Sparrow: Faster Boosting with Smaller Memory

Julaiti Alafate and Yoav Freund
Department of Computer Science and Engineering, University of California, San Diego

Boosting 3-80x faster than XGBoost with limited memory

- 623M examples, 100 GB on disk
- 50M examples, 39 GB on disk

Main Ideas

Focus on the hard part
- Sample the examples close to the decision boundary

Stop early
- Find a good enough rule rather than the best one

Asynchronous
- Reduce worker stalling time using asynchronous parallelism

Sequential Analysis

When can we tell a weak rule is "good"?

Boosting iterations search for a rule with \( \text{err} < 0.5 \)

- Best weak rule
- Good weak rule

Early stopping bound
- Standard approach: scan all data and output the best
- Sequential analysis

System Design and Scalability

Performance on Large-Memory Machines

- i3.large: 61 GB mem, $0.624/$0.25 per hour
- i3.xlarge: 244 GB mem, $2.496/$0.75 per hour

Weighting + Sampling

The "effective" number of examples

\[ n_{\text{eff}} = \frac{C}{\sum_{i=1}^{C} w_i} \]

When the weights of the examples are skewed to a small fraction, the training is prone to overfitting to the small group of heavy weight examples.

Asynchronous Parallelism

Bulk synchronous parallelism:

- All workers must wait the slowest worker

Tell Me Something New:

- Workers communicate only if something new is found