PREDICTIVE ANALYSIS IN NETWORK FUNCTION VIRTUALIZATION

PRESENTATION BY IVAN KHAFFAJI

ABSTRACT

- Recent deployments of Network Function Virtualization (NFV) architectures have gained tremendous traction. While virtualization introduces benefits such as lower costs and easier deployment of network functions, it adds additional layers that reduce transparency into faults at lower layers. To improve fault analysis and prediction for virtualized network functions (VNF), we envision a runtime predictive analysis system that runs in parallel with existing reactive monitoring systems to provide network operators timely warnings against faulty conditions. In this paper, we propose a deep learning-based approach to reliably identify anomaly events from NFV system logs, and perform an empirical study using 18 consecutive months in 2016-2018 of real-world deployment data on virtualized provider edge routers. Our deep learning models, combined with customization and adaptation mechanisms, can successfully identify anomalous conditions that correlate with network trouble tickets. Analyzing these anomalies can help operators to optimize trouble ticket generation and processing rules in order to enable fast, or even proactive actions against faulty conditions.

ABSTRACT: DEFINING SOME TERMS I

- NVF -- Network Function Virtualization
  - Network function virtualization (also network function virtualization or NFV) is a network architecture concept that uses the technologies of IT virtualization to virtualize entire classes of network node functions into building blocks that may connect, or chain together, to create communication services.
- Runtime Predictive Analysis
  - A technique to detect potential concurrency errors in a system by observing its execution traces; the analyzed execution traces may not necessarily hit the errors directly. Typical concurrency errors that can be predictively detected include dataraces and deadlocks, both notoriously hard to find by just ordinary testing.
- Provider Edge Routers
  - A Provider Edge Router (PE router) is a router between one network service provider's area and areas administered by other network providers.

ABSTRACT: DEFINING SOME TERMS II

- Virtualization (In reference to Networking)
  - Reducing the scale, diversity, and cost of the hardware to only what is necessary; and using software for network functions, so if business needs change, providers can easily update the software instead of the whole system's hardware.
- Trouble Ticket
  - A trouble ticket is the result of an end user submitting a help request via an issue tracking system, and it typically contains elements detailing the exact nature of the problem the end user is having with a specific network component.

PROS AND CONS OF VNF

PROS

- Easily deploy new technology
- Easier management
- Lower cost/better scalability
- Streamlines new integrations

CONS

- Probability of failure is higher
- More abstraction and less “visibility” into lower layers
- Requires reorganization of IT personnel and altering existing processes

CHALLENGES THAT WE FACE

Since traditional systems can provide up-time of up to 99.999% (up to “5 9’s”), it is reasonable to want VNFs to generally rival that sort of system stability.

Therefore, failures are somewhat rare, and we need mass amounts of data in order to further balance our results. Sample size will be too small otherwise.

Typically, machine learning gets more effective as more information is used to train the model, so with little data, it will be more difficult to achieve accurate results.

Since most networks are highly specialized dependent on their respective applications, one model might not work effectively on every network.

What is the goal?

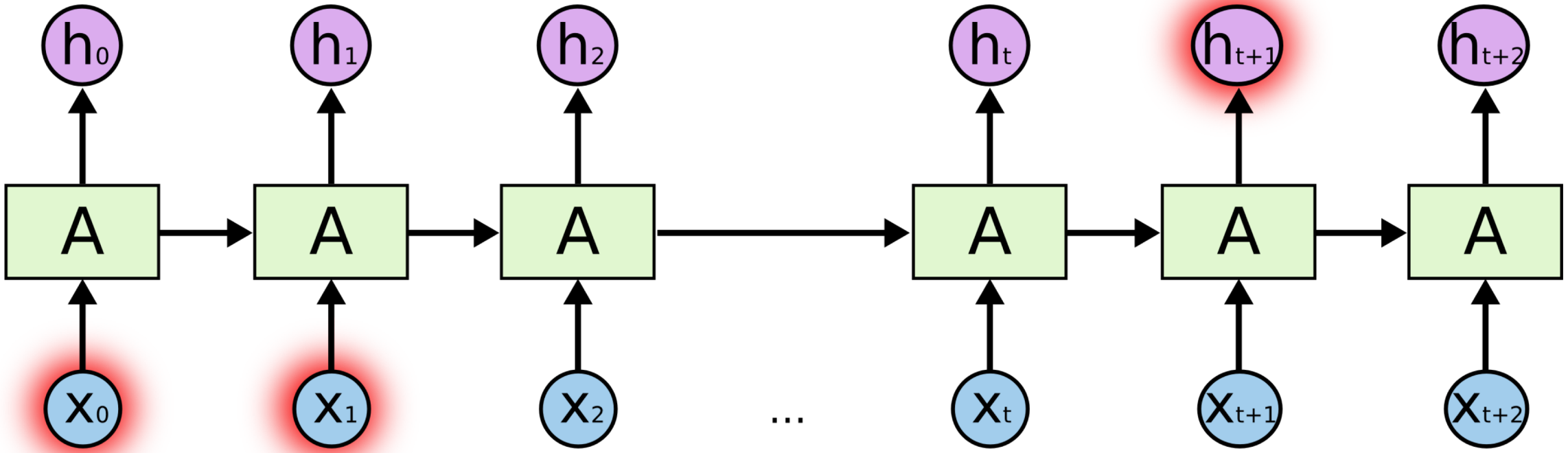
- The goal is to focus on a very important type of VNF: virtualized Provider Edge routers, referenced as vPE from now on.
  - ISPs, etc…

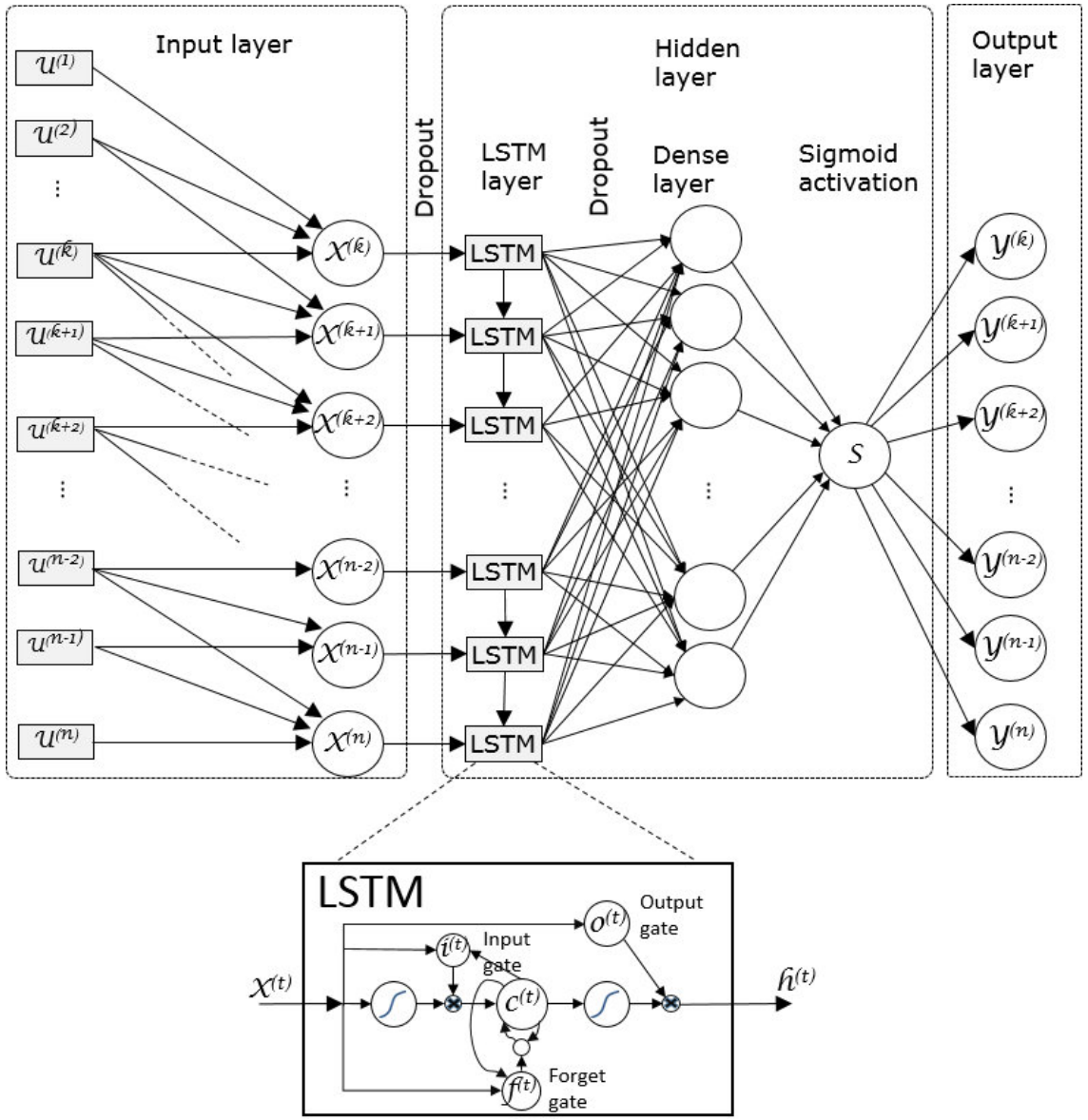
How will it be achieved?

- Using LSTM (Long short-term memory) deep learning model
- Using clustering to aggregate syslogs of similar systems
- Using “transfer learning” to bootstrap a model quickly

WHAT IS A LSTM NETWORK?

- LSTM stands for Long-short term memory and is a type of Recurrent Neural Network (RNN)
  - Recurrent Neural Networks are networks with loops in them, allowing information to persist within them.
  - “Humans don’t start their thinking from scratch every second. As you read this statement, you understand each word based on your understanding of previous words. You don’t throw everything away and start thinking from scratch again. Your thoughts have persistence.”
  - Human thoughts are sequential, and build off each other, rather than each thought being unrelated and independent of one another.
  - If you tell your friend that you are from France, if you are talking to them, and say, “I speak another language,” your mind automatically thinks “French”. This is exactly what an LSTM network accomplishes, by using past information to “learn” and predict information from that.

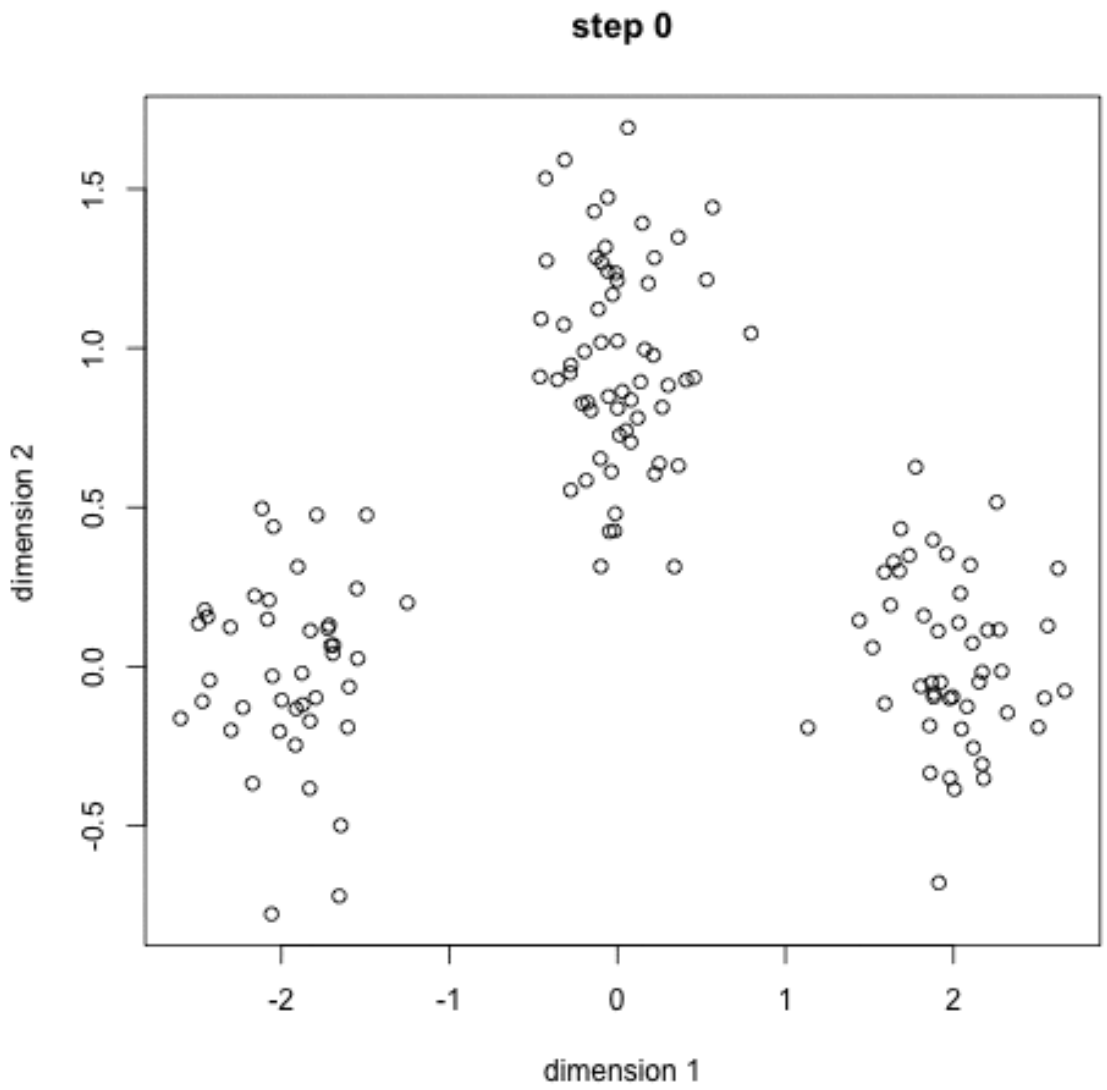

h0
h1
h2
ht
ht+1
ht+2
A
A
A
A
A
A
x0
x1
x2
...
xt
xt+1
xt+2



A DIAGRAM OF A LSTM MODEL

WHAT IS CLUSTERING?

- Clustering is a Machine Learning technique that involves the grouping of data points. Given a set of data points, we can use a clustering algorithm to classify each data point into a specific group. In theory, data points that are in the same group should have similar properties and/or features, while data points in different groups should have highly dissimilar properties and/or features.
- Many different clustering algorithms exist, but we will be focusing on K-Means Clustering, the most popular clustering algorithm
  - K-Means Clustering is the most common clustering algorithm



WHAT IS CLUSTERING: K-MEANS CLUSTERING

- Groups data points into distinct categories
- Calculates a center point for this group of data along a Cartesian plane
- Then, it calculates the distance between each data point and the center of each center point previously calculated and assigns each point to a group in the dataset.
- Finally, it recalculates the center point of each group after all data points are assigned, or until the center point does not change at all, or changes very little.
- Time complexity: $O(n)$
  - *Dependent on how many groups, or K. In this diagram, $k=3$, and $O(3n)$.*

PROS AND CONS OF K-MEANS CLUSTERING

PROS

- Low time complexity
- Minimal calculations
- Easy to understand
- Most common algorithm of clustering

CONS

- Randomly assigned initial center points
- K must be assigned, not automatically deduced
- Possibly less consistent than other clustering methods due to randomly generated center points.

FOR OUR RESEARCH…

IS THERE A BETTER ALGORITHM TO USE RATHER THAN K-MEANS CLUSTERING?

TRANSFER LEARNING

Transfer learning (TL) is a research problem in machine learning (ML) that focuses on storing knowledge gained while solving one problem and applying it to a different but related problem. For example, knowledge gained while learning to recognize cars could apply when trying to recognize trucks. This area of research bears some relation to the long history of psychological literature on transfer of learning, although formal ties between the two fields are limited. From the practical standpoint, reusing or transferring information from previously learned tasks for the learning of new tasks has the potential to significantly improve the sample efficiency of a reinforcement learning agent.

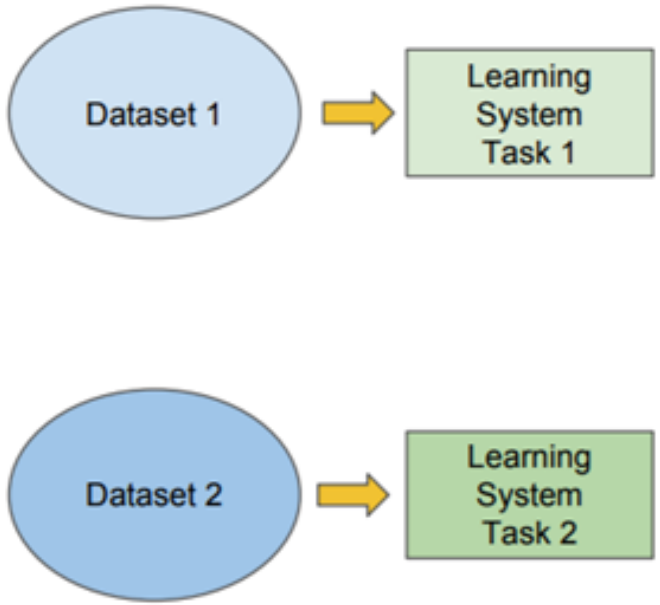
This has great implications for VNFs

No need to build a model from scratch—can use transfer learning in order to bootstrap a model for a VNF

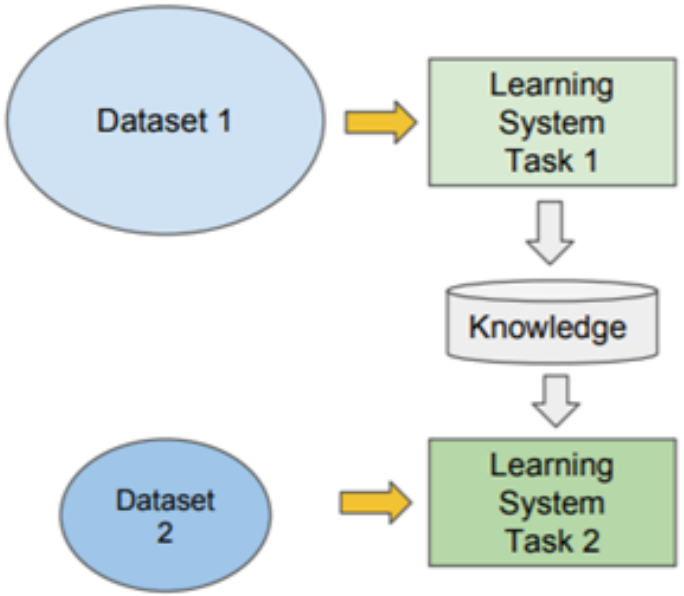
Saves time having to build upon repetitive steps

Traditional ML vs Transfer Learning

- Isolated, single task learning:
  - Knowledge is not retained or accumulated. Learning is performed w.o. considering past learned knowledge in other tasks



- Learning of a new tasks relies on the previous learned tasks:
  - Learning process can be faster, more accurate and/or need less training data



WHY IS TRANSFER LEARNING BENEFICIAL?

- Uses similar approach to LTSM networks: using previously-deduced knowledge in order to help increase accuracy of your model in a shorter time span than would be normally required
- Due to the nature of VNFs and their rapidly updated software, it is crucial to deploy these models as quickly as possible in order to maximize the amount of data that is gathered to train for that specific network
  - Each time the network updates, a new model must be employed, since incremental updates on the software will influence the accuracy of these predictive models
  - This is where transfer learning will be useful: the model can be "bootstrapped" in order to quickly start gathering data to train the model.
  - The more data the model can attain, the better the accuracy of the model in predictions, therefore, it is imperative to collect as much data as possible in order to properly train the model.

WHAT IS THIS LTSM NETWORK-BASED APPROACH BEING USED ON?

- In the paper that was given to me by Dr. Tang, network trouble tickets are aggregated and analyzed in an 18-month period on vPE (virtualized Provider Edge) routers that are deployed in different networks
- By analyzing the data using the aforementioned techniques, network faults can be accurately predicted, and certain failure signatures can be identified and mitigated, resulting in increased uptime and decreased network faults for providers
- This approach utilizes an anomaly detection approach

WHAT IS ANOMALY DETECTION?

- Anomaly detection (also outlier detection) is the identification of rare items, events or observations which raise suspicions by differing significantly from the majority of the data. Typically the anomalous items will translate to some kind of problem such as bank fraud, a structural defect, medical problems or errors in a text. Anomalies are also referred to as outliers, novelties, noise, deviations and exceptions.
- Since fault tickets are so rare in the data sample, an LSTM RNN Network will be utilized in order to ascertain what “normal operation” of the VNF is. Once this information is derived, it will be used to predict signatures of proper operation.
- Any piece of data that falls outside of this prediction will be considered a potential network fault, and its traits will be analyzed in order to find patterns for these network faults.
- This method of fault detection is ideal for VNFs due to the rarity of fault events.

IF WE'RE ALREADY USING LSTM NETWORKS, WHY DO WE NEED CLUSTERING AND TRANSFER LEARNING?

- Both Clustering and Transfer Learning reduce the amount of training data needed in order to customize each LSTM model to each vPE
  - This is beneficial so that less time is needed to set up LSTM Networks and more time can be dedicated to training each LSTM rather than setting up.

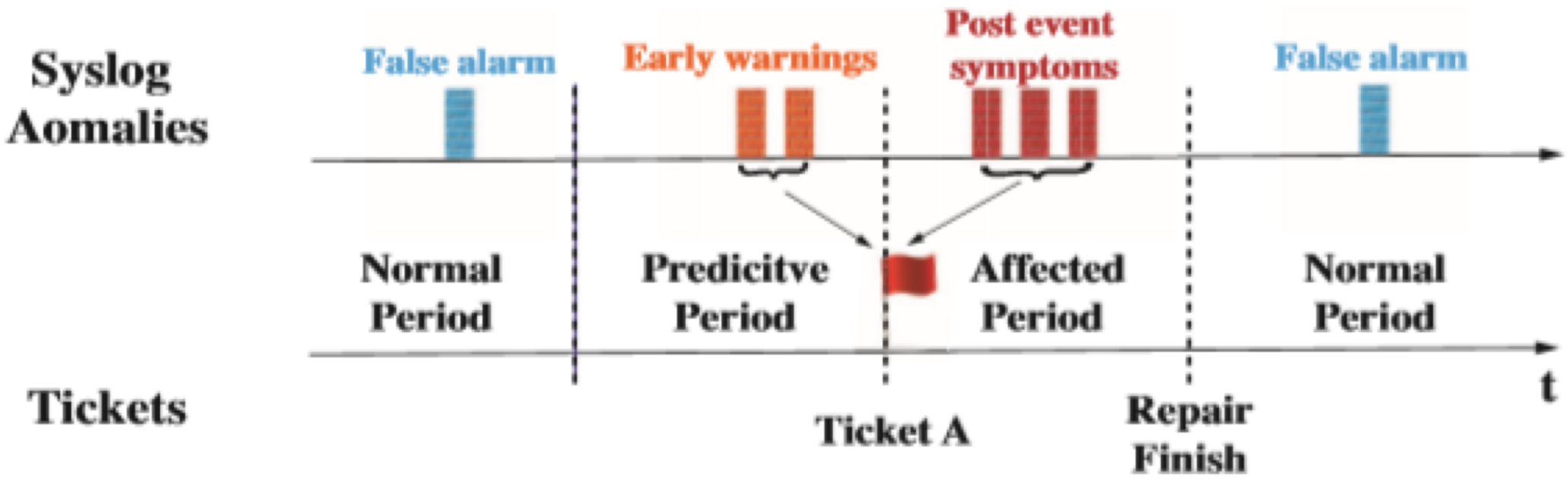
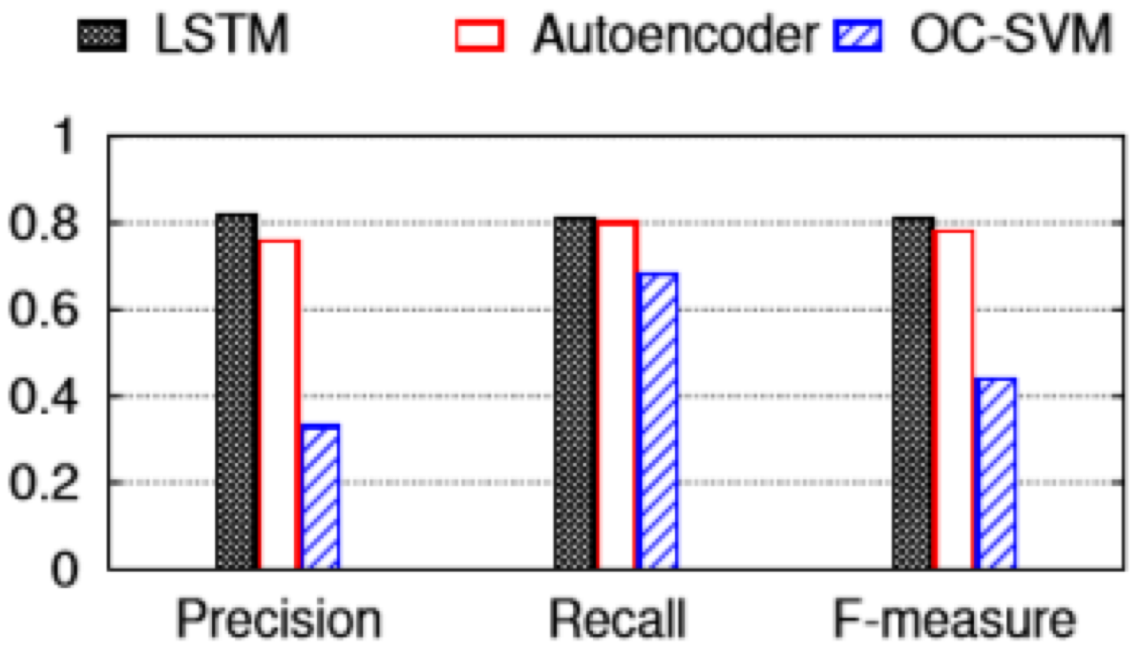


Figure 4: Mapping syslog anomalies to trouble tickets.



RESULTS OF THREE TECHNIQUES FOR PREDICTIVE ANALYSIS