

Interactive Discovery of Chemical Structure in ChemCam Targets Using Gaussian Graphical Models

Diane Oyen and Nina Lanza

Los Alamos National Laboratory, Los Alamos, New Mexico
doyen [at] lanl.gov

Abstract

As the Curiosity rover continues to collect data from the surface of Mars, interactive semi-automated methods for analyzing the breadth of information are needed. Onboard Curiosity, the ChemCam instrument collects data on the chemical composition of the near-surface of rocks using laser-induced breakdown spectroscopy (LIBS). Thus ChemCam observes geochemical surface features that give clues to the past environmental conditions of the planet. Given the high number of ChemCam observations to date ($\sim 300,000$ spectra) and the high dimensionality of LIBS data (~ 6000 channels), it is not straightforward to evaluate trends in chemical composition to discover signatures of geological surface features. We learn a Gaussian graphical model (GGM) to identify chemical depth trends in the near-surface of a rock. The GGM gives a good summary of trends, but a domain expert needs additional information and control over the learning of the GGM to make interesting discoveries about the elements in the rock responsible for the GGM. We develop an interactive approach to learning a GGM which provides information about which data samples contribute the most to the learned structure and allows the user to adjust the relative importance of each sample to learn a new structure from a weighted sample covariance matrix. In the case of ChemCam data, this is equivalent to adjusting the importance of various elements when analyzing the relationships among LIBS spectra. We show that this approach is related to the problem of learning using privileged information and give other potential applications for our specific algorithm of learning a GGM using privileged information about sample importance.

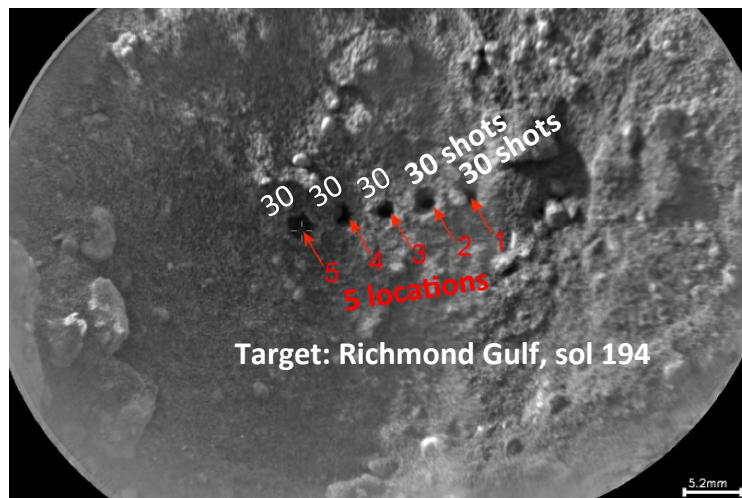
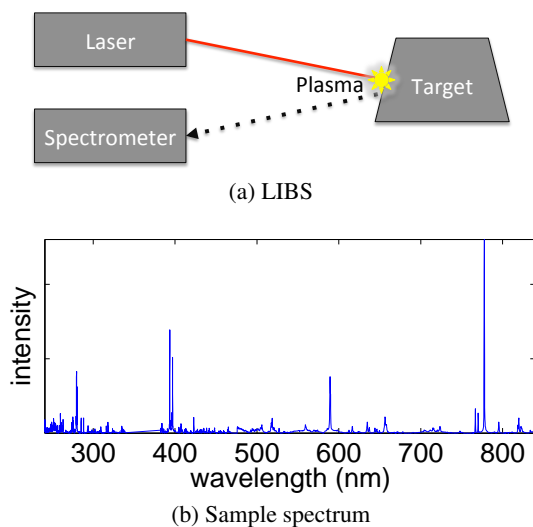
1 Introduction

The Curiosity rover landed in Gale Crater on Mars in 2012 and has been collecting a variety of data ever since. In particular, the ChemCam instrument samples the chemical composition of rocks and soils using laser-induced breakdown spectroscopy (LIBS) (see Figure 1 for a description of LIBS and

ChemCam) [Wiens *et al.*, 2012]. The chemical analysis of surface layers on rocks can give clues to the past conditions of the planet by revealing geological features such as surface coatings (from interaction with the environment), weathering rinds (from interaction with the atmosphere), and thin stratigraphic layers (rock deposit layers from sedimentation or volcanic activity). The ChemCam instrument provides the necessary data of chemical depth profiles with successive laser shots [Lanza *et al.*, 2015]. Machine learning approaches for analyzing this data are needed as the volume and complexity of the data are too large to analyze manually, but the learning objective is ill-defined because the goal of the analysis is open-ended scientific discovery [Wagstaff *et al.*, 2013]. To analyze depth trends in martian rocks, we apply graphical model structure learning to identify patterns representative of geological signatures. The approach that we take is to develop interactive machine learning algorithms where interaction requires two key components: transparency (providing explanations to the user) and user input.

In unsupervised learning, the goal is to discover interesting patterns in large and/or complex data. In scientific discovery, the domain experts know a lot about their data, even before applying any learning algorithms to it. We have found that for machine learning to be used in scientific discovery, transparency in the learning algorithm is critical. We provide transparency through (1) explanations about how the model was obtained from the data and (2) providing control to the end user over the learning process. Because the domain expert knows their data better than the learning algorithm, it is especially important for the end user to be able to explore how various inputs, including the training data itself, affect learned models. Therefore, we treat the problem of scientific discovery as a form of learning using privileged information [Vapnik and Vashist, 2009], in which there are data samples in the form of pairs (X_i, X_i^*) and we learn a generative model for $X \sim P(\theta)$ based on the samples X_i while utilizing the privileged information X_i^* about the samples. In this paper, we investigate the use of importance weighting of samples as the privileged information.

In many interactive machine learning scenarios, the user provides labels or in the case of graphical models, prior beliefs about the presence or absence of certain structures (usually edges) that should appear in the learned model [Mohan *et al.*, 2012; Werhli and Husmeier, 2007]. While that informa-



(c) ChemCam remote microimager (RMI) mosaic overlays a post-LIBS image on a pre-LIBS image.

Figure 1: (a) The ChemCam LIBS instrument shoots a laser at a target, creating a plasma. (b) Depending on the elements in its chemical composition, the plasma gives off different wavelengths of light which are measured by the spectrometer, producing an observed spectrum. (c) Each shot ablates the target surface, leaving a small pit. Typically, sequences of 30 or more shots are fired in each of several locations on a single rock (or soil) target.

tion can help, at times the type of information about the data that the expert can provide is quite different. In this work, we take the novel approach of providing a mechanism for interaction over the weighting of training samples in Gaussian graphical model structure learning. We learn Gaussian graphical models (GGM) interactively with input from the domain expert about which wavelengths are of highest interest. We start initially with a GGM learned from the full spectrum of wavelength bands returned by the ChemCam instrument. The initial GGM gives a summary of the geochemical gradients that produce the structure of the graphical model. However, some wavelengths tend to dominate the spectrum depending on the response of various elements (some elements burn “brightly” or are simply abundant in martian rocks). Often the domain expert would like to examine the geochemical patterns of other wavelength bands representing trace elements or elements that just do not show up as strongly in the spectral response.

While this work is motivated by a specific problem, it represents important research into interactive machine learning using privileged information and for transparency of machine learning algorithms. When learning complex solutions, such as probabilistic graphical model structure, from data; too often, a major limitation in the usefulness of such solutions is in helping the domain expert to understand what the solution means. By providing an interactive interface in which privileged information about the data being used to generate the model can be modified, we provide both a method to refine results toward shedding light on patterns in the most important data, as well as providing a means for the domain expert to investigate how data selection affects the learned model. Our interactive system is composed of these parts:

1. Learn a GGM and display it.

2. Display data used to produce the GGM and highlight which samples contribute most to the learned structure.
3. Collect feedback from user about which samples should be weighted more (or less) heavily.
4. Re-iterate from Step 1.

Other applications for this method would be in learning graphical model structure in biomedical domains. Often a large set of patients or samples have been collected, from which a graphical model will be learned. However, particularly in studies where it is impossible to control for all non-relevant variables (such as age, gender, medications of patients, etc), the domain scientist wants to see how the learned model differs when different subsets of data are used to learn the model [Liu *et al.*, 2010b; Oyen and Lane, 2012]. For example, a graphical model representing the functional brain network of patients with schizophrenia is often confounded by the variability of medications (and non-medical drugs) that schizophrenia patients typically take. Removing a subset of patients on a certain medication from the dataset would change the learned graphical model structure — or if the model structure does not change, that would also be useful information for the domain expert. Another application is learning relationships among malware samples in cybersecurity. The learned graphical model gives insights into how malware is developed including clues about the behavior and purpose of newly discovered malware code by relating it to existing malware that has already been evaluated manually. However, knowing which parts of the malware produce the discovered relationships is critical for speeding the evaluation of new malware, in addition to providing clues about the key components of malware that would likely continue to appear in future versions of malware.

2 Related Work

Our interactive approach is in the spirit of the observation from [Amershi *et al.*, 2011] that “treating a person simply as an oracle neglects human ability to revise and experiment. We therefore propose that a person instead consider ‘*how will different labels for these objects impact the system in relation to my goals?*’” However, instead of looking at labels of objects, we allow a person to investigate how selection and weighting of training samples impact the system.

The problem of providing transparency in machine learning algorithms has gained increasing attention in the past few years, particularly in unsupervised learning scenarios such as anomaly detection, often under the name of providing *explanations* [Senator *et al.*, 2013; Wagstaff *et al.*, 2013; Michelson and Macskassy, 2009]. Other methods for providing transparency focus on refinement of the learning objective through interactive control of learning parameters [Kapoor *et al.*, 2010; Oyen and Lane, 2014].

Interactive machine learning has also been studied, but most often the end goal is increased accuracy of supervised learning [Stumpf *et al.*, 2009; Abu-Mostafa, 1995] or to add supervision to unsupervised learning problems [Dubey *et al.*, 2010; Fogarty *et al.*, 2008; Kunapuli *et al.*, 2013]. Our work is most similar in spirit to [Dy and Brodley, 2000] which introduces interactive feature selection for clustering; although our work is in interactive sample selection for learning graphical model structure.

3 Geochemical Trends as Graphical Models

First we present our initial approach to modeling geochemical trends with standard graphical model structure learning. The ChemCam instrument onboard the Curiosity rover collects observations of the chemical composition of rock targets using laser-induced breakdown spectroscopy (LIBS) [Wiens *et al.*, 2012]. With each laser shot, the rock surface is ablated and therefore ChemCam produces a sequence of samples at increasing depth; revealing compositional trends such as coatings and weathering rinds (from interaction with water or atmosphere); and thin stratigraphic layers (from sedimentation or volcanic activity) that give clues about the past atmospheric and aqueous (water) conditions of Mars [Lanza *et al.*, 2015].

As shown in Figure 1, each LIBS shot produces a spectral observation consisting of 5810 wavelength bands between 224nm and 850nm. The spectral response is given as a table of intensity values for each wavelength band for each shot. A typical sequence of shots includes 30 - 150 shots on a fixed location. We model the correlations of rock chemistry among these shots, as measured by the sample covariance matrix calculated from the observed spectra. As pre-processing, we remove data for wavelengths above 850 nm, set all negative values to zero and normalize the values for each of ChemCam’s three component spectrometers separately. To investigate shot-to-shot correlations, shots are the vertices in the graph while the 5810 wavelength bands are treated as data samples.

Gaussian graphical models (GGM) [Zhao *et al.*, 2012; Meinshausen and Bühlmann, 2006; Friedman *et al.*, 2008]

estimate a sparse set of conditional dependency relationships (partial correlations) among a set of variables. A partial correlation between shot X and shot Y is the residual correlation after accounting for all other shots. If the partial correlation between X and Y is 0 then X and Y are conditionally independent and no edge is drawn between vertices X and Y in the graphical model.

A GGM is estimated from a data matrix \mathbf{X} , where each column $X_{\cdot j}$ is a shot j with spectral values X_{ij} for i in $\{1, \dots, n\}$ wavelength bands. The sample covariance matrix, Σ , is calculated from \mathbf{X} as

$$\Sigma_{jk} = \frac{1}{n} \sum_{i=1}^n (X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k), \quad (1)$$

where $\bar{X}_j = \frac{1}{n} \sum_{i=1}^n X_{ij}$ is the column mean. Then the sparse approximation, Θ , to the precision matrix for a given sparsity constraint, λ , is estimated according to the objective function

$$\hat{\Theta} = \arg \max_{\Theta > 0} [\log \det \Theta - \text{tr}(\Sigma^{-1} \Theta) - \lambda \|\Theta\|_1], \quad (2)$$

where, $\|\Theta\|_1 = \sum_{ij} |\theta_{ij}|$, is the ℓ_1 norm. The sparsity of $\hat{\Theta}$ is controlled by the value of λ , which can be any non-negative real number. If $\lambda = 0$, then all values in $\hat{\Theta}$ are non-zero. As λ increases, the more zeros that $\hat{\Theta}$ contains. Choosing an optimal value for λ is not straightforward, as many procedures for automatically choosing λ produce overly dense graphs [Liu *et al.*, 2010a]. We take the approach of using many values of λ and drawing a learned graph with edges weighted proportionally to how sparse a graph the edge appears in, providing more information to the end-user about the most robust edges, rather than selecting a single model under any one setting of the λ parameter.

Using ChemCam terminology as shown in Figure 1c, ChemCam typically fires 30 *shots* on a single *location* (producing a small pit with shots at increasing depth), and often several locations in a raster pattern per *target*. A ChemCam target is a rock, soil or calibration target; typically identified with a target name and the *sol* (martian day) on which it is sampled. Each shot ablates $\sim 0.3 - 0.5 \mu\text{m}$ of material [Lanza *et al.*, 2015; Wiens *et al.*, 2012].

As just one example of these targets, Figure 3 shows the graph estimated for sampling location 1 on the rock target Bell Island on sol 113 (pictured in Figure 2). It shows that the first several shots are directly dependent on the prior shot and the following shot in the sequence, creating a “chain” from Shot 1 through Shot 10. Without any prior information about the shot sequence, the GGM is able to reconstruct the shot sequence on the first 10 or so shots, which is consistent with observations that these first shots may differ from the underlying rock. In this case, the surface of Bell Island is extremely dusty, making dust a likely contributor to the spectra of the initial shots [Lasue *et al.*, 2014]. Additionally, the first sampling location of Bell Island appears to be a thin, sulfate-rich vein [Nachon *et al.*, 2014]. Our results suggest that in its 30-shot analysis at this location, ChemCam sampled through surface dust into the sulfate vein below. Examination of the spectral data confirm that the elements enriched in dust (e.g.

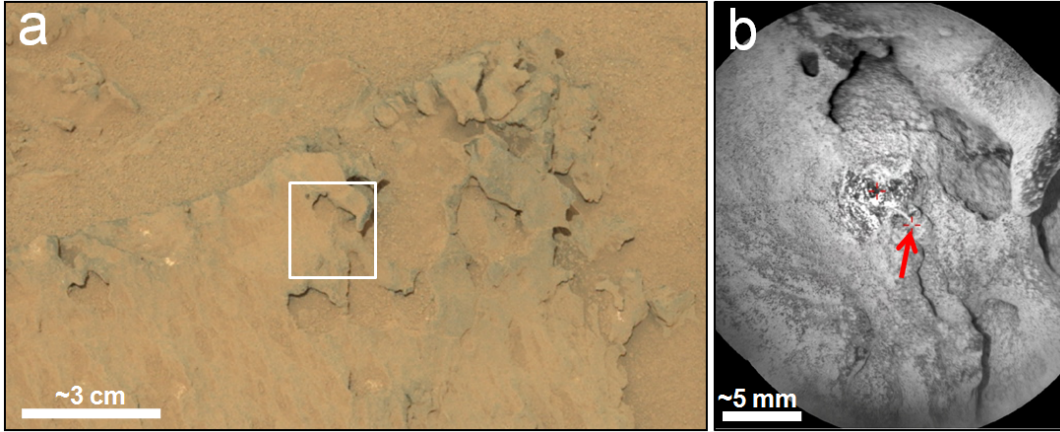


Figure 2: ChemCam target Bell Island, sampled on sol 113 (ccam02113). (a) Mastcam image of rock overview; the white box represents the ChemCam sampling region (mcam00694). (b) ChemCam remote microimager (RMI) mosaic of Bell Island. The red arrow indicates sampling location 1.

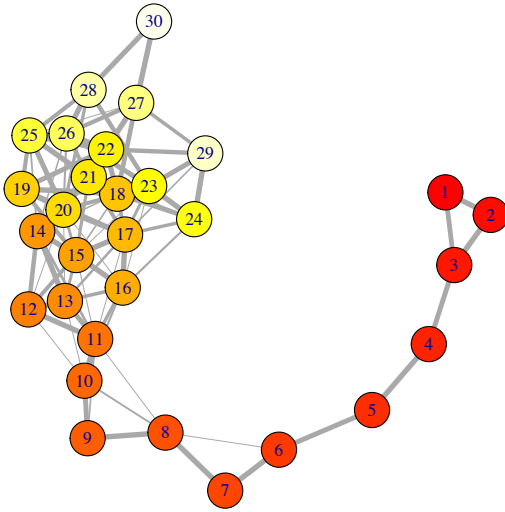


Figure 3: Graph of the shot-to-shot relationships in location 1 on Bell Island (30 shots, sol 113, ccam02113, pictured in Figure 2). Vertices of graph represent shots, numbered 1 through 30. Vertices are colored by shot number, starting with dark red at shot 1 and getting lighter to light yellow at shot 30 to aid visualization.

magnesium [Lasue *et al.*, 2014] decrease with depth (shot number) while sulfur and calcium increase.

4 Interactive Learning of Graphical Models

Our interactive approach includes two key components, one for giving an explanation of the learned model and the other component for allowing the user to change the relative weight of training samples.

4.1 Providing explanations

The learned graphical model gives a good summary of overall shot-to-shot trends in the chemistry of LIBS targets. How-

ever, to make this information useful for a domain scientist to make scientific discoveries, information about the specific elements contributing to these trends is necessary. Therefore, we provide information about which samples (in the LIBS case this translates to wavelengths) that contribute most heavily to the co-variance among connected vertices in the learned model. The contribution, c_i , of sample i is calculated as:

$$c_i(E) = \sum_{e_{jk} \in E} \phi_{jk}(X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k), \quad (3)$$

where $\phi_{jk} \geq 0$ is the learned weight of edge e_{jk} . For more detailed information, the user can select one or more edges in the learned model, defining E^* , a subset of all the learned edges E , and then calculate the contribution of samples to the local covariance in the neighborhood of the selected edges:

$$c_i(E^*) = \sum_{e_{jk} \in E^*} \phi_{jk}(X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k), \quad (4)$$

The largest contributors are highlighted, as shown in Figure 4, with gray rectangles behind the values associated with that sample (the greater the contribution, the darker the rectangle). For ChemCam LIBS data, the samples (x -axis) are wavelength bands observed by the spectrometer and therefore it makes sense to display the data this way. Samples from other data sources may need to be visualized differently, according to the form in which domain experts typically visualize their data.

4.2 User-defined sample weights

Often when learning a graphical model from data, the domain expert would like to explore what the graphical model would look like given a specific subset of data samples, or a non-uniform weighting on those samples. For example, in the neuroimaging domain, when learning a functional brain network, it might be apparent from the learned graph on the full sample that subjects taking a certain medication are affecting the results. By removing those subjects

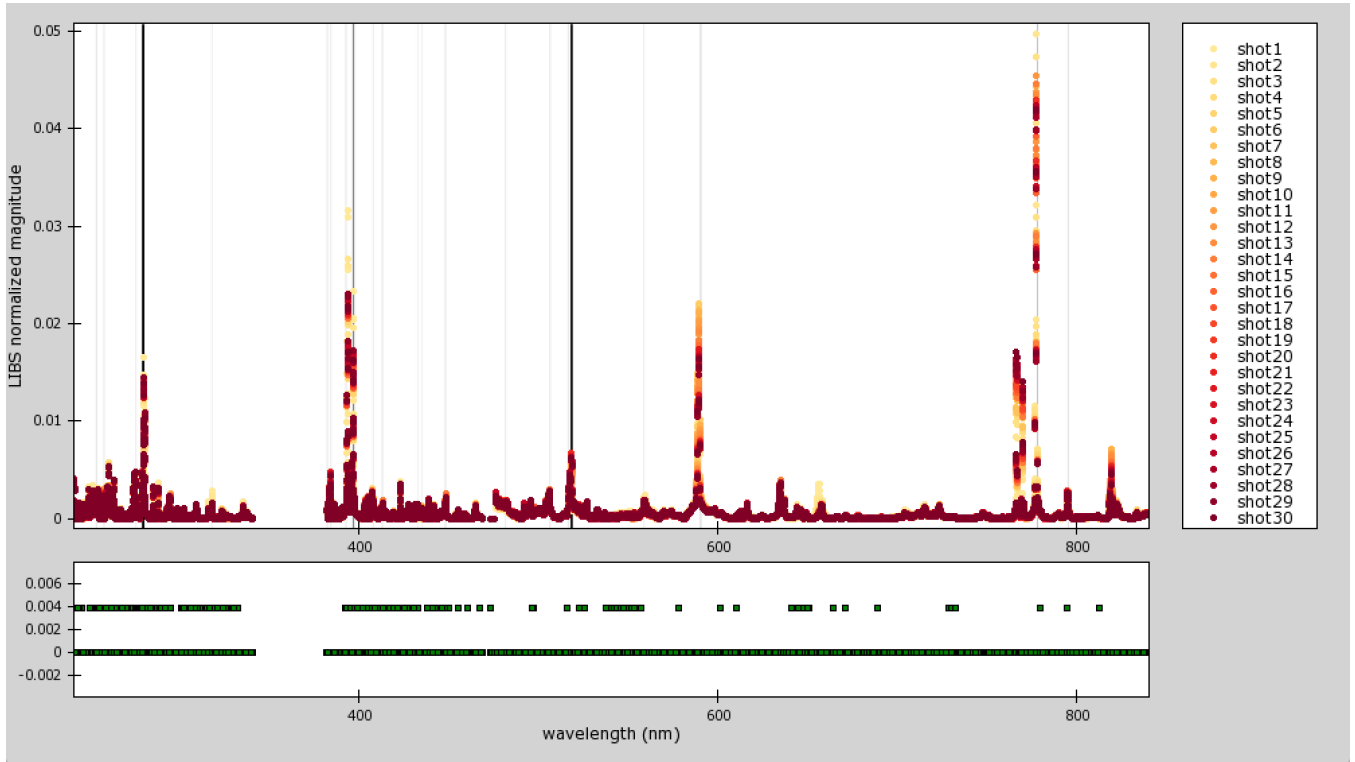


Figure 4: Interactive visualization of the samples that are the biggest contributors to the covariance calculation among adjacent vertices in the learned GGM. The bottom panel of the visualization allows the user to change the weighting of each sample. In this example, the wavelengths associated with six trace elements are evenly weighted while all others are zero-weighted.

from the sample, we learn a new graph [Liu *et al.*, 2010b; Oyen and Lane, 2012]. Similarly, in learning flavor (ingredient) networks from a collection of recipes, it might be desirable to remove or re-weight dessert recipes or recipes from specific geographic regions because they would produce different results [Ahn *et al.*, 2011; Oyen and Lane, 2014]. The end goal is to explore the relationships among the variables and how the set of samples affect these relationships. In our specific goal of this paper, we investigate chemical depth trends in rocks. Some elements are quite abundant in rocks on Mars or burn *brightly* in LIBS (the presence of the element produces a high magnitude in the spectrum disproportionate to the relative amount of the element in the rock sample). The domain expert would like to re-weight the relative importance of various elements (mapping to wavelength samples) to tailor the depth trend analysis, such as for finding trends in trace elements that would otherwise be overshadowed by more abundant and brighter elements.

Defining weights w_i as the privileged information X_i^* available for each data sample, we replace Equation (1) with a weighted sample covariance matrix, Σ' , with elements calculated as:

$$\Sigma'_{jk} = \sum_{i=1}^n w_i (X_{ij} - \bar{X}'_j)(X_{ik} - \bar{X}'_k), \quad (5)$$

where normalized weights w_i for $i \in \{1, \dots, n\}$ are provided interactively from an end user such that $\forall i, w_i \geq 0$

and $\sum_{i=1}^n w_i = 1$. The weighted mean is used, $\bar{X}'_j = \sum_{i=1}^n w_i X_{ij}$. The weighted mean and covariance are maximum likelihood estimators of normal distributions with different reliabilities of estimates for each sample. Thus, the weights are equivalent to judgments of reliability of the measurement. This is appropriate for estimating a graph with some high-variance signals down-weighted and for discounting the contribution of samples that are less important while increasing the weight of the contribution for the most important samples. The weighted covariance matrix is now used in the typical objective function, Equation (2), for estimating a GGM.

4.3 Interactive learning system

After incorporating user-defined weights, w_i , the new contribution of each sample is calculated as:

$$c'_i(E) = \sum_{e_{jk} \in E} w_i \phi_{jk}(X_{ij} - \bar{X}'_j)(X_{ik} - \bar{X}'_k), \quad (6)$$

and displayed in the visualization tool (Figure 4) along with the GGM learned from the weighted covariance matrix. The user can continue to explore how various wavelengths affect the depth trends as represented by the learned GGM.

We demonstrate the use of interactive GGM learning on the Mars rock target Stephen sampled by ChemCam (sol 619). Stephen is a particularly interesting target because close examination of its geochemistry indicates that there is a surface

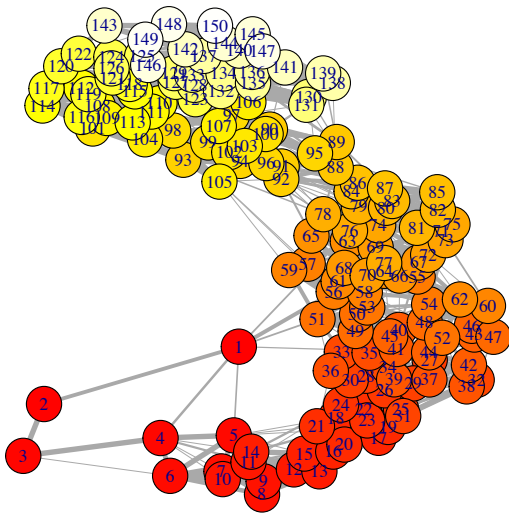


Figure 5: GGM of Stephen (150 shots, sol 619, ccam03619). Vertices are colored and labeled like Figure 3.

coating high in manganese, a geological phenomenon indicative of past atmospheric conditions that may have been habitable [Lanza *et al.*, 2015]. The goal of our interactive GGM learning is to discover such signatures more quickly than is possible by looking through the LIBS data manually.

Figure 5 shows the original GGM learned from the full set of data (unweighted), while Figure 6 shows the GGM learned when the samples are weighted so that a set of wavelengths associated with six trace elements (lithium, chromium, manganese, rubidium, strontium, and barium) are equally weighted and all other wavelengths are zero-weighted. Interesting patterns present themselves in the original unweighted data (Figure 5). The geochemistry shows a general depth trend along the 150 shots into the rock; while there seems to be a stratigraphic layer at about shot 89 where the geochemistry changes more rapidly. When the domain expert investigates how trace elements affect these depth trends, we see that the changes in trace elements are systematically consistent along most of the 150 shots into the rock. This finding is consistent with the observation that the trace element manganese, is present as a varnish on the outer rock and diminishes in abundance as samples are collected further into the rock [Lanza *et al.*, 2016].

5 Future Work

In problems in which the data set is comprised of subsets of interesting data, we could partition the data into a set of data *views* and use multi-view learning [Lian *et al.*, 2015; Anderson *et al.*, 2014]. The standard GGM learning model would need to be updated to handle multi-view data. The standard multi-view approach automatically learns to weight the contribution of each view in order to maximize some objective. In interactive graph discovery, we would allow the domain expert to adjust these weights interactively, similar to our current approach, but the partitioning of data into views significantly reduces the space of variables that the user

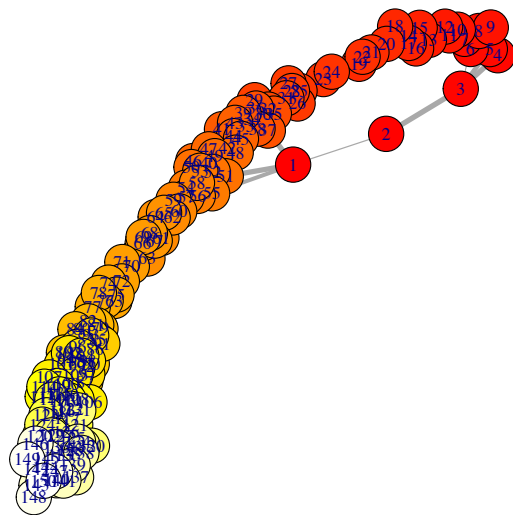


Figure 6: GGM of Stephen (150 shots, sol 619, ccam03619) when only the spectral bands of trace elements are considered. Colors and labels like Figure 3.

would need to explore. Data partitions would represent subsets of data that are meaningful to the domain expert.

The automatic partition of datasets for learning context-specific graphical models has been studied under constrained conditions [Liu *et al.*, 2010b]. Our current work provides a step toward an interactive approach that is more tractable than the fully-automated objective. We believe that a combination of automatic partitioning of data and user-interaction will ultimately give the best solution to this problem. This paper provides an interactive method for learning just one GGM at a time, in the future we will combine this with our multitask methods to learn multiple graphical models with the data partitioning determined interactively.

6 Conclusions

Interactive learning of graphical model structure allows a domain expert to explore relationships among variables in their data while capitalizing on the user’s knowledge about the samples within the dataset. This learning paradigm is the unsupervised learning equivalent of learning using privileged information. The use of machine learning in scientific discovery for uncovering signatures of scientific value will greatly benefit from interactive tools that provide explanations about how the learned model derives from the data and also gives some control to the end user to explore alternative models.

References

- [Abu-Mostafa, 1995] Y. Abu-Mostafa. Hints. *Neural Computation*, 7:639–671, July 1995.
- [Ahn *et al.*, 2011] Y.-Y. Ahn, S. E. Ahnert, J. P. Bagrow, and A.-L. Barabási. Flavor network and the principles of food pairing. *Scientific Reports*, 1, 2011.
- [Amershi *et al.*, 2011] S. Amershi, J. Fogarty, A. Kapoor, and D. Tan. Effective end-user interaction with machine learning. In *Twenty-Fifth AAAI Conference on Artificial Intelligence*, 2011.

- [Anderson *et al.*, 2014] B. Anderson, T. Lane, and C. Hash. Malware phylogenetics based on the multiview graphical lasso. In *Advances in Intelligent Data Analysis XIII*, pages 1–12, 2014.
- [Dubey *et al.*, 2010] A. Dubey, I. Bhattacharya, and S. Godbole. A cluster-level semi-supervision model for interactive clustering. *Machine Learning and Knowledge Discovery in Databases*, pages 409–424, 2010.
- [Dy and Brodley, 2000] J. Dy and C. Brodley. Visualization and interactive feature selection for unsupervised data. In *Sixth International Conference on Knowledge Discovery and Data Mining*, pages 360–364, 2000.
- [Fogarty *et al.*, 2008] J. Fogarty, D. Tan, A. Kapoor, and S. Winder. CueFlik: Interactive concept learning in image search. In *Computer Human Interface Conference on Human Factors in Computing Systems*, pages 29–38, 2008.
- [Friedman *et al.*, 2008] J. Friedman, T. Hastie, and R. Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.
- [Kapoor *et al.*, 2010] A. Kapoor, B. Lee, D. Tan, and E. Horvitz. Interactive optimization for steering machine classification. In *Proceedings of the 28th international conference on Human factors in computing systems*, pages 1343–1352, 2010.
- [Kunapuli *et al.*, 2013] G. Kunapuli, P. Odom, J. W. Shavlik, and S. Natarajan. Guiding autonomous agents to better behaviors through human advice. In *Data Mining (ICDM), IEEE 13th International Conference on*, pages 409–418, 2013.
- [Lanza *et al.*, 2015] N. L. Lanza, A. M. Ollila, A. Cousin, R. C. Wiens, S. Clegg, N. Mangold, N. Bridges, D. Cooper, M. Schmidt, J. Berger, et al. Understanding the signature of rock coatings in laser-induced breakdown spectroscopy data. *Icarus*, 2015.
- [Lanza *et al.*, 2016] N. L. Lanza, R. C. Wiens, R. E. Arvidson, B. C. Clark, W. W. Fischer, R. Gellert, J. P. Grotzinger, J. A. Hurowitz, S. M. McLennan, R. V. Morris, M. S. Rice, J. F. Bell, J. A. Berger, D. L. Blaney, N. T. Bridges, F. Calef, J. L. Campbell, S. M. Clegg, A. Cousin, K. S. Edgett, C. Fabre, M. R. Fisk, O. Forni, J. Frydenvang, K. R. Hardy, C. Hardgrove, J. R. Johnson, J. Lasue, S. L. Mouélic, M. C. Malin, N. Mangold, J. Martín-Torres, S. Maurice, M. J. McBride, D. W. Ming, H. E. Newsom, A. M. Ollila, V. Sautter, S. Schröder, L. M. Thompson, A. H. Treiman, S. VanBommel, D. T. Vaniman, and M.-P. Zorzano. Oxidation of manganese in an ancient aquifer, Kimberley formation, Gale crater, Mars. *Geophysical Research Letters*, 2016.
- [Lasue *et al.*, 2014] J. Lasue, S. Maurice, A. Cousin, O. Forni, P. Meslin, W. Rapin, S. Schroeder, A. Ollila, G. Berger, N. Bridges, et al. ChemCam analysis of martian fine dust. In *Lunar and Planetary Science Conference*, volume 45, 2014.
- [Lian *et al.*, 2015] W. Lian, P. Rai, E. Salazar, and L. Carin. Integrating features and similarities: Flexible models for heterogeneous multiview data. In *AAAI Conference on Artificial Intelligence*, 2015.
- [Liu *et al.*, 2010a] H. Liu, K. Roeder, and L. Wasserman. Stability approach to regularization selection (STARS) for high dimensional graphical models. In *Neural Information Processing Systems*, 2010.
- [Liu *et al.*, 2010b] H. Liu, X. Chen, J. Lafferty, and L. Wasserman. Graph-Valued regression. In J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems 23*, pages 1423–1431, 2010.
- [Meinshausen and Bühlmann, 2006] N. Meinshausen and P. Bühlmann. High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, 34(3):1436–1462, June 2006.
- [Michelson and Macskassy, 2009] M. Michelson and S. A. Macskassy. Layered, multivariate anomaly explanations: A first look. In *International Workshop on Statistical Relational Learning*, 2009.
- [Mohan *et al.*, 2012] K. Mohan, M. Chung, S. Han, D. Witten, S.-I. Lee, and M. Fazel. Structured learning of Gaussian graphical models. In *Advances in Neural Information Processing Systems 25*, pages 629–637, 2012.
- [Nachon *et al.*, 2014] M. Nachon, S. M. Clegg, N. Mangold, S. Schröder, L. C. Kah, G. Dromart, A. Ollila, J. R. Johnson, D. Z. Oehler, J. C. Bridges, S. Le Mouélic, O. Forni, R. Wiens, R. B. Anderson, D. L. Blaney, J. Bell, B. Clark, A. Cousin, M. D. Dyar, B. Ehlmann, C. Fabre, O. Gasnault, J. Grotzinger, J. Lasue, E. Lewin, R. Léveillé, S. McLennan, S. Maurice, P.-Y. Meslin, W. Rapin, M. Rice, S. W. Squyres, K. Stack, D. Y. Sumner, D. Vaniman, and D. Wellington. Calcium sulfate veins characterized by ChemCam/Curiosity at Gale crater, Mars. *Journal of Geophysical Research: Planets*, 119(9):1991–2016, 2014.
- [Oyen and Lane, 2012] D. Oyen and T. Lane. Leveraging domain knowledge in multitask Bayesian network structure learning. In *Twenty-Sixth AAAI Conference on Artificial Intelligence*, 2012.
- [Oyen and Lane, 2014] D. Oyen and T. Lane. Interactive exploration of comparative dependency network learning. In *KDD Workshop on Interactive Data Exploration and Analytics*, 2014.
- [Senator *et al.*, 2013] T. E. Senator, H. G. Goldberg, and A. Memory. Distinguishing the unexplainable from the merely unusual: Adding explanations to outliers to discover and detect significant complex rare events. In *Proceedings of the ACM SIGKDD Workshop on Outlier Detection and Description*, pages 40–45, 2013.
- [Stumpf *et al.*, 2009] S. Stumpf, V. Rajaram, L. Li, W. Wong, M. Burnett, T. Dietterich, E. Sullivan, and J. Herlocker. Interacting meaningfully with machine learning systems: Three experiments. *International Journal of Human-Computer Studies*, 67(8):639–662, August 2009.
- [Vapnik and Vashist, 2009] V. Vapnik and A. Vashist. A new learning paradigm: Learning using privileged information. *Neural Networks*, 22(5):544–557, 2009.
- [Wagstaff *et al.*, 2013] K. Wagstaff, N. Lanza, D. Thompson, T. Dietterich, and M. Gilmore. Guiding scientific discovery with explanations using DEMUD. In *Conference on Artificial Intelligence (AAAI)*, 2013.
- [Werhli and Husmeier, 2007] A. Werhli and D. Husmeier. Reconstructing gene regulatory networks with Bayesian networks by combining expression data with multiple sources of prior knowledge. *Statistical Applications in Genetics and Molecular Biology*, 6(1):Article15, 2007.
- [Wiens *et al.*, 2012] R. C. Wiens, S. Maurice, B. Barraclough, M. Saccoccio, W. C. Barkley, J. F. Bell III, S. Bender, J. Bernardin, D. Blaney, J. Blank, et al. The ChemCam instrument suite on the Mars Science Laboratory (MSL) rover: Body unit and combined system tests. *Space Science Reviews*, 170(1-4):167–227, 2012.
- [Zhao *et al.*, 2012] T. Zhao, H. Liu, K. Roeder, J. Lafferty, and L. Wasserman. The huge package for high-dimensional undirected graph estimation in R. *The Journal of Machine Learning Research*, 13(1):1059–1062, 2012.