研究与开发

基于密度聚类的网络性能故障大数据分析方法

李想，李原，张子飞，杨哲
（中国信息通信研究院，北京 100191）

摘　要：针对层出不穷的网络安全事件，如何快速在海量监测数据中发现异常数据，并开展网络故障分析成为研究难点。针对该问题，提出一种基于密度聚类的网络性能故障大数据分析方法，通过熵权分析、数据清洗与标准化处理实现关键性能特征提取与数据整形，基于参数调优的 DBSCAN 聚类算法提取性能故障异常数据。基于实时采集的全国多家运营商海量骨干网链路性能数据验证该算法，结果表明，与人工标注网络性能异常数据相比，其识别的准确性超过 90%，可满足开展全国网络运行故障分析的需求。

关键词：网络性能；机器学习；密度聚类；测量分析
中图分类号：TP393
文献标识码：A
doi: 10.11959/j.issn.1000−0801.2020270

A density clustering-based network performance failure big data analysis algorithm

LI Xiang, LI Yuan, ZHANG Zifei, YANG Zhe
China Academy of Information and Communications Technology, Beijing 100191, China

Abstract: Facing frequent network security incidents, how to quickly find abnormal data in massive monitoring database and carry out network failure analysis becomes a research difficulty. A density-based network performance failure big data analysis algorithm was proposed, which extracted key performance characteristic indicators through entropy weight analysis, implemented data shaping through data cleaning and standardization, and extracted abnormal performance data on the basis of DBSCAN clustering algorithm. Relying on the real-time massive backbone network link performance data of multiple domestic operators to validated this algorithm, the results shows that compared with the manually manner, the recognition accuracy of the algorithm proposed to the network performance abnormal data is more than 90%, which can well fit for the analysis of real-time Internet network operation failure.

Key words: network performance, machine learning, density clustering, measurement analysis

1　引言

随着网络的飞速发展，用户需求的不断增加，网络应用越来越广泛。与此同时，网络安全事件层出不穷，日益突出，特别是网络层面的故障会导致众多互联网用户无法正常访问网络应用，产生较为严重的影响。例如，2015 年 3 月 20 日，由于中国联通骨干网一台设备发生故障，导致联通

收稿日期：2020−05−12；修回日期：2020−06−01
通信作者：李原，liyuan@caict.ac.cn

网络出现一定范围瘫痪，上海、天津、河北等多个省份均出现不同程度的网络中断现象；2019 年 5 月 20 日，由于网络设备故障原因，致使重庆主城及合川、丰都等区县电信用户受到影响，无法正常使用网络。除了这些网络故障外，通过网络性能监测还会发现一些性能趋势下降的问题，可能是由于带宽利用率较高或路由调度错误导致，也应及时预警提示相关企业关注，避免后期出现更严重问题。面对这些层出不穷的网络安全事件，国内外研究机构或企业已开展了大量监测研究，但国内企业监测范围相对局限，少有对全国骨干网开展监测及故障分析，且实时性较差，特别是如何在海量监测数据中发现异常数据，并开展网络故障分析成为一个难点。因此，本文依托中国信息通信研究院互联网监测分析平台的全国网络监测数据，通过无监督机器学习算法，建立网络性能异常数据提取模型，并针对异常数据进行二次挖掘分析，定位我国互联网网络性能故障，为我国互联网的健康安全运行提供支撑。

2 基于密度聚类的网络性能故障大数据分析方法研究

本文依托互联网监测分析平台提取的2018年我国 31 省 3 家基础电信企业网络性能监测数据，由于地理位置及网络层级等原因，国内端到端性能数据异常门限无法统一设定，例如网络时延同为 80 ms，若为北京到西藏则属于正常范围，而放在北京到天津则可能为异常，故需对每条端到端链路建立模型提取出性能异常数据，再对相关异常数据进行整合分析。

2.1 模型特征选择

常见的机器学习方法包括了监督学习和无监督学习两种，监督学习是从标记的训练数据中来推断一个功能的机器学习任务。无监督学习更适用于具备数据集但没有标签的情况，无须训练过程，而是直接对数据进行建模分析[1]。本文依托平台获取的监测数据记录约 1.5 亿条，并无相关故障标签，且手工打标签工作量巨大，故本文选择无监督的学习方法对数据进行聚类分析。在无监督学习建模过程中，选择合适的、有效的特征可简化建模流程，提升模型输出效果。针对平台提取的性能数据，可利用的特征包括骨干链路的时延、分组丢失率、跳数 3 个指标，本文通过基于信息论的熵权分析方法对 3 个特征指标进行权重分析。

基于信息论的熵权分析是利用熵值法衡量各指标的重要性，其本质是利用该指标信息的价值系数计算权重。若某个指标的信息熵越小，表明指标值的变异程度越大，提供的信息量越多，在综合评价中的权重也就越大[2]，具体如式（1）所示：

$$w_j = \frac{1-\mathrm{e}_j}{\sum_{i=1}^{m}(1-\mathrm{e}_j)} \tag{1}$$

其中：

$$\mathrm{e}_j = -K\sum_{i=1}^{m} y_{ij}\ln y_{ij} \tag{2}$$

$$y_{ij} = \frac{x_{ij}}{\sum_{i=1}^{m} x_{ij}}(0 \leqslant y_{ij} \leqslant 1) \tag{3}$$

本文选取基础电信企业网内、基础电信企业网间、基础电信企业到其他电信企业 3 条有代表性链路的基础数据，并对时延、分组丢失率、跳数 3 个指标进行熵权分析，得出 3 个指标的权重，3 个指标熵权分析结果见表 1。

表 1 3 个指标熵权分析结果

链路方向	时延指标权重	分组丢失率指标权重	跳数指标权重
上海电信-上海电信	0.99	0.005	0.005
上海电信-四川联通	0.011 3	0.987 9	0.000 8
宁夏电信-安徽鹏博士	0.191 5	0.726 4	0.082 1

通过熵权分析，不同链路各指标特征所影响的比重略有不同，综合来看，分组丢失率和时延两个特征重要程度较高，而跳数特征权重占比均

很小。故为简化模型，选取分组丢失率和时延两个特征进行后续的算法建模。

2.2　模型数据处理

数据处理包括数据清洗和数据标准化两部分工作。

（1）数据清洗

数据清洗是对数据进行审查和校验的过程，目的在于删除重复信息、纠正存在的错误，并提供数据一致性。包括缺省值处理、异常值处理、重复值处理以及去周期处理 4 个步骤。

步骤 1　缺省值处理。对数据记录中出现空值数据进行处理，出现缺省值可能是由平台没有正常采集到性能数据导致的。对于时延指标缺省但分组丢失率为 100%的记录，由于网络不通确实无法获取到时延指标，此时采用插补法，将时延值标记为−1；对于其他缺省的记录，考虑到训练样本数据量级大，且经统计此类缺省值出现次数极小，采用直接删除的方法。

步骤 2　异常值处理。对数据中的出现一些异常数据进行处理，如性能数据超出正常范围阈值的记录，采用直接删除的方法将异常值剔除。

步骤 3　重复值处理。对数据中的出现重复现象进行处理，出现重复值可能是平台程序重复启动或在入库阶段出现问题导致。采用合并法，通过判断记录间的属性值是否相等，将相等的记录合并为一条记录。

步骤 4　去周期处理。由于端到端链路性能数据受网络忙闲时影响呈现明显周期性变化，可能会影响性能发展趋势判断，所以需要对数据进行去周期处理。常见去周期化方法有对数变换法、平滑法、差分法、分解法等。本文采用分解法对原始性能数据进行成分分离，形成长期趋势以及周期性趋势。

（2）数据标准化

数据的标准化是将数据按比例缩放，使之落入一个小的特定区间。主要应用于去除数据的单位限制，将其转化为无量纲的纯数值，便于不同单位或量级的指标能够进行比较和加权。其中最典型的就是数据的归一化处理，即将数据统一映射到[0,1]上。数据标准化可以提升模型的收敛速度，还可以提升模型的精度。

本文采用 3 种标准化方法进行实验比较。

- 两个指标均采用 z-score 标准化法。对时延和分组丢失率两个指标采用 z-score 标准化法，根据式（4）进行数据的标准化[3]。

$$x^* = \frac{x-\mu}{\delta} \tag{4}$$

其中，μ 和 δ 分别为原始数据的均值与标准差。

- 两个指标均采用最大最小标准化法。对时延和分组丢失率两个指标采用最大最小值法，将两个指标值映射到[0,1]上。
- z-score 标准化法和最大最小标准化法两种方法混合运用。对时延指标采用 z-score 标准化法，对分组丢失率指标采用最大最小值法。

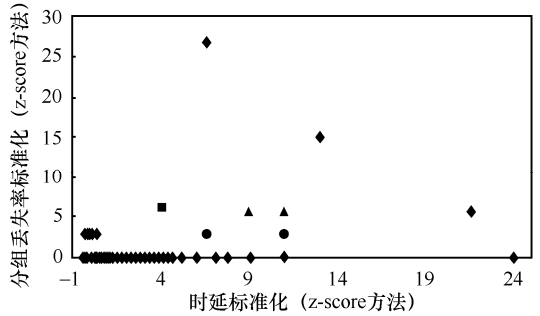
本文选取链路“上海电信–广东移动”进行分析，按照上述 3 种方法标准化后，再做密度聚类，3 种标准化方法聚类效果如图 1 所示。

采用相似性评估方法，对聚类效果的紧密性和间隔性进行比较[4]，并结合人工判断，可知第一种两个指标均采用 z-score 标准化方法得到的效果最好，故在数据处理阶段本文选取两个指标均采用 z-score 标准化法对数据进行标准化处理。

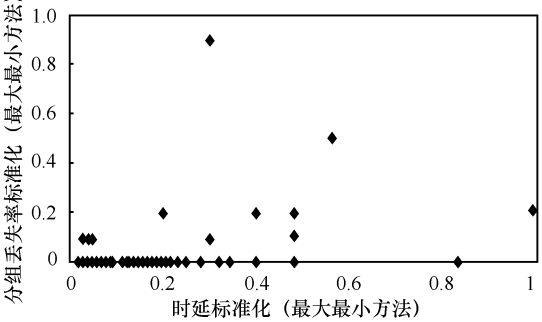
2.3　聚类模型算法及参数调优

常用的无监督学习方法有 *K*-means、层次聚类、DBSCAN、谱聚类等[5]。

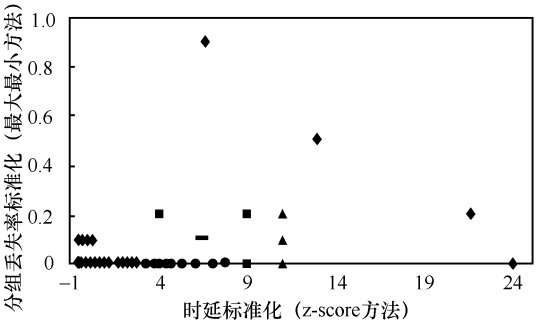
- *K*-means 方法原理简单，计算收敛速度较快，以 k 为参数，把 n 个对象分成 k 个簇，使得每个簇的内部具有较高的相似性[6]。但该算法需要设置参数 k 和初始的聚类中心，且对噪音和异常点比较的敏感。



（a）z-score方法标准化后聚类效果



（b）最大最小方法标准化后聚类效果



（c）z-score方法和最大最小方法混合使用标准化后聚类效果

图 1　3 种标准化方法聚类效果

- 层次聚类方法是把每一个对象看作一个簇，然后根据簇间的距离合并这些原子簇而形成一个大的新簇，直至满足终结条件[7]。但该算法效率相对较低，且结果不具有再分配的能力。
- 具有噪声的基于密度的聚类算法[8]（density-based spatial clustering of applications with noise，DBSCAN）将簇定义为密度相连点的最大集合，能够把具有足够高密度的区域划分为簇。该算法可以对任意形状的稠密数据集进行聚类，并可以在聚类的同时发现异常点，对数据集中的异常点不敏感，但在样本集较大时，该算法聚类收敛时间较长。
- 谱聚类是从图论中演化出来的算法。它的主要思想是把所有的数据看作空间中的点，通过对所有数据点组成的图进行切图，让切图后不同的子图间边权重和尽可能的低，而子图内的边权重和尽可能的高，从而达到聚类的目的[9]。但该算法聚类效果依赖于相似矩阵，不同的相似矩阵得到的最终聚类效果可能不同。

根据平台的数据特点以及上述 4 种聚类算法的特性，本文采用 DBSCAN 算法挖掘网络中的异常性能数据。DBSCAN 算法的重要参数分为两类，一类是 DBSCAN 算法本身的参数，包括了半径和最小样本数，另一类是最近邻度量的参数[10]。

- eps（半径）：表示以给定点 P 为中心的圆形邻域的范围。默认值是 0.5，如果设置过大则更多的点会落在核心对象的邻域，但类别数可能会减少，本来不应该是一类的样本也会被划为一类。反之则类别数可能会增大，本来是一类的样本却被划分开[11]。本文通过实验，初步设定 eps 为 0.8。
- min_samples（最小样本数）：以点 P 为中心的邻域内最少点的数量。默认值是 5，通常和 eps 一起调参。在 eps 一定的情况下，min_samples 过大，则核心对象会过少，此时簇内部分本来是一类的样本可能会被标为噪音[12]。由于本模型算法主要目的是提取出性能异常数据，相对数量记录较少，故此参数不宜设置过大，本文选取算法默认值，设置为 5。
- 最近邻距离度量：可以使用的距离度量方法较多，一般 DBSCAN 算法默认使用欧氏距离方法，此外还有曼哈顿距离、切比雪夫距离等[13]。由于本文已经将数据进行了标准化处理，所以选择欧氏距离，对本模型分类效果最好。

DBSCAN 算法原文提出了一个参数调优的方法，即给定 K 邻域参数 k，对于数据中的每个点，计算对应的第 k 个最近邻域距离，并将数据集所有点对应的最近邻域距离按照降序方式排序，称这幅图为排序的 k 距离图，选择该图中第一个谷值点位置对应的 k 距离值设定为 eps。一般将 k 值设为 4。

本文尝试了若干种数学方法去找到曲线中合适的“拐点”。

（1）曲线多项式[14]拟合并求差值

由于通过模型计算后获得到的距离曲线并不规则，不能得到一个较好的拟合曲线，所以对于拟合效果不太好的曲线计算原始数值与拟合曲线上数值之间差值，将最大差值对应点视作拐点。

（2）指数拟合并求导

对通过模型计算后获得到的距离曲线进行多项式与指数拟合，对拟合点的相邻点进行二次求导取对应的为 0 的点视作拐点。

（3）计算相邻点之间斜率

截取曲线最右侧（距离顺序展示）的 100 个点（根据测试样本数据量大小选择这个数），以第一个点作为参考点，计算各点到参考点之间斜率。比较这 100 个点的斜率，选择最后出现的具有局部最小值特征的点作为拐点。

本文仍然抽取上文所述的基础电信企业网内、基础电信企业网间、基础电信企业到其他电信企业 3 条有代表性链路性能数据，绘制出每条链路 k 曲线，以上文 3 种方式计算得出拐点，得出 DBSCAN 算法的 eps 参数。并针对得出的 eps 参数获取聚类效果，进行评估比较，具体结果见表 2。

表 2　3 种 eps 赋值方法结果

链路方向	eps 值			
	默认值 0.8	曲线多项式拟合并求差值	指数拟合并求导	计算相邻点之间斜率
上海电信–上海电信	0.8	0.127 2	0	0.127 2
上海电信–四川联通	0.8	3.541 5	3.325 3	3.340 7
宁夏电信–安徽鹏博士	0.8	0.864 2	0.746 0	0.842 0

由于宁夏电信－安徽鹏博士链路，3 个方案得出的 eps 值都接近于 0.8，故与默认参数聚类效果几乎一致，偏差率仅为 0.03%。但其他两个方向链路与之前默认参数相差较大，本文继续对聚类效果进行评估。针对上海电信－上海电信链路，根据不同 eps 值，聚类效果明显有差异，偏差率超过 90%，但通过人工核验，明显 eps=0.8 的参数更符合本文算法，上海电信－上海电信不同 eps 聚类效果如图 2 所示。

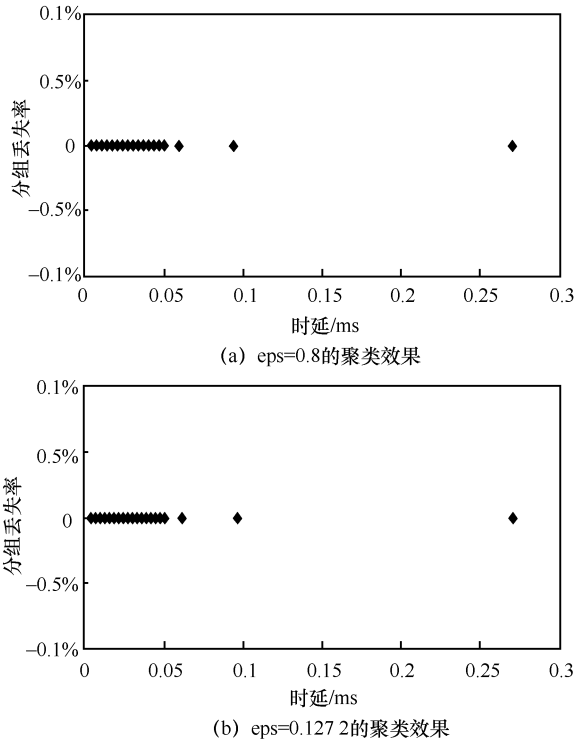
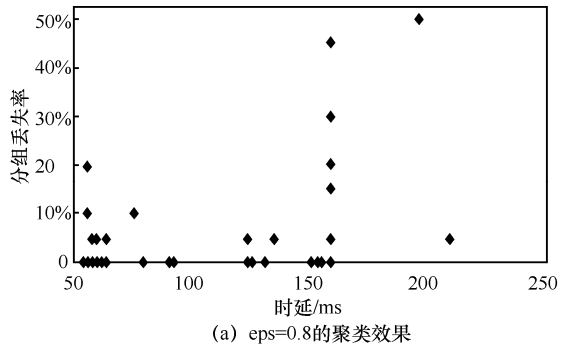


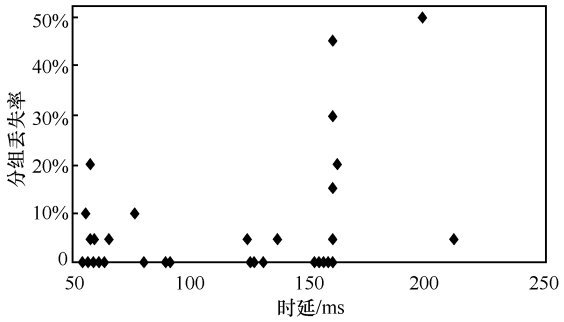
图 2　上海电信－上海电信不同 eps 聚类效果

针对上海电信–四川联通链路，根据不同 eps 值，聚类效果有一定差距，但相差不大，偏差率仅为 0.45%，上海电信－四川联通不同 eps 聚类效果如图 3 所示。

通过选取其他链路进行上述实验，发现对于不同的端到端链路，参数的计算结果有所不同，大多数链路通过 3 种方法计算得出的 eps 参数相差不大，仅有个别链路计算得出的参数带入 DBSCAN 算法发现分类有明显错误。故本文对每条链路主要采用曲线多项式拟合并求差值的方法



(a) eps=0.8的聚类效果



(b) eps=3.541 5/3.325 3/3.340 7的聚类效果

图 3　上海电信－四川联通不同 eps 聚类效果

计算 eps 参数，但针对个别链路该方法出现异常时，采用默认值 0.8 进行补偿计算。

2.5　模型实验结果

本文抽取了 5 条端到端链路性能监测数据依托模型开展异常数据标记，并结合人工经验进行对比验证，模型标注与人工标注对比结果见表 3。

由于性能数据没有故障标签，即使人工判断也可能因为不同研究人员的经验导致不同的标注结果，但整体上通过 DBSCAN 聚类算法，异常数据标注和人工标注相比，准确性在 90%以上。

此外，本文根据样本数据基本特征，选取 *K*-means++ 算法、凝聚的层次聚类方法（agglomerative nesting， AGNES）算法进行结果对比。*K*-means++算法是 *K*-means 算法的改进版，有效地解决了关于初始质心的选取问题，算法参数 *n*_clusters（生成聚类数）设置为 3，最近邻距离度量参数设置为欧氏距离，最终结果选择性能最差的簇作为性能异常数据；AGNES 算法是一种自底向上的层次算法，将每个对象作为一个簇，然后合并这些原子簇为越来越大的簇，直到某个终结条件被满足，算法参数 *n*_clusters 也设置为 3，最近邻距离度量参数设置为欧氏距离，最终结果选择性能最差的簇作为性能异常数据。3 种算法输出结果与人工标签校对后对比见表 4。

通过上述对比发现，由于每条端到端链路性能数据范围不同，散点呈现图形效果也不相同，*K*-means++算法和 AGNES 算法统一参数并不适用于每条端到端链路，特别是当某条链路数据值区间相差较小时，两种算法采用默认的参数设置得出的结果噪声较大，故如果选择其他算法，还需针对端到端链路数据进行个性化分析，设置不同参数，而本文采用的 DBSCAN 算法，已经针对每条端到端链路采用不同的参数值，可以满足后续故障分析研究需求。

表 3　模型标注与人工标注对比结果

对比项	湖南电信–安徽移动	江苏移动–浙江电信	辽宁移动–浙江电信	安徽电信–浙江电信	吉林联通–山西电信
机器识别异常记录数	93/个	5/个	25/个	10/个	49/个
人工判断异常记录数	102/个	5/个	28/个	8/个	45/个
识别准确率	91.2%	100%	89.2%	80%	91.8%

表 4　3 种算法准确率比较结果

对比项	湖南电信–安徽移动	江苏移动–浙江电信	辽宁移动–浙江电信	安徽电信–浙江电信	吉林联通–山西电信
DBSCAN 算法	91.2%	100%	89.2%	80%	91.8%
K-means++算法	88.2%	70%	75%	47.5%	93.75%
AGNES 算法	89.2%	70%	92.8%	42.1%	91.8%

3　网络性能故障分析实例研究

根据 DBSCAN 性能异常提取模型，本文对 2019 年全国网络性能监测数据进行建模，对每条链路开展聚类分析。按照指标影响权重，优先考虑分组丢失率指标，再考虑时延指标，对各类聚类分组数据标签进行重排序，并按照时间维度和地理维度两个方向进行故障分析。

3.1　按时间维度分析研究

针对每条端到端链路，根据标签重新排序后，生成时间性能矩阵，纵坐标为日期，横坐标为小时，以湖南电信−安徽移动 6—7 月份数据为例，时间维度性能矩阵图如图 4 所示。通过图 4 可以直观发现在 6 月 29 日晚上开始到 7 月 9 日上午没有监测数据，经平台核实该段时间受湖南电信监测服务器设备下架机房搬迁影响，虽然不是安全故障导致，但分析结果可以协助运维人员梳理平台自身数据完整性情况。此外不难发现在 6 月部分日期晚上性能较差，可能存在故障隐患。

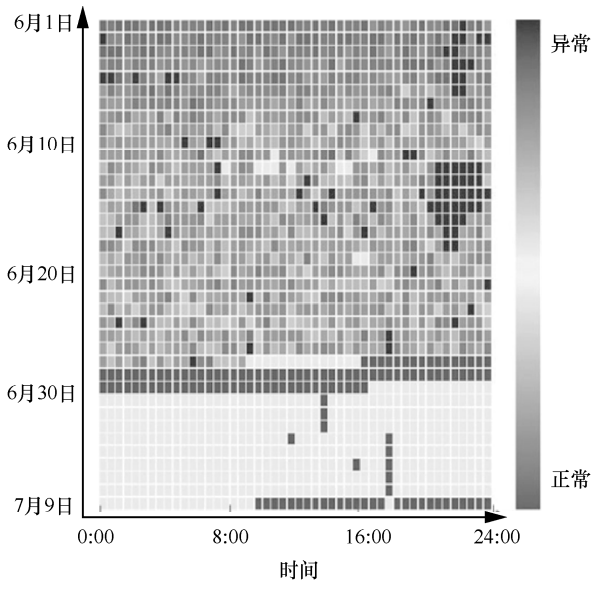


图 4　时间维度性能矩阵

按时间维度输出的矩阵可适用于查看链路历史故障情况，直观反映故障持续时间，但不适用于故障定位。

3.2　按地理维度分析研究

针对某一时间点，可以按地理位置为横纵坐标，形成省到省访问的性能矩阵，以 5 月 20 日晚上 8:00 的电信网内性能矩阵为例，地理维度性能矩阵如图 5 所示。从图 5 中可以发现一条比较明显的“故障十字”，是西藏电信到大部分省份电信互访均出现了故障，推测应该是到西藏方向某个设备或某段链路故障导致。

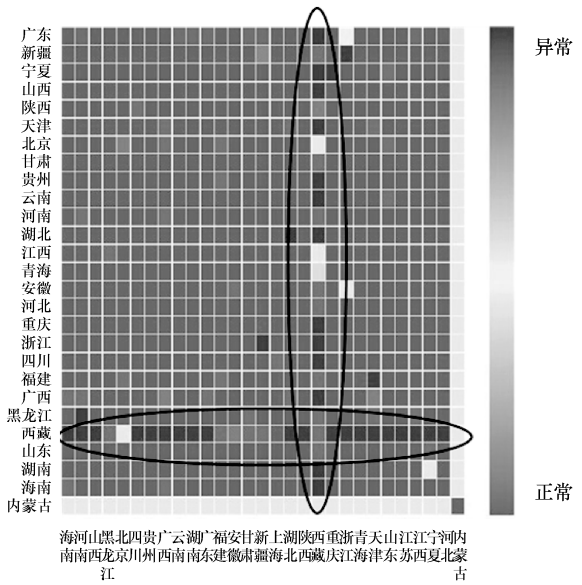


图 5　地理维度性能矩阵

地理位置的性能矩阵可以结合平台常规的路由监测数据进行综合推断，通过当天路由数据发现故障链路中云南、湖北、贵州、江苏、浙江、广东、海南、广西、湖南等省份主要通过昆明中转然后访问西藏、北京、内蒙古、新疆、青海、河北等省份主要通过兰州中转然后访问西藏。按照每条故障链路权重均等，链路通过的所有地理位置权重均等的方法，对故障出现地理位置进行赋权，西藏、云南、甘肃出现故障概率最大，分别为 34%、22%、12%。最终经和运营商核实，确认了昆拉（昆明至拉萨）系统 5 月 20 日和 5 月 21 日两天晚上有过中断，造成拉萨出省电路部分中断，同时备用电路较为拥塞，此结论和本文推测西藏、云南故障概率最高一致。

4 结束语

为了剔除网络性能异常数据及评估网络运行故障，本文提出一种基于密度聚类的网络性能故障大数据分析方法，通过熵权分析提取关键性能特征指标，通过数据清洗与标准化处理实现数据整形，采用基于 DBSCAN 算法对端到端链路监测异常数据进行提取识别，基于全国海量现网实时采集数据实验，该方法的异常数据标注成功率超过 90%。此外针对端到端性能异常数据开展了时间维度和地理纬度的分析，分析故障持续时间、故障初步定位等问题。但目前故障定位仅能推测故障点概率，还需结合人工分析以及运营商核实反馈，未来还需进一步研究，提升自动化故障定位能力及效率。

参考文献：

[1] 李玥. 机器学习的分类、聚类研究[J]. 电脑知识与技术, 2020, 16(4) : 161-162
LI Y. Research on classification and clustering of machine learning[J]. Computer Knowledge and Technology, 2020, 16(4) : 161-162

[2] 蔡志荣. 基于熵权的模糊综合评价法在学习质量评价中的应用[J]. 计算机时代, 2018. (12): 75-77.
CAI Z R. Application of entropy weight based fuzzy comprehensive evaluation method in learning quality evaluation[J]. Computer Era, 2018. (12): 75-77.

[3] 刘竞妍, 张可, 王桂华. 综合评价中数据标准化方法比较研究[J]. 数字技术与应用, 2018, 36(6): 84-85
LIU J Y, ZHANG K, WANG G H. Comparative study on data standardization methods in comprehensive evaluation[J]. Digital Technology & Application, 2018, 36(6): 84-85

[4] 展金梅, 陈君涛. 聚类集成算法中度量方法[J]. 电子技术与软件工程, 2020(3): 170-171
CHEN J M, CHEN J T. Measurement method in clustering integration algorithm[J]. Electronic Technology & Software Engineering, 2020(3): 170-171

[5] 贺玲, 吴玲达, 蔡益朝. 数据挖掘中的聚类算法综述[J]. 计算机应用研究, 2017, (1): 10-13
HE L, WU L D, CAI Y C. Survey of clustering algorithms in data mining[J]. Application Research of Computers, 2017(1): 10-13

[6] 周爱武, 于亚飞. *K*-means 聚类算法的研究[J]. 计算机技术与发展, 2011, 21(2): 62-65
ZHOU A W, YU Y F. The research about clustering algorithm of *K*-means[J]. Computer Technology and Development, 2011, 21(2): 62-65

[7] 闫玮. 基于多种层次聚类的算法研究[D]. 西安: 西安电子科技大学, 2019.
YAN W. Algorithms research based on multiple hierarchical clustering [D]. Xi'an: Xidian University, 2019.

[8] ESTER M, KRIEGEL H P, XU X. A density-based algorithm for discovering clusters adensity-based algorithm for discovering clusters in large spatial databases with noise[C]//Proceedings of International Conference on Knowledge Discovery & Data Mining. New York: ACM Press, 1996.

[9] 谢娟英, 丁丽娟. 完全自适应的谱聚类算法[J]. 电子学报, 2019, 47(5): 1000-1008
XIE J Y, DING L J. The true self-adaptive spectral clustering algorithms [J]. Acta Electronica Sinica, 2019, 47(5): 1000-1008.

[10] 孙鹏, 韩承德, 曾涛. S-DBSCAN:一种基于 DBSCAN 发现高密度簇的算法[J]. 高技术通讯, 2012, 22(6): 589-595.
SUN P, HAN C D, ZENG T. S-DBSCAN: an algorithm for finding high density clusters based on DBSCAN[J]. Chinese High Technology Letters, 2012, 22(6): 589-595.

[11] SHAH G H. An improved DBSCAN, a density based clustering algorithm with parameter selection for high dimensional data sets[C]//Proceedings of Nirma University International Conference on Engineering. Piscataway: IEEE Press, 2013.

[12] 冯振华. 基于 DBSCAN 聚类算法的研究与应用[D]. 无锡: 江南大学, 2016.
FENG Z H. Research and application of clustering algorithm based on DBSCAN[D]. Wuxi: Jiangnan University, 2016.

[13] SUNITA J, PATAG K. Algorithm to determine ε-distance parameter in density based clustering[J]. Expert Systems With Applications, 2014, (6): 2939-2946.

[14] 冯万兴, 朱晔, 郭钧天, 等. 基于改进的 DBSCAN 方法和多项式拟合的雷电短时预测[J]. 计算机工程与科学, 2014, 36(10): 2028-2033.
FENG W X, ZHU Y, GUO J T, et al. Lightning forecast based on the improved DBSCAN and polynomial fitting [J]. Computer Engineering and Science, 2014, 36(10): 2028-2033.

[作者简介]

李想（1986− ），男，中国信息通信研究院高级工程师，主要研究方向为互联网监测分析、域名系统等。

李原（1976− ），男，博士，中国信息通信研究院高级工程师，主要研究方向为互联网网络架构、互联网测量分析、下一代互联网、国际通信等。

张子飞（1987− ），男，中国信息通信研究院工程师，主要研究方向为互联网域名系统、网络性能与业务体验分析等。

杨哲（1990− ），男，中国信息通信研究院工程师，主要研究方向为宽带网络分析、互联网网络等。