v-CAT: A Cyberlearning Framework for Personalized Cognitive Skill Assessment and Training

Michalis Papakostas
University of Texas at Arlington
Computer Science
Arlington, TX, USA
michalis.papakostas@mavs.uta.edu

Konstantinos Tsiakas
University of Texas at Arlington
Computer Science
Arlington, TX, USA
konstantinos.tsiakas@mavs.uta.edu

Maher Abujelala
University of Texas at Arlington
Computer Science
Arlington, TX, USA
maher.abujelala@mavs.uta.edu

Morris Bell
Yale University
Pshyciatry
New Haven, CT, USA
morris.bell@yale.edu

Fillia Makedon
University of Texas at Arlington
Computer Science
Arlington, TX, USA
makedon@uta.edu

ABSTRACT

Recent research has shown that hundreds of millions of workers worldwide may lose their jobs to robots and automation by 2030, impacting over 40 developed and emerging countries and affecting more than 800 types of jobs. While automation promises to increase productivity and relieve workers from tedious or heavy-duty tasks, it can also widen the gap, leaving behind workers who lack automation training. In this project, we propose to build a technologically based, personalized vocational cyberlearning training system, where the user is assessed while immersed in a simulated workplace/factory task environment, and the system collecting and analyzing multisensory cognitive, behavioral and physiological data. Such a system, will produce recommendations to support targeted vocational training decision-making. The focus is on collecting and analyzing specific neurocognitive functions that include, working memory, attention, cognitive overload and cognitive flexibility. Collected data are analyzed to reveal, in iterative fashion, relationships between physiological and cognitive performance metrics, and how these relate to work-related behavioral patterns that require special vocational training.

CCS CONCEPTS

• **Human-Centered Computing**; • **Artificial Intelligence**; • **Knowledge Representation and Reasoning**;

KEYWORDS

User Modeling, Computational Cognitive Modeling, Cyberlearning, Vocational Assessment

PETRA '18, June 26–29, 2018, Corfu, Greece

ACM ISBN 978-1-4503-6390-7/18/06…$15.00
https://doi.org/10.1145/3197768.3201545

ACM Reference Format:
Michalis Papakostas, Konstantinos Tsiakas, Maher Abujelala, Morris Bell, and Fillia Makedon. 2018. v-CAT: A Cyberlearning Framework for Personalized Cognitive Skill Assessment and Training. In *PETRA '18: The 11th PErvasive Technologies Related to Assistive Environments Conference, June 26–29, 2018, Corfu, Greece.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3197768.3201545

1 INTRODUCTION

Workplace cognitive training for automation has many challenges, due to the complexity and variability of human nature. The aim of this project is to develop a prototype that can "unlock" and analyze complex cognitive human behavior as manifested in automated work environments. We propose a set of generalizable cognition-recognition tools, that can be applied across different domains and focus on cognitive assessment and training. In contrast to traditional vocational training done with manual paper and pencil recording, we propose to build an intelligent, data driven modular cyberlearning prototype system, called v-CAT (Vocational Cognitive Assessment and Training). While decreasing the expert's workload and maximizing learning efficiency via user personalization, v-CAT becomes progressively autonomous, offering reconfigurable training options. The goal of v-CAT is to design autonomous training systems that are tailored to the special needs of each individual user. For this reason, v-CAT's first phase is a cognitive assessment phase that provides immersive hands-on experiences to the user through a battery of context-aware tasks that are used to collect multisensory data.

This immersive task-engagement is done through a set of carefully designed workplace task simulations during which we monitor the user's actions and reactions, and collect and analyze a variety of data with different sensors. While the user executes an assigned task, we collect and analyze three types of data, behavioral, physiological and human input data, which are used to produce recommendations for personalized training to support targeted decision-making. To do the analysis of the task-derived data, we use state-of-the-art technologies in Human Computer Interaction (e.g., AR, VR, sensors) and, more importantly, Machine Learning (ML). In particular, we apply Deep Learning (DL), and Reinforcement Learning (RL) to identify, compare and prioritize specific training needs for working in highly technological environments. This includes being able to collaborate with intelligent machines and robots and to respond effectively, without errors and delays.

2 MOTIVATION & RELATED WORK

Unlike early fully automated robotic systems that were "caged" for safety, now humans need to work next to robots and learn how to alternate actions in accomplishing a certain task. This requires personalized training specifically designed to assess a user's individual weaknesses given their current skill level, which is a type of training beyond the capabilities of known Vocational Education and Training (VET) practices [11]. The importance of cognitive assessment and its close relation to learning and education outcomes has been thoroughly researched through the years [2]. The functioning assessment provides key information on the relationship between impairments and functional limitations and thus ascertains a claimant's work disability. Non-fatal brain injuries, such as Post-Traumatic Stress Disorder (PTSD) , can cause communication, executive functioning, and behavioral/emotional problems, which put these persons at risk of losing their job or inability to get a new one [12]. Other examples of people needing special vocational training are people with hearing, visual impairments, chronic illness, or sustained injury that leaves them unable to perform manual tasks [8].

In order for machines and humans to interact in an effective and intuitive manner in such environments, vocational training needs to be able to detect, monitor and predict changes in human affective states implicitly [17]. Recent advances in artificial intelligence have shown that extracting such affective indicators, related to working memory, cognitive load, inattention, fatigue, stress, frustration, etc. can be efficiently implemented using multisensory behavioral and physiological activity data. EEG signals are well known for their potential to provide meaningful indicators related to cognitive load, engagement as well as stress, attention levels and emotion in real time [14, 19]. In addition to EEG other sensing modalities, focusing on the user's behavioral characteristics, have been proven valuable in addressing similar problems. Specifically, body positioning, speech analysis, facial expressions and other general interaction patterns such as eye-gaze, keystroke dynamics and mouse tracking have been used to infer cognitive metrics measuring emotion, stress or cognitive [13]. It has been shown by various multi-disciplinary researches that integrating cognitive based information in the learning process can significantly increase the education outcomes [20]. Adaptive, user-training systems, is a topic that has gained a lot of attention in recent years [15]. Reinforcement learning and Interactive Machine Learning approaches are currently the state-of-the-art technologies for designing such adaptive training scenarios, mainly by focusing into a human-in-the-loop approach [7]. However, despite its proven significance, the research that relates cognitive evaluation to training is still at its infancy and limited [9]. This fact motivates our efforts on designing personalized and adjustable training scenarios that adapt their parameters according to subject's cognitive characteristics.

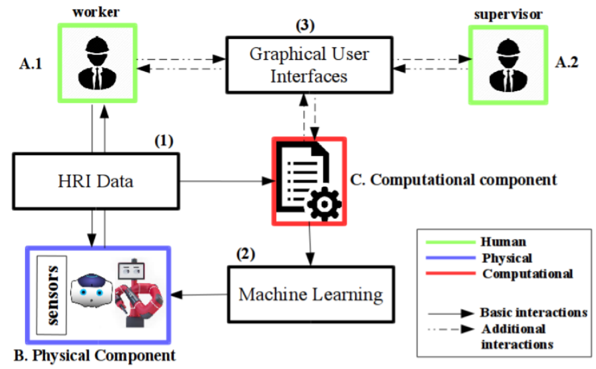

Figure 1: The proposed system architecture, following a CyberPhysical System (CPS) approach. The figure shows how the different components (human, physical, computational) interact with each other.

3 SYSTEM MODEL & ARCHITECTURE

To build v-CAT, we follow a human centric Cyber Physical Systems (CPS) approach. This Human-Centric Cyber-Physical System (HCPS) has three components, shown in Figure-1: (A) the human component (green squares), (B) the physical component (blue rectangle), and (C) the computational component (red rectangle). Figure-1 shows how the different components are connected and share information. We have two types of human components: users who interact and provide input to the system: (A1) the primary user (trainee) and (A2) the secondary user (trainer, supervisor). The physical component is the set of physical entities of this CPS system (robots, sensors, physical setups) with which the user interacts. The computational component (C) collects and analyzes this information and performs the appropriate system adjustments, following a traditional CPS-type closed loop architecture. The primary user (A1) interacts with the physical component B (e.g., robot), doing a work task interaction (cognitive task). During the assessment and training phases, behavioral and physiological data are used by the computational component (C) to update its parameters. The secondary user (A2) can monitor the interaction through the system's Graphical User Interface (GUI) (3) and intervene when needed to improve the system's decisions. This results in a progressively autonomous system for vocational task assessment and training that utilizes behavioral, physiological and expert guidance to provide personalized training. Thus, the proposed cyberlearning prototype integrates these components to investigate the importance of workplace cognitive factors through iterative user studies and evaluation. Moreover, v-CAT serves as a decision support system, to train experts on how to monitor and control automation and robot-based training sessions and enhance their decision-making.

4 VOCATIONAL COGNITIVE ASSESSMENT & TRAINING: SYSTEM PHASES AND COMPONENTS

v-CAT has the following phases and components for Cognitive Assessment and Training: assessment, analysis, recommendation, training, human supervision. The system starts with the user going

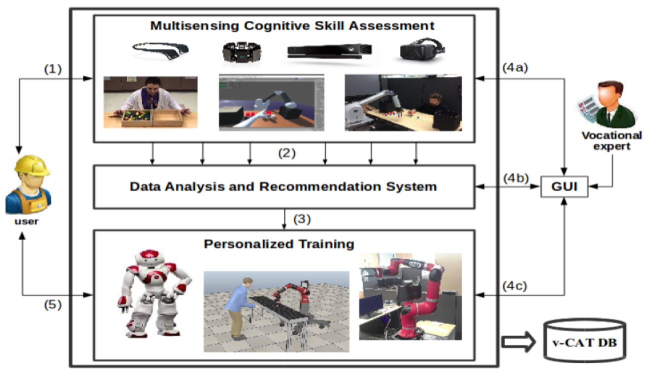

Figure 2: The different phases of the proposed v-CAT system. The system includes the following phases: assessment (top box) , recommendation (intermediate box), training (bottom-box) and human expert supervision (through the GUI component). Data from assessment and training sessions, as well as human intervention data are stored and used for personalization.

through the multisensing cognitive skill assessment phase (1) that has the user go through a set of work task simulations, with the aim to extract specific needs/metrics,related to attention, cognitive workload, working memory, and cognitive flexibility, using multiple sensors. Behavioral and physiological data are analyzed to represent and detect individual differences and weaknesses in cognitive functioning (see 2 in Figure-2). This analysis is then used to recommend a specific set of training tasks to provide personalized and adaptive training (see 3 in Figure-2). Human experts, trainers or supervisors have the ability to monitor and intervene at any phase of the system (see 4 in Figure-2), by using an intelligent Graphical User Interfaces. The user performs the proposed training tasks (see 5 in Figure-2) and information is stored in a database for system improvement and evaluation. Combining Human-in-the-Loop & Interactive Learning, both humans and systems can learn from each other and improve over time, in what is expected to become a lifelong learning setup. The modular closed-loop architecture has a built-in evaluation mechanism because it evaluates user performance during each system iteration. The proposed system's innovation integrates several important aspects of human activity together: using advanced machine learning methods to process multimodal human sensing, behavior monitoring, user profiling and personalization, and to evaluate user performance in a user-specific cyberlearning approach. In order to develop effective and appropriate assessment and training tasks, we follow the guidelines, requirements and metrics of the official NIH toolbox Cognition Battery [5].

5 COGNITIVE FACTORS

The user's task execution activity and other data are analyzed with respect to four critical vocational cognitive functionalities: *working memory, attention, cognitive flexibility and cognitive load.* Cognitive flexibility is important because a worker has to be able to shift smoothly from one task to another and to recognize that different situations call for different solutions. The opposite, rigidity, is especially common in impaired individuals, particularly when frustrated. In conjunction with these neurocognitive measures, it is important to note that there may be other related physiological measures (or factors), including the role that arousal plays in the workplace (possibly causing stress, distraction or even accidents). Our system can test additional factors besides the four mentioned above, such as *processing speed, cognitive reserve, response time, delays*, and other metrics. Based on the analysis of the data we derive from given tasks, recommendations for personalized training will be used to provide just-in-time assistance, (e.g., audio-visual cues, virtual reality (VR), augmented reality (AR) or robot support). The training phase follows the assessment phase. It uses carefully selected tasks to address any cognitive needs perceived during the assessment phase, such as improving attention. Once the training phase ends, the system goes back to the assessment phase with more assessment tasks selected to refine the assessment process. To model the system, we use a closed feedback loop architecture based on CyberPhysical Systems (CPS) methodology. The CPS approach has several advantages, including a data-driven self-learning methodology, iterative evaluation and monitoring longitudinal learning outcomes.

6 EXPERIMENTAL TESTBEDS

For evaluation purposes we have developed several different scenarios where user's have to perform various tasks in collaboration to a computer or robot based system. MAGNI, an upper-limb robotic system for training and assessment, uses virtual reality (VR) tasks and computer vision to evaluate a user's arm control and smoothness and adjust the exercise difficulty level accordingly [10]. The Box and Blocks (BaB) task is an occupational therapy assessment system that evaluates upper limb mobility, concentration, vision, and working memory. The virtual version of BaB has sensors that allow us to evaluate how well a worker can move his hand and joints [6]. MyoLearn is a physical and behavioral data collection and evaluation method which can assess worker safety and performance, as well as extract meaningful human-centric information during assemble tasks [4]. Similarly to MyoLearn the assembly training task proposed in [3] is a task where users need to follow instructions given by robot and assembly specific lego-parts as shown in a display. Other tasks include cognitive based assessment tasks such as the sequence learning [16] and the Towers of Hanoi tasks [1] where user's problem solving abilities and working memory are tested and evaluated. For all the aforementioned testbeds, performance related metrics are closely monitored. Our goal is to associate specific performance patterns with different behaviors observed through multimodal user monitoring and adapt system's parameters on the fly in order to increase the outcomes of both training and assessment. In Figure-3 we show some of the experimental testbeds described above.

7 INTERACTIVE PERSONALIZATION FOR ROBOT-ASSISTED TRAINING

As our team showed in [17] RL-based adaptive multimodal interactive systems can adjust their behavior, to match a current user's preferences and needs, in order to maximize the efficiency of the

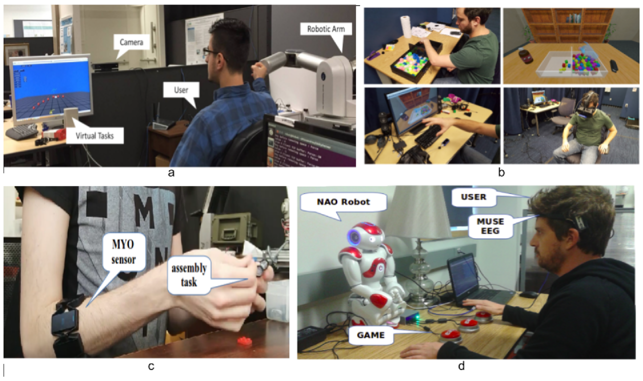

Figure 3: a) The MAGNI, an upper-limb robotic system for training and assessment, b) physical and VR versions of the Box and Blocks game , c) the MyoLearn assembly training task d) The Sequence Learning cognitive task

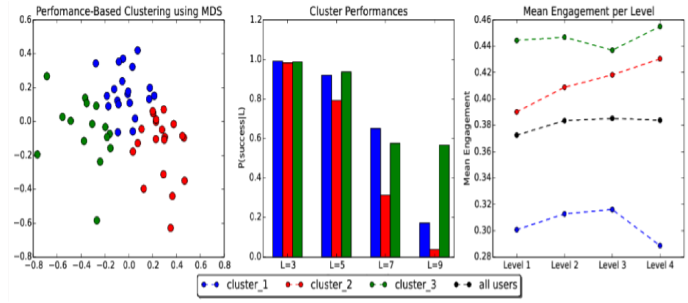

Figure 4: Visualizing task performance and engagement patterns during the sequence learning task. Such patterns can be used to inform personalized training strategies.

interaction. Based on the Interactive Learning and Adaptation (ILA) framework proposed in [17] we plan to evaluate the v-CAT cyberlearning framework with regard to interactive personalization by using Interactive Reinforcement Learning (IRL) techniques. In particular, our efforts focus towards developing a real-time IRL system that utilizes prior knowledge from previous users, human-generated data, informative user feedback and expert interventions as guidance for real time personalization. Human feedback can be provided through multisensory data information implicitly, to indicate level of attention, task engagement, attention, emotion etc., or explicitly, in the form of expert guidance or intervention. Such interaction data are collected and used to learn personalized training policies for different user skills. Our preliminary results [14, 18] highlight the potentials of this framework on the Sequence-Learning task. More specifically as shown in (Figure-4-a) users can be clustered into different groups based on their engagement level while playing the SL task. Indication metrics related to user's engagement and attention can be extracted directly and in real-time from the EEG signals of the user. It is proven also cross-validated by our experiments that engagement and attention levels highly correlate to the final performance of the user when performing any kind of cognitive task. Figure-4-b shows the different probabilities of performance for each user-group when we trained the IRL framework based on user's engagement measurements. Engagement was extracted from raw EEG in real time, for different difficulty levels of the SL task. Figure-4-c. Shows the how engagement varies across different levels and different clusters of users. The three graphs indicate that EEG signals can describe general user behaviors, despite the fact that engagement is an exclusively subjective measure.

8 CONCLUSIONS

In this paper we propose v-CAT, a vocational tool for Cognitive Assessment and Training. v-CAT takes as input multimodal sensory data and creates complex user models describing user's physio-cognitive state. To do so, our system utilizes advanced ML and DL techniques to analyze the sensory data and takes advantage of IRL approaches to achieve personalization and tailor the system parameters to the special needs of each individual. Our initial results, along with multidisciplinary research done under the cyberlearning domain highlight the potentials of our framework and show the great correlation between user performance and specific cognitive aspects such as working memory, attention, cognitive flexibility and cognitive load.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under the Grants CHS 1565328 and PFI 1719031

REFERENCES

[1] Ashwin Ramesh Babu, Akilesh Rajavenkatanarayanan, Maher Abujelala, and Fillia Makedon. 2017. Votre: A vocational training and evaluation system to compare training approaches for the workplace. In *International Conference on Virtual, Augmented and Mixed Reality*. Springer, 203–214.

[2] Morris D Bell, Wayne Zito, Tamasine Greig, and Bruce E Wexler. 2008. Neurocognitive enhancement therapy with vocational services: work outcomes at two-year follow-up. *Schizophrenia research* 105, 1 (2008), 18–29.

[3] Christopher Collander, Joseph Tompkins, Alexandros Lioulemes, Michail Theofanidis, Ali Sharifara, and Fillia Makedon. 2017. An Interactive Robot-based Vocational Assessment Game using Lego Assembly. In *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments*. ACM, 346–353.

[4] Dylan Ebert and Fillia Makedon. 2017. MyoLearn: Using a multimodal armband sensor for vocational safety problem identification. In *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments*. ACM, 248–249.

[5] Richard C Gershon, Molly V Wagster, Hugh C Hendrie, Nathan A Fox, Karon F Cook, and Cindy J Nowinski. 2013. NIH toolbox for assessment of neurological and behavioral function. *Neurology* 80, 11 Supplement 3 (2013), S2–S6.

[6] Shawn N Gieser, Caleb Gentry, James LePage, and Fillia Makedon. 2016. Comparing objective and subjective metrics between physical and virtual tasks. In *International Conference on Virtual, Augmented and Mixed Reality*. Springer, 3–13.

[7] Andreas Holzinger. 2016. Interactive machine learning for health informatics: when do we need the human-in-the-loop? *Brain Informatics* 3, 2 (2016), 119–131.

[8] Azfar Khalid, Pierre Kirisci, Zied Ghrairi, Jürgen Pannek, and Klaus-Dieter Thoben. 2017. Safety Requirements in Collaborative Human–Robot Cyber-Physical System. In *Dynamics in Logistics*. Springer, 41–51.

[9] Séverin Lemaignan, Mathieu Warnier, E Akin Sisbot, Aurélie Clodic, and Rachid Alami. 2017. Artificial cognition for social human–robot interaction: An implementation. *Artificial Intelligence* 247 (2017), 45–69.

[10] Alexandros Lioulemes, Michail Theofanidis, Varun Kanal, Konstantinos Tsiakas, Maher Abujelala, Chris Collander, William B Townsend, Angie Boisselle, and Fillia Makedon. 2017. MAGNI Dynamics: A Vision-Based Kinematic and Dynamic Upper-Limb Model for Intelligent Robotic Rehabilitation. *World Academy of Science, Engineering and Technology, International Journal of Medical, Health, Biomedical, Bioengineering and Pharmaceutical Engineering* 11, 4 (2017), 158–167.

[11] AR Masalimova and AS Chibakov. 2016. Experimental Analytical Model of Conditions and Quality Control of Vocational Training of Workers and Specialists.

IEJME-Mathematics Education 11, 6 (2016), 1796.
[12] Susan R McGurk, Kim T Mueser, Haiyi Xie, Karin Feldman, Yaniv Shaya, Leslie Klein, and Rosemarie Wolfe. 2016. Cognitive remediation for vocational rehabilitation nonresponders. *Schizophrenia research* 175, 1 (2016), 48–56.
[13] Michalis Papakostas, Evaggelos Spyrou, Theodoros Giannakopoulos, Giorgos Siantikos, Dimitrios Sgouropoulos, Phivos Mylonas, and Fillia Makedon. 2017. Deep Visual Attributes vs. Hand-Crafted Audio Features on Multidomain Speech Emotion Recognition. *Computation* 5, 2 (2017), 26.
[14] Michalis Papakostas, Konstantinos Tsiakas, Theodoros Giannakopoulos, and Fillia Makedon. 2017. Towards predicting task performance from EEG signals. *2017 IEEE International Conference on Big Data (Big Data)* (2017), 4423–4425.
[15] John ER Staddon. 2016. *Adaptive behavior and learning.* Cambridge University Press.
[16] Konstantinos Tsiakas, Cheryl Abellanoza, Maher Abujelala, Michalis Papakostas, Tasnim Makada, and Fillia Makedon. [n. d.]. Towards Designing a Socially Assistive Robot for Adaptive and Personalized Cognitive Training. ([n. d.]).
[17] Konstantinos Tsiakas, Maher Abujelala, Alexandros Lioulemes, and Fillia Makedon. 2016. An intelligent interactive learning and adaptation framework for robot-based vocational training. In *Computational Intelligence (SSCI), 2016 IEEE Symposium Series on.* IEEE, 1–6.
[18] Konstantinos Tsiakas, Maria Dagioglou, Vangelis Karkaletsis, and Fillia Makedon. 2016. Adaptive robot assisted therapy using interactive reinforcement learning. In *International Conference on Social Robotics.* Springer, 11–21.
[19] Xiao-Wei Wang, Dan Nie, and Bao-Liang Lu. 2014. Emotional state classification from EEG data using machine learning approach. *Neurocomputing* 129 (2014), 94–106.
[20] Yun Zhou, Tao Xu, Yanping Cai, Xiaojun Wu, and Bei Dong. 2017. Monitoring Cognitive Workload in Online Videos Learning Through an EEG-Based Brain-Computer Interface. In *International Conference on Learning and Collaboration Technologies.* Springer, 64–73.