Data Filtering for Automatic Classification of Rocks from Reflectance Spectra

Jonathan Moody
Computer Science Department
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA
jwmoody+@cs.cmu.edu

Ricardo Silva
Center for Automated Learning and Discovery
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA
rbas+@cs.cmu.edu

Joseph Vanderwaart
Computer Science Department
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA
joev+@cs.cmu.edu

ABSTRACT

The ability to identify the mineral composition of rocks and soils is an important tool for the exploration of geological sites. For instance, NASA intends to design robots that are sufficiently autonomous to perform this task on planetary missions. Spectrometer readings provide one important source of data for identifying sites with minerals of interest. Reflectance spectrometers measure intensities of light reflected from surfaces over a range of wavelengths. Spectral intensity patterns may in some cases be sufficiently distinctive for proper identification of minerals or classes of minerals. For some mineral classes, carbonates for example, specific short spectral intervals are known to carry a distinctive signature. Finding similar distinctive spectral ranges for other mineral classes is not an easy problem. We propose and evaluate data-driven techniques that automatically search for spectral ranges optimized for specific minerals. In one set of studies, we partition the whole interval of wavelengths available in our data into sub-intervals, or bins, and use a genetic algorithm to evaluate a candidate selection of subintervals. As alternatives to this computationally expensive search technique, we present an entropy-based heuristic that gives higher scores for wavelengths more likely to distinguish between classes, as well as other greedy search procedures. Results are presented for four different classes, showing reasonable improvements in identifying some, but not all, of the mineral classes tested.

Categories and Subject Descriptors

I.5.2 [**Pattern Recognition**]: Design methodology; I.5.4 [**Pattern Recognition**]: Applications

Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining San Francisco, California USA

General Terms

Algorithms, Experimentation, Performance

1. INTRODUCTION

Reflectance spectrometers have been used for identification of mineral composition of rocks and soil samples with varying degrees of success. This kind of spectrometer measures the amount of sunlight reflected by a rock or soil sample over a range of wavelengths. The reflectance obtained under different wavelengths can then be used to predict which minerals are present in that sample.

For instance, NASA intends to design robots for planetary exploration that would be sufficiently autonomous to interpret spectrometer data and report only the results back to Earth. Robots equipped with automatic classifiers of rocks would also be useful for automatically planning which different regions of a geological site would be more promising for prospecting certain classes of minerals in a more efficient way.

The data sets collected by spectrometers consist of levels of reflectance intensity of a given rock at different wavelengths. The intensity data are typically measured relative to a reference surface in order to be invariant with respect to the total amount of sunlight in the environment.

The usual approach taken by someone interested in building a predictive model out of this data is running a regression model for each rock or soil sample, where the dependent variable is the reflectance intensity of the unknown rock and the independent variables are the reflectance intensities of a variety of different pure minerals that are possible components of the rock, measured over the same wavelengths. Libraries of such pure mineral spectra exist; in particular, the Jet Propulsion Laboratory has produced a library of spectra for 135 different pure minerals [6], each containing reflectance intensities for 826 different wavelengths between 0.4 and 2.5 μm.

Assuming that the intensity of the rock is a linear combination of the intensity of its components, a regression model is built using each reflectance value at a wavelength as a data point. Then, only those minerals whose coefficients on the regression model pass a given test of statistical significance are considered components of the rock. A successful learn-

ing algorithm should commit as few errors as possible, where an error is accepting a given mineral as part of a rock when this is not true, and rejecting a given mineral as part of a rock when in fact it is.

Ramsey, *et al.* present evidence in [13] that a modified Bayesian network learning algorithm, the PC algorithm [15], performs better than simple linear regression for classifying carbonates. The PC algorithm tests partial correlation between a rock spectrum and different subsets of the library of mineral spectra, eliminating library spectra (hypothesized components) which are not correlated with the input. The advantage over regression is a more refined search that calculates partial correlations conditional on reduced subsets of the remaining variables, instead of considering all of them at once, as in regression. The remaining library spectra are assumed to be components of the input, and a classification can be performed based on the minerals that were not discarded, as explained in the previous paragraph. All of our work described below is based on this classification procedure.

2. DESCRIPTION OF THE PROBLEM

Experiments with the specific class of carbonates have shown that restricting the input of the PC algorithm to a smaller region of the spectrum can improve accuracy. In particular, a region suggested by prior expert knowledge (a region used by experts to identify carbonates) produces much better results than allowing the algorithm to consider the entire spectrum. In other words, the filtered spectrum does not include noninformative or noisy data that could confound mineral identification. This is a promising result that arguably can be extended to other classes.

Carbonates show a very typical curve on the spectral region between 2.0 and 2.5 μm, which motivated the scientists to focus on this region. However, coming up with a good range of wavelengths is not an easy task because little is known for other mineral classes. No automated method has been applied by the authors of [13] to find subintervals that would be more appropriate for identifying given classes and subclasses of minerals.

Our goal is to find intervals of the spectrum, specific to each class of minerals, for which the PC algorithm performs better than the same algorithm using the entire spectrum. This is a search problem that complements other data preprocessing issues described in Section 4.We tried several methods, a collection representative of both heuristic and computational intensive approaches that also bear relation with feature selection techniques.

3. DATA FILTERING TECHNIQUES

Finding an appropriate subset of the spectrum range can be cast as a problem of search among the space of possible subsets. Since we have over 800 available channels, an exhaustive search is infeasible. Also, a larger number of evaluated candidates increases the chance of overfitting [2]. One must decide how to trade-off the complexity of the search space depending on the chosen search algorithm, the available computational resources, and the amount of data available.

By the terminology used in feature selection research, as described in [9], we are basically building wrappers over the PC algorithm. Four algorithms were tried: a computationally demanding genetic algorithm, two greedy hill-climbing algorithms and a simple grid search strategy over a rather reduced number of parameters of a customized evaluation function.

The data filtering methodologies described here should be applied to each class of minerals at a time, since a interval that is suitable to one class is unlikely to be useful to other. Each experiment is therefore a binary classification problem.

One general property of the data that is assumed in this work is a relative locality of importance for the reflectance signal. In other words, the informative spectrum for each rock and mineral must be smooth enough so that grouping the whole range into a reasonably small number of subintervals does not harm the predictive accuracy of the signal.

3.1 Genetic algorithm

A genetic algorithm is an algorithm for combinatorial optimization [5], which is directly related to the task of finding useful subsets of the spectra. The most straightforward representation of a candidate is through a string of 826 bits, where a positive bit represents that the respective channel will be used. However, due to the reasons explained in the beginning of this section, we divided the spectrum into a fixed number of blocks, each represented by a bit. Thus, all channels in the same block are selected or not selected at the same time.

The evaluation function is very time-consuming: it consists in running the modified PC algorithm over a whole set of rock samples. The fitness of a candidate is the proportion of rocks that are correctly classified as containing or not containing the respective mineral. On our available implementation, it takes about 30 seconds to evaluate a single candidate feature mask on a Pentium III 733MHz processor.

3.2 Bitwise hill-climbing

We also used a greedy, hill-climbing algorithm that uses the same representation for search states and the same evaluation function. On the initial state, all bits are activated. The next states are generated from the current state by setting to zero one of the currently activated bits. If the current candidate has n activated bits, it will generate n new candidates. The candidate with the highest evaluation value is chosen to be the next state.

3.3 "Peeling" algorithm

This is another greedy algorithm that is also used for rule induction over continuous/ordered attributes [4]. It consists of trimming the extremes of an interval by some percentage of the data and evaluating the new interval obtained. A typical strategy starts with the complete interval and, at each subsequent step, generates three new candidates: the current interval with the bottom α% of the ordered data discarded, the current interval with the upper α% of the ordered data discarded, and an interval constructed by dropping the bottom and upper $\frac{\alpha}{2}$% from the current interval.

The underlying assumption of this algorithm is that interesting intervals are continuous. Unlike the previous algorithms, all selected subintervals are of the form $[a, b]$, where a and b are points of the original interval. It may clearly result in suboptimal selections, at the advantage of being much less time demanding.

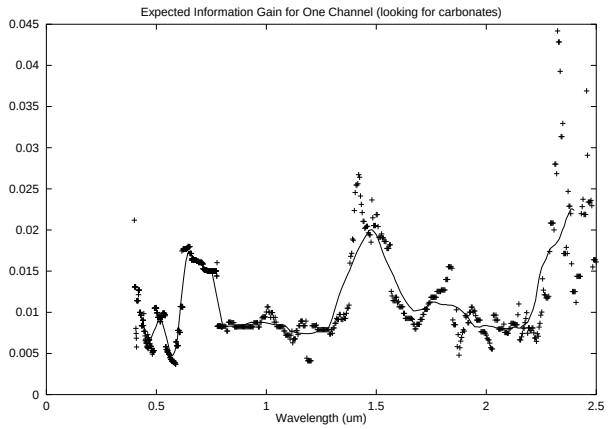

Figure 1: The average information gain per each channel using four bins with respect to the carbonate class. Notice that the highest gains lie on the upper region of the spectra, as suggested by expert knowledge. The solid line is an interpolation over these points, added as an aid to visualization.

3.4 Information gain heuristic

A more straightforward approach would be to construct a "relevance" heuristic, rank the channels accordingly, and select those with relevance above a threshold. Intuitively, we wish to discover those channels that carry a large amount of information relevant to the question of whether a certain class of minerals is present. Therefore, we used information gain, a quantity based on entropy, for our relevance heuristic.

The information gain algorithm for selecting a channel mask is as follows. For each channel, we divide the intensity range into some number of bins. Then for every spectrum in the reference library we look at the intensity at the current channel and take note of which bin it occupies and whether or not it is a member of the target class. When we have finished doing this for a given channel, we calculate the fraction of samples in each bin that are in the desired class; this number is used to calculate an entropy value for that bin. A weighted sum of the entropies of the bins gives the expected entropy given the intensity of a particular channel; subtracting this from a constant gives the *expected information gain* associated with that channel.

When we have calculated the expected gain for each channel, we create a channel mask by looking for intervals where the expected gain is higher than average. Specifically, we divide the spectrum into blocks and calculate the average expected gain in each block. Then blocks whose average expected gain exceeds the global average by some margin are selected for use in classification.

Under this technique, we optimize the number of bins and threshold parameters by performing a grid search over a given interval of possible values. The grid search consists in evaluating each pair of values (*number of bins*, *threshold*) over a predefined interval. The selection that gives the best classification accuracy for the training set is used. Figure 1 shows how it is possible to visualize promising regions using this evaluation function.

4. EXPERIMENTS

For our experiments we used the NASA Jet Propulsion Laboratory (JPL) data set as a reference library, attempting to classify the rocks in the Johns Hopkins University (JHU) data set, a library of reflectance spectra for a variety of solid and powdered rock samples. Each mineral on JPL was measured with different grain sizes. We used the largest grain size, which should give a closer approximation to rocks found on test fields. The data set was processed to treat issues such as making measures of relative reflectance with respect to a white surface, and so subtract the effect of environment luminosity. It was necessary to interpolate the measures of the JHU spectra in order to match the same wavelengths found on the JPL library.

Also, most features of spectra which are diagnostic of the chemical structure of minerals are small scale "dips", or deviations, from the overall background shape of the spectrum, with a width on the order of 1 to 50 μm. By taking the hull difference of a spectrum [6], variations due to the large-scale shape of the spectrum are reduced or eliminated and variations due to these smaller, typically more diagnostic, variations are enhanced. The idea is that a hull is fitted to the spectrum and then the differences between the spectrum and the hull for each wavelength are recorded. This set of differences as a function of wavelength is the hull differenced data. On the following experiments, we refer to data treated by the hull difference process as the "processed data", while "raw data" will refer to spectra without this modification. For further information on these data sets, see Ramsey *et al.* [13].

We performed experiments using four of the mineral classes available in the JPL library. These minerals were chosen according to the number of rocks present in the JHU data set that were reported to have these minerals: it would be unreliable to try to find intervals for a class underrepresented in the available data. Among all 192 JHU rocks, 92 have carbonates, 121 have phyllosilicates, 100 have oxides and 84 have inosilicates.

Tables 1 and 2 show the results for running the modified PC algorithm using the intervals selected by variations of each algorithm described on the previous sections. For each mineral class, we ran a five-fold cross-validation. The accuracy measure is the number of correctly classified rocks (true positives plus true negatives) divided by the number of rocks on the corresponding sample.We opted for 5 folds instead of the usual 10 folds because:

- the genetic algorithm is computationally intensive;
- we wanted a reasonable amount of data on both training and test sets. Using a high number of folds can in fact lead to worse generalization estimates when we have few data points and the prediction error is high, as it is typical of this domain [14].

The whole spectrum interval was divided in 15 subintervals of equal size. For the genetic algorithm, for example, this means we are using 15 genes per individual. The reason for this choice was to allow approximately 50 wavelengths per cell and to avoid introducing too much variation on the search for selected intervals. A more extensive experimental analysis could include this choice as a parameter to be optimized.

Table 1: Mean and standard deviation for classification accuracy (in %) obtained by using the raw data. GA stands for genetic algorithms, HC for the general hill-climbing algorithm, PEEL for the "peeling" procedure and IG is the label for the information gain results. The first column represents the results obtained when all the spectrum is used.

	None	GA	HC	PEEL	IG
Carbonates	56.3 ± 8.6	64.0 ± 5.0	62.0 ± 4.2	52.7 ± 13.0	66.2 ± 6.9
Inosilicates	61.4 ± 8.3	69.7 ± 6.7	65.5 ± 7.5	60.0 ± 7.5	70.0 ± 8.1
Oxides	56.3 ± 6.7	58.7 ± 6.7	49.9 ± 3.1	48.9 ± 7.5	56.3 ± 4.0
Phyllosilicates	56.3 ± 7.5	57.1 ± 3.8	50.2 ± 5.6	55.7 ± 2.1	50.0 ± 5.2

Table 2: The results obtained for the processed data.

	None	GA	HC	PEEL	IG
Carbonates	63.4 ± 7.0	68.3 ± 3.9	66.1 ± 5.3	65.5 ± 5.2	61.4 ± 7.6
Inosilicates	61.4 ± 4.3	66.3 ± 4.3	68.4 ± 4.0	66.1 ± 11.1	60.0 ± 10.9
Oxides	49.3 ± 6.1	48.5 ± 1.4	53.7 ± 9.6	50.0 ± 5.1	50.9 ± 5.9
Phyllosilicates	54.1 ± 6.0	52.7 ± 6.1	53.7 ± 7.5	53.7 ± 6.8	59.4 ± 3.1

For the genetic algorithm, we used 35 individuals. The training proceeded for at most 40 generations. In all cases, by the last generation the pool of individuals was almost completely dominated by copies of a single individual (and in many cases, all individuals were identical), suggesting that further optimization would not improve the result obtained significantly. The code of the genetic algorithm was adapted from [10], with its default parameters: 0.6 chance of crossover and a low (0.0001) chance of mutation.

We also used cached statistics to scale up the algorithm: instead of passing through all the data points when computing an element of the correlation matrix (as required by the PC algorithm), we precomputed the summations and inner products of variables for the data falling under each block. Getting a new element of the correlation matrix required only a pass over these cached statistics. This procedure reduced the computational time by over 30%.

For the standard hill-climbing search, we adopted the following stopping criterion: as a trade-off to avoid bad local maxima without searching till the last state, the search stopped when we did not get improved results for five consecutive states. The best selection on this search path was the output.

For the peeling algorithm, we used a value of 5% for α. We used the same stop criterion applied on the previously described hill-climbing technique.

To find appropriate parameter values for the entropy heuristic, each training set was used to evaluate the masks produced by several different parameter settings. In particular, all possible combinations of 3, 4 or 5 bins with thresholds of 0.1, 0.2, 0.3, 0.4 or 0.5 standard deviations above the mean gain were tried. For each training set, the mask that produced the best accuracy was selected as the optimal mask and its fitness was measured with the corresponding test set.

Using the interval selected by experts for carbonate classification, we get an accuracy of 67.7% for the raw data and 66.1% for the processed data. By comparison with the results obtained, it is clear that some of our approaches were overall able to find selections with similar performance, but unable to significantly improve over it. We should not forget, however, that these results were attained without relying on background knowledge and hence provide evidence that for cases where this knowledge is actually unavailable this set of approaches can be a useful tool.

The data pre-processing by taking hull differences can help in some occasions, as it was the case for carbonates. For the inosilicates, however, reasonable better results were obtained using the raw data. As any smoothing procedure, it can be useful in some situations, but not always. The information gain heuristic proved especially sensitive to this technique.

While our performance on carbonates and inosilicates improved relative to the baseline of enabling all channels, we got unimpressive results with phyllosilicates and oxides. It was expected that for some classes the reflectance spectrum information is not sufficient to provide a good separation between those classes and the remaining ones. In this ill-defined situation, data filtering would not be able to help much. Notice that taking hull differences actually harmed the predictive accuracy of our classifier for these cases.

The variance of the results is due not only to sample variance, but also to the variance of the underlying classifier, a simplified PC algorithm. Depending on the data selection algorithm, we have also small or big variance on the selected intervals. Figure 2 depicts the number of times each cell was chosen for some of the algorithms on the raw data. Due to its simplicity and reduced number of parameters, the entropy heuristic was the most stable.

Also, it is interesting to point out that simple algorithms such as the hill climbing algorithms were competitive when compared with the genetic algorithm. Since our data sets were small, computational time was not a major issue, but in applications where a larger number of measurements is performed, they may be viable solutions.

Under the assumption that cross-validation is a valid mechanism for estimation of generalization error, these experiments can be used to decide which algorithm should be trained in the whole data set in order to be used in a real world application. For example, the genetic algorithm and information gain can be applied to the whole JHU data set and generate one mask for future classification of inosilicates using the raw data, since they have very close accuracy, but are both reasonably superior to the non-filtered data. The model generated from the whole data set would be the final model[1].

[1]One alternative would be to combine masks generated for

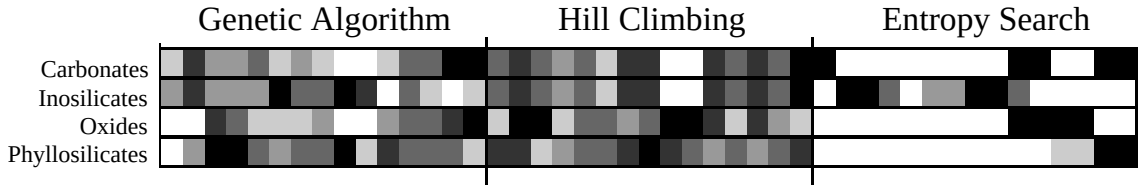

Figure 2: This figure depicts the number of times each of the fifteen cells was chosen across the five training sets used in the cross-validated experiments. A cell that is totally black was chosen every time, while a white space represents a cell that was never chosen. It is interesting to notice that the entropy search was the most stable, but that the exact selection of a given mask is not required for reasonably good generalization.

5. RELATED WORK

Ramsey, *et al.*[13] discuss extensive experiments analyzing the performance measure of different classification algorithms, including decision trees that already carry out an entropy-based selection of data. However, none of these results performed better than the baseline for the PC algorithm with no wavelength selection depicted in Tables 1 and 2. Our approach can be interpreted as an effective combination of different inductive biases that works better than standard decision trees for the case of entropy selection.

The techniques applied in this work are related to the areas of feature selection and data cleaning. Wettschereck, Aha and Mohri [17] formulate a framework for feature weighting methods under the context of lazy learning.[2] Even though in a strict sense the wavelength channels are in fact rows of our data set, not attributes, in principle one can use these techniques to weight the relevance of each data point (or intervals for practical purposes). According to the categories of Wettschereck *et al*'s framework, the genetic algorithm and hill-climbing approaches would be classified as having:

- a performance bias, since we use the actual results of classification for deciding the selection;
- a binary weight space (*i.e.*, 0/1 weights);
- a transformed representations, since we divide the data into blocks;
- a global weighting, because the same intervals are selected for all minerals;
- knowledge-poor, since we did not use prior knowledge in our experiments. Hull differences help in some situations.

The performance bias is also commonly described as a wrapper approach [9]: our selection policies use the modified PC algorithm as a black box that outputs a measure of performance.

each fold and count the different votes in an independent test set. However, since our main purpose in this paper was to provide a more reliable comparison of different approaches with respect to the baseline achieved by using the whole wavelength range, and our sample of rocks was small, we opted to use the whole sample for cross-validation.

[2]In this survey, the authors do not compare different batch optimization techniques: among this class of learning algorithms, only a gradient-based one is used.

Unlike general feature selection problems, we do not have the concern of selecting features that present fewer missing values on the available data bases, nor do we have to consider which are more expensive to measure (e.g., some medical exams for diagnosis problems). That makes our fitness function even simpler than most ones used in feature selection literature [11, 16, 18]. These approaches are virtually identical to the genetic algorithm for data selection described in this work, where the difference is mainly a more complicated evaluation function. Demiroz and Guvenir [3] also describe mechanisms for learning continuous weights between 0 and 1, which arguably are not very useful for our problem, where we have too little data to accomodate such a precise tuning of parameters.

In contrast, the information gain heuristic operates as hybrid between a wrapper and a filter approach. The filter approach applies for each feature a measure of importance that is independent of the learning algorithm that will be used. Hall [8] provides a comparison of filters and wrappers, as well as an overview of feature selection. He favors the filter approach due to its much higher scalability, but in his discussion it is mentioned that ideally the features themselves should be a function of the bias of the learning algorithm that will be used. An intermediate approach such as using the entropy measurements to search for a combination of prominent intervals, which can then be successfully used by the modified PC algorithm, is a way to trade-off these issues.

Entropy measures are commonly related to the degree of unexpectedness of a pattern, and such a characteristic has been explored for data set cleaning. Guyon, Matic and Vapnik [7] describe different ways of using information theoretical measures to identify outliers or highly informative examples. Data points are ranked according to information gain and then submitted to a expert that will classify them as outliers or representative examples. Guyon *et al.* warn against the risk of getting improved results during training by dropping the most difficult examples and then achieving bad generalization accuracy.

Another application of information theoretical measures for data cleaning is discussed by Pyle [12], where it is also described how to find ill-defined regions of a function by checking symmetries between the input and output variables. This specially affects inverse function estimators. Pyle also describes what he calls "attention processing" of data: how to efficiently perform data surveying in a large combinatorial space of potentially problematic regions of the data.

6. SUMMARY AND FUTURE WORK

The most important lesson from this study is that signal-processing algorithms can benefit from search procedures that automatically decide which parts of the signal must be taken into account when making a decision. From the initial hypothesis that the behavior observed in carbonates could be replicated by automated procedures, our experiments were able to achieve a similar performance from scratch. An interesting experimental hypothesis is applying a similar framework for other domains. It must be emphasized that this approach is a complement to other smoothing procedures, as illustrated by running experiments along with the application of hull differences. Even though our main techniques are straightforward, to the best of our knowledge there are no experiments exploring similar ideas for this problem. One of the main goals of this work is pointing out simple yet effective alternative approaches of tackling spectroscopic analysis of material composition and possibly component detection problems of other blind source separation domains.

However, sampling variability may be a concern and the fact that the underlying classifier provides its own source of variability may amplify this problem. Kohavi and George [9] report that feature selection algorithms may overfit easily. Approaches to minimize this problem and perform more reliable performance assessment include resampling techniques such as bootstrapping [1].

For example, it may be possible that more robust masks of selected intervals can be obtained by the combination of different masks. One simple policy is obtaining multiple masks by resampling and then giving to each bin a weight proportional to the number of times each one appears.

This improved reliability does not come for free, and more computational time is required. For instance, Punch *et al.* [11] reported experiments with genetic algorithms for feature selection that took 14 days. In this case, one might not want genetic algorithms, since the difference in accuracy when compared with other approaches may not be great enough to justify the extra effort. Alternatively, one could just gather more labelled data. For example, the U.S. Geological Survey has produced a data set of about 400 labelled rocks. However, some of these labels are wrong, or inconsistent with the classification scheme of the JPL data set. Before combining these data with the JHU data, additional preprocessing would be required.

Concerning the variability of the underlying classifier, a straightforward way to alleviate this problem is to modify the evaluation function of the search algorithms to consider the outcome of an ensemble of classifiers. Future experiments may include this approach.

7. ACKNOWLEDGMENTS

We would like to thank Joseph Ramsey for his valuable help during the preparation of this work.

This work was partially funded by the National Aeronautics & Space Administration, grant number NAG5-9309.

8. ADDITIONAL AUTHORS

Clark Glymour, Philosophy Department, Carnegie Mellon University and Inst. for Human and Machine Cognition, University of West Florida, email: cg09@andrew.cmu.edu

9. REFERENCES

[1] Cohen, P. (1995). Empirical Methods for Artificial Intelligence. MIT Press.

[2] Cohen, P. and Jensen, D. (1997). "Overfitting Explained". Preliminary Papers of the Sixth International Workshop on Artificial Intelligence and Statistics, 115-122.

[3] Demiroz, G. & Guvenir, H. (1996). "Genetic algorithms to learn feature weights for the nearest neighbor algorithm". In Proceedings of BENELEARN-96, 117-126.

[4] Friedman, J. & Fisher, N. (1999). "Bump hunting in high-dimensional data". Statistics and Computing, 9, 123-143.

[5] Goldberg, D. (1989). Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley.

[6] Grove, C. I., S. J. Hook, and E. D. Paylor II. 1992. "Laboratory Reflectance Spectra of 160 Minerals, 0.4 to 2.5 Micrometers". JPL-Publication 92-2.

[7] Guyon, I.; Matic, N. & Vapnik, V. (1995). "Discovering Informative Patterns and Data Cleaning". In: Advances in Knowledge Discovery and Data Mining, 181-204. AAAI Press.

[8] Hall, M. (1999). Correlation-based Feature Selection for Machine Learning. PhD thesis, University of Waikato, Computer Science Department. Hamilton, New Zealand.

[9] Kohavi, Ron & John, George H. (1998) "The Wrapper Approach". In H. Liu and H. Motoda (Eds.), Feature Selection for Knowledge Discovery in Databases. Springer-Verlag.

[10] Masters, T. (1993). Neural Network Recipes in C++. Academic Press.

[11] Punch, W.; Goodman, E.; Pei, M.; Chia-Shun, L.; Hovland, P. & R. Enbody (1993). "Further research on feature selection and classification using genetic algorithms". Proceedings of the International Conference on Genetic Algorithms 93, 557-564.

[12] Pyle, D. (1999). Data Preparation for Data Mining. Morgan-Kaufmann.

[13] Ramsey, Joseph; Gazis, Paul; Roush, Ted; Spirtes, Peter; Glymour, Clark. (2000). "Automated Remote Sensing with Near Infrared Reflectance Spectra: Carbonate Recognition". Dept. of Philosophy, Carnegie Mellon University.

[14] Sarle, W. (2000). The Neural Network FAQ.ftp://ftp.sas.com/pub/neural/FAQ.html.

[15] Spirtes, P., Glymour, C., & Scheines, R. (2000). Causation, Prediction and Search, 2nd edition. MIT Press.

[16] Vafaie, V. & DeJong, K. (1998). "Feature space transformation using genetic algorithms". IEEE Transactions on Intelligent Systems, 13(2), 57-65.

[17] Wettschereck, D., Aha, D. W., & Mohri, T. (1997). "A review and comparative evaluation of feature weighting methods for lazy learning algorithms". Artificial Intelligence Review, 11, 273-31.

[18] Yang, J. & Honavar, V. (1998). "Feature subset selection using a genetic algorithm". In Feature Extraction, Construction, and Subset Selection: A Data Mining Perspective. Motoda, H. and Liu, H. (Ed.) New York: Kluwer.