Best Arm Identification for Variable Selection

Presentation by: Eddie (Yidi) Wu

Brown University

November 28, 2024

Overview

Economic Importance of Variable Selection

- In addition to predictive performance, researchers and decision makers are often interested in which variables are the most influential on outcomes.
- Uncovers economic insights and improves policy effectiveness by targeting factors that are the most effective in addressing specific issues, enabling the design of cost-effective policies.
- Drop unimportant variables when covariates are high-dimensional, circumventing the need for machine learning methods in causal inference.

Multi-Armed Bandits (MAB) and Best Arm Identification (BAI)

- Suppose there are p actions: $\{a_k, k = 1, ...p\}$.
- Each action yields a reward $\gamma_k \sim \mathcal{P}(\theta_k)$.
- At each time step, the agent can choose one action to play.
- MAB: maximize cumulative reward over time (or minimize regret)
- BAI: identify which action yields the highest mean reward in the most efficient way, either fixed confidence or fixed budget.
- Binary bandit is a canonical example, but also the most relevant to this research.

Bandit Algorithms

- Upper confidence bound (e.g. Lai & Robbins 1985): computes UCB based on past actions and rewards, plays the arm with the highest UCB.
- Thompson sampling (e.g. Thompson 1933, Agrawal & Goyal 2012): assumes prior distributions on θ_k, updates the distributions of θ_k at each time step, and chooses which action to play by sampling from the distributions of θ_k.
- Thompson sampling is optimal in terms of the bounds on regret, but suboptimal in terms of the rate of convergence of posterior probabilities.

Thompson Sampling for Variable Selection

- Suppose our data consists of X_i (p $\times$ 1) and y_i where X_i is high-dimensional and

$$y_i = f(X_{i1}, X_{i2}, ...X_{iq}) + \epsilon_i$$

  where $q << p$ and $\epsilon \sim N(0, \sigma^2)$.
- A combinatorial bandit problem: let each variable X_{ik} be an action with reward $\gamma_k \sim Ber(\theta_k)$, and each θ_k has a Beta distribution prior.
- At every time step, we sample a set of X_{ik} to estimate our model, and observe reward realization γ_k^t, and update the distributions of θ_k.

Overview

Rockova and Liu (2021, JASA)

- Fuses **combinatorial bandit** with **Bayesian model selection with spike-and-slab priors**:
  - Combinatorial bandit: choose m_t arms at every time step and receive global reward depending on the set of arms chosen.
  - Spike-and-slab priors: for each coordinate $1 \leq k \leq p$, one assumes a latent indicator γ_k for whether x_k is active and assigns a prior on γ_k:

$$P(\gamma_k = 1|\mu) = \mu$$

    updates the distribution of γ_k using a linear model and from the posterior distribution of γ_k, and obtain the optimal set of variables.

Rockova and Liu (2021, JASA)

- It is standard practice to report the optimal set of variables as:

$$\hat{S} = argmax_S \left\{ \prod_{k \in S} \pi_k \prod_{k \notin S} (1 - \pi_k) \right\} = \{k : \pi_k \geq 0.5\}$$

where π_k is $P(\gamma_k = 1 | Data, \mu)$.

Rockova and Liu (2021, JASA)

- Uses binary combinatorial bandits and Bayesian additive regression tree (BART).
- Algorithm input: M for the number of BART MCMC iterations, a_{k0}, b_{k0} for the Beta prior of θ_k.
- For each t = 1,...T, repeat:
  - Choose: Sample $\theta_{kt} \sim Beta(a_{kt}, b_{kt})$ and play S_t as $\{k : \theta_{kt} \geq 0.5\}$.
  - Reward: Define γ_k^t to be 1 if the M^{th} sample from the BART posterior splits on the variable indexed as k.
  - Update: Update a_{kt} and b_{kt}.
- Obtain the optimal set of variables by $\{k : \pi_k \geq 0.5\}$

Russo (2016, COLT)

- Proposes a class of top-two algorithms for allocating measurement efforts adaptively and proves that the convergence of posterior probability occurs at an exponential rate in a fixed confidence setting.
- The original Thompson sampling: play the action with the highest sampled θ_k.
- Top-two Thompson sampling:
  - With probability β, play the action with the highest sampled θ_k.
  - With probability $1 - \beta$, sample again until we obtain a $\theta_j \neq \theta_k$ and play the action j.
- Tradeoff between exploitation and exploration.

Overview

1. Background
2. Motivation
3. My Progress So Far
4. Experimental Results
5. To-do

Potential Improvements

- Posterior probability estimates are noisy, with many probabilities close to 0.5, on real world datasets, i.e. false negative rate might not be small.
- Computationally intensive: at each Thompson sampling step, we need to run a BART which requires MCMC posterior sampling.

My Contributions So Far

- Developed Thompson LASSO and Thompson Random Forest algorithms:
  - Experimentally shown to improve over both LASSO and random forest, and for now comparable to Rockova&Liu(2021).
- Incorporated two-top Thompson sampling and experimentally shown to achieve more accurate and efficient selection of variables.
- Incorporated bootstrap aggregation (bagging) at each time step which led to greater accuracy, lower variance and higher efficiency in variable selection.

Thompson LASSO

- Algorithm input: a_{k0}, b_{k0} for the Beta prior of θ_k.
- For each t = 1,...T, repeat:
  1. Sample $\theta_{kt} \sim Beta(a_{kt}, b_{kt})$ and define S_t as $\{k : \theta_{kt} \geq 0.5\}$.
  2. Sample $\beta_t \sim Ber(0.5)$. If $\beta_t \leq 0.5$, play S_t. Else, repeat step 1 until we obtain $S_t' \neq S_t$, and play S_t'.
  3. For each l = 1,...5: bootstrap a sample $(X_i, y_i)_l$, compute optimal LASSO penalty by k-fold cross validation, and fit LASSO using the optimal penalty.
  4. Define $\gamma_k^t = 1$ if variable k has $|\beta_k| \geq \epsilon$ in all 5 bootstrap iterations, 0 otherwise.
  5. Update $a_{k,t+1} = a_{k,t} + \gamma_k^t$ and $b_{k,t+1} = b_{k,t} + (1 - \gamma_k^t)$.

Thompson Random Forest

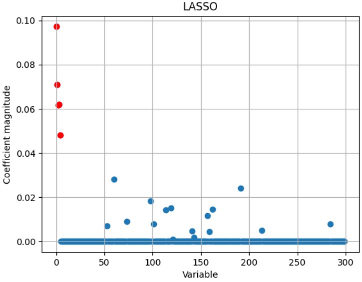
- Algorithm input: a_{k0}, b_{k0} for the Beta prior of θ_k, M for the number of trees.
- For each t = 1,...T, repeat:
  1. Sample $\theta_{kt} \sim Beta(a_{kt}, b_{kt})$ and define S_t as $\{k : \theta_{kt} \geq 0.5\}$.
  2. Sample $\beta_t \sim Ber(0.5)$. If $\beta_t \leq 0.5$, play S_t. Else, repeat step 1 until we obtain $S'_t \neq S_t$, and play S'_t.
  3. Randomly split $(X_i, y_i)_{i=1}^n$ into 80% training and 20% test set. Fit random forest on the training set and compute the average permutation importance of each variable over 10 iterations.
  4. Define $\gamma_k^t = 1$ if variable k has permutation importance greater than or equal to ϵ, 0 otherwise.
  5. Update $a_{k,t+1} = a_{k,t} + \gamma_k^t$ and $b_{k,t+1} = b_{k,t} + (1 - \gamma_k^t)$.

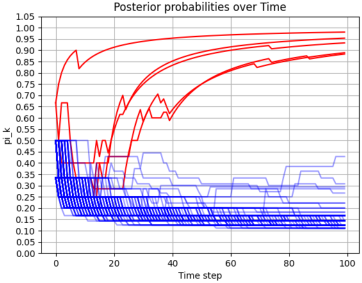
Overview

Data Generating Process: Linear Case

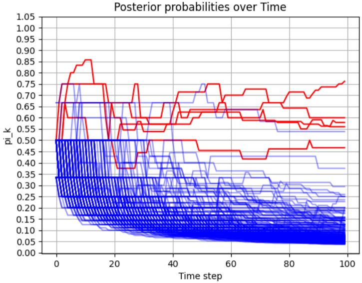
- $X_i \sim Unif[0,1]^p$
- $y_i = 1 + \frac{1}{2}\sum_{j=1}^{q}(-1)^j X_{i,j} + \epsilon_i$
  where $\epsilon \sim N(0, 0.5^2)$.
- n = 300, p = 300, q = 5.

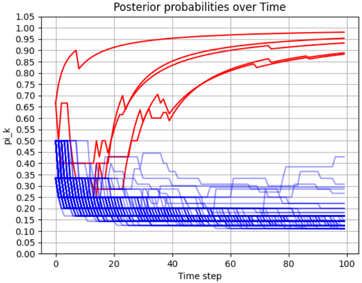
Comparing to the Original LASSO





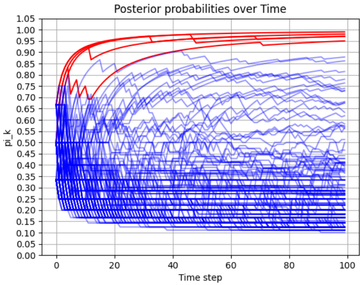
Comparing to Random Variable Subset Selection

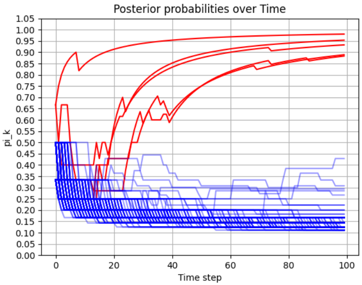




Left: at each time step, choose a random subset of size *m*. Right: at each time step, sample from the posterior distributions of θ_k.

Ablation Study: Remove Bagging



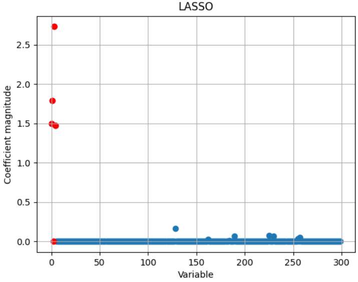


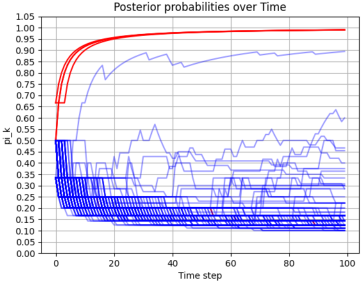
Left: no bootstrap aggregation. Right: with bootstrap aggregation.

Data Generating Process: Non-linear Case

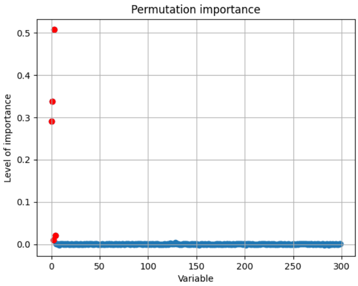
- Based on the Friedman 1991 benchmark dataset used in Rockva&Liu2021.
- $X_i \sim Unif[0,1]^p$
- $y_i = 10\sin(\pi X_{i,1} X_{i,2}) + 20(X_{i,3} - 0.5)^2 + 10X_{i,4} + 5X_{i,5} + \epsilon_i$ where $\epsilon \sim N(0, 0.5^2)$.
- n = 300, p = 300

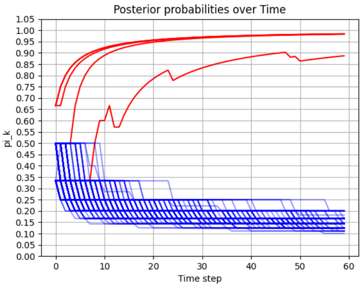
LASSO-Based Bandit



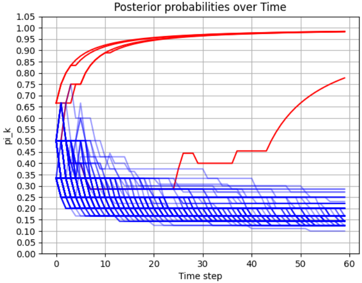


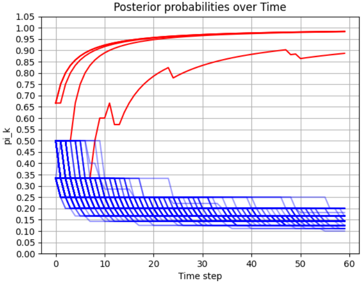
Random Forest-Based Bandit





Ablation Study: Remove Top-Two Sampling





Left: use Thompson sampling in choosing S_t. Right: use top-two Thompson sampling in choosing S_t.

Overview

Theoretical Aspects

- Try to prove the rate optimality of the class of top-two sampling algorithms for best m-arms identification, which is left as a conjecture by Russo (extension from Russo 2016).
- Try to show the regret bound on my algorithm, which should be a function of β, the probability of not playing the first Thompson sampling choice (extension from Wang&Chen 2018 and Rockova&Liu 2021).
- Look at the optimal tradeoff between efficiency and regret to decide on an optimal β in variable selection (extension from Qin&Russo 2024).
- Try to prove the consistency of my algorithms in posterior probability converging on the true θ_k.

Applied Aspects

- Explore the other DGPs in Rockova&Liu2021 and compare to their results.
- Compare to the causal forest method of Athey, Wager and Tibshirani.
- Alternative designs of global and local rewards based on random forest or gradient boosting algorithms.
- Apply to real world datasets and compare to existing results in literature.

Thank you.