Applying Incremental EM to Bayesian Classifiers in the Learning of Hyperspectral Remote Sensing Data

X. Rosalind Wang,
Centre for Autonomous Systems
ACFR (JO4), The University of Sydney
NSW 2006, Australia
Email: r.wang@cas.edu.au

Adrian J. Brown
Australian Centre for Astrobiology
Macquarie University
NSW 2109, Australia
Email: abrown@els.mq.edu.au

Ben Upcroft
Centre for Autonomous Systems
ACFR (JO4), The University of Sydney
NSW 2006, Australia
Email: b.upcroft@cas.edu.au

***Abstract*— In this paper, we apply the incremental EM method to Bayesian Network Classifiers to learn and interpret hyperspectral sensor data in robotic planetary missions. Hyperspectral image spectroscopy is an emerging technique for geological investigations from airborne or orbital sensors. Many spacecraft carry spectroscopic equipment as wavelengths outside the visible light in the electromagnetic spectrum give much greater information about an object.**

The algorithm used is an extension to the standard Expectation Maximisation (EM). The incremental method allows us to learn and interpret the data as they become available. Two Bayesian network classifiers were tested: the Naïve Bayes, and the Tree-Augmented-Naïve Bayes structures. Our preliminary experiments show that incremental learning with unlabelled data can improve the accuracy of the classifier.

***Index Terms*— Bayesian networks, Incremental EM, Hyperspectral Imaging**

I. Introduction

Remote Sensing uses instruments to observe the electromagnetic spectra from an area under study. It has great implications in geology and mineralogy in that it can study a large area quickly and cheaply. In planetary geology and astrobiology, remote sensing is almost the only means by which we can learn the composition, environmental condition and other information about objects in the solar system. The exception being the Apollo Moon missions, the Mars Rover missions and a couple of other recent space missions.

Hyperspectral image spectroscopy is an emerging technique that obtains data over a large spectral bandwidth of an area. Spectral analysis is the extraction of quantitative or qualitative information from spectra based on the wavelength-dependent reflectance properties of materials [10]. Hyperspectral sensors are characterised by the very high spectral resolution that usually results in hundreds of observed wavelength channels. These channels permit very high discrimination capabilities in the spectral domain including material quantification and target detection. Recently, there has been an increased application of infrared reflectance spectroscopy in the field with Short-Wave Infrared (SWIR, light from 1300 to 2500 nanometre in wavelength) spectroscopy allows recognition of subtle mineralogic and compositional variation [11].

Presently, there is no single, universally accepted methodology for spectral analysis of remote sensing data. The most effective techniques, and those that are most readily transported from one field site or data set to another employ basic principles of reflectance spectroscopy.

Currently, the process of hyperspectral analysis is user intensive, requiring a large amount of data analysis, and expert input. The hyperspectral data is often represented as a "cube" of information where the layers of the cube are the image at the spectral bands. Systems are available for the simultaneous viewing of this information in the spatial and spectral domains [10]. The user is able to select points on the image and the program displays the spectrum at the specified location and the closest spectral match to it. The user then applies his/her own knowledge and other methods to interpret the data.

While most methods of spectral analysis require a large amount of knowledge and understanding in spectroscopy and field sites, there are some that do not and can be applied in a relatively straightforward manner. The most widely used of these methods is principle component analysis (PCA) [10]. PCA finds a correlation matrix between variables from which eigenvalues and eigenvectors are extracted. The amount of spectral variability contained in each component is given by the eigenvalue, and the relative proportion or contribution of each band to that component is given by the eigenvector. The method is scene specific, and thus unable to transport to different regions and environmental conditions.

In this paper, we present the Bayesian Incremental EM algorithm as described by Neal and Hinton [9] as a method for analysing hyperspectral data. The Bayesian method has the advantage of learning from data and improving its accuracy as information accumulates. The graphic nature of the Bayes network allows for the dependencies of the random variables, which can be the individual wavelengths in the hyperspectral data and any other relevant information, to be taken into account in the analysis.

This paper is organised as follows. In Section II, we present the data collected for the study. In Section III, we describe the procedure we used for analysing this data. Finally in Section IV, we present the experimental results and discuss the shortcomings of the algorithm, and any improvements that can be achieved in the future.

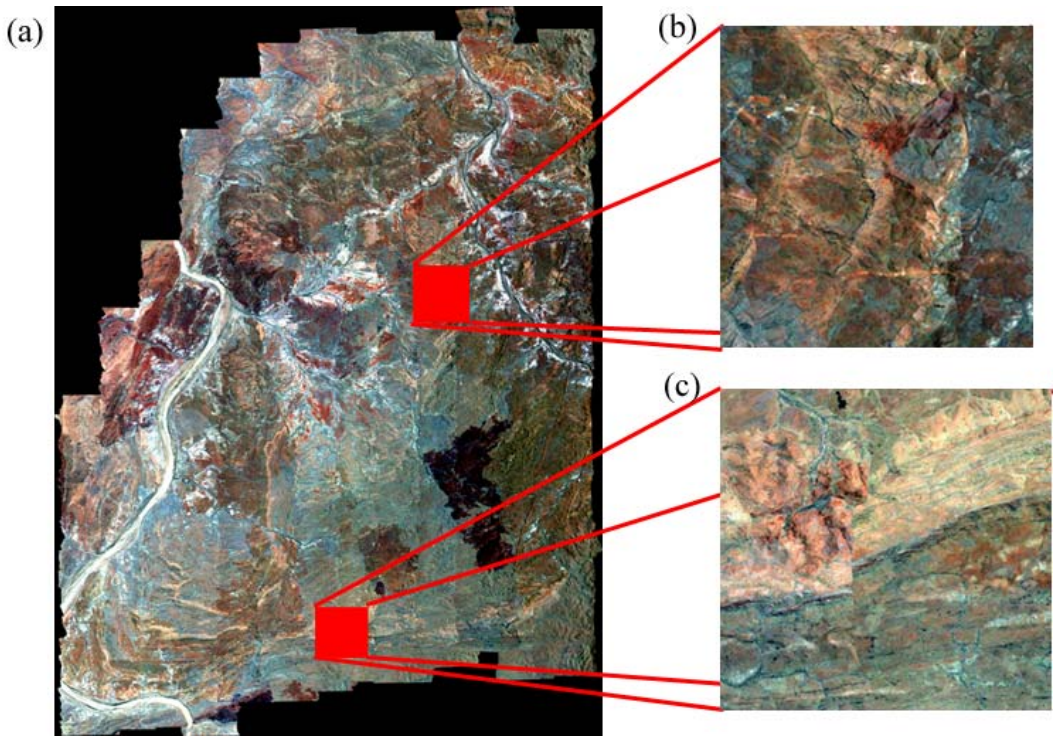

Fig. 1. (a) North Pole Dome dataset showing coverage of the Dome, north is up, image is approximately 27km across. Study region is indicated by red squares. (b),(c) Approximate study area images, size is 500×500 pixels.

II. Data collection

As part of a cooperative research project between the Commonwealth Scientific and Industrial Research Organisation (CSIRO) of Australia and the Australian Centre of Astrobiology at Macquarie University, a hyperspectral visible-near infrared and short wave infrared (400 - 2500 nm) dataset was collected over the North Pole Dome region of the East Pilbara in Western Australia. The dataset (Fig. 1) was collected using the 126 band HyMap instrument [4] at an altitude of approximately $2.5km$ covering approximately $600km^2$ with a resolution of $5m$ per pixel. The dataset was collected between 1000-1300 hours on 22 October 2002, at the end of the dry season in order to minimise vegetation coverage.

The North Pole Dome region of the East Pilbara Granite Greenstone terrain is covered by the volcanic rocks of the Warrawoona Group. These rocks have been dated at approximately 3.5 billion years old, making them similar in age and composition to flood basalts on the surface of Mars. Thus a perfect text ground for future spacecraft missions to Mars. Extensive hydrothermal alteration has taken place in at least two events prior to 3 billion years ago, it is also speculated that similar event could have taken place on Mars around the same time [12]. Hydrothermal systems are usually identifiable by zones of water-altered minerals such as muscovite, chlorite, pyrophyllite and kaolinite. These hydrothermal minerals generally contain the hydroxyl (OH) ion in their structure, which has very prominent spectral features in the short wave infrared range of the electromagnetic spectrum.

The two test areas for the method selected are shown on the right of Fig. 1. These particular areas were used as good exemplars of hydrothermally altered regions. The northern-most region is criss-crossed by volcanic veins surrounded by wide halos of hydrothermal alteration. The southernmost region shows evidence for hydrothermal alteration beneath an impermeable cap, indicated by the presence of the high temperature alteration mineral pyrophyllite.

III. Approach to the problem

Our approach to the problem of classifying the hyperspectral data into their respective mineral class is to use Bayesian Networks Classifiers. Bayesian Networks are graphical representations of multivariate joint probability distributions that exploit the dependency structure between distributions, describing them in a compact and natural manner [7]. In the case of image spectroscopy, the variables could be the data at each wavelength and classes of the data.

The procedure in analysing the hyperspectral image is:

1) Select pixels from the image to be used as training data. For each pixel, classify the spectrum according to expert knowledge to obtain a labelled training data set.
2) Using the labelled training data, learn the classifiers.
3) Apply the learnt classifiers on the entire image while utilising the incremental EM method on each pixel to learn from new unlabelled data, thus updating the parameter of the classifier.
4) Test the original and the new classifiers on a second

image, while also apply incremental EM to the new classifiers on the second image.

A. Bayesian network

In a Bayesian network, variables that are not linked directly in the graph are conditionally independent of each other. This independence can be exploited since the number of parameters needed to characterise the network is reduced. Thus it is possible to efficiently compute marginal probabilities of nodes in the structure given some evidence or observations of the other variables. For example, given the reflectance value at each observed wavelength, we can calculate the probability of a mineral's presence. Using the independence statements encoded in the network, the joint probability distribution is uniquely determined by these local conditional distributions [6], [8].

Formally, a Bayesian network for $\mathbf{U} = \{A_1, \ldots, A_n\}$ is a pair $B = \langle G, \Theta \rangle$. The first component, G, is a directed acyclic graph (DAG) whose vertices correspond to the random variables $A_1, \ldots, A_n$, and whose edges represent direct dependencies between the variables. The graph G encodes the set of independence assumptions of the variables. In our case, the set of variables are defined as $\mathbf{U} = \{C, A_1, \ldots, A_n\}$, where C is the class of the spectrum, and $A_1, \ldots, A_n$ are the reflectance values at each channels of the IR sensor.

The second component, Θ, represents the set of conditional probabilities that quantify the attributes of the network. It contains a parameter for each possible value a_i of the attribute A_i, and Π_{a_i} of Π_{A_i}, where Π_{A_i} denotes the set of parents of A_i in G. A Bayesian network B defines a unique joint probability distribution over $\mathbf{U}$ and is given by

$$P_B(A_1, \ldots, A_n) = \prod_{i=1}^{n} P_B(A_i | \Pi_{A_i}). \tag{1}$$

Learning the probabilities of attributes and the structure of a Bayesian network from data can be a form of unsupervised learning, in the sense that the learner does not distinguish the class variables from the attributes. The objective is to deduce a network (or a set of networks) that best describes the probability distribution over the training data $D = \{\mathbf{u}_1, \ldots, \mathbf{u}_N\}$ of instances of $\mathbf{U}$ [6]. In our case, the training data would be the reflectance values at each wavelength plus the classification of that spectra thus called labelled data. The objective is to recover the dependencies between the channels and the structure of the network, thus determining the classes of the spectrum recorded at the pixels.

1) Using Naïve Bayes as the classifier: The Naïve Bayes Classifier incorporates constraints in the graph. The main assumption underlying the classifier is that every attribute is conditionally independent from the rest of the attributes, given the state of the class variable [6].

In a Naïve Bayes Network, the class variable is the root of the network, and each attribute has the class variable as its unique parent, namely, $\Pi_C = 0$ and $\Pi_{A_i} = \{C\}$ where

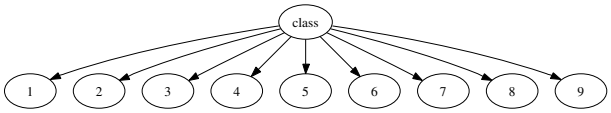

Fig. 2. The structure for the Naïve Bayes network for our example, each numbered node represent an IR channel. The wavelengths of each channel are respectively: 2093.2, 2111.2, 2129.0, 2146.5, 2163.9, 2181.3, 2198.3, 2217.1, 2234.1 nm.

$1 \leq i \leq n$. For this type of graph structure, (1) gives:

$$\begin{aligned} P(C, A_1, \ldots, A_n) &= P(C | A_1, \ldots, A_n) \\ &= \alpha \cdot P(C) \cdot \prod_{i=1}^{n} P(A_i | C) \end{aligned} \tag{2}$$

where α is a normalisation constant. The network structure for the Naïve Bayes classifier with the attributes mentioned above is shown in Fig. 2. We use maximum likelihood (ML) to calculate the network parameters.

2) Using TAN as the classifier: The second initial graph structure we used to test is the Tree-Augmented-Naïve Bayes (TAN) network, which encodes conditional dependencies between its attributes [6]. The Naïve Bayes structure is augmented with edges among the attributes when needed. In an augmented structure, an edge from attribute A_i to A_j implies that the influence of A_i on the assessment of the class variable also depends on the value of A_j.

Like the Naïve Bayes Network, the class variable in the TAN classifier is the root variable and thus has no parents. The attributes have the class variable as a parent like the Naïve Bayes, but each attribute will also have at most one other attribute as its parent. Thus each attribute can have one augmenting edge pointing to it. Mathematically, a Bayesian network B is defined to be a TAN model if $\Pi_C = \emptyset$ and there is a function π that defines a tree over $A_1, \ldots, A_n$ such that $\Pi_{A_i} = \{C, A_{\pi(i)}\}$ if $\pi(i) > 0$, and $\Pi_{A_i} = \{C\}$ if $\pi(i) = 0$.

The procedure for learning the augmented edges in the TAN classifier is based on the method by Chow and Liu [2], which applies the notion of tree dependence to Bayesian structures. In the learning process, the conditional mutual information (CMI) is found between each pair of the attributes as described by Friedman et al. The network is then optimised by finding a tree that maximises the sum of the CMI [6].

B. The Incremental EM algorithm

Correctly labelled data is always hard to obtain as it requires a vast amount of expert input. The application of the standard Expectation Maximisation (EM) algorithm [5] for learning parameters in a fixed parametric model however, allows the use of unlabelled data in the learning process.

The EM algorithm provides a general approach to maximum-likelihood parameter estimate when the observation can be viewed as incomplete data. EM is an iterative algorithm, each iteration of the algorithm consists of an expectation step (E-step) followed by a maximisation step (M-step). In the E step, the values of the unobserved latent, or hidden, variables

are essentially "filled in", where the filling-in is achieved by calculating the probability of these variables, given the observed variables and the current values of the parameters. In the M step, the parameters are adjusted based on the filled-in variables, which is basically that of the Maximum Likelihood estimate when all the variables are observed.

Mathematically, let x be the values of the observed variables and z those of the hidden variables, and the parameters of the variables are represented by θ. The steps of the EM algorithm are:

$$\begin{aligned} \textbf{E step} \quad & q^{(t+1)} = \arg\max_q \mathcal{L}(q, \theta^{(t)}), \\ \textbf{M step} \quad & \theta^{(t+1)} = \arg\max_\theta \mathcal{L}(q^{(t+1)}, \theta). \end{aligned} \tag{3}$$

The function $\mathcal{L}(q,\theta) \triangleq \sum_z q(z|x) \log \frac{p(x,z|\theta)}{q(z|x)}$ is a lower bound for the log likelihood of $l(\theta; x)$ for an arbitrary distribution $q(z|x)$.

Generally, the EM algorithm is applied to the entire training data set that has both labelled and unlabelled data. In the incremental version of the EM algorithm, the distribution of the variables are calculated and updated as new data becomes available. The M step is then performed to re-estimate the parameters before performing the E step for the next data point. To avoid looking at all the components of q in the M step, we summarise the complete data using a vector of sufficient statistics that can be incrementally updated.

Let the vector of sufficient statistics be $s(z,x) = \sum_i s_i(z_i, x_i)$, the iteration of incremental EM can be implemented by incrementally maintaining s, starting with an initial guess of $\tilde{s}_i^{(0)}$. The steps for the incremental version of the EM algorithm as given by Neal and Hinton [9] are:

$$\begin{aligned} \textbf{E step} \quad & \text{Choose some data item, } i, \text{ to be updated,} \\ & \tilde{s}_j^{(t+1)} = \tilde{s}_j^{(t)} \text{for } j \neq i, \\ & \tilde{s}_i^{(t+1)} = E_{q_i}[s_i], \text{where } q_i(z_i) = p(z_i|x_i, \theta^{(t)}), \\ & \tilde{s}_i^{(t+1)} = \tilde{s}^{(t)} - \tilde{s}_i^{(t)} + \tilde{s}_i^{(t+1)} \\ \textbf{M step} \quad & \theta^{(t+1)} = \arg\max_\theta \mathcal{L}(q^{(t+1)}, s), \end{aligned} \tag{4}$$

where $E_q[\cdot]$ denotes expectation with respect to s given by q. Neal and Hinton showed that the incremental version of the EM algorithm gives faster convergence with the same result as the general EM.

C. Training and testing data for the classifier

As discussed previously, the hydroxyl (OH) ion is an indication of past hydrothermal systems, and they are very prominent in the SWIR region of the spectrum. The two minerals selected for classification are kaolinite and pyrophyllite. Kaolinite is a sedimentary clay mineral as well as being a hydrothermal mineral, so its presence can indicate low and high temperature chemical weathering. Pyrophyllite requires higher temperature, acidic waters and is typically found in proximal regions of hydrothermal zones. The Kaolinite spectrum has a well defined doublet structure, with two OH related absorptions at around 2.16 and 2.2 microns. Pyrophyllite has a single OH related

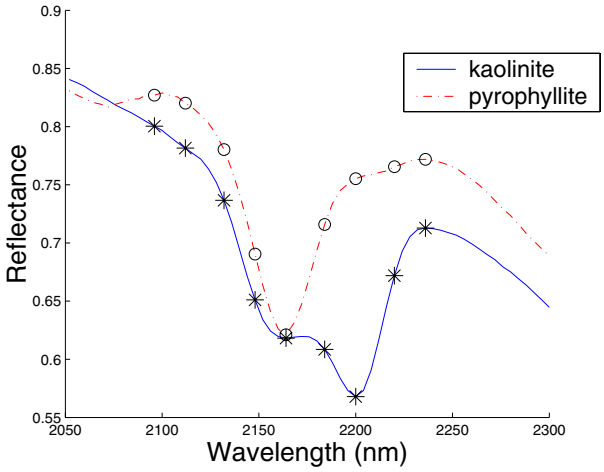

Fig. 3. Reflectance spectra of kaolinite and pyrophyllite, the black $*$ shows the channels of the IR sensor used. [1]

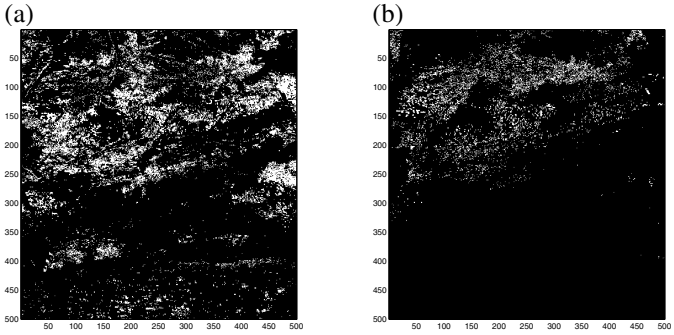

Fig. 4. The resultant classifications of kaolinite (left) and pyrophyllite (right) in test image 1 of Fig. 1(c), using an expert system approach.

absorption at around 2.16 microns. Fig. 3[1] shows the spectra of the two minerals. The wavelengths from the IR sensor that we used for the attributes in the structure are: 2093.2, 2111.2, 2129.0, 2146.5, 2163.9, 2181.3, 2198.3, 2217.1, 2234.1nm.

The two test images as shown on Fig. 1 are 500×500 pixels each in size, thus making it difficult to classify all pixels manually to obtain the ground truth. To obtain the training and testing data, an expert system method is applied on the test area 1 as shown in Fig. 1(c). The approach is intended to detect the centres of overlapping absorption bands, and then compare them to a library of minerals with their absorption band centres [3]. The resultant map of kaolinite is shown on Fig. 4(a), and the corresponding map of pyrophyllite is shown on Fig. 4(b). Similarly, an expert result of the kaolinite map in test image 2 is also obtained, and is shown on Fig. 5. There were no pyrophyllite found in the area using this method.

These results are only a guideline as this approach used a threshold to determine whether a pixel is one class or not. As a result, the maps of Fig. 4 are discontinuous, where

[1]Reproduced from the ASTER Spectral Library through the courtesy of the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California. Copyright ©1999, California Institute of Technology. ALL RIGHTS RESERVED.

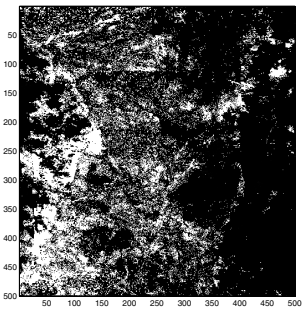

Fig. 5. The resultant classification of kaolinite in test image 2 of Fig. 1(b), using the expert system approach.

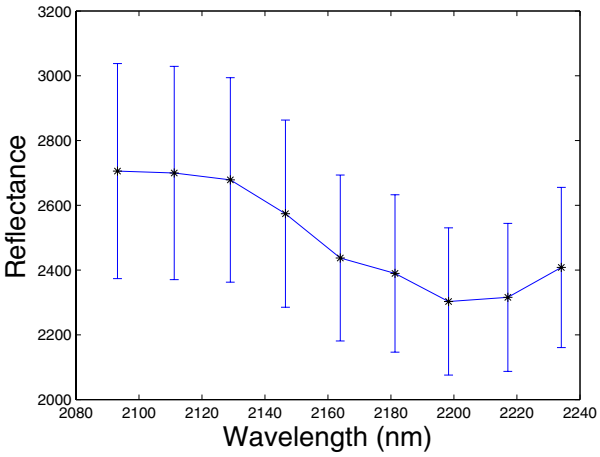

Fig. 6. Spectrum profile of kaolinite for the training data, shown here are the mean and standard deviation of all 2600 pixels used for training.

two pixels that are next to each other, having very similar spectrum profiles but just slightly different in reflectance, can be classified as different classes even when both should be positive identifications. Furthermore, due to the similarity in the profiles of kaolinite and pyrophyllite as shown in Fig. 3, areas on the maps of Fig. 4 overlap when there shouldn't be any. Therefore, before selecting the training data we needed to verify the accuracy of the results. In selecting the training data, we picked areas from Fig. 4 that we knew were accurate classification. In that way, we selected close to 8000 labelled training data.

Fig. 6 shows the spectrum profile of kaolinite used for training. A little over 2600 samples in the training data set were of this mineral, this figure shows the mean and standard deviation of all the kaolinite training data. Note this figure has only slight resemblance with the library spectrum as shown on Fig. 3, this is because the latter is from laboratory results where the samples are pure and the spectrometer has a much greater resolution in wavelength. In determining the training data, we looked for an asymmetry about the 2200 nanometre band. When analysing the result from the classifier, the spectrum at each pixel need to be compared to the spectrum of the minerals used for the training data.

TABLE I

ACCURACY OF THE CLASSIFIERS AS COMPARED WITH THE 'EXPERT' RESULTS

		NB	TAN
Image 1	Orig. Classifier	42%	74%
	IEM Classifier	34%	74%
Image 2	Orig. Classifier	-	45%
	IEM Classifier	-	58%

IV. RESULTS AND DISCUSSION

In this section we report the results of applying incremental EM in the Bayesian network classifiers on the hyperspectral data. Two network structures; the Naïve Bayes Classifier and the TAN Classifier are used. Each of the nodes in the network, apart from the class node, represent a separate channel of the IR sensor. Since the relative reflectance can be any value between 0 and 1, these attribute nodes are continuous and we represent them using Gaussian distributions. The class node takes the discrete values of class number, $C = \{1, 2, 3\}$, where we have the classes of kaolinite identification, pyrophyllite identification and other.

To test the accuracy of the classifiers and the effectiveness of the EM methods, we used the results shown on Fig. 4 and Fig. 5, both comprised of 500×500 pixels. We determine the classification of a pixel if the probability of the data being in a certain class is greater than 50%. The resulting class value of each data from the Bayes Network is then compared with the class value results from the expert results.

It should be pointed out and as can be seen from Fig. 4 that some of the expert results overlap. Therefore, if a pixel is labelled as either kaolinite or pyrophyllite by the expert, then the pixel will be considered as correctly classified if our classifier labels it as either class. Conversely, if the probabilities of the data as calculated by the classifiers in all classes are below 0.5, then we will have a disagreement with the expert result.

We run the experiment in the following order: First, the two classifiers learned from the training data are tested against the expert results. We will refer to these classifiers later on as *'original'* classifiers. Then, the two original classifiers are applied to the entire image 1 of the two test areas. Using incremental EM methods, we learn and interpret the data one pixel at a time and the results are tested again the expert ones. Finally, we run the original classifiers and the updated classifiers on image 2 of the test areas, where the incremental EM method will be applied to the second set of classifiers. We will then compare the probability maps as generated by the classifiers on the second image.

Table I shows the accuracies of the various classifiers as compared with the 'expert' results. However, as noted in the previous section the expert results are not always accurate, thus, we need to analyse the results from comparing the probability of mineral identification with the spectrum at the pixels. The following sections will discuss in detail the accuracy of

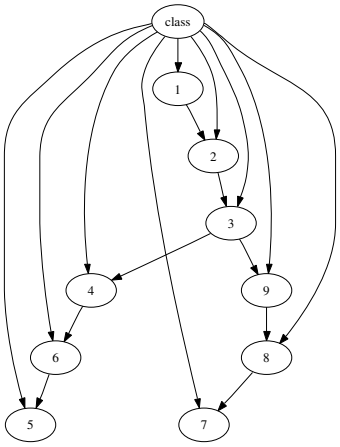

Fig. 9. The structure for the TAN network learnt using CMI method for our example, each numbered node represents the same IR channel as that of Fig. 2.

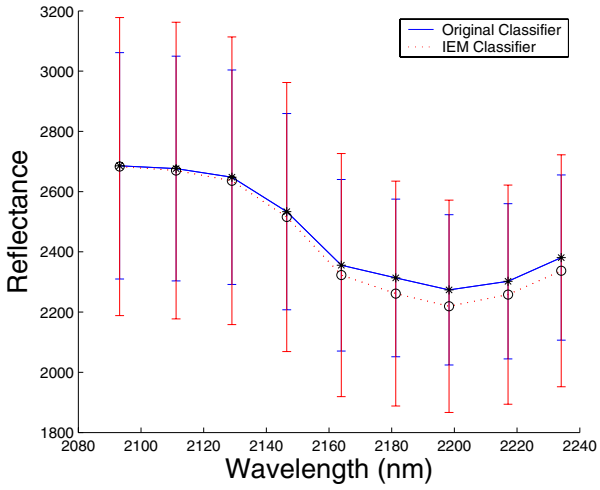

Fig. 10. Spectrum profiles of the kaolinite minerals identified by the TAN classifiers for test image 1. The two profiles represents respectively: using the original classifier, results from the map show in Fig. 7(c); and using the classifier as determined by the incremental EM method, results from the map show in Fig. 8(d).

the classifiers. Due to the large number of samples, it is not possible to show the spectrum at each pixel, and respective resulting probabilities. We will therefore pick a fraction of the resulting data that are calculated by the classifiers as positive identification of a mineral and show the mean and standard deviation of the absolute reflectance at each wavelength. This is because if the mean spectral profile shows a trend for a mineral identification, then the majority of the data in the sample are true positives.

A. *Learning from training data*

1) The Naïve Bayes Classifier: Due to the constraints placed upon the Naïve Bayes network (Fig. 2), the maximum likelihood (ML) process is used to determine the parameters that governs the network. Fig. 7 (a) and (b) show the resultant probability maps of kaolinite and pyrophyllite. Positive identification of minerals seems almost random. The maps do not correlate well with those on Fig. 4. The map of pyrophyllite (Fig. (b)) especially shows the inadequacy of the Naïve Bayes classifier, as we know that the lower half of the map does not contain any such minerals. The failure here is mainly due to the independency assumption placed in the network structure.

2) The TAN Classifier: Fig. 3 shows that bands in the spectrum are dependent on each other in the identification of minerals. Furthermore, in remote sensing situations, environmental condition such as shadows, direction of sunlight and mixture of grains will affect the overall reflectance values greatly. Thus the relative reflectance between bands are much more important. Therefore, the network structure need to reflect the dependencies between the attributes.

Unlike the Naïve Bayes network, the learning process for the TAN network incorporates learning of network structures as well as parameters of the structure. Fig. 9 shows the learnt structure of the TAN network for the training data, and Figure 7 (c) and (d) shows the resultant probability maps of kaolinite and pyrophyllite. Comparing these to the maps of Fig. 4, we can see that the general areas of the mineral identification correlate with one another. In addition, the maps from the TAN classifier also 'fills in' some of the gaps from the expert results making the map more continuous, as expected in real geological settings. The major difference between the two sets of data are in the upper left corner, where the expert map shows a presence of kaolinite when the TAN results don't. In analysing the individual pixels, we found that there is no kaolinite in the area thus the TAN results are more accurate.

B. *Learning from image 1*

Now we look at the results of applying the incremental EM method to the original classifiers. Figure 8 shows the probability map as determined by the two classifiers. The map from the Naïve Bayes classifier shows that most of the decrease in correlation comes from the reclassification of pyrophyllite in the lower half of the image. This shows that while updating the parameters as new data becomes available is a good idea, if we don't have a good model for the classifier, then the result will only becomes worse.

Sub-figures (c) and (d) show the resultant map from the TAN classifier, which we know from the previous results is a reasonable model. We did not therefore incrementally update the structure of the network due to computation time. We can see from the maps that there are slight differences between these results and those found by the original classifiers, thus we need to analyse the spectral profiles of the pixels where the minerals are identified.

Fig. 10 shows the profiles of a random 3000 samples from the positive identification of kaolinite from the two maps, this represents an approximate 10% of the total pixels found to show kaolinite spectral profile. The two spectra both look similar to that of the training profile for kaolinite as shown on Fig. 6, thus showing the majority of these data are accurate.

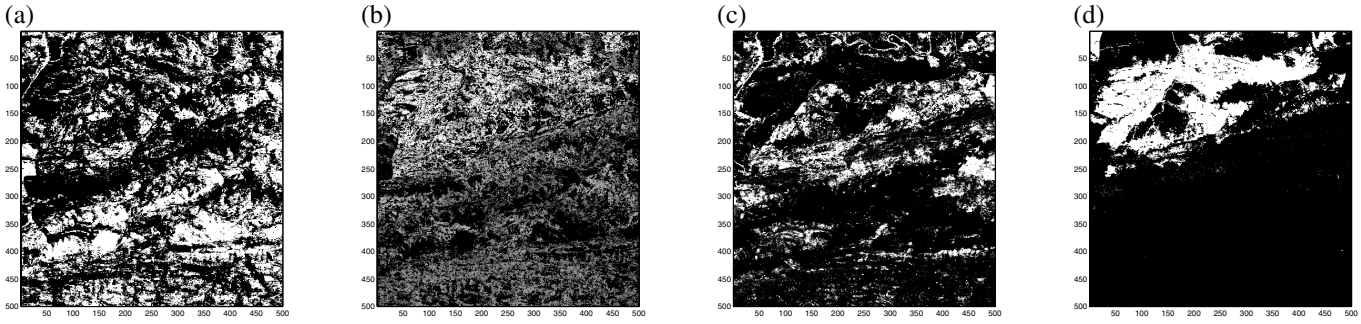

Fig. 7. The resultant probability maps of test image 1 using (a) Naïve Bayes classifier showing kaolinite distribution, (b) Naïve Bayes classifier showing pyrophyllite distribution, (c) TAN classifier showing kaolinite distribution, and (d) showing pyrophyllite distribution. The white area representing 100% probability and black representing 0% probability.

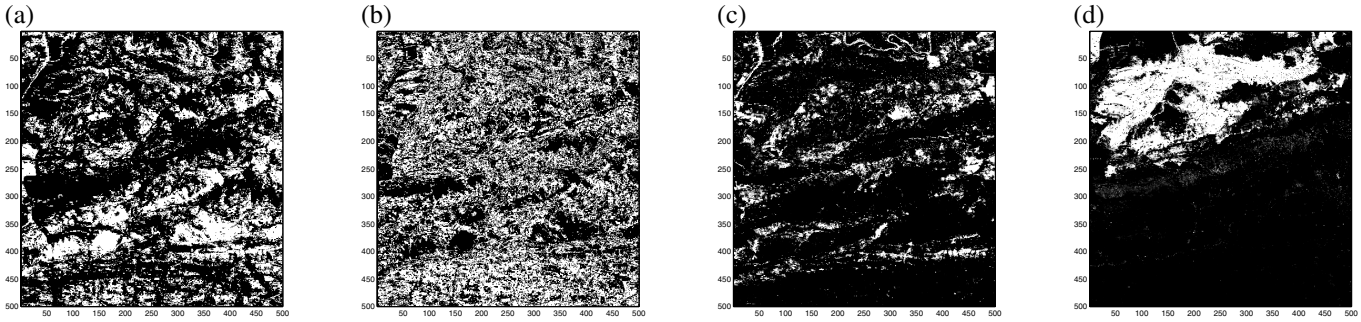

Fig. 8. The resultant probability maps of test image 1 applying the incremental EM method. The sub-figures are in the same order as those in Fig. 7.

The pixels where there is a disagreement in kaolinite identification between the two classifiers comprised of approximately 10% of the total positive kaolinite found in both sets of results. The majority of these pixels also show similar profile to that of Fig. 6, which means there is an equal number of false negatives in both results. This shows that if the training data is representative of the data in the image, then applying incremental EM to the classifier does not affect the result much.

C. Learning from image 2

The previous results show that the Naïve Bayes Classifier is not a good classifier for actual hyperspectral image data. Therefore, we will not be using it to classify the second test image as shown on Fig. 1(b), rather we used the TAN classifier for this image only.

Figures 11 and 12 shows the resultant probability maps of the minerals in test image 2. The former is the result from using just the original classifier, and the latter is the result from employing the incremental EM method on the classifier learnt from the test image 1.

Both result maps show that there is little pyrophyllite present in the area, which is consistent with the result from the expert system. Comparing the maps of kaolinite with that of the expert result from Fig. 5 shows that the map generated by the incremental EM method is closer to that of the expert system. However, we still need to analyse the spectrum profiles of the data to assess the effectiveness of the classifiers.

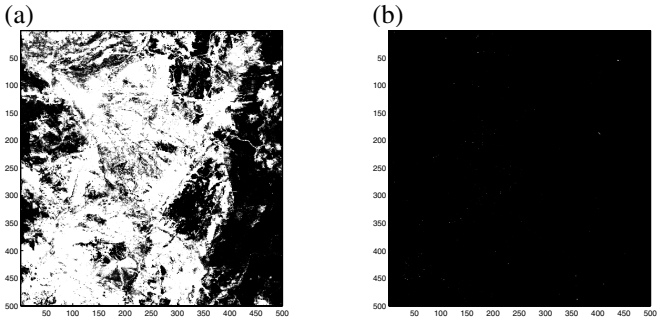

Fig. 11. The resultant probability maps of test image 2 using the original TAN classifiers: (a) kaolinite, (b) pyrophyllite. The scale of the results are the same as those in Fig. 7.

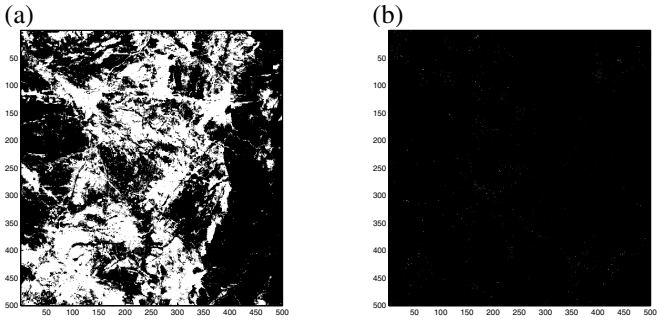

Fig. 12. The resultant probability maps of test image 2 applying the incremental EM method to the TAN classifiers from Fig. 8: (a) kaolinite, (b) pyrophyllite. The scale of the results are the same as those in Fig. 7.

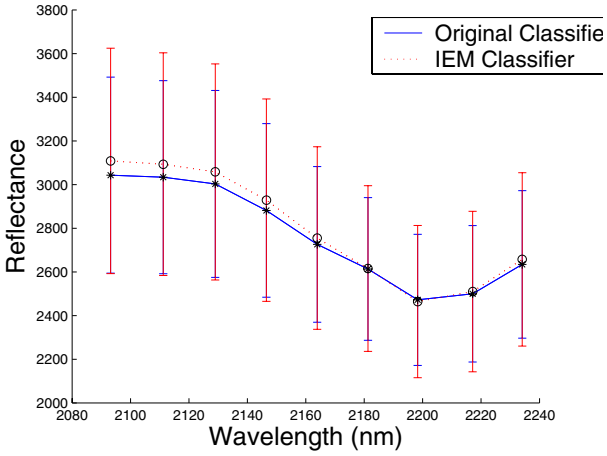

Fig. 13. Spectrum profile showing the kaolinite mineral identified by the TAN classifiers for test image 2. The two profiles represents respectively: using the original classifier, results from the map show in Fig. 11(a); and using the classifier as determined by the incremental EM method, results from the map show in Fig. 12(a).

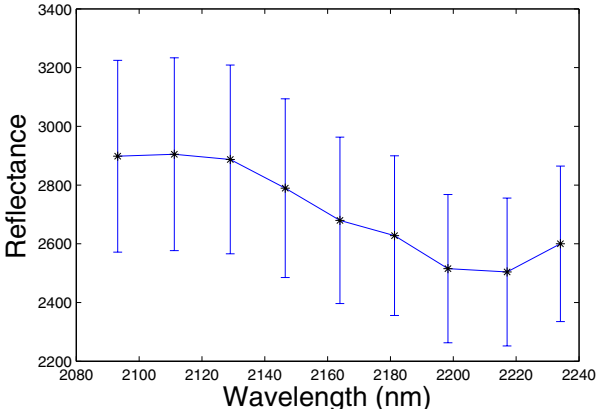

Fig. 14. Spectrum profile showing the mean and standard deviation of samples where there is a disagreement in the identification of the kaolinite minerals between the original classifier and the classifier through the incremental EM method for test image 2. Showing here are the samples where the original classifier found to be of positive identification.

Figure 13 shows the spectrum profile of approximately 4000 random samples of positive identification by the the two classifiers, representing approximately 5% of the total identification. The figure shows that the mean of the positive identifications showing prominent asymmetries at the 2200nm band, as identified on the training data.

In the two resulting maps shown on Fig. 11 and 12 there is approximately 30 000 pixels classified as positive identification of kaolinite in Fig. 11(a), which uses the original classifier, that are not identified as kaolinite in Fig. 12(a), which apply the IEM method. Fig. 14 shows the spectrum profile of approximately 2000 of these data. We can see from this figure, that the majority of the differences do not have the asymmetry at 2200nm band as defined by the training data. Furthermore, there are approximately 1800 pixels that are found to be positive identification of kaolinite by the IEM method that are not identified by the original classifier, and this data show the asymmetry at the 2200nm band. Therefore, on applying to a different image than where the original training data comes from, the incremental EM method gives a more accurate result.

V. Conclusion

In this paper, we applied the incremental EM method to Bayesian Network Classifiers, to allow continuous update of hyperspectral data and the classifier parameters. We found that if the training data is representative of the data in the area, then applying the incremental EM method on the image where the training data is taken does not affect the result. However, on a different image tens of kilometre apart from the training data, applying the incremental EM method to the classifier improves classification accuracy.

Acknowledgment

This work is supported by the ARC Centre of Excellence programme, funded by the Australian Research Council (ARC) and the New South Wales State Government.

References

[1] ASTER Spectral Library, http://speclib.jpl.nasa.gov/

[2] C. K. Chow, C. N. Liu, *Approximating Discrete Probability Distributions with Dependence Trees*, IEEE Transations on Information Theory, vol. IT-14, no. 3, pp. 462-467, May 1968

[3] R. N. Clark, *et al. Imaging Spectroscopy: Earth and Planetary Remote Sensing with the USGS Tetracorder and Expert Systems*, Journal of Geophysical Research, vol. 108, pp. 5131-5175, 2003

[4] T. Cocks, R. Jenssen, A. Stewart, I. Wilson, and T. Shields, *The HyMap airborne hyperspectral sensor: the system, calibration and performance*, 1st EARSeL Converence, pp. 37-42., 1998

[5] A. P. Dempster, N. M. Laird, and D. B. Rubin, *Maximum likelihood from incomplete data via the EM algorithm*, Journal of the Royal Statistical Society B, vol. 39, pp. 1-39, 1977

[6] N. Friedman, D. Geiger, and M. Goldszmidt, *Bayesian network classifiers*, Machine Learning, vol. 29, pp. 131-163, 1997

[7] N. Friedman, and D. Koller, *Being Bayesian About Network Structure: A Bayesian Approach to Structure Discovery in Bayesian Networks*, Machine Learning, vol. 50, pp. 95-126, 2003,

[8] F. V. Jensen, *Bayesian Networks and Decision Graphs*, Springer-Verlag, New York, 2001

[9] R. M. Neal, and G. E. Hinton. *A new view of the EM algorithm that justifies incremental and other variants*, In M. Jordan, editor, Learning in Graphical Models. MIT Press, 1998.

[10] A. N. Rencz, *Remote Sensing for the Earth Sciences*, 3rd ed., John Wiley & Sons, Inc., 1999

[11] A. J. B. Thompson, P. L. Hauff, and A. J. Robitaille, *Alteration mapping in exploration: application of short-wave infrered (SWIR) spectroscopy*, Socity of Economic Geologists Newsletter, no. 39, October 1999

[12] M. R. Walter, and D. J. Des Marais, *Preservation of Biological Information in Thermal Spring Deposits: Developing a Strategy for the Search for Fossiel Life on Mars*, Icarus, Vol. 101, No. 1, pp. 129-143, 1993