Paper：Rapid Deployment of Anomaly Detection Models for Large Number of Emerging KPI Streams

快速部署大量新兴KPI流异常检测的模型，无需手动算法选择，参数调整或新的任何新出现的异常KPI流标记。我们提出第一个框架解决问题的ADS（通过自我训练进行异常检测）上述问题，通过聚类和半监督学习。

方法：KPI 流曲线形状有很多相似之处，先识别类似流形状的KPI，并将大量流分组到少数集群，这样就减少了异常检测的开销。文章采用ROCKA ，一种快速聚类

KPI流的算法基于它们的形状。它应用移动平均线来提取成功减少的基线噪音和异常的特殊情况。

ADS提出（1）将所有现有/历史KPI流集群到集群中，（2）手动标记所有聚类质心的异常，（3）将每个新出现的KPI流归类到现有的一个群集，以及（4）合并新KPI流（未标记）的数据和它的集群质心（标记），并使用半监督学习为每个新KPI流训练新模型。

关于时间序列分类：T. W. Liao, “Clustering of time series data—a survey,” Pattern recognition,vol. 38, no. 11, pp. 1857–1874, 2005.

Paper：A review on time series data mining

时序数据挖掘综述

总结了一个关于时间序列分类的大量方法，其中大多数是专为平滑和理想化的数据而设计。 但是，大KPI流中的峰值，下降和级别变化的数量可以显着改变KPI流的形状，这些方法不适用。

Paper：Unsupervised Anomaly Detection via Variational Auto-Encoder for Seasonal KPIs inWeb Applications

一种基于VAE的异常检测算法。

Paper：Robust and Rapid Clustering of KPIs forLarge-Scale Anomaly Detection

Paper 1的KPI 聚类方法

KPI 检测现有代码库：

<https://github.com/haowen-xu/donut> 基于VAE 的KPI异常检测

<https://github.com/hamzakeurti/anomaly-detectives-kpi> 清华课程中的KPI 异常检测

多维异常检测算法：<https://github.com/ymirsky/KitNET-py> 基于自动编码器集合的在线异常检测器