Neuro-Cognitive Science Inspired Directions in Learning for Control

Workshop on Cognition and Control
ACC 2021

Deepan Muthirayan and Pramod P. Khargonekar

Department of Electrical Engineering and Computer Science
University of California, Irvine

24 May 2021

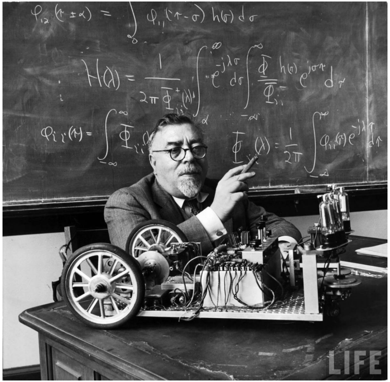
This presentation contains numerous hyperlinks (in blue) as pointers for further study and exploration.

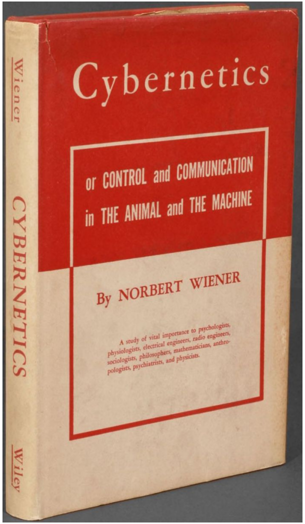
Outline

Dedicated to the Memory of Dr. Kishan Baheti

Thanks to the National Science Foundation for support through
Grant No. ECCS-1839429

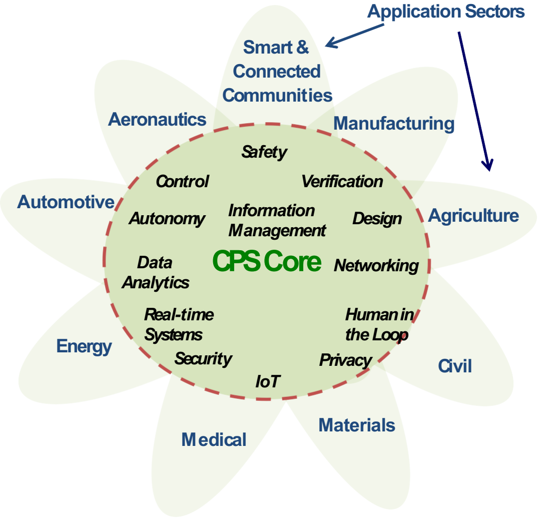
Wiener, Cybernetics, and Macy Conferences





How would the pioneers of cybernetics and AI envision the future of CPS?

Cyber-Physical Systems



Application Domains

Transportation

- Faster and safer vehicles (airplanes, cars, etc)
- Improved use of airspace and roadways
- Energy efficiency
- Manned and un-manned

Energy and Industrial Automation

- Homes and offices that are more energy efficient and cheaper to operate
- Distributed micro-generation for the grid

Healthcare and Biomedical

- Increased use of effective in-home care
- More capable devices for diagnosis
- New internal and external prosthetics

Critical Infrastructure

- More reliable power grid
- Highways that allow denser traffic with increased safety

Source: NSF

Aspirational and Emerging Applications: Examples

- Smart-X
  1. Smart manufacturing
  2. Smart grid
  3. Smart transportation
  4. Smart cities
  5. Smart health
- Autonomous systems
  1. Unmanned air vehicles
  2. Self-driving cars
  3. Autonomous robots

Human individual and group behavior are central in many of these applications:
Smart Cyber-Physical-Human Systems (CPHS).

Smart-X: Conceptual View

Sensors

Communications

Internet-of-things

Automation

Data analytics

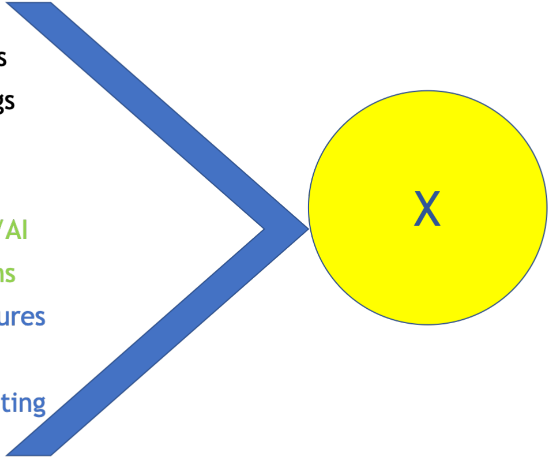
Machine learning/AI

Control/Decisions

Software Architectures

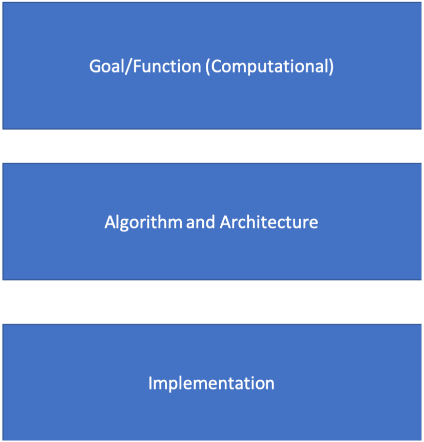
Cybersecurity

Cloud/Edge Computing



Cognitive Cyber-Physical Systems

Marr's 3 Levels of Analysis and Cognitive Science



Marr's 3 Levels of Analysis and Cognitive Science, Peebles and Cooper (2015)

Cognition - Definitions and Characteristics

- "All processes by which the sensory input is transformed, reduced, elaborated, stored, recovered, and used." — Neisser, Cognitive Psychology, 1967.
- Important role of in-built capacity in the brain from genetics and evolution, e. g., symmetry, intuitive physics.
- Key Cognitive Functions
  1. Perception
  2. Attention
  3. Memory
  4. Reasoning
  5. Problem solving
  6. Knowledge representation

Cognitive Psychology, Neisser (1967)

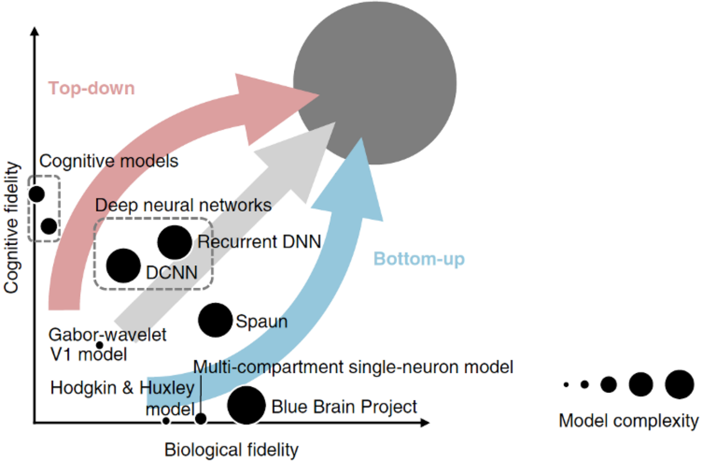
Mind as Machine: A History of Cognitive Science, Boden (2006)

Cognitive CPS - Key Principles

- Working Definition: CPS that have *cognitive functions and capabilities*.
- CPS can be explicitly designed and/or can learn to possess cognitive functions.
- Need for specific cognitive functions and capabilities will depend on the problem.
- Cognitive CPS's may learn from each other, from humans, and also form collaborative networks.
- Hypothesis: Cognitive CPS will be better able to augment humans and lead to human flourishing.

Cognitive CPS concept offers the most expansive and ambitious program for integrating ML/AI with CPHS for realizing Smart-X Systems.

Cognitive Models and Biological Fidelity



Cognitive Computational Neuroscience, Kriegeskorte and Douglas (2018)

Symbolic vs. Neural Connectionist Approaches

- Historical and ongoing debate on the nature of human cognition and the structure of the brain.
- Key topic in cognitive science: neuroscience, ML/AI, psychology, linguistics.
- Three major components:
  - Computational logic systems
  - Connectionist neural network models
  - Models and tools for uncertainty
- Pragmatic approach: combine connectionist, logic and probabilistic approaches to achieve desired system goals and objectives.

Neural-Symbolic Learning and Reasoning: A Survey and Interpretation, Besold et al. (2017)

Cognitive Models

- Production systems (Newell and Simon):
  1. If-then rules, logic, symbols
  2. Goals and subgoals, conflict resolution mechanisms
  3. Example: ACT-R, SOAR
- Reinforcement learning based models
  1. Actions, states, rewards
  2. Perception and motor modules
  3. Value and policy based approaches
  4. Three modes: Model-free, model-based, and episodic
  5. Brain combines all three of these modes but it is not known how this is done.
- Bayesian probabilistic models

Free Energy Principle

- A most ambitious principle for brain function due to K. Friston
- Brain seeks to minimize surprise
- Bayesian brain hypothesis: brain has an internal model that allows for computation of state estimate from sensory observations using Bayes rule
- Agent chooses action policy to maximize "information gain" (KL divergence or relative entropy)
- Free energy principle: minimize expected free energy under future observations and future states
- Connections to statistical mechanics, predictive coding, risk sensitive control, ...

The free-energy principle: a unified brain theory?, Friston (2010)

Perception in ML

- Deep learning is revolutionizing perception
- Compositionality is built-in
- Examples of very impressive progress in:
  - Computer vision
  - Speech recognition and processing
  - Language translation
- Architectures:
  - Convolutional neural networks
  - Long Short Term Memory (LSTM) recurrent neural networks

Perception in CPS

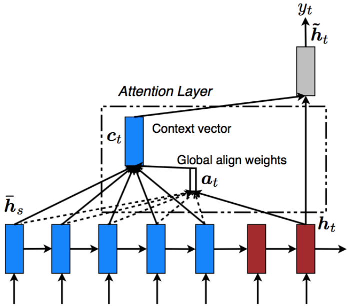
- CPS with multiple, distributed sources of sensed information
- Immediately possible to leverage DL advances
- Prior knowledge plays a very large role in cognitive theories of perception
- Neural network techniques could be combined with relational prior knowledge for improved context awareness in sensor rich CPS
- Potential tools and techniques for relational priors:
  1. Neural networks with symbolic front ends with priors to learn the symbolic front end
  2. Graph networks

Computational Models of Attention

- Vision (human, robot, driving) has been a major focus for modeling of attention
- Feature integration theory, guided search model, CODE theory of visual attention, signal detection theory, ...
- Computational models:
  1. Itti's model: color, intensity, orientation
  2. Bayesian models of attention
  3. Decision theoretic models
  4. Information theoretic models
  5. Graphical models
  6. Spectrum analysis models

Attention in ML

- Attention is the key to focusing on the most relevant information from multiple distributed sources of information
- Examples:
  - Recurrent Models of Visual Attention, Mnih et al. (2014)
  - Effective Approaches to Attention-based Neural Machine Translation, Luong et al. (2015)
  - Show, Attend and Tell: Neural Image Caption Generation with Visual Attention, Xu et al. (2015)
  - Self-attention Generative Adversarial Networks (GANs), Zhang et al (2019)



Attention based Machine Translator

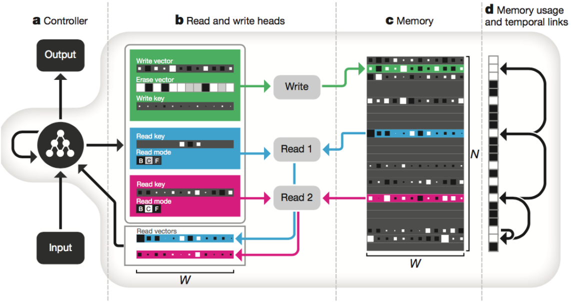
Possible Routes to Attention in CPS

- Two levels of attention:
  - First level - selection and focus on a particular task
  - Second level - top-down search for relevant information
- Attention for detecting changing conditions and contexts.
- Attention for fault detection and/or resilience.
- Attention models that are hierarchical and programmable will be required for CPS
- Examples of programmable attention:
  1. Self-attention models of deep learning
  2. Non-local neural networks for image recognition
  3. Attentive meta learners

Memory

- Memory is central to intelligent behavior.
- Multiple memory mechanisms in human cognition:
  - short-term
  - long-term
  - episodic (content-addressable)
  - semantic
- LSTM - excellent example of use of memory in machine learning
- Experience replay - a key innovation in Deep RL breakthroughs
- Differentiable neural computer by Graves et al. (2016)
- Sparse distributed representations. Examples: hierarchical temporal memory, sparsey

Differentiable Neural Computer



Hybrid computing using a neural network with dynamic external memory, Graves et al. (2016)

Memory, Attention, and Composition Cell Architecture

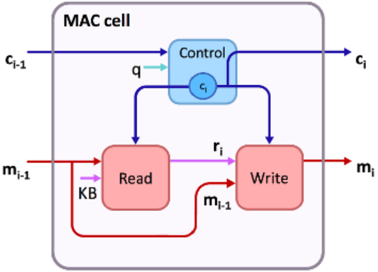


Figure 3: The MAC cell architecture. The MAC recurrent cell consists of a control unit, read unit, and write unit, that operate over dual ***control*** and ***memory*** hidden states. The **control unit** successively attends to different parts of the task description (question), updating the control state to represent at each timestep the reasoning operation the cell intends to perform. The **read unit** extracts information out of a knowledge base (here, image), guided by the control state. The **write unit** integrates the retrieved information into the memory state, yielding the new intermediate result that follows from applying the current reasoning operation.

Compositional Attention Networks for machine Reasoning, Hudson and Manning (2018)

Example of Memory in CPS: Episodic Control

- Episodic control - re-enact successful episodes from memory storage.
- Episodic control has potential relevance to "small data" learning and control.
- Example: Model-free episodic control, Blundell et al. (2016)
- Model-free episodic control – recorded experiences are used as value function estimators.
- Neural episodic control – combining deep learning model and lookup tables of action values.
- Hierarchical episodic control – episodes as options.

Selected Methodological Challenges

- There are numerous major challenges:
- Approaches for combining model-based and model-free techniques.
- Approaches to combine hierarchical and distributed architectures and algorithms.
- Reducing the need for large amounts of data: few-shot learning, one-shot learning
- Bringing meta learning paradigm for achieving autonomy: “learning to learn”.

Combining Model-based and Model-free Approaches

- Model free ML based approaches for sensing, perception, memory and model-based for planning, safety and closing the loop
- Model predictive control and reinforcement learning – compute action sequence based on the model via MPC (model based), update the model via reinforcement learning and supervised learning
- Guided policy search – robust local policies are derived from local models; local policies used to guide a global policy

Hierarchical Control

- Hierarchical structures appropriate and necessary for control and management of Smart-X
- Optimal behavioral hierarchy, Solway et al. (2014)
- Hierarchical control for sparse reward settings: meta controller sets the intermediate goal/sub-tasks and a lower level controller achieves the goal
  Example: Hierarchical DQN
- Hierarchical control provides scalable methods for large state-action spaces.
  Examples:
  - Options framework – temporally extended sequence of actions to simplify the learning process
  - Feudal RL – Higher level task is divided into a hierarchy of tasks
  - MAXQ framework: extension of the Q learning framework for the hierarchical setting

Our Recent Work

- External memory architectures and algorithms for adaptive control
- Regret guarantees for online learning for control
- Reinforcement learning for matching markets with applications to smart grids
- Meta learning
- Smart-X applications:
  - Anomaly detection in smart grids and manufacturing
  - Graph network techniques for decision making in autonomous vehicles

Meta Learning Paradigm

- Meta Learning as a paradigm for dealing with new environments by "learning to learn" approaches
- Learning from task properties, *transfer learning* from prior models, ...
- Meta learning principles and approaches could be leveraged for autonomy and control under uncertainty
- Central question: can the experience from learning in one setting to improve learning in another?
- Meta-learning is relevant in scenarios where the environment is different in each learning or control episode.
- Our Goals:
  - To provide a framework for meta-learning in a control setting
  - To provide a benchmark for finite episode meta-learning guarantees

Problem Setting

- N episodes of length T
- The environment draws an arbitrary $\theta = [A, B] \in \Theta$ in each episode
- System dynamics within each episode:

$$x_{t+1} = Ax_t + Bu_t,\ \ y_t = x_t + \epsilon_t,\ \ x_1 = x_s,\ \ \epsilon_t \text{ is noise.}$$

- Control input constraints: $u_t \in \mathcal{U},\ \forall\ t,\ \mathcal{U} = \{u | F_u u \leq b_u\}$ = a bounded polytope
- Control cost function: $c_t(x_t, u_t)$.
- Information:

Known	Observable
Θ, $\{c_s(.,.)\}_{s\geq 1}$ (limited preview of future cost)	$\{y_s\}_{s\leq t}, \{u_s\}_{s\leq t-1}$

Meta-Learning Architecture

superscript i: episode, subscript k: time index within an episode

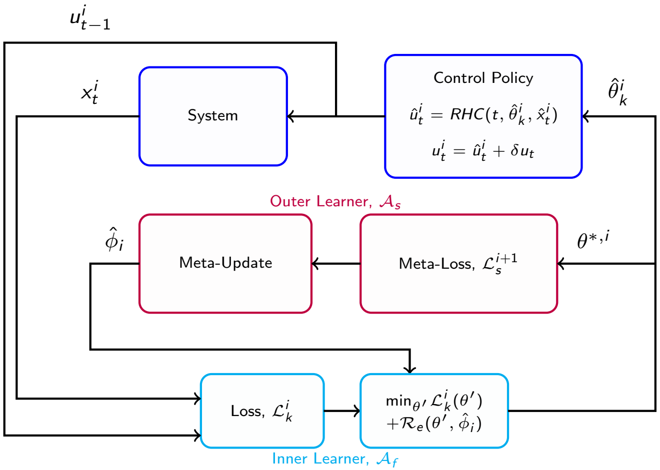
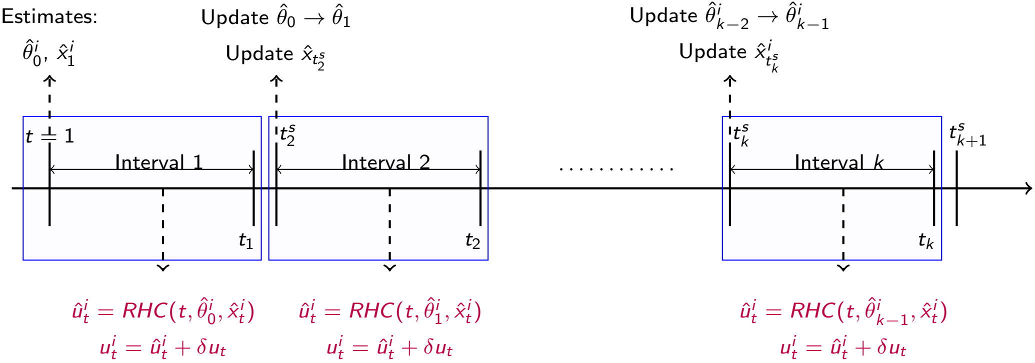


Figure: Online Model-based Meta-learning Control Architecture

Online Control Algorithm



Length of interval k: $t_k - t_k^s + 1 = 2^{k-1}H$.

Inner Learner $\mathcal{A}_s$

- Loss $\mathcal{L}_k^i$,

$$\mathcal{L}_k^i(\hat{\theta}) = \sum_{j=1}^{t_k} l_{\hat{\theta},j}^i, \;\; l_{\hat{\theta},j}^i = \left\| y_{j+1}^i - \hat{\theta}[(y_j^i)^\top, (u_j^i)^\top]^\top \right\|_2^2 .$$

- Least-squares estimate:

$$\hat{\theta}_{l,k}^i = \arg\min_{\hat{\theta}} \mathcal{L}_k^i(\hat{\theta}) + \mathcal{R}_e(\hat{\theta}, \hat{\phi}_i),$$

$$\mathcal{R}_e(\hat{\theta}, \hat{\phi}_i) = \lambda \left\| \hat{\theta} - \hat{\phi}_i \right\|_F^2 .$$

Control Policy: $RHC(t, \hat{\theta}_t^i, \hat{x}_t^i)$

- Input: Horizon M, $\{c_k\}_{t \leq k \leq t+M-1}$, Output: $\hat{u}_t^i$
- $RHC(t, \hat{\theta}_t^i, \hat{x}_t^i)$: (Optimizes the cost-to-go for the estimated dynamics)
  1. $U^* = \arg\min_U \sum_{k=0}^{M-1} c_{k+t}(\tilde{x}_k, w_k)$

     s.t. $\tilde{x}_{k+1} = \hat{A}\tilde{x}_k + \hat{B}w_k,\ \hat{\theta}_t^i = [\hat{A}, \hat{B}],\ w_k \in \mathcal{U},\ \tilde{x}_0 = \hat{x}$
  2. $\hat{u}_t^i = w_0^*$ (current RHC control input)

Perturbation

- The RHC approach requires persistence of excitation for parameter estimation. This requires perturbation along certain directions. One of the key contributions of this work: balancing exploration and exploitation in online RHC.

- Perturbation may violate control input constraints. The control is designed so that while balancing exploration and exploitation constraint violation is bounded.

Perturbation δu_t

- Perturbation by δu_t guarantees that

$$\sum_{j=1}^{t_k} \begin{bmatrix} x_j \\ u_j \end{bmatrix} \left[x_j^\top, u_j^\top\right] \geq O(\sqrt{t_k}) \text{ (persistent excitation)}$$

(The specific rate of growth balances exploration and exploitation!)

Outer Learner $\mathcal{A}_f$

- Outer learner update

$$\psi_{i+1} = \hat{\phi}_i - \eta_i \nabla l_i^o(\hat{\phi}_i),\ \eta_i = \frac{1}{\sqrt{i}},\ l_i^o(\hat{\phi}) = \left\| \hat{\theta}^{*,i} - \hat{\phi} \right\|_F,$$

$$\hat{\theta}^{*,i} = \text{best inner learner estimate in episode } i$$

$$\hat{\phi}_{i+1} = \text{Proj}_{\Theta}(\psi_{i+1})$$

Regret for Cost

- Regret

$$R_T^i = \left[\mathcal{C}^i(\mathcal{H}) - \mathcal{C}^{i,*}\right], \text{ where } \mathcal{C}(\mathcal{H}) = \sum_{j=1}^{T}[c_j(x_j^i, u_j^i)],$$

$\mathcal{C}^{i,*}$ cost with complete knowledge of system and state

- Average regret across N episodes:

$$\overline{R} = \frac{1}{N}\sum_{i=1}^{N} R_T^i$$

Performance for Constraint Violation

- Constraint violation in episode i,

$$\mathcal{V}^i = \sum_{t=1}^{T} \left(\sum_{s} \{F_u u_t^i - b_u\}_{s,+} \right), \; U_{1:T}^i = \{u_1^i, u_2^i, ..., u_T^i\}$$

where $\{.\}_l$ denotes the l-th component of a vector. The subscript $\{.\}_+$ is a shorthand notation for $\max\{., 0\}$

- Average constraint violation across N episodes:

$$\overline{\mathcal{V}} = \frac{1}{N} \sum_{i=1}^{N} \mathcal{V}^i(.)$$

Key Results

- Per episode regret and constraint violation: Under suitable technical conditions, for δ arbitrarily small, with probability greater than $1 - \mathcal{O}(\delta)$:

$$R_T \leq \tilde{\mathcal{O}}\left(T^{3/4}\right), \ \mathcal{V} \leq \tilde{\mathcal{O}}\left(T^{3/4}\right).$$

- Early result: Under the same technical conditions, $N \geq T$, for δ arbitrarily small, with probability greater than $1 - \mathcal{O}(\delta)$

$$\overline{R} \leq \tilde{\mathcal{O}}\left(\left(1 + \frac{1}{\sqrt{N}}\right) T^{3/4}\right), \ \overline{\mathcal{V}} \leq \tilde{\mathcal{O}}\left(\left(1 + \frac{1}{\sqrt{N}}\right) T^{3/4}\right).$$

Contributions

- Key contribution: Comparison with respect to receding horizon controller with complete knowledge of system and state. Prior online control works analyse regret w.r.t linear feedback controllers.
- Novel approach to balance exploration and exploitation in online RHC
- First finite time regret guarantee for online RHC
- First finite-time guarantee for meta-learning in a control setting

Publications for More Details I

1. D. Muthirayan and P. P. Khargonekar, "Working Memory Augmentation for Improved Learning in Neural Adaptive Control," IEEE Conference on Decision and Control, pp. 6785-6792, 2019.
2. D. Muthirayan, and P. P. Khargonekar, "Memory Augmented Neural Network Adaptive Controllers: Performance and Stability", arXiv preprint arXiv:1905.02832, 2019.
3. D. Muthirayan, and P. P. Khargonekar, "Memory Augmented Neural Network Adaptive Controller for Strict Feedback Nonlinear Systems," arXiv preprint arXiv:1906.05421, 2019.
4. D. Muthirayan, S. Nivison and P. P. Khargonekar, "Improved Attention Models for Memory Augmented Neural Network Adaptive Controllers," arXiv preprint arXiv:1910.01189, 2019, Proceedings of American Control Conference, pp. 639-646, 2020.
5. D. Muthirayan, J. Yuan and P. P. Khargonekar, "Regret Guarantees for Online Receding Horizon Learning Control," arXiv preprint arXiv:2010.07269, 2021.
6. D. Muthirayan, and P. P. Khargonekar, "Meta-Learning Guarantees for Online Receding Horizon Learning Control," arXiv preprint arXiv:2010.11327, 2021.
7. D. Muthirayan, J. Yuan, P. P. Khargonekar, "Adaptive Gradient Online Control", arXiv preprint arXiv:2103.08753, 2021.

Publications for More Details II

8. M. Majidi*, D. Muthirayan*, M. Parvania, P. P. Khargonekar, "Dynamic Matching Markets in Power Grid: Concepts and Solution using Deep Reinforcement Learning", arXiv preprint arXiv:2104.05654, 2021.
9. D. Muthirayan, M. Parvania, P. P. Khargonekar, "Online Algorithms for Dynamic Matching Markets in Power Distribution Systems", IEEE Control Systems Letters, pp. 995-1000, 2020.
10. A. Barua, D. Muthirayan, P. P. Khargonekar, M. A. Al. Faruque, "Hierarchical Temporal Memory based One-pass Learning for Real-Time Anomaly Detection and Simultaneous Data Prediction in Smart Grids", IEEE Transactions on Dependable and Secure Computing, 2020
11. A. V. Malawade, N. D. Costa, D. Muthirayan, P. P. Khargonekar, M. A. Al. Faruque, "Neuroscience-inspired algorithms for the predictive maintenance of manufacturing systems", IEEE Transactions on Industrial Informatics, 2021, early access.
12. S. Y. Yu, A. V. Malawade, D. Muthirayan, P. P. Khargonekar, M. A. Al. Faruque, "Scene-graph augmented data-driven risk assessment of autonomous vehicle decisions", IEEE Transactions on Intelligent Transportation Systems, 2021, early access.

Thank you!

email: pramod.khargonekar@uci.edu
website: https://faculty.sites.uci.edu/khargonekar/