Maintenance Training with Digital Twins and Structured Machine Learning

J.T. Folsom-Kovarik, Robert A. Sottilare
Soar Technology, Inc.
Orlando, FL
jeremiah@soartech.com, bob.sottilare@soartech.com

Ray S. Perez
Office of Naval Research
Arlington, VA
ray.perez@navy.mil

ABSTRACT

The US military requires technical personnel capable of diagnosing and resolving operational system issues as part of their maintenance duties. Schoolhouse training needs to deliver both familiarity with fundamental principles and readiness to maintain specific assets. The usefulness of the training after deployment is challenged by skill decay, systems updates, and idiosyncratic systems that behave differently from a general model because of usage or wear.

A digital twin provides a simulation that models one individual asset, rather than a general model of an idealized asset. The increasing use of digital twins to model military and industrial hardware provides a wealth of data with the potential to create a high-fidelity experience to accelerate learning, minimize skill decay, improve transfer of skills from the schoolhouse to the operational platform, and support on-the-job training once deployed.

The paper addresses two of the challenges in using digital twins for training. First, a *building block* approach interprets recorded data in terms of underlying principles to help author focused training and help learners generalize training. Second, building blocks help implement interactive training and training scenarios. The structure of the building blocks lets a digital twin training system focus on key parts of a process, accept interactive learner input, accurately predict asset response, and assess learner performance for feedback to learners or instructors.

The approach to implement digital twin training uses structured machine learning to predict the behavior of a single asset, in this case a commercial jet engine. The building blocks add structure to guide the machine learning and accurately predict how an individual jet engine will respond to new conditions that were never fed into machine learning. As a result, the digital twin can present both realistic and “what-if” training scenarios with automated instructional feedback.

ABOUT THE AUTHORS

Dr. Jeremiah T. Folsom-Kovarik is the Senior Scientist for Artificial Intelligence Exploration research within the Intelligent Systems business area at Soar Technology, Inc (SoarTech). His research combines modern data science and machine learning approaches with SoarTech’s 20+ years of experience in modeling expert knowledge and human reasoning. He earned a Ph.D. in computer science from the University of Central Florida (UCF).

Dr. Robert Sottilare is the Director of Learning Sciences at Soar Technology, Inc. He came to SoarTech in 2018 after completing a 35-year federal career in both Army and Navy training science and technology organizations. His doctorate is in Modeling & Simulation with a focus in Intelligent Systems from UCF. At the US Army Research Laboratory, he led the adaptive training science and technology program where the focus of his research was automated authoring, instructional management, and analysis tools and methods for intelligent systems and standards. He is the father of the Generalized Intelligent Framework for Tutoring (GIFT), an award-winning open source, AI-based adaptive instructional architecture. Dr. Sottilare’s current research includes the application of AI to enhance human performance, adapt/tailor instruction, automate processes, mine data, and model digital twins.

Dr. Ray S. Perez is a program officer at the Office of Naval Research, where he manages the Cognitive Science of Learning Program. His previous experience includes working as bench Scientist at the US Army Research Institute for the Behavioral Science, Dr. Perez’s research interest lies at the intersection of Cognitive Science, Computer Science, and Cognitive Neural Science. His research interests include; Individual Differences, Training Technologies, and Neural Biology of Learning. Currently, his research focus is leveraging emerging technologies like immersive environments (Augmented and Virtual Reality) to build adaptive training and next-generation measurement systems. He received a Ph.D. in Cognitive Psychology from the University of California Los Angeles.

Maintenance Training with Digital Twins and Structured Machine Learning

J.T. Folsom-Kovarik, Robert A. Sottilare
Soar Technology, Inc.
Orlando, FL
jeremiah@soartech.com, bob.sottilare@soartech.com

Ray S. Perez
Office of Naval Research
Arlington, VA
ray.perez@navy.mil

INTRODUCTION

The military requires qualified technical personnel who can detect, diagnose and solve system problems as part of their maintenance duties. Technical training takes place in settings such as a classroom for learning fundamentals, a physical or digital simulation for learning applied and hands-on skills, and finally with actual equipment. A high-fidelity simulation aims to prepare personnel to use and maintain actual equipment. However, actual equipment can present something of a moving target—which can lead to gaps in training simulations.

Gaps can occur between simulation-based training and actual equipment for several reasons. Individual assets may have idiosyncratic usage, repair, and upgrade histories that change their performance. Complex interactions with the environment may be streamlined to train the most common cases. Depending on the acquisition pipeline, new operational equipment may be introduced quickly or in multiple configurations that simply outpace old training models. As a result of these gaps between training simulations and actual equipment, maintenance personnel often require significant retraining to become familiar with specific actual equipment after they deploy.

The fielding of new equipment is particularly challenging because the opportunities to gain experience or learn about it prior to deployment are limited. The new equipment may respond to usage, maintenance, and environment in ways that appear subtly difficult to predict or unexpected when more reliance is placed on designs, manuals, and training materials than on familiarity with actual operation. Errors can cause dramatic departures from required performance in expensive new equipment where relatively few personnel are ready to respond. The gravity of consequences requires maintenance personnel to be both proficient and efficient before they can independently perform duties with actual equipment.

One approach to improve the skills of qualified technical personnel is through an apprenticeship model where qualified personnel mentor the next generation of maintainers both prior to their accessing the physical system and then continuing their coaching while deployed with the physical system. The apprenticeship approach has proven to be a highly effective method for military training domains, but it requires large time investments to deliver one-to-one tutoring (Sottilare & Proctor, 2012).

To further enhance each maintainer's exposure to expert support, accelerate their learning, minimize gaps between simulation and the operational equipment, and support on-the-job training once deployed, digital twin technology can help create accurate, interactive simulations of idiosyncratic individual assets and new equipment. The digital twin will form the basis of an adaptive, interactive instructional tool that will augment the current apprenticeship model from the schoolhouse to the initial deployment and beyond.

Digital Twins

A digital twin is a virtual model of a system or other physical entity where data are transmitted seamlessly between the physical entity and virtual model. The entity and model then simultaneously co-exist as identical entities through real-time updates (open system) or batch updates (closed system). Sensors and controls on the real asset, and optionally the training software, can participate in the internet of things (IoT). The resulting twin or virtual system model may then be accurately visualized on any device with Internet access or, with cached data, in an offline mode anytime and anywhere (Datta, 2016). Accurately reflecting one real asset is its primary advantage over wholly simulated systems.

Three uses of digital twins are product design, planning for manufacturing and production, and analysis of system performance. Two of these, digital twins for design and planning, are typically focused on prediction and they require high-fidelity physical simulations often combined with machine learning to interpolate and extrapolate data. However, these types of digital twins are unlikely to be available after design and during the operation of real assets. Two reasons are that the design and planning models represent a costly, proprietary investment to collect experimental data under a range of conditions, and they model performance with highly detailed physics engines running on high-performance computing clusters. As a result, companies that create these types of digital twins are unlikely to share them if it would reduce the business advantage of their proprietary models. Also, these types of models require large computing resources and will not run in deployed settings on commodity laptops or mobile devices.

There is a third type of digital twin which is a better fit for training simulations. These digital twins support ongoing analysis of system performance by capturing, analyzing, and acting on sensor data from individual operational assets. Typical non-training uses of operational asset digital twins include monitoring asset performance or enabling predictive maintenance. For example, operational digital twins might model a thousand individual aircraft engines in a commercial fleet, or one factory with a thousand components, to identify departures from performance and possible reasons in the environment (Dawes et al., 2019). Some digital twins that analyze operational assets can rely less on proprietary models or large computational resources. They reflect the actuality of an operating asset as-built, not the idealized behavior as-designed. The as-built, operational models are therefore well aligned with simulation training.

Digital Twin Training Desiderata

Digital twins offer potential to provide excellent maintenance training that is better aligned than ever before with real systems. Learning science shows that when training is well aligned with the final performance, there is increased opportunity for training to influence performance through transfer of training (Holding, 1965). The kinds of alignment that impact training influence both *near and far transfer* (Perkins & Salomon, 1992). Near transfer mainly improves performance in contexts that are similar to training, and far transfer improves performance in new contexts.

For near transfer, aligning sensory cues between training and performance (Hull, 1943) can call for accurate visualization of the asset or the contextual elements like the work tasks and work environment. Digital twins enable learners to connect the training to the real asset and can make a specific asset seem familiar to a learner even before encountering it. To meet near transfer goals, the digital twin training should reflect real asset performance and response to learner inputs in control presentation, appearance, sound, timing, and so on (e.g., Hontvedt & Øvergård, 2020).

For far transfer, learners must encode a general and reusable concept representation during training (Clark & Voogel, 1985). If successful, they can recall and apply the general understanding to address new situations, such as troubleshooting an unexpected fault. Digital twins enable learners to explore underlying principles in several ways: by directly presenting a principle in simplified visualizations, by presenting different models or types of devices that work on the same general principles, and by letting trainees explore the workings of an asset under many conditions, with appropriate guidance to construct an understanding of the underlying principles (e.g., Evans & Johri 2008). Importantly for maintenance training needs, learners can experiment on an asset and try many maintenance actions in many conditions, including extrapolated “what-if” scenarios, without costly downtime or endangering a real asset.

Effective training incorporates assessment and feedback to keep learners on track and let instructors monitor progress. Digital twins have the potential to automate data-driven feedback that reflects learner performance and learning progress. Properly connected to digital twin data streams, an adaptive instructional system (AIS) can guide training to provide feedback to each learner. As a result, digital twin maintenance training offers the potential to support many forms of learning from anytime, anywhere self-directed interactive training (SDIT) or instructor-led training (ILT) prior to deployment to performance support (PS) or structured on-the-job training (OJT) while deployed in the field.

Digital twins that capture data from a real asset provide a strong starting point for fidelity to a specific asset. Technical challenges remain in transforming raw digital-twin data into training that lets learners generalize their knowledge and become ready for more than rote replication of past recorded situations. First, the digital twin data need to be filtered and selected to focus on subsets and scenarios that have training impact. Second, the recorded playback of the digital twin must be made interactive so that training can respond realistically to learner inputs, even ones not previously recorded. The experimental approach described in the next section offers tools to filter digital twin data via building blocks and to extrapolate realistic training in conditions that have not been recorded via structured machine learning.

Example Digital Twin for Training

An aircraft gas turbine engine is a suitable asset for demonstrating principles of digital twin training. Data from dozens of sensors embedded throughout the engine are constantly updated during a flight and can be monitored using an operational digital twin. For maintenance purposes, a likely use case might be accumulating engine sensor data during each sortie and downloading the data in a batch when the aircraft is on the ground. In this setting, maintenance personnel may use the digital twin for non-training purposes to include identifying possible performance issues, diagnosing faults, estimating metrics like mean time to failure and remaining useful life, and predicting future maintenance and logistics needs to support the equipment.

A dataset containing sensor data for individual gas turbine engines was developed and released to the public by NASA (Saxena et al., 2008). The dataset contains 61,249 rows of data representing 249 individual engines of the same make, model and type. Data was generated with the validated, high-fidelity Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) (Liu et al., 2012). Data in the selected dataset focuses on 24 separate sensors relevant to engine maintenance and failure prediction. These include temperature readings at five engine locations, pressure readings at four locations, and rotational speed readings at six points in the engine. There are also three sensors for operating conditions such as altitude and Mach speed. Based on this dataset, a digital twin could collect information that lets maintainers find individual differences between the engines that develop during their usage.

Importantly, the NASA dataset included one type of data that is usually never available in an operational digital twin. All of the individual engines were run to failure in this dataset. The engine failures were caused by a fault either in the fan or in the high-pressure compressor (HPC), and each fault was allowed to continue without maintenance to show the range of effects that can be seen in the engine sensors. Under normal operation, an aircraft engine would be repaired or replaced before its performance degraded and became dangerous. Therefore, an operational digital twin would usually not contain data representing conditions of extreme wear, only data that reflects normal operation. The extra data provided ground truth in a test of machine learning to accurately predict how training should respond to learners.

Data from an individual aircraft engine was used to implement an example of digital twin training design and the design steps that could be automated with structured machine learning. The data was used to implement an interactive training simulation to accurately show learners the observable effects of performance degradation and multiple duty cycles without maintenance (Figure 1). The training could be viewed on a range of devices, from a low-cost tablet to a touch-interactive virtual reality experience. The training captured learner inputs and provided accurate real-time engine responses. The responses were both interactive and up to date with new recorded data from a digital twin.

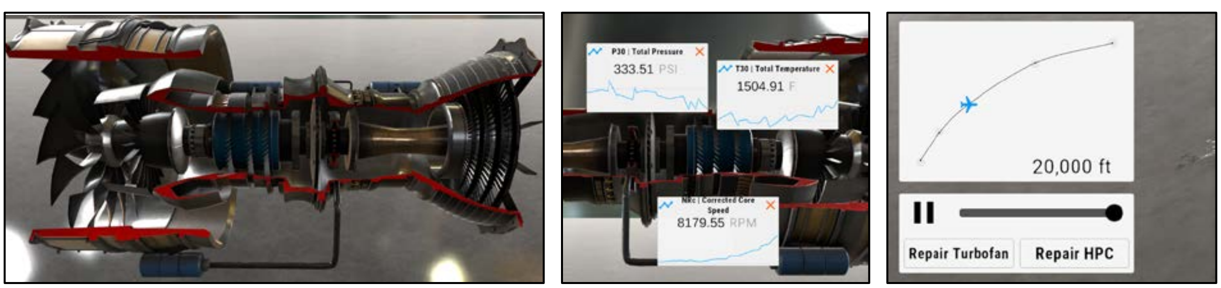

Figure 1: Interactive digital twin training shows learners accurate interaction responses of a specific asset.

Digital twin data was transformed into training as follows. First, the available recorded data was manually inspected using knowledge of the fault type that would occur during training and knowledge of how that fault type impacts the digital twin sensors. In the future, this step could be replaced by the instructional design and structured machine learning contributions discussed in the following sections. Second, the sensors needed to identify the fault and to rule out other faults were embedded in a digital twin visualization. The visualization used a commercially available 3D model of a jet engine and was rigged and animated to display sensor readings and trends when selected by the learner. Third, controls were added to let learners change the operating conditions (altitude, speed, and duty cycles elapsed) and observe the effects on the selected sensors. The ability of the training to display accurate sensor readings for operating conditions that had never been observed was the result of the various machine learning components described below. Finally, the training was instrumented to monitor the learner's interactions in a way that could be assessed in a real training setting. For example, the choice of sensors to check, the choice of which component is faulty, and the time elapsed could all help to assess the learner's exploration, familiarity, or decision making.

TRAINING DESIGN

Digital twins contain a potentially large amount of data, and the data can be captured at different levels ranging from a single sensor reading, a component with several sensors, or an entire system. Therefore, digital twin training must be easy to control in terms of focusing on a few sensors out of many, highlighting a particular underlying principle, or setting the conditions for a particular scenario. These are tasks for instructional designers and, to a lesser extent, instructors, as they identify which scenarios to offer learners and monitor how learners perform in the selected training.

The discrete event simulation (Misra, 1986) provides a structure to identify what data is important for training. A discrete event simulation represents the physical states of systems, sub-systems, and components for end users without the need for computer programming. Instead, components are related to each other with causal rules which describe how a change anywhere in the system updates the other components. Within a discrete event simulation, reusable building blocks can represent various mechanical, electrical, chemical, and thermodynamic processes. The configuration of various reusable components can be made internally accurate and detailed with system performance data extracted from the digital twin, while remaining abstracted for end users.

The specific approach to instructional authoring and instructional delivery was based on the adaptive instructional system GIFT. GIFT, the Generalized Intelligent Framework for Tutoring, is a powerful, deployed software system that can interface with training simulations and other learner-facing software to monitor, assess, and respond to learner performance (Sottilare et al., 2012; Sottilare et al., 2017). A training design typical for GIFT focuses on assessing learner competency and adapting instruction with instrumented training tools such as interactive simulations.

Authoring Training with Building Blocks

Reusable building blocks can enable non-programmer training authors such as subject matter experts (SMEs) and instructors to: 1) create objects and define their attributes, 2) build compound objects to create systems and subsystems, 3) create scenarios, and 4) create learning objectives, measures of success, and assessment mechanisms.

An operational system is composed of working machines that are made up of physical objects (subsystems and objects), processes and sensors. In their simplest form, the systems' physical objects come in three general types: 1) *structural components* which are the elemental building blocks of machines (e.g., frames, members and other rigid bodies, bearings, axles, splines, fasteners, seals and springs), 2) *mechanisms* that control movement (e.g., motors, generators, engines, gears, wheels, belts, chain drives, linkages, cams, brakes, turbines, pumps and clutches) and 3) *controls* which regulate system and subsystem behaviors (e.g., buttons, switches, actuators, throttles, displays and computer controllers). The example training replicates systems and subsystem level behaviors using compound physical objects (e.g., jet engine turbofan or jet engine HPC).

The system's physical objects incorporate mechanical, electrical, fluid and or thermal processes with measures that change over time. Sensors are devices that detect or measure a physical property of a system, subsystem or object in order to describe mechanical, electrical, fluid, or thermal processes. A digital twin differs from an idealized simulation model of a system in that it accounts for the physical properties and limitations (e.g., failures over time, work hardening of metals) of a specific operational system throughout its lifecycle. Digital twin training requires the ability 1) to model the physical processes from sensor readings in the operational system over time, and 2) to represent the physical properties of the system (its controls, physical objects, and performance model) with respect to time. Maintenance training also requires an ability 3) to inject faults or create scenarios where maintenance or repair is needed.

Each reusable building block to author training encodes inputs, outputs, and a modeled process. The example jet engine maintenance training included an implemented building block, *remaining useful life (RUL)*, to show how building blocks impact authoring and accurate machine learning. RUL represents an interpretive layer on top of the digital twin data. The building block has the input of elapsed time (duty cycles) and the output of an exponential curve degrading component performance. The model within each building block like RUL consists of *structure* provided by expert knowledge, such as the fact that degradation is an exponential function on a time input, and *parameters* learned from sensor data, such as the shape of a degradation curve for a specific component. The RUL building block lets authors create training where RUL affects performance, extending the digital twin data to new settings not yet recorded. The effects on sensors of change over time and context like speed control inputs and operational conditions (altitude) are also represented in building block inputs the SME or instructor can manipulate. Machine learning reconstructs system processes from sensor data, which is why accurate machine learning is needed.

Reusable building blocks can model any physical process in an operational system, and machine learning can update these models as follows. Operational data is used to train multiple models which may be linked with a system, subsystem or part. The models are validated and deployed as part of authoring training. The models are monitored and reinforced to ensure accuracy over time. As operational system data continues to flow into the digital twin, the model is retrained and updated to reflect the most current available performance data of the operational system it represents.

The input-process-output structure in building blocks provides encapsulation and composability that support the feasibility of authoring training. Building on discrete event models and on past work in computational scenario sequences and assessment (e.g., Folsom-Kovarik et al., 2015), authoring training in interactive scenarios is expected to include selecting and sequencing a timeline of events that provide information about the context and conditions that exist before, during, and after each event. Events can be represented as changes in model variables and may be categorized as performance monitoring, routine maintenance or troubleshooting (failure analysis). An example event in a maintenance scenario is a system, subsystem or component failure which results in abnormal performance.

SMEs and instructors can author digital twin training with reusable building blocks that specify what is important in training. An SME will be able to reconfigure or copy and modify these objects in a graphical dashboard without computer programming. Examples of configuration in a dashboard could include associating an object with sensor data from the digital twin, selecting a level of detail, defining scenario starting conditions and events, and linking building blocks to interactive controls. SMEs and instructors can already design and control information of this type with graphical interfaces that configure automated assessment and feedback from an AIS such as GIFT (Davis, Riley, & Goldberg, 2018) or in the Cognitive Tutor Authoring Tools (CTAT) (Aleven et al., 2015). Therefore, integration with GIFT as an instance of the end-user dashboard is expected to enable digital twin training to identify building block changes with learning objectives, measures, assessments and feedback. SMEs and instructors who wish to create maintenance training experiences could use GIFT dashboards to author and configure training, track learner progress, and see feedback in real time or in an after-action report at the end of the lesson.

In summary, digital twins alone record data about sensors, controls, and environmental contexts as a valuable source of accurate playback and frequent updates. However, reusable building blocks can enhance recorded data. Building blocks provide structure to the data that creates training content with relevant data selection, interactive processes that respond to learner inputs, scenarios and events, and learning objectives for assessment and feedback.

Digital Twin Training Instructional Design

This section discusses learning science recommendations for building blocks representing scenario events (e.g., degraded performance and failures) in order to support adaptive instruction of maintenance tasks. Adaptive instruction with a digital twin can let maintenance training capture the idiosyncrasies of a specific system and aid transfer of training knowledge and skills to operational contexts.

Instructional design for digital twin training, as with all effective training, includes three fundamental elements:

- Experience – presentation and immersion in a scenario that applies the trainee's knowledge and skill to the accomplishment of a task to be learned
- Measures – information used to assess the learner's level of success in accomplishing the task to be learned
- Feedback – information provided with the goal of improving future performance of the task to be learned

The instructional experience considers what is to be presented to the learner, how the experience relates to the instructional goals and learning objectives, the pace and difficulty of the content presented, and maintaining the relationship between the learner's domain competency and the difficulty level of the material and concepts presented (Vygotsky, 1978). First, digital twin training can present excellent immersive experiences by leveraging the fidelity of the digital twin data to reflect the real asset. Training should be presented in context to build on immersion, including mission situation or briefings. Second, the relevance of the experience, its ability to exercise the knowledge and skill of the trainee with respect to real job tasks, can be maximized when digital twin training lets learners interact with a digital twin model that responds realistically to their inputs. Finally, linking pace and difficulty of training to the needs of each learner relies on collecting accurate measures during training. Meeting learner needs is the key source of efficiency and effectiveness demonstrated by adaptive instruction.

Measures in the instructional design sense must be thoughtfully centered on what learners are expected to know before training, what they are expected to learn, and how learning will be demonstrated in an observable manner (Grafinger, 1988). Digital twin training is likely to be well suited to knowledge measures and cognitive skill measures, such as troubleshooting, sensemaking, and applying knowledge to a maintenance task. If digital twin training is paired with a realistic 3D model reflecting real controls and interfaces, it will also be able to support perceptual-cognitive skills such as recognition of states (Klein, 1993). It is also very interesting to speculate about whether digital twins can help create a level of familiarity that enables training higher-level perceptual-cognitive skills such as anticipation or intuition (Patterson & Eggleston, 2017).

Overall, digital twin measures are likely to focus on the cognitive aspects of maintenance training. Cognitive measures usually include response correctness and delay measured at each step (Newell & Simon, 1972), as well as behavioral evidence of supporting process such as perseverance or exploration for higher-order cognitive skills (e.g., Folsom-Kovarik, Boyce, & Thompson, 2018). Enabling cognitive assessment by measuring learner behavior is a key requirement for building blocks, which can enhance raw digital twin data with interpretation such as acceptable performance, tolerances, and response times. Digital twin training is not likely to replicate physical actions in a maintenance process. Instead, digital twin training can cognitively prepare learners for effective psychomotor training later in a physical simulator, a lab setting, or on actual machines.

Feedback design should include both real-time feedback and delayed feedback at the end of a training interaction. This feedback can include direct messages to a learner or instructor as well as tailoring the training to match the learner's improving knowledge and skill. An advantage of digital twin training is providing practice that aligns with newly acquired knowledge to increase training transfer. The main constraint on effective feedback is efficiency in terms of automated measure assessment and automated response to assessments. Feedback may be efficiently automated if the building blocks representing components and scenario events incorporate triggers that control direct display of feedback or progression of scenario events. These triggers should be configurable in the end-user dashboard.

Digital twin training can provide an immersive experience, measures that assess learner interaction, and built-in controllable feedback to learners and instructors. These advantages build on the groundwork of existing success with adaptive instruction using GIFT with interactive simulations. The ability of the digital twin to provide training beyond merely recorded data is enabled by authoring with building blocks. The next section demonstrates how machine learning can help create those building blocks without increasing workload on subject-matter experts and instructors.

STRUCTURED MACHINE LEARNING

The feasibility of simulation building blocks as a basis for digital twin training lies in machine learning to construct the building blocks and keep them up to date using recorded data. Machine learning has the potential to reduce authoring effort by automating the construction of the input, process, and output that defines each building block.

Within the example digital twin training for jet engine maintenance, a building block was created to reflect remaining useful life (RUL). Using the building block, the example training could give learners accurate responses for combinations of controls and conditions that were not recorded in the raw digital twin data. This capability can be carried out with various interpolation algorithms, depending on the level of fidelity needed, from simple linear interpolation to splines and derivatives. Whereas interpolation fills in gaps between known datapoints, another approach is needed to accurately model an asset under conditions that have never been recorded at all.

Machine learning was used to extrapolate asset performance under degraded conditions. In a real operational digital twin, it is possible that no recording has ever been made of extreme usage patterns or component degradation. Usually such conditions are avoided because they threaten the real asset. However, maintenance personnel need to train with "what-if" scenarios that prepare them to recognize and respond to extreme conditions or degraded performance caused by otherwise dangerous component faults. Extrapolation from normal performance data can let digital twin training accurately reflect degraded performance even though it has not been recorded.

An experiment was carried out to measure the accuracy of machine learning in predicting degraded performance for a specific, individual jet engine. A stock, general-purpose machine learning algorithm was compared against the same algorithm enhanced with the building block structure. The experiment showed that the structured machine learning delivered training that more accurately reflected how the individual jet engine performed as its HPC failed.

Method

A target task was designed to test machine learning for digital twin training. A single jet engine was selected from the source dataset and its eventual failure (due to HPC degradation) was identified. The machine learning was then trained with only a subset of the source data from the early part of the data, when performance was least degraded. After training, the machine learning was tested using the latter part of the data, when performance was most degraded. Better machine learning would be able to more accurately predict degraded performance based on normal performance.

First, a general-purpose machine learning algorithm using a neural network was applied to learn the underlying physical principles of the raw engine performance data. Neural networks have been shown to learn physics principles in past research (Raissi, Perdikaris, & Karniadakis, 2019). The topology of the general-purpose neural network was one that empirically could reconstruct the desired response curve. Furthermore, the same topology was used for both the general-purpose machine learning and the structured machine learning. The neural network had one input per sensor in the digital twin, one hidden layer of ten rectified linear units (RELUs), and one output corresponding to the compressor core rotational speed, which was the digital twin sensor most affected by the HPC degradation. If machine learning was successful, then the neural network would show the core rotational speed changing in an accurate, valid simulation of the sensor under different commands (speed) and conditions (altitude).

Second, general-purpose machine learning was compared to structured machine learning. The same neural network was trained and tested using the RUL building block as a precursor input. The RUL building block had two effects on the raw input data. First, it contained domain knowledge specifying which inputs and outputs are related to RUL in the digital twin. Second, it contained a bias to produce a positive exponential curve because of the knowledge that performance degrades exponentially. The rate of change or curve parameters were not set before machine learning.

Using domain knowledge to guide machine learning is a technique which can increase the accuracy and decrease the training data requirements of general-purpose machine learning (Haley et al., 2018; Wray et al., 2019). Structured machine learning has the potential to combine the power of modern deep learning and neural network approaches with the expert knowledge captured in a structure that influences and directs the machine learning. As an added benefit for digital twin training, the knowledge structure can align with the building blocks that authors use to construct scenarios. The main challenges are identifying what knowledge should influence machine learning and how the knowledge structure can be implemented to redirect or influence a machine learning algorithm.

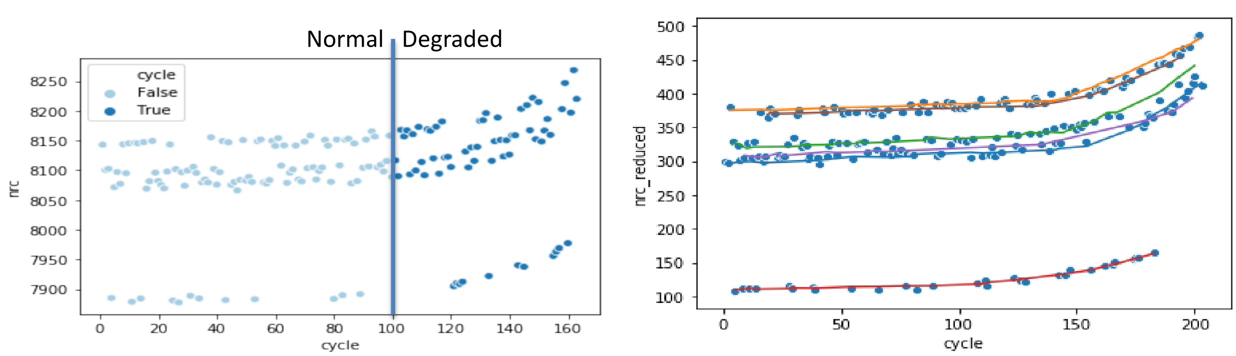

Figure 2: Machine learning can predict asset performance from digital twin data without expert programming.

Figure 2 reflects the training and testing data for both machine learning algorithms on the left. The x-axis represents time, measured in duty cycles. The boundary between normal performance and degraded performance shows the subtle difference that maintenance personnel must be able to detect and remediate before the jet engine fails entirely.

The right-hand side of Figure 2 demonstrates a good fit to data that *both* algorithms can achieve when trained on the full data. The different colored lines represent different control inputs the learner can select in the simulator. General-purpose machine learning finds this model quickly and does not require expertise to use, meaning that it is *theoretically* possible for instructional designers who are not subject-matter experts to load the digital twin data and use it for training. However, the general-purpose algorithm only works well when all data is available. The good fit to the full data is equivalent to an interpolation problem. It does not represent extrapolation to model a new failure condition.

Experimental Results

Both general-purpose machine learning and structured machine learning were used to predict degraded performance. The accuracy of the predictions was compared to ground truth as available in the full data. Typical results are shown in Figure 3. When given less data (light curves), only structured machine learning was able to extrapolate the shape and magnitude of the curve that learners need to see during training. General-purpose machine learning was inaccurate in all cases except the darkest line, because it required the full training data to make an accurate model.

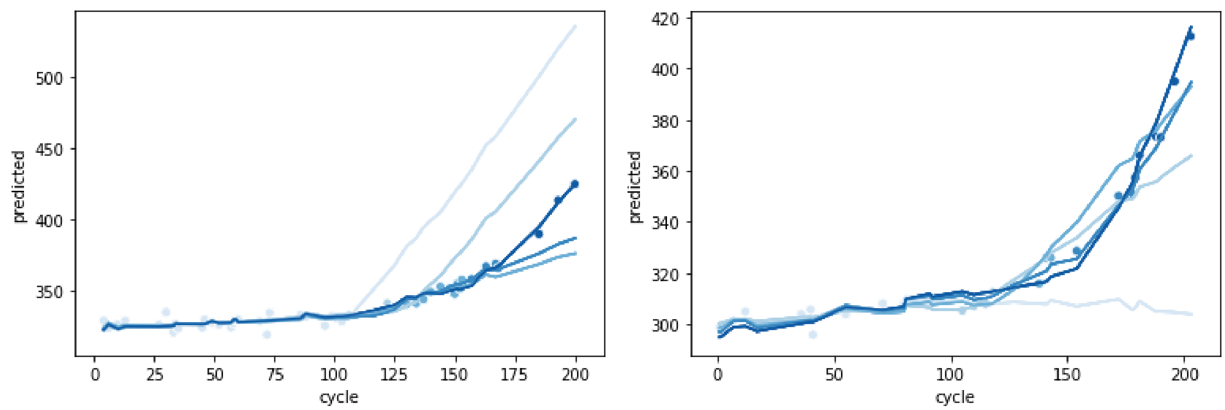

Figure 3: General machine learning (left) extrapolates less accurately than structured machine learning (right).

Finally, the machine-learning experiment was repeated 100 times to account for variability in initializing the machine learning. Figure 4 compares general-purpose machine learning (blue bars) to structured machine learning (orange bars).

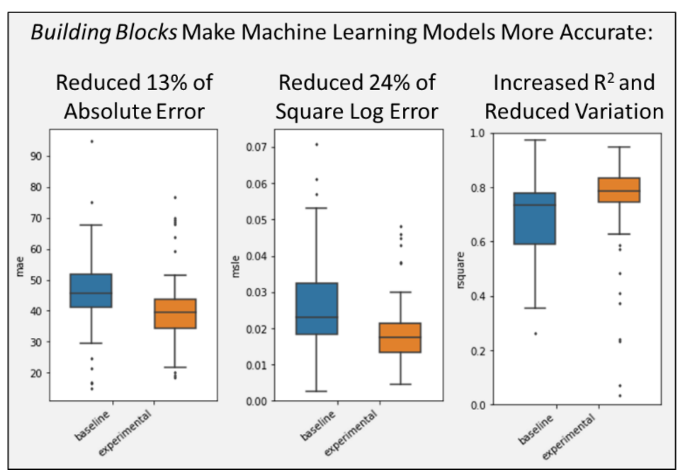

Figure 4: Structured machine learning enables more accurate interactive training with more consistency and less variation.

Structured machine learning extrapolated digital twin behavior more accurately according to several measures. Compared to the general-purpose baseline, adding structure reduced the median value of mean absolute error across 100 trials from 45.7 to 39.5, an improvement of 13%. The mean square log error, a measure appropriate for exponential functions, was reduced from 0.0231 to 0.0175, a 24% improvement. Adding structure also increased model fit and ability to explain the variation in the ground truth data (R^2) from 0.732 to 0.785. Finally, the structure reduced variation compared to general-purpose machine learning as measured by interquartile range. Variation describes the spread of outcomes achieved by machine learning across the different runs, and lower variation suggests the structured machine learning produced more consistent predictions.

The experimental results quantify the accuracy improvements that are also visible in the qualitative differences between the prediction graphs. As a result of increased accuracy, structured machine learning enables interactive digital twin training to respond realistically to learner interactions.

CONCLUSIONS

In order to provide high-fidelity training in multiple modes and devices, digital twins can form a basis for training simulations that reflect idiosyncratic assets and stay up to date with actual performance characteristics of a real asset.

Digital twins are a promising new source of training, so recommendations for digital twin training design were presented. Digital twin training can be made interactive, with accurate responses to learner inputs, using building blocks that encode underlying principles and extend the digital twin data beyond simple playback of recordings. The building blocks can support effective training with an immersive experience, measures and assessments, and automated feedback in an adaptive instructional system.

The benefits of building blocks can be obtained through machine learning that extends the digital twin data by interpolating missing data or extrapolating to new conditions that have not been recorded. General-purpose machine learning is easy to use and can accurately interpolate accurate responses in cases where the digital twin data does not exactly match learner inputs but the recorded conditions are similar in most respects.

The most promising machine learning makes use of the structure within building blocks to enable extrapolating from digital twin data to entirely new, “what-if” scenarios that let learners safely explore the workings of an asset under untested conditions. Structured machine learning uses the expert knowledge in building blocks to describe devices, components, and physical laws and make the extrapolation more accurate than general-purpose machine learning.

With additional work to continue implementing and evaluating new building blocks, a reusable library of digital twin training can be created to help maintain a valuable asset throughout its lifecycle.

ACKNOWLEDGEMENT

This material is partially based upon work supported by the Office of Naval Research (ONR) under contract N68335-19-C-0843. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Department of Defense or the Office of Naval Research. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

REFERENCES

Aleven, V., Sewall, J., Popescu, O., van Velsen, M., Demi, S., & Leber, B. (2015). Reflecting on twelve years of ITS authoring tools research with CTAT. *Design recommendations for adaptive intelligent tutoring systems, 3*, 263-283.

Clark, R. E., & Voogel, A. (1985). Transfer of Training Principles for Instructional Design. *Educational Communication and Technology*, 113-123.

Datta, S.P.A. (2016). Emergence of digital twins. *arXiv preprint arXiv:1610.06467*.

Davis, F.C., Riley, J.M., & Goldberg, B.S. (2018). *Iterative development of the GIFT wrap authoring tool.* In the Proceedings of the Sixth Annual GIFT Users Symposium.

Dawes, W.N., Meah, N., Kudryavtsev, A., Evans, R., Hunt, M., & Tiller, P. (2019). *Digital Geometry to Support a Gas Turbine Digital Twin.* Paper presented at the AIAA Scitech 2019 Forum.

Evans, M.A., & Johri, A. (2008). Facilitating guided participation through mobile technologies: designing creative learning environments for self and others. *Journal of Computing in Higher Education, 20*(2), 92-105.

Folsom-Kovarik, J.T., Boyce, M.W., & Thomson, R.H. (2018). *Perceptual-cognitive Training Improves Cross-cultural Communication in a Cadet Population*. Presented at the 6th Annual GIFT Symposium, Orlando, FL.

Folsom-Kovarik, J.T., Woods, A., Jones, R.M., & Wray, R.E. (2015). *Narrative representation for training cross-cultural interaction: Lessons learned and recommendations.* Paper presented at the 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015), Las Vegas, NV.

Grafinger, D.J. (1988). Basics of instructional systems development. *Infoline: Tips, tools, and intelligence for trainers, 8803*.

Haley, J., Hung, V., Bridgman, R., Timpko, N., & Wray, R. (2018). *Low Level Entity State Sequence Mapping to High Level Behavior via a Deep LSTM Model.* Paper presented at the 20th International Conference on Artificial Intelligence, Las Vegas.

Holding, D. H. (2013). *Principles of training: the commonwealth and international library: psychology division*. Elsevier.

Hull, C.L. (1943). *Principles of Behavior*. New York: Appleton-Century-Crofts, Inc.

Hontvedt, M., & Øvergård, K.I. (2020). Simulations at Work—a Framework for Configuring Simulation Fidelity with Training Objectives. *Computer Supported Cooperative Work (CSCW), 29*(1), 85-113.

Klein, G. A. (1993). A recognition-primed decision (RPD) model of rapid decision making. *Decision making in action: Models and methods*, *5*(4), 138-147.

Liu, Y., Frederick, D.K., DeCastro, J.A., Litt, J.S., & Chan, W.W. (2012). *User's Guide for the Commercial Modular Aero-Propulsion System Simulation (C-MAPSS): Version 2* (NASA/TM—2012-217432). Cleveland, OH: Glenn Research Center.

Misra, J. (1986). Distributed discrete-event simulation. *ACM Computing Surveys (CSUR), 18*(1), 39-65.

Newell, A., & Simon, H. A. (1972). *Human problem solving* (Vol. 104, No. 9). Englewood Cliffs, NJ: Prentice-Hall.

Patterson, R.E., & Eggleston, R.G. (2017). Intuitive Cognition. *Journal of Cognitive Engineering and Decision Making, 11*(1), 5-22.

Perkins, D.N., & Salomon, G. (1992). Transfer of learning. *International encyclopedia of education*. Oxford: Pergamon Press.

Saxena, A., Goebel, K., Simon, D., & Eklund, N. (2008). *Damage propagation modeling for aircraft engine run-to-failure simulation.* Paper presented at the 2008 International Conference on Prognostics and Health Management.

Sottilare, R.A., Brawner, K.W., Goldberg, B.S., & Holden, H.K. (2012). *The generalized intelligent framework for tutoring (GIFT).* Orlando, FL: US Army Research Laboratory Human Research & Engineering Directorate.

Sottilare, R.A., Brawner, K.W., Sinatra, A.M., & Johnston, J.H. (2017). *An updated concept for a Generalized Intelligent Framework for Tutoring (GIFT)*. Orlando, FL: US Army Research Laboratory.

Sottilare, R.A., & Proctor, M. (2012). Passively classifying student mood and performance within intelligent tutors. *Journal of Educational Technology & Society, 15*(2), 101-114.

Vygotsky, L.S. (1978). *Mind and society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wray, R., Haley, J., Bridgman, R., & Brehob, A. (2019). *Comparison of Complex Behavior via Event Sequences.* The 2019 International Conference on Computational Science and Computational Intelligence, Las Vegas.