

Can Bigdata and Machine Learning really improve Network Operations?

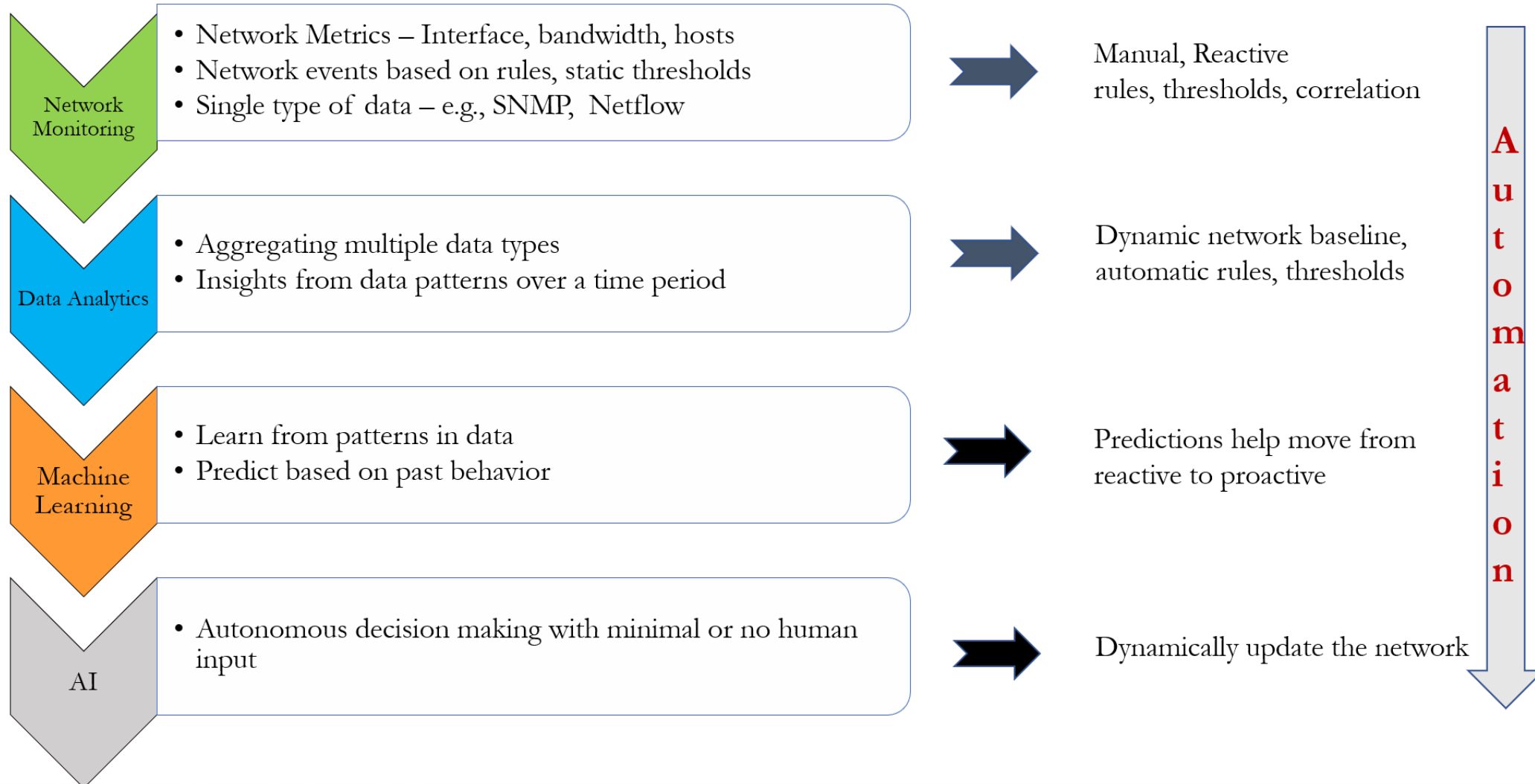
Srividya Iyer

Co-Founder and CEO – CANIV TECH INC.

What this talk is about

- How can data analytics and Machine Learning help your network operations ?
- How can you iteratively implement analytics within your operations without tearing down your existing systems and processes?
- How do you integrate with current infrastructure and supporting systems?
 - What processes need to be adapted to support this?
 - What is the additional cost to implement this?
 - What are the skills required and what kind of training you would need to provide your staff and management to effectively implement and manage the data analytics.

Network Analytics to Automation



Implementing Analytics and Machine Learning

An iterative approach

Analytics/Machine Learning checklist

Clearly defined Usecases/Problem statement

Right type of data collection

Data processing and storage

Machine Learning Framework, Approach, Algorithm

Clearly defined usecases

- Networks are complex, so before you embark on adding data analytics define the problem you are trying to solve.
- Problem domain can be sliced by one of the following (not exhaustive)
 - Application protocols (DNS, HTTP, SMTP)
 - Transport protocols (all TCP connections)
 - Geographic region
 - IP address/subnet ranges
 - Specific hosts or servers (database server, GIS server)
 - Specific devices (Edge router, core router, firewalls)
 - Traffic types (East-West traffic, WAN traffic, wireless)
 - Content type (video, voice)

Collect the right data

- Network data is inherently noisy.
- Devices where data is collected (Host, Server, Router , Switch, Firewall etc)
 - Active measurement – Ping, Traceroute
 - Passive traffic data
 - Flow data - Netflow, IPFIX, sFlow,
 - Packet captures (headers, full capture),
 - Logs (syslog, firewall logs, wifi logs etc.)
 - Routing information, Topology, device configurations
- Depends on your usecase.
- Add data sources/types incrementally (for e.g., DNS Server logs first and then packet capture)

Storing and processing the data

- Data processing and storage depends on what you are trying to do with the data. There are too many options.
- Some general criteria to think about before selecting a solution
 - In house or cloud storage and processing.
 - Stream or batch processing of data (Apache Spark, Kafka)
 - Search and indexing (ELK stack, Splunk)
 - Persistent data storage and retrieval (Postgres, MongoDB)
 - Simple data storage and retrieval using scripts (Python/Bash) - CSV, JSON files

Machine Learning Framework/Approach/Algorithms

- ML Framework
 - Open Source build it yourself tools (Python/Scikit Learn, Keras, Tensorflow)
 - Cloud provider tools (e.g., Google's AutoML)
- ML Approach
 - This depends on what your usecases are. There are two main approaches – Supervised Learning and Unsupervised Learning.
- ML Algorithms
 - Algorithms are selected based on your approach. There are some general guidelines based on the problem. Have to do some trial and error.

Integrating with current operations

Infrastructure, Process and Skills

What do you need to consider ?

- Review your current data collection infrastructure and policies. Is the data collected and stored in a form that can be used for analytics? What are the additional data infrastructure needs ?
- Review your current network monitoring/NOC process. How would you use the results from the analytics within your operations?
- Review additional cost of implementation
 - Buying vs Building
 - Open Source vs Commercial solutions
 - Additional staffing needs.

What do you need to consider?

- These are some of the skills that are required to successfully implement data analytics and Machine Learning.
 - A good understanding of data sources, types, formats
 - Knowledge of data collection tools and how to configure them.
 - Working knowledge of TCP/IP protocol stack
 - Knowledge of bash/python scripting
 - A basic understanding of Machine Learning algorithms.

A quote from the Google Developers site – “do machine learning like the great engineer you are, not like the great machine learning expert you aren’t”

Intangibles ...

- Trusting Machine Learning
 - It is not autonomous, so you have to give it the right data and instructions.
 - Initially the predictions have to be verified.
 - Efforts are underway to “standardize” trust in AI/ML (NIST)
- Job Security
 - You are automating your own expertise on the network. This is not making your skills irrelevant. You will still need domain knowledge. The biggest change you will see is that you'll have more time to strategize instead of constantly fixing the same problem.
- Management and Staff buy in
 - Management has to communicate the business needs clearly and allocate time for training and implementation.
 - Staff has to feel that they are being supported when things don't go according to plan.

To wrap up...

Define your problem clearly

Start small and build on your success

Email: siyer@caniv-tech.com

Twitter: @siyeratwork