A UNIFIED FRAMEWORK FOR KNOWLEDGE INTENSIVE GRADIENT BOOSTING

Leveraging Human Experts for Noisy Sparse Domains

Harsha Kokel, Phillip Odom, Shuo Yang, Sriraam Natarajan
AAAI 2020
Motivation
Qualitative influences

1. Wellman AI 1990
2. Robertson, Wright, and Dykstra 1988, Altendorf et al. UAI 2005, Yang and Natarajan ECML-PKDD 2013

\[ x_1 < x_2 \Rightarrow P(Y \leq k|p_{a_{y1}}) \geq P(Y \leq k|p_{a_{y2}}) \]
Qualitative Constraint

-order-restricted constraints

\[ x_0 < x_1 \Rightarrow \psi(x_0) \leq \psi(x_1) \]

-conditional-probability constraints\(^2\)

\[ x_0 < x_1 \Rightarrow P(Y \leq y_1|x_0) \geq P(Y \leq y_1|x_1) \]

\[ x_1 < x_2 \Rightarrow P(Y \leq y_1|x_1) \geq P(Y \leq y_1|x_2) \]

\[ x_1 < x_2 \Rightarrow P(Y \leq k|pa^{x_1}_y) \geq P(Y \leq k|pa^{x_2}_y) \]

\(^2\)Altendorf et al. UAI 2005, Yang and Natarajan ECML-PKDD 2013

*caeteris paribus*
Knowledge-intensive Gradient Boosting

\[ X^Q Y \]

\[
x_1 < x_2 \Rightarrow \mathbb{E}_{\psi}[x_1 | \ldots ] \leq \mathbb{E}_{\psi}[x_2 | \ldots ]
\]

\[
\mathbb{E}_{\psi}[n_L] \leq \mathbb{E}_{\psi}[n_R] + \varepsilon
\]

\[
\zeta_n = \begin{cases} 
\mathbb{E}_{\psi}[n_L] - \mathbb{E}_{\psi}[n_R] - \varepsilon < 0 
\end{cases}
\]

\[
\arg\min \sum_{i=1}^{N} (y_i - \psi_t(x_i))^2 + \frac{\lambda}{2} \sum_{n \in N(x_c)} \max (\zeta_n \cdot |\zeta_n|, 0)
\]

loss function w.r.t data

\[
\sum_{n \in N(x_c)} \max (\zeta_n \cdot |\zeta_n|, 0)
\]

loss function w.r.t advice
KiGB

- Leaf update equation

\[
\psi_t^\ell(x) = \frac{1}{|\ell|} \sum_{i=1}^{N} \tilde{y}_i \cdot \mathbb{I}(x_i \in \ell) + \left\{ \begin{array}{l}
\frac{\lambda}{2} \sum_{n \in N(x_c)} \mathbb{I}(\zeta_n > 0) \zeta_n \cdot \left( \frac{\mathbb{I}(\ell \in n_R)}{|n_R|} - \frac{\mathbb{I}(\ell \in n_L)}{|n_L|} \right) \end{array} \right.
\]

penalty for advice violation
Monotonic Trees Ensemble

- Usually for classification tasks
- Focus on global monotonicity
  - *prune*
  - *preprocessed data by reweighting*
  - *voting mechanism*
  - *restrict split criteria*
- *Monoensemble* converts trees to rules and recalculate leaf values

---

3Cano et al. Neurocomputing 2019, 4Dembczynski et al. 2009
5Ke et al. NIPS 2017 (LightGBM), Chen et al. KDD 2016 (XGBoost), 6Bartley et al. AAAI 2019
Sparse data

\( a^{Q+y} \)
Sparse data

KiGB with $\lambda = 0$
standard boosting

KiGB with $\lambda = 3$

KiGB with $\lambda = 8$
Overfitting by monotonic function
## Experiments

### Standard baselines

| Dataset     | SKiGB | SGB | Dataset     | SKiGB | SGB |
|-------------|-------|-----|-------------|-------|-----|
| Adult       | 0.855 | 0.853 | Cleveland   | 0.737 | 0.677 |
| Australian  | 0.855 | 0.83 | Ljubljana   | 0.696 | 0.621 |
| Car         | 0.984 | 0.982 |
| Abalone     | 5.377 | 5.491 | CPU         | 0.185 | 0.204 |
| Auto mpg    | 9.793 | 13.623 | Crime       | 2.211 | 2.296 |
| Auto price  | 8.866 | 8.945 | Red wine    | 0.381 | 0.419 |
| Boston      | 24.065 | 21.493 | Whitewine   | 0.426 | 0.439 |
| California  | 47.159 | 47.468 | Windsor     | 3.9 | 4.626 |

**KiGB:** ours with S/L  
**SGB:** Scikit-learn gradient boosting  
**LGBM:** LightGBM  
**LMC:** LightGBM with monotonic constraints  
**MONO:** Monoensemble

### Monotonic baselines

#### Classification Task (Accuracy)

| Dataset     | SKiGB | MONO | LKiGB | LMC |
|-------------|-------|------|-------|-----|
| Adult       | 0.855 | 0.857 | 0.865 | 0.863 |
| Australian  | 0.855 | 0.884 | 0.878 | 0.867 |
| Car         | 0.984 | 0.765 | 0.972 | 0.959 |
| Cleveland   | 0.737 | 0.74 | 0.757 | 0.73 |
| Ljubljana   | 0.696 | 0.611 | 0.721 | 0.718 |

#### Regression Task (Mean-Squared Error)

| Dataset     | LKiGB | LMC |
|-------------|-------|-----|
| Abalone     | 4.786 | 4.797 |
| Auto mpg    | 8.047 | 8.33 |
| Auto price  | 14.953 | 15.614 |
| Boston      | 15.496 | 16.292 |
| California  | 48.517 | 50.94 |
Experiments

Real datasets

| Dataset            | LKiGB | LGBM  | LMC  |
|--------------------|-------|-------|------|
| Logistics (mse)    | 1.851 | 1.898 | 1.889|
| Dataset            | SKiGB | SGB   | MONO |
| HELOC (accuracy)   | 0.717 | 0.7   | 0.688|

Logistics : Turvo
HELOC : FICO xML challenge
THANKS