Keynote Presentation:

Resilient Cyber-Physical-Human Systems and Role of Machine Learning

Prof. Azad M. Madni
Systems Architecting and Engineering

2018 Annual Research Review
Center for Systems and Software Engineering
University of Southern California

March 14, 2018
Outline

- Cyber-Physical-Human Systems
- Human Roles in CPH Systems
- Dealing with Disruptions
- Resilient CPH Systems
- Machine Learning Opportunities
- Exemplar Resilient CPH System
- Key Takeaways
- Q & A
Cyber-Physical-Human Systems

- A class of safety-critical applications in which interactions between the physical system and cyber elements that control its operation are influenced by human agent(s) and system objectives are achieved through purposeful interactions between the three

- A CPH system comprises:
  - Physical model of the system(s) to be controlled
  - Cyber elements (i.e., communication links and software)
  - Human agents who monitor CPS operation and intervene as needed

- Human intervention consists of:
  - providing information that CPS needs
  - taking over partial or complete control in contingency situations
Exemplar CPH Systems

- Self-Driving Vehicles
- Smart Buildings
- Smart Manufacturing
- Medical Devices
- Unmanned Aerial Vehicles
CPH Systems Are:

- Socio-technical systems
  - exist at multiple scales (CP elements, humans)
  - ability to continuously improve (learning)
  - capable of mutual adaptation (changing context)

- Complex engineered systems
  - built from computational algorithms, physical components, and human agents
  - performance depends on shared context and mutual predictability especially in the face of disruptions
  - difficult to assess their long-term impact (hidden interactions, change cascades)
Resilient CPH Systems: Dealing with Disruptions

- Monitor, anticipate and **circumvent** disruption and **learn** from it (fault/disruption avoidance)

- **Withstand** disruption within performance envelope without reconfiguration or reorganization (robustness)

- **Extend capacity** to cope with unavoidable disruption outside performance envelope and **learn** from it (resilience)

- **Rapidly** and **fully/partially recover** from unavoidable disruption outside performance envelope and **learn** from it (resilience)
Resilient CPH System Characteristics

- Learn human information seeking preference
  - function of context

- Infer human intent
  - from noisy signals, incomplete information and context

- Share perceptual tasks
  - context-driven localization vs. object recognition within scenes

- Share decision making tasks
  - creative option generation (human), objective structuring (machine)

- Jointly adapt to new circumstance
  - e.g., new context (disruption, contingency)

- Transfer control (pass “conn” between captain and OOD)
  - requires context awareness
  - requires understanding of human and CPS strengths and limitations
Exploiting Team Construct

- View CPS and Human within a Team construct
  - teamwork key to sustained high performance

- What Makes a High-Performance Team?
  - capitalize on each other’s strengths
  - circumvent each other’s limitations
  - maintain shared context and mutual predictability
  - jointly adapt to contingencies and respond to disruptions
  - rapidly restore cognitive coupling post-disruption

- Limits/Constraints
  - human adaptation limits – set upper bound on CPS adaptation rate
  - CPS technology limitations – e.g., legacy, interoperability, speed
Modeling the Human in Resilient CPH Systems

- Humans have:
  - kinematic constraints
  - cognitive limitations
  - decision making and control behaviors

- **Question:** what aspects of humans should be modeled (i.e., represented) for a specific CPH system?
  - only a subset typically needed

- **Question:** is there a methodological basis for determining appropriate sparse representation of a human for a particular class of CPH activities?
Machine Learning in Resilient CPH Systems

- Multiple sources of learning
  - sensors, networks, people

- Complicating factors
  - partial observability
  - noisy sensors
  - disruptive events

- Machine learning options
  - supervised learning
  - unsupervised learning
  - reinforcement learning
Conceptual Schema for High-Performance CPH System

- **Goals**
  - High-Performance Human-Machine Teams
  - Operational Context
    - Co-located/Distributed
    - Interchangeable Actors
    - Collaborative Work
    - Need for Mutual Adaptation

- **Characteristics**
  - Highly Trained
  - Shared Conceptual Models
  - Mutual Predictability
  - Mutual Learning

- **Collaborative/Social Human-Machine Processes**
  - Interactive
  - Interdependent Activities
  - Ongoing Adaptations

- **Simulation/Virtual World**
  - Instrumented
  - Researcher-Controlled

- **Metrics**
  - Neurophysiological Signals
  - Behaviors
    - Human Error Rate
    - Machine Error Rate
    - Cycle Time
    - Resource Utilization
    - Recovery Time

USC Viterbi
School of Engineering
Systems Architecture and Engineering

Copyright © 2012 – 2018 Azad M. Madni
Mutual Adaptation in High-Performance CPH System
Human Roles

- Observer (outside control loop)
- Execution Monitor w/over-ride privileges (inside control loop)
- Decision Maker/Approver (inside control loop)
- Associate: Shared task performer (inside control loop)
- Interactive Sensor/Actuator Tasker (inside control loop)
CPS Role: Monitor and Correlator

- Humans can’t control many physiological reactions (e.g. heart rate)
- What should/can CPS monitor to determine human state or question human when in doubt
  - e.g. can CPS know difference between an elevated heart rate due to stress versus excessive caffeine?
- Can correlating with other sensors help distinguish stress from caffeine?
- How should machine behavior/actions change based on such evaluations?
- Papers on use of HMM in adaptive control systems to:
  - monitor human behavior
  - change control functions based on predicted state
Shared Decision Making and Dynamic Allocation of Responsibility

- Humans good at:
  - understanding complex situations (contextualization)
  - making high-level decisions

- CPS good at:
  - implementation of high level decisions (local level)
  - embedded and remote sensing
  - fast computation and multi-source information aggregation

- Task performed by human and CPS can be re-allocated based on context (poor performance, inability to perform, request for help)

- Need a way to assess human-CPS interactions so that right (safe, correct, efficient) allocation of functions is achieved

- Boundaries of responsibilities can change (performance, context)
  - e.g., if human drowsy, CPS takes over
  - e.g., if CPS headed into trouble or signals handoff, human takes over
Types of Human and CPS Interactions

- Direct human control of CPS (primarily supervisory control)
  - human intervenes in control algorithms and adjusts set points
  - CPS carries out commands, reports results, awaits next command

- CPS passively monitors human; takes action if needed
  - open loop or closed loop

- Open loop example: sleep tracking device (sleep quality)
  - also monitors sound, light, temperature and motion sensors to record environmental conditions during sleep (i.e., context);
  - this information is presented to user on a tablet/smart phone to inform them of possible causes of disruption
  - human is in the loop, but does not directly control system
  - system does not take any proactive action to improve sleep quality because open loop system; if it did, it would be a closed loop system
Types of Human and CPS Interaction (cont’d)

- Closed loop example: smart thermostat (human-in-the-loop)
  - uses sensors to detect occupancy and sleep patterns in home
  - uses patterns to proactively turn off HVAC system to save energy

- Human monitors CPS which is in control; human decides to take back control (by taking back the “conn”)
  - key construct: passing the “conn” – a naval metaphor for control between the captain and Officer of the Deck (OOD)
  - Captain can retain the conn, or give it to the OOD
  - Captain can command through the OOD or monitor OOD’s decisions/actions
  - if captain does not like OOD’s commands to subsystems, captain can revoke “conn” and assume command; then can command through OOD
Use of Human Behavior Models

- State-of-the-art techniques to model human behavior
  - Range from general to specific
- For example, Smart Thermostat uses Hidden Markov Model to model occupancy and sleep patterns of residents to save energy
  - It captures human behavior at a very high level
- For example, impulsive injection of insulin uses math models
  - For diabetes mellitus
  - Very specific model determines need for insulin injection by monitoring glucose level relative to threshold level for administering insulin
Disruption Circumvention: Key Questions

- How does CPH system avoid disruption?
- What/who is responsible?
- What needs to be anticipated and circumvented to enable CPH to continue operating?
Failure Detection: Key Questions

- How does CPH system detect failure?
- What/who is responsible?
- What needs to be fixed to either enable CPH to continue operating, or to place the CPH into a safe mode?
- POMDP concepts can apply here with some modifications
  - need to go beyond autonomous actions
  - need hierarchical construct for dynamic allocation of control either to human or CPS depending on context (i.e., safety concerns, risk, availability)
Modularity and Re-configurability

- Creation of reusable building blocks that can be adapted to multiple CPH system applications
  - hardware, software, interfaces, sensors, actuators
- Primarily an engineering question
- However, open research questions when CPH systems involve networks, WiFi, etc.
Exemplar Resilient CPH System Problem: Security of Parked C-130

- **Context: Forward Base Operations**
- **C-130** parked on an unsecured landing strip adjacent to semi-urban environment with sparse roads
- **Parked C-130** offers adversaries ample attack opportunities
- Perimeter security is provided by video cameras and LWIR mounted on built-up structures in the vicinity and UGS that are augmented with additional units strewn around the aircraft by members of disembarking troops that stay behind to patrol area
- **Aircraft security commander** has a quick set-up laptop with:
  - wireless connection to sensors, and human/robotic sentries
  - real-time monitoring dashboard with facilities for anomaly detection, machine learning, selective region monitoring, and dynamic resource allocation
Security of Parked Aircraft
Illustrative Example: Human Roles and CPS Functions

- CPH system comprises
  - **physical**: laptop running smart dashboard software, sensors, actuators (robotic sentries); wireless connection to building mounted sensors and unattended ground sensors (UGS)
  - **cyber**: monitoring, planning, visualization, resource allocation, and machine learning software
  - **human**: commander in charge of maintaining aircraft security

- Human roles
  - Supervisor, sensor tasking (what region to surveil); robotic sentry tasking (what region to patrol and patrol type), human patrol tasking, intrusion monitoring, re-planning perimeter defense based on incoming intel

- CPS Functions:
  - learn commander preferences and priorities in various contexts (ML); learn normal traffic and intruder ingress patterns (ML), offer plans and patrol schemes; generate context-sensitive visualization; issue alerts upon intrusion detection; auto-reconfigure perimeter in response to changes to environment including intrusion detection (standing orders)
Sample UI and User System Interaction

- Monitoring system uses 2 monitors (UI for humans-in-the-loop)
- Monitors provide real-time state and status info in color-coded format
  - color-coded status of regions R1 through R8 is presented through the UI
  - green region means region is safe, red means an intruder has been detected
  - real-time views of the regions are providing data from security cameras
  - detailed information on each region is provided based on actions on users
- Monitor #1 provides overall status of individual regions, and real-time view based on updates from security cameras
- Monitor #2 changes based on actions of human user taken through UI
  - when a region’s display turns red, human clicks on that region’s icon to acquire details
  - monitor #2 provides detailed information on sensors, motion, location, camera, etc. - options: calling a security crew, turning on alarms, and reporting the incident
Exemplar UI for Monitoring Aircraft and Surrounding Area

User Interface Of The Monitoring System
UI for Area Monitoring, Issuing Commands/Alerts

Region 1
Region Sensors
7 14 15
Sensor Coordination
(X1,Y1)
(X2,Y2)
(X3,Y3)
Area Covering Camera
1
Camera Coordination
(X4,Y4)
Object Detected
Human Intruder

CALL SECURITY CREW
REPORT INCIDENT
TURN ON ALARMS
Layered Architecture of CPH System
Machine Learning Techniques for Anomaly Detection

**Supervised Learning:**
- Requires labeled data
- Assumes a priori knowledge of behavioral classes for both normal and abnormal activities (e.g. running, holding a gun)
- Examples:
  - Wireless and sensor network data: K-nearest neighbors (KNNs), support vector machines (SVMs), Regularized linear and quadratic discriminant analysis (LDA and QDA), and single classifiers, such as decision trees and Naïve Bayes networks
  - Video data: SVMs, HMMs, Gaussian Mixture Models (GMMs), and decision trees

**Unsupervised Learning:**
- Learns models/patterns of behaviors
- Creates clusters for data samples with special characteristics (e.g. special behavior patterns)
- Compares data samples with created models
- Examples:
  - Wireless network: graph-based outlier detection algorithms and clustering approaches, such as: K-means
  - Video data: distance-based likelihood ratio test-based clustering methods, dynamic Bayesian networks, and artificial neural networks

**Reinforcement Learning:**
- Learning behavior through trial-and-error interactions with dynamic environment
- Overall tendency: Increase long-run sum of values of reinforcement signals (e.g. collect maximum rewards with providing the correct action)
- Examples:
  - Autonomous vehicles, self-driving cars, and anomaly detection from sensory data: Partially Observable Markov Decision Processes (POMDPs)
Development Environment: Workflow

Simulation Platform (e.g. AnyLogic)

Software Developing Platform (e.g. Java)

Machine Learning Platform (e.g. Python)
Key Takeaways

- Cyber-Physical-Human Systems are safety-critical systems
- Humans can play a variety of roles in CPH systems – leads to increased complexity and vulnerability to cyber-attacks
- Resilient CPH systems are able to operate in uncertain, unpredictable environments prone to disruptions
- System design tools that individually address cyber, physical, and human elements are inadequate for CPH system design
  - need to focus on interactions and changes in interactions
  - requires holistic thinking, understanding of change propagation
- Example shows how Machine Learning can play an important role in resilient CPH system performance improvement
References

- Madni, A.M. and Sievers, M. Model Based Systems Engineering: Motivation, Current Status and Research Opportunities, in Systems Engineering, 2018
- Munir, S., Stankovic, J.A., Liang, C.M., and Lin, S. Cyber Physical System Challenges for Human-in-the-Loop Control
Thank You