Clinical Decision Making under Uncertainity

A Bootstrapped Counterfactual Inference Approach

Anirudh Choudhary, Hang Wu and May Wang

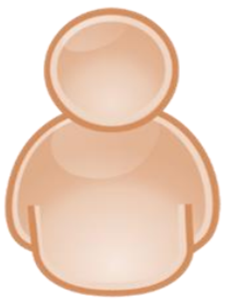
27 Feb 2020

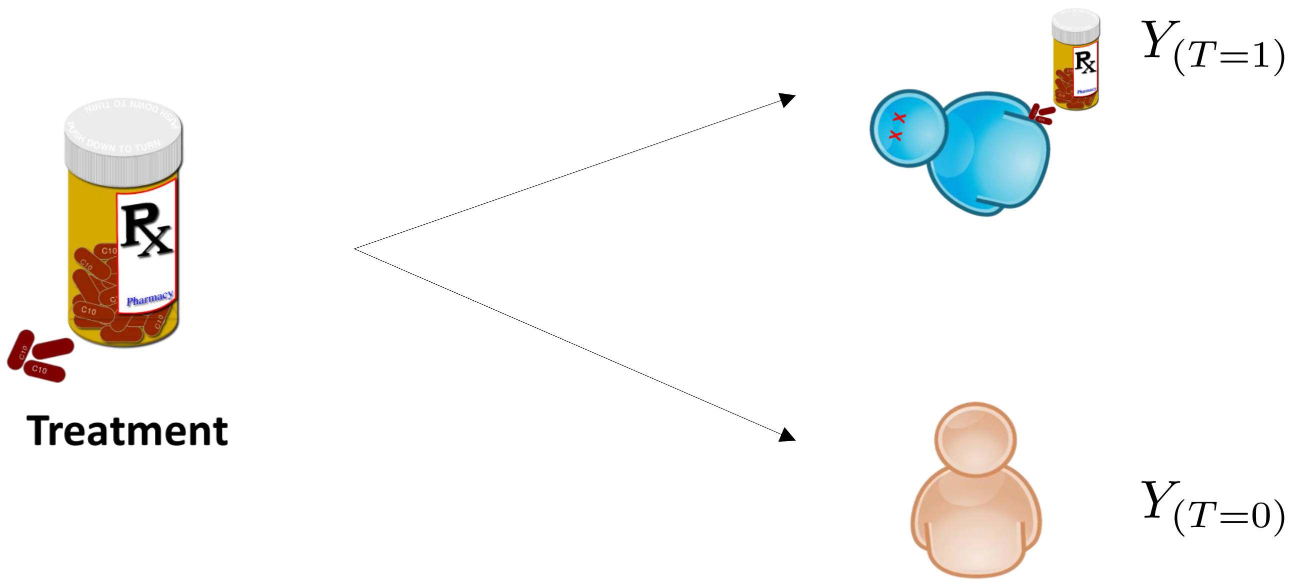
CSL Student Conference

Georgia Institute of Technology
Bio - MIBLab
BioMedical Informatics & Bio-imaging Laboratory
EMORY UNIVERSITY

Clinical Decision-Making

Patient ‘P’ comes in with hypertension → The doctor prescribes a medication → The patient responds to the treatment.





Clinical Decision-Making

- Clinical Decision Support Systems (CDSS) learn policy for choosing targeted treatments for patients
- However, this is not a typical supervised learning problem

Person	T	$Y_{(T=1)}$	$Y_{(T=0)}$
P1	1	0.5	0.3
P2	0	0.9	0.5
P3	1	0.2	0.3
P4	1	0.5	0.5
P5	0	0.3	0.1
P6	0	0.6	0.5

Counterfactual outcome is not observed!!

- Only one of all possible outcomes is observed
- Loss function unknown at training time

Counterfactual Learning

Reason about a world that does not exist

Clinical Records

Clinician’s decision policy

Counterfactual World

Ideal decision policy

- What if the patient was put on ventilation early?
- What if the patient was admitted longer in ICU?
- What if I gave a drug to a patient?

Contextual Bandits

At each round ‘t’ ($1 \leq t \leq T$)

The world produces some context (x_t) → **The learner chooses an action (a_t)** → **The world reacts with reward (r_t)**

Patient P with hypertension (symptoms, medical history, test results etc.)

a1. Calcium channel blocker
a2. ACE inhibitor

Patient Survives/Recovers (r = 1)
Health deteriorates (r = 0)

- x_t : Drawn i.i.d from unknown P(X)
- a_t : Selected by existing system following policy π_0: $X \rightarrow A$
- r_t : Feedback from unknown function r_t : $X * A \rightarrow R$

Goal: Learn a good policy π for choosing actions given context
(maximize cumulative reward for all patients)

Offline Learning

Given observation data for ‘n’ patients collected under a policy π_0

- $D = (x_1,a_1,r_1),\ldots\ldots\ ,\ (x_n,a_n,r_n)$

Clinical Records

Goals

Evaluation: Estimate reward $R(\pi)$ of an alternate policy π offline

Optimization: Find new policy $\pi(\theta)$ that improves performance over π_0

- Directly testing the policy π in real-world (online) is not possible
- Policy learning depends on how confidently we can evaluate π given π_0

Outline

- Off-Policy Evaluation
- Motivation
- Proposed Method
- Experiments
- Results

Off-Policy Evaluation

- Inverse Propensity Score (IPS) Estimator

Evaluation Policy

$$R_{ips}(\pi) = \frac{1}{n}\sum_{i=1}^{n} \frac{p(a_i|x_i)}{p_0(a_i|x_i)} r_i$$

Behavior Policy
(Generally known)

- Unbiased estimate
- Prone to high variance

$$p_0(a_i|x_i) \approx 0$$

- **p & $\mathbf{p_0}$** : Probabilities of selecting the action a_i using policies π & π_0 respectively
- p_0 also known as 'propensity score'
- IPS weighs unlikely actions in observed data more compared to likely actions

Motivation

- In clinical settings, propensity score is typically unknown and is imputed by training a model

Evaluation Policy

$$R_{ips}(\pi) = \frac{1}{n}\sum_{i=1}^{n} \frac{p(a_i|x_i)}{p_0(a_i|x_i)} r_i$$

- Unbiased estimate
- Prone to high variance when $p_0(a_i|x_i) \approx 0$

Behavior Policy (Clinician)

Challenge:
- Model uncertainty (our ignorance about the correct model that generated p_0)
- Significant variability in patient-specific predictions and optimal decisions
- Uncertainty in modeling p_0 introduces bias & variance in reward estimates

Uncertainty of Predictive Models

- Where does uncertainty arise from in machine learning?

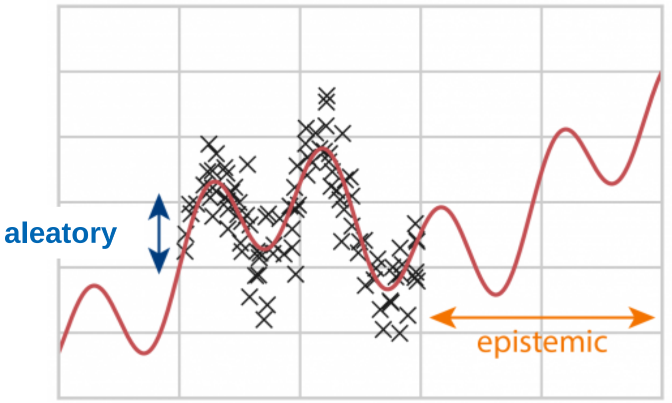
Expected-Loss Minimization

$$min_\theta\left[E_{\mathbf{x},a}[l(f(\mathbf{x};\theta),a)]\right] \approx \frac{1}{n}\sum_{i=1}^{n}(l(f(\mathbf{x}_i;\theta),a_i)$$

Data uncertainty (epistemic)

Model Uncertainty (aleatory)

Our focus



*https://www.inovex.de

- How to tackle uncertainty? - Bootstrapping

M. W. Dusenberry, D. Tran, E. Choi, J. Kemp, J. Nixon, G. Jerfel,K. Heller, and A. M. Dai, "Analyzing the Role of Model Uncertainty for Electronic Health Records." http://arxiv.org/abs/1906.03842

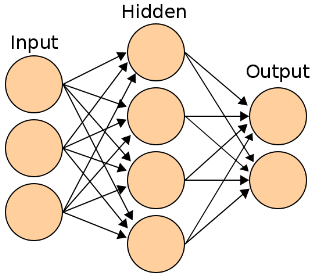
Model Uncertainty

- Multiple ways to characterize uncertainty in neural networks

Deep Ensemble

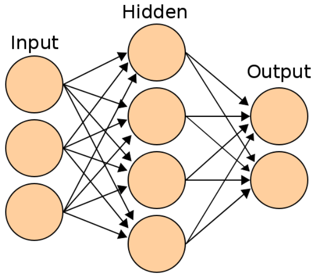
Weights as Point Estimates

Initialize 'M' networks with varying seeds



NN_1

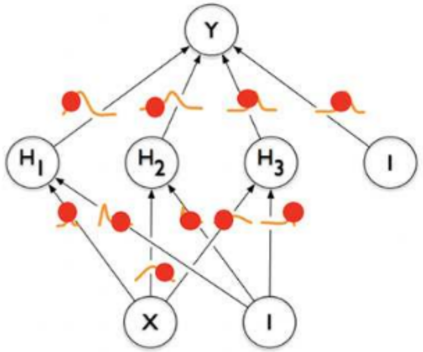
....



NN_M

Bayesian Neural Networks

Weights represented by probability distribution



Sample M networks during test time

$$\hat{Y} = \frac{1}{M} \sum_{M}^{m=1} NN_m(x) \qquad p(y|x) = \frac{1}{M} \sum_{1}^{m} p(y|x, w_i)$$

1. Lakshminarayanan, Balaji, Alexander Pritzel, and Charles Blundell. "Simple and scalable predictive uncertainty estimation using deep ensembles." NIPS. 2017.
2. Blundell, Charles, et al. "Weight uncertainty in neural networks." arXiv preprint (2015)

Proposed Method

We propose **bootstrapping-based** counterfactual inference framework

Bootstrap 'M' models to impute clinician's action propensity

'M' propensity scores ($\pi_0[1]$, $\pi_0[2]$ …… $\pi_0[M]$)

Incorporate into vanilla offline policy estimators (IPS[1], Doubly-Robust[2])

Policy Evaluation

$$\hat{R}_{inv} = \frac{1}{M}\sum_{i=1}^{M} E[\frac{\pi(a|x)}{\pi_0[i](a|x)} r]$$

$$\hat{R}_{avg} = E[\frac{\pi(a|x)}{\frac{1}{M}\sum_{i=1}^{M}\pi_o[i](a|x)} r]$$

Policy Learning

$$\pi^*_{inv} = argmax\frac{1}{M}\sum_{i=1}^{M} E[\frac{\pi}{\pi_0[i]} r]$$

$$\pi^*_{avg} = argmax E[\frac{\pi}{\frac{1}{M}\sum_{i=1}^{M}\pi_0[i]} r]$$

$$\pi^*_{max} = argmax_i E[\frac{\pi}{\pi_0[i]} r]$$

1. Strehl, Alex, et al. "Learning from logged implicit exploration data." NIPS 2010.
2. Dudík, Miroslav, et al. "Doubly robust policy evaluation and optimization." Statistical Science 29.4 (2014): 485-511.

Experiments

Clinical Setting

Warfarin Dosing

Warfarin is a widely-prescribed oral anticoagulant agent

Challenges

Therapeutic dosage varies widely across patients; incorrect dose leads to adverse side effects

Physicians currently follow fixed-dosage strategy (base dosage followed by adjustments)

PharmaGKB dataset
~5300 patients
Demographic, physiological & genotype features with **ideal dosage** for each patient

https://www.pharmgkb.org/page/iwpc

Warfarin Dosing

- **Action Space**: Discretize therapeutic dose into low, medium and high
- **Policy Task**: Predict correct therapeutic dosage for each patient

However, we have access to counterfactuals in the original dataset !!

Warfarin Dataset

Person	X	$Y_{(low)}$	$Y_{(med)}$	$Y_{(high)}$
P1	X1	0	1	0
P2	X2	1	0	0
P3	X3	0	1	0
P4	X4	0	0	1
P5	X5	1	0	0
P6	X6	0	0	1

Behavior Policy

→

Picks one out of low/med/high stochastically

Synthetic Bandit Dataset

Person	X	A	R
P1	X1	med	1
P2	X2	high	0
P3	X3	low	0
P4	X4	high	1
P5	X5	med	0
P6	X6	low	0

Experimental Setup

- Create bandit dataset using behavior policy (20 simulations)
  - PHARMA : Choose action using WPDA* with probability ‘p’; otherwise choose randomly
  - LR : Train logistic regression model on 5% of classification dataset
- Train a classifier on full Warfarin dataset (evaluation policy $\boldsymbol{\pi}$)
- Bootstrap 10 models for $\boldsymbol{\pi_0}$ using Bayesian NN and ensemble methods
- Evaluate $\boldsymbol{\pi}$ using proposed framework (R_{avg}, R_{inv}) and compare with vanilla IPS and DR estimators
- Learn $\boldsymbol{\pi}$ using proposed framework (π_{avg}, π_{inv}, π_{max}) and compare with vanilla IS and DR learners

*WPDA (Warfarin Pharmacogenetic Dosing Algorithm) is deterministic algorithm proposed by IWPC

Results – Policy Evaluation

Baseline (p_0 estimated using single network) | p_0 bootstrapped from 10 networks

Behaviour Policy	Methods	Vanilla IPS/DR	Bayesian NN(1)	Bayesian NN(10)	Deep Ensemble Model	Deep Ensemble Model + Data
LR (0.7017)	IPS – avg	0.7129 ± 0.0182	0.6811 ± 0.0478	0.6318 ± 0.0064	**0.6986 ± 0.0048**	0.7269 ± 0.0057
	IPS - inv			0.6818 ± 0.0122	0.7101 ± 0.0054	0.7442 ± 0.0065
	DR - avg	0.7304 ± 0.0032	0.6997 ± 0.0107	0.6935 ± 0.0006	0.7302 ± 0.0005	0.7175 ± 0.0017
	DR - inv			**0.6985 ± 0.0016**	0.7303 ± 0.0006	0.7183 ± 0.0016
PHARMA (0.7031)	IPS – avg	0.7295 ± 0.0149	0.7290 ± 0.0450	0.6571 ± 0.0119	0.6998 ± 0.0033	0.6948 ± 0.0034
	IPS - inv			**0.7046 ± 0.0128**	0.7266 ± 0.0051	0.7374 ± 0.0074
	DR - avg	0.7009 ± 0.0253	0.6893 ± 0.0088	0.6833 ± 0.0016	0.6889 ± 0.0013	0.6906 ± 0.0017
	DR - inv			0.6920 ± 0.0015	**0.6982 ± 0.0211**	0.7133 ± 0.0121

True reward of
policy evaluated

Reward estimates $\hat{R}$ (mean ± std. dev.)

Policy evaluated: Classifier trained on original Warfarin dataset
Reward Estimators: IPS – Inverse Propensity Score; DR – Doubly-Robust Estimator

Results – Policy Learning

Baseline (p_0 estimated using single network) **p_0 bootstrapped from 10 networks**

Behavior Policy	Methods	Vanilla	Bayesian NN(1)	Bayesian NN(10)	Bootstrapping	
					Model	Model + Data
LR	IPS – avg			**0.6439 ± 0.0064**	0.6384 ± 0.0120	0.6335 ± 0.0106
	IPS – inv	0.6378 ± 0.0124	0.6359 ± 0.0082	**0.6432 ± 0.0085**	0.6377 ± 0.0135	0.6329 ± 0.0115
	IPS – max			0.6467 ± 0.0064	0.6440 ± 0.0111	**0.6502 ± 0.0062**
	DR – avg			**0.6737 ± 0.0057**	0.6728 ± 0.0035	0.6710 ± 0.0058
	DR – inv	0.6726 ± 0.0032	0.6626 ± 0.0132	0.6682 ± 0.0063	0.6721 ± 0.0030	0.6706 ± 0.0066
	DR – max			0.6721 ± 0.0058	0.6753 ± 0.0032	**0.6768 ± 0.0049**
PHARMA	IPS – avg			0.6480 ± 0.0022	**0.6493 ± 0.0040**	0.6344 ± 0.0065
	IPS – inv	0.6469 ± 0.0061	0.6191 ± 0.0266	0.6373 ± 0.0017	**0.6487 ± 0.0042**	0.6302 ± 0.0075
	IPS – max			0.6461 ± 0.0036	**0.6544 ± 0.0023**	0.6552 ± 0.0050
	DR – avg			0.6626 ± 0.0011	**0.6634 ± 0.0028**	0.6575 ± 0.0047
	DR – inv	0.6633 ± 0.0027	0.6588 ± 0.0025	0.6633 ± 0.0013	**0.6636 ± 0.0030**	0.6525 ± 0.0073
	DR – max			0.6649 ± 0.0009	0.6674 ± 0.0018	**0.6680 ± 0.0037**

Actual reward of learnt policy (mean ± std. dev.)

Reward Estimators: IPS – Inverse Propensity Score; DR – Doubly-Robust Estimator

Policy Learning – MIMIC

- Clinical records of ~40000 critical care patients
- Includes demographics, laboratory tests, vital signs, medications and more

- Task: Recommend length of stay for patient on arrival in the ICU
  - **# patients selected** : ~12000
  - **Action** : Length of stay buckets (2-3, 3-5, 5-8, 8+)
  - **Reward** : 0 if re-admitted within 30 days, else 1

- ~10% patients are readmitted

- Balanced sub-sampling to counter imbalance in reward

Methods	Reward
IPS	0.5279 ± 0.0209
IPS - avg	0.5303 ± 0.0077
IPS - inv	0.5328 ± 0.0103
IPS - max	$\mathbf{0.5541} \pm \mathbf{0.0129}$
DR	0.5129 ± 0.0070
DR - avg	0.5131 ± 0.0066
DR - inv	0.5125 ± 0.0082
DR - max	$\mathbf{0.5284} \pm \mathbf{0.0065}$

Reward estimates $\hat{R}$

Takeaways

- Bootstrapping leads to **lower variance** and **improved policy learning**
  - Policy with highest reward among bootstrapped samples has lower variance
- Bayesian Neural Networks achieve lower variance during learning
- R_{inv} policy evaluator performs better than R_{avg}
- Our approach can be used to derive action confidence bounds for each patient before policy deployment

Can we explore other paradigms to ensure robustness of policy π ?

Adversarial Policy Optimization

Optimize π for worst-case propensity scoring model π_0

$$R(\theta_1, \theta_2) = \min_{\pi_0} \max_{\pi} \sum_{i=1}^{n} \frac{\pi(a = a_i | x_i, \theta_1)}{\pi_0(a = a_i | x_i, \theta_2)} r_i + \lambda * Loss(\pi_0, \pi_0^t)$$

L

R

IPS-based Policy Learner

Cross-entropy loss for modeling clinician's action policy

Minimize R;
Minimize L

π_0

Adversarial

π

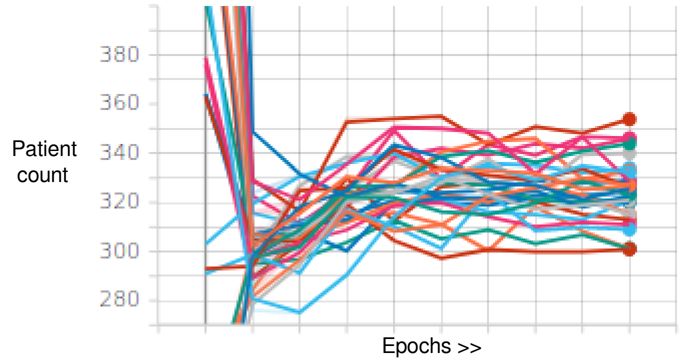
Maximize R

Preliminary Results – Warfarin Dosing

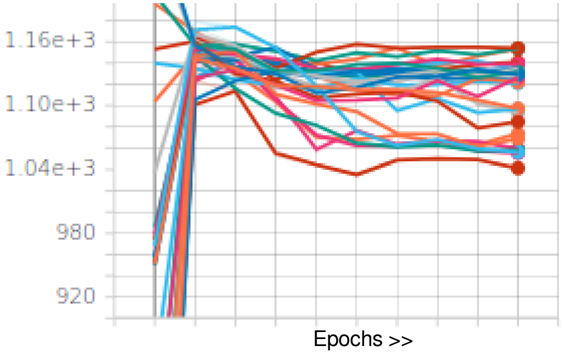
Adversarial Learning leads to lesser variance in recommended actions, particularly for high dosage actions

Vanilla IPS (Learn π_0 ; Independently learn π)

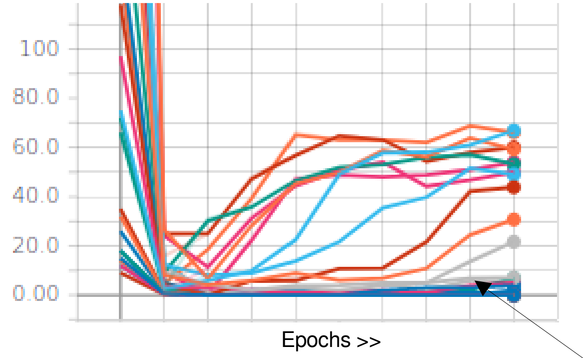
patients recommended low dosage



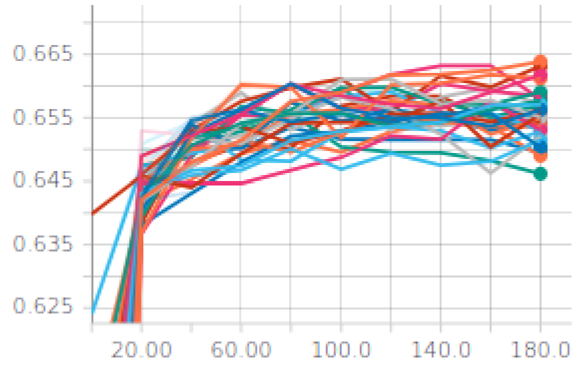
patients recommended medium dosage



patients recommended high dosage



Reward

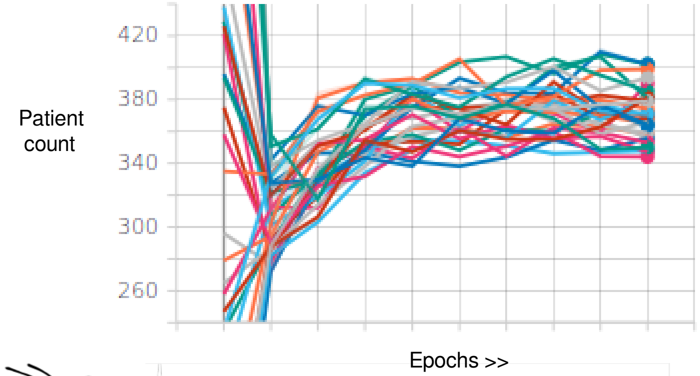


Adversarial IPS

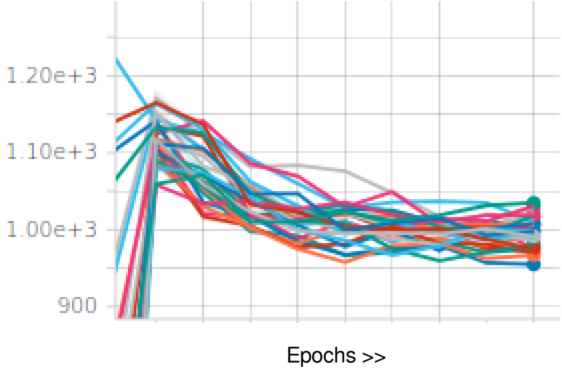
Lower variance

Slightly better rewards

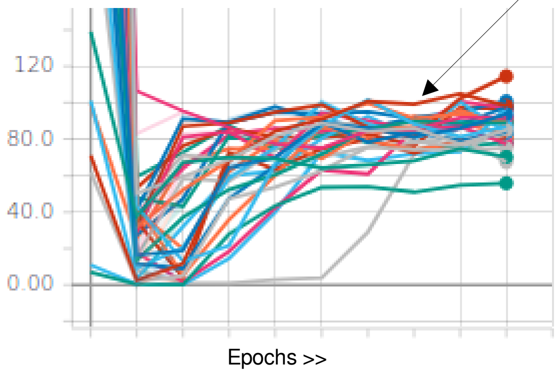
patients recommended low dosage



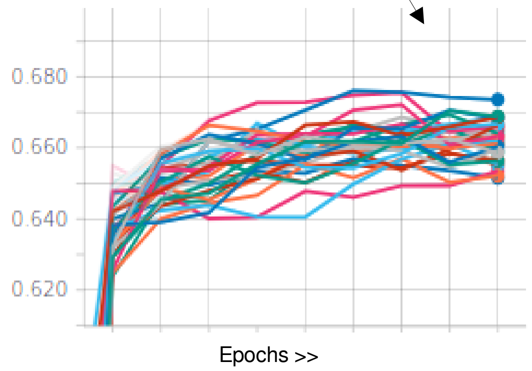
patients recommended medium dosage



patients recommended high dosage



Reward



Next Steps

- Bootstrapping
  - Analyze patient-wise action uncertainty distribution for different learnt policies
  - Policy learning and evaluation on eICU dataset
- Adversarial Learning
  - Evaluate on MIMIC and eICU datasets

Thank You!

Anirudh Choudhary

Email : achoudhary46@gatech.edu