Learning sparse symbolic policies for sepsis treatment

Jacob F. Pettit [*1] **Brenden K. Petersen** [1] **Chase Cockrell** [2] **Dale B. Larie** [2]
Felipe Leno Silva [1] **Gary An** [2] **Daniel M. Faissol** [1]

Abstract

Sepsis is a life-threatening organ dysfunction caused by a dysregulated host response to infection. Despite its severity, no FDA-approved drug treatments exists. Recent work controlling sepsis simulations with deep reinforcement learning have successfully discovered effective cytokine mediation strategies. However, the performance of these neural-network based policies comes at the expense of their deployability in clinical settings, where *sparsity* and *interpretability* are required characteristics. To this end, we propose a pipeline to learn *simple, sparse symbolic policies* represented by constants and/or succinct, human-readable expressions. We demonstrate our approach by learning a sparse symbolic policy that is efficacious on simulated sepsis patients.

1. Introduction

Sepsis is a life threatening condition wherein the immune response to infection or injury becomes dysregulated and paradoxically leads to tissue damage and organ failure. The condition has a mortality rate between 28 and 50 percent and approximately 1 million people are diagnosed with sepsis each year (Singer et al., 2016). Therefore, even small reductions in the mortality rate of the disease will potentially save hundreds to thousands of lives. Effectively treating sepsis is still an elusive goal and there is no FDA approved drug treatment (Wood & Angus, 2004). The current management of sepsis is based on controlling the infection with antibiotics and providing physiological support of failing organs until the patient's immune system sufficiently readjusts and recovers from its disordered state.

The inability to translate basic knowledge of the mechanisms that drive sepsis is due in large part to the complexity of the interactions in the face of a paucity of clinical data. Therefore, we contend that the future effective control of sepsis will require an enhanced ability to identify finer-grained differences between patient disease trajectories, that, by necessity, must be generated *in silico*; additionally, we contend that the ability to manage the combinatorial challenge is associated with the need to potentially manipulate multiple mediators at a given time.

We approach sepsis treatment as a sequential decision-making problem, in which an agent decides which cytokine-mediating drugs to administer and at what dosages at clinically relevant time scales. A simulation of sepsis is a necessary tool to open the disease up to computational analysis and control. Thus, we leverage the Innate Immune Response Agent-Based Model (IIRABM) (An, 2004), a widely used sepsis simulation (Petersen et al., 2018; Cockrell & An; 2018; 2021), as a starting point for our studies.

Previous works have successfully applied deep reinforcement learning to similar sequential decision-making problems, including simulations of sepsis (Petersen et al., 2018); however, the resulting policies (treatments) are represented by neural networks (NNs), which are notoriously difficult to interpret (Montavon et al., 2018). While effective, such black-box models are especially undesirable in health-related domains, where interpretability and safety are crucial design requirements. Failure modes can also occur unexpectedly and inexplicably, and assessing such risks is challenging if not infeasible. For these reasons, treatment policies based on black-box models are unlikely to be deemed acceptable for deployment on real patients.

A related problem is that black-box (e.g. NN-based) models tend to be *dense*: that is, all actions are used at each time step, and all observations affect the choice of all actions. However, many healthcare applications (and simulations thereof) exhibit many plausible drug candidates, e.g. cytokine-regulating drugs for sepsis. Indeed, a large part of the challenge is learning *which* drug or combination of drugs to use. Thus, a key design criteria for clinically adoptable policies is that they are *sparse* in their selection of drugs.

Inspired by recent advances in learning *symbolic control policies* (Landajuela et al., 2021), we address the above

[1]Computational Engineering Division, Lawrence Livermore National Laboratory, Livermore, California, USA [2]Department of Surgery, University of Vermont, Burlington, Vermont, USA. Correspondence to: Jacob F. Pettit <pettit8@llnl.gov>.

Workshop on Interpretable ML in Healthcare at International Conference on Machine Learning (ICML).

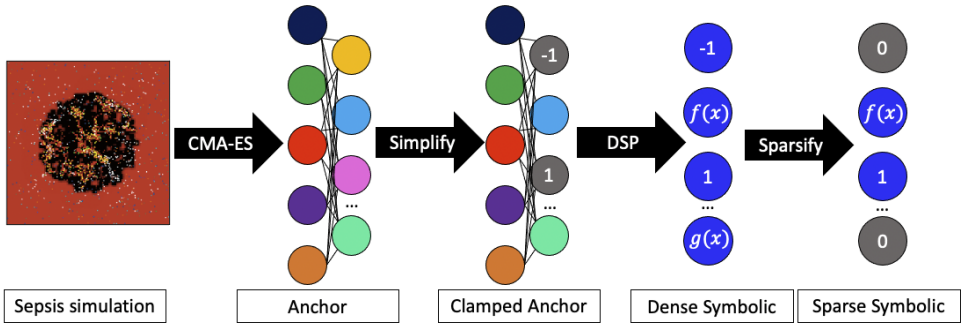

Figure 1. Overview of our pipeline for learning sparse, fully symbolic policies for a reinforcement learning environment.

challenges by proposing a pipeline for learning *sparse, interpretable* symbolic policies for sepsis treatment. Unlike previous works, our method produces policies that are *sparse*, in that they only use a subset of available actions (drugs), and *interpretable*, in that outputs are human-readable expressions, from which clinical insights can be gleaned by domain experts simply by inspection.

We summarize our contributions as: (1) a pragmatic pipeline for discovering sparse, interpretable, symbolic policies for novel treatment discovery, (2) a sparse symbolic policy that halves the baseline sepsis mortality rate for in silico experiments on the IIRABM, and (3) a clinically meaningful interpretation of the discovered symbolic policy.

2. Background & Related Work

To represent sepsis *in silico*, we utilize the innate immune response agent-based model (IIRABM). The IIRABM represents the human endothelial-blood interface that incorporates the primary drivers of innate immunity and endothelial remodeling post injury, including endothelial cells, macrophages (of multiple polarities), polymorphonuclear leukocytes, T_h0, T_h1, and T_h2 cells, as well as the associated precursor cells. Previous work has shown that the aggregate output of the model can be considered to be a random dynamical system (Cockrell & An) which captures the clinical heterogeneity seen in the septic population.

We highlight the most critical details of the IIRABM here (see An (2004) and Petersen et al. (2018) for further details). The environment observations comprise a vector of various cytokine (small signaling molecule) concentrations, along with a measure of aggregate tissue damage. Actions represent putative cytokine-mediating drugs. Specifically, each of the 11 cytokines c has an action $a_c \in [-1, 1]$ that either augments ($a_c > 0$) or inhibits ($a_c < 0$) the effects of that cytokine's interactions within the IIRABM. Actions are selected every 12 hr simulated time, reflecting a clinically relevant timescale for changing dosages.

Initial attempts at identifying therapeutic control strategies utilizing the IIRABM demonstrated that simple treatment regimes (e.g. give one/a few drugs intermittently) are ineffective (An, 2004). Cockrell & An (2018) leveraged genetic algorithms to treat simulated septic patients, significantly lowering mortality rate. However, upon examination of the *in silico* patients that did not heal, it was observed that the genetic algorithm-based policy led the patient into a configuration from which it was *unable* to heal. Thus, while the policy was successful for the majority, it was harmful to a minority, motivating the need for an *adaptive* policy.

Reinforcement learning (RL) has been very successful in similar settings, including in medical and health applications (Yu et al., 2020). Various RL algorithms have been applied to clinical care, including cancer chemotherapy drug dosage, (Ahn & Park, 2011; Zhao et al., 2009; Padmanabhan et al., 2017) and sepsis treatment (Komorowski et al., 2016; 2018; Raghu et al., 2017; Petersen et al., 2018). However, most of the proposed methods for sepsis treatment rely on pre-collected clinical care datasets such as MIMIC-III (Johnson et al., 2016). This means that they cannot leverage a simulator for testing hypothetical interventions, such as the application of a new drug or drug combination.

Similarly to our approach, Petersen et al. (2018) leveraged an earlier version of the IIRABM to train and evaluate RL policies. However, the learned NN-based policies were dense and difficult to interpret. Recent works propose methods such as saliency maps to partially interpret NN models (Douglas et al., 2019; Fan et al., 2020). However, these methods do not provide the level of interpretability required to build sufficient confidence in medical domains. For this reason, interpretable tree-based models have been used in medicine despite achieving worse performance than NN models (Laber & Zhao, 2015).

3. Methods

Our pipeline, illustrated in Figure 1, comprises four main steps: (1) train a NN-based policy[1], (2) iteratively simplify the NN-based policy, (3) distill the NN-based policy into a symbolic policy, and (4) iteratively sparsify the symbolic policy.

Training a NN-based policy. We train a neural-network based policy using the Covariance Matrix Adaptation Evolutionary Strategy (CMA-ES) (Igel et al., 2007; Hansen, 2016) implementation in ESTool (Ha, 2017). We train on a single IIRABM parameterization, only changing the random seed each episode. The policy network is composed of one input layer with 12 units, two hidden layers each with 64 units, and one output layer with 11 units. Hyperbolic tangent activations are applied at each layer. Hereafter, we refer to this NN-based policy as **Anchor**.

Simplifying the NN-based policy. After obtaining **Anchor**, we seek to *simplify* it by determining which action dimensions can be locked or "clamped" to either extrema of their allowed range, instead of using the dynamic value prescribed by the NN. We determine which actions can be set to constants by iterative over each action dimension in **Anchor** and evaluating it with each action clamped to +1 or -1. If the performance decreases by less than 5%, then that action is locked or clamped to +1 or -1. Hereafter, we refer to this simplified NN-based policy as **Clamped Anchor**.

Algorithm 1 Policy simplification

for $i \in 1, \ldots, N$ **do**
  for $j \in -1, 1$ **do**
    Action $a_i = j$
    Performance P gathered from evaluation run.
    if P within 5% of normal performance **then**
      Action "clamped" at $a_i = j$.
      break

Searching for a symbolic policy. To distill **Clamped Anchor** into a symbolic policy, we leverage Deep Symbolic Policies (DSP) (Landajuela et al., 2021). DSP begins with a pre-trained neural network "anchor" policy, then employs neural-guided search to directly search the space symbolic expressions, learning one action dimension at a time. See Landajuela et al. (2021) for details. Beginning with **Clamped Anchor**, we perform DSP on each non-constant action dimension, resulting in a fully symbolic policy. Hereafter, we refer to this symbolic policy as **Dense Symbolic**.

Sparsifying the symbolic policy. To increase the simplicity and interpretability of the resulting symbolic policy, we investigate "sparse policies." These policies aim to replace as many actions as possible with a zero without having an adverse effect on performance.

The final policy resulting from this pipeline consists entirely of constant values and parsimonious symbolic representations of a subset of the action dimensions. Hereafter, we refer to this sparsified symbolic policy as **Sparse Symbolic**.

Algorithm 2 Policy sparsification

for $i \in 1, \ldots, N$ **do**
  Action $a_i = 0$
  Performance P gathered from evaluation run.
  if P within 5% of normal performance **then**
    Action locked at $a_i = 0$.

4. Results

Experimental setup. The anchor policy is trained using CMA-ES. For DSP, we use default hyperparameters (Landajuela et al., 2021), with a library of allowable symbols $\{+, -, \times, \div, \sin, \cos, \exp, \log, -1, 0.1, 1, 5, s_1, \ldots, s_n\}$, where s_i is the ith observation dimension.

We evaluate each algorithm by running it over 100 different patient parameter sets, each with 3 seeds, taken from a hold-out set not seen during training. IIRABM patient parameters include a measure of host resilience, two meausures of microbial virulence (invasiveness and toxivenesis), a measure of environmental toxicity/contamination, and an initial injury severity. Each patient parameter set has an associated *baseline mortality rate* that occurs when there is no cytokine intervention. Our evaluation set of patients has an average baseline mortality rate of 55.73%, ranging from 1% to 99%.

Final sparse symbolic policy. The final sparse symbolic policy comprises 2 expressions, 6 zero-valued actions (i.e. that drug is unused), and 3 non-zero constants. The two expressions are: $a_{\text{TNF}} = \sin(\text{IL8}) - \sin(\text{sIL1r})$ and $a_{\text{IL8}} = \cos(\text{IL8} - \text{IL1})$. The zero-valued actions are: $a_{\text{sTNFr}} = a_{\text{GCSF}} = a_{\text{IFN}\gamma} = a_{\text{IL1}} = a_{\text{IL4}} = a_{\text{IL12}} = 0$. The non-zero constant actions are: $a_{\text{PAF}} = a_{\text{sIL1r}} = 1$ and $a_{\text{IL10}} = -1$. Note that a_c refers to mediation of cytokine c; positive values augment (up-regulate) c and negative values inhibit (down-regulate) c.

Notably, while both symbolic policies leverage periodic functions, the periodicity is *not* observed within the ranges of observations the environment takes on. For example, the cosine operator in a_{IL8} can be replaced with up to the quadratic term of the Taylor expansion, and still achieves the same performance.

Performance evaluation. For empirical analysis, we evaluate policies at each of the four stages of our pipeline, as well as a policy that prescribes no intervention ($a_c = 0$ for all cytokines). Table 1 shows the results achieved by each pol-

[1] In general, this step can use a policy of *any* form. For simplicity, we assume NN-based policies throughout.

	None	Anchor	Clamped Anchor	Dense Symbolic	Sparse Symbolic
Mortality	55.7%	10.3%	37.0%	33.0%	29.0%
Symbolic?	N/A	×	×	✓	✓
Sparse?	N/A	×	×	×	✓

Table 1. Performances of a no-intervention policy (**None**), NN-based policy (**Anchor**), simplified NN-based policy with clamped actions (**Clamped Anchor**), a symbolic policy generated using DSP (**Dense Symbolic**), and a sparsified symbolic policy with most dimensions fixed to zero (**Sparse Symbolic**).

icy. **Anchor** has a mortality rate of 10.3% over the 100 test patients; thus, it generalizes well across patients. The final **Sparse Symbolic** policy has a mortality rate of 29%. Thus, we demonstrate a performance-simplicity trade-off between dense NN-based policies and sparse symbolic policies.

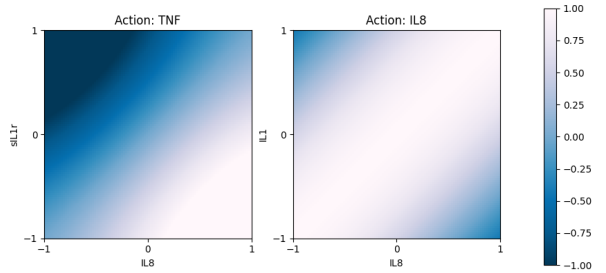

Figure 2. Heatmap illustrating the only two non-constant dimensions of our sparsified symbolic policy.

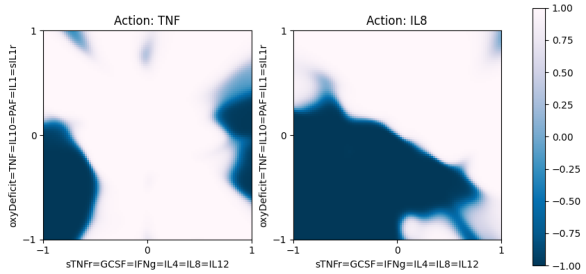

Figure 3. Heatmap representing the trained anchor policy. Notice that a heatmap was generated for each action used by the symbolic policy for comparison purposes, but the anchor policy also uses the other 9 actions.

5. Discussion

Our sparse symbolic policy approximately halves the mortality rate, from a baseline of 56% without intervention to 29% with intervention. While not as effective as the 10.3% mortality of the NN-based anchor policy, our method highlights the trade-off between performance and three key desirable features: simplicity, sparsity, and interpretability.

First, the sparse symbolic policies are dramatically simpler than their NN counterparts. Our discovered expressions contain only a handful of mathematical operators; in contrast, NN policies consist of thousands of real-valued parameters, requiring several matrix multiplications wrapped by nonlinearities. To illustrate this, we juxtapose a visualization of the symbolic policy (Figure 2) with the NN policy (Figures 3). Notably, the NN policy is so high-dimensional that we cannot even visualize it without looking at high-dimensional slices of the observation space.

Second, our symbolic policies are sparse. That is, they contain many zero-valued action dimensions, meaning that many of the possible drug candidates are *never* administered by the policy. This is a highly desirable feature in the clinical setting, as using fewer drugs can reduce side effects and the chances of drug-drug interactions.

Third, by virtue of being symbolic, we can glean clinical understanding by analyzing the expressions. To this end, we provide a clinical interpretation of the discovered policy. The policy's manipulation of the IL8 pathway is consistent with its known role of signaling distress of damaged endothelial cells to inflammatory cells (Harada et al., 1994). In episodes where the patient healed under the policy, we observe near maximal augmentation of the IL8 protein synthesis pathway, increasing the spatial range from which a distressed cell can recruit help. The role of IL1 is to propagate the inflammatory response. When IL8 and IL1 begin to diverge, it indicates the inflammation is (dangerously) propagating faster than the innate immune system's ability to contain it, which the policy mitigates by downregulating IL8 in response to this increasing divergence ($a_{\text{IL8}} = \cos(\text{IL8} - \text{IL1})$ when $0 < (\text{IL8} - \text{IL1}) < \sim 0.5$). The constant augmentation of PAF (Zimmerman et al., 2002) serves a similar purpose as that molecule attracts neutrophils to sites of infection, effectively increasing the immune response to microbial infection. This specific version of the IIRABM was specialized to simulate a hypoinflammatory/immunocompromised patient. The function of IL10 (Saraiva & O'garra, 2010) is to limit inflammation; in the immunocompromised patient, the ability to generate inflammation is inherently limited, and thus the policy's constant inhibition of IL10 is clinically realistic.

6. Conclusion

We propose a pipeline to discover sparse symbolic policies for sepsis treatment via cytokine modulation. Beginning with a high-dimensional NN-based policy, our pipeline first simplifies the policy by clamping a subset of action dimensions to constant values. We then leverage Deep Symbolic Policy to learn symbolic representations of the remaining actions. Finally, we sparsify the symbolic policy be replacing a subset of action dimensions with zero (no action). We demonstrate a proof-of-concept of this approach by applying it to an existing sepsis simulation. We find that sparse

symbolic policies can still be effective in treating simulated patients, while achieving several major benefits: sparsity, simplicity, and interpretability. In the future, we plan to expand the set of patient parameters to span a range of clinical heterogeneity and hope that methods for sparse symbolic policy discovery will enable learning interpretable treatment strategies that directly lead to actionable clinical insights.

7. Acknowledgements

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC.

References

Ahn, I. and Park, J. Drug scheduling of cancer chemotherapy based on natural actor-critic approach. *BioSystems*, 106: 121–129, 2011.

An, G. In silico experiments of existing and hypothetical cytokine-directed clinical trials using agent-based modeling. *Critical Care Medicine*, pp. 2050–2060, 2004.

Cockrell, C. and An, G. Sepsis reconsidered: Identifying novel metrics for behavioral landscape characterization with a high-performance computing implementation of an agent-based model. *Journal of theoretical biology*, 430:157–168.

Cockrell, C. and An, G. Examining the controllability of sepsis using genetic algorithms on an agent-based model of systemic inflammation. *PLoS Computational Biology*, 2018.

Cockrell, C. and An, G. Utilizing the heterogeneity of clinical data for model refinement and rule discovery through the application of genetic algorithms to calibrate a high-dimensional agent-based model of systemic inflammation. *Frontiers in Physiology*, 2021.

Douglas, N., Yim, D., Kartal, B., Hernandez-Leal, P., Maurer, F., and Taylor, M. E. Towers of saliency: A reinforcement learning visualization using immersive environments. In *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces*, pp. 339–342, 2019.

Fan, F., Xiong, J., and Wang, G. On interpretability of artificial neural networks. *CoRR*, abs/2001.02522, 2020. URL http://arxiv.org/abs/2001.02522.

Ha, D. Evolving stable strategies. *blog.otoro.net*, 2017. URL http://blog.otoro.net/2017/11/12/evolving-stable-strategies/.

Hansen, N. The cma evolution strategy: A tutorial. *arXiv preprint arXiv:1604.00772*, 2016.

Harada, A., Sekido, N., Akahoshi, T., Wada, T., Mukaida, N., and Matsushima, K. Essential involvement of interleukin-8 (il-8) in acute inflammation. *Journal of Leukocyte Biology*, 56(5):559–564, 1994.

Igel, C., Hansen, N., and Roth, S. Covariance matrix adaptation for multi-objective optimization. *Evolutionary Computation*, 15:1–28, 2007.

Johnson, A. E., Pollard, T. J., Shen, L., Li-Wei, H. L., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L. A., and Mark, R. G. Mimic-iii, a freely accessible critical care database. *Scientific data*, 3(1):1–9, 2016.

Komorowski, M., Gordon, A., Celi, L., and Faisal, A. A markov decision process to suggest optimal treatment of severe infections in intensive care. *Neural Information Processing Systems Workshop on Machine Learning for Health*, 2016.

Komorowski, M., Celi, L., Badawi, O., Gordon, A., and Faisal, A. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. *Nature Medicine*, 24:1716, 2018.

Laber, E. B. and Zhao, Y.-Q. Tree-based methods for individualized treatment regimes. *Biometrika*, 102(3):501–514, 2015.

Landajuela, M., Petersen, B. K., Kim, S., Santiago, C. P., Glatt, R., Mundhenk, N. T., Pettit, J. F., and Faissol, D. M. Discovering symbolic policies with deep reinforcement learning. *Proc. of the International Conference on Machine Learning*, 2021.

Montavon, G., Samek, W., and Müller, K.-R. Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, 73:1–15, 2018. ISSN 1051-2004. doi: https://doi.org/10.1016/j.dsp.2017.10.011. URL https://www.sciencedirect.com/science/article/pii/S1051200417302385.

Padmanabhan, R., Meskin, N., and Haddad, W. Reinforcement learning based control of drug dosing for cancer chemotherapy treatment. *Mathematical biosciences*, 293: 11–20, 2017.

Petersen, B. K., Yang, J., Grathwohl, W. S., Cockrell, C., Santiago, C., An, G., and Faissol, D. M. Deep reinforcement learning and simulation as a path toward precision medicine. *Journal of Computational Biology*, pp. 597–604, 2018.

Raghu, A., Komorowski, M., Ahmed, I., Celi, L., Szolovits, P., and Ghassemi, M. Deep reinforcement learning for sepsis treatment. *arXiv preprint arXiv:1711.09602*, 2017.

Saraiva, M. and O'garra, A. The regulation of il-10 production by immune cells. *Nature reviews immunology*, 10(3): 170–181, 2010.

Singer, M., Deutschman, C., and Seymour, C. e. a. The third international consensus definitions for sepsis andseptic shock (sepsis-3). *Journal of the American Medical Association*, pp. 801–810, 2016.

Wood, K. and Angus, D. Phamacoeconomic implications of new therapies in sepsis. *Pharmacoeconomics*, pp. 895–906, 2004.

Yu, C., Liu, J., and Nemati, S. Reinforcement learning in healthcare: A survey. *arXiv preprint arXiv:1908.08796v4*, 2020.

Zhao, Y., Kosorok, M., and Zeng, D. Reinforcement learning design for cancer clinical trials. *Statistics in Medicine*, 28:3294–3315, 2009.

Zimmerman, G., McIntyre, T., Prescott, S., and Stafforini, D. The platelet-activating factor signaling system and its regulators in syndromes of inflammation and thrombosis. *Critical Care Medicine*, 30:294–301, 2002.