·安全技术·　　文章编号：1000-3428(2019)10-0122-08　　文献标志码：A　　中图分类号：TP309

基于改进高斯核函数的 BGP 异常检测方法

戴仙波[1,2]，王　娜[1,2]，刘　颖[1,2]

（1. 信息工程大学 密码工程学院，郑州 450001；2. 河南省信息安全重点实验室，郑州 450001）

摘　要： 通过将边界网关协议（BGP）更新报文激增异常问题抽象为二分类问题，提出一种基于改进高斯核函数的 BGP 异常检测（IGKAD）方法。采用 FMS 特征选择算法，选择能同时最大化类间距离和最小化类内距离的特征，得到度量分类能力的特征权值。利用基于 Manhattan 距离与特征权值的改进高斯核函数构造支持向量机（SVM）分类模型，并结合基于网格搜索与交叉验证的参数寻优方法，提高 SVM 模型分类准确率。通过设计特征效率函数，给出最优特征子集构造方法，从而选取最优特征子集作为训练数据集。实验结果表明，当训练集包含 TOP10 和 TOP8 特征时，IGKAD 方法的分类准确率分别为 91.65% 和 90.37%，相比基于机器学习的 BGP 异常检测方法分类性能更优。

关键词： 高斯核函数；边界网关协议；异常检测；支持向量机；机器学习

开放科学（资源服务）标志码（OSID）：

中文引用格式： 戴仙波，王娜，刘颖. 基于改进高斯核函数的 BGP 异常检测方法[J]. 计算机工程，2019，45(10)：122-129.

英文引用格式： DAI Xianbo，WANG Na，LIU Ying. BGP anomaly detection method based on improved Gauss kernel function[J]. Computer Engineering，2019，45(10)：122-129.

BGP Anomaly Detection Method Based on Improved Gauss Kernel Function

DAI Xianbo[1,2]，WANG Na[1,2]，LIU Ying[1,2]

（1. College of Cipher Engineering，Information Engineering University，Zhengzhou 450001，China；
2. Henan Key Laboratory of Information Security，Zhengzhou 450001，China）

【Abstract】 By abstracting the Border Gateway Protocol（BGP）update message augmentation anomaly problem into a two-class problem，an Improved Gaussian Kernel Function-based BGP Anomaly Detection（IGKAD）method is proposed. The Fisher-Markov Slector（FMS）feature selection algorithm is used to select the feature that can simultaneously maximize the distance between classes and minimize the distance within the class，and obtain the feature weights of metric classification ability. The improved Gaussian kernel function based on Manhattan distance and feature weight is used to construct the Support Vector Machine（SVM）classification model，and the parameter optimization method based on grid search and cross-validation is combined to improve the classification accuracy of SVM model. By designing the feature efficiency function，the optimal feature subset construction method is given，which is selected as the training dataset. Experimental results show that when the training set contains TOP10 and TOP8 features，the classification accuracy of the IGKAD method is 91.65% and 90.37%，respectively. Compared with the machine learning-based BGP anomaly detection method，the classification performance is better.

【Key words】 Gauss kernel function；Border Gateway Protocol（BGP）；anomaly detection；Support Vector Machine（SVM）；machine learning

DOI：10.19678/j.issn.1000-3428.0052713

0　概述

边界网关协议（Border Gateway Protocol，BGP）[1]作为互联网域间路由标准协议，承担着建立、维持所有自治系统（Autonomous System，AS）之间网络可达性的功能，对整个互联网的可靠稳定运行起到至关重要的作用。但错误配置、链路故障甚至病毒爆发等均容易导致 BGP 异常事件的发生，使得数据流被

基金项目： 国家重点研发计划（2018YFB0803603）；国家自然科学基金（61802436，61502531）；河南省自然科学基金（162300410334）。
作者简介： 戴仙波（1994—），男，硕士研究生，主研方向为网络与信息安全；王　娜，副教授、博士；刘　颖，副教授。
收稿日期： 2018-09-20　　**修回日期：** 2018-10-31　**E-mail：** beyond_dxb@126.com

重定向,形成流量黑洞或在极短时间内产生大量的 BGP 更新报文,破坏全球互联网的可达性和稳定性[2-3]。近年来,BGP 异常事件频繁发生。例如,2012 年 8 月,加拿大运营商 Dery Telecom(AS46618)由于错误配置将其全部路由表泄露给了提供商 Bell Canada(AS577),导致产生“BGP updates storm”,破坏了全球互联网的稳定性和部分网络的可达性[4];2017 年4 月,瑞士运营商 SwissIX(AS200759)因配置错误发起前缀劫持攻击,影响了全球 576 个自治系统和3 431 个前缀的可达性,甚至包括 Google、Amazon、Twitter、Apple 等[5]。针对 BGP 异常,如果能够及时准确地检测出异常事件的发生,通过采取有效的应对措施,则将显著降低 BGP 异常事件对全球互联网可达性与稳定性的影响。

本文提出一种基于改进高斯核函数的 BGP 异常检测(Improved Gauss Kernel Function-based BGP Anomaly Detection, IGKAD)方法。IGKAD 方法采用 FMS(Fisher-Markov Selector)特征选择算法、基于 Manhattan 距离和特征权值的改进高斯核函数以及基于网格搜索和交叉验证的模型参数寻优方法,实现 BGP 更新报文激增情况的异常检测。

1　相关工作

根据事件后果,BGP 异常可分为数据流劫持异常和更新报文激增异常。数据流劫持异常会导致受害网络数据流的重定向,形成流量黑洞,破坏受害网络的可达性。更新报文激增异常会导致在极短时间内产生大量的 BGP 更新报文,破坏全球互联网的稳定性。本文主要研究 BGP 更新报文激增异常的检测。目前 BGP 异常检测方法一般分为 5 类,分别是基于统计模式识别的方法、基于历史 BGP 更新报文的方法、基于可达性验证的方法、基于时间序列分析的方法和基于机器学习的方法。基于统计模式识别的方法[6]采用统计概率论进行模式识别,根据模式之间的距离函数来判定异常,能够同时检测数据流劫持异常和更新报文激增异常。但该方法正确估计高维数据分布困难、检测速度慢,实际应用中需人工确定模型参数的阈值。基于历史 BGP 更新报文的方法[7]和基于可达性验证的方法[8]仅能检测数据流劫持异常,前者利用历史数据对 BGP 异常路由进行检测,后者根据目标前缀的可达性验证结果进行异常检测。基于时间序列分析的方法和基于机器学习的方法能够检测更新报文激增异常。基于时间序列分析的方法[9]将 BGP 更新报文视为一个多维的时间序列,通过选择合适的滑动时间窗口实现异常检测。但该方法难以确定时间窗口的大小,时间窗口过小会导致模型可利用的信息量不够,时间窗口过大又会导致模型对局部异常不敏感,使得漏报率上升。近年来,基于机器学习的方法在 BGP 异常检测领域得到一定应用。从机器学习角度来看,BGP 异常检测问题可抽象为二分类问题[10],目的是将未知的 BGP 更新报文识别为正常报文或异常报文,从而实现 BGP 异常检测。如文献[11]从公开路由数据集中提取出 37 个特征,通过支持向量机算法和隐马尔科夫模型,达到 81.5% 的分类准确率。文献[12]利用决策树和模糊粗糙集方法进行特征选择,然后基于特征子集构造极限学习机模型,达到 80.08% 的分类准确率。需要指出的是,目前基于机器学习的 BGP 更新报文激增异常检测方法在分类准确率上有待进一步提高。

2　FMS 特征选择算法

特征是机器学习的核心,数据和特征决定了机器学习的上限,会直接影响模型的分类性能。因此本文采用特征提取将原始 BGP 更新报文转换为一组具有明显统计意义的特征,这些特征能够较好地描述原始 BGP 更新报文的内部结构。经由特征提取[8]过程得到的 37 个特征具体定义如表 1 所示。

表 1　特征具体定义

特征	定义
1	类型字段为声称的更新报文数量
2	类型字段为撤销的更新报文数量
3	声称的 NLRI 前缀数量
4	撤销的 NLRI 前缀数量
5	平均 AS 路径长度
6	最大 AS 路径长度
7	平均唯一 AS 路径长度
8	双重声称数量
9	双重撤销数量
10	隐式撤销数量
11	平均编辑距离
12	最大编辑距离
13	间隔时间
14 ~ 24	最大编辑距离 m,$m=7,8,\cdots,17$
25 ~ 33	最大 AS 路径长度 n,$n=7,8,\cdots,15$
34	IGP 包数量
35	EGP 包数量
36	INCOMPLETE 包数量
37	包大小

由于冗余特征的存在,基于高维特征构造分类模型会增大计算开销,并且噪声数据会降低模型分类准确率,因此在数据预处理阶段经由特征提取[13-14]过程得到相应特征集合的基础上,需要进一步删除冗余和不相关特征,寻找区分类别能力最优

的特征子集,达到降低特征矩阵的维数和计算复杂度,同时提高模型分类准确率的目的。FMS 特征选择算法[15]能够根据 Fisher 线性分析和 Markov 随机场技术选择出能同时最大化类间距离和最小化类内距离的特征。采用 FMS 算法的特征选择过程与分类过程相互独立,根据数据内在属性来衡量相关性,并对每一个特征的权值大小进行排序,权值越大表示该特征区分类别的能力越强。该方法效率较高,不仅可以保证全局最优,而且计算复杂度较低,适合处理大规模数据。

将训练数据表示为 $\{(\boldsymbol{x}_k,\boldsymbol{y}_k)\}_{k=1}^{n}$,$\boldsymbol{x}_k\in\boldsymbol{R}^p$ 表示 p 维特征向量,$\boldsymbol{y}_k\in\{\omega_1,\omega_2,\cdots,\omega_g\}$ 表示类标签,$\boldsymbol{C}_i(i=1,2,\cdots,g)$ 表示第 i 个类,每一个类 $\boldsymbol{C}_i$ 有 n_i 个样本。

定义 1 类内散布矩阵、类间散布矩阵和总体散布矩阵分别记为 $\boldsymbol{S}_{\mathrm{w}}$、$\boldsymbol{S}_{\mathrm{b}}$ 和 $\boldsymbol{S}_{\mathrm{t}}$。

$$\boldsymbol{S}_{\mathrm{w}}=\frac{1}{n}\sum_{j=1}^{g}\sum_{i=1}^{n_j}(\boldsymbol{x}_i^{(j)}-\mu_j)(\boldsymbol{x}_i^{(j)}-\mu_j)^{\mathrm{T}}$$

$$\boldsymbol{S}_{\mathrm{b}}=\frac{1}{n}\sum_{j=1}^{g}n_j(\mu_j-\mu)(\mu_j-\mu)^{\mathrm{T}}$$

$$\boldsymbol{S}_{\mathrm{t}}=\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i-\mu)(\boldsymbol{x}_i-\mu)^{\mathrm{T}}=\boldsymbol{S}_{\mathrm{w}}+\boldsymbol{S}_{\mathrm{b}}\tag{1}$$

其中,$\boldsymbol{x}_i^{(j)}$ 表示 $\boldsymbol{C}_j$ 类的第 i 个样本,$\mu_j=\sum_{i=1}^{n_j}\boldsymbol{x}_i^{(j)}/n_j$ 表示 $\boldsymbol{C}_j$ 类的样本均值,$\mu=\sum_{i=1}^{n}\boldsymbol{x}_i/n$ 表示总体均值。

定义 2 从输入空间 $\boldsymbol{R}^p$ 到核空间 $\boldsymbol{R}^D$ 的非线性映射 $\phi(\cdot)$ 定义如下:

$$\phi:\boldsymbol{R}^p\rightarrow\boldsymbol{R}^D\tag{2}$$

定义 3 核函数 $k(\cdot,\cdot)$ 满足:

$$<\phi(\boldsymbol{x}_1),\phi(\boldsymbol{x}_2)>=k(\boldsymbol{x}_1,\boldsymbol{x}_2)\tag{3}$$

其中,运算 $<\cdot,\cdot>$ 代表核空间下的点积。

定义 4 $\boldsymbol{K}$ 和 $\boldsymbol{K}^{(i)}$ 分别为 n 阶和 n_i 阶方阵,且满足:

$$\{\boldsymbol{K}\}_{kl}=k(\boldsymbol{x}_k,\boldsymbol{x}_l)$$

$$\{\boldsymbol{K}^{(i)}\}_{uv}=k(\boldsymbol{x}_u^{(i)},\boldsymbol{x}_v^{(i)})\tag{4}$$

其中,$k,l\in\{1,2,\cdots,n\}$,$u,v\in\{1,2,\cdots,n_i\}$,$i=1,2,\cdots,g$。

定义 5 $\boldsymbol{S}_{\mathrm{w}}$、$\boldsymbol{S}_{\mathrm{b}}$ 和 $\boldsymbol{S}_{\mathrm{t}}$ 在核空间下分别记为 $\tilde{\boldsymbol{S}}_{\mathrm{w}}$、$\tilde{\boldsymbol{S}}_{\mathrm{b}}$ 和 $\tilde{\boldsymbol{S}}_{\mathrm{t}}$,对应的迹分别为:

$$\mathrm{Tr}(\tilde{\boldsymbol{S}}_{\mathrm{w}})=\frac{1}{n}\mathrm{Tr}(\boldsymbol{K})-\frac{1}{n}\sum_{i=1}^{g}\frac{1}{n_i}\mathrm{Sum}(\boldsymbol{K}^{(i)})$$

$$\mathrm{Tr}(\tilde{\boldsymbol{S}}_{\mathrm{b}})=\frac{1}{n}\sum_{i=1}^{g}\frac{1}{n_i}\mathrm{Sum}(\boldsymbol{K}^{(i)})-\frac{1}{n^2}\mathrm{Sum}(\boldsymbol{K})$$

$$\mathrm{Tr}(\tilde{\boldsymbol{S}}_{\mathrm{t}})=\frac{1}{n}\mathrm{Tr}(\boldsymbol{K})-\frac{1}{n^2}\mathrm{Sum}(\boldsymbol{K})\tag{5}$$

其中,$\mathrm{Sum}(\cdot)$ 表示计算矩阵所有元素的总和。

定义 6 定义特征选择向量为:

$$\boldsymbol{\alpha}=[\alpha_1,\alpha_2,\cdots,\alpha_p]^{\mathrm{T}}\in\{0,1\}^p\tag{6}$$

其中,$\alpha_k=1$ 表明第 k 个特征被选中,$\alpha_k=0$ 表明第 k 个特征没有被选中。

特征向量 $\boldsymbol{x}$ 选定的特征由 $\boldsymbol{x}(\boldsymbol{\alpha})=\boldsymbol{x}\odot\boldsymbol{\alpha}$ 给出,$\odot$ 表示 Hadamard 积。因此,可将最大化类分离的特征选择准则转化为一个无约束最优化问题。

$$\underset{\boldsymbol{\alpha}\in\{0,1\}^p}{\mathrm{argmax}}\ \mathrm{Tr}(\tilde{\boldsymbol{S}}_{\mathrm{b}})(\boldsymbol{\alpha})-\gamma\mathrm{Tr}(\tilde{\boldsymbol{S}}_{\mathrm{t}})(\boldsymbol{\alpha})\tag{7}$$

其中,γ 是自由参数,分析表明 $\gamma\leqslant 0$ 能得到一个更好的分类效果,在 γ 的合理波动范围内,分类器的实验效果对 γ 不敏感[15]。为能处理线性不可分离和冗余特征带来的高噪声数据集,加入 l_0 范数,即特征选择准则优化为:

$$\underset{\boldsymbol{\alpha}\in\{0,1\}^p}{\mathrm{argmax}}\{\mathrm{Tr}(\tilde{\boldsymbol{S}}_{\mathrm{b}})(\boldsymbol{\alpha})-\gamma\mathrm{Tr}(\tilde{\boldsymbol{S}}_{\mathrm{t}})(\boldsymbol{\alpha})-\beta\|\boldsymbol{\alpha}\|_0\}\tag{8}$$

其中,正则因子 β 表示全局阈值。

考虑如下线性核函数:

$$k_1(\boldsymbol{x}_1,\boldsymbol{x}_2)(\boldsymbol{\alpha})=1+<\boldsymbol{x}_1\odot\boldsymbol{\alpha},\boldsymbol{x}_2\odot\boldsymbol{\alpha}>=1+\sum_{i=1}^{p}\boldsymbol{x}_{1i}\boldsymbol{x}_{2i}\alpha_i\tag{9}$$

将式(9)代入到式(8)得到:

$$\underset{\boldsymbol{\alpha}\in\{0,1\}^p}{\mathrm{argmax}}\sum_{j=1}^{p}(\theta_j-\beta)\alpha_j\tag{10}$$

θ_j 定义如式(11)所示。

$$\theta_j=\frac{1}{n}\sum_{i=1}^{g}\frac{1}{n_i}\sum_{u,v=1}^{n_i}\boldsymbol{x}_{uj}^{(i)}\boldsymbol{x}_{vj}^{(i)}-\frac{\gamma}{n}\sum_{i=1}^{n}\boldsymbol{x}_{ij}^2+\frac{\gamma-1}{n^2}\sum_{u,v=1}^{n}\boldsymbol{x}_{uj}\boldsymbol{x}_{vj}\tag{11}$$

其中,θ_j 用来衡量特征在类分离判别中的重要程度,即特征权值。θ_j 值越大,表明第 j 个特征越重要。

对于给定的 β 和 γ,结合式(10),FMS 特征选择算法得到一个最优的特征选择向量 $\boldsymbol{\alpha}^*\in\{0,1\}^p$,满足:

$$\theta_j>\beta\Leftrightarrow\alpha_j^*=1\tag{12}$$

若 $\alpha_j^*=1$,则表明第 j 个特征被选中;若 $\alpha_j^*=0$,则表明第 j 个特征未被选中。

FMS 特征选择算法的伪代码如算法 1 所示,该算法的计算复杂度为 $O(n^2p)$。

算法 1 FMS 特征选择算法

输入 特征矩阵 $\boldsymbol{X}=[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n]\in\mathbb{R}^{p\times n}$,类标签向量 $\boldsymbol{Y}=[\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_n]$,$\boldsymbol{y}_k\in\{\boldsymbol{w}_1,\boldsymbol{w}_2,\cdots,\boldsymbol{w}_g\}$,$k=1,2,\cdots,n$,参数 γ、β

输出 特征选择向量 $\boldsymbol{\alpha}^*=[\alpha_1^*,\alpha_2^*,\cdots,\alpha_p^*]^{\mathrm{T}}\in\{0,1\}^p$

```
1. begin
2. for j = 1→p do
3. 由式(11)计算 θ_j
4. if θ_j > β then
5. α_j^* = 1
6. else
7. α_j^* = 0
8. end if
9. return α_j^*, θ_j
10. end for
11. end
```

3　基于改进高斯核函数的 SVM 模型

3.1　SVM 模型与高斯核函数

SVM[16]是在统计学习理论的 VC 维理论和结构风险最小化原理基础上发展起来的一种机器学习方法。SVM 在很大程度上克服了“维数灾难”“过学习”等问题,在解决非线性及高维模式识别问题中表现出特有优势,已成为机器学习的主流技术之一。

对于输入空间中线性不可分的情形(如图 1 所示),SVM 的基本思想是利用满足 Mercer 条件的核函数将输入空间映射到一个高维特征空间,在高维特征空间中找到一个划分超平面,使得样本能够线性可分。表 2 列出了 4 类常用的核函数。Mercer 条件使得相应的优化问题转化为凸问题,因此没有局部最小。模式识别理论已经证明:如果输入空间是有限维,即特征维数有限,那么一定存在一个高维特征空间使得样本可分。

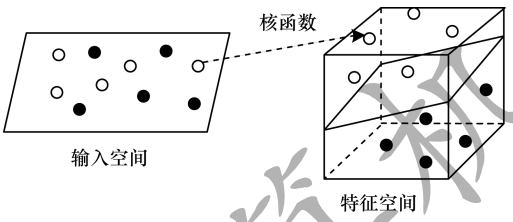

图 1　输入空间中线性不可分的情形

表 2　常用核函数

函数名称	表达式
线性核函数	$k(\boldsymbol{x}_i,\boldsymbol{x}_j)=\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j$
多项式核函数	$k(\boldsymbol{x}_i,\boldsymbol{x}_j)=(\gamma\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j+r)^d,\gamma>0$
高斯核函数	$k(\boldsymbol{x}_i,\boldsymbol{x}_j)=\exp(-\gamma\parallel\boldsymbol{x}_i-\boldsymbol{x}_j\parallel^2),\gamma>0$
Sigmoid 核函数	$k(\boldsymbol{x}_i,\boldsymbol{x}_j)=\tanh(\gamma\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j+r)$

在表 2 中,$\boldsymbol{x}_i$、$\boldsymbol{x}_j$表示输入向量,γ 表示斜率,d 表示多项式度数,r 表示常数项。高斯核函数可将样本非线性地映射到更高维空间,因此与线性核不同,可以处理类标签和特征之间的关系是非线性的情况。文献[17]指出线性核是高斯核的一种特殊情况,因为带惩罚参数的线性核与带参数的高斯核具有相同的性能。多项式核比高斯核具有更多的超参数,而参数数量影响模型复杂程度,且考虑到多项式核的表达式,可能会导致数值计算困难。例如,当度数 d 很大且$(\gamma\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j+r)<1$ 时,值趋于 0;当度数 d 很大且$(\gamma\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j+r)>1$ 时,值趋于无穷大。此外,文献[18]研究表明,Sigmoid 核在特定参数下性能与高斯核几乎相同,而且由 Sigmoid 核计算得到的核矩阵不一定是正定矩阵,由它构造的 SVM 模型在准确性上通常也不如高斯核。因此,本文选择基于高斯核函数的 SVM 模型。

3.2　基于 Manhattan 距离与特征权值的高斯核函数

传统高斯核函数采用 Euclidean 距离度量两个向量之间的距离,但是 Euclidean 距离在一定程度上会增强误差较大的元素在距离计算中的作用,影响 SVM 的分类准确率。基于此,本文采用 Manhattan 距离作为高斯核函数中衡量两个向量之间的距离测度方法。Manhattan 距离中各元素的误差对整体距离的影响相同,使得数值更具可比性,并且运算量较低。

如果在距离计算中能够体现出各特征对分类的贡献程度,则将会使分类方法更贴合 BGP 的数据特点,进一步提高分类准确率。据此,引入特征权值来度量特征对分类的贡献程度,提出基于 Manhattan 距离与特征权值的改进高斯核函数,记为 $k'(\boldsymbol{x},\boldsymbol{y})$,如式(13)所示。

$$k'(\boldsymbol{x},\boldsymbol{y})=\exp(-\gamma\delta(\boldsymbol{x},\boldsymbol{y})) \tag{13}$$

其中,$\delta(\boldsymbol{x},\boldsymbol{y})$ 表示两向量间的 Manhattan 距离,如式(14)所示。

$$\delta(\boldsymbol{x},\boldsymbol{y})=\sum_{i=1}^{n}\theta_i\left|\boldsymbol{x}_i-\boldsymbol{y}_i\right| \tag{14}$$

根据平移不变核的定义[19],可以证明改进高斯核函数 $k'(\boldsymbol{x},\boldsymbol{y})$ 的有效性。平移不变核是一类重要的核函数,如高斯核函数。

定义 7　平移不变核形如 $k(\boldsymbol{x},\boldsymbol{y})=f(\boldsymbol{x}-\boldsymbol{y})$,其中 $f:\boldsymbol{X}\rightarrow\boldsymbol{R}$ 是有界可积连续函数,则 $k(\boldsymbol{x},\boldsymbol{y})$ 是核函数的充分必要条件为 $f(0)>0$,且其 Fourier 变换满足:

$$f(\omega)=\int_{-\infty}^{+\infty}f(\boldsymbol{x})\mathrm{e}^{-\mathrm{i}<\omega,x>}\mathrm{d}\boldsymbol{x}\geqslant 0 \tag{15}$$

命题　$k'(\boldsymbol{x},\boldsymbol{y})$ 是有效的核函数。

证明　因为 $f(\boldsymbol{x})=\exp(-\gamma\delta(\boldsymbol{x}))$,所以 $f(\boldsymbol{x})$ 是有界可积连续函数,且 $f(0)=1$。又因为 $f(\boldsymbol{x})\mathrm{e}^{-\mathrm{i}<\omega,x>}=\mathrm{e}^{-\gamma\delta(x)-\mathrm{i}<\omega,x>}\geqslant 0$,所以 $f(\omega)\geqslant 0$ 恒成立。综上,$k'(\boldsymbol{x},\boldsymbol{y})$ 是有效的核函数,命题证毕。

3.3　基于网格搜索与交叉验证的参数寻优方法

SVM 模型的性能依赖一对重要的参数(C, *gamma*),其中 C 为惩罚因子,表示对误差的容忍度。C 越大,表明模型越不能容忍出现误差,越容易导致模型过拟合;相反地,C 越小,越容易导致模型欠拟合。C 过大或过小,均会降低模型的泛化能力,因此参数 C 的适当取值对模型分类准确率和泛化能力的提升具有重要意义。*gamma* 是多项式核、高斯核以及 Sigmoid 核中的一个参数,它决定了数据映射到新的特征空间后的分布。*gamma* 值越大,则支持向量越少,*gamma* 值越小,则支持向量越多,支持向量的个数会影响模型训练与预测的速度。

考虑到训练数据集两类样本的不平衡性(如表 3所示),以总体分类准确率为评价目标的传统分类算法会过多关注多数类样本,从而使得少数类样本的分类性能下降。为此,在(C, *gamma*)的寻优过程中,需要充分考虑少数类样本数据,使两类样本在训练过程中具有相同的“话语权”。本文按照两类样本数目大小比值的反比分别为两类样本赋予权值,这样可以有效解决数据不均衡问题。

表 3 训练数据集样本设置

类别	标签	样本数量	权值
异常类	+1	4 392	2.28
正常类	-1	10 008	1.00

核参数的选取是一个难点,目前还没有国际公认的普适性方法,实际应用中只能依靠实验比较或经验所得,因此本文在不平衡数据集约束下结合网格搜索与交叉验证进行参数寻优(如图 2 所示),将(C,$gamma$)的搜索范围按照取值划分成网格,网格中的每个点代表一种参数组合方案。网格搜索范围满足式(16),步长为 1,即 $C\in\{2^{-5},2^{-4},\cdots,2^{5}\}$ 且 $gamma\in\{2^{-4},2^{-3},\cdots,2^{0}\}$。

$$\begin{cases}\mathrm{lb}\ C\in[-5,5],\text{且取值为整数}\\ \mathrm{lb}(gamma)\in[-4,0],\text{且取值为整数}\end{cases}\tag{16}$$

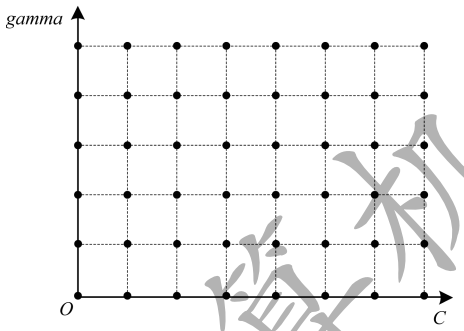

图 2 网格搜索过程

每一个网格点按如下流程进行交叉验证:将总的训练集平均分成 N 个子集,其中 $N-1$ 个作为训练集,余下 1 个作为测试集。每次用测试集去测试训练后的模型会得到一个分类准确率,当 N 个子集均做过测试集后,取 N 折交叉验证分类准确率的平均值。这样遍历网格内所有点,取分类准确率平均值最大的点即为对应的性能最优的(C,$gamma$),流程如图 3 所示。需要指出的是,本文采用 5 折交叉验证,且由于(C,$gamma$)在搜索过程中均选取搜索范围有限且离散的值,因此(C,$gamma$)可能仅为局部最优解。

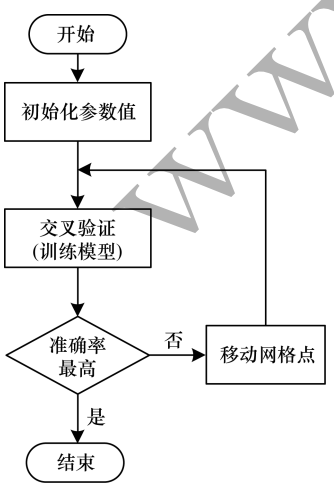

图 3 参数寻优流程

4 特征效率函数

基于 FMS 特征选择算法可以得到各特征的权值,将各个特征按权值降序排列,根据排序结果将特征依次加入模型训练集中。实验发现,因为前期加入训练集中特征的权值较大,模型的分类准确率会逐渐提高,但随着后期权值较低特征的加入,且数据集中噪声和冗余数据的存在,此时模型分类准确率的增长速度会放缓甚至使准确率下降。但与此同时,SVM 模型的训练时间一直会随着特征数量的增加而增加。因此,持续增加特征用于模型训练并不合适。基于此,本文提出特征效率函数的概念,以度量模型分类准确率和模型训练时间之间的关系。根据特征效率函数最大值构建特征子集,此时的特征子集为最优特征子集,能够使模型性能(即模型的分类准确率和训练时间)达到综合最优。

定义 8 函数 $f(n)$($n\in\mathbb{Z}$)是模型分类准确率关于特征数量 n 的函数。

定义 9 函数 $g(n)$($n\in\mathbb{Z}$)是模型训练时间关于特征数量 n 的函数。

由上述定义可知,函数 $f(n)$ 和 $g(n)$ 分别描述了模型训练集包含一定数量的特征时,模型的分类准确率与模型训练时间的大小。为评价模型的最优综合性能,定义特征效率函数。

定义 10 $h(n)$ 是关于特征数量 n 的特征效率函数,其表达式为:

$$h(n)=\frac{f(n)}{g(n)},n\in\mathbb{Z}\tag{17}$$

其中,$h(n)$ 描述了单位时间内分类准确率的大小,若单位时间内分类准确率越大,即 $h(n)$ 越大,则模型综合性能越优。

定义 11 将 $h(n)$ 取得最大值的点 n_0 作为模型的最优点。

最优点描述了当 $n=n_0$ 时,模型在单位时间内取得的最大分类准确率,此时模型综合性能达到最优。根据特征权值排序,基于 TOP n_0 构建的特征子集,能使模型性能达到综合最优。

5 实验结果与分析

5.1 实验数据集与数据标准化

5.1.1 实验数据集

实验选取 AS513(RIPE RIS、rcc04、CIXP、Geneva)提供的 Slammer[20]、Nimda[21] 以及 Code Red I[22] 爆发时的 BGP 更新报文作为 BGP 异常数据集,其中,Slammer 和 Nimda 为训练集,Code Red I 为测试集。利用 libBGPdump 工具[23] 将路由数据从 MRT 格式[24] 转换为 ASCII 格式,而后基于 C#编写的工具解析 ASCII 文件并提取特征统计信息。5 天内每间隔 1 min 抽样统计 1 次特征值,从而获得每个异常事件的 7 200 个样本,经过数据标准化及特征选择后可被

用作分类模型的输入。每个事件前后 2 天的样本被认为是正常数据集,第 3 天的样本是每个异常活动的高峰期。

5.1.2　数据标准化

为消除量纲和数值大小的影响,需要进行数据标准化处理,从而使得不同特征能够进行比较和加权。本文采用 Z-score 标准化方法,如式(18)所示。

$$x' = \frac{x - \mu}{\sigma} \tag{18}$$

其中,μ 表示总体均值,σ 代表总体标准差。

本文考虑到 BGP 数据集来自于抽样统计,使用样本均值代替总体均值,用样本标准差代替总体标准差,处理方法如式(19)所示。

$$x' = \frac{x - \bar{x}}{S} \tag{19}$$

其中,$\bar{x}$ 代表样本均值,S 代表样本标准差。

5.2　评估指标

由于训练数据集是不平衡数据集,仅由分类准确率这一指标来衡量模型的分类效果并不合适,因此为全面准确地评价模型性能,本文综合分类准确率和 F1-Score 两个评价指标来全面衡量模型性能。

5.2.1　准确率

基于分类混淆矩阵(如表 4 所示)计算分类准确率。对于二分类问题,将模型预测的类别与样本真实类别进行组合,分为 *TP*、*FN*、*FP*、*TN* 4 种情形,其中:*TP* 表示样本是正例并且预测结果也是正例;*FN* 表示样本是正例,却错误预测为负例;*FP* 表示样本是负例,预测结果为正例;*TN* 表示样本是负例,预测结果也是负例。

表 4　分类混淆矩阵

真实类别	预测类别	
	正例	负例
正例	*TP*	*FN*
负例	*FP*	*TN*

将分类准确率(*Accuracy*)定义为:

$$Accuracy = \frac{TP + TN}{TP + FN + FT + TN} \tag{20}$$

5.2.2　F1-Score

F-Score 是精确率(Precision)和召回率(Recall)的加权平均调和值,精确率反映了测试集样本被模型预测的正例样本中真实正例样本所占的比例,召回率反映了测试集样本中被模型正确预测的正例样本占总的正例样本的比例,精确率与召回率在多数情况下两者相互制约。精确率、召回率和 F-Score 分别定义如下:

$$Precison = \frac{TP}{TP + FP} \tag{21}$$

$$Recall = \frac{TP}{TP + FN} \tag{22}$$

$$F\text{-}Score = (1 + \beta^2)\frac{Precision \times Recall}{\beta^2(Precison + Recall)} \tag{23}$$

其中,$\beta > 0$ 衡量了精确率和召回率的重视程度,$\beta < 1$ 时对精确率有更大影响,$\beta > 1$ 时对召回率有更大影响。本文令 $\beta = 1$,表示对精确率和召回率同等重视,此时的 F-Score 又称为 F1-Score,计算公式为:

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precison + Recall} \tag{24}$$

5.3　实验结果

实验采用 SVM 模式识别与回归软件包 libsvm-3.22[25],具有 C、Java、Matlab、Python 等多种语言版本,代码开源,易于扩展。本文选择在 Intel© Core™ i7-6500U CPU @ 2.50 GHz 四核处理器、4.00 GB 内存、运行64 位 Windows 10 系统的 PC 机上进行实验。

5.3.1　特征选择算法实验结果

基于已构造的实验训练集,采用 FMS 特征选择算法计算各特征的重要性指数(即特征权值)。计算结果如图 4 所示。TOP10 特征及其权值如表 5 所示。

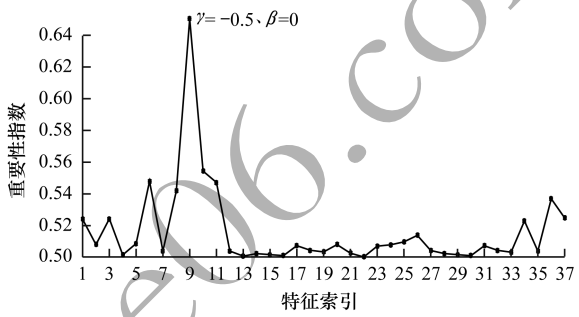

图 4　特征权值分布

表 5　TOP10 特征及其权值

特征排序	特征索引	特征权值
1	9	0.650 3
2	10	0.554 4
3	6	0.547 9
4	11	0.547 2
5	8	0.541 9
6	36	0.536 9
7	37	0.524 9
8	3	0.524 3
9	1	0.524 1
10	34	0.523 0

5.3.2　改进高斯核函数实验结果

本文对改进高斯核函数和传统高斯核函数的分类准确率进行对比实验。如图 5 所示,当训练集为 TOP10 特征时,改进高斯核函数的分类准确率达到 91.65%,优于传统高斯核函数,达到较好的分类效果。可以看出,随着后续低权值特征的加入,分类准确率有下降趋势,而模型训练时间却不断增长,如图 6所示。训练时间为模型 10 次运行的 CPU 平均耗时。

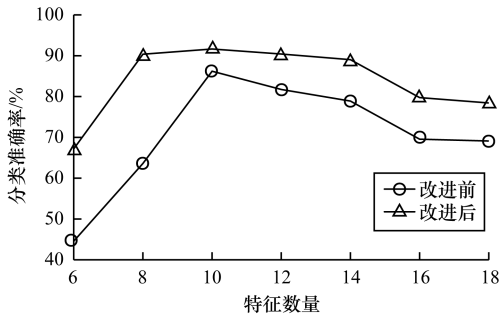

图 5 高斯核函数改进前后的分类准确率对比

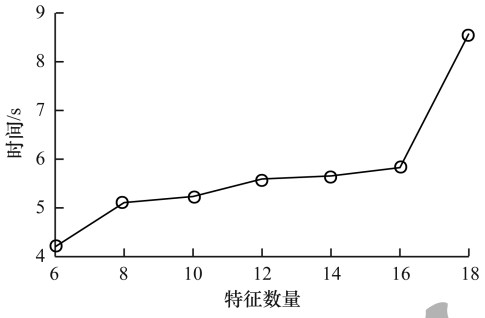

图 6 改进高斯核函数的模型训练时间

5.3.3 特征效率函数实验结果

实验验证特征效率函数的有效性并确定最优特征子集,实验结果如图 7 所示。可以看出,当特征数量为 8 时,特征效率函数取得最大值。由定义 11可知,最优点为 8。当特征子集包含 TOP8 特征时,模型性能达到综合最优,此时的分类准确率为 90.37%。

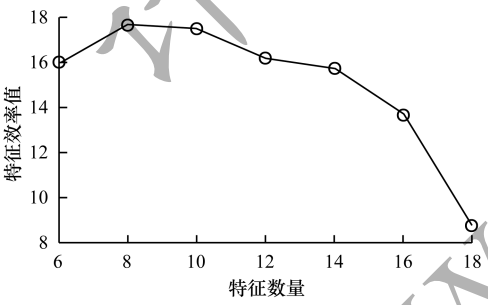

图 7 特征效率随特征数量的变化趋势

5.3.4 与现有研究的比较结果

在训练集和测试集相同的情况下,基于分类准确率和 F1-Score 两项评价指标,本文对比了极限学习机(ELM)[12]、决策树(DT)[12]、长短时记忆网络(LSTM)[14]、朴素贝叶斯(NB)[26-27]、支持向量机(SVM)[27]和本文提出的 IGKAD 方法,结果如表 6 所示,其中“—”表示没有该评价指标的实验结果。可见,在分类准确率上,本文 IGKAD 方法为 91.65%,仅次于文献[27]中基于 NB 方法 2 的 97.5%;在 F1-Score 上,本文方法为 94.27%,优于其他方法。

表 6 与现有方法的性能比较结果 %

方法	准确率	F1-Score
ELM 方法[12]	80.08	—
DT 方法[12]	78.80	—
LSTM 方法[14]	89.58	23.33
NB 方法 1[26]	74.10	52.80
SVM 方法[27]	79.10	80.10
NB 方法 2[27]	97.50	48.00
本文 IGKAD 方法	91.65	94.27

5.3.5 综合实验结果

根据特征选择情况并通过选取不同的核函数,可构造出 12 个 SVM 模型,并对这 12 个 SVM 模型进行实验测试。实验结果表明(如表 7 所示):1)采用 FMS 特征选择算法训练的模型要优于未进行特征选择的模型,证明了 FMS 特征选择算法在提高分类准确率的同时减少了模型训练时间;2)改进高斯核函数优于线性核函数、多项式核函数及 Sigmoid 核函数;3)采用 FMS 特征选择算法和改进高斯核函数的模型 SVM_5(即 IGKAD 方法)取得了最优的分类准确率(91.65%)和 F1-Score(94.27%)。

表 7 本文方法在不同参数下的综合性能比较结果

模型	特征	核函数	寻优参数		准确率/%	F1-Score/%	CPU 平均耗时/s	
			C	*gamma*			训练集	测试集
SVM_1	1~10	线性核函数	16.000 00	—	66.21	78.38	10.69	0.89
SVM_2	1~10	多项式核函数(d=2)	32.000 00	0.250 0	88.76	93.87	66.56	1.36
SVM_3	1~10	多项式核函数(d=3)	0.062 50	0.125 0	88.39	93.72	9.41	1.44
SVM_4	1~10	传统高斯核函数	16.000 00	1.000 0	86.21	90.34	8.59	1.63
SVM_5	1~10	改进高斯核函数	32.000 00	0.062 5	91.65	94.27	5.23	1.51
SVM_6	1~10	Sigmoid 核函数	0.031 25	0.062 5	77.76	86.88	6.41	2.14
SVM_7	1~37	线性核函数	0.031 25	—	83.50	90.34	19.78	3.06
SVM_8	1~37	多项式核函数(d=2)	32.000 00	0.250 0	84.53	91.09	75.41	2.75
SVM_9	1~37	多项式核函数(d=3)	0.062 50	0.250 0	79.18	87.39	13.94	2.47
SVM_{10}	1~37	传统高斯核函数	16.000 00	0.125 0	76.47	85.64	11.75	2.42
SVM_{11}	1~37	改进高斯核函数	8.000 00	0.062 5	79.99	87.98	9.24	2.39
SVM_{12}	1~37	Sigmoid 核函数	0.031 25	0.062 5	84.50	91.80	14.22	3.72

6　结束语

针对 BGP 更新报文激增情况的异常检测问题,本文提出一种基于改进高斯核函数的 BGP 异常检测方法。实验结果表明,当训练数据集包含 TOP10 特征时,该方法的分类准确率为 91.65%,F1-Score 为 94.27%,分类性能优于基于机器学习的 BGP 更新报文激增异常检测方法。下一步将结合时间序列属性,分析历史 BGP 报文对 BGP 异常检测的影响,进一步提高异常检测准确率。

参考文献

[1] REKHER Y,LI T,HARES S. Border gateway protocol4: RFC 4271[R]. Fremont,USA:IETF,2006:1-31.

[2] MURPHY S. BGP security vulnerabilities analysis: RFC 4272[R]. Fremont,USA:IETF,2006:1-21.

[3] Al-MUSAWI B, BRANCH P, ARMITAGE G. BGP anomaly detection techniques: a survey[J]. IEEE Communications Surveys and Tutorials, 2017, 19(1): 377-396.

[4] TOONK A. A BGP leak made in Canada[EB/OL]. (2012-08-08)[2018-05-03]. http://www.bgpmon.net/a-bgp-leak-made-in-canada.

[5] TOONK A. Large hijack affects reachability of high traffic destinations[EB/OL]. (2016-04-22)[2018-05-03]. http://bgpmon.net/large-hijack-affects-reachability-of-high-traffic-destinations/.

[6] THEODORIDIS G, TSIGKAS O, TZOVARAS D. A novel unsupervised method for securing BGP against routing hijacks[M]//GELENBE E,LENT R. Computer and Information Sciences III. Berlin,Germany:Springer, 2013:21-29.

[7] SHI Xingang, XIANG Yang, WANG Zhiliang, et al. Detecting prefix hijackings in the Internet with argus[C]// Proceedings of 2012 Internet Measurement Conference. New York,USA:ACM Press,2012:15-28.

[8] ZHANG Zheng,ZHANG Ying, HU Y C, et al. iSPY: detecting IP prefix hijacking on my own[J]. IEEE/ACM Transactions on Networking, 2010, 18(6): 1815-1828.

[9] CHENG Min,XU Qian,LÜ Jianming,et al. MS-LSTM:a multi-scale LSTM model for BGP anomaly detection[C]// Proceedings of the 24th International Conference on Network Protocols. Washington D. C.,USA:IEEE Press, 2016:1-6.

[10] 张蕾,崔勇,刘静,等. 机器学习在网络空间安全研究中的应用[J]. 计算机学报,2018,41(9):1943-1975.

[11] AL-ROUSAN N M, TRAJKOVIĆ L. Machine learning models for classification of BGP anomalies[C]//Proceedings of IEEE International Conference on High Performance Switching and Routing. Washington D. C., USA:IEEE Press,2013:103-108.

[12] LI Yan,XING Hongjie,HUA Qiang,et al. Classification of BGP anomalies using decision trees and fuzzy rough sets[C]//Proceedings of IEEE International Conference on Systems, Man and Cybernetics. Washington D. C., USA:IEEE Press,2014:1312-1317.

[13] BATTA P, SINGH M, LI Zhida, et al. Evaluation of support vector machine kernels for detecting network anomalies[C]//Proceedings of IEEE International Symposium on Circuits and Systems. Washington D. C., USA:IEEE Press,2018:1-7.

[14] DING Qingye, LI Zhida, BATTA P, et al. Detecting BGP anomalies using machine learning techniques[C]// Proceedings of IEEE International Conference on Systems, Man,and Cybernetics. Washington D. C.,USA:IEEE Press, 2017:3352-3355.

[15] CHENG Qiang, ZHOU Hongbo, CHENG Jie. The Fisher-Markov selector: fast selecting maximally separable feature subset for multiclass classification with applications to high-dimensional data[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2011,33(6):1217-1233.

[16] CORTES C, VAPNIK V. Support-vector networks[J]. Machine Learning,1995,20(3):273-297.

[17] KEERTHI S S, LIN C J. Asymptotic behaviors of support vector machines with Gaussian kernel[J]. Neural Computation,2003,15(7):1667-1689.

[18] LIN H T,LIN C J. A study on Sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods[J]. Neural Computation,2003,15(1):15-23.

[19] 王国胜. 核函数的性质及其构造方法[J]. 计算机科学,2006,33(6):172-174.

[20] MOORE D,PAXSON V,SAVAGE S,et al. Inside the slammer worm[J]. IEEE Security and Privacy,2003, 99(4):33-39.

[21] MACHIE A,ROCULAN J,RUSSELLU R,et al. Nimda worm analysis[EB/OL]. [2018-05-03]. http://dpnm.postech.ac.kr/research/04/nsri/papers/010919-Analysis-Nimda.pdf.

[22] MOORE D, SHANNON C. Code-Red: a case study on the spread and victims of an Internet worm[C]// Proceedings of the 2nd ACM SIGCOMM Workshop on Internet Measurement. New York, USA: ACM Press, 2002:273-284.

[23] LibBGPdump[EB/OL]. [2018-05-03]. http://bitbucket.org/ripenccwiki/.

[24] MANDERSON T. Multi-threaded routing toolkit Border Gateway Protocol (BGP) routing information export format with geo-location extensions: RFC 6397[R]. Fremont,USA:IETF,2018:1-8.

[25] CHANG C C,LIN C J. LIBSVM:a library for support vector machines[J]. ACM Transactions on Intelligent Systems and Technology,2011,2(3):1-27.

[26] AL-ROUSAN N M, TRAJKOVIĆ L. Feature selection for classification of BGP anomalies using Bayesian models[C]//Proceedings of International Conference on High Machine Learning and Cybernetics. Washington D. C., USA:IEEE Press,2013:103-108.

[27] AL-ROUSAN N M. Comparison of machine learnig models for classification of BGP anomalies[C]//Proceedings of the 13th International Conference on High Performance Switching and Routing. Washington D. C.,USA:IEEE Press,2012:1-6.

编辑　陆燕菲