Australasian Conference on Robotics and Automation (ACRA), December 2-4, 2009, Sydney, Australia

Spectral Feature Selection for Automated Rock Recognition using Gaussian Process Classification

H. Zhou, S. T. Monteiro, P. Hatherly, F. Ramos, E. Nettleton and F. Oppolzer
Australian Centre for Field Robotics, The University of Sydney, Australia
{h.zhou, s.monteiro, p.hatherly, f.ramos, e.nettleton, f.oppolzer}@acfr.usyd.edu.au

Abstract

A spectral feature selection scheme is proposed for multi-class automated rock recognition from real world drilling data using Gaussian Process classification. This work is part of a larger project aimed at surface mine automation. The motivation for this research is to investigate which combination of drilling data measurements is most relevant for rock recognition. We conduct feature selection in the frequency domain where characteristics are more distinguishable. In particular, we extended the spectral feature selection method from binary classification to multi-class classification by decomposing the multi-class classification dataset into a series of one versus one binary classification datasets. A non-uniform discrete Fourier transform (NDFT) is then applied to data on each of the binary classification features, where the features with the most consistent "major bandwidth" across all decomposed binary classifications are selected. The approach has been applied on multi-class rock recognition (based on drilling data) and results are presented on real world drilling data.

1 Introduction

Automated rock recognition is used to classify properties such as rock type and strength from blast hole drilling data on an autonomous drill rig. The system operates using measurements while drilling (MWD) data (e.g., rotation speed, pull down rate and pull down pressure, etc.) with a prior obtained from geophysical logging. This work is a key element of a larger project aimed at developing a fully autonomous, remotely operated mine. The rock recognition results are highly desired by the mining industry as they provide information that can be used in the optimization of the mine operations as well as mine planing and design [Gonzalez, 2007]. For instance, the rock strength can assist automatic loading of explosives in the charging process and the rock boundary map is useful for further blast hole design as well as long term strategic planning.

Figure 1: An autonomous blast hole drill rig used for collecting experimental MWD data in this paper.

The MWD data used for rock recognition are the measurements (called features in classification) collected from sensors equipped on large drill rigs used in mining for blast hole drilling. The MWD data in this work is from an autonomous blast hole drill rig (as shown in Figure 1). Each measurement corresponds to one of the sensors. They are primarily used to control and monitor the drilling process. We use the MWD data to relate the drill performance to the physical properties of the rocks being drilled. It is necessary to understand how the MWD measurements contribute to the classification output of the rock recognition. This facilitates a better understanding of how the system functions, and also provides insight as to what new information may be useful for improving performance. We resolve this problem by means of feature selection in Gaussian Process classification. This is a state-of-the-art classification approach [Rasmussen *et al.*, 2006] which is flexible and works especially well on high dimensional data (which is the case for rock recognition with numerous measurements).

Previously, feature selection has been done intuitively, i.e., to choose part of the measures that appear to be relevant to the classification. It is necessary to choose the features in a more meaningful and systematical way. In this paper, we analyze this problem in the frequency domain where some of the underlying characteristics of the data can be better revealed. We then propose to choose the features that have the most consistent frequency domain measure and discard the rest.

The study of classifying rock types from drilling data is not new. Machine learning methods have been applied on drilling data based rock recognition [Itakura *et al.*, 1997] [King *et al.*, 1993] [Utt, 1999] [Gonzalez, 2007]. Among these approaches, [Gonzalez, 2007] tried to apply Principal Component Analysis (PCA) to the feature data for the purpose of feature selection. However, the author does not seem to show an effective and consistent rule for PCA feature selection. Essentially, PCA aims to move as much variance as possible into the first few dimensions and retain these few dimensions (principal components) so that features with the greatest variance are kept, but this does not guarantee to maintain (or even enhance) the performance in supervised learning.

Mackay and Neal [Neal, 1996] used Automatic Relevance Determination (ARD) to assign each input a weight value (ranging from 0 to 1). The weight values have independent Gaussian prior distributions with standard deviation given by the corresponding hyper-parameter with some prior. Then, the posterior distribution of the hyper-parameters is calculated given the training data. The values of the hyper-parameters are proportional to the corresponding input weight values. An issue for ARD is: what prior should be used for the hyper-parameters? (as it will have a significant impact on the accuracy of estimation results of the hyper-parameters). Zhou et al. [Zhou *et al.*, 2009] eliminated inconsistent features for Gaussian Process classification by choosing the features with the most consistent spectral measure. The advantage of spectral feature selection is efficient and direct, avoids the need to choose priors for the hyper-parameters as it does for feature selection using ARD. We adopted spectral feature reduction method in [Zhou *et al.*, 2009], extended it for using in multi-class classification and applied to the feature selection for rock recognition.

The remainder of the paper is organized as follows. A brief introduction to Gaussian Process classification is given in Section II. Details of spectral feature analysis for multi-class classification are described in Section III. In Section IV, experimental results are presented and discussed, followed by a summary of the main conclusions in Section V.

2 Gaussian Process Classification

In this paper, we consider using the state-of-the-art Gaussian Process model for classification.

A Gaussian Process is a collection of random variables, any finite number of which have a joint Gaussian distribution [Rasmussen *et al.*, 2006]. A GP is fully specified by its mean function $\mu(\mathbf{x})$ and kernel function $k(\mathbf{x}, \mathbf{x}')$, i.e., $f \sim \mathrm{GP}(\mu, k)$. With the prior represented by the GP kernel function, GP classification models the posterior directly [Rasmussen *et al.*, 2006]. The kernel function's hyper-parameters can be learned from the training data. The kernel function studied in this paper is the Radial Basis Function (RBF).

Assume we have a data set $\mathcal{D}$ with n observations $\mathcal{D} = \{(\mathbf{x}_i, y_i), i = 1, 2, \cdots, n\}$, where $\mathbf{x}$ is the input vector of dimension d and y is the class label $+1/-1$. The input $d \times n$ matrix is denoted as X. Predictions for new inputs $\mathbf{x}'$ are made out of this given training data using the GP model. As described in [Rasmussen *et al.*, 2006], GP binary classification is done by first calculating the distribution over the latent function f corresponding to the test case

$$p(f'|X, y, \mathbf{x}') = \int p(f'|X, \mathbf{x}', f)p(f|X, y)df \qquad (1)$$

where $p(f|X, y) = p(y|f)p(f|X)/p(y|X)$ is the latent variable posterior, $p(f'|X, \mathbf{x}', f)$ is the predictive posterior wrt possible latent functions, and the values of this could lie anywhere within the range of $(-\infty, +\infty)$. So the probabilistic prediction is made by

$$\bar{\pi}' = p(y' = +1|X, y, \mathbf{x}') = \int s(f')p(f'|X, y, \mathbf{x}')df' \quad (2)$$

where s can be any sigmoid function that 'squashes' the prediction output to guarantee a valid probabilistic value within the range of $[0, 1]$.

For the multi-class classification problem with c classes (such as rock recognition), we turn it into a series of $(c(c-1)/2)$ one versus one two-class problems and apply binary classification individually to each of them, followed by max vote to assign the class labels [Friedman, 1996].

3 Spectral Feature Selection for Multi-Class Gaussian Process Classification

Following [Zhou *et al.*, 2009], we use spectral feature selection to choose the features. The advantages of spectral feature selection are that it is efficient and direct. It also avoids the need to choose priors for the hyper-parameters which was an issue with feature selection using Automatic Relevance Determination (ARD) [Neal, 1996].

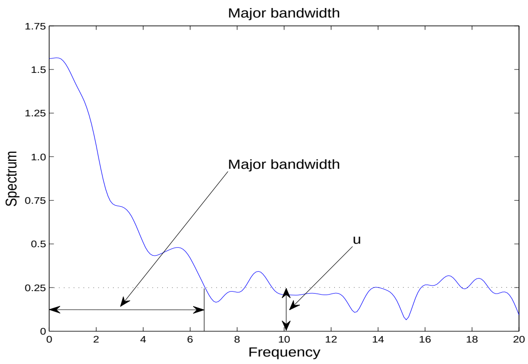


Figure 2: Major bandwidth of GP data, which is the range of frequencies from zero to the first level u crossing.

Table 1: Spectral feature selection algorithm.

Input: $N \times m$ data matrix $[x_j]$ and $N \times 1$ label matrix y, where each x_j is a $N \times 1$ vector which is the data on the j th feature, $j = 1...m$. N is the total number of data points and m is the total number of features (dimensions).

y is the $N \times 1$ label matrix with c classes.

(1) Derive a total of S one versus one $N_i \times m$ matrices $[vx_j^i]$ and the corresponding $N_i \times 1$ label matrices y^i from $[x_j]$ and y, where $S = c * (c - 1)/2$, $i = 1...S$, $y_n^i \in [-1, 1], n = 1...N_i$.

(2) For the ith matrix pair $[vx_j^i]$ and y^i, apply NDFT to data on each feature vx_j^i to calculate the data frequency content $F(f)_j^i$.

(3) Estimate the major bandwidth MBW_j^i of each feature on the ith matrix pair using equations 3 and 4. Then calculate the ith matrix pair major bandwidth average $\overline{MBW^i}$

$$\overline{MBW^i} = \sum(MBW_j^i)/m, \; j = 1...m$$

(4) For the jth feature, compute
num_j=total number of MBW_j^i
where $MBW_j^i >= \eta * \overline{MBW^i}$
$i = 1...S$, η is set to be 0.25.

(5) $ox_{j'} = x_j$ if
$num_j > 0.5 * S$

Output: $N \times m'$ data matrix $[ox_{j'}]$, where m' is the number of chosen features.

As the feature data are usually unevenly spaced high dimensional data, non-uniform discrete Fourier transform (NDFT) [Bagchi *et al.*, 1999] is applied to the data on each feature to calculate the data frequency content (which is the NDFT magnitude). From this, a metric called "major bandwidth" (MBW) is extracted.

Intuitively, MBW is the bandwidth range where the major energy of the frequency content is concentrated. It is precisely defined as:

Major bandwidth: In a Gaussian Process data frequency content, the range of frequencies from zero to the first level u crossing (see Figure 2).

MBW value should be the smallest frequency value satisfying

$$F(f) > u * max(F) \qquad f < MBW \tag{3}$$

$$F(f) \leq u * max(F) \qquad f = MBW \tag{4}$$

where F is the data frequency content (NDFT result) of the given data. f is the frequency variable and u is the level where the data frequency content first cross, which is used for extracting the central lobe and filtering the noisy side lobes. u is empirically set to be 0.25.

Since the Gaussian Process kernel is isotropic [Rasmussen *et al.*, 2006], i.e., rotation invariant (both on signal domain and frequency domain), the main idea behind spectral feature selection is to choose the features that have the most consistent spectral metric, which is MBW in our approach.

As indicated in Section 2, we convert the multi-class rock recognition problem to a binary classification problem by extracting a series of one versus one data matrices from the original data. For each of the one versus one data matrix, NDFT is applied to data of each feature individually to estimate the corresponding MBW values. We then apply max vote to the MBW values of each feature across all one versus one division, i.e., the feature will be chosen only when a majority number of one versus one MBW values is greater than or equal to the preset threshold. The rest of the features will be discarded. In this way, the chosen features have a more consistent MBW, which can fit better to the isotropic Gaussian Process kernel. Details of the spectral feature selection algorithm are summarized in Table 1. Binary GP classification is implemented using Lawrence's fast GP classification approach, which makes the calculation significantly faster and close to real time (scaling at most $O(n \cdot d^2)$, where n is the training dataset size and d is the active set size which is much smaller than n) [Lawrence *et al.*, 2003].

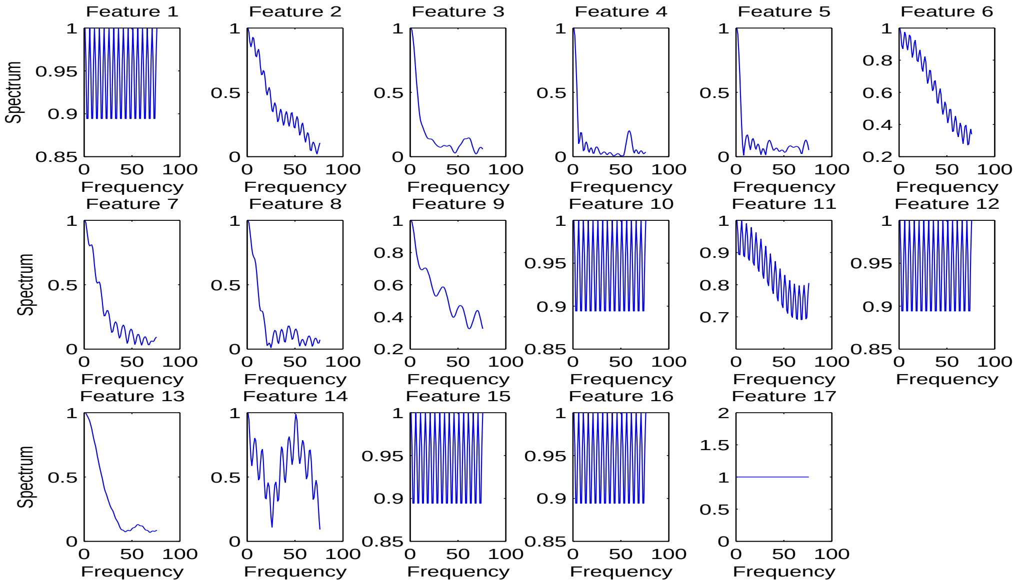


Figure 3: Data frequency contents (frequency unit is Hz) of all features in the 3 class rotary dataset.

Table 2: Features used for rotary drilling.

Feature index	Feature name
1	Time
2	Rotation speed
3	Pull down rate
4	Drill depth
5	Head position
6	Target depth
7	Rotation pressure
8	Pull down pressure
9	Bit air pressure
10	Rotation relief
11	Pull down relief
12	Rotation output
13	Pull down output
14	Water injection flow
15	Water on
16	Air on
17	Hold back

Table 3: Features used for percussion drilling.

Feature index	Feature name
1	Time
2	Rotation speed
3	Pull down rate
4	Drill depth
5	Rotation pressure
6	Pull down pressure
7	Bit air pressure
8	Hydraulic case pressure
9	Right tram pressure
10	Rotation pressure
11	Aux hydraulic pressure
12	Hydraulic charge pressure
13	Compressor air tank pressure
14	Auto lube pressure
15	Engine oil pressure
16	Compressor oil pressure
17	Water injection pressure

4 Experiments and Results

Tests were carried out using various real world MWD datasets collected from 135 blast holes at an iron ore open pit mine in Western Australia. The data includes information from both rotary and percussion drilling modes. Each blast hole is approximately 10m deep, and the drilling measurements in each hole are downsampled at 10cm interval. The drilling measurements collected for rotary and percussion drilling are shown in Table 2 and Table 3 respectively. The geologists label and correspond the drilling measurements to several lithological rock types. Two labeling schemes are applied. The first uses 3 class labeling which divides the rocks to shale, ore and bif. The second uses 5 class labeling which categorizes the rocks as shale, low grade medium ore, high grade medium ore, high grade soft ore and bif.

In the work presented here, only five of the drilling measurements were selected for use in the rock recog-

Table 6: Illustration of spectral feature selection on 3 class percussion drilling dataset.

Feature index	MBW of class 1 vs 2	If chosen (Y/N)	MBW of class 1 vs 3	If chosen (Y/N)	MBW of class 2 vs 3	If chosen (Y/N)	Total vote	Final decision
No. 1	27	N	55	N	74	N	0	N
No. 2	266	Y	301	Y	702	Y	3	Y
No. 3	262	Y	490	Y	752	Y	3	Y
No. 4	61	N	68	N	65	N	0	N
No. 5	166	N	125	Y	99	N	1	N
No. 6	743	Y	497	Y	54	N	2	Y
No. 7	607	Y	395	Y	65	N	2	Y
No. 8	105	N	378	Y	1028	Y	2	Y
No. 9	689	Y	264	Y	495	Y	3	Y
No. 10	166	N	125	Y	100	Y	2	Y
No. 11	2655	Y	1200	Y	429	Y	3	Y
No. 12	935	Y	340	Y	474	Y	3	Y
No. 13	664	Y	357	Y	61	N	2	Y
No. 14	257	Y	516	Y	1524	Y	3	Y
No. 15	2707	Y	430	Y	736	Y	3	Y
No. 16	488	Y	369	Y	77	N	2	Y
No. 17	584	Y	309	Y	69	N	2	Y
Mean	670		366		400			
		class 1: shale	class 2: ore	class 3: bif				

Table 4: Number of features selected in drilling datasets for classification.

Drilling data type	Empirically chosen features	Spectrally chosen features
Rotary - 3 classes	5	11
Percussion - 3 classes	5	14
Rotary - 5 classes	5	11
Percussion - 5 classes	5	15

nition system. These five measurements are: rotation speed, pull down rate, rotation pressure, pull down pressure and bit air pressure. The selection of these measurements was done empirically by picking out those that are considered relevant and meaningful to the rock types.

Following our proposed method in Section 3, spectral feature selection is applied to all the drilling datasets with the selected feature numbers shown in Table 4. As an example, the details of spectral feature selection on 3 class rotary data are shown in Figure 3 and Table 5, where Figure 3 shows the data frequency contents of each individual features and the selected feature details are listed in Table 5. It can be seen that only those features with a similar major bandwidth are selected and the rest discarded.

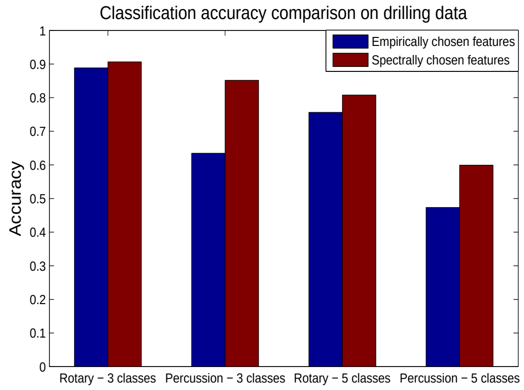
Table 6 illustrates how the multi-class spectral feature selection algorithm in Table 1 works. 3 one versus one data matrices are extracted from the original data. NDFT is applied to data of each feature in every one

Table 5: Selected feature details for 3 class rotary drilling dataset.

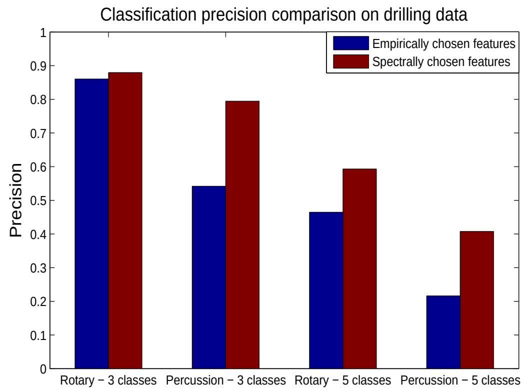
Feature index	Feature name
2	Rotation speed
3	Pull down rate
4	Drill depth
5	Head position
6	Target depth
7	Rotation pressure
8	Pull down pressure
9	Bit air pressure
11	Pull down relief
13	Pull down output
14	Water injection flow

versus one data matrix individually to estimate the corresponding MBW values. With each group of one versus one MBW values, initial judgements are made on each feature to decide if it should be chosen. For those initially selected features, the relevant MBW value must be greater than a preset threshold which is the mean MBW values (of all features within each one versus one group) weighted with a coefficient. Then max vote is applied to the initial feature selection results from all one versus one data, from which the final feature selection decision is made.

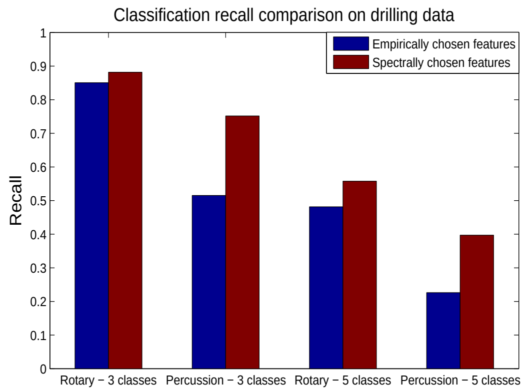
Data with selected features are classified with GP classification using k-fold cross validation. The classification



(a) Classification accuracy comparison on varied MWD data using empirically chosen features and spectrally chosen features.



(b) Classification precision comparison on varied MWD data using empirically chosen features and spectrally chosen features.



(c) Classification recall comparison on varied MWD data using empirically chosen features and spectrally chosen features.

Figure 4: Classification results comparison between empirically chosen features and spectrally chosen features of varied MWD data.

results are evaluated by calculating accuracy, precision and recall, which reflect the classification performance from varied aspects. Accuracy is the percentage of all correct predictions (both positive and negative), precision is the ratio of correct labels among the positive predictions and recall is the percentage of the positive labels that has been correctly predicted.

The classification results are shown in Figure 4. It can be seen that in all configurations of drilling mode and labeling scheme, the classification evaluation results (accuracy, precision and recall) of the data with spectrally chosen features constantly outperforms the classification results on the data with empirically chosen features. It should also be noted that there is a larger performance leap on the percussion drilling data compared with the rotary drilling data. This difference shows that spectral feature selection can effectively capture the most prominent features and considerably improve the classification performance over the raw percussion drilling data which are normally less distinguishable (compared with rotary drilling data).

5 Conclusions

The development of new sensing and classification systems to support large scale mine automation is a key challenge. This work has begun to address this by proposing a spectral feature selection method for multi-class GP classification of rock types from real world drilling data (MWD data). The main contribution of this paper is generalizing the spectral feature selection method from binary classification to multi-class classification and applying to rock recognition. By analyzing feature data on the frequency domain, a measure called "major bandwidth" is defined so that features with similar "major bandwidth" across all classes are chosen and the rest discarded. Although the proposed approach was applied to MWD data classification, it is equally applicable to varied classification / pattern recognition applications. The experimental results have shown the effectiveness of our approach.

In future work, it is worthwhile to compare the proposed spectral feature selection based GP classification with some other relevant approaches, e.g., decision tree learning [Mitchell, 1997].

Acknowledgments

This work has been supported by the Rio Tinto Centre for Mine Automation and the ARC Centre of Excellence programme, funded by the Australian Research Council (ARC) and the New South Wales State Government. The authors also acknowledge the support of Annette Pal, James Batchelor and Charles McHugh of Rio Tinto.

References

[Rasmussen *et al.*, 2006] C.E. Rasmussen and C.K.I. Williams. *Gaussian Processes for Machine Learning.* The MIT Press, 2006.

[Friedman, 1996] J. Friedman. Another Approach to Polychotomous Classification. Technical Report, Stanford University, 1996.

[Bagchi *et al.*, 1999] S. Bagchi and S. K. Mitra. *The Nonuniform Discrete Fourier Transform and Its Applications in Signal Processing.* Kulwer Aacdemic Pulishers, 1999.

[Lawrence *et al.*, 2003] N. Lawrence, M. Seeger and R. Herbrich. Fast Sparse Gaussian Process Methods: The Informative Vector Machine. *Advances in Neural Information Processing Systems*, 2003.

[Neal, 1996] R. M. Neal. *Bayesian Learning for Neural Networks.* Springer-Verlag New York, 1996.

[Zhou *et al.*, 2009] H. Zhou, L. Wang and D. Suter. Human Action Recognition using Feature - Reduced Gaussian Process Classification. *Pattern Recognition Letters*, April 2009.

[Gonzalez, 2007] Jorge Eduardo Jose Martin Gonzalez. Application of Pattern Recognition Techniques to Monitoring-While-Drilling on a Rotary Electric Blasthole Drill at an Open-Pit Coal Mine. *Thesis for Master of Science*, Queen's University, Canada, 2007.

[Scoble *et al.*, 1989] M. J. Scoble, J. Peck and C. Hendricks. Correlation between Rotary Drill Performance Parameters and Borehole Geophysical Logging. *Mining Science and Technology*, 8, pp. 301-312, 1989.

[Teale, 1965] R. Teale. The Concept of Specific Energy in Rock Drilling. *International Journal of Rock Mechanics and Mining Sciences*, 2, pp. 57-73, 1965.

[Itakura *et al.*, 1997] K. Itakura, K. Sato, Y. Ichihara, G. Deguchi, H. Matsumoto and H. Eguchi. Development of a Roof Logging System by Rock Bolt Drilling. *Transactions Institute of Mining and Metallurgy*, Section A 106, pp. 118-123, 1997.

[King *et al.*, 1993] R. L. King, M. A. Hicks and S. P. Signer. Using Unsupervised Learning for Feature Detection in a Coal Mine Roof. *Engineering Applications of Artificial Intelligence*, 6(6), pp. 565-573, 1993.

[Utt, 1999] W. K. Utt. Neural Network Technology for Strata Strength Characterization. *In Proceedings of International Joint Conference on Neural Networks*, 6, pp. 3806-3809, 1999.

[King *et al.*, 1993] R. L. King, M. A. Hicks and S. P. Signer. Using Unsupervised Learning for Feature Detection in a Coal Mine Roof. *Engineering Applications of Artificial Intelligence*, 6(6), pp. 565-573, 1993.

[Mitchell, 1997] T. M. Mitchell. *Machine Learning.* McGraw Hill, 1997.