

1. Tracking down the Villains: Outlier Detection at Netflix

This post describes how the reliability team cluster analysis, which is an unsupervised machine learning technique, determines which servers are not performing as well as others, which may cause streaming to fail and trigger a flood of calls from customers.

An unhealthy server will respond to health checks and show normal system-level metrics, but will still operate in a sub-optimal state. A slow or unhealthy server is worse than an offline server because its effects can be small enough to stay within the monitoring system's tolerances and be overlooked by the on-call engineer on duty looking at graphs.

The goal of cluster analysis is to group objects in such a way that objects in one cluster are more similar to each other than in other clusters. The advantage of using an unsupervised method is that they do not need to have labeled data, i.e. they do not need to create a training dataset containing outlier examples.

Their implementation is based on a mini-batch approach where they collect a window of data and use this to make a decision. Compared to the real-time approach, this has the drawback that the outlier detection time is closely related to the window size: too small and has noise, too big and detection time suffers.

To evaluate the effectiveness of this method, they evaluated the results of a production service with outlier detection enabled. Using one week's worth of data, they manually determined whether to classify the server as an outlier and remediated. They then matched those servers against the results of outlier detection system in Netflix. It revealed 90% of accurate results are guaranteed, which is not a little when there are tens of thousands of servers in the company. In addition, Netflix's cloud infrastructure is growing every day, and automation of operational decisions allows for increased availability and reduced human intervention.

Using machine learning and automated responses to improve the efficiency of site reliability engineers and on-duty developers. Server outlier detection is one example of such automation.

2. The development of technology is of great public value, which allows government departments to use data between agencies and IT departments.

These technologies, whether data mining or sophisticated middleware, are making three transformational changes—they can make it easier for citizens to access services, they make it easier to solve problems for field workers, and they provide the foundation for answering big, predictive questions through data analytics.

The federal and state governments have recognized that the collection of personal data carries the risk of being mishandled, used, or even stolen. To protect personal

privacy, there are a number of rules governing the sharing of various data sets such as medical history, educational records, and personal identifiers.

Part of the problem is not that integrating services through technology is actually against the rules, but the simple perception that privacy laws can be violated.

In 2007, New York City launched an initiative HHS-Connect that aimed to reduce bureaucracy for people applying for services and assist city agencies with providing services. The idea is that people can go to various social services agents without having to enter information and fill out duplicate paperwork. And a federal agency can save the time, cost, and resources needed to create a system that meets requirements by deploying a single system that all agencies can implement immediately. Using a single system also eliminates the need to integrate different systems and test them separately.

For example, a homeless shelter does not need to resubmit personal information if the person is already in the system. CARES is a unified case management system designed to enable DHS and its providers to serve and anticipate clients' prospective referrals prior to placement in a shelter and back to the community. The admissions officer uses CARES as the source of client data for processing applications for the shelter. As soon as the client is in the admissions department, they are interviewed by a social worker who uses CARES to initiate the application for temporary housing. Daily paper records of attendance are maintained by shelter providers and updated later in CARES. Once a client is placed in permanent housing and leaves the shelter, DHS staff updates the client's information in CARES. The system also works with partner search agencies, New York City and the Department of Children's Services.

CARES are extremely important for processes occurring in one system and the consolidation of doubly recurring systems, as well as for supporting information flow management processes using various input options. CARES allows the caseworker to enter information electronically, scan the necessary documents, the manager and other responsible person to create standard reports that extract information from CARES.

3. Context: In my daily life, I use many different Internet resources where I need to enter login and password data. I use a single system for storing and registering passwords in my Google account

Purpose: to reduce the registration time as well as the time to select the correct login and password

Data: login and password for certain services.

Description: Using this function, I can quickly register on most Internet resources using a Google account, which reduces the time for registration and allows you to store data in one place