IPv6 远程监控网络下无状态通信数据的多尺度离群点挖掘算法

刘琨，张晓涵，曹汝坤，李帅
（中国绿发投资集团有限公司，北京 100020）

摘　要：为了准确挖掘离群点，降低离群点对通信数据造成的影响，对 IPv6 远程监控网络无状态通信数据多尺度离群点挖掘算法进行了研究。通过 IPv6 远程监控网络获得无状态通信数据，依据提取的无状态通信数据的季节性、趋势性和自相似性特征，运用傅里叶变换将无状态通信数据分为两类。再用 *K* 均值法对两类分别进行聚类，确定无状态通信数据的邻域，将其作为基础。采用卷积神经网络对无状态通信数据进行离群点挖掘，初始化卷积神经网络；根据卷积神经网络输出值，判别该网络是否符合停止条件，反复重复卷积神经网络的运算步骤，挖掘全部离群点，实现无状态通信数据多尺度离群点挖掘。实验结果表明，无状态通信数据类别的个数越少，挖掘效率越高；所提方法能准确挖掘 IPv6 远程监控网络无状态通信数据多尺度离群点的个数，准确分析离群原因。

关键词：IPv6；远程监控网络；无状态；通信数据；多尺度；离群点挖掘
中图分类号：TP181
文献标志码：A
doi: 10.11959/j.issn.1000−0801.2023149

Multi-scale outlier mining algorithm for stateless communication data under IPv6 remote monitoring network

LIU Kun, ZHANG Xiaohan, CAO Rukun, LI Shuai
China Green Development Investment Group Co., Ltd., Beijing 100020, China

Abstract: In order to accurately mine outliers and reduce the impact of outliers on communication data, a multi-scale outlier mining algorithm for stateless communication data in IPv6 remote monitoring network was investigated. The stateless communication data were obtained through an IPv6 remote monitoring network, and based on the seasonality, trend, and self-similarity characteristics of the extracted stateless communication data, the Fourier transform was used to divide the stateless communication data into two classes. Then, the *K*-mean method was used to cluster the two classes to determine the neighborhood of the stateless communication data, which was used as the basis for outlier mining using a convolutional neural network on the stateless communication data. The convolutional neural network was initialized, and according to the output value of the convolutional neural network, it was determined whether the network met the stopping condition. The operation steps of the convolutional neural network were repeated, all the outlier points were mined, and the multi-scale outlier mining of stateless communication data was

收稿日期：2023−02−09；修回日期：2023−07−19

achieved. The experimental results showed that the fewer the number of stateless communication data categories, the higher the mining efficiency; the proposed method can accurately mine the number of multiscale outliers of stateless communication data in IPv6 remote monitoring network and accurately analyze the reasons for the outliers.

Key words: IPv6, remote monitoring network, stateless, communication data, multi-scale, outlier mining

0　引言

IPv6 是最新版的互联网协议，可以自动配置 IP 地址，能够增强实时业务，移动性较好，安全性较高[1]。远程监控的意思是本地计算机通过互联网、无线传感器网络等网络设备监测远端区域并控制远端设备，对其进行设备开关、维护等。远程监控可以监控无人区或者环境较为恶劣的区域，实时传送数据，节约了大量的人力物力。将 IPv6 和远程监控的无线传感器组成 IPv6 远程监控网络，结合两者的优点对远程监控地区实时进行监测和控制。无状态是指无论接收方第几次发送同样的请求，请求方都视为第一次接收请求，将所有通信数据发送给接收方。而为了提升网络数据的安全性，保障通信网络数据的合理占比，应开展针对无状态通信的研究。无状态通信是指网络通信过程中，在传输数据的同时不需要保留连接状态信息的一种通信方式。相比保持连接状态的有状态通信，无状态通信能够大大减少网络中的数据传输开销和网络资源占用。在大规模数据传输的场景下，保持连接状态的有状态通信需要在服务器端占用大量内存空间，而无状态通信则可以通过分摊状态信息减轻服务器的负担，提高网络的可扩展性，降低在网络中传输数据的开销和网络中被攻击的风险。多尺度即在不同角度得到通信数据的特征。离群点就是系统在传递数据时受外部因素影响出现各种偏差，出现极大值或者极小值，影响分析数据的结果。离群点挖掘便是通过一系列的方法挖掘离群数据并进行分析[2-3]。IPv6 远程监控网络无状态通信数据多尺度离群点挖掘可有效发现网络通信状态数据的异常情况，保证 IPv6 远程监控网络安全稳定运行。

康耀龙等[4]提出谱聚类算法挖掘通信数据离群点，提取数据的高维特征，研究离群点的来源。根据信息熵确定数据流的关联性，运用亲和矩阵得到离群点和数据样本之间的距离，挖掘该数据的离群点。该算法的执行时间较短，可以根据数据流的维度与大小有效发现其中的异常数据。但在对数据流进行合并降维处理的过程中，干扰因子的处理量较大，离群点位置判别效果不好，导致挖掘离群点的效率不高。王晓辉等[5]运用邻域密度挖掘数据离群点，根据节点运算性能进行区域划分，结合各维度数据的分布情况，计算数据的邻域密度和离群度，以此在异构数据中挖掘数据离群点。利用该算法挖掘数据离群点的准确度，不受数据类型和复杂程度影响，具有良好的检测性能。但在计算数据离群度的过程中，没有从不同角度分析通信数据的特征，这影响数据挖掘的结果。

针对上述数据处理量大、挖掘离群点的效率不高、没有考虑多尺度的数据特征，从而影响数据挖掘结果的问题，本文对 IPv6 远程监控网络无状态通信数据多尺度离群点进行挖掘，使挖掘结果更准确。

1　无状态通信数据多尺度离群点挖掘算法

1.1　IPv6 远程监控网络

IPv6 远程监控网络融合了 IPv6 无线传感器网络和无线通信技术码分多址（code division multiple access，CDMA）。IPv6 远程监控网络将不同种类的传感器节点分布在检测区域内，传感器采集到无状态通信数据后，先对数据进行初步处理再传送到网关节点，最后由 Internet 和 CDMA 将无状态通信数据传送到服务器。管理员可以设置环境参数的范围，若传感器采集的数据超过这个

范围，服务器的终端会发出警示[6]。管理员通过服务器观察远程监测区域的状况，相应的设备可以自动停止或者启动，完成对无状态通信数据的远程监控。但 IPv6 无线传感器网络的使用不局限于和无线通信 CDMA 技术的融合，而是可以根据实际情况和需求与多种无线通信技术结合使用，一般来说，IPv6 无线传感器网络可以使用多种无线通信技术，如 Wi-Fi、蓝牙、ZigBee 等，这些技术都有自己的特点和应用场景，可以根据不同的需求进行选择和应用。IPv6 远程监控网络构架如图 1 所示。

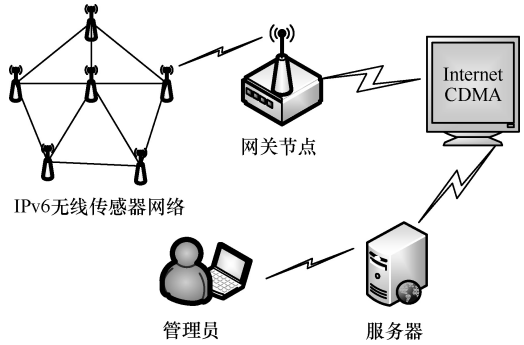


图 1　IPv6 远程监控网络构架

1.2　无状态通信数据聚类

1.2.1　提取无状态通信数据的特征

（1）提取季节性和趋势性特征

假设从 IPv6 远程监控网络中获取的无状态通信数据序列为 $X_t=\{x_1,x_2,\cdots,x_N\}$，选择季节−趋势时序分解法将该通信数据序列分成 3 个分量：

$$X_t=T_t+S_t+E_t \tag{1}$$

其中，E_t 为随机性因素，T_t 为趋势性因素，S_t 为季节性因素。季节性指标为：

$$s_{\deg}=1-\frac{\mathrm{var}(E_t)}{\mathrm{var}(E_t)-\mathrm{var}(T_t)} \tag{2}$$

趋势性指标为：

$$t_{\deg}=1-\frac{\mathrm{var}(E_t)}{\mathrm{var}(X_t)-\mathrm{var}(S_t)} \tag{3}$$

（2）提取自相似性特征

无状态通信数据的自相似性值是该数据序列的长期依赖性。序列零均值表达式为：

$$X_t'=X_t-\mathrm{mean}(X_t) \tag{4}$$

其中，X_t' 为均值化后序列。

设 Y_t 为均值化后序列累积，y_i 为第 i 个元素序列，则 y_i 为：

$$y_i=\sum_{k=1}^{i}x_i' \tag{5}$$

Y_t 的取值范围为 R，运用赫斯特指数表征该数据的自相似性，其定义为：

$$\frac{R}{\sigma}=\left(\frac{N}{2}\right)^{\xi} \tag{6}$$

其中，σ 为无状态通信数据序列的标准差，ξ 为赫斯特指数，N 为序列的长度。用 ξ 表征数据自相似性，如式（7）所示。

$$\xi=2/N\lg(R/\sigma) \tag{7}$$

通过赫斯特指数值判断自相似性特征，当 $\xi=0.5$ 时，表示时间序列可以用随机游走（布朗运动）；当 $0<\xi<0.5$ 时，表示记忆的转弱（反持续性），即均值回复过程；当 $0.5<\xi<1$ 时，表示记忆增强（持续性），即暗示长期记忆的时间序列。

1.2.2　无状态通信数据聚类方法

本文针对无状态通信数据进行聚类，去除相似的数据。首先用傅里叶变换将第 1.2.1 节提取的无状态通信数据特征划分为两类，再用 K 均值法对两类分别进行聚类。

IPv6 远程监控网络的数据是无状态的通信数据，对该数据进行观察，发现无状态通信数据的时序存在一定的周期形态，因此，本文将无状态通信数据的时序经过傅里叶变换划分成周期形态和无明显的周期形态，在这两类形态中结合结构特征聚类方法[7]，细分为更小类，此种分为不同等级处理无状态通信数据的方法使聚类的复杂程度降低。傅里叶变换取得频率幅值谱的表达式如下：

$$\left|F[d]\right|=\left|\sum_{n=0}^{N-1}x(n)\mathrm{e}^{-\mathrm{j}\frac{2\pi}{N}dn}\right| \tag{8}$$

其中，d 是频率。

傅里叶变换的幅值谱均值为 $|F|_{\mathrm{mean}}$、标准差

为$|F|_{\text{std}}$、最大值为$|F|_{\max}$，这 3 个值满足式（9）。

$$|F|_{\text{mean}}+c\times|F|_{\text{std}}<|F|_{\max} \tag{9}$$

其中，c 为一个大于或等于 3 的系数，假设设定值为 1，$|F|_{\max}$ 相对的周期等于 1，这个无状态的通信数据的时序为周期形态。

通过 K 均值聚类对提取的无状态通信数据特征进行分析。首先设置 n 个无状态通信数据样本，将其分成 k 个聚类，K 均值聚类能够准确找到 k 个聚类的均值，即中心[8]。所有的通信数据跟聚类中心之间的距离最小，其计算式为：

$$\arg_{\mu_i}\min=\sum_{i=1}^{k}\|k\mu_i\| \tag{10}$$

其中，μ_i 表示该通信数据的序列的均值。

K 均值聚类的具体流程为：在无状态通信数据集中任意挑选 k 个通信数据样本当作 k 个簇的中心，算出 k 个通信数据样本和 k 个簇中心之间的距离，将每个数据样本都分配到离它最近的中心的簇中。一个簇中心及分配过来的通信数据样本就组成了一个聚类。如果所有的数据样本都被分配完毕，所有簇中心会根据现有的通信数据样本重新运算[9]，这个聚类过程将不断重复，直到收敛，最后输出每个聚类的中心和相对应的类别标签。聚类流程如图 2 所示。

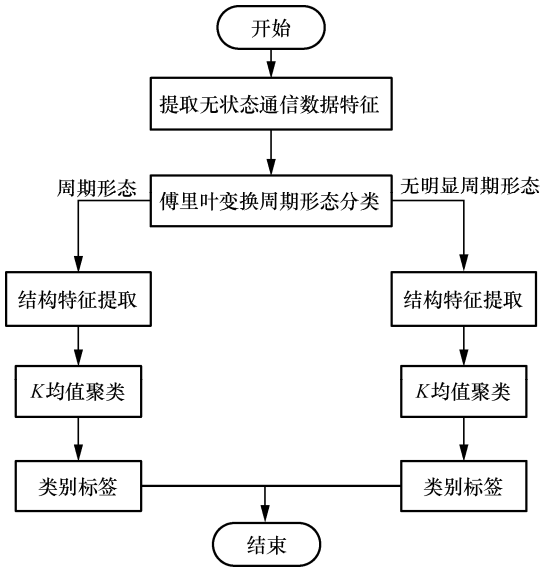


图 2　聚类流程

上述聚类过程将不同的无状态通信数据划分到相应的类别中，从而确定该通信数据的邻域，将邻域内的数据相互比较，作为无状态通信数据多尺度离群点挖掘的基础。

1.3　基于卷积神经网络离群点挖掘

基于第 1.2 节确定无状态通信数据的邻域，挖掘该通信数据的离群点。对该通信数据进行分析时，易将异常的通信数据错认为无效通信数据或者错误的通信数据，降低挖掘该数据的准确度，提高分析该数据的难度。因此本文运用卷积神经网络挖掘 IPv6 远程监控网络无状态通信数据多尺度离群点。

卷积神经网络的表征学习能力较强，它能够提取无状态通信数据所有的特征[10-11]。卷积神经网络模型由很多层组成，卷积神经网络结构如图 3 所示。

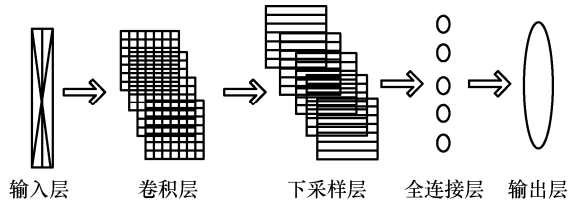


图 3　卷积神经网络结构

预先处理无状态通信数据，将该数据处理成卷积神经网络相关矩阵 $\boldsymbol{X}_{nj}$。预先处理传感器传送的历史数据，处理后的值为 Z_j，其表达式为：

$$Z_j=\text{mod}\,\theta\frac{\left(Z_j-\text{Avg}(Z_j)\right)\times\varphi}{Z_{\max}} \tag{11}$$

其中，φ 是放大函数，组距为 θ，神经元为 j。将 Z_j 值代入 $\boldsymbol{X}_{nj}$ 中，则 $\boldsymbol{X}_{nj}$ 为：

$$\boldsymbol{X}_{n,j}=\begin{bmatrix}Z_1 & Z_2 & \cdots & Z_j\\ 0 & 0 & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & 0\end{bmatrix} \tag{12}$$

将 $\boldsymbol{X}_{n,j}$ 作为原始矩阵输入卷积层，卷积神经网络特征矩阵的 i 层神经元输出值为 Q_i，则 Q_i 产

生过程为：

$$Q_j = f\left(W_{jQ}\sum_{i=1}^{n}Q_i + b_j\right) \tag{13}$$

其中，输入层为Q，神经元输出值是Q_j、偏置值为b_j；激励函数是$f(\cdot)$；W_{jQ}表示连接神经元j和输入层Q的权重值。接下来是下采样层，下采样层对无状态通信数据的特征矩阵进行降维，且该特征矩阵的走向趋势不会发生改变。假定下采样层为O_i：

$$O_i = \text{subsampling}(Q_i - 1) \tag{14}$$

无状态通信数据特征矩阵先经过卷积层运算，再经过下采样层运算，全连接层再把该矩阵按照各自特征分类[12-13]，其分布概率为V。通过式（13）卷积运算原始无状态通信数据矩阵$\boldsymbol{Q}_0$，得到新的数据特征V为：

$$V(i) = Q\left(S = g_i \,|\boldsymbol{Q}_0 : (W,b)\right) \tag{15}$$

其中，第i个类别标签为g_i，最小损失函数为(W,b)，损失函数为：

$$\text{NLL}(W,b) = -\sum_{i=1}^{|V|}\lg V(i) \tag{16}$$

在训练时，如果模型对通信数据描摹得太细致，会导致异常的无状态通信数据或其他新的通信数据的误差值增加[14-15]，泛化能力变弱，出现过拟合现象。将L_2作为范数输入损失函数，以此控制过拟合，该表达式为：

$$E(W,b) = \lambda \times \frac{1}{2}\boldsymbol{W}_i^{\mathrm{T}} + L(W,b) \tag{17}$$

其中，λ为拟合参数，能够控制过拟合强度。前向传导矩阵$\boldsymbol{X}_{n,j}$时，损失函数值和期望值会产生残差。在训练时，运用随机梯度下降的方法反向传播残差，更新每层的训练参数(W,b)，过程为：

$$\boldsymbol{W}_i = \frac{\partial E(W,b)}{\partial W} \times \mu \tag{18}$$

$$b_i = b_i - \frac{\partial E(W,b)}{\partial b_i} \times \mu \tag{19}$$

其中，μ代表学习强度的速率，可以控制方向传播的学习强度。

基于卷积神经网络数据多尺度离群点挖掘流程如下。

步骤 1 对卷积神经网络各层的偏差和权值初始化，无状态通信数据集中类别的数量决定输入层中神经元的数量，获得该通信数据的邻域范围。

步骤 2 通过训练集得到输入向量和输出向量。

步骤 3 确定节点的数量，将其隐藏并输出，获取卷积神经网络输出值。

步骤 4 该输出值体现通信数据集具体的分布情况，通过卷积神经网络运算得到的值，根据熵值判断异常的无状态通信数据。熵值越高，说明通信数据样本中的不确定性越高，越容易出现异常情况。当熵值不在给定的阈值范围内时，无状态通信数据样本为异常点。

设定阈值为E，阈值范围为0～1，评价函数$E(\cdot)$为：

$$E = E(ap_{\mathrm{r}}, bp_w) \tag{20}$$

其中，权值分别为a、b，样本分类正确概率为p_{r}，分类错误概率为p_{w}。通过E判断异常点的有效性。则挖掘通信数据异常点的效率$P(E)$为：

$$P(E) = -ap_{\mathrm{r}} + bp_{\mathrm{w}} \tag{21}$$

其中，$P(E)$值越小，挖掘通信数据异常点的效果越差；$P(E)$值越大，挖掘通信数据异常点的效果越好。

根据卷积神经网络实际输出值和期望值，计算 IPv6 远程监控网络输出无状态通信数据的误差，判别该网络是否符合停止条件。如果不符合，则返回上述过程重新训练；如果符合，则停止训练，不再评价卷积神经网络对无状态通信数据多尺度离群点。最后，评价挖掘出的离群点，明确该通信数据的离群原因。

反复迭代卷积神经网络的运算过程，直到挖掘完全部的离群点，停止计算，以此完成对 IPv6 远程监控网络无状态通信数据多尺度离群点的挖

掘，无状态通信数据多尺度离群点挖掘流程如图 4 所示。

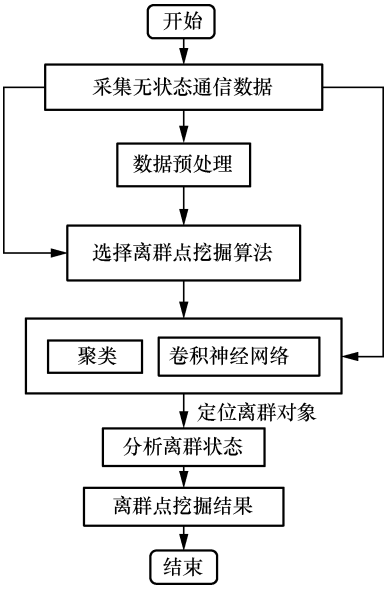


图 4　无状态通信数据多尺度离群点挖掘流程

2　实验结果

本次实验 IPv6 远程监控网络覆盖面积为 79 万平方米，监控点位 598 个，每个传感器探测覆盖面积为 1.4 万平方米，实验所用通信数据均由该网络获取，由于该数据较多，采用随机取样的方法，抽取 15%的通信数据样本，在此基础上进行实验。

本文方法在运算的过程中有个重要的参数 λ。该参数会控制过拟合现象，影响本文挖掘结果，所以对此进行分析。本实验所用 λ 的取值范围为 0～2.4，λ 值变化对无状态通信数据多尺度离群点挖掘结果的影响如图 5 所示。

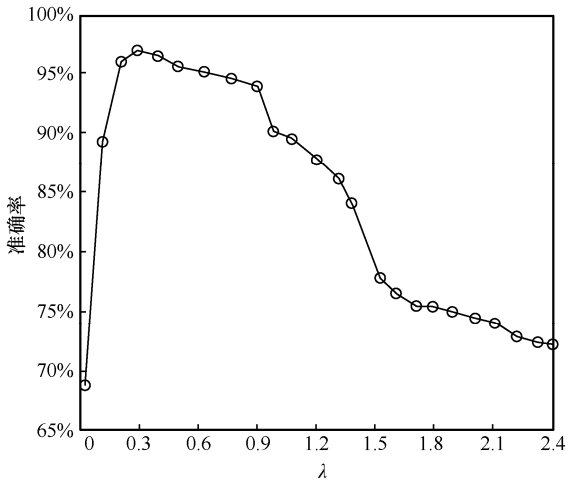


图 5　λ 值变化对无状态通信数据多尺度离群点挖掘结果的影响

由图 5 可知，λ 值超过 0.9 后，无状态通信数据多尺度离群点数据挖掘结果的准确率明显降低到 90%及以下，因此，为了维持挖掘结果的准确率，λ 值的取值范围应为 0.2～0.9。当 λ 值为 0.3 时准确率高达 97.5%，因此本文方法选取 0.3 作为最佳 λ 值。

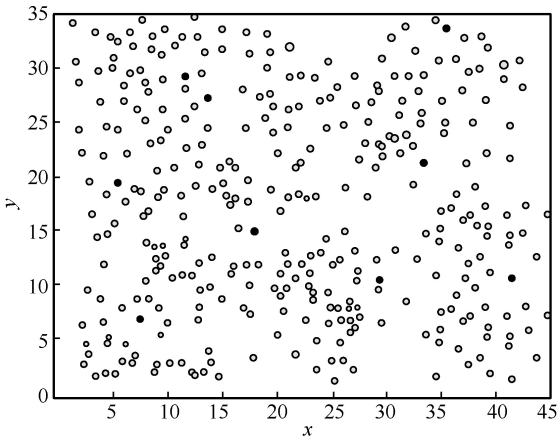
运用本文方法、文献[4]方法划分无状态通信数据的类别，按照类别划分的无状态通信数据集的大小会影响其离群点挖掘结果，无状态通信数据类别对离群点挖掘结果的影响见表 1。

表 1　无状态通信数据类别对离群点挖掘结果的影响

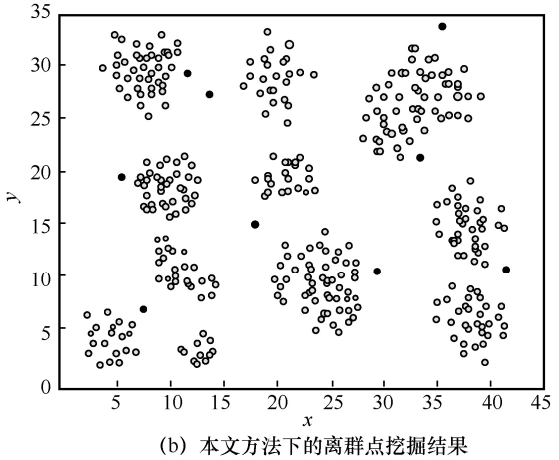
数据集/万	类别个数/个	本文方法挖掘次数比	本文方法时间比	文献[4]方法挖掘次数比	文献[4]方法时间比
10	30	1.298	2.10	1.775	3.02
	20	1.134	1.09	1.593	2.16
	10	1.008	1.01	1.487	2.02
30	30	1.197	3.25	1.768	3.46
	20	1.090	1.43	1.647	2.98
	10	1.015	1.05	1.319	2.61
50	30	1.101	4.78	1.541	5.88
	20	0.981	2.42	1.316	3.29
	10	0.981	1.08	1.149	3.02
70	30	1.098	5.58	1.532	6.46
	20	0.987	3.06	1.311	5.13
	10	0.889	1.14	1.135	3.64

由表 1 可知，随着无状态通信数据集数量的增加，类别个数影响离群点的挖掘次数和时间。本文方法的挖掘次数比整体上小于文献[4]方法，在时间比上整体低于文献[4]方法。通信数据类别越多，挖掘该通信数据离群点所耗用的时间越多，挖掘的次数也越多。实验结果表明，无状态通信数据类别的个数也越少，挖掘离群点所用的时间越短、挖掘的次数越少，挖掘效率越高。

为了验证本文方法在无状态通信数据分布复杂情况下的离群点挖掘性能，选取了一组未进行离群点挖掘的无状态通信数据集，该数据集中有 587 个正常数据和 9 个离群数据，运用本文方法对其离群点进行挖掘，无状态通信数据离群点挖掘结果如图 6 所示。



(a) 无状态通信数据集



(b) 本文方法下的离群点挖掘结果

图 6 无状态通信数据离群点挖掘结果

图 6 中，本文方法下的无状态通信数据集中的正常数据被分为 11 簇，有 9 个离群点被挖掘出来。该通信数据具体无状态通信数据离群原因见表 2。

表 2 无状态通信数据离群原因

离群点	离群原因
1	环境变化导致原本的数据不在传感器收集范围内
2	外部噪声太大导致数据异常
3	暴雨天气使传感器收集通信数据造成误差
4	数据获取来源不同
5	数据存储混乱
6	数据记录错误
7	远程监控系统出现损坏
8	服务器接收错误
9	计算错误

根据表 2 可知，本文方法能准确挖掘 IPv6 远程监控网络无状态通信数据多尺度离群点的个数，对离群原因进行准确分析。

为了验证本文方法的可行性，运用本文方法在远程监控通信数据系统中对异常的通信数据进行离群点挖掘，远程监控系统离群点挖掘如图 7 所示。

图 7 远程监控系统离群点挖掘

图 7 中，指示灯绿灯亮时，该通信数据未出现异常；指示灯红灯亮时，该通信数据出现异常，开始对通信数据的离群点进行挖掘，挖掘准确的离群点个数。实验结果表明，本文方法在远程监控通信数据系统中能够快速挖掘该通信数据的离

群点并发出警示，远程监控人员能及时对离群点进行分析。

3　结束语

本文方法对 IPv6 远程监控网络无状态通信数据多尺度离群点进行挖掘，通过实验对不同参数值进行比较、比较数据集的大小对挖掘无状态通信数据离群点效率的影响，研究本文方法的挖掘性能，在远程监控通信数据系统中，验证本文方法的有效性。通过实验说明，本文方法针对 IPv6 远程监控网络无状态通信数据多尺度离群点挖掘效率高、准确率高。远程监控人员能够及时获取离群点，分析该通信数据离群原因，了解远程的无状态通信数据具体状况。

参考文献：

[1] 丑义凡, 易波, 王兴伟, 等. IPv6 网络中基于 MF-DL 的 DDoS 攻击快速防御机制[J]. 计算机学报, 2021, 44(10): 2047-2060.
CHOU Y F, YI B, WANG X W, et al. A rapid defense mechanism based on MF-DL for DDoS attack in IPv6 networks[J]. Chinese Journal of Computers, 2021, 44(10): 2047-2060.

[2] 罗晓媛, 赵丽艳, 刘君, 等. 神经网络技术下多尺度时序数据离群点挖掘[J]. 计算机仿真, 2021, 38(1): 231-235.
LUO X Y, ZHAO L Y, LIU J, et al. Mining outliers in multi-scale time series data based on neural network technology[J]. Computer Simulation, 2021, 38(1): 231-235.

[3] 张倩倩, 于炯, 李梓杨, 等. 基于近邻传播的离群点检测算法[J]. 计算机应用研究, 2021, 38(6): 1662-1667.
ZHANG Q Q, YU J, LI Z Y, et al. Outlier detection algorithm based on affinity propagation[J]. Application Research of Computers, 2021, 38(6): 1662-1667.

[4] 康耀龙, 冯丽露, 张景安, 等. 基于谱聚类的高维类别属性数据流离群点挖掘算法[J]. 吉林大学学报(工学版), 2022, 52(6): 1422-1427.
KANG Y L, FENG L L, ZHANG J A, et al. Outlier mining algorithm for high dimensional categorical data streams based on spectral clustering[J]. Journal of Jilin University (Engineering and Technology Edition), 2022, 52(6): 1422-1427.

[5] 王晓辉, 宋学坤, 王晓川. 基于邻域密度的异构数据局部离群点挖掘算法[J]. 计算机仿真, 2021, 38(7): 281-285.
WANG X H, SONG X K, WANG X C. Local outlier mining algorithm for heterogeneous data based on neighborhood density[J]. Computer Simulation, 2021, 38(7): 281-285.

[6] 梁勇, 刘承启. IPv6 远程监控网络无状态双向通信方法[J]. 计算机仿真, 2021, 38(2): 119-123.
LIANG Y, LIU C Q. IPv6 remote monitoring network stateless two-way communication method[J]. Computer Simulation, 2021, 38(2): 119-123.

[7] 蔡顺, 耿豪鹏, 郑炜珊, 等. 基于傅里叶变换的谷间距特征信息提取及其影响因素研究[J]. 地球信息科学学报, 2020, 22(3): 399-409.
CAI S, GENG H P, ZHENG W S, et al. Valley spacing character information and its influencing factors based on the Fourier transform[J]. Journal of Geo-Information Science, 2020, 22(3): 399-409.

[8] 孟笑天, 徐艳蕾, 王新东, 等. 基于改进 *K* 均值特征点聚类算法的作物行检测[J]. 农机化研究, 2020, 42(8): 26-30.
MENG X T, XU Y L, WANG X D, et al. Crop line detection based on improved *K*-means feature point clustering algorithm[J]. Journal of Agricultural Mechanization Research, 2020, 42(8): 26-30.

[9] 马永军, 汪睿, 李亚军, 等. 利用聚类分析和离群点检测的数据填补方法[J]. 计算机工程与设计, 2019, 40(3): 744-747, 761.
MA Y J, WANG R, LI Y J, et al. Data filling using cluster analysis and outlier detection[J]. Computer Engineering and Design, 2019, 40(3): 744-747, 761.

[10] 崔斌, 张永红, 闫利, 等. 一种基于卷积神经网络的 SAR 变化检测方法[J]. 测绘科学, 2019, 44(6): 170-175, 186.
CUI B, ZHANG Y H, YAN L, et al. A method of SAR change detection based on convolutional neural networks[J]. Science of Surveying and Mapping, 2019, 44(6): 170-175, 186.

[11] 马京晖, 潘巍, 王茹. 基于 *K*-means 聚类的三维点云分类[J]. 计算机工程与应用, 2020, 56(17): 181-186.
MA J H, PAN W, WANG R. 3D point cloud classification based on *K*-means clustering[J]. Computer Engineering and Applications, 2020, 56(17): 181-186.

[12] 李晓莉, 韩鹏, 李晓光. 基于典型样本的卷积神经网络技术[J]. 计算机工程与设计, 2020, 41(4): 1113-1117.
LI X L, HAN P, LI X G. Convolutional neural network based on typical samples[J]. Computer Engineering and Design, 2020, 41(4): 1113-1117.

[13] 许璟琳, 彭阳, 余芳强. 基于 *K*-means 聚类和离群点检测算法的医院建筑节能诊断方法[J]. 计算机应用, 2021,41(S1): 288-292.
XU J L, PENG Y, YU F Q. Energy-saving diagnosis method for hospital buildings based on *K*-means clustering and outlier detection algorithm[J]. Journal of Computer Applications, 2021, 41(S1): 288-292.

[14] 神显豪, 李驰, 桂琼, 等. 基于卷积神经网络的网络节点异常数据检测方法[J]. 机床与液压, 2020, 48(22): 18-23.

SHEN X H, LI C, GUI Q, et al. A method for detecting abnormal data of network nodes based on convolutional neural network[J]. Machine Tool & Hydraulics, 2020, 48(22): 18-23.

[15] 赵晓永, 王宁宁, 王磊. 基于主动学习的离群点集成挖掘方法研究[J]. 计算机工程与应用, 2020, 56(12): 112-117.
ZHAO X Y, WANG N N, WANG L. Research of outlier ensemble mining based on active learning[J]. Computer Engineering and Applications, 2020, 56(12): 112-117.

[作者简介]

刘琨（1970− ），男，中国绿发投资集团有限公司高级工程师，主要研究方向为电力工程。

张晓涵（1988− ），女，中国绿发投资集团有限公司工程师，主要研究方向为数字化。

曹汝坤（1988− ），男，博士，中国绿发投资集团有限公司工程师，主要研究方向为环境工程及产业分析。

李帅（1986− ），男，中国绿发投资集团有限公司工程师，主要研究方向为数字化、信息化。