智能运维系列（三）| 浅析智能异常检测：“慧识图”核心算法

智能运维系列专题简介：智能运维（AIOps），根据 Gartner 的最新阐释，意指整合大数据和机器学习能力，通过松耦合、可扩展方式去提取和分析数据量（volume）、种类（variety）和速度（velocity）这三个维度不断增长的 IT 数据，进而为 IT 运维管理产品提供支撑。在此，微众银行智能运维团队根据一线工作的实践经验与心得体会，特别撰写了《智能运维系列》文章，敬请持续关注。

[点击回顾：智能运维系列（一） | AIOps 的崛起与实践](https://www.infoq.cn/article/fqUfkjhecOla1zKUKycN)

[点击回顾：智能运维系列（二）| 智能化监控领域探索](https://www.infoq.cn/article/Qta6VCyjvHdoiJg5wKze)

微众银行在智能化运维领域的最初目标是异常事件的自动发现和预警，彻底摆脱根据人工经验定义异常阈值，使用机器学习算法实现无阈值 KPI 曲线异常识别。根据这一目标，微众银行规划了智能监控系统识图模块。

“微众银行智能监控系统识图模块”是针对业务四大黄金指标而设计的智能曲线异常检测系统。四大黄金指标包括交易量 (业务实时产生的交易量)、业务成功率 (业务成功量 / 交易量)、系统成功率 (系统成功量 / 交易量, 业务成功量和系统成功量的区别在于是否明确捕捉到系统异常)、平均时延 (交易的平均耗时)。这四大黄金指标都是分钟级数据，因为过去只要影响到业务层面的故障，最终都会在这四个指标上有所体现。因此只要能准确捕捉到这四个指标的异常波动，就可以检测到所有影响业务的异常。这四个指标统计维度不同，波动规律也有所差别，因此需要用不同的算法检测。

识图的检测方法主要有三种：

* 基于 LSTM 与高斯分布的检测，这个算法主要用于交易量和时延的检测。大部分的曲线突变都能准确检测到，但算法的死角在于小幅度长时间的缓慢变化容易被漏掉。
* 基于 k-means 算法的特征检测，主要用于填补第一种算法的盲区, 在交易量缓慢变化的案例效果较好。
* 基于概率密度的检测，主要用于业务成功率和系统成功率的曲线，因为成功率曲线的背后隐藏着无数的可能，需要用一个更接近本质的量来衡量异常的程度。

而以上三种方法都有一个共同的判断原则——少见即异常。在我们确立了无监督为主的大前提下，异常检测的问题转换成了如何衡量当前的情况是否“少见”的问题。

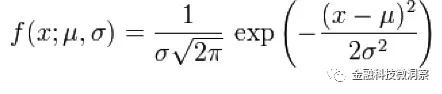
**基于 LSTM 与高斯分布的检测**

该方法分两步：1. 曲线预测 2. 异常判断。曲线的预测，在我们的调研中尝试了 ARIMIA、HOLT-WINTER、LSTM。可能是由于可以学习到长距离信息的缘故,LSTM 在我们的业务指标数据集中表现特别好，对于归一化后的数据，大部分 Loss 在 0.0001 左右。如下图所示，业务正常的情况下, 预测线与实际十分贴合。因此选择了 LSTM 作为我们的预测模型。



图 1  LSTM 拟合效果图

因为预测线是学习了历史数据的规律，因此对于稳定的业务，正常情况下预测线与实际曲线是十分吻合的。当实际曲线偏离预测线时，说明实际曲线已经偏离了历史的规律, 那么这个时候，很大概率就是发生了异常。因此，这里需要一个方法来判断实际曲线是否偏离预测曲线。而实验中发现，正常情况下预测线与实际曲线之间的差呈高斯分布。因此，前文说到的少见的数据就是高斯分布概率的边缘数据。高斯分布的概率密度函数如下：



假设某个 x 计算出来结果特别小，接近 0，那么可以认为该值的出现是一个不可能事件。而当我们用正常规律计算出来的不可能事件出现了，很可能就是出异常了。

上面是基于 LSTM 和高斯分布的检测，该方法在识别短时间内突变情况表现很好，如下图：

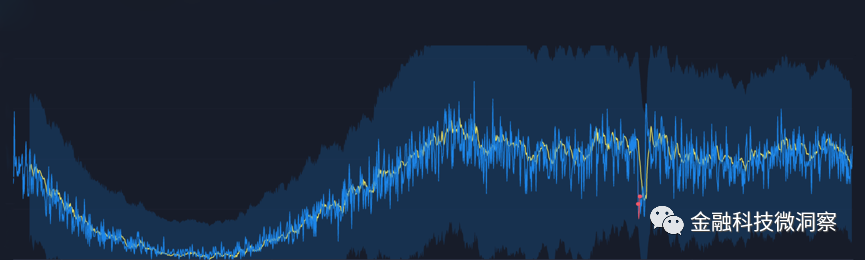


图 2  基于 LSTM 和高斯分布的检测

但是对于幅度较小，延续时间较长的异常识别能力较弱，主要的原因在于 LSTM 的预测线跟得太紧，一旦异常持续时间稍长，预测线会被带偏，这个时候基于 LSTM 和高斯分布的检测的方法就失效了，如下图：

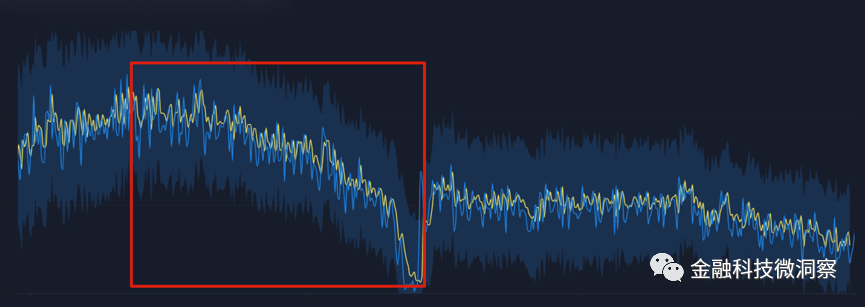


图 3 缓慢下降后异常识别

因此需要有一种可调节时间窗口的算法来弥补前面的缺陷，于是有了第二种补充算法，基于 k-means 的特征聚类检测。

基于 k-means 算法的特征检测

曲线异常检测方面，在业界比较流行的一种做法是提取曲线特征当作输入，然后使用运维同学的标注作为标签训练判断模型。据了解，该方法在完美的数据集中表现很好，但却很少有公司真正把该方法落地使用，主要原因还是因为这个算法需要巨大的标注量。如果要模型复用，减少标注量，需要用到复杂的迁移学习。

因此，这里对该方法做了一个小修改，变成一个超级简化版，用于弥补缓慢变化的盲区。这里提取了四个重要特征：均值、斜率、零值率、一阶差分的均方差。当指标为中高频指标时，均值、斜率在缓慢变化的异常中会体现较为明显。而当指标为低频指标时，零值率、一阶差分的均方差更为明显。当某个窗口的特征明显偏离正常规律, 则该窗口很可能产生异常。如何判断特征偏离正常规律，这里采用的是 K-MEANS 算法。K-means 算法原理是提前设置聚类的个数 k，然后通过多次迭代，最终聚成符合预设个数的 k 个簇。检测的时候，如果当前窗口值是异常，那把当前窗口值与邻近窗口、历史同期窗口一起用 k-means 聚类，设置 k 值为 2，最后的结果很可能如下图：

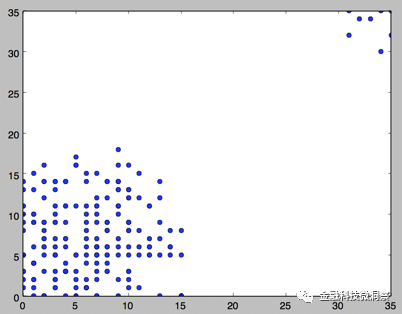


图 4 设置 k=2 的 k-means 聚类



图 5 缓慢下降后的异常识别

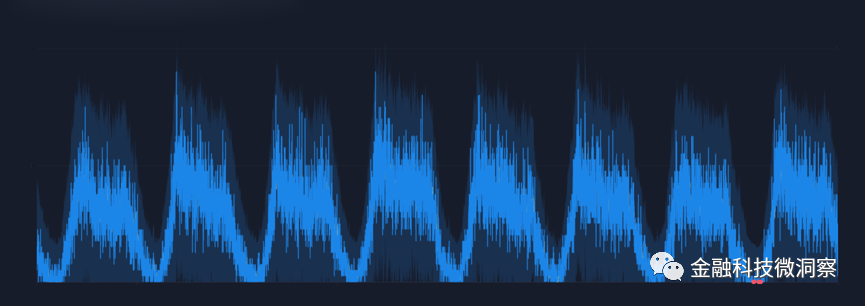


图 6 低交易量成功率检测异常

当前窗口的特征被归到一个簇，邻近窗口及历史窗口被归到另一个簇，根据少见即异常为原则判断该窗口为异常，反之则说明无异常。上述两种检测, 对于大部分规律的业务曲线有较好的检测效果, 但是对于成功率曲线的检测却差强人意，因此我们设计了一种专门针对成功率曲线的检测方法。

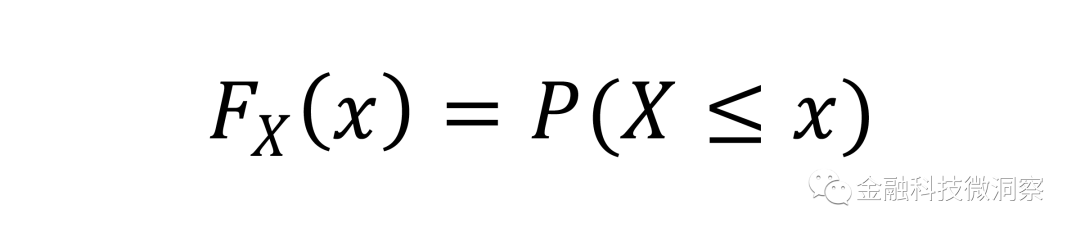
基于概率密度的检测

前两种算法对成功率曲线检测效果不好的原因是，单独一个成功率的值并不能体现出它背后的所有可能，如下图:



图 7 低交易量成功率检测异常

该业务日常成功率为 95%。如果当前发生 1 笔交易，失败了，那么当前成功率是 0%，按照过往规律计算，这是可能发生的，说明不了问题。但如果当前发生 30 笔交易，失败了 15 笔，当前成功率是 50%，但按照过往规律计算，这几乎是不可能事件，极大可能是异常。如果不结合交易量来检测，我们会以为这个 0% 比 50% 更接近异常，但事实并非如此。以下算法便是结合量解决这一问题，计算成功量小于某个值的累积概率。



例如某业务的成功率为 R=95%，当前发生 30 笔交易，其中成功笔数为 x，概率分布如下图所示：

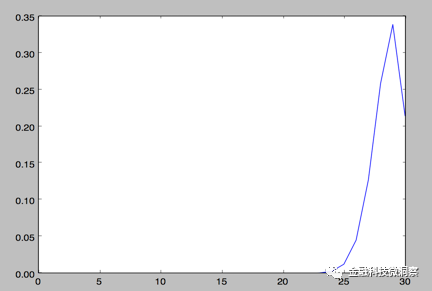


图 8  成功率 R=95% 的概率分布

从图中可以看到, 成功量最大可能是在 29 左右波动。而当成功量掉到 15，可能性已经接近 0。如果出现按照日常规律计算几乎不可能出现的事件，那么最大的可能就是出现异常了。

以上便是微众银行智能异常检测的主体框架，都遵从少见即异常原则。小概率发生的事件就判断为异常，如何划定小概率事件在日常的运维工作中也很重要，宽松了容易漏告，严格了又会造成骚扰。