

While hydrocyclones are used in the mineral processing industry for particle separation, an exact model of a hydrocyclone is difficult to derive due to their highly non-linear characteristics and the large number of parameters involved. As efficient operation of a hydrocyclone is important in improving system performance, it is essential that the model should be able to provide non-linear matching between the multi-dimensional system inputs and outputs. Although the collection of the data in this field is different compared to well log data analysis, they both fall into the same category of inferential data analysis problem. Therefore, the methodology used in the previous section can be duplicated and used in this field. Research results show that SOM methods can also increase the prediction reliability [4].

7. CONCLUSIONS

This use of SOMs offers advantages in framing the interpretation model in intelligent data analysis. SOM-based intelligent data analysis approach in three significant applications areas verifies the value of the method. This paper has shown that SOM can assist in ensuring the generalisation ability of the BPNN by splitting the available data. SOM has also shown that it can give some kind of indication to the analyst on how similar is the testing data as compared to the training data. When the available data is large, the MNN that make use of SOM could generate a few BPNNs with each handling a small portion of the data. Together with other data-analysis tools, SOM has shown to be a useful approach to improve the performance of the data-analysis process.

8. REFERENCES

[1] Barker, D.A., Vuori, T.A., and Myers, D.G., "The Use of Ray Parameters for the Discrimination of Australian Wheat Varieties," Plant Varieties and Seeds, No. 5, pp. 35-45, 1992.

[2] Bradley, D., The Hydrocyclone, Pergamon Press Ltd, 1965.

[3] Crocker, H., Fung, C.C., Wong, K.W., "The STAG Oilfield Formation Evaluation: A Neural Network Approach", APPEA'99 Journal, pp. 451-460, 1999.

[4] Eren, H., Fung, C.C., Wong, K.W., and Gupta, A., "Artificial Neural Networks in Estimation of Hydrocyclone Parameter d50c with Unusual Input Variables," IEEE Transactions on Instrumentation & Measurement, 46(4), pp. 908 - 912, 1997.

[5] Fung, C.C., Truong, M.S. and Myers, D.G., "An Unsupervised Learning Algorithm for the Classification and Determination of Proportion Mix of Wheat Varieties", Proceedings ANZIIS-93, pp. 462-466, 1993.

[6] Fung, C.C., Wong, K.W., and Eren, H., "Determination of a Generalised BPNN using SOM Data-splitting and Early Stopping Validation Approach," Proceedings of Eighth Australian Conference on Neural Network (ACNN'97), pp. 129 – 133, 1997.

[7] Fung, C.C., Wong, K.W., Eren, H., Charlebois, R. and Crocker, H., "Modular Artificial Neural Network for Prediction of Petrophysical Properties from Well Log Data," IEEE Transactions on Instrumentation & Measurement, 46(6), pp. 1259-1263, 1997.

[8] Kohonen, T., "The Self-Organising Map", Proceedings of the IEEE, Vol. 78, No. 9, pp. 1464-1480, 1990.

[9] Kohonen, T., Kangas, J., Laaksonen J. and Torkkola K., "LVQ_PAK: A program package for the correct application of Learning Vector Quantization algorithms", Proceedings of the International Joint Conference on Neural Networks, pp. I-725-730, 1992.

[10] Kohonen, T., Self-Organising Map, Springer-Verlag, 1995.

[11] Mendenhall, W., and Sincich, T., Statistics for Engineering and the Sciences, 3rd Edition, Dellen Publishing Company, 1992.

[12] Myers, D.G. and Edsall, K.J., "The Application of Image Processing Techniques to the Identification of Australian Wheat Varieties," Plant Varieties and Seeds, No. 2, pp. 109-116, 1989.

[13] Myers, D.G., and Vuori, T., "A Comparison of Classification Techniques for the Identification of Australian Wheat Varieties," Proceedings of the SPIE, v2345, pp 104-109, 1994.

[14] Rider, M., The Geological Interpretation of Well Logs, Second Edition, Whittles Publishing, 1996.

[15] Rumelhart, D.E., Hinton, G.E. and Williams, R.J., "Learning Internal Representation by Error Propagation" Parallel Distributed Processing, Vol. 1, Cambridge MA: MIT Press, 1986, pp. 318-362.

[16] Vuori, T., "A Pattern Recognition approach to the Identification of Australian Wheat Varieties," Ph.D Thesis, Curtin University of Technology, 1995.

[17] Weiss, S.M., and Kulikowski, C.A., Computer Systems That Learn, Morgan Kaufmann, 1991.

[18] Wong, K.W., Fung, C.C., and Eren, H., "A Study of the Use of Self-Organising Map for Splitting Training and Validation Sets for Backpropagation Neural Network," Proceedings of IEEE Region Ten Conference (TENCON) - Digital Signal Processing Applications, pp. 157 - 162, 1996.