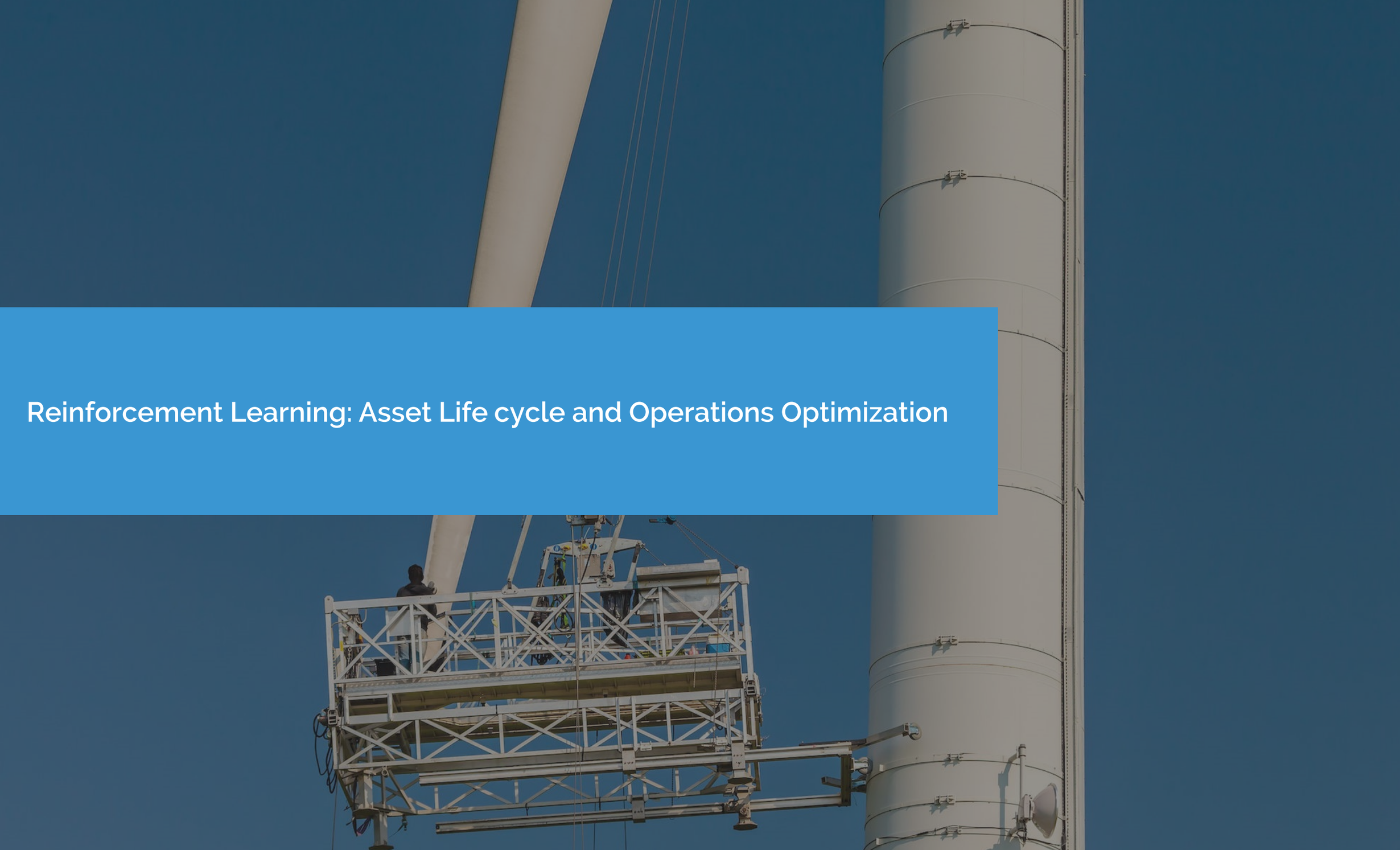


Reinforcement Learning: Asset Life cycle and Operations Optimization

Table of Contents

Introduction....2

Superhuman Performance with Machine Learning....3

Reinforcement Learning Tackles Real World Challenges....3

Five Ways Reinforcement Learning Optimizes Your Business....4

Wind Farm Operations and Maintenance....7

Conclusion....9

Introduction

Every day, your organization handles large, complex, asset-related decisions. Maybe you're delivering millions of packages, or building highway bridges, or keeping the lights on for a city. Thanks to the Industrial Internet of Things (IIoT) and big data, the scale, speed, and complexity of these decisions, along with the consequences for mistakes, continue to grow beyond the limits of human analysis and decision making.

Even if you could find and hire more people with the right skills, experience, and judgment, it wouldn't be enough to keep up. How are you going to reach the next level of efficiency, cost savings, and risk reduction? How are you going to design, launch, and support new products, new services, and new business models?

If your organization is struggling to make optimal decisions in the face of overwhelming speed, scale, and complexity, it's time to employ machine learning. Machine learning is a computer science approach that gives computers the ability to learn new information and solve new problems without being explicitly programmed. It's becoming the required approach for optimization in the age of IIoT and big data.

This paper provides an introduction to a method of machine learning called reinforcement learning. It briefly discusses how reinforcement learning can be used to optimize some fundamental aspects of business, then discusses in some detail how reinforcement learning optimizes a common business process such as scheduling and routing a field asset maintenance workforce.

Superhuman Performance with Machine Learning

Machine learning tackles problems characterized by:

- Massive scale
- Massive complexity
- Massive number of constraints
- High speed of calculation
- High dynamism
- High collaboration

This combination of characteristics places these problems beyond the capabilities of human decision making.

Methods of machine learning include:

Supervised learning which creates a model where the outcomes are known. The system is trained using real or example data sets, creating a model from which it can then respond appropriately to new data. For example, systems that recommend products, movies, or music you might like can be built with supervised machine learning. The system has access to lots of training data from the profiles, preferences, and past actions of existing users. Once it's trained with this data, the system then matches your profile and preferences with recommended products and choices.

Unsupervised learning which uses different algorithms to find hidden structures and relationships in unlabeled data without knowing what the possible outcomes could be. It's good for finding clusters of similar items, detecting anomalous data, and performing multivariate analysis. For example, predicting asset failure times can be accomplished with unsupervised machine learning. This is explored in the SpaceTime white paper "Machine Learning Model for Predicting Asset Failure."

Reinforcement learning which features a software agent that learns from the outcomes of the actions it takes to maximize future rewards. Different changes in the state of the environment are associated with different levels of rewards. Typically, its actions influence subsequent decisions, as the agent seeks the highest level of rewards over a given time period (the time period may be infinite). For example, a field maintenance optimization system that seeks to lower operational costs can be built with reinforcement learning. Over time the system learns to make the optimal decisions in scheduling crews, routing crews, maintaining parts inventory, and more. Optimizing these decisions contributes to reducing crew hours and equipment downtime, thus increasing the reward of lower operational costs and lower down time.

Reinforcement Learning Tackles Real World Challenges

In reinforcement learning, a software agent generates a set of actions to interact with an environment, with the goal of maximizing an objective function and amassing maximum reward for each step taken. The objective function can relate to many types of business objectives such as lower costs, lower defect rates, or fewer accidents. The software agent learns what actions to take to maximize the objective function by finding an optimal policy such that the agent accumulates the most rewards over many steps; the steps can be continuous and the time period over which it optimizes the policy may be infinite.

The software agent isn't programmed to take specific actions in response to specific situations or input. Rather, it learns the best possible actions to take given a certain state by observing the reward of those actions. Each action the software agent takes generates data about the change in the environment

Objective Function

An objective function can be either a loss function or a reward function that maps the values of one or more variables in an event onto a real number that represents the value of some "cost" (in the case of loss function) or "profit" (in the case of reward function) of that event. Minimizing the loss or maximizing the reward is the objective of an optimization model.

resulting from the action. Each action also generates some level of reward.

As it learns, the agent discovers which actions generate the most long-term reward either by trial-and-error, or by more sophisticated Monte Carlo learning methods. The agent tries an action and then observes the result. The other approach is exploiting an action that it already knows has worked in the past, then observing how well it works in a similar situation. Through applying exploration and exploitation and then observing the results, the software agent improves its performance and maximizes its target value. This combination of exploration and exploitation is typical for reinforcement learning. As with our own decisions in the real world, we explore and then we have to make decisions, often based on experience in similar situations, and we learn from the consequences of those decisions.

To apply exploration and exploitation, reinforcement machine learning needs ways to observe and influence the environment in which it is working. This makes it an ideal match for sensors and automation. In other words, reinforcement learning is a great complement to the Industrial Internet of Things.

In most business contexts, the software agent must use exploration and exploitation in the face of uncertainty. One type of uncertainty relates to the environment in which the software agent is operating. Environmental variables can be forecast but not guaranteed. For instance, a software agent optimizing driving routes may use real-time sensor data about forecasted traffic and weather conditions to make decisions. Traffic and weather conditions are, however, ultimately unpredictable.

Another type of uncertainty relates to the outcome of actions. Because the agent is acting in and on an uncertain and complex environment, the outcome of actions is variable and not guaranteed. In our driving route example, the software agent must decide when and where to stop for fuel. To reduce time in transit, the agent may decide to pass up a fueling station because there's another station along a shorter route that is currently in range. However, at the time of the decision there is no guarantee that the next station will be operational or have enough fuel to offer when the truck arrives.

Uncertainty makes finding optimal decisions impossible for humans, and difficult even for reinforcement machine learning applications. These variables maybe be unpredictable, but they often follow laws of probability. Using probabilities rather than definitive values for these unpredictable variables — an approach known as **stochastic optimization** — gives reinforcement learning applications a way to find optimal solutions that allow for the level of uncertainty.

Monte Carlo Methods

A Monte Carlo method describes a class of algorithms that performs risk analysis using random values drawn from a probability distribution for problems with uncertain inputs, such as a complex business problem. The model is run repeatedly to generate a detailed graph of output values that provide not only a prediction of what will happen, but how likely it is to happen.

Five Ways Reinforcement Learning Optimizes Your Business

Because it discovers what works best as it interacts with its environment, reinforcement learning is well suited for tackling the real world IIoT problems of today and tomorrow.

Reinforcement learning works on complex, dynamic problems under uncertainty. It seeks to maximize its objective across a long series of interactions, instead of analyzing just the current conditions.

Reinforcement learning makes real-world decisions to balance risks and rewards at a speed and scale not possible with human-centered decision making. It goes beyond predicting what will happen next to discovering

and implementing the optimal long-term course of action.

Optimizing Asset Operating Hours

Because there is a time value to money, you want to extend the operating life of assets and delay spending on repair or replacement whenever practical. As we explored in an earlier white paper, the decision involves balancing the cost of a false negative prediction of failure against the cost of capital and investment in replacing the asset too soon. Waiting too long to repair or replace the asset conserves money but risks system downtime, customer complaints, safety hazards, and even compliance problems.

Methods exist for predicting the failure times for specific assets, including using unsupervised machine learning to predict asset failure times. Predicting failure alone, however, is not sufficient to optimize the repair or replacement of those assets.

A predictive maintenance application can incorporate reinforcement learning to consider the possible outcomes resulting from the decision to repair or replace an asset at any given point in time and prescribe the optimal decision that represents the highest financial reward.

Optimizing Modes of Work

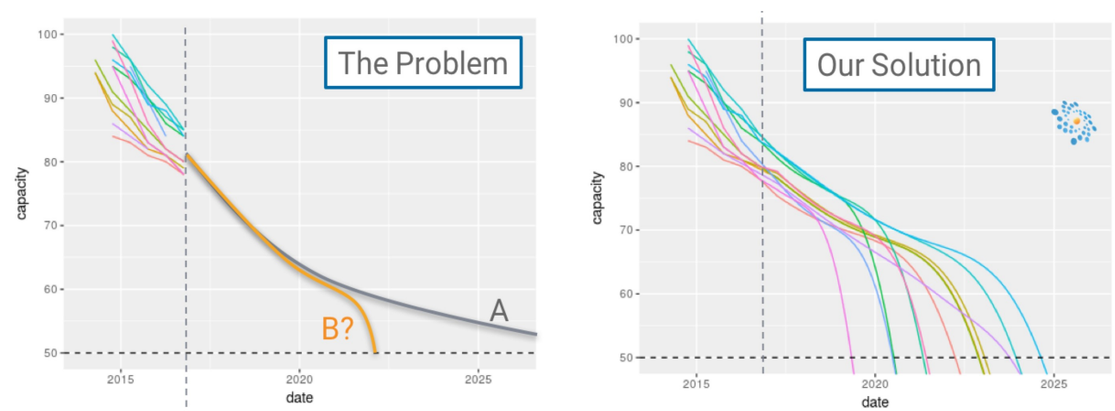
If you have multiple methods and assets for accomplishing a task, how do you choose the optimal mode? For example, package delivery companies use two primary modes to move packages: trucks and planes. Trucks cost less to operate, but hold less cargo and move more slowly than planes. Each mode has its own network of routes and hubs.

The reinforcement machine learning application ingests, analyzes, and learns from data related to shipping requirements, delivery times, route schedules, vehicle capacities, traffic, and weather to adjust the modes used to move a package. It determines the best combination of modes and routes to deliver a package on time at the lowest cost.

Optimizing Asset Risk

Every asset on which your organization relies carries some level of risk. You don't try to eliminate all risk, because doing so is expensive and practically impossible. Instead, you balance an acceptable level of risk and performance with the financial reserves necessary to cover the costs of any projected asset failures.

The stakes are high for financial reserves related to big ticket items. Suppose your company manufactures industrial battery energy storage units and offers a ten-year warranty that those units will maintain charge capacity of at least fifty percent. Typically you accrue for warranty claims based on actual historical failure rates, but in the absence of real-world failures, a conservative financial approach dictates the accrual of tens of millions of dollars for claims and the corresponding financial reserves to replace failed units.



Using failure time prediction and reinforcement learning, a battery manufacturer can determine the best combination of warranty coverage and financial accruals to optimize returns.

A realistic estimate of how many units will fail each year is a financial necessity.

Batteries degrade over time, but do so in a non-linear fashion, and the degradation rate can vary considerably from unit to unit. Further, most of these units have not been in service anywhere near the ten-year warranty window, so there is no historical data on which to base estimates of warranty claims. You can extrapolate from short-term degradation rates, but lab tests demonstrate that such projections are inaccurate.

Using reinforcement machine learning you can better predict failure times and understand the probable number of warranty claims per year. This more accurate picture would free up millions of dollars in financial reserves to be invested elsewhere.

The reinforcement machine learning application ingests, analyzes, and learns from data about battery capacity, supported loads, charging and discharging cycles, warranty claims and more. It determines the best combination of warranty coverage and financial accruals to optimize financial returns for the battery business.

The application optimizes decisions about individual battery installations and battery pricing options based on multiple factors such as:

- Manufacture and in-service dates
- Component vendors
- Location of battery installation
- Usage (commercial, industrial, utility, residential)
- Maintenance and repair record

In the end, the battery company improves both service levels and profitability, while customers benefit from highly reliable batteries backed by appropriate and affordable warranties. SpaceTime Insight will detail its battery degradation prediction capabilities in its upcoming white paper Predicting Battery Degradation.

Optimizing Your Hub Operations

You can apply reinforcement learning to increase capacity at a distribution hub by minimizing package routing time, routing errors, and handling costs.

This machine learning application ingests, analyzes, and learns from data about hub personnel, operations, and conditions. It determines the best way to unload, sort, and load packages to reach their destination with minimal delay and cost.

The application optimizes decisions about individual package routing based on multiple factors such as:

- Package origin and destination
- Package arrival and departure time at the hub
- Level of service (overnight, two-day shipping, etc.)
- Hub staffing and safety records
- Time of day and day of the week
- Local weather and traffic conditions
- On-time delivery record for packages passing through the hub

In the end, hub capacity can be increased, avoiding the expense of expanding the hub or building new ones.

Optimizing Your Asset Maintenance Workforce

Equipment downtime directly affects revenue generation, so proper maintenance is critical as is rapid response to unplanned downtime. However, your maintenance workforce is an expensive asset to operate, and total crew hours need to be tightly controlled. How you schedule and route your asset maintenance and repair crews directly affects your bottom line.

Deciding how to schedule your workforce encompasses many variables and constraints, including worker availability and skills, pay rates, parts availability, travel times, labor laws, union rules, regulatory compliance, weather, safety concerns, and service level agreements, as well as the probability that a given asset is about to fail.

There's also an opportunity cost in selecting the maintenance and repair of assets. Working on one asset means you're missing the opportunity to work on another asset that may also need to be repaired or replaced. Plus, as noted earlier, assets exist and operate in the context of your entire business. Deciding whether and when to service an asset affects multiple other assets. Optimizing this complex decision is a job for reinforcement learning.

Reinforcement Learning and Wind Farm Operations and Maintenance

The drive for clean, affordable, renewable energy is driving strong demand for wind farms. A wind farm may contain hundreds of wind turbines, with each wind turbine costing $3-4 million to install. Crews that maintain and repair turbines have a unique, and valuable, mix of technical and physical skills.

Scheduling a handful of crews to maintain and repair hundreds of turbines requires managing multiple variables, some of which, such as the weather, traffic, and real-time levels of power generation and consumption, are constantly shifting. Making optimal real-time decisions with this much uncertainty requires manipulating more variables, constraints, and tolerances than the human mind can handle.

You can apply reinforcement machine learning to scheduling wind farm maintenance crews to minimize operations and maintenance costs, turbine downtime, and safety incidents, thereby maximizing profits.

Stochastic Optimization

If you could create a model of your business where you could accurately assume certain values for all input parameters, you could develop a deterministic approach to decision making. Unfortunately, such certainty is rarely the case in a real-world operation. There is always a level of uncertainty in both the inputs and the outcomes of the decisions.

A better, but still deterministic, model would be a “what if” decision tree, where you iterate through a set of possible inputs, accounting for the probability of different values for a set of discrete variables, and map the possible cost outcomes of your decisions to derive a

set of possible results, representing a probability distribution that allows you to approximate an optimal decision.

An even better model would make use of stochastic, not deterministic, calculations, to explicitly account for the uncertainty inherent in the system by using continuous variables (a variable that can take on infinitely many, uncountable values) in your calculations. Using stochastic optimization, you look at the possible costs of your initial decision plus the costs of subsequent or "recourse" decisions made to adjust for the uncertainty of the outcome of the initial decision. Using random sets of values for the continuous variables and running your simulation many times you'll develop an understanding of not only the most likely outcome of a decision but the probability of that and all other possible outcomes.

By applying reinforcement learning on continuous variables (which is better described by stochastic optimization) you can calculate a marginal on all of the attributes of the problem, like the price of gasoline, the number of people per crew, work hours, travel time, spare parts percentage (one spare, two spares, or more), market prices for your product, and so on. If you calculate costs using marginal probabilities, you are simulating the marginal of profits for each data variable, and you can then make a calculation using these marginals that gives you the profits when asking, for example, if you should increase total crew hours per week or whether your marginal profits will decrease if you increase crew hours to reduce the time it takes to complete repairs. Using marginals you may see that your profit peaked at a certain staff level and is actually decreasing as you add crew hours.

Online Learning

Another aspect of machine learning relevant to the example is online or continuous learning. Some machine learning techniques (e.g. deep learning) use batch learning, where they are trained on a large set of data all at once to learn the best predictors of future data. Online learning, by contrast, operates on a continuous, sequential stream of data and updates its best predictor for future data at each step. This method is particularly useful in the stochastic optimization problem since crews operate in an environment that is stochastic (e.g. changing weather and traffic conditions) and the application is receiving a steady stream of time-series data (e.g. state of the equipment, location of crews, status of ongoing repairs and crew availability, weather reports, and traffic reports).

Combining stochastic optimization and online learning techniques allows us to build a reinforcement learning model to optimize continuously maintenance and repair crews to minimize crew hours and reduce downtime.

SpaceTime's reinforcement learning optimization model offers this type of simulation to optimize expected profitability (in the example case), allowing the planner to take the best decision or even to automate that decision; in our example, scheduling more hours or not. Other questions that could be put to the simulation might be: "should I increase the number of crew locations?," "what is the optimal number of parts to keep in inventory?," "if a repair crew spends time climbing a turbine to fix an unplanned issue should they perform other maintenance while they are there or return to their original schedule?"

Sensor and Automation: The IIoT Advantage

One major factor driving decisions beyond human capacities is the Industrial Internet of Things: connecting physical devices to information and communications networks. Connected devices allow your organization to collect unprecedented amounts of sensor and process data, and to exercise a new level of control and automation with manufacturing processes and assets in the field.

As explained in this paper, reinforcement learning applications require the ability to interact with an environment and to receive feedback about the impact of their interaction on the environment. Sensors and automation technology coupled with communications networks afford applications the ability to interact with an environment by starting, stopping, and modifying processes. Sensors provide applications with data about those processes. Comparing current to past process data tells applications what impact their actions had, and thus the reward associated with the action's outcome. Over time, the reinforcement learning application works towards optimizing the environment by maximizing rewards.

Using reinforcement learning within complex and real-time IIoT environments means you can help operations staff make better decisions, or even take people out of the loop entirely for decisions that are too massive, complex, rapid, and dynamic to make optimally in real time.

Conclusion

Machine learning is revolutionizing our ability to make sense of the enormous amounts of data generated by modern industrial operations. Reinforcement learning is an indispensable tool for modeling real-world business problems and optimizing and automating operations for maximum financial and other benefits.

If superhuman problems prevent you from further optimizing your costs, risks, and operations, consider putting reinforcement learning on your team. You'll move from wrestling with understanding your problems to implementing optimal solutions that save you money, lower your risk, and increase your competitive advantage and differentiation.