Data mining　report4　　205713B 朝比奈太郎 2022/07/28

Level1

(1) 資料 1 の「Types of Bias」まで取り組め。Reporting Bias、Selection Bias、Group Attribution Bias、Implicit Bias について解説せよ。各々数行程度で良い。

・Reporting Bias

「Reporting Bias」はデータセットに含まれる事象、特性、結果の頻度が現実の頻度を正確に反映していない場合に起こる。このようなバイアスが生じるのは、人は普通でない状況や特に記憶に残る状況を記録することに重点を置き、普通のことは「当然である」と思い込む傾向があるためである。

・Selection Bias

「Selection Bias」は、データセットの例が実際の分布を反映していない方法で選ばれている場合に発生する。「Non-response bias (or participation bias)」,「Coverage Bias」,「Sampling Bias」のような様々な形で発生する。

・Group Attribution Bias

「Group Attribution Bias」は、個人に当てはまることをその人が属する集団全体に一般化する傾向のことである。このバイアスの主な現れとして、「In-group Bias」,「Out-group Bias」の 2 つが挙げられる。

・Implicit Bias

「Implicit Bias」は、自分のメンタルモデルや個人的な経験に基づいて、必ずしも一般的には当てはまらない仮定がなされることである。「Implicit Bias」の一般的な形態は「confirmation Bias」であり、モデル構築者は無意識のうちに既存の信念や仮説を肯定するような方法でデータを処理する。これは「experimenter's Bias」と呼ばれ、モデル構築者が自分の仮説と一致する結果が出るまでモデルを訓練し続けることがある。

(2) 教材には書かれていないバイアスの具体例を述べよ。この際、(a) バイアスを検討したテーマそのものについて述べるとともに、(b) バイアスについて述べよ。

(a) バイアスを検討したテーマ: 課題 1 の beer-data-set において「Reporting Bias」の検討
Beer-data-set は、"Brew No", "OG", "ABV", "pH", "IBU の 5 要素からなる、サンプル数 150 のデータセットである。クラスは"IPA", "Light Lager", "Premium Lager"の 3 つで構成され

ている。
(b)「Reporting Bias」とは、回答者が意図的に解答を変えてしまうバイアスである。Beer-data-set において、ラベルの”IPA”, “Light Lager”, “Premium Lager”は 50 個ずつ、合計 150 個あるが、機械学習の結果、約 50 個ずつクラスタリングされるのではなく、100 個、0 個、50 個といった偏りが大きい場合に起こりうるバイアスであると考える。

Level2
資料 1 の「Programming Exercise」まで取り組め。そのうえで、自分自身の興味のあるテーマについて fairness の視点から想定される問題点について論じよ。

テーマ: UK police are using AI to inform custodial decisions – but it could be discriminating against the poor

・テーマに関する説明
イギリスでは、危害評価リスクツール(HART)を開発している。この AI は、容疑者が今後２年間でさらなる犯罪を犯す可能性が低い、中程度、高いかどうかを判断するように設計されている。判断基準として、人の年齢、性別、問題のある履歴等の 34 の異なるカテゴリのデータを用いている。HART は機械学習によって、容疑者を拘留するかどうかの判断を警官が下す手助けをするために開発されている。しかし、実際のところアルゴリズムはブラックボックスであり、決定を下す方法を完全に説明をすることができない。これらの観点から、容疑者を裁く場合に AI を用いることは正当性があるのかを説いている。

・考察
私は、HART や COMPAS が多要素のカテゴリデータを用いて危害評価を行い、警官の判断を促すために使用されることには賛成である。しかし、HART においては、34 のカテゴリデータで評価を行うため、1 つ 1 つのカテゴリが本当に評価基準として適切かどうかを議論する必要があると考える。2017 年からは特定の地域に住む人々に対する偏見を強化しないように内容を変更したとあるが、これは以前、郵便番号というカテゴリつまり居住地によって、容疑者の犯罪リスクが高いかどうかを判断要素の 1 つにしていたということである。確かに地域によって自治が異なるため、犯罪率は大きく異なることは考えられる。しかし、郵便番号カテゴリを用いた場合、犯罪の多い地域は貧困層が多いと考えられるため、必然的に貧困層を差別することに繋がってしまう。Fairness の観点から見ると、居住地によって結果が異なる可能性があることは良いことではない。「犯罪の多い地域」、「容疑者」は必要条件の関係であり、「犯罪の多い地域」であるため、「容疑者」が有罪であるといった十分条件は存在しないと考える。
機械学習のメリットとして、バイアスを最小限にし、最適解を導くことのできる点が挙げら

れるが、「郵便番号」といった人間のバイアスが入ったカテゴリを用いることによって機械学習のメリットを享受できないのは惜しい。
従って、私は機械学習を行う上で、再度カテゴリを見直し人間のバイアスを全て取り除いたAIを導入するべきであると考える。