

PARTIALLY-OBSERVED MODELS FOR CLASSIFYING MINERALS ON MARS

Murat Dundar, Lin Li

Bartek Rajwa

IUPUI
Indianapolis, IN 46202

Purdue University
West Lafayette, IN 47906

ABSTRACT

The identification of phyllosilicates by NASA's CRISM (Compact Reconnaissance Imaging Spectrometer for Mars) strongly suggests the presence of water-related geological processes. A variety of water-bearing phyllosilicate minerals have already been identified by several research groups utilizing spectral enrichment techniques and matching phyllosilicate-rich regions on the Martian surface to known spectra of minerals found on earth. However, fully automated analysis of the CRISM data remains a challenge for two main reasons. First, there is significant variability in the spectral signature of the same mineral obtained from different regions on the Martian surface. Second, the list of mineral confirmed to date constituting the set of training classes is not exhaustive. Thus, when classifying new regions, using a classifier trained with selected minerals and chemicals, one must consider the potential presence of unknown materials not represented in the training library. We made an initial attempt to study these problems in the context of our recent work on partially-observed classification models and present results that show the utility of such models in identifying spectra of unknown minerals while simultaneously recognizing spectra of known minerals.

Index Terms— semi-supervised learning, unknown class discovery, mars, crism, mineralogy, phyllosilicates

1. INTRODUCTION

The Compact Reconnaissance Imaging Spectrometer for Mars (CRISM) is a hyperspectral imaging spectrometer on board the Mars Reconnaissance Orbiter (MRO), and measures the visible and infrared electromagnetic radiation from 0.4 to 4.0 μm using two detectors (S and L) that cover the spectral region from 0.4 to 1.0 μm and 1.0 to 4.0 μm , respectively. CRISM operates in two modes: multispectral untargeted and hyperspectral targeted. When operating in the targeted mode CRISM can collect data in 544 channels with a spatial resolution of 15-38 m/pixel. Radiance data measured by CRISM were converted to I/F data by computing the ratio of the radiance to the solar irradiance at Mars [1].

CRISM images acquired by the L detector were previously analyzed to identify minerals that can be chemically al-

tered by water (iron bearing minerals or iron oxides) or those that form in the presence of water (phyllosilicates and carbonates) [2, 3]. We investigate the utility of a partially-observed classification model in discovering new minerals on the Martian surface with the CRISM image. Simple atmospheric and photometric corrections were applied to all data cubes using the CRISM Analysis Toolkit developed by the CRISM science team [4]. Only the spectral channels that cover the spectral region from 1.0 to 2.6 μm (248 channels) are used in experiments performed in this study. The channels corresponding to the remaining part of the spectrum (2.6 to 4.0 μm) were excluded because they do not offer much additional information for classifying phyllosilicates and show low data quality and residual artifacts.

We use images of the Nili Fossae region analyzed in [3] to construct a labeled data set. Several subregions containing Fe/Mg smectite, zeolite, kaolinite, chlorite, carbonate, K mica, hydrated silica, and serpentine have already been reported in [3]. We expand these subregions by including new ones, which are initially identified with the help of summary parameters [5] and then confirmed by spectral enrichment followed by spectral matching. In addition to several subregions identified as belonging to the previously reported mineralogy, we identified two subregions with almost identical spectral signatures, but that do not match any of the eight previously reported mineral types for these images.

There are nine different mineral types in our labeled data set, eight of which are known and the remaining one is unknown. The list of mineral classes along with the number of subregions and pixels for each class are shown in Table 1. In addition to the one unknown mineral component, available we will treat serpentine as a *known unknown* and sequester subregions belonging to these two types from the rest of the data. Serpentine is chosen as a *known unknown* because it was reported on the Martian surface for the first time in the Nili Fossae region [3], the region from which the labeled data set used in this study is derived. The remaining subregions covering seven different mineral types were split into two groups for training and testing. Subregions belonging to serpentine and to the unmatched mineral class were merged with the test set. This way a test set that contains two additional classes than those observed in the training data set is constructed. The classification task involves classifying pixels of observed

This is the author's manuscript of the article published in final edited form as:

Mineral Class	Number of Subregions	Number of Pixels
K Mica	7	69
Fe/Mg Smectite	20	2166
Carbonate	7	599
Chlorite	5	301
Kaolinite	3	147
Silica	4	338
Zeolite	4	121
Serpentine	8	310
Unmatched	2	57

Table 1. The list of mineral classes and the number of subregions and pixels available from each class. Unobserved classes are shown in boldface font.

classes to their respective classes while discovering and modeling the two classes not observed during training. The isometric log-ratio transform [6] is applied to each pixel data before using this data for classification.

The rest of this paper is organized as follows. In Section 2 we review the partially-observed hierarchical Dirichlet process. In Section 3 we present and discuss experimental results. In Section 4 we conclude with the strengths and limitations of a partially-observed classification model and present future directions for research.

2. PARTIALLY-OBSERVED CLASSIFICATION MODELS

The unobserved class problem is traditionally studied within the scope of anomaly/novelty detection problems, where samples of unobserved classes are treated as anomalies/novelty techniques. Although these techniques show some promise in distinguishing samples of unobserved classes from observed ones, they cannot differentiate between samples of multiple unobserved classes and may not be readily useful for simultaneous discovery of multiple classes.

In this section we briefly describe partially-observed classification models [7]. In this framework each class is modeled by a Gaussian mixture model (GMM) with an unknown number of components. A partially-observed classification model arises when the training data set is non-exhaustively defined in terms of the set of classes or the set of components for some or all of the classes or both. The test data may contain samples from classes and subclasses that are not observed in the training data. Under such settings classification with a fixed model becomes impractical, as samples of classes and components in the test data that are not observed during training will be misclassified with certainty. A dynamic model that can adjust itself by adding new classes and components as needed can better accommodate test data.

Toward achieving this end we define a hierarchical Dirichlet process (HDP) [8] over component distributions to dynamically model the number of classes and their compo-

nents. HDP extends Dirichlet process (DP) [9], which is mainly used in clustering and density estimation problems as a non-parametric prior defined over the number of mixture components. A DP can be considered as a distribution over distributions. HDP models each group of data in the form of a Dirichlet process mixture (DPM) model, where DPM models across different groups are connected together through a higher level DP. We use the notation $x_{ji} \in \mathcal{R}^d$, $i = \{1, \dots, n_j\}$, $j = \{1, \dots, J\}$ to identify sample i in class j , where n_j denotes the number of samples in class j , J is the total number of classes, and θ_{ji} defines the parameters of the mixture component associated with x_{ji} . Each x_{ji} is associated with a mixture component defined by the parameter θ_{ji} , which is generated i.i.d. from a Dirichlet process as follows:

$$\begin{aligned} x_{ji} | \theta_{ji} &\sim p(\cdot | \theta_{ji}) && \text{for each } j, i \\ \theta_{ji} | G_j &\sim G_j && \text{for each } j, i \end{aligned} \quad (1)$$

where G_j are random probability measures distributed i.i.d. according to a DP with base distribution G_0 and precision parameter α . In this framework, unlike continuous distributions, the probability of sampling the same θ_{ji} twice is not zero. Thus G_j is considered a discrete distribution. The precision parameter, α , controls the prior probability of assigning a new sample to a new component and thus, plays a critical role in the number of components generated. In the HDP model the base distribution G_0 is distributed according to a higher level DP with a base distribution H and parameter γ . This hierarchical model couples G_j and allows for sharing of mixture components within and between groups. The HDP model is completed as follows:

$$\begin{aligned} G_j | G_0, \alpha &\sim DP(G_0, \alpha) && \text{for each } j \\ G_0 | H, \gamma &\sim DP(H, \gamma) \end{aligned} \quad (2)$$

The sharing mechanism inherent with the HDP model not only mitigates the curse of dimensionality problem but also connects observed classes with unobserved ones through sharing of their parameters. To avoid unidentifiable mixture components, we limit sharing with covariance matrices of components while leaving their mean vectors free. The component membership of training samples and component and class membership of test samples are jointly inferred by a collapsed Gibbs sampler as described in detail in [7]. Each sweep of the Gibbs sampler also involves sampling γ and α values using the technique described in [10].

Component data are distributed according to a Gaussian distribution with mean vector μ and a covariance matrix Σ . For the base distribution H we define a conjugate prior:

$$H = p(\mu, \Sigma) = \underbrace{\mathcal{N}\left(\mu | \mu_0, \frac{\Sigma}{\kappa}\right)}_{p(\mu | \Sigma)} \times \underbrace{W^{-1}(\Sigma | \Sigma_0, m)}_{p(\Sigma)} \quad (3)$$

where μ_0 is the prior mean and κ is a scaling constant that controls the deviation of the mean vectors of mixture components from the prior mean. The smaller the κ , the larger the scattering between the components will be. The parameter Σ_0 is a positive definite matrix that encodes our prior belief about the expected Σ . The parameter m is a scalar that is negatively correlated with the degrees of freedom. In other words, the larger the m is the less Σ will deviate from Σ_0 and vice versa. The parameters $(\Sigma_0, \mu_0, \kappa)$ are estimated using training samples in the same way as described in our earlier work [11] and m is vaguely defined as $1.5d$, where d is the dimensionality of the data.

To evaluate the collapsed Gibbs sampler we need the predictive distribution $p(x|\bar{x}, S)$ for each component, where \bar{x} and S denotes sample mean and covariance matrices, respectively. Under the prior model considered above $p(x|\bar{x}, S)$ turns out to be a multivariate Student-t distribution.

3. EXPERIMENTAL RESULTS

We evaluate the performance of the partially-observed HDP (PO-HDP) for classifying test samples of mineral classes shown in Table 1. The model is trained in a semi-supervised fashion using samples from training and test data sets together with the model blinded to test labels. Once the Gibbs sampler converges to the target distribution we record the Gibbs state that maximizes the mode and use this state for evaluation purposes.

Components containing at least one training sample from an observed class are considered observed and are readily associated with that class. Components containing only test samples are considered unobserved. These may point to unobserved components of observed classes but may also point to components of unobserved classes. To associate these components with observed and unobserved classes we use class labels of test samples and associate each component with the class that has the most number of samples in that component. After all components are associated with classes the classification performance on the test data is evaluated by F1 scores computed separately for each of the nine classes.

The performance of the PO-HDP model is compared against a support vector machine (SVM) classifier as well as an anomaly detection technique based on support vector domain description (SVDD). For the SVM classifier a one-against-all multi-class classification approach is adopted. The parameters of the SVM are tuned to maximize the average F1 score on the validation data set, which is produced from the training data set by randomly sequestering 35% of the subregions available in that set. Once the parameters are tuned, the final classifier is trained using all subregions in the original training data. This classifier is evaluated on the test data and F1 scores are recorded.

The SVDD approach fits a tight hypersphere on to each class training data. During the testing phase the distances

Mineral Class	F1 Scores		
	PO-HDP	SVM	SVDD
K Mica	0.64	0.42	0.54
Fe/Mg Smectite	0.84	0.68	0.61
Carbonate	0.88	0.72	0.68
Chlorite	0.91	0.79	0.00
Kaolinite	1.00	0.90	0.00
Silica	1.00	0.71	0.91
Zeolite	1.00	0.90	0.00
Serpentine	0.85	0.00	0.00
Unmatched	0.88	0.00	0.00

Table 2. F1 scores obtained on the test data set for the nine mineral classes using PO-HDP, SVM, and SVDD. Unobserved classes are shown in boldface font.

from test samples to the center of each hypersphere are computed. Samples whose distances to the center of a hypersphere are less than a pre-optimized threshold, i.e., some multiples of the radius of that hypersphere, are assigned to the class associated with that hypersphere. Samples that are not assigned to any of the classes are identified as samples of unknown classes. The parameters are optimized similar to the tuning process described above for SVM.

The results of the classification for all three techniques are shown in Table 2. PO-HDP shows great promise not only in classifying samples of observed classes by more accurately modeling their underlying class densities but also in discovering and recovering components of unobserved classes with fairly good accuracy. SVM performs slightly worse than PO-HDP on observed classes. Since SVM cannot discover new classes, samples of the two unobserved classes are misclassified into one of the observed classes noticeably reducing F1 scores for some of these classes. SVDD is included in this analysis to serve as a benchmark for detecting samples of unobserved classes. Despite extensive tuning, SVDD fails to identify any samples from these two classes. Its performance on the observed classes is not very promising either.

The number of observed and unobserved components generated by the PO-HDP model for each of the nine classes are shown in Table 3. A total 47 components are generated, of which 36 are observed and 11 are unobserved. Three of the unobserved components are associated with observed classes, possibly indicating that the test data set contains spectral variants of these classes different from those included in the training data set.

4. CONCLUSIONS

Automated analysis of images acquired by planetary orbital spectrometers is a challenge not only because of the many natural and instrumental limitations we face but also owing to our inability to define beforehand what we should be looking at in these images. Partially-observed classification models can be used for more effective analysis of these images for

Mineral Class	Number of Components	
	Observed	Unobserved
K Mica	4	0
Fe/Mg Smectite	17	2
Carbonate	5	0
Chlorite	3	0
Kaolinite	2	1
Silica	1	0
Zeolite	4	0
Serpentine	0	7
Unmatched	0	1

Table 3. Number of components generated by PO-HDP. Unobserved classes are shown in boldface font.

discovery of unknown rock/mineral types while classifying regions with matching spectra into their respective classes. Experimental results on a small mineral data set acquired by the CRISM spectrometer suggest that there is promise in such endeavors.

Although convergence of the Gibbs sampler was not an issue for the small data set considered in this study, scaling the PO-HDP model to larger images will require developing more efficient sampling schemes. This is a very general problem that afflicts most Markov Chain Monte Carlo (MCMC) techniques. Particle filters, sequential Monte Carlo samplers, and variational Bayes techniques can be explored toward this end with some success. In most practical settings finding out whether an unobserved component belongs to an observed or unobserved class will be a challenge. However, if a training library with most common spectral variants of observed classes can be constructed, new compositional components can be more confidently assigned to unobserved classes, which can then be identified by comparing their image spectra with laboratory-measured spectra of minerals found on earth. The proposed framework facilitates such a verification task by on-the-fly clustering of samples not fitting into one of the observed classes and allows for verification at the cluster level, as opposed to the sample level as would otherwise be done with most anomaly/novelty detection algorithms.

Acknowledgment

This research was in part sponsored by the National Institute of Allergy and Infectious Diseases (NIAID) under Grant Number 5R21AI085531 and by the National Science Foundation (NSF) under Grant Number IIS-1252648 (CAREER). The content is solely the responsibility of the authors and does not necessarily represent the official views of NIAID or NSF.

5. REFERENCES

- [1] S. Murchie et al., “Compact reconnaissance imaging spectrometer for mars (crism) on mars reconnaissance orbiter (mro),” *Journal of Geophysical Research: Planets*, vol. 112, no. E5, 2007.
- [2] Scott L. Murchie et al., “A synthesis of martian aqueous mineralogy after 1 mars year of observations from the mars reconnaissance orbiter,” *Journal of Geophysical Research: Planets*, vol. 114, no. E2, 2009.
- [3] Bethany L. Ehlmann et al., “Identification of hydrated silicate minerals on mars using mro-crisp: Geologic context near nili fossae and implications for aqueous alteration,” *Journal of Geophysical Research: Planets*, vol. 114, no. E2, 2009.
- [4] Frank Morgan et al., “Crisp data users’ workshop cat tutorial,” http://pds-geosciences.wustl.edu/missions/mro/CRISM_Workshop_090322_CAT_MFM.pdf, March 2009.
- [5] S. M. Pelkey et al., “Crisp multispectral summary products: Parameterizing mineral diversity on mars from reflectance,” *Journal of Geophysical Research: Planets*, vol. 112, no. E8, 2007.
- [6] J. J. Egozcue, V. Pawlowsky-Glahn, G. Mateu-Figueras, and C. Barceló-Vidal, “Isometric logratio transformations for compositional data analysis,” *Mathematical Geology*, vol. 35, no. 3, pp. 279–300, 2003.
- [7] Ferit Akova, Yuan Qi, Bartek Rajwa, and Murat Dundar, “Self-adjusting models for semi-supervised learning in partially-observed settings,” in *Proceedings of the IEEE International Conference on Data Mining (ICDM’12)*, 2012.
- [8] Y.W. Teh, M.I. Jordan, M.J. Beal, and D.M. Blei, “Hierarchical dirichlet processes,” *Journal of the American Statistical Association*, vol. 101, no. 476, pp. 1566–1581, 2006.
- [9] Thomas S. Ferguson, “A Bayesian analysis of some nonparametric problems,” *Annals of Statistics*, vol. 1, no. 2, pp. 209–230, 1973.
- [10] Michael D. Escobar and Mike West, “Bayesian density estimation and inference using mixtures,” *Journal of the American Statistical Association*, vol. 90, pp. 577–588, 1994.
- [11] Murat Dundar, Ferit Akova, Yuan Qi, and Bartek Rajwa, “Bayesian nonexhaustive learning for online discovery and modeling of emerging classes,” in *Proceedings of the International Conference on Machine Learning (ICML’12)*. 2012, pp. 113–120, Omnipress.

- [1] S. Murchie et al., “Compact reconnaissance imaging spectrometer for mars (crism) on mars reconnaissance