Mars Target Encyclopedia: Text-Based Machine Learning for Planetary Science

Kiri L. Wagstaff
(Jet Propulsion Laboratory, California Institute of Technology),
Raymond Francis (JPL), Thamme Gowda (ISI/USC/JPL),
You Lu (JPL), Ellen Riloff (University of Utah),
Karanjeet Singh (USC/JPL), Nina Lanza (LANL),
and Tom Stein (Washington University, St. Louis)

ASU SESE

November 1, 2018

© 2018, All rights reserved. CL#18-6220
This work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA.
Collect data

Uplink commands

Downlink data
What have we learned?

MSL Data Archive

MSL: 905K Mastcam, 1.7M Hazcam, 8.2M Navcam

2000+ per-sol summaries of data collected

What have we learned?
IRON METEORITE CANDIDATES WITHIN GALE CRATER, MARS, FROM MSL/MASTCAM MULTISPECTRAL OBSERVATIONS. D. F. Wellington (dfwellin@asu.edu), J. R. Johnson, P.-Y. Meslin, J. F. Bell III, Arizona State Univ., Johns Hopkins Univ., APL, IRAP, UPS-CNRS, Univ. Toulouse

Introduction: The Mastcam instruments on the MSL Curiosity rover each include an 8-position filter wheel that allows a wavelength range of 400-1000 nm. From observations made on the surfaces of Gale Crater, where water may have existed, both organic and inorganic deposits are possible. The presence of these materials could provide evidence for past habitability on Mars.

NEW ANALYSIS OF HYDROGEN ABUNDANCE IN THE SHEEPBED MEMBER OF YELLOWKNIFE BAY USING IN SITU GEOCHEMICAL DATA. S. Zbarneki, C. Hardgrove, T. S. J. Gabriel, M. Litvak, I. Mitrofanov, D. Lisov, S. F. Nowicki, W. Rapin, Arizona State University, Tempe, AZ, sczbarneki@asu.edu, Space Research Institute, RAS, Moscow, Russia, Los Alamos National Laboratory, Los Alamos, NM, Université de Toulouse, UPS-OMP, Toulouse, France, Institut de Recherche en Astrophysique et Planetologie, CNRS, UMR 5727, Toulouse, France.

Introduction: The Mars Science Laboratory (MSL) rover, Curiosity, entered a fluvio-lacustrine depression named Yellowknife Bay (YKB) in December 2012. Curiosity completed a detailed observational campaign as it traversed across several geologic units of the YKB formation, including the Sheepbed Member (SBM). Along the traverse, the Mastcam (MST) instrument acquired images of the YKB surface. The images were acquired with the Day-Night (DAN) instrument, a neutron monitor measuring the hydrogen (H) content of the subsurface. To compare DAN data to Mastcam observations, we analyzed Mastcam observations with user-created geochromatography data. The results of our analysis showed that the H content of the subsurface varies significantly across the YKB formation.

Previous studies have reported H abundances as wt. % Water-Equivalent-Hydrogen (WEH), assuming all H is bound in water, a convention we follow here. Litvak et al. (2014) reported WEH based on 2-layer DFN active modeling using variable Cl and top layer depth.

DETECTING AND CHARACTERIZING COMPRESSION-RELATED ARTIFACTS IN MARS SCIENCE LABORATORY MASTCAM IMAGES. H. R. Kerner, J. F. Bell III, H. Ben Amor, Arizona State University School of Earth and Space Exploration, Tempe, AZ 85251 (hkerner@asu.edu); Arizona State University School of Computing, Informatics, and Decision Systems Engineering, Tempe, AZ 85251.

Introduction: The Mastcam color imaging system on the Mars Science Laboratory Curiosity rover acquires images within Gale crater for a variety of geologic and atmospheric studies. Images are often JPEG compressed onboard the rover before being downlinked to Earth. While critical for transmitting images on a low-bandwidth connection, this compression style can result in small image artifacts most noticeable as anomalous brightness or color changes within or near 8x8 JPEG compression block boundaries. In high-frequency detail regions of some images, for example in regions showing fine layering or laminated in sedimentary rocks, the image must be retransmitted losslessly to avoid introducing problems in the scientific interpretation of the data. The process of identifying which images have been adversely affected by such compression artifacts is performed manually by the Mastcam science team, costing significant ex-
DRILL SITES AT GALE CRATER
Mars Science Laboratory

Aeolis Palus
John Klein
Cumberland
Windjana

Confidence Hills
Mojave
Telegraph Peak

Buckskin
Big Sky
Greenhorn

Lubango
Okoruso

Oudam

Drill hole diameter is ~1.6 cm.
High-Temperature, Perhaps Silicic, Volcanism on Mars Evidenced by Tridymite Detection in High-SiO2 Sedimentary Rock at Gale Crater, Mars
RV Morris, DT Vaniman, DF Blake, R Gellert... - 2016 - oro.open.ac.uk
... Curiosity at Gale Crater: The Mars Science La- boratory (MSL) rover, Curiosity, has been exploring sedimentary rocks within Gale crater since landing in ... of Aeolis Mons (aka Mount Sharp), drill powder was collected from a high-silica (74 wt% SiO2) outcrop named Buckskin (BK ... All 3 versions  Cite  Save

Alteration Mineralogy and Geochemistry at Swansea, Arizona: A Potential Analog for Brine-Basalt Interaction on Mars
JR Michalski, SJ Reynolds, TG Sharp... - AGU Fall Meeting ..., 2002 - adsabs.harvard.edu
... has been documented in the Harcuvar Mountains to the east, where deep (~10 km) crustal brines that migrated along the Tertiary Buckskin-Rawhide detachment fault ... Future work will include the development of a model for alteration of basalt by brine on the Earth and Mars. ...
Cite  Save

Silicic volcanism on Mars evidenced by tridymite in high-SiO2 sedimentary rock at Gale crater
RV Morris, DT Vaniman, DF Blake... - Proceedings of the ..., 2016 - National Acad Sciences
... Tridymite, a low-pressure, high-temperature (>870 °C) SiO2 2 polymorph, was detected in a drill sample of laminated mudstone (Buckskin) at Marias Pass in Gale crater, Mars, by the Chemistry and Mineralogy X-ray diffraction instrument onboard the Mars Science Laboratory ...
Cited by 2  Cite  Save

Is Tridymite at Gale Crater Evidence for Silicic Volcanism on Mars?
RV Morris, DT Vaniman, DW Ming, TG Graff, RT Downs... - 2016 - ntrs.nasa.gov
... The X-ray diffraction (XRD) instrument (CheMin) onboard the MSL rover Curiosity detected 17 wt% of the SiO2 polymorph tridymite (relative to bulk sample) for the Buckskin drill sample (73 wt% SiO2) obtained from sedimentary rock in the Murray formation at Gale Crater, Mars. ...
Cite  Save  More
What if I want to know:

- Which targets contain fluorine?
- Which targets contain hematite?
- Is there consensus on target X’s composition?

HELP!
Mars Target Encyclopedia

- Collect all published knowledge about every target on Mars
- Provide search access
- Answer questions
- Inspire new investigations and hypotheses
Can computers do the reading?

From flickr user Atomic Taco (CC BY-SA)
Can computers do the reading?

- How hard is it?

Johnson et al., 2016

Mountain in CA

Resort in MT, 2015 movie

“Hedging”

[Johnson et al., 2016]
Can computers do the reading?

- How hard is it?

[Johnson et al., 2016] Not even in the same sentence!
The documents

- Lunar and Planetary Science Conference
  - Three years
  - 5,920 documents
  - 2-page abstracts
  - 7.2M words
Mars Target Encyclopedia

Information Extraction

Entities
Find Elements, Minerals, Targets

Relations
Classify pairs of Target + (Element or Mineral)

MTE Database

User queries via web

Kiri Wagstaff et al.
1. Find Targets, Elements, Minerals

- Use known lists
  - What about newly discovered targets?

- Machine learning
  - Given example texts, learn word patterns
    - "The Big Sky tailings were spectrally flat…"
  - Stanford CoreNLP system [Finkel et al., 2005]
    - Word sequences, parts of speech, word “shape”
    - ML: Conditional Random Field (CRF) model
Entity extraction performance

Train on 62 docs from 2015, test on 35 docs from 2016

<table>
<thead>
<tr>
<th>Element</th>
<th>Mineral</th>
<th>Target</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lists</td>
<td>ML</td>
<td>ML+lists</td>
<td></td>
</tr>
<tr>
<td>0.825</td>
<td>0.825</td>
<td>0.825</td>
<td>1.000</td>
</tr>
</tbody>
</table>

F1 measure
2. Find relationships

- Predict whether there is a "contains" relation for each [Target, Element] or [Target, Mineral] pair
- Machine learning: jSRE [Giuliano et al., 2006]
  - Words, positions, endings, parts of speech
  - ML: Support Vector Machine (SVM) model

Example:

The mineralogy of Confidence Hills (Table 1) is dominated by plagioclase, augite and hematite.
Relation extraction performance

Train on 429 pairs from 2015, test on 173 pairs from 2016

F1 measure

Precision

0.7

0.525

0.35

0.175

0

All-yes

SVM

All-yes

SVM
Mars Target Database

<table>
<thead>
<tr>
<th>Content type</th>
<th>Manual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Documents</td>
<td>118</td>
</tr>
<tr>
<td>Elements</td>
<td>2,224</td>
</tr>
<tr>
<td>Minerals</td>
<td>1,456</td>
</tr>
<tr>
<td>Targets</td>
<td>916</td>
</tr>
<tr>
<td>Relations</td>
<td>696</td>
</tr>
</tbody>
</table>
# Mars Target Database

<table>
<thead>
<tr>
<th>Content type</th>
<th>Manual</th>
<th>Automatic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Documents</td>
<td>118</td>
<td>5,920</td>
</tr>
<tr>
<td>Elements</td>
<td>2,224</td>
<td>48,614</td>
</tr>
<tr>
<td>Minerals</td>
<td>1,456</td>
<td>34,287</td>
</tr>
<tr>
<td>Targets</td>
<td>916</td>
<td>3,255</td>
</tr>
<tr>
<td>Relations</td>
<td>696</td>
<td>1,412</td>
</tr>
</tbody>
</table>
## Mars Target Database

<table>
<thead>
<tr>
<th>Content type</th>
<th>Manual</th>
<th>Automatic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Documents</td>
<td>118</td>
<td>5,920</td>
</tr>
<tr>
<td>Elements</td>
<td>2,224</td>
<td>48,614</td>
</tr>
<tr>
<td>Minerals</td>
<td>1,456</td>
<td>34,287</td>
</tr>
<tr>
<td>Targets</td>
<td>916</td>
<td>3,255</td>
</tr>
<tr>
<td>Relations</td>
<td>696</td>
<td>1,412</td>
</tr>
<tr>
<td>Time (per document)</td>
<td>~30 mins</td>
<td>~5 seconds</td>
</tr>
</tbody>
</table>
Manual review of relations

- Relations from non-training documents
- ~5 sec per document

- Correct: 332
- Partial: 108
- Non-Mars: 87
- Wrong: 65
Example extractions – correct

- Link contains potassium
  - “Link, which was one of the first K-rich conglomerate targets observed with ChemCam, whereas felsic group 5 shows a higher Na/K ratio.”

- Link contains hydrogen
  - “Both of these are good candidates since, in Link at least, the hydrogen signature is relatively prominent.”

- JK/CB and olivine/magnetite
  - “The RN crystalline component is depleted in MgO and FeO relative to JK and CB because of the absence of olivine and enrichment of magnetite in the latter.”
Example extractions – incorrect

- Not a target (Mars region)
  - “While limited detections of phyllosilicates and hydrated silica are found in Acidalia and Utopia Planitia…”

- Not an element (Fm)
  - “The results indicate that the dip of the Shoemaker Fm impactite section…”

- Meteorite (not Mars, but same name as Mars target)
  - “Finally, the Bilanga diogenite has a model age that seems older but still similar within the error than basaltic and cumulative eucrites.”
Map display of search results

Thanks: Fred Calef (MMGIS)
Integration with MSL Analyst’s Notebook

https://an.rsl.wustl.edu/msl/
Integration with MSL Analyst’s Notebook

<table>
<thead>
<tr>
<th>Mineral in lit ref</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>akaganeite</td>
<td>2</td>
</tr>
<tr>
<td>alkali</td>
<td>1</td>
</tr>
<tr>
<td>anhydrite</td>
<td>2</td>
</tr>
<tr>
<td>apatite</td>
<td>1</td>
</tr>
<tr>
<td>augite</td>
<td>3</td>
</tr>
<tr>
<td>basanite</td>
<td>2</td>
</tr>
<tr>
<td>ca_sulfate</td>
<td>3</td>
</tr>
<tr>
<td>carbonate</td>
<td>1</td>
</tr>
<tr>
<td>chlorite</td>
<td>1</td>
</tr>
<tr>
<td>chloritesmectite</td>
<td>0</td>
</tr>
<tr>
<td>clinopyroxene</td>
<td>3</td>
</tr>
<tr>
<td>cristobalite</td>
<td>0</td>
</tr>
<tr>
<td>fayalitemagnetite</td>
<td>1</td>
</tr>
<tr>
<td>fe_smectite</td>
<td>2</td>
</tr>
<tr>
<td>fe_sulfate</td>
<td>1</td>
</tr>
<tr>
<td>feldspar</td>
<td>5</td>
</tr>
<tr>
<td>fluorapatite</td>
<td>1</td>
</tr>
<tr>
<td>forsterite</td>
<td>1</td>
</tr>
<tr>
<td>hematite</td>
<td>5</td>
</tr>
</tbody>
</table>

**Sol**

1 to 2002

Only 0
Integration with MSL Analyst’s Notebook
Integration with MSL Analyst’s Notebook

https://an.rsl.wustl.edu/msl/
Integration with MSL Analyst’s Notebook

https://an.rsl.wustl.edu/msl/

**Target Big Sky**
Target defined on Sol 1114 Site 50 / 592

**Compositional references - Mineral**

**Anhydrite**
"In comparison, the Big Sky sandstone has strikingly similar abundances of feldspar, pyroxene, amorphous constituents, anhydrite and other phases."

**Feldspar**
"In comparison, the Big Sky sandstone has strikingly similar abundances of feldspar, pyroxene, amorphous constituents, anhydrite and other phases."

**Hematite**
"The Big Sky tailings were spectrally flat (similar to Telegraph Peak) likely from the presence of magnetite, and include a weak downturn > 750 nm, possibly from minor hematite."
Summary

- Information extraction for scientific publications
- Enable searches not previously possible
- Facilitate scientific progress and exploration using artificial intelligence methods
- Make discoveries accessible to everyone

https://an.rsl.wustl.edu/msl/

Thank you: JPL MGSS program, NASA Planetary Data System, and MSL project