Artificially Intelligent Cyber Security: Reducing Risk and Complexity

John N. Carbone and James A. Crowder

1 Introduction: Non-Linearity and Complexity

Traditionally, much research exists for analysis, characterization, and classification of complex heterogeneous non-linear systems and interactions which historically have been difficult to accurately understand and effectively model [1, 2]. Systems that are nonlinear and dynamic generally comprise combinatorial complexity with changes in variables over time [3]. They may appear chaotic, unpredictable, or counterintuitive when contrasted with much simpler linear systems. Complex system interrelationships and chaotic behavior of these systems can sometimes be perceived as random. However, they are not. Simple changes in one part of a nonlinear system can produce complex effects. Widespread nonlinearity exists in nature as well [4]. Micro- and macroscopic examples include collision-based particles in motion; common electronic distortion; chemical oscillations; weather modeling; and, for this paper, complex relationships between vast volumes of seemingly ambiguous and superficially relatable binary cyber event data from a wide array of systems, users, and networks.

Additionally, advanced cyber research shows that meaningful enhancements are required to mitigate the ever-increasing volume of intrusion detection system (IDS) alerts overwhelming human analysts today, thousands of which are received per day, and 99% of which are false indications [5]. The currently well-known cyber security

J. N. Carbone
Department of Electrical and Computer Engineering, Southern Methodist University, Dallas, TX, USA
e-mail: john.carbone@forcepoint.com

J. A. Crowder (✉)
Colorado Engineering, Inc., Colorado Springs, CO, USA
e-mail: jim.crowder@coloradoengineering.com

H. R. Arabnia et al. (eds.), *Advances in Artificial Intelligence and Applied Cognitive Computing*, Transactions on Computational Science and Computational Intelligence, https://doi.org/10.1007/978-3-030-70296-0_38

landscape is wrought with many types of cyber security attacks where individuals and nation states employ a wide array of tactics to attack at every available data, network, and system access location. Mitigation research appears to be focused on using network-based intrusion detection systems and fusing their outputs to gain a more comprehensive understanding of undesired activities on the network [6]. While there has been some success, overall awareness of current network status and future adversarial action prediction has still not been achieved [6]. Although analysts are undeniably capable of performing difficult cyber security mitigation tasks, our understanding of the cognitive processes required by analysts, under current conditions, to produce more effective protection is very limited and requires new research [7, 8]. Therefore, as many complex problems generally involve more than one solution and as innovation is found at the intersection of multiple disciplines [9], our approach will be multi-, trans-, disciplinary to improve critical cyber situational awareness, classification, contextual understanding, and decision support [7]. We survey and heuristically decompose cyber security, axiomatic design, complexity theory, and novel new AI/ML/ITM learning techniques.

While many disciplines are known to have well-known mathematical formalisms for describing, predicting, and approximating solution;, historically, however, non-linear solution accuracy and analysis have generally been problem dependent and have required significant added effort to simplify and bound [10]. For example, many current cyber event classification-based techniques rely on an experts' extensive knowledge of network attack characteristics. Once provided to a detection system, an attack with a known pattern can then be more rapidly detected. Literature shows that the ability to mitigate a cyber event becomes highly dependent upon the fidelity of the attack's signature [11, 12]. Hence, systems often detect only what is known and are therefore significantly vulnerable to an environment of continuously adaptive attacks and vectors. Even if new attack signatures are rapidly incorporated for improving mitigation, the initial loss is irreplaceable and repair procedures costly. Mitigations are made even more difficult by the indeterminate validity of an inherently stochastic nonlinear system attack [13].

Recent advances in artificial intelligence (AI) and machine learning (ML) research are well known to be benefitting many disciplines struggling with rapidly increasing velocity, volume, and complexity of data and systems, and for improving timely generation of qualitative readily consumable knowledge [14]. However, AI/ML classification-based approaches generally rely on what is considered normal or standard baseline data traffic activity profiles generally built and stored over time. Comparative activities then explore for anomalies which deviate from captured baseline profiles. However, as vast volumes and high velocity binary data are individually captured, subsequent significant processing is then required to correlate/fuse potential complex non-linear inter- and intra-data relationships [15].

The difficult objective is to understand what is considered normal traffic, what is not already included in an ever-increasing ambiguous cyber knowledge store, and if an obscure event is nominal, a risk, or an attack. Today, this must all be accomplished within a given day's context challenged cyber environments where 99% of Intrusion Detection System (IDS) events can be potentially considered inadvertent false alarms [5]. Fortunately, the use of iterative methods, followed by

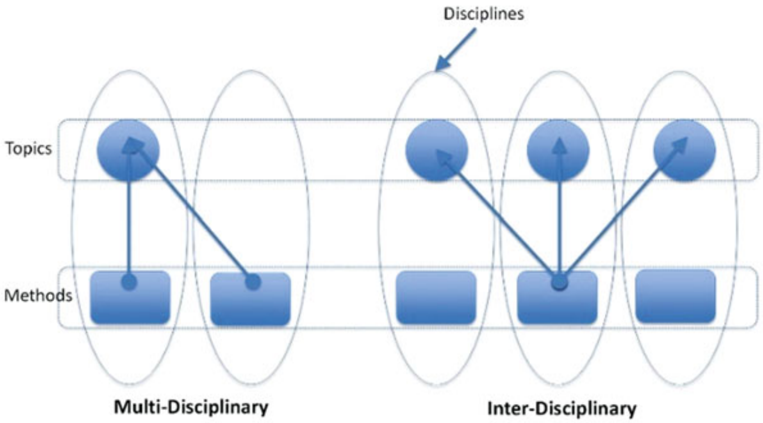

Fig. 1 Cross discipline engineering

domain−/discipline-specific solution sets, support effective solution generation for nonlinear problems. The addition of relatable data domains and data sources can potentially also add context to mitigating cyber threats [16, 17].

As part of our transdisciplinary approach, we therefore first employ axiomatic design's iterative design matrix optimization mechanisms to manage, understand, and simplify existing and future cyber systems, which generally comprise multifaceted processing methods and ambiguously related data. We also decrease cyber security risks by reducing information content with AD analysis processes, which in turn helps to reduce false positives and false negatives.

In Fig. 1, *multi-disciplinary* engineering involves engaging methods from other disciplines to solve a problem within one discipline, *inter-disciplinary* involves choosing methods from one discipline and applying it to one or more disciplines. In contrast, *transdisciplinary* engineering is the field of study which supports simplification and optimization by engineering common solutions across many disciplines [18]. Computer science, mathematics, and cyber security are examples of disciplines which provide capabilities for many disciplines. Even single-discipline engineering solutions can be complicated. Multi- and/or trans-disciplinary engineering solutions can increase the complexity and non-linearity even further, requiring more robust engineering principles and increased efforts for standard or more critical implementations.

1.1 Complexity

Exacerbating the problematic design of nonlinear systems are challenging levels of ambiguity and intricacy. Specifically, complex nonlinear relationships exist between vast volumes of seemingly ambiguous, independent cyber events and their potential relationships across multi-domain data. Additionally, modern manufacturing systems are increasingly required to adapt to changing market demands, their structural and operational complexity increases, creating a negative impact on system performance [19]. Similarly, cyber security systems suffer from increasingly adaptive adversaries [20] and must innovate in-kind to adapt to the asymmetric assault on system, data, and knowledge integrity. Significant research and patents exist to improve knowledge and context reliability with the correlation and fusion of big data [21]. Thus, the employment of complexity theory and applications of axiomatic design potentially decrease cyber system and data ambiguity and enable cyber security systems and their algorithms to become increasingly adaptive.

Axiomatic design (AD) research originated, in the 1990s with Nam P. Suh, within the Massachusetts Institute of Technology (MIT) Mechanical Engineering school. AD has been widely deployed for optimizing industrial and manufacturing applications. Complexity theory is applied for simplifying and optimizing system designs for mechanical, manufacturing, industrial, economic, social, and other systems. We propose that cyber security system designs can similarly benefit from AD ultimately improving cyber system adaptability and resiliency.

Suh describes that significant confusion exists within the definition of what is complex and explains that many more attempts have been made to understand complexity in terms of physical entities stead focusing on what is to ultimately be achieved. Suh describes complexity as computational, algorithmic, and probabilistic [3] and employs an approach comprising four complexity types: Time-independent real/imaginary complexity and time-dependent combinatorial/periodic complexity. Suh mitigates overall system complexity by employing optimizing actions: Reduce the time-independent real complexity, eliminate time-independent imaginary complexity where possible, and transform time-dependent combinatorial complexity into time-dependent periodic complexity for decomposing complex systems into smaller, more easily comprehensible, operable units. Suh describes this action as functional periodicity domain determination (e.g., biological, thermal, circadian, temporal, geometric, etc.).

1.1.1 Managing Complexity

Suh stipulates that complexity must be viewed within the functional domain. Therefore, fundamental management of complexity within a discipline, project, need, or gap is a focused process which defines what we want to achieve or understand within the functional domain. Managing complexity in the physical domain begins with creating system definitions using the voice of the customer known as customer needs (CN), which are subsequently translated into functional requirements (FR). To achieve optimized system design goals a set of design parameters (DP) is iteratively defined. DPs are analyzed for their individual probability of satisfying FR based system designs as shown in Fig. 2, probability density function (PDF). For example, if all required functional cyber system's DPs are completely encompassed within

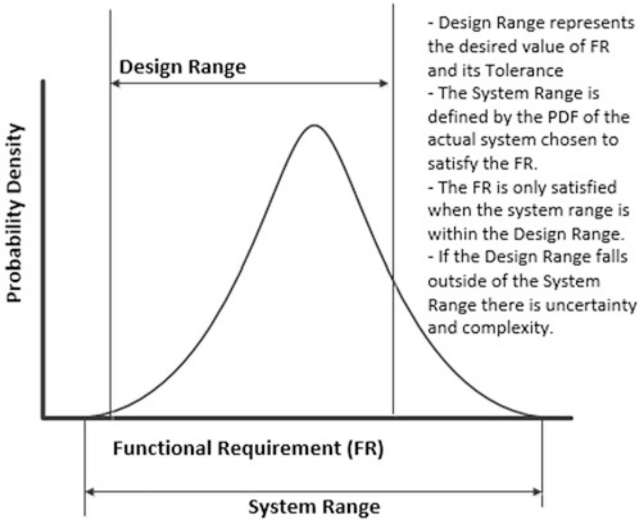
Design Range
Probability Density
Functional Requirement (FR)
System Range
- Design Range represents the desired value of FR and its Tolerance
- The System Range is defined by the PDF of the actual system chosen to satisfy the FR.
- The FR is only satisfied when the system range is within the Design Range.
- If the Design Range falls outside of the System Range there is uncertainty and complexity.

Fig. 2 Design vs. system PDF

the cyber system range, then we know that a cyber system's FRs are very likely to be achieved within a bounded design. Therefore, as an existing design or actively designed system becomes iteratively more well defined, it becomes less complex and less ambiguous. This is achieved through iterative FR-DP mapping and design matrix decoupling efforts where system optimization is driven through utilization of an independence axiom which drives orthogonality analysis. The ultimate design objectives are to achieve, as close as possible, a completely orthogonal uncoupled FR-DP design matrix, and the reduction of information content by minimizing DP random variations which reduce the variance of cyber system function and consequently reduces time dependent combinatorial complexity and non-linearity [3].

Moreover, when a system range extends outside of a design range boundary, a design and the satisfaction of functional requirements become more difficult to achieve and more complex. Axiomatic design processes achieve optimized functional designs through the aforementioned matrix decomposition process by traversing from "what we want to achieve" to "how we hope to satisfy" cyber system functionality across the following domains: customer, functional, physical, and process. Two important axioms, eluded to earlier, drive optimized cyber system functionality. First, the independence axiom helps maintain design independence and orthogonality of functional requirements and supports minimization of design overlap where possible to drive solutions to improve minimization and increase cost effectiveness. Second, the information axiom drives minimizing information content throughout the iterative design process to provide continuous simplification. The ultimate cyber system complexity reducing objective is complete uncoupled design relationships where all functional requirements are satisfied by independent design parameters and functions.

Analogously, for managing and securing cyber systems, software, and data, matrix decomposition provides logical decoupling of dependent interfaces and supports development of common normalized functional requirement inputs(what) and design parameter output(how) mappings. Consequently, upon design completion, physical and logical complex cyber system component tasks are effectively abstracted and simplified, and their data minimized, increasing successful understanding and improved correlation/fusion, thereby optimizing cyber sensing, improving scalability through orthogonal/independent design features, and reducing overall complexity and cost.

Analogously, supervised machine learning is a process, derived from statistical learning theory, also used for mapping inputs to outputs by learning unknowns from example pairs of related information. Hence, machine learning and axiomatic design are generally well aligned as they both support problem bounding and correlation-based discovery of unknowns. Each utilizes vectorized learning across a spectrum of widely varying characteristics. Comparably, empirical risk minimization (ERM) is a principle of statistical learning which helps bound algorithm performance and can be characterized by using joint probability distributions. Similarly, in axiomatic design, a system range correlates to a pre-defined bounded data range where a design range correlates to an actual range of each design component's datum. Simply, if a datum **X** represents a cyber event which occurred at time $\mathbf{t_1}$, while datum **Y** represents a cyber event occurrence at time $\mathbf{t_2}$, if Y occurred within the time range for cyber event **X**, then **Y** is within the range of **X** or $[(\mathbf{t_2}–\mathbf{t_1})$ and $(\mathbf{t_1} + \mathbf{t_2})]$, there exists a PDF which can represent the overlap of the system vs design range. Therefore, expanding upon the ML-AD parallels, the proceeding sections examine combining advanced novel machine-based learning techniques for increasing data correlation/fusion to further reduce cyber risk and complexity.

2 Artificially Intelligent Cyber

Discipline breadth is required for developing AI systems making AI research and education inherently multi- and trans- disciplinary [18, 22]. Machine learning is also supporting the evolution of cognition-based learning within these many domains [22, 23]. Cognition research employs computer software design analogous to components of the human brain combined with varying advances in artificial neural networks (ANN). ANNs are deemed one of the hallmarks of machine learning and designed to operate synonymously as neurons within a human brain, and as a group of interconnected nodal relations. ANNs were inspired by neuroscience-based biological neural networks which are the structure by which chemical-based ionic electrical signals are passed throughout the body [24]. They are therefore at the core of probabilistically relating changing levels of data relationships and recent active research supports their use in developing automated machine- based pattern recognition in general and for enhancing the cyber security domain [25, 26]. Thus, we describe the combined use of ITMs and cognitive-based learning methods to support the challenging processing of high volume, high velocity data, and improving opportunities for autonomous operations within an overwhelmed cyber user environment, minimal contextual security, and lacking system stability, thus improving understanding, transmitting less, and enabling individuals to more effectively utilize their cyber systems to better control their heterogeneous systems and data environments.

2.1 Cyber Data Ingest

Moving toward more autonomous self-learning operations requires a more intelligent, normalized, and optimized set of data ingress. The relatively standard extract transform load (ETL) tasks listed below, among others, are used throughout the data industry for processing passive and active streaming data (e.g., sensor data). Generally, learning from any data, including a cyber security scenario, involves a collection of each varying set of resting or streaming sensor data which is then compared, contrasted, associated, and normalized against previously captured content. The efficacy of algorithms being used to try and understand the data, relies on, among others, the ability to detect change. In short, these changes evolve into patterns which support improved learning fidelity over time. Ultimately these processes support learning continuously and evolving toward eventual autonomous learning and autonomous improvement of system functionality:

- Collect & verify data.
- Analyze for missing data.
- Cleanse data.
- Tag data.
- Organize/correlate/classify data.
- Compare, contrast, normalize with existing data.
- Verify, remember, store analytical results.
- Transmit/display results.

Traditionally, ETL functionality is also included for improving scaling of input source types (e.g., TCP/UDP Sockets, Filesystems) by understanding and classifying content prior to, and for improved, system processing. It is well known that current industry data volume, velocity, and variety vary greatly and can require significant processing for discovering patterns and context. However, a perception of high complexity exists in cyber security data, primarily because cyber data pattern analysis has traditionally been wrought with false positives and false negatives stemming from minimally included and minimally derivable context. This challenge increases the difficulty in determining nominal, unscrupulous, or accidental behavior from the many data, network, or user-based cyber events. As an example, a distributed denial of service (DDoS) attack is a high volume and velocity attack attempting to impact user services and is usually directed from many distributed locations. Understanding whether malicious or accidental is many times difficult. Other attacks: Phishing, JavaScript, SQL injections are generally small in data volume and velocity. This type of cyber data can be characterized as more passive (files per day) or active (streaming sensor data) but can still be difficult to classify. Important however is that correlation of high volume and/or passive data requires the proper infrastructure to support collection and processing of both.

As systems and sensors scale up or down, we therefore propose employing the logical and physical efficiency benefits of a well-known common ingest and processing architecture known as Lambda Architecture [27]. Lambda is well-known within many high-volume data architectures for processing individualized passive/batch and/or active/streaming data. Therefore, cyber sensor input data, like many varying data types, can be transformed per more well-defined flows and through a scalable orchestrated ETL process. This ensures proper a priori ingest curation and formatting required for subsequent cyber-based analytical processing algorithms, which is common for most modern ETL environments. Once data ingest analytics have curated and appropriately tagged the input, the resulting curated output is subsequently correlated, normalized, and/or fused with data within parallel streams of data and/or passive data including previous results. The algorithmically infused output is then believed to be of enhanced value (learning from the data) and becomes synonymous with terms like adaptive system learning, machine learning, and algorithmic domain terms (e.g., anomaly detection, cyber behavioral learning, intelligence, surveillance reconnaissance (ISR) fusion, molecular effect correlation, etc.).

2.2 Machine Learning and Cyber Security

It is well known that machine learning algorithms derive from statistical learning theory, are therefore inherently statistical in nature, and require significant initial training within a given bound or context to become qualitatively relevant. Similarly, finite element analysis within engineering disciplines and many mathematical concepts have historically supported non-linear solution accuracy and analysis by also providing problem-dependent bounding and thus simplification [10]. It is also well known that machine learning algorithm training, to be of consequence, is time consuming and generally considered a fine art to suitably discover and train with a sufficiently related problem data. Similarly, qualitative mitigation of cyber security risk requires proper human and data training to improve anomaly detection techniques and to adequately build normal activity profiles [28].

The effectiveness of the cyber algorithm training (as for most applications) then depends greatly upon the availability of completely normalized cyber traffic datasets, which, in practice, are rare, extremely difficult to keep up to date, and thus expensive to obtain, especially attack-free instances [5]. It is well known that for ML algorithms to be of benefit a significant amount of work must be achieved early in just understanding the data through a process of problem definition, data analysis, and data preparation. Understanding the context around the problem, constraints, assumptions, and motivations around who, what, where, and how the data was captured is critical to successful and useful subsequent modeling and application. A large pool of well-known machine learning algorithms and classification-based anomaly detection techniques are available today for computer vision, natural language processing, dimensionality reduction, anomaly detection, time series analysis, prediction, and recommender systems. Although multi-model ML shows promise in pattern analysis by creating ensemble outputs of multiple classifiers, herein we discuss traditional single model machine learning examples, their pitfalls, and subsequently propose information theoretic mechanisms to decrease cyber risk by significantly improving data-context understanding and thereby also improving autonomy when analyzing complex cyber systems.

2.2.1 Machine Learning: Value, Characteristics, and Limitations

Traditional machine learning is divided into two groups known as *Supervised* and *Unsupervised.* Additionally, we discuss our proposal for the additional application of AD. The objective is to drive more mindful initial and continuous interpretation of data, to help optimize development of common processing, simplification of well-defined data-analytics pairing, as well as, optimized frameworks for the improved processing of high volume, high velocity ML data.

ML processing flows can be resource intensive and can come in the form of *data collection exemplars, learning approximations, learning associations, striving for equality/specificity using sensitivity, and the use of optimization strategies.* Considering recent research into ML, what becomes apparent is that ML's usefulness is measured in a few different ways. ML algorithms are generally employed to sift through massive volumes of data looking for patterns. Some challenging characteristics of ML processing can include significant time consumption when compared against traditional data processing, ambiguous output, improved only with significant a priori data analysis, and the perceived complexity of data dependencies. Benefits of ML can include significant mitigation of the relative difficulty and/or inability of less automated approaches for determining discriminatory separation and classification of data.

We propose that in order to more fully understand the potential benefits and drawbacks of ML and in order to significantly improve valuable in-context affiliation between cyber data, one must account for the context of human interactions taking place between systems, data, algorithms, and applications. This includes capturing specific human cognitive states, and simultaneous and continuous correlation of all information, recursively. The implication is that "ones" and "zeroes" by themselves are most often analyzed "out-of-context" and hence provide much less is discernible meaning than when also compared to additive valuable contextual characteristics.

These value-based data characteristics become more visible when considering the subtle differences between the concepts of *presentation* and *representation*, well researched within the domains of information theory and physics [22] and used for improving human decision-making. *Representation* is simply defined as the underlying simple and complex relationships that are represented by mathematics, protocols, and formats. However, *representation* of simple and complex data and relationships between data are often not as readily discernible, when data relationships reach higher dimensions [29]. Hence, a potential use for ML pattern learning and algorithmic association designed specifically for this purpose.

Pattern analysis performance on high dimensional, most often initially unknown and unrelated data, is one key factor generally used to determine the value of an algorithm. Proper *presentation*/visualization of ML output is also required within the context of how data is used to support perceived benefits and/or drawbacks of a given method. These perceived benefits derive from the subjective quality of decision value derived from both, the a priori knowledge/context of the processed data, and the expression quality of the output. Perceived value inherently derives from how separable and discernible the data is and how well expressed the context around the data becomes after processing. As an example, parallel coordinates, as described by Inselberg, is a renowned method for discerning multi-dimensional data relationships through novel visualization for significantly improved decision making [22]. The objective of multi-dimensional visualization is to vastly improve the ability to perform multi-dimensional comparisons and context development whether simple or non-linear and then rapidly transform the visual presentation of higher order complex mappings into more simply discernible dimension reducing sets of two-dimensional relationships.

A quick review of ML: *Supervised* learning (SL), takes as input, data which has been previously learned and tagged, also known as "labeled." As an SL algorithm iterates over and processes each piece of new data, it compares each to previously learned data. As comparisons are made, exactly or partially, within a certain range or approximation of an a priori defined boundary, then the resulting response/decision variable is placed in the appropriate group. Hence, if the SL objective is to estimate the value of a given variable, then a simple "*curve fit approximation*" would be the recommended approach. If the SL algorithm is attempting to discover discrete categories within the data, then *Decision Tree* methods are preferred for organizing decisions, strategies, and possible outcomes.

Unsupervised learning (UL) treats all data equally and its prime directive is not to estimate the value of a variable but rather to simply look for patterns, groupings, or other ways to characterize the data that may lead to understanding of the way data interrelates. Hence, cluster analysis, K-means, correlation, hidden Markov models, and even factor analysis (principal components analysis) and statistical measures are examples of unsupervised learning. Unsupervised techniques essentially attempt a "partitioning" of a large pool of data into smaller chunks of data that are more related to members of the partition than data outside the partition. Different methods and different disciplines have varying names for this partitioning, whether for simple organization like concept maps or more algorithmic, like clustering, a term frequently used and commonly associated with methods such as (K-means). "Chunking" and "kriging" are terms for methods that handle the data differently but strive for the same organization. There are also methods such as Voronoi maps, Self-Organizing-Maps, Isoclines, Gradients, and many other approaches that also strive for separation of data into partitions without regard to what to call (labeling) members of each partition. It is this lack of requiring an external "label" as criterion which drives "unsupervised" partitioning.

Therefore, one of the most useful applications of statistical analysis is the development of estimators and function approximators (not models) to *visually* explain (*present*) the relationship between many data items (variables). Thus, many types of estimators have been developed (e.g., linear and nonlinear regression (function fitting), discriminant analysis, logistic regression, support vector machines, neural networks, and decision trees). However, Wulpert's no free lunch (NFL) theorem describes that each method has advantages only within a particular use case, and therefore, great care must be taken to understand each use case thoroughly [30].

ML algorithms focus upon similar varying optimization strategies where no single estimation or approximating method is best for all applications. Therefore, in order to improve the fidelity of machine-based learning and go further than current ML allows, we propose that it is uniquely important to understand the difference between memorizing (not accepted within academia as learning) discovered patterns and comparing/estimating how well those patterns compare/relate to recent and well-known information theory innovations and information theoretic methods (ITM) [31]. These methods strive to "explain" data at higher fidelity for rounding out the expression of details/context in every case, focusing upon enabling better realizations and decision-making. Analogously, imagine employing a new, more highly expressive method that enabled the realization of the exact location of electrons in space and time as opposed to Shroedinger approximations. The scientific applications and optimizations possible with this new expressive information would be considered revolutionary because of the wealth of new opportunities and vastly improved decision-making. Therefore, we discuss objectives to significantly improve understanding of ambiguous cyber security event data, optimization, and innovation, by providing higher fidelity-based information context and insights and for ultimately creating higher quality cyber security mitigation decision-making.

2.2.2 Currently Employed Supervised and Unsupervised Cyber Security Machine Learning Approaches

Most recently Ferrag et al. [32] analyzed 35 of the most well-known cyber security data sets with ten significant machine learning algorithms (Fig. 3 [32]) against performance efficiency for binary and multi-class classification, as well as accuracy, false alarm rate, and detection rate. Each data set-ML paired analysis yielded a best algorithm. FFDNN: Feed forward deep neural network; CNN: Convolutional neural network; DNN: Deep neural network; RNN: Recurrent neural network; DBN: Deep belief net- work; RBM: Restricted Boltzmann machine; DA: Deep auto-encoder; DML: Deep migration learning; STL: Self-taught learning; ReNN: Replicator neural network. Additionally, Eskin describes the use of unsupervised.

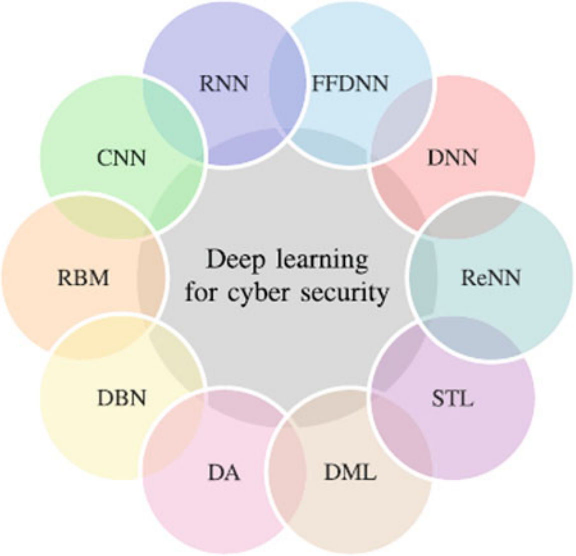

Fig. 3 Current cyber deep learning approaches [32]

SVM for detecting anomalous cyber events. Example SVM usage approaches: Registry anomaly detection (RAD) for monitoring Windows registry queries [33], Robust SVM (RSVM) anomaly detection ignores noisy data [34], confident anomaly detection [35]. A core assumption of SVMs is that all sample data for training are independently and identically distributed. Additionally, in ML practice, training data is often noisy, thereby often invalidating results and driving standard results into highly non-linear decision boundaries, leading to poor generalization. Research is also lacking to optimize and reduce anomaly detection and SVM runtime processing [35]; like many ML methods, SVM benefits and/or drawbacks are highly tied to a priori, well-defined boundaries and to the homogeneity of data mapped to equivalent small or large numbers of false detections, hence, our objective to provide improved machine learning fidelity and efficacy.

2.2.3 Improving Machine Learning Fidelity: Information Theory and Information Theoretical Methods

Vast systemic complexity and multi-dimensionality issues retard, impede, and provide significant friction to building advanced systems more capable of managing and protecting valuable commercial and government assets. Therefore, among a treasure trove of issues managing and understanding current and exponentially expanding Big Data and system endpoints, as well as, globally distributed computing, this chapter confronts the major issue regarding *data and system fidelity* along with prescribed solutions.

Historically, the fidelity of information content combined with *representation* and *presentation* clarity lends itself to improved insights and thus greater potential for improved decision-making and efficient actions [22]. In the complexity section above, the information and independence axioms were outlined expressly to provide background on mechanisms used today that support the optimization and simplification of information content relationships. The definition of fidelity is defined herein as the degree of exactness, accuracy, or precision. Historically, in biochemistry the ability to see and understand finer grained cellular interactions increases our ability to make improved decisions on curing disease or the manufacture of specifically targeted rather than generalized drugs. The lack of fidelity of understanding can result in dire consequences. In nuclear physics the quest for higher fidelity understanding of the universe drives the search for even smaller particles (e.g., Higgs Boson) which provide even more globally impacting insights. However, our systems, software, processing, and hardware are forever built by a third party and most often are not developed to a common standard. Software intra- and inter- dependencies are not effectively known or managed. Hence, the continuous lack of fidelity and lack of real-time insight into comprehensive system, system processing, application, and data dependencies is one of several core reasons we continue down a path of cyber frustration and insecurity. Therefore, along with employing axiomatic design for improving data and system understanding, and dependency mapping for simplifying system construction, we propose the application of advanced ITM approaches. The objective is to potentially achieve significantly higher fidelity system dependency and data anomaly detection/classification understanding for today's complex multi-dimensional data issues and cyber security challenges.

2.2.4 Cyber Data: Reducing High Dimensionality and Complexity of Machine Learning

Remember that data representation is defined as the underlying simple and complex relationships represented as collected, ingested, correlated, protocols, and data formats. Subsequent fusion of context from additive data sources increases relational complexity, ranging from clear simple comparisons to more ambiguous, higher dimensional, complex interrelationships. Stated earlier, parallel coordinates (PC) [28] provide for visually deriving clarity from complex higher dimensional relations. However, as in supervised learning, PC assume a priori parameterization has been normalized across data types, and initial interrelationships have been a priori defined. As stated above, SL algorithms iterate over and process each piece of new data, comparing each to previously learned data, thus iteratively adding complexity and higher dimensionality to data interrelationships.

Therefore, we propose an introductory data context leveling step where we apply axiomatic design principles for improving ML fidelity of a given SL application. First, AD should be applied to determine the type of complexity (e.g., time-dependent, time-independent, complexity, imaginary complexity, combinatorial complexity) [3] surrounding the creation of a collected data set driving a priori

SL data labeling and boundary approximation decision-making. AD supports the development of better indicators for unambiguously attributing truth to applied labels which increases data definition quality through added context and discernibility. Specifically, since SL processes perform data comparisons within specific ranges or approximations of a priori defined boundaries, AD's system design range correlation mechanism enhances understanding of the types of complexity which can drive SL range approximations. The added knowledge of which type of complexity was involved in data creation provides added insight into the value of a given SL algorithm on said data. Thus, this combined approach provides a candidate complexity reducing tool for reducing cyber and/or other data relationship complexities and improved efficacy of SL algorithm utilization.

2.2.5 Increasing Cyber Security Event Understanding with Information Theoretic Methods (ITM)

Novel Data Characterization Using Fractals

For understanding complex data and relationships, Jaenisch et al. [36] describe how to apply continuous wavelet transform (CWT) to discrete nth order integrable and differentiable data, and how to automate derivation of analytical functions which explicitly describe the data, directly from the data. They prove mathematically how to automate modeling of disparate data types using a new concept of univariate fluxion, coined as Unifluxion (UF) [36]. The UF formulation employs adaptive partitioning of data into fractals and then derives a fractal wavelet parent function for each partition. This enables automated wavelet transformation (integration or differentiation) across all partitions in a common repeatable manner, across the complete set of provided time series data [36]. Jaenisch et.al also compare UF to classical techniques and provide details on enhanced performance [36]. Hence, we propose that correlated and sequenced time series-based cyber, network, and user event data can be similarly described and modeled using fractals and UF.

Jaenisch et al. show how the unique formulation of $U(f(x))$ enables an automated data model transformation into either an integral or differential model of any order (an automatically derived differential equation) [36]. They describe how UF is defined to be a data model because it provides a continuous predictive model that can be both integrated and differentiated to any order. UF is also derived incrementally, as each measurement point is piecewise collected, although the final result is a continuous analytical function across all the time series data [36].

Subsequently, adding credence to the use fractals for improved cyber understanding, Jaenisch et al. [37] provide research examining the hypothesis that decision boundaries between malware and non-malware is fractal. They characterized the frequency of occurrence and distribution properties of malware functions compare them against non-malware functions [37]. They then derived data model–based classifiers from identified features to examine the nature of the parameter space classification boundaries between families of existing malware and the general non-malware category. Their preliminary results strongly supported a fractal boundary hypothesis based upon analyses at the file, function, and opcode levels [37].

Security information and event management (SIEM) data and systems are used extensively throughout industry, incrementally capturing many system-wide time series events: Data, Network, User Behavioral, Endpoint, Email, Web, etc. Additionally, security analytics in big data environments present a unique set of challenges, not properly addressed by the existing SIEM systems that typically work with a limited set of those traditional data sources [38]. Thus, we propose the reasonable application of UF to cyber SIEM time series data. UF's decomposition of cyber event content into fractal partitions enables the application of wavelet transformation which increases fidelity of cyber inter-, intra-, data relationships through the derived analytical functions which explicitly describe the cyber data, directly from the cyber data. Simultaneously, UF's continuous analytical function provides rapid predictive integration and/or differentiation, thereby improving the speed of cyber event relationship prediction and understanding. Hence, potentially traditional time-consuming machine learning classification would not be necessary (k-means, SVMs, etc.)

Spatial Voting (SV) for High Fidelity Data Characterization

Spatial voting is a multidimensional clustering and grouping algorithm. Individual spatial measurements (e.g., latitude and longitude) are stacked onto a coarse resolution SV grid. Similar or closely related points are organized into the same or within neighboring cell locations on the SV grid. The input features for the SV grid form the x and y axes of the SV grid. Once measurements are stacked on the SV grid, if required, a 2-D spatial convolution kernel is used to smooth the stack values in the landscape and connect isolated regions together into regions (subgroups) [39]. As proposed by Jaenisch et al., SV provides an analog data modeling approach to provide a solution to the "object to assembly" aggregation problem [40]. Generally, "object to assembly" refers to the perception of objects within their given spatial relations. For example, information design in advanced augmented reality (AR) applications requires support for assembly with high demands of spatial knowledge. SV is based upon combined principles of voting, geometry, and image processing using 2D convolution [40]. Voting is defined as a democratic process where the majority decision rules. Votes equate to hard decisions. For example, if sensors observe a phenomenon and identify it and then rank based upon each different hypothesis, then as one sums the number of sensors that declare a hypothesis to be true, then the largest sum becomes the winner [40]. Hence, voting reduces to probabilities, and typically, this is where Bayesian and other probabilistic analysis methods are generally used [40]. A presentation summary of the SV process is shown in Fig. 4. A shows an example plots of.

spatial events(e.g., UAV locations), B shows the detection grid output after initial ellipse stacking has been performed, C depicts the identification relationships of the candidate sub-frames, D represents a graphic output after additional feature

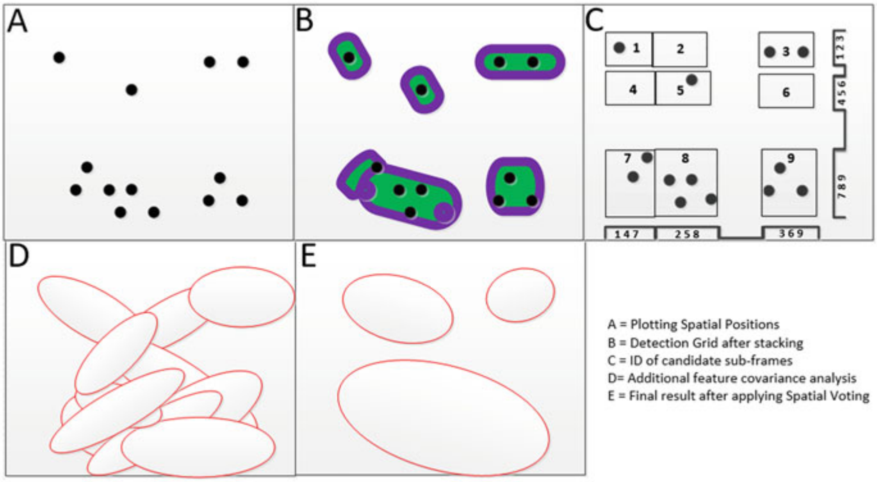
A
B
C
1
2
3
4
5
6
7
8
9
123
456
789
147
258
369
D
E
A = Plotting Spatial Positions
B = Detection Grid after stacking
C = ID of candidate sub-frames
D= Additional feature covariance analysis
E = Final result after applying Spatial Voting

Fig. 4 Spatial voting process summary

covariance analysis has been performed (e.g. sensor reporting locations), and E depicts the final results after spatial voting was applied. It should be noted that additional intermediary feature analysis can be applied to further the exactness of the final output.

Hence, the difference is that conceptually SV and ITMs strive to "*explain*" rather than "*learn*" flash card style as machine learning aka machine remembering does. The emphasis is that memorization is not considered in academia as a measure of learning at all. Learning requires understanding which requires insight, the ability to synthesize (i.e., restate differently), and then generalize to conclusions. Neither machine learning nor artificial intelligence is focused on a path to achieve this. Hence, SV employs a matrix/grid where spatial artifacts are captured and marked/classified based upon their unique identifying characteristics [40]. Objects are characterized using recursive forms of higher order highly descriptive parametric and non-parametric (fractal) features and statistics, and the classifiers are derived for discrimination and classification. Characterizations and formulations are highly object and situation dependent. The next step is to create a classifier to practically associate the individual features. As always classifiers and algorithms which support it must continue to be chosen carefully. Their research has shown that SV provides significant characterization, discrimination, and performance benefits for improving object context in physical operational environments [36, 37, 40]. In the proceeding sections, we will show that spatial constructs can be extended to high fidelity explanation and characterization of varying types of digital artifact interrelationships and highly contextual knowledge creation.

High Fidelity Object/Data Relationship Modeling

When very few examples are available to discern statistical behavior, data models can be employed. Generally, data models are constructed using a bottom-up approach in the form of a knowledge base to collect, capture, identifies, encodes each example encountered. Data models within systems where little intelligence initially exists and into which intelligence is injected can contain many data models. Of course, these myriads of models have varying structures, formats and datum, all created under the influential context of the data model author. Relating similar and alien/abstract models and context has historically been challenging for many reasons. Hence, an inductive self-organizing approach, polynomial neural networks (PNN) or group method of data handling (GMDH), has been applied in a great variety of areas for deep learning and knowledge discovery, forecasting and data mining, optimization, and pattern recognition [41]. GMDH algorithms provide value in automatically finding interrelations in data, in order to select an optimal model or network structure and to increase algorithm accuracy. GMDH focus is to minimize modeling influence of the author and enables the computer to find optimal model structure or laws of the system acting upon the data.

Extending Spatial Constructs to System Learning and System Knowledge Development

Improving decision quality autonomy of artificially intelligent systems requires reliable information discovery, decomposition, reduction, normalization, and context-specific knowledge recall [22]. Hence, capturing the essence of any given set of information content is paramount. When describing how science integrates with information theory, Brillouin [42] defined knowledge as resulting from exercising thought. Knowledge was mere information without value until a choice was made based upon thought. Additionally, Brillouin concluded that a hundred random sentences from a newspaper, or a line of Shakespeare, or even a theorem of Einstein have exactly the same information value. He concluded that information content had "no value" until it had been thought about and turned into knowledge.

Artificially infused robotic systems must be able to integrate information into their cognitive conceptual ontology [43] in order to be able to "think" about, correlate, and integrate information. Humans think to determine what to do or how to act. It is this decision-making that can be of great concern when processing ambiguity because of the sometimes-serious ramifications which occur when erroneous inferences are made. Often there can be severe consequences when actions are taken based upon incorrect recommendations. Inaccurate inferences can influence decision-making before they can be detected or corrected. Therefore, the underlying challenge is to reliably understand the essence of a situation, action, and activity and to significantly increase capability and capacity to make critical decisions from a complex mass of real-time information content. Harnessing actionable knowledge from these vast environments of exponentially growing structured and unstructured sources of rich interrelated cross-domain data is imperative [44] and a major challenge for autonomous systems that must wrestle with context ambiguity without the advantage of human intervention [23]. The next section comprises combining ITMs with enhancing understanding ambiguous characteristics using knowledge relativity threads (KRT) [22]. As SV is employed to "explain" spatial characteristics, KRTs extend mature physical spatial mechanics for defining adaptive knowledge object presentation and representation.

2.2.6 Physical Representation of Meaning

Research shows that the community of disciplines researching how humans generate knowledge has traditionally focused upon how humans derive meaning from interactions and observations within their daily environments, driving out ambiguity to obtain thresholds of understanding. With similar goals, spatial voting, information theory, and complexity theory, as described earlier, focus more closely on explaining actual information content. Zadeh pioneered the study of mechanisms for reducing ambiguity in information content, informing us about concepts in "fuzzy logic" and the importance of granular representations of information content [45], and Suh focused upon driving out information complexity via the use of axiomatic design principles [3]. Hence, a vast corpus of cognitive-related research continually prescribes one common denominator, representation of how information content, knowledge, and knowledge acquisition should be modeled. Gardenfors [46] acknowledges that this is the central problem of cognitive science and describes three levels of representation: symbolic—Turing machine like computational approach; associationism—different types of content relationships which carry the burden of representation; and thirdly, geometric—structures which he believes best convey similarity relations as multi-dimensional concept formation in a natural way; learning concepts via similarity analysis has proven dimensionally problematic for the first two and is also partially to blame for the continuing difficulties when attempting to derive actionable intelligence as content becomes increasingly distended, vague, and complex.

Historically, there are many examples and domains, which employ concepts of conceptual representation of meaning as geometric structures (e.g., cognitive psychology [47], cognitive linguistics [48–50], transdisciplinary engineering [22], knowledge storage [14], computer science, e.g., entity relationship, sequence, state transition, and digital logic diagrams, Markov chains, neural nets, and many others. It should be noted here that there is not one unique correct way of representing a concept. Additionally, concepts have different degrees of granular resolution as Zadeh [45] describes in the fuzzy logic theory. However, geometric representations can achieve high levels of scaling and resolution [46] especially for n-dimensional relations, generally difficult if not impossible to visualize above the fourth dimension. However, high dimensionality can be mathematically represented within systems in several ways. Hence, mature mathematics within the physical domain allows this freedom. Therefore, we show the overlay of physics-based mathematical characteristics to enhance relational context and develop a unifying underlying knowledge structure within information theory. We employ knowledge relativity threads (KRT) [22] to minimize ambiguity by developing detailed context and for conveying knowledge essence simply and robustly. The next section describes presentation formation, representation, and the process of organization of n-dimensional contextual relationships for humanistic prototypical object data types with application of the common denominators: time, state, and context.

2.2.7 Knowledge Relativity (KR)

Knowledge relativity threads (KRT) [22] primarily originate from computational physics concepts as an analogy to Hibeller [51] where the concept of relating the motion of two particles is a frame of reference and is measured differently by different observers. Different observers measure and relate what they behold, to a context of what has been learned before and what is being learned presently. The reference frame of each knowledge building action contains the common denominators of time, state, and context—the point in time and all the minutia of detailed characteristics surrounding and pronouncing the current captured state of all related context. Historically, organization, presentation, representation of knowledge, and context have been researched across many disciplines (e.g., psychology, computer science, biology, and linguistics) because of the primal need to survive, understand, and make sense of a domain. However, most systems we engineer today are increasingly incapable of processing, understanding, presenting, and structurally representing the volume, velocity, variety, and complexity of content because first, they are not built to learn, only to store [52], and second, the content systems store and filter are what is generally or explicitly known to be true, not the more valuable and higher fidelity tacit knowledge that is context specific to each frame of reference or situation [53].

Therefore, we build KRTs upon the concept of "occam learning" [54] to construct continually renegotiable systems [55] with the simplest portrayal (e.g., present and represent) capable of encapsulating complex n-dimensional causal structures, within and between the complex data generated from the observed/captured behavior [14].

The KRT concept was developed to take advantage of mature physical universe n-dimensional relationship calculations relating any celestial object to another regardless of size or composition. KRTs extend physics space-time mathematics and apply to information theory to increase contextual knowledge understanding through a concept of recombinant knowledge assimilation (RNA) [22] or recursive spatial data model representations consisting of information object relationships. The logical concept develops from the following:

- *An infinite amount of data and data relationships exists in the universe.*
- *The infinite amount of data doesn't increase or decrease; it simply changes in form (knowledge increases).*

- *Fundamental increases in data volume increases decision points, which fundamentally should result in, but does not guarantee, increased data maturity and increased quality decision making.*
- *As data is consumed or processed, an increase in data quality and maturity appears, if and ONLY if enough relationships are known or can be discerned, can be captured, and reused to inform.*
- *Information explosion has always provided a human challenge.*
- *Humans, by themselves, are not physically capable of rapidly comprehending vast complex data sets, and then providing associated solutions to complex problems, the human brain can only handle approximately 7 events at a time [56].*
- *Can we capture and establish understanding of the fundamental relationships among all types structured and unstructured data?*
- *Can we extend or abstract premier concepts used to capture physical universe relationships?*

2.2.8 High Fidelity Fusion Using Concepts of Space-Time

In the general theory of relativity, the relationship depends upon the observer. This is similarly the case for fusing data relationships per Joint Directors of Laboratories (JDL) Fusion Level 5 [57], where the user serves a primary observer role in support for "decision-making" and defining actionable relationships. Different observers of the same data may apply different relationships. This is not dissimilar in Einstein's theory of special relativity where it is demonstrated that a "correct" answer is measured differently by different observers [58] for any independent event. Therefore, any observer of n-data of n-types can have n-independently observable relationships. Today, complex systems of data stores are developed from significant research across many different scientists/observers. Research data considered mundane or unrelated to one observer might be the ultimate piece of the puzzle or major discovery for another.

Hence, beginning with Reimann, space-time mathematics with respect to relativity has been in development for more than a century. Extending n-dimensional relationship mathematics principles to correlate and fuse non-physical data seemed intuitive. Here we describe Hendrik Lorentz's and Schrödinger's use of manifolds [59, 60] for application to n-related data objects in a linear or non-linear space. In systems biology, a cell is made up of many things. A strand of deoxyribonucleic acid (DNA) is made up of numerous bits and pieces of information which define a genetic blueprint, as well as, the DNA helix like physical shape. A space-time object in a Lorentzian manifold can be defined by the tangent vectors or signatures on the curved manifold. These objects can be represented as tensors or metric tensors which comprise a vector of eigenvalue attributes which define an object signature in n-dimensions [59]. Multivariate analysis in n-dimensions is applied to aspects of complexity in the information theory as well [32, 39]. Employing tensors as vector relationships to systems biology data, we can describe DNA attribute relationships mathematically and can store them in a common manner. Their pedigree is maintained mathematically as a tensor attribute per data element. Metric tensors or tangent vectors where a manifold intersects a spheroid at a single point represent the data that describes the attributes of the intersection. Hence, if the spheroid space represents a locale where data exists and the point on the curved surface is a datum, then the vector of coefficients and attributes reflects the characteristics of that datum.

Lorentzian manifolds also have the concept of causal structure. Causal structure in space-time can describe ordering, past and future, designated by the directionality of each tensor vector. Consequently, this can also be applied to digital data. For example, biology can have time-dependent ordering when describing the time lapse yeast growth characteristics in cell array experimentation, just as described earlier in Unifluxion and spatial voting time series analysis. Hence, metric tensors or "vectors of knowledge" intersect at their point of relation, Fig. 5. Hence, as one vector's directionality disperses or moves away from another vector, one can logically deduce that the strength of that relationship decreases or increases. Therefore, Newton's law of gravitation is used as an analogy to compare, contrast, associate, and normalize representations of relationship strength for any type information (analog or digital) artifact type. Figure 6 depicts an example representation of two information objects being compared, contrasted, and associated based upon user and/or system definable characteristics: importance, closeness. Newtonian gravity defines that the force of two bodies is proportional to their mass and inversely proportional to the square of their separation. The calculation is then multiplied by the universal gravitational constant to achieve the final force of gravity result between two bodies. For application to information theory, Newtonian mass is extended to denote relative information *importance* provided by a person or system as it pertains to a specific knowledge object to knowledge object comparison or any other smaller or larger context. The importance measure is a user r system defined scalar diameter in order to provide relative radius to the separation denominator. The square of separation is analogous to how close (*closeness*) the two pieces of information are relative to the overarching context. Figure 6, knowledge object (KO) #1, depicts the concept of two internal sub-nested objects of information which if reviewed would show additional context for KO1. The attractive force, $\boldsymbol{A}$, of the two pieces of information in Fig. 6 is shown to equal 10. Lastly, a user or system can also employ a balance variable as an analogy to the Newtonian universal gravitational constant multiplier. This type of constant is considered a balance factor/variable of proportionality would be dependent upon user or system situational context which KO1 and KO2 are part of.

.

2.2.9 Conclusions and Discussion

This is preliminary work and significant research is still required. Here we have presented adaptive learning methods for enhancing cyber security risk reduction

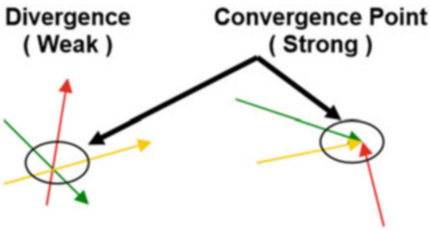

Fig. 5 Vector-based data relationships

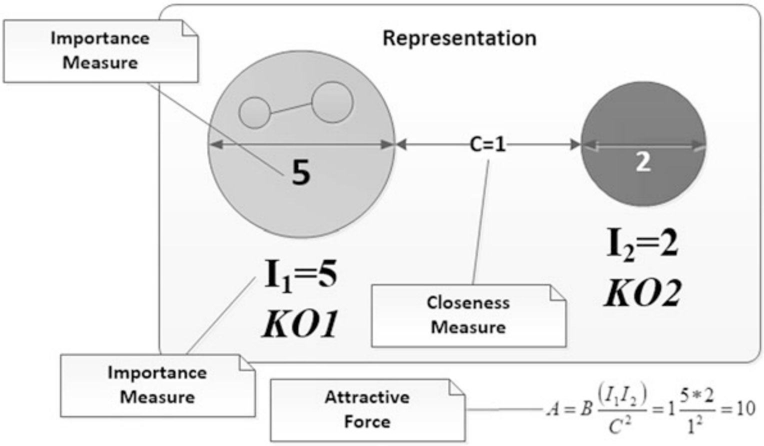

Fig. 6 Knowledge relativity threads

by improving knowledge density (KD) (how much do we know about a given event or subject) and knowledge fidelity (KF) (how well do we know) to improve cyber event context and decision quality for improved and more autonomous action. Axiomatic design axioms and design features were recommended for adding to our KD data analysis prior to machine learning application. High fidelity Unifluxion fractals, spatial voting research, and improved performance analysis of time series data was provided. Spatial voting was shown to also operate against cyber digital time series data (e.g., malware detection) and their benefit comparisons to cyber machine learning data classification was also provided.

Tensor vector data and knowledge relativity threads (KRT) were shown to provide spatial constructs for explaining high fidelity relationships and the use of manifolds and metric tensor vectors as attribute descriptors was also described. These methods were combined with axiomatic design concepts to organize and supply complexity reduction techniques for reducing traditional time-consuming machine learning classification. Together, it was shown that these capabilities potentially produce support for more efficient decision actuation, due to improved explanation and relationship context of data and higher data analysis performance, thereby providing added insights to cyber systems and analysts to reduce security risk and reduced non-linearity and complexity. Suggested next steps should prototype, design, and implement an architecture to learn on large cyber datasets. Future papers will present progress and results as available.

References

1. S.S. Zhou, G. Feng, C.B. Feng, Robust control for a class of uncertain nonlinear systems: adaptive fuzzy approach based on back- stepping. Fuzzy Sets Syst. **151**(1), 1–20 (Apr. 2005)
2. W.S. Yu, C.J. Sun, Fuzzy model based adaptive control for a class of nonlinear systems. IEEE Trans. Fuzzy Syst. **9**(3), 413–425 (2001)
3. N. Suh, *Complexity Theory and Applications* (Oxford University Press, 2005)
4. G. Nicolis, *Introduction to Nonlinear Science, DI-Fusion* (Cambridge University Press, 1995)
5. J.R. Goodall, W.G. Lutters, A. Komlodi, I know my network: collaboration and expertise in intrusion detection, in *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work*, ed. by J. Herbsleb, G. Olson, (ACM, New York, 2004), pp. 342–345
6. N.A. Giacobe, Application of the JDL data fusion process model for Cyber Security. *Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2010*, vol. 7710. International Society for Optics and Photonics (2010)
7. P.C. Chen, P. Liu, J. Yen, T. Mullen, Experience-based cyber situation recognition using relaxable logic patterns. *In Proceedings of the 2012 IEEE international multi-disciplinary conference on cognitive methods in situation awareness and decision support (CogSIMA)*, pp. 243–250, IEEE (2012)
8. A. Joinson, T. van Steen, Human aspects of cyber security: behaviour or culture change? Cyber Secur. Peer-Reviewed J. **1**(4), 351–360 (2018)
9. S.A. Zahra, L.R. Newey, Maximizing the impact of organization science: theory-building at the intersection of disciplines and/or fields. J. Manag. Stud. **46**(6), 1059–1075 (2009)
10. D.V. Hutton, *Fundamentals of Finite Element Analysis* (McGraw-Hill, 2017)
11. A. Aziz, Prospective client identification using malware attack detection. U.S. Patent No. 9,027,135. 5 May 2015
12. D. Clark, J. Strand, J. Thyer, Active attack detection system. U.S. Patent No. 9,628,502. 18 Apr. 2017
13. S. Liu, G. Wei, Y. Song, Y. Liu, Extended Kalman filtering for stochastic nonlinear systems with randomly occurring cyber-attacks. Neurocomputing **207**, 708–716 (2016)
14. J. Crowder, J. Carbone, *The Great Migration: Information to Knowledge Using Cognition-Based Frameworks* (Springer Science, New York, 2011)
15. I. I. Liggins, D. H. Martin, J. Llinas (eds.), *Handbook of Multisensor Data Fusion: Theory and Practice* (CRC Press, 2017)
16. G. Bello-Orgaz, J.J. Jung, D. Camacho, Social big data: recent achievements and new challenges. Inform. Fusion **28**, 45–59 (2016)
17. D. Quick, K.K.R. Choo, Digital Forensic Data and Open Source Intelligence (DFINT+OSINT). In: *Big Digital Forensic Data. Springer Briefs on Cyber Security Systems and Networks*. Springer, Singapore (2018)
18. A. Ertas, M.M. Tanik, T.T. Maxwell, Transdisciplinary engineering education and research model. J. Integr. Design Proc. Sci. **4**(4), 1–11 (2000)
19. P. Nyhuis (ed.), *Wandlungsfähige Produktionssysteme* (GITO mbH Verlag, 2010)
20. R. Colbaugh, K. Glass, Predictability-oriented defense against adaptive adversaries. *Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference on. IEEE* (2012)
21. J. Lee, B. Bagheri, H.-A. Kao, Recent advances and trends of cyber-physical systems and big data analytics in industrial informatics. *International proceeding of int conference on industrial informatics (INDIN)* (2014)

22. J. Carbone, A framework for enhancing transdisciplinary research knowledge. Texas Tech University (2010)
23. J.A. Crowder, J.N. Carbone, S.A. Friess, *Artificial Cognition Architectures* (Springer, New York, 2014)
24. J. Crowder, S. Friess, Artificial neural diagnostics and prognostics: self-soothing in cognitive systems. *Proceedings of the 12th annual International Conference on Artificial Intelligence*, Las Vegas, NV (2010)
25. W. Liu et al., A survey of deep neural network architectures and their applications. Neurocomputing **234**, 11–26 (2017)
26. S.S. Roy, et al., A deep learning based artificial neural network approach for intrusion detection. *International Conference on Mathematics and Computing*, Springer, Singapore (2017)
27. N. Marz, J. Warren, Big data: principles and best practices of scalable real-time data systems. Manning (2013)
28. S. Sridhar, M. Govindarasu, Model-based attack detection and mitigation for automatic generation control. IEEE Trans. Smart Grid **5**(2), 580–591 (2014)
29. A. Inselberg, Parallel coordinates, in *Encyclopedia of Database Systems*, (Springer, Boston, 2009), pp. 2018–2024
30. D.H. Wolpert, W.G. Macready, No free lunch theorems for optimization. IEEE Trans. Evol. Comput. **1**(1), 67–82 (1997)
31. K.P. Burnham, D.R. Anderson, Practical use of the information-theoretic approach, in *Model Selection and Inference*, (Springer, New York, 1998), pp. 75–117
32. M.A. Ferrag et al., Deep learning for cyber security intrusion detection: approaches, datasets, and comparative study. J. Inform. Secur. Appl. **50**, 102419 (2020)
33. K.A. Heller, et al., One class support vector machines for detecting anomalous windows registry accesses. *Proc. of the workshop on Data Mining for Computer Security*, vol. 9 (2003)
34. W. Hu, Y. Liao, V. Rao Vemuri, Robust Support Vector Machines for Anomaly Detection in Computer Security. ICMLA (2003)
35. I. Balabine, A. Velednitsky, Method and system for confident anomaly detection in computer network traffic. U.S. Patent No. 9,843,488. 12 Dec. 2017
36. H.M. Jaenisch, J.W. Handley, N. Albritton, Converting data into functions for continuous wavelet analysis. *Independent Component Analyses, Wavelets, Neural Networks, Biosystems, and Nanoengineering VII*, vol. 7343. International Society for Optics and Photonics (2009)
37. H.M. Jaenisch, et al., Fractals, malware, and data models. Cyber Sensing 2012, vol. 8408. International Society for Optics and Photonics (2012)
38. R. Zuech, T.M. Khoshgoftaar, R. Wald, Intrusion detection and big heterogeneous data: a survey. J. Big Data **2**(1), 3 (2015)
39. H. Jaenisch, Spatial voting with data modeling for behavior based tracking and discrimination of human from fauna from GMTI radar tracks. *Unattended Ground, Sea, and Air Sensor Technologies and Applications XIV*, vol. 8388. International Society for Optics and Photonics (2012)
40. H.M. Jaenisch, et al., A simple algorithm for sensor fusion using spatial voting (unsupervised object grouping). *Signal Processing, Sensor Fusion, and Target Recognition XVII*, vol. 6968. International Society for Optics and Photonics, 2008
41. T. Aksenova, V. Volkovich, A.E.P. Villa, Robust structural modeling and outlier detection with GMDH-type polynomial neural networks. *International Conference on Artificial Neural Networks*. Springer, Berlin, Heidelberg, 2005
42. L. Brillouin, *Science and Information Theory* (Dover, 2004)
43. J. Crowder, V. Raskin, J. Taylor, Autonomous creation and detection of procedural memory scripts, in *Proceedings of the 13th Annual International Conference on Artificial Intelligence*, (Las Vegas, 2012)
44. J. Llinas, et al., Revisiting the JDL data fusion model II. *Space and Naval Warfare Systems Command San Diego CA* (2004)

45. L.A. Zadeh, A note on web intelligence, world knowledge and fuzzy logic. Data Knowl. Eng. **50**(3), 291–304 (2004)
46. P. Gärdenfors, *Conceptual Spaces: The Geometry of Thought* (MIT Press, 2004)
47. P. Suppes, Current directions in mathematical learning theory, in *Mathematical Psychology in Progress*, (Springer, Berlin, Heidelberg, 1989), pp. 3–28
48. R.W. Langacker, *Foundations of Cognitive Grammar: Theoretical Prerequisites*, vol 1 (Stanford University Press, 1987)
49. G. Lakoff, Z. Kövecses, The cognitive model of anger inherent in American English, in *Cultural Models in Language and Thought*, Cambridge University Press, (1987), pp. 195–221
50. L. Talmy, Force dynamics in language and cognition. Cogn. Sci. **12**(1), 49–100 (1988)
51. R.C. Hibbeler, *Engineering mechanics* (Pearson Education, 2001)
52. D. Ejigu, M. Scuturici, L. Brunie, Hybrid approach to collaborative context-aware service platform for pervasive computing. JCP **3**(1), 40–50 (2008)
53. I. Nonaka, H. Takeuchi, *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation* (Oxford University Press, 1995)
54. M.J. Kearns, U.V. Vazirani, U. Vazirani, An *Introduction to Computational Learning Theory* (MIT Press, 1994)
55. T. Gruber, Collective knowledge systems: Where the social web meets the semantic web. J Web Semantics **6**(1), 4–13 (2008)
56. J.C. Platt, Fast training of support vector machines using sequential minimal optimization, in *Advances in Kernel Methods*, MIT Press, Cambridge, MA, (1999), pp. 185–208
57. E.P. Blasch, S. Plano, JDL Level 5 fusion model: user refinement issues and applications in group tracking, SPIE Vol. 4729, Aerosense (2002)
58. A. Einstein, Relativity: the special and general theory: a popular exposition, authorized translation by Robert W. Lawson: Methuen, London (1960)
59. A. Hendrik Lorentz, Considerations on Gravitation. In: *KNAW, Proceedings*, 2, 1899–1900, Amsterdam (1900)
60. M.S. Alber, G.G. Luther, J.E. Marsden, Energy Dependent Schrodinger Operators and Complex Hamiltonian Systems on Riemann Surfaces, August 1996