Learning from an expert

Pierre Humbert
CMLA, ENS Paris-Saclay
humbert@cmla.ens-cachan.fr

Julien Audiffren
CMLA, ENS Paris-Saclay
audiffren@cmla.ens-cachan.fr

Clément Dubost
Hôpital d'instruction des armées Bégin
COGNAC G, Université Paris Descartes
clément.dubost@intradef.gouv.fr

Laurent Oudre
L2TI, Université Paris 13
COGNAC G, Université Paris Descartes
laurent.oudre@univ-paris13.fr

Abstract

In this work, we study the problem of designing an automated procedure which aims at controlling the concentration of drugs during a global anesthesia. First, we select easy to measure physiological variables, whose choice is validated using a Hidden Markov Model. Then, we use expert trajectories provided by anesthesiologists to train an agent, whose anesthesia policy shows promising result.

1 Introduction

The progress of medical research in the last century has permitted to efficiently diagnose increasingly complex diseases and conditions, which often require general anesthesia in order to perform advanced surgeries. Therefore, a full understanding of those techniques and its mechanisms is one of the central unsolved problem of today's neurosciences and is key to improve medical treatments.

One of the main challenge of anesthesia is the identification and the quantification of the depth of anesthesia. An accurate evaluation would allow a better use of drugs and therefore decrease the risk of complications, such as intraoperative awareness or other side effects [Sessler et al., 2012].

Currently, the gold-standard to quantify the depth of anesthesia remains the electroencephalogram (EEG). However, the sensors used to record EEG are expensive and suffer many drawbacks. An alternative is the Bispectral Index (BiS), whose usefulness during total-intravenous anesthesia has been shown by several studies (see e.g. [Leslie et al., 2010]). Nevertheless, recent studies have questioned the efficiency of the BiS, for instance in the case of volatile anesthetic use (see e.g. [Whitlock et al., 2011]).

Besides the EEG, many different relevant physiological variables – such as heart rate or blood pressure – are monitored during an anesthesia. One of the most difficult task for the anesthesiologist is to understand and combine all these variables to maintain good working condition of the vital organs (homeostasis).

Despite several studies, the assessment of the depth of anesthesia using physiological variable remains arduous. Recently, [Moore et al., 2014] have successfully applied Reinforcement Learning to a simple anesthesia model relying on BiS. While effective, the reliance of this model on BiS limits its potential application, due to the drawbacks of the measure.

In this work, we propose a different approach by introducing a simple model for anesthesia based on easy to measure physiological variables. This choice of the variables is evaluated using a Hidden Markov Model (see Section 2). Then, we use expert trajectories provided by anesthesiologists to train an agent, and compare its anesthesia policy with actual actions undertaken by the anesthesiologist (see Section 3).

30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

2 Modelizing Anesthesia : Hidden Markov Model

This section details the medical setting of this work, anesthesia, and the model we used to encode it.

2.1 The challenge of Anesthesia

The main objective of anesthesia, in our framework, is to maintain the patient into unconsciousness during surgery. However, the crucial difficulty of the process is that even slight overdoses of the anesthetic drug lead to increased danger of complications and side-effects. To avoid these drawbacks, it is necessary to maintain the right level of consciousness by carefully choosing the dosage of the anesthetic drug at any time throughout the surgery.

Unfortunately, the depth of anesthesia is not directly observable, and can only be estimated from the measure of many different physiological variables. A common way to proceed is based on the electroencephalogram (EEG) signals. For instance, [Bennett et al., 2009] showed that specific frequencies of the EEG signals are strongly correlated with the current general state of the patient : Awake (before the anesthesia), Loss Of Consciousness (LOC – transition period between Awake and Anesthesia), Anesthesia, Recovery Of Consciousness (ROC) and Emergence.

However, EEG require expensive sensors and are time consuming to setup. One of the crucial aim of this work was to measure Anesthesia state without EEG, and to that end, we chose to use three physiological variables that are assumed to be related to the level of consciousness : Heart Rate (HR), Mean Blood Pressure (MeanBP) and Respiratory Rate (RR).

Dataset Our dataset is composed of 2 women and 8 men aged 24 to 65 suffering from an inguinal hernia, who underwent the appropriate surgery with sevoflurane, a volatile anesthetic agent. Patient data consist in 2 EEG signals sampled at 100 Hz, and the three variables previously cited HR, MeanBP, RR as well as Inspired Fraction of AA (AAFi) – which is used to measure the current dosage of anesthetic drug– all sampled at 1 Hz.

2.2 Estimating the Anesthesia state

To verify if the selected physiological variables were a relevant indicator of the anesthesia state of the patient, we proceeded as follows.

The anesthesia state was encoded using a Hidden Markov Model (HMM) $(\mathcal{S}, T, \mathcal{O}, B, \pi)$ where $\mathcal{S} = \{0, 1, 2, 3, 4\}$ ={Awake, LOC, Anesthesia, ROC, Emergence} are the hidden states. In the dataset, we used the EEG to label the hidden state. Each observation $\overrightarrow{\mathcal{O}} \in \mathcal{O}$ consists of triplets (HR, MeanBP, RR), with each coordinate taking value in $\{0, 1, \dots, 4\}$ – where 0 represents very low values, and 4 very high values (see Table 1 for the threshold values). π, the initial distribution, was set to $\{1, 0, 0, 0, 0\}$ since the patient is awake at the beginning of the surgery. Finally, T and B (respectively the transition matrix and the belief matrix) were estimated using a standard maximum likelihood approach.

Thresholds

HR (/min)	MeanBP (mmHg)	RR (/min)	AAFi (%)
58	65	14	1.47
61	72	15	1.89
65	80	16	2.03
70	86	17	2.14

Table 1: Thresholds for each physiological variables.

To evaluate the model, we first used the Viterbi algorithm ([Viterbi, 1967]) in order to compute the most likely sequence of hidden states. Figure 1 displays one of those sequences. Then, we compared the predictions to the labels computed using the EEG, by considering the ratio (Number of good hidden states prediction) over (Number of hidden states). The average of this ratio being 0.907, the three physiological variables HR, MeanBP and RR seem to be promising indicators of depth of anesthesia.

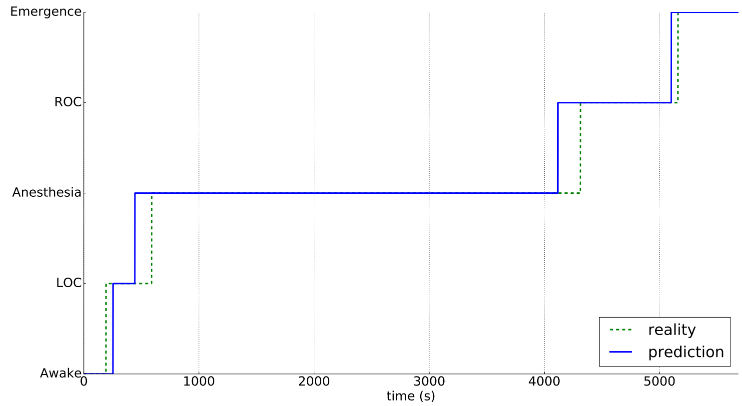

Figure 1: Real states of one patient given by analysis of EEG signals (green line) and the predicted states by HMM (blue line) during the time.

3 Learning from an expert

3.1 Setting

The main purpose of this study is to train an agent to mimic the behavior of an expert – here an anesthesiologist – during the first part of the anesthesia, i.e. Awake, LOC and Anesthesia states. To train the agent, we used an approach inspired by Apprenticeship Learning (see e.g. [Abbeel and Ng, 2004]). In this setting, we consider the actions performed by the anesthesiologist during the surgeries as expert trajectories. The goal is then to learn a policy (i.e. a function from state to action) that is able to reproduce the strategy of the expert.

Previous works have addressed the reinforcement learning problem of anesthesia's automation using the BiS (see e.g. [Martín-Guerrero et al., 2009], [Moore et al., 2014], [Nemati et al., 2016]). However, in our case, the complexity of the biological process involved in anesthesia make it extremely difficult to create a simple yet effective MDP model to encode this problem.

Consequently, we chose to address a simpler problem. Given $\mathcal{O}$, the set of observations, $\mathcal{A}$, the set of actions, and $\mathcal{H}$, trajectories of an expert, we aim to learn a policy $\hat{\pi}_*$ that mimic the expert behavior reasonably well. In particular, it is important to note that the transition matrix, T, is unknown in our setting.

In the following, state are defined as quadruplets $O =$(HR, MeanBP, RR, AAFi) discretised using thresholds defined in Table 1. Possible actions are $\mathcal{A} = \{0, 1, 2\} =$ {Reduce drug dose, Do nothing, Add drug dose}.

The policy $\hat{\pi}_*$ was learned in two steps. First $\hat{\pi}_*$ was initialized using bayesian inference with uniform prior, using the expert trajectories. Then, the information was propagated to unseen states, i.e. states which were not observed in the expert trajectories, using the following similarity measure s

$$s(O_1, O_2) = \exp\left(-\theta^T(|O_1 - O_2|)\right)$$

where | denotes the absolute value element-wise, and

$$\theta = (1, 1, 1, 100).$$

3.2 Experimental results

To evaluate the performance of the agent, we defined the global agreement rate as follow

$$\text{global agreement rate} = \frac{\text{number of agent's actions that matches the expert's action}}{\text{number of actions}}$$

On the test set considered, the average agreement rate of the agent was 61.90%, with individual values ranging from 54.8 to 71.7%. Also, it is important to remark that a significant part of the error was due to small time latency – the agent taking action a few seconds before or after the expert. While those actions would have likely produced results similar to the expert actions, we chose to classify them as error in this work.

We also evaluated our policy qualitatively, by comparing the concentration of the anesthetic drug resulting from the agent actions, and the concentration resulting from the expert actions. Figure 3.2 illustrates those results. It is worth noting that the resulting policy $\hat{\pi}_*$ is performing surprisingly well, considering the simplicity of the model used in our setting.

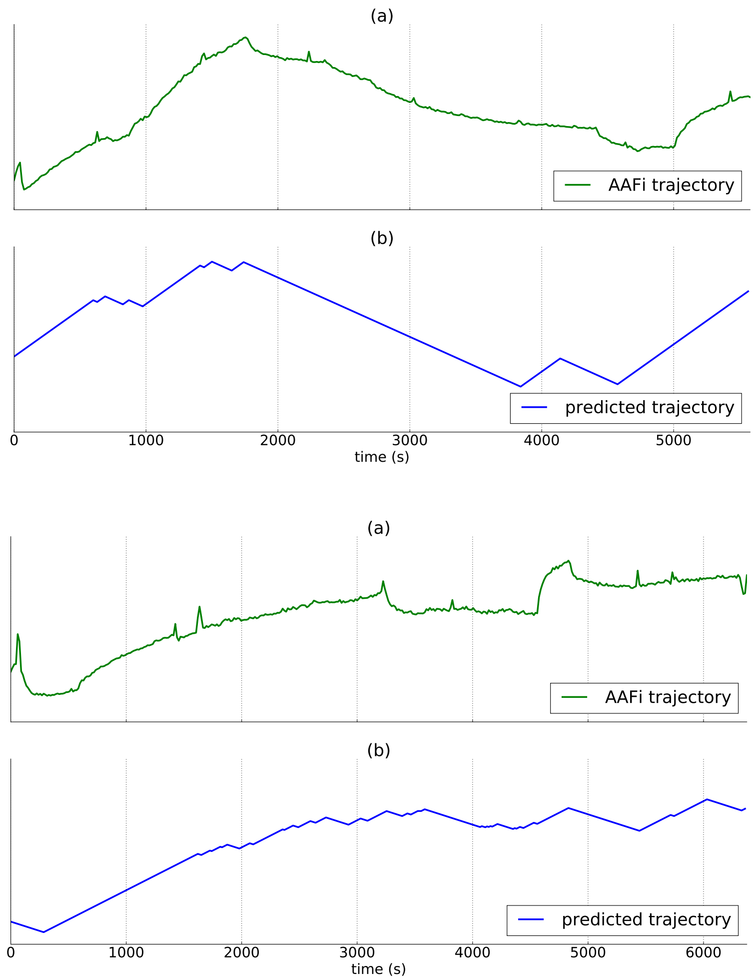

Figure 2: Examples of the evolution of the concentration of anesthetic drug during the surgery (a) resulting from the expert actions and (b) resulting from the agent actions.

4 Conclusion

In this work, we presented a simple model to encode the anesthesia process. We then learned a policy on this model, using expert trajectories. The performances of the resulting policy results look promising, and future research directions might include the creation of a more complex model, where inverse reinforcement learning could be used.

References

Pieter Abbeel and Andrew Y Ng. Apprenticeship learning via inverse reinforcement learning. In *Proceedings of the twenty-first international conference on Machine learning*, page 1. ACM, 2004.

Cambell Bennett, Logan J Voss, John PM Barnard, and James W Sleigh. Practical use of the raw electroencephalogram waveform during general anesthesia: the art and science. *Anesthesia & Analgesia*, 109(2):539–550, 2009.

Kate Leslie, Paul S Myles, Andrew Forbes, and Matthew TV Chan. The effect of bispectral index monitoring on long-term survival in the b-aware trial. *Anesthesia & Analgesia*, 110(3):816–822, 2010.

José D Martín-Guerrero, Faustino Gomez, Emilio Soria-Olivas, Jürgen Schmidhuber, Mónica Climente-Martí, and N Víctor Jiménez-Torres. A reinforcement learning approach for individualizing erythropoietin dosages in hemodialysis patients. *Expert Systems with Applications*, 36 (6):9737–9742, 2009.

Brett L Moore, Larry D Pyeatt, Vivekanand Kulkarni, Periklis Panousis, Kevin Padrez, and Anthony G Doufas. Reinforcement learning for closed-loop propofol anesthesia: a study in human volunteers. *Journal of Machine Learning Research*, 15(1):655–696, 2014.

Shamim Nemati, Mohammad M Ghassemi, and Gari D Clifford. Optimal medication dosing from suboptimal clinical examples: A deep reinforcement learning approach. In *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, pages 2978–2981. IEEE, 2016.

Daniel I Sessler, Jeffrey C Sigl, Scott D Kelley, Nassib G Chamoun, Paul J Manberg, Leif Saager, Andrea Kurz, and Scott Greenwald. Hospital stay and mortality are increased in patients having a "triple low" of low blood pressure, low bispectral index, and low minimum alveolar concentration of volatile anesthesia. *The Journal of the American Society of Anesthesiologists*, 116(6):1195–1203, 2012.

Andrew Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE transactions on Information Theory*, 13(2):260–269, 1967.

Elizabeth L Whitlock, Alexander J Villafranca, Nan Lin, Ben J Palanca, Eric Jacobsohn, Kevin J Finkel, Lini Zhang, Beth A Burnside, Heiko A Kaiser, Alex S Evers, et al. Relationship between bispectral index values and volatile anesthetic concentrations during the maintenance phase of anesthesia in the b-unaware trial. *The Journal of the American Society of Anesthesiologists*, 115 (6):1209–1218, 2011.