Continuous Perceiver Latent Space: Interpretability Through MDP State Mapping

Poster author: Charles VIELZEUF, Statistical Genetics Team, RIKEN AIP (supervised by R. Sakurai)

Introduction

In this work we focus on the **Medical Information Mart for Intensive Care** (MIMIC-III) clinical dataset [1, 2] and **mortality prediction** in Intensive Care Unit (ICU) stays.

Using the first 48 hours after ICU admission, we train various models (baselines, transformers, including **Neural Ordinary Differential Equations** (NODEs)) to predict the patient's in-hospital mortality likelihood. [3, 4]

In doing so, we obtain compact **intermediate embeddings** for patient trajectories. To interpret these embeddings, we map them to a **Markov Decision Process (MDP)** state space built following the **AI-Clinician** framework [5].

The goal was to see if models, while learning to predict mortality, are also able to learn about the patient's clinical dynamics and state transitions.

While doing so, we also explore some **reinforcement learning** methods [6] on the rebuilt MDP to **evaluate expert treatments** and the impact of various medically-informed reward functions on the policy quality and survival-oriented evaluation metrics.

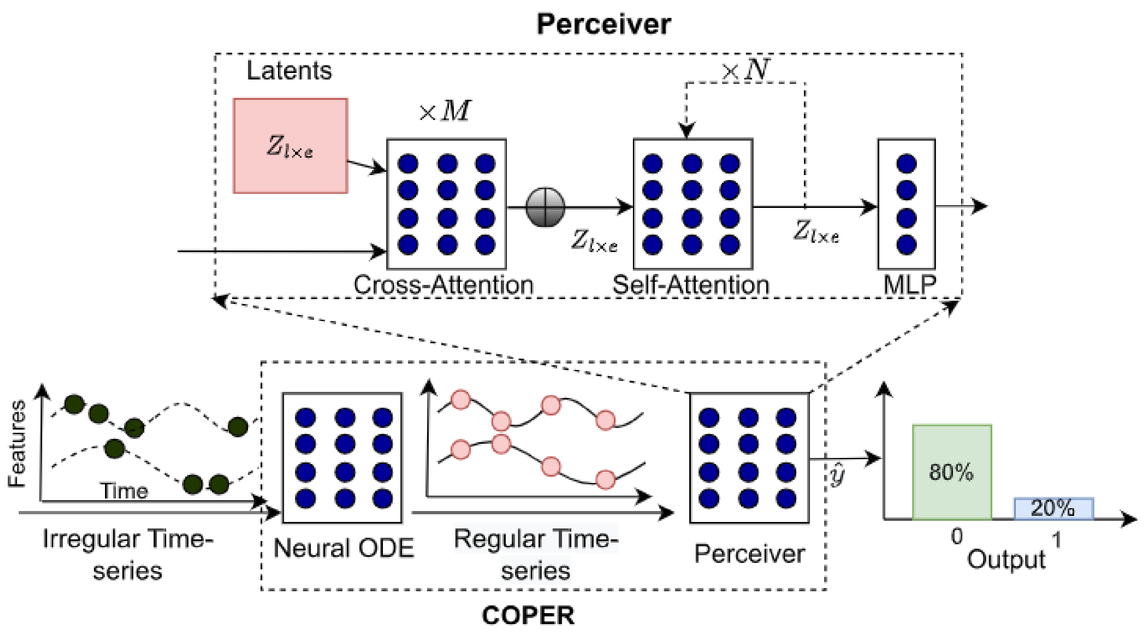
MIMIC-III Dataset, Sepsis cohort and patient trajectories

We extract from the **MIMIC-III dataset** around **20,000 ICU stays** after filtering on the sepsis cohort (trajectory includes at least one **ICD-9 diagnosis code** consistent with sepsis), and some other criteria such as duration (at least 48 hours), age, **Sequential Organ Failure Assessment** (max(SOFA) ≥ 2 (Sepsis-3 style, baseline SOFA assumed 0). The pipeline also applies cohort-cleaning exclusions (e.g., implausible fluid/urine outliers and early-withdrawal death patterns). [5, 6]

A **patient trajectory** consists of a sequence of input vectors (one hour timesteps for 48h after admission) made of **76 variables** (both clinical and mask channels, renormalized), including examples such as heart rate, mean arterial pressure, respiratory rate, temperature, oxygen saturation, creatinine, bilirubin, lactate, white blood cell count, urine output... using **carry forward and masking techniques** to handle missing values. [3]

COPER and other models for predicting mortality

Architecture overview. Both COPER and Transformer process the same 48×76 hourly ICU tensor, but differ in the attention pattern and latent bottleneck. COPER uses **cross-attention** from learned latent queries to the patient sequence, then **self-attention** on latent slots. The Transformer baseline applies **self-attention** directly on the patient sequence (no latent-query bottleneck). Neural ODE dynamics are inserted before and/or after attention blocks.



Results for mortality task. We train a few models on the same dataset (train/val/test), until convergence for small models and after 10 epochs for larger models (transformer/COPER).

model	tr. acc	val acc	test acc	val AUC	test AUC
logistic (L2)	0.9877	0.6687	0.6845	0.6601	0.6481
logistic (L1)	0.8930	0.7026	0.6893	0.6956	0.6788
random forest	1.0000	0.7381	0.7896	0.7634	0.7573
LSTM	0.7768	0.7288	0.7880	0.7449	0.7418
COPER (1 NODE) (10 ep)	0.7081	0.6626	0.6796	0.7263	0.7218
COPER (2 NODE) (10 ep)	0.6818	0.6749	0.6748	0.7300	0.6926
Transformer (baseline) (10 ep)	0.6835	0.6826	0.6618	0.7291	0.7230
Transformer (1 NODE) (10 ep)	0.7000	0.6826	0.6974	0.7370	0.7252

Logistic models and random forest **strongly overfit** and cannot capture temporal clinical dynamics. Attention-based models are overall **more consistent across settings, especially when we add noize to datas**. [4]
Among non-overfitting models, the 2024 Transformer variant reaches the best validation/test AUC. NODEs help for the continuity of the latent dynamics, but roughly double the training time.

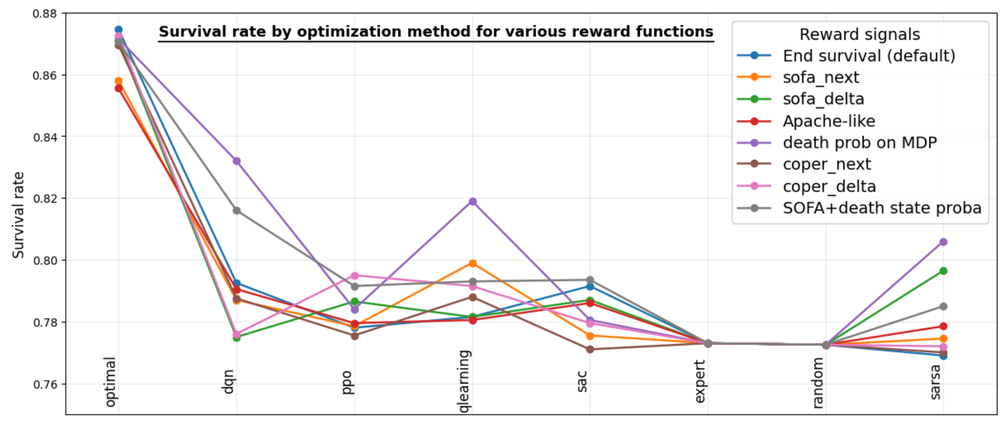
To further investigate on those models, we **study the latent learnt representations** (last latent vector in the case of COPER) of each patient trajectory. One easy to interpret tool are MDPs, as states are clinically meaningful.

Markov Decision Process definition

We use an AI-Clinician-style tabular MDP [5]. At each time step t, the patient physiological vector is assigned to a discrete clinical state using **KMeans**; this defines a state space of 750 states (arbitrary value).
The action space is a **5**×**5** grid combining **fluid and vasopressor** intensity levels, giving 25 discrete actions. For each state–action pair, **transition dynamics are estimated from observed trajectories**: $C(s, a, s')$ is the number of observed transitions from state s to state s' after action a, and $C(s, a)$ is the total number of transitions observed from (s, a). To keep transitions statistically robust, we keep as admissible in state s only actions with enough support ($C(s, a) \geq 20$). Transition probabilities $T(s, a, s')$ are then estimated empirically from these counts; for rare/non-admissible actions, we use the average dynamics of admissible actions in the same state.
The **expert policy** (empirical clinician behavior) is defined as: $\pi_{\text{expert}}(a \mid s) = \frac{C(s,a)}{\Sigma_{\bar{a}} C(s,\bar{a})}$, i.e., the observed action frequency distribution in each state.

MDP policy training and evaluation

Following the work done in [6], we train a few reinforcement learning policies on the MDP. Given that the MDP is tractable, we can use **value iteration** to find the optimal policy for a given reward function. Previous works only considered sparse survival rewards at the end of the trajectory; to evaluate the reliability of optimization methods, we tried different reward functions and RL methods.



End survival (default):

$$r_{survival}(s, a, s') = \begin{cases} +1 & s' = s^{\text{surv}} \\ 0 & s' \in \{s^{\text{death}}, s_\infty\} \\ 0 & \text{else} \end{cases}$$

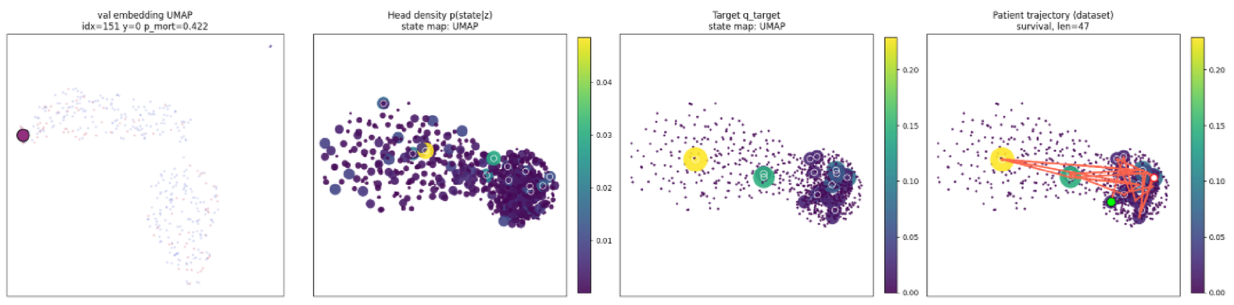
SOFA shaping (λ=0.05,; shaping guides optimization; evaluation uses r_{packaged} at absorption):

$$r_{sofa\_next}(s, a, s') = \begin{cases} -\lambda\,\text{SOFA}(s')/S_{\max} & s, s' \text{ transient} \\ r_{\text{survival}}(s, a, s') & \text{else} \end{cases}$$

This comparison of **optimal policy** (value iteration), **expert clinical policy**, and RL methods (**DQN, PPO, Q-Learning, SAC, SARSA**) shows the variability under multiple reward definitions (based on COPER mortality score on a state, SOFA score, or directly mortality deduced from the MDP transition dynamics). This last reward is the most successful, as it closer aligns with the evaluation metric (survival probability estimated by sampling trajectories in the same MDP).

Conclusion: latent to MDP state distribution mapping

After freezing the models, a small heads learns to map **COPER patient-level last latent vector** used for mortality prediction to a **distribution over MDP states** deduced from the patient's observed trajectory (weights are occupation times in each state).



From left to right, MDP states colored by SOFA score, density predicted by the head, true time-averaged target density deduced from the true trajectory dynamics, and the true trajectory itself; all panels correspond to a same patient from test set.

The **latent-to-state mapping quality** indicates that mortality prediction is not based on a shallow snapshot but on a rather **trajectory-aware compressed representation**. When removing the **neural ODE component**, the mapping is less successful, which is consistent with the hypothesis that **continuous-time modeling** improves the faithfulness of latent clinical dynamics.

The mapping isn't perfect, as we can see that the head is not often confident and most of the time predicts a distribution that is not concentrated on the target states. Instead of using only the last latent vector, used for mortality prediction, using the whole latent representation also helps improving the mapping quality.

Ways of improving the predictions could be to **add text-free information** to the latent embeddings to add context for instance. The code is available at https://github.com/charles-vzf/COPER-latents.

References

[1] Alistair E. W. Johnson et al. Mimic-iii, a freely accessible critical care database. *Scientific Data*, 3:160035, 2016. doi: 10.1038/sdata.2016.35. URL https://physionet.org/content/mimiciii/1.4/.

[2] Alistair E. W. Johnson et al. Deidentification of free-text medical records using pre-trained bidirectional transformers. In *Proceedings of the ACM Conference on Health, Inference, and Learning (CHIL '20)*, Toronto, Ontario, Canada, April 2020. Association for Computing Machinery. ISBN 978-1-4503-7046-2. doi: 10.1145/3368555.3384455. URL https://doi.org/10.1145/3368555.3384455.

[3] Vinod Kumar Chauhan et al. Coper: Continuous patient state perceiver. In *2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, pages 1–4, 2022. doi: 10.1109/BHI56158.2022.9926807. URL https://arxiv.org/abs/2208.03196.

[4] Vinod Kumar Chauhan et al. Continuous patient state attention model for addressing irregularity in electronic health records. *BMC Medical Informatics and Decision Making*, 24(1):117, 2024. URL https://github.com/jmdvinodjmd/COPER. Reference details mirrored from the COPER project repository.

[5] Matthieu Komorowski et al. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. *Nature Medicine*, 24(11):1716–1720, 2018. doi: 10.1038/s41591-018-0213-5. URL https://doi.org/10.1038/s41591-018-0213-5.

[6] Kartik Choudhary et al. Icu-sepsis: A benchmark mdp built from real medical data. *arXiv preprint arXiv:2406.05646*, 2024. URL https://arxiv.org/abs/2406.05646.