HARVARD John A. Paulson School of Engineering and Applied Sciences
MICHIGAN MEDICINE UNIVERSITY OF MICHIGAN
COMPUTER SCIENCE & ENGINEERING UNIVERSITY OF MICHIGAN

This work is funded by NSF and NIH-NLM.

Leveraging Factored Action Spaces for Efficient Offline RL in Healthcare

Shengpu Tang, Maggie Makar, Michael W. Sjoding, Finale Doshi-Velez, Jenna Wiens

NeurIPS 2022

Action Spaces in Clinical Problems

Commonly exhibit combinatorial structures

Acute Dyspnea

(ongoing project at UM)

$|A| = 2^5 = 32$

- {0,1} Antibiotics
- {0,1} Anticoagulants
- {0,1} Fluids
- {0,1} Diuretics
- {0,1} Steroids

Mech Vent Weaning

(Prasad et al., UAI 2017)

$|A| = 2 \times 4 = 8$

MV setting $a[0] \in \{0, 1\}$

Sedation level $a[1] \in \{0, 1, 2, 3\}$

$$\mathcal{A} = \left\{ \begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}0\\2\end{bmatrix}, \begin{bmatrix}0\\3\end{bmatrix}, \begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\2\end{bmatrix}, \begin{bmatrix}1\\3\end{bmatrix} \right\}$$

AI Clinician / MIMIC-sepsis

(Komorowski et al., Nature Medicine 2018)

$|A| = 5 \times 5 = 25$

		Dose of vasopressor				
		1	2	3	4	5
Dose of i.v. fluid	1	1	2	3	4	5
	2	6	7	8	9	10
	3	11	12	13	14	15
	4	16	17	18	19	20
	5	21	22	23	24	25

Factored Action Spaces

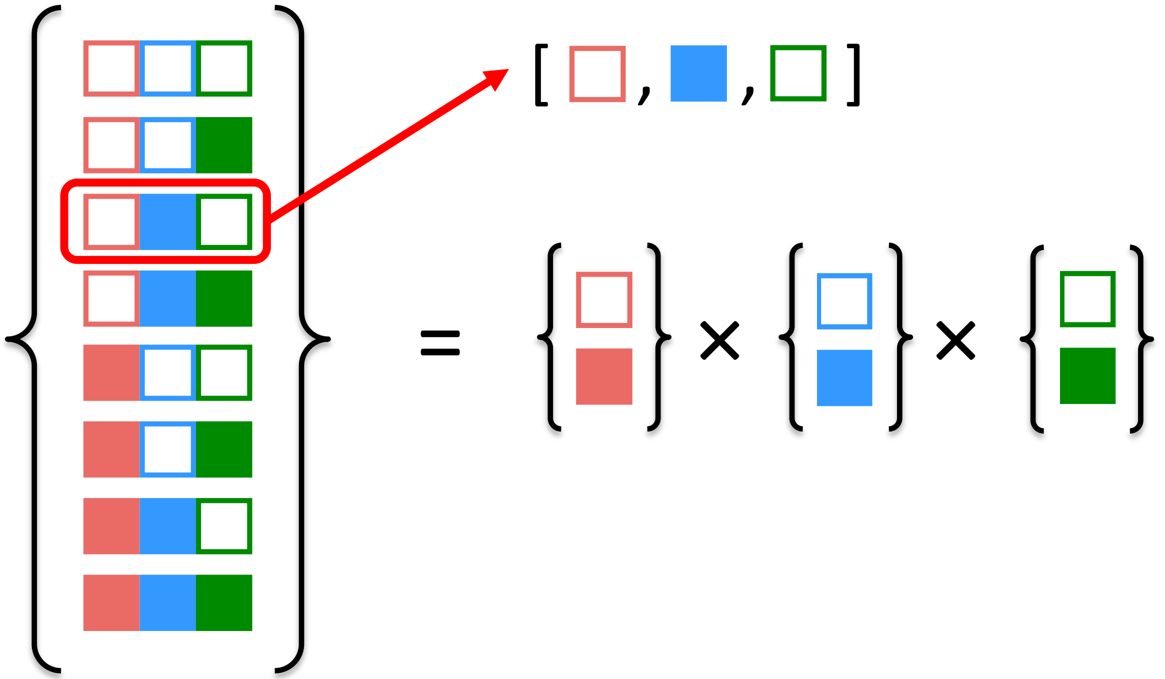
$$\mathcal{A} = \mathcal{A}_1 \times \cdots \times \mathcal{A}_D$$

Overall action space is a
Cartesian product of *D* sub-action spaces

$$\boldsymbol{a} = [a_1, \ldots, a_D] \in \mathcal{A}$$

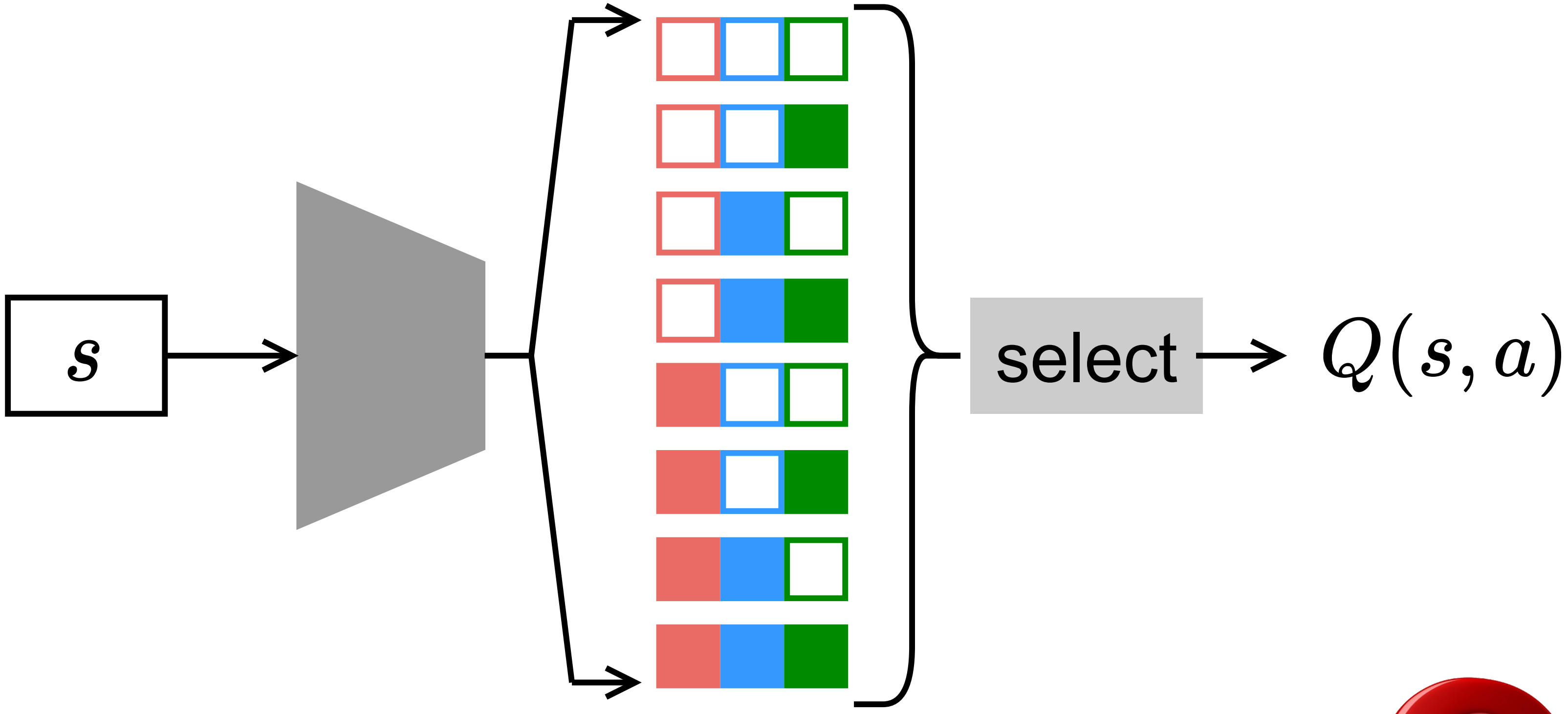
$$a_d \in \mathcal{A}_d$$

Each action is a **vector** of
D sub-actions



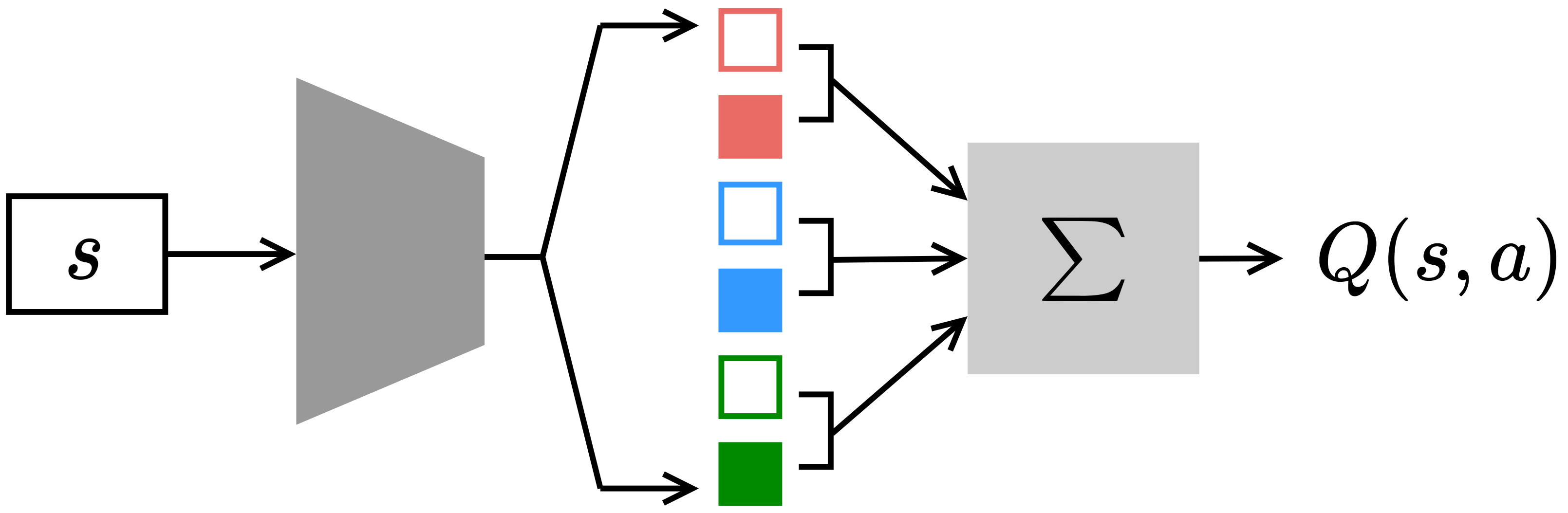
Combinatorial action space $\rightarrow$ Typical Q function

See paper for details on related works

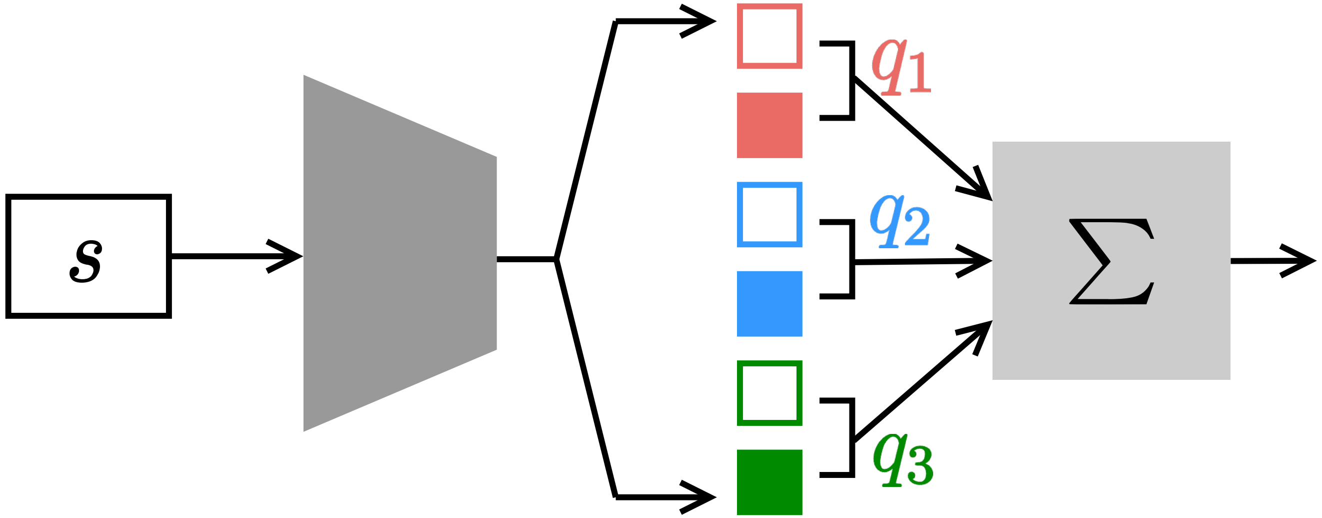


Inefficient?

Factored action space $\longrightarrow$ Linear Q decomposition



Factored action space $\longrightarrow$ Linear Q decomposition



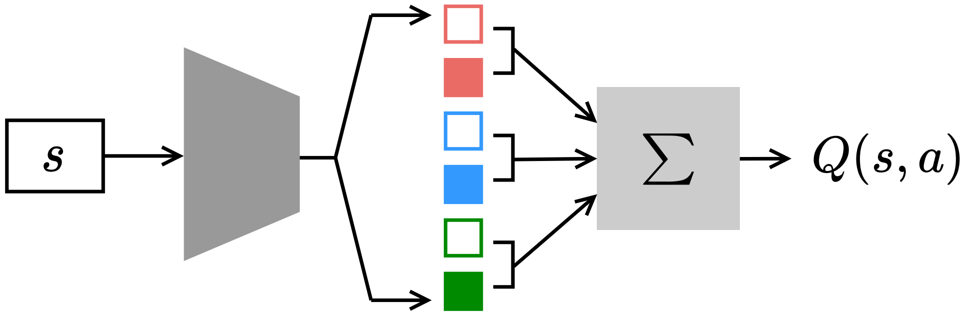
$$Q(s, \boldsymbol{a}) = \sum_{d=1}^{D} q_d(s, a_d)$$

$$Q(s, \square\blacksquare\square) = q_1(s, \square) + q_2(s, \blacksquare) + q_3(s, \square)$$

Our Contributions

linear Q-function decomposition

$$Q^{\pi}(s, \boldsymbol{a}) = \sum_{d=1}^{D} q_d(s, a_d)$$



We develop an approach for offline RL with **factored action spaces** by learning **linearly decomposable** Q-functions.

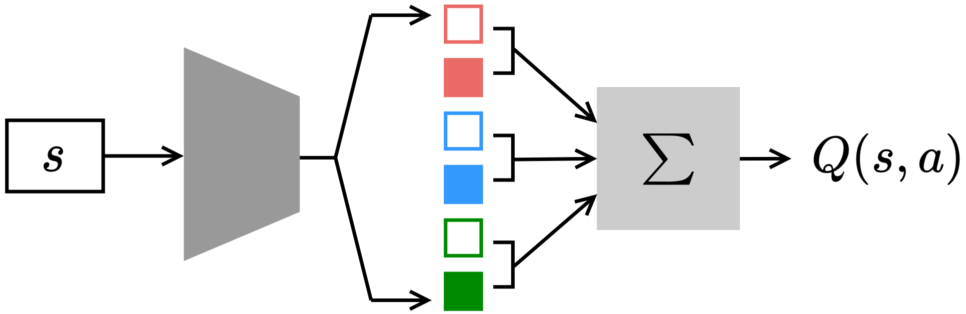
- Provide new ***theoretical insights*** on its applicability
- Conduct ***empirical evaluations*** in the context of <u>offline RL for healthcare</u>

Theoretical Insights

See paper for details

linear Q-function decomposition

$$Q^\pi(s, \boldsymbol{a}) = \sum_{d=1}^{D} q_d(s, a_d)$$



"When does it work?"

Does linear decomposition always exist? Will using linear decomposition introduce bias?

Sufficient Conditions for Zero Bias

…yet are not necessary

D "parallel" MDPs $\rightarrow$ implicitly factorized MDP via state abstractions

Theoretical Insights

See paper for details

Outside the regime of theoretical guarantees --

Implication of linear approximation on bias, variance, and policy optimality

Reduced Variance

Bias-Variance Trade-off

The number of free parameters of tabular MDP

$$|\mathcal{S}||\mathcal{A}| = |\mathcal{S}|(\prod_{d=1}^{D} |\mathcal{A}_d|) \quad \rightarrow \quad |\mathcal{S}|\left((\sum_{d=1}^{D} |\mathcal{A}_d|) - D + 1\right)$$

Bias $\nRightarrow$ Suboptimality

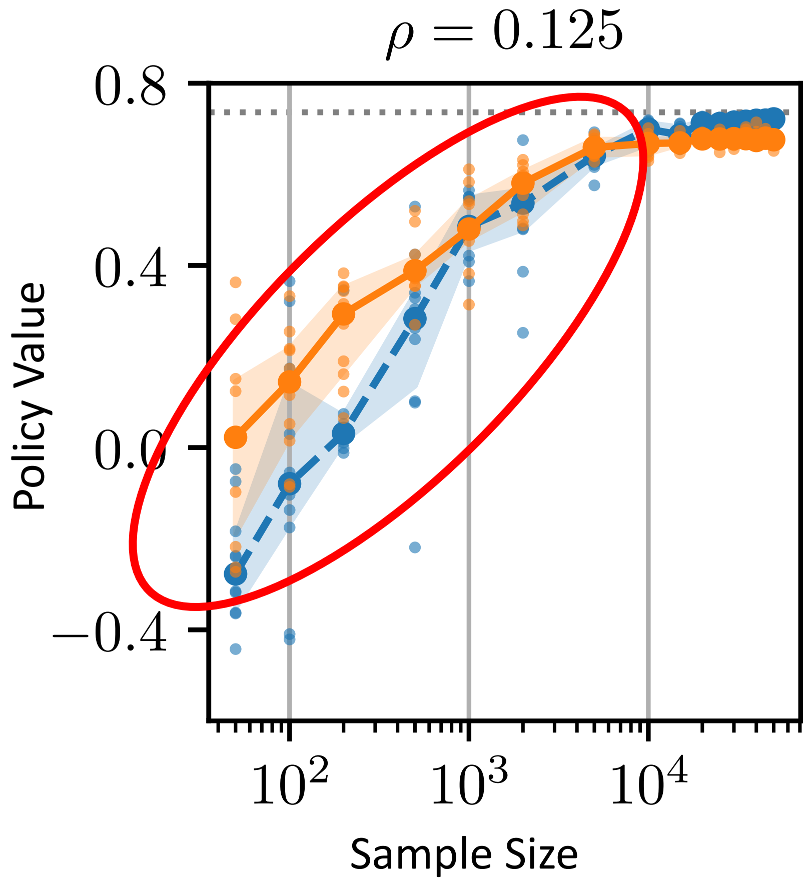
e.g., when two sub-actions "reinforce" their independent effects

Demonstrate, with examples, how **domain knowledge** may be used to inform its **applicability** in real-world problems (e.g., healthcare, education)

Experiment: Sepsis Simulator

Simulator based on Oberst & Sontag, ICML 2019.

Action Space: $\mathcal{A} = \mathcal{A}_{\text{abx}} \times \mathcal{A}_{\text{vaso}} \times \mathcal{A}_{\text{mv}}$ $\quad |\mathcal{A}| = 2^3 = 8$



Behavior policy takes the optimal action with probability ρ

Proposed approach…

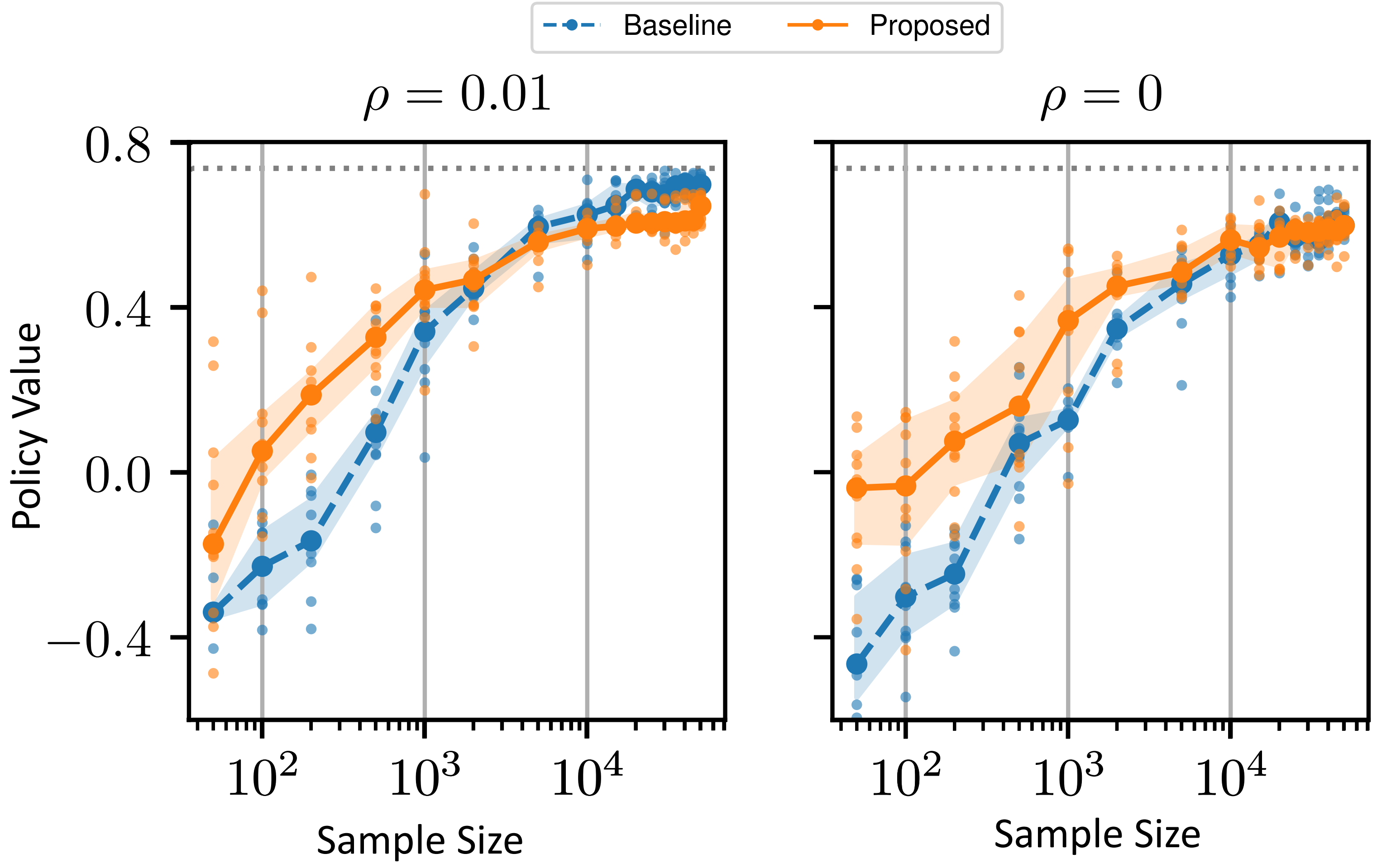
Outperforms **baseline** for small sample sizes

Closely matches **baseline** for large sample sizes

Experiment: Sepsis Simulator

Simulator based on Oberst & Sontag, ICML 2019.

Action Space: $\mathcal{A} = \mathcal{A}_{\text{abx}} \times \mathcal{A}_{\text{vaso}} \times \mathcal{A}_{\text{mv}}$ $\quad |\mathcal{A}| = 2^3 = 8$



Behavior policy takes the optimal action **less than random**

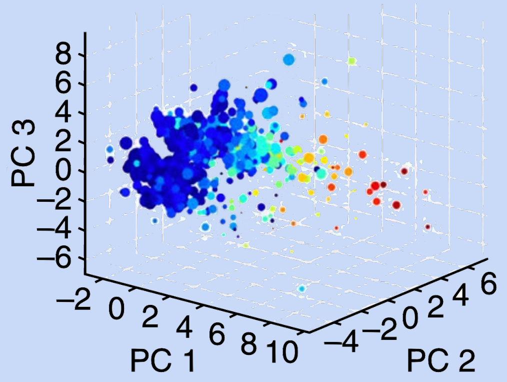
Proposed approach better at inferring **underexplored** actions

Experiment: Sepsis Treatment in MIMIC-III

Problem setup based on Komorowski et al., “AI Clinician”, *Nature Medicine* 2018.

State Space

Derived from 48 physiological signals



Action Space

$$\mathcal{A} = \mathcal{A}_{\text{vaso}} \times \mathcal{A}_{\text{fluid}}$$
$$|\mathcal{A}| = 5 \times 5 = 25$$

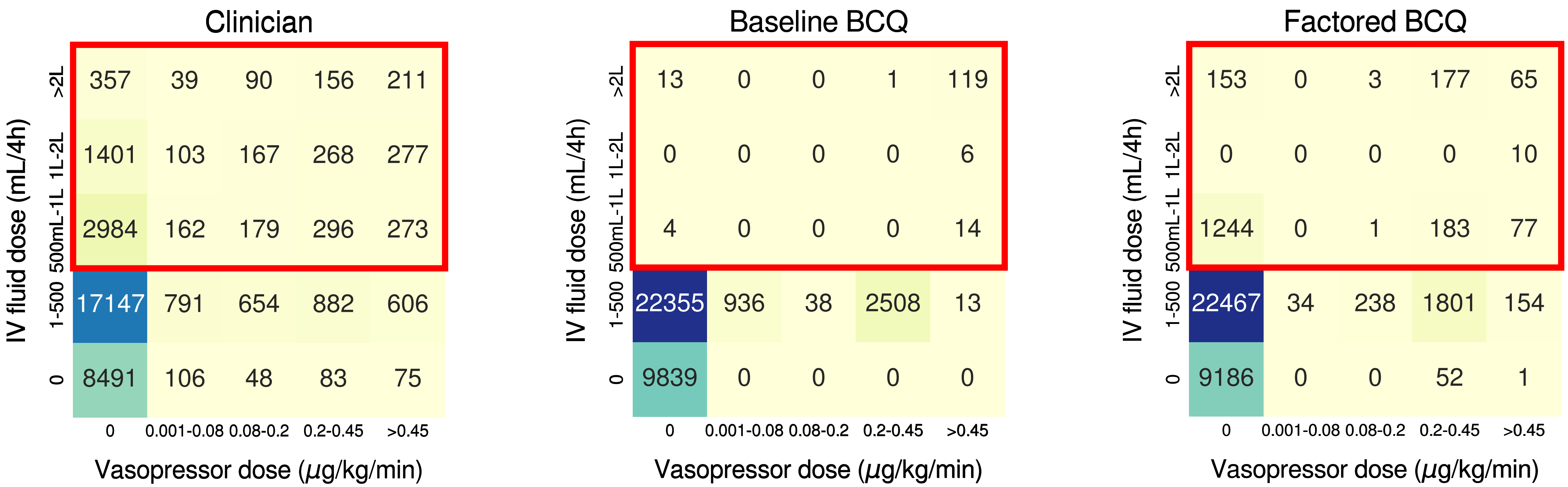
		Dose of vasopressor				
		1	2	3	4	5
Dose of i.v. fluid	1	1	2	3	4	5
	2	6	7	8	9	10
	3	11	12	13	14	15
	4	16	17	18	19	20
	5	21	22	23	24	25

Policy	Baseline BCQ	Factored BCQ	Clinician
Test WIS	90.44 ± 2.44	91.62 ± 2.12	90.29 ± 0.51
Test ESS	178.32 ± 11.42	178.32 ± 11.96	2894

Better performance at same effective sample size

Experiment: Sepsis Treatment in MIMIC-III

See paper for details



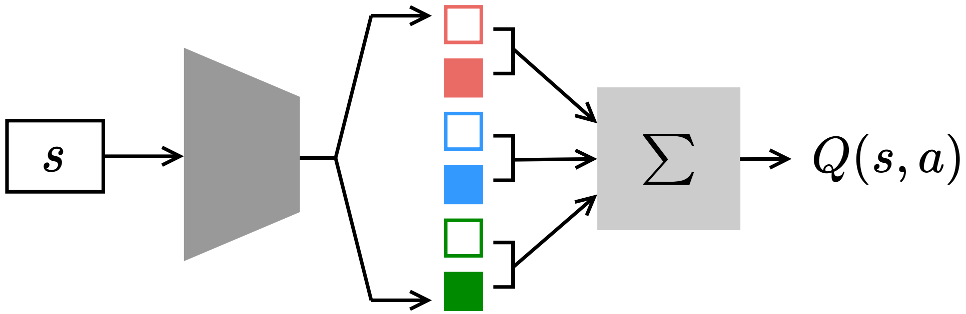
For less frequently observed / underexplored treatment combinations

Proposed approach captures their effects better

Takeaways

We develop an approach for offline RL with **factored action spaces** by learning **linearly decomposable** Q-functions.

- Leverage domain knowledge when available
- Identify scenarios when approximation bias does not lead to suboptimal performance
- Could apply more broadly to help scale RL methods in other applications involving combinatorial action spaces



S. Tang M. Makar M.W. Sjoding F.Doshi-Velez J. Wiens

This work is funded by NSF and NIH-NLM.

https://github.com/MLD3/OfflineRL_FactoredActions