Adaptive Treatment of Epilepsy via Batch Mode Reinforcement Learning

Arthur Guez, Robert D. Vincent and Joelle Pineau
School of Computer Science, McGill University

Massimo Avoli
Montreal Neurological Institute

20th Conference on Innovative Applications of Artificial Intelligence
17 July 2008
Epilepsy

- Epilepsy is a common disorder of the nervous system.
  - Affects ~0.6% of all humans worldwide.
  - Up to 75% successfully treated with drug therapies.
  - A few others are candidates for surgery.

- Causes: Varied

- Symptoms: Recurrent unprovoked seizures.
Deep-brain stimulation (DBS)

- Electrical stimulation of the brain or vagus nerve is an attractive alternative treatment.

http://www.medtronic.com/
Evidence in support of DBS

Fixed frequency stimulation in brain slices decreases seizure duration

Responsive stimulation for epilepsy

• The current paradigm: Open loop

• The emerging paradigm: Responsive
  
  – Seizure prediction or detection: Algorithms measure electrical activity (e.g.) to determine patient's state.
  
  – Seizure control: Prevent seizures through electrical stimulation.
To create a stimulation device which is:

- **Adaptive**: strategy evolves as a function of the observation.
- **Automatic**: stimulation strategy learned from data.
- **Optimal**: maximize seizure reduction and minimize stimulation.
Four reasons why this is a hard problem

1. Large amount of information available at each decision point.

2. Large variance in the disease (within patient, between patients, and from animal to patient).

3. No available generative model of epilepsy.

4. Data acquisition (for exploration and validation) is “expensive.”
Human seizure variation

Two seizures recorded on the same patient:
Four reasons why this is a hard problem

1. Large amount of information available at each decision point.

2. Large variance in the disease (within patient, between patients, and from animal to patient).

3. No available generative model of epilepsy.

4. Data acquisition (for exploration and validation) is “expensive”.

![Image of healthcare technology]
Four reasons why this is a hard problem

1. Large amount of information available at each decision point.

2. Large variance in the disease (within patient, between patients, and from animal to patient).

3. No available generative model of epilepsy.

4. Data acquisition (for exploration and validation) is “expensive”.
Reinforcement Learning

• We model an agent interacting with a stochastic environment:
  – Discrete time: $t \in \{0 \ldots T\}$
  – State variables: $s \in S$
  – Actions: $a \in A$
  – Transition dynamics: $P(s'|s,a)$
  – Scalar reward function: $R(s,a)$

• The agent's goal is to learn a policy that maximizes the expected total discounted reward:

$$R_T = \sum_t \gamma^t R(s_t,a_t)$$

• The value of each state/action pairing can be expressed as:

$$Q'(s,a) = R(s,a) + \gamma \sum s' P(s'|s,a) \max_{a'} Q^{t+1}(s',a')$$
Batch RL using animal data

- Side-step exploration problem by using data from a small set of fixed (hand-selected) policies.

- Data collected using an *in-vitro* animal model of epilepsy.
  - Rat brain slices from mesial temporal structures.
  - Seizure-like discharges induced by 4AP added to medium.
  - Discharges monitored with field potential extracellular recordings.
  - Stimuli delivered using a square pulse, constant current, pattern.
Recordings of electrical activity

- Recorded from single sensing electrode placed *in-vitro*.
- Raw data consists of a set of traces of electrical field activity in the sample (> 50 million data points)
- Each trace contains several minutes with each fixed stimulation strategy.
Converting raw data to a state representation

- We need to extract: \( D = \{ <s_t, a_t, r_t, s_{t+1}>, t=1 \ldots |D| \} \)

- Divide signal into overlapping windows of different length.

\( s_t = 114 \) real-valued features
mean, range, energy,
frequency spectra
Adding actions and rewards

- We need to extract: \( D = \{ <s_t, a_t, r_t, s_{t+1}>, \ t=1…|D| \} \)

- Divide signal into overlapping windows of different length.

\[ r_t = \{ \text{normal, stimulation, seizure} \} \text{ hand-labeled for each frame.} \]

- Seizure (large negative reward)
- Stimulation (small negative reward)
- Other (zero reward)

\[ a_t = \{ 0 \text{ Hz, 0.2 Hz, 0.5 Hz, 1.0 Hz} \} \text{ stationary stimulation frequencies.} \]
Estimating the value function

• Recall the value function: \[ Q'(s,a) = R(s,a) + \gamma \sum_s P(s'|s,a) \max_{a'} Q^{t+1}(s',a') \]

• Don’t know \( P(s'|s,a) \) or how it may be affected by treatment.

• Cannot represent \( Q(s,a) \) in tabular form.

• Need a good **regression function**, appropriate for:
  – Batches of continuous-valued data.
  – High-dimensional input.
  – Limited domain knowledge.
Fitted-Q iteration algorithm

- Estimate Q-function via a sequence of supervised learning trials

Given data set: \( D = \{ <s_t, a_t, r_t, s_{t+1}>, t=1 \ldots |D| \} \)

For \( n = 1 : N \)

Form a training set \( D_n = \{ (i^l, o^l) \}, l = 1, \ldots, |D| \) \)

\( i^l = (s^l_t, a^l_t) \)

\( o^l = r^l_t + \gamma \max_{a \in A} Q_{n-1} (s^l_{t+1}, a) \)

Estimate \( Q_n \) from \( D_n \) using any regression algorithm.

Gordon 1999; Ormoneit and Sen 2002
Regression via extremely randomized trees

- Build $M$ trees independently from the complete training set $D$.
- Generate tests randomly:
  - Select $K$ candidate attributes ($f_i$).
  - Choose cut points ($t_i$) uniformly.
  - Keep the test with highest computed score.
- Predicted value is the mean of $M$ trees.

This method (and its relatives) appear to be good supervised learning algorithms for high-dimensional data.

*Ernst et al. 2005; Geurts et al. 2004*
Evaluating performance

• Now we can learn!

• How do we know if our learned policy is good?
  – We can’t compare to the optimal stimulation strategy.
  – Limited data means no optimality guarantees.
  – Data collected under different (fixed) policy.

• But we have some indicators that suggest we are not doing too badly.
  – Perform leave-one-out cross validation.
  – We use rejection sampling to simulate the policy.
Comparison to fixed stimulation strategies

Proportion of seizure states:

![Bar chart showing proportion of seizure states for different stimulation frequencies: Control, 0.2 Hz, 0.5 Hz, 1 Hz, and TBRL. The y-axis represents the proportion of seizure states, ranging from 0 to 0.15, and the x-axis represents the different stimulation frequencies.]
Comparison to fixed stimulation strategies

Number of stimulation actions:
Comparison to fixed stimulation strategies

Empirical return:

![Bar chart showing empirical return for different stimulation strategies (Control, 0.2 Hz, 0.5 Hz, 1 Hz, TBRL)].
Example trace
Future challenges

- Now conducting experiments in active slice models.
- Exploration: We need a safe and efficient method for learning online.
- Evaluation: We need better methods for evaluating the policy.
- Knowledge transfer / Multi-task learning:
  
  \[
  \text{Slice} \rightarrow \text{Slice} \\
  \text{Slice} \rightarrow \text{Animal} \\
  \text{Animal} \rightarrow \text{Human}
  \]
- Applications to other diseases (e.g. depression, schizophrenia).
We present a framework for automatically learning stimulation strategies for the treatment of epilepsy.

Preliminary results using cross-validation imply both improved seizure suppression and reduced stimulation.

Now evaluating by learning policy from batch data and then applying the policy to an active slice.
Acknowledgements

- Collaborating researchers:
  - Keith Bush, Ph.D., SOCS
  - Philip de Guzman, Ph.D., MNI
  - Gabriella Panuccio, Ph.D. Student, MNI

- Funding
  - Natural Sciences and Engineering Research Council
  - Canadian Institutes for Health Research