Machine Learning with Humans in the Loop

Chien-Ju (CJ) Ho

Wikipedia
The Free Encyclopedia

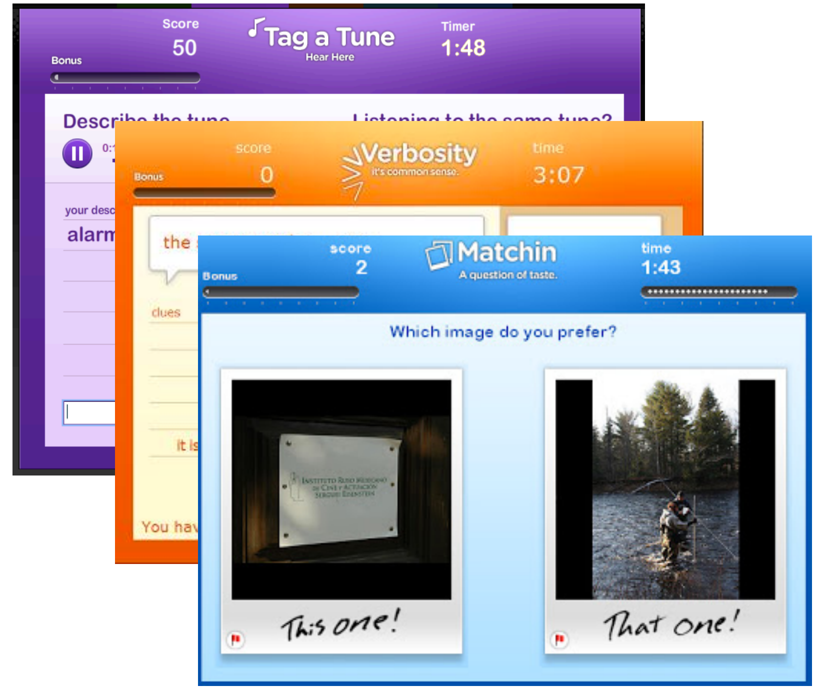
Washington University in St. Louis

From Wikipedia, the free encyclopedia

Washington University in St. Louis (also referred to as **WashU**, or **WUSTL**) is a private research university located in St. Louis, Missouri, United States. Founded in 1853, and named after George Washington, the university has students and faculty from all 50 U.S. states and more than 120 countries.[6] Twenty-five Nobel laureates have been affiliated with Washington University, nine having done the major part of their pioneering research at the university.[7] Washington University's undergraduate program is ranked 19th by *U.S. News & World Report*[8] and 11th by the Wall Street Journal in 2016.[9] The university is ranked 23rd in the world in 2016 by the Academic Ranking of World Universities.[10]

Washington University is made up of seven graduate and undergraduate schools that encompass a broad range of academic fields.[11] To prevent confusion over its location, the Board of Trustees added the phrase "in St. Louis" in 1976.[12]

gwap



fold.it
Solve Puzzles for Science

HEALTHY LIVING 09/19/2011 03:37 pm ET | **Updated** Nov 19, 2011

Gamers Decode AIDS Protein That Stumped Researchers For 15 Years In Just 3 Weeks

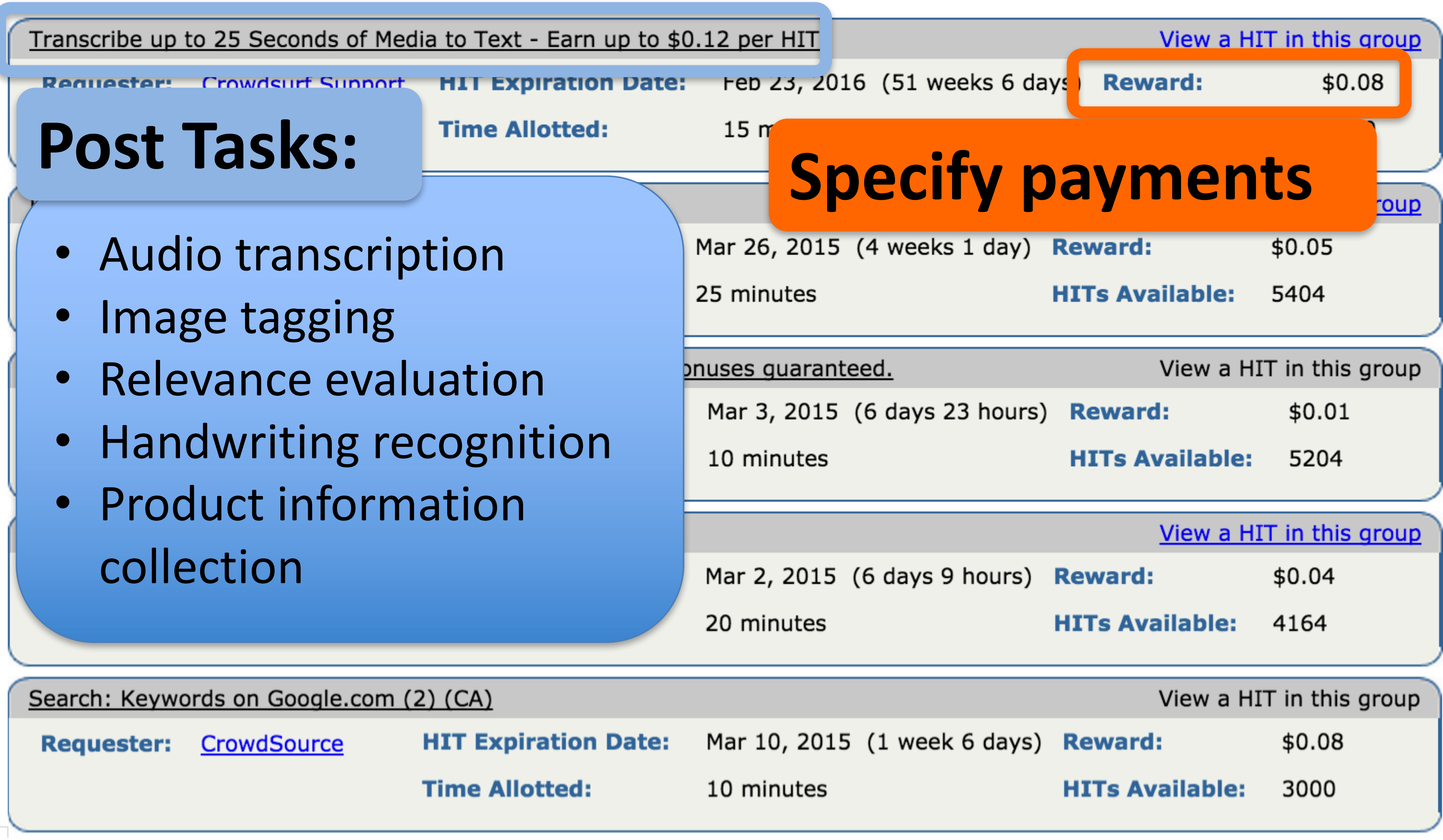
amazon mechanical turk™

Artificial Artificial Intelligence

Post Tasks:

- Audio transcription
- Image tagging
- Relevance evaluation
- Handwriting recognition
- Product information collection

Specify payments



Google

twitter

yelp

Quora

facebook

YAHOO! ANSWERS

gwap

duolingo

stackoverflow

reCAPTCHA™

PredictWise

amazon mechanical turk™
Artificial Artificial Intelligence

Predictious

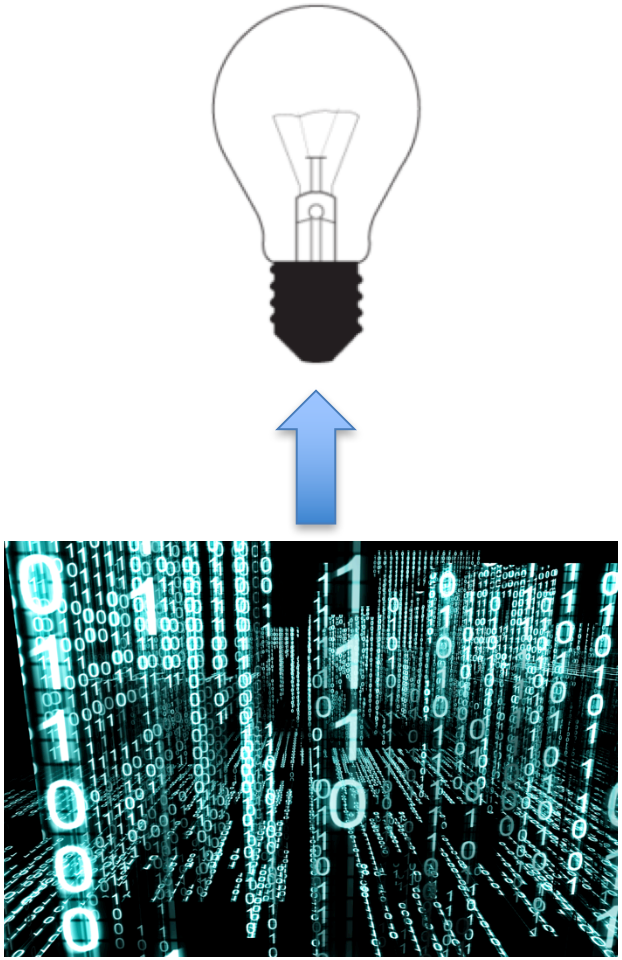
99 designs®

kaggle™

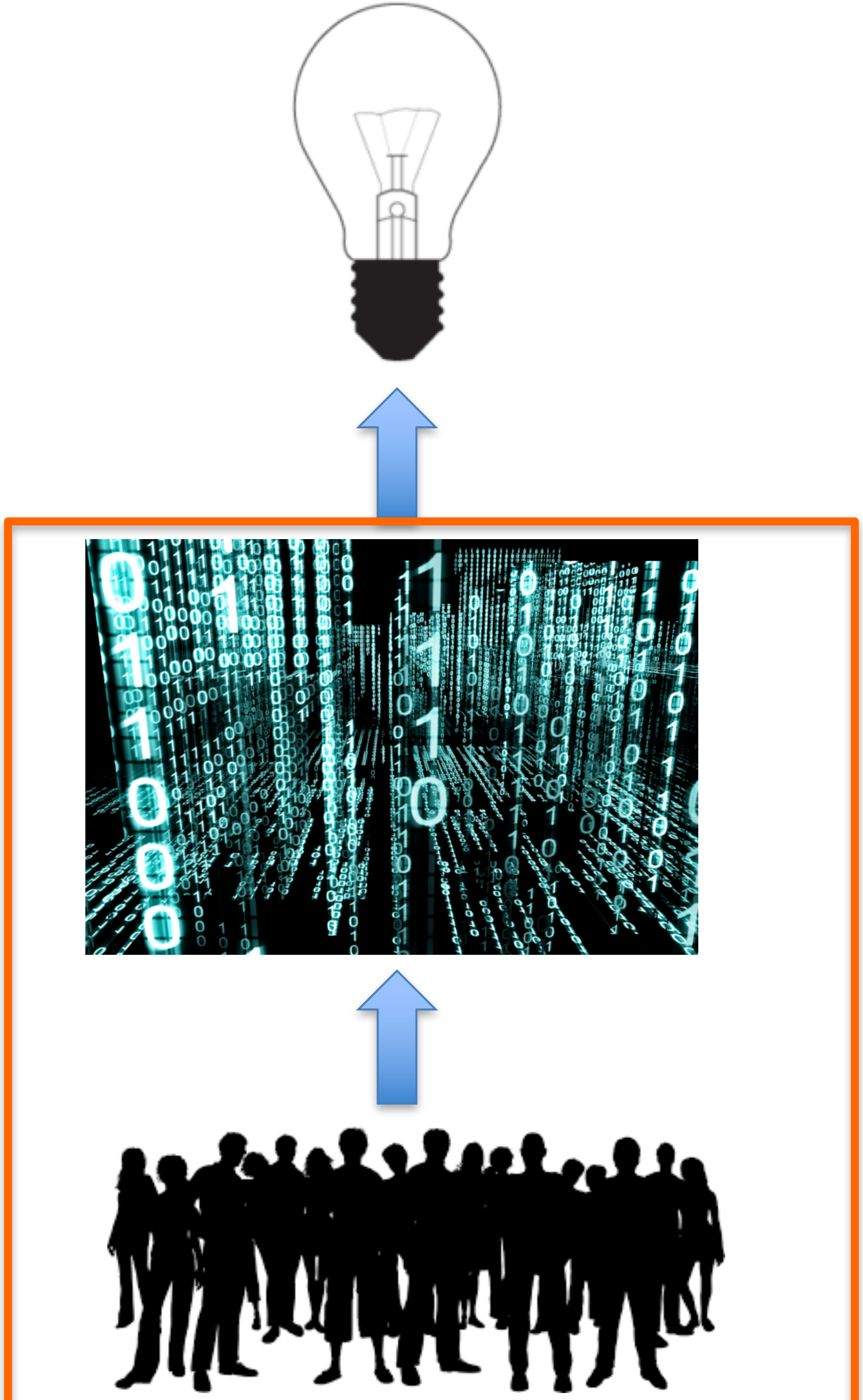
up

[topcoder]™

CrowdFlower

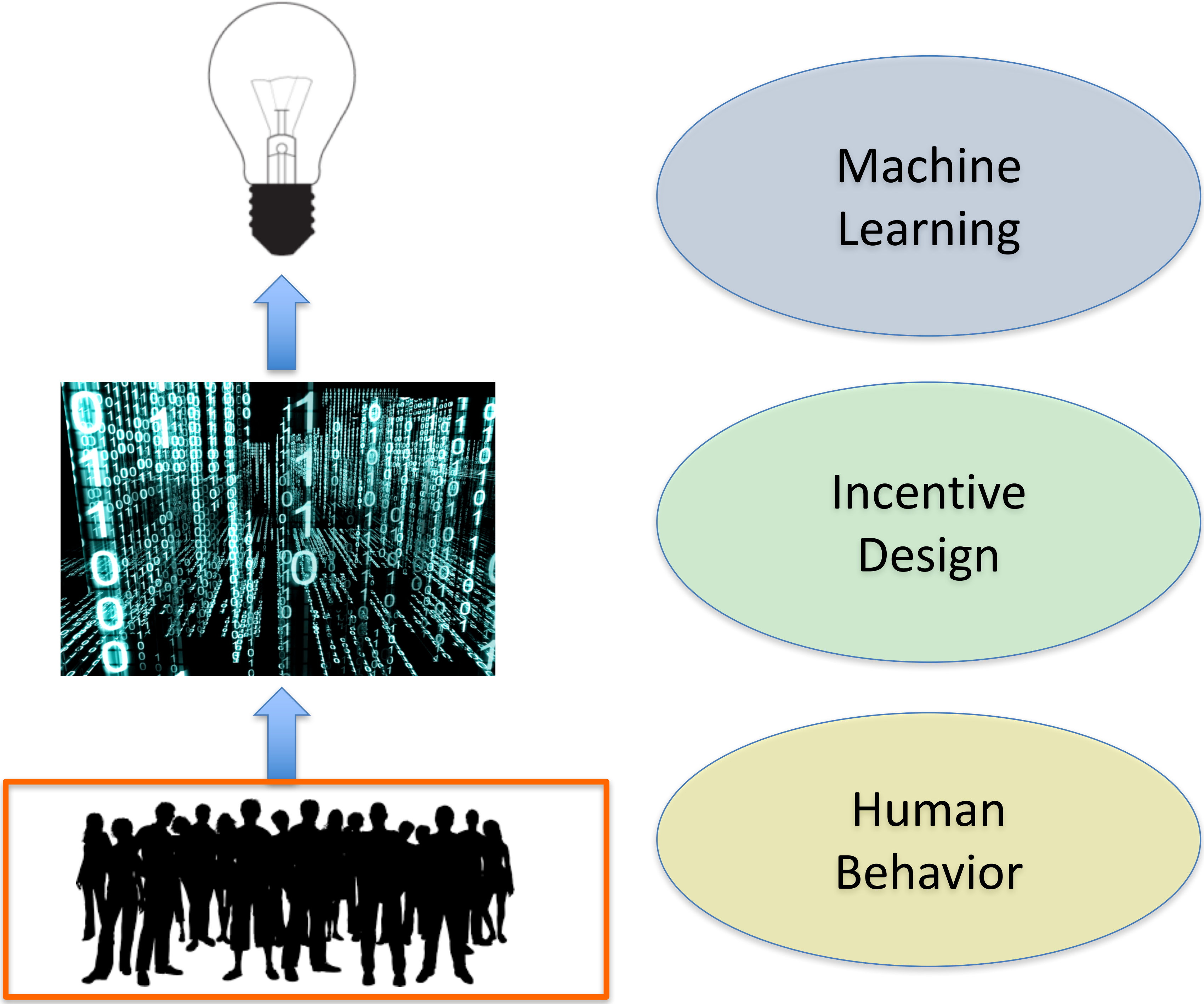


Machine Learning



Machine Learning

Incentive Design


Machine Learning
Incentive Design
Human Behavior

Outline

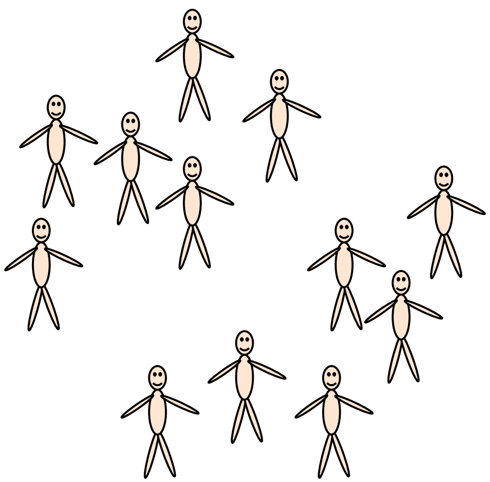
- A sample of my past research
  - Active buying data for machine learning
- Ongoing research directions
  - Bandit learning with human feedback
  - Leveraging communications in crowdsourcing
  - Online resource allocations (with discussions on fairness and equity)
- Summary and discussion

Active Buying Data for Machine Learning

Joint work with
Jake Abernethy, Yiling Chen, and Bo Waggoner

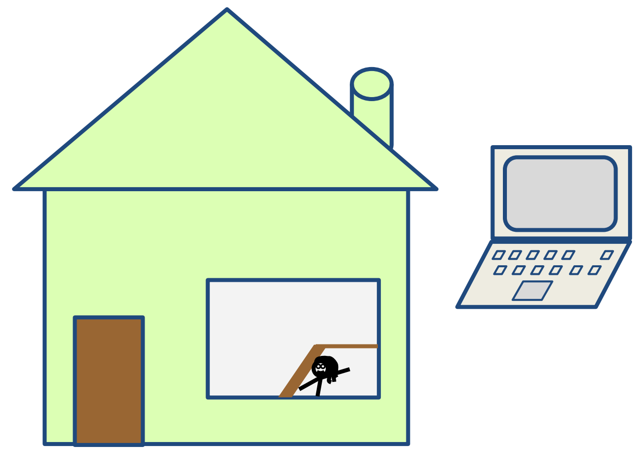
Learning via Buying Data from Humans

data-holders



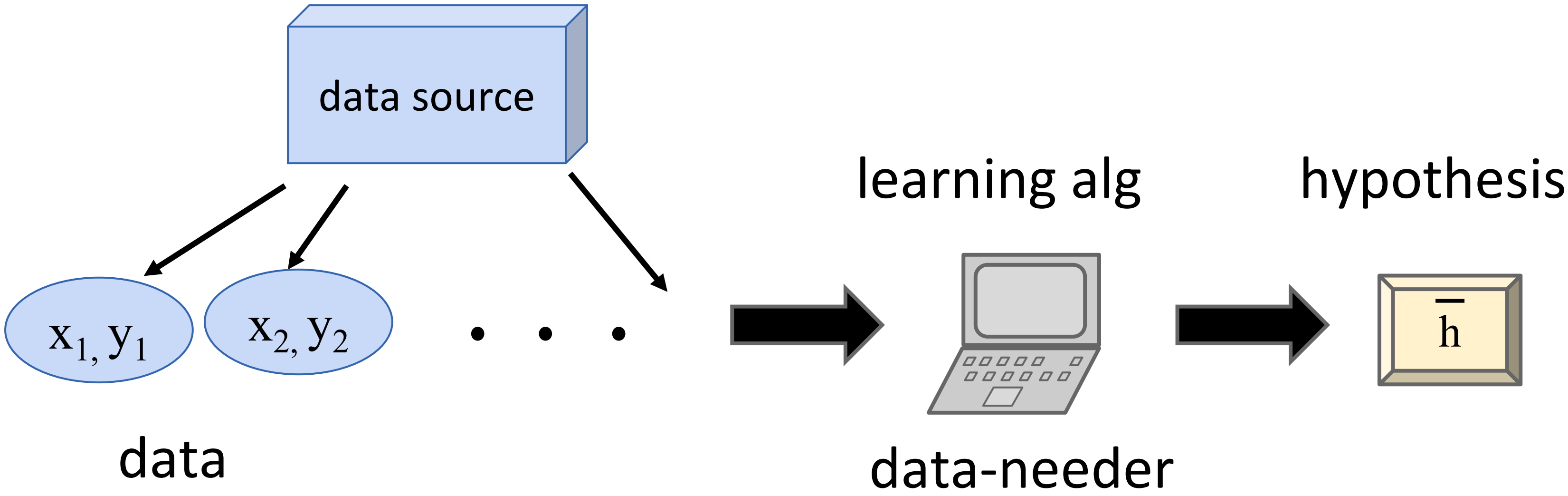
ex: medical data

data-needers

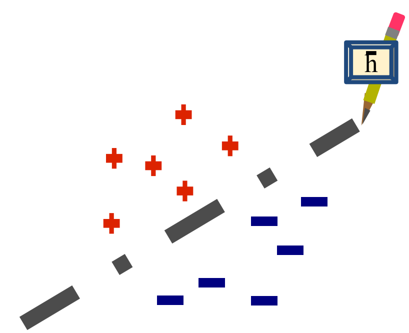


ex: pharmaceutical co.

(Traditional) Learning Problems

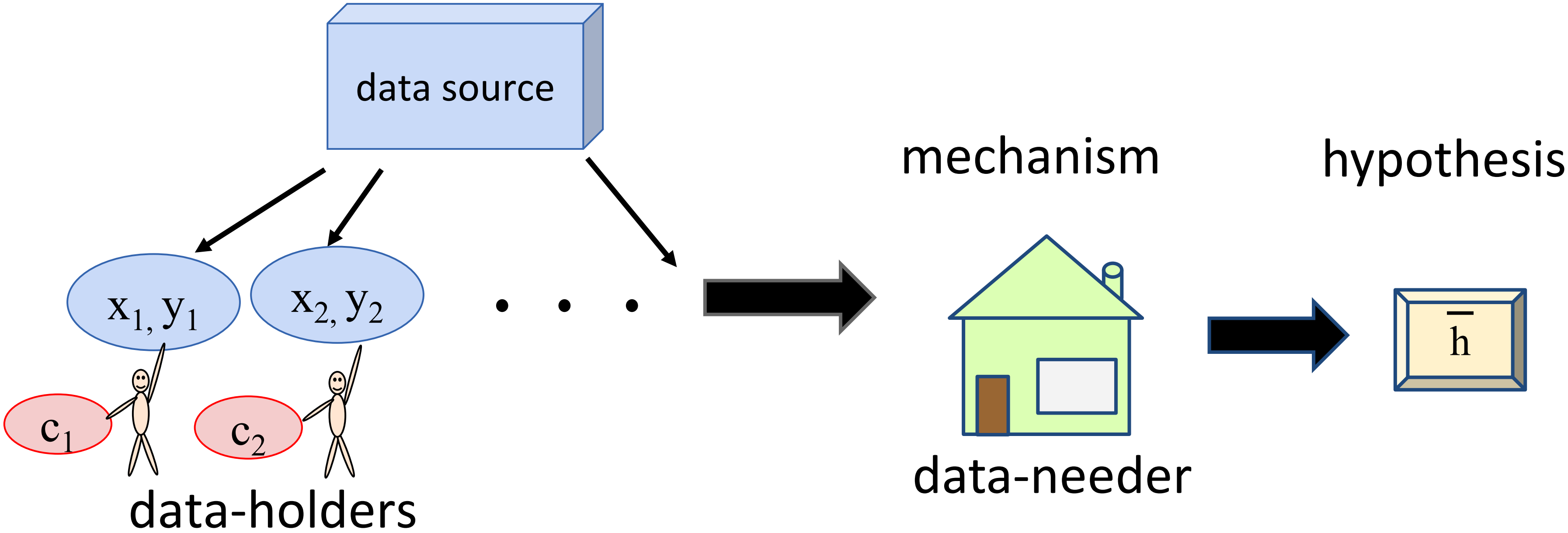


Goal:
Learn a **good** hypothesis *h* with **few** data points



Example: Classification
Data: (point, label) where label is + or −
Hypothesis: hyperplane separating the two types

Our Setting: Data are Held by Humans



Goal:

Learn a **good** hypothesis *h* with **small** budgets

Assumptions:

data cannot be fabricated

costs are **unknown** to the data-needer and **bounded**

costs can be arbitrarily **correlated** with data

In this Work

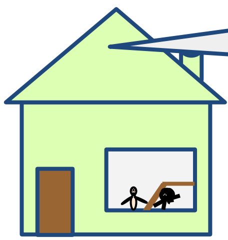
1. **Interface with existing ML algorithms**
   Understand how value derives from learning alg.
   Toward black-box use of learners in mechanisms.

2. **Prove ML-style risk or regret bounds**
   ML-style approach: understand error rate as function of budget and problem characteristics.

3. **Online data arrival**
   Active-learning approach

What can we do?

Want to learn a classifier for HIV
(the maximum cost is $1000)



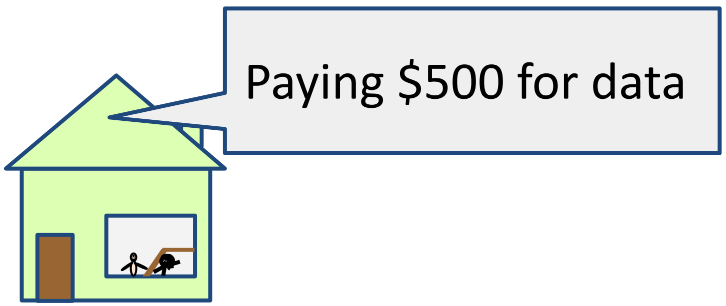
Paying $1000 for everyone
until the budget is exhausted.

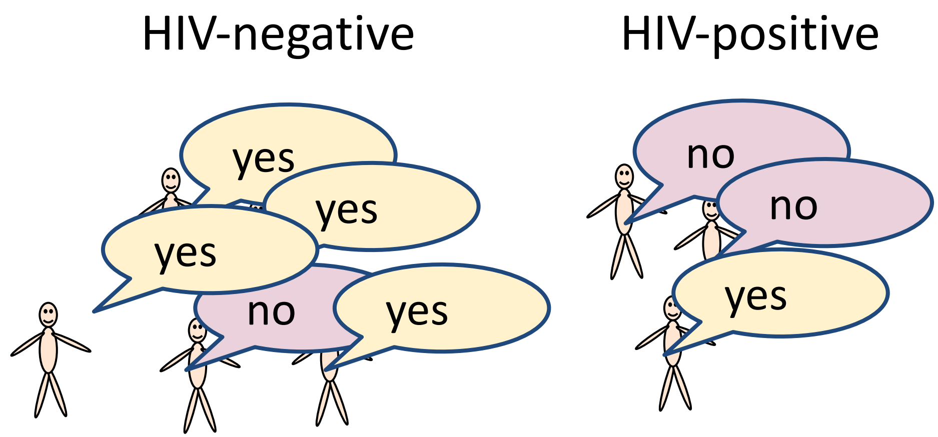
Pro: We can apply standard learning algorithms

Con: Waste a lot of money

What can we do?

Want to learn a classifier for HIV
(the maximum cost is $1000)





Challenge 1: How to deal with **biases**?

Challenge 2: Which data is more **useful**?

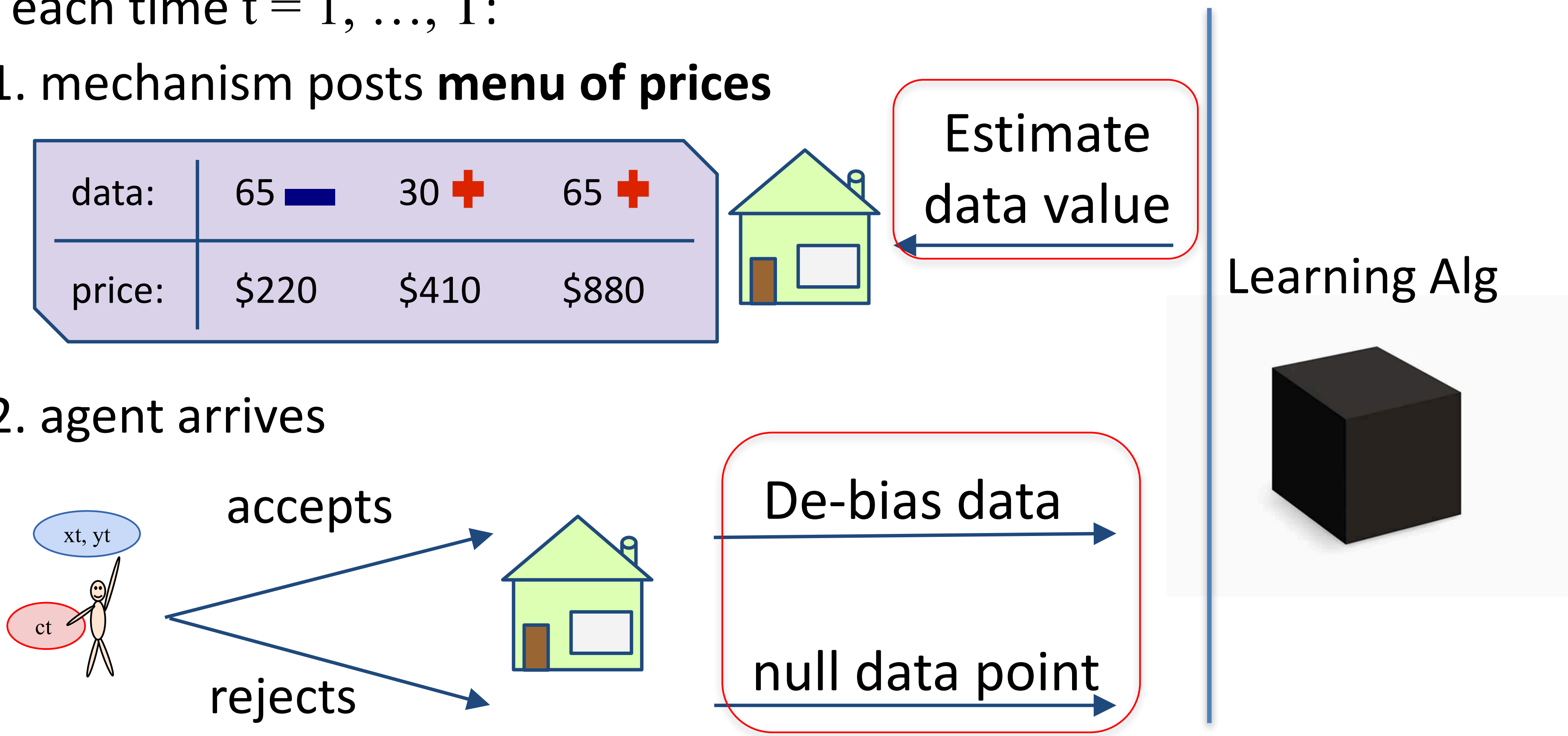
Key Ideas

- Interfacing with existing ML algorithms
- Active learning -> active buying
- De-biasing via importance weighting

At each time $t = 1, \ldots, T$:

1. mechanism posts **menu of prices**

2. agent arrives

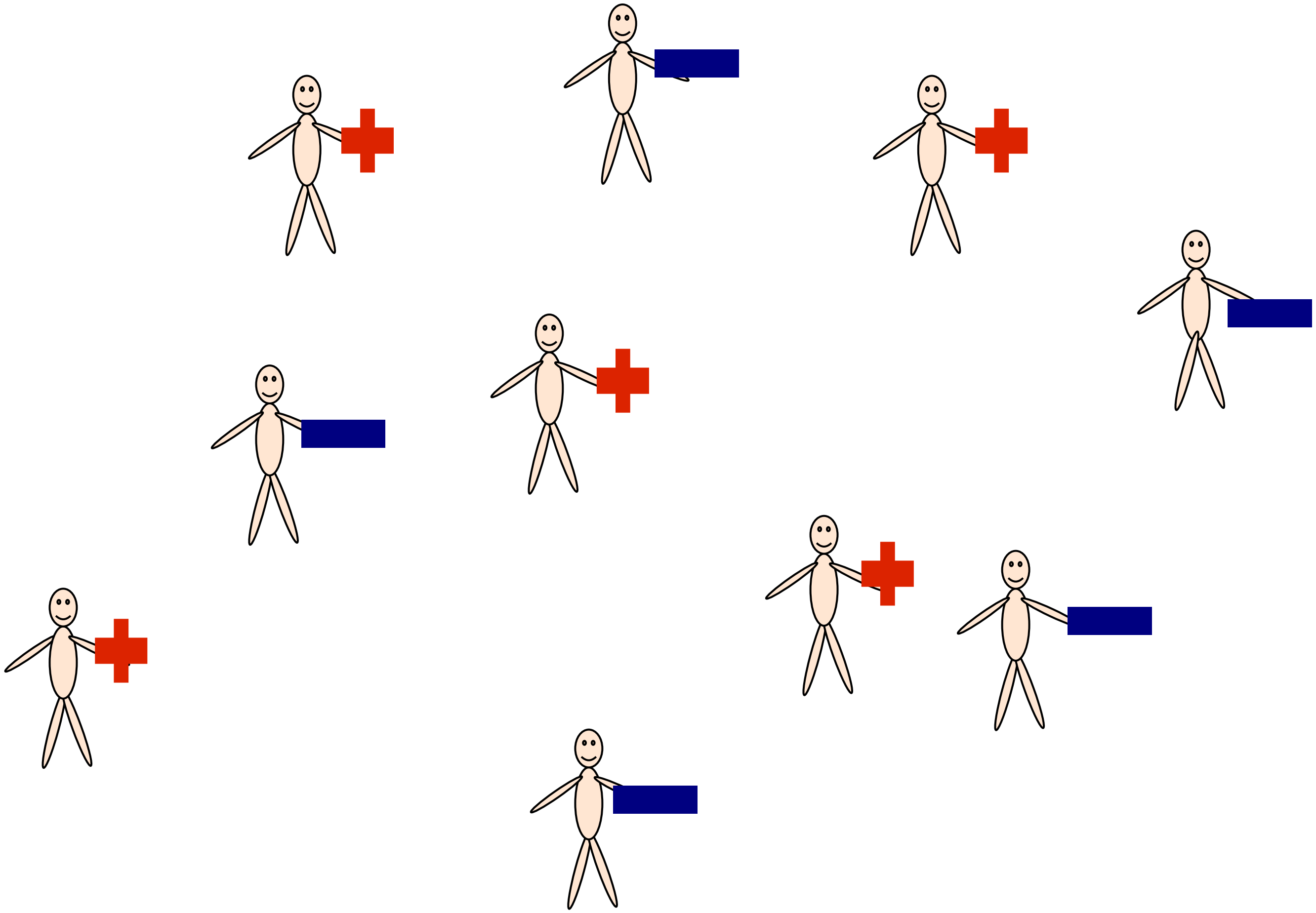


Intuition – How to Debias

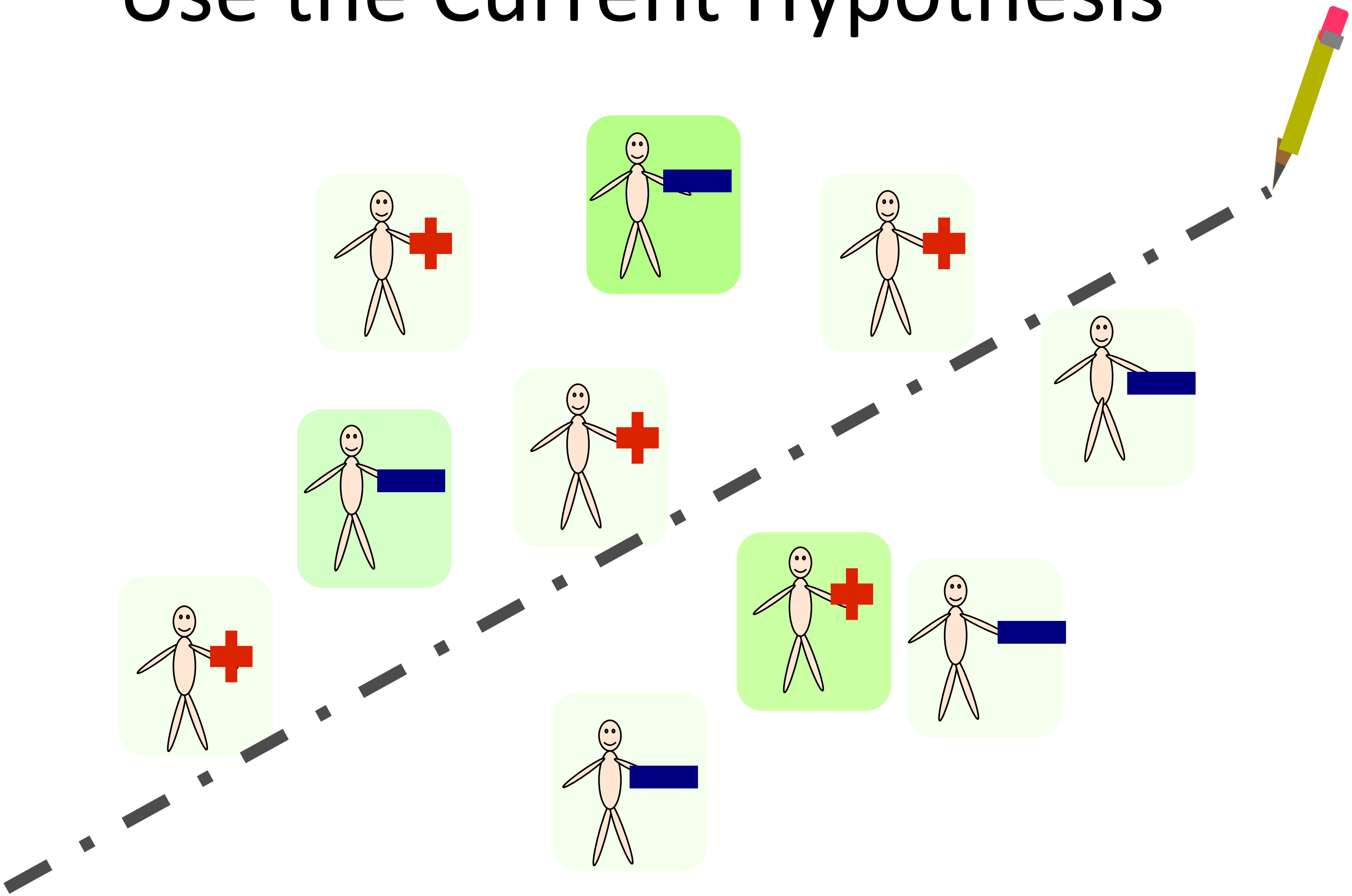
- Estimate the probability of getting data
  - Higher price -> higher sampling probability
- De-bias via importance weighting
  - Double the weights for points with ½ sampling prob

$$E\left[\sum_{i=1}^{n} x_i\right] = E\left[\sum_{i=1}^{n} \frac{x_i}{p_i} \mathbb{1}_{\{x_i \text{ is sampled}\}}\right]$$

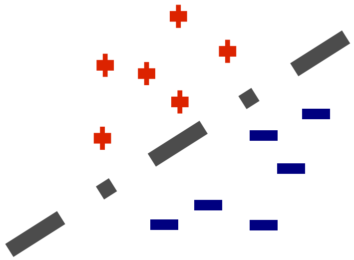
How to Assess Value/Price of Data?



Use the Current Hypothesis



Intuitive Example



- Perceptron algorithm [Rosenblatt, 1958]
  - An online algorithm for learning the linear classifier
  - For each arriving point:
    - If the current hypothesis is right, do nothing
    - If the current hypothesis is wrong, update the hypothesis
  - If there exists a perfect hypothesis
    - The algorithm makes at most ***$1/(margin)^2$*** mistakes
- Pay for mistakes!

Extending to General Learning Alg

- Follow the regularized leader (FTRL)
  - Including online gradient descent, multiplicative weights updates, etc

- Given the de-biased data points, we can calculate the optimal sampling probability for a data point:

$$q_t \propto \frac{\Delta_{h_t, f_t}}{\sqrt{c_t}}$$

cost of data point

$\Delta_{h,f} := \|\nabla f(h)\|_\star$

Difficulties of arriving data points:

How much the arriving points update the current hypothesis

- Design randomized pricing to achieve the above sampling probability

Main Result

- For a general class of learning algorithms (FTRL, e.g., online gradient descent, and multiplicative weight updates), our mechanism achieve

measure of "problem difficulty", in $[0,1]$.

measure of how "good" our mechanism is

$$E \operatorname{loss}(\bar{h}) \le E \operatorname{loss}(h^*) + O\left(\sqrt{\frac{\gamma}{B}}\right)$$

our hypothesis

optimal hypothesis

Budget

Main Result

- For a general class of learning algorithms (FTRL, e.g., online gradient descent, and multiplicative weight updates), our mechanism achieve

measure of how “good” our mechanism is

measure of “problem difficulty”, in $[0,1]$.

$$E\ \mathrm{loss}\left(\bar{h}\right) \leq E\ \mathrm{loss}\left(h^{*}\right) + O\left(\sqrt{\frac{\gamma}{B}}\right)$$

our hypothesis

optimal hypothesis

Budget

- For any mechanism,

$$E\ \mathrm{loss}\left(\bar{h}\right) \geq E\ \mathrm{loss}\left(h^{*}\right) + \Omega\left(\frac{\gamma}{\sqrt{B}}\right)$$

Summary

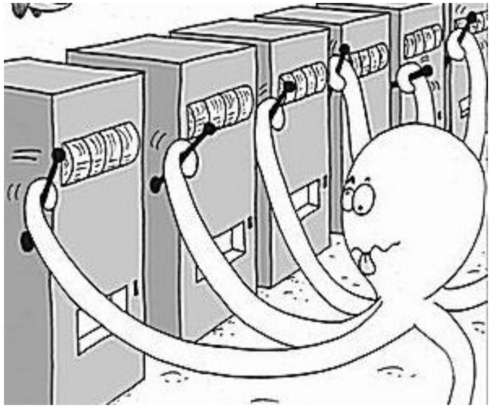
- Explore a new class of machine learning problems with humans in the loop.
- Derive a theoretical bound which parallels the standard bound in machine learning.
- My research interests
  - explore various human-AI interactions and quantify the effects
  - address ethic issues such as fairness, privacy, etc

Recent Research Directions

- Bandits learning with human feedback
- Leveraging worker communications in crowdsourcing
- Online resource allocation (with discussion on fairness and equity)

Multi-Arm Bandit (MAB) Framework

- MAB is a decision making & learning framework
  - Make a sequence of decision on selection, when facing multiple options with unknown statistics.
  - **Q**: which one to select next
  - **Goal**: Maximize total payoff returned by the choice; or regret minimization



Regret:
Utility(OPT) - Utility(ALG)

The reward for each arm is often assumed to be "independently" drawn

Upper Confidence Bound (UCB)

- An index-based method for stochastic bandits
  - Maintain an index for each arm *k* at every time *t*
  - Select the arm with the largest index

$$I_k(t) = \bar{X}_k(t) + \sqrt{\frac{L \log t}{n_k(t)}}, \forall k.$$

Empirical mean: exploitation

Confidence interval: exploration

  - UCB achieves regret bound *O(log T)* in stochastic settings!


Made With
VivaVideo

- This happens everyday…

NETFLIX yelp

amazon

YouTube Quora

- And more…
  - Where to send polices to patrol in different areas
  - How to present food items at school cafeterias

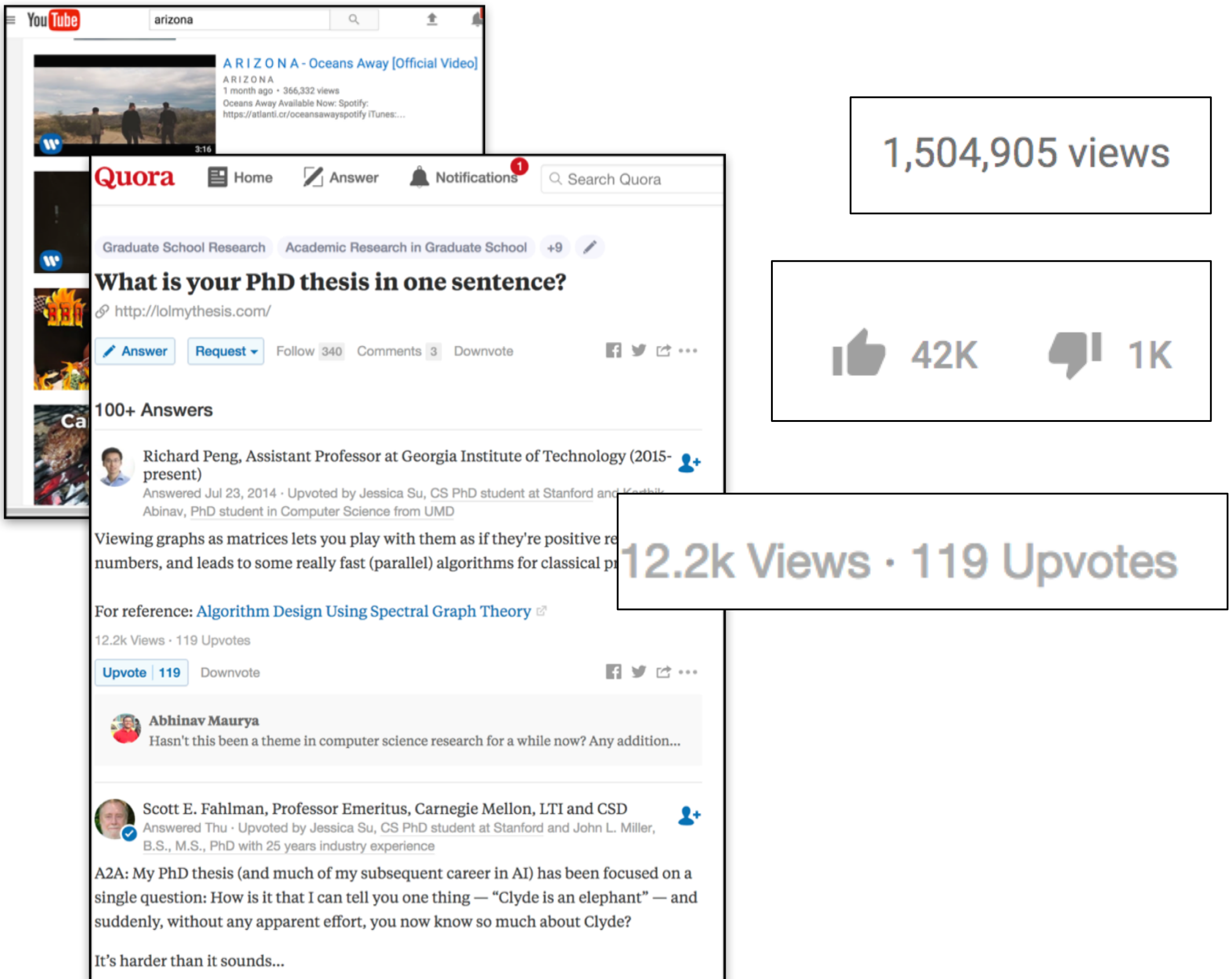
New Arm Generation in Bandit Learning:

An application to User Generated Content Platforms

Joint work with Yang Liu

AAAI'18

User Generated Content Platforms



A Bandit Formulation

- When each new user arrives
  - Show the user some (set of) content
  - Obtain feedback (upvotes, likes, shares, etc) from the user

Assume the user feedback is "unbiased".

- Goal:
  - maximize the total number of positive feedback (user happiness)

- A standard bandit learning problem.

Users are Both Raters and Contributors

- When each new user arrives
  - Show the user some (set of) content
  - Obtain feedback (upvotes, likes, shares, etc) from the user
  - The user decides whether to contribute new content

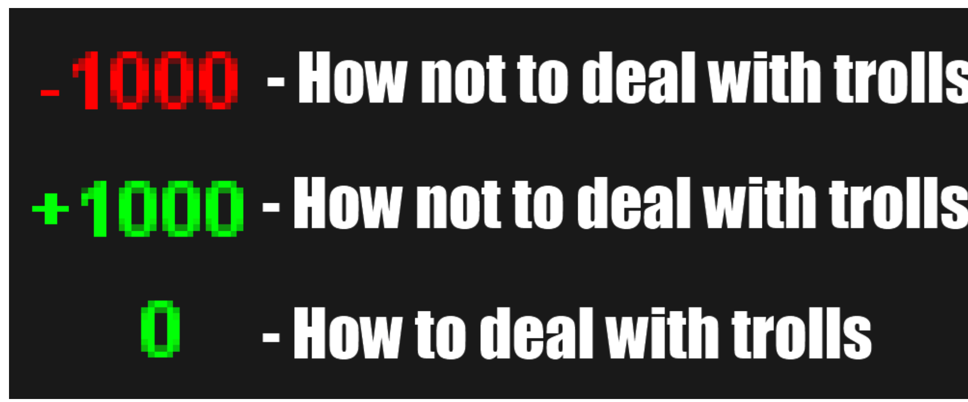
Assume the user feedback is "unbiased".

- Goal:
  - maximize the total number of positive feedback (user happiness)

- A standard bandit learning problem.

Why Users Contribute?

- Model:
  - Users likes attention (e.g., attention => money)



  - Users aim to maximize
    (Total # views of their content) – **(Cost for contributing)**

Curse of Exploration

- Theorem:
  - When T goes to infinity, no standard bandit algorithms will work (impossible to achieve sublinear regret).

- Intuition:
  - We need to sample each content enough number of times to make sure it's not one of the best
    - enough number of times => in the order of log T
  - When T goes large, every user will decide to contribute

Key Question: can we reduce the amount of explorations?

RandUCB: Randomly Dropping Arms

- Run (almost) standard UCB algorithm
- When a new arm is contributed at t, we only include it with probability p_t

---

Algorithm 1: `Rand_UCB`

Input: $\{p_t : t = 1, \ldots, T\}$
for $t = 1, \cdots, T$ **do**
    select arms to display according to UCB1.
    if a new arm is contributed **then**
        add the new arm in $A(t+1)$ with probability p_t
    end if
end for

Incentive Properties of RandUCB

- Set $p_t = \min\{1, C/t\}$
- Theorem
  - If a user has *good content* to contribute, she will always contribute
  - If a user only has bad content to contribute
    - If she arrives before some $\mathrm{t} = \Theta(\log T)$, she will contribute
    - Otherwise, she won't contribute
- RandUCB encourages high-quality contributions

Regret Analysis

Lemma 6. *At any time t, we have*

$$\textit{Regret}_{\mathcal{A}}(t) \leq 16\sqrt{C}\sqrt{t}\log t + O(\sqrt{t}).$$

Asymptotically, RandUCB achieves OPT

Do we really want to randomly drop arms?

- Soft-version of RandUCB
  - Each arm is guaranteed to obtain a small constant number of explorations
  - The dropping decision is based on the small explorations

- Incorporating information other than user votes
  - E.g., apply NLP algorithm to learn the quality of the Quora answer
  - These additional information can be considered as “free explorations”
  - Need a **perfect** ML algorithm to get rid of the curse of exploration entirely
    - Finite T rounds
    - Contextual bandit -> Learning from user feedback

The key is to reduce the amount of exploration

Bandit Learning with Biased Feedback

joint work with Wei Tang

A Bandit Formulation

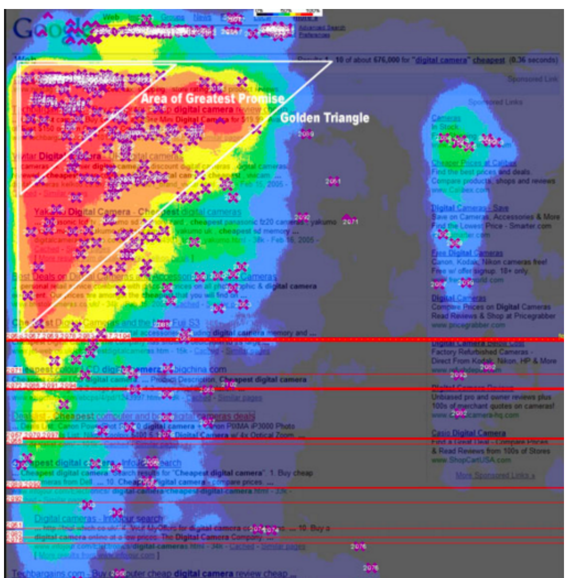
- When each new user arrives
  - Show the user some (set of) content
  - Obtain feedback (upvotes, likes, shares, etc) from the user

Assume the user feedback is "unbiased".

- Goal:
  - maximize the total number of positive feedback (user happiness)

Users' feedback might be biased

- Bias in selecting items
- Bias in voting
- and others…



Feedback Model 1

- Model
  - Users feedback depends on
    - their own experience
    - average feedback of others
- Positive results
  - Collectively, users are performing online gradient descent on a latent function.
  - **Sublinear regret is achievable** under mild conditions using techniques from online optimization.

Feedback Model 2

- Model
  - Users feedback depends on
    - their own experience
    - average feedback of others
    - length of the feedback history

- Impossibility results
  - The average feedback converges to a random variable with non-zero variance.
  - **No algorithm can achieve sublinear regrets.**

Summary and Future Work

- A small deviation of human behavior could lead to very different outcomes for machine learning.

- Future/ongoing directions
  - Information design to induce different behavior.
  - Learning how to “nudge” humans in decision making.

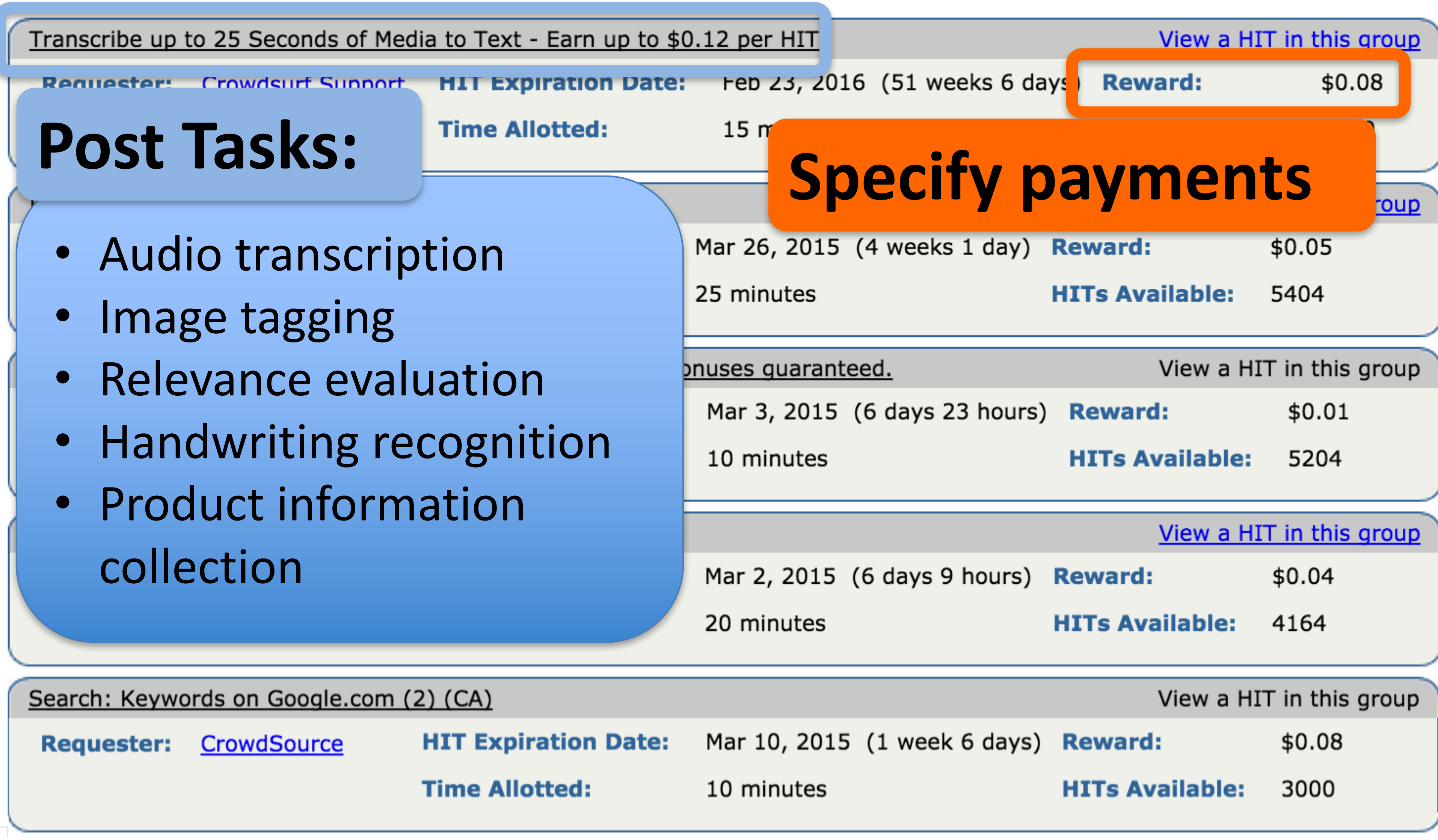
Recent Research Directions

- Bandits learning with human feedback

- Leveraging worker communications in crowdsourcing

- Online resource allocation (with discussion on fairness and equity)

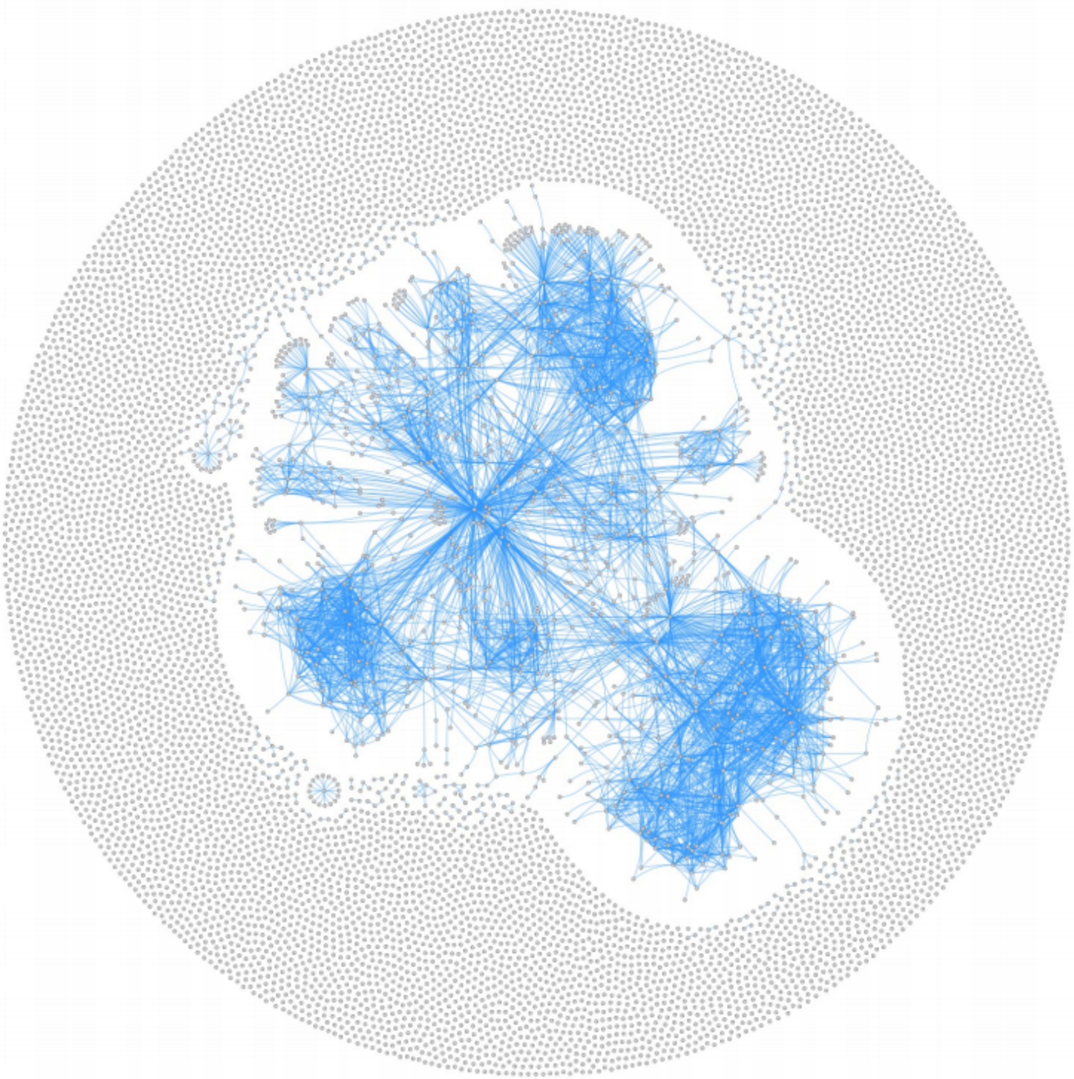
amazon mechanical turk™
Artificial Artificial Intelligence



Post Tasks:

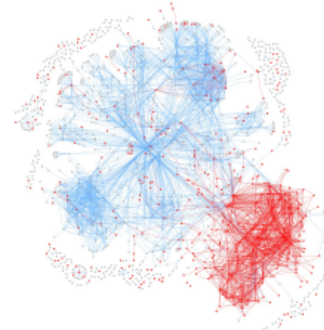
- Audio transcription
- Image tagging
- Relevance evaluation
- Handwriting recognition
- Product information collection

Specify payments

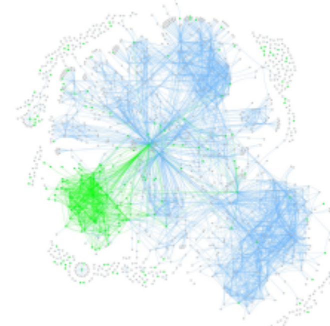


(a) The communication network

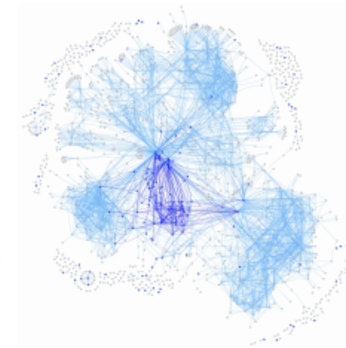
(b) Reddit HWTF



(c) MTurkGrind



(d) TurkerNation



(e) Facebook

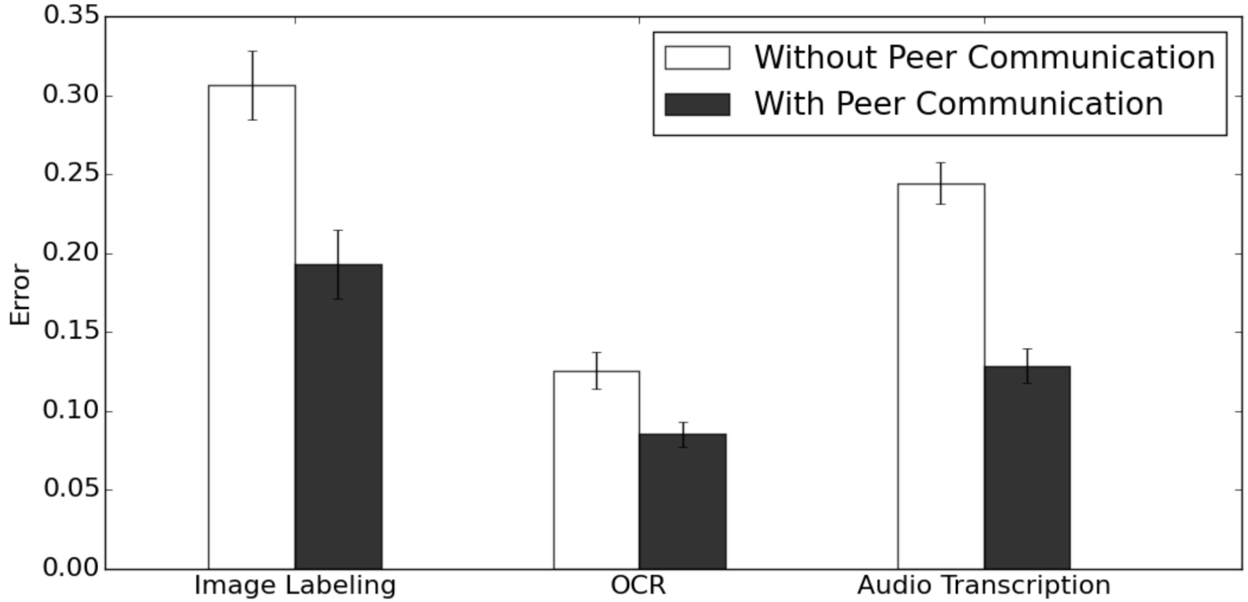
(f) MTurkForum

Leveraging Peer Communication to Enhance Crowdsourcing

joint work with Wei Tang and Ming Yin

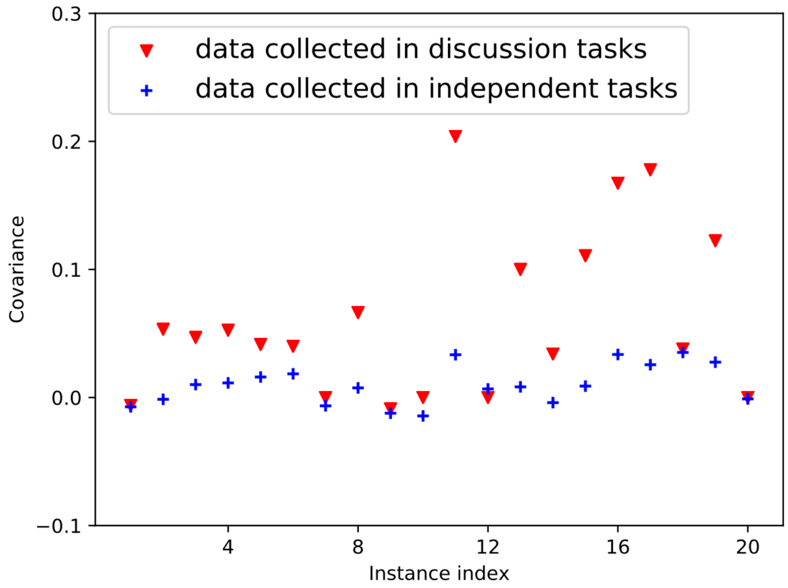
The Effects of Communications

- Peer communication:
  – Ask a pair of workers to work independently, then discuss, and then submit final answers.
- Experiments on >1000 online workers



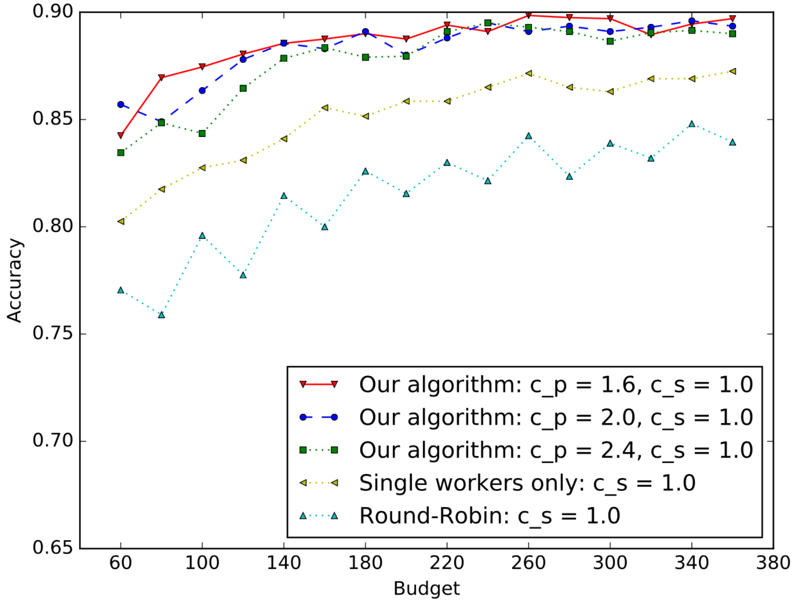
OK, but.....

- Are two correlated data points better than two independent but lower-quality data points?
- Hiring two workers to communicate might be more costly than hiring two independent workers



Leveraging Correlated Data

- Derive an aggregation rule that achieves maximum likelihood aggregation
- Propose a MDP framework to determine which task and whether to use communications.



Summary and Future Directions

- Enabling communications between users could lead to better-quality data

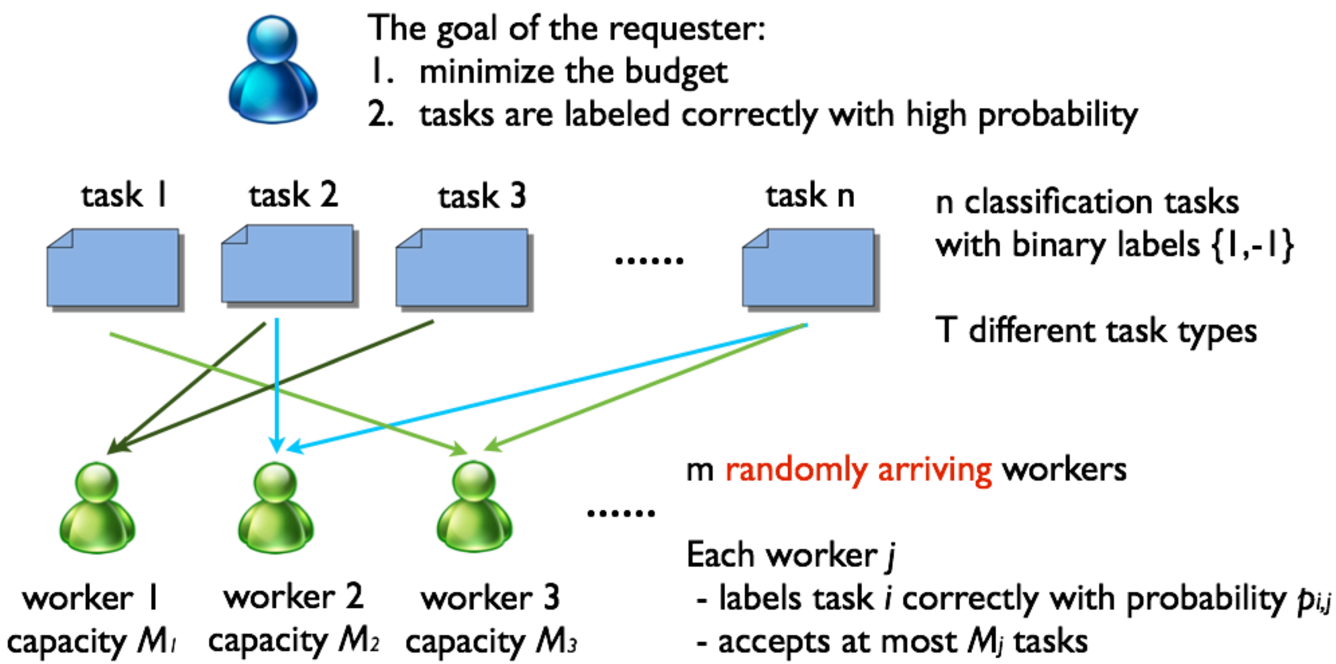
- Future work
  - Opinion formation (subjective tasks)
  - Network manipulation
  - Adversarial attack

Recent Research Directions

- Bandits learning with human feedback
- Leveraging worker communications in crowdsourcing
- Online resource allocation (with discussion on fairness and equity)

Online Task Assignment

joint work with Shahin Jabbri and Jennifer Wortman Vaughan



- **Our approach**
  - Propose a (non-trivial) integer program formulation
  - Estimate workers' skills with gold standard tasks
  - Find the assignment using online primal-dual techniques
- **Our result**
  - A near-optimal online assignment algorithm

Online Primal-Dual Matching Algorithms: An Application to Kidney Exchange

joint work with
Kelsey Lieberman, William Macke, Zhuoshu Li, and Sanmay Das

Kidney Exchange – An online resource allocation problem

- A primal-dual formulation
  - The dual formulation is helpful for dealing with the dynamic nature of kidney exchange.
  - The dual space is useful in quantifying whether we are "fair" for different population.

Primal

$$\max \sum_{n=1}^{N} \sum_{i=0}^{I} w_{n,i} x_{n,i}$$

$$\text{s.t.} \sum_{i=0}^{I} x_{n,i} \leq 1, \forall n \in [T]$$

$$\sum_{n=1}^{N} x_{n,i} + \sum_{j=1}^{I} x_{T+i,j} \leq 1, \forall i \in [I]$$

$$x_{n,i} \in \{0,1\}, \forall n \in [N], \forall i \in [I]^*$$

Dual

$$\min \sum_{t=1}^{T} \alpha_t + \sum_{i=0}^{I} \beta_i$$

$$\text{s.t. } w_{t,i} - \alpha_t - \beta_i \leq 0, \forall t \in [T], i \in [I]^*$$

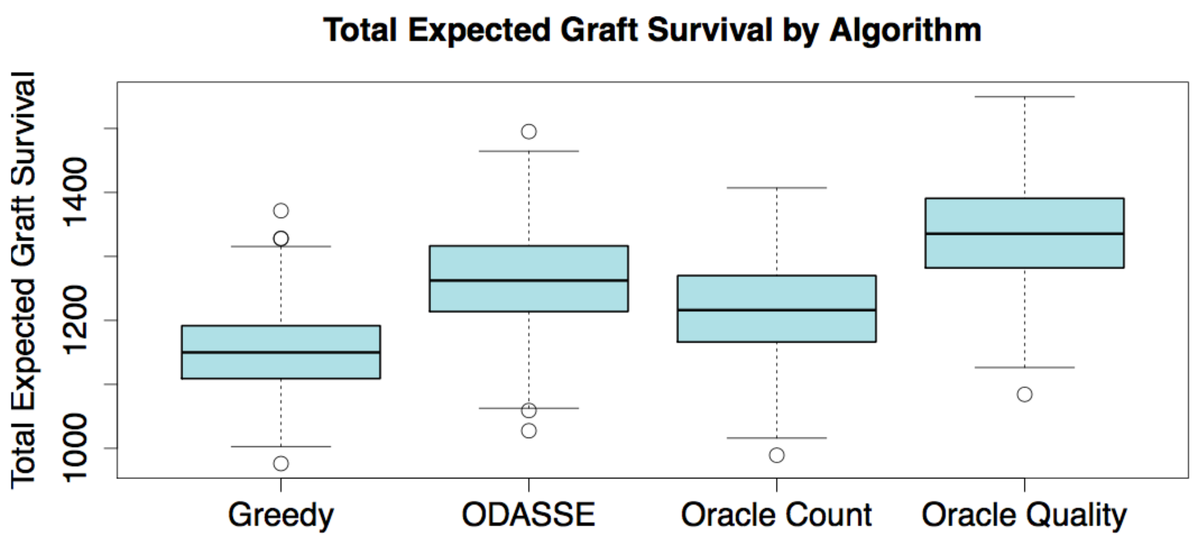
$$w_{t+j,i} - \beta_j - \beta_i \leq 0, \forall i \in [I], j \in [I]$$

$$\alpha_t, \beta_i \geq 0, \forall t \in [T], i \in [I]$$

$$\beta_0 = 0$$

Results

- Overall utility is higher than greedy algorithms



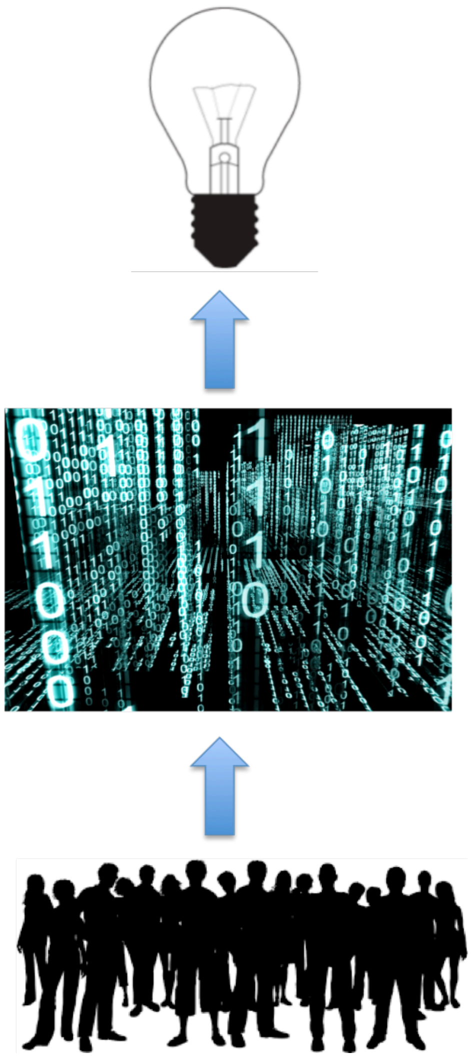
- More fair to hard-to-match groups than greedy algorithms

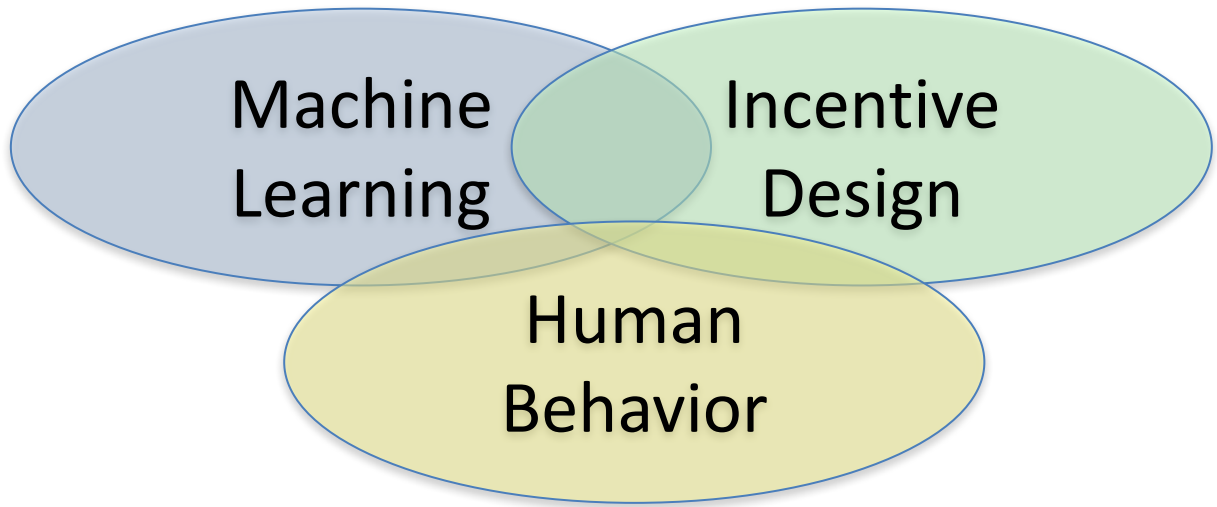
Summary and Future Work

- Online resource allocation is ubiquitous and has direct impacts to humans' welfare.
- Propose a primal-dual framework that's more efficient and don't sacrifice the welfare of sub-populations.

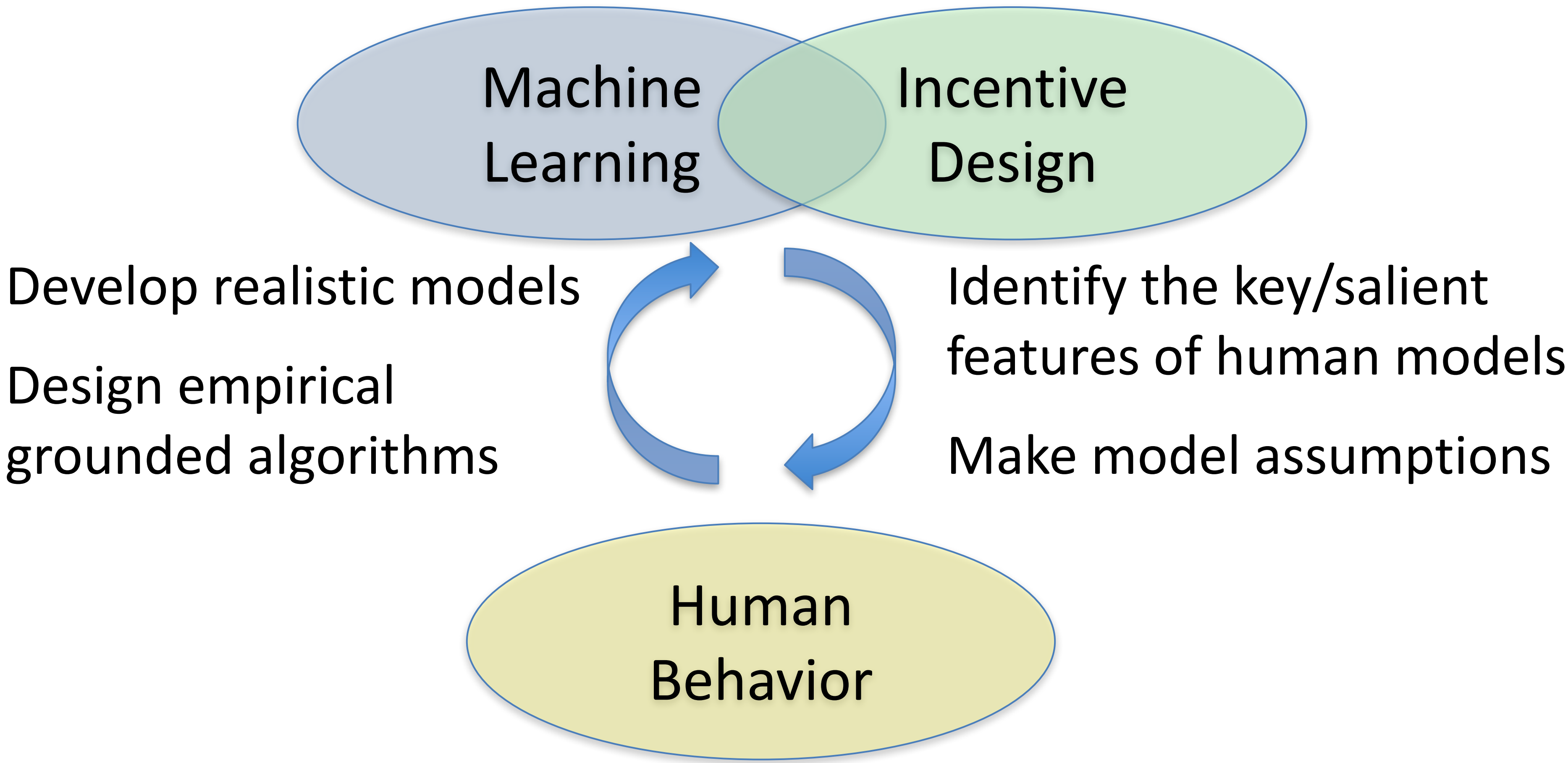
- Future direction
  - Explore the effects of dynamic process in the allocation problem.
  - Characterizing and understanding "fairness" notions in the dual space

Summary of My Research





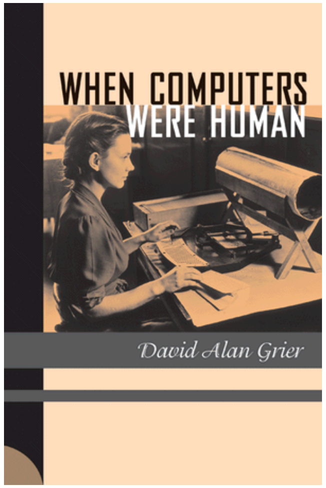
Research Directions – Closing the Loop

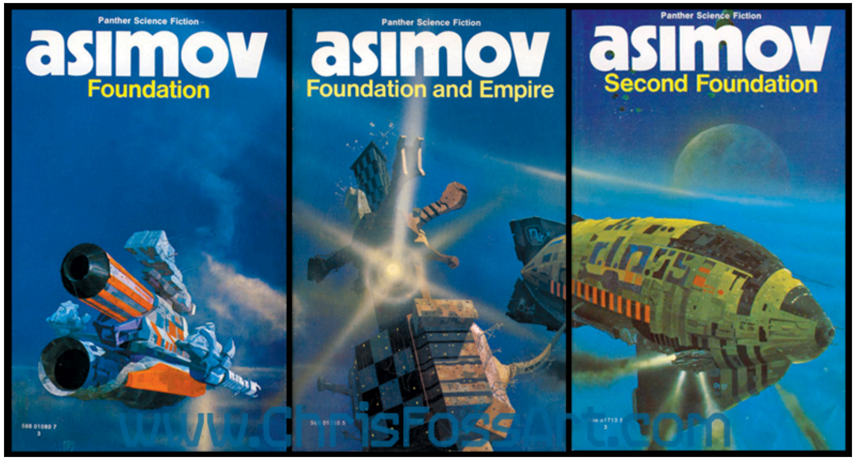


Human-AI Interaction

Vision

- Develop formal computational frameworks with humans in the loop.
  - Quantify the performance/costs of human algorithms
  - Address ethical issues (fairness, privacy, etc)
- Predict the “future” (e.g., the outcome and evolution of the platforms with humans involved).





Questions?