

NEURAL NETWORKS AND THE CLASSIFICATION OF MINERALOGICAL SAMPLES USING X-RAY SPECTRA

Marcus Gallagher and Peter Deacon

School of Information Technology and Electrical Engineering
University of Queensland, Brisbane Qld 4072, Australia
marcusg@itee.uq.edu.au, Peter.Deacon@mincom.com

ABSTRACT

The automatic classification of large numbers of mineral samples is a practical problem in mining research. A system currently in use is based on simple statistical tests. Although the system performs well under typical conditions, the data collection procedure can be very time-consuming. This time can be significantly reduced, but at a cost of introducing noise into the data, leading to a degradation in classification performance. This paper reports on an initial investigation into the application of neural network techniques to the mineral identification task, and compares the performance of these methods to the current system. The results are very encouraging and suggest that a more powerful classifier might allow the data collection process to be significantly sped up without significant loss of classification accuracy for the overall system.

1. INTRODUCTION

In mining research and industry, an important task is the automated analysis of large numbers of grain-structured mineral particles. Energy-dispersive X-ray analysis is used to identify different mineral types that may be present in the samples. This analysis is able to generate large amounts of data using large amounts of particles, however there is a time factor associated with the collection and quality of this data. Because of the large volume of particles that are analysed, any reduction in this time factor would become significant, assuming there is no trade-off with the other system components.

This paper describes our approach to mineral classification using artificial neural network models. Samples are classified on the basis of data obtained from a scanning electron microscope. The following section details the mineral classification problem domain and data collection process. Section 3 describes the data pre-processing and feature extraction techniques that were used in our experiments as

The authors would like to acknowledge Dr. Ying Gu from the Julius Kruttschnitt Mineral Research Centre for the initiation of this project and for providing the mineralogical data.

well as the neural network models used. Results are discussed in Section 4, and Section 5 provides some conclusions.

2. CLASSIFYING MINERAL SAMPLES USING X-RAY SPECTRA DATA

The identification of specific minerals within particles is a laborious task that is machine and computer-automated. A system currently in use at the University of Queensland's Julius Kruttschnitt Mineral Research Centre¹ utilizes a scanning electron microscope and energy dispersive analysis to provide data for the identification of a given particle. Fig.1 shows an image of a typical particle. The particle will be scanned at each point on the grid marked by an "x". Points A and B on the particle correspond to different minerals, while point C indicates a mixture of these two minerals.

The mineral liberation analyser contains a scanning electron microscope that fires a focused beam of electrons onto a mineral sample. Back scattered electrons are emitted from the sample as a result, at an intensity that is a function of the atomic number of the particular sample. The energy released in this process falls in the X-ray spectrum. An X-ray analyser is able to collect X-ray spectra by recording the intensity and energies of the X-rays emitted from the sample. Fig.2 shows an example spectra from a mineral sample. The collected X-ray spectra depend on the amount of time that the electron beam is focused on a specific point in the sample. In addition, electron beams slow down when they hit a sample, resulting in a degree of noise in the X-ray spectra (i.e, X-rays present that are not characteristic of the given sample). Fig.3 is a spectra collected from the same sample as that of Fig.2, but at a shorter collection time (1759 back-scattered "hits" are recorded in Fig.2, compared with only 884 from Fig.3. It can be seen that the significance of the noise in the spectra increases as the time spent scanning (and hence the back-scattered hit count) decreases. Following collection, X-ray spectra are classified on the basis of

¹<http://www.jkmrc.uq.edu.au>

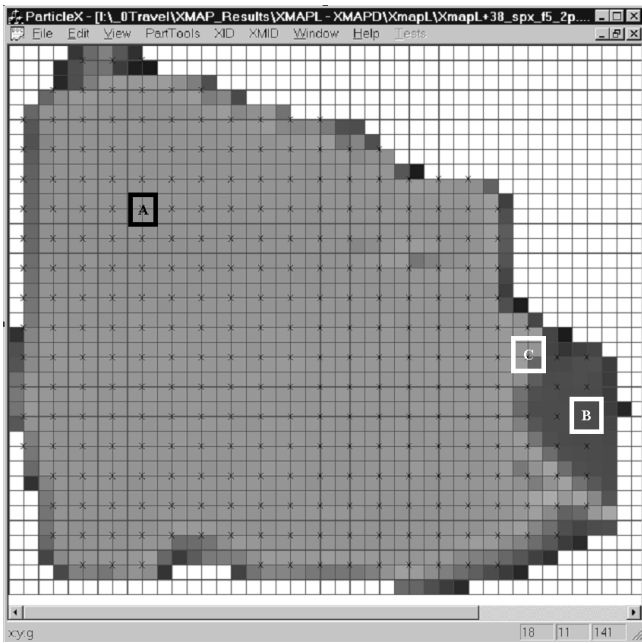


Figure 1: Grey-scale image produced from scanning of a mineral particle with different types of minerals present.

statistical tests into a number of given classes of minerals (since the theoretical number of expected hits in each bin along the spectra can be specified).

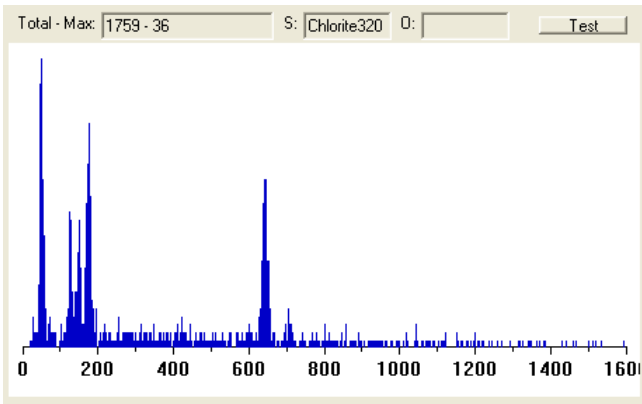


Figure 2: An example of X-ray spectra from a mineral sample with 1759 hits (back-scattered electron count), from a 320ms scan time.

Spectral imaging and analysis provides a wealth of data analysis tasks and challenges for the machine learning and neural network communities (for an overview of applications in the domain of remote-sensing imaging, see [1]). The mineral identification system described above is relatively specialized and few applications of machine learning techniques to this particular problem have been reported. Rui-

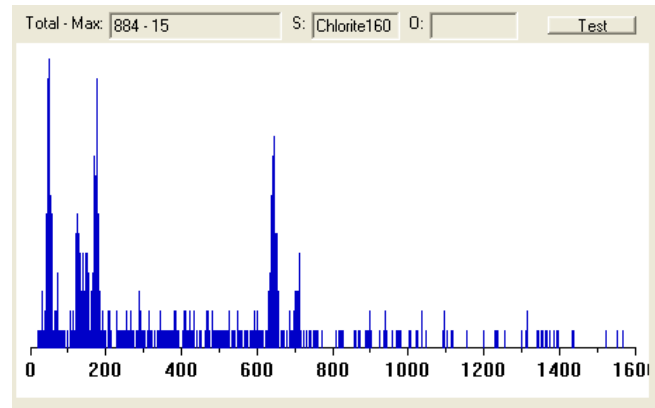


Figure 3: X-ray spectra for the same sample used to produce Fig.x, but with only 884 hits from a 160ms scan time.

sanchez et. al [2] use a 3D (15-15-15) Kohonen neural network to classify minerals into 12 different classes on the basis of 15 different intensity levels in X-ray spectra. Very good results are reported on a small data set (100% classification accuracy after training on 94 samples, measured on a test set of 95 samples). In a similar study, Magallanes and Vazquez [3] classify 19 different types of steel on the basis of X-ray data using a Kohonen network. Perfect (100%) accuracy was also obtained on 95 test samples, after training with 95 more samples. Mayo et. al [4] hold a patent on a system which uses multi-layer perceptron neural networks to identify explosive substances in material using X-ray data. As well as raw intensity data, a set of user-intuitive features and a set of “cepstrum” coefficients (based on a Fourier transformation of the data to the frequency domain) are utilized.

In this paper we discuss a much larger mineral classification problem in terms of the number of intensity bands measured (dimensionality of the raw data) and the number of data samples. Additionally, the level of noise present as a result of the data collection procedure is of particular interest, as discussed in the following section.

3. EXPERIMENTAL METHODOLOGY

Raw X-ray spectrum data is obtained in the form of a histogram (e.g Figs.2 and 3 above). X-rays are detected across 2000 different channels (wavelengths) and in each channel an intensity value is obtained. As mentioned above, the intensity values are dependent on the time interval for each measurement. Longer times produce spectra that are more representative of the “true” characteristics of the given mineral, while for data collected at shorter time intervals, noise becomes more significant (being a more significant proportion of the total intensity measurement across all channels). For our experiments 100 spectra were collected for each

Table 1: Percentage of variance in the mineral data captured by Principal Components.

No. PCs	1	2	3	4	5	10	15
Variance	29.1	36.7	42.2	46.2	48.7	54.2	54.6

mineral at 20 different time intervals (from 150ms to 340ms at 10ms intervals), resulting in 2000 complete spectra (2000-dimensional data points). Each sample was one of 10 different mineral classes: quartz, pyrrhotite, pyrite, pentlandite-pyrrhotite (i.e a commonly found mixture in samples), pentlandite, magnetite, galena, chlorite, chalcopyrite and apatite. Hence the final raw dataset contained 20000 points.

The magnitude of components of each data point varies according to the time interval used during data collection. This effect was removed by normalizing each raw data point in the range [0,1]. Because the data is of very high dimensionality, two pre-processing methods were employed to reduce its dimensionality and produce good features for classification prior to training.

Firstly, a set of explicit features was identified based on human experience and inspection of the spectra. One obvious feature is the peaks of the spectra; the energy levels and corresponding intensity levels from the five largest peaks were selected to represent this information. In addition, the spectra was divided into 10 equal energy bands. From each band the centroid and average intensity were used, making a total of 30 variables.

Secondly, Principal Component Analysis (PCA)[5], see also [6] was conducted on the (normalized) raw data. The cumulative percentage of variance in the data that was captured by (up to) the first 15 principal components is shown in Table 1. These 15 principal components were used as a second set of features for classification. This projection of the 2000-dimensional data into 15 dimensions results in a significant loss of variance information. Nevertheless, it is expected that features derived from PCA are likely to differ substantially from the explicit feature set. Combining the explicit and PCA features, the networks used had a total of 45 inputs.

A number of different neural networks were employed on this data. Firstly, multi-layer perceptron (MLP) networks were implemented using a single hidden layer of logistic sigmoidal units, together with a layer of 10 output units with a softmax activation function. The cross-entropy error function was used to train the networks. The dataset was divided into 60% (training), 20% (testing) and 20% (validation) subsets. Training was halted using a simple early stopping method (i.e when error was observed to increase on the validation set). Three different training algorithms were used: backpropagation, Quasi-Newton with BFGS search-direction update [6, 7] and quickprop. Reasonable values for learning rate parameters were obtained in preliminary

testing; the results reported below used a value of 0.3 for backpropagation (with a momentum term also equal to 0.3) and a value of 0.1 for quickprop. Weights were initialized randomly in the range $[-1, 1]$.

A Kohonen SOM network was also implemented, using LVQ to produce supervised learning-classification [6]. Note that the SOM has a natural feature of producing an “unknown” classification when a test input is not sufficiently close to an identified output/cluster centre. Again, the learning parameters were determined by trial-and-error: training ran for 200 epochs, and over this time the learning rate was reduced from 0.9 to 0.05 geometrically.

4. RESULTS

We experimented with a number of different MLP configurations. A sample of these results are shown in Table 2. Overall, it proved relatively easy to produce networks that classified 100% of the test data correctly (the exceptions shown in the table being the Backprop (3) and the Quasi-Newton (3,5,7) nets). The Quasi-Newton algorithm was consistently unable to reach a low error value, presumably due to local minima or an ill-conditioned error surface. It was found that 5 hidden units was enough to obtain perfect performance on the test data, though good results were maintained for larger numbers of hidden units. Typically 100-200 epochs were completed before performance on the validation set halted the training process.

The SOM network produced an average error of 0.518 and a standard deviation of 0.008 over ten training runs. This resulted in 94.6% correct classification, 2.7% incorrect and 2.7% unknown on average. The performance of the SOM was worse than the MLPs in general, though more extensive experiments would be required to qualify this observation.

In normal day-to-day operation, the simple statistical tests used are able to produce near-perfect performance. However in that situation the data used is produced at scan times not below 300ms. For comparison, we performed the statistical tests on a subset of data collected at short scan times (150-180ms). This data was a subset of that used to train the neural networks above. Note also that the statistical tests work with the raw spectra data directly. The results are shown in Table 3. It is clear that many pyrrhotite samples are incorrectly classified as the pentlandite-pyrrhotite mixture, and that the test is unable to significantly identify other samples (notably some chlorite samples).

5. DISCUSSION

We have shown the ability of neural networks to provide good results on a large-scale mineral classification task. The

Table 2: Classification performance of methods used on mineral data.

Algorithm	No. Hidden Units	Error (ave.)	Error (std. dev.)
Backprop	3	0.771	0.312
Backprop	5	<0.001	<0.001
Backprop	7	<0.001	<0.001
Backprop	13	<0.001	<0.001
Quasi-Newton	3	0.617	0.421
Quasi-Newton	5	0.204	0.154
Quasi-Newton	7	0.261	0.146
Quasi-Newton	13	0.083	0.066
Quickprop	5	0.011	0.024
Quickprop	7	<0.001	<0.001
Quickprop	13	<0.001	<0.001

Table 3: Classification performance of currently-used statistical tests on noisy data.

Mineral Name	Classification	Classification without unknowns
Quartz	399	400
Pyrrhotite	209	209
Pyrite	400	400
Pentlandite-Pyrrhotite	591	591
Pentlandite	400	400
Magnetite	400	400
Galena	400	400
Chlorite	339	400
Chalcopyrite	400	400
Apatite	400	400
Unknown	62	0

pre-processing, feature extraction and classification methods used lead to reliable performance on data having a significant amount of noise present due to short scan times. Shorter scan times mean an increase in the efficiency of the mineral analysis procedure.

We were primarily interested in the feasibility of neural network techniques for performing this large classification task adequately. Therefore a number of different methods were used, without a rigorous experimental comparison being performed between them. Preliminary experiments indicated that using both the explicit features and the PCA features provided better performance than using either feature set individually. Further experiments and alternative feature extraction techniques are required to explore this aspect of the application.

The inclusion of an "unknown" classification category deserves particular attention in future work, particularly if it can be integrated into a practical analysis method (e.g. it may be possible to manually examine or rescan samples that are classed as unknown by a first-level classifier. It is also possible to increase the number of data samples and the number of mineral types to be identified. The main thrust of future work is to determine how much sample scan times can be

reduced before the impact on neural network classification becomes significant. Data needs to be generated to facilitate such work.

6. REFERENCES

- [1] E. Merényi, "The challenges in spectral image analysis: an introduction, and review of ANN approaches," in *Proceedings of the 7th European Symposium on Artificial Neural Networks, ESANN99*, M. Verleysen, Ed., 1999, pp. 93–98.
- [2] I. Ruisanchez, P. Potokar, and J. Zupan, "Classification of energy dispersion X-ray spectra of mineralogical samples by artificial neural networks," *Journal of Chemical Information and Computer Sciences*, vol. 36, pp. 214–220, 1996.
- [3] J. F. Magallanes and C. Vazquez, "Automatic classification of steels by processing energy-dispersive X-ray spectra with artificial neural networks," *Journal of Chemical Information and Computer Sciences*, vol. 38, 1998.

- [4] W. E. Mayo, Z. Kalman, M. C. Croft, J. Wilder, R. Mammone, and A. B. Fineberg, "Analysis methods for energy dispersive X-ray diffraction patterns," Tech. Rep. US Patent no. US6118850, Rutgers, The State University, 1999.
- [5] I. T. Jolliffe, *Principal Component Analysis*, Springer-Verlag, New York, 1986.
- [6] S. Haykin, *Neural Networks A Comprehensive Foundation*, Prentice-Hall, Upper Saddle River, NJ, second edition, 1999.
- [7] R. D. Reed and R. J. Marks II, *Neural Smothing: supervised learning in feedforward artificial neural networks*, The MIT Press, Cambridge, MA, 1999.