**REAL-TIME DATA SELECTION AND ORDERING FOR COGNITIVE BIAS MITIGATION**

Sara Mourad, Ahmed Tewfik

---

**Biased Information Processing Model**

- **Cognitive bias**: Deviation from rational judgement yielding potentially incorrect or damaging inference/decision (Tversky and Kahneman 1972).
- **Sources of bias**: Limited human mental processing capacity, cognitive shortcuts (heuristics), social context, emotions, etc.
- **Examples**: - **Confirmation bias**: searching for or interpreting information in a way that supports a preconception. Observed in many settings, e.g., trials.
  - **Anchoring bias**: Over reliance on one piece of information (usually the first). Observed in many settings, e.g., negotiations.
  - **Framing**: drawing conclusion from information depending on how it is presented. Observed in many settings, e.g., marketing, media, politics.

- **Experiment**: - Humans receive training on how to classify objects from two classes.
  - Human classification performance depends on order in which items are presented to humans.

**Problem formulation**

- **Binary hypothesis detection problem**: $H_0$: $Y_n = W_n$ 
  $H_1$: $Y_n = m + W_n$

  where $W_n \sim N(0, \delta^2)$ are i.i.d. and $m$ is the difference in the means under the two hypothesis.

- **Proposed Model for human decision-making under cognitive biases**:

  $$L_k = L_{k-1} + p_k l_k \quad (1)$$

  where $l_k = \log \frac{f(y_k|m)}{f(y_k|\theta)} = \frac{2m^2 - m^{2+k}}{2k}$ and $p_k$ the adjustment weight that the subject gives to the new observation.

**Goal**: Select $N'$ out of $N$ total observations to show to Bob so that his decision performance is within a desired distance from Alice’s.

**Find in polynomial time a subset $K \subset [N], |K| \leq N'$, such that**

$$|T - L_K| = \min, \text{ where } T = \sum_{i=1}^{N'} f[y_i|H_0] + E[L_{N'}|H_0]$$

**Bob (Actual decision maker)**

- **Unbiased agent uses model (1)**
- **Has access to N’ out of N observations**

**Alice (Ideal reference)**

- **Ideal reference**
- **Unbiased agent**
- **Has access to N observations**

---

**Proposed approximate solution: an extension of the approximate subset sum algorithm**

- **The extension of the approximate subset sum algorithm is used to find the closest $L_{N'}$ to the target $T$**

  **Algorithm 1 Modified approximate subset algorithm $(S_1, T, \epsilon)$**

  1. $n \leftarrow |S_1|$
  2. $R_0 \leftarrow \{0\}$
  3. $G_0 \leftarrow \{0\}$
  4. for $i \leftarrow 1$ to $n$ do
  5.   $R_i \leftarrow \text{MergeLists}(R_{i-1}, R_{i-1} + w_j | i_i)$
  6.   $G_i \leftarrow \text{MergeLists}(G_{i-1}, G_{i-1} + 1)$
  7.   $(R_i, G_i) \leftarrow \text{Trim}(R_i, G_i, N', \epsilon/2n)$
  8. end for
  9. return the closest element in $R_n$ to $T$ with size in $G_n$ less than or equal to $N'$

**S\leftarrow\{1, 2, ..., n\}, S1\leftarrow[S,S,...,S] N'\text{'s times} > \text{To account for permutations of same subset}

**G Tracks sizes of each element in R**

**Trimming function modified to keep only smallest size subset**

**Returned subset sum has restricted size \( \leq N' \)**

**Performance guarantee – Running time**

**Let $L_{N'}$ be the closest subset sum to $T$, such that**

$$T - \frac{\delta}{2} \leq L_{N'} \leq T + \frac{\delta}{2}$$

$$Q(\lambda_A - E[L_N|H_1] + \frac{\sigma_N}{2} \sqrt{\frac{\lambda_N}{N}}) \leq P_d(B) \& P_f(B)$$

$$Q(\lambda_A - E[L_N|H_0] - \frac{\sigma_N}{2} \sqrt{\frac{\lambda_N}{N}}) \leq Q(\lambda_A)$$

**where $P(d(B) \& P_f(B))$ is the probability of detection (false alarm) of $\lambda_A$, the threshold used by Alice when using Neyman-Pearson test.**

**Proposed algorithm returns $y^*$**

$$y^* = \left(1 - \frac{1}{\epsilon}\right) L_{N'} \leq y^* \leq (1 + \epsilon) L_{N'}$$

**Running time of algorithm: $O(NN' \log N' N \& O(1/\epsilon)$**

---

**Results for confirmation bias towards $H_0$**

- Near optimal performance of algorithm in region of interest
- For low probability of detection, accuracy of algorithm is more critical

**Results for anchoring bias (Emphasis on first few observations)**

- Near optimal performance in region of interest
- For low probability of detection and high probability of false alarm, accuracy of algorithm is more critical

---
