Onboard detection of jarosite minerals with applications to Mars

Benjamin Bornstein, Rebecca Castaño
Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Drive
Pasadena, California 91109
818-393-5349, 818-393-5344
{ben.bornstein, rebecca.castaño}@jpl.nasa.gov

Martha S. Gilmore, Matthew Merrill, James P. Greenwood
Dept. of Earth and Environmental Sciences
Wesleyan University
265 Church St.
Middletown, Connecticut 06459
860-685-3129, 860-685-4477, 860-685-2545
{mgilmore, mmerrill, jgreenwood}@wesleyan.edu

Abstract—We have developed a highly accurate Support Vector Machine (SVM) based detector capable of identifying jarosite (K, Na, H\textsubscript{2}O)Fe\textsubscript{3}(SO\textsubscript{4})\textsubscript{2}(OH)\textsubscript{6}) in the visible/NIR (350–2500 nm) spectra of both laboratory specimens and rocks in Mars analogue field environments. To keep the computational complexity of the detector to a minimum, we restricted our design to an SVM with a linear kernel and a small number of support vectors. We used our generative model to create linear mixtures of end-member library spectra to train the SVM. We validated the detector on museum quality laboratory samples (97% accuracy) and field rock samples measured in both the laboratory and the field (both 88% accuracy). In the interest of technology infusion, the detector has been integrated into the CLARAty autonomous mobile robotics software architecture\textsuperscript{1,2}.

TABLE OF CONTENTS
1. INTRODUCTION.............................................................. 1
2. SUPPORT VECTOR MACHINES................................. 1
3. GENERATIVE MODEL............................................... 2
4. DETECTOR CONSTRUCTION...................................... 3
5. DETECTOR TESTING............................................... 3
6. RESULTS...................................................................... 4
7. CLARATY INTEGRATION........................................ 4
8. CONCLUSIONS............................................................ 5
9. FUTURE DIRECTIONS................................................ 5
10. ACKNOWLEDGEMENTS.......................................... 5
11. REFERENCES.......................................................... 5
12. BIOGRAPHIES....................................................... 7

1. INTRODUCTION

Data collection capabilities of Mars rovers far surpass downlink rates, which can result in a critical loss of valuable science data. Onboard techniques for identifying geologically relevant information and prioritizing its downlink have the potential to dramatically increase the science return of future rover missions. For these reasons, we have developed methods to construct mineral detectors capable of running on current and future rover and orbital hardware. Carbonate and sulfate minerals are of particular geologic importance because they can signal the presence of water which is necessary for life. In our previous work, we reported on the creation and testing of a calcite (a carbonate mineral) detector [3, 8, 9]. In this work, we describe our jarosite (a sulfate mineral) detector. Our focus on jarosite is driven by the recent discoveries of the mineral at the Eagle and Endurance craters in Meridiani Planum by the Mars Exploration Rover (MER) Opportunity [15].

The jarosite detector was developed using a support vector machine (SVM), a type of supervised machine learning classifier. A brief introduction to SVMs is given in Section 2. In Section 3, we describe the source of data used to train the detector. We then describe the development of the jarosite detector in Section 4. In Sections 5 and 6, we assess the performance of the detector on museum quality laboratory samples and field samples (measured in both the laboratory and field). Integration of the detector with the Coupled Layer Architecture for Robotic Autonomy (CLARAty) and the detector’s low computational complexity per analysis of spectral measurement is discussed in Section 7. Finally, in Sections 8 and 9, we draw conclusions and offer directions for future work, including the use of ensemble classifiers and the applicability of our techniques to hyperspectral images taken from airborne and spaceborne instruments such as AVIRIS, Hyperion, OMEGA and CRISM.

2. SUPPORT VECTOR MACHINES

Support Vector Machines (SVMs) are a family of classifiers that identify the optimal linear separator between classes in a (possibly) high dimensional space [7]. There are two aspects of SVMs that make them appealing. First, SVMs belong to a category of machine learning techniques that produce empirically derived classification algorithms by explicitly attempting to maximize the margin, the boundary that separates one type of classified data from another. Maximizing the margin prevents the algorithm from overfitting the training data, which can lead to poor algorithm performance on new input data sets. Thus, maximizing the margin leads to better generalization of the classifier.
Second, nonlinear decision boundaries are supported by mapping the input feature space to a higher dimension (possibly infinite) where the features are linearly separable. The key is that the mapping is carried out implicitly so that all that is necessary is to compute dot products in the original input feature space.

A primary disadvantage of support vector machines is that the time to classify points is proportional to the number of support vectors, which for complicated problems can be as large as the number of training examples. Since we wanted to train on thousands of examples, this would result in unreasonably slow classification. Instead, we decided not to use a kernel function, leaving us with a linear support vector machine that can be evaluated quickly.

3. GENERATIVE MODEL

Our generative model was first used to create synthetic spectra to train our carbonate mineral detector [3]. We describe the model again here because it is integral to our detector creation methodology. Readers familiar with our previous work may wish to skip this section.

A potential drawback of many types of machine learning classifiers is that large numbers of training examples are often required for the network to converge (learn). Our initial experiments showed that tens of spectra were insufficient to train SVMs to predict the presence or absence of mineral end-members. This presented a problem since collecting only a few dozen samples often requires days of fieldwork followed by lab work to verify the modal mineralogy of each sample. While it is possible to use spectral libraries to alleviate the labor burden of sample collection and analysis, for our purposes, spectral libraries have two drawbacks: i) they often contain ten or fewer spectra per sample and ii) the spectrum is usually measured in the laboratory, which yields data cleaner than that collected in the field. For an SVM detector to perform well in field, it must be trained on field data or something that closely resembles it.

In an effort to inexpensively obtain spectral data with many of the characteristics of field data, we have begun to develop a generative model for spectra. The model starts with a database populated from two well-know spectral libraries: JPL’s Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) library [12] and the U.S. Geological Survey’s (USGS) speclib04 [6]. To this database we add virtual rocks, defined as containing one or more constituent minerals in specific percentages or percentage ranges. For data generation, all percentages are chosen using a uniform pseudorandom number generator.

In addition to a percentage (range), for each virtual rock description, we classify each mineral as essential, accessory, or accidental. Essential minerals are required to occur in a rock at the percentage (or within the range) specified. Accessory minerals, which may occur in the virtual rock, are added to a rock if the chosen percentages for essential minerals do not sum to 100. Finally, accidental minerals, which rarely occur in the virtual rock, are added to rocks with low probability (usually less than 10% of the rocks generated) (Table 1). The compositions of the virtual rocks were selected to represent rocks known (from meteorite specimens and orbital spectroscopy) and predicted to occur on Mars. This includes basic volcanic rocks (i.e., basalts) and dust that comprise the majority of the surface as well as minerals of interest (minerals associated with water such as carbonates and evaporates). The classification of minerals as essential, accessory, or accidental and their percentages are constrained to be geologically reasonable.

<table>
<thead>
<tr>
<th>Mineral</th>
<th>% in rock</th>
<th>Mineral</th>
<th>% in rock</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labradorite</td>
<td>40–55</td>
<td>Hematite</td>
<td>1–10</td>
</tr>
<tr>
<td>Augite</td>
<td>31–45</td>
<td>Magnetite</td>
<td>1–10</td>
</tr>
<tr>
<td>Forsterite</td>
<td>0–20</td>
<td>Ilmenite</td>
<td>1–10</td>
</tr>
<tr>
<td>Labradorite</td>
<td>40–55</td>
<td>Hematite</td>
<td>1–10</td>
</tr>
<tr>
<td>Augite</td>
<td>31–45</td>
<td>Magnetite</td>
<td>1–10</td>
</tr>
<tr>
<td>Forsterite</td>
<td>0–20</td>
<td>Ilmenite</td>
<td>1–10</td>
</tr>
<tr>
<td>Labradorite</td>
<td>40–55</td>
<td>Hematite</td>
<td>1–10</td>
</tr>
<tr>
<td>Augite</td>
<td>31–45</td>
<td>Magnetite</td>
<td>1–10</td>
</tr>
<tr>
<td>Forsterite</td>
<td>0–20</td>
<td>Ilmenite</td>
<td>1–10</td>
</tr>
</tbody>
</table>

To generate spectra for each virtual rock, we take the rock’s constituent mineral spectra and corresponding percentages and apply a mixing model. The simplest is a linear mixing model. Let \( r(m, b) \) be the reflectance for end-member \( m \) at bandwidth \( b \). Then, given the mix percentage for each end-member \( w_m \), the mixed reflectance for band \( b \), \( R_b \), is:

\[
R_b = \sum_m w_m r(m,b)
\]

Thus, under the linear mixing model, mixed reflectances are simply weighted, linear combinations of end-member spectra. Recall, however, the spectra are drawn from spectral libraries, which contain predominately laboratory spectra. To simulate instrument and other noise encountered when taking field measurements, we add pseudorandom Gaussian noise with mean zero and variance \( \sigma \):

\[
R_b = \sum_m w_m r(m,b) + N[0,\sigma]
\]

While such a mixing model is simplistic, it has allowed us to setup and test our generative model framework and it can provide a wealth of spectra with many subtle variations to train SVMs and other machine learning techniques. We are currently developing a set of richer, nonlinear mixing models based on the reflectance and refraction models of Hapke [11].
4. DETECTOR CONSTRUCTION

To reduce computation time, we first limit the detector input to those regions of the spectrum that contain characteristic features of jarosite. We also sought to avoid noise due to atmospheric water vapor at 1400 and 1900 nm that would be encountered during field tests. Jarosite typically has the following spectral features: a steep slope from 350–700 nm which is the edge of a charge transfer band, ferric crystal field transition bands at 430 nm and ~930 nm, and bound water vibration bands ~1470, 1850, 2250 and 2500 nm [13, 14, 22, 19, 2] (Figure 1). While the 430 nm band is specific to jarosite, this band is lost in noise in some of our spectra and omitted in this initial detector. We selected two spectral intervals over which the detector will operate: 500–1350 nm and 2050–2380 nm. Compositional variability in jarosite (natrojarosite NaFe3(SO4)2(OH)6, plumbojarosite PbFe3(SO4)4(OH)12, hydronium jarosite (H3O)Fe(SO4)2(OH)6) are found to have minimal effect on band positions [16, 2], and thus the detector should apply to these end-members and their solid solutions.

A generative model was used to create linear mixtures of spectral library data for SVM training [3] (Section 3). Nine different jarosite spectra representing the K, Na and H2O varieties of the mineral from the Unites States Geologic Survey (USGS) Speclib04a database were introduced as “jarosites” for training purposes [6]. Training spectra for the non-jarosite class includes minerals that are often found in the presence of jarosite and which are consistent with martian petrology. (Table 2) [17]. Minerals that occur terrestrially with jarosite were also included. In total 100 spectra were created to train the SVM. Of these, 54 were jarosites and 46 were non-jarosite with the later group comprising equal amounts of pure end member, binary, tertiary and quaternary spectra mixtures.

Table 2. Minerals used for the “non-jarosite” class while training the SVM detector.

<table>
<thead>
<tr>
<th>Martian igneous and weathering product minerals</th>
<th>Minerals often associated with jarosite</th>
</tr>
</thead>
<tbody>
<tr>
<td>Augite</td>
<td>Alunite</td>
</tr>
<tr>
<td>Chromite</td>
<td>Anhydrite</td>
</tr>
<tr>
<td>Clinochlore</td>
<td>Epsomite</td>
</tr>
<tr>
<td>Kaolinite</td>
<td>Ferrhydrite</td>
</tr>
<tr>
<td>Montmorillonite</td>
<td>Goethite</td>
</tr>
<tr>
<td>Olivine</td>
<td>Gypsum</td>
</tr>
<tr>
<td>Pigeonite</td>
<td>Hematite</td>
</tr>
<tr>
<td>Siderite</td>
<td>Lepidocrocite</td>
</tr>
<tr>
<td>Talc</td>
<td>Magnetite</td>
</tr>
</tbody>
</table>

Upon completion of the original detector, an additional detector was created with inputs tailored to avoid spectral absorptions expected in the Martian atmosphere. The atmosphere of Mars contains 95% CO2, 0.13% O2, 0.07% CO and 0.03% H2O [21], resulting in characteristic absorptions in the VIS/NIR [1]. The presence of CO2 and CO bands limits the input ranges of the new detector to 500–1260 and 2185–2310 nm.

5. DETECTOR TESTING

After training, the detector was tested on laboratory spectra of known jarosite samples from Wesleyan University's Peoples Museum. Spectra were collected with an ASD FieldSpec® FR operating over 350–2500 nm. Spectra were taken of multiple locations on each sample to include variations in crystal size and color. Additionally, 200 non-jarosite (evaporites and iron oxides) spectra of 22 pure mineral specimens were collected in the laboratory and analyzed to determine the detector's ability to correctly reject samples.

To better assess the sensitivity of the jarosite detector to mineral assemblages more typical of what is seen in the field, laboratory spectra were collected of samples taken from the Sulphur Springs hydrothermal field in St. Lucia in June and November 2004, where jarosite typically occurs as a hydrothermal alteration product [10]. Finally, the detector was run on spectra of Sulphur Springs samples directly taken in the field to test the sensitivity of the detector to changing light and atmospheric conditions and consequent signal to noise ratio. Jarosite was confirmed in a subset of the samples by X-ray Diffraction (XRD) and Scanning Electron Microscope Energy Dispersive Spectrometry (SEM-EDS), the remaining spectra were inspected and compared to published spectra [14, 13, 6, 12].
6. Results

Jarosite detector performance is indicated in Table 3. These results pertain to the averaged spectra of 201 samples, where 10–20 spectra were averaged. The jarosite detector performed well on spectra of museum quality samples taken under laboratory conditions. These samples are near pure jarosites and therefore most similar to the training spectra. The detector also performed well on the spectra of rock samples collected under both laboratory and field conditions, although detector outputs for data collected in the field have greater variation ($\sigma = 0.16$) than laboratory spectra of field samples ($\sigma = 0.024$).

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Category</th>
<th>Earth Atmosphere (% correct)</th>
<th>Mars Atmosphere (% correct)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Museum Quality Samples (in Lab)</td>
<td>Jarosite detection</td>
<td>8/9 (89)</td>
<td>7/9 (78)</td>
</tr>
<tr>
<td></td>
<td>non-jarosite rejection</td>
<td>22/22 (100)</td>
<td>20/22 (91)</td>
</tr>
<tr>
<td>Field Samples (in Lab)</td>
<td>Jarosite</td>
<td>21/29 (72)</td>
<td>23/29 (79)</td>
</tr>
<tr>
<td></td>
<td>non-jarosite</td>
<td>62/69 (90)</td>
<td>62/69 (90)</td>
</tr>
<tr>
<td>Field Samples (in Lab)</td>
<td>Jarosite</td>
<td>17/21 (81)</td>
<td>19/21 (91)</td>
</tr>
<tr>
<td></td>
<td>non-jarosite</td>
<td>46/51 (90)</td>
<td>43/51 (84)</td>
</tr>
<tr>
<td>Overall Performance</td>
<td></td>
<td>176/201 (88%)</td>
<td>174/201 (87%)</td>
</tr>
</tbody>
</table>

There was some consistency in the spectra misclassified by the detector. False positives generally tend to exhibit a large Fe$^{3+}$ absorption at 920 nm or a steep positive slope at 700 nm, both of which are characteristics of jarosite spectra. These features are associated with several iron oxides. Biological materials containing chlorophyll also commonly produced false positives due to a steep positive slope in this 700 nm region. Spectra that were incorrectly identified as non-jarosite by the detectors typically have either low reflectance and/or small band depths in the detector input ranges (500–1380 and 2050–2380 nm). The latter is indicative of mineral mixing where jarosite is at low concentration or fine grain size which increase scattering and therefore reduce the strength of characteristic absorptions.

While the Earth-based and Mars-based jarosite detectors performed similarly overall, the detectors classified spectra differently. The Mars-based detector was less likely than the Earth-based detector to assign false positives to biological samples, yet more likely to assign false positives to goethite spectra (absorption at ~900 nm). False negative assignment errors were similar between the two detectors.

These results do show the ability of the detector to recognize jarosite in natural mixtures both in the laboratory and field conditions. In these samples, jarosite is often found in intimate contact with alunite (typically (K,Na)Al$_3$(SO$_4$)$_2$(OH)$_6$), goethite FeO(OH) and gypsum CaSO$_4$$\cdot$2H$_2$O [10] (Figure 1). Spectra from the field samples frequently contain absorptions from two or more of these minerals. The detector is successful in distinguishing jarosite from this hydrated assemblage of sulfates and oxide.

7. CLARATY Integration

The Coupled Layer Architecture for Robotic Autonomy (CLARAty) [20] is a software infrastructure for autonomous mobile robotics. CLARAty provides an environment for researchers to develop, validate, and share robotics software components. The number and variety of highly specialized software components required to create autonomous mobile robots makes component sharing not only an effective form of software reuse, but nearly essential. By pooling software components in a single place and working to ensure consistent data structures and interfaces among them, researchers can leverage the work of one another. For instance, the developer of a path-planning module will likely require basic locomotion and machine vision capabilities. Rather than developing custom implementations of these modules from scratch, they can instead make use of the ones already present in CLARAty. Similarly, others can make use of their path-planning module. The Jet Propulsion Laboratory, NASA Ames, Carnegie Mellon University, and the University of Minnesota develop CLARAty with additional contributions from the University of Washington, the University of Michigan, and the Massachusetts Institute of Technology.

We chose to integrate our jarosite detector with CLARAty to make it available to researchers who may have little knowledge of geology and spectroscopy, but who could use it as a component of a larger rover autonomous traverse science scenario (e.g. [4]). CLARAty defines four levels of integration. At Level 1 (Deposited) a component must be under CLARAty revision control and compile and run as a standalone application. At Level 2 (Encapsulated) a component must integrate with other CLARAty components and run on at least one robot platform. The Level 2 integration requirement helps promote a cohesive and consistent software interface throughout CLARAty. Levels 3 and 4 require the component adhere to a CLARAty Application Programming Interface (API), undergo extensive peer, and run on all supported robot platforms. In addition to levels of integration, CLARAty defines two major layers of operation, the Functional Layer (FL) and the Decision Layer (DL). Components in the FL provide for basic rover navigation, locomotion, and state estimation. Components in the DL are responsible for higher-level decision-making including obstacle avoidance and path planning.
We have integrated the jarosite detector into CLARAty as a Level 2 Decision Layer component. We could not integrate the jarosite detector beyond Level 2 since a CLARAty API for spectral analysis has yet to be defined. The detector makes use of a preexisting CLARAty data structure for storing calibrated reflectance spectroscopy data. The detector interface is written in C++ according to CLARAty coding standards. However, the core of the detector is written in ISO C and has several desirable properties that make it particularly lightweight and amenable to running on flight hardware.

Recall the jarosite detector is an SVM with a linear kernel and nine 120 dimensional support vectors. Using the technique outlined in Section 2, we reduced the detector from nine dot products to a single dot product computation. Thus, only 120 floating-point multiplications, 1 subtraction and a comparison to zero are required per detection. To keep memory requirements to a minimum, the vector to dot is stored in static memory and the dot product is computed using register variables. No dynamic memory allocation is required.

8. CONCLUSIONS

We have developed two SVM-based jarosite detectors, designed to avoid Earth and Martian atmospheric absorption bands respectively. Both detectors correctly identify jarosite minerals in laboratory and field settings with overall accuracies of 88% (Earth-based) and 87% (Mars-based). The success of the detectors on field spectra demonstrates the potential ability of this technique to autonomously identify critical aqueous mineralogies on Mars. To bring these detectors one-step closer to flight hardware and make them available to researchers at several NASA centers and universities, we integrated them into the CLARAty autonomous mobile robotic software architecture. Finally, we demonstrated that the same detector construction techniques could be applied to hyperspectral imaging spectrometer data such as AVIRIS.

9. FUTURE DIRECTIONS

While our detectors were created to operate on ground-based point spectra, we believe similar detector construction techniques will yield detectors that perform well on hyperspectral images. We are currently attempting to create detectors tuned to a variety of airborne and spaceborne instruments, including AVIRIS and Hyperion at Earth and OMEGA and CRISM at Mars. There are at least two other systems for identifying minerals in hyperspectral images, Hypereye [18] and Tetracorder [5]. Our approach offers advantages over both. First, Hypereye makes use of unsupervised clustering to identify groups of similar spectra. It is still necessary for a domain expert to examine each spectral group and determine its end-member minerals. In contrast, since we make use of supervised machine learning techniques we can create detectors to target specific minerals. Second, while Tetracorder can target specific minerals, its expert system relies on rules that look for diagnostic absorption bands. To search for new minerals, new rules and reference spectra must be added, tested, and refined, which can be a labor-intensive process. Finally, we believe our linear mixing model is particularly apt for hyperspectral mineral detectors, since at orbital resolutions, mixing is well approximated as linear [23].

10. ACKNOWLEDGEMENTS

This research is supported by the NASA Applied Information Systems Research Program. XRD and SEM-EDS analyses were conducted at Yale University. Field assistance by J. Andrew Gilmore is appreciated. Special thanks to the Sulphur Springs Park in St. Lucia managed by the Soufriere Foundation.

Richard Maclin contributed to the development of the code used for the creation of the training data. In part, this work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

11. REFERENCES


12. Biographies

Ben Bornstein is a research programmer in the Machine Learning Systems group at the Jet Propulsion Laboratory in Pasadena, CA. He enjoys bringing machine learning techniques and considerable hacking (programming) skills to bear to solve problems in geology, bioinformatics, and systems biology. He has designed and implemented software systems for several Caltech biology labs, the Institute for Genomics and Bioinformatics (IGB) at UC Irvine, USC Children’s Hospital, and JPL’s Mars Exploration Rover (MER) project. He is also the inventor of and lead developer for LIBSBML, an open-source library for the Systems Biology Markup Language (SBML). Ben received a B.Sc. in Computer Science from the University of Minnesota Duluth in 1999 and is currently pursuing a M.Sc. in Computer Science at the University of Southern California.

Rebecca Castaño Supervisor, Machine Learning Systems (MLS) group at JPL. She received her Ph.D. in Electrical Engineering from the University of Illinois with her dissertation in the area of computer vision. Dr. Castaño has been advancing the state of the art in onboard science analysis methods for the past five years and has been lead author on numerous publications in this field. From 1999 - 2001, Dr. Castaño served as the application lead for phenomenological computational field geology efforts in the MLS Group. She is currently the technology lead for science data processing for the Autonomous Sciencecraft Experiment on the New Millenium Program’s ST6 project. Dr. Castaño is also the Team Lead of the Onboard Autonomous Science Investigation System (OASIS) project and the subtask lead for Data Analysis on the automated Multi-rover Integrated Science Understanding System (MISUS). Her research interests include machine learning, computer vision and pattern recognition.

Martha Gilmore is an Assistant Professor of Earth and Environmental Sciences at Wesleyan University. Prior to this she was a postdoctoral researcher in the Earth and Planetary Science Division at the Jet Propulsion Laboratory. Her primary research involves the use of images of the surfaces of Mars and Venus to interpret geologic process and history. Dr. Gilmore is also involved in the future of planetary exploration including strategies for the return of samples from Venus and Mars, the exploration of Venus with balloons, and the use of artificial intelligence to improve the geological capabilities of Mars rovers. She previously served as member of the Mars Surveyor 2001 lander team and has contributed to site selection for the Mars Exploration Rovers. Dr. Gilmore is a member of the Space Studies Board’s Committee for Planetary and Lunar Exploration. She received her Ph.D. in Geological Sciences from Brown University.

Matthew Merrill received his M.A. in Earth Sciences at Wesleyan University in 2005 and currently works at the USGS, Reston.

James Greenwood is a Research Assistant Professor of Earth and Environmental Sciences at Wesleyan University. Dr. Greenwood is an expert on petrology of martian and chondritic meteorites, with extensive experience using electron and ion microprobes. Recent work includes study of the isotopic signatures of microbes and their detectability at Mars. He was a member of the Mars Pathfinder mineralogy team and performs research on the nature and chemistry of martian soils. He received his Ph.D. in Geological Sciences from Brown University.