PRIVACY PROTECTION IN LEARNING FAIR REPRESENTATIONS

Yulu Jin, Lifeng Lai

University of California, Davis

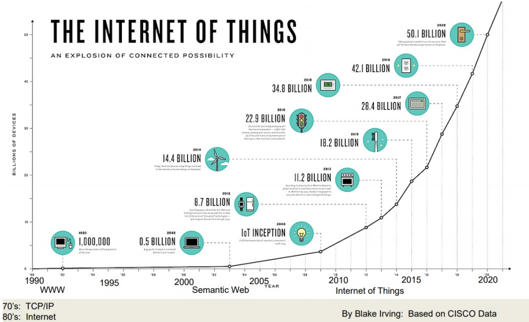
May 2022

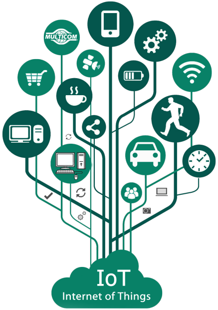
Overview

Outline

Internet of Things

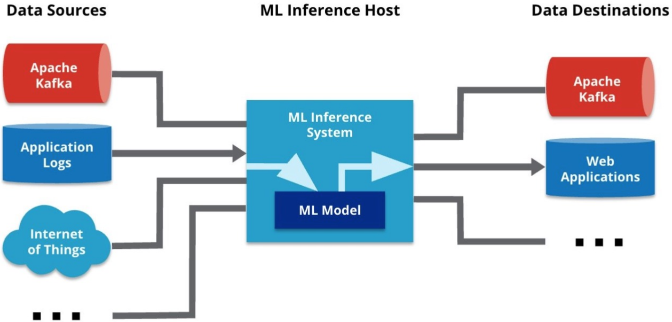
- The Internet of Things (IoT) devices.





Inference as a service

- The Internet of Things (IoT) devices.
- Inference as a service (IAS).



- However, IAS brings privacy issues.

Fairness issue

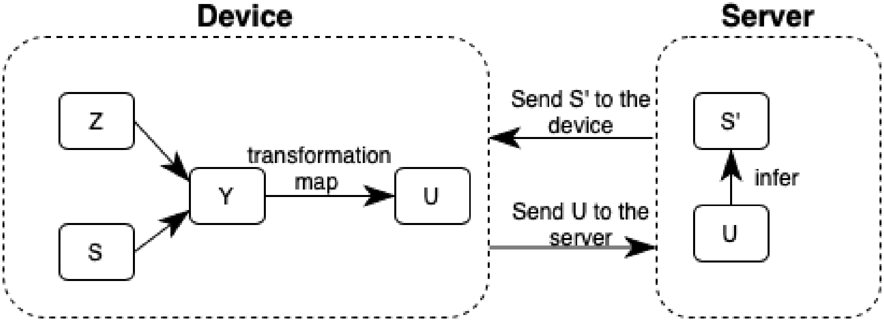
- Main purpose: ensure that the inference decisions do not reflect discriminatory behavior toward certain groups or populations.
- Example: Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), a software that measures the risk of a person to recommit another crime.
- Potential sources of unfairness: those arising from biases in the data and those arising from the algorithms.
- A variety of methods have been proposed that satisfy some of the fairness definitions or other new definitions depending on the application

Our goal

- Our goal is to address the fairness and privacy issues simultaneously in the IAS design.
- Instead of sending data directly to the server, we preprocess the data through a transformation map.
- Analyze the trade-off among data utility, fairness representation and privacy protection.
- Formulate an optimization problem to find the optimal transformation map.

Outline

Problem Statement and Notations



The optimization problem is

$$\max_{P_{U|Y}} \ \mathcal{F}[P_{U|Y}] \triangleq I(S;U) - \beta \mathbb{E}_{Y,U}\left[f\left(\frac{p(u|y)}{p(u)} \right) \right] - \alpha I(Z;U),$$

$$\text{s.t.} \ \ p(u|y) \geq \epsilon, \forall y, u, \sum_{u} p(u|y) = 1, \forall y \in \mathcal{Y}.$$

Problem Statement and Notations

$$\max_{P_{U|Y}} \; \mathcal{F}[P_{U|Y}] \triangleq I(S;U) - \beta \mathbb{E}_{Y,U}\left[f\left(\frac{p(u|y)}{p(u)} \right) \right] - \alpha I(Z;U), \tag{1}$$

$$\text{s.t.} \; p(u|y) \geq \epsilon, \forall y, u, \sum_u p(u|y) = 1, \forall y \in \mathcal{Y}, \tag{2}$$

where $d(y,u) = f(\frac{p(y)}{p(y|u)})$ and f is a continuous function defined on $(0, +\infty)$.

- The proposed framework in (1) is general with respect to the privacy metric. For $f(\cdot) = \log(\cdot)$, we have

$$\begin{aligned} \mathbb{E}_{Y,U}[d(y,u)] &= \sum_{y,u} p(y)p(u|y) \log \left(\frac{p(u)}{p(u|y)} \right) \\ &= -\sum_y p(y) D_{KL}[p(u|y) \parallel p(u)] = -I[U;Y]. \end{aligned}$$

As the result, we will use mutual information between U and Y to measure information leakage.

Alternating optimization

Lemma 1

$$I(S;U) = I(S;Y) - \sum_{u,y} p(y)p(u|y)D_{KL}[p(s|y) \parallel p(s|u)].$$

Then the objective function defined in (1) can be written as

$$\begin{aligned}\mathcal{F}[P_{U|Y}, P_U, P_{Z|U}, P_{S|U}] &= I(S;Y) + \beta\mathbb{E}_{Y,U}[d(y,u)] \\ &\quad - \sum_{u,y} p(y)p(u|y)D_{KL}[p(s|y) \parallel p(s|u)] - \alpha I(Z;U).\end{aligned}$$

Alternating optimization

Lemma 1

$$I(S;U) = I(S;Y) - \sum_{u,y} p(y)p(u|y)D_{KL}[p(s|y) \parallel p(s|u)].$$

Then the objective function defined in (1) can be written as

$$\begin{aligned}\mathcal{F}[P_{U|Y}, P_U, P_{Z|U}, P_{S|U}] &= I(S;Y) + \beta\mathbb{E}_{Y,U}[d(y,u)] \\ &\quad - \sum_{u,y} p(y)p(u|y)D_{KL}[p(s|y) \parallel p(s|u)] - \alpha I(Z;U).\end{aligned}$$

For consistency, we require the following equations to be satisfied simultaneously

$$p(u) = \sum_{y} p(u|y)p(y), \forall u, \tag{3}$$

$$p(z|u) = \frac{\sum_y p(u|y)p(z,y)}{p(u)}, \tag{4}$$

$$p(s|u) = \frac{\sum_y p(u|y)p(s,y)}{p(u)}. \tag{5}$$

Concavity

$$
\begin{aligned}
\max_{P_{S|U}} \max_{P_{Z|U}} \max_{P_U} \max_{P_{U|Y}} \mathcal{F}[P_{U|Y}, P_U, P_{Z|U}, P_{S|U}]. \\
\text{s.t.} \quad & p(u|y) \geq \epsilon, \forall y, u, \ \sum_u p(u|y) = 1, \forall y, \\
& p(u) > 0, \forall u, \ \sum_u p(u) = 1, (3), \\
& p(z|u) \geq 0, \forall u, z, \ \sum_z p(z|u) = 1, \forall u, (4), \\
& p(s|u) \geq 0, \forall u, s, \ \sum_s p(s|u) = 1, \forall u, (5).
\end{aligned}
$$

- **Lemma 2** Suppose that $f(\cdot)$ is a strictly convex function. Then for given $P_U, P_{Z|U}, P_{S|U}$, $\mathcal{F}[P_{U|Y}, P_U, P_{Z|U}, P_{S|U}]$ is concave in each $P_{U|y_i}, \forall y_i \in \mathcal{Y}$. Similarly, for given $P_{U|Y}, P_{Z|U}, P_{S|U}$, $\mathcal{F}[P_{U|Y}, P_U, P_{Z|U}, P_{S|U}]$ is concave in P_U. For given $P_{U|Y}, P_U, P_{S|U}$, $\mathcal{F}[P_{U|Y}, P_U, P_{Z|U}, P_{S|U}]$ is concave in $P_{Z|U}$. For given $P_{U|Y}, P_U, P_{Z|U}$, $\mathcal{F}[P_{U|Y}, P_U, P_{Z|U}, P_{S|U}]$ is concave in $P_{S|U}$.

Concavity

$$\max_{P_{S|U}} \max_{P_{Z|U}} \max_{P_U} \max_{P_{U|Y}} \mathcal{F}[P_{U|Y}, P_U, P_{Z|U}, P_{S|U}].$$

$$\begin{aligned} \text{s.t.} \quad & p(u|y) \geq \epsilon, \forall y, u, \ \sum_u p(u|y) = 1, \forall y, \\ & p(u) > 0, \forall u, \ \sum_u p(u) = 1, (3), \\ & p(z|u) \geq 0, \forall u, z, \ \sum_z p(z|u) = 1, \forall u, (4), \\ & p(s|u) \geq 0, \forall u, s, \ \sum_s p(s|u) = 1, \forall u, (5). \end{aligned}$$

- The alternating optimization problem can be solved iteratively.

Algorithm

- In the first step, given $P_{S|U}^{(j-1)}$ and $P_{Z|U}^{(j-1)}$, we obtain $P_{U|Y}^{(j)}$ and $P_U^{(j)}$ by solving

$$\begin{aligned}\max_{P_{U|Y}} \max_{P_U} \quad & \mathcal{F}[P_{U|Y}, P_U | P_{S|U}^{(j-1)}, P_{Z|U}^{(j-1)}], \\ \text{s.t.} \quad & p(u|y) \geq \epsilon, \forall y, u, \ \sum_u p(u|y) = 1, \forall y, p(u) > 0, \forall u, \ \sum_u p(u) = 1, \\ & \delta(u) = p(u) - \sum_y p(u|y)p(y) = 0, \forall u.\end{aligned}$$

  - Apply ADMM to solve the problem.
  - The optimization problem can be solved by the iterative procedure,

$$P_{U|y_i}^{t+1} = \arg\max_{P_{U|y_i}} \mathcal{L}[P_{U|y_i}, P_{U|Y^{(i-)}}^{t+1}, P_{U|Y^{(i+)}}^{t}, P_U^t; \Lambda^t], \tag{6}$$

$$P_U^{t+1} = \arg\max_{P_U} \mathcal{L}[P_{U|Y}^{t+1}, P_U; \Lambda^t], \tag{7}$$

$$\Lambda^{t+1} = \Lambda^t - \rho(P_U^{t+1} - (P_{U|Y}^{t+1})^T P_Y). \tag{8}$$

Algorithm

- In the second step, we obtain $P^{(j)}_{Z|U}$ by the consistency equation

$$p^{(j)}(z|u) = \frac{\sum_y p^{(j)}(u|y)p(z,y)}{p^{(j)}(u)}.$$

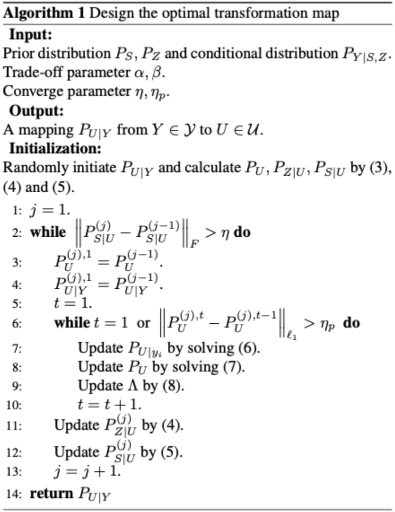
- In the third step, obtain $P^{(j)}_{S|U}$ by solving

$$\begin{aligned} \max_{P_{S|U}} \quad & \mathcal{F}[P_{S|U}|P^{(j)}_{U|Y}, P^{(j)}_{U}, P^{(j)}_{Z|U}], \\ \text{s.t.} \quad & p(s|u) \geq 0, \forall u, s, \sum_s p(s|u) = 1, \forall u, (5), \end{aligned}$$

which has a simple closed form solution

$$p^{(j)}(s|u) = \frac{\sum_y p^{(j)}(u|y)p(s,y)}{p^{(j)}(u)}.$$

Algorithm



Algorithm 1 Design the optimal transformation map

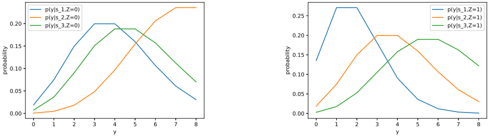
```
Input:
Prior distribution $P_S, P_Z$ and conditional distribution $P_{Y|S,Z}$.
Trade-off parameter $\alpha, \beta$.
Converge parameter $\eta, \eta_p$.
Output:
A mapping $P_{U|Y}$ from $Y \in \mathcal{Y}$ to $U \in \mathcal{U}$.
Initialization:
Randomly initiate $P_{U|Y}$ and calculate $P_U, P_{Z|U}, P_{S|U}$ by (3), (4) and (5).
$j = 1$.
while $\left\| P_{S|U}^{(j)} - P_{S|U}^{(j-1)} \right\|_F > \eta$ do
    $P_U^{(j),1} = P_U^{(j-1)}$.
    $P_{U|Y}^{(j),1} = P_{U|Y}^{(j-1)}$.
    $t = 1$.
    while $t = 1$ or $\left\| P_U^{(j),t} - P_U^{(j),t-1} \right\|_{\ell_1} > \eta_p$ do
        Update $P_{U|y_i}$ by solving (6).
        Update $P_U$ by solving (7).
        Update $\Lambda$ by (8).
        $t = t + 1$.
    Update $P_{Z|U}^{(j)}$ by (4).
    Update $P_{S|U}^{(j)}$ by (5).
    $j = j + 1$.
return $P_{U|Y}$
```

Outline

1. Introduction
2. The Proposed Method
3. Numerical Examples
4. Conclusion

Numerical Examples

- Suppose that $Z \in \{0, 1\}$.
- Set the prior distributions $\boldsymbol{p}_z = \{\frac{1}{4}, \frac{3}{4}\}$.
- Let $|\mathcal{Y}| = 9, |\mathcal{U}| = 11$.
- The conditional distributions $P_{Y|S}(y|s, Z = 0)$ and $P_{Y|S}(y|s, Z = 1)$ are shown below

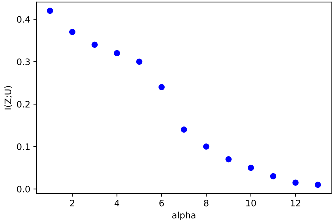


(c) $p(y|s, Z = 0)$ (d) $p(y|s, Z = 1)$

Figure: Conditional distributions

Numerical Examples: relationship between α and degree of fairness

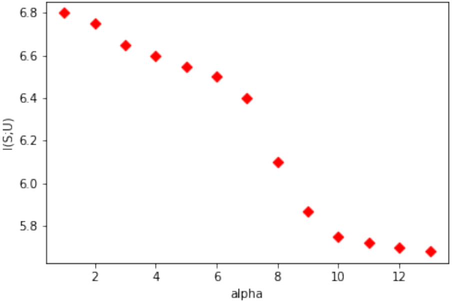
- Set the privacy trade-off parameter $\beta = 7$.
- Randomly initialize $P_{U|Y}$.
- Run the algorithm until it terminates for different αs.
- Repeat 300 times for each α.



- As α increases, the transformed variable provides less information about the sensitive attribute.

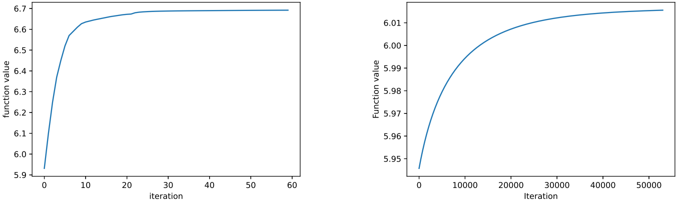
Numerical Examples: relationship between α and information accuracy

- The information accuracy $I(S;U)$ is decreasing as α increases.
- The deduction of $I(S;U)$ is not very large.



Numerical Examples: convergence speed of the proposed algorithm

- The objective function value monotonically increases and converges as the iterative process progresses.
- Algorithm 1 converges within 30 iterations.
- GA is hard to converge. The optimal function value found by GA is always smaller.



(a) Function value of Algorithm 1 (b) Function value of GA

Figure: Function value v.s. iteration

Outline

1. Introduction
2. The Proposed Method
3. Numerical Examples
4. Conclusion

Conclusion

- We have explored the utility, fairness and privacy trade-off in IAS scenarios under sensitive environments.
- We have formulated an optimization problem to find the desirable transformation map.
- We have designed an iterative method to solve this complicated optimization problem.
- The method has better performance than GA.
- Numerical results are provided.