Individual Fair Gradient Boosting

2021/04/13 @読み会

楊明哲

論文情報

個別公平性 + GBDTに注目した研究

- 著者情報
  - Alexander Vargo, Fan Zhang, Mikhail Yurochkin, Yuekai Sun
  - ミシガン大学, 上海科技大学, MIT-IBM Watson AI Lab
- 出典: ICLR2021
- なんで選んだか？
  - 個別公平性+決定木を考えているのは意外に少ない. →これが初めてらしい

イントロダクション

公平性を考慮していかないといけない

- 機械学習(ML)が意思決定の分野で広く使われ始めている
- 特定のグループ(人)に対して不公平な評価をしていけない
  - Amazonの履歴書審査システムで差別が行われていたことが明らかになった.

イントロダクション

今回は個別公平性を対象としていく

- ML界隈では大きく二種類の公平性を考える
  - 個別公平性: 似ている個人は同じ評価を受けるべき
  - 集団公平性: 集団ごとに評価の差別がないようにするべき
- 集団公平性がよく取り上げられている
  - 個人の類似度をきちんと定義することが困難だったから

イントロダクション

勾配ブースティング決定木(GBDT)を対象とする

- 表データにGBDTを用いるのが主流になってきている.
- 従来のFair-awarness MLではnon-smoothなモデルやノンパラメトリックMLではあまり良い効果が得られていなかった.

イントロダクション

貢献

- 個別公平性を対象にしたGBDTによる手法を提案した.
- モデルのそれぞれの公平性を証明することが可能.
- 個別公平性だけでなく集団公平性を向上させつつ，精度を維持する手法になっていることを実験的に示した.

準備

使う記号を定義する

- 入力: $\mathcal{X} \in \mathbb{R}^d$, 出力: $\mathcal{Y} = \{0,1\}$
- 保護する属性: $\mathcal{Z} = \mathcal{X} \times \{0,1\}$ いわゆるセンシティブ属性
- サンプルごと公平指標: d_x これはサンプルが近いほど似ている
- 目標:

  サンプルごとに公平なモデル$f : \mathcal{X} \to \{0,1\}$を獲得すること

既存手法はどうだったの？

Non-smoothなモデルではうまくいかなかった.

- 敵対学習によって達成する方法は存在している
  - 学習が入力に対して滑らかであることが前提になっている
- 滑らかでないモデル（決定木とか）に対しても敵対学習を行えるようにしたい！
  - **制限付き敵対的コスト関数**を定義したよ！

準備

サンプルごとに公平なモデルを学習したい

- Transport cost function: 個別のサンプルが近いほど小さい

$$c\left(\left(x_1, y_1\right), \left(x_2, y_2\right)\right) \triangleq d_x^2\left(x_1, x_2\right) + \infty \cdot \mathbf{1}_{\{y_1 \neq y_2\}}$$

- Zの確率分布上の最適輸送距離W: 分布の近さを考えている

$$W\left(P_1, P_2\right) \triangleq \inf_{\Pi \in C\left(P_1, P_2\right)} \int_{\mathcal{Z} \times \mathcal{Z}} c\left(z_1, z_2\right) d\Pi\left(z_1, z_2\right)$$

準備

敵対的リスク関数を定義する.

$$L_r(f) \triangleq \sup_{P:W(P,P_*)\leq\epsilon} \mathbb{E}_P[\ell(f(X), Y)]$$

- $P_\star$はデータ生成分布，$\epsilon > 0$の微小な許容パラメータ

- 標本空間上で1) データ生成分布が近い

  2) MLモデルの損失を大きくなるもの を探したい

準備

ロバストで公平な分布を得たい！

- 類似したサンプルに対してモデルの性能差を見つけられる
  - 性能差を探索することで分布に対して頑健な公平性だと捉えられる.
- 現状だとまだsmoothなモデルの勾配しか得られない.

提案手法

制限を加えてnon-smoothのために工夫する.

- データセットを拡張する:

$$\mathscr{D}_0 \triangleq \left\{ \left(x_i, y_i\right), \left(x_i, 1 - y_i\right) \right\}_{i=1}^{n}$$

- 最適輸送関数に制限を加える: 違いは上のデータセットかどうか

$$W_{\mathscr{D}}\left(P_1, P_2\right) \triangleq \inf_{\Pi \in C_0\left(P_1, P_2\right)} \int_{\mathscr{Z} \times \mathscr{Z}} c\left(z_1, z_2\right) d\Pi\left(z_1, z_2\right)$$

提案手法

やっと非平滑に適用できるよ

- データセットを加えることで上界はD_0に指示された分布に制限される
- これによって有限次元線形計画法によって解けるようになる.
- 損失は$\ell\left(f\left(x_i\right), y_i\right)$ and $\ell\left(f\left(x_i\right), 1-y_i\right)$にしか依存してないから非平滑なモデルでも適用できる.

提案手法

勾配ブースティング木でも使えるようにする

勾配ブースティングでは$\frac{\partial L}{\partial \hat{y}}$を求める必要がある.

ダンスキンの定理を用いると勾配は,

$$\frac{\partial L}{\partial \hat{y}_i} = \frac{\partial}{\partial f(x_i)} \left[\sup_{P:W_{\mathscr{D}}(P,P_n)\leq\epsilon} \mathbb{E}_P \left[\ell\left(f(x_i), y_i\right) \right] \right] = \sum_{y\in\mathscr{Y}} \frac{\partial}{\partial f(x_i)} \left[\ell\left(f(x_i), y\right) \right] P^*(x_i, y)$$

提案手法

関数勾配を考える

- 先述の勾配では，モデルを微分する必要がないから非平滑なモデルでも関数勾配を評価することができる！
- あとは$\mathbf{P}_\star$を求めれば良い.
- 線形計画法によって$\mathbf{P}_\star$を求める方法を提案する.

提案手法

$\mathbf{P}_\star$を線形計画法で求める

- D_0による任意の分布Pに対して，$P_{i,k} = P(\{(x_i, k\}), k \in \{0,1\}$
  とすると$W_D(P, P_n) \leq \epsilon$は次のような行列$\Pi$で表せる.

$$1.\Pi \in \Gamma \text{ with } \Gamma = \left\{ \Pi \mid \Pi \in \mathbb{R}_+^{n \times n}, \langle C, \Pi \rangle \leq \epsilon, \Pi^T \cdot \mathbf{1}_n = \frac{1}{n}\mathbf{1}_n \right\}$$

$$2.\Pi \cdot y^1 = \left(P_{1,1}, \ldots, P_{n,1}\right), \text{ and } \Pi \cdot y^0 = \left(P_{1,0}, \ldots, P_{n,0}\right)$$

提案手法

さらに定義していくよ

- 行列$R_{i,j} = l(f(x_i), y_j)$ → ラベルjであるサンプルjがサンプルiになったときの損失

- 求めたい行列$\Pi_\star$は次のようになる

$$\Pi^\star \in \underset{\Pi \in \Gamma}{\arg\max} \langle R, \Pi \rangle$$

提案手法- まとめ

これで非平滑な関数にも適用できる！

- 結局最後の$\Pi_\star$を求めることができれば良い.

- 求めるにあたって，関数Fには何も仮定を置いていないので，非平滑な関数にも適用できる.

Algorithm 1 Fair gradient boosting

Input: Labeled training data $\{(x_i, y_i)\}_{i=1}^n$; class of weak learners $\mathcal{H}$; initial predictor f_0; search radius ϵ; number of steps T; sequence of step sizes $\alpha^{(t)}$; fair metric d_x on $\mathcal{X}$
Define the matrix C by $C_{i,j} \leftarrow d_x^2(x_i, x_j)$.
for $t = 0, 1, \ldots, T-1$ **do**
    Define the matrix R_t by $(R_t)_{ij} = \ell(f_t(x_i), y_j)$
    Find $\Pi_t^* \in \arg\max_{\Pi \in \Gamma} \langle R_t, \Pi \rangle$; and set $P_{t+1}(x_i, k) \leftarrow (\Pi_t^* \cdot y^k)_i$
    Fit a base learner $h_t \in \mathcal{H}$ to the set of pseudo-residuals $\{\frac{\partial L}{\partial f_t(x_i)}\}_{i=1}^n$ (see (2.6)).
    Let $f_{t+1} = f_t + \alpha_t h_t$.
end for
return f_T

実験

- 3つのデータセット(German Credit, Adult, COMPASS)で検証
- 提案手法で用いる決定木アルゴリズムは，XGBoostとする.
- 損失関数はロジスティック損失を用いる.

実験

公平性指標について(個別のサンプルに関して)

- Yurochikinらのを利用する: $d_x^2 = (x_1 - x_2, Q(x_1 - x_2))$
  - Qはセンシティブ部分空間と直行する射影行列
- 保護されるセンシティブ属性以外の情報が同じなら同等に扱われるべきであるという考えから作られた.

実験

対抗手法について

- 決定木手法に関しては，対抗がないためバニラを用いる.
- データの前処理を用いる手法と比較する
  - 保護属性をなくし，部分空間に投影する(Yurochkin et al., 2020)
  - 個人に異なる重みを適用してバランスをとる(Kamiran & Calders, 2011)

実験

評価について(既存手法に対し優劣がないように加工をする)

- 保護されるている属性と相関がある属性(e.g. 夫か?妻か?)をずらすことで反事実の人物を作成.
  - →ほぼ同じ人物だから同じ評価をされるべき
- 保護属性ごとのTPR,TNRの差(GAPMax)→モデルの公平性指標
- 保護属性ごとのRMSEの差(GAPRMSE) →モデルの予測性能

実験結果

① German Credit

- 年齢をセンシティブ属性に設定 → 米国では年齢をつけて与信判断するのは違憲

Method	BAcc	Status cons	Age gaps	
			GAP_{Max}	GAP_{RMS}
BuDRO	.715	**.974**	**.185**	.151
Baseline	**.723**	.920	.310	.241
Project	.698	.960	.188	**.144**
Baseline NN	.687	.826	.234	.179

- 射影による前処理は提案ほど個人の公平性を向上させなかった.

実験結果

② Adult

Method	BAcc	Individual fairness S-cons	Individual fairness GR-cons	Gender gaps GAP_{Max}	Gender gaps GAP_{RMS}	Race gaps GAP_{Max}	Race gaps GAP_{RMS}
BuDRO	.815	**.944**	.957	.146	.114	.083	.072
Baseline	**.844**	.942	.913	.200	.166	.098	.082
Project	.787	.881	**1**	**.079**	.069	.064	.050
Reweigh	.784	.853	.949	.131	.093	**.056**	**.043**
Baseline NN	.829	.848	.865	.216	.179	.105	.089
SenSR	.789	.934	.984	.087	**.068**	.067	.055
Adv. Deb.	.815	.807	.841	.110	.082	.078	.070

- 提案手法はGBDTの性能の良さを引き継ぎつつ，公平なモデルになっていた！

実験結果

③COMPASS

Table 3: COMPAS: average results over 10 splits into 80% training and 20% test data.

Method	Acc	Individual fairness		Gender gaps		Race gaps	
		G-cons	R-cons	GAP_{Max}	GAP_{RMS}	GAP_{Max}	GAP_{RMS}
BuDRO	0.652	**1.000**	**1.000**	**0.099**	**0.124**	0.125	0.145
Baseline	0.677	0.944	0.981	0.180	0.223	0.215	0.258
Project	0.671	0.874	**1.000**	0.150	0.190	0.185	0.230
Reweigh	0.666	0.788	0.813	0.207	0.245	**0.069**	**0.092**
Baseline NN	**0.682**	0.841	0.908	0.246	0.282	0.228	0.258
SenSR	0.652	0.977	0.988	0.130	0.167	0.159	0.179
Adv. Deb.	0.670	0.854	0.818	0.219	0.246	0.108	0.130

- NNモデルの方が精度は基本的に良かった.
- しかし公平性については，提案の方が良かった.

まとめ

個別公平性＋決定木の手法を提案したぞ

- 個別公平性を達成する課題をMLモデルの性能差を探索できないこと → 探索空間を有限区間に制限することで克服した.

- 今回設定した制限付き敵対損失関数は他のnon-smooth手法(ランダムフォレスト)などにも適用できるかもしれない.

- 実感として，NNモデルよりも決定木ベースのほうが精度＋公平性を達成できそう.

感想

きちんと理論を追える数学力が欲しい

- 作者が示している理論をちゃんと理解できなくてくやしい.
- 個別公平性を考えている論文を読めて良かった.