Personalised Medicine in Critical Care Using Bayesian Reinforcement Learning

Chandra Prasetyo Utomo[1](✉), Hanna Kurniawati[1], Xue Li[1,2], and Suresh Pokharel[1]

[1] The University of Queensland, Brisbane, QLD 4072, Australia
{c.utomo,hannakur,s.pokharel}@uq.edu.au
[2] Neusoft Education Technology Group, Dalian 116023, People's Republic of China
lixue@neusoft.edu.cn

Abstract. Patients with similar conditions in the intensive care unit (ICU) may have different reactions for a given treatment. An effective personalised medicine can help save patient lives. The availability of recorded ICU data provides a huge potential to train and develop the systems. However, there is no ground truth of best treatments. This makes existing supervised learning based methods are not appropriate. In this paper, we proposed clustering based Bayesian reinforcement learning. Firstly, we transformed the multivariate time series patient record into a real-time Patient Sequence Model (PSM). After that, we computed the likelihood probability of treatments effect for all patients and cluster them based on that. Finally, we computed Bayesian reinforcement learning to derive personalised policies. We tested our proposed method using 11,791 ICU patients records from MIMIC-III database. Results show that we are able to cluster patient based on their treatment effects. In addition, our method also provides better explainability and time-critical recommendation that are very important in a real ICU setting.

Keywords: Personalised medicine · Treatment recommendation · Intensive care unit · Time series · Bayesian reinforcement learning

1 Introduction

Personalised treatment recommendation is one of the most desired applications of intensive care unit (ICU) decision support systems. Research in sepsis related patients shows that ICU patients in a similar condition had difference responses to a set of Vasopressor treatments [10]. Some patients responded properly (getting a better condition), other set of patients had complications (getting a worse condition), and the remaining patients did not respond at all. Every four years, ICU community updates their best practise guidelines to deal with sepsis based on recent evidence based medicine research [12]. This process is expensive, time consuming, and possibly has conflicting results.

The availability of ICU database has opened the opportunity to develop a data driven approach for personalised medicine. We aim to develop a personalised

J. Li et al. (Eds.): ADMA 2019, LNAI 11888, pp. 648–657, 2019.
https://doi.org/10.1007/978-3-030-35231-8_47

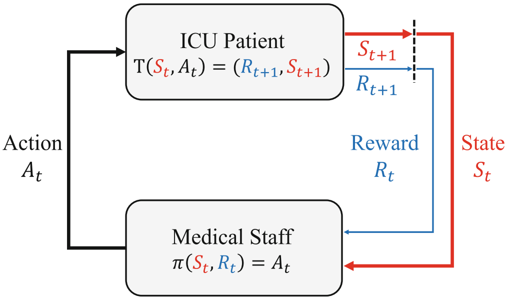


Fig. 1. The interaction between ICU patient and medical staff modeled in reinforcement learning. At a timestamp t, medical staff analyse the current state of patient S_t and reward R_t (scalar value indicating how good the patient condition). Based on a medical policy π, they perform action A_t. The patient response the treatment on current state by revealing the next state S_{t+1} and reward R_{t+1}.

treatment recommendation for sepsis related patients using existing ICU data. We choose sepsis related patients because sepsis is the most frequent cause of death in ICU [13]. In addition, both the definition of sepsis [15] and its effective treatments are still subject of future medical research.

There are two main challenges to do precision medicine in ICU. Firstly, there is no ground truth of best treatments. The current recorded treatments are subject to evaluation for future best practise guidelines. This makes existing supervised learning based recommendation systems [3] are not appropriate. Secondly, the ICU record is a multivariate time series data with common missing values and no proper alignment. Existing recommendation systems [11] were using a fix interval to sequence these data. However, this approach is not applicable as ICU demands a real-time decision making process.

In this paper we proposed personalised medicine based on Bayesian reinforcement learning. Firstly, we developed a real-time multivariate time series sequencing technique called Patient Sequence Model (PSM). Figure 1 illustrates our sequence model. After that, we calculated the likelihood of treatment responses to generate a meaningful feature representation for each patient. Then, we clustered the patient based on their treatment responses. Finally, we define a mathematical model to compute a personalised policy.

There are three main contributions of this paper.

- **Explainable Sequence Model**. We are able to visualise a meaningful state transition diagram represented patient dynamics in ICU.
- **Discovery of Patient Clusters**. We are able to cluster the patients into several groups based on patient responses to the treatments.
- **Personalised Policy Computation**. We define a framework to recommend a personalised treatment as combination of policies from all clusters.

The rest of the paper is organised as follow. Section 2 describes related work. The proposed methods is explained in Sect. 3. Evaluation is provided in Sect. 4. Finally, Sect. 5 gives concluding remark and direction of future work.

2 Related Work

The first set of attempts in treatment recommendation were based on expert systems [1,5]. However, it is difficult and costly to get knowledge from domain expert. As the availability of electronic health records (EHR), data driven method became a reasonable direction. The first set of methods are supervised learning based methods. Hu et al. [6] used similarity method, Cheerla et al. [4] integrated genomics data, and Bajor et al. [3] proposed deep model. However, these methods are not suitable in ICU as ground truth of treatment is unclear. A more reasonable data driven approach is using reinforcement learning. Tsoukalas et al. [16] modeled the treatment recommendation problem as Partially Observable Markov Decision Process (POMDP) while Nemati et al. [9] and Raghu et al. [11] modeled it as deep reinforcement learning. However, these methods sequenced the ICU patients based on a fix length interval, e.g. 4 h. This is not applicable in real ICU setting as it demands real-time decision making. In addition, their models are not explainable to domain expert which is important in this sensitive application.

3 Proposed Methods

Markov Decision Process (MDP) is defined by tuple ($\mathcal{S}$, $\mathcal{A}$, T, R) where:

- $\mathcal{S} : S_1 \times S_2 \times \ldots \times S_{ns}$, is the set of states of the system; $S_1, \ldots, S_{ns}$ correspond to the domain of the ns state variables (features). We define qSOFA variables ($ns = 3$) as state variables. So that, we have $S_1 = ABPSystolic$, $S_2 = RespiratoryRate$, and $S_3 = Mentation$. We added two terminal states *survived* and *dead* to the state space.
- $\mathcal{A} : A_1 \times A_2 \times \ldots A_{na}$, is the set of actions that can be performed by the agent. We define vasopressor treatments ($na = 5$) as the action set. Here, we have $A_1 = Epinephrine$, $A_2 = Dopamine$, $A_3 = Phenylephrine$, $A_4 = Norepinephrine$, and $A_5 = Vasopression$.
- $T : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$ is the transition function, where $T(s, a, s') = P(s'|s, a)$ represents the conditional probability of moving to state $s' \in \mathcal{S}$ if the agent executes action $a \in \mathcal{A}$ in state $s \in \mathcal{S}$.
- $R : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow \mathbb{R}$, the reward function, encodes a reward earned when state s' is reached after executing action a in state s. We defined $R(s, a, survival) = +100$, $R(s, a, death) = -100$, and $R(s, a, s') = -qSOFA(s')$ for all non-terminal states s'.

The goal of the MDP agent is to find an action selection strategy $\pi*$, called a policy, that maximises its long-term expected rewards. The optimal action to take in a state s is defined via the optimal value function V^* representing the return obtained by the optimal policy starting in state s:

$$V^*(s) = \max_{a \in A} \left[R(s, a, s') + \sum_{s' \in S} T(s, a, s') V^*(s') \right]. \qquad (1)$$

The optimal action in s is obtained by taking the *argmax* instead of the max in the Eq. 1.

3.1 Sequence Model

We proposed a real-time multivariate time series sequencing called Patient Sequence Model (PSM). Each ICU stay $I^{(i)} = [S_1, A_1, R_2, S_2, A_2, \ldots, R_{T_i}, S_{T_i}]$ is represented by a sequence of states, actions, and rewards. In this model, we did not divide timestep based on a fix length interval, e.g. 4 h. Instead, we extract a new timestep when there is a new observation. We believe that in the real ICU settings, medical staff makes decision based on new observation. They will not wait until a given fix interval to take action.

We discretised both state and action to have a better explainability. We discretised state variables into binary values *normal* and *abnormal*. We followed medical literature to define thresholds for all variables. As we have 3 state variables and 2 values for each variables, we have $2^3 = 8$ states. Since we added 2 terminal states, the total number of states in our state space is 10.

For action variables, we discretised into *true* (if the drug was administered) and $false$ (otherwise). As we have 5 action variables and 2 values for each variables, the total number of actions in our action space is $2^5 = 32$. Table 1a and b shows the combination of variables in state and action space, respectively. We left the definition of reward R_t the same.

3.2 Feature Engineering

We want to divide the whole patient cohort into several subgroups. Patients within a same subgroup should have a common responses of a set of treatments. On the other hand, patients from two different subgroups should have different responses. By correctly identifying subgroups of patients, we will be able to deliver a more personalised treatment for a new patient.

We designed a feature that reflects patient responses for all treatments at any given state. For each ICU stay $I^{(i)}$, we calculated feature representation $X^{(i)}$ as

$$X^{(i)} = T(S_t, A_t, S_{t+1}) = P(S_{t+1}|S_t, A_t). \tag{2}$$

$X^{(i)} \in \mathbf{R}^{|S_t| \times |A_t| \times |S_{t+1}|}$ is a transition function of current state S_t, action A_t, and next state S_{t+1}. We can unfold the first dimension into $X^{(i)} = [X_1^{(i)}, X_2^{(i)}, \ldots, X_8^{(i)}]$, where

Table 1. State definition (a) and action definition (b). The reward function is associated with the qSOFA values defined in table (a).

States	ABP Systolic	Respiratory Rate	Mental Status	qSOFA
s_1	normal	normal	normal	0
s_2	normal	normal	abnormal	1
s_3	normal	abnormal	normal	1
s_4	normal	abnormal	abnormal	2
s_5	abnormal	normal	normal	1
s_6	abnormal	normal	abnormal	2
s_7	abnormal	abnormal	normal	2
s_8	abnormal	abnormal	abnormal	3

(a) State space

Actions	Epinephrine	Dopamine	Phenylephrine	Norepinephrine	Vasopressin
a_0	false	false	false	false	false
a_1	false	false	false	false	**true**
a_2	false	false	false	**true**	false
a_3	false	false	false	**true**	**true**
a_4	false	false	**true**	false	false
a_5	false	false	**true**	false	**true**
a_6	false	false	**true**	**true**	false
a_7	false	false	**true**	**true**	**true**
...	...	...	...	...	...
a_{31}	**true**	**true**	**true**	**true**	**true**

(b) Action space

$$X_j^{(i)} = P(S_{t+1}|S_t = s_j, A_t). \tag{3}$$

Here, $X_j^{(i)}$ is the conditional probability of i^{th} ICU stay in state s_j. The index j can take value from 1 to 8 because we have $|S_t| = 8$ non-terminal states in our definition (see Table 1a).

For each ICU stay $I^{(i)}$, we calculated the likelihood of conditional probability $\hat{P}(S_{t+1} = s_l|S_t = s_j, A_t = a_k)$ as

$$\hat{P}(s_l|s_j, a_k) = \frac{1}{N(S_t = s_j, A_t = a_k)} \sum_{t=1}^{T} 1(S_t, A_t, S_{t+1} = s_j, a_k, s_l). \tag{4}$$

Figure 3 shows the conditional probability of ICU ID = 204176 in state s_4 (abnormal respiratory rate and abnormal mental status). We may see that the patient had been given 7 different treatments when she/he was in this state.

3.3 Personalised Policy

We flattened three dimensional matrix $X^{(i)}$ into a single long vector $\hat{X^{(i)}}$. Number of non-terminal states $|S_t| = 8$ (see Table 1a), number of action states $|A_t| = 32$ (see Table 1b), and number of all possible states $S_{t+1} = 10$ (8 non-terminal states and 2 terminal states). So that, $\hat{X}^{(i)} \in \mathbf{R}^{2,560}$. To have a better visualisation of all patients in the experimental data, we reduced the dimension using PCA and projected in the first three dimensions in Fig. 5a.

The existing methods were calculated a single general policy π_g^* from the whole training data. Then, they applied this policy to the all test data. We believe that the patients can be segmented into some subgroups with a common similarity. We applied k-means clustering to the extracted feature $\hat{X^{(i)}}$ for all patient in training data. The clustering result $(c_1, c_2, \ldots, c_{nc})$ will reflect subgroups of patient with similar responses of treatments.

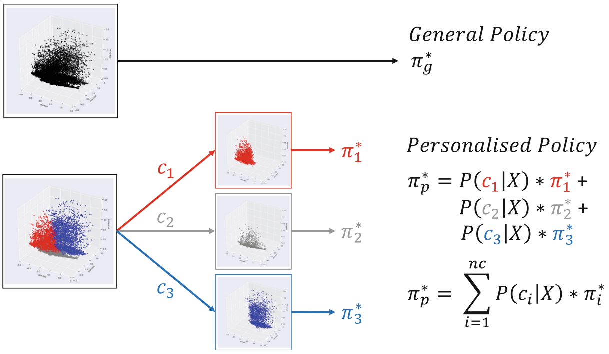

General Policy
π_g^*
c_1
c_2
c_3
π_1^*
π_2^*
π_3^*
Personalised Policy
$\pi_p^* = P(c_1|X) * \pi_1^* + P(c_2|X) * \pi_2^* + P(c_3|X) * \pi_3^*$
$\pi_p^* = \sum_{i=1}^{nc} P(c_i|X) * \pi_i^*$

Fig. 2. General policy versus personalised policy

We computed a dedicated policy π_i^* for subgroup c_i. We defined personalise policy π_p^* as

$$\pi_p^* = P(c_1|X) * \pi_1^* + P(c_2|X) * \pi_2^* + \ldots + P(c_{nc}|X) * \pi_{nc}^*. \tag{5}$$

Here, $P(c_i|X)$ is the believe that a new patient X is belong to cluster c_i. Figure 2 illustrates the different between general policy and personalised policy. Suppose we cluster the training data into 3 clusters, then we will have 3 subgroup policies. For n_c number of clusters, we can write the personalised policy as

$$\pi^* = \sum_{i=1}^{nc} P(c_i|X) * \pi_i^*. \tag{6}$$

4 Evaluation

4.1 Dataset

We used real-world ICU dataset from MIMIC-III (Medical Information Mart for Intensive Care) database [8]. We follow six exclusion criteria [7] in order to extract experimental data for optimal treatment recommendation related to sepsis. Firstly, we only considered records of patient admitted from 2008. After that, we excluded non-adults, non-primary admissions, cardiothoracic surgical service admissions, and admissions with missing data. Finally, we removed patient records with suspected of infection more than 24 h before and more than 24 h after ICU admission. The final patient cohort contained 11,791 patients.

4.2 Experimental Design

We divided our experiments into two main parts. Each part serves a dedicated purpose. In the first part, we want to test the explainability of our proposed sequence model. This is the key feature to discuss with domain experts, further our research in the right direction, and increase its applicability. We compute the likelihood of state transition probability $\hat{P}(S_{t+1}|S_t)$ using all dataset. After that, we visualise the result and generate state transition diagram.

In the second part, we want to test the effectiveness of our proposed feature representation. We need to see whether represent each patient based on their response to treatment is useful. We compute pairwise distance of all patients then group them into several number of clusters. We compare the clustering results with existing concepts in medical domain knowledge such as sepsis.

4.3 Result and Analysis

We computed probability of state transition from a current state S_t to a next state S_{t+1}. Figure 4a shows likelihood probability $\hat{P}(S_{t+1}|S_t)$ computed from all patient records in the experimental data. We can see that the diagonal of the figure is mostly dark blue ($\hat{P}(S_{t+1} = s_i|S_t = s_i)$ is close to 1. It suggests

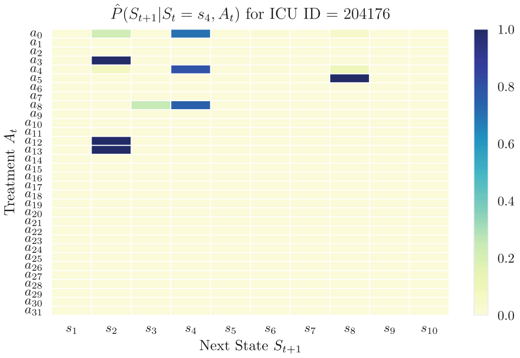
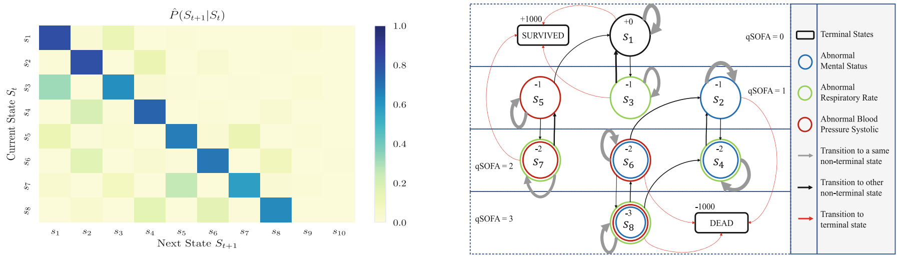


Fig. 3. A conditional probability $\hat{P}(S_{t+1}|S_t = s_4, A_t)$ for a patient (ICU ID = 204176) in current state $S_t = s_4$ (abnormal respiratory rate and mental status).


$\hat{P}(S_{t+1}|S_t)$
Current State S_t
s_1 s_2 s_3 s_4 s_5 s_6 s_7 s_8
s_1 s_2 s_3 s_4 s_5 s_6 s_7 s_8 s_9 s_{10}
Next State S_{t+1}
1.0
0.8
0.6
0.4
0.2
0.0
+1000
SURVIVED
+0
s_1
qSOFA = 0
-1
s_5
-1
s_3
-1
s_2
qSOFA = 1
qSOFA = 2
-2
s_7
-2
s_6
-2
s_4
qSOFA = 3
-3
s_8
-1000
DEAD
Terminal States
Abnormal Mental Status
Abnormal Respiratory Rate
Abnormal Blood Pressure Systolic
Transition to a same non-terminal state
Transition to other non-terminal state
Transition to terminal state

(a) State transition probability $\hat{P}(S_{t+1}|S_t)$.

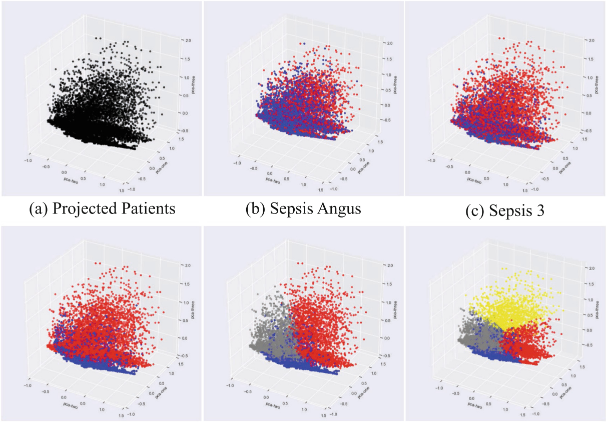
(b) State transition diagram.

Fig. 4. Patients dynamics in transition probability (a) and diagram (b). (Color figure online)

that most of the time, patient will stay in the same state for several timestamps before changing to another states. State s_3 and s_7 are the two states with greater probability to move to other states. This means, intervention in these states will more likely change patient conditions.

To justify personalised medicine idea, we need to compare the general state transition probability with a particular patient. For that reason, we observed a patient who received the most number of unique treatments as an extreme case. In our experimental data, a patient with ICU ID = 204176 received 18 treatments which is the highest of all patients. Figure 3 shows the conditional probability $\hat{P}(S_{t+1}|S_t = s_4, A_t)$ for the patient in the current state $S_t = s_4$.

We analyse some key features in Fig. 3. The patient received 7 type of treatments $\{a_0, a_3, a_4, a_5, a_8, a_{12}, a_{13}\}$ in state s_4 (abnormal respiratory rate and mental status). The Patient 100% went to a better state s_2 (abnormal respiratory) when given treatments $\{a_3, a_{12}, a_{13}\}$ and 100% went to a worse state s_8 (abnormal blood pressure, respiratory, and mental status) when given treatments $\{a_5\}$. On the other hand, the patient was most probably stay in the state s_4 when give



(a) Projected Patients (b) Sepsis Angus (c) Sepsis 3

(d) Clustering (*nc* =2) (e) Clustering (*nc* =3) (f) Clustering (*nc* =4)

Fig. 5. Patients clustering (Color figure online)

treatments $\{a_0, a_4, a_8\}$ with some small probability moved to other states. These unique features are significantly different with the overall case shown in Fig. 4a where mostly patient in state s_4 are stay in the same state. This suggests that developing treatment recommendation systems using overall dynamics from all patients will not be effective.

We visualise the likelihood of state transition probability into state transition diagram in Fig. 4b. The vertices are eight non-terminal states (circle) and two terminal states (rectangle). The lower the position of a state in the diagram, the worse its condition with respect to qSOFA score. The directed edges represent state transitions. The edge's width is proportional to its conditional probability $P(S_{t+1}|S_t)$. We filtered the edges with $P(S_{t+1}|S_t) < 0.1$. We added three edges with highest probability for each terminal states.

In this filtered diagram, we can see the dynamics such as common survival and mortality models. Most of the survived patients had normal mental status (no blue circle) in ICU discharge. We can also see that most of the patients we getting better during ICU stays. Between non-terminal states, there are 7 edges going up (better qSOFA score) and 4 edges going down (worse qSOFA score). This kind of visualization can be useful for better explainability of proposed method with medical practitioners.

We projected all patients into 3 dimensional space using PCA. Figure 5(a) shows the results. Every single dot in the figure represents a single patient. We color the patients based concepts in medical knowledge. In Fig. 5(b), we used the definition of sepsis by Angus et al. [2]. In Fig. 5(c), we used the third definition of sepsis or sepsis-3 [15] established in 2016. In both definitions, we may see that the sepsis patients (red) are somehow well separated with non-sepsis patients (blue). it means that our feature engineering method based one treatment responses is aligned with medical concept.

We want to compare both definitions of sepsis with our clustering results. The latest definition of sepsis has more identified sepsis patients (more red dots) and visually better separated in the figures. This may suggest that medical knowledge with respect to sepsis definition has been improved. Now, lets compare with the clustering results in Fig. 5(d) where they have the same number of groups. Interestingly, they also have a similar clusters shape. The difference is the clustering result have a lot more red cluster and even better separated.

As sepsis definition in critical care society is keep developing (the next is sepsis-4), the clustering result can help to make a better definition. Our medical collaborator is eager to investigate patients within the same group of majority of sepsis patients using data driven method but not identified as sepsis in the current definition, and vice versa. Furthermore, we can cluster the patients into more than two clusters as in Fig. 5(e) and (f). In other diseases, such as cancer, they are able to identify new subtype of disease [14] using clustering result. Further analysis in our clustering results with domain expert is needed to investigate potential new subtype of sepsis.

5 Conclusion

We proposed the Patient Sequence Model (PSM) to transform multivariate time series patient data into explainable and computable representation. The PSM model is able to generate a meaningful state transition diagram. We developed a reasonable feature extraction method based on probability of treatment effect and clustered all patients based on those high dimensional feature. The clustering result aligns with current medical concept and potentially discovers novel subtypes of sepsis. We proposed a novel Bayesian reinforcement learning method to compute personalised policy based on combination of dedicated policies in each cluster. The future work is discussion with medical collaborator with respect to this clustering results. In addition, we will need to define a medically acceptable performance evaluation criteria to compare our proposed personalised policy with general policy and doctor policy.

Acknowledgements. Chandra Prasetyo Utomo is sponsored by Indonesia Endowment Fund for Education (LPDP). This research is partially supported by Australian Research Council (ARC) project ID: DP160104075 and Universitas YARSI, Jakarta. The authors thank Dr. Robert Boots from Royal Brisbane and Women's Hospital (RBWH) for valuable insight in treatment recommendation research.

References

1. Almirall, D., Compton, S.N., Gunlicks-Stoessel, M., Duan, N., Murphy, S.A.: Designing a pilot sequential multiple assignment randomized trial for developing an adaptive treatment strategy. Stat. Med. **31**(17), 1887–1902 (2012)
2. Angus, D.C., van der Poll, T.: Severe sepsis and septic shock. New England J. Med. **369**(9), 840–851 (2013)

3. Bajor, J.M., Lasko, T.A.: Predicting medications from diagnostic codes with recurrent neural networks. In: Proceeding of the 5th International Conference on Learning Representations - (ICLR 2017), pp. 1–19 (2017)
4. Cheerla, N., Gevaert, O.: MicroRNA based pan-cancer diagnosis and treatment recommendation. BMC Bioinf. **18**(1), 32 (2017)
5. Chen, Z., Marple, K., Salazar, E., Gupta, G., Tamil, L.: A physician advisory system for chronic heart failure management based on knowledge patterns. Theory Pract. Logic Program. **16**(5–6), 604–618 (2016)
6. Hu, J., Perer, A., Wang, F.: Data driven analytics for personalized healthcare. In: Weaver, C.A., Ball, M.J., Kim, G.R., Kiel, J.M. (eds.) Healthcare Information Management Systems. HI, pp. 529–554. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-20765-0_31
7. Johnson, A.E.W., et al.: A comparative analysis of sepsis identification methods in an electronic database. Crit. Care Med. **46**(4), 494–499 (2018)
8. Johnson, A.E., Pollard, T.J., Shen, L., et al.: MIMIC-III, a freely accessible critical care database. Sci. Data **3**, 160035 (2016)
9. Nemati, S., Ghassemi, M.M., Clifford, G.D.: Optimal medication dosing from suboptimal clinical examples: a deep reinforcement learning approach. In: Proceeding of the 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2016), Orlando, FL, USA , pp. 2978–2981, August 2016
10. Pollard, S., Edwin, S.B., Alaniz, C.: Vasopressor and inotropic management of patients with septic shock. Pharm. Ther. **40**(7), 438–50 (2015)
11. Raghu, A., Komorowski, M., Ahmed, I., Celi, L., Szolovits, P., Ghassemi, M.: Deep Reinforcement Learning for Sepsis Treatment. In: Proceeding of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA (2017)
12. Rhodes, A., Evans, L.E., Alhazzani, W., et al.: Surviving sepsis campaign: international guidelines for management of sepsis and septic shock: 2016. Intensive Care Med. **43**(3), 304–377 (2017)
13. Sakr, Y., Jaschinski, U., Wittebole, X., et al.: Sepsis in intensive care unit patients: worldwide data from the intensive care over nations audit. Open Forum Infect. Dis. **5**(12), ofy313 (2018)
14. Shen, R., Olshen, A.B., Ladanyi, M.: Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. Bioinformatics **25**(22), 2906–2912 (2009)
15. Singer, M., Deutschman, C.S., Seymour, C., et al.: The third international consensus definitions for sepsis and septic shock (sepsis-3). JAMA - J. Am. Med. Assoc. **315**(8), 801–810 (2016)
16. Tsoukalas, A., Albertson, T., Tagkopoulos, I.: From data to optimal decision making: a data-driven, probabilistic machine learning approach to decision support for patients with sepsis. JMIR Med. Inf. **3**(1), e11 (2015)