ProphetStor

Performance Prediction and Anomaly Detection Using Deep Learning

Performance Prediction and Anomaly Detection Using Deep Learning

Table of Contents

Introduction

As more and more software services and solutions are deployed in the cloud in recent years, not only does it increase the complexity of IT infrastructure design in data centers, but it also increases the difficulty of data center operation and management. With the rapid advance of artificial intelligence (AI) technology, the use of AI can greatly enhance the operation and maintenance of data centers.

With the progress of artificial neural networks (ANN) and deep learning techniques, many applications have been successfully implemented using AI recently, such as image recognition and speech recognition. This article describes how the data science team at ProphetStor uses ANN and deep learning technologies to predict the performance of data center hardware components, virtual components, and applications. The performance prediction model is further used to develop anomaly detection of performance metrics. We will use disk performance metrics as an example in this article.

Dataset of Disk I/O metrics

Figure 1 shows six disk performance metrics of a hard disk used for a digital streaming service. From top to bottom, they are read/write throughput (read bytes, write bytes), read/write IOPS (reads, writes), and read/write latency (read time, write time). The data for each metric is divided into two groups, training data and test data, where training data covers observations for about 30 days and test data covers observations for about seven days. The data size for each observation is generated over one hour from real workload. Each metric's workload is represented as a time series. Given such real workloads and their corresponding time series, the technical challenge is then to find the right model to predict various disk performance metrics accurately.

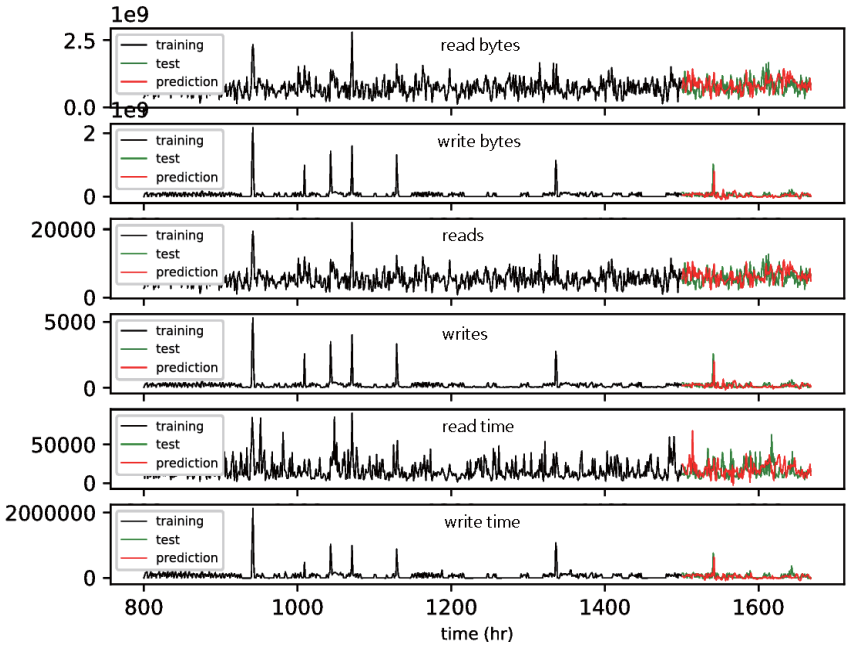

Figure 1: Disk I/O Performance Prediction by LSTM

Performance Prediction

We use the Long-Short Term Memory (LSTM) to predict performance metrics. LSTM is a type of Recurrent Neural Network (RNN) which is usually used to model sequence data. To generate the prediction for a performance metric, a model is first trained by feeding its historical data into the LSTM algorithm. The performance prediction is then generated by feeding new data into the trained LSTM model. In particular, these six metrics are considered as six variables, and the multivariate LSTM model is used to learn the association between the variables to obtain better prediction accuracy. One can observe from the results that our approach performs very well - the RAE (relative absolute error) of the performance results is around 1.

Similar machine learning approaches can be applied to other metrics important to IT operation, such as CPU usage, memory usage, network traffic, and power consumption, etc. Such prediction results can help IT administrators plan and allocate data center hardware resources in a better way.

Performance Anomaly Detection

Furthermore, the prediction results can be used to detect performance anomalies. For example, Figure 2 shows for a given performance metric that an observed value P significantly deviates from its corresponding predicted value. Since the predicted values are inferred from the historical data, the observed value P can be considered an anomaly.

However, this simple approach can generate too many false alarms when only using the difference between the observed value and the predicted value. To avoid this problem, we further enforce the anomaly detection by comparing the difference between the observed value and previously observed values over a certain time interval. For example, as shown in Figure 2, the time index of observed value P is T2, and we compared P with the observed values over a previous time interval (T1, T2) as well. If (1) the difference between the observed value and its corresponding predicted value is sufficiently large, and (2) the difference between the observed value and previously observed values (over a time interval) is sufficiently large, then we determine that the observed value is abnormal, thus reducing frequent false alarms.

For example, as shown in Figure 2, if only the difference between the observed value and the predicted value is used, the observed value P, the observed value Q, and the observed value R are all determined to be abnormal. By further calculating the differences between the foregoing observed values (P, Q, and R) and their corresponding previously observed values, only the observed value P is judged to be abnormal. This is because among P, Q, and R, only P satisfies the criterion that the difference between the observed value and previously observed values is sufficiently large.

Clearly, one can find variations of different criteria in determining when an observed value is abnormal. Our model allows users to have the flexibility of changing to their own criteria easily.

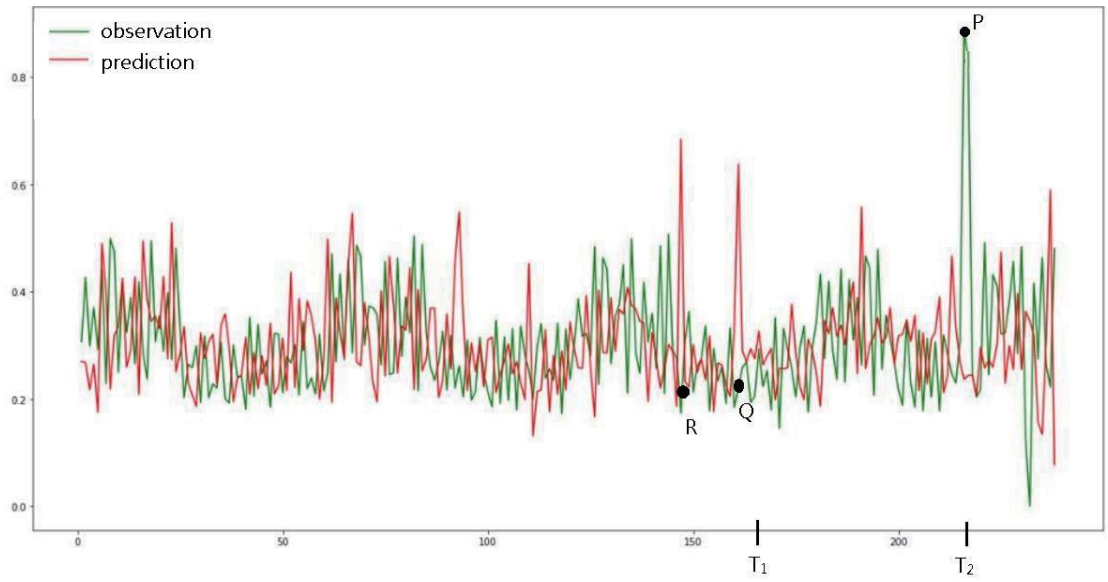

Figure 2: Performance Anomaly Detection

Why we adopt deep learning for performance prediction?

In the era of big data, web-scale services, and microservice applications, IT operation accompanied with the astonishing data growth becomes a bottleneck for a business to grow. It has become a mounting challenge for IT administrators to solve IT environment incidents that happen every day. Many metrics are generated from different IT layers, such as web requests in the application layer, virtual machines in the virtualization layer, and disk health data and physical memory data in the infrastructure layer.

By using only traditional algorithms, such as the MA (moving average) and the ARIMA (Autoregressive Integrated Moving Average) models, dynamic performance metrics cannot be accurately modeled. IT administrators faced the dilemma of handling too many false alarms generated by these algorithms, or they missed the true signals that indicate performance issues after turning off the anomaly detection based on traditional, less-effective algorithms. As the example shows in this article, the deep learning algorithm, enhanced by ProphetStor and implemented in Federator.ai®, not only accurately predicts performance but also effectively reduces false alarms to indicate performance anomalies.

Learn more

To learn more about ProphetStor AIOps solutions, visit us at https://prophetstor.com/federator_ai/