

MEMORANDUM

DATE: 09/23/2024

TO: Faculty and Students

FROM: Professor(s) Robert McCulloch and Yunro Chung
Chair/Co-Chairs of Yaliang Zhang
Defense for the PhD in Statistics
Committee Members Dan Cheng
Shiwei Lan
Shuang Zhou

DEFENSE ANNOUNCEMENT

Candidate: Yaliang Zhang

Defense Date: October 25, 2024

Defense Time: 11:00 AM

Virtual Meeting Link: <https://asu.zoom.us/j/4555134401> or Tempe Campus: WXMLR A113

Title: Bayesian Treed Machine Learning with Applications to Personalized Predictions and Disease Diagnostics

Please share this information with colleagues and other students, especially those studying in similar fields. Faculty and students are encouraged to attend. The defending candidate will give a 40-minute talk, after which the committee members will ask questions. There may be time for questions from those in attendance. However, guests are primarily invited to attend as observers and will be excused when the committee begins its deliberations or if the committee wishes to question the candidate privately.

ABSTRACT
-See next page-

ABSTRACT

Heterogeneity naturally arises in many large-scale data applications, making analysis challenging due to the absence of a common set of features associated with an outcome of interest for all subjects. In this dissertation, we address the classification and regression problems in studies enrolling a heterogeneous population. First, we propose a Bayesian treed model that tailors a univariate feature according to individual profiles. By employing a decision tree framework, this approach partitions the population into distinct subgroups and identifies the most relevant feature for each, taking into account the unique characteristics of the individuals within these subgroups. This allows for more precise and personalized predictions across heterogeneous populations. Second, we present a Bayesian treed machine learning model that advances this univariate approach by recommending feature combinations for each individual. By tailoring feature sets to specific subgroups, this model offers a more comprehensive approach to capturing complex relationships and interactions between variables. It effectively addresses data heterogeneity by selecting optimal feature sets for each group, enhancing both predictive accuracy and interpretability. Extensive simulations show that the proposed method improves prediction accuracy compared to the standard one-size-fits-all machine learning models. Analysis of data from a recent colon cancer biomarker study demonstrates the practical utility of the proposed methodology in making precision diagnostics and personalized predictions.