RL Based Decision Support System for u-Healthcare Environment

Devinder Thapa, In-Sung Jung and Gi-Nam Wang
AJOU University
South Korea

1. Introduction

We can imagine a haywire situation with no healthcare centres nearby. In this situation, a high risk patient, away from the medical healthcare center, may get major heart attack or unpredictable sudden stroke, or some other noxious symptoms. Lack of on-time information, proper diagnosis, and decision making system, may sometimes cause the life of the patient. In order to access the timely information and to employ correct diagnosis at anytime and anywhere, use of ubiquitous technologies is becoming ideal test-beds for u-Healthcare environments. However, using ubiquitous device, it would be one of the most crucial requisites to accumulate accurate signals timely and appropriate processing of those signals during such critical circumstances. Furthermore, lack of proper decision support system may delay the treatment, and it may cost a life of the patient. The effort to rectify any of these issues will minimize the time lag between observation and treatment during the emergency circumstances, and helps to reduce the diagnosis time, that can be better utilize for caring the patient.

The objective of this chapter is to combine the agent based decision support system with ubiquitous artefacts and make it more intelligent, so that it can help the doctors to acquire correct and timely diagnosis information and select appropriate treatment choices. Also, designed is a novel interpretation of Markov decision process, providing clear mathematical formulation to connect reinforcement learning agent system. An attempt is given to supervise the dynamic situation by using agent based ubiquitous artefacts and to find out the appropriate solution for emergency circumstances, providing correct diagnosis and proper treatment in time. The well known reinforcement learning can be utilized to model u-healthcare decision support system. The reason for using the RL (Reinforcement Learning) agent based on MDP (Markov Decision Process) model is because it needs less number of parameters compare to other decision trees it also gives approximation method to make trade off between accuracy and speed, in turn, solve the complex number of cases in less time compare to other decision support system (Milos H., Fraser H., 2000).

Organization of this chapter is as follows. Section 2 is a review of the related works, RL agent, and Markov decision model is also explained. Section 3 describes the details scenario of the proposed approach. Similarly, section 4 discusses the formulation of the model and optimal policy finding algorithm of the RL based decision support system. Finally section 5 & 6 concludes the chapter and contains references.

2. Related works

The concept of ubiquitous healthcare system using agent technology is studied in the reference Jakob E. Bardram (2004). Where the author visualized about highly interactive hospital, it will facilitate the doctors to access relevant medical information at anytime and anywhere, using ubiquitous hardware and software. Similar nature of research is explained in the reference work of (Thapa D., et al., 2005). In the reference work of (Wendelken S.M., et al., 2003), the idea of location finder agent using iterative deepening search is described. It can be utilized to design the patient location finder agent in a u-healthcare system. Similarly, the concept of agent based healthcare environment is explained in the reference (Rodriguez M., et al., 2003). Although, the functional architecture of the reference work is different, the conceptual idea is deployable to our work.

In the reference work of (Milos H., & Fraser H., 2000) partially observable Markov decision process [POMDP] can be deployable utilized for the diagnosis and treatment of ischemic heart disease. This work can be theoretically considered as the part of reinforcement learning agent. Although POMDP is more suitable to observe the hidden state of the patient, its exponential growth is prone to state explosion problem. In our work, we deployed MDP assuming the patient state is within the domain of experts' knowledge. In addition, time complexity of the MDP is lesser than POMDP.

All of the existing works are focused on the exploitation of ubiquitous artefacts for the betterment of healthcare system. However, little attention is given to the concept of developing integrated emergency system using reinforcement learning decision support system in a *u*-healthcare environment. The main objective of this research is to design agent based decision support system using reinforcement learning, to reduce the time lag between the onset of the attack and the time that care is administered. Ubiquitous devices combined with agent technology can reduce the time latency, and they can provide suitable on-time treatment information when the patient is away from the hospital premises.

2.1 Reinforcement learning agents

Reinforcement learning (RL) is based on interaction with an environment, from the consequences of action, rather than from explicit teaching. RL methods are intended to address the kind of learning and decision making problems that people face in their everyday lives. Main elements of Reinforcement learning are states s, actions a, and rewards r as depicted in Fig.1. The reinforcement learning agent (RL-agent) is connected to its environment via sensors. In every step of interaction the agent receives a feedback about the state of the environment s_{t+1} and the reward r_{t+1} of its latest action a_t. The agent chooses an action a_{t+1} representing the output function, which changes the state s_{t+2} of environment. The agent gets a new feedback, through the reinforcement signal r_{t+2}. It involves a decision-making agent interacting with its environment so as to maximize the cumulative reward that it receives over time. The agent perceives aspects of the environment's state and it selects right actions. The agent may estimate a value function and use it to construct better and better decision-making policies over a given time. The main objective of the agent is to maximize the aggregated reinforcement signals. RL could be characterized by a mathematical framework of Markov decision processes (MDPs) (Stuart J. R. and Peter N., 2003).

In medical healthcare system, when we are supposed to take right decision in a right time, reinforcement learning agent, by combining different model and actions, can help a

physician to find out the best diagnosis and treatment options in different states of the patient. By using RL algorithms we can choose the best alternative action during the emergency circumstances which can reduce the total cost (time, treatment complexity etc) of the action (diagnosis & treatment) taken by the physician.
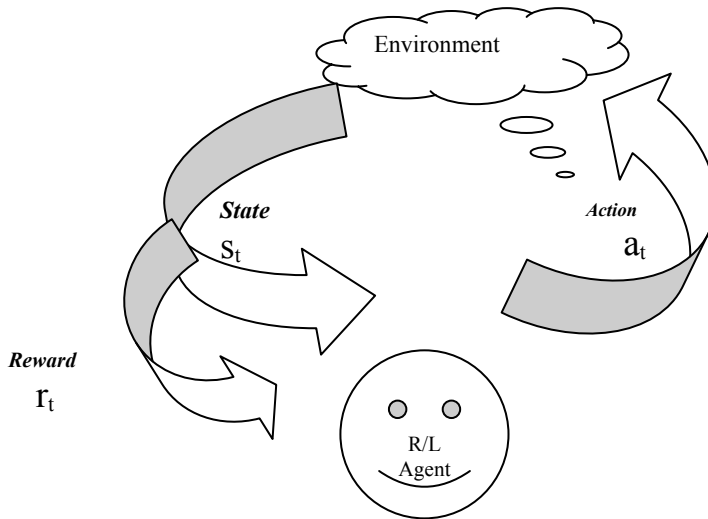

Fig. 1. Reinforcement Learning Agent

2.2 Markov decision process

Reinforcement learning basically uses the MDP concept for implementation. We are using the MDP which is based on shortest route approach to reduce the time latency of the reactive action, as per our approach, time is as much important as efficiency.

An MDP is defined by a set of states *S*, actions *A*, Reward *R*, and transition probabilities *T* (Puterman M. L., 1994). A transition function, T: S x A ->P(S), defines the effects of the various actions on the state of the environment. P(S) represents the set of discrete probability distributions over the set *S*. The reward function, R: S x A→ R, specifies the agent's task to find a policy mapping, choice of action, so as to maximize the expected sum of reward. Two types of decision model can be used to obtain the reward. Infinite horizon uses discount factor 0<γ<1 to control how much effect future rewards have on the optimal decisions, with small values giving significant weight to later rewards. However, finite horizon does not use discount factor and the iteration of certain action is known in advance. This paper deployed the finite horizon decision model.

$$V^*(s) = \max_a (R(s,a) + \sum_{s' \in S} T(s,a,s')V^*(s')), \forall s \in S, \tag{1}$$

$$V^*(s) = \max_\pi E(\sum_{t=0}^{\infty} r_t) \tag{2}$$

$$R(s,a) = \sum_{s' \in S} P(s'|s,a)R(s,a,s')$$ (3)

Implementation of this model in our approach is to ascertain, current state (patient status at the time of incident), action (medication, no action), transition probabilities (between current state and new state), and reward (cost and complexities) certain payoffs related to this transition. The objective of this model is to find out the optimized action to maximize the reward or cost in a finite discounted horizon as shown in equation 1.

Due to the computation complexities of the pure MDP model we use Bellman's value function recursively; it [eqn. 1] calculates the total reward value by adding all the suboptimal values [eqn. 3] at some finite time horizon
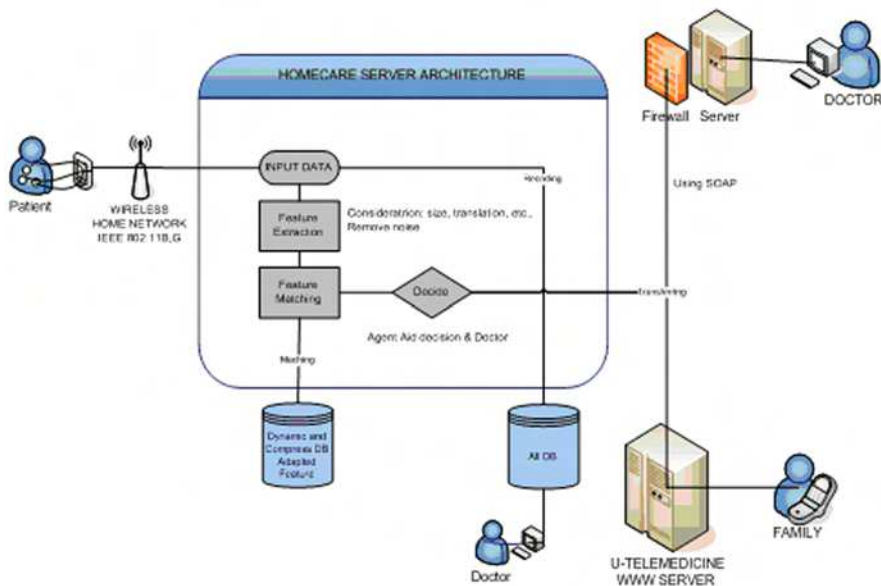
3. Proposed model

Fig. 2. RL based decision support systems in a u-Healthcare system

Assumption of this research test bed has been made in the ubiquitous environment. As shown in Fig.1, when a high risk patient, far from medical facilities, gets some perilous occurrence in their body, the ubiquitous sensor device attached to their body sends bio signals like (digital sounds of lung, SaO2, EKG) etc. to the home medical server using IEEE 802.11b wireless network.

The home medical server is connected to the hospital knowledgebase server through TCP/IP using internet connection. This signal sends the patient current status to the HIS (Hospital Information Server). On the bases of this crucial input data the decision making agent, based on RL model, make inference of the data and provide entire data history of the patient with best alternate action (diagnosis and treatment) to the related department with

minimal time cost. Decision agent also helps to the related physician check his scheduling, and sends the patients profile to the related departments. It helps in timely availability of the crucial information to the right place. The flow of Information processing is depicted in the following figure 2.
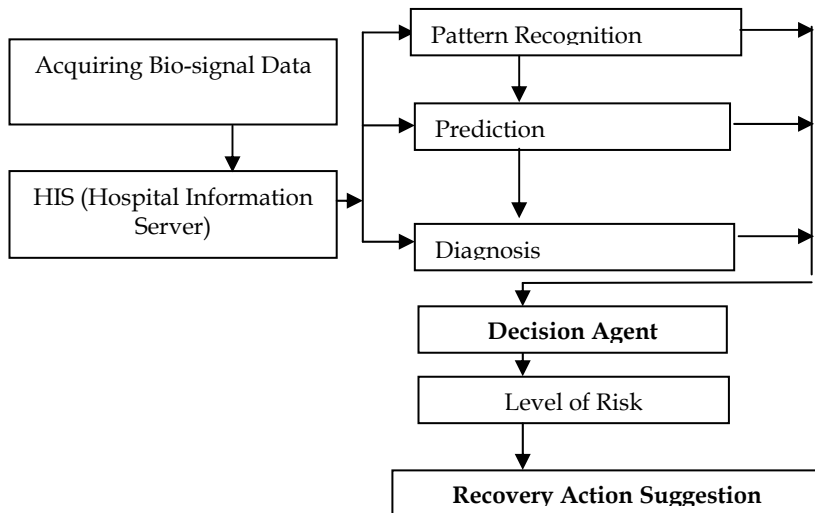

Fig. 3. Flow of information processing in a u-Healthcare system

1. Aquiring bio-signals and pattern recognition can be done by using artificial neural network like back propagation.
2. Prediction & diagnosis can be done by using time series analysis & self organizing feature map (SOM), artificial neural network .
3. RL decision making agent is based on MDP (Markov Decision Process), this chapter particularly describes about this approach.

Final output of the decision making agent will be two fold measurements as shown in fig 4:

*   Patient Current State (Level of risk)
        - Normal
        - Serious
        - Emergent

*   Emergency Measurement
        -If the (State = Emergent)
                Send SMS Message to Doctor
                Send SMS Message to User Relatives
                Send SMS Message to Ambulance

On the basis of the level of risk it suggests best series of action or optimal policy for the better way of diagnosis and treatment.
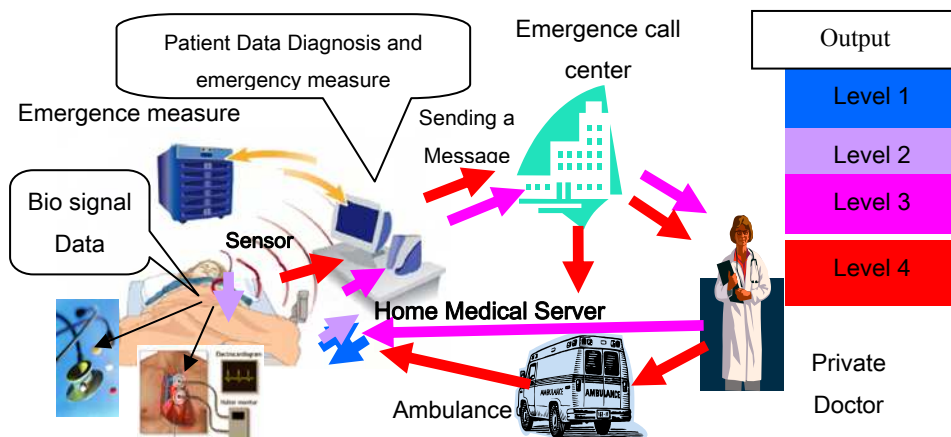
Fig. 4. Complete graphical scenario of RL integrated u-Healthcare System

As soon as the patients reach to the hospital premises information will be ready for the optimal and immediate action by physician. Although this is a conceptual idea, the parallel research on this idea has been going on. Our approach is to make the RL based decision making model more efficient and rational to save the life of the high risk patients in emergency circumstances, with the help of RL integrated ubiquitous artifacts in the u-Healthcare system. The level of risk will be calculated on the basis of the input data described in Table I.

Data	Standard
SaO2	Normal >90% Lack of oxygen: Mild 90~94% Moderate 75-90%

HR
   (Heart Rate)
	Normal 60~100/min
	Tachycardia >100/min
	Paroxysmal Tachycardia:
	                150~250/min

BP
    (Blood
    Pressure)
	Normal : 120/80mm HG
	Hypertension: >140/90mmHG
	Serious Hypertension:
	                    >200/140mmHG
	Hypotonic >100/60mmHG
	Serious Hypotonic <80/60mmHG

Body Temperature	Normal: 36.5~37Degree Slight fever: Morning >37.2 Evening >37.7 High Fever > 38.3

Table I. Parameters for calculating the level of risk (Source: AJOU university hospital)

4. Formulations to a reinforcement learning problem

In this paper, we assumed that patient current state is fully observable; we have used the MDP model with finite discounted horizon. On the basis of his current available data (like heart beat, pulse rate, respiration, chest pain) as shown in table I, and past history of the patient from Hospital Information Server, it makes different combination of the action (medication or no action or emergency measurement) and reinforce (negative and positive) rewards (cost) with every action to go to next state. Finally, it finds out the best action or minimum cost (time taking) solution.
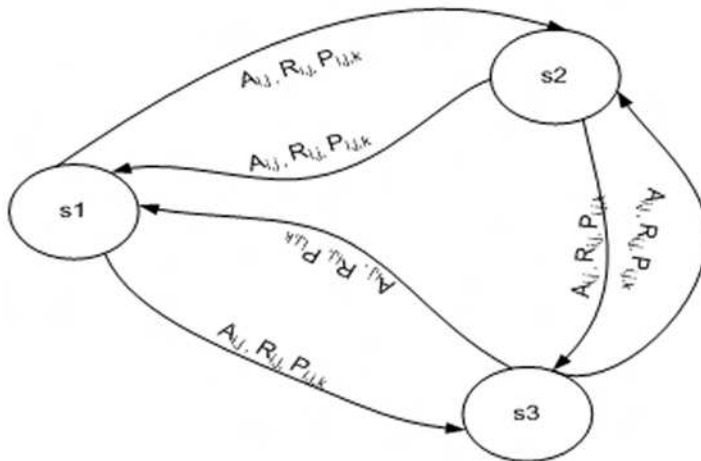

Fig. 5. Symbolic representation of Markov Decision Process

R=Reward, P=Transition probability, A=Action, S=State

Decision Epochs: [Finite time horizon]

$$T=\{1,2,....,N\}, \quad N \leq \infty$$

States(S) = [Normal (s3), Serious (s1), Emergent (s2)]

$$S=\{s1, s2, s3\}$$

Action (A) = [No action(a1), Medication(a2), Medication with emergency measurement(a3)]
$$A = \{a_{i,j} \mid i=1,2,3 \text{ and } j=1,2,3\}, \text{ where } i \text{ refers to the state and } j \text{ refers to action}$$

Rewards(R) = {Cost $(r_{i,j})$ | i=1,2,3 and j= integer }

$$R(S1, a_{i,j})= r_{i,j}$$
$$R(S2, a_{i,j})= r_{i,j}$$
$$R(S3, a_{i,j})= r_{i,j}$$
$$\text{If } N \leq \infty$$

Transition Probabilities (pt): [Effect of diagnosis and treatment (p)]

$$pt(s3|s1, a_{1,2})=p_{1,2,3}$$

For example, if the patient is in the state s1 and the action $a_{1,2}$ will be taken then the probability of patient going to state s3 will be $p_{1,2,3}$. Where p denotes the probability between [0, 1] of transition, and 1, 2, and 3 denotes the action $a_{1,2}$ and state s3 respectively. Calculation of expected Reward or Cost:

Example: $R(S1, a_{1,1})= R(S1, a_{1,1}, S1) \, pt(S2|S1, a_{1,1}) + R(S1, a_{1,1}, S2) \, pt(S2|S1, a_{1,1})$

4.1 Finding the best policy or the minimum cost function using DP (Dynamic programming) approach

Compared with other methods for solving MDPs, DP methods are actually quite efficient. The (worst case) time DP methods take to find an optimal policy is polynomial in the number of states and actions (Stuart J. R. and Peter N., 2003). If n and m denote the number of states and actions, this means that a DP method takes a number of computational operations that is less than some polynomial function of n and m. A DP method is guaranteed to find an optimal policy in polynomial time even though the total number of (deterministic) policies is m^n. In this sense, DP is exponentially faster than any direct search in policy space could be, because direct search would have to exhaustively examine each policy to provide the same guarantee.

choose an arbitrary policy π'

loop

$\pi := \pi'$

compute the value function of policy π :

solve the linear equations

$$V\pi(s) = R(s, \pi(s)) + \sum_{s' \in s} T(s, \pi(s), s')V_\pi(s') \tag{4}$$

improve the policy at each state:

$$\pi'(s) := \arg\min{}_a(R(s, a) + \sum_{s' \in s} T(s, a, s')V_\pi(s')) \tag{5}$$

until $\pi = \pi'$

Denote a policy as Л, where Л=action selected in current state. Where $V\pi(s)$ and $\pi'(s)$ are optimal value and control function. We can take Л' as any random policy and $V\pi(s)$ is reward value starting from current state and following the Л policy. Now we can define another greedy policy in terms of Л'(s) and make iteration of the value function $V\pi(s)$ function until $\pi = \pi'$. We consider whether the value could be improved by changing the first action taken. If it can, we change the policy to take the new action whenever it is in that situation. When $\pi = \pi'$ and no improvements are possible, then the policy is considered to be optimal. This model will be helpful to achieve the objective of finding an action or a sequence of actions that optimizes the time cost of diagnosis and treatment in a given finite horizon under emergency circumstances.

5. Conclusion

This paper presents and describes a Reinforcement Learning agent based model used for information acquiring and real time decision support system at emergency circumstances. The well known reinforcement learning is utilized for modeling emergency u-Healthcare system. Markov decision process is also employed to provide clear mathematical formulation in order to connect reinforcement learning as well as to express integrated agent system. This method will be highly effective for the real time diagnosis and treatment of high risk patient during the emergency circumstances, when they are away from the hospital premises. Looking at the growing increase in the research area of ubiquitous devices this approach seems to be very beneficial and life saving for the high risk patient at the time of emergency circumstances. Further pursuing will be to develop some prototype, and simulate the testing data, planning modules, and find out the actual outcome of this approach.

6. Acknowledgement

7. References

Jakob E. Bardram(2004). The Personal Medical Unit -- A Ubiquitous Computing Infrastructure for Personal Pervasive Healthcare. *UbiHealth 2004: The 3rd International Workshop on Ubiquitous Computing for Pervasive Healthcare Applications*

Jung I., Thapa D. and Wang G. N.,(2005). Neural Network Based Algorithm for Diagnosis and Classification of Breast Cancer Tumor, *LNAI 3801, Springer-Verlag Berlin Heidelberg,* (2005), pp. 107–114.

Leong, T. Y. (1998). Multiple perspective dynamic decision making, *Artificial Intelligence*, Vol.105, No. 1-2 (October 1998), pp. 209-261, ISSN: 0004-3702

Lesile P. K., Michael L. L., Andrew W. M. (1996). Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research, Vol.* 4, (1996), pp. 237-285

Milos H., Fraser H.(2000). Planning Treatment of Ischemic Heart Disease with Partially Observable Markov Decision Process. *Artificial Intelligence in Medicine*, vol. (18), (2000) pp. 221-244

Puterman M. L., (1994). Markov Decision Process: Discrete Stochastic Dynamic Programming. *A Wiley-Interscience publication, John Wiley & Sons, Inc.*, (1994) New York.

Rodriguez M., Favela J., Gonzalez V. and Muñoz M. (2003). Agent Based Mobile Collaboration and Information Access in a Healthcare Environment. *Proceedings of Workshop of E-Health*, Applications of Computing Science in Medicine and Health Care. ISBN: 970-36-0118-9. (December 2003), Cuernavaca, México

Sutton R. S. & Barto A. G. (1998). Reinforcement Learning: An Introduction. *MIT Press* (1998), a Bradford Book, Cambridge, MA

Stuart J. R. and Peter N.(2003). Artificial Intelligence: A Modern Approach IInd edition. *Pearson Education International*, (2003), New Jersey

Thapa D., Jung I., and Wang G.N.(2005). Agent Based Decision Support System   using Reinforcement Learning under Emergency Circumstances, *Spriger Lecture Notes in Computer Science (LNCS)*, 3610, (2005), pp. 888-892.

Ulieru, M.; Geras, A: Emergent, (Nov. 2002). Holarchies for e-health applications: a case in glaucoma diagnosis. *IECON 02 (IEEE 2002 28th Annual Conference of the Industrial Electronics Society)*, Volume: 4, 5-8 (Nov. 2002) pp. 2957 - 2961

Watrous, R.L. and Towell, G.(1995). A Patient-adaptive Neural Network ECG Patient Monitoring Algorithm. *In Proceedings Computers in Cardiology*, Vienna, Austria (1995) pp. 229--232.

Wendelken S.M., McGrath S.P. and Blike G.T.(2003). A medical assessment algorithm for automated remote triage. *International conference of the IEEE EMBS*, Mexico, September 17-21,(2003)

Reinforcement Learning

Edited by Cornelius Weber, Mark Elshaw and Norbert Michael Mayer

Brains rule the world, and brain-like computation is increasingly used in computers and electronic devices. Brain-like computation is about processing and interpreting data or directly putting forward and performing actions. Learning is a very important aspect. This book is on reinforcement learning which involves performing actions to achieve a goal. The first 11 chapters of this book describe and extend the scope of reinforcement learning. The remaining 11 chapters show that there is already wide usage in numerous fields. Reinforcement learning can tackle control tasks that are too complex for traditional, hand-designed, non-learning controllers. As learning computers can deal with technical complexities, the tasks of human operators remain to specify goals on increasingly higher levels. This book shows that reinforcement learning is a very dynamic area in terms of theory and applications and it shall stimulate and encourage new research in this field.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Devinder Thapa, In-Sung Jung and Gi-Nam Wang (2008). RL Based Decision Support System for u-Healthcare Environment, Reinforcement Learning, Cornelius Weber, Mark Elshaw and Norbert Michael Mayer (Ed.), ISBN: 978-3-902613-14-1, InTech, Available from:
http://www.intechopen.com/books/reinforcement_learning/rl_based_decision_support_system_for_u-healthcare_environment

INTECH
open science | open minds