Veena B. Mendiratta

Nokia Bell Labs

Naperville, Illinois, USA

veena.mendiratta@noki-bell-labs.com

phone: +1 630 740 5669

Anomaly Detection in Networks

1/2 day tutorial proposal

Abstract and Goals

The application of analytics methods to data collected from communication networks provides valuable information about the network state, and for detecting and predicting anomalous behavior in the network. Large volumes of operation data, including textual and log files, are collected from communication networks; and the inherent value of this data is recognized by academics, practitioners as well as network operators, who need easy to use and robust methods to detect and analyze anomalies based on the network data. The anomalies are indicators of vulnerabilities in the network. From an operations perspective, it is important to detect the anomalies and correct the problem (based on knowing the root cause) in a timely manner. The goal of the tutorial is to deliver a balanced mix of theory and hands-on practice.

The first part of the tutorial will focus on introducing analytics methods for network anomaly detection. Next, a real-world case study is presented applying non-parametric machine learning techniques and PCA based methods to detect anomalies in wireless networks. Once an anomaly is detected, message patterns are analyzed for root cause analysis by comparing the message patterns of the anomaly data to those of the normal data to determine where the problems are occurring. Neural network based Kohonen Self Organizing Maps (SOMs) and visual analytics are also used for exploring the anomalous behavior. Data from a 4G network will be used for the analyses. The case study is significant, as communications traffic on wireless networks generates large

volumes of log metadata with hundreds of fields including error codes on a continuous basis across the various servers involved in a communication session. The last third (an approximate estimate, can be modified) of the tutorial will provide a hands-on session where attendees will be guided in the analysis of real log data using the techniques described above, in particular, the use of Kohonen SOMs. The hands-on session will focus on exploratory data analysis and modeling approaches using the provided (real) datasets. The hands-on session will be conducted using: the *R* software environment, the *rstudio* user interface for *R*, and various *R packages*.

Outline

1. **Part I — Concepts and Survey**
   - Motivation and introduction - network anomaly detection problem and importance
   - Definition of anomaly and terminology
   - Overview of different anomaly detection techniques with assumptions, advantages, disadvantages, and domain applicability
   - Application of techniques to problem of network anomaly detection
2. **Part II — Case Study**
   - Problem description
   - Data description and EDA
   - Models and results of anomaly detection
   - Visual analytics - display of interactive tool for inferring cause of anomaly
3. **Part III — Hands-on session using real log data and R**
   - Exploratory analysis of the data to understand the data
   - Create SOM models to detect anomaly

Target Audience

The target audience for this tutorial is novice as well as moderately skilled users that have an interest in software failures, anomaly detection, machine learning and/or visual analytics; and are interested in learning to use R for these applications.

Instructor Bio

Veena Mendiratta is the Research Lead for Network Reliability and Analytics at Nokia Bell Labs based in Naperville, Illinois, USA. Her research interests include telecom data analytics, system and network dependability analysis, software reliability engineering, and programmable networks (SDN) resiliency. Current research is focused on network reliability and analytics — architecting and modeling the reliability of next generation programmable networks, and development of analytics-based anomaly detection algorithms for improving network performance and reliability. She has presented tutorials at many IEEE and ACM conferences, most recently at: KDD 2017, ISSRE 2016, 2017 and DSN 2016. She is a member of the SIAM Visiting Lecturer Program, Life Member of SIAM, Senior Member of IEEE, Member of INFORMS; a Fulbright Specialist; and TPC member for several IEEE conferences. She holds a B.Tech in engineering from the Indian Institute of Technology, India, and a Ph.D. in operations research from Northwestern University, USA.

Related Tutorials

1) **Measuring Resiliency through Field Data: Techniques, Tools and Challenges**

Joint with Antonio Pecchia (Critiware) and Marcello Cinque (University of Naples Federico II) at IEEE DSN 2016. Presented the case study described above and led the hands-on R session.

2) **Data Analytics for Software Reliability**

Joint with Catello di Martino (Nokia Bell Labs) at ISSRE 2016. Presented an overview of anomaly detection methods and the case study described above, and led the hands-on R session.

3) Anomaly Detection in Networks

Tutorial presented at ACM KDD 2017, August 2017.

4) Anomaly Detection in Networks

Tutorial presented at ISSRE 2017. The content is similar to that in this proposal; however, for this proposal **the case study is updated** as follows:

- new techniques are introduced for anomaly detection (PCA), and
- new techniques are introduced for root cause analysis (FSM).