Real-time Machine Learning for Assistive Medical Robotics

Patrick M. Pilarski

Reinforcement Learning & Artificial Intelligence Laboratory
Alberta Innovates Centre for Machine Learning

Dept. Computing Science, University of Alberta
Edmonton, Alberta, Canada

Chiba University, Nov. 18, 2011
Outline

The University of Alberta

Toward Intelligent Artificial Limbs

Reinforcement Learning

Real-time Prediction in a Clinical Setting
  \textit{(General Value Functions \& Nexting)}

Results (Able-bodied \& Clinical)
Alberta Nature
The University of Alberta

- Opened in 1908
- 6,000 Graduate students
- 30,000 Undergraduate students
- 1 of the top 100 universities in the world
Department of Computing Science
University of Alberta
Reinforcement Learning & Artificial Intelligence Lab

PIs: Rich Sutton, Csaba Szepesvari, Michael Bowling, Dale Schuurmans
Outline

- The University of Alberta
- Toward Intelligent Artificial Limbs
- Reinforcement Learning
- Real-time Prediction in a Clinical Setting
  (General Value Functions & Nexting)
- Results (Able-bodied & Clinical)
Multifunction Myoelectric Prostheses

Otto Bock’s Dynamic Arm combined with myoelectric wrist rotator and prehensor.

Three Known Barriers

“Three main problems were mentioned as reasons that amputees stop using their ME prostheses: nonintuitive control, lack of sufficient feedback, and insufficient functionality.”

— Peerdeman et al., JRRD, 2011.
Intelligent Interfaces

(Prostheses that approach and someday exceed the abilities of a biological limb.)
Intelligent Interfaces

(Prostheses that approach and someday exceed the abilities of a biological limb.)

DLR Hand Arm System

Image: German Aerospace Center (DLR) & IEEE Spectrum
Intelligent Interfaces

(Prostheses that approach and someday exceed the abilities of a biological limb.)


DLR Hand Arm System
Image: German Aerospace Center (DLR) & IEEE Spectrum
Machine Learning

In the face of growing complexity, learn the correct way to map numerous EMG signals to actuator commands.

*DLR Hand Arm System*
*Image: German Aerospace Center (DLR) & IEEE Spectrum*
State of the Art

- Excellent examples of machine learning work in classifying EMG patterns for use in limb control (e.g. Oskoei and Hu ‘08, Parker et al. ‘06, Sensinger et al. ’09).

- However, most contemporary learning approaches rely on external knowledge of their domain to guide learning, and function primarily in offline or batch learning scenarios.

- This breaks down as the complexity and individuality of the input and output space increases; very hard to determine the “correct” thing to do.
Three Missing Elements

• **Real-time machine learning.**
  *(Online, adaptive algorithms; noted by Sensinger et al. ’09, Scheme & Englehart ‘11)*

• **Generalized interfaces.**
  *(Blank-slate human-machine interaction & collaboration; e.g. Pilarski et al. 2011)*

• **Data-respecting biomedical pattern analysis.**
  *(Complexity is good: interpreting myriad signals without reducing the sensorimotor space)*
Ongoing Projects

• **Real-time control learning** without *a priori* information about a user or device.

• **Prediction and anticipation** of signals during patient-device interaction.

• **Collaborative algorithms** for the online human improvement of limb controllers.
Team

• **RLAI / AICML**: new methods for improved control, feedback, and online interaction.

• **Mec. Eng.**: new mechanical limbs and platforms for amputee training (MTT).

• **Glenrose / Medicine**: new surgeries (TMR & TSR), patients, and clinical expertise.
Setting: TMR/TSR
Useful Predictions

- Assuming we continue as usual (on-policy):
  - What will the force sensor report over the next few seconds? *(Slippage/gripping.)*
  - Where will the limb be in the next 30s? *(Safety; fluid multi-joint motion.)*
  - How strong will each user EMG signal be in 250ms? *(User intent; preemptive motion.)*

* Address key issues, as per Scheme and Englehart, JRRD, 2011; Peerdeman et al., JRRD, 2011.
Outline

- The University of Alberta
- Toward Intelligent Artificial Limbs
- Reinforcement Learning
- Real-time Prediction in a Clinical Setting  
  *(General Value Functions & Nexting)*
- Results  *(Able-bodied & Clinical)*
Reinforcement Learning
An Introduction
Richard S. Sutton and Andrew G. Barto

Reinforcement Learning is an approach to:

- Natural intelligence
- Artificial intelligence
- Optimal control
- Operations research
- Solving partially observable Markov decision processes

(and the perspective that all of these are the same)
Main Ideas

• Reinforcement learning involves an agent and an environment.

• The learning system (agent) perceives the state of the environment via a set of observations and takes actions.

• It then receives a new set of observations and a reward.

• These observations and rewards are used to predict future rewards, and to change the agent’s policy (how it selects actions).

• Key point: A single, scalar reward signal drives learning.
Reinforcement Learning

Intelligent Agent

World (Environment)

ACTION

REWARD

STATE

EXPERIENCE
Number RL Papers per Year

Google scholar hits for the phrase “reinforcement learning”
RL Headlines

• RL is widely used in robotics
• RL algorithms have found the best known approximate solutions to many games (RL is part of the revolution in solving Go)
• RL algorithms are now the standard model of reward processing in the brain
• RL breaks the curse of dimensionality
What is Special About RL?

- Radical generality
- None of the signals are given any interpretation
  ... no reference signals or labels
  ... no human interpretation, no calibration
- Just data in the form of signals
  ... one of which is to be maximized (reward)
Outline

- The University of Alberta
- Toward Intelligent Artificial Limbs
- Reinforcement Learning
- Real-time Prediction in a Clinical Setting
  (General Value Functions & Nexting)
- Results (Able-bodied & Clinical)
Setting

Diagram thanks to K. Ming Chan
Useful Predictions

• Assuming we continue as usual (on-policy):
  - What will the force sensor report over the next few seconds? *(Slippage/gripping.)*
  - Where will the limb be in the next 30s? *(Safety; fluid multi-joint motion.)*
  - How strong will each user EMG signal be in 250ms? *(User intent; preemptive motion.)*

* Address key issues, as per Scheme and Englehart, JRRD, 2011; Peerdeman et al., JRRD, 2011.
Online Nexting

- **General Value Functions.**
  (Sutton et al., 2011, AAMAS)

- GVF forms questions; “what will happen next?” (**Nexting**)

- **In brief:** instead of reward, learn anticipations (expectations of real-valued signals).

- Can learn many temporally extended predictions in parallel.
General Value Functions

- Conventional value functions are predictions w.r.t. the rewards, discount, and terminal values of the problem, for a given policy:

\[ Q^\pi(s, a) = \mathbb{E}[r_1 + \gamma r_2 + \gamma^2 r_3 + \cdots | s_0 = s, a_0 = a, a_{1:\infty} \sim \pi] \]

\[ = \mathbb{E}[r_1 + \cdots + r_k + z_k | s_0 = s, a_0 = a, a_{1:k} \sim \pi, k \sim \gamma] \]

- General value functions are predictions w.r.t. to four given functions:

\[ Q^{\pi, r, \gamma, z}(s, a) = \mathbb{E}[r(s_1) + \cdots + r(s_k) + z(s_k) | s_0 = s, a_0 = a, a_{1:k} \sim \pi, k \sim \gamma] \]

these four functions define the semantics of the prediction

Sutton et al., AAMAS, 2011.
General Value Functions

\[ Q^{\pi,r,\gamma,z}(s,a) = \mathbb{E}[r(s_1) + \ldots + r(s_k) + z(s_k) \mid s_0 = s, a_0 = a, a_1:k \sim \pi, k \sim \gamma] \]

these four functions define the semantics of the prediction

- **Policy** \( \pi : A \times S \rightarrow [0, 1] \)
- **Reward** \( r : S \rightarrow \mathbb{R} \)
- **Termination** \( \gamma : S \rightarrow [0, 1] \)
- **Terminal Value** \( z : S \rightarrow \mathbb{R} \)

*Sutton et al., AAMAS, 2011.*
Why GVFs?

- Thousands of accurate predictions can be made and learned in real time (i.e., 10hz)
- A single state representation be used to accurately predict many different sensors at many different time scales.
- A model-free algorithm that can learn fast enough to be useful.

Sutton et al., AAMAS, 2011.
Massively Parallel Prediction

HUMAN → EMG SIGNALS → CONTROL → ACTUATOR SIGNALS → ROBOT

P1, P2, P3, PN → GVF 1, GVF 2, GVF 3, GVF N

FXN APP

FEEDBACK SIGNALS → CONTROL SIGNALS

STATE
Predictions (Nexting)

\[ p_t^i = f_t^\top w^i = \sum_j f_t(j)w^i(j) \approx r_{t+1}^i + \gamma^i r_{t+2}^i + (\gamma^i)^2 r_{t+3}^i + (\gamma^i)^3 r_{t+4}^i + \cdots \]

- Where each prediction has its own reward \( r_t^i \) and discount rate \( \gamma^i \in [0, 1) \)
- Ideal predictions are the convolution of the reward with an exponential kernel
Learning GVFs

- Temporal-Difference (TD) learning
- Linear TD(\(\lambda\)) (Sutton, 1988)

\[
\begin{align*}
\mathbf{w}_{t+1}^i &= \mathbf{w}_t^i + \alpha \left[ r_{t+1}^i + \gamma^i \mathbf{f}_{t+1}^\top \mathbf{w}_t^i - \mathbf{f}_t^\top \mathbf{w}_t^i \right] \mathbf{e}_t^i \\
\mathbf{e}_t^i &= \gamma^i \lambda \mathbf{e}_{t-1}^i + \mathbf{f}_t
\end{align*}
\]

just \(f_t\) if \(\lambda=0\)
Outline

- The University of Alberta
- Toward Intelligent Artificial Limbs
- Reinforcement Learning
- Real-time Prediction in a Clinical Setting
  (General Value Functions & Nexting)
- Results (Able-bodied & Clinical)
Able-Bodied Study

- Eight runs at 5–10 min (12k–25k timesteps).
- Record EMG signals and joint angles.
- Recording and learning at 40Hz.
- GVF state = TileCoder(8,10) \{C-EMG × 2, elbow joint angle, wrist joint angle\}. 
Prediction Results

For joint angle signals after ten minutes of online learning ($\gamma=0.97$).

ANTICIPATION RESULTS

VALIDATION RESULTS
Prediction Results

For EMG signals after ten minutes of online learning ($\gamma=0.97$).
Learning Curves

Over eight independent runs.

- ACTUATOR SIGNALS
- EMG SIGNALS
Clinical Experiments

• Approximately 20min of patient interaction with the MTT system (~60k timesteps).
• Recorded EMG signals, force signals, joint angle, joint speed, joint temp., joint load.
• Recording samples at 50Hz.
• GVF state = TileCoder(8,10) { EMG x 2, force, joint angle, joint speed}. 
Prediction Results

After three iterations through the training data.

T1 = 0.97  T2 = 0.99

ANTICIPATION RESULTS

VALIDATION RESULTS
Results on Test Data

After three iterations through the training data. Testing data previously unseen by the system; no learning during testing evaluation.
Learning Curves

Over ten iterations through the training data.

**HAND ANGLE**
Active Angular Range: 500–700
Mean Abs. Error on testing data after first iteration: <7%

**CONTACT FORCE**
Sensor Range: 0–3V
Mean Abs. Error on testing data after first iteration: <6%
Learning Curves

Over ten iterations through the training data.

**HAND ANGLE**

Active Angular Range: 500–700
Mean Abs. Error on testing data after first iteration: <7%

**CONTACT FORCE**

Sensor Range: 0–3V
Mean Abs. Error on testing data after first iteration: <6%
Summary

• *Real-time machine learning* can help alleviate barriers to assistive rehabilitation robotics.

• Recent work is on *prediction* and *anticipation* for improving the control of artificial limbs.

• **Results:** successful on-policy nexting for both patient data and able-bodied subject data.

• **Big picture:** artificial limbs that learn and improve through on-going user interaction.
• Dr. Richard S. Sutton, Dr. Thomas Degris
  RLAI, Dept. Computing Science, University of Alberta

• Michael R. Dawson, Dr. Jacqueline S. Hebert, Dr. K. Ming Chan
  Glenrose Rehabilitation Hospital & University of Alberta

• Dr. Jason P. Carey
  Dept. of Mechanical Engineering, University of Alberta

• Funders: Alberta Ingenuity Centre for Machine Learning (AICML), the Natural Sciences and Engineering Research Council (NSERC), Alberta Innovates – Technology Futures (AITF), and the Glenrose Rehabilitation Hospital Foundation.
Questions

... and thank you very much for your hospitality and attention.

patrick.pilarski@ualberta.ca

http://www.ualberta.ca/~pilarski/