Hierarchical Apprenticeship Learning for Disease Progression Modeling

Xi Yang\(^1\)\(^*,\) Ge Gao\(^2\), Min Chi\(^2\)
\(^1\) IBM Research
\(^2\) North Carolina State University
xi.yang@ibm.com, \{ggao5, mchi\}@ncsu.edu

Abstract

Disease progression modeling (DPM) plays an essential role in characterizing patients’ historical pathways and predicting their future risks. Apprenticeship learning (AL) aims to induce decision-making policies by observing and imitating expert behaviors. In this paper, we investigate the incorporation of AL-derived patterns into DPM, utilizing a Time-aware Hierarchical EM Energy-based Subsequence (THEMES) AL approach. To the best of our knowledge, this is the first study incorporating AL-derived progressive and interventional patterns for DPM. We evaluate the efficacy of this approach in a challenging task of septic shock early prediction, and our results demonstrate that integrating the AL-derived patterns significantly enhances the performance of DPM.

1 Introduction

Disease progression modeling (DPM) aims to characterize patients’ longitudinal medical records, identify disease progressive stages, and evaluate factors affecting the progression pathways [Cook and Bies, 2016]. Empowered by the increasing availability of large medical datasets, e.g., electronic health records (EHRs), various deep learning models [Choi et al., 2016; Zhang et al., 2017] have been developed for DPM based on the recorded observations in EHRs, e.g., vital signs and laboratory test results. However, relying solely on monitored observations may not fully capture the complexity of disease progression [Tintinalli et al., 1985]. For example, different patient groups may exhibit varying observations for the same disease, where a normal sign in one group may turn out to be abnormal in another [Baytas et al., 2017]. Some recent DPM works have used subsequence clustering for deriving progressive stages automatically in a data-driven manner from various observations [Yang et al., 2021]. Other than using observations for DPM, interventions also have a significant impact on disease progression [Komorowski et al., 2018; Azizsoltani et al., 2019], and some prior works have taken clinicians’ interventions as additional features for DPM [Esteban et al., 2016; Goh et al., 2021]. Clinical interventions are usually carried out based on prior medical knowledge, which is difficult to be reflected by observations alone. In addition, interventions tend to take hours or days to manifest in observations [Dulac-Arnold et al., 2020]. Furthermore, individual clinical interventions on DPM cannot be fully explained without assessing their effectiveness, since different patient groups may respond to the same treatment differently, and patients at different stages of DPM may respond differently to the same treatment [Clifford et al., 2016]. For this reason, when learning progressive patterns in DPM, it is essential to incorporate the effectiveness of interventions.

Reinforcement learning (RL) has shown considerable potential in learning effective interventional strategies for clinical decision-making [Wang et al., 2018; Yu et al., 2021]. A common challenge when applying RL to healthcare is the design of reward function, which serves as an incentive to induce effective policies [Azizsoltani et al., 2019]. Apprenticeship learning (AL) was proposed to address this issue [Abbeel and Ng, 2004]: instead of taking an explicitly delineated reward function as input, AL learns the policy by imitating the demonstrated behaviors of clinical experts. Recently, an AL method named energy-based distribution matching (EDM), has shown great success in inducing effective clinical interventional policies directly from EHRs [Jarrett et al., 2020]. In EDM, the demonstrations provided by clinicians are assumed to be generated via a uniform policy, latent driven by a single reward function. However, while managing patients under various progressive stages, such as common fevers or severe organ failures, clinicians may adopt varying policies with multiple reward functions [Wang et al., 2021].

In this paper, we leverage an AL named Time-Aware Hierarchical EM Energy-based Subsequence (THEMES) clustering [Yang et al., 2023] for DPM. THEMES is designed to handle the multiple evolving reward functions through two key components: a) subsequence partitioning, which clusters patients’ sequential records into progressive stages, and b) policy induction, which induces interventional policy for each stage. The two components are performed iteratively: referring to the progressive stages learned by subsequence partitioning, fine-grained evolving interventional policies can be induced; meanwhile, these derived interventional patterns will, in turn, refine the partitioned subsequences to indi-
cate more accurate progressive stages. The two components are modeled by a Reward-regulated Multivariate Time-aware Toeplitz Inverse Covariance-based Clustering (RMT-TICC) and an Expectation Maximization EDM (EM-EDM), respectively. RMT-TICC encodes the *interventional* patterns by introducing a reward regulator during the subsequence partitioning, motivated by the fact that interventions are latently driven by rewards; while EM-EDM captures the *progressive* patterns by using an EM to simultaneously cluster the RMT-TICC learned subsequences and induce their respective policies. We incorporate both *progressive* patterns from RMT-TICC and *interventional* patterns from EM-EDM for DPM.

The effectiveness of incorporating THEMES-learned patterns for DPM is evaluated by modeling sepsis, an extremely challenging and life-threatening organ dysfunction that is a leading cause of death worldwide [Singer et al., 2016]. Without timely interventions, patients can progress to the most severe condition of septic shock with a high mortality rate of 50% [Martin et al., 2003], while 80% of sepsis deaths can be prevented with timely diagnosis and treatment [Kumar et al., 2006]. Therefore, modeling sepsis progression is crucial for more accurate early prediction of septic shock. To achieve this goal, we leverage THEMES by incorporating latent *progressive* patterns extracted by RMT-TICC and *interventional* patterns derived by EM-EDM as supplemental features for original observations. Our results demonstrate that incorporating such additional features leads to improved DPM accuracy. The major contributions of this work are three-fold:

- **To our best knowledge, this is the first work incorporating AL-derived *progressive* and *interventional* patterns for DPM.**
- **Our empirical results reveal that the success of incorporating AL for DPM relies on the fact that THEMES can handle time-awareness and evolving reward functions.**
- **We utilize AL for modeling a challenging disease, i.e., sepsis, and predicting its most severe condition, i.e., septic shock, in early stages. This indicates the potential to enhance personalized and timely treatments for patients.**

## 2 Related Work

### 2.1 Disease Progression Modeling

The importance of DPM has been recognized in previous studies [Cook and Bies, 2016; Severson et al., 2020], leading to the usage of deep learning models such as recurrent neural networks [Lipton et al., 2015; Saqib et al., 2018]. Among these models, long short-term memory (LSTM) has shown remarkable success due to its ability to capture sequential patterns from historical records. Typically, these deep learning models take monitored *observations* as input and use neural networks to extract latent progressive patterns for DPM.

An alternative approach to build DPM is through *subsequence clustering*, which captures disease progressive stages by partitioning and clustering subsequences. In general, subsequence clustering can be categorized into distance-based (e.g., dynamic time warping [Giannoula et al., 2018]) and model-based (e.g., Gaussian mixture models [Faruqui et al., 2021] and hidden Markov models [Kwon et al., 2020]). Model-based methods are usually more reliable for better handling noise and outliers. Recently, a model-based approach called Toeplitz inverse covariance-based clustering [Hallac et al., 2017] has gained attention for accurately partitioning subsequences in various applications, such as analyzing physical activities for Alzheimer’s patients [Li et al., 2018] and segmenting critical stages for sepsis patients [Gao et al., 2022]. Building upon this work, a multi-series time-aware Toeplitz inverse covariance-based clustering approach (MT-TICC) [Yang et al., 2021] was proposed, which takes multiple series as input and uses a time-awareness mechanism to handle irregular time intervals in EHRs. The MT-TICC-derived progressive patterns, when used as additional information of observations, enable more accurate early prediction for septic shock, outperforming competitive baselines.

It is important to recognize that clinicians’ *interventions* play a significant role in patient’s disease progression [Kmorowski et al., 2018; Azisoltani et al., 2019], yet this is not sufficiently addressed in the majority of existing works. Though some prior studies have included interventions as additional features in DPM [Esteban et al., 2016; Goh et al., 2021], they do not consider the effectiveness of interventions.

### 2.2 RL & AL for EHRs

Reinforcement learning (RL) has been widely applied in EHRs for dynamic treatment regimes, which aims at inducing a decision-making policy to dictate how the interventions should be executed so that the patients can gain improved outcomes [Yu et al., 2021]. As an input of RL, the reward function plays a critical role in praising/punishing the learning model to derive an optimal policy. However, manually specifying an appropriate reward function is usually expertise-intensive and time-consuming, posing a significant barrier to the broader applicability of RL [Abbeel and Ng, 2004]. *Apprenticeship learning* (AL) [Ng et al., 2000] tackles such problems by directly learning the reward function from experts’ demonstrations. Behavior cloning [Raza et al., 2012] is a classic AL method, which directly learns a mapping from states to actions greedily imitate experts’ demonstrated behaviors [Ross et al., 2011]. Later, various Inverse RL (IRL) [Abbeel and Ng, 2004; Ziebart et al., 2008] and adversarial imitation learning [Ho and Ermon, 2016; Finn et al., 2016] based AL approaches have been proposed, but they are often online, requiring iteratively executing the latest policy to collect data for updating the model. The execution of a bad policy in healthcare is unethical [Levine et al., 2020], making offline AL methods desired. Though some online approaches have been adapted to offline [Chan and van der Schaar, 2021; Kostrikov et al., 2018], they usually rely on off-policy evaluation, which itself is nontrivial with imperfect solutions.

More recently, [Jarrett et al., 2020] introduced an AL approach named energy-based distribution matching (EDM) and evaluated it on various benchmarks, e.g., Acrobat, LunarLander, and BeamRider. Their results demonstrated that EDM outperformed both IRL-based and adversarial imitation learning-based methods [Brockman et al., 2016]. Therefore, we employ EDM as a baseline in this paper. When applying EDM to EHRs for modeling clinicians’ interventions, it makes a strong assumption that all demonstrations are generated with a unified policy, following a *single re-
ward function. However, clinicians commonly execute different policies with multiple reward functions when treating patients under different progressive stages [Wang et al., 2021; Wang et al., 2022]. To handle the multiple reward functions, some improved AL have been proposed [Dimitrakakis and Rothkopf, 2011; Babes et al., 2011]. For example, [Babes et al., 2011] developed an EM-based IRL, which assumes reward functions are diverse across different demonstrations, while within each demonstrated sequence, the reward function is assumed to be unified. To model the multiple reward functions evolving over time, several more recent works have been proposed [Krishnan et al., 2016; Hausman et al., 2017; Wang et al., 2021; Wang et al., 2022]. However, these methods generally partition the demonstrations into fixed-length subsequences to learn their respective reward functions, without considering the irregular intervals during the partitioning.

Furthermore, existing AL methods typically concentrate on learning interventional patterns for either inducing a decision-making policy or evaluating whether interventions are carried out as expected. However, none of these methods have been incorporated to learn the progressive stages during the DPM.

3 THEMES for DPM

Figure 1 provides an illustration of how THEMES-derived patterns can be utilized for DPM. The entire input data $D_{\text{Entire}}$ consists of $L$ sequences represented as $\{D^l\} = \{(x^l_t, a^l_t) | t = 1, \ldots, T^l; l = 1, \ldots, L\}$, where $x^l_t \in \mathbb{R}^m$ is the $t$-th multivariate observed state with $m$ features, and $a^l_t$ is the corresponding action in the $l$-th sequence with the length of $T^l$. In the context of AL, it is typically assumed that the experts’ demonstrations provided as input are optimal or near-optimal following a latent reward function [Abbeel and Ng, 2004]. The quality of these demonstrations plays an important role in inducing more accurate policies. Therefore, we have designed a procedure for determining the experts’ demonstrations (detailed in Section 4.1), based on which a subset of $N (N \leq L)$ sequences $D_{\text{Expert}}$ are selected, denoted as: $\{D^n\} = \{(x^n_t, a^n_t) | t = 1, \ldots, T^n; n = 1, \ldots, N\}$.

Taking $D_{\text{Expert}}$ as input, THEMES aims at learning the multiple underlying reward functions evolving over time. As illustrated in Figure 1 (Left), THEMES has a hierarchical structure, with two major components at its low-level, i.e., RMT-TICC and EM-EDM. Specifically, RMT-TICC operates on the states in $D_{\text{Expert}}$ as input. It partitions and clusters the subsequences, such that each subsequence cluster exhibits the same time-invariant patterns, which is regarded as a high-level state. Then focusing on the state-action pairs over the partitioned subsequences, EM-EDM will cluster and induce their policies, such that each cluster exhibits consistent decision-making patterns, which is referred to as a high-level action. Subsequently, taking the high-level state-action pairs as input, a high-level reward regulator will be learned by IRL and fed back to refine the RMT-TICC. This iterative procedure continues until convergence is achieved.

Once the THEMES model has been learned, we apply it over $D_{\text{Entire}}$ for extracting progressive patterns from RMT-TICC and interventional patterns from EM-EDM. These patterns are then combined with the observations from the original data, which will further serve as an input of LSTM for septic shock early prediction, as depicted in Figure 1 (Right).

3.1 Subsequence Partitioning by RMT-TICC

Preliminaries

Given experts’ demonstrations, denoting the states in $D_{\text{Expert}}$ as $\{x^n_t | t = 1, \ldots, T^n\}$, the objective of subsequence clustering is partitioning and clustering the $\{x^n_t\}$ into subsequences based on their latent time-invariant patterns. It can be achieved by learning a mapping from each state to a specific cluster $\{k | k = 1, \ldots, K\}$. To capture the interdependence among neighboring states, a sliding window of length $\omega \ll T^n$ is employed to incorporate the context information. When assigning a state $x^n_t$ into a cluster $k$, we also consider its preceding states within the sliding window, i.e., $X^n_{t-\omega} = \{x^n_{t-\omega+1}, \ldots, x^n_t\}$, where $X^n_t$ is a random variable of dimension $m\omega$, concatenating the $m$-dim states in $\omega$.

To learn the subsequence clusters, we fit $X^n_t$ into $K$ Gaussian distributions, with each distribution indicating a specific cluster $k$. It can be modeled by Toeplitz inverse covariance-based clustering [Hallac et al., 2017] to learn the mean and inverse covariance matrix for each cluster. More specifically, the mean vectors, $\{\mu_k | k = 1, \ldots, K\}$, are determined by assigning each state to an optimal cluster, resulting in clustering assignments $P = \{P_k | k = 1, \ldots, K\}$, where $P_k \subset \{1, \ldots, T^n\}$ is the indices of states (sliding windows) belonging to cluster $k$. The inverse covariance matrices, $\Theta = \{\Theta_k | k = 1, \ldots, K\}$, are estimated to characterize the time-invariant structural patterns for each cluster. Herein, $\Theta_k$ is constrained to be blockwise Toeplitz, composed of $\omega$ sub-blocks $A^{i} \in \mathbb{R}^{m \times m}$, $i \in [0, \omega - 1]$, where the sub-block $A^{(i)}$ represents the partial correlations among $m$ features between timestamp $t$ and $t+i$. For instance, the $(p, q)$-th element in $A^{(i)}$ represents the partial correlation between the $p$-th feature at $t$ and the $q$-th feature at $t+i$, where $p, q \in \{1, \ldots, m\}$.

A multi-series time-aware Toeplitz inverse covariance-based clustering (MT-TICC) [Yang et al., 2021] was proposed recently, improving the subsequence clustering from two perspectives: First, MT-TICC operates on multi-series data, allowing for improved estimation of the mean and inverse covariance matrix by considering shared patterns across different sequences. This capability is particularly advantageous in real-world scenarios like EHRs, where the data is...
collected from various patients. Second, MT-TICC incorporates time-awareness to handle irregular intervals by using a decay function to constrain the consistency within each cluster. It is highly preferred when dealing with data collected irregularly, such as in healthcare settings. However, when applied to DPM, MT-TICC primarily focuses on the observed states, and it does not fully account for the interventional patterns conveyed by state-action pairs, which can significantly impact the disease progression [Komorowski et al., 2018].

**RMT-TICC**

To incorporate the interventional patterns conveyed by state-action pairs, a Reward-regulated MT-TICC (RMT-TICC) is leveraged in THEMES, considering the fact that the interventions are driven by underlying reward functions. The objective function of RMT-TICC is as follows:

\[
\argmin_{\theta, \rho} \sum_{k=1}^{K} \sum_{n=1}^{N} \left[ -\ell(X^n, \Theta_k) + c(X^n, \Theta_k, \Delta T^n, \Delta r^n) \right] + \lambda \left\| \Theta_k \right\|_1
\]

**Log-likelihood term** measures the probability that \(X^n\) belongs to the cluster \(k\). Specifically, assuming \(X^n\) is a Gaussian distribution with the mean of \(\mu_k\) and the inverse covariance matrix of \(\Theta_k^{-1}\), the log-likelihood term is defined as:

\[
el(X^n, \Theta_k) = -\frac{1}{2} (X^n - \mu_k)^T \Theta_k (X^n - \mu_k) + \frac{1}{2} \log |\Theta_k| - \frac{m}{2} \log (2\pi)
\]

**Reward-regulated Time-aware Consistency term** encourages the consecutive states \(\{X_{t-1}, X_t\}\) to be assigned into the same cluster, by taking account of both the time intervals and the corresponding rewards. It penalizes the neighbor states belonging to different clusters by minimizing Eq.(3):  

\[
c(X^n, \Theta_k, \Delta T^n, \Delta r^n) = \frac{\beta}{\Phi(\Delta T^n, \log(e + \Delta r^n))} + \frac{1}{2} \left\| \Theta_k \right\|_1
\]

where \(\beta\) is a weight parameter. \(1 \{ t \neq t' \in P_k \}\) is an indicator function, with the value of 1 if \(X_{t-1}^n\) and \(X_t^n\) do not belong to the same cluster; otherwise its value is 0. 

In above equation, \(\beta\) is a discount factor. Then, Eq.(1) simplifies to include only the log-likelihood term and the consistency term. It can be solved by dynamic programming to find a minimum cost Viterbi path \([Viterbi, 1967] \); In M-step, by fixing \(\Theta\) to learn \(\Theta\), Eq.(1) reduces to include only the log-likelihood term and the sparsity term. It can be formulated as a typical graphical lasso \([Friedman et al., 2008]\) with a Toeplitz constraint over \(\Theta\) and be solved by an alternating direction method of multipliers \([Boyd et al., 2011]\). We iteratively perform the E-step and M-step until convergence.

### 3.2 Policy Induction by EM-EDM

**Preliminaries**

Energy-based distribution matching (EDM) [Jarrett et al., 2020] is a strictly offline AL method that learns a policy solely from expert demonstrations \(\{D^n\}\) without requiring knowledge about model transitions or off-policy evaluations. It assumes that the \(\{D^n\}\) are carried out with a policy \(\Pi^*\) parameterized by \(\theta\), driven by a single reward function.

For simplicity, we will denote a state-action pair as \((x, a)\), omitting their indexes when it does not cause ambiguity. The occupancy measures for the demonstrations and for the learned policy are denoted as \(\rho_D\) and \(\rho_{\Pi^*}\), respectively. The probability density for each state-action pair can be measured as \(\rho_{\Pi^*}(x, a) = \frac{1}{\sum_{x, a} \gamma^0} \mathbb{E}_{\xi}[\sum_{t=0}^{\infty} \gamma^t I\{x_t = x, a_t = a\}]\), where \(\gamma\) is a discount factor. Then, the probability density for each state can be measured by: \(\rho_{\Pi^*}(x) = \sum_{a} \rho_{\Pi^*}(x, a)\). The goal of inducing the policy \(\Pi^*\) can be achieved by minimizing the KL divergence between \(\rho_D\) and \(\rho_{\Pi^*}\):

\[
\argmin_{\theta} D_{KL}(\rho_D || \rho_{\Pi^*}) = \argmin_{\theta} -\mathbb{E}_{x, a \sim \rho_D} \log \rho_{\Pi^*}(x, a)
\]
Since $\Pi^0(a|x) = \rho_{11v}(x, a)/\rho_{11v}(x)$, the objective function can be reformulated as:

$$\arg\max_{\Pi^0} -E_{x\sim p_D} \log \rho_{11v}(x, a) - E_{x, a \sim p_D} \log \Pi^0(a|x) \quad (6)$$

When it is not possible to execute the policy $\Pi^0$ in an online manner, estimating $\rho_{11v}(x)$ in the first term of Eq.(6) becomes challenging. EDM addresses this issue by employing an energy-based model [Grathwohl et al., 2019].

According to energy-based model, the probability density $\rho_{11v}(x)$ is proportional to $e^{-E(x)}$, where $E(x)$ is an energy function. The occupancy measure for state-action pairs can be obtained by marginalizing out the actions: $\rho_{11v}(x) = \sum_a e^{f_{11v}(x, a)} / Z_{11v}$. Herein, $Z_{11v}$ is a partition function, and $f_{11v}: \mathbb{R}^{11v} \rightarrow \mathbb{R}$ is a parametric function that maps each state to $|A|$ real-valued numbers.

The parameterization of $\Pi^0$ implicitly defines an energy-based model over the states distribution, where the energy function can be defined as: $E_{11v}(x) = -\log \sum_a e^{f_{11v}(x, a)}$. Within the scope of the energy-based model, the first term in Eq.(6) can be reformulated as an occupancy loss:

$$L_{\theta}(\theta) = E_{x\sim p_D} E_{11v}(x) - E_{x\sim p_{11v}} E_{11v}(x) \quad (7)$$

where $\nabla_{\theta} L_{\theta}(\theta) = -E_{x\sim p_{11v}} \nabla_{\theta} \log \rho_{11v}(x)$ can be solved by existing optimizers, e.g., stochastic gradient Langevin dynamics [Welling and Teh, 2011]. Thus, by substituting the first term in Eq.(6) with Eq.(7) via energy-based model, we can derive a surrogte objective function to get the optimal solution without the need for online policy rollouts.

**EM-EDM**

To handle multiple reward functions varying across demonstrations, an EM-based inverse reinforcement learning approach was proposed in [Babes et al., 2011]. It iteratively clusters the demonstrations in $E$-step and induces policies for each cluster by IRL in $M$-step. However, in $M$-step, it relies on IRL methods with discrete states, which may not be scalable for large continuous state spaces like EHRs. To handle the continuous states, motivated by the EM framework and the success of EDM, THEMES employs an extended method—EM-EDM, with EDM in the $M$-step of EM.

Taking subsequences $\{D^n|n = 1, \ldots, N\}$ learned by RMT-TICC as input, where $N$ represents the number of subsequences, the goal of EM-EDM is to cluster these subsequences and learn cluster-specific policies $\{\Pi_g|g = 1, \ldots, G\}$, with $G$ being the number of clusters. The prior probability for each cluster is denoted as $\nu_g$ and the policy parameter is denoted as $\theta_g$. Both $\nu_g$ and $\theta_g$ are randomly initialized. The objective function of EM-EDM is shown in Eq.(8):

$$\arg\max_{\theta_g} \mathcal{L} = \sum_{g=1}^G \sum_{n=1}^N \log(u_{n,g}) \quad (8)$$

where $u_{n,g}$ denotes the probability that subsequence $D^n$ follows the policy of the $g$-th cluster. It is defined in Eq.(9), with $U$ being a normalization factor.

$$u_{n,g} = Pr(D^n|\theta_g) = \prod_{(x, a)\in D^n} \Pi_{g}(x, a) \nu_g / U, \quad (9)$$

During the EM, in the $E$-step, the probability that subsequence $D^n$ belonging to the cluster $g$ is calculated by Eq.(9). Then, in the $M$-step, the prior probabilities are updated as $\nu_g = \sum_{n} u_{n,g} / N$, and the policy parameters $\theta_g$ are learned by EDM. The $E$-step and $M$-step are iteratively executed until convergence. Finally, the output of EM-EDM consists of the clustered subsequences along with their respective policies.

Based on the subsequence clusters learned by RMT-TICC and the decision-making policies learned by EM-EDM in THEMES, we extract progressive patterns and interventional patterns from the entire input data $D^{\text{Entire}}$. For each timestamp $(x_t, a_t)$, based on Eq.(2), we calculate the probabilities that $x_t$ belonging to each progressive stage. Additionally, we calculate the probabilities that $(x_t, a_t)$ follows each policy. These probabilities serve as additional features that are concatenated with the original observations $x_t$, resulting in an augmented input for the downstream early prediction task. In this paper, we fix the early predictor as LSTM [Zhang et al., 2017] and focus on evaluating the effectiveness of feature engineering. We believe that by combining the patterns derived from THEMES with more advanced prediction models (e.g., [Baytas et al., 2017; Zhang et al., 2019], the performance of early prediction can be further enhanced.

**4 Experiments**

To assess the effectiveness of THEMES-derived patterns for DPM, we applied it to an EHRs dataset obtained from the Christiana Care Health System. This dataset spans over a period of two and a half years, where each sequence represents a patient’s visit consisting of a series of observations and corresponding clinicians’ interventions.

**4.1 Data Preprocessing**

Our sepsis-related study cohort comprised 52,919 visits (sequences) with suspected infection, consisting of 4,224,567 timestamps. We conducted data preprocessing as follows:

- **Feature selection**: We consulted clinicians and selected 14 sepsis progression-related features: 1) Vital signs: systolic blood pressure, mean arterial pressure, respiratory rate, oxygen saturation, heart rate, temperature, fraction of inspired Oxygen; and 2) Lab results: white blood cell, bilirubin, blood urea nitrogen, lactate, creatinine, platelet, neutrophils.
- **Missingness handling**: We address the missing data by an expert-suggested forward-filling (8 hours for vital signs and 24 hours for lab tests), with the remaining missing values imputed as the mean. This method has shown robustness, especially for septic shock early prediction [Zhang et al., 2019].
- **Tagging septic shock visits**: Clinical labeling commonly relies on diagnosis codes, such as ICD-9, which are primarily intended for administrative and billing purposes, leading to constrained reliability [Zhang et al., 2019]. To address this issue, our clinicians referred to the Sepsis-3 guidelines [Singer et al., 2016] and established specific rules for identifying septic shock. By combining the ICD-9 codes and clinicians’ rules, we identified 1,869 shock and 23,901 non-shock visits. To handle the highly imbalanced ratio, we employed a stratified random sampling technique over the non-shock visits, which
ensured that the dataset maintained the original distribution of age, sex, ethnicity, and stay duration, while achieving a balanced amount. Thus, the non-shock visits were refined to include 1,869 visits, equaling to the number of shock visits.

- **States & Actions**: a) **States** are defined based on the 14 continuous features that are relevant to the progression of sepsis. In addition to the original features, we calculate the maximum and minimum values observed within the past 1 hour for each feature. This allows us to capture temporal changes and trends in the data. As a result, the states are represented as 42-dimensional vectors. b) **Actions** are binary values indicating whether specific antibiotics (e.g., clindamycin, daptomycin) have been administered or not. It is well established, as suggested in [Gauer, 2013], that antibiotic therapy plays a crucial role in enhancing the clinical outcomes of sepsis patients.

- **Determining Experts’ Demonstrations**: To enhance the selection of high-quality demonstrations for inducing more accurate policies, we developed a strategy to evaluate the effectiveness of interventions. This strategy involved comparing the early prediction of septic shock before the first intervention to the actual outcome at the onset of septic shock, as depicted in Figure 2. To assess the effectiveness of interventions, we trained multiple LSTMs (100 times) using different hyperparameter settings over the data prior to the first intervention for randomly selected visits. For a given visit, if more than 80% of the LSTMs produce the same early prediction result, we compare this prediction with the actual onset of septic shock: If a patient transitions from shock (prior to the first intervention) to non-shock (onset of septic shock), it indicates the interventions were effective, suggesting the visit to be treated as an experts’ demonstration for policy induction by THEMES; Otherwise, the visit would not be included during the THEMES modeling process. Using this approach, we identified 195 sequences as experts’ demonstrations.

4.2 Experimental Settings

We compared the following methods for evaluating the effectiveness of progressive and interventional patterns in DPM:

- **Original**: Except for the Original observed state features, none of other additional information will be incorporated;
- **MT-TICC**: Observed state features are supplemented with additional progressive patterns learned by MT-TICC [Yang et al., 2021]. The additional features are extracted as the probabilities belonging to each subsequence cluster.
- **Action**: Interventional Actions are directly taken as additional features for the observed state features.
- **EDM**: An additional interventional feature is extracted by EDM-learned policy [Jarrett et al., 2020], represented as the probability of following the learned policy.
- **THEMES**: Based on THEMES, the probabilities belonging to RMT-TICC-learned subsequence clusters and the probabilities following EM-EDM-learned policies are respectively taken as additional progressive and interventional features.

As demonstrated in prior works, EDM outperformed competitive AL methods with a single reward function [Jarrett et al., 2020]. Meanwhile, THEMES_0 and THEMES could induce more accurate policies compared to competitive AL baselines with multiple reward functions (e.g., hierarchical IRL [Krishnan et al., 2016] and multi-modal imitation learning [Hausman et al., 2017]) as well as their ablations (e.g., EM-EDM, and learning progressive stages by MT-TICC then inducing cluster-wise policy by EDM) [Yang et al., 2023]. Therefore, in this paper, we employed EDM, THEMES_0, and TMEMES for extracting AL-derived patterns for DPM.

We utilized Keras to implement the LSTM and performed parameter tuning through grid search. The results were measured when $\tau = 12$ and $\tau \in \{12, 36\}$, for evaluating if septic shock prediction could be achieved 12 hours prior to the onset or even earlier by 36 hours. Each method was repeated 10 times, with the data randomly divided into 80% for training and 20% for testing. The evaluation metrics include Accuracy (Acc), Recall (Rec), Precision (Prec), F1-score (F1), and AUC. All parameters in THEMES are determined by 5-fold cross-validation. In RMT-TICC, the cluster number $K$ is set to 11 based on Bayesian information criteria (BIC) [Friedman et al., 2001], the window size $w$ is set to 2, and the sparsity $\lambda$ and consistency $\beta$ coefficients are set to 1e-5 and 4, respectively. In EM-EDM, the cluster number is determined heuristically as 3, by iteratively applying the EM algorithm until

### Table 1: Early prediction in EHRs. The best methods are in bold with *, and the second-best is in bold only.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Hold-off Window $\tau = 12$</th>
<th>Hold-off Window $\tau \in {12, 36}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Acc</td>
<td>Rec</td>
</tr>
<tr>
<td>Original</td>
<td>.754(0.013)</td>
<td>.743(0.012)</td>
</tr>
<tr>
<td>MT-TICC</td>
<td>.803(0.010)</td>
<td>.802(0.012)</td>
</tr>
<tr>
<td>Action</td>
<td>.757(0.013)</td>
<td>.754(0.013)</td>
</tr>
<tr>
<td>EDM</td>
<td>.765(0.013)</td>
<td>.753(0.012)</td>
</tr>
<tr>
<td>THEMES_0</td>
<td>.809(0.011)*</td>
<td>.837(0.012)*</td>
</tr>
<tr>
<td>THEMES</td>
<td>.834(0.011)*</td>
<td>.820(0.012)</td>
</tr>
</tbody>
</table>

Figure 2: Determining experts’ demonstrations by comparing early prediction before the first intervention vs. the onset of septic shock.
empty clusters are generated or the log-likelihood varies less than a predefined threshold. The THEMES approach uses a threshold of 10 iterations, as our observed that the clustering likelihood for both MT-TICC and EM-EDM converges within 10 iterations. To ensure fair comparisons, optimal parameters for other baselines are also determined by cross-validation.

4.3 Results

The experimental results are reported in Table 1. The best results are in bold highlighted with * and the second-best results are in bold only. Additionally, we provide Critical Difference diagrams with Wilcoxon signed-rank tests over F1 and AUC in Figure 3. In the diagrams, unconnected models indicate pairwise significance at a confidence level of 0.05.

According to the results: a) For progressive patterns: Comparing MT-TICC to Original, the incorporation of progressive patterns in MT-TICC yields significant improvements. Moreover, among the four methods incorporating interventions (i.e., Action, EDM, THEMES_0, and THEMES), the two with progressive patterns (i.e., THEMES_0 and THEMES) outperform the others (i.e., Action and EDM). As a result, the inclusion of progressive patterns enables a better capture of progressive stages, leading to improved early prediction. b) For interventional patterns: Comparing Action and EDM vs. Original, the Action yields similar performance to Original, while EDM shows slightly better performance. Additionally, comparing THEMES_0 and THEMES vs. MT-TICC, THEMES_0 is marginally better than MT-TICC, while THEMES further improves the performance. Thus, directly taking actions as additional features via Action cannot fully capture the clinicians’ decision-making patterns, while incorporating AL-learned patterns can better reflect the effectiveness of interventions; c) For both progressive and interventional patterns: THEMES performs better than THEMES_0, which indicates the effectiveness of using interventional patterns to refine the learned progressive stages.

Figure 4 shows the F1 and AUC for septic shock early prediction ($\tau \in [1, 36]$) with additional features learned by different methods.

Figure 5 demonstrates the projection of THEMES-derived features for 100 patients whose interventions best match the learned policies (High Match) and 100 patients with the lowest matching rate (Low Match) on a 2D scatter plot utilizing t-SNE. One shock (red) and one non-shock (green) patient with the same length are randomly sampled. Though the start points for the two patients are close, they drift apart as the non-shock patient’s treatments (green) align with the High Match group, while the shock patient’s treatments (red) align with the Low Match group. It demonstrates that THEMES can capture the effectiveness of treatments for progression to distinguish shock patients from non-shock patients.

5 Conclusions

In this paper, we explore incorporating AL-derived patterns for DPM. Building upon the success of subsequence clustering in extracting progressive patterns for DPM, we leverage an AL approach named THEMES to capture both progressive and interventional patterns. Taking advantage of these patterns, we aim to handle a challenging task for septic shock early prediction. The experimental results demonstrate that the inclusion of THEMES-derived patterns leads to improved accuracy in predicting septic shock at earlier stages. This advancement holds significant potential for assisting clinicians in delivering timely and personalized treatments.

Acknowledgements

This research was supported by the NSF Grants: #2013502, #1726550, #1660878, and #1651909.
References


