SPRINT Multi-Objective Model Racing

Tiantian Zhang, Michael Georgiopoulos, Georgios C. Anagnostopoulos
Machine Learning Lab
University of Central Florida
Overview

• **Racing Algorithms**
• **Multi-objective Model Selection**
• **SPRINT-Race**
  – Motivation
  – Sequential Probability Ratio Test
  – (Non-) Dominance Inference via Dual-SPRT
• **Experimental Results and Discussions**
• **Conclusion and Future Direction**
Racing Algorithm

- Model Selection
  - Consider an ensemble of models and stick with the best
  - Elimination-type Multi-Armed Bandit

1. A set of initial models
2. Sample validation instances
3. Evaluate all models
   - Abandon bad performers
4. Stop?
   - No
   - Yes
5. Return the best performers
Racing Algorithm

- Trade-off
  - Benefit: Identify best model(s)
  - Price to be paid: Computational cost
- RAs trade off *model optimality* vs. *computational effort* by automatically allocating computational resources
## Racing Algorithm

<table>
<thead>
<tr>
<th>Name</th>
<th>Year</th>
<th>Statistical Test</th>
<th>Criterion Of Goodness</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hoeffding</td>
<td>1994</td>
<td>Hoeffding’s inequality</td>
<td>Prediction Accuracy</td>
<td>Classification &amp; Func. Approximation</td>
</tr>
<tr>
<td>BRACE</td>
<td>1994</td>
<td>Bayesian Statistics</td>
<td>Cross-validation Error</td>
<td>Classification</td>
</tr>
<tr>
<td>F-Race</td>
<td>2002</td>
<td>Friedman Test</td>
<td>Function Optimization</td>
<td>Optimal parameter of Max-Min-Ant-System for TSP</td>
</tr>
<tr>
<td>Bernstein Racing</td>
<td>2008</td>
<td>Bernstein’s inequality</td>
<td>Function Optimization</td>
<td>Policy selection in CMA-ES</td>
</tr>
<tr>
<td>S-Race</td>
<td>2013</td>
<td>Sign Test</td>
<td>Prediction Accuracy</td>
<td>Classification</td>
</tr>
</tbody>
</table>
Multi-Objective Optimization

- Pareto Dominance (minimization case)
Multi-Objective Model Selection

• Multi-Criteria Recommendation

Story: 2  
Actors: 5  

Story: 3  
Actors: 4  

Story: 4  
Actors: 3
Multi-Objective Model Selection

- Multi-task Learning
Multi-Objective Model Selection

• Motivation
  – Model selection often involves **more than one criterion** of goodness
  – Examples
    • Generalization performance vs. model complexity
    • Single model addressing multiple tasks
  – How can racing be adapted to a MOMS setting?

• Approaches
  – **Scalarization** of criteria into a single criterion
    • Pros: Existing racing algorithms may be applicable
    • Cons: Good models may be left undiscovered
      (due to non-convexity of the relevant Pareto front)
  – **Vector Optimization**
    • Pros: Selecting models based on Pareto optimality
    • Cons: No racing algorithm available
Multi-Objective Racing Algorithm

- **S-Race**
  - Fixed-Budget
  - Offline
  - Maximize selection accuracy

- **SPRINT-Race**
  - Fixed-Confidence
  - Online
  - Minimize computational cost
SPRINT-Race

• Racing via Sequential Probability Ratio with INdifference zone Test

  ✓ Try to minimize the computational effort needed to achieve a predefined confidence about the quality of the returned models.
  ✓ A near-optimal non-parametric ternary-decision sequential analogue of the sign test is adopted to identify pair-wise dominance/non-dominance relationship.
  ✓ Strictly confine the error probability of returning dominated models and, simultaneously, abandoning non-dominated models at a predefined level.
  ✓ Able to stop automatically.
  ✓ The concept of indifference zone is introduced.
Non-sequential VS Sequential Test

- Non-sequential test
  - Sample size $n$ is fixed
  - Given $\alpha$ and rejection region, $\beta$ is a function of sample size.
  - The Uniformly Most Efficient test – given $n$ and $\alpha$, find the test procedure that minimize $\beta$

- Sequential Test
  - Sample size $n$ is a variable
  - Either we accept $H_0$, or we accept $H_1$, or we continue sampling
  - Properties: OC function (probability of accepting $H_0$), ASN (average sample number)
  - The Uniformly Most Efficient test – given $\alpha$ and $\beta$, find the test that minimize the expected sample size
-- Sequential Probability Ratio Test

- Locally Most Efficient Test procedure
- SPRT with Bernoulli Distribution

  - At n-th step, assume \( \{x_1, x_2, ..., x_n\} \) are i.i.d. samples collected from a Bernoulli distribution with \( P\{x_i = 1\} = p \), and \( d = \#\{x_i = 1\} \)

  - Given two simple hypothesis
    - \( H_0: p = p_0 \)
    - \( H_1: p = p_1 \)
    - \[ \tau_n = \frac{p_1^d (1 - p_1)^{n-d}}{p_0^d (1 - p_0)^{n-d}} \]

  Let \( A = \frac{1-\beta}{\alpha}, B = \frac{\beta}{1-\alpha}. \)

  - if \( \tau_n \leq B \), \( H_0 \) is accepted;
  - if \( \tau_n \geq A \), \( H_1 \) is accepted;
  - otherwise, continue sampling.
(Non-) Dominance Inference

• Given model $C_i$ and $C_j$.
  
  $n_{ij}$ - the number of instances that the performance vector of $C_i$ dominates the performance vector of $C_j$

  $n_{ji}$ - vise versa

  $p$ – the probability that $C_i$ dominates $C_j$

  $n_{ij} \sim Binomial(n_{ij} + n_{ji}, p)$
Dual-SPRT (Sobel-Wald’s test)

- The three-hypothesis problem is composed of two component two-hypothesis tests

\[ \begin{align*}
    \text{SPRT1: } & H_0: p \leq \frac{1}{2} - \delta \\
    & H_1: p \geq \frac{1}{2} \\
    \text{SPRT2: } & H_0: p \leq \frac{1}{2} \\
    & H_1: p \geq \frac{1}{2} + \delta
\end{align*} \]
Dual-SPRT (Sobel-Wald’s test)

- Assume that common $\alpha$ and $\beta$ values are shared between $SPRT1$ and $SPRT2$
## Dual-SPRT Accuracy Analysis

<table>
<thead>
<tr>
<th>Interval</th>
<th>Wrong Decisions</th>
<th>$\gamma(p)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p \leq \frac{1}{2} - \delta$</td>
<td>Accept $H_1$ or $H_2$</td>
<td>$\gamma(p) \leq \alpha$</td>
</tr>
<tr>
<td>$\frac{1}{2} - \delta &lt; p &lt; \frac{1}{2}$</td>
<td>Accept $H_2$</td>
<td>$\gamma(p) \leq \alpha$</td>
</tr>
<tr>
<td>$p = \frac{1}{2}$</td>
<td>Accept $H_0$ or $H_2$</td>
<td>$\gamma(p) \approx \alpha + \beta$</td>
</tr>
<tr>
<td>$\frac{1}{2} &lt; p &lt; \frac{1}{2} + \delta$</td>
<td>Accept $H_0$</td>
<td>$\gamma(p) &lt; \beta$</td>
</tr>
<tr>
<td>$p \geq \frac{1}{2} + \delta$</td>
<td>Accept $H_0$ or $H_1$</td>
<td>$\gamma(p) \leq \beta$</td>
</tr>
</tbody>
</table>

$$\gamma^* = \max\{\gamma(p)\} \sim \alpha + \beta$$
Overall Error Control

• Bonferroni Approach

\[ \Delta \leq \sum_{i=1}^{M \choose 2} \gamma_i^* = \sum_{i=1}^{M \choose 2} (\alpha_i + \beta_i) \]

Let \( \alpha_i = \beta_i = \epsilon \) for \( i = 1, \ldots, (M \choose 2) \), \( \Delta_{\text{max}} \) denote the maximum error probability allowed for SPRINT-Race

\[ \epsilon = \frac{\Delta_{\text{max}}}{M(M-1)} \]
SPRINT-Race

- Initialize $Pool \leftarrow \{C_1, C_2, \ldots, C_m\} (m \geq 2), t = 1$

- Randomly sample a problem instance

  - for each pair $C_i, C_j \in Pool \text{ s.t. } i < j$ do
    - if the corresponding dual-SPRT continues then
      - Evaluate $C_i, C_j$, update $n_{ij}, n_{ji}$
      - if $H_0$ is accepted then
        - $Pool \leftarrow Pool \backslash \{C_i\}$
        - stop all dual-SPRTs involving $C_i$
      - else if $H_2$ is accepted then
        - $Pool \leftarrow Pool \backslash \{C_j\}$
        - stop all dual-SPRTs involving $C_j$
      - else if $H_1$ is accepted then
        - stop all dual-SPRTs involving $C_i, C_j$
    - end if
  - end for

- until All dual-SPRTs are terminated

- return the Pareto front models found
Experiments

- Artificially Constructed MOMS Problem

• Given M models, construct \( \binom{M}{2} \) Bernoulli distributions with known \( p \) values.

• Three performance metrics

\[
R(\text{retention}) = \frac{|P_R \cap P_{PF}|}{|P_{PF}|}
\]

\[
E(\text{excess}) = \frac{|P_R \setminus P_{PF}|}{|P_R|} = 1 - \frac{|P_R \cap P_{PF}|}{|P_R|}
\]

\( P_{PF} \) - Pareto front models

\( P_R \) - models returned by SPRINT-Race

\( T \) – sample complexity
Impact of No. of Objectives

R

E

T

Impact of No. of Objectives

R

E

T

Impact of No. of Objectives

R

E

T
Impact of $\delta$ and $\Delta_{max}$ Values

- Decreasing $\Delta_{max}$ will definitely increase sample complexity with slightly varied R and E values.
Experiments

- ACO Selection for TSPs

- Parameter tuning of ACOs is time consuming

- A pool of 125 candidate models were initialized with diverse configuration in terms on different combinations of three parameters
  - $\alpha_{ACO}$ - the influence of pheromone trials
  - $\beta_{ACO}$ - the influence of heuristic information
  - $\rho_{ACO}$ - the pheromone trial evaporation rate

- Two objectives
  - Minimize the TSP tour length
  - Minimize the actual computation time to find this tour

- ACOTSPJava + DIMACS TSP instance generator
Experiments

- ACO Selection for TSPs
Conclusions

- **SPRINT-Race is a multi-objective racing algorithm**
  - Solving multi-objective model selection
  - Dual-SPRT is adopted for dominance and non-dominance inference
  - The total probability of falsely retaining any dominated model and removing any non-dominated model is strictly controlled
  - Be able to stop automatically with fixed confidence

- **SPRINT-Race is able to return almost exactly the true Pareto front models at a reduced cost**
  - Artificially-constructed MOMS problems
  - Select Pareto optimal ACP parameters for solving TSPs

? **Optimal model configuration**
References


Thank you!

Any Questions?
Backup Slides
Multi-Objective Optimization

- Performance vector of model $C$ – $L$ objectives
  \[ f(C) \triangleq [f_1(C), f_2(C), \ldots, f_L(C)] \]

- (Minimization)
  $C$ dominates $C'$ if and only if
  \[ f_i(C) \leq f_i(C') \ \forall \ i \in \{1, 2, \ldots, L\} \text{ and } \exists \ j \in \{1, 2, \ldots, L\} (j \neq i) \ f_j(C) < f_j(C') \]
Multi-Objective Racing Algorithm

• S-Race (GECCO’13)
  – Pareto-optimality
  – Test Pareto dominance non-parametrically via sign test
  – Consider family-wise testing procedures

• Limitations of S-Race
  x No inference on non-dominance relationship
  x No control over the overall probability of making any Type II errors
  x Sign test is not an optimum test procedure in sequential setting
**Dual-SPRT (Sobel-Wald’s test)**

- The three-hypothesis problem is composed of two component two-hypothesis tests

\[
\begin{align*}
\text{SPRT1: } & H_0: p \leq \frac{1}{2} - \delta & H_1: p \geq \frac{1}{2} \\
\text{SPRT2: } & H_0: p \leq \frac{1}{2} & H_1: p \geq \frac{1}{2} + \delta
\end{align*}
\]

<table>
<thead>
<tr>
<th>if SPRT1 accepts</th>
<th>if SPRT2 accepts</th>
<th>then dual SPRT accepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0: p \leq \frac{1}{2} - \delta$</td>
<td>$H_0: p \leq \frac{1}{2}$</td>
<td>$H_1: p \leq \frac{1}{2} - \delta$ (remove $C_i$)</td>
</tr>
<tr>
<td>$H_1: p \geq \frac{1}{2}$</td>
<td>$H_0: p \leq \frac{1}{2}$</td>
<td>$H_2: p = \frac{1}{2}$ (keep both)</td>
</tr>
<tr>
<td>$H_1: p \geq \frac{1}{2}$</td>
<td>$H_1: p \geq \frac{1}{2} + \epsilon$</td>
<td>$H_3: p \geq \frac{1}{2} + \epsilon$ (remove $C_j$)</td>
</tr>
</tbody>
</table>