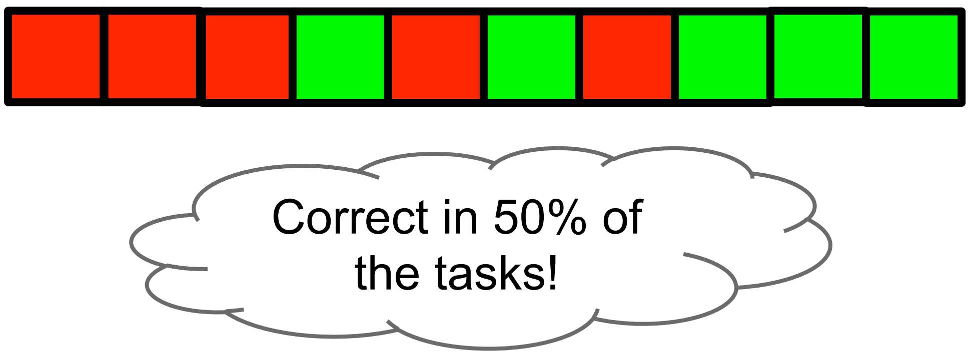
Predicting Crowd Work Quality under Monetary Interventions

Ming Yin, Yiling Chen

HCOMP-16 • Austin, TX • October 31, 2016

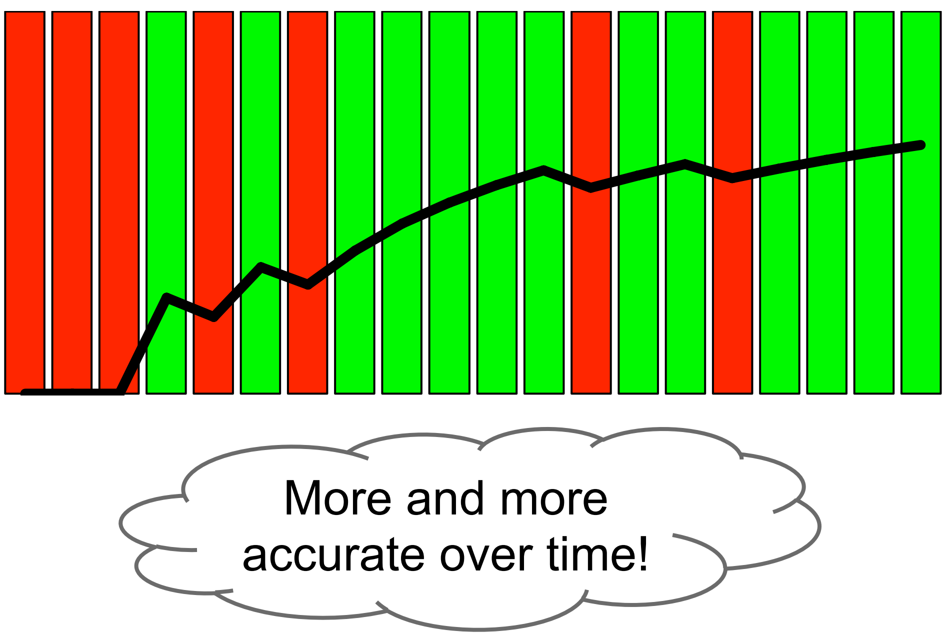
HARVARD
School of Engineering and Applied Sciences

Modeling Worker Performance in Crowdsourcing



Accuracy / Error Rate

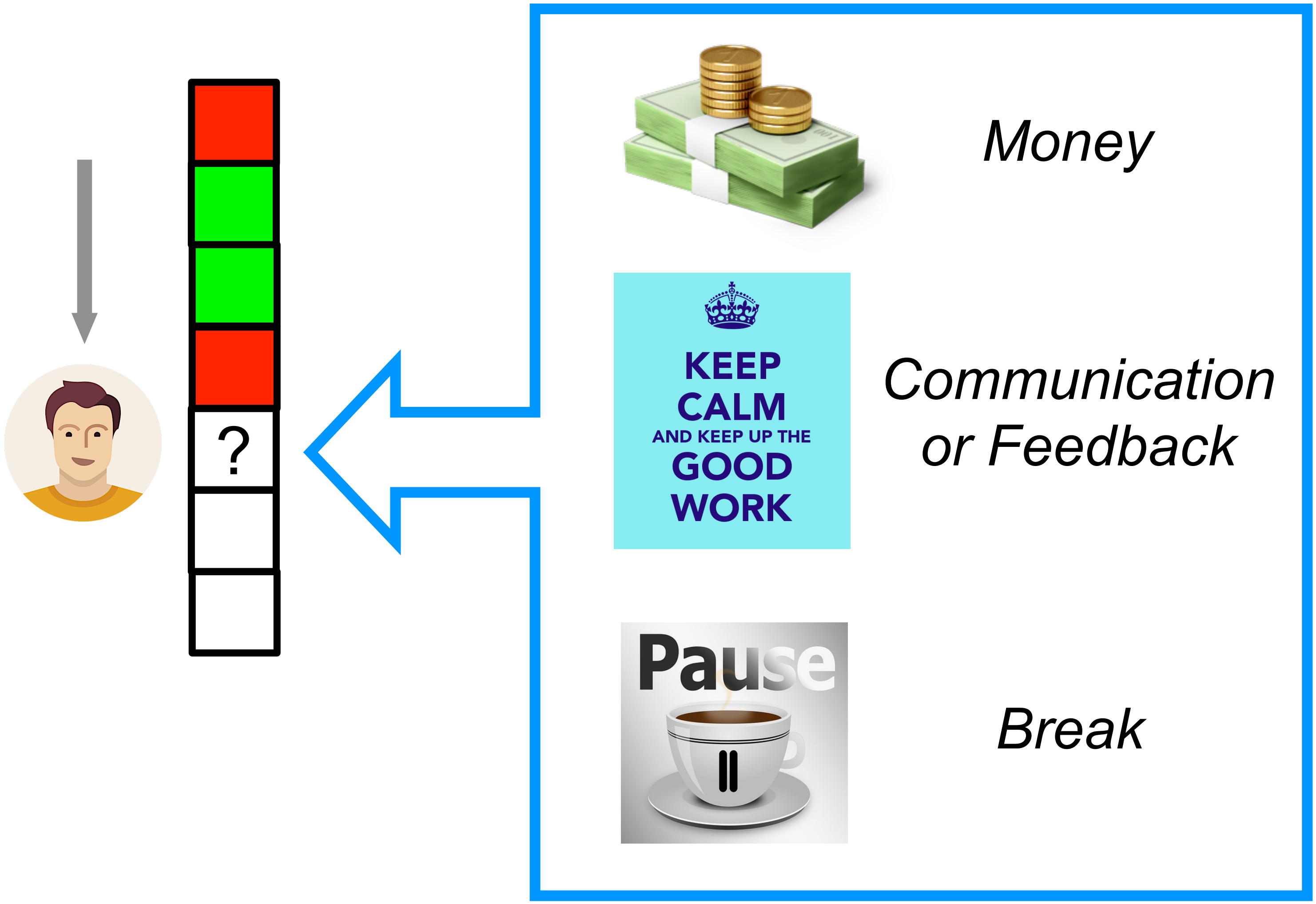
(e.g. Whitehill et. al. 2009)



Temporal Pattern

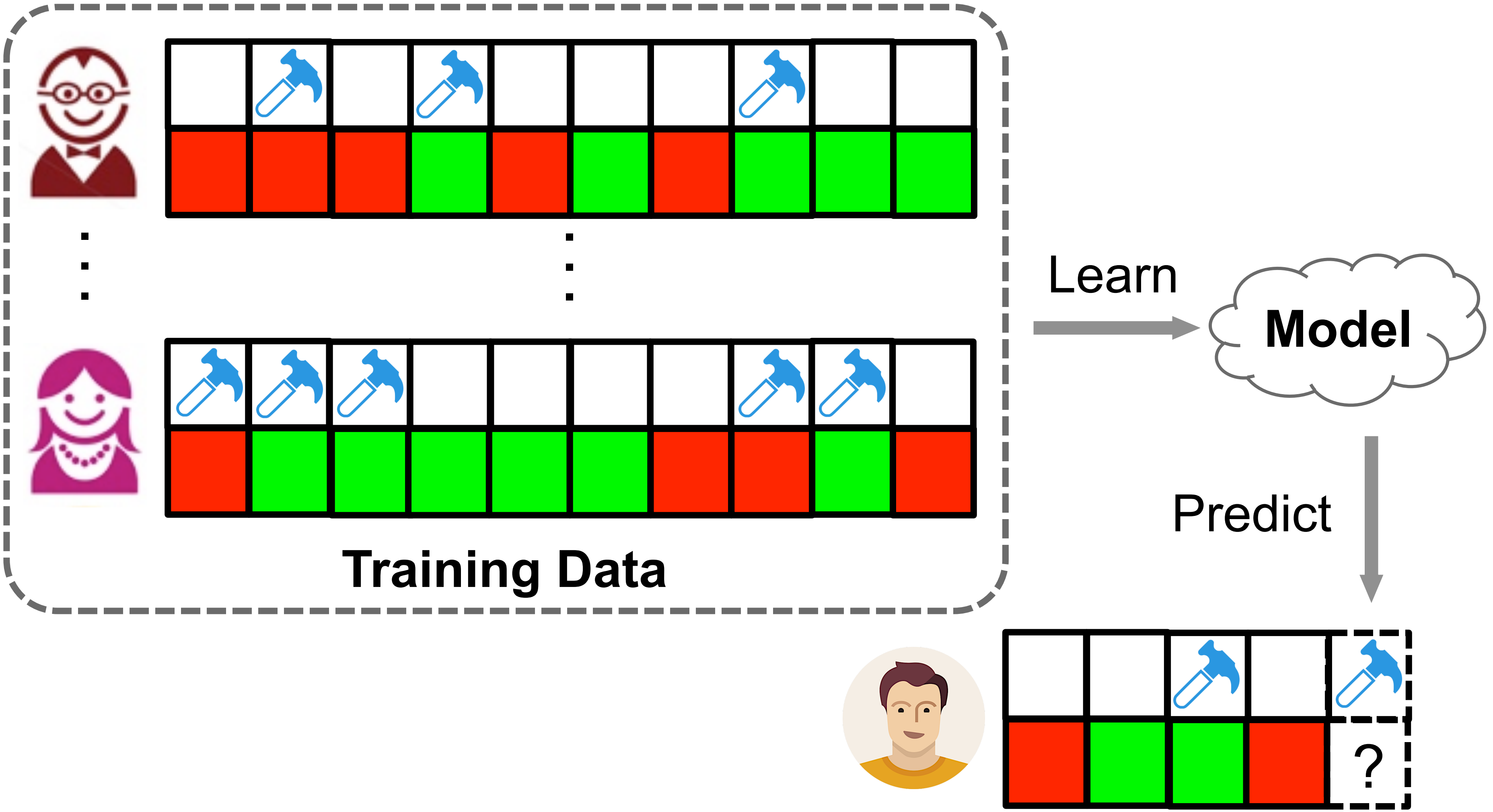
(e.g. Jung, Park & Lease. 2014)

Modeling Worker Performance under Interventions



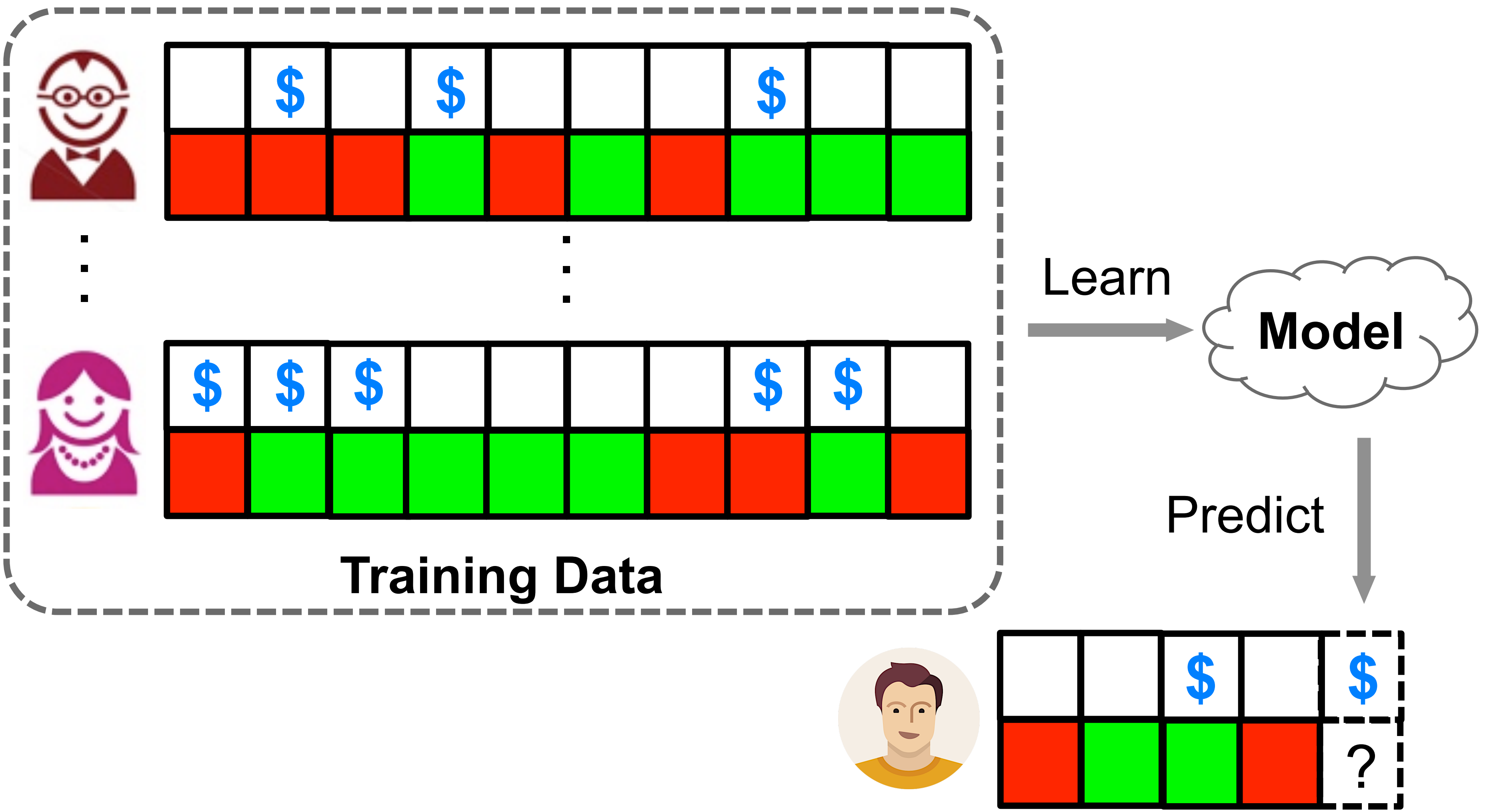
How to capture worker performance under interventions?

A Prediction Perspective



Categorical time series prediction ***with exogenous inputs***!

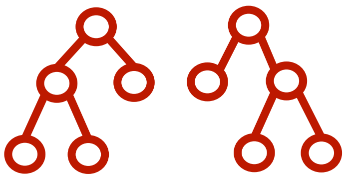
A Prediction Perspective



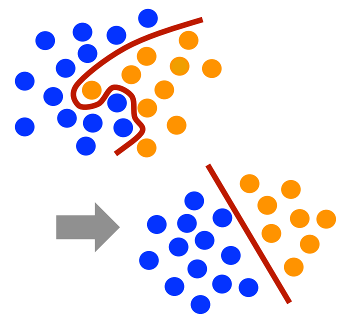
Focus on ***monetary intervention*** in this talk!

An Empirical Comparison

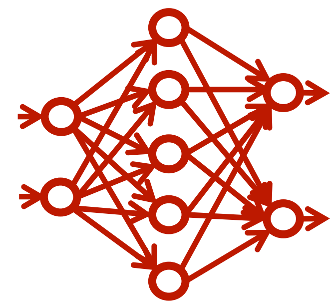
Supervised Learning Models



Random Forests



SVM



Neural Network

Autoregressive Models

$$y_i^t = I_t y_i^{t-D_t} + (1-I_t)e_t$$

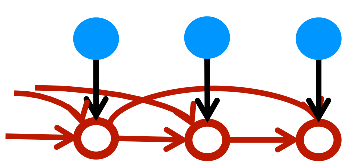
DARX

$$z_i^t = c + \sum_{j=1}^{p} \phi_j z_i^{t-j} + \sum_{j=0}^{q-1} \theta_j a_i^{t-j} + \epsilon_i^t$$

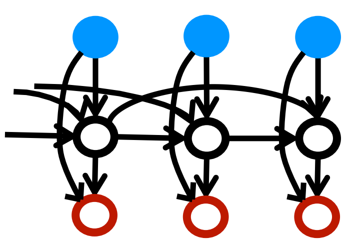
$$Pr(y_i^t = 1) = \frac{1}{1+e^{-z_i^t}}$$

LARX

Markov Models

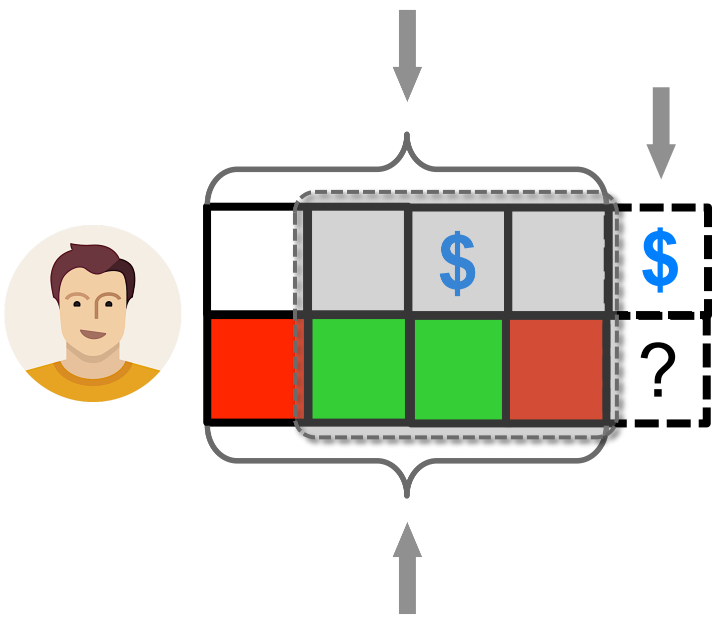


Controlled MC



IOHMM

Supervised Learning Models: Features



Current Intervention Level

Average Intervention Level

Average Performance

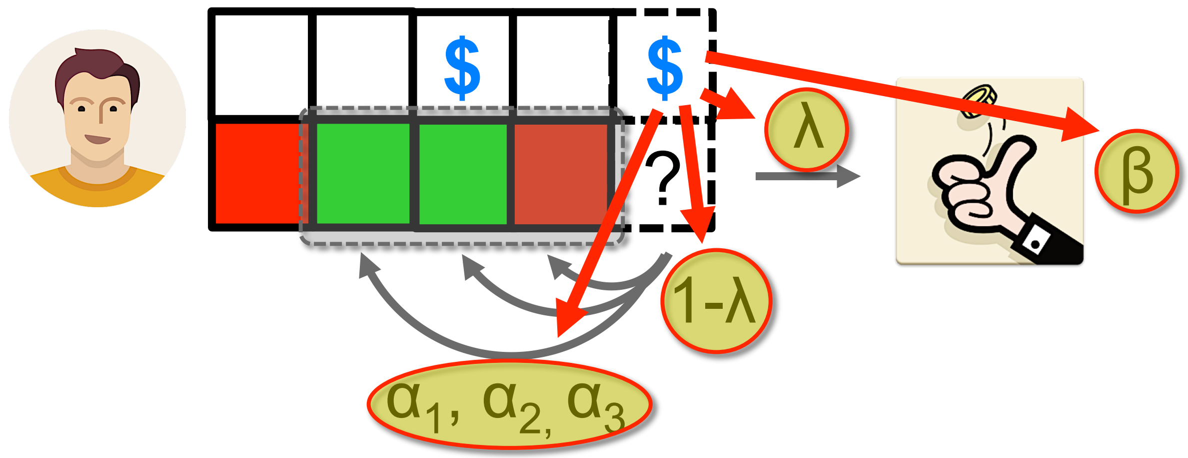
Within a history window of size *L*:

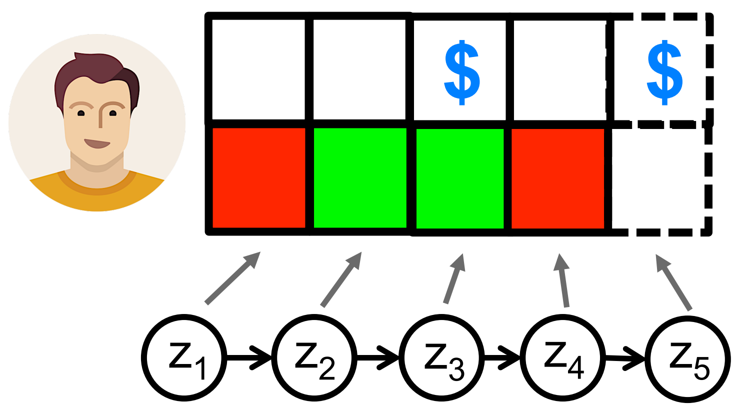
- Historical Intervention Levels
- Historical Performance
- Historical Intervention Changes
- Historical Performance Changes

Random Forests, SVM, Neural Network

Autoregressive Models: Incorporating Exogenous Inputs

DARX: Extended from DAR [Jacobs and Lewis 1983]



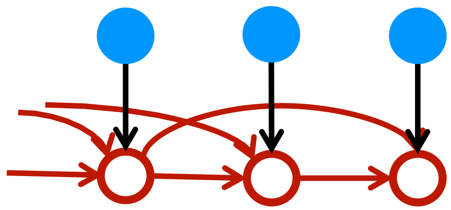


$$z_t = c + \phi z_{t-1} + \varepsilon_t + \theta a_t$$

LARX: Extended from LAR [Jung, Park and Lease 2014]

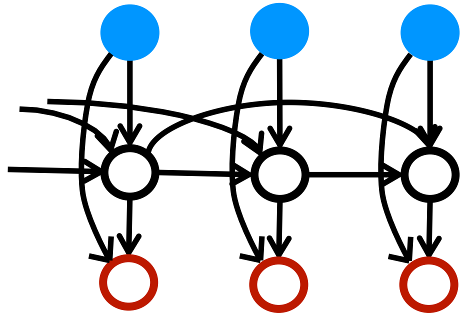
Markov Models: Application

Controlled Markov Chain



Action: Intervention
State: Worker Performance

Input-Output Hidden Markov Model



Inputs: Intervention
Outputs: Worker Performance

Evaluation Datasets

Word Puzzle

300 workers
9 tasks in a session
37% bonus tasks
76.8% high-quality

Butterfly Classification

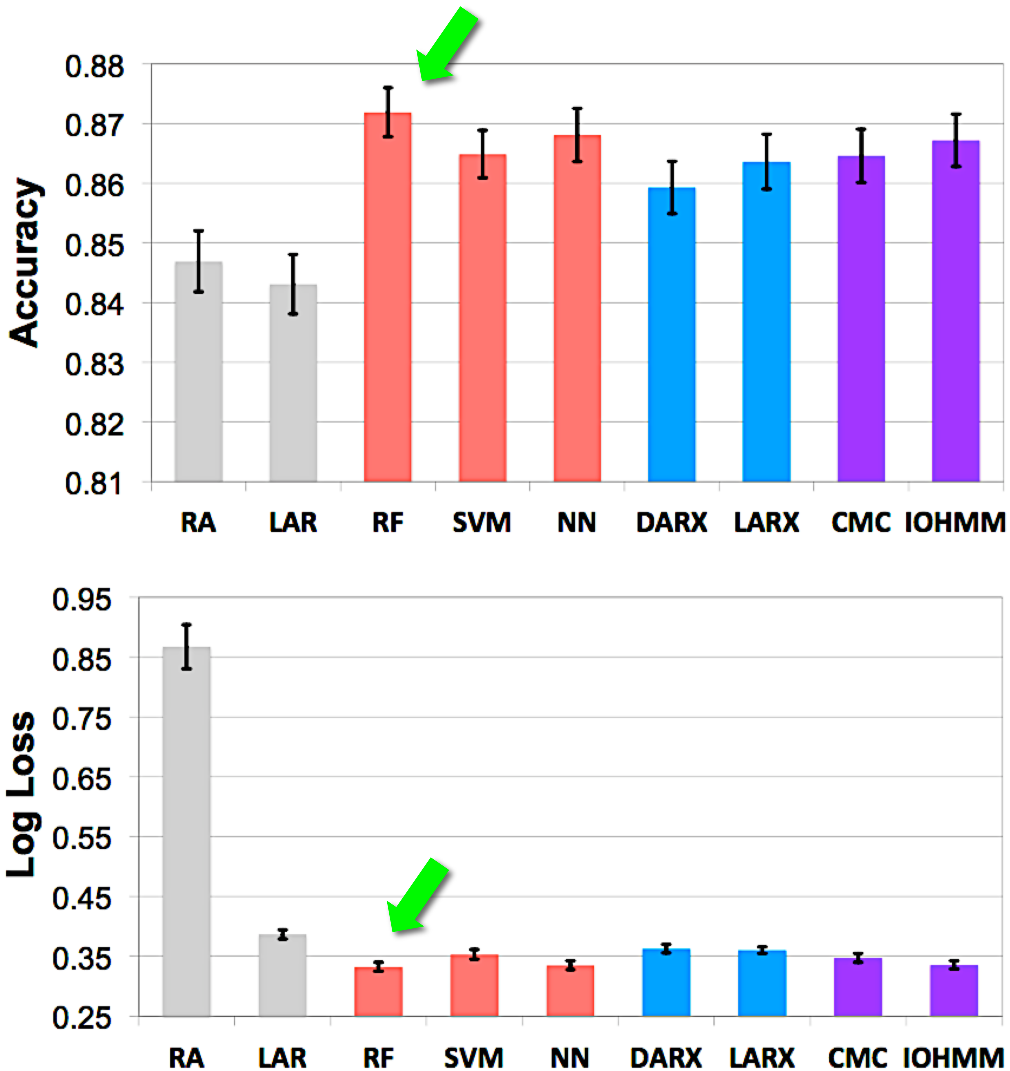
220 workers
10 tasks in a session
29% bonus tasks
55.5% high-quality

Proofreading

80 workers
10 tasks in a session
49% bonus tasks
63.4% high-quality

80% Training | 20% Testing

Baselines: Running Accuracy, LAR

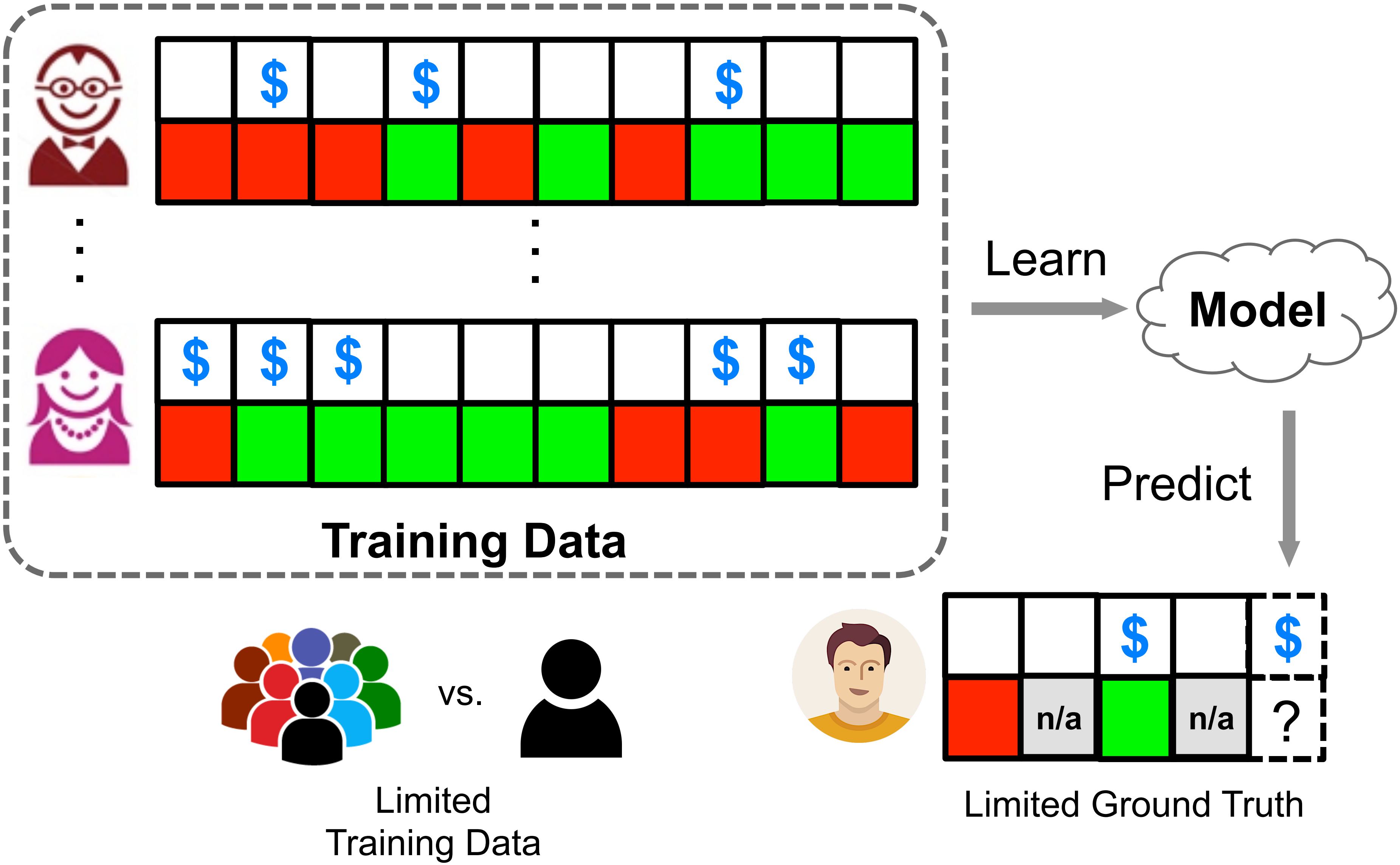


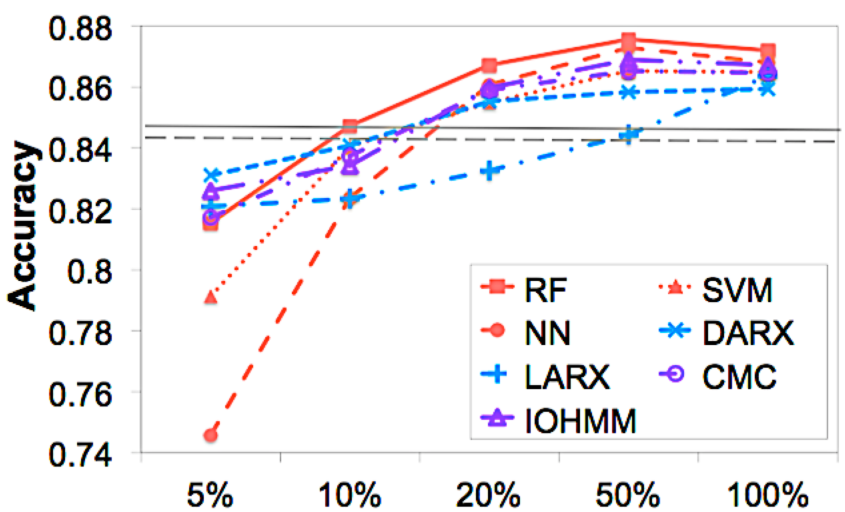
It is ***necessary*** to model the impact of monetary interventions on worker performance.

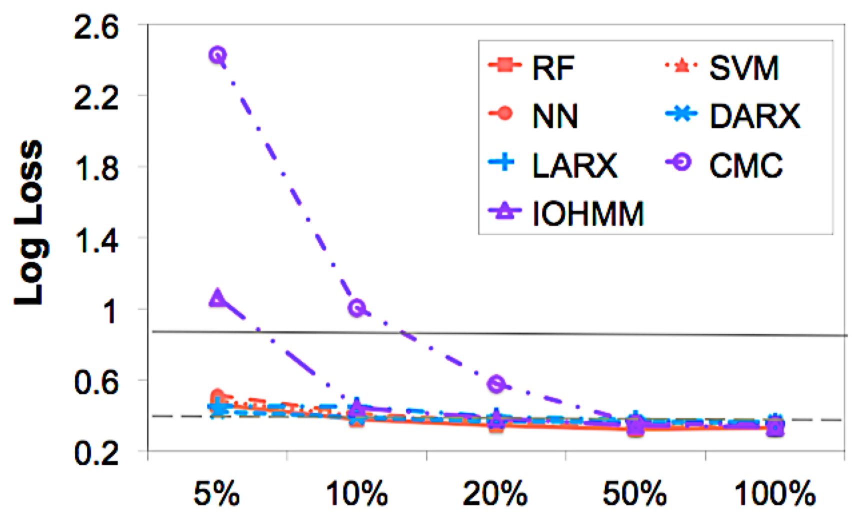
The ***random forest*** model outperform other prediction models! ***(Best model for 7 out of 9 comparisons!)***

Predictive features: average performance; average intervention level.

More Realistic Scenarios





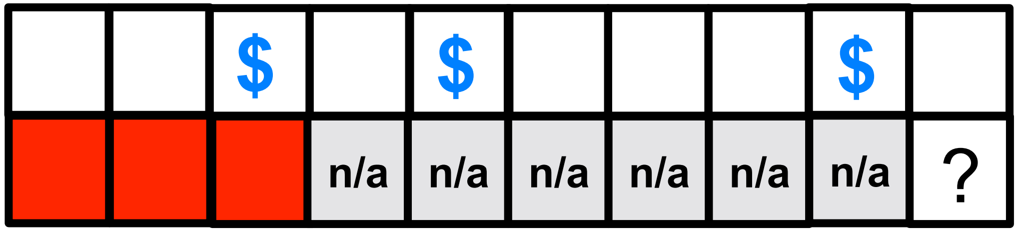


The random forest model is ***relatively robust against limited training data***.

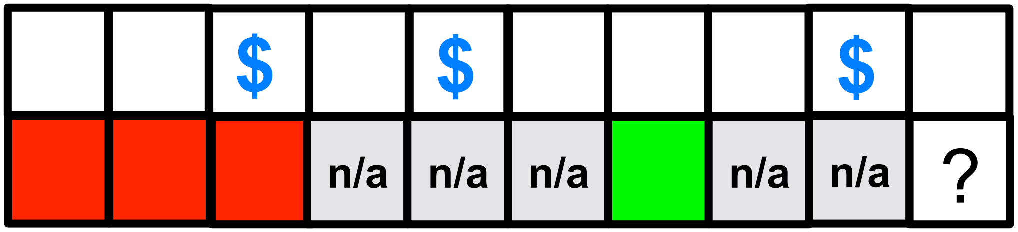
80% Training

20% Testing

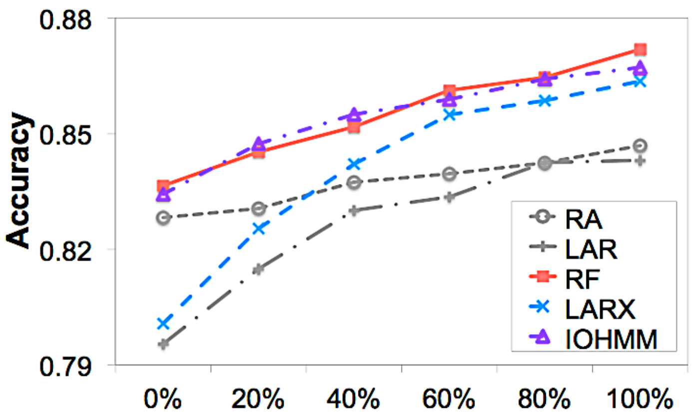
0% Verification

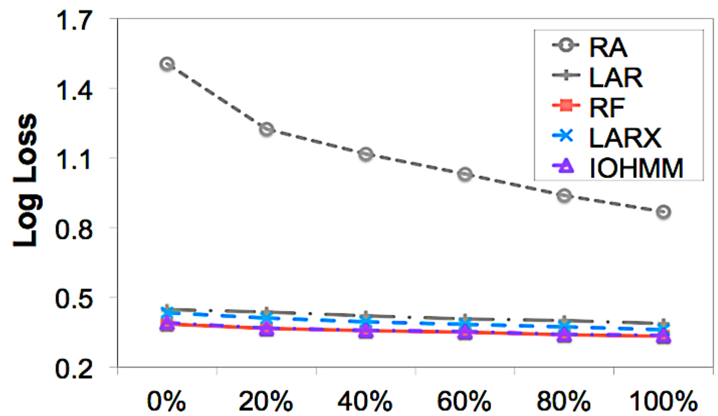


20% Verification



40%, 60%, 80%...





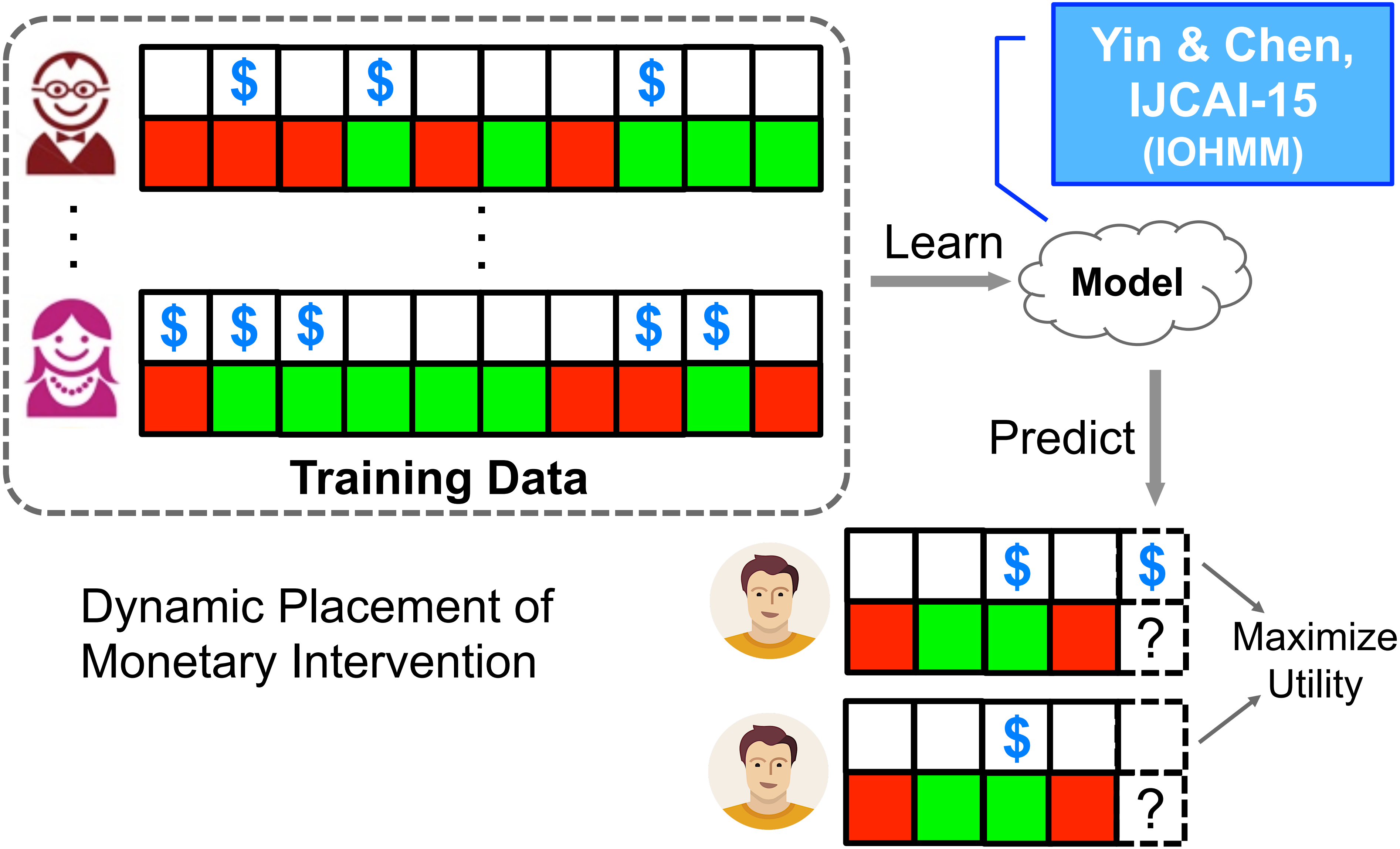
The random forest model (and the IOHMM model) is ***relatively robust against limited access to ground truth***.

Summary

The ***random forest model*** can be a good model to use in practice to predict crowd work quality under monetary interventions, because of its:

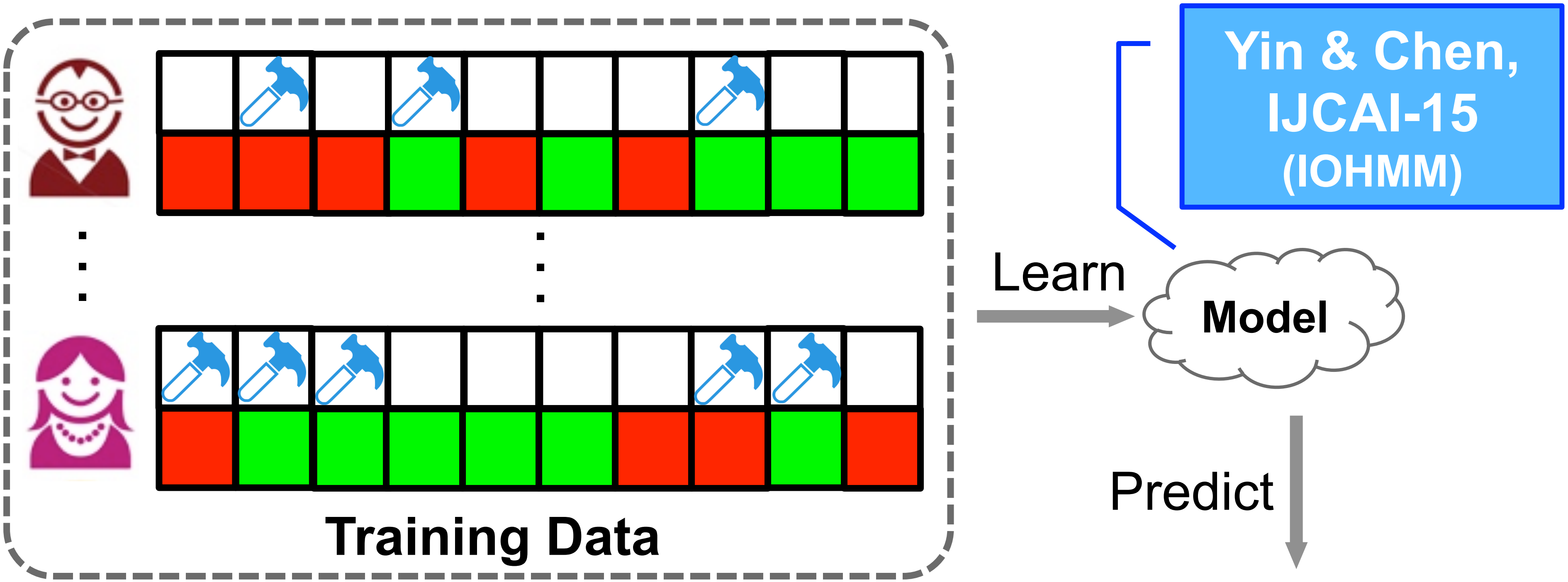
- Accurate predictions with high confidence across different types of tasks
- Robustness against limited training data
- Robustness against limited ground truth

Future Directions



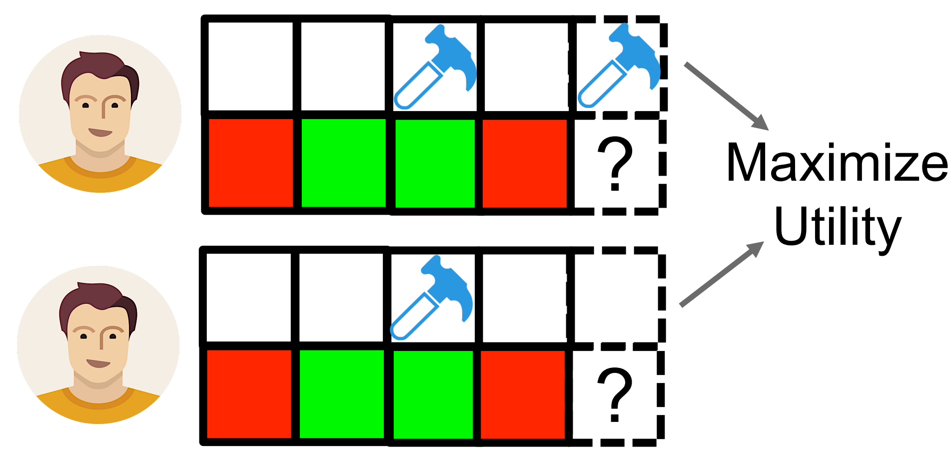
Dynamic Placement of Monetary Intervention

Future Directions



Dynamic Placement of Monetary Intervention

Performance Modeling under Other Interventions



Thank you!