Autonomous Data & Goal Driven Intelligent Inspection for Unknown Precursor Detection

Christopher Bowman, PhD
Data Fusion & Neural Networks (DF&NN)
cbowman@df-nn.com
Introduction

• This brief focuses on *data-driven and goal-driven computational intelligence (CI) systems* that
  – Learn normal behavior so as to affordably detect & characterize unexpected abnormal conditions
  – Learn to adapt behavior to achieve user defined goals
• These types of CI systems are more affordable than traditional model-driven systems for hard problems needing to discover the unknown-unknowns
• A [technical architecture](#) for the development of Data Fusion & Resource Management (DF&RM) CI systems and a [TRL 7 CI system implementation](#) is described
Current Data & Goal-Driven Capability Summary

- Affordably detect abnormal activity patterns of interest in dynamic ‘big data’
- Provide turn-key intelligent data-driven and goal-driven systems
- TRL 7 system delivered to 3 sites that automatically learns normal activities in ‘big’ State of Health (SOH) data sets
- Provides abnormality detection scores in real-time for moving time windows of over 10K measurands on a laptop computer
- Abnormality detections for unexpected ‘unknown-unknowns’ are clustered, classified, and tracked over time
- Desired response for each historical abnormality type is listed
- Temporal pattern recognition tools predict effects of abnormal precursor signatures based upon historical data
Problem-to-Solution Space Mapping Depends upon Problem Difficulty

- **A Priori Model-Driven Systems**: Known Modeled Behavior
- **Contextual Data-Driven Systems**: Consistent Historically-Based Behavior
- **Adaptive Goal-Driven Systems**: Inconsistent [Not Based Upon Historical] Behavior

**Performance (log scale)** vs **Problem Difficulty (log scale)**
First Principles Based Techniques: Rule Based Expert Systems

- Useful in well understood easier problem domains
- Subject Matter Experts (SMEs) formalize their heuristics and best practices for deciding that a situation is abnormal, requires action, etc...
- SMEs configure automated rules that display alerts, flag data, or initiate courses of action based on data criteria
- These rules have the advantage of being intuitive and relatively transparent, but they can be costly and complex to implement and maintain
- *Require time consuming access to well-informed SMEs* in order to extract and formalize the rules – the system is only as powerful as the SMEs can directly make it
- Rule-based systems fail as affordable solutions for hard problems
First Principles Based Techniques: Probabilistic and Possibilistic Algorithms

• These model-driven solutions can implement harder to capture known influences and subtle relationships in data
• They can be easier for SMEs to configure, but still rely on SMEs to identify behavior models, algorithms, & desired relationships in the data
• Possibilistic algorithms are applicable when the uncertainty-in-the-uncertainty is high

Model-Driven Approaches Find the Known Needles in “Big Data” Haystacks, But Are Costly to Maintain
Data-Driven Advantages Over the First Principles Approach

• Data-driven approaches find the “dirty hay in haystack” (i.e., the unexpected & unknown unknowns)

• Data-driven approaches are affordable since do not require experts to solve the problem

• Data-driven analysis can enhance detection of expected abnormal signatures or relationships

• Data-driven analysis detects abnormal activity patterns in “big” context data relevant to baseline fusion system outputs
Goal-Driven Advantages Over Model-Driven Approach

• Goal-driven approach discovers solutions based upon mission goals when system models and historical data do not provide the solution.
• Goal-driven analysis is applicable for inconsistent behavior such as when there are unexpected modes of entity behavior.
• Goal-driven analysis provides adaptive solutions that are not known ahead of time such as controllers that achieve system goals in new unexpected system states.
• Goal-driven techniques can be driven by mission objective functions for each component, then these component outputs integrated based upon overall mission objectives.
• However, goal-driven approaches do not typically achieve the accuracy and Pd/Pfa of model-driven systems when the entity models are known, nor are they as robust as data-driven when behaviors are slowly changing.
CI Requirements for Data-Driven Abnormality Detection

• The DF&NN ANOM must automatically detect events in data when we:
  – Do not know what the signatures that affect our mission will all look like (i.e., unknown dirty hay)
  – Don’t know where or when they will occur
  – Must find abnormal behavior that is “within limits”
  – Must find abnormal temporal & abnormally correlated behavior
  – Must detect in near real-time and be linearly scalable with number of inputs
  – Need to automatically retrain/adapt to changing system dynamics
DF&NN ANOM versus Traditional Systems

- Traditional abnormality detection systems use manufacturer or analyst-derived fixed *red and yellow limits* on key measurands.
- These traditional approaches cannot detect within limits abnormal behavior and the limits go out of date.
- Traditional systems only detect statistically abnormal variations.
- A unique characteristic of the DF&NN ANOM is that the architecture of the ANOM components are automatically adapted based on the data.
- Occam’s Razor is applied to this adaption.
- Sufficient performance is achieved without time consuming system domain expert questioning.
ANOM Detects Abnormal Temporal Changes [such as Abnormal Frequency Behavior within Traditional Red/Yellow Limit Checking]

Small temporal changes are difficult to detect with limit checking.
Primary ANOM Service Components

- Off-Line Learning of Normal Behavior
- On-line Abnormality Detection, Characterization, Response, and Event Tracking
- Detects *Abnormal Correlations* amongst System Measurands
- Automated Detection Sensitivity Performance and Context Assessment
- Data-Driven Condition-Based Health Management (DCPM)
Abnormal Catalog Update (ACU) Automatically Manages over 250K Neural Networks (NNs)

Two Line Element (TLE) Training Data

ANOM NN’s Training

ANOM NNs Testing / Management

Promoted NNs

L0 Abnormality Detection

L1 Event Track & Characterization

ACU Mining & Fusion Process Management

User Goals

TLE Relational Database Mgmt.

TLE Testing Data

TLE Training & Testing Times

ACU Events

User Goals

Promoted NNs

Abnormal TLE Detection & Mgmt.

TLE Mining & Mgmt.
On-Line Abnormality Detection, Characterization, Response, and Event Tracking

• The on-line data drives the trained functional components to generate abnormality scores for each data window.
• Data window records with residual error scores above user defined thresholds are abnormality detections.
• Abnormal events are persistent associated detections.
• The abnormality detections are clustered to previously occurring abnormalities signatures in real-time for cause naming and response description.
• For unmatched (i.e., new) abnormalities ANOM provides automated default names based upon the dominant abnormal measurands and subsystems.
Abnormal Behavior Correlations Detected during ANOM IV&V
(e.g., due to missing correlated value changes)
ANOM Has Been Proven at TRL 7

• The DF&NN ANOM technology has a proven track record for learning what is normal and detecting unexpected abnormally correlated behavior based upon independent performance evaluations.

• The ANOM tools are applicable for a variety of CI applications to include SOH, cyber, aircraft, spacecraft, and RFI problems.

• ANOM computational burden grows linearly with the number inputs per system.

• ANOM has run on over 10,000 measurands on a single computer.

• ANOM computes derived measurands from combinations of inputs to reduce the problem dimensions.

• Known abnormal behavior time intervals can be flagged for enhanced detection.
ANOM Data-Driven Health Management
Existing Tools

• The ANOM health management system is maintained using a hierarchy of management services that are interlaced with the data fusion & mining services

• Examples of these management tools are as follows:
  – *Adaptive On-Line Abnormality Suppression and Enhancement GUI* to immediately get rid of ringing
  – *Off-Line ANOM Component Fast Retraining* to increase sensitivity
  – *Known Abnormality Detection Enhancement* for known abnormal times
  – *Detection Rate Management* to automatically adapt the number of abnormalities reported by ANOM
Automated Detection Sensitivity Performance and Context Assessment

• The ANOM detection sensitivity performance assessment computes the detection onset values for positive & negative variations in each measurand

• This provides the conditions that guide the snipping, replication, and enhancement of measurands and for comparing the performance of alternative trained ANOM solutions

• The culprit assessment tool allows user to drill down to understand the cause of each detection

• The context assessment capability enables the user to find significant nonconforming relevant behavior
DF&RM Dual Node Network (DNN) Technical Architecture

Data Fusion Node Network

- Fusion Level 0: Feature Assessment
- Fusion Level 1: Entity Assessment
- Fusion Level 2: Situation Assessment
- Fusion Level 3: Impact Assessment
- Fusion Level 4: Process Assessment

Resource Management Node Network

- Management Level 0: Resource Signal Management
- Management Level 1: Resource Response Management
- Management Level 2: Resource Relationship Management
- Management Level 3: Mission Objective Management
- Management Level 4: Process/Knowledge Management

Sources:
- Sensor
- Source

Resources:
- Sensors
- CMs
- Comm
- Others

User I/O (at all levels as needed)
Data Fusion & Resource Management (DF&RM) Dual Node Network (DNN) Technical Architecture

Data Fusion

L.0 Feature/Signal Assessment
L.1 Entity Assessment
L.2 Situation Assessment
L.3 Impact Assessment
L.4 System Assessment

Resource Management

L.0 Resource Signal Management
L.1 Individual Resource Management
L.2 Resource Relationship Management
L.3 Mission Objective Management
L.4 System Management

DF Output KB
Model KB

Organic Sources & Resources
External Sources & Resources

Dual Node Network (DNN) DF&RM Technical Architecture
External Data used as Context to Assess and Improve Fusion Products

Environment
- Sensor Report Generation
- Scenario Definition

Baseline DF&RM
- Baseline Fusion System
- Baseline Resource Management System

CAFM
- External Data
- Context Conformity Assessment (CA)
- Process Assessment
- Context Management Models
- Context Conformity Management (CM)
- Process Management

Users
DF & RM Node Duality
Facilitates Technique/Software Reuse

DATA FUSION NODE

- DATA PREPARATION
- HYPOTHESIS GENERATION
- HYPOTHESIS EVALUATION
- HYPOTHESIS SELECTION
- STATE ESTIMATION

• L.0-3 Entity State Estimates
• System State Estimates (L.4)
• Response Plans
• Controls

RESPONSE TEMPLANNING

- TASK PREPARATION
- PLAN GENERATION
- PLAN EVALUATION
- PLAN SELECTION
- TASKING/CONTROL

Higher-Level Mgmt Node

Lower-Level Mgmt Nodes

Sources or Prior DF Nodes

User or Next DF Node
ANOM CI Capability Summary

• Data-driven ANOM is the **affordable** solution to **unexpected abnormality detection & characterization**
• ANOM can support all levels of data fusion as defined by the DF&RM DNN technical architecture such as **abnormal entity relationships** and **COA detection**
• The ANOM affordably solves harder problems to include ISR and **cyber change detection**
• ANOM automatically retrains to learn recent normal health behavior and provide **context for historical abnormalities**
• The system is extensible, scalable, cross-platform, and supports multiple users and roles in Linux and Windows
• ANOM provides **data-driven and goal-driven computational intelligence services**
Computational Intelligence Operational Utility

• Role for CI is for harder problems
• CI learns to solve these problems automatically based upon the data and goals from the user
• Thus CI is affordable and extendable as the baseline system behavior changes
• TRL 7 CI solutions have been delivered to 3 sites
• Lower TRL prototype solutions have been developed and tested for numerous applications
backups
Resource Management Functions

- Determine candidate tasks that meet prioritized needs
- Mediate resource tasks (e.g., common representation & utility)
- Compensate for resource mismodeling and faults
- Output prioritized resource tasks

- Generate feasible resource task plans (e.g., task schedules)
- Score feasible resource plans
- Select, delete, or initiate resource plans
- Output selected resource plan(s)

- Manage resource objectives & relationships
- Task/control resources
  - HW/SW modes, cues, & signals
  - Resolve resource scheduling conflicts
- Feedback resource mgmt. status
- Output L0/1/2/3 resource tasking/controls