



The R Package BHAM: Fast and Scalable Bayesian Hierarchical Additive Model for High-dimensional Data

Boyi Guo

University of Alabama at Birmingham

Nengjun Yi

University of Alabama at Birmingham

Abstract

`\pkg{BHAM}` is a freely available R package that implements Bayesian hierarchical additive models for high-dimensional clinical and genomic data. The package includes functions that generalized additive model, and Cox additive model with the spike-and-slab LASSO prior. These functions implements scalable and stable algorithms to estimate parameters. `\pkg{BHAM}` also provides utility functions to construct additive models in high dimensional settings, select optimal models, summarize bi-level variable selection results, and visualize nonlinear effects. The package can facilitate flexible modeling of large-scale molecular data, i.e. detecting susceptible variables and inferring disease diagnostic and prognostic. In this article, we describe the models, algorithms and related features implemented in `\pkg{BHAM}`. The package is freely available via the public GitHub repository <https://github.com/boyiguoi1/BHAM>.

Keywords: additive model, spike-and-slab LASSO, scalable.

1. Introduction

High-dimensional statistics has been an indispensable area of research for its high impact in molecular and clinical data analysis. In recent year, there are continuous efforts to make high-dimensional models more flexible and interpretable, aiming to capture more complex signals. One particular family of such flexible and interpretable models is the additive models where predictors are included in a model in their functional forms. The additive models can help select predictors who have linear or nonlinear effects and provide more accurate prediction when nonlinear effects exist. Guo et al. developed Bayesian hierarchical additive models to analyze continuous, categorical and survival outcomes, and demonstrated improved prediction performance compare to the state-of-the-art additive models. In this article, we

introduce the R package **BHAM** that implements the spike-and-slab LASSO additive models and computationally efficient algorithms to fit these models.

The package **BHAM** provides functions for setting up and fitting various spike-and-slab LASSO additive models, including generalized additive models for various continuous and discrete outcomes and Cox survival models for censored survival outcomes. These functions are extended from previously published Bayesian Hierarchical linear models **BhGLM**, and develop upon commonly used R functions `s` in **mgcv** to construct additive functions. Hence, the proposed models shares similar syntax from well-developed packages and provide powerful features of these standard tools. The syntax can be easily followed and provide user friendliness. In addition, the algorithms implemented in **BHAM** is easily scalable, particularly suitable for fitting high-dimensional models. In the package, we also provide a series utility functions, for example `.plot`. Hence, **BHAM** provides `.plot` and is helpful for `.plot`.

1.1. Literature Review

We enlist current available packages that have similar functionality, i.e. modeling to the best of our knowledge. To note, we don't list packages that are unable of handling high-dimensional data, for example the well known R package **mgcv**, and high-dimensional packages that requires extra steps to construct the design matrix of functional form of predictors (Such implementation can be found with grouped sparse models, for example **SGL**.)

? Summarized the software development of additive models in high-dimensional data analysis before 2013.

Generalized Additive Model

- **COSSO**
- **spikeSlabGAM**
- **sparseGAM**

Additive Cox Proportional Hazard Model

- **COSSO**
- **tfCox**

The **BHAM** package provides a scalable solution for fitting high-dimensional generalized additive model and additive Cox model using spike-and-slab LASSO priors or other regularized priors, including continuous spike-and-slab priors, Student' T priors and double exponential priors. It fits linear, logistic, poisson and Cox regression models. The specification of the additive functions follows a popular syntax implemented in **mgcv**. Ancillary functions are provided, including cross-validation, model summary, and visualization.

In this article, we focus on the packages that can directly construct additive models for high-dimensional data analysis, instead of requiring additional step of constructing design matrix of functional form of the variables before fitting a sparse model.

There are other methods to model survival outcome and provides proportional hazards interpretation, for example `?` provides a link-based survival additive model for mixed censoring in package `GJRM`.

2. Models and algorithms

3. Features

4. Reference

Affiliation:

Boyi Guo
University of Alabama at Birmingham
1665 University Blvd
Birmingham, AL 35294-0002 USA
E-mail: boyigu01@uab.edu
URL: <http://boyigu01.github.io>

Nengjun Yi
University of Alabama at Birmingham
1665 University Blvd
Birmingham, AL 35294-0002 USA
E-mail: nyi@uab.edu