Engng Applic. Artif. Intell. Vol. 6, No. 6, pp. 565–573, 1993
Printed in Great Britain

0952-1976/93 $6.00 + 0.00
Pergamon Press Ltd

Contributed Paper

Using Unsupervised Learning for Feature Detection in a Coal Mine Roof

ROGER L. KING
Mississippi State University

MICHAEL A. HICKS
Mississippi State University

STEPHEN P. SIGNER
U.S. Bureau of Mines, Spokane

(Received April 1993)

The use of an unsupervised learning technique for classifying geological features in the roof overlying an underground coal mine is described. The technique uses torque, thrust, drill speed, penetration rate, and drill position data from a roof bolter as inputs for the classification. Data were obtained from an underground coal mine in the western United States and initially classified using clustering. Some of the available approaches for clustering are reviewed and the rationale used in selecting the chosen approach is discussed. The cluster centers, or exemplars, obtained from this approach can be used to train two supervised neural networks involving the back-propagation of error learning algorithm.

Keywords: Clusters, feature detection, artificial neural networks, underground mining, unsupervised learning.

1. BACKGROUND

The simplest classification of a mine roof is one in which the roof falls into one of two different classes or states—safe or unsafe. From this elementary classification, one can conclude that if the mine roof is safe, then the roof support plan can be implemented as prescribed. However, if the mine roof is unsafe, then the plan must be modified immediately. The deficiency in this simplistic approach lies in its inability to discern how "unsafe" the mine roof is. A metric for "unsafe" is necessary before a cost-effective supplemental support scheme can be recommended.

A natural extension of the two-state classifier is one that would indicate not only that the roof is either safe or unsafe, but also why the roof is either degrading or improving. To accomplish this, it would be necessary to know something about the structure of the mine roof. Once the structure is determined, then it is just a matter of determining the best supplemental roof support plan for that structure. Depending upon the classification, the system could recommend anything from additional posts and cribs to changes in entry width.

The most likely place to obtain the real-time data necessary to make decisions about a mine roof is from a roof bolter. Research by the U.S. Bureau of Mines and its contractors have shown this to be true. Information that could be provided to a system from a bolter are:

- 1. Drilling parameters—position, penetration rate, thrust, torque, and rotation rate.
- 2. Amount of reflected light from cuttings in the vacuum line.
- 3. Vibration of the roof at some distance from the drill.

Other potential sources of input data are instrumented bolts, geophones, vibrating wire stress meters, flatjacks, etc. In other words, almost any geotechnical monitoring device could provide input to the system.

The U.S. Bureau of Mines has been collecting data from roof bolters as the basis for developing "smart drills," i.e. autonomous roof bolting machines.[1] The data collection phase of this research is proceeding successfully, and it was appropriate to endeavor to use data from an underground coal mine in the western

Correspondence should be sent to: Dr R. L. King, Department of Electrical and Computer Engineering, Mississippi State University, P.O. Drawer EE, Mississippi State, MS 39762, U.S.A.

United States in the development of an intelligent system for geologic feature detection. The data were from only one section of the mine where the roof was predominately sandstone. Five core holes had already been drilled and logged along 15.24 m (50 ft) of the section. Five holes were then drilled around each core hole with an instrumented drill to record torque, thrust, drill speed, penetration rate, and drill bit position. In addition, the drill was equipped with a void detector that gave an audible alarm that was recorded by the drill operator.

The obvious step was to use the drilling parameter data in some sort of pattern recognition scheme. Statistical approaches were inappropriate, because data were limited to only one section. However, the abundance of data from the section and its quality indicated that an artificial neural network approach to classifying the data should be pursued. However, the data sets had not been labeled, which made it impossible to use a supervised learning approach. Therefore, it was decided to use an unsupervised learning approach to discover patterns within the data that could be used as labels.

2. UNSUPERVISED LEARNING

Unsupervised learning is the clustering of patterns in feature space. The operation takes a "follow the leader" approach. That is, when a new pattern is presented to the network, it will associate with an existing cluster unless the new pattern is too dissimilar. In that case, the new pattern will begin a new cluster. Classification occurs when nearly all patterns within a cluster are of one class, so discovery of a cluster provides the basis for classification. The learning is "unsupervised" because the clusters are formed and learned without recourse to class labels or to information on attributes. An advantage of unsupervised learning is its innate ability to compress data, which results because the cluster center can serve as the prototype of a very large number of patterns in that cluster.

In this problem, the training set data were not classified. All that was available for classification and interpretation were core holes and their logs drilled within 0.61 m (2 ft) of the roof bolt holes (Fig. 1). Figure 2 illustrates the actual drilling parameter data from a roof bolt hole drilled near the core hole shown in Fig. 1. Since an accurate method of matching drilling parameter data with core log information did not exist, unsupervised learning appeared to be the only logical approach to classifying the drill data into a finite number of classes that would represent roof lithology.

Before unsupervised classification was attempted, a few important assumptions were made. It was decided that a microscopic picture of the mine roof (like the detailed core log of Fig. 1) was not necessary. What was needed was a more macroscopic view of the mine

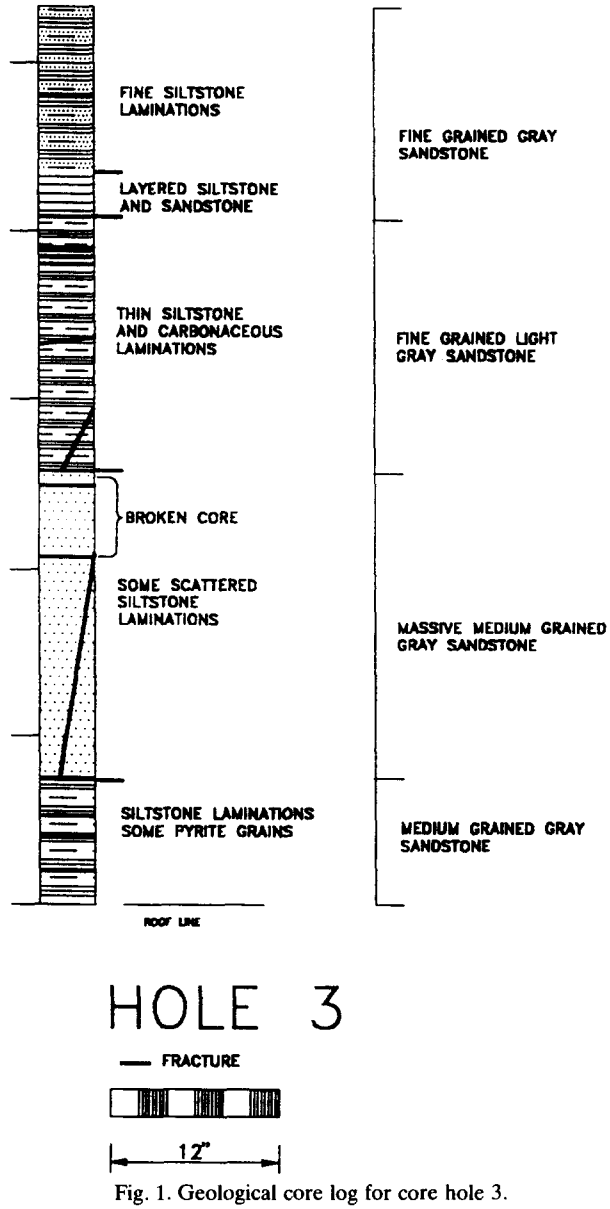

Fig. 1. Geological core log for core hole 3.

roof—the location of bedding planes, voids, etc. It was theorized that if the bolter were drilling within a given stratum of a certain average strength, a pattern should emerge that would describe a relationship between the different variables being monitored. Then, when the drill penetrated a new stratum or feature, a new pattern would emerge assuming the new stratum's strength was different from the one previously drilled. Therefore, it would only be necessary to classify a limited number of patterns representing the major features within a mine roof.

Unsupervised learning algorithms operate by determining the similarity of a new object to existing cluster centers, or exemplars. A training set is used to create the exemplars, and new objects are then compared to all existing exemplars. Training is done off-line and more effort may be expended in computation than is desirable for on-line recognition purposes. The objects are typically parameters arranged as vectors. The com-

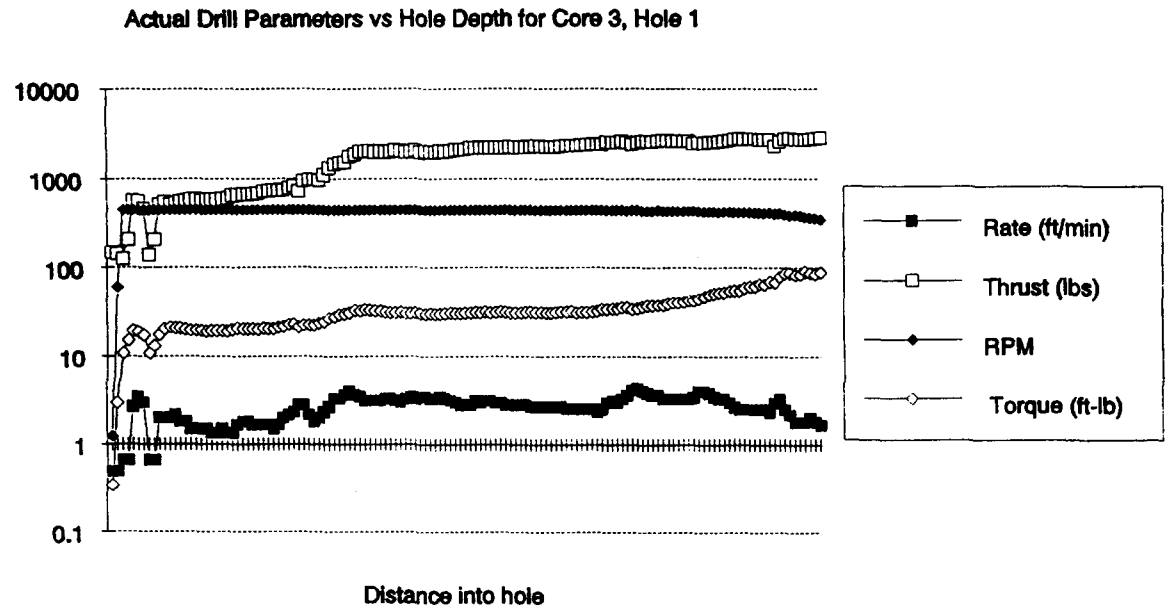

Fig. 2. Drilling parameters for core 3, hole 1. Note logarithmic vertical scale.

parison is then a vector distance, or metric, of some sort. The effectiveness of the clustering algorithm is completely dependent on the distance function used.

Because of the importance of correctly classifying the data from the drill holes, it was decided to review various similarity and distance metrics that could be used for clustering. It was also decided to explore a method of visualizing the clustered data in a two- or three-dimensional graph. Since the original data was four-dimensional (torque, thrust, drill speed, and penetration rate), it was thought that this would aid in ascertaining the validity of the clusters. In the following section of the paper, several distance metrics are compared for use in a clustering algorithm and a nonlinear mapping algorithm, and the efficiency of computation and discrimination of the metric are reviewed.

2.1. Euclidean distance

Euclidean distance is perhaps the simplest and most efficient metric in common use. Equation (1) shows the formula for computing this metric.

$$d(x,y) = \|x-y\| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}. \tag{1}$$

The Euclidean distance is the 2-norm of the vector difference between the cluster center and the sample vector. The metric is easily computed, although it requires a square root. Using the Euclidean distance in determining clustering of data geometrically corresponds to placing an n-suface in n-space. Any data points inside the n-surface are considered as being within the cluster. The radius of the n-surface is the threshold of the clustering algorithm. A disadvantage of using the Euclidean distance metric is that the shape of the volume included tends to exclude objects that are in elongated or ellipsoid clusters.

2.2. Minkowski distance

The Minkowski distance is directly related to the Euclidean distance and has some of the same advantages and disadvantages. It uses the p-norm of the vector difference.

$$d(x,y) = \left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{1/p}. \tag{2}$$

The Minkowski distance is computationally more difficult because it requires raising each vector difference to the pth power and then taking the pth root of the sum. Overflow must be managed carefully in the calculations. Using the pth power magnifies the differences and makes this metric more discriminating than the Euclidean. The shape of the volume included is the same as that of the Euclidean distance and has the same disadvantages.

2.3. Hamming distance

The Hamming distance (shown in equation 3) is most commonly used to find the distance between binary-represented values. It is used in cluster analysis for nonnumeric data represented by feature vectors. The presence of a feature is represented by a 0 and its absence by a 1.

$$d_H(x,y) = \sum_i \delta(x_i, y_i), \text{ where } \delta(x_i, y_i) = \begin{cases} O \text{ iff } x_i = y_i \\ 1 \text{ otherwise.} \end{cases} \tag{3}$$

The Hamming distance is just the sum of feature mismatches in the data. Computation of the Hamming distance is extremely simple, but data must be amenable to representation as binary values to give mean-

ingful results. This is not typical of real-world data so forcing the data into this representation can lead to artificial and/or meaningless conclusions.

2.4. Mahalanobis distance

The Mahalanobis distance is based on the generalized Gaussian distribution for multidimensional space. A single-point centroid exists for some set of data. The metric is then the distance between the ith sample, $\mathbf{x}$, and the centroid, $\mathbf{y}$, given by

$$d_m(x,y) = +\sqrt{(x-y)^T \sum^{-1} (x-y)}, \tag{4}$$

where Σ is the covariance matrix of the data set. Since the covariance matrix is rarely known, it must be estimated from a sample of the data set of size n. The revised metric is

$$d_m(x,y) = +\sqrt{(x-y)^T S^{-1} (x-y)}, \tag{5}$$

where

$$S = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - y)^T (x_i - y). \tag{6}$$

The geometric representation of the Mahalanobis distance model is an ellipsoid-shaped cluster with the population mean as its centroid and its size defined by the value of chi square at a specified confidence level with k degrees of freedom.[2] A sample is defined as within the cluster if its Mahalanobis distance falls inside a sufficiently high probability level.

The computational load for the Mahalanobis distance is high, requiring two vector-matrix multiplies and a matrix inverse. The Mahalanobis metric has the advantage of taking into account the correlation among the data and is unaffected by a change of scale of the data.[3]

2.5. Correlation

The correlation is a measure of the similarity of the object to the exemplar rather than the distance from the exemplar. A higher correlation value denotes more similarity. The correlation coefficient r, normally defined and used in statistics, is given in equation (7)

$$r^2 = \frac{\left[\sum_i (x_i - \bar{x})(y_i - \bar{y})\right]^2}{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}. \tag{7}$$

Usually the samples are normalized to the average of all the samples, allowing the use of the simpler equation (8).

$$r^2 = \frac{\left(\sum_i x_i y_i\right)^2}{\sum_i x_i^2 \sum_i y_i^2} \tag{8}$$

If each of the exemplars is normalized so that

$$\sum_i x_i^2 = 1 \quad \text{and} \quad \sum_i y_i^2 = 1, \tag{9}$$

then the similarity measure becomes just

$$r^2 = \left(\sum_i x_i y_i\right)^2. \tag{10}$$

This quantity can be recognized as the inner (dot) product of two vectors and is often used without normalizing the vectors. The coefficient produced may be larger than 1 and has a more complex dependency on the similarity of the vectors.

A means of normalizing the metric derived in equation (10) is to divide by the product of the norms of the two vectors involved. The measure thus derived corresponds to the cosine of the angle between the vectors, as seen in equation (11).

$$\cos\theta = \frac{(x \cdot y)}{\|x\| \, \|y\|}. \tag{11}$$

This metric has a maximum value of 1 and is easily computed. The discrimination of the measure is small since the metric normalizes out the vector lengths. One concern with the correlation metric is that two sets of vectors may have the same metric, but the vector endpoints can be widely separated if the vectors have different lengths.

2.6. Tanimoto

The Tanimoto similarity measure, as seen in equation (12), is a generalization of its basic form.[4]

$$S_T(x,y) = \frac{(x,y)}{\|x\|^2 + \|y\|^2 - (x,y)}. \tag{12}$$

The Tanimoto metric is typically used in the comparison of nonnumeric sets. Equation (13) shows that the similarity of two sets, A and B, may be defined as a ratio of the number of their elements in common to the number of elements that are different.

$$S_T(x,y) = \frac{n(A \cap B)}{n(A \cup B)} = \frac{n(A \cap B)}{n(A) + n(B) - n(A \cup B)}. \tag{13}$$

2.7. Clustering and nonlinear mapping algorithms

Promising distance and similarity metrics were used in a simple clustering algorithm to compare their com-

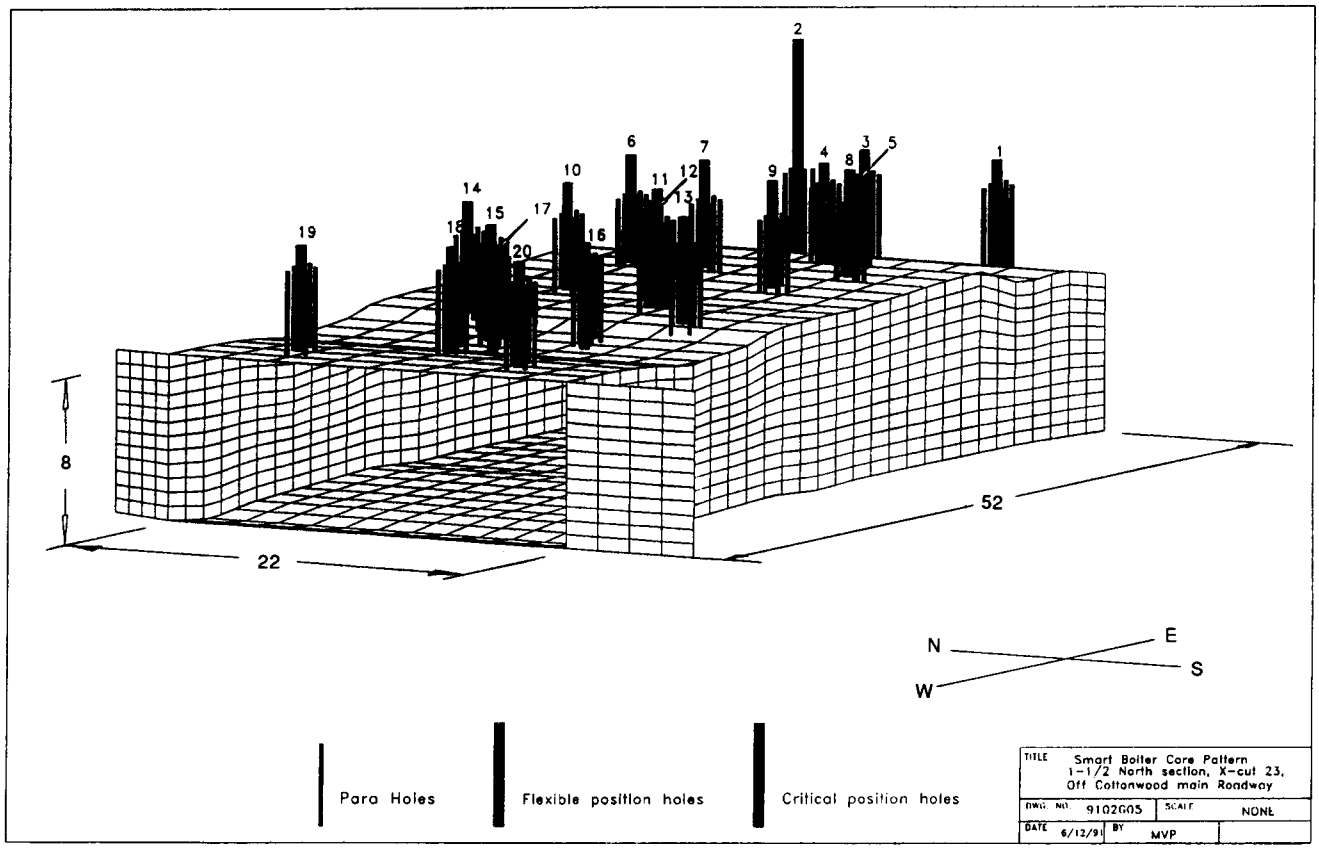

Fig. 3. Three-dimensional depiction of sandstone section of mine showing various drill holes.

putational efficiency and discrimination properties for the roof bolter data. Then a nonlinear mapping algorithm was used to visualize the data.

The clustering algorithm used was quite simple. The first data vector was taken as an exemplar and the second data vector was compared to it. If the distance function had a value less than some given threshold, the second vector became part of the cluster. When a new vector was found to be within the cluster, the cluster's center was updated and this became the new exemplar. If a new data vector had a distance value above the threshold, a new cluster was formed with the data vector as the first exemplar. All subsequent data vectors were compared to each cluster exemplar with a minimum distance used to determine membership in the cluster.

The threshold may be adjusted to be so small that each pattern forms a unique cluster, or so large that all patterns fall into one cluster, or anywhere in between. The clusters formed by the algorithm must be inspected by a human expert after they are grouped to ensure they are reasonable. The program used in this research was adapted from one originally developed by M.S. Klassen and Yoh-Han Pao.[5]

The purpose of the nonlinear mapping algorithm[6] was to reduce multidimensional data to a lower dimensional space (specifically two- and three-dimensional spaces) where structure or clustering may be identified visually. The structure of the original multidimensional data was approximately preserved by fitting the points in the lower dimensional space (l-space) so that the distances approximated the interpoint distances in the higher dimensional space (M-space). An initial location for the points was produced by projecting the original M-space data vectors orthogonally onto the l-space. An error term E was calculated from the mean squared error between the l-space distances and the M-space distances. A steepest descent procedure was then used to adjust the points so as to decrease the error.

2.8. Results

As mentioned previously, five core holes were drilled and logged along 15.24 m (50 ft) of the section. Then five roof bolt holes were drilled around each core hole. Figure 3 depicts a three-dimensional view of the mine entry and the drill holes. Table 1 shows how the drilling

Table 1. Data from drill hole 1-1 during hole collaring

Position (mm)	Penetration rate (mm/s)	Thrust (N)	Drill speed (rpm)	Torque (J)
183.39	0.9	−285	356.2	81.2
183.90	5.1	−0.5	356.2	78.8
187.71	6.8	165	356.7	77.3
192.79	11.9	1421	357.3	80.2
202.95	12.8	1855	359.3	83.1
210.06	11.9	1690	357.9	81.1
215.65	11.0	1689	357.5	81.6
222.25	9.4	957	357.3	77.9
225.30	4.3	544	358.1	69.2

Table 2. Variation in data from drill hole 1-1

	Penetration rate (mm/s)	Thrust (N)	Drill speed (rpm)	Torque (J)
Mean	12.1	6187	366.7	76.6
Standard deviation	3.7	2435	4.5	3.8
Minimum	0.9	−285	355.4	69.2
Maximum	20.4	14,283	374.7	92.4

parameter data change during drilling and what occurs when the drill is first starting to collar a hole. Note the negative thrust values recorded and the variations in the magnitudes of recorded data. Table 2 shows variation in drilling parameters during the drilling of a 1.22 m-long (4 ft) roof bolt hole.

Initially, the drilling parameter data from five different drill holes (one from each core hole location) were individually clustered using the Euclidian distance metric in conjunction with the clustering algorithm. This resulted in mapping the individual drilling parameter data into a few patterns. Figure 4 shows the results of clustering actual torque, thrust, speed, and penetration rate data using the hole 3-1 data of Fig. 2.

Although a reduction in the randomness of the data and a grouping into specific clusters were achieved using unsupervised learning, it was still difficult to visualize the results. It was necessary to compare the clusters with the core logs to determine whether or not the thresholds selected for learning were making reasonable cluster centers. If the clusters had not been reasonable, then another threshold value could have been selected and the clustering process repeated.

To permit the comparison, it was decided to fuse or combine the clustered drill parameters into a relative measure of specific energy.[7] This measure is described in the relationship:

$$e=\left(\frac{F}{A}\right)+\left(\frac{2\pi}{A}\right)\left(\frac{NT}{u}\right), \tag{14}$$

where e = specific energy (J/m^3), F = thrust (N), A = area of the hole (m^3), N = rotation speed (rpm), T = torque (J), u = penetration rate (m/s).

It should be noted that the specific energy has two components—a thrust component and a rotational component. No attempt was made to directly relate the specific energy term to the strength of the rock types being drilled. The specific energy was used only to order cluster numbers from the lowest value of specific energy (cluster 0) to the highest value of specific energy (cluster n). Thus, cluster numbers served as a relative strength index; i.e. the geologic feature represented by cluster 3 was easier to drill through than the geologic feature represented by cluster 5.

After several attempts at selecting threshold values, favorable results were obtained (Fig. 5), and it was decided the unsupervised learning approach to classification of the drilling parameter data should be continued. These initial tests were conducted using the Euclidian distance metric. As a result of this success, all data from five individual holes (1-1, 2E-1, 3-1, 4-1, and 5E-1) were grouped into one large data set of 617 patterns and the impact of the various distance metrics on the clustering was explored. The hole notation x–y refers to a specific core hole, x, and a particular roof bolt drill hole, y, drilled in the vicinity of core hole x.

Figures 6–9 illustrate various distance metrics applied to the 617-pattern data set and subsequent nonlinear mapping of the results to a two-dimensional plot. Figure 6 represents nonlinear mapping of the actual four-dimensional data into two dimensions. Note that the units for the axes will not be of any great significance. The reader is encouraged to inspect the mappings with a view toward visualizing clusters of data. Figure 7 is a mapping of the data after a Euclidian metric has been applied, and Fig. 8 represents a Minkowski metric where $p=1$. Arrows have been added to help in visualizing clusters that the authors believe they see. Figure 9 uses the correlation measure

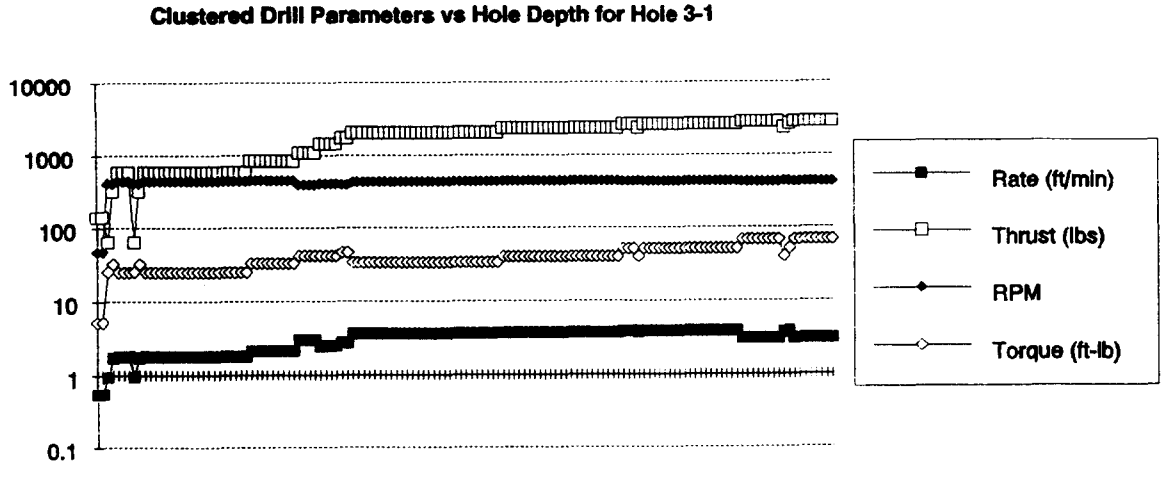

Fig. 4. Clustered drilling parameters for core 3, hole 1.

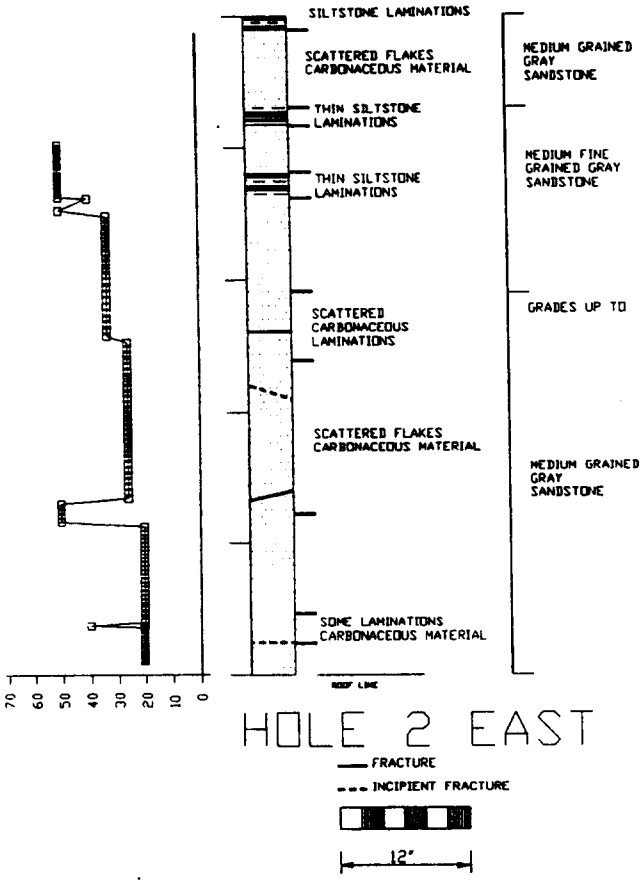

Fig. 5. Mapping of hole 2E-1 based on clustering of drilling parameter data.

of similarity. No arrows were added to this figure so that readers would have the opportunity to visualize clusters.

After the various metrics applicable to real valued data were all tried with the nonlinear mapping technique, no significant differences were noted in the various metrics. Therefore, it was appropriate to continue using the Euclidian distance metric for the data because this metric is easier to program and more widely accepted. It was speculated that since all the data were from sandstone of fairly similar strength characteristics, the separation in the data was not large

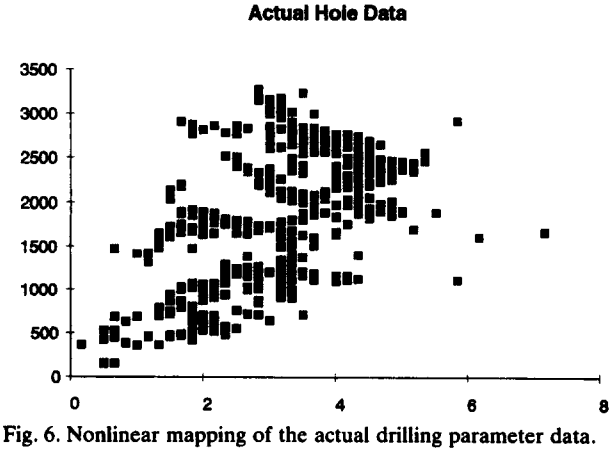

Fig. 6. Nonlinear mapping of the actual drilling parameter data.

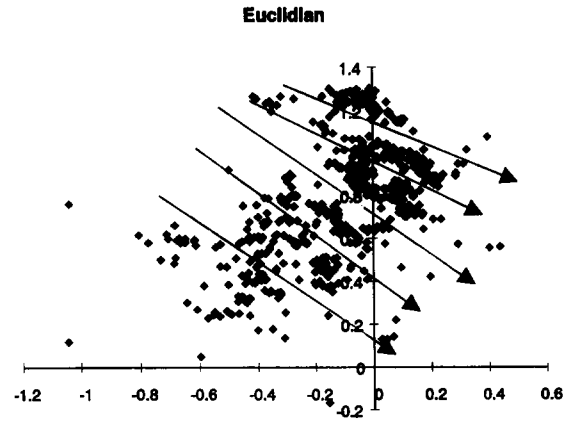

Fig. 7. Nonlinear mapping of drilling parameter data using the Euclidian metric.

enough to allow the various metrics to discern significant differences. However, if research continues in this area and data sets that are composed of a variety of different rock types become available, it would be useful to again attempt clustering the data with the various distance and similarity measures to see if anything different can be learned.

After several trials, the 617 data sets were clustered into 16 different classes. Determination of the threshold was based primarily on how the clustering algorithm handled what was obviously bad data (for example, entries 1, 2, and 3 in Table 1). It was logical that the unsupervised learning should be able to classify bad or unrealistic data into a sort of "trash" category. By adjusting the threshold of learning for the algorithm and then reviewing how the system handled the known "bad" data, the cluster centers in Table 3 were eventually obtained. Using equation 14, the cluster numbers were ordered from the least amount of energy required for drilling (cluster 0) to the most (cluster 15). The actual values of specific energy of drilling ranged from 3750.75 to 618521.78 kJ/m^3 (544–89,709 in-lb/in^3). Unconfined compressive strength tests on sandstone

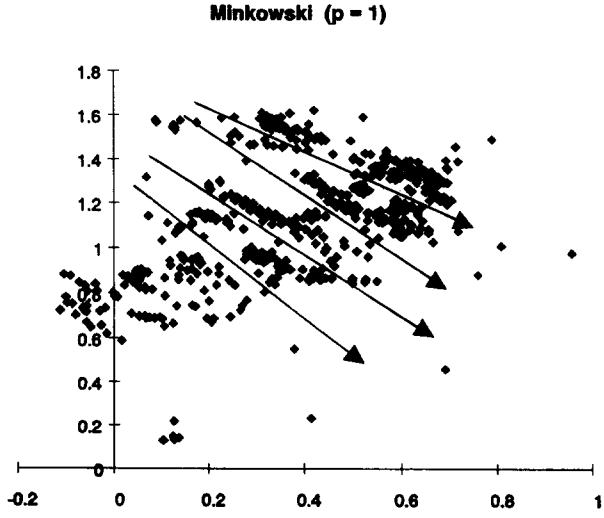

Fig. 8. Non-linear mapping of drilling parameter data using the Minkowski metric with $p = 1$.

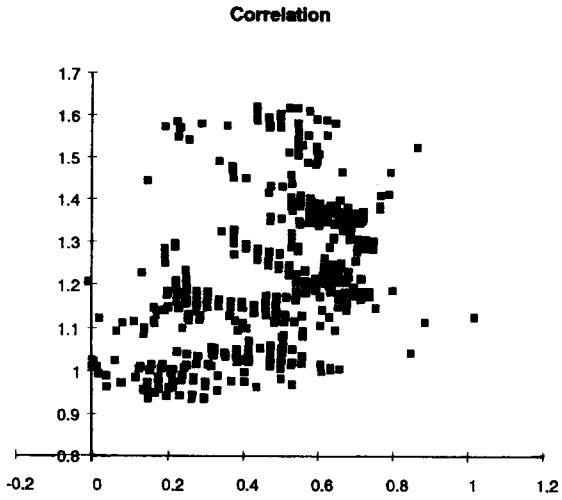

Fig. 9. Nonlinear mapping of drilling parameter data using the correlation measure of similarity.

core samples from the roof ranged from 87.15 to 177.20 MPa (12,640 to 25,700 psi). Obviously, no direct relationship exists between the two measures. However, as stated previously, the specific energy measure was only used to order the cluster numbers into a relative strength index.

Two different classifications of the data were then developed to permit a comparison of their efficiencies

Table 3. Sixteen-exemplar set obtained from 617 unclassified training patterns

Cluster no.	Penetration rate (mm/s)	Thrust (N)	Drill speed (rpm)	Torque (J)
0	2.6	1879	1.2	0.5
1	2.8	622	47.3	7.1
2	14.4	3776	82.8	20.8
3	13.6	5088	171.3	27.9
4	18.6	9230	434.2	46.3
5	18.8	10,537	438.3	54.5
6	15.3	4926	395.2	57.0
7	9.3	2603	438.1	33.6
8	19.6	11,652	434.8	69.6
9	14.0	7759	394.8	63.3
10	12.8	6559	407.4	56.6
11	10.9	3765	447.6	45.4
12	8.8	1443	409.5	43.8
13	15.8	12,648	425.7	92.5
14	15.2	14,018	436.3	91.8
15	4.8	289	414.5	34.5

Table 4. Nine-exemplar set obtained from 617 unclassified training patterns

Cluster no.	Penetration rate (mm/s)	Thrust (N)	Drill speed (rpm)	Torque (J)
0	2.7	1251	24.3	3.8
1	14.4	3776	82.8	20.5
2	18.6	9793	426.5	49.2
3	14.0	4875	402.7	55.7
4	9.5	2335	428.2	37.9
5	14.3	7655	398.3	62.3
6	18.2	11,875	428.9	76.1
7	15.8	13,525	439.7	95.2
8	5.3	549	416.6	36.4

in classifying the validation patterns. Table 4 lists the clusters of the second classification. The threshold of the clustering algorithm was adjusted so that it grouped the 617 patterns into nine geologic features.

3. CONCLUSIONS

The use of unsupervised learning to discover patterns in roof bolter drill data was successful. The Euclidian distance metric was found to be a satisfactory measure of similarity. Two major results were achieved. First, the 617 drill parameter training patterns were classified (labeled) and can be used in the development of supervised learning classification networks. Second, the 617 data points representing the roof lithology have been reduced (data compression) to just a few (9 or 16) unique patterns or features representing major geologic features within a mine roof.

REFERENCES

1. Frizzell E. M., Howie W. L. and Briggs R. C. 'Roof drill—new source of low cost roof strata information. *Proc. 31st U.S. Symposium on Rock Mechanics*, 505–510, Golden, CO (1990).
2. Shah N. K. and P. Gemperline. Combination of the mahalanobis distance and residual variance pattern recognition techniques for classification of near-infrared reflectance spectra. *Analyt. Chem.* **62**, 466–467 (1990).
3. Everitt B. *Cluster Analysis*. Wiley, New York (1974).
4. Kohonen T. *Self-Organization and Associative Memory*, 3rd edn. Springer, New York (1989).
5. Pao Y. H. *Adaptive Pattern Recognition and Neural Networks*. Addison-Wesley, Reading, MA (1989).
6. Sammon J. W. Jr A nonlinear mapping for data structure analysis. *IEEE Trans. Comput.* **C-18**, 401–409 (1969).
7. Teale R. The concept of specific energy in rock drilling. *Int. J. Rock Mech. Mining Sci.* **2**, 57–73 (1965).

AUTHORS' BIOGRAPHIES

Roger L. King received the B.S.E.E. degree from West Virginia University, the M.S.E.E. degree from the University of Pittsburgh, and the Ph.D. degree from the University of Wales in Cardiff, Wales. His area of research at Mississippi State University is intelligent control. Before joining Mississippi State University he worked for 14 years at the Bureau of Mines. He is a member of IEEE, Eta Kappa Nu, Sigma Xi, AAAI and INNS. He is listed in Who's Who in the Computer Industry and Who's Who in Science and Engineering.

Michael A. Hicks received the B.S.E.E. from Mississippi State University and the M.S.E.E. from Syracuse University. He is presently pursuing the Ph.D. degree at Mississippi State University. He has been previously employed with General Electric Co. in Syracuse, NY and with Harris Corp. in Palm Bay, FL. His research interests include neural networks, asynchronous processing and parallel architectures. He is an IEEE Member and member of Eta Kappa Nu and Tau Beta Pi.

Stephen P. Signer is a project leader for the U.S. Bureau of Mines at the Spokane Research Center. For the past 10 years, his research has been in the area of coal mine ground control and has focused on the interaction mechanics involving artificial supports and rock structures. He received the 1990 U.S. Rock Mechanics Award in the Case Histories category for research on grouted roof bolts. Recently he has been developing knowledge-based systems and neural networks to improve support selection in coal mines. He has a B.S. degree in Civil Engineering from Gonzaga University.