Correlative Analytic Methods in Large Scale Network Infrastructure

Hariharan Krishnaswamy
Senior Principal Engineer
Dell EMC
Data Center Network Characteristics

- Continuous growth in scale & complexity
- 24/7 business workload
- Addition, removal of infrastructure components
- Changing execution environment
- Dynamic workload patterns – varies by day of the week, hour of the day
- Dynamic application arrivals/departures
- Dynamic updates to firmware/software
BALANCE THE DEMANDS OF AVAILABILITY & EFFICIENCY

Data Center Availability
- Uptime Management
- Service Level Management
- Performance Management

Business Efficiency
- Capacity Optimization
- Infrastructure Management
- Operational Cost Management
- Capital Cost Management
Network Outage & Consequences

- 400+ network failures occur each year in data centers
- Network outages leads to extensive losses due to lack of responsiveness or availability

- Predictive intelligence LEADS TO Reduced downtime & Maximum efficiency
  A. Predict Failure/outage in advance
  B. Proactively mitigate ill-effects of the anomalous behavior
Data Center Failures

- Device failures
  - Server Host, NIC, HBA, CNA
  - Router, Switch, Load Balancer, Firewall
  - Storage Controller, Disk Array
  - Cable/Optics/Media
- System/Protocol software failures
- Application failures
- Network failures
- Data Traffic issues
  - Latency increase/Throughput decline
  - Sustained unexpected long term traffic load
Data from monitoring instrumentation
Huge volumes of Historic Data & Current Data
- Events
- Alerts
- Traps
- Syslog
- Counters
- Packet traces
- Debug dumps
Objectives of AI & Analytic methods

- Improve the ‘signal-to-noise’ ratio to a large extent
- Model development from available historic data
- Blend and ingest a variety of structured, semi-structured and unstructured data
- Find hidden patterns & correlations relating the device / network behavior
Symptoms & Anomalous conditions

Application symptoms
- Backup application failure
- Huge delay in storage access
- Streaming video stall

Anomalous conditions
- BGP anomaly (BGP flapping, leaks, table clears, etc.)
- Queue full/buffer depletion
- STP interop problem
- Transceiver/optics/Cables issues
- QOS misconfiguration
AI methods for network analytics

- Association Rule Mining
- Supervised Machine Learning – Regression
- Supervised Machine Learning – Classification
- Unsupervised Machine Learning - Clustering
Association Rule Mining

- Important data mining concept
- Uncover mutual connection between data items in the massive data set
- Discover credible and representative rules
- Algorithms: Apriori, Partition, Pincer-Search, Incremental, and Border algorithm
- Most Popular algorithm: Apriori
Apriori Algorithm

- Algorithm for mining frequent itemsets for boolean association rules.
- Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time (a step known as candidate generation) and groups of candidates are tested against the data.
- Apriori is designed to operate on dataset containing transactions
- Used extensively in Retail Analytics (Market Basket Analysis)
Frequent Item Set – (applied to event logs)

- Given a set $B$ of Events called the item base and a large database $T$ of event logs, itemset $I \subseteq B$. The support $S_T(I)$ of an item set $I \subseteq B$ is the number of event logs in the database $T$.

- With a specified minimum-support $S_{\text{MIN}}$, an item set $I$ is called frequent in $T$ iff $S_T(I) \geq S_{\text{MIN}}$

- The goal is to identify all item sets $I \subseteq B$ that are frequent in a given event log database $T$.
<table>
<thead>
<tr>
<th>Transaction</th>
<th>Event Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>{ev-1, ev-3, ev-4}</td>
</tr>
<tr>
<td>02</td>
<td>{ev-2, ev-3, ev-5}</td>
</tr>
<tr>
<td>03</td>
<td>{ev-1, ev-2, ev-3, ev-5}</td>
</tr>
<tr>
<td>04</td>
<td>{ev-2, ev-5}</td>
</tr>
</tbody>
</table>
APRIORI ALGORITHM EXAMPLE (Contd..)

Minimum support is 50%

<table>
<thead>
<tr>
<th>Eventset</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>{ev-1}</td>
<td>2</td>
</tr>
<tr>
<td>{ev-2}</td>
<td>3</td>
</tr>
<tr>
<td>{ev-3}</td>
<td>3</td>
</tr>
<tr>
<td>{ev-4}</td>
<td>1</td>
</tr>
<tr>
<td>{ev-5}</td>
<td>3</td>
</tr>
</tbody>
</table>

The eventset {ev-4} has less than minimum support. Hence is discarded

<table>
<thead>
<tr>
<th>Eventset</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>{ev-1}</td>
<td>2</td>
</tr>
<tr>
<td>{ev-2}</td>
<td>3</td>
</tr>
<tr>
<td>{ev-3}</td>
<td>3</td>
</tr>
<tr>
<td>{ev-5}</td>
<td>3</td>
</tr>
</tbody>
</table>
APRIORI ALGORITHM EXAMPLE (Contd..)

Two item eventsets

<table>
<thead>
<tr>
<th>Eventset</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>{ev-1, ev-2}</td>
<td>1</td>
</tr>
<tr>
<td>{ev-1, ev-3}</td>
<td>2</td>
</tr>
<tr>
<td>{ev-1, ev-5}</td>
<td>1</td>
</tr>
<tr>
<td>{ev-2, ev-3}</td>
<td>2</td>
</tr>
<tr>
<td>{ev-2, ev-5}</td>
<td>3</td>
</tr>
<tr>
<td>{ev-3, ev-5}</td>
<td>2</td>
</tr>
</tbody>
</table>

The eventsets \{ev-1, ev-2\} and \{ev-1, ev-5\} have less than minimum support. Hence are discarded.

<table>
<thead>
<tr>
<th>Eventset</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>{ev-1, ev-3}</td>
<td>2</td>
</tr>
<tr>
<td>{ev-2, ev-3}</td>
<td>2</td>
</tr>
<tr>
<td>{ev-2, ev-5}</td>
<td>3</td>
</tr>
<tr>
<td>{ev-3, ev-5}</td>
<td>2</td>
</tr>
</tbody>
</table>
APRIORI ALGORITHM EXAMPLE (Contd..)

Three item eventsets

<table>
<thead>
<tr>
<th>Eventset</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>{ev-1, ev-2, ev-3}</td>
<td>1</td>
</tr>
<tr>
<td>{ev-2, ev-3, ev-5}</td>
<td>2</td>
</tr>
</tbody>
</table>

The eventsets \{ev-1, ev-2, ev-3\} has less than minimum support. Hence are discarded.

Frequent Eventset:

<table>
<thead>
<tr>
<th>Eventset</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>{ev-2, ev-3, ev-5}</td>
<td>2</td>
</tr>
</tbody>
</table>
Rule base Creation Algorithm

- Event sequence model
- Apply frequent itemset mining to extract frequently event sets
- Among the frequent item sets, select the ones with failure events
- Form Rule with Preceding events --> Failure event
- Ex: \{ ev-A,ev-X,ev-T,ev-R\} -> Failure event ev-F
- Iterate through all the frequent itemsets
TIMELINE OF FAILURE
PREDICTION BASED ON ASSOCIATION RULE MINING - WORKFLOW

Historic Data → Association Rule Mining

CURRENT DATA → RULE BASE

RULE BASE → PREDICTION-1, PREDICTION-2, PREDICTION-3
PREDICTION WITH RULEBASE

- Certain sequence of events LEAD TO specific Failures.

- \{ev-1, ev-2, ev-3, ev-4, ev-5, ev-6, \ldots \ldots \ldots ev-FAIL\}

- Specific points in the event sequence could indicate probabilities of specific Failures

- Additional statistics & support data could help improve prediction confidence
Supervised Machine Learning

- Requirement: Labeled historic data
- Predictor variables and their interactions are the key
- Machine learning algorithms learn from the labeled historic data
- Failure prediction is a CLASSIFICATION task
- Traffic demand forecasting is a REGRESSION task
Feature Extraction

- Data: Statistics, Events, Logs, Alerts, Traps, Counters, Packet traces, Debug dumps
- Data volume is huge (significant noise component too)
- Hence increased processing time
- Features may be highly correlated
- Not all features may contribute to prediction
- Solution: Dimensionality Reduction
  - Remove highly correlated features
  - PCA – Principal Component Analysis
Solution: Dimensionality Reduction

- Remove highly correlated features by feature-wise correlation analysis

- PCA – Principal Component Analysis
Feature Extraction - Principal Component Analysis

- New set of features called components, which are composites of the original features, but are uncorrelated with one another.
- First principal component accounts for the largest possible variability in the data, the second component the second most variability, and so on.
HISTORIC DATA

Data Cleansing
Feature extraction

Feature representation

Learning Algorithm

Model

CURRENT DATA

Feature extraction & Representation

PREDICTION

Supervised Learning Workflow
Classification Algorithms – Machine Learning

- Naive Bayes Classifier
- Nearest Neighbor
- Support Vector Machines
- Decision Trees / Boosted Trees
- Random Forest
- Neural Networks
## Dynamic Time Series of Feature Vectors

<table>
<thead>
<tr>
<th>TIME</th>
<th>feature f1</th>
<th>f2</th>
<th>f3</th>
<th>...</th>
<th>fn</th>
<th>PREDICTED OUTCOME</th>
</tr>
</thead>
<tbody>
<tr>
<td>time t-k</td>
<td>&lt;&gt;</td>
<td>&lt;value&gt;</td>
<td>&lt;value&gt;</td>
<td></td>
<td>&lt;value&gt;</td>
<td>NO ANOMALY</td>
</tr>
<tr>
<td>...</td>
<td>&lt;value&gt;</td>
<td>&lt;value&gt;</td>
<td>&lt;value&gt;</td>
<td></td>
<td>&lt;value&gt;</td>
<td>NO ANOMALY</td>
</tr>
<tr>
<td>time t</td>
<td>&lt;value&gt;</td>
<td>&lt;value&gt;</td>
<td>&lt;value&gt;</td>
<td></td>
<td>&lt;value&gt;</td>
<td>ANOMALY</td>
</tr>
<tr>
<td>...</td>
<td>&lt;value&gt;</td>
<td>&lt;value&gt;</td>
<td>&lt;value&gt;</td>
<td></td>
<td>&lt;value&gt;</td>
<td>NO ANOMALY</td>
</tr>
<tr>
<td>time t+x</td>
<td>&lt;value&gt;</td>
<td>&lt;value&gt;</td>
<td>&lt;value&gt;</td>
<td></td>
<td>&lt;value&gt;</td>
<td>ANOMALY</td>
</tr>
<tr>
<td>time t+y</td>
<td>??</td>
<td>??</td>
<td>??</td>
<td></td>
<td>??</td>
<td>HOW TO PREDICT OUTCOME FOR FUTURE TIME ????</td>
</tr>
</tbody>
</table>
Predicting future outage/failure conditions?

- No feature vector available for time “t+y”
- Solution: Combine Time Series Regression & Classification
  - First predict the feature vectors at time “t+y” (by Time Series Regression)
  - Then predict the future outcome (by Prediction Model)
Characteristics of Time Series Data

- Trend over time (Ex: Gradual increase/decrease of activity over time)
- Seasonal trend or cycle (Ex: increases in the morning hours, peaks in the afternoon and declines late at night)
- Seasonal variability. (Ex: Fluctuations wildly minute by minute during the peak hours of 4-9 pm, and declining to nearly zero by 1 am)
- Need to account for the Trend & Seasonality in the dataset
Handling Trends & Seasonality

- **Additive Holt-Winters method**
  
  Used for time series with constant (additive) seasonal variations

- **Multiplicative Holt-Winters method**
  
  Used for time series with increasing (multiplicative) seasonal variations
Steps in Time Series Prediction

1. Apply Holt Winters smoothing
2. Time series regression to forecast feature vector at time “t+y”
3. Predict the future outcome using feature vector

<table>
<thead>
<tr>
<th>Time</th>
<th>f1</th>
<th>f2</th>
<th>…</th>
<th>…</th>
<th>fn</th>
<th>PREDICTED OUTCOME</th>
</tr>
</thead>
<tbody>
<tr>
<td>time t+y</td>
<td>&lt;value&gt;</td>
<td>&lt;value&gt;</td>
<td>&lt;value&gt;</td>
<td>&lt;value&gt;</td>
<td>&lt;value&gt;</td>
<td>&lt;PREDICTION&gt;</td>
</tr>
</tbody>
</table>
Analytics System Architecture

- Text Mining Engine
- Time Series Processing Engine
- Association Rule Engine

Insight Engine
Insight Engine

- Context-aware Intelligent engine
- Meaningful insights & predictions from the data
- Dynamic learning
Proactive Mitigation

- Outage/failure predicted in advance
- How to perform mitigative action?
- Reaction time should be very fast
- Human intervention reaction time is too high
- Automated, pre-defined, established actions through software
  “Event-Driven Programmability”
QUESTIONS & DISCUSSION

???
THANK YOU VERY MUCH !!!
Metrics (System level)

- **CPU**
  Average CPU usage - historical average
  CPU usage - current

- **Memory**
  System memory - historical average
  System memory usage - current

- **Disk**
  Disk space usage - historical average
  Disk space usage - current

- **File systems**
  File system/Descriptors - historical average
  current usage
Metrics (Environmental)

- Temperature sensor
- Power-supply
- Fan Trays
- Voltage-sensor
- Optics characteristics
Metrics (Data path related)

- inbound(rx) packet errors in percentage
- inbound(rx) packets discarded in percentage
- inbound(rx) traffic, measured in Kb/s
- incoming(rx) bandwidth in use in percentage
- incoming(rx) packets discarded because of an unknown or unsupported protocol
- incoming(rx) ucast pkts
- incoming(rx) mcast pkts
- protocol specific stats
Metrics (Data path related)

- outbound(tx) packet errors in percentage
- outbound packets discarded in percentage
- outbound traffic, measured in Kb/s
- outgoing bandwidth in use in percentage
- incoming(rx) ucast pkts
- incoming(rx) mcast pkts
- protocol specific stats
Metrics (Data path related)

- ICMP round-trip-time and packet loss
- Iperf data
- QoS Profiles
- BGP Sessions
- OSPF Neighbor states
- Historic throughput profile
- Historic latency profile
Data Correlation

Correlate symptoms to specific events, data and patterns

- Symptom – Symptom correlation
- Symptom – Event correlation
- Event – Event correlation (Events consistently occurring within a predefined time threshold of each other)
- Event – Data correlation
- Data – Data correlation
- Pattern – Pattern correlation
SUPPORT VECTOR MACHINES