MULTI-DATA-SOURCE AI TRAINING MECHANISM AND THE REVELATION PRINCIPLE

Abstract.

1. Introduction

2. The Proof of the Revelation Principle

Definition 2.1. Let $\vec{v} = (v_1, ..., v_n)$ be an n-dimentional vector. We will denote the $(n-1)$-dimentional vector in which the i-th coordinate is removed by $\vec{v_{-i}} = (v_1, ..., v_{i-1}, v_{i+1}, ..., v_n)$. Thus we have three equivalent notations:$\vec{v} = (v_1, ..., v_n) = (v_i, \vec{v_{-i}})$.

Definition 2.2. An original AGI mechanism is a choice function f and a vector of utility functions $u_1^f, ..., u_n^f$

$$f : V_1 \times ... \times V_n \to A \tag{2.1}$$

$$u_i^f : V_1 \times ... \times V_n \to \mathbb{R} \tag{2.2}$$

Definition 2.3. An original AGI mechanism $(f, u_1^f, ..., u_n^f)$ is called incentive compatible if $\forall i, \exists v_i \in V_i, \forall v_i' \in V_i$,

$$u_i^f(v_i, v_{-i}) \geq u_i^f(v_i', v_{-i}) \tag{2.3}$$

Definition 2.4. A trained AGI with(independent private values and) strict incomplete information for a set of n trainers is given by the following ingredients:

(i)For every trainer i , a set of *actions* X_i.

(ii)For every trainer i , a set of *data* T_i . A value $t_i \in T_i$ is the private information that i has.

(iii)For every trainer i , a $utility\ function$ $u_i : T_i \times X_1 \times ... \times X_n \to \mathbb{R}$, where $u_i(t_i, x_1, ..., x_n)$ is the utility achieved by player i , if his type is t_i, and the profile of actions taken by all trainers is $x_1, ..., x_n$.

Definition 2.5. (i)A strategy of a trainer i is a function $s_i : T_i \to X_i$.

(ii)A strategy s_i is a (weakly) dominant strategy if for every t_i we have that the action $s_i(t_i)$ is a dominant strategy in the full information training defined by t_i . Formally:For all t_i, allx_{-i} and all x_i' we have that

$$u_i(t_i, s_i(t_i), x_{-i}) \geq u_i(t_i, x_i', x_{-i}) \tag{2.4}$$

A profile $s_1, ..., s_n$ is called a dominant strategy equilibrium if each s_i is a dominant strategy.

Definition 2.6. (i)A synthetic training for n trainers is given by

(a)trainers' data spaces $T_1, ..., T_n$,

(b)trainers' action spaces $X_1, ..., X_n$,

(c)an alternative set A ,

(d)an outcome function $a : X_1 \times ... \times X_n \to A$ and ,

1991 *Mathematics Subject Classification.* .

The AGI with strict incomplete information induced by the synthetic training is given by using the data spaces T_i , the action spaces X_i , and the utilities $u_i(t_i, x_1, ..., x_n)$.

(ii)The synthetic training implements a choice function $f : T_1 \times ... \times T_n \to A$ in dominant strategies if for some dominant strategy equilibrium $s_1, ..., s_n$ of the induced game , where $s_i : T_i \to X_i$, we have that for all $t_1, ..., t_n$, $f(t_1, ..., t_n) = a(s_1(t_1), ..., s_n(t_n))$.

Proposition 2.1 (Revelation Principle)**.** *If there exists an arbitrary synthetic training AGI that implements f in dominant strategies , then there exists an incentive compatible original AGI that implements f.*

Proof. The new AGI will simply simulate the equilibrium strategies of the players . That is , let $s_1, ..., s_n$ be a dominant strategy equilibrium of the synthetic training AGI , we define a new direct revelation AGI :

$$f(t_1, ..., t_n) := a(s_1(t_1), ..., s_n(t_n)) \tag{2.5}$$

$$u_i^f(t_1, ..., t_n) := u_i(t(i), s_1(t_1), ..., s_n(t_n)), \text{ where } t(i) := t_i \tag{2.6}$$

. Now since each s_i is a dominant strategy for player i , then for every t_i, x_{-i}, x_i' we have that

$$u_i(t(i), s_i(t_i), x_{-i}) \geq u_i(t(i), x_i', x_{-i}) \tag{2.7}$$

$$u_i(t(i), s_i(t_i), s_{-i}(t_{-i})) \geq u_i(t(i), s_i(t_i'), s_{-i}(t_{-i})) \tag{2.8}$$

$$u_i^f(t_i, t_{-i}) \geq u_i^f(t_i', t_{-i}) \tag{2.9}$$

, which gives the definition of incentive compatibility of the original AGI $(f, u_1^f, ..., u_n^f)$. □

tips:if $\forall s_i$ is an injection , then the original AGI is equal to the synthetic training AGI given by:

$$\forall x_i \in Ims_i, a(x_1, ..., x_n) := f(s_1^{-1}(x_1), ..., s_n^{-1}(x_n))$$

3. Preorder on Syntheticness

Definition 3.1. *Measured data* is defined as data collected directly from measurements or experiments without any modifications from algorithms.

Definition 3.2. *Synthetic data* is defined as data that is *generated algorithmicly with prompt* or through algorithmic simulations based on models rather than being directly measured or collected from real-world events.

However, things are not purely black-and-white in real world. Even if some data came right out from a measurement instrument, it still may not be "purely-measured", because certain instruments have built-in denoise filter algorithms. Thus, the distinction between measured data and synthetic data is practically "fuzzy", with the following preorder on syntheticness.

Definition 3.3. For certain algorithm f,

$$a \succeq_f b \iff a = f(b) \tag{3.1}$$

In particular, if f is not bijective, then $a \succ_f b$. If f is bijective, then f is called a transform and $a \sim_f b$.

Theorem 3.1. *Suppose synthetic data is generated from two distinct AGI models a_1, a_2, with synthesis function $s_1(t_1, a_1), s_2(t_2, a_2)$. Then $\exists f$ as a training mechanism, such that $f(....) = a_0$ with $a_0 \succeq a_1, a_2$.*

Proof, hint an AGI model is a loss-less compression (bijective) on some observed data t.

References