Predictive, Proactive, and Collaborative ML with IT Service Intelligence

Not a Science Project – Creating actionable events through Analytics

Nate Smalley | Staff SE
Andrew Stein | Analytical Architect
Sept 26, 2017 | Washington, DC
Forward-Looking Statements

During the course of this presentation, we may make forward-looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC.

The forward-looking statements made in this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward looking statements we may make. In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.

Splunk, Splunk>, Listen to Your Data, The Engine for Machine Data, Splunk Cloud, Splunk Light and SPL are trademarks and registered trademarks of Splunk Inc. in the United States and other countries. All other brand names, product names, or trademarks belong to their respective owners. © 2017 Splunk Inc. All rights reserved.
Nate Smalley

- IT Operations Technologist (Reformed Security Guy)
- Former Technical Director of Security & Monitoring Tools Team – Apollo Group (University of Phoenix)
- Currently Splunk Staff Sales Engineer supporting Large Businesses in FinServ & Manufacturing
- Enjoy Long walks across SNMP and Candle light dinners while fighting Operational Outages
Andrew Stein

- Splunk Global Analytical Architect
- 17 years creating mathematically modeled solutions
- I spend 80 percent of time spent preparing data and 20 percent of time complaining about the need to prepare data.
Problem Statement

Operations Teams need more time between an alert and a failure that has Availability impacting ramifications. The introduction of Machine Learning is a have to have in order to predict these failures. These notable events must be able to pushed and collaborated on by via teams in various tools.
Agenda

- What data do we need
  - Indicators Matter
  - Data Where to Get it
  - Demo

- We Need Machine Learning
  - What can we use
- Real use case Example
- Let Dive into “How it works”
- Next Steps
Indicators Matter
Indicators

▶ Key Performance Indicators that is
  • Defined - **metrics** that are used to evaluate the overall status of a service.

▶ Leading Indicators – Drivers of a Result
▶ Lagging Indicators – Outcome of the Result

▶ Example Scenario
  • DB Runs out of Space
  • KPI Storage value = 100% <- Leading Indicator
  • KPI User Response time value = 2000+ secs <- Lagging Indicator
KPI’s so what?

Understanding the historical break out of Leading vs Lagging KPI’s and the association to Services is critical in understanding how to predict a good outcome vs bad outcome.

Understanding these key parts from the Service Experts is critical:

- Business Priority
- Use Case Data Needed
- The “RIGHT” KPI’s to measure
- What decision happens when a Bad outcome Occurs
Data Where to Get it

I need it STAT
Where Does Data Come From?

**Machine Data**
- Synthetic APM
- Byte Code Instrumentation
- ML/Adaptive Thresholding
- Server
- Storage
- Network

**Human Data**
- Change Records
- Server Changes
- Application Code
- Code Deployment

**Service Intelligence**
- METRICS
- IP Creation
- EVENTS
What Do I do with it Now?

▶ Use IT Service Intelligence
▶ Dynamic Service Structure
  • Provide Flexible Dependency Service mapping for interactions at Scale
▶ Build Key Performance Indicators
  • Leverage a Platform to build out KPI’s
  • Ensure repeatability amongst Services for Consistency
  • Aggregation and Per Entity KPI Values
▶ Ease the burden of Cleaning Data (Schema at search time?)

“Data scientists spend 60% of their time on cleaning and organizing data. Collecting data sets comes second at 19% of their time, meaning data scientists spend around 80% of their time on preparing and managing data for analysis.”
Real Use Case Example

Lets Get the data
We Need Machine Learning

Do you want to watch it or let it go?
Custom Machine Learning – Success Formula

**Domain Expertise** (IT, Security, …)

- Identify use cases
- Drive decisions

**Data Science Expertise**

- Statistics/math background
- Algorithm selection
- Model building

**Splunk Expertise**

- SPL
- Data prep
- Operational success

Set business/ops priorities

What KPIs matter to me

-Splunk ML Toolkit facilitates and simplifies via examples & guidance
Overview of ML at Splunk

- **CORE PLATFORM SEARCH**
- **PACKAGED PREMIUM SOLUTIONS**
- **MACHINE LEARNING TOOLKIT**

**splunk** Platform for Operational Intelligence
Three Types Of Machine Learning

Supervised Learning:

Unsupervised Learning:

Reinforcement Learning:
Deviation from past behavior
Deviation from peers
(aka Multivariate AD or Cohesive AD)
Unusual change in features

ITSI MAD Anomaly Detection

**Anomaly detection**

Predict Service Health Score
Predicting Churn
Predicting Events
Trend Forecasting
Detecting influencing entities
Early warning of failure – predictive maintenance

**Predictive Analytics**

Identify peer groups
Event Correlation
Reduce alert noise

ITSI Event Analytics

**Clustering**
Splunk Machine Learning Toolkit

**Algorithms**
- 25+ standard algorithms available prepackaged with the toolkit

**Assistants**
- Guide model building, testing, & deployment for common objectives

**Showcase**
- Interactive examples for over 25 typical IT, security, business, IoT use cases

**New commands**
- to fit, test and operationalize model

**Python for Scientific Computing Library**
- 300+ open source algorithms available for use

**MLib integration**
Real Use Case Example

Let's use ML to get predictive.
Event Analytics

BUILT IN - IT Service Intelligence Machine Learning
Notable Events are key

• We created a notable for Web Store Service – Called webstore_health_alert
• Lets see how Splunk can cluster this with other events to show other Notable Events that are attributed

Smart Mode is actually Smart

• It uses clustering to group events remember unsupervised ML at the beginning
The Diamonds in the Rough
Clustering Events into actionable alerts

- Take 1000’s or 100’s of 1000’s of alerts and connect them
- Use ML prediction to improve correlation
- Boil the events down to reasonable count of Actionable Events
Real Use Case Example

Predictive Analytics in Real-time
Lets Get Others Involved

ChatOps - Anyone Slack?
Slack

- Instant Message
- Easy Channel Setup
- Fast time to create incident working groups
- Other Products just as easy HipChat, Skype etc...
ChatOps = Splunk & Slack

▶ Slack Setup
  • Install the Alert Action
    • https://splunkbase.splunk.com/app/2878/
  • For Core Searches – Configure as alert action in Alerts
  • For ITSI
    – Enable Slack in notable_event_actions.conf
  • Lets add it to our Policy
    • Configure -> Notable Event Aggregation Policy -> Create New
Real Use Case Example

Predictive Analytics in Real-time
ITSI allows for Rapid Service, KPI, Entity creation and alerting

Machine Learning through MLTK provides a repeatable ability to Predict Service Health

IT Service Intelligence Event Analytics enables teams to cluster Notable events together to have 1 Actionable Alert

Splunk IT Service Intelligence provides extendable capabilities to provide immediate notification to Chat Groups, Ticketing system and other communication platforms to improve the Mean Time to Remediate Availability Impacting Situations
I fell asleep. Where else can I read about this?

Go see these Talks!
This is where the subtitle goes

► David Vueve
  • Splunk Ninja Skills
  • 190+ slides of SPL and guidance.

► Xander and (fred?)
  • Deep dive into MLTK
  • MLTK API for importing algorithms

► Phillip Drieger
  • DGA Analysis
  • End to End MLTK example

► Making ML Solutions
  • Deep dive into ML process
  • Customer use cases explored
Join the Pony Poll

ponypoll.com/***
Questions?

Please Feel Free
Want to Learn More About ITSI at .conf2017?

- **Ready, Set, Go! Learn From Others - The First 30 Day Experiences of ITSI Customers:** Tuesday, September 26th, 2017 12:05 PM - 12:50 PM Room Salon C
- **Splunk ITSI Overview:** Tuesday, September 26th, 2017 1:10 PM - 1:55 PM Room 147 AB
- **PWC: End-to-End Customer Experience:** Tuesday, September 26th, 2017 2:15 PM - 3:00 PM Room 143ABC
- **RSI: Operational Intelligence: How to go From Engineering to Operationalizing IT Service Intelligence Where the Rubber Meets the Road:** Tuesday, September 26th, 2017 2:15 PM - 3:00 PM Room 147AB
- **Cardinal Health: Ensuring Customer Satisfaction Through End-To-End Business Process Monitoring Using Splunk ITSI:** Tuesday, September 26th, 2017 2:15 PM - 3:00 PM Room 147AB
- **ITSI in the Wild - Why Micron Chose ITSI and Lessons Learned From Real World Experiences:** Tuesday, September 26th, 2017 4:35 PM - 5:20 PM Room Salon C
- **Event Management is Dead. Time Series Events are the Means to the End, not the End Itself. See How Event Analytics is Revolutionizing IT:** Wednesday, September 27th, 2017 11:00 AM - 11:45 AM Ballroom C
- **Triggering Alerting (xMatters) and Automated Recovery Actions from ITSI:** Wednesday, September 27th, 2017 1:10 PM - 1:55 PM Room Salon C
- **Leidos - Our Journey to ITSI:** Wednesday, September 27th, 2017 2:15 PM - 3:00 PM Room 147AB
- **How Rabobank's Monitoring Team Got a Seat at the Business Table by Securing Sustainability on Competitive Business Services Build on Splunk's ITSI:** Wednesday, September 27th, 2017 2:15-3:00pm Room 147AB
- **Here Comes the Renaissance: Digital Transformation of the IT Management Approach:** Wednesday, September 27th, 2017 3:30 PM - 4:15 PM Room Salon C
- **The ITSI ‘Top 20’ KPI’s:** Thursday, September 28th, 2017 10:30 AM - 11:15 AM Room Salon C
- **Automation of Event Correlation and Clustering with Machine Learning Algorithms – An ITSI Tool:** Thursday, September 28th, 2017 11:35 AM - 12:20 PM Room Salon C
- **Event Management is Dead. Time Series Events are the Means to the End, not the End Itself. See How Event Analytics is Revolutionizing IT:** Thursday, September 28th, 2017 11:35 AM - 12:20 PM in Ballroom B
- **IT Service Intelligence for When Your Service Spans Your Mainframe and Distributed ITSI:** Thursday, September 28th, 2017 1:20 PM - 2:05 PM Room Salon C
Hidden Gems

- Remember that the MLTK will see any numeric field as a number, not as an entity, so if we want to use `date_hour` (a number between 0 and 23) we need to change the value into a string (`|eval date_hour_string= date_hour ."."."’_t` for example)

- Remember we are predicting the future, so we need to move the target of our regression through time.

- Remember that your data retention policy means you will lose past data at some point – consider making two models, one with `partial_fit` and one without and comparing the results.