



基于大数据的电力信息网络流量异常检测机制

姜红红¹, 张涛², 赵新建¹, 钱欣¹, 赵天成¹, 高莉莎¹

(1. 国网江苏省电力公司南京供电公司, 江苏 南京 210019;

2. 北京邮电大学网络与交换技术国家重点实验室, 北京 100876)

摘要:随着智能电网建设的加强, 电力信息网络及其承载的业务系统得到迅猛发展, 网络业务流量的检测和预警具有重要的安全意义。针对目前电力信息网络缺乏处理流量异常问题的有效技术手段, 提出了一种基于大数据的电力信息网络流量异常检测机制, 并通过对改进的局部异常因子(M-LOF)和支持向量域数据描述(SVDD)两种常用异常检测算法的对比分析, 总结出适合电力信息网络的流量异常检测方法。

关键词: 电力信息网络; 流量异常检测; 局部异常因子; 支持向量域数据描述

中图分类号: TM744

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017031

A big data based flow anomaly detection mechanism of electric power information network

JIANG Honghong¹, ZHANG Tao², ZHAO Xinjian¹, QIAN Xin¹, ZHAO Tiancheng¹, GAO Lisha¹

1. Jiangsu Nanjing Power Supply Company, Nanjing 210019, China

2. National Key Lab of Networking and Switching Technology,

Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract: With the construction of smart grid, the electric power information network and its business system get rapid development. The early flow anomaly detection and warning are significant to the safety of network. Due to the lack of efficient measuring means to handle the flow abnormal problems, a flow anomaly detection mechanism based on big data for the electric power information network was proposed. Through the comparative analysis of two common anomaly detection algorithms, the improved local outlier factor algorithm (M-LOF) and the support vector data description (SVDD) algorithm, the suitable flow anomaly detection method for electric power information network was summarized.

Key words: electric power information network, flow anomaly detection, local outlier factor, support vector data description

1 引言

随着电力信息网络规模不断扩大、复杂性不断增加, 基于信息网络的各类应用业务也越来越广泛, 产生的数据量

庞大, 这就导致网络出现异常的可能性大大增加^[1-3]。而且在信息网络中, 很难找到发生问题的根源, 未被诊断的网络异常会进一步传播并影响网络的正常信息传输。因此, 如何对网络流量进行实时的监测和评估, 并及时发现网络的异常,

对提高网络的稳定性和安全性具有重要意义^[4]。

目前,网络异常诊断比较常规的做法是根据网络运行质量监测和评估的指标来观察网络的规律,并进一步发现和排除异常。但是目前国内针对电力信息网络的流量分析手段处理能力弱、颗粒度大,这些问题造成了电力信息网络的数据盲视和流量数据信息黑洞,同时在流量饱和或网络拥塞等情况发生时,无法进一步对异常源进行深层次排查分析^[5]。另外,网络管理中心需要处理网络中各种各样的故障信息,这些大多需要历史经验来指导,然而庞大的数据量不可能都由人工完成。因此,目前电力信息网急需一种流量问题的预警机制,能够在第一时间发现异常,并让管理中心快速地通过流量管理手段解决问题^[6-8]。在智能电网运行过程中,为了及时、全面、准确地了解网络的实时状况,智能电网系统中安装了很多流量探针、传感器等采集设备。系统会实时监听和采集到海量的数据信息并上传,电力大数据信息集合由此产生。截至 2015 年,国家电网已完成招标智能电表 4.3 亿台,实现用电信息采集用户 4.5 亿户,其采集系统将产生以 PB 级计的数据。电力系统运行过程中产生的数据具备大数据典型的“4V”特征,即规模性(volume)、多样性(variety)、高速性(velocity)和价值性(value)。智能电网中的这些大数据蕴藏着很多有价值的信息,但这些数据价值密度不高,并不能直接运用于网络的快速判稳,需要从实际的系统特性出发,研究数据内部的规律,使用在线的数据挖掘手段来检测网络的异常。

现有基于大数据的网络流量异常检测方法通常使用基于知识的检测方法,如基于事例的推理、模糊逻辑、粗糙集方法、人工神经网络、支持向量机以及信息融合故障检测方法等^[9]来构建。基于机器学习的方法^[10,11]是现在异常检测技术的主流,常用的异常检测方法中主要是处理分类问题,一类(目标)样本充分采样,另一类(异常)进行欠采样,一般从已知的正常类数据中进行学习,建立正常的学习模型用以进行数据分类。而在电力数据网中,首先在网络中采集一些样本并进行异常标注,然后利用这些样本来训练模型并将模型运用于实际网络的异常检测中,这是一种有监督的学习方法。但是,实际网络的情况十分复杂,数据量也远远超过人工所能标注的范围,因此很难精确指出哪些流量为异常流量。

相对于上述有监督的学习方法,无监督的学习方法可以省去数据标注的过程,从海量数据集中学习正常的模式,成为目前网络流量异常检测的重要方法。提出一种基于大数据的电力信息网络流量异常检测机制,从实际应用

出发,重点进行海量在线数据的特征量选择、样本预处理并引入两种不同的数据挖掘方法来处理电力网流量异常问题:基于密度的局部异常因子(local outlier factor, LOF)学习方法和基于距离的支持向量域数据描述(support vector domain description, SVDD)学习方法,这两种方法都是无监督学习方法。

2 流量异常检测机制

基于数据挖掘方法的网络流量异常检测机制如图 1 所示。首先从网络中进行样本数据的采集,并保存到数据库中。然后对数据进行预处理,包括数据筛选、数据降维和数据标准化,之后输入数据挖掘算法模型中,通过使用合适的机器学习算法给出异常检测结果。

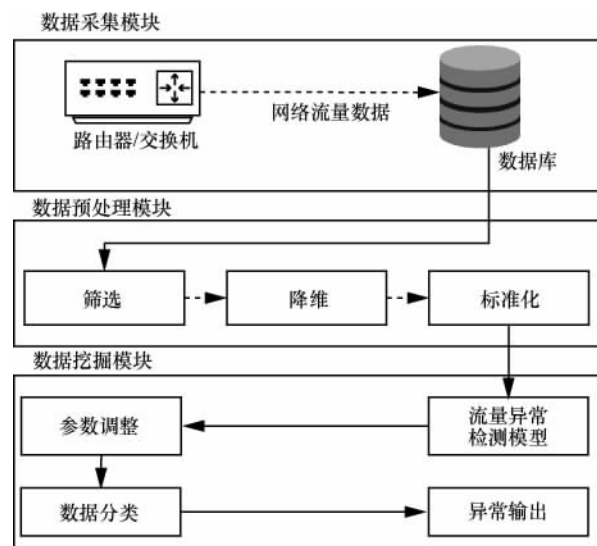


图 1 基于数据挖掘方法的网络流量异常检测机制

2.1 样本数据采集

在电力信息网中,数据主要来自智能电表、相量测量单元(power management unit, PMU)以及各种传感器设备的采集信息,这些数据的数据分组很小,但数据规模大、结构复杂、传输速度快。为了准确实时地获取这些信息,通过部署在网络节点(交换机或者路由器)上的流量采集设备采集样本数据,采集到的数据分组包括分组头各个字段的信息,具体字段信息见表 1。

2.2 数据预处理

数据预处理是进行数据挖掘前很重要的一步,不仅需要数据转换成模型要求的向量格式,还需要清除数据源中的脏数据、重复数据等^[12]。通过特征选取、数据筛选、降维和标准化完成数据的预处理工作。



表 1 原始流量数据分组字段

字段名	说明
RecordTime	记录时间
LinkID	链路 ID
SourceIP	源 IP 地址
DestinationIP	目的 IP 地址
SourcePort	源端口号
DestinationPort	目的端口号
ProtocolID	传输层的协议类型
PacketsIn	输入报文数
PacketsOut	输出报文数
BytesIn	输入字节数
BytesOut	输出字节数
TocIn	流中输入数据的一个字段
TocOut	流中输出数据的一个字段
AppType	流的协议号
FlowDirection	服务器相关
StartTime	流创建时间
EndTime	流结束时间
FlowID	流标识
DeveiceID	设备标识
ProtocolGroupID	协议分组 ID
UserGroupID	用户组 ID
UserID	用户 ID
RegionID	地域 ID
CarrierID	运营商 ID
Custom	备用预留字段

2.2.1 特征选取

电力信息网络流量数据的维数较大,为了直观显示数据的分布,使用平行坐标的方法来观察数据,可以剔除无关维度。平行坐标是一种通用的可视化方法,用于对高维几何和多元数据的可视化,图 2 中每一个纵轴都是数据的一个维度。

从图 2 中可以直观地看到数据的整体分布情况。由于电力信息网中的流量分布特点与业务种类和时间有关,所以对每时刻每端口发送的数据分组进行统计,特征选取维度见表 2。

表 2 特征选取维度

维度字段名	说明
业务种类	业务种类由端口号来区分
源 IP 地址的分布空间	单位时间内访问的不同源 IP 地址的数目
目的 IP 地址的分布空间	单位时间内进行数据传输的不同目的 IP 地址的数目
目的地址的最大访问次数	单位时间内单个目的地址被重复发送的最大次数
数据分组发送个数	单位时间内该业务端口数据分组发送的总个数
数据分组接收个数	单位时间内该业务端口数据分组接收的总个数
数据分组发送字节数	单位时间内该业务端口数据分组发送的总字节数
数据分组接收字节数	单位时间内该业务端口数据分组接收的总字节数

实验过程中,特征提取的单位时间粒度取值为 10 min,这样能包含足够的信息判断流量异常。如果时间粒度过小,则会造成检测算法的误报率上升,产生虚警。

2.2.2 数据筛选与降维

多维尺度分析(multi-dimensional scaling,MDS)是一种常用的数据分析方法,MDS 可以利用成对样本间的相似性来构建合适的低维度空间,并且能保证样本在低维度空间的距离相似性和高维度的一致性^[3]。具体描述如下。

首先,对于要降维的 t 维数据,构建一个 $t \times t$ 的相异度矩阵 Δ^x ,MDS 期望寻找 m 维的 t 个向量 $\psi_1, \dots, \psi_t \in R^N$ 组成的矩阵 Δ^ψ ,使得向量间的距离与 Δ^x 中向量距离相似。在经

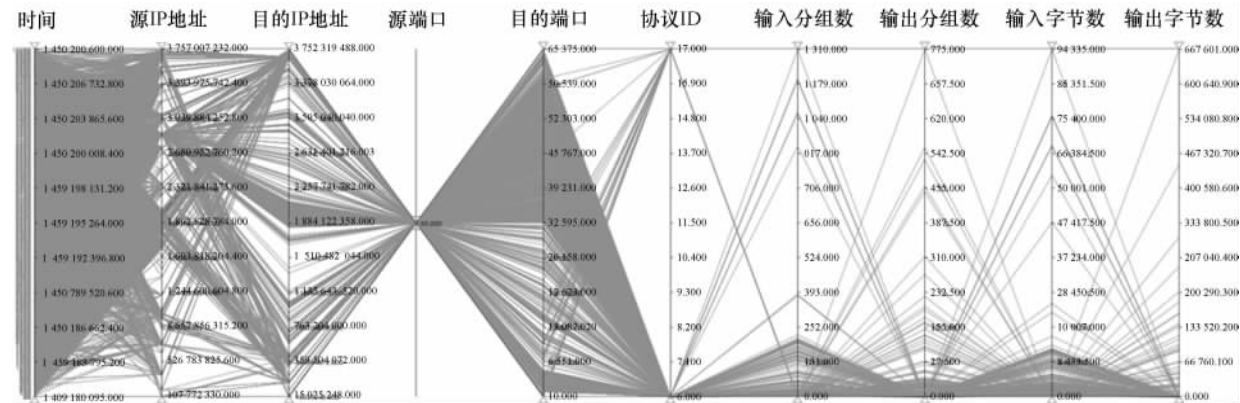


图 2 网络流量数据平行坐标

典的多维尺度分析中,该距离指的是欧氏距离。欧式距离可以被任意旋转和变换,且这些变换不会改变样本间的距离。

所以多维尺度分析相当于一个优化问题:

$$\min_{\psi} \sum_{i=1}^l \sum_{j=1}^l (\delta_{ij}^X - \delta_{ij}^{\psi})^2 \quad (1)$$

其中, $\delta_{ij}^X = \|x_i - x_j\|^2$, $\delta_{ij}^{\psi} = \|\psi_i - \psi_j\|^2$, 式(1)可以转换为矩阵表达形式:

$$\min_{\psi} \sum_{i=1}^l \sum_{j=1}^l (x_i^T x_j - \psi_i^T \psi_j)^2 \quad (2)$$

实际数据采集,由于操作人员的记录错误或者采集设备的误差等原因,会出现个别样本点与其他点偏差很大的情况,这类点被称为离群点或者野值。从图2中很容易找到这些点,并予以剔除。

2.2.3 数据标准化

由于向量空间上各个维度的单位不一致,如输入分组数、输出分组数与输入字节数、输出字节数的单位不一致,需要对数据进行标准化处理,本文使用离差标准化的方法:

$$S_j' = \frac{S_j - \min(S_j)}{\max(S_j) - \min(S_j)} \quad (3)$$

通过式(3),让各个维度的数据都转换为[0,1]区间上的值。

2.3 基于大数据的异常检测算法

基于机器学习的分类方法被广泛应用于异常检测领域,引入两种不同类型的机器学习方法。

2.3.1 改进的局部异常因子检测算法

在无监督异常检测方法中,由于仅有一类样本可供学习,因此最简单也最直接的方式就是通过参数化方法或非参数化方法来估计训练样本的密度模型并设置密度阈值,小于该阈值的即被认为异常。

参数化方法的原理是假定样本数据符合某项分布,如多元高斯模型,以测试样本与均值的距离来判定其是否为异常。该方法比较简单,但是存在维度灾难问题,即时间成本与样本数成指数关系,当样本数过高时,时间成本将十分巨大。而由于电力信息网的高可靠性要求,对业务流量异常检测的时效性要求较高,因此参数化方法不适于电力信息网的流量异常检测。

选用的局部异常检测算法是一种非参数化的无监督异常检测算法,其设计原理是根据样本点与其局部邻域样本点分隔程度的局部异常度来进行异常检测。

由于网络节点的故障异常是具有局部性的,提出了一

种改进的 LOF 算法——M-LOF(mean local outlier factor),该方法在 LOF 算法的 k 距离和 k 近邻的基础上,提出了 m 距离和 m 近邻的概念,并基于 m 距离和 m 近邻对异常点进行检测,具体算法描述如下。

定义 1 对象 p 的 k 距离 $d_k(p)$, 表示对象 p 与距离它第 k 近的邻居的距离,表示如下。

至少有 k 个对象,满足:

$$d(o_i, p) \leq d(o_k, p) \quad (4)$$

最多有 $k-1$ 个对象,满足:

$$d(o_i, p) < d(o_k, p) \quad (5)$$

定义 2 对象 p 的 k 近邻 $N_{d_k(p)}(p)$, 表示由所有与 p 之间距离小于 $d_k(p)$ 的对象组成的集合。

定义 3 对象 p 的 m 距离,为了降低传统的 LOF 算法的 k 取值敏感性,M-LOF 算法提出 m 距离的概念,即将对象 p 到 k 近邻的距离求平均,得到 p 的 m 距离:

$$m_d_k(p) = \varepsilon + \left[\frac{\sum_{o \in N_{d_k(p)}(p)} d(p, o)}{|N_{d_k(p)}(p)|} \right] \quad (6)$$

其中, ε 是用来提高准确度的常量。

定义 4 对象 p 的 m 近邻,针对 m 距离,可以得到对象 p 的 m 近邻: $N_{m_d_k(p)}(p)$, 表示所有与 p 之间距离小于 m 距离的对象组成的集合。

定义 5 对象 p 相对于对象 o 的可达距离 $reach_dist_m(p, o)$, 表示 p 的 m 距离和 p 与 o 之间距离的最大值:

$$reach_dist_m(p, o) = \max\{m_d_k(o), d(o, p)\} \quad (7)$$

定义 6 对象 p 的局部密度,计算式如下:

$$lrd_m(p) = \frac{|N_{m_d_k(p)}(p)|}{\sum_{o \in N_{m_d_k(p)}(p)} reach_dist_m(p, o)} \quad (8)$$

定义 7 对象 p 的局部异常因子,计算式如下:

$$LOF_m(p) = \frac{\sum_{o \in N_{m_d_k(p)}(p)} \frac{lrd_m(o)}{lrd_m(p)}}{|N_{m_d_k(p)}(p)|} \quad (9)$$

M-LOF 算法流程如下。

输入数据流 $p = \{p_j\} j=1, 2, \dots, N$

For $j=1, 2 \dots N$: do

Find $d_k(p_j)$ and $m_d_k(p_j)$

Find $N_{d_k(p)}(p)$ and $N_{m_d_k(p)}(p)$

Calculate $reach_dist_m(p_j, p_j)$

Calculate $lrd_m(p_j)$



Calculate $LOF_m(p_j)$

End For

2.3.2 支持向量域描述算法

支持向量机(support vector machine, SVM)是一种主流的两类分类方法,其通过寻求一个超平面将两类样本以最大间隔分开。但在流量异常检测领域,由于一般仅有一类样本,故原有 SVM 不再适用。

支持向量域描述(SVDD)是一种单类(one class)SVM,可用来解决异常检测领域中仅有一类样本的问题。SVDD 可简单描述为:设输入空间中的一组数据为: $\{x_i\}, x_i \in R^d, i=1, \dots, N$, 该数据被映射到一个高维度的向量空间上,然后在这个空间里期望找到一个包含该组数据的半径最小的超球体来对其描述。为了更好地进行描述,常常使用一个非线性映射将数据映射到高维空间中,并引入一个核函数 $k(x, y) = \phi(x) \cdot \phi(y)$ 来代替内积。为了减少奇异点的影响,又引入了松弛变量 ξ_i 将奇异点排除在超球体的外面,即求解下面的二次规划问题:

$$\begin{aligned} \min(R^2 + C \sum_i \xi_i), \\ \text{s.t. } (\phi(x_i) - a)^T (\phi(x_i) - a) \leq R^2 + \xi_i \\ \xi_i \geq 0, \forall i \end{aligned} \quad (10)$$

SVDD 被用来处理线性分类的问题,核函数可以将高维空间的内积运算转化为低维输入空间的核函数计算,从而巧妙地解决了在高维特征空间中计算的“维数灾难”等问题。常用的核函数有:线性核函数、多项式核函数、径向基核函数、sigmoid 核函数和复合核函数等^[14]。本文选用线性核函数(linear)、径向基核函数(radial basis function, RBF)和 sigmoid 核函数(sigmoid)3种进行比较,分析不同核函数处理流量异常问题的优劣。

3 实验结果与对比分析

3.1 实验结果

通过流量探测工具获取某电力公司信息网 2016 年 3 月内的连续流量数据分组,并在实验平台中进行处理和分析。实验平台基于 MATLAB 开发,在实验平台中实现了 M-LOF 算法,并使用 LibSVM 工具包进行二次开发完成 SVDD。经过预处理后的数据作为实验平台的输入,通过不同算法的处理来对算法表现进行评估。

3.1.1 M-LOF 算法结果

传统的 LOF 算法对算法中邻居 k 的取值比较敏感,当数据量很大时,如果 k 取值过小,会将很多比较集中的异常

点判断为正常点,如果 k 取值过大,又会使正常点受影响被判断为异常点。而提出的 M-LOF 算法虽然也会受 k 值影响,但对 k 值的敏感性比 LOF 降低很多,因为如第 2.3.1 节的描述, M-LOF 算法计算的 m 距离是对 k 近邻的 k 距离的平均,所以 M-LOF 算法更加关注于局部性。 $k=10$ 和 $k=100$ 时 10 000 连续流量分组检测结果如图 3 和图 4 所示。

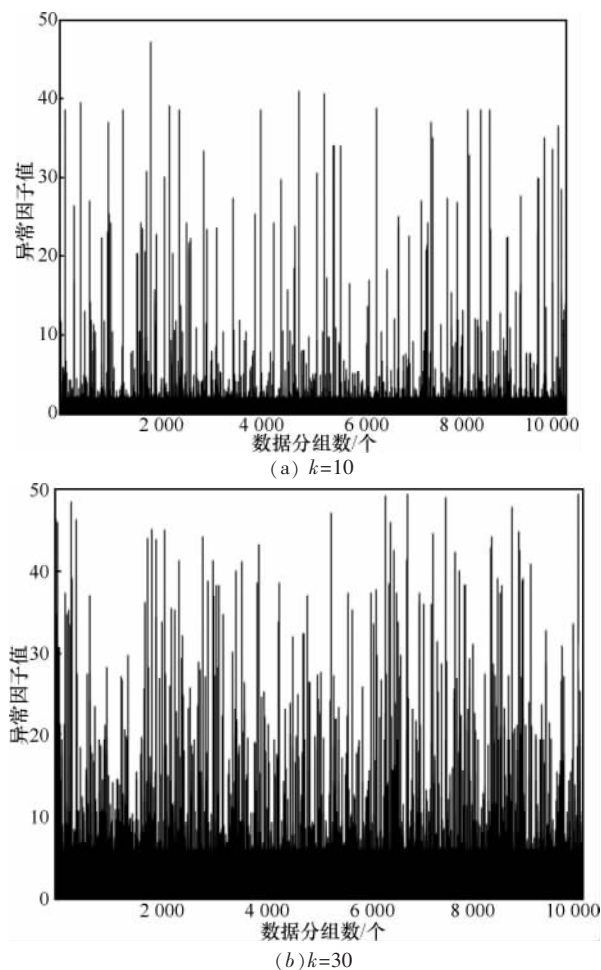


图 3 10 000 个连续流量分组检测结果

M-LOF 算法的异常判决方法是根据 M-LOF 值是否大于某个异常门限值来进行判决的。观察 $k=10$ 和 $k=30$ 两种不同情况下, M-LOF 的计算结果。在 $k=10$ 的实验中,由于异常因子值的均值在 2.5 左右, 95% 的数据都小于 2.5, 可以将异常的阈值设置为 2.5, 即所有大于 2.5 的点都判为异常。实际使用中,阈值的设置是通过经验数据来调整的,为了提高算法召回率可以将门限提高。同样方法,在 $k=30$ 的实验中,将异常的阈值设置为 5。通过比较发现,在 $k=30$ 的实验中, 90% 的数据都在门限值以下,精度要比 $k=10$ 的实验低 5%。这表明当 k 值较大时,会使小部分正常点被误

判为异常点。所以,在本实验中,选用 $k=10$ 的 M-LOF 算法作为实验模型。

作为比较分析,为了测试本算法在不同数据集时的准确度,通过设置数据集的大小,从 1 000,2 000,⋯,9 000,10 000 条数据量,对 M-LOF 算法和 LOF 算法的运行情况进行了分析比较,结果如图 4 所示。

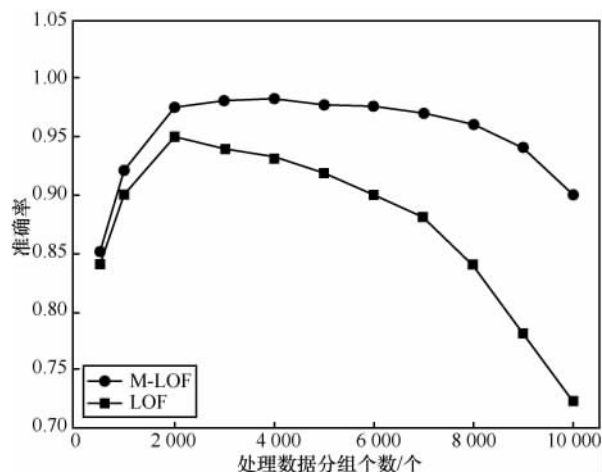


图 4 两种算法准确率对比

从图 4 可以看出,数据集较小时,两种方法的准确率都较低,这是因为数据量不够,各个数据间的距离仍然比较稀疏,基于密度的算法不能从稀疏的数据集中很好地区分异常点和正常点。当数据集大于 2 000 时,虽然随着数据集个数的增加,两种算法的效率都会下降,但是 M-LOF 算法的检测精度明显高于传统的 LOF 算法,更适合在数据量大的网络流量数据集中作为异常检测的算法模型。

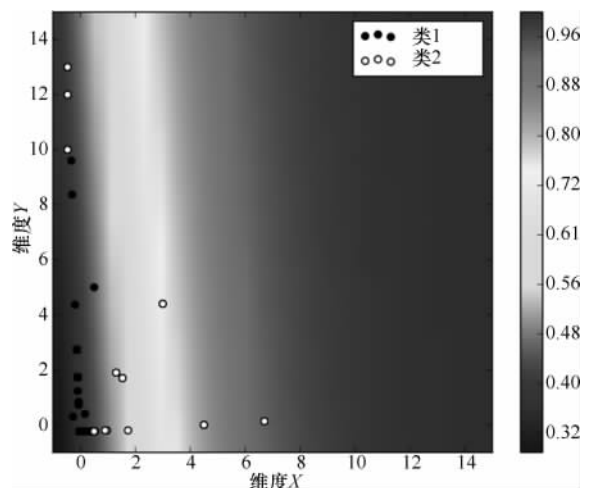
3.1.2 SVDD 算法结果

基于 SVDD 的分类算法利用 LibSVM 工具包对 2016 年 3 月 15 日的流量数据进行训练学习,在训练时,每加入一个点需要在所有已有数据集上重新训练 SVDD,最后利用 3 月 18 日的流量数据作为测试集进行实验。为了做横向比较,实验时分别对 3 种核函数(sigmoid、RBF 和 linear)进行实验作为对比,图 5 给出了分类概率密度,颜色表示该位置点为异常的概率。

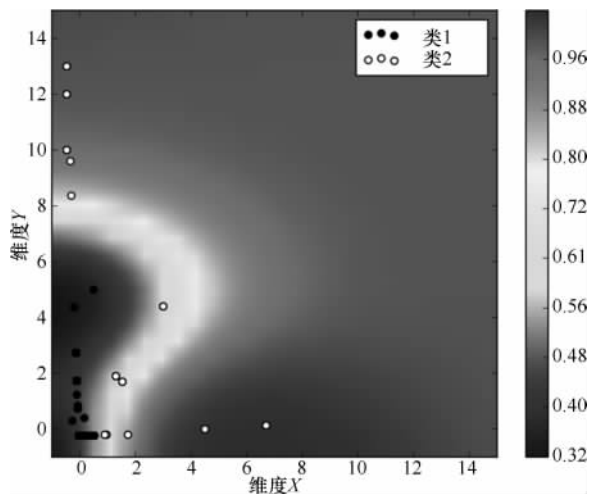
3 种方法处理 10 000 条数据集的表现见表 4。准确率(accuracy)反映了分类系统对整个样本的判定能力,能将正样本判定为正,负样本判定为负;召回率(recall)反映了被正确判定的正例占总正例的比重。

对于异常问题的检测,LibSVM 对非监督分类器的结果评价主要考虑两个方面,一个是采用半径尽可能小的超

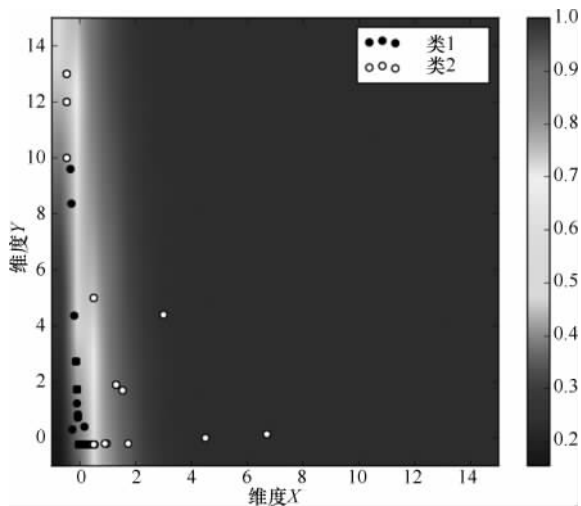
球面覆盖尽可能多的正常数据点(准确率),另一个是对异常的区分能力。



(a) sigmoid



(b) RBF



(c) linear

图 5 3 种核函数得到的 SVC 结果



表 4 3 种方法结果比较

	sigmoid-SVDD	RBF-SVDD	linear-SVDD
准确率	67.4%	96.1%	82.5%
召回率	62.1%	88.3%	76.3%

对比以上 3 种核函数的结果,可以发现 RBF 核函数的分类效果明显高于其他两个核函数,准确率达到 90%,另外从图 5(b)可以看出,RBF 给出了一个近似圆形的分界线,因为本次试验所处理的流量异常问题是非线性的,RBF 核函数可以通过非线性变换将非线性的输入数据转化为线性的输出数据,从而给出正确的分类。因此,利用 RBF 作为核函数在处理网络流量数据异常时具有很好的分类效果。

3.2 对比分析

对比 M-LOF 和 SVDD 两种无监督异常检测方法,对于处理 5 000 条和 10 000 条数据的结果见表 5 和表 6。

表 5 处理 5 000 条数据两种不同方法精度对比

	M-LOF	RBF-SVDD
准确率	97.0%	86.1%
召回率	87.4%	82.1%

表 6 处理 10 000 条数据两种不同方法精度对比

	M-LOF	RBF-SVDD
准确率	90.0%	93.1%
召回率	82.1%	88.3%

从表 5 可以看出,在数据量为 5 000 条时,提出的 M-LOF 算法具有较高的准确率,远高于 SVDD 算法的准确率。这是因为 M-LOF 是针对局域距离来计算异常点的,在数据量不大时,可以很好地利用数据局部相似性的特点。但当数据量达到 10 000 条时,由于数据流时间跨度增加,数据的局部相似性减少,因此 M-LOF 算法的精度会有所下降,而 SVDD 会有一定上升,这是因为 SVDD 算法可以持续学习,不断更新调整分类器参数来满足数据的分布状态,但缺陷是这种重复的学习过程在一定程度上会增加算法的复杂度。

综上分析,M-LOF 算法在对短时间跨度内的较小数据量进行异常检测时,检测准确率较高且复杂度低,适于对时间跨度小的较小数据量进行检测。同时针对时间跨度大的大数据量进行检测时,也可以利用 M-LOF 算法的局部特性,采用多时间窗口来处理数据,充分利用异常的时间相关性,以提高检测准确率。

相比较 M-LOF 算法,SVDD 算法具有一种持续学习的能力,但数据量越大,时间复杂度越高,适用于对实时性要求较低的时间跨度大的大数据量进行异常检测。

4 结束语

随着智能电网的不断革新发展,电力信息网络规模不断扩大,网络中出现故障的可能性也不断增加。针对电力信息网流量异常检测领域处理手段较少等问题,提出了一种基于大数据方法的异常检测机制,分别采用两种不同的无监督机器学习方法对现网中的流量数据做了挖掘分析,一种是经过改进后的 M-LOF 算法,另一种是以 RBF 为核函数的 SVDD 算法,经比较分析,两种算法各自适于不同的数据量和不同的应用环境。

综上所述,给出了基于大数据方法的流量异常检测机制,详细给出了两类算法流程。通过大量的真实电力信息网流量数据验证了上述方法具有较高的准确率和较低的误报率,能快速地对网络中的流量异常发出提前预警,从而进一步提高网络质量,增加电力信息网络的稳定性。

参考文献:

- [1] 孟小峰,慈祥. 大数据管理概念、技术与挑战[J]. 计算机研究与发展, 2013, 50(1): 146-169.
MENG X F, CI X. Big data management: concepts, techniques, challenges[J]. Journal of Computer Research and Development, 2013, 50(1): 146-169.
- [2] 张东霞,苗新,刘丽平等. 智能电网大数据技术发展研究[J]. 中国电机工程学报, 2015, 35(1): 2-12.
ZHANG D X, MIAO X, LIU L P, et al. Research on development strategy for smart grid big data[J]. Proceedings of the CSEE, 2015, 35(1): 2-12.
- [3] 宋亚奇,周国亮,朱永利. 智能电网大数据处理技术现状与挑战[J]. 电网技术, 2013, 37(4): 927-935.
SONG Y Q, ZHOU G L, ZHU Y L. Present status and challenges of big data processing in smart grid [J]. PowerSystem Technology, 2013, 37(4): 927-935.
- [4] 孙宏斌,胡江溢,刘映尚,等. 调度控制中心功能的发展——电网实时安全预警系统[J]. 电力系统自动化, 2004, 28(15): 1-6.
SUN H B, HU J Y, LIU Y S, et al. Development of the power dispatching control center-real time power security early warning system[J]. Automation of Electric Power Systems, 2004, 28(15): 1-6.
- [5] 严剑峰,于之虹,田芳,等. 电力系统在线动态安全评估和预警系统[J]. 中国电机工程学报, 2008, 28(34): 87-93.

YAN J F, YU Z H, TIAN F, et al. Dynamic security assessment and early warning system of power system[J]. Proceedings of the CSEE, 2008, 28(34): 87-93.

- [6] 汤涌, 王英涛, 田芳, 等. 大电网安全分析, 预警及控制系统的研发[J]. 电网技术, 2012, 36(7): 1-11.

TANG Y, WANG Y T, TIAN F, et al. Research and development of stability analysis, early-warning and control system for huge power grids[J]. Power System Technology, 2012, 36(7): 1-11.

- [7] 蔡斌, 吴素农, 王诗明, 等. 电网在线安全稳定分析和预警系统[J]. 电网技术, 2007, 31(2): 36-41.

CAI B, WU S N, WANG S M, et al. Power grid on-line security and stability analysis and forewarning system[J]. Power System Technology, 2007, 31(2): 36-41.

- [8] 荆铭, 邱夕兆, 延峰, 等. 电力调度数据网安全技术及其应用[J]. 电网技术, 2008, 32(26): 173-176.

JIN M, QIU X Z, YAN F, et al. Security technology of electric power dispatching data network and its application[J]. Power System Technology, 2008, 32(26): 173-176.

- [9] 许涛, 贺仁睦, 王鹏, 等. 基于统计学习理论的电力系统暂态稳定评估[J]. 中国电机工程学报, 2003, 23(11): 51-55.

XU T, HE R M, WANG P, et al. Power system transient stability assessment based on statistical learning theory[J]. Proceedings of the CSEE, 2003, 23(11): 51-55.

- [10] 黄天恩, 孙宏斌, 郭庆来, 等. 基于电网运行大数据的在线分布式安全特征选择[J]. 电力系统自动化, 2016, 40(4): 32-40.

HUANG T E, SUN H B, GUO Q L, et al. Online distributed security feature selection based on big data in power system operation [J]. Automation of Electric Power Systems, 2016, 40(4): 32-40.

- [11] 穆瑞辉, 付欢. 浅析数据挖掘概念与技术 [M]. 北京: 机械工业出版社, 2008.

MU R H, FU H. Analyze the concept of data mining and technology[M]. Beijing: China Machine Press, 2008.

- [12] 郑黎明, 邹鹏, 贾焰. 网络流量异常检测中分类器的提取与训练方法研究[J]. 计算机学报, 2012, 35(4): 719-729.

ZHENG L M, ZOU P, JIA Y. How to extract and train the classifier in traffic anomaly detection system[J]. Chinese Journal of Computers, 2012, 35(4): 719-729.

- [13] 曲朝阳, 陈帅, 杨帆, 等. 基于云计算技术的电力大数据预处理属性约简方法[J]. 电力系统自动化, 2014, 38(8): 67-71.

QU Z Y, CHEN S, YANG F, et al. An attribute reducing method for electric power big data preprocessing based on cloud computing technology[J]. Automation of Electric Power Systems, 2014, 38(8): 67-71.

- [14] 奉国和. SVM 分类核函数及参数选择比较[J]. 计算机工程与应用, 2011, 47(3): 123-124.

FENG G H. Parameter optimizing for support vector machines classification[J]. Computer Engineering and Applications, 2011, 47(3): 123-124.

[作者简介]



姜红红(1984-),女,博士,国网江苏省电力公司南京供电公司信息工程师,主要研究方向为智能电网、网络服务质量管理、电力信息化。



张涛(1992-),男,北京邮电大学网络与交换技术国家重点实验室网络管理研究中心硕士生,主要研究方向为智能电网、自组织网络和无线网络管理技术。



赵新建(1988-),男,国网江苏省电力公司南京供电公司信息工程师,主要研究方向为智能电网、电力信息化。



钱欣(1989-),女,国网江苏省电力公司南京供电公司信息工程师,主要研究方向为智能电网、电力信息化。



赵天成(1990-),男,国网江苏省电力公司南京供电公司信息工程师,主要研究方向为智能电网、电力信息化。



高丽莎(1982-),女,国网江苏省电力公司南京供电公司信息工程师,主要研究方向为智能电网、电力信息化。