Explaining potential epistasis in genomic data using symbolic representations of complex black box models

AAKARSH ANAND¹, PRATEEK ANAND¹, Boyang Fu², Sriam Sankararaman²,³,⁴,⁵

¹BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA ²Department of Computer Science, UCLA
³Department of Human Genetics, David Geffen School of Medicine, UCLA ⁴Department of Computational Medicine, David Geffen School of Medicine, UCLA
⁵Bioinformatics Interdepartmental Graduate Program, UCLA

Abstract

Epistasis, known as the interaction among genetic variants, has long been hypothesized to play a major role in explaining missing heritability. Though recent studies have found many candidate variants demonstrating epistasis signals in the UKBiobank, it remains a controversial question how to interpret the findings. Nonlinear models have shown potential in capturing these signals, but we require additional explanation methods to understand the projected relationships. Here, we utilize symbolic pursuit, a form of symbolic regression that provides a closed-form, interpretable model which generalizes first order explanations. Furthermore, we extend this study by applying Taylor expansions to the model, balancing interpretability with performance while improving its generalizability. We found the method performed reliably and was consistent with other methods across a variety of simulated data. This work contains strong implications for its use on large genomic datasets and its ability to capture nonlinear interactions without prior knowledge of the genetic architecture.

Methodology

Symmetric Pursuit (SP)

Symbolic Pursuit provides the user with importance values that convey how much each feature contributes to the output. In critical domains such as medicine and defense, explanation methods are commonly used as a means of understanding the decisions made by black box models to gain trust in their underlying algorithms. For their ease of implementation, 1st order explanations which consider the marginal contribution of each feature are commonly used.

By only considering marginal contributions, 1st order explanations are insufficient to explain nonlinear models. Some issues:

1. Unable to account for nonlinearity in data
2. Inaccurate, non-generalizable local fidelity
3. Misleading feature importances

Symbolic Pursuit (SP)

Why can’t we modify existing methods to explain nonlinear models?

For example, using the gold standard (SHAP), can we manually construct “interaction” features and obtain their importances? There are two problems with this approach:

1. Which features are interacting with each other, and to what extent?
   a. In the real world, we have no prior knowledge of the data, so this is impossible.
2. How will this affect runtime?
   a. Adding custom features will cause scalability issues, as SHAP values require iterating through the power set of size 2ⁿ feature coalitions to obtain fair rankings.

Symbolic metamodels explain complex models in an interpretable way.

Simulated Data (used to train 200 models for each order across which all results are averaged)

X = (100 instances, 3 features) drawn from random distribution 0 - 1
β drawn from random distribution 0 - 1

1. 1st order: y = Xβ
2. 2nd order: y = Xβ + X²β
3. 3rd order: y = Xβ + X²β + X³β

Metrics of Interest

Model performance measured using:

1. Mean Squared Error (MSE) between SP predictions and targets
2. Kendall Tau Correlation (r) between feature importances and known weights (β)

Linear Comparison

To verify the potential of SP, we first compare its performance to other notable linear explainers on a 1st order dataset. We note a high average correlation between original and predicted importances for all explainers across 200 trials.

Taylor expansions control the method’s overfitting and improve generalizability.

While SP importance attributions effectively model data similar to its training environment, the complex nature of Meijer-G functions often causes the final expression to be overfitted, uninformative, and non-generalizable. This does not help with the real world case where we wish to explain instances similar to those already learned:

\[ Y_{new} = X_{new} \beta + f(X_{new}) \]

We perturb our input data as \( X_{new} = X + \epsilon \), \( \epsilon \sim N(0, \sigma) \), and observe the average performance of Taylor expansions vs. the original SP model. We note a significant decrease in MSE of Taylor Expansions for increasing noise across 1st and 2nd order data, indicating high generalizability.

Conclusions/Future Directions

The above findings show theoretical promise in the underlying strategies of SP. However, the impact on performance when analyzing extremely large datasets remains to be seen. As the high dimensionality of genetic data could cause problems for SP, one idea we hope to explore is using an ensemble of neural networks to summarize groups of features before generating explanations. Regardless, the current performance of SP shows strong potential for effective complex model interpretations which have never been seen before.

References