Multilayer General Value Functions for Robotic Prediction and Control

Craig Sherstan
Department of Computing Science
University of Alberta
Edmonton, Alberta, Canada
Email: sherstan@ualberta.ca

Patrick M. Pilarski
Division of Physical Medicine and Rehabilitation
University of Alberta
Edmonton, Alberta, Canada
Email: pilarski@ualberta.ca

***Abstract*—Predictions are a key component to intelligence and necessary for accurate motor control. In reinforcement learning, such predictions can be made through general value functions (GVFs). This paper introduces prosthetic arms as a domain for artificial intelligence and discusses the role that predictions play in prosthetic limb control. We explore the use of multilayer predictions, that is, predictions based on predictions, using robotic and simulation experiments. From these experiments two observations are made. The first is that compound predictions based on GVFs are viable in a robotic setting. The second, is that strong GVF predictors can be built from weaker ones with different input and target signals, similar to boosting. Finally, we theorize how such topologies might be used in transfer learning and in the simultaneous control of multiple actuators. Our approach to integrating machine intelligence with robotics has the potential to directly improve the real-world performance of bionic limbs.**

I. Gentle Integration

When combining machine intelligence systems with electromechanical devices such as mobile or mounted robots, it is natural to think of the machine intelligence as providing most or all of the key aspects of the robot's control system. Integration of this kind is often challenging—it simultaneously addresses many important barriers faced by our computing technology—but is incredibly fruitful for both the fields of robotics and artificial intelligence. Another, complementary approach is the use of machine intelligence to supplement an existing control system or sensorimotor interface. Machine learning and artificial intelligence (AI) can augment the capacity of existing systems in small but important ways. While more modest in its aims, this kind of staged deployment is well suited to the refined study of individual machine learning methods as they impact real-world domains of use. It further provides a smooth pathway to machine intelligence seeing practical use within complete, deployed systems.

In this paper we look specifically at the second, more gentle approach to integrating machine intelligence within a robotic device. In particular, we highlight one area where our group has made recent progress: improving robotic artificial limbs (Fig. 1) through real-time learning and utilization of temporally extended predictions. This setting lends itself well to translating algorithmic and conceptual advances into tangible benefit within a deployed environment; machine learning can improve the ability of people with amputations to control their bionic limbs. Sharing the challenges and opportunities of prosthetics as a domain for AI Robotics is the first contribution of our paper. We present a brief overview of our machine learning work within the prosthetic domain, and follow on this overview with a concrete example on a simple robotic platform of how real-time predictions can be beneficially combined into a learning hierarchy. Lastly, we discuss how multilayer predictions can be integrated back into prosthetic control approaches to further extend their practical reach.

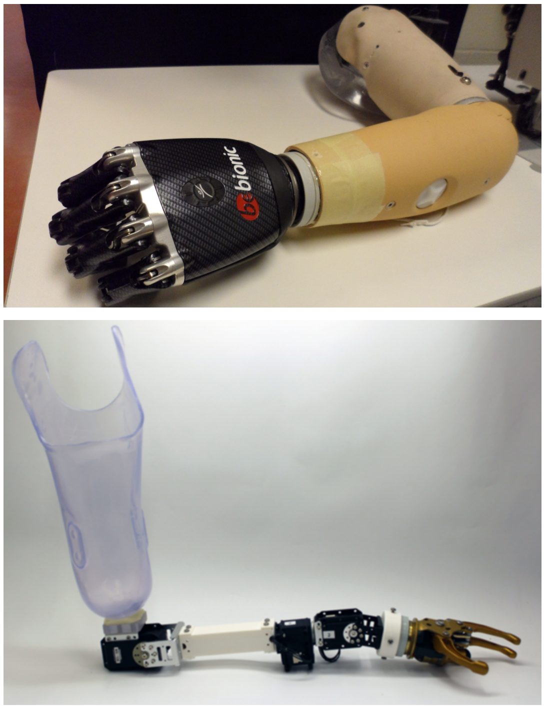

Fig. 1. Augmentative and restorative prosthetics are of specific interest for incrementally integrating AI into a robotic setting. *Top:* commercial limb system prescribed to an amputee for use during daily life. *Bottom:* research robot limb system with direct access to a rich sensorimotor stream [4].

II. Bionic Limbs

Bionic limbs are robotic devices fixed directly to the body of someone with a motor impairment or complication (e.g, someone with an amputation), or for the purposes of extending or augmenting the abilities of healthy individuals. These devices have multiple actuators and sensors, both on and off the human body, and use this sensorimotor information to interpret a user's intent and actuate the joints of the robot limb accordingly. Despite the growing availability of dexterous robotic prosthetic arms, amputees often reject these arms due to the difficulty they find in their control [1]–[3]. The most common approach to controlling such arms is the use of electromyographic signals (EMG), which are the electrical activities of muscles. Unfortunately, the number of control signals available from EMG is much lower than the control space of the robot arms, creating a large gap between user intent and achievable motor outcomes. There are a number of techniques people have tried to address this gap, some on the software side, such as pattern recognition [2], and some on the clinical side, such as targeted muscle reinnervation[1] [5]. However, control remains difficult and indeed, there will almost always be a disparity between signal and control spaces.

Our goal is to apply artificial intelligence to the control of these arms, in such a way as to make using them more intuitive and functional for the users [6]–[9]. We propose that a more complete way to think about the prosthetic control problem is that we are looking to create an assistive, context aware robot, which happens to be a prosthetic arm. The techniques we are developing here are also applicable beyond the scope of prosthetic arms. Our approach has been to incrementally apply AI techniques to existing control schemes for other assistive and augmentative devices. One of the great benefits of working with prosthetic devices is that the users of these devices have clear objectives that they need the prostheses to address, and concrete measures for the success of the system. Additionally, there is a clear path to commercial and clinical use.

III. Improvement from Ongoing Experience

Making forward predictions is believed to be a key component in making accurate motor commands [10]–[12]. Further more, predictions have been shown to be an important way to think about and formalize the state information being provided to a learner (for example, predictive representations of state [13]). By learning and maintaining predictions in real time, it is possible for a robotic system to acquire and self-verify small pieces of knowledge in an autonomous fashion as it interacts with the world [14]–[17].

Incremental, ongoing knowledge can be acquired using techniques known as general value functions (GVFs), a generalization of the reward-based value functions common in reinforcement learning (RL) [14]. While other forms of machine learning might be used for prediction, RL algorithms are somewhat unique in their ability to learn online and continuously in a computationally efficient manner. In GVFs, replacing reward with a target signal allows a system to learn either cumulative, Eq. (1), or instantaneous, Eq. (2), predictions for any scalar signal. For example, we can ask "How much total current will the shoulder servo use in the next 10s?" or "What will the light sensor read in 3s?" GVFs can also be used to give the probability of a binary event occurring, e.g., "What is the probability of colliding with the wall in the next 5s?" GVFs can be thought of as representing temporally extended knowledge about a robot, its environment, and the interaction between the two.

The GVF algorithm is composed of three main steps: calculation of the temporal difference (TD) error (Eq. 1, 2), calculation of traces (Eq. 3), and weight vector update (Eq. 4). Note that Eq. (3) shows the form of replacing traces used in the experiments described in this paper, but other types of traces may be used (we suggest the approach of Van Siejen and Sutton [18]).

$$\delta_{t+1} = r_{t+1} + \gamma\phi_{t+1}^T\theta_t - \phi_t^T\theta_t \tag{1}$$

$$\delta_{t+1} = \beta r_{t+1} + \gamma\phi_{t+1}^T\theta_t - \phi_t^T\theta_t \tag{2}$$

where

δ temporal difference error

r in GVFs this represents the target signal to be predicted

γ continuation probability, # timesteps lookahead$= 1/(1-\gamma)$

ϕ input feature vector

θ learned weight vector

β termination probability $= 1-\gamma$

$$e_{t+1} = \lambda\gamma e_t + \frac{\alpha\phi_t}{max(1, ||\phi_t||_0)} \tag{3}$$

where

e is the eligibility trace

λ trace decay rate (amount of bootstrapping)

α learning rate

$$\theta_{t+1} = \theta_t + \delta_{t+1}e_t \tag{4}$$

GVFs have seen some promising application with robots. Sutton et al. demonstrated that GVFs were able to simultaneously learn to predict large numbers of sensorimotor signals in an online fashion on a mobile robot [14], [15]. Some studies have also looked at using GVFs in control. In particular, Modayil and Sutton have used prediction with a nexting approach to control a simple mobile robot [16], such that, when a prediction exceeds a threshold the robot will activate and follow a fixed, hand-coded behavior. Specifically, when their mobile robot predicted a future over-current condition it would shut off the motors. This approach is similar to many prediction-based reflexive reactions found in humans and other animals [11], [12].

The idea to make a predictive link to known control behaviors also fits well within the domain of artificial limbs. A

[1] Targeted Muscle Reinnervation is a surgical procedure where the nerves that would have gone to the missing limb are transplanted into new host muscle tissue in the residual limb, such as the biceps, triceps or pectoral muscles [5].

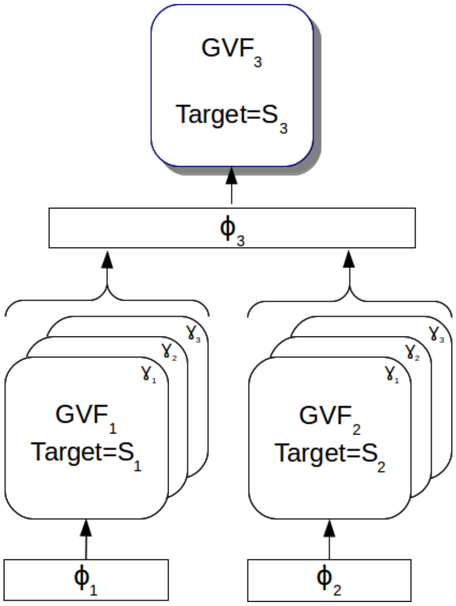

Fig. 2. Topology. From bottom up: ϕ_1 and ϕ_2 are primary layer feature vectors, which may or may not be the same, depending on the experiment. Primary layer GVFs are grouped by the target signal, with one or more lookahead values (γ). The output of the primary layers are then used as input, possibly with other inputs, as features to a secondary layer GVF.

typical control setup for using EMG to control a prosthetic arm is to use two EMG signals to proportionally control the velocity of one joint at a time. Active joint selection is performed by toggling through a fixed joint list via another EMG signal or a mechanical switch. As one can imagine, this is a very tedious way to control an arm. Edwards et al. have demonstrated improved task performance using an adaptive switching order based on learned predictions. When an amputee user begins a toggle sequence, the joints are selected in the order that the learner predicts will be most likely needed at the moment; this was found to reduce the number of voluntary switching interactions needed to complete a simple manipulation task, and thus also the time needed to complete the task [8]. Users appeared to be happy with the improvement and to develop increased trust in the system. Additionally, Pilarski et al. controlled the wrist joint of a 3 DOF robot arm where the objective was to have the controller place the wrist in the position it predicted it should go in the near future, given the current state [7]. This study demonstrated the ability to use GVF predictions as direct target signals for control as well as in combination with actor-critic RL agents (e.g., as predictive state information).

Our present paper now proposes an extension of these examples through the use of multilayer predictions, i.e. predictions based on predictions (Fig. 2). Predictions such as these represent compound knowledge about the environment and in some cases can be thought of as hierarchies where each layer represents a level of abstraction. Imagine that we have two predictors: 1. “Is a Tiger nearby?”, 2. “Will I have an asthma attack in the near future?” A very important prediction to make is, “Am I in danger?”, for which the previous two predictions would be valuable. In the context of a robotic arm we can imagine similar scenarios. For example, in a prosthetic task we could structure a set of GVF predictions as follows:

Fig. 3. Create recording session.

- Where is the elbow moving to? Where is the shoulder moving to? → Where should the wrist move to?
- Where is my hand moving to? Does Joe want coffee? Is there coffee in front of Joe? → Should I open my hand?

IV. Experiments

To examine the feasibility of multilayer GVF predictions in prosthesis use, we first performed a set of preliminary tests on a more controlled experimental setting. Architectures like those shown in Fig. 2 were tested in several contexts: first in simulation with a series of deterministic square pulses, next with a series of stochastic square pulses, and finally on an iRobot Create mobile robot (Fig. 3). However, for the sake of brevity, only a subset of these experiments are discussed here.

GVF learning was conducted as described in prior work [6], [15], [16]. The feature vectors used in temporal difference learning of the GVFs had the following form. Each feature vector had a bias unit of 1, followed by a signal transform, which was some representation of the input signals for that layer. Lastly, a representation of history of the signal transform was optionally appended depending on the experiment:

1	Signal Transform	History ...

No recurrence was used, that is, the output of a GVF was not used as input to itself. The representations used meant that the feature vector was completely binary, i.e. every feature was 0 or 1. Additionally, the length of the feature vectors were constant. Scalar signals, such as sensor values and the GVF outputs, were converted to binary features using tile coding function approximation (e.g., as done in [15]), which allows for nonlinear transforms on the signal space.

A. Create Robot

The Create robot (iRobot, Inc.) is a very simple mobile robot with a limited number of sensors, similar to a Roomba vacuum. The sensors used in this experiment were:

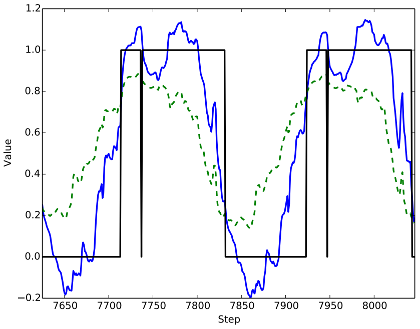

Fig. 4. Create predicting the thresholded right cliff sensor at different timescales. *Black line:* Actual Right cliff sensor, *Blue line:* Prediction at 4 timesteps, *Dashed Green Line:* Prediction at 30 timesteps

- 4 downward facing cliff sensors along the front edge. Thresholded to a binary on/off signal.
- 1 forward facing wall sensor. Value range: 0 to 4095
- wheel speed sensors. Value range: -500 to 500

The goal for this experiment was to accurately predict the turning on and off of three of the cliff sensors (Left, Front-Left, Front-right) at the primary layer and then make predictions about the fourth cliff sensor (Right) based only on the outputs of the primary layer. The Create rotated counter-clockwise for 20 minutes, randomly changing speed every 2 minutes. As the Create spun it passed over various surfaces: black tape, blue tape, beige tiles and black tiles. Additionally, objects were placed around the Create, which gave readings for the forward facing wall sensor. Under this behavior the Right cliff sensor would be the last sensor in the sequence to pass over a given surface. Fig. 3 shows the experimental setup.

Control and data recording was performed on a Raspberry Pi running the Robotic Operating System (ROS). Prediction was performed offline after recording.

The outputs of the primary layer GVFs were tile coded individually, and in pairs. Additionally, a history was used in the input feature vector to the secondary layer GVF.

At the primary layer, reasonable predictions of the Left, Front-Left, Front-right, and Right (used for comparison with the secondary layer) sensors were learned. Using the representation described, it was indeed possible to learn to make predictions for the Right cliff sensor using the outputs of the primary layer GVFs (excluding the reference primary Right cliff sensor predictors) as shown in Fig. 4. While these results are expected, it is important to establish the validity of the topology and implementation used. In comparing the predictions of the secondary GVF and the reference primary GVF for the Right cliff sensor, the secondary GVF actually had a lower error, as compared against the ideal predictor, than the primary. It is important to note that the representations used produced a significantly larger feature vector for the secondary layer, which likely accounts for the lower error observed, i.e., the secondary layer GVF had a higher resolution view into the data than did the primary layer.

B. *Combining Weak Predictors to Produce a Stronger Predictor*

We also examined whether combining weak predictors could produce a stronger predictor, akin to the concept of boosting in machine learning [19]. This scenario was tested using three square pulses of the same size, but different temporal offsets, as targets and input signals. In this setting GVF_1(Target=S_1, timescales=2,4,8,10 timesteps) and GVF_2(Target=S_2, timescales=2,4,8,10 timesteps) used an impoverished feature space that was not sufficient to predict the signal. Each of ϕ_1 and ϕ_2 contained only a bias feature, the target signal, and the inverse of the target signal (1-Target). The output of the primary layer GVFs was then tile coded individually and in pairs. GVF_3's target was the third square pulse, S_3, and was predicted at a lookahead of 4 timesteps. No history was used in the features at either level. The best that the primary layer GVFs could do with such an inadequate state space was to chase the signal. Despite this, GVF_3 was able to learn to accurately predict S_3 as shown in Fig. 5.

Essentially, the output of the primary GVFs served as a form of history for the two signals, providing more information about the signals than was directly available from the representation used. It seems reasonable to conclude that we should expect this sort of boosting behavior as long as the primary GVFs are at least somewhat temporally correlated with that secondary layer target.

V. Moving Forward: Opportunities for Integration

We believe that the use of GVFs and hierarchies of GVFs will prove beneficial to the simultaneous multi-joint control of prosthetic arms. In particular, we propose two applications that go beyond what has already been demonstrated with the adaptive switching work demonstrated by Edwards et al. [8].

A. *Transfer Learning between Simulation and Real-world*

Learning on a robot is expensive in terms of time and risk to hardware. For these reasons it is very desirable to be able to train in simulation and then transfer what is learned to the real world. One approach in RL is to learn a policy in simulation and then use that learned policy in the real world, although this has had limited success [20]. The topologies of GVFs presented in this paper suggest an approach like the one shown in Fig. 6. In this scenario, a GVF learns to predict some signal, such as joint angle, in simulation. In the real world, another GVF learns to predict the same signal, using the output of the simulation learned GVF as an adviser, in coordination with other input data. In theory, this should allow for more rapid learning in the real world, with GVFs that are already partially learned. In reality, we do not expect that the transferred GVF should predict that well, given the difficulty of accurately simulating. However, the results presented in this paper, where a strong predictor was based on weak ones, lead us to believe that we should still see some benefit using this technique. Our hope is that this will greatly reduce the amount of time needed for an amputee to train their prosthetic. Additionally, this

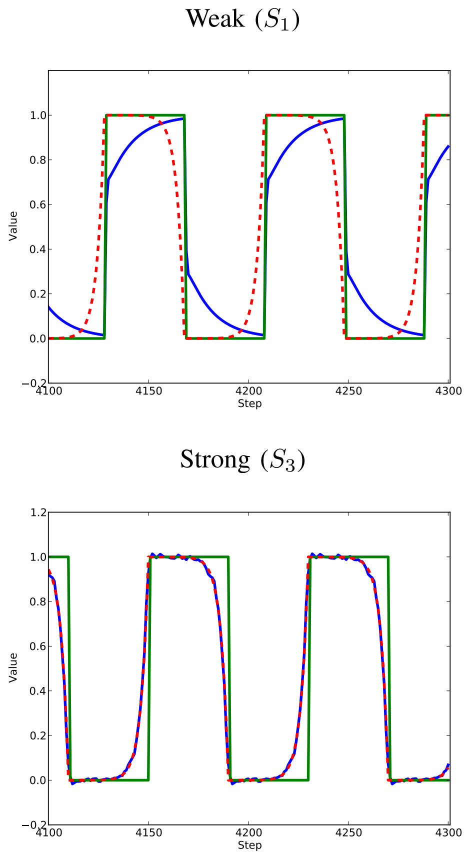

Fig. 5. Weak and Strong predictors. *Top*: A Weak predictor, target=S_1, which is only able to chase the signal, not predict it. The distinctive shape of the blue line is a clear indicator that the state space is insufficient. The graph for target S_2, not shown, is the same, just shifted in time. *Bottom*: A Strong predictor, target=S_3 is learned with only weak predictors as input. *Solid Green:* Target signal, *Solid Blue:* Prediction at 4 timesteps, *Dashed Pink:* Ideal prediction for 4 timesteps calculated from post-processing .

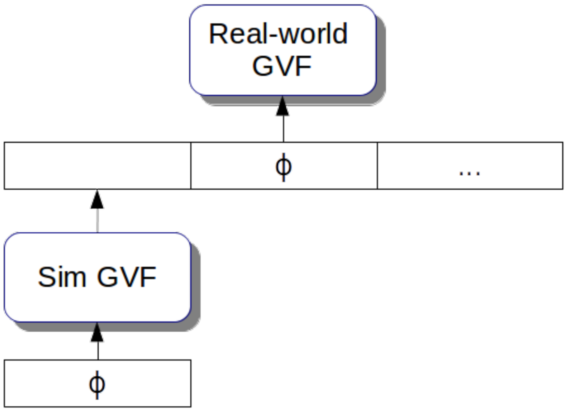

Fig. 6. Transfer learning using a multilayer topology of GVFs.

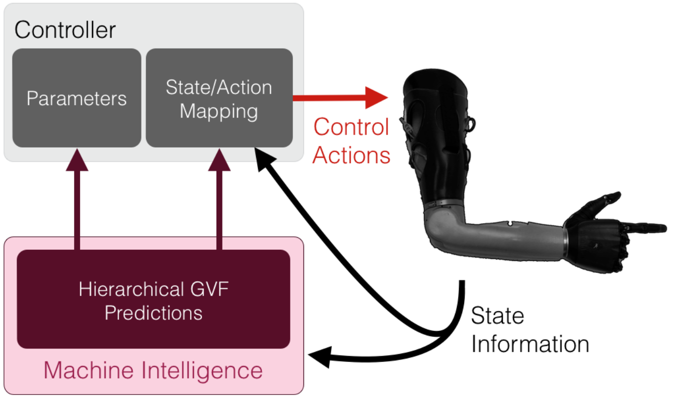

Fig. 7. *Integration approach:* machine intelligence and automatically acquired knowledge—in this case a multilayer topology of predictions—is used to extend the capacity of conventional control systems within an artificial limb.

technique could also be used with aggregated learning where the adviser is a GVF representing the cumulative predictive advice learned by many robots or from interactions with many users.

B. *Predictions for Simultaneous Multi-joint Control*

Ultimately, our aim is to use predictions for control (Fig. 7). One particular challenge of interest is the simultaneous control of multiple joints of a prosthetic limb via limited input channels—an open issue in the prosthetic domain [2].

As was mentioned, prior studies have shown clear, task specific ways of basing control on predictions [7]–[9], [16]. Predictions represent a type of temporal forward model, which are useful to thinking agents, be they biological or mechanical, and are a necessary component in developing good motor control, asking questions like, “Where is my hand moving?”, “Am I going to collide with something?”, and “What direction will I be heading 3s from now?” They are also important for higher levels of intelligence and control. For an intelligent assistive robot, such as the prosthetics we are creating, understanding a user, their environment and the current situation are long term goals. In order to do this, higher level predictions are necessary, such as, “Is the user upset?”, “Is the user hungry?”, “Is the user in danger?”, “Which object might the user want to grab?”. At both levels predictions are useful and understanding them at the more primitive level is an incremental step towards understanding the more complicated types of predictions needed for the higher level.

For low-level control there are specific ways in which we might leverage hierarchies of predictors. For example, it may be useful to make a prediction about the target position of the wrist given predictions about the target positions of all the other joints in a robot arm. Additionally, by using hierarchies of predictions we have the potential benefit of speeding learning, where, under certain circumstances, we can imagine a reduction in state space at the secondary or higher levels of predictors. Finally, under certain circumstances, we would expect to see a gain computationally where a particular prediction might be leveraged in many hierarchies. This would be more efficient than having each of the secondary layers calculating predictions directly from the data themselves, each performing the same calculations.

VI. Conclusion

As a first contribution of this work, we identified one domain—that of robotic artificial limbs—where the integration of machine intelligence with robotic systems has both clear utility and immediate areas for incremental progress. The second contribution of this work was to examine the use of multilayer topologies of prediction learners, particularly as they would apply to robots. Two main results were observed from these experiments. The first is that it is possible to learn a reliable prediction during robot operation when using the output of other predictors as input. To our knowledge, this is the first example of multilayer GVFs being applied during robot control. The second result is that it is possible to combine the output of weak GVF predictors with different target signals and input spaces to create a strong predictor of a third target signal. These two results will be useful in developing robust control methods for prosthetic robots; as a final contribution of this paper, we suggested two ways that multilayer predictions could be beneficially deployed within bionic limbs and other robotic applications. Future work in this area promises to benefit both the users of human-machine interfaces and researchers seeking to better understand the links that can be made between robot control and advances in machine intelligence.

Acknowledgments

The authors acknowledge support from the Alberta Innovates Centre for Machine Learning (AICML), Alberta Innovates – Technology Futures (AITF), and the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

[1] B. Peerdeman, D. Boere, H. Witteveen, R. Huis in 't Veld, H. Hermens, S. Stramigioli, H. Rietman, P. Veltink, and S. Misra, "Myoelectric forearm prostheses: State of the art from a user-centered perspective," *J. Rehab. Res. Dev.*, vol. 48, no. 6, pp. 719–738, 2011.

[2] E. Scheme and K. B. Englehart,"Electromyogram pattern recognition for control of powered upper-limb prostheses: State of the art and challenges for clinical use," *J. Rehab. Res. Dev.*, vol. 48, no. 6, pp. 643–660, 2011.

[3] L. Resnik, M. R. Meucci, S. Lieberman-Klinger, C. Fantini, D. L. Kelty, R. Disla, and N. Sasson, "Advanced upper limb prosthetic devices: implications for upper limb prosthetic rehabilitation," *Arch. Phys. Med. Rehabil.*, vol. 93, no. 4, pp. 710–717, 2012.

[4] M. R. Dawson, C. Sherstan, J. P. Carey, J. S. Hebert, P. M. Pilarski, "Development of the bento arm: An improved robotic arm for myoelectric training and research," in *Proc. of the Myoelectric Controls Symposium (MEC'14)*, Fredericton, New Brunswick, August 18–22, 2014, pp. 60–64.

[5] J. S. Hebert, K. Elzinga, K. M. Chan, J. Olson, and M. Morhart, "Updates in targeted sensory reinnervation for upper limb amputation," *Curr. Surg. Rep.*, vol. 2, no. 3, art. 45, pp. 1–9, 2014.

[6] P. M. Pilarski, M. R. Dawson, T. Degris, J. P. Carey, K. M. Chan, J. S. Hebert, and R. S. Sutton, "Adaptive artificial limbs: A real-time approach to prediction and anticipation" *IEEE Robot. Autom. Mag.*, vol. 20, no. 1, pp. 53–64, 2013.

[7] P. M. Pilarski, T. B. Dick, and R. S. Sutton, "Real-time prediction learning for the simultaneous actuation of multiple prosthetic joints," *Proc. of the 2013 IEEE International Conference on Rehabilitation Robotics (ICORR)*, Seattle, USA, June 24–26, 2013, pp. 1–8.

[8] A. L. Edwards, M. R. Dawson, J. S. Hebert, R. S. Sutton, K. M. Chan, and P. M. Pilarski, "Adaptive switching in practice: Improving myoelectric prosthesis performance through reinforcement learning," in *Proc. of the Myoelectric Controls Symposium (MEC'14)*, Fredericton, New Brunswick, August 18–22, 2014, pp. 69–73.

[9] P. M. Pilarski, M. R. Dawson, T. Degris, J. P. Carey, and R. S. Sutton, "Dynamic switching and real-time machine learning for improved human control of assistive biomedical robots," in *Proc. 4th IEEE RAS & EMBS Int. Conf. Biomedical Robotics and Biomechatronics (BioRob)*, Roma, Italy, 2012, pp. 296–302.

[10] D. M. Wolpert, Z. Ghahramani, and J. R. Flanagan, "Perspectives and problems in motor learning," *Trends Cogn. Sci.*, vol. 5, no. 11, pp. 487–494, 2001.

[11] A. D. Redish, *The Mind Within the Brain: How We Make Decisions and How those Decisions Go Wrong*. New York: Oxford University Press, 2013.

[12] D. J. Linden, "From molecules to memory in the cerebellum," *Science*, vol. 301, pp. 1682–1685, 2003.

[13] M. L. Littman, R. S. Sutton, and S. Singh, "Predictive representations of state," in *Advances in Neural Information Processing Systems 14*, pp. 1555–1561, MIT Press, 2002.

[14] R. S. Sutton, J. Modayil, M. Delp, T. Degris, P. M. Pilarski, A. White, and D. Precup, "Horde: a scalable real-time architecture for learning knowledge from unsupervised sensorimotor interaction," in *Proc. 10th Int. Conf. Autonomous Agents and Multiagent Systems (AAMAS)*, Taipei, Taiwan, 2011, pp. 761–768.

[15] J. Modayil, A. White, and R. S. Sutton, "Multi-timescale nexting in a reinforcement learning robot," *Adaptive Behavior*, vol. 22 no. 2, pp. 146–160, 2014.

[16] J. Modayil and R. S. Sutton, "Prediction driven behavior: Learning predictions that drive fixed responses," in *The AAAI-14 Workshop on Artificial Intelligence and Robotics*, Quebec City, Quebec, Canada, July 27, 2014.

[17] A. White, J. Modayil, and R. S. Sutton, "Surprise and curiosity for big data robotics," *AAAI-14 Workshop on Sequential Decision-Making with Big Data*, Quebec City, Quebec, Canada, July 28, 2014.

[18] H. Van Siejen and R. S. Sutton, "True online TD(λ)," in *Proc. of the 31st International Conference on Machine Learning*, vol. 32, Beijing, China, 2014, pp. 692–700.

[19] R. E. Schapire, "The strength of weak learnability," *Machine Learning*, vol. 5, iss. 2, pp. 197–227, June 1990.

[20] J. Kober, J. a. Bagnell, and J. Peters, "Reinforcement learning in robotics: A survey," *Int. J. Rob. Res.*, vol. 32, no. 11, pp. 1238–1274, 2013.