Machine Learning in Predicting Multi-Component Mineral Compositions in Gale Crater, Mars

SHAUNNA M. MORRISON,[1] FEIFEI PAN,[2] ANIRUDH PRABHU,[2] AHMED ELEISH,[2] PETER FOX,[2] OLIVIER C. GAGNE,[1] ROBERT T. DOWNS,[3] THOMAS F. BRISTOW,[4] ELIZABETH B. RAMPE,[5] DAVID F. BLAKE,[4] DAVID T. VANIMAN,[6] CHERIE N. ACHILLES,[7] DOUGLAS W. MING,[5] ALBERT S. YEN,[8] ALLAN H. TREIMAN,[9] RICHARD V. MORRIS,[5] STEVE J. CHIPERA,[10] PATRICIA I. CRAIG,[9] VALERIE M. TU,[5] NICHOLAS CASTLE,[9] PHILIPPE C. SARRAZIN,[11] DAVID J. DES MARAIS,[4] ROBERT M. HAZEN[1]

[1]Carnegie Geophysical Laboratory, Washington, DC
[2]Rensselaer Polytechnic Institute (RPI) Troy, NY
[3]University Of Arizona, Tucson, AZ
[4]NASA Ames Research Center, Moffett Field, CA
[5]NASA Johnson Space Center, Houston, TX
[6]Planetary Science Institute, Tucson, AZ
[7]NASA Goddard Space Flight Center, Greenbelt, MD
[8]Jet Propulsion Laboratory, Caltech, Pasadena, CA
[9]Lunar and Planetary Institute - USRA, Houston, TX
[10]Chesapeake Energy Corporation, Oklahoma City, OK
[11]SETI Institute, Mountain View, CA

To better understand the formational conditions and geologic history of the minerals found in by NASA MSL rover *Curiosity* in Gale crater, Mars the CheMin X-ray diffractometer team developed a crystal-chemical method to predict limited chemical compositions of the minerals observed in the CheMin samples [1,2]. In this study, we adapt a machine learning technique, Label Distribution Learning (LDL) [3], to predict multicomponent chemical compositions of Gale crater mineral phases, thereby allowing for more detailed petrologic interpretation of the geologic history of the martian surface.

LDL is a novel framework for classification problems with small datasets and has been widely applied to facial recognition problems such as age estimation. In this study, we adapt the LDL algorithm such that it can predict chemical elements (labels) and their abundances (degrees) for each martian mineral sample, based on crystallographic parameters. We evaluate performance using distance and similarity between label distributions as well as mean square error and also compare the results to traditional machine learning methods.

[1] Morrison et al. (2017) Am Min, 103(6): 848-856 [2] Morrison et al. (2017) Am Min, 103(6): 857-871 [3] Geng (2016) IEEE Transactions on Knowledge and Data Engineering, 28(7), 1734-1748