Review: Optimal Medication Dosing from Suboptimal Clinical Examples: A Deep Reinforcement Learning Approach

presented by unt ko

5.12.2024

https://ieeexplore.ieee.org/document/7591355

1 Background

Medication dosing in intensive care units (ICUs) presents a significant challenge, particularly for drugs with sensitive therapeutic windows like heparin. Incorrect dosing can result in adverse effects, extended hospital stays, and increased healthcare costs. While established clinical protocols exist, deviations are common and often necessary as clinicians adapt treatments to individual patient needs. This creates an opportunity for machine learning approaches to learn optimal dosing strategies from real-world clinical data.

2 Research Framework

The authors present a clinician-in-the-loop sequential decision-making framework utilizing deep reinforcement learning (RL) to develop individualized dosing policies. The system learns from retrospective electronic medical record (EMR) data, capturing the complex relationships between patient characteristics, medication doses, and outcomes. This framework integrates state estimation with policy learning through end-to-end optimization.

3 Methodology

Objective

The goal is to learn a dosing policy that maximizes the overall fraction of time a given patient stays within their therapeutic aPTT range.

The approach combines four key components:

1. A discriminative hidden Markov model (DHMM) for state estimation
2. Deep Q-learning for policy optimization
3. End-to-end training to simultaneously optimize state estimation and policy learning
4. clinician-in-the-loop policy: a control policy that takes into account the action of a clinician as well as the patient's response when suggesting a new action. In this setting, a clinician may choose to approve or overwrite the action suggested by the RL agent at any given point in time.

4 Data Source and Selection

- **Database**: Multiparameter Intelligent Monitoring in Intensive Care MIMIC II ICU database
- **Cohort Size**: 4,470 patients
- **Time Window**: 48 hours of data per patient from first heparin administration
- **Split**: 80% training, 20% testing

Data Types

- **Time-series features** (measured hourly):
  - Heparin dose levels (past 4 hours), activated partial thromboplastin time (aPTT, past 4 hours), arterial carbon dioxide level (CO_2), albumin, bilirubin, creatinine, hematocrit, hemoglobin, INR, platelet count, prothrombin time, troponin, urea, white blood cell count (WBC), blood pH.
- **Vital signs**:
  - Heart rate (HR), systolic blood pressure (SBP), diastolic blood pressure (DBP), respiration rate, oxygen saturation (SaO2), body temperature.
- **Scores**:
  - Glasgow Coma Scale (GCS), Sequential Organ Failure Assessment (SOFA) score (daily).
- **Dichotomous Features (Binary Attributes)**:
  - Ethnicity (white vs. non-white), ICU service type (surgical vs. medical), gender, presence of pulmonary embolism, presence of obesity.

5 System Architecture

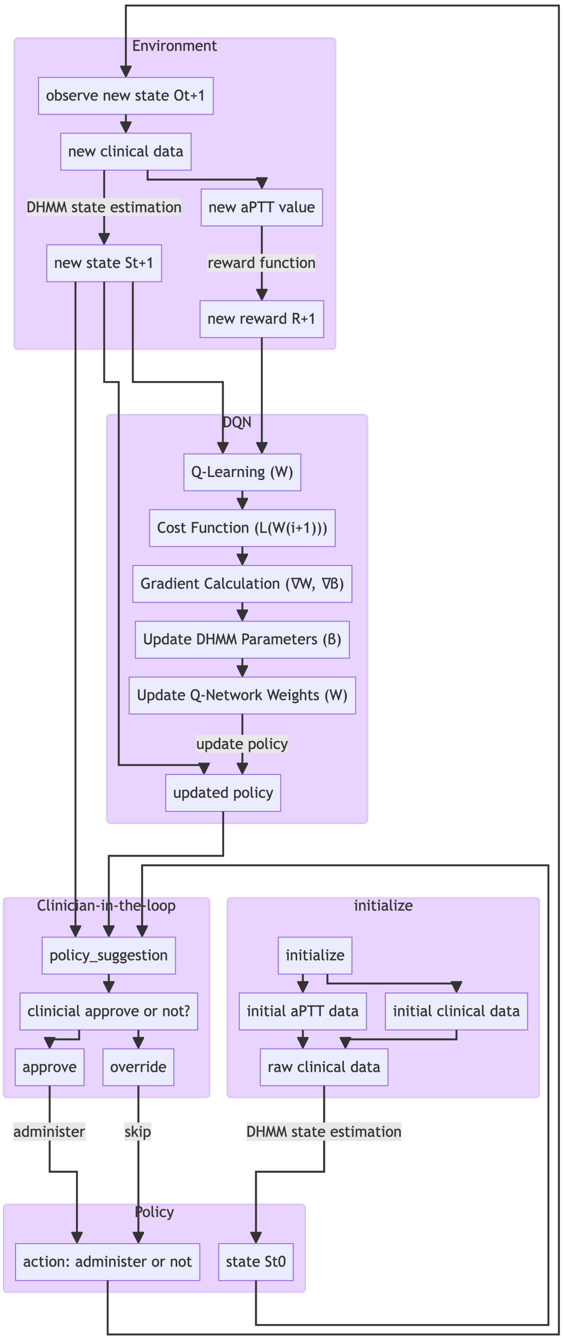

Figure 1: RL architecture

6 RL Components

Component	**Description**
Environment	ICU setting with patient physiological parameters
State	Patient condition estimated by DHMM from clinical measurements
Action	Six discrete heparin dosage levels based on quantile intervals
Reward	Function based on aPTT levels: +1 for therapeutic range (60-100s), decreasing to -1 outside range
Policy	the agent to provide an optimized heparin dosing and stay within the theraputic range

Table 1: Reinforcement Learning Components

7 Reinforcement Learning for Heparin Dosing

We describe the application of reinforcement learning (RL) for personalized heparin dosing based on the provided information.

7.1 Environment Definition

The environment is defined as follows:

- **State (s):** A representation of the patient's clinical phenotype at a given time. This includes:
  - Last measured activated Partial Thromboplastin Time (aPTT) value.
  - Other relevant clinical variables (e.g., age, weight, comorbidities).
- **Action (a):** The heparin dose to administer. Heparin values are discretized into six quantile intervals to define a set of actions.
- **Reward (r):** A function reflecting how well the chosen action maintains the patient's aPTT within the therapeutic range (60-100). The reward function is:

  $$r_t = \frac{2}{1 + e^{-(\mathrm{aPTT}(t)-60)}} - \frac{2}{1 + e^{-(\mathrm{aPTT}(t)-100)}} - 1$$

  This function assigns a maximal reward of one when aPTT is within the therapeutic window and diminishes rapidly towards -1 as the distance from the window increases.

7.2 Agent Initialization

The RL agent is initialized with:

- **Policy (π):** A neural network mapping states to actions. It takes the patient's state as input and outputs the recommended heparin dose.
- **Deep Hidden Markov Model (DHMM):** Used for state estimation to handle high-dimensional clinical observations and provide a compact state representation.

7.3 Training Loop

The training loop proceeds as follows:

1. **For each patient (episode):**
   (a) Initialize the patient's state (s).
   (b) **While the episode is not terminated (patient requires heparin):**
      i. Use the DHMM to estimate the current state (s) from raw clinical data.
      ii. Select an action (a) based on the current policy: $a = \pi(s)$.
      iii. Administer the chosen heparin dose (a).
      iv. Observe the new state (s') and reward (r).

v. Update the policy (π) using a deep reinforcement learning algorithm (e.g., Deep Q-Network or an actor-critic method) to maximize expected future rewards.

The Q-learning algorithm updates weights W by minimizing the cost function:

$$L(W_{i+1}) = \frac{1}{2|N_i|} \sum_{n \in N_i} \sum_{t=1}^{T^{(n)}} [v_t^{(n)}(W_i) - Q(s_t^{(n)}, a_t^{(n)}; W_{i+1})]^2 \quad (1)$$

(c) End episode when the patient no longer needs heparin.

7.4 Evaluation

The trained policy is evaluated as follows:

- Test the trained policy on a separate dataset of patients.
- Compare the outcomes (e.g., percentage of time within the therapeutic aPTT range) achieved by the RL agent to those achieved by standard clinical guidelines.

Results

- Clinicians initially overdosed the patient, leading to worsened aPTT levels until corrective actions were taken (6-15 hours). This included discontinuing heparin temporarily and subsequently tapering off dosing, which ultimately resulted in underdosing.
- In contrast, the RL agent's recommendations began slightly above the population mean dose and converged to the population mean, achieving quicker alignment with the therapeutic range.
- Patients whose heparin dosages closely followed the RL agent's recommendations demonstrated improved long-term outcomes, achieving positive rewards after a few adjustments.
- Clinically administered doses deviating from RL recommendations were associated with poorer outcomes, confirming the RL agent's superior performance in optimizing heparin dosing.

Performance Analysis

- Patients whose heparin dosages closely followed the RL agent's recommendations demonstrated improved long-term outcomes, achieving positive rewards after a few adjustments.
- Clinically administered doses deviating from RL recommendations were associated with poorer outcomes, confirming the RL agent's superior performance in optimizing heparin dosing.

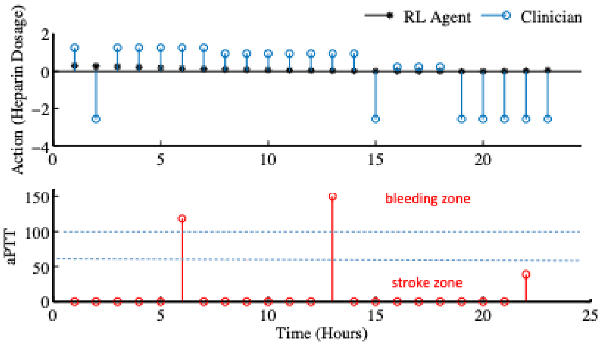

Figure 2: An example of heparin dosing (mean-normalized) by a clinician, and the corresponding recommended dosing of the RL agent.

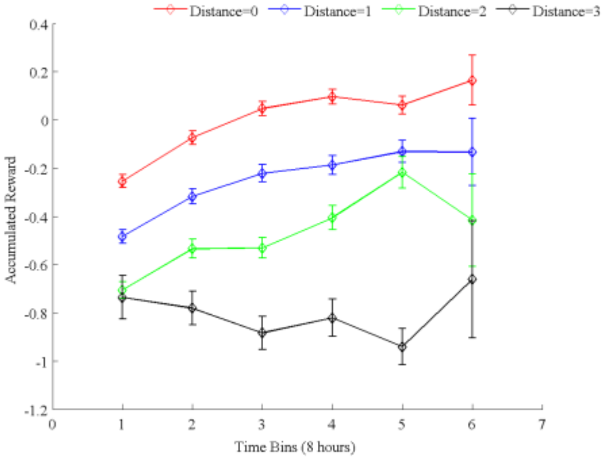

Figure 3: On average and consistently over time, following the recommendations of the RL agent (red line) results in the best long-term performance.

8 Conclusion

The study demonstrates that deep reinforcement learning can effectively learn optimal medication dosing policies from retrospective clinical data. Key findings include:

1. The RL agent's recommendations consistently achieved better outcomes than clinical guidelines. The trained RL agent's recommendation starts slightly above the population mean for heparin and then converges to the population mean, which is likely to bring patients within their therapeutic range more quickly.

2. Following the RL policy led to positive rewards after fewer adjustments

3. The combination of state estimation and RL training via end-to-end optimization proved crucial for handling high-dimensional clinical data

4. The approach shows promise for developing data-driven, individualized treatment policies

While the study focuses on heparin, the framework could potentially be extended to other medications with sensitive therapeutic windows. The authors note that this represents an early step in applying deep reinforcement learning to clinical decision support, with potential implications for precision medicine and learning healthcare systems.