Copula-based anomaly scoring and localization of high-dimensional data with application in telecommunication networks

Gábor Horváth[1], Edith Kovács [1,2], Roland Molontay [1,2], Szabolcs Nováczki[3]

[1]Budapest University of Technology and Economics, Budapest, Hungary

[2]University of Debrecen, Debrecen, Hungary

[3]Nokia Bell Labs, Budapest, Hungary

For decades, telecommunication operators have relied on network domain experts to report problems that affect the network performance and customer experience, with the aim of performance monitoring tools. The conventional approach of the tools is to pre-select a set of Key Performance Indicators (KPIs), based on human knowledge and experience of the domain experts. The trending of these KPIs are closely monitored, based on pre-defined single-value or multi-value thresholds, and/or pre-built single variate or multivariate time series profiles. If the KPI values exceed the thresholds/profiles, then alarms will be raised to trigger investigation, mitigation and on-site support. All the processes are manual or semi-auto with limited support of tools. The traditional tools are mainly based on human knowledge, and that the complex and hidden rules in the telecommunication network are not easy to fully detect, capture and utilize.

Thus there is a strong need to revolutionize network management with AI/machine learning technology. Automated solutions are needed that are capable to analyze the raw data and draw conclusions, generate actionable insights using AI technology. One important area in this field is the predictive detection of anomaly patterns that appear in the data.

Here we present an anomaly detection method that does not only indicate whether an observation is anomalous or not but also tells us what exactly makes an anomalous observation unusual. The proposed approach is model-based; it relies on the multivariate probability distribution associated with the observations. Since the rare events are present in the tails of the probability distributions, we use copula functions, that are able to model the fat-tailed distributions well. The presented procedure scales well; it can cope with a large number of high-dimensional samples.

We demonstrate the usability of the method through a case study, where we analyze a large data set consisting of the performance counters of a real mobile telecommunication network. Since such networks are complex systems, the signs of sub-optimal operation can remain hidden fora potentially long time. With the proposed procedure, many such hidden issues can be isolated and indicated to the network operator.

Keywords: anomaly detection, telecommunication network

REFERENCES

[1] C Chow and Cong Liu. 1968. Approximating discrete probability distributions with dependence trees. IEEE Transactions on Information Theory 14, 3 (1968), 462–467.
[2] Rémi Domingues, Maurizio Filippone, Pietro Michiardi, and Jihane Zouaoui. 2018. A comparative evaluation of outlier detection algorithms: Experiments and analyses. Pattern Recognition 74 (2018), 406–421.
[3] Maria Erdmann. 2018. Unsupervised Anomaly Detection in Sensor Data used for Predictive Maintenance. Ph.D. Dissertation.
[4] Edith Kovács and Tamás Szántai. 2017. On the connection between cherry-tree copulas and truncated R-vine copulas. Kybernetika 53, 3 (2017), 437–460.
[5] Hans-Peter Kriegel, Peer Kröger, and Arthur Zimek. 2009. Clustering high-dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering. ACM Transactions on Knowledge Discovery from Data (TKDD) 3, 1 (2009), 1.
[6] Arthur Zimek, Erich Schubert, and Hans-Peter Kriegel. 2012. A survey on unsupervised outlier detection in high-dimensional numerical data. Statistical Analysis and Data Mining: The ASA Data Science Journal 5, 5 (2012), 363–387.