Rule Extraction and Visualization from Data Traces

Paul Mario Koola	Satheesh Ramachandran	Kalyan Vadakkeveedu
Knowledge Based Systems Inc.	**Knowledge Based Systems Inc.**	**Knowledge Based Systems Inc.**
College Station, Texas	**College Station, Texas**	**College Station, Texas**
pkoola@kbsi.com	sramachandran@kbsi.com	kvadakkeveedu@kbsi.com

ABSTRACT

In this paper we present methods, process and algorithms to decipher the hidden rules or logic of complex traffic patterns – traces of assets, people and equipment -- and demonstrate technologies for Automated Rule Learning from Data Traces. We call this system for automated rule learning from data traces - TraceLogic.

We demonstrate this technology on aircraft carriers. Operations on aircraft carriers are often described as controlled chaos. Complex organizational systems, such as those that govern and effect aircraft movement onboard carriers, do not operate on predetermined low level rules but rather adapt and learn based on high-level guidelines. Critical situations that are often new or unique are best tackled by adaptive and well-trained organizations.

We present typical examples of data statistics, visual analytics, and rule and pattern analysis from carrier deck notional data. These techniques help analysts explore the data and improve the operational process. These tools could also support Validation and Verification (V&V) of the rules engine of the simulator.

The algorithms and logic developed in TraceLogic are scalable across domains. The TraceLogic methodology, algorithms and tools have a number of potential customers in the Navy community, the broader Department of Defense community and the commercial sector. This technology will scale to manage any asset moving in groups subject to complex spatio-temporal constraints. Some of the possible application domains are depot maintenance, manufacturing shop floor, airport traffic management and even theme parks where people move in groups.

ABOUT THE AUTHORS

Paul Mario Koola, Ph.D. is a research scientist at Knowledge Based Systems Inc., Texas, USA and a principal investigator managing AI and machine learning projects.

Satheesh Ramachandran, Ph.D. is a research scientist at Knowledge Based Systems Inc., Texas, USA and an experienced independent practitioner in the application of machine learning systems across many industries.

Kalyan Vadakkeveedu is a researcher at Knowledge Based Systems Inc., Texas, USA and works on machine learning and AI projects.

Rule Extraction and Visualization from Data Traces

Paul Mario Koola

Knowledge Based Systems Inc.

College Station, Texas

pkoola@kbsi.com

Satheesh Ramachandran

Knowledge Based Systems Inc.

College Station, Texas

sramachandran@kbsi.com

Kalyan Vadakkeveedu

Knowledge Based Systems Inc.

College Station, Texas

kvadakkeveedu@kbsi.com

INTRODUCTION

In this paper we present methods, process and algorithms to decipher the hidden rules or logic of complex traffic patterns – traces of assets, people and equipment -- and demonstrate technologies for Automated Rule Learning from Data Traces. We call this system for automated rule learning from data traces - TraceLogic.

We demonstrate this technology on aircraft carriers. Operations on aircraft carriers are often described as controlled chaos. Complex organizational systems, such as those that govern and effect aircraft movement onboard carriers, do not operate on predetermined low level rules but rather adapt and learn based on high-level guidelines (Rochlin et. al., 1987). Critical situations that are often new or unique are best tackled by adaptive and well-trained organizations.

Aircraft handling aboard the flight and hangar decks of aircraft carriers is a series of complex processes in a constrained environment which involves recovery, refueling, payload loading, servicing and maintenance, manning, and positioning for the next launch. As unmanned air systems begin to integrate in greater numbers (Barno et. al 2014, US Navy 2016, CVN-78 2013) it will be critical to understand this traffic flow better to allow for the development of wholly new paradigms for future carriers to support 1) use of both manned and unmanned systems, 2) increased tempo of operations and 3) maintain sortie generation rate requirements under degraded conditions.

The key challenge posed by our work is to discover and codify the rules of behavior "encoded" into the patterns of performance of aircraft handlers and their leadership. Based on our research we show that algorithms are capable of generating a large rule set based on statistical properties of the data. It is essential that domain experts and algorithm experts research the rules generated by the algorithms and further refine the methodology for processing this rule set. This methodology will be used to formulate a design for developing tools that provide automated aids for processing this vast set of rules generated. These tools will help sift out domain relevant, but undiscovered, patterns/rules for better understanding the complexities of operations on an aircraft carrier. Eventually, the new knowledge can be used to further improve operations. These technologies are being developed using realistic simulated notional data in order to facilitate transition to actual sensor data when that becomes available in the near future. The "notional" input data for this analysis has been provided by Stoller Newport News Nuclear "Next Gen Carrier" simulator.

The algorithms and logic developed in TraceLogic are scalable across domains. The TraceLogic methodology, algorithms and tools have a number of potential customers in the Navy community, the broader Department of Defense community and the commercial sector. This technology will scale to manage any asset moving in groups subject to complex spatio-temporal constraints. Some of the possible application domains are depot maintenance, manufacturing shop floor, airport traffic management and even theme parks where people move in groups.

METHODOLOGY

Conceptual Architecture

The conceptual architecture of TraceLogic applied to aircraft carriers is illustrated in Figure 1. The overall solution can be viewed as an adaptive feedback learning system that extracts operational data, and performs automated discovery of operational patterns. These patterns can eventually be coded within autonomous systems. These technologies also can be used as a design exploration tool for aircraft carrier air operations optimization.
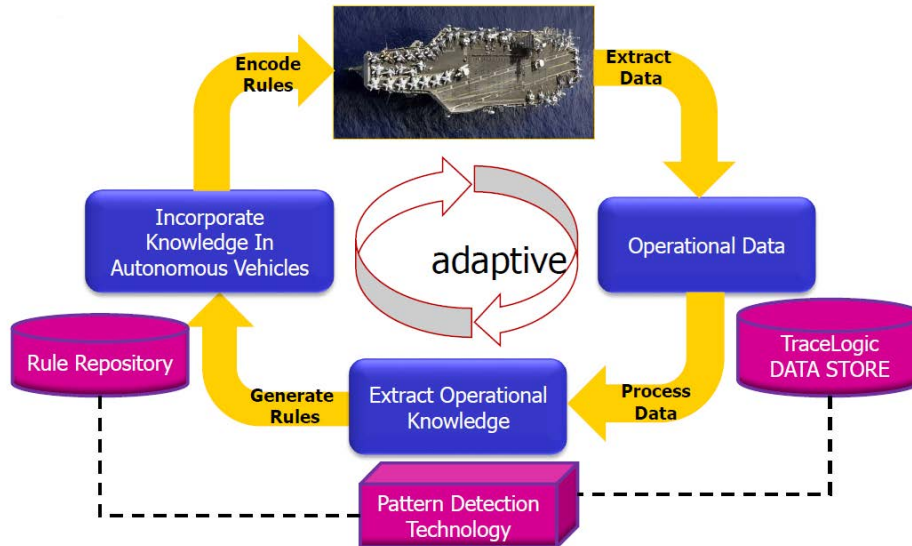
Figure 1: TraceLogic Conceptual Architecture

The input to this architecture (Figure 1) is the domain from which data is extracted. Raw data is then transformed into a common operational data format that is internally stored in the data store. Pattern detection and machine learning algorithms are used to discover the knowledge patterns from operations data. Eventually, the discovered rules can be used for improving existing operational processes or ultimately incorporated to control autonomous vehicles.

Data Statistics

The first step is to provide some insights into the essential characteristics of the dataset, including the dimensions in terms of number and types of aircrafts, locations, and activities, as well as some statistical properties of the dataset. The five fixed wing aircraft types using the shuttles are classified as:

➢ two 10-aircraft squadrons of Fighter 1,
➢ one 12 aircraft squadron of Fighter 2,
➢ one 12 aircraft squadron of Fighter 3,
➢ one 5 aircraft squadron of Electronic Warfare and
➢ one 5 aircraft squadron of Electronic Attack
➢ one 6 aircraft squadron of Helicopters
➢ one 5 aircraft squadron of Helicopters

One of the data sets included 3323 launches, and represented approximately 350 events. A typical event could include different number of launches (as shown in the count distribution Figure 2), and each of these launches could be from any of the five launching spots (including four shuttle spots shown in Distribution of Launch Spots Figure 2). Shuttle 2 is used for 37% of the launches.
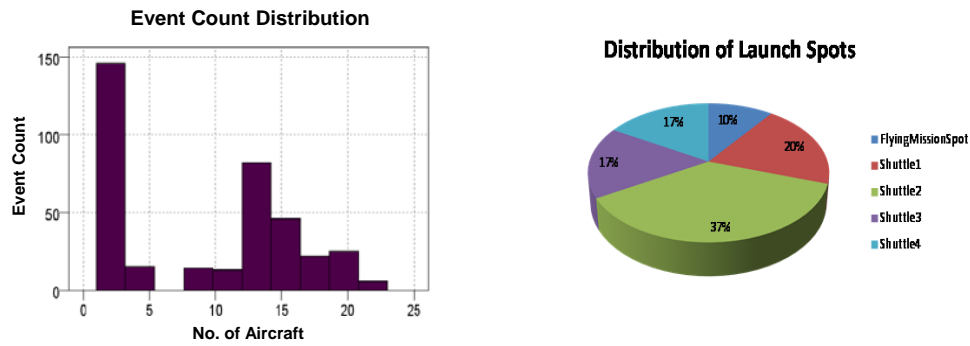
Figure 2: Summary Stats

Value	Proportion	%	Count ▽
Shuttle2		36.59	1216
Shuttle1		19.89	661
Shuttle3		16.73	556
Shuttle4		16.61	552
FlyingMissionSpot		10.17	338

Figure 3: Shuttle Usage by Aircraft

The shuttles are used by all types of fixed-wing aircraft with no overarching bias towards a particular shuttle for any aircraft type. Figure 3 shows the proportion of the five aircraft types that use a particular shuttle. In addition to the fixed wing aircraft there are two batches of helicopters as shown in the last row of Figure 3 using a different spot for lift off.

Visual Analytics

We applied clustering algorithms to the landing time versus launch time data for all 30 days of operations. The blue ellipses identifies the clusters. The markers inside the ellipses represent different types of aircraft. If an aircraft launches more than 30 minutes before or after all previous flight groups, it was considered as part of a new flight group. Additionally, if an aircraft lands 12 minutes before or after all previous flight groups, it is considered part of a new cluster. Figure 4 shows a zoomed in view of the flight group clustering results. There are clusters of various types of aircraft and numbers. Some of the cluster groups are well separated, whereas some others are not. There does not appear to be any noticeable ordering of individual aircraft types within similar clusters. A well-defined sequence of aircraft launch, can be explained based on real world knowledge.
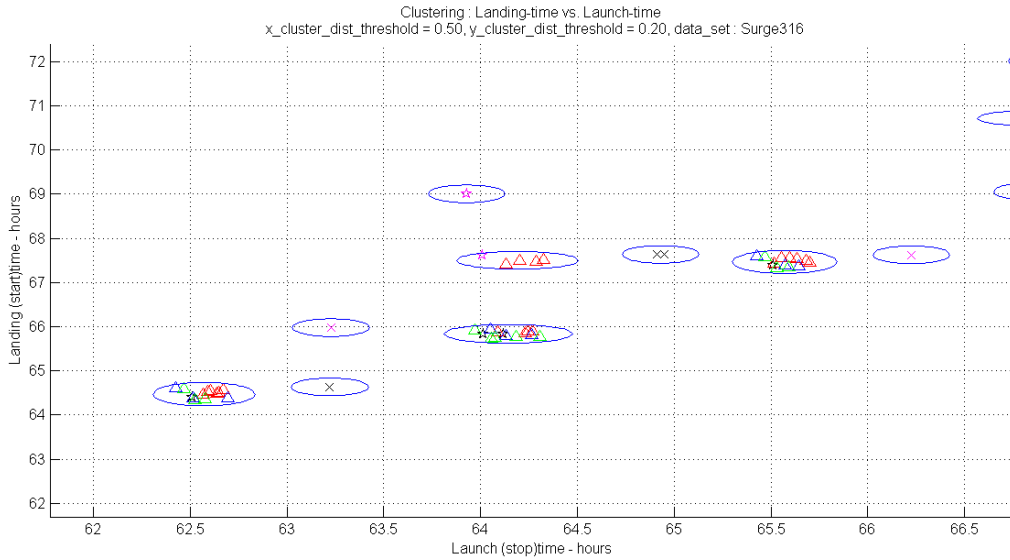
Figure 4: Flight Group Clustering

Figure 5 shows the global view of all aircraft on the carrier deck over a 30 day operational cycle. Each row represents an aircraft and each column a second in time. Hence every row has about 2.5 plus million pixels representing every second. The black regions are inactive regions. We can quickly determine the group of aircraft that are simultaneously active and the dead time zones where there is no activity. This global view of activity across all aircraft and across all time can be filtered and displayed for every one of the hundred plus activities provided in the data corresponding to recovery, refueling, payload loading, servicing and maintenance, manning, and positioning for the next launch. Standard zooming and panning is provided to dig into and explore this data in more detail. The size of this image is limited by the computational resources available especially memory to handle the data.
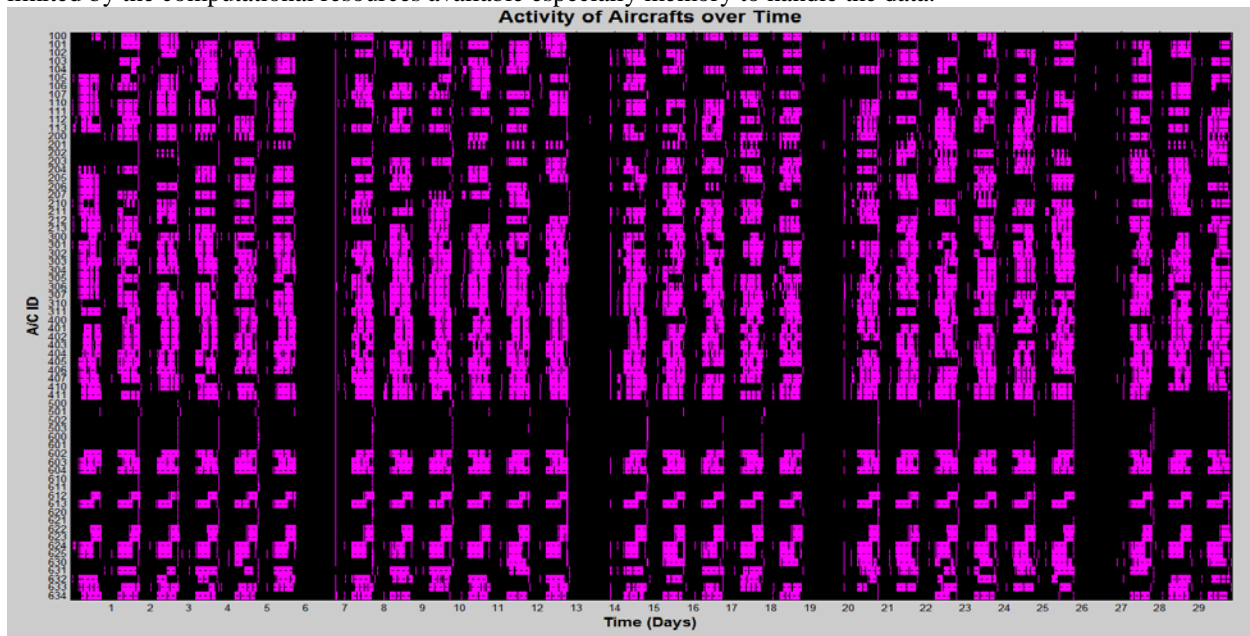

Figure 5: Global Activity over Time

Figure 6 is an example of a visualization, overlaid on the physical domain, the deck of an aircraft carrier. Figure 6 shows the paths taken by all aircraft in a given aircraft type, for multiple simulation runs. The yellow color depicts the last simulation run. The thickness of the lines indicate the amount of repetition for the same path. To those knowledgeable of this domain, these path patterns can be very useful. For instance this aircraft type uses catapult C3 extensively and catapults C1 and C2 sparingly. Catapult C4 was never used as it could have been under maintenance.

Aircraft are typically parked at the edge of the carrier deck. The thin magenta lines shows the path from parking to launch. The thick lines indicate that these aircraft are mainly in a high tempo operation landing and taking off, in a continuous loop. Just as with Figure 5, these paths can be filtered based on individual aircraft or aircraft types for different operational conditions as provided in the data. Figure 5, and Figure 6 can be thought of as visual query mechanisms with filtering capability.
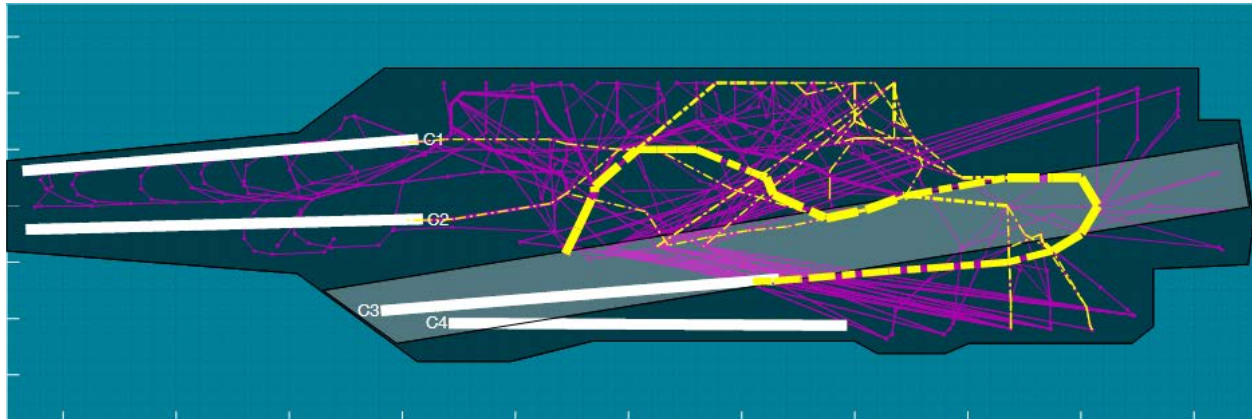

Figure 6: Path Visualization

In Figure 7 we present a summary and analysis of unique path analysis across all scenarios of catapult failures. This figure depicts the percent unique paths for every aircraft versus its type designation on the abscissa for different scenarios. In addition we compute a simple metric "Mean Path Uniqueness," as the average path uniqueness across all aircrafts. As can be seen, for catapult 2 failure "Mean Path Uniqueness" value stands at 70, different than all other scenarios approximately in the 50 range. The vertical line bounded by a red upper marker and the yellow bottom marker, indicates maximum and minimum values for each aircraft. The green circles are the mean value for that aircraft across 20 repeat simulations. Another very interesting aspect of the 20 repeat simulations is that the repeats have a lot of variations. The complexity of the entire rule set gives rise to this behavior. This does not look like a simple discrete event simulation generating data from a standard distribution function. Such analysis is thus helpful for Validation and Verification (V&V) of the rules engine.
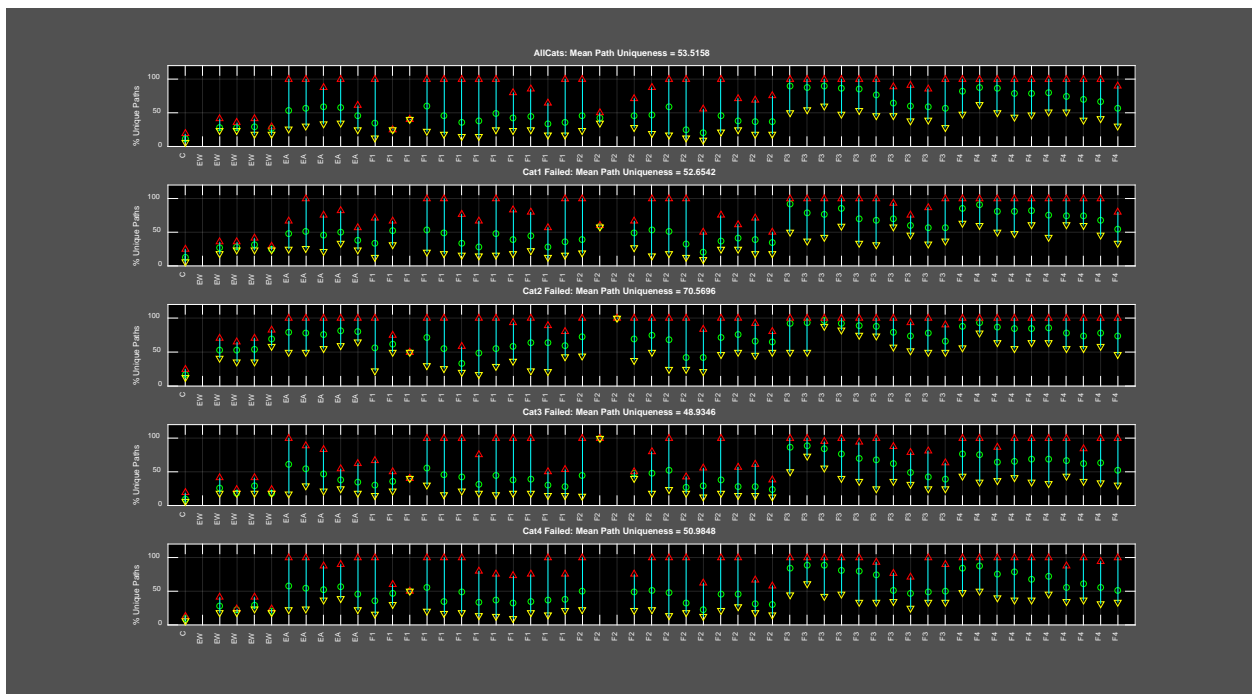

Figure 7: Summary of Unique Paths across Scenarios

Rules and Pattern Analysis

Based upon the statistics, the data offers sufficient variance that allow for investigating the key factors that influence different metrics such as throughput rate. There are a variety of factors that can potentially influence efficiency such as the availability of the shuttles, composition of ACs that are being launched, the parking places that are feeding the events, the specific navigational paths that are available, etc. Our approach was to leverage machine learning algorithms to extract patterns that encapsulate the relationship between domain factors and efficiency. The conceptual approach is depicted in Figure 8 below.
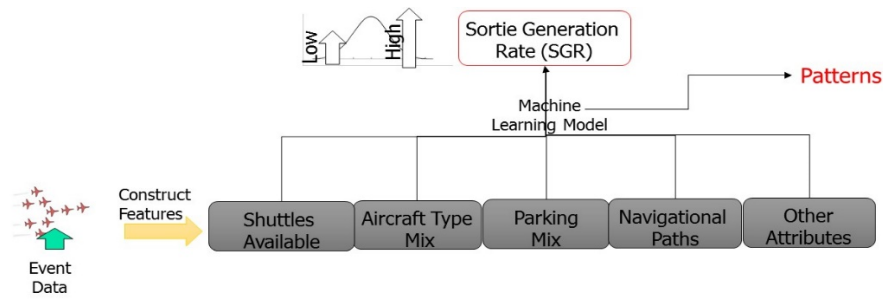
Figure 8: Conceptual Depiction of a Throughput Prediction Model

For the many events from the dataset, an event vector can be computed where the event vector includes the Shuttle usage, the AC mix, the parking places from where the ACs are sourced, the routes that the AC take, etc. The event vector represents the independent variables.

Next, we present some of the machine learning models that predict Sortie Generation Rate (SGR) and the patterns that they represent. While there are many modeling algorithms that are available to model the relationship between deck/domain attributes and throughput metrics such as SGR, we chose regression trees as the appropriate modeling algorithm primarily due to the easy interpretability of the modeling outputs. Classification And Regression Tree (CART) (Breiman et al. 1984) models were calibrated for different operational scenarios. For this problem, these tree structures get very complex and hence have to be run by domain experts to pick out those that are useful. Figure 9 shows the partial view of a decision tree. In this specific example, the sortie generation rate (SGR) is a function of the fighter F1 squadron aircraft depicting the split at the first node. If F1 is present, the tree branches to the right, then bifurcates at cargo aircraft. If F1 is not present, the tree branches to the left, then bifurcates at rotary wing HMS. Presence of F1 indicates higher sortie generation rate.
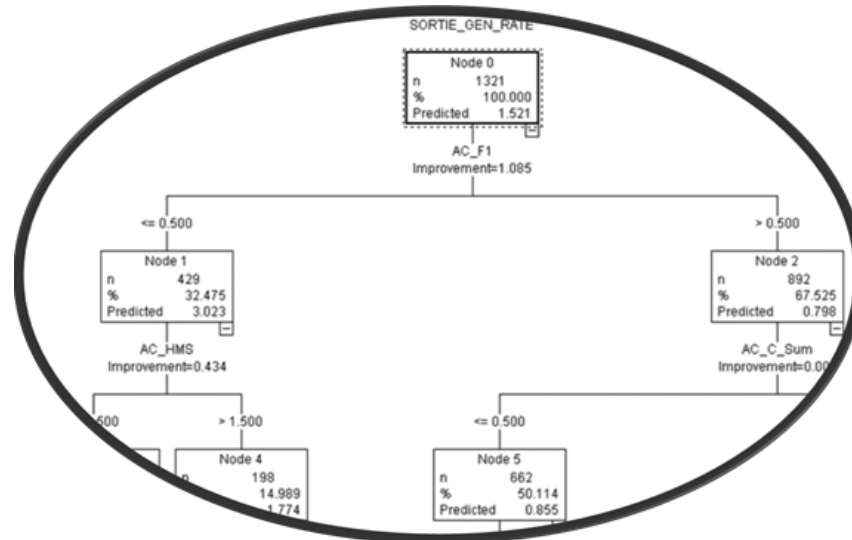
Figure 9: Decision Tree Rule Extraction

Another example of a decision tree for aircraft state changes is shown in Figure 10. For a closer look at how these rules are interpreted and validated, consider the following rule concerning "Launch" and "Observe.FixedWing.TakeOff" as the focus activity (Figure 10).

Precedent	Consequent	Rule Support	Rule Confidence
Arm Forward Firing Weapons.FixedWingTakeoff > TensionCat > FinalChecks > DelayBeforeLaunch	Launch	42.157	100
InstallLaunchBar > DirectToShuttle > RaiseJBD > Arm Forward Firing Weapons.FixedWingTakeoff	Observe.FixedWingTakeoff	42.25	100

Figure 10: Selected Rules

These rules can be elaborated in terms of domain language as follows:
- Row 1 Figure 10: When the following steps have been observed:
    - "Arm Forward-Firing Weapons for Fixed-Wing Takeoff"
    - "Apply Tension to Catapult"
    - "Final Checks"
    - "Delay Before Launch"
  the "Launch" step should be expected with 100 percent certainty

- Row 2 Figure 10: When the following steps have been observed:
    - "Install Launch Bar"
    - "Direct To Shuttle"
    - "Raise Jet Blast Door"
    - "Arm Forward-Firing Weapons For Fixed-Wing Takeoff"
  the "fixed-wing takeoff" step should be expected with 100 percent certainty

Next Steps

Given the portfolio of algorithms, visualizations and analytics, we are packaging those that are useful into a software tool that can be used by domain experts to explore the data set and to suggest operational improvements. We also believe this tool could be used for Validation and Verification (V&V). Our ultimate goal is to extract sufficient rules to embed into an autonomous aircraft. The rules engine will be probabilistic and will predict the next step, the autonomous aircraft should take. With the aid of additional sensors on-board the aircraft we expect to make the predictions more robust. We hope this technology will further the prospects of integrating unmanned vehicles into the current mix of manned aircraft.

This technology will scale to manage any asset moved in groups subject to complex spatio-temporal constraints. Some of the possible application domains such as depot maintenance, manufacturing shop floor, airport traffic management and even theme parks where people move in groups are our next target. In each of the target domains, the best use of this technology might be different. Unlike autonomous control on a carrier deck, in theme parks, we believe, tracking groups can help reduce ride congestion by giving groups incentives via a mobile app messages.

SUMMARY

TraceLogic -- Automated Rule Learning from Data Traces, has demonstrated methods, process and algorithms to decipher the hidden rules or *logic* of complex traffic patterns – *traces* of aircraft, people and equipment on Aircraft Carriers. Typical examples of data statistics, visual analytics, and rule and pattern analysis have been presented. Based on the success of these techniques a tool is being built to help analysts explore the data and improve the process. These tools could also support in Validation and Verification (V&V) for the rules engine of the simulator.

ACKNOWLEDGEMENTS

REFERENCES

Rochlin, Gene I., Todd R. La Porte, and Karlene H. Roberts. (1987). "The Self-Designing High-Reliability Organization: Aircraft Carrier Flight Operations at Sea." Naval War College Review. http://govleaders.org/reliability.htm, visited April 6, 2016

Breiman, L., J.H. Friedman, R.A. Olshen, and C.J. Stone. (1984). Classification and Regression Trees. Belmont, CA: Wadsworth.

David Barno, Nora Bensahel and M. Thomas Davis (2014). The Carrier Air Wing of the Future

US Navy (2016), *"Aircraft Carriers - CVN"*. *Fact File. United States Navy. 12 January 2016. Retrieved 13 April 2016.* http://www.navy.mil/navydata/fact_display.asp?cid=4200&tid=200&ct=4

CVN-78 (2013), *FY2013 Annual Report for the Office of the Director, Operational Test & Evaluation - CVN-78 Gerald R. Ford Class Nuclear Aircraft Carrier*