Medical Dead-ends and Learning to Identify High-risk States and Treatments

M. Fatemi, T. W. Killian, J. Subramanian, M. Ghassemi

Microsoft · University of Toronto · Vector Institute · CSAIL · MIT · IMES

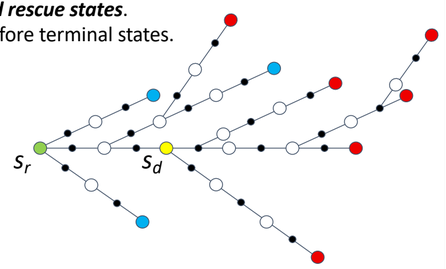
What is a medical dead-end?

- **Undesired** terminal state (e.g., patient death)
- **Desired** terminal state (e.g., patient recovery)
- **Dead-end**: all trajectories starting from s_d reach an undesired terminal state w.p.1
- **Rescue**: from s_r a desired terminal state is reachable w.p.1

➢ We want to identify all **dead-ends**, and the **treatments** that lead to them so they can be avoided

NOTE:

➢ Terminal states are assumed to be signaled when entered.
➢ This is ***NOT*** the case for ***dead-end and rescue states***.
➢ Dead-end and rescue may exist far before terminal states.



A Paradigm Shift for Offline RL in *Safety-Critical* Environments

- Learning optimal treatment strategies from observational clinical data is *offline + off-policy*

➢ **Inability to explore**
➢ **Limited data diversity**

} Severely complicates the development of RL algorithms ***that suggest what to do***

Rather than the *infeasible* task of learning optimal policies that suggest ***what to do***, we learn ***what treatments to avoid***

Major research questions:

➢ Can dead-ends be identified in clinical data?
➢ Is there anything that can signal the occurrence of a dead-end?
➢ Were alternative treatments available that could have been selected so as to avoid the patient entering a dead-end state?

Establishing Treatment Security

Treatment Security (intuitively):

➢ We update the security condition from Fatemi, et al. (ICML'19) to constrain the policy to avoid treatments which lead to dead-ends

Definition: A policy π is secure if for any $\lambda \in [0,1]$

$$P_D(s,a) + F_D(s,a) \geq \lambda \implies \pi(s,a) \leq 1 - \lambda$$

➢ Of course, it is not possible to secure "all" policies. Also, inferring the maximal λ for all (s,a) pairs is intractable.

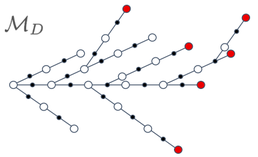
Challenges:

1. Dead-ends are not known *a priori*
2. Transition function T underlying P_D and F_D is often unknown

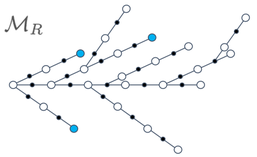
We develop a **learning** framework to satisfy this treatment security condition **from data**

Fundamental Value Functions

➢ Construct two independent MDPs to assign value to the observed terminal outcomes in the offline data. Both are identical to the original MDP with the following specifications:



- $r_D = \begin{cases} -1, & \text{if an undesired terminal state is reached} \\ 0, & \text{otherwise} \end{cases}$
- No discounting: $\gamma_D = 1$ ⇒ $Q_D^* \in [-1, 0]$



- $r_R = \begin{cases} 1, & \text{if a desired terminal state is reached} \\ 0, & \text{otherwise} \end{cases}$
- No discounting: $\gamma_R = 1$ ⇒ $Q_R^* \in [0, 1]$

➢ We prove an important **basic property**:

$$-Q_D^*(s,a) = P_D(s,a) + F_D(s,a) + M_D(s,a)$$

➢ This property assigns a **special physical meaning** to $-Q_D^*(s,a)$: It corresponds to the **minimum probability of a negative outcome**

➢ Equivalently, $1 + Q_D^*(s,a)$ is the **maximum hope of a positive outcome**

Dead-end Discovery (DeD)

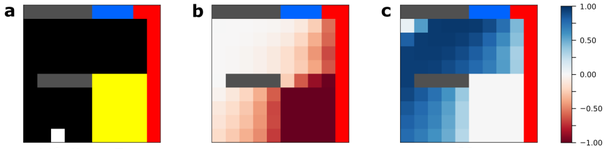
➢ We further prove the following (see the paper for the formal exposition):

1. If $\pi(s,a) \leq 1 + Q_D^*(s,a)$ and $P_D(s,a) + F_D(s,a) \geq \lambda$ then $\pi(s,a) \leq 1 - \lambda$ for all λ
2. If $\pi(s,a) \geq Q_R^*(s,a)$ and $P_R(s,a) + F_R(s,a) \geq \lambda$ then $\pi(s,a) \geq \lambda$ for all λ
3. There exists a threshold $\delta_D \in (-1, 0)$ independent of states and treatments that separates dead-end states from the rest
4. There exists a threshold $\delta_R \in (0, 1)$ independent of states and treatments that separates rescue states from the rest

➢ Hence, for treatment security it is sufficient to abide by the maximum hope of recovery

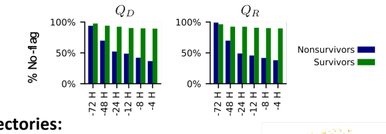
➢ These four main results ground the DeD method

Demonstrating DeD – LifeGate



Applying DeD to Sepsis Treatment (MIMIC-III)

- **Dead-end Identification:**



- **Analyzing the First Flag and Individual Trajectories:**