Learning and Strongly Truthful Multi-Task Peer Prediction

A Variational Approach

Grant Schoenebeck, University of Michigan

Fang-Yi Yu, Harvard University

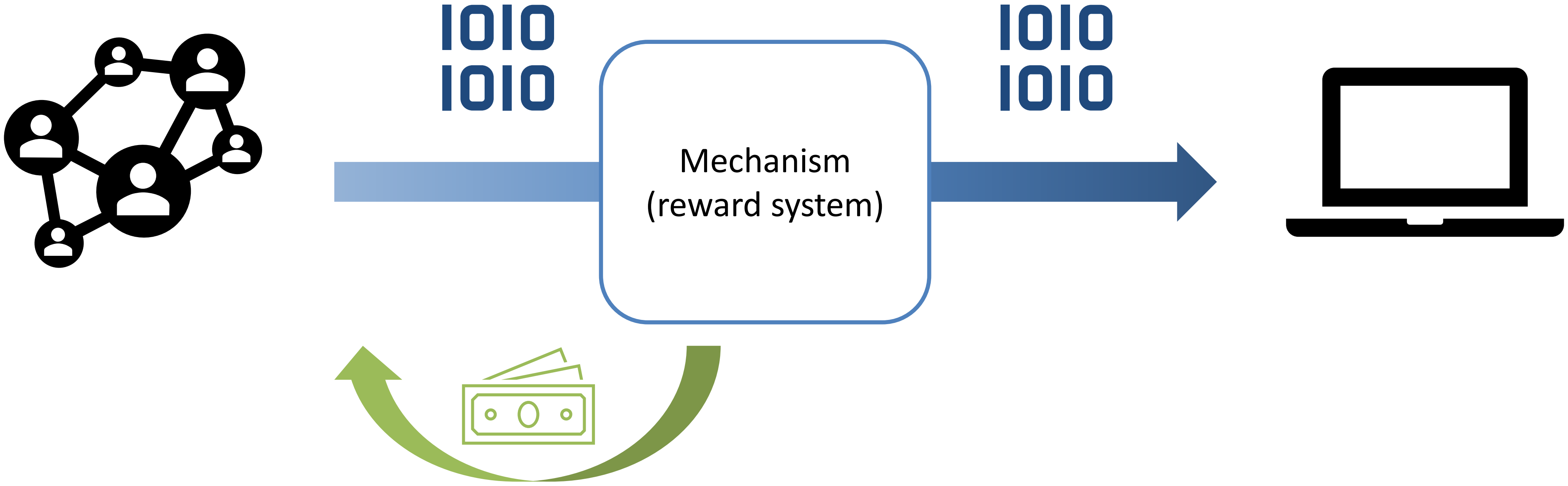
Elicit Information from Crowds

- Subjective
  - Are you happy?
  - Do you like the restaurant?
- Private
  - What is your commute time?

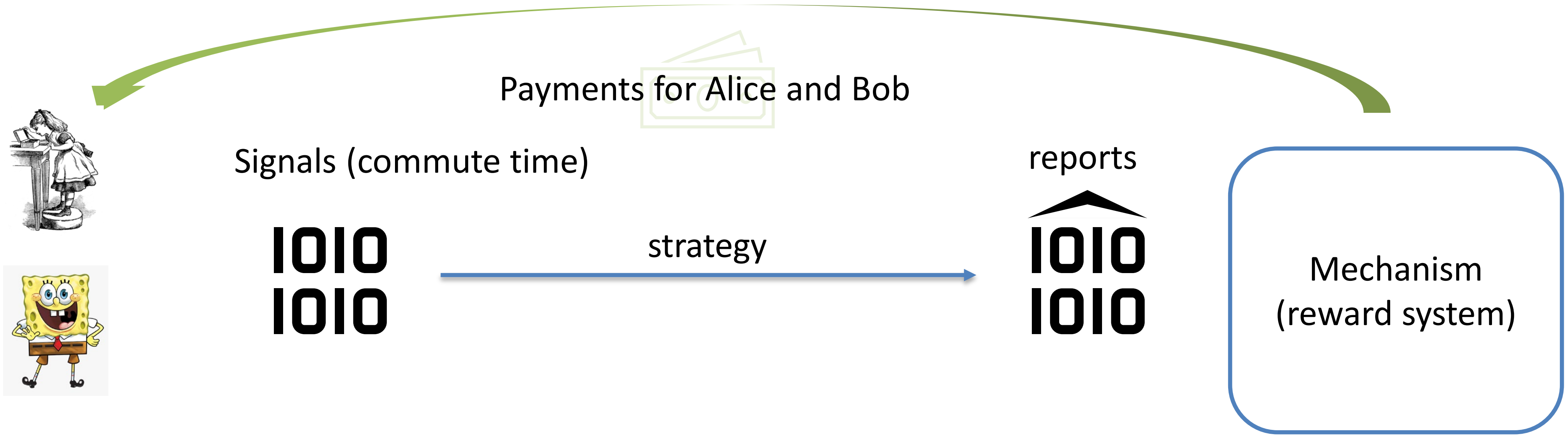
Cannot verify!

Data from strategic agents

Information elicitation



Setting of information elicitation



Setting of information elicitation

$M_A(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}), M_B(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$

Day 1, … , Day m

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \dots, \begin{pmatrix} x_m \\ y_m \end{pmatrix} \xrightarrow{\text{strategy}} \begin{pmatrix} \hat{x}_1 \\ \hat{y}_1 \end{pmatrix}, \dots, \begin{pmatrix} \hat{x}_m \\ \hat{y}_m \end{pmatrix}$$

Mechanism

$M_A, M_B: \mathcal{X}^m \times \mathcal{Y}^m \to \mathbb{R}$

A mechanism is truthful if truth telling maximizes the rewards of the both.

Multi-task information elicitation

$M_A(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}), M_B(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$

Uniform strategy

$$\binom{x_1}{y_1}, \dots, \binom{x_m}{y_m} \xrightarrow[\theta_B:\, \mathcal{Y} \to \mathcal{Y}]{\theta_A:\, \mathcal{X} \to \mathcal{X}} \binom{\hat{x}_1}{\hat{y}_1}, \dots, \binom{\hat{x}_m}{\hat{y}_m}$$

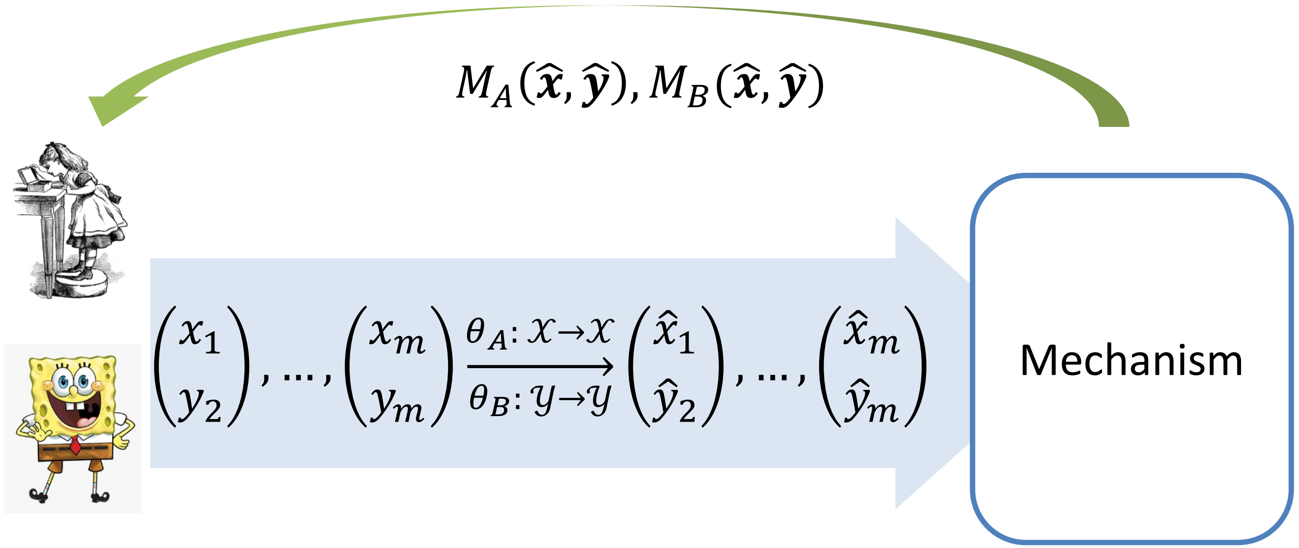
Mechanism
$M_A, M_B : \mathcal{X}^m \times \mathcal{Y}^m \to \mathbb{R}$

Each day's signals are sampled from $P_{X,Y}$
a joint distribution on $\mathcal{X} \times \mathcal{Y}$

A mechanism is (strongly) truthful if $\mathbb{E}[M_A(\boldsymbol{x}, \boldsymbol{y})] > \mathbb{E}[M_A(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})]$ and $\mathbb{E}[M_B(\boldsymbol{x}, \boldsymbol{y})] > \mathbb{E}[M_B(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})]$ for any nontruthful θ_A or θ_B

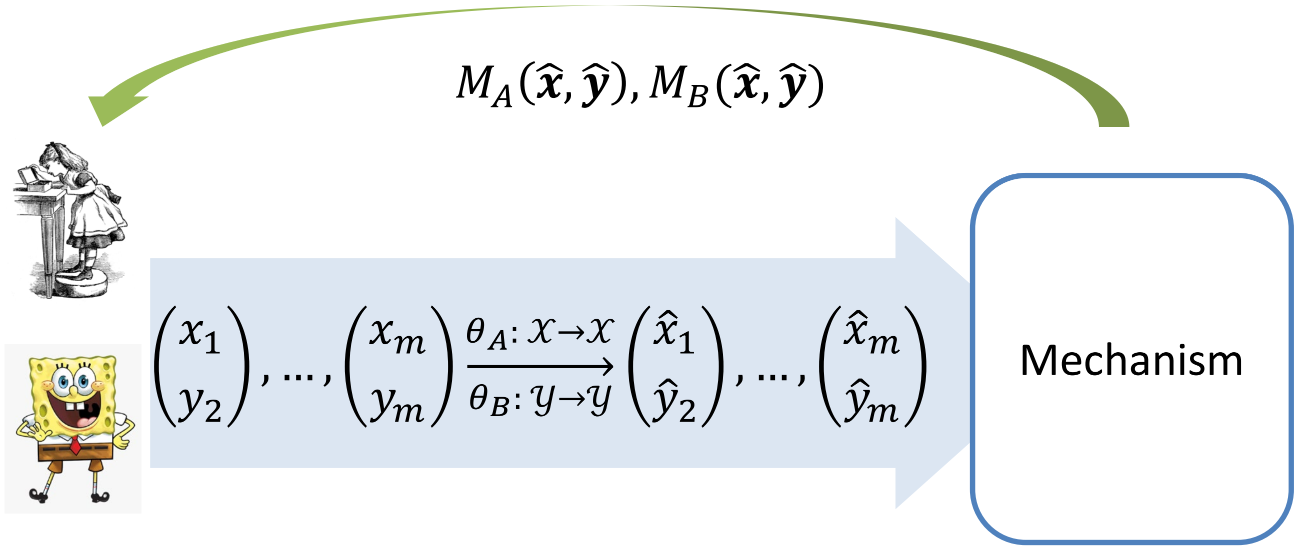
Goal of information elicitation

- Truthful > any nontruthful
  $\mathbb{E}[M_A(\boldsymbol{x},\boldsymbol{y})] > \mathbb{E}[M_A(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}})]$ and
  $\mathbb{E}[M_B(\boldsymbol{x},\boldsymbol{y})] > \mathbb{E}[M_B(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}})]$



Goal of information elicitation

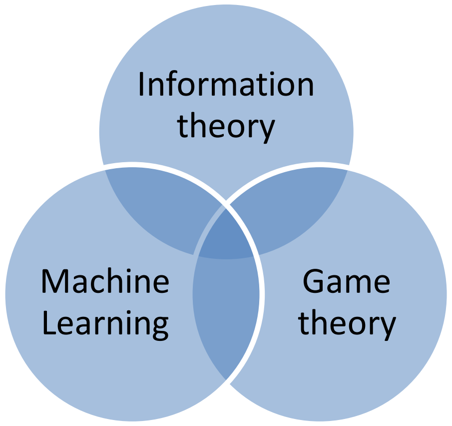
- Truthful > any nontruthful
- No verification
  - Private: What is your commute time?
  - Subjective: Do you like the restaurant?
- No knowledge about P_{XY}



Contributions

Propose pairing mechanisms

1. Elicit truthful reports from strategic agents even for general signal spaces, $\mathcal{X}$ and $\mathcal{Y}$
2. Generalize previous mechanisms
3. Connect information elicitation mechanism design to learning



Outline

- Model and our contributions
- From mechanism to learning
  - **Three observations**
  - Challenges for learning from strategic agents
  - Pairing mechanisms
- Connection to previous mechanisms

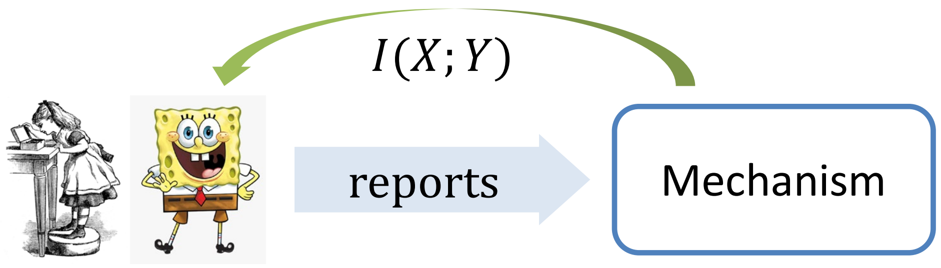
Three observations

1. Correlated signals P_{XY}
2. Strategy = data processing

$$Y \xrightarrow{P_{X|Y}} X \xrightarrow{\theta_A} \hat{X}$$

3. Data processing ineq. for mutual information $I(X;Y) = \mathbb{E}_{P_{XY}}\left[\ln \frac{P_{XY}}{P_X P_Y}\right]$

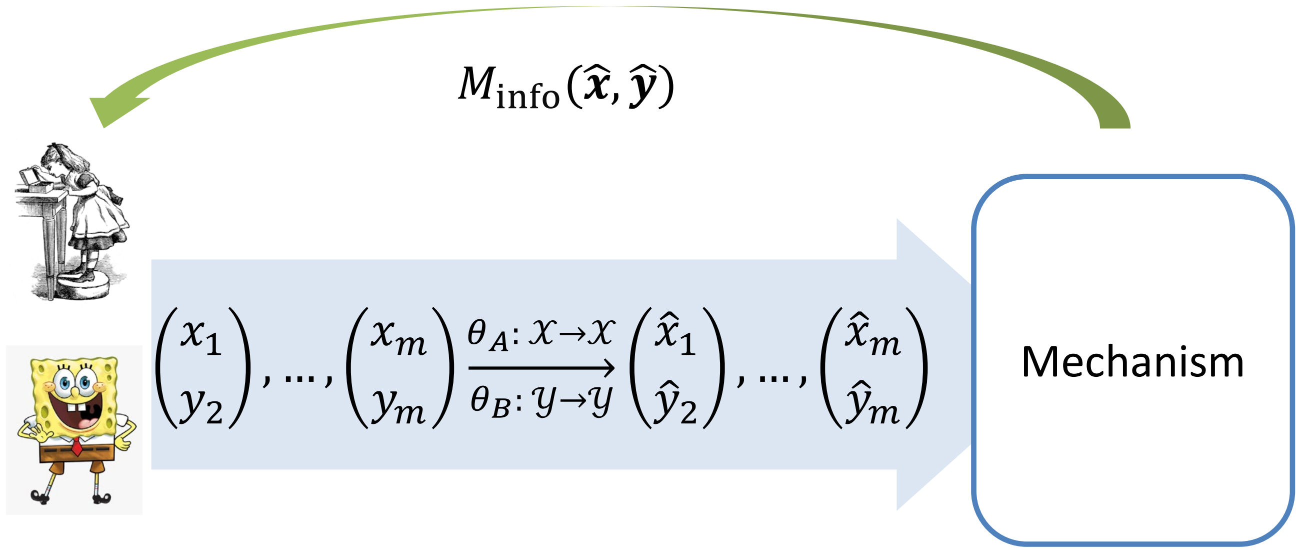
$$I(Y;X) \geq I(Y;\hat{X})$$



Mechanism to learning

- Approx. mutual information is approx. truthful

$$M_{\text{info}}(\boldsymbol{x}, \boldsymbol{y}) \approx I(X;Y) \geq I(\hat{X};\hat{Y}) \approx M_{\text{info}}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$$



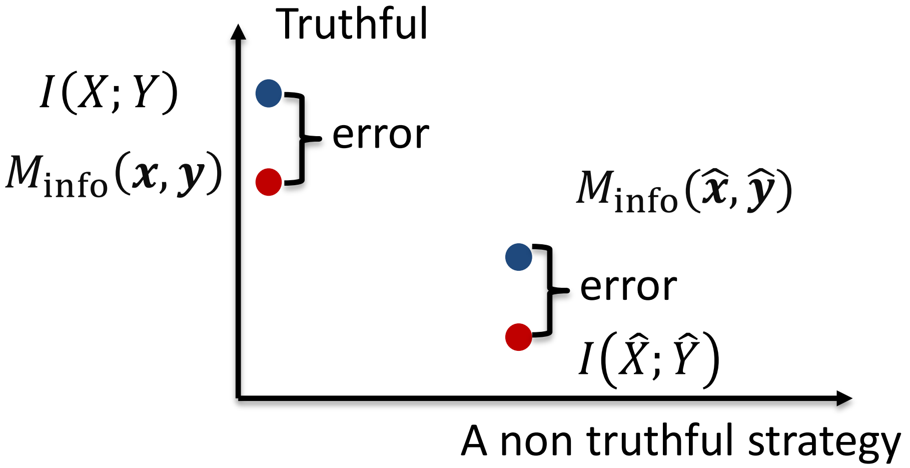
Outline

- Model and our contributions
- From mechanism to learning
  - Three observations
  - **Challenges for learning from strategic agents**
  - Pairing mechanisms
- Connection to previous mechanisms

Challenge for learning from strategic agents

- Approx. truthful

$$M_{\text{info}}(\boldsymbol{x},\boldsymbol{y}) \approx I(X;Y) \geq I(\hat{X};\hat{Y}) \approx M_{\text{info}}(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}})$$

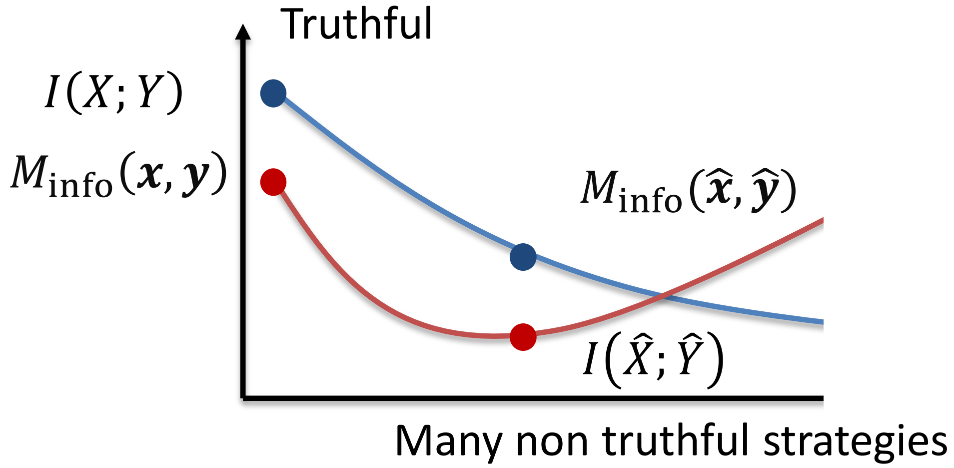


Challenge for learning from strategic agents

- Approx. truthful

$$M_{\text{info}}(\boldsymbol{x}, \boldsymbol{y}) \approx I(X;Y) \geq I\left(\hat{X};\hat{Y}\right) \approx M_{\text{info}}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$$

requires uniform estimate error bound for all strategies



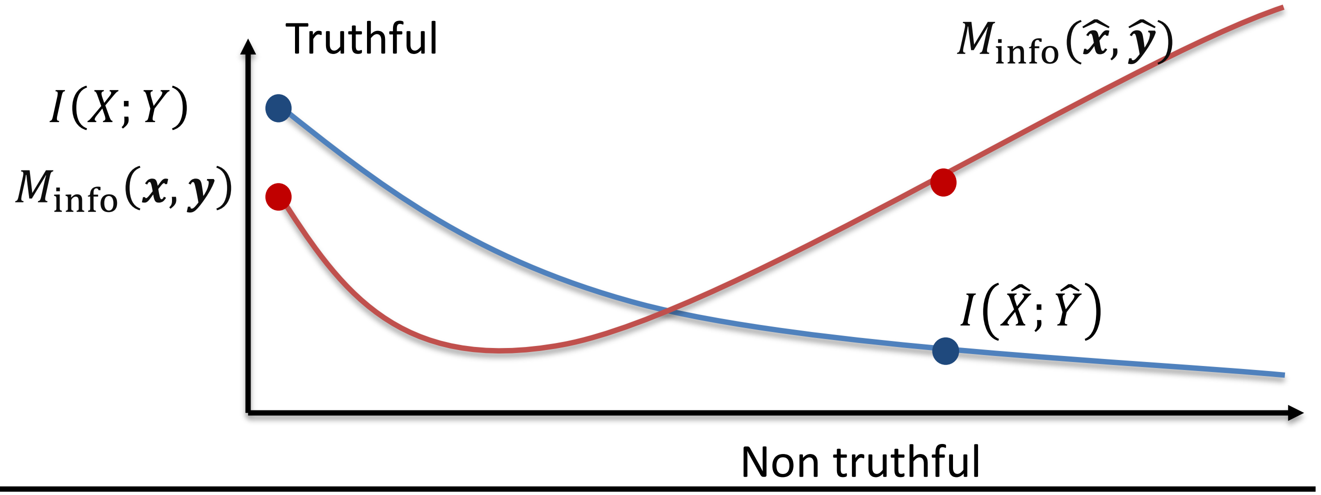
Challenge for learning from strategic agents

- Approx. truthful

$$M_{\text{info}}(\boldsymbol{x}, \boldsymbol{y}) \approx I(X;Y) \geq I(\hat{X};\hat{Y}) \approx M_{\text{info}}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$$

requires uniform estimate error bound for all strategies

- Strategic agents
- Large signal space $\mathcal{X} \times \mathcal{Y}$



Outline

- Model and our contributions
- From mechanism to learning
  - Three observations
  - Challenges for learning from strategic agents
  - **Pairing mechanisms**
- Connection to previous mechanisms

Pairing Mechanism

Suppose we have a scoring function $K: \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ mapping a pair of reports to a score. The pairing mechanism M_{pair}^{K}

1. Samples a pair on a common task, (x_b, y_b),
2. Samples a pair on distinct tasks, (x_p, y_q), and
3. Pays Alice and Bob

$$K(x_b, y_b) - \exp K(x_p, y_q) + 1$$

x_1	x_2	…	x_b	…	x_p	…	x_q	…	…	…	…	…	…	…	…	…	…	…	x_m
y_1	y_2	…	y_b	…	y_p	…	y_q	…	…	…	…	…	…	…	…	…	…	…	y_m

Tasks for payment

Connection of mutual information

- Given K, the expected payment is $\mathbb{E}\left[K(x_b, y_b) - \exp K\left(x_p, y_q\right)\right] + 1$

$$= \mathbb{E}_{P_{XY}}[K(X,Y)] - \mathbb{E}_{P_X P_Y}\left[\exp\left(K(X,Y)\right)\right] + 1$$

- Variational presentation of mutual information $I(X;Y) = \mathbb{E}_{P_{XY}}\left[-\ln \frac{P_X P_Y}{P_{XY}}\right]$

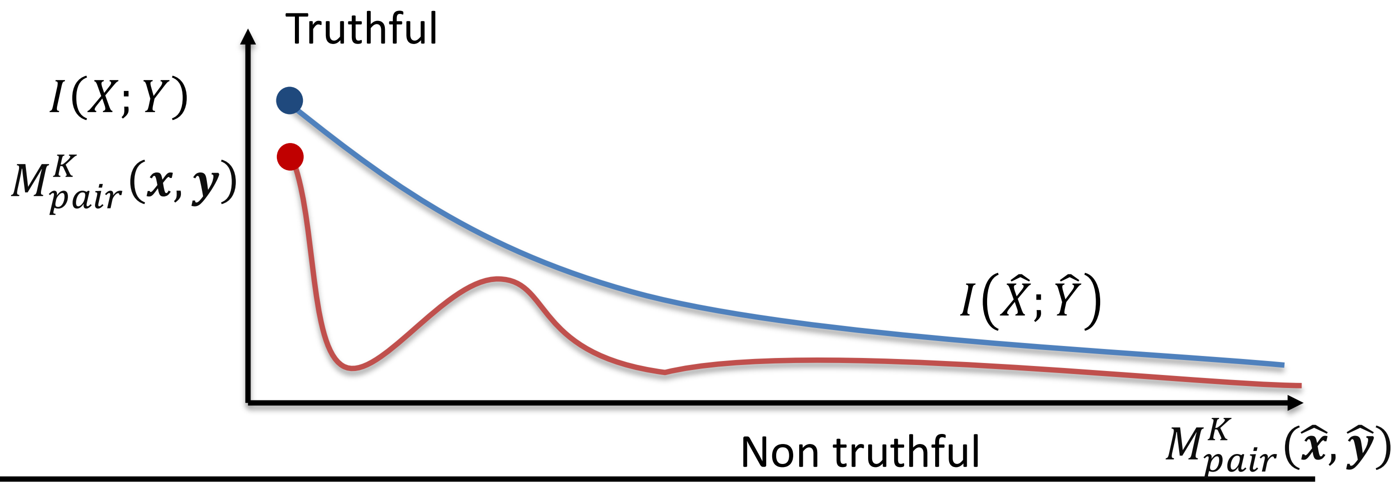
$$= \sup_{L:\mathcal{X}\times\mathcal{Y}\to\mathbb{R}} \mathbb{E}_{P_{XY}}[L(X,Y)] - \mathbb{E}_{P_X P_Y}\left[\exp\left(L(X,Y)\right)\right] + 1$$

and maximum happens at $L = K^{\star} = \ln\left(\frac{P_{XY}}{P_X P_Y}\right)$

Agent's Manipulation

$$M_{pair}^{\hat{K}}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) = \mathbb{E}\left[\hat{K}(\hat{x}_b, \hat{y}_b) - \exp\left(\hat{K}\left(\hat{x}_p, \hat{y}_q\right)\right)\right] + 1 \leq I\left(\hat{X}; \hat{Y}\right) \leq I(X; Y)$$

- Maximum happens only if both
  - $\hat{K} = K^\star$
  - truthful report
- Approx. truthful only requires error bound at the truthful strategy



Pairing Mechanism

Given a scoring function $K: \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, the pairing mechanism M_{pair}^{K}

1. Sample a pair on a common task, (x_b, y_b) .
2. Sample a pair on distinct tasks, (x_p, y_q).
3. Pay Alice and Bob

$$K(x_b, y_b) - \exp K(x_p, y_q) + 1$$

x_1	x_2	…	x_b	…	x_p	…	x_q	…	…	…	…	…	…	…	…	…	…	…	x_m
y_1	y_2	…	y_b	…	y_p	…	y_q	…	…	…	…	…	…	…	…	…	…	…	y_m

Tasks for payment

Pairing Mechanism (conti.)

1. Estimate ideal scoring rule $K^\star$ from tasks for learning.
2. Sample a pair on a common task, (x_b, y_b) .
3. Sample a pair on distinct tasks, (x_p, y_q).
4. Pay Alice and Bob

$$K(x_b, y_b) - \exp K(x_p, y_q) + 1$$

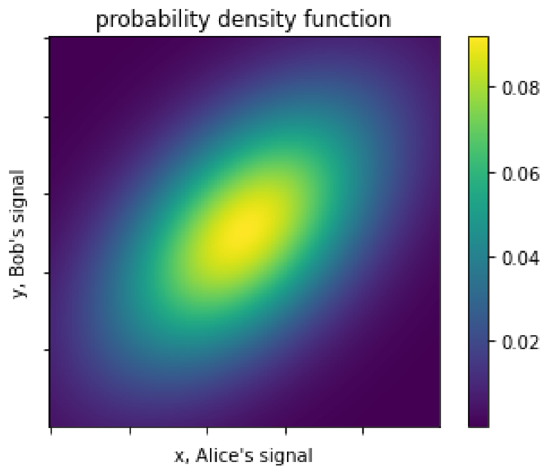
x_1	x_2	…	x_b	…	x_p	…	x_q	…	…	…	…	…	…	…	…	…	…	…	x_m
y_1	y_2	…	y_b	…	y_p	…	y_q	…	…	…	…	…	…	…	…	…	…	…	y_m

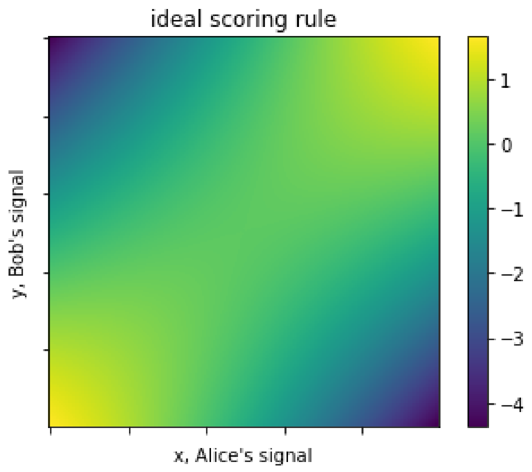
Tasks for payment — Tasks for learning

Pairing mechanism (conti.): $K^\star = \log \frac{P_{XY}}{P_X P_Y}$

- Plug-in estimator
- Optimization $K^\star = \text{argmax}_K \left\{ \mathbb{E}_{P_{X,Y}}[K(x,y)] - \mathbb{E}_{P_X P_Y}[\exp K(x,y)] \right\}$
  - Empirical risk minimization
  - Standard optimization
  - Deep neural network…

$\rho_{XY} = 1/2$





Variational method for strategic learning

Challenges of strategic learning

$$M_{\text{info}}(\boldsymbol{x}, \boldsymbol{y}) \approx I(X;Y) \geq I(\hat{X};\hat{Y}) \approx M_{\text{info}}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$$

requires uniform error bound

1. Strategy spaces are large
2. Agents are strategic

Variational representation

$$I(X;Y) = \sup_L \mathbb{E}_{P_{XY}}[L] - \mathbb{E}_{P_X P_Y}[e^L] + 1$$

becomes learning ideal scoring rules $K^\star$

1. Sufficient to bound the error at the truthful strategy
2. Agents want to help us to learn.

Outline

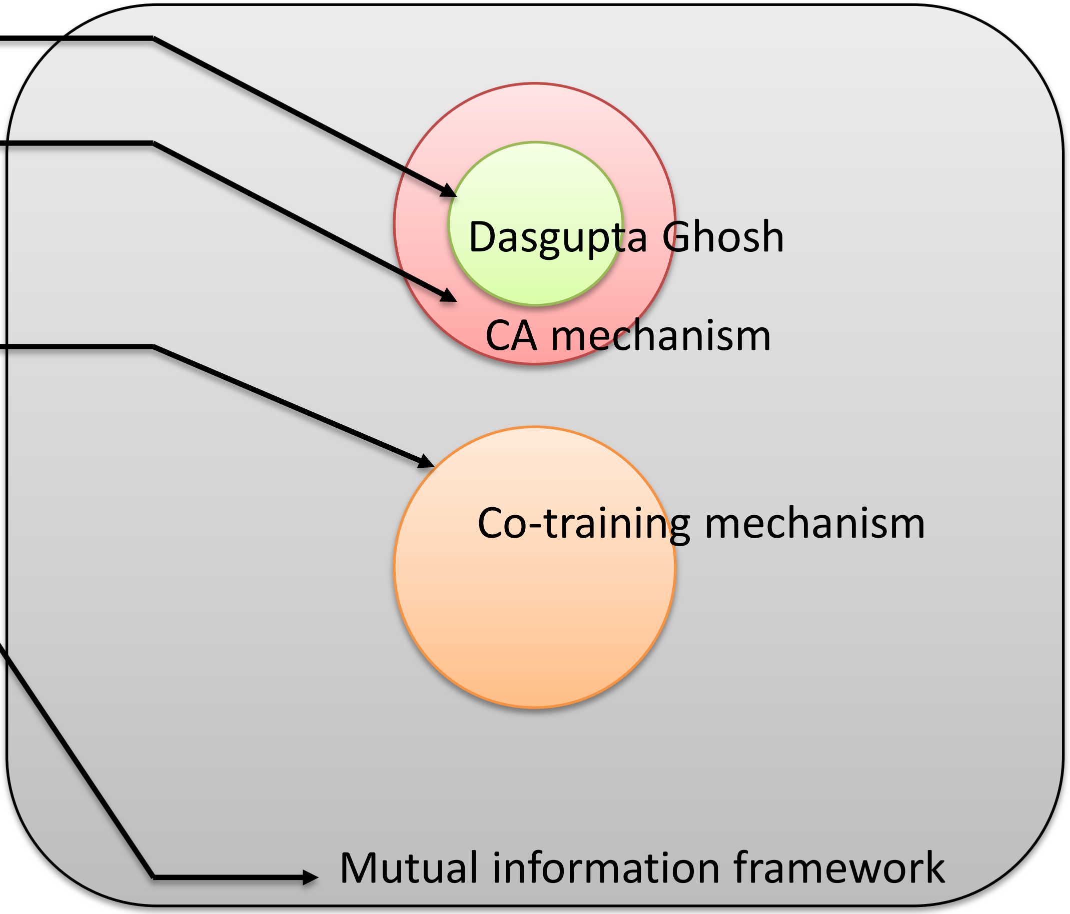
- Model and our contributions
- From mechanism to learning
  - Three observations
  - Challenges for learning from strategic agents
  - Pairing mechanisms
- **Connection to previous mechanisms**

Related Works in Multi-task Peer Prediction

- Mutual information framework
  - Binary positive correlated signals [Dasgupta, Ghosh 2013]
  - Correlated agreement mechanism [Shnyder et al 2016; Agarwal et al 2017]
  - Co-training mechanism [Kong, Schoenebeck 2018]
  - Mutual information framework [Kong, Schoenebeck 2019]
- Others
  - Determinant mechanism [Kong 2020]
  - Surrogate scoring rule mechanism [Chen et al 2020]

Mutual information framework

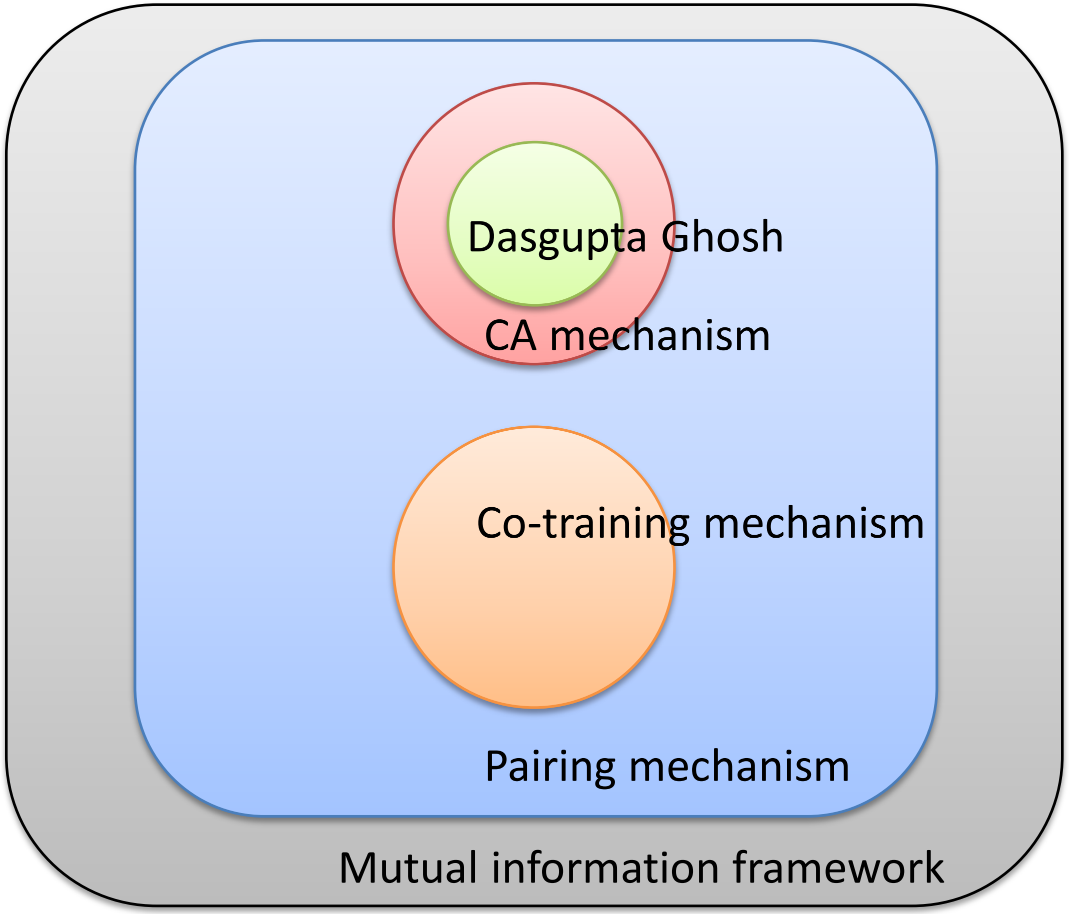
- Binary positive correlated signals
- Correlated agreement mechanism
- Co-training mechanism
- Mutual information framework



Contributions

If $P_{X,Y}$ is stochastic relevant, our pairing mechanism can elicit agents to report truthfully.

- General signal spaces, $\mathcal{X}$ and $\mathcal{Y}$
- Mechanism design to learning reduction



Special cases $\Phi(a) = |a - 1|$

- Total variational distance
  - $\Phi(a) = |a - 1|$
  - $\Phi^*(b) = b$ if $|b| \leq 1; \infty$, o.w.
  - $\Phi'(a) = \text{sign}(a - 1)$
- Pairing mechanism
  - Payment

$$K(\hat{x}_b, \hat{y}_b) - K(\hat{x}_p, \hat{y}_q)$$

  - Φ-ideal scoring function

$$\Phi'\left(\frac{dP_{XY}}{dP_X P_Y}\right) = \text{sign}(P_{XY} > P_X P_Y)$$

Special cases $\Phi(a) = |a - 1|$

CA mechanism

- Total variational distance
  - $\Phi(\mathrm{a}) = |a - 1|$
  - $\Phi^*(b) = b$ if $|b| \leq 1; \infty$, o.w.
  - $\Phi'(\mathrm{a}) = \mathrm{sign}(a - 1)$
- Pairing mechanism
  - Payment
    $$\boldsymbol{K}(\widehat{\boldsymbol{x}}_{\boldsymbol{b}}, \widehat{\boldsymbol{y}}_{\boldsymbol{b}}) - \boldsymbol{K}(\widehat{\boldsymbol{x}}_{\boldsymbol{p}}, \widehat{\boldsymbol{y}}_{\boldsymbol{q}})$$
  - Φ-ideal scoring function
    $$\Phi'\left(\frac{dP_{XY}}{dP_X P_Y}\right) = \mathrm{sign}(P_{XY} > P_X P_Y)$$

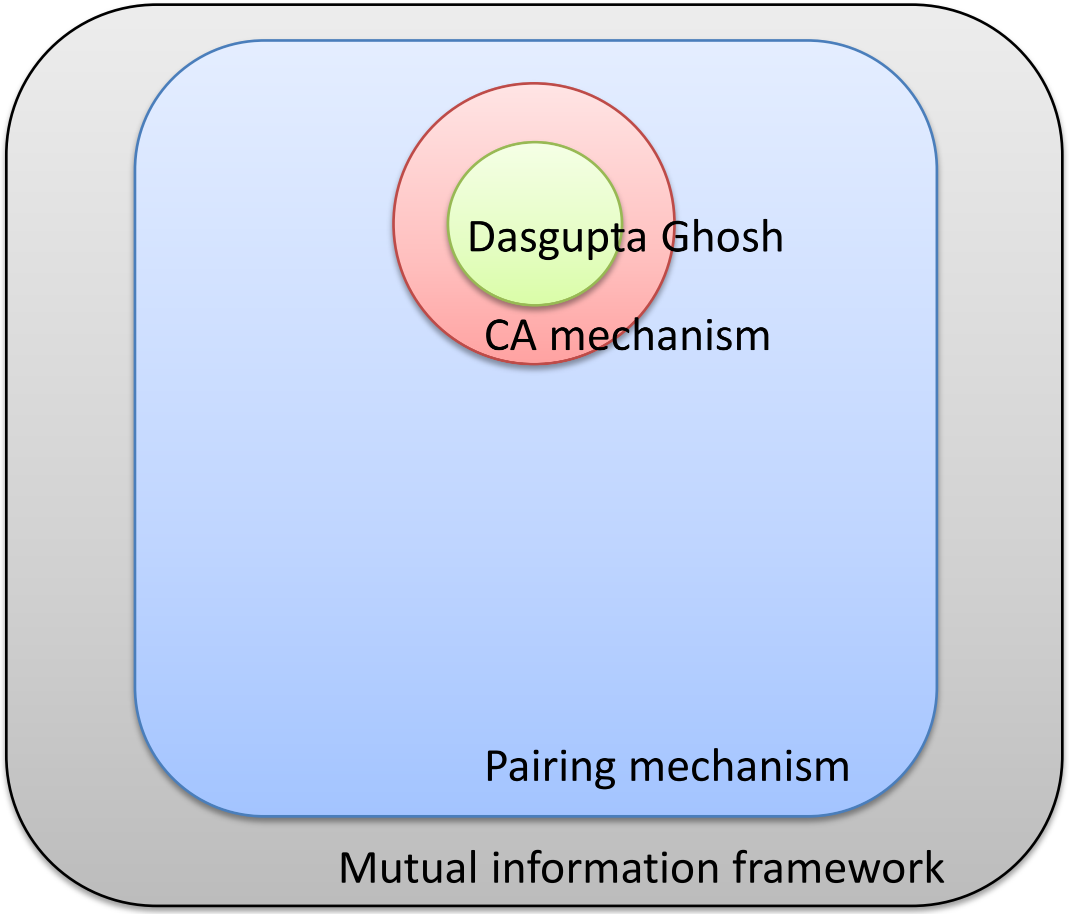
Dasgupta Ghosh

- Binary and positive correlated signals. For all $z = 0,1$
  $$P_{XY}(z, z) > P_X(z) P_Y(z)$$
- Pairing mechanism
  - Payment
    $$\mathbf{2}\left(\mathbf{1}[\widehat{\boldsymbol{x}}_{\boldsymbol{b}} = \widehat{\boldsymbol{y}}_{\boldsymbol{b}}] - \mathbf{1}[\widehat{\boldsymbol{x}}_{\boldsymbol{p}} = \widehat{\boldsymbol{y}}_{\boldsymbol{q}}]\right)$$
  - $K^\star(x, y) = \mathrm{sign}\big(P_{XY}(x, y) > P_X(x) P_Y(y)\big) = 2 \cdot 1[x = y] - 1$

Comparisons of Peer Prediction mechanisms

- Pairing mechanism pays an approximation of mutual information
- CA mechanism is the pairing mechanism with

$$\Phi(a) = |a - 1|$$



Comparisons of Peer Prediction mechanisms

- Pairing mechanism pays an approximation of mutual information
- CA mechanism is the pairing mechanism with

$$\Phi(\mathrm{a}) = |a - 1|$$

- Co-training mechanism

$$\frac{dP_{XY}}{dP_X P_Y} = \sum_w \frac{P(W|X)P(W|Y)}{P(W)}$$

$$K^\star = \Phi'\left(\frac{dP_{XY}}{dP_X P_Y}\right)$$