Artificial Intelligence in Medicine 62 (2014) 47–60

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

journal homepage: www.elsevier.com/locate/aiim

Optimization of anemia treatment in hemodialysis patients via reinforcement learning

Pablo Escandell-Montero[a,*], Milena Chermisi[b], José M. Martínez-Martínez[a], Juan Gómez-Sanchis[a], Carlo Barbieri[b], Emilio Soria-Olivas[a], Flavio Mari[b], Joan Vila-Francés[a], Andrea Stopper[b], Emanuele Gatti[b,c], José D. Martín-Guerrero[a]

[a] Intelligent Data Analysis Laboratory, University of Valencia, Av. de la Universidad, s/n, 46100 Burjassot (Valencia), Spain
[b] Healthcare and Business Advanced Modeling, Fresenius Medical Care, Else-Kröner-Strasse 1, 61352 Bad Homburg, Germany
[c] Centre for Biomedical Technology at Danube, University of Krems, Dr.-Karl-Dorrek-Strasse 30, 3500 Krems, Austria

ARTICLE INFO

Article history:
Received 3 November 2013
Received in revised form 23 June 2014
Accepted 11 July 2014

Keywords:
Reinforcement learning
Markov decision processes
Fitted Q iteration
Chronic kidney disease
Renal anemia
Darbepoietin alfa

ABSTRACT

Objective: Anemia is a frequent comorbidity in hemodialysis patients that can be successfully treated by administering erythropoiesis-stimulating agents (ESAs). ESAs dosing is currently based on clinical protocols that often do not account for the high inter- and intra-individual variability in the patient's response. As a result, the hemoglobin level of some patients oscillates around the target range, which is associated with multiple risks and side-effects. This work proposes a methodology based on reinforcement learning (RL) to optimize ESA therapy.
Methods: RL is a data-driven approach for solving sequential decision-making problems that are formulated as Markov decision processes (MDPs). Computing optimal drug administration strategies for chronic diseases is a sequential decision-making problem in which the goal is to find the best sequence of drug doses. MDPs are particularly suitable for modeling these problems due to their ability to capture the uncertainty associated with the outcome of the treatment and the stochastic nature of the underlying process. The RL algorithm employed in the proposed methodology is fitted Q iteration, which stands out for its ability to make an efficient use of data.
Results: The experiments reported here are based on a computational model that describes the effect of ESAs on the hemoglobin level. The performance of the proposed method is evaluated and compared with the well-known Q-learning algorithm and with a standard protocol. Simulation results show that the performance of Q-learning is substantially lower than FQI and the protocol. When comparing FQI and the protocol, FQI achieves an increment of 27.6% in the proportion of patients that are within the targeted range of hemoglobin during the period of treatment. In addition, the quantity of drug needed is reduced by 5.13%, which indicates a more efficient use of ESAs.
Conclusion: Although prospective validation is required, promising results demonstrate the potential of RL to become an alternative to current protocols.

1. Introduction

Anemia is a common complication characterized by a reduced concentration of hemoglobin (Hb) that occurs in over 90% of patients undergoing hemodialysis [1]. Hemodialysis is the most common treatment for patients in advanced stages of chronic kidney disease (CKD), particularly in its end state, commonly referred as end-stage renal disease (ESRD). In the last years the prevalence of ESRD has increased substantially, reaching more than 1000 per million population in most of the developed countries [2]. In some countries, such as USA and Japan, the current prevalence is over 2000 per million [2]. ESRD involves a gradual loss of kidney function over time, which produces, among other health problems, a poor production of erythropoietin (EPO). This hormone regulates the red blood cell (RBC) production, a class of cells rich in Hb. Low Hb levels are associated with heart disease, poorer overall quality of life, and increased mortality [3,4].

Current standard treatment of anemia consists mainly of the administration of erythropoiesis-stimulating agents (ESAs). The response to this kind of drugs is known to have a large inter- and intra-interindividual variability due to differences in background

* Corresponding author. Tel.: +34 963543421.
E-mail address: pablo.escandell@uv.es (P. Escandell-Montero).

http://dx.doi.org/10.1016/j.artmed.2014.07.004
0933-3657/

characteristics, disease severity, comorbidities and concurrent medications [5,6]. Although there exist protocols to help physicians determine the appropriate dose, achieving stable Hb levels within the target range can be complex and often requires dose titration. Results from several studies suggest that a phenomenon known as Hb cycling is a common occurrence in ESA-treated patients [7,8]. Hb cycling is defined as the cyclical, repeated, up and down movement of Hb levels during ESA treatment. The exact causes of Hb cycling are not yet completely understood; however, a number of possible reasons have been proposed. Fishbane and Berns [9] suggested two ESA management practices as major causes. First, the use of rigid dose adjustment protocols that do not account for the high heterogeneity in patient response. Second, narrow Hb target ranges recommended in clinical guidelines [10,11], which need frequent dose changes. The effect of an ESA dose change does not reach a steady state until 70–120 days (RBC lifespan). When doses are changed frequently, it is difficult to take into account the long-term effects of each dose, and often they are ignored [12,13]. The link between Hb cycling and the development of several diseases [7] together with the high cost of the treatment (e.g., around $2.3 billions per year in USA [2]) justifies the need to improve current protocols.

The widespread use of electronic medical records is giving rise to large amounts of data that could be useful to reduce medical errors, improve treatments and minimize side effects and costs [14]. This work proposes a methodology based on reinforcement learning (RL) to optimize ESA therapy. RL is a data-driven approach for solving sequential decision-making problems that are formulated as Markov decision processes (MDPs) [15]. Computing optimal drug administration strategies for chronic diseases is a sequential decision-making problem in which the goal is to find the best sequence of drug doses. MDPs are particularly suitable for modeling these problems due to their ability to capture the uncertainty associated with the outcome of the treatment and the stochastic nature of the underlying process [16–18]. The standard approach to solve MDPs is dynamic programming (DP); however, the practical application of DP is limited because it cannot deal with large-scale problems and requires full knowledge of the MDP model, including the transition probability function. In contrast, RL (also known as approximate dynamic programming (ADP)) uses function approximation to address large-scale problems and the data sampled from the process to implicitly represent the transition function [19]. RL can exploit the information contained in medical records to compute policies of ESA administration tailored to the individual characteristics of each patient. In addition, the optimization process is made over sequences of doses instead of isolated doses, which is crucial to include the drug long-term effects.

The methodology proposed in this work uses the algorithm fitted Q iteration to learn a policy of ESA administration from a set of medical records. The features employed to define the MDP model are extracted in part from the laboratory tests and in part from a clustering procedure of the patient's main attributes. In order to test the methodology, a series of experiments has been conducted using a computational model that simulates the response of the patients. The performance has been assessed against the algorithm Q-learning and a standard protocol of dose adjustment.

The rest of the paper is organized as follows. Next section provides a brief review of related work in this domain. Section 3 introduces the necessary background in RL and briefly explains the algorithms employed in the experiments, namely, Q-learning and fitted Q iteration. The latter algorithm makes use of extremely randomized trees, a supervised learning method that is described in Section 4. Section 5 discusses the computational model used in the experiments to simulate patients' response to ESA. Anemia management formulation using the MDP framework is presented in Section 6. Experiments carried out are detailed in Section 7. Section 8 shows and discusses the achieved results. Finally, conclusions and proposals for further work are given in Section 9.

2. Literature review

The idea of using a data-driven method to optimize ESA administration is not new. Artificial neural networks have been used by several authors during the last decade to individualize ESA doses [20–22]. In general, those methods used current and previous Hb levels, ESA doses, and other variables that describe the patient's condition, in order to predict the next Hb level. The goal of those previous works was to select the optimal ESA dose in order to achieve a given Hb level. This approach is suitable only when the optimization horizon is the next time step. On the contrary, the aim of ESA therapy is the long-term Hb stabilization. The same idea has been applied using other machine learning techniques, such as fuzzy logic [23,24], support vector machines [25] or Bayesian networks [26].

Model predictive control (MPC) is a method of process control whose main advantage is that it incorporates a finite time-horizon in the optimization process. Gaweda et al. [27] showed that MPC may result in improved anemia management. A major difficulty of MPC is the requirement of an accurate system model. Even if the system model is available, RL has shown to be competitive with MPC [28].

RL in the context of anemia management was previously studied by Gaweda et al. [29] and Martín-Guerrero et al. [30]. Both agree in the potential of RL to become an alternative to currently used protocols. The algorithm employed in those works was the popular Q-learning [31]. This algorithm has been widely used in some fields as robotics because it requires little computation and can work in real time. However, Q-learning makes an inefficient use of the data, thus, it is not suitable for problems in which acquiring data is costly [32]. Fitted Q-iteration (FQI) [33] is a relatively new RL algorithm that significantly reduces the quantity of data required to learn useful policies. Recently there has been a growing interest in applying FQI to optimize the treatment of several diseases including HIV/AIDS [34], psychiatric disorders [35], epilepsy [36,37], schizophrenia [38,39] or smoking addiction [40]. To the authors' knowledge, this is the first work that applies FQI to the optimization of anemia treatment.

3. Reinforcement learning

Reinforcement learning (RL) is a general class of algorithms in the field of machine learning for solving decision-making problems where decisions are made in stages [41]. Such problems are present in a wide range of fields, including operations research [42,43], artificial intelligence [44,45], automatic control [46], or medicine [30]. The standard RL setting consists of an agent (or controller) in an environment (or system). Each decision (also called action) produces an immediate reward. The agent learns to perform actions in order to maximize the reward collected over time. The goal is defined by the user through the reward function. Contrary to other approaches, RL does not rely on a mathematical model of the system, but is based on experience (or data). The agent obtains experience interacting with the environment. Fig. 1

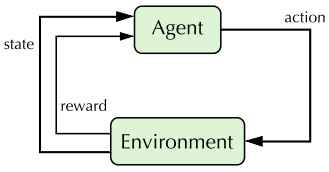

Fig. 1. Elements of RL and their flow of interaction.