PAC Learning from a Strategic Crowd

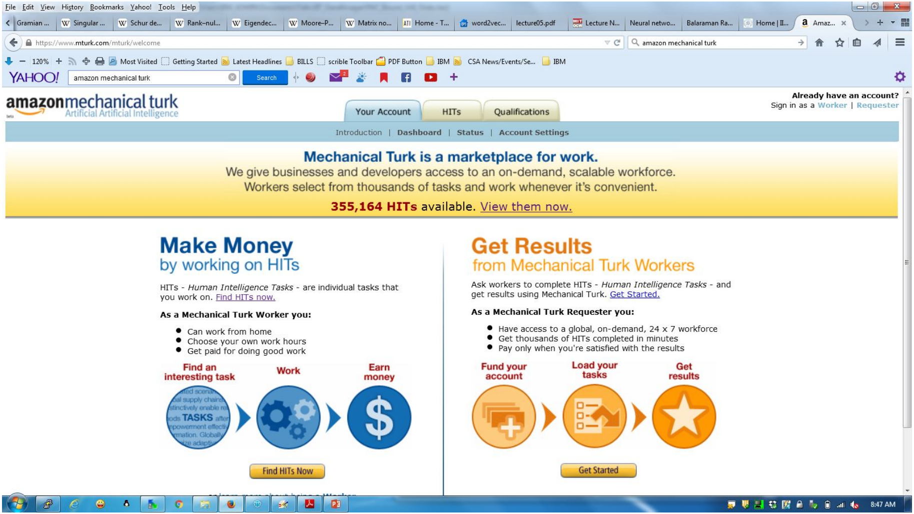
Dinesh Garg
IBM Research - Bangalore
Joint work with
Sourangshu Bhattacharya, S. Sundararajan, and Shirish Shevade

March 17, 2016

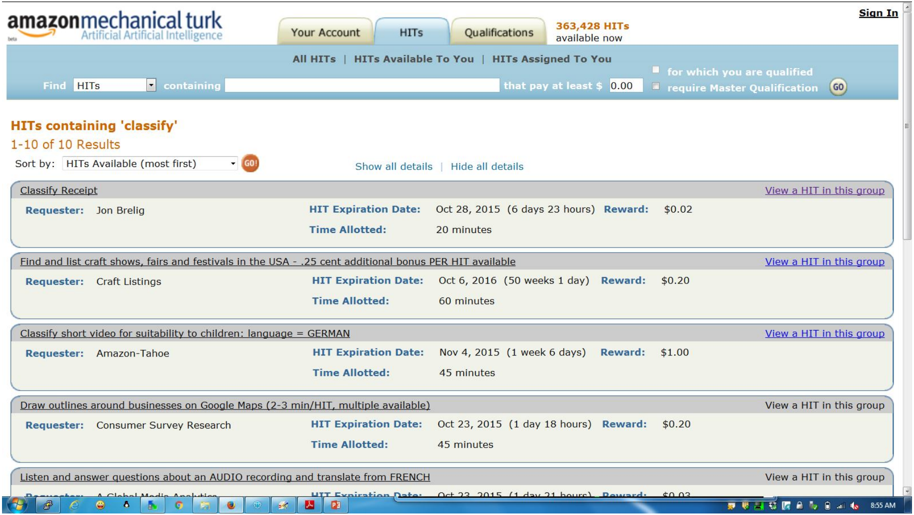
Data is New Natural Resource

- *Ginni Rometty, CEO, IBM*

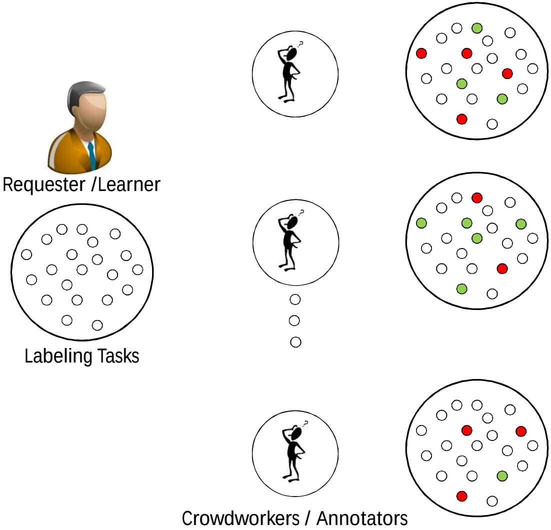
Amazon's Mechanical Turk (M-Turk)



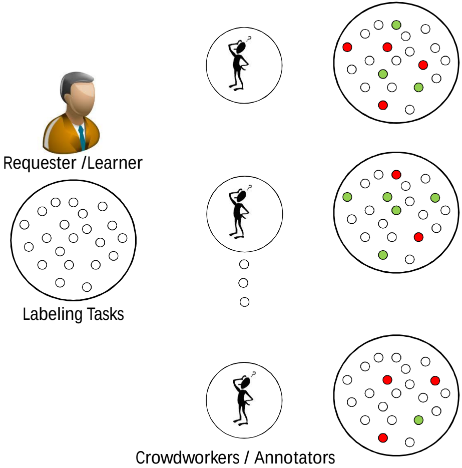
Human Intelligence Tasks (HITs)



Data Labeling: Not a Child's Play

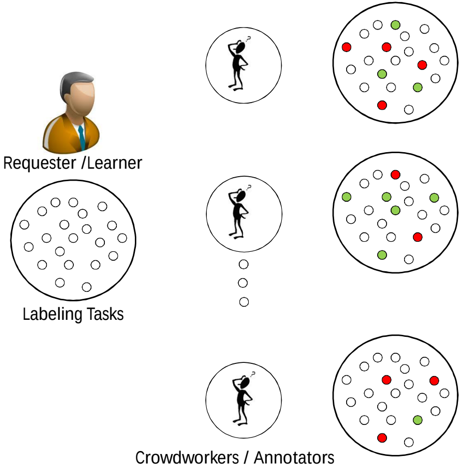


Data Labeling: Not a Child's Play



Data / Annotators	x_1	x_2	....	x_m	True Label
A_1	+1	?	....	-1	?
A_2	-1	+1	....	-1	?
A_3	?	+1	....	?	?
......	....	....	....	....	....
A_n	?	+1	....	?	?

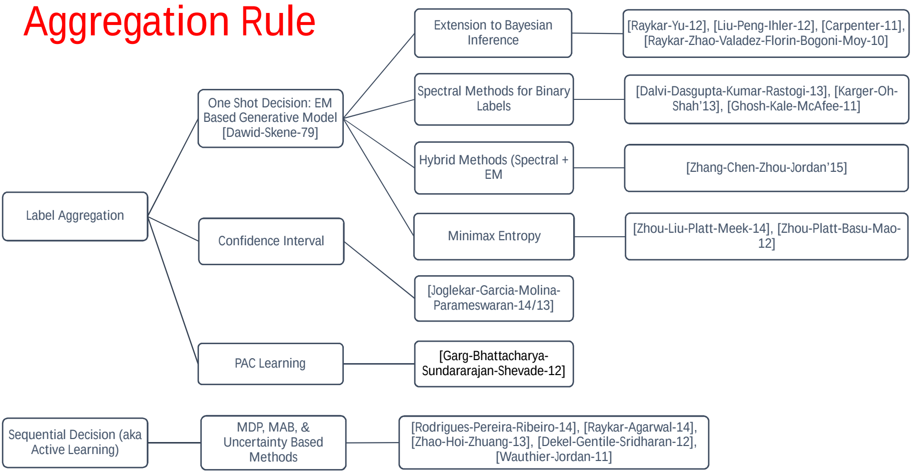
Data Labeling: Not a Child's Play



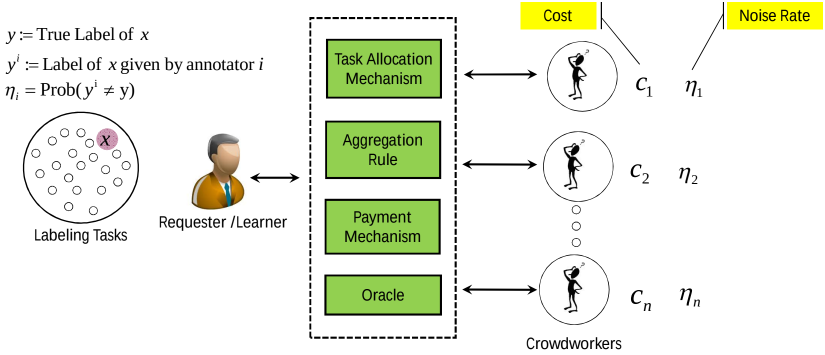
Data / Annotators	x_1	x_2	....	x_m	True Label
A_1	+1	?	....	-1	?
A_2	-1	+1	....	-1	?
A_3	?	+1	....	?	?
......	....	....	....	....	....
A_n	?	+1	....	?	?

- How to aggregate the labels ?
- Who should annotate what?
- How much to pay for?

Prior Work


Aggregation Rule
Label Aggregation
One Shot Decision: EM Based Generative Model [Dawid-Skene-79]
Confidence Interval
PAC Learning
Extension to Bayesian Inference
Spectral Methods for Binary Labels
Hybrid Methods (Spectral + EM
Minimax Entropy
[Joglekar-Garcia-Molina-Parameswaran-14/13]
[Garg-Bhattacharya-Sundararajan-Shevade-12]
[Raykar-Yu-12], [Liu-Peng-Ihler-12], [Carpenter-11], [Raykar-Zhao-Valadez-Florin-Bogoni-Moy-10]
[Dalvi-Dasgupta-Kumar-Rastogi-13], [Karger-Oh-Shah'13], [Ghosh-Kale-McAfee-11]
[Zhang-Chen-Zhou-Jordan'15]
[Zhou-Liu-Platt-Meek-14], [Zhou-Platt-Basu-Mao-12]
Sequential Decision (aka Active Learning)
MDP, MAB, & Uncertainty Based Methods
[Rodrigues-Pereira-Ribeiro-14], [Raykar-Agarwal-14], [Zhao-Hoi-Zhuang-13], [Dekel-Gentile-Sridharan-12], [Wauthier-Jordan-11]

Binary Labeling: A Mental Model



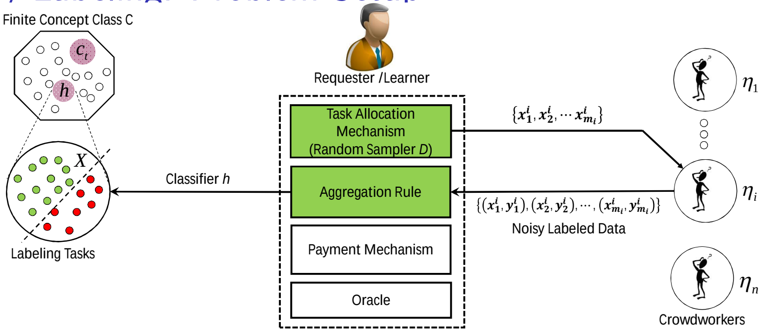
Annotators:

- Multiple noisy human annotators
- Noise could be due to human error, lack of expertise, or even intentional
- Expertise level of an annotator can be expressed by its noise rate
- Each annotator needs to be paid

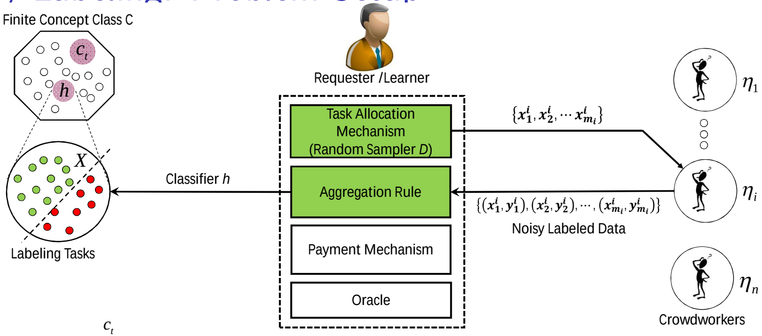
Learner:

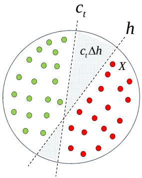
- Goal is to obtain good quality labels at minimum cost

Binary Labeling: Problem Setup



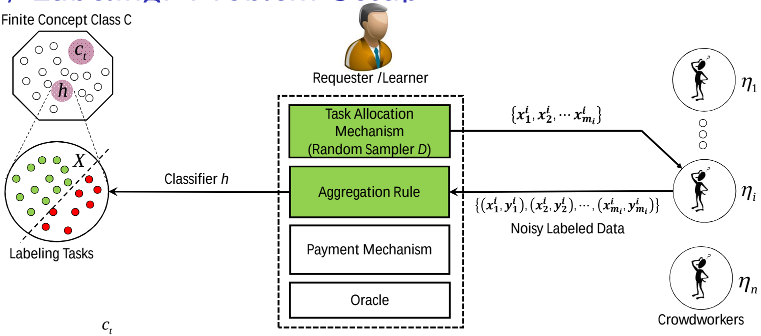
Binary Labeling: Problem Setup

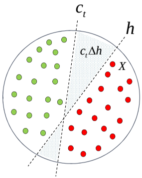




- annotation plan $\mathbf{m} := (m_1, m_2, \ldots, m_n)$
- error rate of $h := \mathbf{Pr}^D(c_t \Delta h)$
- ϵ-bad hypothesis $:= \mathbf{Pr}^D(c_t \Delta h) > \epsilon$
- PAC Bound $:= \mathbf{Pr}^{\mathbf{m}}(\mathbf{Pr}^D(c_t \Delta h) > \epsilon) < \delta$

Binary Labeling: Problem Setup





- annotation plan $\mathbf{m} := (m_1, m_2, \ldots., m_n)$
- error rate of $h := \mathbf{Pr}^D(c_t \Delta h)$
- ϵ-bad hypothesis $:= \mathbf{Pr}^D(c_t \Delta h) > \epsilon$
- PAC Bound $:= \mathbf{Pr}^{\mathbf{m}}(\mathbf{Pr}^D(c_t \Delta h) > \epsilon) < \delta$

Goal: Design an (1) Aggregation Rule and an (2) Annotation Plan to ensure PAC bound for the learned classifier h at (3) Minimum Cost.

[1] L.G. Valiant, "A Theory of Learnable", Communications of the ACM, 27:1134-1142, 1984.

(1) Aggregation Rule: *Minimum Disagreement Algorithm*

Input: Labeled examples from n annotators.
Output: A hypothesis $h^* \in \mathscr{C}$
Algorithm:

1. Let $\{(x_j^i, y_j^i)\ \ i = 1, 2, \ldots, n;\ j = 1, \ldots, m_i\}$ be the labeled examples.
2. Ouput a hypothesis h^* that minimally disagrees with the given labels (use any tie breaking rule). That is,

$$h^* \in \arg\min_{h \in \mathscr{C}} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \mathbf{1}(h(x_j^i) \neq y_j^i)$$

Properties of the MDA

- Does not require the knowledge of annotators' noise rates η_i (Analysis would require !!)
- Does not require the knowledge of sampling distribution D

(2) Annotation Plan for MDA *[Complete Info. Setting]*

(2) Annotation Plan for MDA *[Complete Info. Setting]*

Learner's Problem: "Which annotation plan would guarantee me (ϵ, δ) PAC bound?"

(2) Annotation Plan for MDA *[Complete Info. Setting]*

Learner's Problem: "Which annotation plan would guarantee me (ϵ, δ) PAC bound?"

Assumption: Learner precisely knows the noise rate η_i of every annotator i

(2) Annotation Plan for MDA *[Complete Info. Setting]*

Learner's Problem: "Which annotation plan would guarantee me (ϵ, δ) PAC bound?"

Assumption: Learner precisely knows the noise rate η_i of every annotator i

Theorem (Feasible Annotation Plan for MDA)

The MDA will satisfy PAC bound if the annotation plan $\mathbf{m} = (m_1, m_2, \ldots, m_n)$ *satisfies:*

$$\log(N/\delta) \leq \sum_{i=1}^{n} m_i \psi(\eta_i) \tag{1}$$

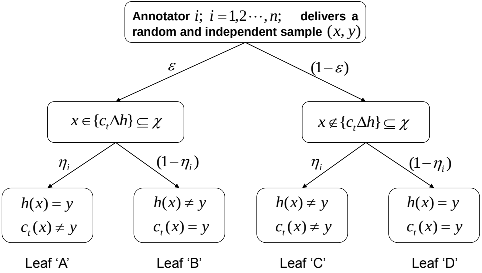
where concept class is finite, i.e. $N = |\mathscr{C}| < \infty$ *and* $\forall i = 1, 2, \ldots, n$*, we have*

- $0 < \eta_i < 1/3$
- $\psi(\eta_i) = -\log\left[1 - \epsilon\left(1 - \exp\left(\frac{3\eta_i - 1}{8}\right)\right)\right]$.

D. Garg, S. Bhattacharya, S. Sundararajan, S. Shevade, "Mechanism Design for Cost Optimal PAC Learning in the Presence of Strategic Noisy Annotators", Uncertainty in Artificial Intelligence (UAI), 275-285, 2012.

Proof Sketch

Probability of an ϵ-bad hypothesis h having lower empirical error than c_t

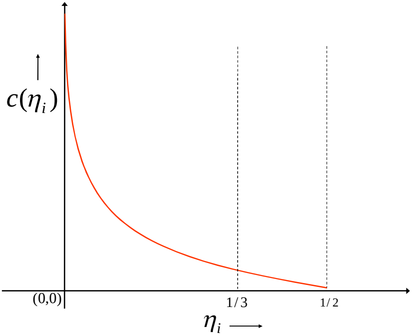


$$\mathbf{Pr}^{(m_1,\ldots,m_n)}[L_e(h) \leq L_e(c_t)] = \mathbf{Pr}\{\#\text{ samples under leaf A} \geq \#\text{ samples under leaf B}\}$$

(3) Cost of Annotation

Assumptions:

- Each annotator i incurs a cost of $c(\eta_i)$ for labeling one data point
- The cost function $c(\cdot)$ is the same for all the annotators
- $c(\cdot)$ is bounded, continuously differentiable, and strictly decreasing function
- Function $c(\cdot)$ is a common knowledge



- A more competitive annotator i means low η_i
- He can earn more by selling his services (time)
- It means his internal cost of annotation is high

(1-2-3) Putting It All Together *[Complete Info Setting]*

Learner's Problem:

- Learner is using MDA as an aggregation rule to learn a binary classifier
- Learner precisely knows the cost (equivalently, noise rates η_i) of each annotator i
- Learner wants to ensure PAC learning with parameters (ϵ, δ)
- Learner wants to minimize the cost of a feasible annotation plan

(1-2-3) Putting It All Together *[Complete Info Setting]*

Learner's Problem:

- Learner is using MDA as an aggregation rule to learn a binary classifier
- Learner precisely knows the cost (equivalently, noise rates η_i) of each annotator i
- Learner wants to ensure PAC learning with parameters (ϵ, δ)
- Learner wants to minimize the cost of a feasible annotation plan

Relaxed Primal Problem

$$\underset{m_1, m_2, \ldots m_n}{\text{Minimize}} \quad \sum_{i=1}^{n} c(\eta_i) m_i$$

$$\text{subject to} \quad \log(N/\delta) \leq \sum_{i=1}^{n} \psi(\eta_i) m_i$$

$$0 \leq m_i \ \forall i$$

(1-2-3) Putting It All Together *[Complete Info Setting]*

Learner's Problem:

- Learner is using MDA as an aggregation rule to learn a binary classifier
- Learner precisely knows the cost (equivalently, noise rates η_i) of each annotator i
- Learner wants to ensure PAC learning with parameters (ϵ, δ)
- Learner wants to minimize the cost of a feasible annotation plan

Relaxed Primal Problem

$$\underset{m_1, m_2, \ldots m_n}{\text{Minimize}} \quad \sum_{i=1}^{n} c(\eta_i) m_i$$

$$\text{subject to} \quad \log(N/\delta) \leq \sum_{i=1}^{n} \psi(\eta_i) m_i$$

$$0 \leq m_i \ \forall i$$

Relaxed Dual Problem

$$\underset{\lambda}{\text{Maximize}} \quad \lambda \log\left(\frac{N}{\delta}\right)$$

$$\text{subject to} \quad \lambda \leq \frac{c(\eta_i)}{\psi(\eta_i)} \ \forall i$$

$$0 \leq \lambda$$

(1-2-3) Putting It All Together *[Complete Info Setting]*

Learner's Problem:

- Learner is using MDA as an aggregation rule to learn a binary classifier
- Learner precisely knows the cost (equivalently, noise rates η_i) of each annotator i
- Learner wants to ensure PAC learning with parameters (ϵ, δ)
- Learner wants to minimize the cost of a feasible annotation plan

Relaxed Primal Problem

$$\begin{aligned} \underset{m_1, m_2, \ldots m_n}{\text{Minimize}} \quad & \sum_{i=1}^{n} c(\eta_i) m_i \\ \text{subject to} \quad & \log(N/\delta) \leq \sum_{i=1}^{n} \psi(\eta_i) m_i \\ & 0 \leq m_i \ \forall i \end{aligned}$$

Relaxed Dual Problem

$$\begin{aligned} \underset{\lambda}{\text{Maximize}} \quad & \lambda \log\left(\frac{N}{\delta}\right) \\ \text{subject to} \quad & \lambda \leq \frac{c(\eta_i)}{\psi(\eta_i)} \ \forall i \\ & 0 \leq \lambda \end{aligned}$$

Definition (Near Optimal Allocation Rule - NOAR)

Let i^* be the annotator having minimum value for *cost-per-quality* given by $c(\eta_i)/\psi(\eta_i)$. The learner should buy $\lceil \log(N/\delta)/\psi(\eta_{i^*}) \rceil$ number of examples from such an annotator.

(1-2-3) Putting It All Together *[Complete Info Setting]*

Theorem

Let $COST$ *be the total cost of purchase incurred by the* Near Optimal Allocation Rule*. Let* OPT *be the optimal value of the ILP. Then,*

$$OPT \leq COST \leq OPT\left(1 + \frac{1}{m_0}\right)$$

where $m_0 = \log\left(\frac{1}{1-\epsilon}\right)$

Proof:

$$\begin{aligned} COST &= c(\eta_{i^*})\lceil \log(N/\delta)/\psi(\eta_{i^*})\rceil \\ &\leq \log(N/\delta)c(\eta_{i^*})/\psi(\eta_{i^*}) + c(\eta_{i^*}) \\ &\leq OPT + c(\eta_{i^*}) \\ &\leq OPT + m_0 c(\eta_{i^*})/m_0 \\ &\leq OPT + OPT/m_0 \end{aligned}$$

Back to Binary Labeling Problem: *Incomplete Info Setting*

Back to Binary Labeling Problem: *Incomplete Info Setting*

Let us Face the Reality

Back to Binary Labeling Problem: *Incomplete Info Setting*

Let us Face the Reality

- Learner does not know the cost (equivalently, noise rate) of any annotator

Back to Binary Labeling Problem: *Incomplete Info Setting*

Let us Face the Reality

- Learner does not know the cost (equivalently, noise rate) of any annotator

So What?

Back to Binary Labeling Problem: *Incomplete Info Setting*

Let us Face the Reality

- Learner does not know the cost (equivalently, noise rate) of any annotator

So What?

- Learner can not compute the PAC annotation plan because $\psi(\eta_i)$ is required for this: $\log(N/\delta) \leq \sum_{i=1}^{n} \psi(\eta_i) m_i$

Back to Binary Labeling Problem: *Incomplete Info Setting*

Let us Face the Reality

- Learner does not know the cost (equivalently, noise rate) of any annotator

So What?

- Learner can not compute the PAC annotation plan because $\psi(\eta_i)$ is required for this: $\log(N/\delta) \leq \sum_{i=1}^{n} \psi(\eta_i) m_i$

Options Available with Learner

Back to Binary Labeling Problem: *Incomplete Info Setting*

Let us Face the Reality

- Learner does not know the cost (equivalently, noise rate) of any annotator

So What?

- Learner can not compute the PAC annotation plan because $\psi(\eta_i)$ is required for this: $\log(N/\delta) \leq \sum_{i=1}^{n} \psi(\eta_i) m_i$

Options Available with Learner

- Estimation

Back to Binary Labeling Problem: *Incomplete Info Setting*

Let us Face the Reality

- Learner does not know the cost (equivalently, noise rate) of any annotator

So What?

- Learner can not compute the PAC annotation plan because $\psi(\eta_i)$ is required for this: $\log(N/\delta) \leq \sum_{i=1}^{n} \psi(\eta_i) m_i$

Options Available with Learner

- Estimation

  - Overestimation $\Rightarrow$ Excess examples procured by NOAR $\Rightarrow$ Higher COST

Back to Binary Labeling Problem: *Incomplete Info Setting*

Let us Face the Reality

- Learner does not know the cost (equivalently, noise rate) of any annotator

So What?

- Learner can not compute the PAC annotation plan because $\psi(\eta_i)$ is required for this: $\log(N/\delta) \leq \sum_{i=1}^{n} \psi(\eta_i) m_i$

Options Available with Learner

- Estimation

  - Overestimation $\Rightarrow$ Excess examples procured by NOAR $\Rightarrow$ Higher COST

  - Underestimation $\Rightarrow$ **Pr**(ϵ-bad hypothesis gets picked by NOAR) $> \delta$

Back to Binary Labeling Problem: *Incomplete Info Setting*

Let us Face the Reality

- Learner does not know the cost (equivalently, noise rate) of any annotator

So What?

- Learner can not compute the PAC annotation plan because $\psi(\eta_i)$ is required for this: $\log(N/\delta) \leq \sum_{i=1}^{n} \psi(\eta_i) m_i$

Options Available with Learner

- Estimation
  - Overestimation $\Rightarrow$ Excess examples procured by NOAR $\Rightarrow$ Higher COST
  - Underestimation $\Rightarrow$ **Pr**(ϵ-bad hypothesis gets picked by NOAR) $> \delta$
- Elicitation

Back to Binary Labeling Problem: *Incomplete Info Setting*

Let us Face the Reality

- Learner does not know the cost (equivalently, noise rate) of any annotator

So What?

- Learner can not compute the PAC annotation plan because $\psi(\eta_i)$ is required for this: $\log(N/\delta) \leq \sum_{i=1}^{n} \psi(\eta_i) m_i$

Options Available with Learner

- Estimation

  - Overestimation $\Rightarrow$ Excess examples procured by NOAR $\Rightarrow$ Higher COST

  - Underestimation $\Rightarrow$ **Pr**(ϵ-bad hypothesis gets picked by NOAR) $> \delta$

- Elicitation

  - Invite annotators to report (bid) their costs (equivalently, noise rates)

Back to Binary Labeling Problem: *Incomplete Info Setting*

Let us Face the Reality

- Learner does not know the cost (equivalently, noise rate) of any annotator

So What?

- Learner can not compute the PAC annotation plan because $\psi(\eta_i)$ is required for this: $\log(N/\delta) \leq \sum_{i=1}^{n} \psi(\eta_i) m_i$

Options Available with Learner

- Estimation

  - Overestimation $\Rightarrow$ Excess examples procured by NOAR $\Rightarrow$ Higher COST

  - Underestimation $\Rightarrow$ $\mathbf{Pr}(\epsilon\text{-bad hypothesis gets picked by NOAR}) > \delta$

- Elicitation

  - Invite annotators to report (bid) their costs (equivalently, noise rates)

  - Setup an auction to decide the work (contract) size and payment for annotators

Back to Binary Labeling Problem: *Incomplete Info Setting*

Let us Face the Reality

- Learner does not know the cost (equivalently, noise rate) of any annotator

So What?

- Learner can not compute the PAC annotation plan because $\psi(\eta_i)$ is required for this: $\log(N/\delta) \leq \sum_{i=1}^{n} \psi(\eta_i) m_i$

Options Available with Learner

- Estimation
  - Overestimation $\Rightarrow$ Excess examples procured by NOAR $\Rightarrow$ Higher COST
  - Underestimation $\Rightarrow$ **Pr**(ϵ-bad hypothesis gets picked by NOAR) $> \delta$
- Elicitation
  - Invite annotators to report (bid) their costs (equivalently, noise rates)
  - Setup an auction to decide the work (contract) size and payment for annotators
  - Challenge: If annotators misreport noise rates, we are *back to square one!!*

Back to Binary Labeling Problem: *Incomplete Info Setting*

Let us Face the Reality

- Learner does not know the cost (equivalently, noise rate) of any annotator

So What?

- Learner can not compute the PAC annotation plan because $\psi(\eta_i)$ is required for this: $\log(N/\delta) \leq \sum_{i=1}^{n} \psi(\eta_i) m_i$

Options Available with Learner

- Estimation
  - Overestimation $\Rightarrow$ Excess examples procured by NOAR $\Rightarrow$ Higher COST
  - Underestimation $\Rightarrow$ **Pr**(ϵ-bad hypothesis gets picked by NOAR) $> \delta$
- Elicitation
  - Invite annotators to report (bid) their costs (equivalently, noise rates)
  - Setup an auction to decide the work (contract) size and payment for annotators
  - Challenge: If annotators misreport noise rates, we are *back to square one!!*

Back to Binary Labeling Problem: *Incomplete Info Setting*

Let us Face the Reality

- Learner does not know the cost (equivalently, noise rate) of any annotator

So What?

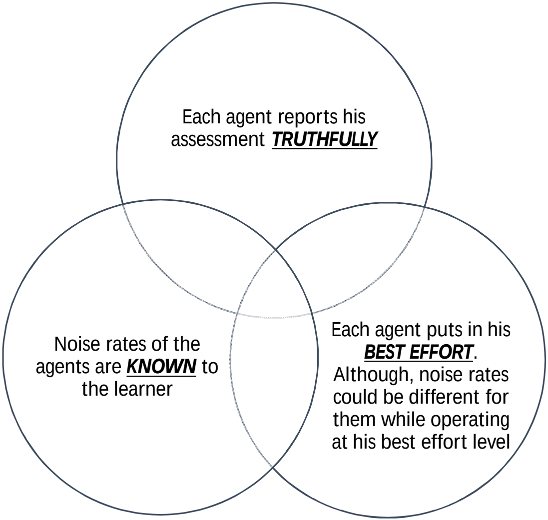
- Learner can not compute the PAC annotation plan because $\psi(\eta_i)$ is required for this: $\log(N/\delta) \leq \sum_{i=1}^{n} \psi(\eta_i) m_i$

Options Available with Learner

- Estimation
  - Overestimation $\Rightarrow$ Excess examples procured by NOAR $\Rightarrow$ Higher COST
  - Underestimation $\Rightarrow$ **Pr**(ϵ-bad hypothesis gets picked by NOAR) $> \delta$
- Elicitation
  - Invite annotators to report (bid) their costs (equivalently, noise rates)
  - Setup an auction to decide the work (contract) size and payment for annotators
  - Challenge: If annotators misreport noise rates, we are *back to square one!!*

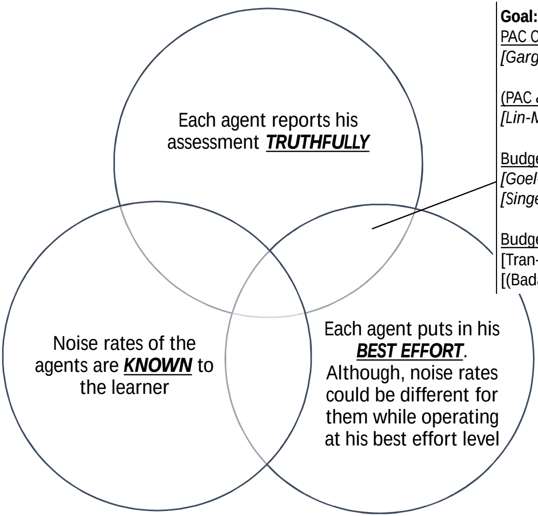
Goal: Design a Truthful & Cost Optimal Auction for PAC Learning via MDA.

Prior Work



Payment Mechanisms

Prior Work



Goal: Whom to hire?
PAC Constraints + Solicit Bids:
[Garg-Bhattacharya-Sudararajan-Shevade-12],

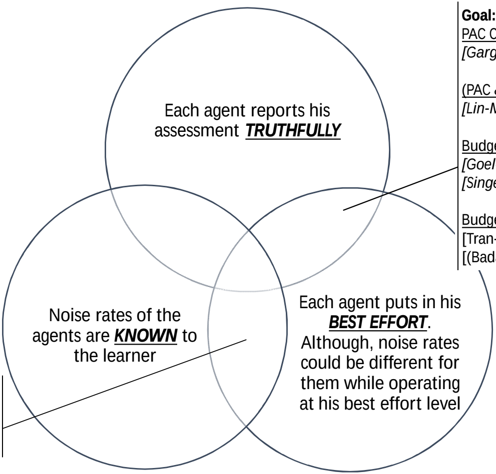
(PAC & Budget) Constraint + Same Noise
[Lin-Mausam-Weld-14]

Budget Constraint + Online+ Solicit Bids:
[Goel-Nikzad-Singla-14], [Singla-Krause-13], [Singer-Mittal-11],

Budget Constraint+ No Bidding+ MAB:
[Tran-Thanh-Stein-Rogers-Jennings-12],
[(Badanidiyuru-Kleinberg-Singer-12]

Payment Mechanisms

Prior Work



Goal: Whom to hire?
PAC Constraints + Solicit Bids:
[Garg-Bhattacharya-Sudararajan-Shevade-12],

(PAC & Budget) Constraint + Same Noise
[Lin-Mausam-Weld-14]

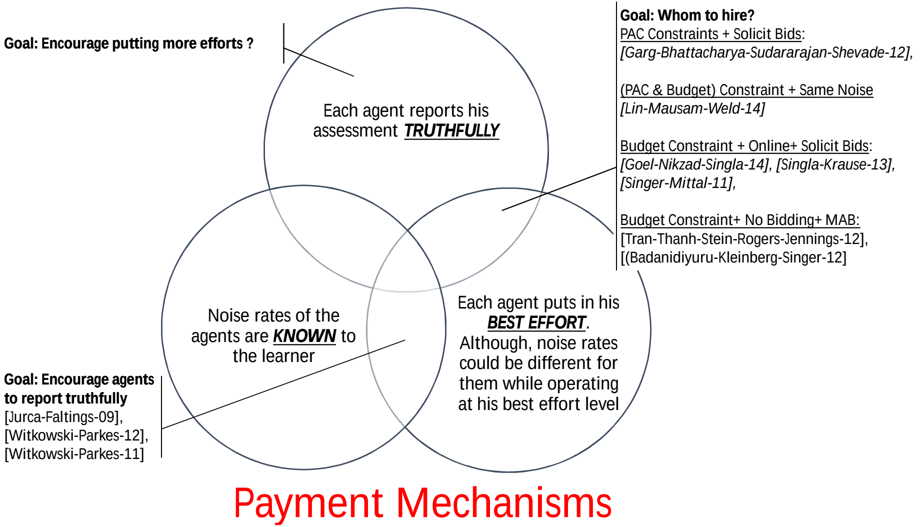
Budget Constraint + Online+ Solicit Bids:
[Goel-Nikzad-Singla-14], [Singla-Krause-13], [Singer-Mittal-11],

Budget Constraint+ No Bidding+ MAB:
[Tran-Thanh-Stein-Rogers-Jennings-12],
[(Badanidiyuru-Kleinberg-Singer-12]

Goal: Encourage agents to report truthfully
[Jurca-Faltings-09],
[Witkowski-Parkes-12],
[Witkowski-Parkes-11]

Payment Mechanisms

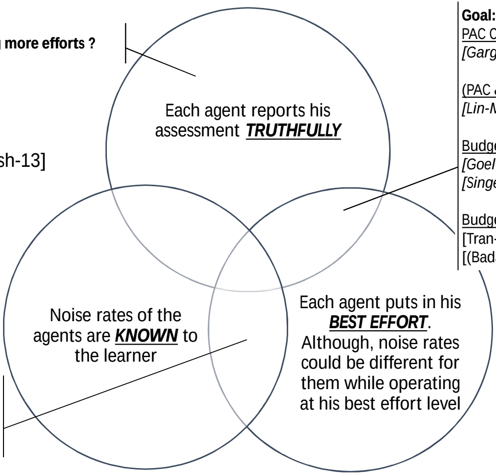
Prior Work


Goal: Encourage putting more efforts ?
Each agent reports his assessment TRUTHFULLY
Goal: Whom to hire?
PAC Constraints + Solicit Bids:
[Garg-Bhattacharya-Sudararajan-Shevade-12],
(PAC & Budget) Constraint + Same Noise
[Lin-Mausam-Weld-14]
Budget Constraint + Online+ Solicit Bids:
[Goel-Nikzad-Singla-14], [Singla-Krause-13],
[Singer-Mittal-11],
Budget Constraint+ No Bidding+ MAB:
[Tran-Thanh-Stein-Rogers-Jennings-12],
[(Badanidiyuru-Kleinberg-Singer-12]
Noise rates of the agents are KNOWN to the learner
Each agent puts in his BEST EFFORT. Although, noise rates could be different for them while operating at his best effort level
Goal: Encourage agents to report truthfully
[Jurca-Faltings-09],
[Witkowski-Parkes-12],
[Witkowski-Parkes-11]
Payment Mechanisms

Prior Work

Goal: Encourage putting more efforts ?

[Dasgupta-Ghosh-13]



Goal: Whom to hire?
PAC Constraints + Solicit Bids:
[Garg-Bhattacharya-Sudararajan-Shevade-12],

(PAC & Budget) Constraint + Same Noise
[Lin-Mausam-Weld-14]

Budget Constraint + Online+ Solicit Bids:
[Goel-Nikzad-Singla-14], [Singla-Krause-13], [Singer-Mittal-11],

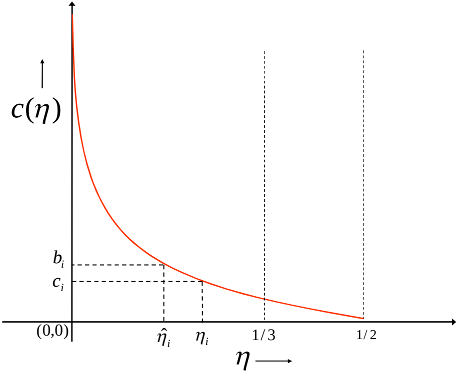
Budget Constraint+ No Bidding+ MAB:
[Tran-Thanh-Stein-Rogers-Jennings-12], [(Badanidiyuru-Kleinberg-Singer-12]

Goal: Encourage agents to report truthfully
[Jurca-Faltings-09], [Witkowski-Parkes-12], [Witkowski-Parkes-11]

Payment Mechanisms

Auction Framework for Incomplete Info Setting

- Bids
  - Annotator i bids b_i (could be different than his true cost c_i)
  - Bids are translated into equivalent noise rates: $\hat{\eta}_i = c^{-1}(b_i) \in I_i = [0, 1/3]$
  - Let $I = I_1 \times I_2 \ldots \times I_n$
  - The bid vector is given by $\hat{\eta} = (\hat{\eta}_1, \hat{\eta}_2, \ldots, \hat{\eta}_n) \in I$



Auction Framework for Incomplete Info Setting

- Task Allocation Mechanism $a(\cdot)$
  - Learner uses an allocation rule $a : I \mapsto \mathbb{N}_0^n$ to award the contracts
- Payment Mechanism $p(\cdot)$
  - Learner uses a payment rule $p : I \mapsto \mathbb{R}^n$ to pay the annotators
- Mechanism $\mathcal{M}$
  - A pair of allocation and payment mechanisms is called mechanism $\mathcal{M} = (a, p)$
- Utilities
  - Annotator i accumulates following utility when bid vector is $\hat{\eta}$

$$u_i(\hat{\eta}; \eta_i) = p_i(\hat{\eta}) - a_i(\hat{\eta})c(\eta_i)$$

  - To compute this utility, annotator i must know the bids of others

Common Prior Assumption and Expected Utility

Assumptions (IPV Model):

- Noise rate η_i gets assigned via an independent random draw from interval $[0, 1/3]$
- $\phi_i(\cdot)$ and $\Phi_i(\cdot)$ denote the corresponding prior density and CDF respectively
- The joint prior $\left(\phi(\cdot) = \prod_{i=1}^{n} \phi_i(\cdot)\right)$ is a common knowledge

- Expected Allocation Rule $\alpha_i(\cdot)$

$$\alpha_i(\hat{\eta}_i) = \int_{I_{-i}} a_i(\hat{\eta}_i, \hat{\eta}_{-i})\phi_{-i}(\hat{\eta}_{-i})d\hat{\eta}_{-i}$$

- Expected Payment Rule $\pi_i(\cdot)$

$$\pi_i(\hat{\eta}_i) = \int_{I_{-i}} p_i(\hat{\eta}_i, \hat{\eta}_{-i})\phi_{-i}(\hat{\eta}_{-i})d\hat{\eta}_{-i}$$

- Expected Utility $U_i(\cdot)$

$$U_i(\hat{\eta}_i; \eta_i) = \pi_i(\hat{\eta}_i) - \alpha_i(\hat{\eta}_i)c(\eta_i)$$

Optimal Auction Design for Incomplete Info Setting

$$\underset{a(\cdot),p(\cdot)}{\text{Minimize}} \quad \Pi(a,p) = \sum_{i=1}^{n} \int_0^{1/3} \pi_i(t_i)\phi_i(t_i)dt_i \text{ (Procurement Cost)}$$

$$\text{Subject to} \quad \log(N/\delta) \leq \sum_i a_i(\eta_i, \eta_{-i})\psi(\eta_i) \; \forall(\eta_i, \eta_{-i}) \in I \text{ (PAC Constraint)}$$

$$(a,p) \text{ satisfies } BIC \text{ (BIC Constraint)}$$

$$\pi_i(\eta_i) \geq \alpha_i(\eta_i)c(\eta_i) \; \forall \eta_i \in I_i, \forall i \text{ (IR Constraint)}$$

A Mechanism is said to be

- **Bayesian Incentive Compatible (BIC)** if for every annotator i, $U_i(\cdot)$ is maximized when $\hat{\eta}_i = \eta_i$, i.e., $U_i(\eta_i;\eta_i) \geq U_i(\hat{\eta}_i;\eta_i) \; \forall \hat{\eta}_i \in I_i$.
- **Individually Rational (IR)** if no annotator loses (in expected sense) anything by reporting true noise rates, i.e., $\pi_i(\eta_i) - \alpha_i(\eta_i)c(\eta_i) \geq 0 \; \forall \; \eta_i \in I_i$.

BIC Characterization: *Myerson's Theorem*

An allocation rule a is said to be Non-decreasing in Expectation (NDE) if we have $\alpha_i(\eta_i) \geq \alpha_i(\hat{\eta}_i)\ \forall \eta_i > \hat{\eta}_i$

Theorem (Myerson 1981)

Mechanism $\mathcal{M} = (a, p)$ is a BIC mechanism iff

1. *Allocation rule $a(\cdot)$ is NDE, and*
2. *Expected payment rule satisfies:*

$$
\begin{aligned}
U_i(\eta_i) &= U_i(0) - \int_0^{\eta_i} \alpha_i(t_i)c'(t_i)dt_i \\
\Rightarrow \pi_i(\eta_i) &= \alpha_i(\eta_i)c(\eta_i) + U_i(0) - \int_0^{\eta_i} \alpha_i(t_i)c'(t_i)dt_i
\end{aligned}
$$

Roger Myerson (Winner of 2007 Nobel Prize in Economics)

[1] R. B. Myerson. Optimal Auction Design. Math. Operations Res., 6(1):58 -73, Feb. 1981.

Back to Optimal Auction Design

$$\underset{a(\cdot),p(\cdot)}{\text{Minimize}} \quad \Pi(a,p) = \sum_{i=1}^{n} \int_0^{1/3} \pi_i(t_i)\phi_i(t_i)dt_i \text{ (Procurement Cost)}$$

$$\text{Subject to} \quad \log(N/\delta) \le \sum_i a_i(\eta_i, \eta_{-i})\psi(\eta_i) \ \forall(\eta_i, \eta_{-i}) \in I \text{ (PAC Constraint)}$$

$$\alpha_i(\cdot) \text{ is non-decreasing (BIC Constraint 1)}$$

$$\pi_i(\eta_i) = \alpha_i(\eta_i)c(\eta_i) + U_i(0) - \int_0^{\eta_i} \alpha_i(t_i)c'(t_i)dt_i \text{ (BIC Constraint 2)}$$

$$\pi_i(\eta_i) \ge \alpha_i(\eta_i)c(\eta_i) \ \forall \eta_i \in I_i, \forall i \text{ (IR Constraint)}$$

Back to Optimal Auction Design

$$\begin{aligned}
&\underset{a(\cdot),p(\cdot)}{\text{Minimize}} && \Pi(a,p) = \sum_{i=1}^{n} \int_0^{1/3} \pi_i(t_i)\phi_i(t_i)dt_i \text{ (Procurement Cost)} \\
&\text{Subject to} && \log(N/\delta) \leq \sum_i a_i(\eta_i,\eta_{-i})\psi(\eta_i)\ \forall(\eta_i,\eta_{-i}) \in I \text{ (PAC Constraint)} \\
& && \alpha_i(\cdot) \text{ is non-decreasing (BIC Constraint 1)} \\
& && \pi_i(\eta_i) = \alpha_i(\eta_i)c(\eta_i) + U_i(0) - \int_0^{\eta_i} \alpha_i(t_i)c'(t_i)dt_i \text{ (BIC Constraint 2)} \\
& && \pi_i(\eta_i) \geq \alpha_i(\eta_i)c(\eta_i)\ \forall \eta_i \in I_i, \forall i \text{ (IR Constraint)}
\end{aligned}$$

Insights:

- If (BIC Constraint 2) is satisfied then (IR Constraint) is satisfied iff $U_i(0) \geq 0$

Back to Optimal Auction Design

$$\underset{a(\cdot),p(\cdot)}{\text{Minimize}} \quad \Pi(a,p) = \sum_{i=1}^{n} \int_0^{1/3} \pi_i(t_i)\phi_i(t_i)dt_i \text{ (Procurement Cost)}$$

$$\text{Subject to} \quad \log(N/\delta) \leq \sum_i a_i(\eta_i, \eta_{-i})\psi(\eta_i) \ \forall(\eta_i, \eta_{-i}) \in I \text{ (PAC Constraint)}$$

$$\alpha_i(\cdot) \text{ is non-decreasing (BIC Constraint 1)}$$

$$\pi_i(\eta_i) = \alpha_i(\eta_i)c(\eta_i) + U_i(0) - \int_0^{\eta_i} \alpha_i(t_i)c'(t_i)dt_i \text{ (BIC Constraint 2)}$$

$$\pi_i(\eta_i) \geq \alpha_i(\eta_i)c(\eta_i) \ \forall \eta_i \in I_i, \forall i \text{ (IR Constraint)}$$

Insights:

- If (BIC Constraint 2) is satisfied then (IR Constraint) is satisfied iff $U_i(0) \geq 0$
- Because our goal is to minimize the objective function, we must have $U_i(0) = 0$

Back to Optimal Auction Design

$$\underset{a(\cdot),p(\cdot)}{\text{Minimize}} \quad \Pi(a,p) = \sum_{i=1}^{n} \int_0^{1/3} \pi_i(t_i)\phi_i(t_i)dt_i \text{ (Procurement Cost)}$$

$$\text{Subject to} \quad \log(N/\delta) \leq \sum_i a_i(\eta_i,\eta_{-i})\psi(\eta_i) \ \forall(\eta_i,\eta_{-i}) \in I \text{ (PAC Constraint)}$$

$$\alpha_i(\cdot) \text{ is non-decreasing (BIC Constraint 1)}$$

$$\pi_i(\eta_i) = \alpha_i(\eta_i)c(\eta_i) + U_i(0) - \int_0^{\eta_i} \alpha_i(t_i)c'(t_i)dt_i \text{ (BIC Constraint 2)}$$

$$\pi_i(\eta_i) \geq \alpha_i(\eta_i)c(\eta_i) \ \forall\eta_i \in I_i, \forall i \text{ (IR Constraint)}$$

Insights:

- If (BIC Constraint 2) is satisfied then (IR Constraint) is satisfied iff $U_i(0) \geq 0$
- Because our goal is to minimize the objective function, we must have $U_i(0) = 0$
- Using (BIC Constraint 2), objective becomes $\Pi(a,p) = \int_I \left(\sum_{i=1}^{n} v_i(x_i)a_i(x)\right)\phi(x)dx$
- $v_i(\eta_i) := c(\eta_i) - \frac{1-\Phi_i(\eta_i)}{\phi_i(\eta_i)}c'(\eta_i)$ is virtual cost function (Note $v_i(\eta_i) \geq c(\eta_i)$)

Reduced Problem

Overall Problem

$$\underset{a(\cdot),p(\cdot)}{\text{Minimize}} \quad \Pi(a,p) = \int_I \left(\sum_{i=1}^n v_i(x_i)a_i(x)\right)\phi(x)dx \text{ (Procurement Cost)}$$

$$\text{Subject to} \quad \log(N/\delta) \leq \sum_i a_i(\eta_i,\eta_{-i})\psi(\eta_i)\ \forall(\eta_i,\eta_{-i}) \in I \text{ (PAC Constraint)}$$

$$\alpha_i(\cdot) \text{ is non-decreasing (BIC Constraint 1)}$$

Insights:

- Keep aside (BIC Constraint 1) for the moment
- It suffices to solve following problem for every possible profile η

Instance Specific ILP

$$\underset{a_1(\eta),\ldots,a_n(\eta)}{\text{Minimize}} \quad \sum_{i=1}^n v_i(\eta_i)a_i(\eta) \text{(Procurement Cost for profile } \eta)$$

$$\text{Subject to} \quad \log(N/\delta) \leq \sum_i \psi(\eta_i)a_i(\eta)\ \forall(\eta_i,\eta_{-i}) \in I \text{ (PAC Constraint)}$$

$$a_i(\eta) \in \mathbb{N}_0\ \forall i$$

Solution Via Instance Specific ILP

- Instance specific ILP is similar to Primal Problem in complete info setting (replace $c(\eta_i)$ with $v_i(\eta_i)$)
- Instance specific ILP can be relaxed and solved approximately just like NOAR

Definition (Minimum Allocation Rule)

Let i^* be the annotator having minimum value for cost-per-quality given by $v_i(\eta_i)/\psi(\eta_i)$. The learner should buy $\lceil \log(N/\delta)/\psi(\eta_{i^*}) \rceil$ number of examples from such an annotator.

Theorem

Let COST be the total cost of purchase incurred by the Minimum Allocation Rule. Let OPT be the optimal procurement cost. Then,

$$OPT \leq COST \leq OPT + c(\eta_{i^*}) \leq OPT(1 + 1/m_0)$$

where $m_0 = \log[1 - \epsilon]^{-1}$

What About (BIC Constraint 1) ?

Regularity Condition: $v_i(\cdot)/\psi(\cdot)$ is a non-increasing function.

If Regularity Condition is satisfied, then under the minimum allocation rule

- As η_i increases, the annotator i remains the winner if he/she is already the winner (with an increased contract size) or becomes the winner
- The allocation rule satisfies ND property (hence, NDE)
- The payment of annotator i is given by

$$p_i(\eta_i, \eta_{-i}) = a_i(\eta_i, \eta_{-i})c(\eta_i) - \int_0^{\eta_i} a_i(t_i, \eta_{-i})c'(t_i)dt_i$$

- Winning annotator gets positive payment and others get zero payment

Near Optimal Auction Mechanism for PAC Learning

Under regularity condition of $v_i(\cdot)/\psi(\cdot)$ being a non-increasing function of η_i

$$a_i(\eta) = \begin{cases} \lceil \log(N/\delta)/\psi(\eta_i) \rceil & : \text{ if } \frac{v_i(\eta_i)}{\psi(\eta_i)} \le \frac{v_j(\eta_j)}{\psi(\eta_j)} \; \forall j \neq i \\ 0 & : \text{ otherwise} \end{cases}$$

$$p_i(\eta) = \begin{cases} \left\lceil \frac{\log(N/\delta)}{\psi(\eta_i)} \right\rceil c(q_i(\eta_{-i})) & : \text{ for winner} \\ 0 & : \text{ otherwise} \end{cases}$$

$$q_i(\eta_{-i}) = \inf \left\{ \hat{\eta}_i \mid \frac{v_i(\hat{\eta}_i)}{\psi(\hat{\eta}_i)} \le \frac{v_j(\eta_j)}{\psi(\eta_j)} \; \forall j \neq i \right\}$$

$=$ smallest bid value sufficient to win the contract for annotator i

Theorem

Suppose Regularity Condition holds. Then, above mechanism is an approximate optimal mechanism satisfying BIC, IR, and PAC constraints. The approximation guarantee of this mechanism is given by $ALG \le OPT + v_{i^*}(\eta_{i^*}) \le OPT(1 + 1/m_0)$.

Conclusions

- Analyzed the PAC learning model for noisy data from multiple annotators
- Analyzed complete and incomplete information scenarios
- Essentially, we identify the annotator whose (cost/quality) ratio is the least
- Surprisingly, greedily buying all the examples from such an annotator is near optimal

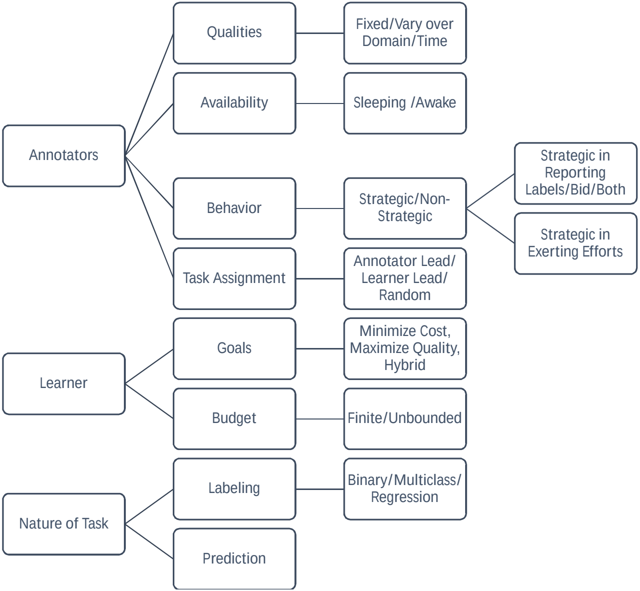
Future Extensions

- What if the cost function $c(\cdot)$ is not a common knowledge?
- What if the cost function $c(\cdot)$ is different for different annotators?
- Annotators having a capacity constraint and/or learner having a budget constraint
- Work with general hypothesis class (e.g. linear models of classification)
- Other learning tasks - *multiclass/multilabel classification*, *regression*
- What about *sequentially deciding the tasks assignments*?

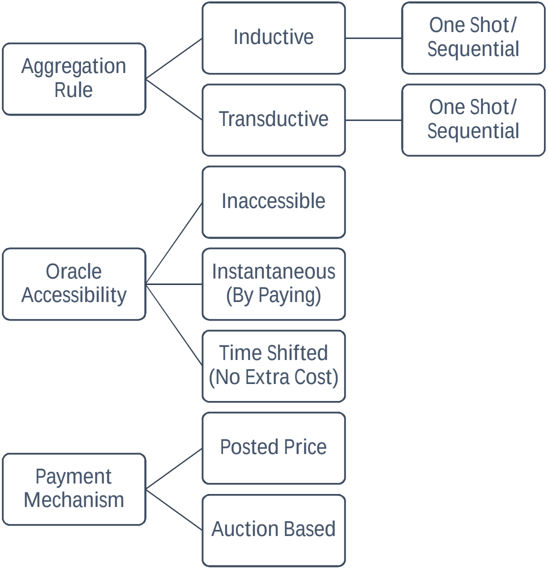
Thank You!!

Backup Slides

Aspects of Crowdsorcing Systems



Aspects of Crowdsorcing Systems



Proof Sketch

Events

- $E_1(h, m_1, \ldots, m_n)$: The empirical error of a given hypothesis $h \in \mathscr{C}$ is no more than the empirical error of the true hypothesis c_t, i.e. $L_e(h) \leq L_e(c_t)$.
- $E_2(h, m_1, \ldots, m_n)$: The empirical error of a given hypothesis $h \in \mathscr{C}$ is the minimum across all hypotheses in the class $\mathscr{C}$, i.e. $L_e(h) \leq L_e(h') \; \forall h' \in \mathscr{C}$.
- $E_3(h, m_1, \ldots, m_n)$: MDA outputs a given hypothesis h.
- $E_4(\epsilon, m_1, \ldots, m_n)$: MDA outputs an ϵ-bad hypothesis.

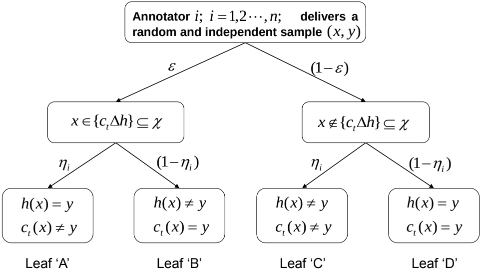
Observations

- $E_3(h, m_1, \ldots, m_n) \subseteq E_2(h, m_1, \ldots, m_n) \subseteq E_1(h, m_1, \ldots, m_n)$
- $\mathbf{Pr}^{(m_1,\ldots,m_n)}[E_4(\epsilon)] \leq (N-1) \times \left[\max_{h \in \mathscr{C},\, h \text{ is } \epsilon\text{-bad}} \mathbf{Pr}^{(m_1,\ldots,m_n)}[E_1(h)] \right]$
- If annotation plan $(m_1, \ldots, m_n)$ satisfies the following condition, then MDA will satisfy PAC bound.

$$\left[\max_{h \text{ is } \epsilon\text{-bad}} \mathbf{Pr}^{(m_1,\ldots,m_n)}[E_1(h)] \right] \leq \delta / N \qquad (2)$$

Proof Sketch

Probability of an ϵ-bad hypothesis h having lower empirical error than c_t



$\mathbf{Pr}^{(m_1,\dots,m_n)}[L_e(h) \leq L_e(c_t)] = \mathbf{Pr}\{\#\text{ samples under leaf A} \geq \#\text{ samples under leaf B}\}$

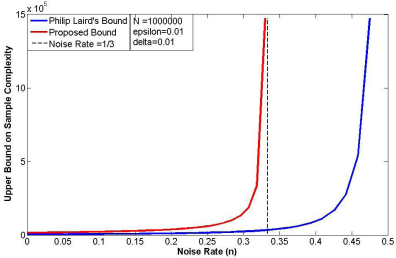
Special Case: Single Annotator

When $\eta = 0$

- Easy to show that sample complexity m_0 satisfies $m_0 \leq \log(N/\delta)/\log[1-\epsilon]^{-1}$
- The range of η_i in previous theorem can be extended to include $\eta_i = 0$ by having $\psi(0) = \log[1-\epsilon]^{-1}$

When $\eta = 1/3$

- Angluin and Laird proposed following bound for single annotator, for $0 \leq \eta < 1/2$ $\psi(\eta_i) = \log\left[1 - \epsilon\left(1 - \exp\left(-(1-2\eta_i)^2/2\right)\right)\right]^{-1}$
- The range of η_i in previous theorem can be extended to include $\eta_i = 1/3$ by having $\psi(1/3) = \log[1-\epsilon(1-\exp(-1/18))]^{-1}$



[1] Dana Angluin and Philip Laird. Learning from noisy examples. Machine Learning, 2(4):343-370, 1988.

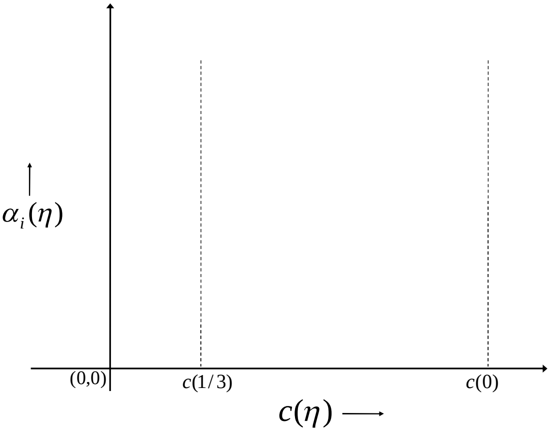
Understanding Myerson's Theorem

$$
\begin{aligned}
\pi_i(\eta_i) &= \alpha_i(\eta_i)c(\eta_i) + U_i(0) + \int_{\eta_i}^{0} \alpha_i(t_i)c'(t_i)dt_i \\
\Rightarrow \pi_i(\eta_i) &= \alpha_i(\eta_i)c(\eta_i) + \pi_i(0) - \alpha_i(0)c(0) + \int_{\eta_i}^{0} \alpha_i(t_i)d\left[c(t_i)\right]
\end{aligned}
$$

Understanding Myerson's Theorem

$$\pi_i(\eta_i) = \alpha_i(\eta_i)c(\eta_i) + U_i(0) + \int_{\eta_i}^{0} \alpha_i(t_i)c'(t_i)dt_i$$

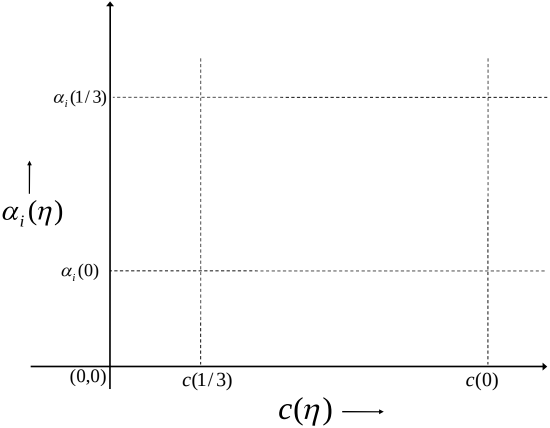
$$\Rightarrow \pi_i(\eta_i) = \alpha_i(\eta_i)c(\eta_i) + \pi_i(0) - \alpha_i(0)c(0) + \int_{\eta_i}^{0} \alpha_i(t_i)d\left[c(t_i)\right]$$



Understanding Myerson's Theorem

$$\pi_i(\eta_i) = \alpha_i(\eta_i)c(\eta_i) + U_i(0) + \int_{\eta_i}^{0} \alpha_i(t_i)c'(t_i)dt_i$$

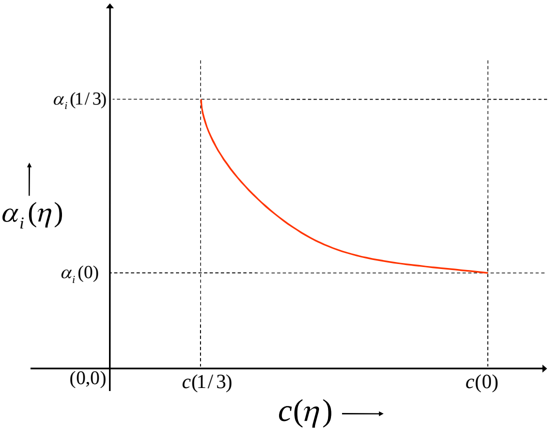
$$\Rightarrow \pi_i(\eta_i) = \alpha_i(\eta_i)c(\eta_i) + \pi_i(0) - \alpha_i(0)c(0) + \int_{\eta_i}^{0} \alpha_i(t_i)d\left[c(t_i)\right]$$



Understanding Myerson's Theorem

$$\pi_i(\eta_i) = \alpha_i(\eta_i)c(\eta_i) + U_i(0) + \int_{\eta_i}^{0} \alpha_i(t_i)c'(t_i)dt_i$$

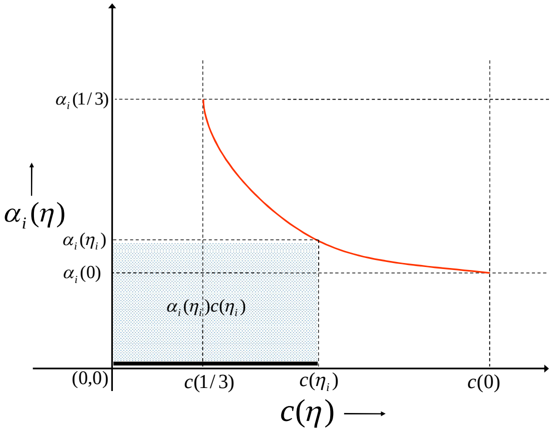
$$\Rightarrow \pi_i(\eta_i) = \alpha_i(\eta_i)c(\eta_i) + \pi_i(0) - \alpha_i(0)c(0) + \int_{\eta_i}^{0} \alpha_i(t_i)d\left[c(t_i)\right]$$



Understanding Myerson's Theorem

$$\pi_i(\eta_i) = \alpha_i(\eta_i)c(\eta_i) + U_i(0) + \int_{\eta_i}^{0} \alpha_i(t_i)c'(t_i)dt_i$$

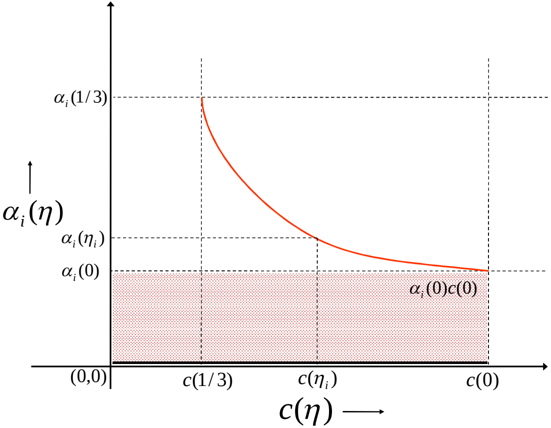
$$\Rightarrow \pi_i(\eta_i) = \alpha_i(\eta_i)c(\eta_i) + \pi_i(0) - \alpha_i(0)c(0) + \int_{\eta_i}^{0} \alpha_i(t_i)d\left[c(t_i)\right]$$



Understanding Myerson's Theorem

$$\pi_i(\eta_i) = \alpha_i(\eta_i)c(\eta_i) + U_i(0) + \int_{\eta_i}^{0} \alpha_i(t_i)c'(t_i)dt_i$$

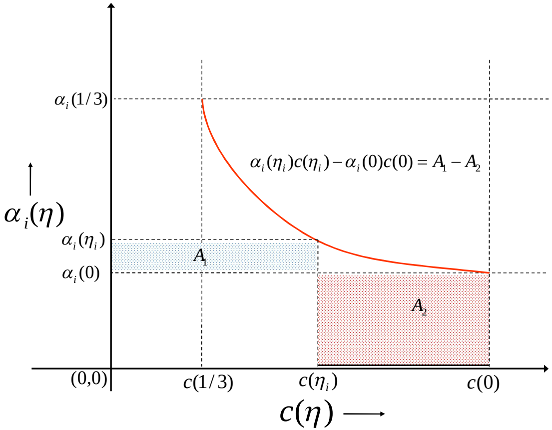
$$\Rightarrow \pi_i(\eta_i) = \alpha_i(\eta_i)c(\eta_i) + \pi_i(0) - \alpha_i(0)c(0) + \int_{\eta_i}^{0} \alpha_i(t_i)d\left[c(t_i)\right]$$



Understanding Myerson's Theorem

$$\pi_i(\eta_i) = \alpha_i(\eta_i)c(\eta_i) + U_i(0) + \int_{\eta_i}^{0} \alpha_i(t_i)c'(t_i)dt_i$$

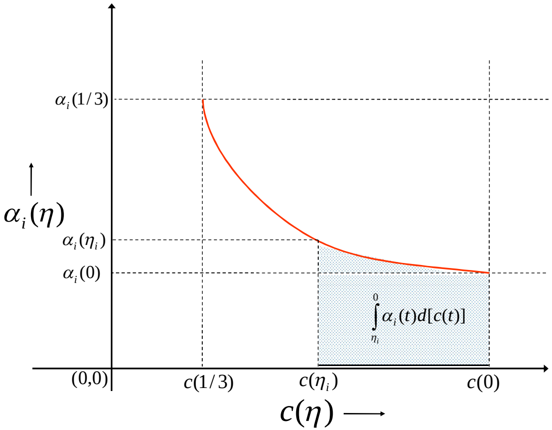
$$\Rightarrow \pi_i(\eta_i) = \alpha_i(\eta_i)c(\eta_i) + \pi_i(0) - \alpha_i(0)c(0) + \int_{\eta_i}^{0} \alpha_i(t_i)d\left[c(t_i)\right]$$



Understanding Myerson's Theorem

$$\pi_i(\eta_i) = \alpha_i(\eta_i)c(\eta_i) + U_i(0) + \int_{\eta_i}^{0} \alpha_i(t_i)c'(t_i)dt_i$$

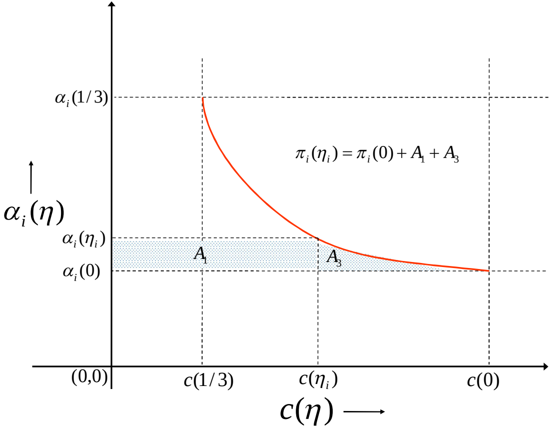
$$\Rightarrow \pi_i(\eta_i) = \alpha_i(\eta_i)c(\eta_i) + \pi_i(0) - \alpha_i(0)c(0) + \int_{\eta_i}^{0} \alpha_i(t_i)d\left[c(t_i)\right]$$



Understanding Myerson's Theorem

$$\pi_i(\eta_i) = \alpha_i(\eta_i)c(\eta_i) + U_i(0) + \int_{\eta_i}^{0} \alpha_i(t_i)c'(t_i)dt_i$$

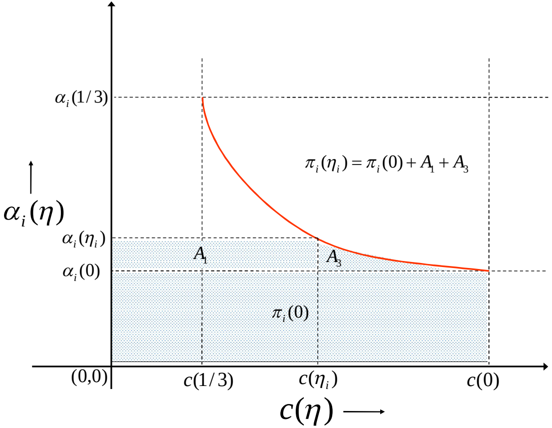
$$\Rightarrow \pi_i(\eta_i) = \alpha_i(\eta_i)c(\eta_i) + \pi_i(0) - \alpha_i(0)c(0) + \int_{\eta_i}^{0} \alpha_i(t_i)d\left[c(t_i)\right]$$



Understanding Myerson's Theorem

$$\pi_i(\eta_i) = \alpha_i(\eta_i)c(\eta_i) + U_i(0) + \int_{\eta_i}^{0} \alpha_i(t_i)c'(t_i)dt_i$$

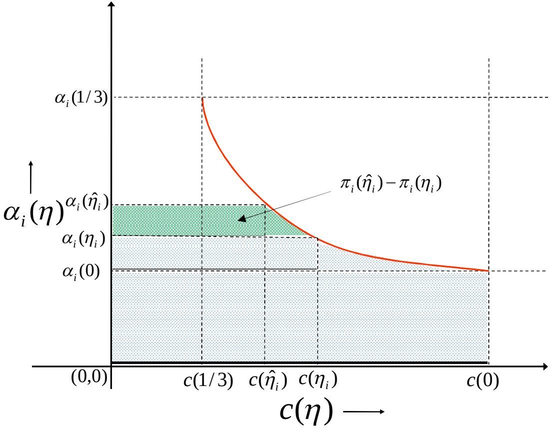
$$\Rightarrow \pi_i(\eta_i) = \alpha_i(\eta_i)c(\eta_i) + \pi_i(0) - \alpha_i(0)c(0) + \int_{\eta_i}^{0} \alpha_i(t_i)d\left[c(t_i)\right]$$



Understanding Myerson's Theorem

$$\pi_i(\eta_i) = \alpha_i(\eta_i)c(\eta_i) + U_i(0) + \int_{\eta_i}^{0} \alpha_i(t_i)c'(t_i)dt_i$$

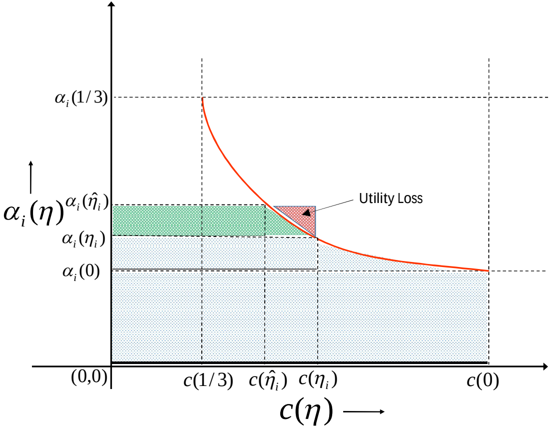
$$\Rightarrow \pi_i(\eta_i) = \alpha_i(\eta_i)c(\eta_i) + \pi_i(0) - \alpha_i(0)c(0) + \int_{\eta_i}^{0} \alpha_i(t_i)d\left[c(t_i)\right]$$



Understanding Myerson's Theorem

$$\pi_i(\eta_i) = \alpha_i(\eta_i)c(\eta_i) + U_i(0) + \int_{\eta_i}^{0} \alpha_i(t_i)c'(t_i)dt_i$$

$$\Rightarrow \pi_i(\eta_i) = \alpha_i(\eta_i)c(\eta_i) + \pi_i(0) - \alpha_i(0)c(0) + \int_{\eta_i}^{0} \alpha_i(t_i)d\left[c(t_i)\right]$$



Understanding Myerson's Theorem

$$\pi_i(\eta_i) = \alpha_i(\eta_i)c(\eta_i) + U_i(0) + \int_{\eta_i}^{0} \alpha_i(t_i)c'(t_i)dt_i$$

$$\Rightarrow \pi_i(\eta_i) = \alpha_i(\eta_i)c(\eta_i) + \pi_i(0) - \alpha_i(0)c(0) + \int_{\eta_i}^{0} \alpha_i(t_i)d\left[c(t_i)\right]$$