SYM-AM-21-089

Excerpt from the Proceedings of the Eighteenth Annual Acquisition Research Symposium

Leverage Artificial Intelligence to Learn, Optimize, and Wargame (LAILOW) for Navy Ships

May 11–13, 2021

Published: May 10, 2021

Approved for public release; distribution is unlimited.

Prepared for the Naval Postgraduate School, Monterey, CA 93943.

The research presented in this report was supported by the Acquisition Research Program of the Graduate School of Defense Management at the Naval Postgraduate School.

Leverage Artificial Intelligence to Learn, Optimize, and Wargame (LAILOW) for Navy Ships

Ying Zhao—is a research professor at the Naval Postgraduate School (NPS). Her research focused on data sciences, machine learning, and artificial intelligence methods for defense military applications such as semantic and social networks, common tactical air pictures, combat identification, logistics, and mission planning. Since joining the NPS, Zhao has been a principal investigator (PI) on many projects awarded for DoD research projects. She received her PhD in mathematics from MIT and is the co-founder of Quantum Intelligence, Inc. [yzhao@nps.edu]

Maj. Gabe Mata—is with Marine Operations Analysis Program Office, USMC, responsible for various PMW programs including Tactical Logistics Systems, Logistics Integrated Information Solutions-MC, and their modernizations. Mata is also a long-distance student at the NPS, working on his master's degree on data sciences and operation research. [gabino.mata@usmc.mil]

Erik Hemberg—is a research scientist in the AnyScale Learning For All (ALFA) group at Massachusetts Institute of Technology Computer Science and Artificial Intelligence Lab (MIT-CSAIL). He has a PhD in computer science from University College Dublin, Ireland, and undergraduate degree from Chalmers University of Technology, Sweden. His work focuses on ALFA methods to cyber security on novel ways to cyber hunt, make defenses more robust, understand what makes attacks effective, and model attack-defense coevolution. He also works in domains such as program synthesis, tax compliance, online learning, and neuronal development. [hembergerik@csail.mit.edu]

Una-May O'Reilly—is a Principal Research Scientist at the MIT-CSAIL. O'Reilly is the leader of ALFA Group at MIT-CSAIL, which conducts wide range of research applying ALFA methods, especially genetic and evolutionary computation to cyber security, software analysis, MOOC technology, and medical technology. She has served as Vice-Chair of the ACM Special Interest Group on Genetic and Evolutionary Computation (SIG-EVO). O'Reilly is a Fellow of International Society for Genetic and Evolutionary Computation (ISGEC). [unamay@csail.mit.edu]

Nate Derbinsky—is an Associate Teaching Professor, as well as the Assistant Dean for Teaching Faculty, in the Khoury College of Computer Sciences at Northeastern University. His research focuses on artificial intelligence, drawing inspiration and techniques from cognitive science, database systems, and optimization. Derbinsky earned his PhD in computer science and engineering from the University of Michigan and was a Postdoctoral Associate at Disney Research. [n.derbinsky@northeastern.edu]

Bruce Cormany—is the Integrated Product Support/Logistics Lead, Fleet Readiness Directorate (FRD) at the NAVWARSYSCOM. [bruce.cormany@navy.mil]

Joy Allen—is the C4I Systems RM&A Branch Head, Acquisition Readiness Department at the U.S. Navy NSWC Corona. [joy.b.allen@navy.mil]

LCDR Andrew Haley, USN—is a U.S. Navy Supply Officer, currently serving at Naval Supply Systems Command (NAVSUP). He has served aboard USS *Springfield* (SSN 761) and USS *Ronald Reagan* (CVN 76) and has experienced the full range of afloat operations, from deployments to shipyard availabilities. He received his Master of Science degree in Operations Research from the Naval Postgraduate School in Monterey, CA. The views expressed in this paper do not necessarily represent the views of the Navy or the United States. [andrew.r.haley1@navy.mil]

LCDR Adam Hilliard, USN—is a U.S. Navy Supply Officer, currently serving in the Office of the Chief of Naval Operations (N412). He has served multiple tours onboard ships undergoing shipyard maintenance. He received his Master of Science degree in Operations Research from the Naval Postgraduate School in Monterey, CA. The views expressed in this paper do not necessarily represent the views of the Navy or the United States. [adam.hilliard@navy.mil]

Abstract

Navy ships are complex enterprises comprised of multiple organizations that must interact smoothly and interface externally without threats to efficiency and combat-readiness. As logistical challenges increase and technology pushes response times, it is critical to introduce state of the art computational methods for analyzing the interlocked systems and training for different events. To address these challenges in this context, we introduce a framework called LAILOW: learn, optimize, and wargame. LAILOW exploits data arising from multiple sources in a complex enterprise by offering data mining, machine learning, and predictive algorithms that can be used for analysis and discovery of patterns, rules, and anomalies. LAILOW's output can then be used to optimize business processes and course of actions.

We show three use cases of using the of LAILOW framework. We show the whole LAILOW framework to search for vulnerability of a major Marine equipment's maintenance and supply system for difficult tests and evolve resilience and novel solutions accordingly. We show using of lexical link analysis (LLA) as part of LAILOW to improve the prediction accuracy of probability of failure of critical Navy Ship parts, related to C4I systems, for NAVWARSYSCOM's Predictive Risk Sparing Matrix (PRiSM) product. We also show the comparison of LLA prioritizing items in the Financially Restricted Work Que (FRWQ) with the baseline calculation.

Introduction

Leveraging deep analysis such as LAILOW for the U.S. Navy Ships is motivated by current challenges and needs. The Navy Ships conduct their activities based on the concepts of operations (CONOPS). For example, Distributed Maritime Operation (DMO) is a CONOP for the Navy; Expeditionary Advanced Base Operation (EABO) is a CONOPS for the U.S. Marine Corps (USMC). These CONOPS require capabilities, manpower, maintenance, and supply among other resources to be carefully analyzed, planned, and executed to complete missions successfully. Meanwhile, military sensors have been constantly collecting big data in many readiness components to facilitate decision-making and improve courses of actions. All the activities require data analytics. Deep analytics for big data and business intelligence including machine learning (ML), and artificial intelligence (AI) algorithms such as deep learning algorithms (LeCun et al., 2015), and game theory (Brown & Sandholm, 2017; Silver et al., 2017) have performed benchmark tasks and demonstrated the superb performance to human. It is imperative to adopt these analytics and tools to understand the entire spectrum of the Navy Ships related to complex enterprises including capabilities, manpower, maintenance, supply, transportation, health services, general engineering, and finance. The paper addresses the needs on the Navy and Marine logistics value chain, where there is a need to consider uncertainty, disruption, and perturbation that can impact the logistics plans as a whole. For example, uncertainty factors related to the environment in wide geographic areas, such as weather change, mission change from a peace time to a conflict time, or a sudden event can cause a perturbation and disruption for previous logistics and supply plans. Previously high-impact but low-failure parts may suddenly become in high demand.

The LAILOW framework focuses on three pillars of deep analytics, that is, machine learning, optimization, and wargame as shown in Figure 1. When there are data from various sources, data mining, machine learning, and predictive algorithms are often used to analyze data and discover patterns, rules, and anomalies that can later be used to optimize business processes and course of actions. New requirements have emerged in recent years that emphasize greater complexity in uncertainty, unknowns, and unexpected situations for Navy Ships. More importantly is the risk of adversarial novelty: adversaries might work on the scenarios and situations that Navy Ships might never encounter before or there are no data are available for decision-making. These requirements and concerns motivate wargame simulations that could generate synthetic data, perform what-if analyses, and explore analyses of alternatives (AoAs). The ultimate goal is to enhance total force readiness and project combat power across the whole range of military operations and spectrum of conflict at any time.

LAILOW Framework

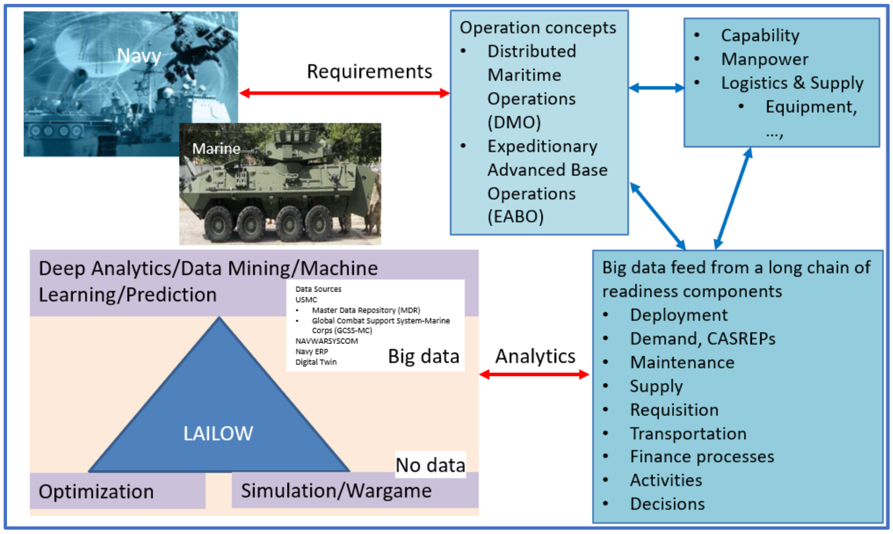

Figure 1. The LAILOW Framework Applied to U.S. Ships

The technical concept of this paper is to leverage artificial Intelligence (AI) to learn, optimize, and wargame (LAILOW) for a complex enterprise, customized to Navy Ships, especially to the logistics value chain and readiness of Navy and Marine enterprises as shown in Figure 1. The components of LAILOW are as follows:

- Component 1—Learn: When there is certain amount of data available, LAILOW performs data mining, machine learning, and learns predictive, associative, and sequential patterns from historical data.

The motivation for Component 1 is the great value of prediction. Once specific way prediction helps is to guide anticipatory preparations. For example, predictive maintenance reduces downtime and indicates which spare parts should be proactively prepared. Predictions of maintenance-related parameters (e.g., MTBF—mean time between failures, probability of failure, probability of demand) enable forecasts of parts' lifetimes under potential circumstances and scenarios. Predications of customer waiting time—the days between a trouble ticket opened and closed—reveal potential bottlenecks or ensure accurate customer expectations. This general ML capability indirectly helps extend asset product life and reduces total ownership costs.

In our use cases, we apply open source ML and data mining toolkit (i.e., Orange software by University of Ljubljana, 1996–2021) to data sets. Orange features a visual programming front-end for explorative rapid qualitative data analysis and interactive data visualization. It contains supervised ML algorithms of logistic regression, decision trees, naïve Bayes, random forest, k-nearest neighbors, and neural networks; and unsupervised ML algorithms such as principal component analysis (PCA) and k-means algorithms among others. Orange algorithms are wrapped from the python machine learning library scikit-learn (Pedregosa et al., 2011).

We also apply Soar (Laird, 2012; Laird et al., 2012), a cognitive architecture that scalably integrates a rule-based AI system and reinforcement learning (RL; Sutton & Barto, 1998, 2014). Soar-RL has advantages for defense applications over other ML/AI algorithms because it is rule-based and explainable, providing reasons for prediction, classification, and anomaly detection results. Rules can include existing tactical knowledge and rules of engagement. New rules can also be discovered from big data via online, on-policy, and continual learning. Soar has been used in modeling large-scale complex cognitive functions for warfighting processes in kill chain applications such as Combat Identification (Zhao et al., 2018) and Battle Readiness Engagement Management (BREM) wargame (Zhao et al., 2020)

The machine learning component also includes unsupervised ML algorithms. We often perform clustering, unsupervised neural network, or lexical link analysis (LLA) from the LAILOW framework to improve prediction, detect anomalies, and sort/rank important information. LLA is an unsupervised ML method and describes the characteristics of a complex system using a list of attributes or features, or specific vocabularies or lexical terms. Because the potentially vast number of lexical terms from big data, the model can be viewed as a deep model for big data. For example, we can describe a system using word pairs or bi-grams as lexical terms extracted from text data. LLA automatically discovers word pairs, and displays them as word pair networks. This innovative configuration of LLA allows us to use it to discover and rank high-value information such as attributes and factors that correlate to the measures of performance of a complex enterprise from both unstructured and structured data. Bi-grams allow LLA to be extended to numerical or categorical data. For example, using structured data, such as attributes from maintenance and supply chain databases, we discretize numeric attributes and categorize their values to word-like features. The word pair model can further be extended to a context-concept-cluster model (Zhao & Zhou, 2014). A context can represent a location, a time point, or an object shared across data sources. Figure 2 shows an output of word networks from LLA for an unstructured data of a ship corrosion patent. Each node represents a word and each link represents how likely (the strength of the link) two words are next to each other as a bi-gram phrase. For example, “polystyrene dish” has a strength 242.5. An example of LLA for unstructured data is shown in Figure 2.

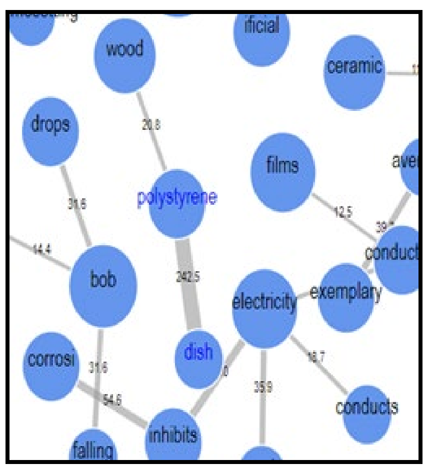

Figure 2. Examples of LLA from Unstructured Data

In this paper, we use LLA for the structured data. With the bi-gram representation and context-concept-cluster models, LLA can be used as a “market basket analysis,” where items appear together in the same context are considered associated or linked. Such association and link patterns can be used to improve prediction (Use Case 1 and 2) and rank items (Use Case 3).

Causal learning is also important for process improvement and quality control of a complex enterprise such as Navy Ships. The common consensus is that data-driven analysis or data mining can discover initial statistical correlations and associations from big data. Decision-makers and engineers often need to validate causes behind any observable effects. This calls a systematic approach of causal machine learning. The key factors for causal learning include the three layers of a causal hierarchy—association, intervention, and counterfactuals (Mackenzie & Pearl, 2018). A typical causal machine learning needs to select a cause (*C)* that maximizes the counterfactual difference *Probability (E|C) – Probability(E|Not C),* where the effect *E* is observable data and cause *C* is actionable and controllable variable*.* If deep analytics can reason and detect the cause for good or bad events (effects), engineers and decision-makers can fix the cause, avoid bad events/effects, and achieve desired effects. LLA potentially allows such a causality analysis. Causality analysis is related to the counterfactual regret minimization (CFR) that becomes important for many ML/AI applications.

- Component 2—Optimize: Based on the patterns, LAILOW optimizes the measures of effectiveness (MOEs) or the measures of performances (MOPs), defined by business decision-makers, by searching through all possible courses of actions to improve performance. The MOPs can be probability of failure (POF), probability of demand (POD), cost, time, and total readiness *(*e.g., a system uptime).

After machine learning algorithms in LAILOW discover associations, patterns, and rules, optimization algorithms can use them to search for decisions, course of actions, configurations, and combinations to optimize predicted MOEs and MOPs. LAILOW draws upon evolutionary algorithms for optimization. Evolutionary algorithms are genetic algorithms, which integrate the metaphor of genetic reproduction of selection, mutation, and crossover where the objective function's derivatives are not easy to compute.

- Component 3—Wargame: The LAILOW framework can be set up as a wargame with two players in order to test the quality or performance capabilities of a complex enterprise. One player, "self," represents the complex enterprise. The "opponent" of "self" evaluates the complex enterprise for robustness and resilience under stress, for example, when environmental factors, such as mission requirements, weather, emergency events, natural disasters, or adversaries, (who may deliberately generate disruption and exploit the vulnerability of the self-player), come into consideration or suddenly emerge.

The wargame serves the purpose of improving a real-time and dynamic operational environment through adaptative modeling. The opponent generates new operation conditions and events that might challenge the whole value chain and readiness measures in an intent to either improve the complex enterprise or disrupt the complex enterprise. The self-player adapts to optimize the actions and solutions to counter the opponent's actions/decisions. The whole process iterates and escalates due to each player adapting to the other.

To create a wargame environment, LAILOW uses coevolutionary algorithms (O'Reilly & Hemberg, 2018; Popovici et al., 2012). These are related to evolutionary algorithms and genetic algorithms (Back, 1996; Goldberg, 1989) and provide search, adaptation, and optimization mechanisms for two populations that engage to test and solve problems respectively. Coevolutionary algorithms explore domains in which the quality of a candidate solution (e.g., an action combination) is determined by its ability to successfully pass some set of tests (attackers), for example, solutions (defenders) in a logistics chain need to pass the known difficult or adversarial tests (attacks). Competitive coevolutionary algorithms are used to solve minmax problems, similar to those encountered by generative adversarial networks (GANs; Arora et al., 2017; Goodfellow et al., 2014), where adversarial engagements of opponents can be computationally modeled. Competitive coevolutionary algorithms take a population-based (parallel) approach to iterative adversarial engagement. In this competitive setting, the test (attacker) and solution (defender) strategies can lead to an arms race between the players, both adapting or evolving while pursuing conflicting objectives.

In summary, Navy Ships including USMC need to constantly perform a type of what-if and AoA wargame simulations in order to get ready for the unknown situations and perform in a contested environment. There are many challenges for Navy Ships and global materiel distribution that often require such wargame simulations. Forward deployed Navy Ships, particularly in the high operating tempo (OPTEMPO) areas such as the Seventh and Fifth Fleets, have challenges that arise in receiving logistical support when parts failures occur. These failures manifest as either a demand on the supply system, a casualty report (CASREP), or a request for technical assistance, which can cause a "redline," or a failure that stops the unit from being able to complete the whole mission until the problem can be resolved. Limited manpower, funding, storage space, and resources for repair are all in high demand (Stevens & Zhao, 2021). A good system needs to be in place to determine the most efficient and effective method of stocking, forward staging, or contracting for the materials that have the highest likelihood of demand and balance with the potential impact of failure. LAILOW can support these system requirements because it exploits algorithms for learning, optimization and wargaming.

Use Case 1: Marine Maintenance and Supply System

As part of Navy Ships, the USMC maintenance and supply chain is a complex enterprise and exemplifies socio-technological infrastructures that require continuous learning, optimizing, and wargaming. To show the feasibility of the whole LAILOW framework, we first fuse and synthesize seven years of maintenance and supply time series data for a Marine equipment, namely, Land Armored Vehicles (LAV), including maintenance, supply, and equipment usage from the database Global Combat Support System-Marine Corps (GCSS-MC). We then aggregate the data for each maintenance and supply ticket as shown in Figure 3(a). There are about 500 aggregated variables representing states and actions for both the self-player and opponent when applying LAILOW. The sample data set contains ~11% tickets that have the days between deadlined (i.e., the Marine term for "redlined") and closed date more than 32 days (32 days is the mean of the days between the deadlined and closed dates for the data set).

As shown in Figure 3(b), we first apply Orange's predictive algorithms to predict the target variable "days between deadlined and closed" for each ticket. We add LLA to improve predictive models. We also add Soar-RL as another predictive algorithm outside Orange to predict the same target variable which result in comparable predictive accuracy. Finally, we divide all the variables into two groups: Attackers and Defenders, shown in Figure 3(d), and apply the coevolutionary algorithm using the predictive rules generated using Soar-RL. The predictive rules are generated for both Attacker variables and Defender variables to predict the target variable or fitness function in opposite directions. During the wargame phase, the Attacker variables change their values to increase the Attackers' fitness, or increase the days between deadlined and closed; while the Defender variables change their values to increase the Defenders' fitness or decrease the days between deadlined and closed.

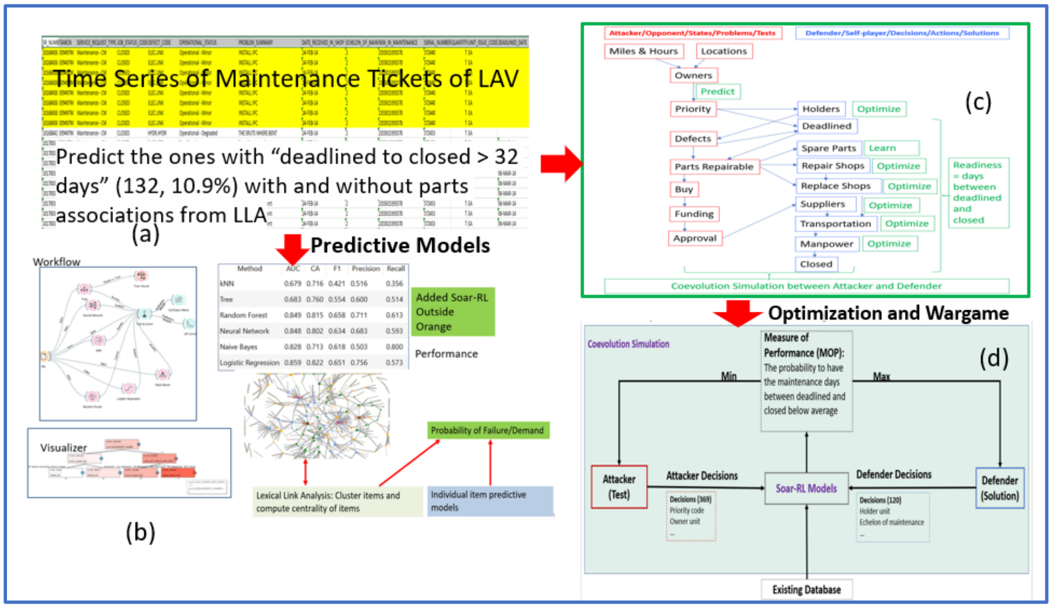

Figure 3. The LAILOW Framework Applied to the Marine Use Case

The Soar-RL and coevolutionary algorithms thus systematically simulate and discover possible new tests or “vulnerabilities” for a complex system and evolve solutions accordingly. For example, the evolved Attacker “d284e4” in Figure 4(c), which is a specific combination of Attacker variables, has an improved fitness -0.204 from where it starts from the database, i.e., -0.34 for “10fe75,” in Figure 4(a), against the best Defender “b642cf.” Such an attacker can potentially present a challenge or vulnerability to the current logistics solution system, because it is difficult for the defender to come up with a better solution than “b642cf.” Of course, the feasibility of such an Attacker configuration needs to be considered as well.

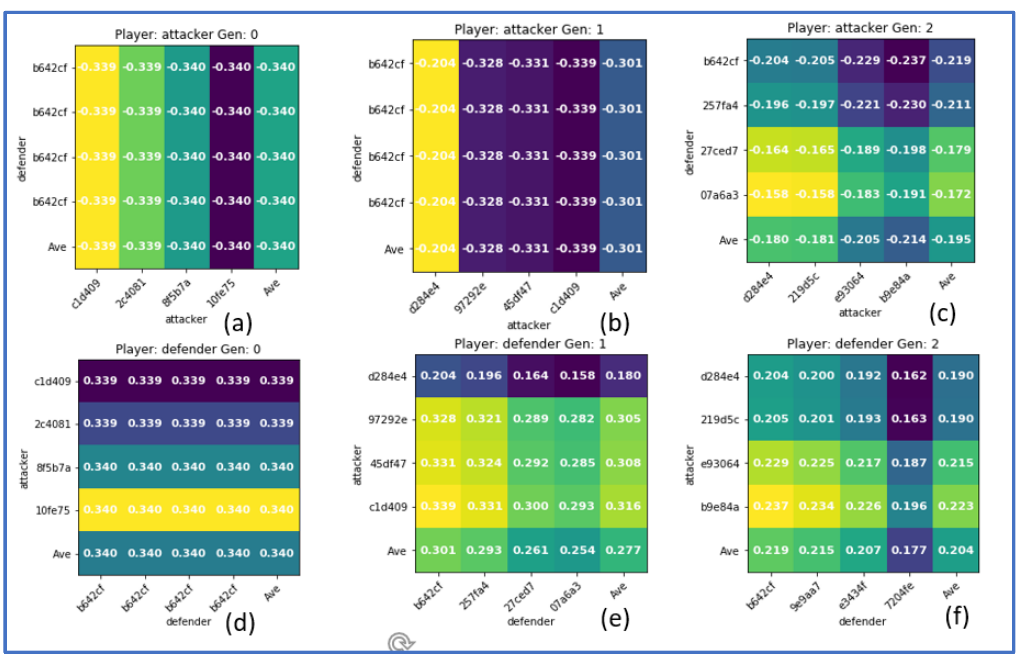

Figure 4. The Evolution Process for Attackers and Defenders.

The evolved Attacker “d284e4” in Figure 4(c), which is a specific combination of Attacker variables, has an improved fitness -0.204 from where it starts from the database, i.e., -0.34 for “10fe75,” in Figure 4(a), against the best Defender “b642cf.” Such an attacker can potentially present a challenge or vulnerability to the current logistics solution system, because it is difficult for the defender to come up with a better solution than “b642cf.”

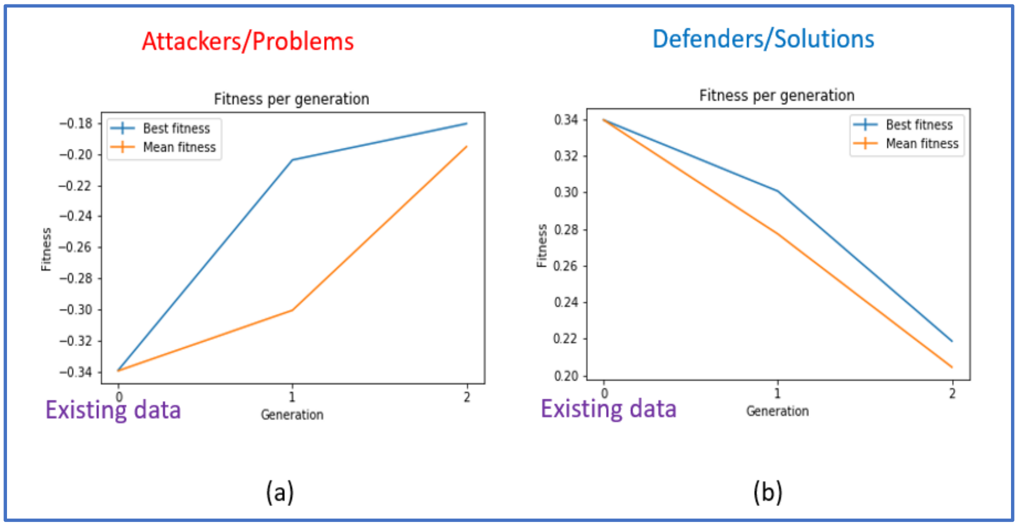

Figure 5. The Evolution Process for Attackers and Defenders

Figures 5(a) and 5(b) show the Attackers’ and Defenders’ mean and best fitness values changing for three generations in the coevolutionary algorithms, respectively. The trends validate the results and analyses that the self-player or Defender, representing the logistics solutions, gets worse on average while the opponent or Attacker, representing logistics tests, gets better on average in the coevolution simulation.

In summary, we show a use case of LAILOW which is capable of evolving, searching, simulating, and performing what-if analyses that reveal the new tests and solutions, possible vulnerability of the logistics system. The simulation can also suggest novel and more powerful solution (defender) configurations to handle new tests (attacker) that are never seen before.

Use Case 2: Predictive Risk Sparing Matrix (PRiSM)

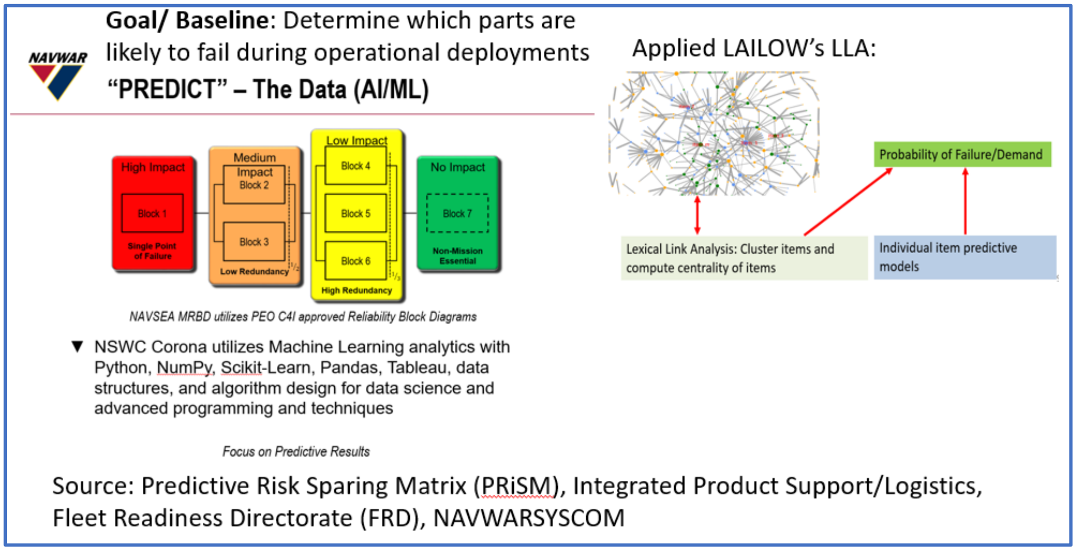

Figure 6. Predictive Risk Sparing Matrix (PRiSM)

NAVWARSYSCOM's Predictive Risk Sparing Matrix (PRiSM) as shown in Figure 6 is a product to accurately predict which critical Navy ship parts, related to Command, Control, Communications, Computers, and Intelligence (C4I) systems, are likely to fail during an operational deployment. If the ship parts' failure can be correctly predicted, the parts can be proactively staged aboard Carrier Strike Group (CSG) and Amphibious Ready Group (ARG) platforms. Advanced analytics is needed for decision-making to replace parts with low remaining service life or pre-position parts afloat or ashore or accept risk with no parts support. The decisions are then to use for course of actions, for example, to build and fill AT-5 TYCOM Allowances onboard CSG/ARG ships, transfer parts to areas of responsibility (AOR) DD site, and monitor results during the deployments. This provides support from their advanced training phase (COMTUEX) to end of deployment in increasing mission readiness by reducing Mean Logistics Delay Time (MLDT) and relative Mean Down Time (MDT). Using various key data points, PRiSM utilizes ML analytics with Python, NumPy, Scikit-Learn, Pandas, Tableau, data structures, algorithm design for data science, and advanced programming and techniques to build a baseline of prediction performance. It has been previously funded by Commander, U.S. Pacific Fleet and developed by NAVSEA, NSWC Corona. In 2020, we tested the LLA algorithm with the PRiSM product. LLA uses the parts association patterns (e.g., what parts are likely to fail with what other parts) to improve the probability of failure. For the test using real-time USS *Boxer* (BOXER ARG) and USS *Theodore Roosevelt* Carrier Strike Group (TR CSG) deployment data, the predictive accuracy improved from ~60% to ~80% by adding LLA. PRiSM and LLA are complementary and use different information to pick up different types of failure, which made the improvement possible.

Data Set

The following detail of LLA to PRiSM is showing using the data from the TR CSG. Like many ML algorithms, LLA first sifts through a so-called train data set to extract patterns (i.e., failure association patterns for patterns), network models, and visualizations, and then apply the patterns to a test/validation data set as shown in Figure 7. Both data sets were extracted from the current PRiSM application, reflecting the real-time event of the TR CSG deployment.

- Train data: Failure data 2 years prior to the TR CSG deployment

- Test/validation data: Failure data during the TR CSG deployment

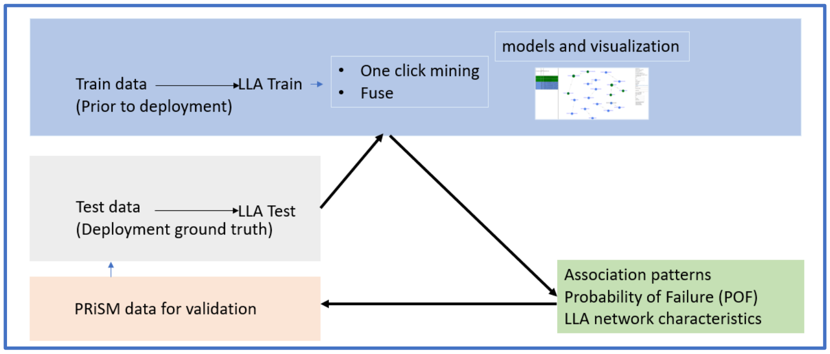

Figure 7. LLA Process Overview

LLA Application

We first use LLA to compute associations to see if two items or parts (used interchangeably below) are linked in the sense if one item fails, the other item also fails. The steps are shown below:

1. In order to compute statistically significant association patterns, we first group the failed parts into "baskets." Each basket is identified as a combination of the location of a failed part in the system (e.g., "object index," "unit id") and timestamp of the failure event (e.g., "year and month of event start date time") as shown in Figure 8. We choose the failure event time "month" for the basket combination for associated failed items which fail in a period of month sequentially. Figure 8 also shows an example of basket and item pairs.

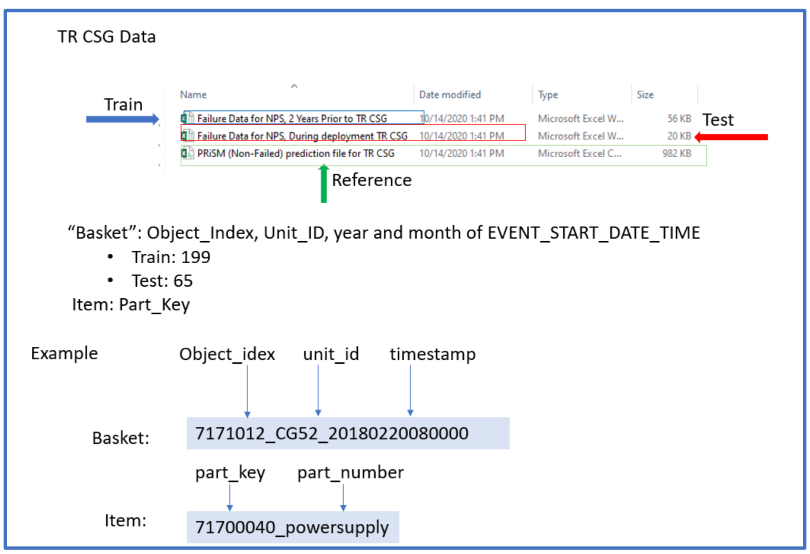

Figure 8. Basket and Item Mapping for the TR CSG Sata

2. LLA applies causal learning and compute counterfactual proportion difference, i.e.,

$$cf = [P(B|A) - P(B|\text{Not } A)] * (\text{pooled sample size}) \qquad (1)$$

to compute the strength of the association of two parts as *cf*, where P(B|A) is the probability of part B fails within the same basket (i.e., fails at the same location and time frame) if part A fails. The pooled sample size is a pooled number of historical failure of item A and B based on their pooled historical failure probabilities. *cf* is a z-score (PSU, 2021) and we use *cf >1.96* for *p-value < 0.05* as the statistical significance for the associations. In Figure 9, item A could be more important than item B for causality, because A's failure might cause B's failure although A has fewer failure than B in total. The step generates an item network for all historical failed parts for the train data set.

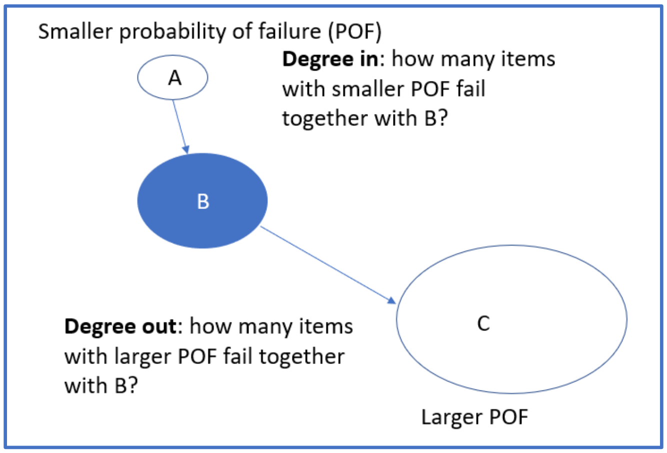

Figure 9. Causal Learning in LLA

3. Once a network of items is generated using the measure defined in Equation (1) , a community finding algorithm (Girvan & Newman, 2002) is performed to cluster items into groups or communities, and then compute centralities such as degree in, degree out, and betweenness scores.
4. Outputs and centralities from LLA
    a. The probability of failure (POF): the percentage of baskets containing an item (e.g., item B in Figure 8)
    b. Degree in: the number of items with smaller POF (e.g., item A in Figure 9) fail together (i.e., in the same basket) with item B
    c. Degree in weight: degree in/average (cf), total estimated causal impact from other items
    d. Degree out: the number of items with bigger POF (e.g., item C in Figure 9) fail together (i.e., in the same basket) with item B
    e. Degree out weight: degree out/average (cf), total estimated causal impact to other items if item B fails
    f. Betweenness: the number of items to which item B links are in the different groups from the one of item B based on the community finding algorithm output.

Results

The outputs and centralities from LLA are used to improve the prediction for PRiSM. Fifty-one high impact C4I parts that were predicted failed and had actually, matched with either the LLA predictions or PRiSM predictions, improved from 36 matched from the PRiSM predictions alone. The total failed number of items failed is 64. PRiSM and LLA are complementary in terms of predicting failed items in this use case.

Use Case 3: Readiness Impacts of Underfunding Spares Backlogs

Navy Ships' aviation and maritime units order spares from their general funds to fill modeled allowances. If there is not enough funding to buy all modeled allowances, spares requirements accumulate in a Financially Restricted Work Que (FRWQ) awaiting resourcing. In the meantime, the systems with these parts support are still fielded, and the Fleet still generates requirements to replace these parts.

The goal of this use case is to conduct a comparison of Fleet Demands against requirements in a FRWQ and assess these requirements with high priority demands linked to maritime units' CASREP and aviation unit's casualties, i.e., Non-Mission Capable Supply (NMCS). The result will help improve and determine the efficient and effective method of prioritizing materials that have the highest likelihood of demand balanced with their impact to readiness as a whole.

Baseline

The current tool and methodology of scoring and prioritizing the items are based on the DoD Manual 4140.01-V2 (DoD, 2018) in a FRWQ, with respect to their impact to the weapon system and aviation readiness data from CASREP and NMCS. The DoD Manual 4140.01-V2 describes that two categories of measures, i.e., the weapons system criticality and fleet demand, are needed to prioritize items as shown in Figure 10. The weapons system criticality is measured by the Item Mission Essentiality Code (IMEC) or Weapon System Group (WSG) code. We use IMEC in this paper as follows:

- IMEC Points: IMEC=5, 100 points awarded; IMEC=4, 80 points awarded; IMEC=3, 60 points awarded; IMEC=2, 40 points awarded; and IMEC=1, 20 points awarded. Figure 11 shows the detail of IMEC points calculation.

The fleet demand is based on the two criteria of intermittency and correlation variance (CV) for scoring and prioritizing items

- CV Points: calculated as the ratio of the standard deviation of the demand to the average demand and normally expressed as a percentage. A "low variance" is less than 75%; "median variance" is between 75% and 125%; and "high variance" is greater than 125%.
- Intermittency Points: calculated as the percentage of total historical demand periods (e.g., months in a year) that have non-zero demand. A "continuous" intermittency for an item means it is needed greater than 85% of 12 months (i.e., at least 11 of 12 months), while "limited" is less than 10%; "uneven" is between 10% and 60%; "erratic" is between 60% and 85%. Figure 12 show the details of CV and Intermittency calculations.
- Platform/Type Points: In addition to the IMEC, CV, and Intermittence points, platform and hull type are also used in the baseline calculation for total points. The overarching idea is that afloat units are given higher priority than shore units. Within afloat, we can prioritize further by leaning on U.S. Fleet Forces Command, which releases a fleet priority list on a semi-annual basis. Figure 13(a) shows the detail of the points calculation. This is a very insightful list, as it shows which units are highest priority (deploying soon) vs. low priority units (those in extended shipyard avails). But even the lowest priority afloat unit would still receive slightly higher points than a shore unit. The platform/type ranking shown in Figure 13(b) is just for CVNs. Other platforms and types receive their own rankings.

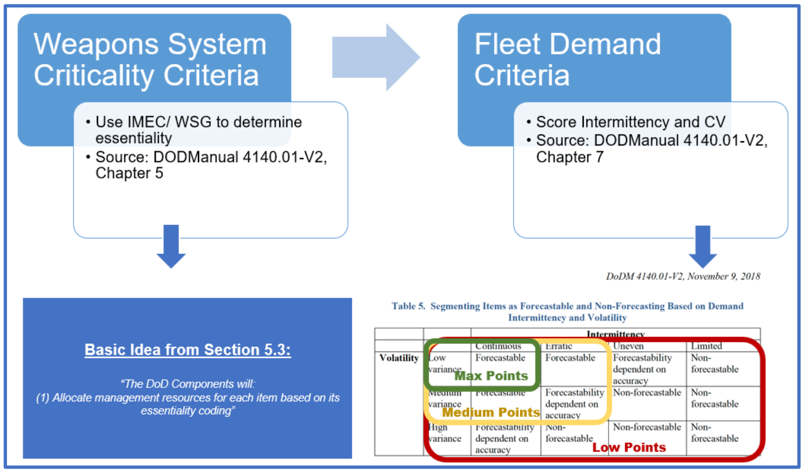

Figure 10. Current Methods of Prioritizing Items for a FRWQ

IMEC Score Legend

IMEC_CATEGORY	IMEC Meaning	POINTS AWARDED
5	Critical	100
4	UNK	80
3	UNK	60
2	UNK	40
1	Non-Critical	20
Unknown/Missing	N/A	60

Figure 11. IMEC Points Calculation (DoD, 2018)

Platform Type Score Legend

TYPE_NAME	TYPE_CATEGORY	POINTS AWARDED
CVN/LDECK	AFLOAT [AVIATION]	55-100 (USFF Sliding Scale)
All Maritime Units	AFLOAT [MARITIME]	51-100 (USFF Sliding Scale)
MALS	SHORE	50
NAS	SHORE	50
Unknown/Missing	N/A	50

CVN Name	Type Points
DWIGHT D. EISENHOWER	100
THEODORE ROOSEVELT	95
RONALD REAGAN	90
CARL VINSON	85
HARRY S. TRUMAN	80
NIMITZ	75
ABRAHAM LINCOLN	70
GEORGE H.W. BUSH	65
GEORGE WASHINGTON	60
JOHN C. STENNIS	55

Figure 13. Platform/Type Points Calculation

In the base line calculation, total points are the sum of CV points, Intermittency points, IMEC points, and Platform/Type Points. The maximum possible score is 400 points, and the minimum possible score is 115 points.

Data Sets

There are two data sets used in the use case:

- Data set 1: Historical raw demand for items related to aviation readiness and NMCS
- Data set 2: Historical raw demand for items related to maritime parts and CASREPs

The baseline points calculation examples:

1. An aviation example for NIIN 015761575 associated with R2010393330312 (LHA 6 Document)
   - CV Points: low variance (100 points awarded)
   - Intermittency Points: continuous (100 points awarded)
   - IMEC Points: 4 (critical system; 80 points awarded)
   - Platform/Type Points: LDECK (afloat type, highest LDECK Priority for USFF; 100 points awarded)
   - Grand Total: 380 points (highest priority for FRWQ Investment)
2. A maritime example: NIIN 005181789 associated with N2194510161535 (DDG 71 Document)
   - CV Points: low variance (100 points awarded)
   - Intermittency Category: continuous (100 points awarded)
   - IMEC Category: 5 (critical system; 100 points awarded)
   - Platform/Type Points: DDG (afloat type, 4th highest CG/DDG priority for USFF; 98.5 points awarded)
   - Grand Total: 398.5 Points (highest priority for FRWQ Investment)

LLA Application

LLA can be applied to the maritime data set 2 more meaningfully since it contains a data attribute JCN that is used to group the items into a same requisition time or “basket.” The items are the National Item Identification Numbers (NIIN) and hull type. There are 611,335 unique baskets and 280,762 unique items in this data set; 2,093,633 statistically significant associations are found. The hypothesis is that items that appear together in the same baskets in the raw data, historical data might they be associated with a same cause so they are demanded together.

In general, LLA and network theory are potentially to provide a network and centrality view of the items/parts generated from raw demand data, which is related to the network analysis applications, for example, ranking people in a social network or ranking an object such as a biological gene in an environment where such a baseline ranking is not available.

However, application of LLA in this use case do not conclude better and more meaningful rankings than the existing methods. The correlations of LLA scores and baseline points are shown in Table 1. LLA suggests using “Degree out weight” scores as the total estimated impact to other items as the scores for the item’s importance, which has the least correlation of the total points. One of association patterns discovered is meaningful as shown in Figure 14. POD has a correlation 0.34 with the total points. This indicates LLA’s centrality measure “Degree out weight” does not use demand as signals for deciding the importance of an item. This may indicate the argument of causality learning that one item’s demand might cause another item’s demand may not fit to this problem, and the low-demand and high-impact items may not exist in the FRWQ data.

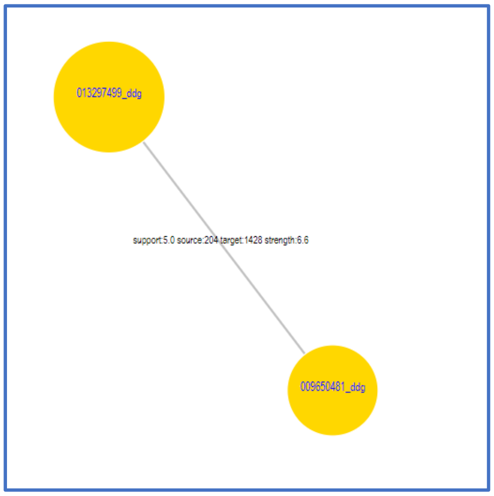

Figure 14. Example of LLA Associations Between Two Items

This calls for further research of LLA, possibly applications of other LAILOW methods such as supervised ML methods. For example, feedback data needs to be collected for consequences to understand how item prioritizing decisions and resource allocation decisions' impact future readiness. Future work also includes a review of business processes at a holistic level and consideration to plan for a whole class of ships or a whole fleet for a period of time (e.g., the CVN-74, USS *John C. Stennis* group for last a few years).

Table 1. Correlations of LLA and Baseline Points

	Total Points	CV	Intermittency	IMEC	Platform/Type
POD	0.34	0.34	0.36	0.09	-0.09
Degree in weight	0.20	0.10	0.25	0.16	-0.11
Degree out weight	0.08	0.06	0.19	-0.019	-0.06
Degree	0.07	0.03	0.15	0.012	-0.04
Betweenness	0.17	0.07	0.19	0.18	-0.13

Conclusions

In this paper, we show the LAILOW framework provides a holistic predictive and simulation platform to improve the readiness of Navy Ships. The Soar-RL, comparable to other predictive machine learning algorithms, rule-based, and explainable, can be integrated with the coevolutionary algorithm to conduct a wargame for a Navy complex enterprise. The wargame simulates and discovers possible new tests or "vulnerabilities" of a value chain for U.S. Ships and related complex enterprises, and evolve solutions or "resiliency" accordingly for uncertain and new conditions.

Recommendations and Future Work for Navy Ships

In some use cases, LAILOW methods may require to collect more right data for deep analytics such as feedback data to be collected for consequences to understand how item prioritizing decisions and resource allocation decisions' impact future readiness. Navy Ships need to adopt more deep analytics, machine learning and AI algorithms for big data or no data and focus on the entire spectrum or end-to-end (E2E) logistic planning.

References

Arora, S., Ge, R., Liang, Y., Ma, T., & Zhang, Y. (2017). Generalization and equilibrium in generative adversarial nets (GANs). *Proceedings of the 34th International Conference on Machine Learning*, 70, 224–232. http://proceedings.mlr.press/v70/arora17a.html

Brown, N., & Sandholm, T. (2017). Safe and nested endgame solving for imperfect information games. *Proceedings of the AAAI workshop on Computer Poker and Imperfect Information Games*. https://arxiv.org/pdf/1705.02955.pdf

DoD. (2018). *DoD supply chain materiel management procedures: Demand and supply planning* (DoD Manual 4140.01-V2). https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/414001m/414001m_vol02.pdf

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, *99*(12),7821–7826.

Goldberg, G. (1989). *Genetic algorithms in search, optimization, and machine learning*. Addison Wesley.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 2672–2680. https://papers.nips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf

Laird, J. E. (2012). *The Soar cognitive architecture*. MIT Press.

Laird, J. E., Derbinsky, N., & Tinkerhess, M. (2012). Online determination of value-function structure and action-value estimates for reinforcement learning in a cognitive architecture. *Advances in Cognitive Systems*, 2, 221–238.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521 (7553), 436–444. https://doi.org/10.1038/nature14539

Loch, J., & Satinder, S. (1998). Using eligibility traces to find the best memoryless policy in partially observable Markov decision processes. *Proceedings of the 15th International Conference on Machine Learning*, 323–331. Morgan Kaufmann Publishers.

Mackenzie, D., & Pearl, J. (2018). *The book of why: The new science of cause and effect*. Penguin.

O'Reilly, U., & Hemberg E. (2018). An artificial coevolutionary frame work for adversarial AI. *Proceedings of AAAI Fall Symposium: Adversary-aware Learning Techniques and Trends in Cybersecurity (ALEC)*. http://ceur-ws.org/Vol-2269/FSS-18_paper_37.pdf

Pearl, J. (2019). The seven pillars of causal reasoning with reflections on machine learning. *Communications of the ACM.* https://ftp.cs.ucla.edu/pub/stat_ser/r481.pdf

Pedregosa, F., et al. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12, 2825–2830. http://scikit-learn.org/

Popovici, E., Bucci, A., Wiegand, R. P., & De Jong, E. D. (2012). *Coevolutionary principles.* In G. Rozenberg, T. Back, & J. N. Kok (Eds.), *Handbook of natural computing* (pp. 987–1033). Springer.

Silver, D., et al. (2017). Mastering the game of Go without human knowledge. *Nature*, *550*(7676), 354–359. https://doi.org/10.1038/nature24270

Stevens, E., & Zhao, Y. (2020). Using lexical link analysis (LLA) as a tool to analyze a complex system and improve sustainment. *Proceedings of AAAI Symposium on the 2nd Workshop on Deep Models and Artificial Intelligence for Defense Applications: Potentials, Theories, Practices, Tools, and Risks*. http://ceur-ws.org/Vol-2819/session1paper3.pdf

Sutton, R. S., & Barto, A. G. (1998, 2014). *Reinforcement learning: An introduction*. MIT Press.

University of Ljubljana. (1996–2020). *Orange software.* https://orange.biolab.si/

Penn State University (PSU). (2021). *Online statistics: Normal approximation method formulas*. https://online.stat.psu.edu/stat200/lesson/9/9.1/9.1.2/9.1.2.1

Zhao, Y., & Zhou, C. (2014). System and method for knowledge pattern search from networked agents (U.S. Patent No. 8,903,756). U.S. Patent and Trademark Office. https://www.google.com/patents/US8903756

Zhao, Y., Derbinsky, N., Wong, L., Sonnenshein, J., & Kendall, T. (2018, December 2). Continual and real-time learning for modeling combat identification in a tactical environment. *Proceedings of the NIPS 2018 Workshop on Continual Learning.* Montreal, Canada. https://marcpickett.com/cl2018/CL-2018_paper_77.pdf

Zhao, Y., Nagy, B., Kendall, T., & Schwamm, R. (2020, November). Causal learning in modeling multi-segment war game leveraging machine intelligence with EVE structures. *Proceedings of AAAI Symposium on the 2nd Workshop on Deep Models and Artificial Intelligence for Defense Applications: Potentials, Theories, Practices, Tools, and Risks* [Virtual]. http://ceur-ws.org/Vol-2819/session3paper3.pdf

Acquisition Research Program
Graduate School of Defense Management
Naval Postgraduate School
555 Dyer Road, Ingersoll Hall
Monterey, CA 93943

www.acquisitionresearch.net