Human Allied Artificial Intelligence

Harsha Kokel, Sriraam Natarajan

StARLinG LAB

STARLING LAB

Lab Members

Who we are!

Sriraam Natarajan

Yuqiao Chen

Justin G. Dula

Alexander L. Hayes

Bhagirath Athresh Karanam

Harsha Kokel

Saurabh Mathur

Brian Ricks

Michael Skinner, MD

Siwen Yan

Who we are!

Alumni (PhD)
Nandini Ramanan, Navdeep Kaur, Srijita Das, Devendra Singh Dhami, Mayukh Das, Phillip Odom, Shuo Yang, Tushar Khot

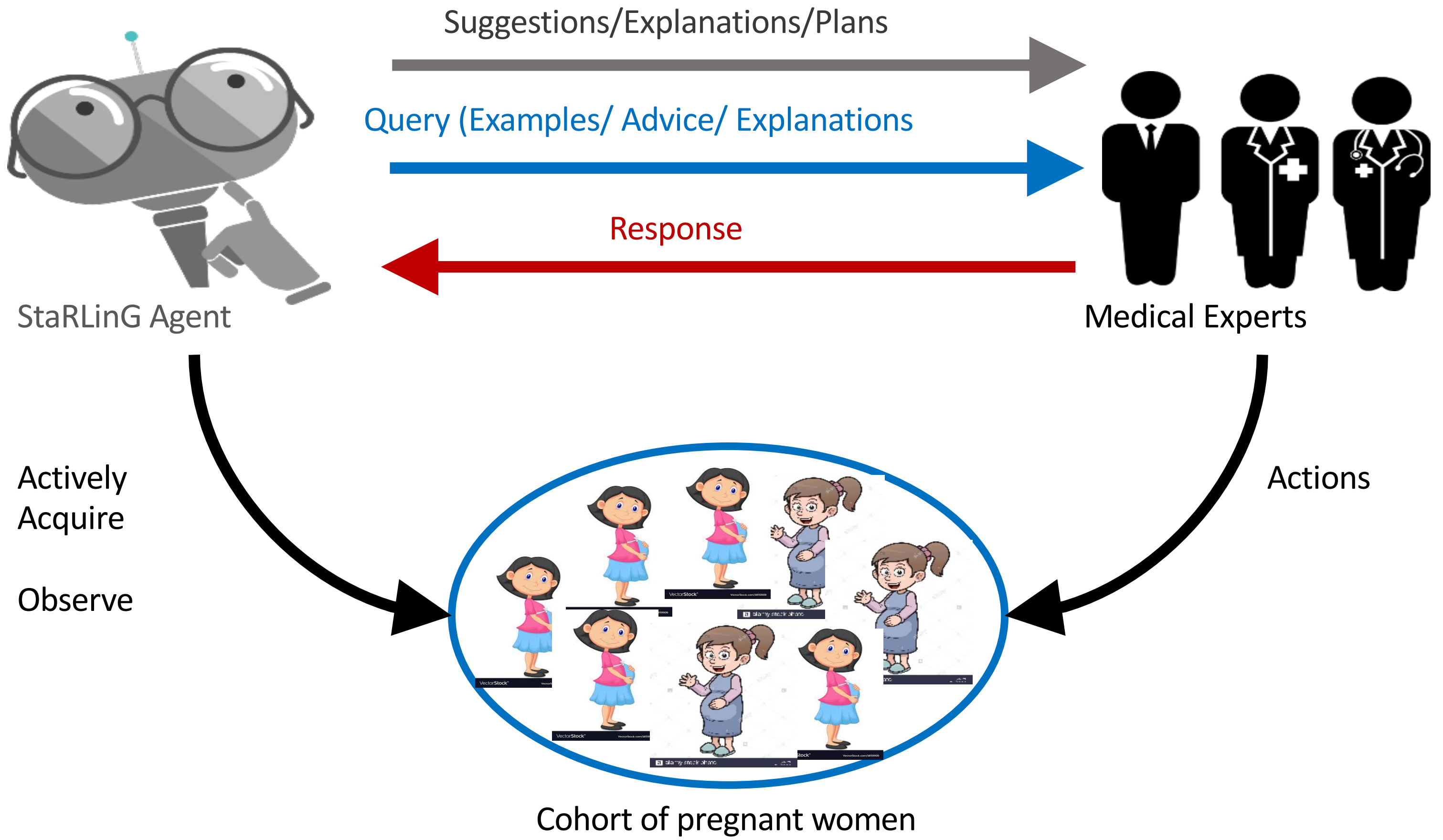
Key Collaborators
Kristian Kersting, Jude Shavlik, Gautam Kunapuli, Prasad Tadepalli, David Page, Dan Roth, Jana Doppa, Ron Parr, Predrag Radivojac, William Cohen, David Poole, Kay Connelly, Balaraman Ravindran, Clinical collaborators ….

Funding agencies
DARPA (Minerva, CwC, DEFT & Machine Reading), NSF (EAGER & SCH), AFRL, ARO (YIP, STIR), AFOSR (SBIR), NIH (R01), Indiana (Precision Medicine), XEROX PARC, Amazon, Intel, TURVO and Verisk Inc.

What is Human Allied AI?

Can we build systems that can seamlessly interact with, learn from, collaborate with and potentially teach the human expert?

What is Human Allied AI?



Fern et al. IJCAI 07; Natarajan et al. ILP08, ILP09; Natarajan et al. KAIS 11; Fern et al. JAIR 14; Kunapuli et al. ICDM 13; Odom et al. AAAI 15, AAMAS 16, ECML 16, ILP 16, AIME 15; Yang et al. ICDM 14, ECML 13; Macleod et al. CHASE 16; Natarajan et al. IJCAI 18; Das et al. AAAI 19, HMCL WS 17, AAMAS 18; KBS 18; Ramanan et al. BIBM 17, KR 18; Dhami et al. AIME 17, Smart Health 18, AI for Good 19; Kaur et al. ILP 17,19, IJAR 20; Hayes et al. KCAP 17; Kokel AAAI 20, ICAPS 21; Das et al. 20; Karanam AIME 21; Dhami AIME 21

(Our) 3 Steps to HAAI

Effective Learning

Learn "only" from data
Effective
Efficient
Generalizable
Personalized
Explainable
…
Ignores human knowledge

Advice taking

Allow "**richer**" human inputs

More than a "**mere labeler**"

Take advice and guidance

Allows for robust learning

Active advice seeking

Close the "**loop**"

Knows what it knows

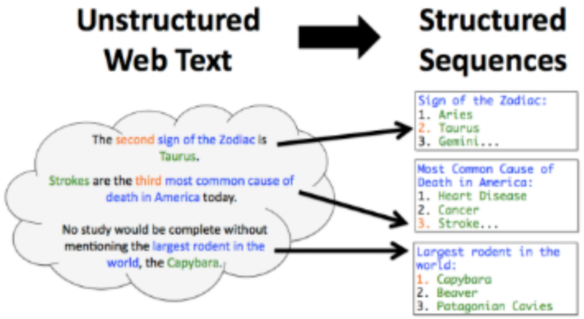
Asks what it does **not** know

Student-teacher interaction

Teach the human!

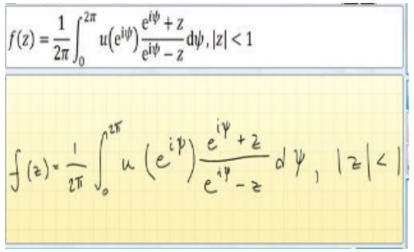
Several **Real** Applications

Logistics Domains



Information Extraction

Games



Handwriting Recognition

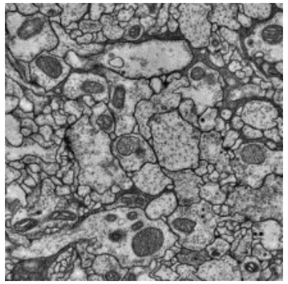
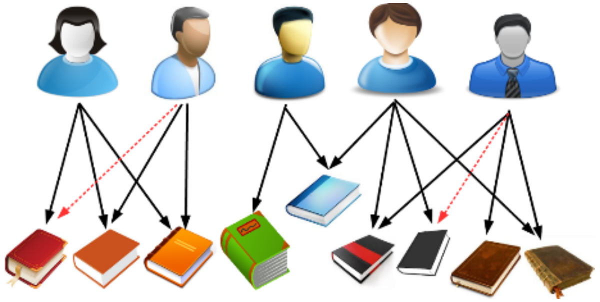
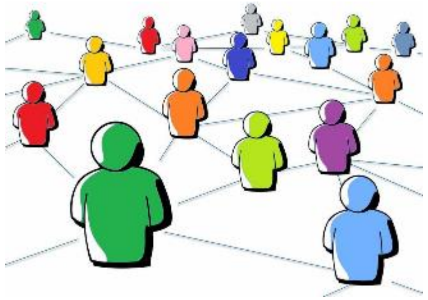


Image Segmentation and Classification



Recommendation Systems



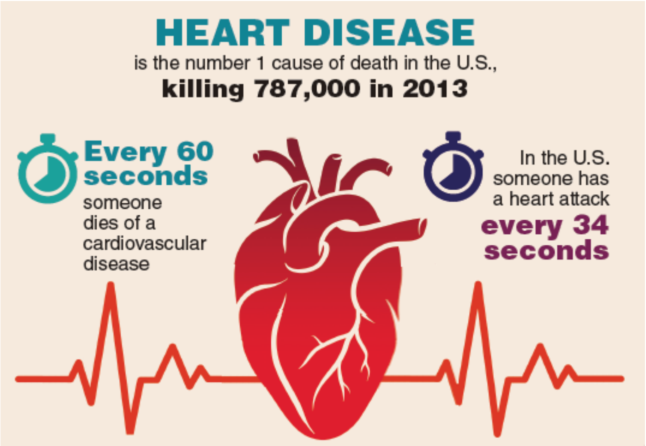
Social Network Analysis

UTSouthwestern Medical Center

VANDERBILT UNIVERSITY MEDICAL CENTER

The promise of discovery

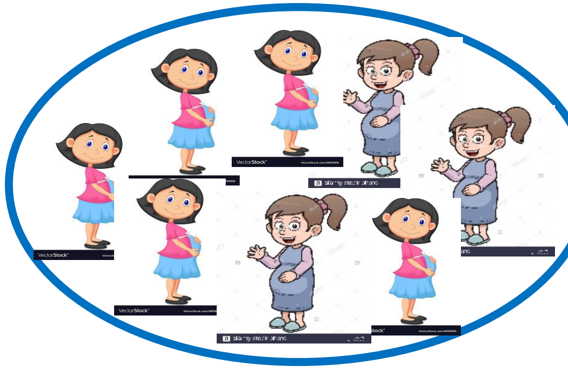
Wake Forest® Baptist Health



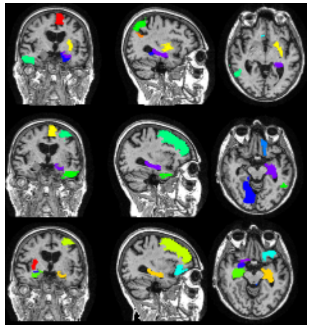
Cardiovascular Events Prediction and Treatment

Predicting rare diseases, post-partum depression from survey data

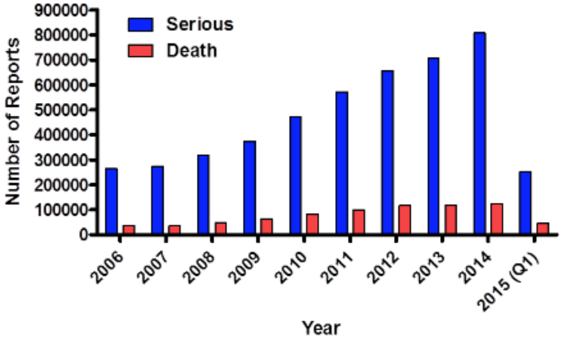
Predicting diabetes / cognition from sensors



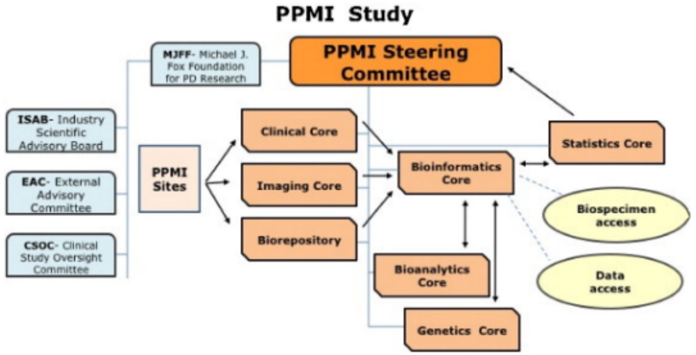
Qualitative Knowledge Extraction



Alzheimer's disease prediction



Predicting the side-effects of drugs



Parkinson's disease prediction

Regenstrief Institute

Better Care. Better Health.

Marshfield Clinic®

Research Institute

(Our) 3 Steps to HAAI

Effective Learning

Advice taking

Active advice seeking

(Our) 3 Steps to HAAI

Active advice seeking

Advice taking

Effective Learning

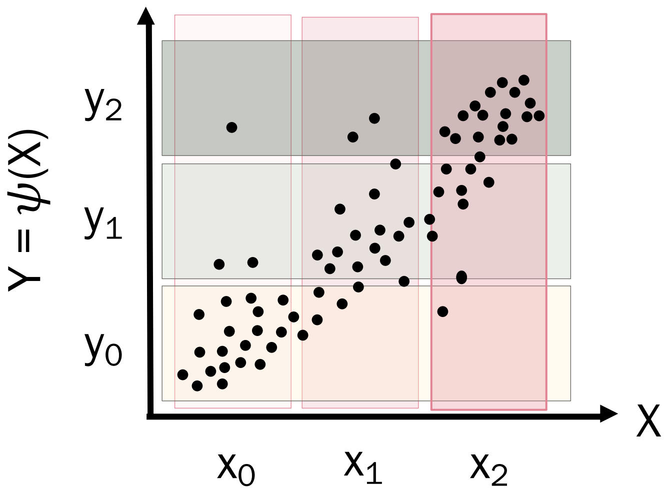
Allow "**richer**"
human inputs

More than a
"**mere labeler**"

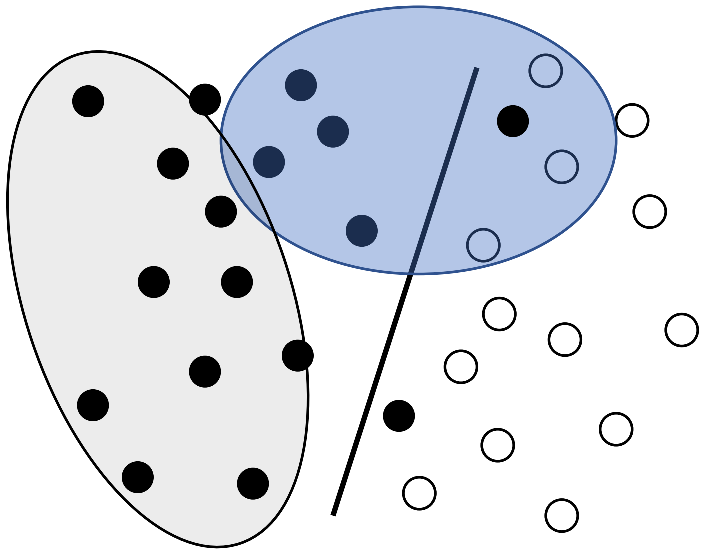
Take advice and guidance

Allows for robust learning

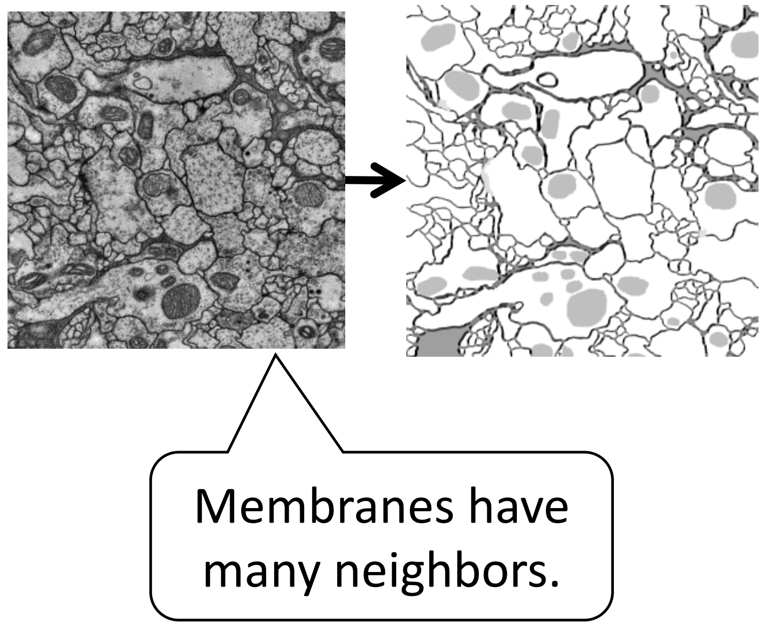
Types of Advice



Monotonicity
As x ↑, y ↑
Synergy



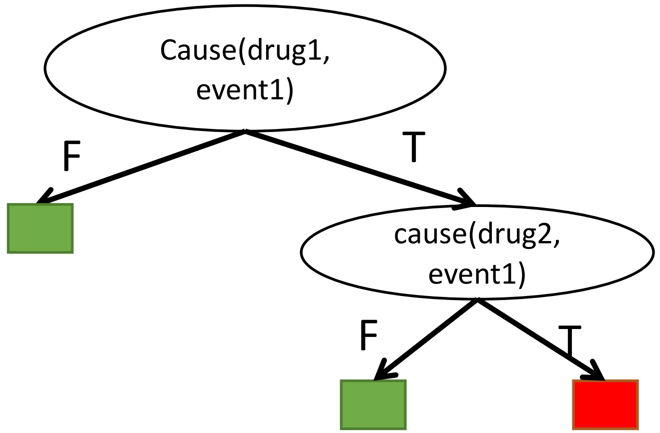
Precision/Recall
Tradeoff



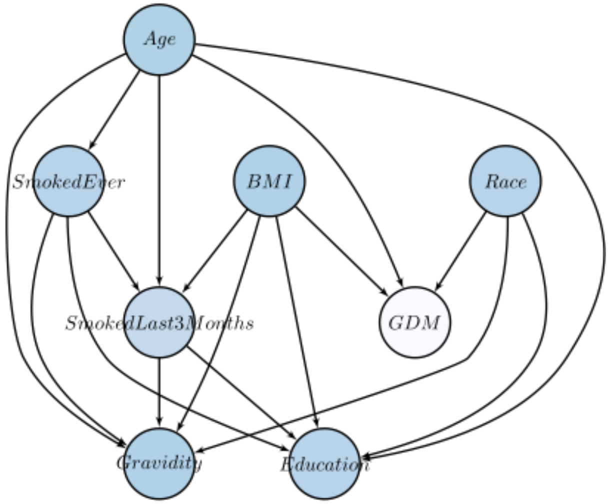
Preference information

Yang 13, 14; Odom 15, 18; Das 18 ;Kokel 20

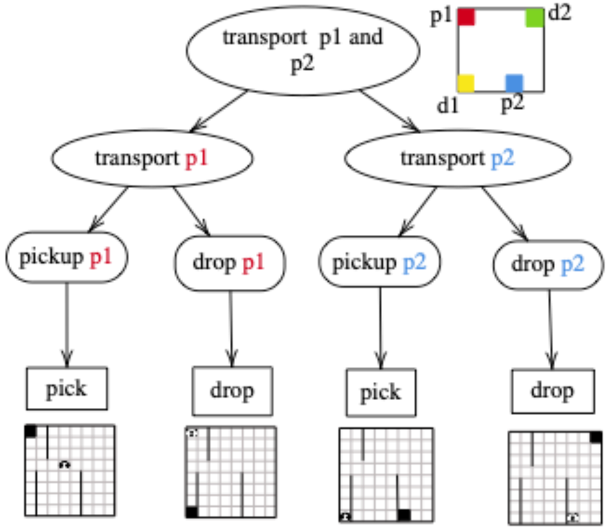
Types of Advice



Rules / Horn clauses



Causal Knowledge/
Influence information



Task hierarchy

Example Frameworks for Advice Taking

- Probabilistic Graphical Models (Yang and Natarajan 2013, 2014; Ramanan 2018, 2021)
- Relational Probabilistic Models (Odom and Natarajan, 2016, 2018; Das et al 2018, 2021)
- Gradient Boosting (Odom & Natarajan 2018; Yang et al 2014; Kokel et al 2020)
- Deep Learning (Dhami et al 2021)
- Hierarchical Planning (Das et al 2018)
- Inverse Reinforcement Learning (Odom et al 2016)
- Imitation Learning (Odom and Natarajan 2018)
- Probabilistic Planning (Das et al 2018, 2019)
- Task-specific Abstractions (Kokel et al 2021)

Gradient Boosting

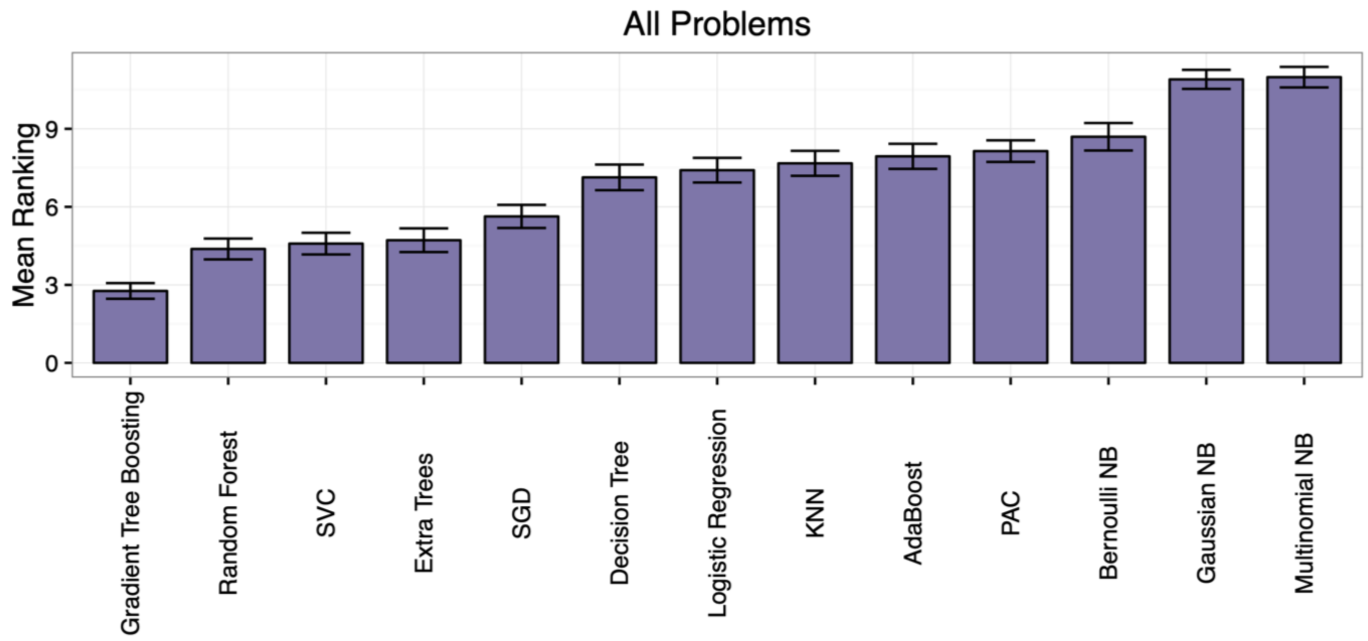
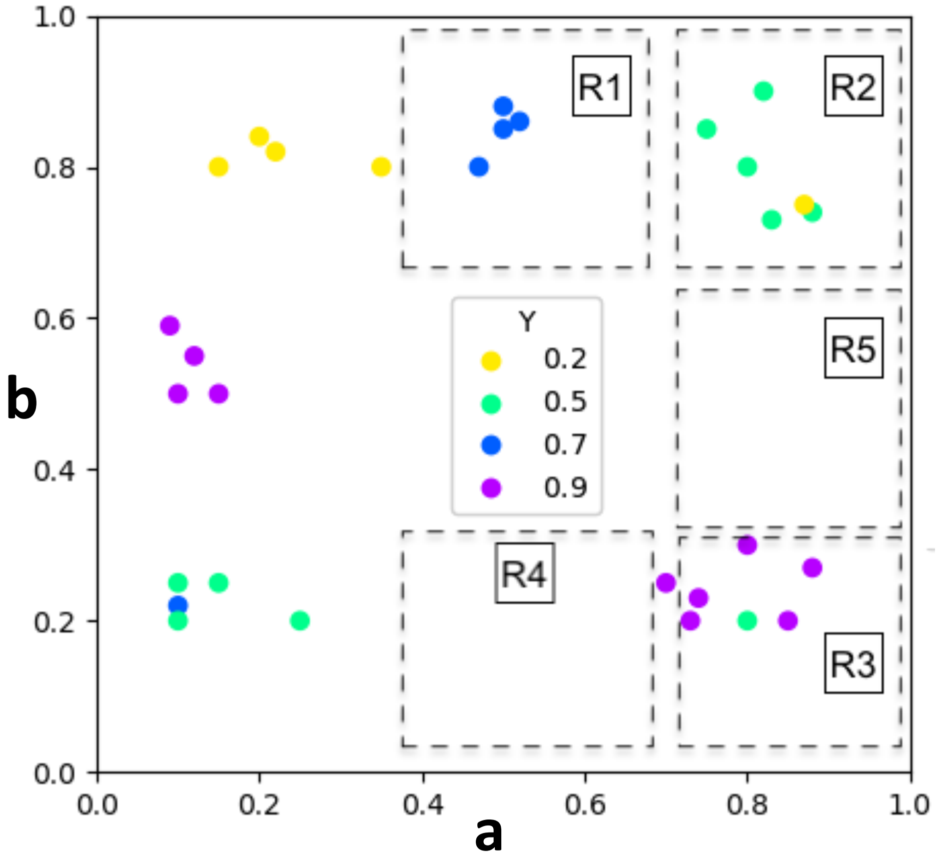
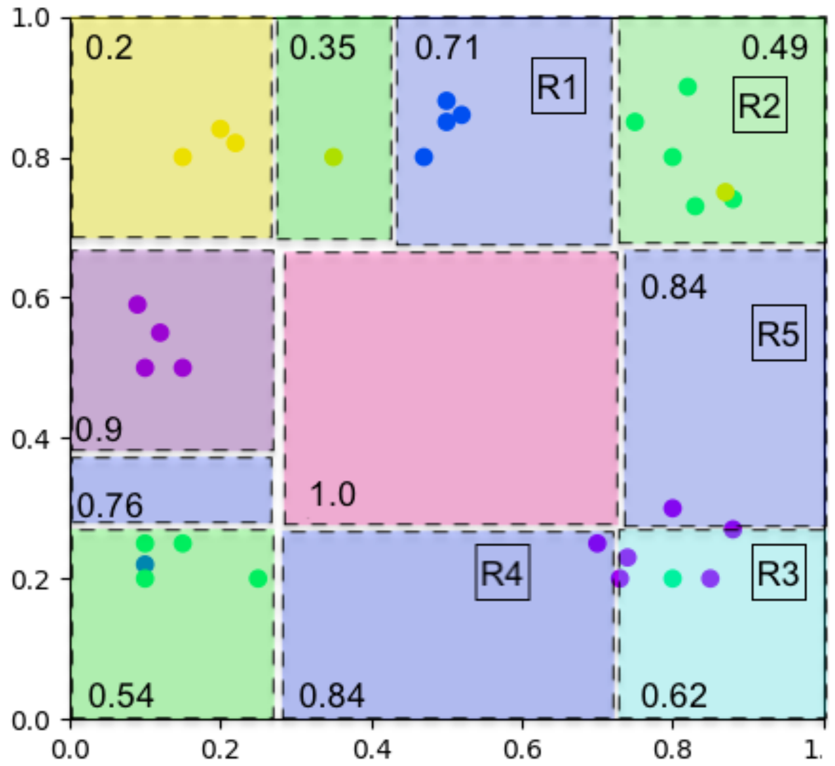


Fig. 1. Average ranking of the ML algorithms over all datasets. Error bars indicate the 95% confidence interval. Source: Olson et al. 2018 PSB

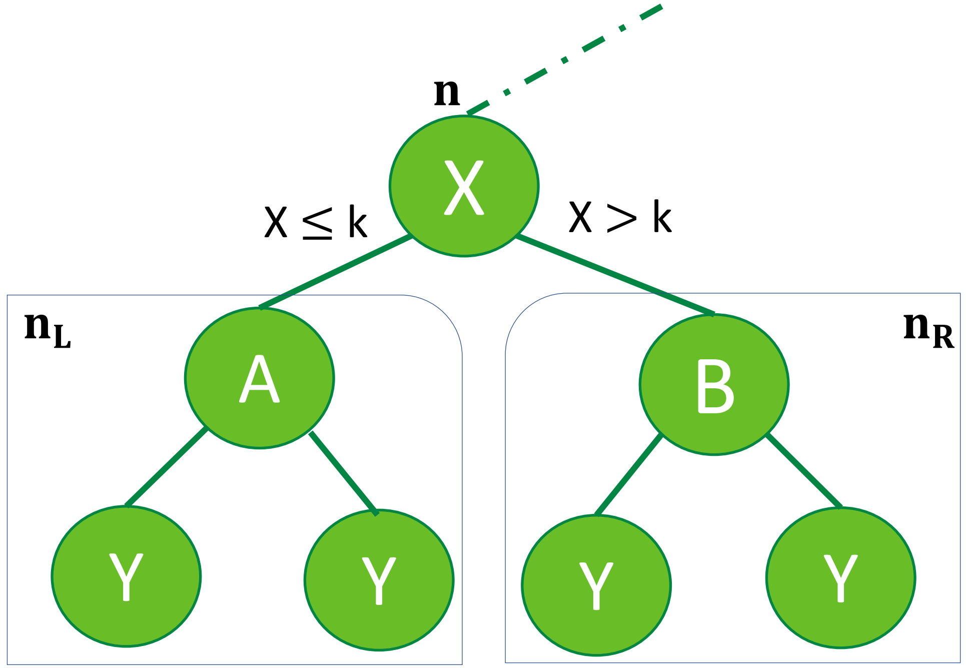
Sparse pockets





Boosting with Qualitative Constraints

$X \overset{Q+}{\prec} Y$



$$x_1 < x_2 \Rightarrow \mathbb{E}_{\psi}[x_1] \leq \mathbb{E}_{\psi}[x_2]$$

$$\mathbb{E}_{\psi}[\mathbf{n_L}] \leq \mathbb{E}_{\psi}[\mathbf{n_R}] + \varepsilon$$

$$\zeta_n \left\{ \quad \mathbb{E}_{\psi}[\mathbf{n_L}] - \mathbb{E}_{\psi}[\mathbf{n_R}] - \varepsilon < 0 \right.$$

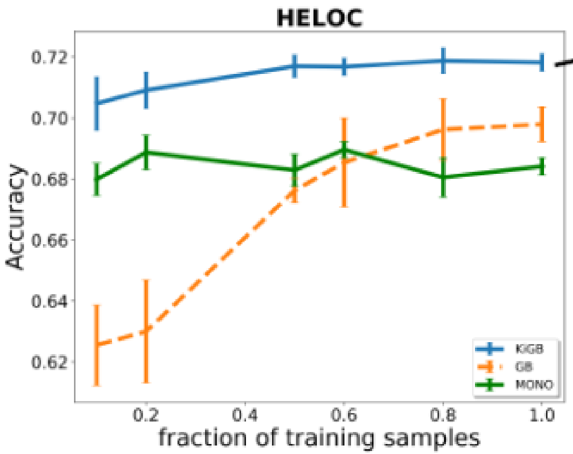
$$\text{argmin} \underbrace{\sum_{i=1}^{N} (y_i - \psi_t(x_i))^2}_{\text{loss function w.r.t data}} + \underbrace{\frac{\lambda}{2} \sum_{\mathbf{n} \in \mathcal{N}(\mathbf{x}_c)} \max(\zeta_{\mathbf{n}} \cdot |\zeta_{\mathbf{n}}|, 0)}_{\text{loss function w.r.t advice}}$$

Boosting with Qualitative Constraints

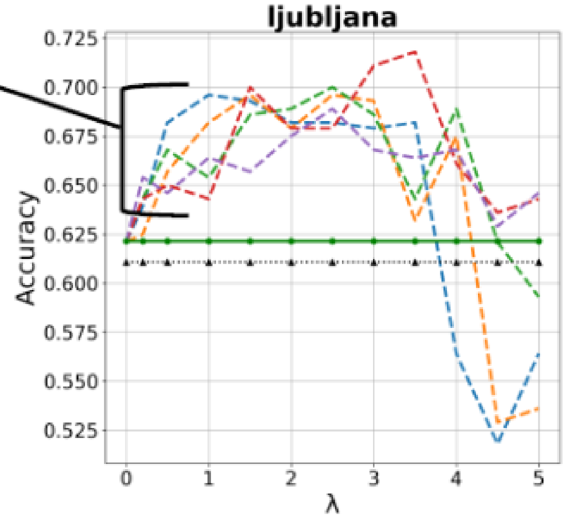
Objective

$$\underbrace{\operatorname*{argmin}_{\psi_t} \sum_{i=1}^{N} \left(\tilde{y}_i - \psi_t(x_i)\right)^2}_{\text{loss function w.r.t data}} + \underbrace{\frac{\lambda}{2} \sum_{\mathbf{n}\in\mathcal{N}(\mathbf{x_c})} \max\left(\zeta_{\mathbf{n}} \cdot |\zeta_{\mathbf{n}}|, 0\right)}_{\text{loss function w.r.t. advice}}$$

Results



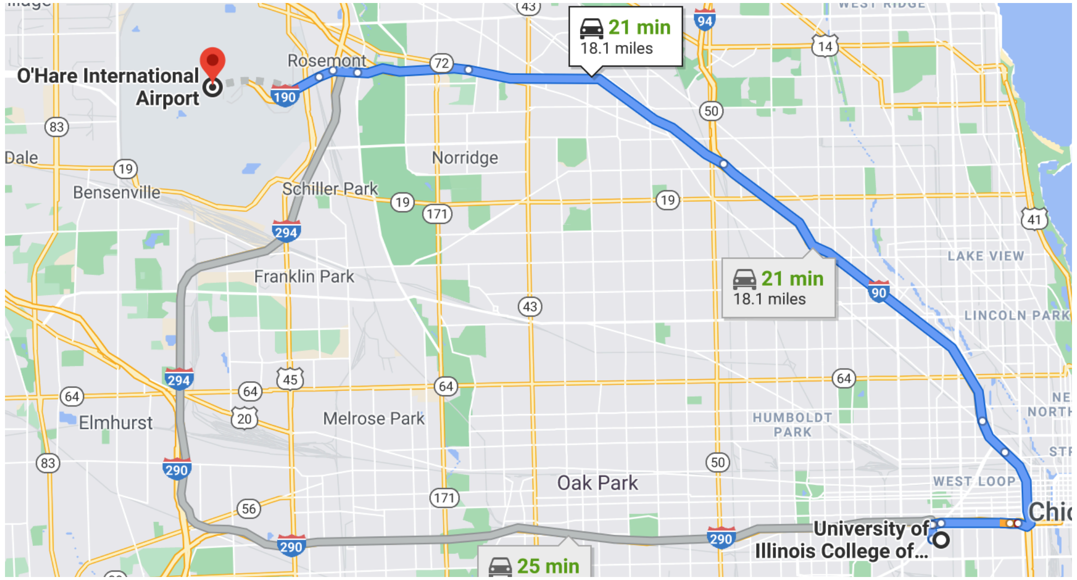
our method



Kokel et al. AAAI 2020

Sequential Decision Making

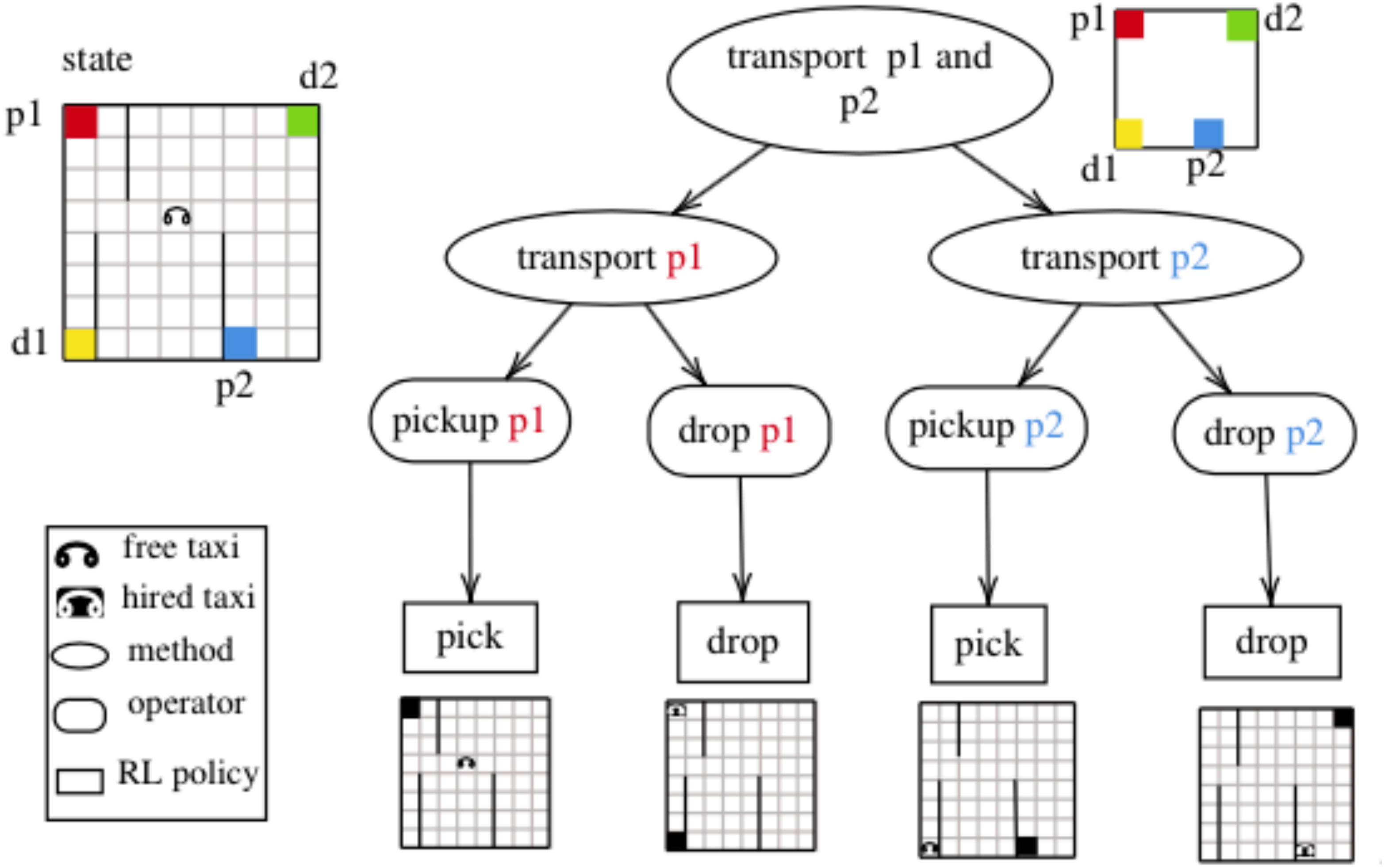
Abstractions



Planning

Execution

Influence information for Task-specific Abstraction



Influence information for Task-specific Abstraction

Given

State: $\{\mathrm{at(p1,r), taxi\text{-}at(13), dest(p1,y), \neg\, at\text{-}dest(p1), \neg\, in\text{-}taxi(p1), at(p2,b), dest(p2,g), \neg\, at\text{-}dest(p2), \neg\, in\text{-}taxi(p1)}\}$

subtask: $\langle\ \mathrm{pickup}(P), \{P/p1, L/r\}\ \rangle$

D-FOCI:

$\{\text{taxi-at(L1), move(Dir)}\} \xrightarrow{+1} \text{taxi-at(L2)}$

$\{\text{taxi-at(L1), move(Dir)}\} \longrightarrow R$

pickup(P):

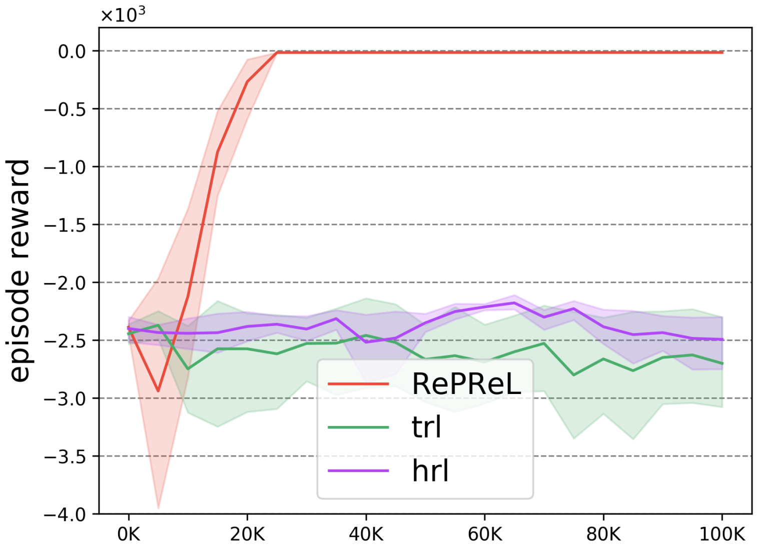
$\{\text{taxi-at(L1), at(P, L), in-taxi(P)}\} \xrightarrow{+1} \text{in-taxi(P)}$

pickup(P): $\text{in-taxi(P)} \longrightarrow R_o$

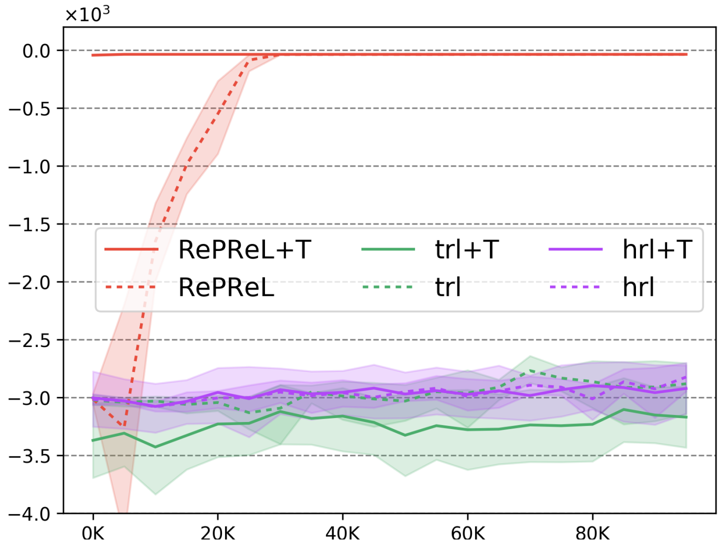
Get

Abstract state: $\{\mathrm{at(p1,r), taxi\text{-}at(13), \neg\, in\text{-}taxi(p1), move(Dir)}\}$

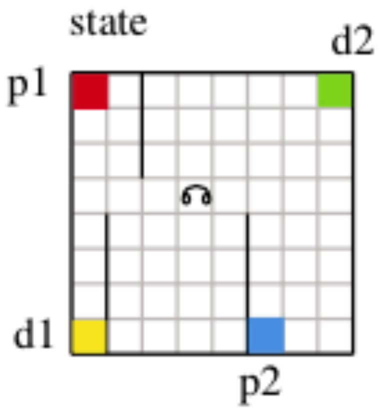
Experiments



transport p1



transport p1 and p2



Sample efficiency
Transfer across task
and
Generalization across objects

(Our) 3 Steps to HAAI

Effective Learning

Advice taking

Active advice seeking

Close the "**loop**"

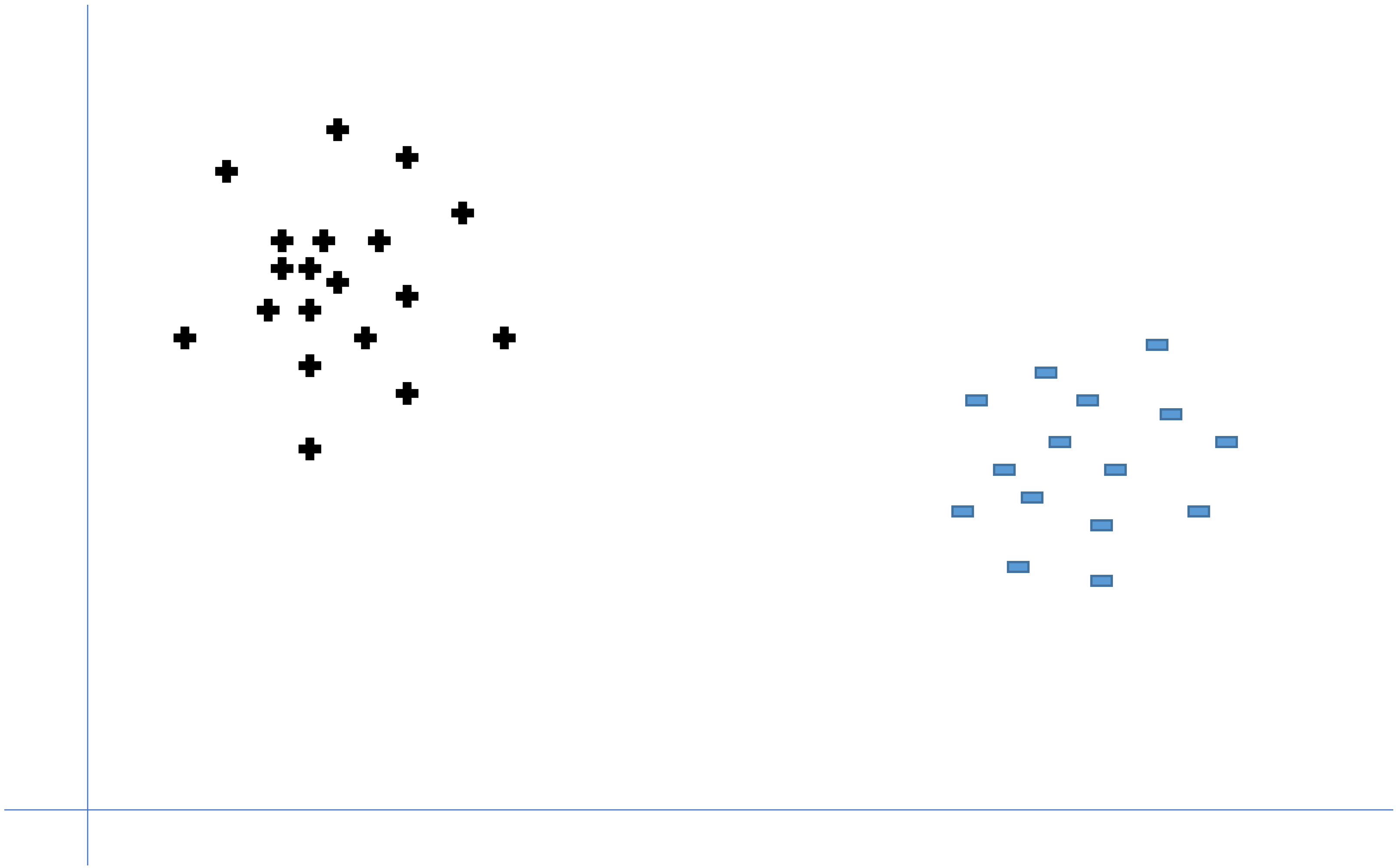
Knows what it knows

Asks what it does **not** know

Student-teacher interaction

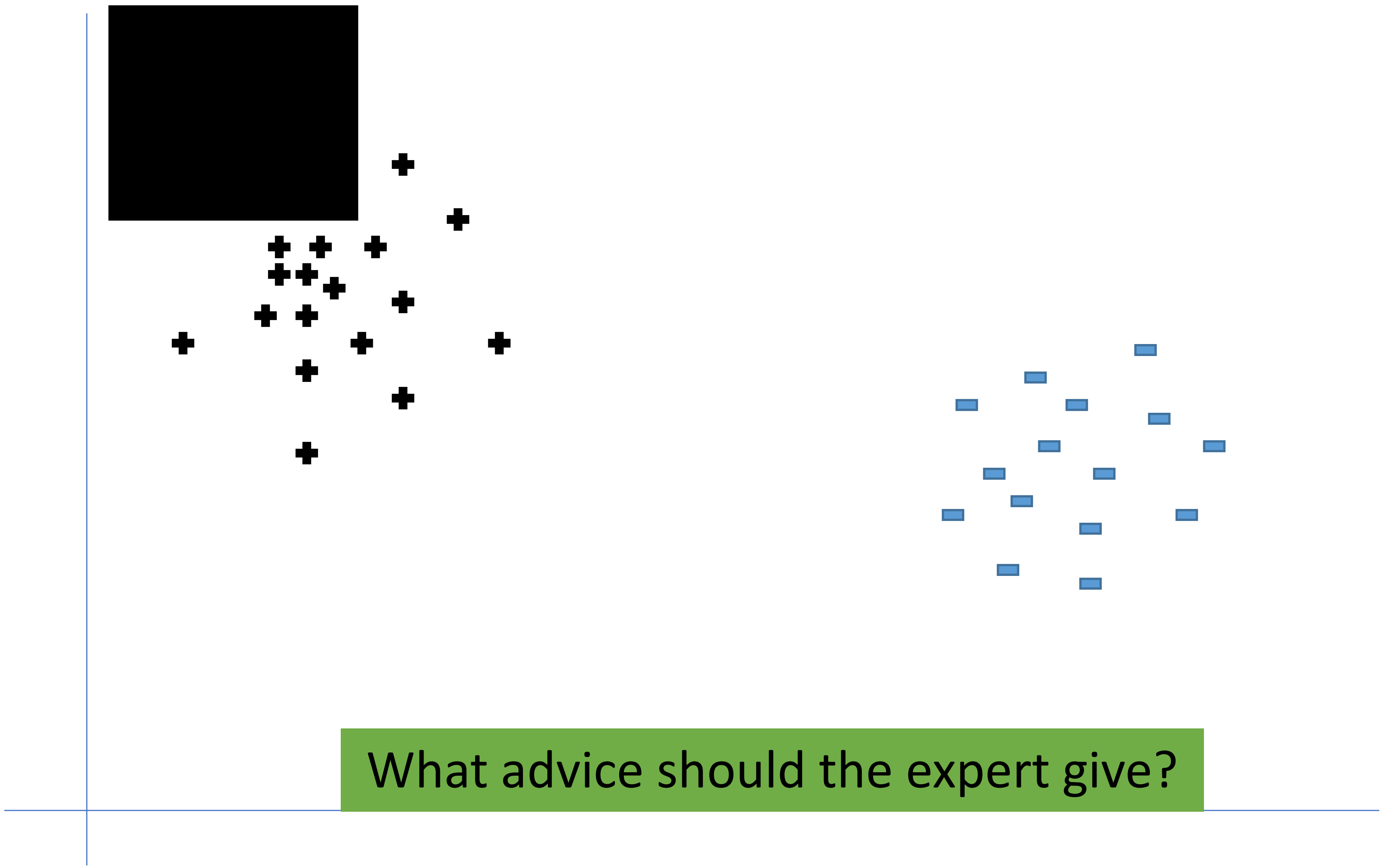
Teach the human!

Knowledge-Based Learning



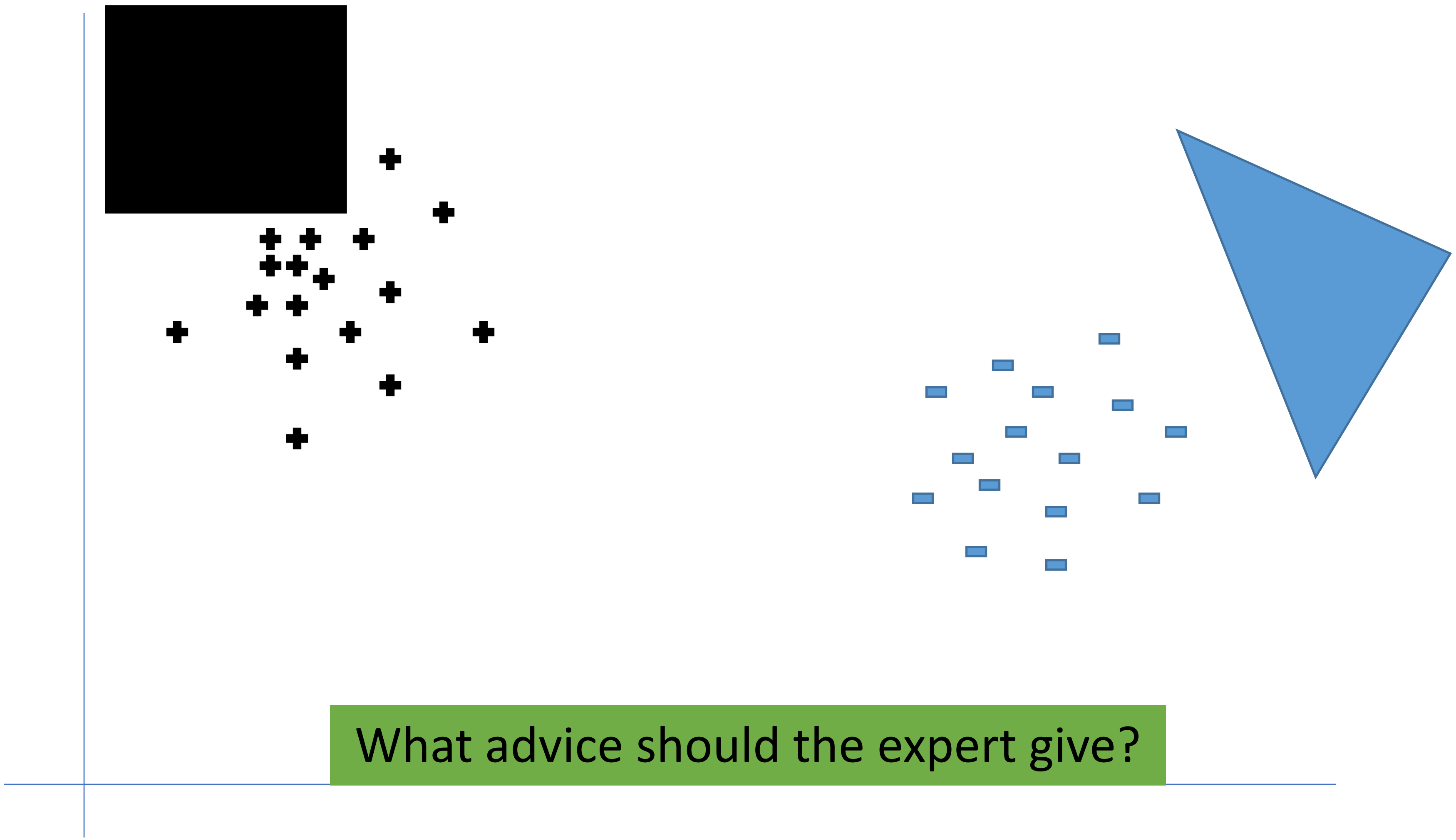
Fung et al. 02, Boutilier 02, Torrey et al. 05, Wiewiora et al. 03; Mangasarian et al 04, Kunapuli et al. 10,13

Knowledge-Based Learning



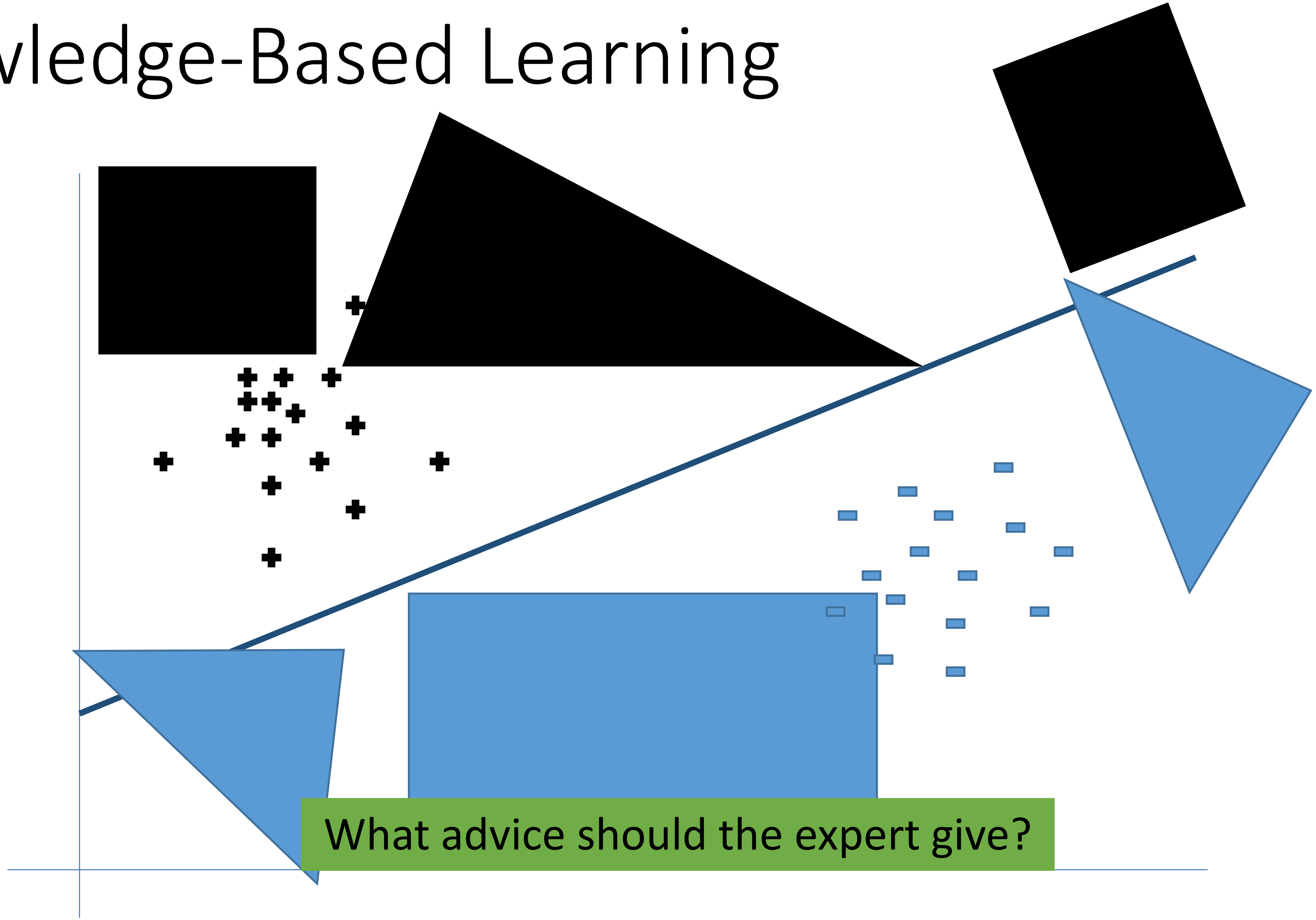
Fung et al. 02, Boutilier 02, Torrey et al. 05, Wiewiora et al. 03; Mangasarian et al 04, Kunapuli et al. 10,13

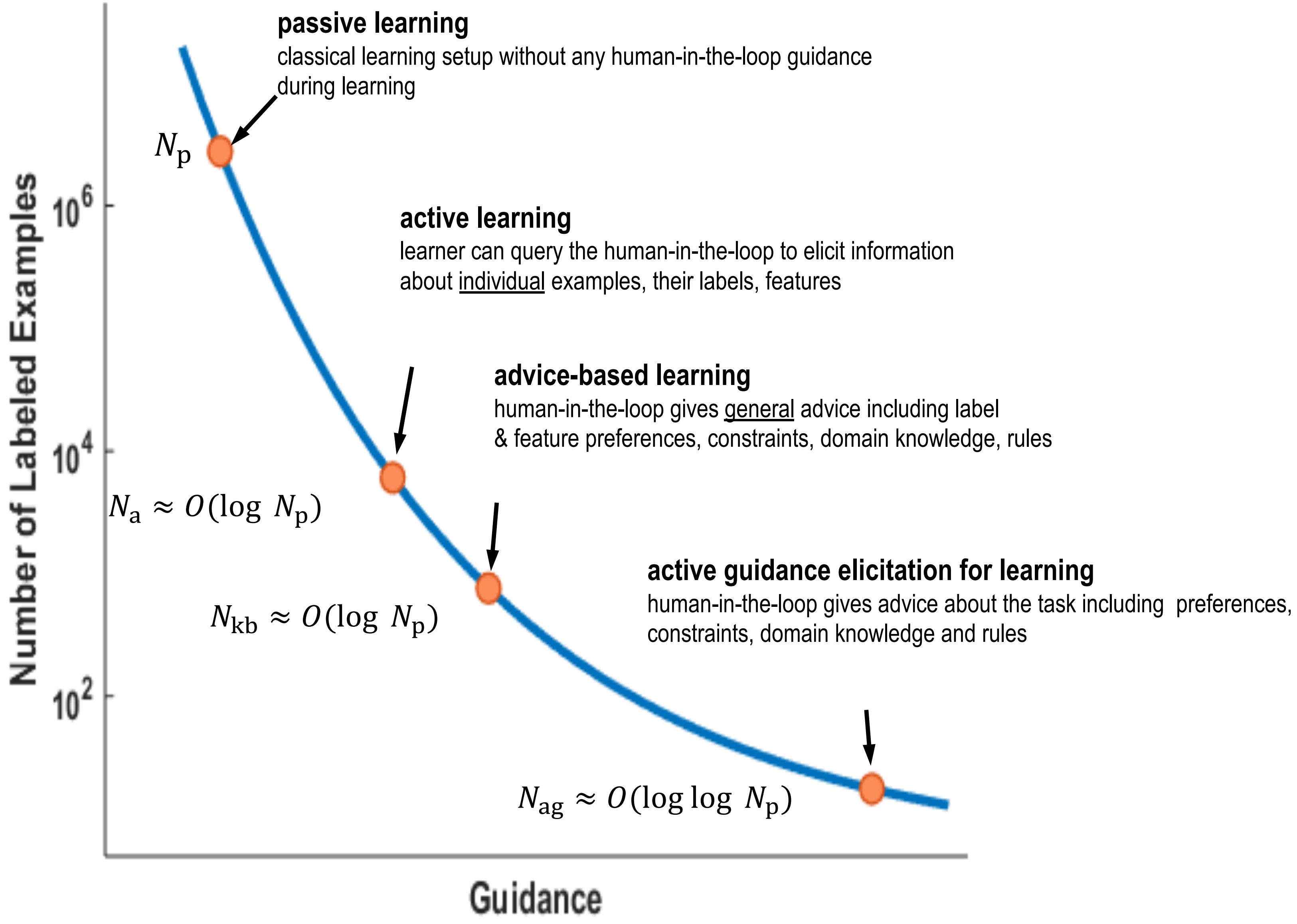
Knowledge-Based Learning



Fung et al. 02, Boutilier 02, Torrey et al. 05, Wiewiora et al. 03; Mangasarian et al 04, Kunapuli et al. 10,13

Knowledge-Based Learning

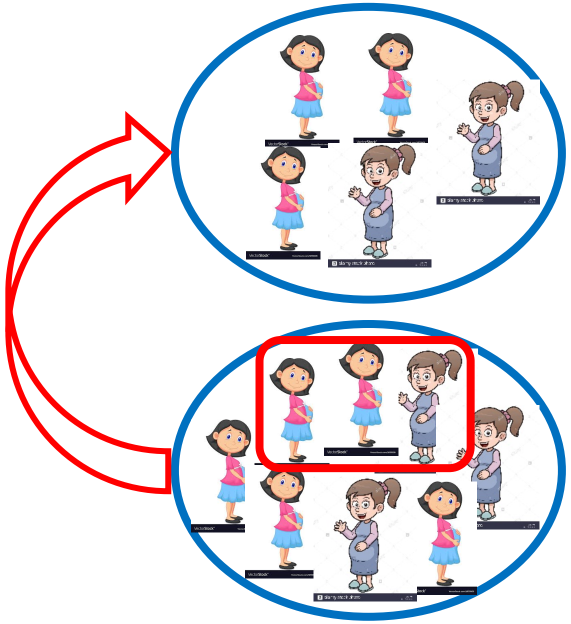


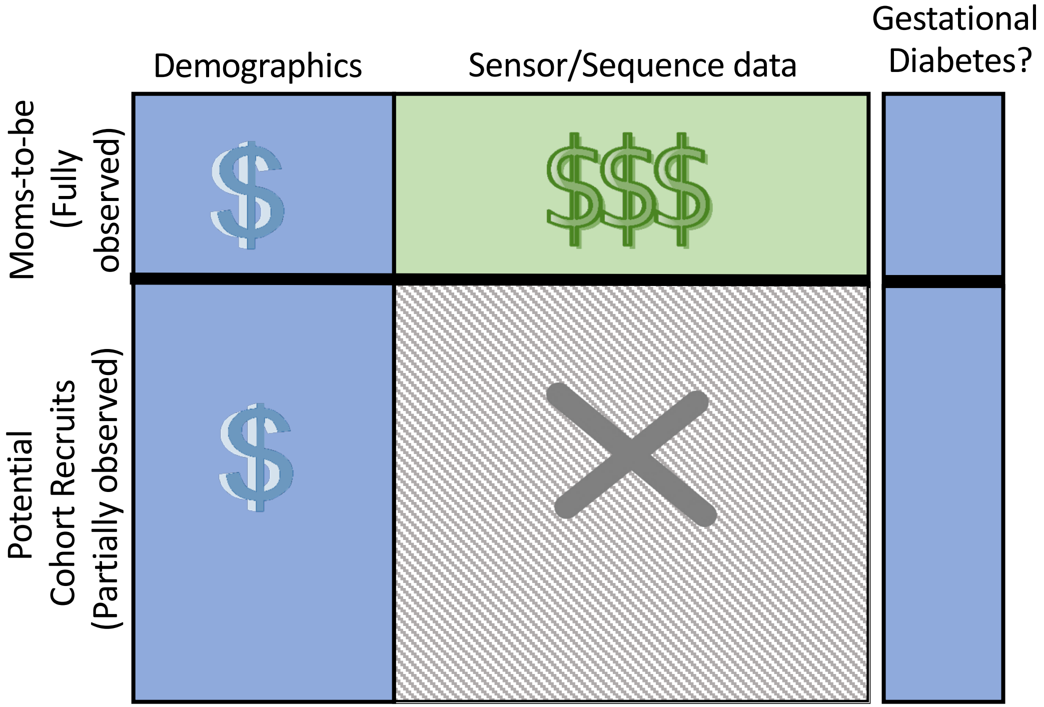


passive learning
classical learning setup without any human-in-the-loop guidance during learning
N_p
active learning
learner can query the human-in-the-loop to elicit information about individual examples, their labels, features
advice-based learning
human-in-the-loop gives general advice including label & feature preferences, constraints, domain knowledge, rules
$N_a \approx O(\log N_p)$
$N_{kb} \approx O(\log N_p)$
active guidance elicitation for learning
human-in-the-loop gives advice about the task including preferences, constraints, domain knowledge and rules
$N_{ag} \approx O(\log\log N_p)$
10^6
10^4
10^2
Number of Labeled Examples
Guidance

Active Feature Elicitation

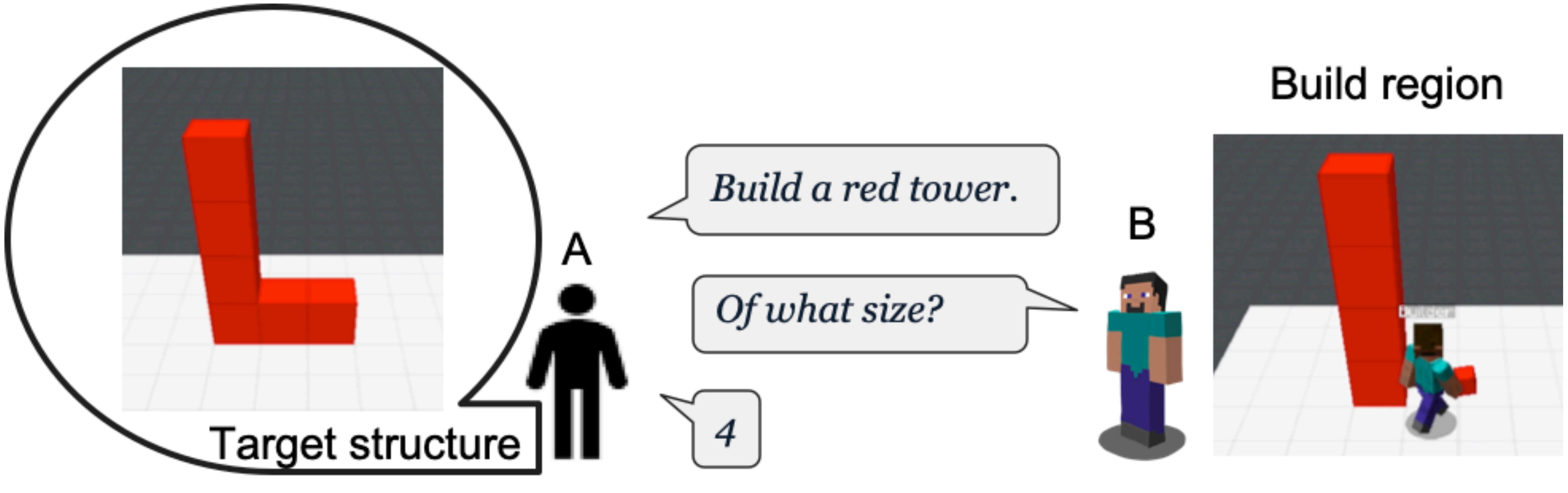
Whom should we recruit?





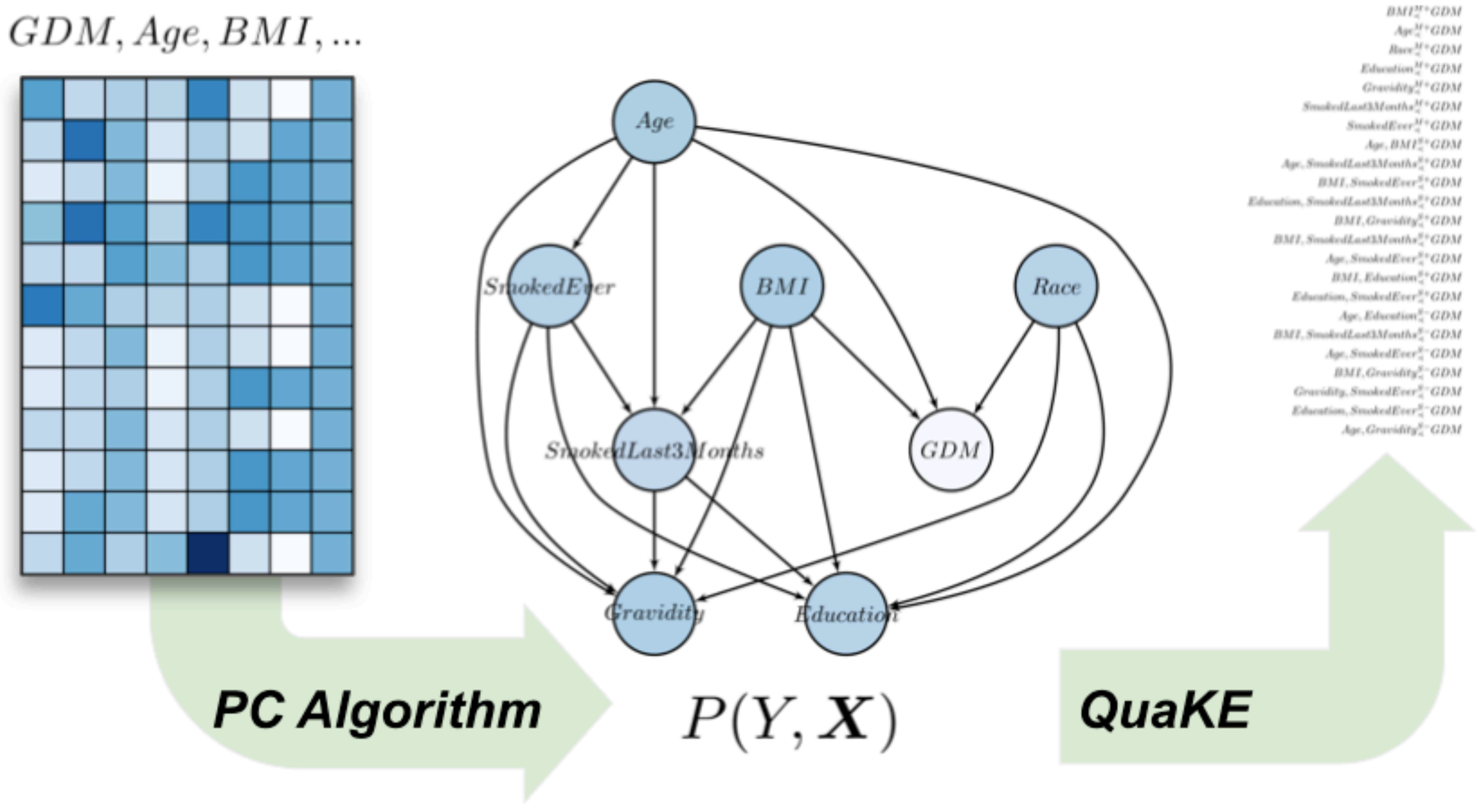
Natarajan et al. IJCAI '18; Das et al 2021

Collaborative Problem Solving



Achieve Generalization in One shot Learning

Extract Knowledge from Data



Miles to go before we sleep!

- **Ensuring Human Trust** – explain decisions and solicit feedback *Always include humans in decision-making*
- **Enabling Machine Fairness** – avoid bias in learning *(social/economic/religious)* impossible to maximize all notions of fairness
- **Handling Ethical Issues** – white lies to make us eat healthy vs negotiation for profit
- **Data vs Knowledge** – what if the evidence is contrary to human perception?
- **Optimal/Rational vs. Human-like**

Questions?

https://starling.utdallas.edu

@starling_lab
@Sriraam_utd

Thanks!