

University of Massachusetts Medical School
eScholarship@UMMS

University of Massachusetts Medical School Faculty Publications

2020-06-28

Scalable learning of interpretable rules for the dynamic microbiome domain [preprint]

Venkata Suhas Maringanti
University of Massachusetts - Dartmouth

Et al.

Let us know how access to this document benefits you.

Follow this and additional works at: https://escholarship.umassmed.edu/faculty_pubs

 Part of the [Bioinformatics Commons](#), and the [Microbiology Commons](#)

Repository Citation

Maringanti VS, Bucci V, Gerber GK. (2020). Scalable learning of interpretable rules for the dynamic microbiome domain [preprint]. University of Massachusetts Medical School Faculty Publications. <https://doi.org/10.1101/2020.06.25.172270>. Retrieved from https://escholarship.umassmed.edu/faculty_pubs/1779

Creative Commons License



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](#).
This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in University of Massachusetts Medical School Faculty Publications by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.

Scalable learning of interpretable rules for the dynamic microbiome domain

Venkata Suhas Maringanti¹ Vanni Bucci² Georg K. Gerber³

Abstract

The microbiome, which is inherently dynamic, plays essential roles in human physiology and its disruption has been implicated in numerous human diseases. Linking dynamic changes in the microbiome to the status of the human host is an important problem, which is complicated by limitations and complexities of the data. Model interpretability is key in the microbiome field, as practitioners seek to derive testable biological hypotheses from data or develop diagnostic tests that can be understood by clinicians. Interpretable structure must take into account domain-specific information key to biologists and clinicians including evolutionary relationships (phylogeny) and dynamic behavior of the microbiome. A Bayesian model was previously developed in the field, which uses Markov Chain Monte Carlo inference to learn human interpretable rules for classifying the status of the human host based on microbiome time-series data, but that approach is not scalable to increasingly large microbiome datasets being produced. We present a new fully-differentiable model that also learns human-interpretable rules for the same classification task, but in an end-to-end gradient-descent based framework. We validate the performance of our model on human microbiome data sets and demonstrate our approach has similar predictive performance to the fully Bayesian method, while running orders-of-magnitude faster and moreover learning a larger set of rules, thus providing additional biological insight into the effects of diet and environment on the microbiome.

1. Introduction

The human microbiome refers to the collection of trillions of micro-organisms living on and within us. The microbiome is inherently dynamic, changing due to factors including maturation of the gut in childhood, diet, environmental exposures, and medical interventions (Gerber, 2014). Recent studies have investigated the microbiome longitudinally in the context of various human diseases (e.g., diabetes (Kostic et al., 2015) or mortality after hematopoietic-cell transplantation (Peled et al., 2020)). Analyzing such time-series data is key to robustly linking the microbiome to disease and uncovering potentially causative relationships.

However, extracting meaningful information from these data sets is challenging. Human microbiome time-series studies have relatively few subjects, exhibit high degrees of subject-to-subject variability, and are typically imbalanced and irregularly temporally sampled (Gerber, 2014). Inherent aspects of microbiome data present additional challenges, including high-dimensionality, compositionality, zero inflation and complex dependencies among variables (Tsilimigras & Fodor, 2016).

Model interpretability is also very important in microbiome studies, because researchers usually seek to generate testable biological hypotheses or inform clinical decisions. A key component of interpretability in this domain is phylogeny, or information about the evolutionary relationships among microbes, which is typically encoded in a tree structure with extant microbes or Operational Taxonomic Units [OTUs] forming the leaves of the tree. Such information is used by practitioners to interpret the relatedness of microbes identified through sequencing data, which allows one to make educated guesses about the phenotypic properties of the microbes (e.g., their potential to produce immune-modulating compounds) and to define conditions for isolating the microbes via culture-based techniques. Moreover, for microbiome data that is dynamic, meaningful interpretation must also take into account temporal information, e.g., *when* in the first year of life do we see changes in the microbiome that differentiate breast-fed versus formula-fed infants?

A state-of-the-art method in the microbiome field that addresses the above-mentioned challenges is MITRE (Bogart et al., 2019), which uses a fully Bayesian model that learns human-interpretable rules to classify the host's status (e.g.,

¹Department of Computer and Information Science, University of Massachusetts Dartmouth, MA, USA ²Department of Microbiology and Physiological Systems, University of Massachusetts Medical School, Worcester, MA, USA ³Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. Correspondence to: Vanni Bucci <vanni.bucci2@umassmed.edu>, Georg Gerber <ggerber@bwh.harvard.edu>.

Scalable learning of interpretable rules for the dynamic microbiome domain

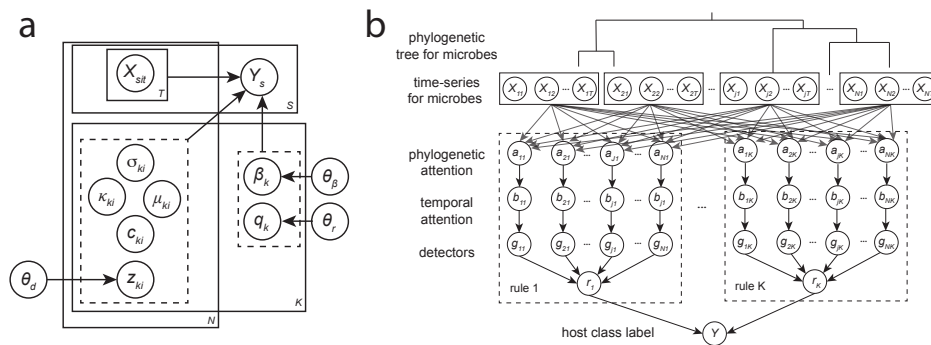


Figure 1. (a) Our probabilistic model. The binary class label Y_s for each subject is dependent on the input data X_{sit} (microbiome time-series) and variables for K rules, each comprised of variables for N detectors. Indicator variables q and z select which rules or detectors are active and have priors with parameters θ_r and θ_d respectively. Rule weights β have a prior with parameters θ_β . Detector variables μ and σ specify temporal attention filter centers and bandwidths respectively; κ specify phylogenetic attention filter bandwidths; and c specify detector thresholds. **(b) Layered representation of our model.** Our model can be viewed as a 5-layer Neural ILP-type model that assigns labels to hosts given input microbial time-series data and a graph structure based on phylogenetic relationships between bacteria. The outputs of each layer are composed to form the final output of human-interpretable rules that produce the classification label.

healthy or diseased) based on microbiome time-series data. MITRE rules consist of conjunctions of *detectors* that handle dependencies in both microbiome and time-series data. These detectors are of the form: “*TRUE if the aggregated abundance of organisms in phylogenetic subtree A within time window T is above threshold B.*” The rules are combined via logistic regression to model the probability of the classification label. This approach, which performs a set of nonlinear but interpretable and domain-specific transformations on the covariates, was demonstrated to perform generally on par or often outperform “black-box” machine learning methods (e.g., Random Forests) and produce biologically meaningful results on validation data sets. To perform inference, MITRE uses a Markov Chain Monte Carlo algorithm with custom Metropolis Hastings moves that operate combinatorially on a large space of pre-computed features (e.g., abundance thresholds and phylogenetic subtrees.) Consequently, MITRE’s runtime can be very slow on even moderately sized data sets, in some cases exceeding a day.

To address the challenge of scalability to large data sets while maintaining interpretability, we developed a novel 5-layer Neural Inductive Logic Programming (ILP)-type model with domain-specific microbiome and temporal attention mechanisms that outputs human-interpretable classification rules. Our method is inspired by MITRE and can be viewed as a fully-differentiable relaxation amenable to fast gradient descent-based inference. Several relaxation techniques are employed including the concrete distribution for binary selector variables (Maddison et al., 2016) to encode model sparsity and approximation of the Heaviside function to encode domain-specific attention mechanisms. We use an efficient gradient descent-based inference ap-

proach, and demonstrate on three human microbiome data sets that our model’s classification performance is similar to that of MITRE, while maintaining interpretability and running orders of magnitude faster. Moreover, our model learns a richer set of classification rules that suggest further biological insights into the effects of diet and environment on the microbiome.

2. Related work

Machine learning for microbiome time-series data. Although a variety of machine learning approaches have been applied to static microbiome data (Vangay et al., 2019; Haran et al., 2019; Knights et al., 2011a;b), there remains a dearth of microbiome time-series specific methods. Typical analysis methods in the literature include performing statistical tests for differential abundance at each time-point individually or over manually-selected temporal windows of interest for each taxonomic level to find differences between two cohorts (Arrieta et al., 2015). Such approaches obviously do not effectively use temporal information and moreover cannot find interactions between microbes. Mixed-effects models have been developed for analyzing microbiome data with repeated measurements (Chen & Li, 2016; Bokulich et al., 2018). However, these approaches do not model interactions or other nonlinearities between microbes. As noted, phylogenetic structure is important for microbiome data analysis and various models have been developed that take into account phylogeny (e.g., (Tang et al., 2017)), but these methods are generally not applicable to time-series data. Our work is directly inspired by MITRE (Bogart et al., 2019), which is currently the state-of-the-art for microbiome time-series specific classification methods. However, MITRE’s MCMC-based inference algo-

Scalable learning of interpretable rules for the dynamic microbiome domain

rithm suffers from scalability issues, effectively sampling over a discrete combinatoric space of phylogenetic subtrees, time windows and abundance thresholds that must be pre-enumerated. In contrast, our approach relaxes the problem to a continuous space, thus enabling use of highly efficient optimization tools.

Differentiable rule learning. Our approach is related to recent work that seeks to unify Inductive Logic Programming (ILP) with deep learning architectures amenable to gradient-based optimization, including (Evans & Grefenstette, 2018; Yang et al., 2017; Madsen & Johansen, 2020). ILP is a broad field concerned with learning logical rules from data, often with a goal of inferring human interpretable representations. For example, Differentiable Inductive Logic Programming (Evans & Grefenstette, 2018) proposes a framework to learn symbolic rules on small noisy data sets and can also generalize to non-symbolic domains such as images. Other examples include Neural Logic Programming (Yang et al., 2017), a framework that learns first-order logical rules for mining knowledge bases and Neural Arithmetic Units (Madsen & Johansen, 2020) that learns basic arithmetic operations that can extrapolate well to numbers out of the range of training data. These methods share a common theme with our model in that they induce direct and task-specific inductive biases with the goal of improving generalization, robustness and interpretability. In our case, we incorporate into our model inductive biases via temporal and phylogenetic attention mechanisms that leverage the domain-specific structure of microbiome time-series data.

Interpretability in deep models. Deep learning models have achieved state-of-the-art predictive performance on many tasks in fields such as computer vision, natural language and speech processing. Although predictive performance is an important metric, model interpretability is also essential as these systems are increasingly being deployed in high-risk scenarios for augmenting human decision making (Lipton, 2016; Doshi-Velez & Kim, 2017). A variety of *post hoc* interpretation methods or methods that remove features and evaluate the impact on the model have been described, but the effectiveness of these methods remains unclear, e.g., (Hooker et al., 2019). Our work differs from these methods by constructing a model encoding a specific structure that is inherently interpretable throughout the layers and in the final output.

3. Model

3.1. Overview

Our model classifies hosts (e.g., individual human subjects) given temporal, high-dimensional measurements of their microbes, while incorporating domain-specific inductive biases to facilitate interpretability. This structure includes

aggregation of covariates into features that preserve phylogenetic and temporal information in the data and selection of a sparse set of features. Our model learns which covariates to attend to for the classification task, with attention influenced by the dependence structure among the covariates.

Figure 1A shows our model as a graphical (plate) representation and Figure 1B shows it in layered form. Let Y_s denote the binary label for subject s . Let X_{sit} denote a microbiome measurement from subject s at time t for OTU i of an N -dimensional data source (e.g., relative abundances of OTUs from 16S rRNA amplicon or metagenomic shotgun sequencing). The task is then to output the probability of the label for each subject given the data for that subject, e.g., $P(Y_s | \mathbf{X}_s)$. We further assume there exists a graph structure on the OTUs, such that d_{ij} specifies the weight of the edge between OTU i and j (i.e., derived from a phylogenetic tree.)

We model $P(Y_s | \mathbf{X}_s)$ as a logistic regression over rules. Each rule is comprised of conjunctions of *detectors* of the form “(SOFT) TRUE If the aggregated abundance of organisms in phylogenetic cluster A within time region T is above threshold B .” Note that in the original MITRE model that served as an inspiration for our model, strict phylogenetic subtrees are used, whereas in our model we use clusters (e.g., nearby OTUs in phylogenetic space) to facilitate differentiability. Our model additionally relaxes strict time windows to time regions and hard thresholds to soft decision boundaries. Associated with each relaxation is a sharpness parameter. These parameters are annealed during our gradient descent-based learning algorithm, to arrive at detectors that approximate the discrete nature of the original MITRE detectors. We implemented our model in PyTorch (Paszke et al., 2019); source code is available at <https://github.com/gerberlab/mditre>.

3.2. Rules

We assume a maximum of K rules with each rule comprised of a maximum of N detectors. Let $g_{ki}(\cdot)$ denote binary valued detector i in rule k and z_{ki} denote a binary variable that selects whether the detector is active (value of 1) in the rule. A binary valued *hard* rule $r_k(\cdot)$ would be the logical conjunction of active detectors:

$$r_k(\cdot) = \bigwedge_{i=1}^N g_{ki}(\cdot)^{z_{ki}} \quad (1)$$

In order to make our model differentiable, we relax the logical conjunction using the following approximation:

$$\tilde{r}_k(\mathbf{X}_s) = \prod_i^N (1 - z_{ki}(1 - g_{ki}(\mathbf{X}_s))) \quad (2)$$

This formulation is inspired by the Neural Multiplication Unit from (Madsen & Johansen, 2020). Additionally, we

Scalable learning of interpretable rules for the dynamic microbiome domain

relax z_{ki} by modeling it as a binary concrete (Maddison et al., 2016) variable.

As described below, the detector functions $g_{ki}(\cdot)$ will effectively learn focused features from the data: (a) localizing regions in the phylogenetic space and, (b) time-periods in the time-series.

3.3. Detectors

3.3.1. PHYLOGENETIC ATTENTION MECHANISM

We use a relatively sharp (but smooth) averaging filter that serves as an attention mechanism over the graph on OTUs:

$$a_{ki}(\mathbf{X}_s) = \sum_j^N u_{kij} \mathbf{X}_{sj} \quad (3)$$

$$u_{kij} = \frac{h(\kappa_{ki} - d_{ij}, m)}{\sum_j^N h(\kappa_{ki} - d_{ij}, m)} \quad (4)$$

Here, u_{kij} is a normalized weight for OTU j and κ_{ki} controls the filter bandwidth. The function h is an analytic approximation to the Heaviside step function:

$$h(x, m) = \frac{1}{1 + e^{-2mx}} \quad (5)$$

The annealing parameter m controls the sharpness of the response, i.e., in the limit of large m the detector will include all OTUs within a radius κ . Intuitively, this filter “pulls in” other OTUs phylogenetically close to central OTU i .

3.3.2. TEMPORAL ATTENTION MECHANISM

We similarly use a relatively sharp (but smooth) averaging filter which serves as an attention mechanism over time:

$$b_{ki}(\mathbf{X}_s) = \sum_t^T v_{kit} a_{kit}(\mathbf{X}_s) \quad (6)$$

$$v'_{kit} = h(t - \mu_{ki} + \sigma_{ki}/2, m) - h(t - \mu_{ki} - \sigma_{ki}/2, m) \quad (7)$$

$$v_{kit} = \frac{v'_{kit}}{\sum_t^T v'_{kit}} \quad (8)$$

Here, μ_{ki} controls the center of the filter and σ_{ki} its bandwidth. As the annealing parameter m increases, this will approach a time window of duration σ and center μ .

3.3.3. THRESHOLDS

We model the response of detector g_{ki} as:

$$g_{ki}(\mathbf{X}_s) = h(b_{ki}(\mathbf{X}_s) - c_{ki}, m) \quad (9)$$

Here, c_{ki} specifies the value for a soft-threshold. Again, as the annealing parameter m increases, this will approach a hard threshold.

3.4. Classifier

Finally, we model the class probability for subject s via logistic regression with the rules as covariates:

$$\hat{y}_s = P(Y_s = 1 | \mathbf{X}_s) = \frac{1}{1 + e^{-\beta_0 - \sum_{k=1}^K q_k \beta_k \bar{r}_k(\mathbf{X}_s)}} \quad (10)$$

Here q_k is a binary concrete (Maddison et al., 2016) variable that selects which rules to include in the model.

3.5. Priors and Loss Function

To encourage parsimonious classifiers, we place priors on the number of rules and detectors. Specifically, we place Negative Binomial Distributed (NBD) priors on the total number of active rules $\sum_k^K q_k$ and the number of active detectors per rule $\sum_i^N z_{ki}$ with mean 0 and variance 100 in both cases, thus encoding sparsity with a prior belief of no rules or detectors. We also place a diffuse Gaussian prior on the regression coefficients β_k with mean 0 and variance 10^5 ; for other parameters we assume improper flat priors.

Our loss function is then equal to the negative log likelihood of the posterior and we perform maximum a posteriori (MAP) estimation:

$$L = L_{ce}(\hat{\mathbf{Y}}, \mathbf{Y}) + L_{nb}^{(r)}\left(\sum_k^K q_k\right) + \sum_k^K L_{nb}^{(d)}\left(\sum_i^N z_{ki}\right) + L_g(\beta) \quad (11)$$

Here, L_{ce} is the standard cross-entropy loss, $L_{nb}^{(r)}$ and $L_{nb}^{(d)}$ are the negative logs of the NBD priors and L_g is the negative log of the Gaussian prior.

4. Experiments

4.1. Data sets

We evaluated the performance of our model on human microbiome data sets originally assessed by MITRE (Bogart et al., 2019), using the same preprocessing and class labels as in the original study, in order to provide comparable results. These data sets consist of 16S rRNA amplicon sequencing data, which identifies the relative abundances of OTUs in the sample. David et al. (David et al., 2014) (20 subjects, 185 OTUs, class label = plant vs. animal diet) studied the effects of short-term dietary changes on the human gut microbiome. They tracked the gut microbiome composition of healthy adults before, during, and after a 5-day period of consuming exclusively plant-based or exclusively animal-based diets. Bokulich et al. (Bokulich et al., 2016) (35 subjects, 60 OTUs, class label = formula vs. breastfed diet) studied the effects of antibiotic exposures, cesarean section, and formula feeding on the gut microbiome composition during the first 2 years of infancy. Vatanen et al. (Vatanen et al., 2016) (113 subjects, 237 OTUs, class label

Scalable learning of interpretable rules for the dynamic microbiome domain

Table 1. Cross-validated F1 score and runtime comparison for MITRE and our model (DR = differentiable rules) on the data sets analyzed. We report the mean and standard deviation of F1 score for our model on 10 random seeds, with the runtime given in parenthesis. Scores and runtime of MITRE are taken directly from (Bogart et al., 2019), which provided only mean performance statistics.

MODEL	DAVID (DAVID ET AL., 2014)	VATANEN (VATANEN ET AL., 2016)	BOKULICH (BOKULICH ET AL., 2016)
MITRE	0.77 [23 HRS]	0.83 [64.50 HRS]	0.64 [30 HRS]
DR	0.81 ± 0.07 [0.45 HRS]	0.82 ± 0.05 [0.83 HRS]	0.58 ± 0.1 [1 HR]

= Russian vs. Estonian/Finnish nationality) tracked the gut microbiome composition from birth to 3 years of age in cohorts of children at high risk for autoimmune disease.

4.2. Model training details

We set $K = 10$ and $N = \# \text{ OTUs}$ for the maximum number of rules and detectors per rule respectively. We used full-batch gradient descent with the Adam optimizer (Kingma & Ba, 2014) and cosine learning rate annealing schedule over 1000 training epochs. We used a nested 5-fold cross-validation scheme to determine the number of epochs to train. The sharpness parameter m of the Heaviside step function approximation and temperature of the binary concrete variables are linearly annealed during training. Our model was run on a single Tesla V100 GPU with 32-bit precision training. Please see our source code for full details on training, including the initialization used.

4.3. Performance Results

We used cross-validated F1 scores (harmonic mean of precision and recall), as in the original MITRE paper, to assess generalization performance. Table 1 shows the performance and runtime comparison between MITRE and our model. Our results indicate that our differentiable model performs quite similarly to the fully Bayesian method, while running approximately 30-70X faster, suggesting our approach provides greater scalability without sacrificing performance.

4.4. Model Interpretability

We show that our model learns rules that are human-interpretable and moreover provide insights in the microbiome domain not seen with the fully Bayesian method. We note also that our model learned a different number of rules and average number of detectors per rule on each of the data sets (see below), suggesting that it automatically expands model capacity as needed. This capability is part of the inference procedure and does not require separate hyperparameter tuning sweeps, which would be very expensive over the large combinatorial space.

David (David et al., 2014). Our model learned 2 rules with 1 detector per rule. Both detectors are temporally mean-

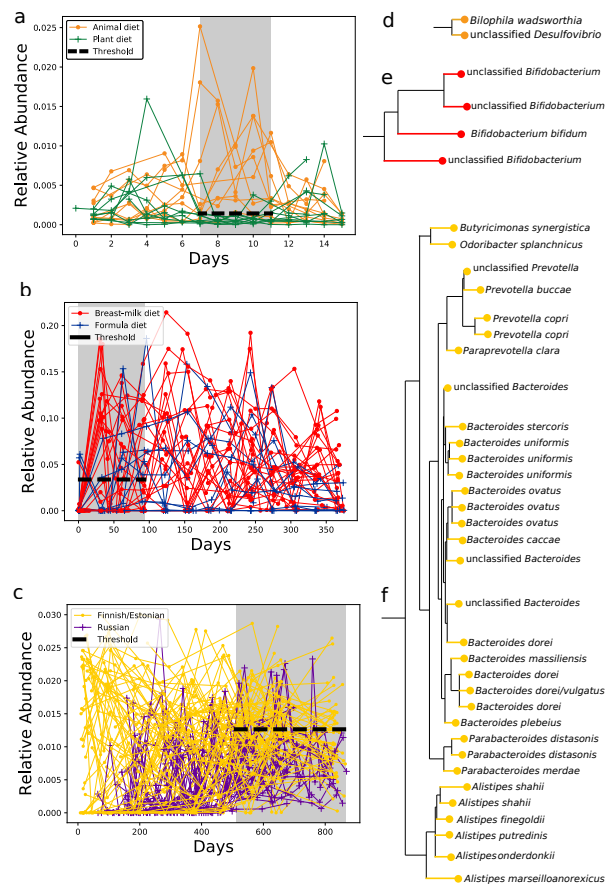


Figure 2. Visualization of example detectors. Panels (a-c) show detector time-windows (shaded region) and thresholds (black dashed line) superimposed on data (aggregated relative abundances over time for Operational Taxonomic Units [OTUs] selected by the detector.) Panels (e-f) show the OTUs selected by the detector and their phylogenetic context. Figures (a) and (d): detector discriminating study participants eating plant versus meat-based diets, (b) and (e): detector discriminating infants on breast-milk versus formula predominant diets, (c) and (f): detector discriminating children of Russian versus Finnish/Estonian nationalities.

ingful: the learned time windows occur during the dietary intervention (day 5 to day 10). Moreover, the detectors are biologically meaningful and provide phylogenetic context

Scalable learning of interpretable rules for the dynamic microbiome domain

that greatly facilitates interpretation. The first rule/detector reads: “*TRUE if the aggregated relative abundance of OTUs [Operational Taxonomic Units] in the family Eubacteriaceae and in the genera Roseburia is above 0.49% between day 7 and day 10*” and is similar to one found by MITRE. The second rule/detector (Figure 2a) reads: “*TRUE if the aggregated relative abundance of OTUs from the species *Bilophila wadsworthia* and genera *Desulfovibrio* is above 0.31% between day 7 and 11.*” This detector was not found by MITRE, but is clearly biologically meaningful. The detector differentiates subjects on the animal-based diet by increased abundances of two taxa belonging to the species *Bilophila wadsworthia* and to the genus *Desulfovibrio*. *B. wadsworthia* is a sulfite-reducing pathobiont shown to increase in abundance with dietary lipids (Devkota et al., 2012) and *Desulfovibrio* is a sulfate-reducing bacterium shown to increase in abundance with chondroitin sulfate (an important component of animal cartilage) (Rey et al., 2013).

Bokulich (Bokulich et al., 2016). Our model learned 5 rules with an average of 1.5 detectors per rule. In the interest of space, we discuss one of the single detector rules that is clearly biologically relevant and was not found by MITRE. This rule is true mostly for infants on a predominantly breast-milk diet and reads, “*TRUE if the aggregated relative abundance of OTUs in the genus *Bifidobacterium* is above 3% between day 0 and day 93.*” This rule is clearly biologically meaningful, as bacteria in the genus *Bifidobacterium* have been shown to express the necessary enzymes to digest human milk oligosaccharides (Lawson et al., 2020).

Vatnen (Vatnen et al., 2016). Our model learned 3 rules with 1 detector per rule. The first rule/detector that reads, “*TRUE if the aggregated relative abundance of OTUs from the genera *Lactobacillus*, *Streptococcus*, *Granulicatella*, *Enterococcus* and *Staphylococcus* is above 0.17% between day 208 and day 466*” is similar to one found by MITRE. The second two rules/detectors were distinct from ones learned by MITRE. These rules/detectors, which are true predominantly for subjects of Russian nationality read “*TRUE if the aggregated relative abundance of OTUs from the genera *Alistipes*, *Bacteroides*, *Butyricimonas*, *Parabacteroides*, *Prevotella* is below 0.9% between day 385 and day 659*” and “*TRUE if the aggregated relative abundance of OTUs from the genera *Alistipes*, *Bacteroides*, *Butyricimonas*, *Parabacteroides*, *Prevotella* is below 1.2% between day 511 and day 886.*” These rules indicate relatively depleted levels of *Bacteroides* and related bacteria in the second year of life, which may reflect different degrees of plant fiber consumption between children in Russia vs. Finland/Estonia after switching to exclusively solid foods (Vatnen et al., 2016).

5. Conclusion

We have presented a fully differentiable model with domain-specific microbiome/temporal inductive biases that learns classifier rules and achieves competitive generalization performance with a fully Bayesian model, MITRE, while running 30-70X faster. We have further demonstrated that our method learns human-interpretable rules that are biologically meaningful, including rules not found by MITRE. A possible reason that our model learns a richer set of rules is that MITRE effectively treats the problem combinatorially, i.e., sampling from different discrete thresholds, phylogenetic subtrees, etc. In contrast, our model relaxes the problem and has a much faster run-time, which appears to allow broader exploration of the space of rules, while still learning sharp/almost discrete structure. Of note, our model automatically expands capacity (number of detectors and rules) as part of the core inference procedure, without requiring exogenous and expensive hyperparameter tuning. Directions for future work include specifying additional priors in the model, for instance to encourage time windows or phylogenetic aggregation based on prior beliefs of biologist colleagues. Another direction would be to incorporate different types of detectors, including rate-of-change of abundance (as in the original MITRE model) or types of detectors tailored to new datatypes such as metabolomic or metatranscriptomic information. Also, although our MAP inference algorithm performed well, it would be interesting to explore Variational Inference to obtain approximations to the model posterior. Finally, we intend to perform experiments on additional microbiome data sets that are already published as well as new data that are currently being collected from large cohorts of hospitalized patients.

6. Funding

V.S.M., G.K.G and V.B. received support from NIH NIGMS 1R01GM130777-01. G.K.G. received support from Brigham and Women Precision Medicine, the Brigham and Women President’s Scholar Award and Harvard Catalyst. We acknowledge the use of the computational resources of the Center for Scientific Computing and Visualization Research at UMass Dartmouth. In particular, the CARNIE cluster which was funded by ONR DURIP Grant No. N00014181255.

7. Competing interests

V.B. receives support from a Sponsored Research Agreement from Vedanta Biosciences, Inc. G.K.G. is a Strategic Advisory Board member and shareholder of Kaleido Biosciences, Inc., and a Scientific Advisory Board member and shareholder of ParetoBio, Inc. G.K.G.’s financial interests were reviewed and are managed by Brigham and Women’s

Scalable learning of interpretable rules for the dynamic microbiome domain

Hospital and Partners Healthcare in accordance with their conflict of interest policies. The remaining authors declare that they have no competing interests. None of the work in this study was supported by commercial interests.

References

- Arrieta, M.-C., Stiemsma, L. T., Dimitriu, P. A., Thorson, L., Russell, S., Yurist-Doutsch, S., Kuzeljevic, B., Gold, M. J., Britton, H. M., Lefebvre, D. L., et al. Early infancy microbial and metabolic alterations affect risk of childhood asthma. *Science translational medicine*, 7(307):307ra152–307ra152, 2015.
- Bogart, E., Creswell, R., and Gerber, G. K. MITRE: inferring features from microbiota time-series data linked to host status. *Genome Biol.*, 20(1):186, 09 2019.
- Bokulich, N. A., Chung, J., Battaglia, T., Henderson, N., Jay, M., Li, H., D Lieber, A., Wu, F., Perez-Perez, G. I., Chen, Y., Schweizer, W., Zheng, X., Contreras, M., Dominguez-Bello, M. G., and Blaser, M. J. Antibiotics, birth mode, and diet shape microbiome maturation during early life. *Sci Transl Med*, 8(343):343ra82, 06 2016.
- Bokulich, N. A., Dillon, M. R., Zhang, Y., Rideout, J. R., Bolyen, E., Li, H., Albert, P. S., and Caporaso, J. G. q2-longitudinal: Longitudinal and paired-sample analyses of microbiome data. *mSystems*, 3(6), 2018. doi: 10.1128/mSystems.00219-18. URL <https://msystems.asm.org/content/3/6/e00219-18>.
- Chen, E. Z. and Li, H. A two-part mixed-effects model for analyzing longitudinal microbiome compositional data. *Bioinformatics*, 32(17):2611–2617, 09 2016.
- David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., Ling, A. V., Devlin, A. S., Varma, Y., Fischbach, M. A., Biddinger, S. B., Dutton, R. J., and Turnbaugh, P. J. Diet rapidly and reproducibly alters the human gut microbiome. *Nature*, 505(7484):559–563, Jan 2014.
- Devkota, S., Wang, Y., Musch, M. W., Leone, V., Fehlner-Peach, H., Nadimpalli, A., Antonopoulos, D. A., Jabri, B., and Chang, E. B. Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in *Il10*^{-/-} mice. *Nature*, 487(7405):104–108, Jul 2012.
- Doshi-Velez, F. and Kim, B. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*, 2017.
- Evans, R. and Grefenstette, E. Learning explanatory rules from noisy data. *Journal of Artificial Intelligence Research*, 61:1–64, 2018.
- Gerber, G. K. The dynamic microbiome. *FEBS Lett.*, 588(22):4131–4139, Nov 2014.
- Haran, J. P., Bhattarai, S. K., Foley, S. E., Dutta, P., Ward, D. V., Bucci, V., and McCormick, B. A. Alzheimer’s disease microbiome is associated with dysregulation of the anti-inflammatory p-glycoprotein pathway. *mBio*, 10(3), 2019. doi: 10.1128/mBio.00632-19. URL <https://mbio.asm.org/content/10/3/e00632-19>.
- Hooker, S., Erhan, D., Kindermans, P.-J., and Kim, B. A benchmark for interpretability methods in deep neural networks. In *Advances in Neural Information Processing Systems*, pp. 9737–9748, 2019.
- Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization, 2014. URL <http://arxiv.org/abs/1412.6980>. cite arxiv:1412.6980Comment: Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015.
- Knights, D., Costello, E. K., and Knight, R. Supervised classification of human microbiota. *FEMS Microbiol. Rev.*, 35(2):343–359, Mar 2011a.
- Knights, D., Parfrey, L. W., Zaneveld, J., Lozupone, C., and Knight, R. Human-associated microbial signatures: examining their predictive value. *Cell Host Microbe*, 10(4):292–296, Oct 2011b.
- Kostic, A. D., Gevers, D., Siljander, H., Vatanen, T., Hyötyläinen, T., Hämäläinen, A.-M., Peet, A., Tillmann, V., Pöhö, P., Mattila, I., et al. The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. *Cell host & microbe*, 17(2):260–273, 2015.
- Lawson, M. A. E., O’Neill, I. J., Kujawska, M., Gowrinadh Javvadi, S., Wijeyesekera, A., Flegg, Z., Chalklen, L., and Hall, L. J. Breast milk-derived human milk oligosaccharides promote *Bifidobacterium* interactions within a single ecosystem. *ISME J*, 14(2):635–648, 02 2020.
- Lipton, Z. C. The mythos of model interpretability. *CoRR*, abs/1606.03490, 2016. URL <http://arxiv.org/abs/1606.03490>.
- Maddison, C. J., Mnih, A., and Teh, Y. W. The concrete distribution: A continuous relaxation of discrete random variables. *CoRR*, abs/1611.00712, 2016. URL <http://arxiv.org/abs/1611.00712>.
- Madsen, A. and Johansen, A. R. Neural arithmetic units. *arXiv preprint arXiv:2001.05016*, 2020.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga,

Scalable learning of interpretable rules for the dynamic microbiome domain

- L., et al. Pytorch: An imperative style, high-performance deep learning library. In *Advances in neural information processing systems*, pp. 8026–8037, 2019.
- Peled, J. U., Gomes, A. L., Devlin, S. M., Littmann, E. R., Taur, Y., Sung, A. D., Weber, D., Hashimoto, D., Slingerland, A. E., Slingerland, J. B., et al. Microbiota as predictor of mortality in allogeneic hematopoietic-cell transplantation. *New England Journal of Medicine*, 382(9): 822–834, 2020.
- Rey, F. E., Gonzalez, M. D., Cheng, J., Wu, M., Ahern, P. P., and Gordon, J. I. Metabolic niche of a prominent sulfate-reducing human gut bacterium. *Proc. Natl. Acad. Sci. U.S.A.*, 110(33):13582–13587, Aug 2013.
- Tang, Z. Z., Chen, G., Alekseyenko, A. V., and Li, H. A general framework for association analysis of microbial communities on a taxonomic tree. *Bioinformatics*, 33(9): 1278–1285, 05 2017.
- Tsilimigras, M. C. and Fodor, A. A. Compositional data analysis of the microbiome: fundamentals, tools, and challenges. *Ann Epidemiol*, 26(5):330–335, 05 2016.
- Vangay, P., Hillmann, B. M., and Knights, D. Microbiome Learning Repo (ML Repo): A public repository of microbiome regression and classification tasks. *Gigascience*, 8(5), 05 2019.
- Vatanen, T., Kostic, A. D., d’Hennezel, E., Siljander, H., Franzosa, E. A., Yassour, M., Kolde, R., Vlamakis, H., Arthur, T. D., Hämäläinen, A.-M., et al. Variation in microbiome lps immunogenicity contributes to autoimmunity in humans. *Cell*, 165(4):842–853, 2016.
- Yang, F., Yang, Z., and Cohen, W. W. Differentiable learning of logical rules for knowledge base reasoning. In *Advances in Neural Information Processing Systems*, pp. 2319–2328, 2017.