doi:10.11918/j.issn.0367-6234.2015.05.018

ODA-IPNMF：一种在线全网络流量异常检测方法

柏 骏[1,2]，夏靖波[1]，吴吉祥[3]，鹿传国[4]

(1.空军工程大学 信息与导航学院,710077 西安;2. 95852 部队,572600 海南 东方;
3.空军大连通信士官学校,116600 辽宁 大连;4. 95806 部队,100076 北京)

摘 要：为实时、高效地检测网络流量异常,提出一种基于增量投影非负矩阵分解(IPNMF)的全网络流量异常检测方法(ODA-IPNMF).提出一种增量投影非负矩阵算法,该算法不仅具有与PCA相同的表达形式,还能以增量的方式构建正常子空间和异常子空间,进而利用Shewhart控制图实现全网络流量异常的在线检测.理论分析表明,该方法计算开销远小于NMF-NAD,具有更高的实用价值;模拟网络数据以及实测网络数据实验表明,基于NMF异常检测方法(NMF-NAD和ODA-IPNMF)的检测性能优于PCA方法;本文所提ODA-IPNMF与NMF-NAD网络异常检测效果相当,且可在线检测网络异常.

关键词：网络异常检测;流量矩阵;增量投影非负矩阵分解;在线检测

中图分类号：TP393 **文献标志码**：A **文章编号**：0367-6234(2015)05-0104-06

ODA-IPNMF: An Online Anomaly Detection Algorithm

BAI Jun[1,2], XIA Jingbo[1], WU Jixiang[3], LU Chuanguo[4]

(1. Institute of Information and Navigation, AFEU, 710077 Xi'an, Shanxi, China; 2. Unit 95852, 572600, Dongfang, Hainan, China; 3. Air Force Dalian Communications Noncommissioned Officers School, 116600, Dalian, Liaoning, China; 4. Unit 95806, 100076 Beijing, China)

Abstract: An online anomaly detection algorithm based on incremental projective non-negative matrix factorization is proposed to detect the network anomaly real-timely and efficiently. Firstly, an incremental projective non-negative matrix factorization is given, which has the same expression with PCA, and is able to construct normal and abnormal subspace to detect network-wide anomalies online by Shewhart control chart. Theoretic analysis indicates that, the proposed algorithm computation is far smaller than NMF-NAD. In addition, traffic matrix datasets analyzing for internet and simulation results show that the network anomalies detection algorithms based on NMF(such as NMF-NAD and ODA-IPNMF) performs better than that based on PCA, and the proposed ODA-IPNMF has comparable network anomaly detection by NMF-NAD, which the ability to detect the network anomaly online.

Keywords: network anomalies detection; traffic matrix; incremental projective non-negative matrix factorization; online detection

随着互联网技术的迅猛发展,各类网络异常活动如影随形,网络安全成为各方关注的焦点.异常检测对于维护网络安全、高效、可靠运行具有重要意义,是网络安全领域的研究热点[1-9].

从全网的视角实时检测网络异常对于维护网络稳定运行极为重要.文献[3-4]提出基于流量矩阵子空间映射的全网络异常检测方法,该方法利用流量矩阵的空间相关性和时间相关性,从全网络的视角检测异常行为. 文献[5]指出基于PCA异常检测方法对于主成分数量、检测阈值等参数的选取非常敏感,而且连续的或者足够大的网络异常可能毒害正常子空间,并致使误报率增加.文献[6]利用奇异值分解(SVD)实现PCA的在线更新,进而实现全网络的多元在线异常检测.

收稿日期：2014-06-19.
基金项目：国家自然科学基金(61272486);陕西省科技计划自然基金重点项目(2012JZ8005).
作者简介：柏 骏(1985—),男,博士研究生;
夏靖波(1963—),男,教授,博士生导师.
通信作者：柏 骏,peking1985-2005@163.com.

文献[7]先利用小波去除数据中的噪声，以减轻大的异常对于正常子空间的毒害，再使用 PCA 将流量矩阵映射到正常和异常子空间.文献[8]亦做了类似的工作，将 PCA 和小波分析结合起来以提高流量异常检测的精度.针对基于 PCA 的异常检测方法难以解释矩阵分解过程中出现的负值以及无法连续检测异常的问题，文献[9]提出利用非负矩阵分解检测网络异常，将原始流量矩阵与重构矩阵的差视为噪声和异常，并通过 Shewhart 控制图捕捉网络异常，但该文献并未对方法的合理性做出解释，并且只能离线检测网络异常.

本文采用与 PCA 表达形式相同的投影非负矩阵分解算法(PNMF)作为检测工具，提出了一种基于 IPNMF 的在线全网络流量异常检测方法(ODA-IPNMF).将 PNMF 引入到基于流量矩阵的全网络异常检测中，并提出了一种增量投影矩阵分解算法——IPNMF；提出一种基于 IPNMF 的全网络异常检测方法，该方法以增量的方式对流量矩阵进行投影，构建正常子空间和异常子空间，可实时检测网络异常.

1　流量矩阵及 PNMF 算法

1.1　流量矩阵

流量矩阵 TM(traffic matrix)是全网络流量概览，可全面、准确地描述网络流量特征(流量大小、字节数、IP 地址、端口等)的分布情况[10].矩阵中的元素代表网络中某一时刻(时间间隔)源节点与目的节点之间(OD 对)的流量，可表示为一个 $d \times p$ 的矩阵 $\boldsymbol{X}$.其中，d 为测量的周期数；p 为 n 个网络节点的 OD 对数，有 $p = n \times n$(或 $p = n \times (n-1)$).$\boldsymbol{X}_{ij}$ 为在第 i 个测量周期时第 j 个 OD 对的流量特征数据的测量值.根据选择的网络节点类型不同，可以分为链路级、路由级和 PoP(point of presence)级流量矩阵.本文以 PoP 级流量矩阵作为研究对象.

1.2　PNMF 算法

为应对高维数据“维度诅咒”问题，多采用矩阵分解来分析实际问题中的高维数据，如 PCA、ICA、SVD、VQ 等.然而上述方法分解的矩阵元素可能出现负数往往无法得到合理的解释，因此，文献[11]的非负矩阵分解应运而生.

对任意给定的非负矩阵 $\boldsymbol{X} \in R_+^{d\times p}$ 进行分解，试图寻找最优解 W 和 H 使得

$$\boldsymbol{X} \approx \boldsymbol{WH}. \tag{1}$$

并通过迭代使得二者之间的近似误差最小，用欧式距离衡量，有

$$D_{EU}(\boldsymbol{X} \parallel \boldsymbol{WH}) = \sum_{ij} [\boldsymbol{X}_{ij} - (\boldsymbol{WW}^{\mathrm{T}}\boldsymbol{X})_{ij}]^2. \tag{2}$$

式中：$\boldsymbol{W} \in \mathbf{R}_+^{r\times p}$ 称为基矩阵，$\boldsymbol{H} \in \mathbf{R}_+^{d\times r}$ 称为系数矩阵，且 $r \ll n$.

PNMF[12]是一个改进的 NMF 算法，它试图寻找一个非负的投影矩阵使得

$$\boldsymbol{X} \approx \hat{\boldsymbol{X}} = \boldsymbol{PX}. \tag{3}$$

式中的投影矩阵可进一步表示为

$$\boldsymbol{P} = \boldsymbol{WW}^{\mathrm{T}}. \tag{4}$$

式中 $\boldsymbol{W} \in R_+^{r\times p}$，相比于式(1)，系数矩阵为 $\boldsymbol{H} \approx \boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}$.

PNMF 算法为求得最佳 $\boldsymbol{W}$，定义了如下优化问题：

$$\min F(\boldsymbol{W}) = \frac{1}{2} \parallel \boldsymbol{X} - \boldsymbol{WW}^{\mathrm{T}}\boldsymbol{X} \parallel^2. \tag{5}$$

对 $\boldsymbol{X}$ 求偏导，有

$$\frac{\partial F(\boldsymbol{W})}{\partial W_{ik}} = -2(\boldsymbol{XX}^{\mathrm{T}}\boldsymbol{W})_{ik} + (\boldsymbol{WW}^{\mathrm{T}}\boldsymbol{XX}^{\mathrm{T}}\boldsymbol{W})_{ik} + (\boldsymbol{XX}^{\mathrm{T}}\boldsymbol{WW}^{\mathrm{T}}\boldsymbol{W})_{ik}. \tag{6}$$

令

$$\eta_{ik} = \frac{W_{ik}}{(\boldsymbol{WW}^{\mathrm{T}}\boldsymbol{XX}^{\mathrm{T}}\boldsymbol{W})_{ik} + (\boldsymbol{XX}^{\mathrm{T}}\boldsymbol{WW}^{\mathrm{T}}\boldsymbol{W})_{ik}}. \tag{7}$$

带入上式，得到

$$W'_{ik} = W_{ik} - \eta_{ik}\frac{\partial F(\boldsymbol{W})}{\partial W_{ik}}. \tag{8}$$

进而得到更新方程：

$$W'_{ik} = W_{ik}\frac{2(\boldsymbol{XX}^{\mathrm{T}}\boldsymbol{W})_{ik}}{(\boldsymbol{WW}^{\mathrm{T}}\boldsymbol{XX}^{\mathrm{T}}\boldsymbol{W})_{ik} + (\boldsymbol{XX}^{\mathrm{T}}\boldsymbol{WW}^{\mathrm{T}}\boldsymbol{W})_{ik}}, \tag{9}$$

$$\boldsymbol{W} \leftarrow \boldsymbol{W} / \parallel \boldsymbol{W} \parallel. \tag{10}$$

较其他 NMF 算法，PNMF 更适用于流量矩阵的网络异常在线检测.首先，PNMF 与与基于 PCA 的异常检测算法[3]具有相同的表达形式，可将流量矩阵向正常子空间和异常子空间上进行投影；其次，通过 PNMF 不仅可得到非负的稀疏矩阵，而且只需考虑更少的参数，就能产生更稀疏的矩阵，便于在线流量异常检测；而且，PNMF 在不断迭代的过程中，可将正常与异常子空间的差异最大化，具备更佳的检测效果.

2　ODA-IPNMF 算法

采用子空间分析方法[3]检测流量异常的前提是分离正常子空间和异常子空间，并对异常子空间进行分析以检测网络是否出现异常.上述 PNMF 算法只能对流量矩阵进行离线的批量处理，无法以增量的方式分离正常、异常子空间.因此，为克服 PNMF 只能进行离线数据处理的不足，提出一种增

量投影非负矩阵分解算法(IPNMF).

增量计算是一个不断利用前面处理结果进行后续增量运算的过程.那么增量过程的关键就是如何利用已得到的计算结果.文献[13]指出,增量非负矩阵分解过程中,基向量代表了之前数据的所有信息.基于此,将新增样本数据 $\boldsymbol{D}(t)$ 添加到上一时刻的基向量 $\boldsymbol{W}(t-1)$ 中作为 IPNMF 算法的新的输入向量 $\boldsymbol{X}(t)$,记为 $\boldsymbol{X}(t)=[\boldsymbol{W}(t-1),\boldsymbol{D}(t)]$;然后将 $\boldsymbol{X}(t)$ 代入式(9)和式(10)中,不断迭代直至收敛.在迭代过程中,主要的计算是式(9)中的矩阵乘法,为减少运算量,令

$$\boldsymbol{Q}=\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}. \tag{11}$$

下面给出 IPNMF 算法的伪代码:

输入: $\boldsymbol{X}(0)$; $\boldsymbol{D}(t)$, $t=0,1,2,3\cdots$
输出: $\boldsymbol{W}(t)$
初始化: $\boldsymbol{W}(0)=\mathrm{random}(p,r)$;
$\boldsymbol{W}(0)=$ PNMF($\boldsymbol{X}(0),\boldsymbol{W}(0)$);% 调用PNMF求解 $\boldsymbol{W}$
$\boldsymbol{Q}(0)=\boldsymbol{X}(0)\boldsymbol{X}(0)^{\mathrm{T}}\boldsymbol{W}(0)$;
while($\boldsymbol{D}(t)$ 不为空)
  $\boldsymbol{X}(t)=[\boldsymbol{W}(t-1),\boldsymbol{D}(t)]$;
  $\boldsymbol{Q}(t)=\boldsymbol{X}(t)\boldsymbol{X}(t)^{\mathrm{T}}\boldsymbol{W}(t)+\boldsymbol{Q}(t-1)$;
  while(不满足收敛条件):
    $\boldsymbol{W}_{ik}\leftarrow\boldsymbol{W}_{ik}\dfrac{2\boldsymbol{Q}_{ik}}{(\boldsymbol{W}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{Q})_{ik}+(\boldsymbol{Q}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W})_{ik}}$;
    $\boldsymbol{W}(t)=\boldsymbol{W}(t)/\|\boldsymbol{W}(t)\|$;
  end while
  $t++$;
end while

基于以上分析,本文采用子空间分析方法,利用 IPNMF 将流量矩阵映射到正常子空间和异常子空间中,然后利用 Shewhart 控制图捕捉异常子空间中的流量异常.该方法步骤如下:

1)流量矩阵重构.对于流量矩阵 $\boldsymbol{X}$,利用 IPNMF 算法得到基矩阵 $\boldsymbol{W}$,进而重构流量矩阵 $\hat{\boldsymbol{X}}=\boldsymbol{W}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}$,即为流量矩阵 $\boldsymbol{X}$ 经 $\boldsymbol{W}\boldsymbol{W}^{\mathrm{T}}$ 投影的正常子空间;

2)残余矩阵获取.残余矩阵亦可称为异常子空间,有 $\tilde{\boldsymbol{X}}=\boldsymbol{X}-\hat{\boldsymbol{X}}=(\boldsymbol{I}-\boldsymbol{W}\boldsymbol{W}^{\mathrm{T}})\boldsymbol{X}$;

3)异常流量捕获.为了对残余矩阵进行分析,本文引入残余流量向量中所有元素的平方和(SSE)来衡量残余向量的大小,并利用 Shewhart 控制图来捕捉异常子空间中的流量异常.

利用上述的 IPNMF 增量算法思想实现全网络流量异常的在线检测——ODA-IPNMF,算法描述如下:

输入: $\boldsymbol{X}(0)$; $\boldsymbol{D}(t)$, $t=0,1,2,3\cdots$
输出: NA(网络异常)
初始化: $\boldsymbol{W}(0)=\mathrm{random}(p,r)$;
$\boldsymbol{X}(1)=\boldsymbol{D}(0)$;
$\boldsymbol{W}(0)=\mathrm{PNMF}(\boldsymbol{X}(0),\boldsymbol{W}(0))$;
$\boldsymbol{Q}(0)=\boldsymbol{X}(0)\boldsymbol{X}(0)^{\mathrm{T}}\boldsymbol{W}(0)$;
while($\boldsymbol{D}(t)$ 不为空)
  $\boldsymbol{X}(t)=[\boldsymbol{W}(t-1),\boldsymbol{D}(t)]$;
  $\boldsymbol{Q}(t)=\boldsymbol{X}(t)\boldsymbol{X}(t)^{\mathrm{T}}\boldsymbol{W}(t)+\boldsymbol{Q}(t-1)$;
  while(不满足收敛条件):
    $\boldsymbol{W}_{ik}\leftarrow\boldsymbol{W}_{ik}\dfrac{2\boldsymbol{Q}_{ik}}{(\boldsymbol{W}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{Q})_{ik}+(\boldsymbol{Q}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W})_{ik}}$;
    $\boldsymbol{W}(t)=\boldsymbol{W}(t)/\|\boldsymbol{W}(t)\|$;
  end while
  $\tilde{\boldsymbol{X}}=\boldsymbol{X}-\hat{\boldsymbol{X}}=(\boldsymbol{I}-\boldsymbol{W}\boldsymbol{W}^{\mathrm{T}})\boldsymbol{X}$;
  [residual, UCL, LCL] = Shewhart($\tilde{\boldsymbol{X}}$);
  if ($R(i)>UCL$ or $R(i)<LCL$)
    i 时刻出现异常;
    添加 i 到 NA;
  end if
  $t++$;
end while

ODA-IPNMF 方法的计算开销主要在于增量数据 $\boldsymbol{D}(t)$ 的 IPNMF 分解和基于 Shewhart 的异常检测.增量数据 $\boldsymbol{D}(t)$ 的 IPNMF 的时间复杂性为 $O(pr(r+s)k)$,其中 p 为 OD 对数,r 是基矩阵维数,s 为增量数据维数,k 为迭代次数.而基于 Shewhart 的异常检测的时间复杂性仅与增量数据维数 s 相关,为 $O(s)$,远小于前者.因此,ODA-IPNMF 方法的时间复杂性为 $O(pr(r+s)k)$,相比于文献[9]的 $O(pdrk)$,其中 $(r+s)\ll d$,那么所提方法的运算效率有明显提高.

3 实验分析及评价

为对所提 ODA-IPNMF 算法性能进行分析、评价,本文基于 Matlab 实验平台,以检测率和误检率作为评价指标,采用模拟实验分析与因特网实测数据分析相结合的方式来比较异常检测算法的检测性能.

3.1 模拟数据实验及其分析

3.1.1 模拟数据生成

首先产生近似周期性的正常成分、高斯噪声成分和异常成分[14],再按适当的比例人工合成网络流量.具体步骤:1)通过将 3 种不同周期、随机初始相位的正弦波叠加来模拟 OD 流,进而构成基准流量矩阵,图 1(a);2)在 OD 流上加入零均

值的高斯噪声，得到不含异常的基准流量矩阵，见图 1(b)；3) 以一定规则注入各类典型异常，见图 1(c).采用间隔为 5 min，共模拟一个星期的网络流量，最终生成包含 2 016 个测量周期、121 条网络流的流量矩阵 $\boldsymbol{X}$.

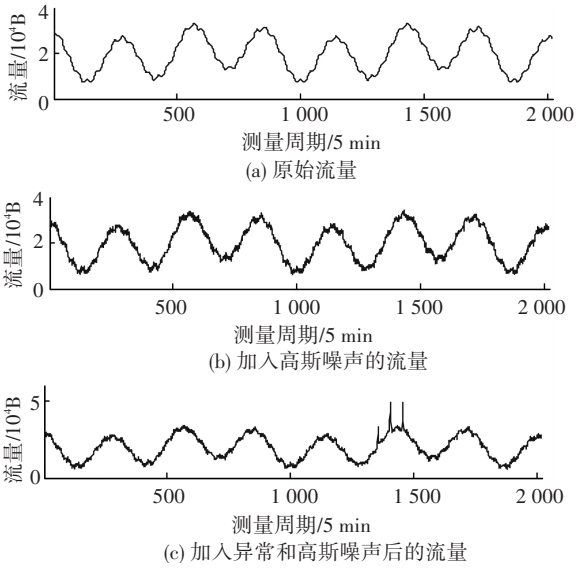

图 1　模拟网络流量数据合成

本文所提方法是从流量大小的角度去检测异常，故在此考虑 4 种最常见的、引起流量剧烈变化的典型异常：阿尔法(alpha)异常、(分布式)拒绝服务攻击(DoS, DDoS)、突发流(flash crowd)、入口/出口移动(ingress/egress shift)异常，并利用持续时间、流量变化大小、源-目的数以及形状函数等 4 个参数来描述.上述异常的具体行为特征参见文献[6].

网络异常注入过程如下：从第 250 个测量周期到第 500 个测量周期，每 50 个测量周期注入一次持续 30 min(6 个测量周期)阿尔法攻击；从第 750 个测量周期到第 1 000 个测量周期，每 50 个测量周期注入一次持续时间为 30 min(6 个测量周期)DDoS 攻击；从第 1 250 测量周期到 1 500 测量周期，每 50 个测量周期注入一次突发流；在第 1 900 测量周期到 1 950 测量周期内注入一次入口/出口移动异常.

3.1.2　IPNMF 算法性能分析及参数选取

首先对 IPNMF 算法的性能及参数选取进行研究.非负矩阵分解算法最重要的两个性能指标是收敛性(convergence)和矩阵还原性(recovery)，前者与迭代次数和收敛时间有关，后者可用欧式来衡量，有距离 $d=\sqrt{\dfrac{\sum_{ij}\left[\boldsymbol{X}_{ij}-(\boldsymbol{W}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X})_{ij}\right]^2}{mn}}$(其中 mn 为矩阵元素个数). 在以下实验中，IPNMF 算法或 ODA-IPNMF 方法中 $\boldsymbol{D}(t)$ 的维数为 6，即每 6 个测量周期(半个小时)更新一次数据.

为比较本文所提 IPNMF 算法与 PNMF 算法的收敛性以及矩阵还原性，在固定基矩阵维数 $r=3$ 的条件下，利用上述两种算法对模拟实验数据重复 10 次，得到的实验结果见表 1.

表 1　IPNMF 算法与 PNMF 算法性能比较结果

算法	迭代次数	收敛时间/s	d
PNMF	176.10	72.55	596.28
IPNMF	43.11	42.14	622.43

从表 1 中看出 IPNMF 算法对收敛性的明显改善，矩阵分解过程中的迭代次数以及收敛时间都大幅减少.一方面原因是 IPNMF 算法的时间复杂性低于 PNMF 算法；另一方面则是 IPNMF 算法所需存储的最大矩阵规模为 $m\times(r+s)$，远小于后者的 $m\times(n+s)$，节省了大量存储资源.此外，两种投影非负矩阵分解算法的 d 值相差无几，因此二者具有相当的矩阵还原性.

非负矩阵分解算法的一个非常重要的参数是基矩阵维数 r.为确定基矩阵维数，本文设置最大迭代次数为 500，另 r 从 2 递增至 20，以寻求最优的矩阵还原性. 实验重复 10 次，实验结果见图 2.

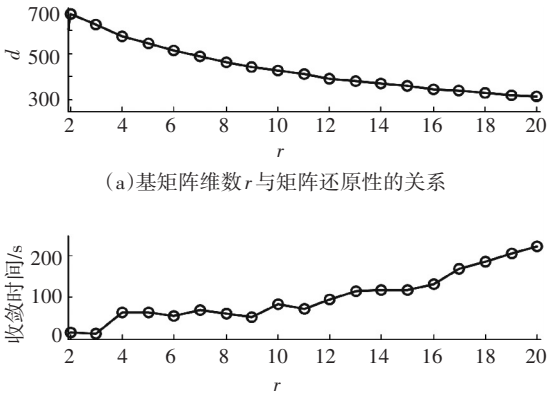

图 2　基矩阵维数 r 与算法性能的关系

3.1.3　ODA-IPNMF 算法性能分析与比较

将本文所提 ODA-IPNMF 与 MOADA-SVD 方法[6]和 NMF-NAD 方法[9]进行比较.MOADA-SVD 是基于奇异值分解的在线网络异常检测方法，采用 Q 统计量的方法检测异常；NMF-NAD 是基于 NMF 的离线网络异常检测方法，采用 Shewhart 控制图的方法进行异常检测.上述 3 种算法异常时刻凸显的结果见图 3.

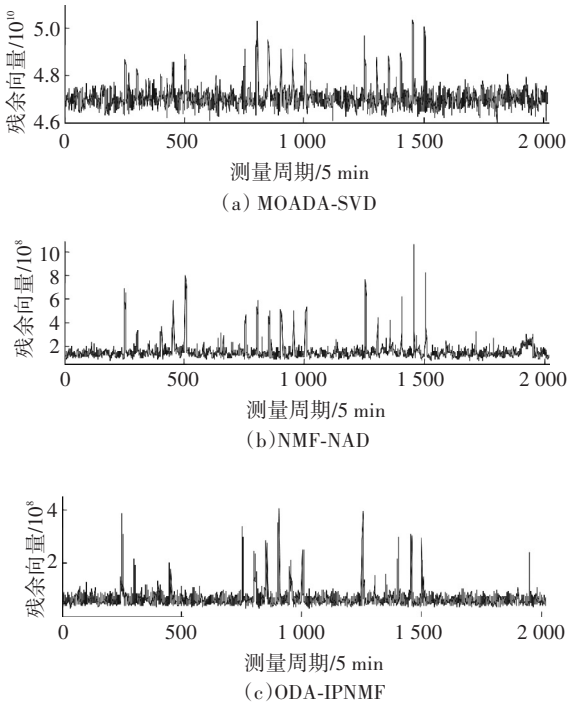

图 3　3 种方法的异常凸显对比

由图 3 可发现,相较于 MOADA-SVD 方法,基于 NMF 的异常检测方法(NMF-NAD 和 ODA-IPNMF)对网络异常更为敏感,其异常/正常流量的幅值比明显大于前者,更易被检测,且误检率更低.究其原因,MOADA-SVD 仍是基于 PCA 算法,它力求寻找均方误差最小意义下的最优线性映射投影,却忽略了正常流量与异常流量的属性差异,导致可能包含的重要信息丢失;而 NMF 在不断迭代的过程中将这种差异最大化,因而具有更佳的异常检测效果.

在图 4 中,利用 Shewhart 对 ODA-IPNMF 与 NMF-NAD 的检测结果进行比较,其中 R 为流量变化范围,在[LCL, UCL]区间内即为正常.结合图 3(b)和图 3(c),可以看出,上述两种方法对全网络异常的凸显几乎是一样的,唯一不同在于第 1 900 到 1 950 测量周期的入口/出口移动异常检测结果,NMF-NAD 方法表现出连续异常,而本文算法表现为突发异常.究其原因,本文方法采用增量分解算法,以 6 个测量周期进行矩阵分解,残余矩阵中的连续流量信息容易被忽略,导致连续异常未能被充分检测.此外,ODA-IPNMF 算法以增量的方式实时检测网络异常,每次需处理的数据仅与时间间隔 t 有关,数据量远远小于一次性分解全流量矩阵,所占用的存储空间以及计算资源要远远小于 NMF-NAD 算法,因而实际应用价值更高.

重复上述实验过程 10 次,3 种算法的检测结果见表 2.通过比较发现,NMF-NAD 和 ODA-IPNMF 方法无论是检测率还是误检率都相差无几,且明显优于 MOADA-SVD 方法,而后者与文献[3]的结果吻合.实验表明,NMF 具备更优异的异常检测能力.

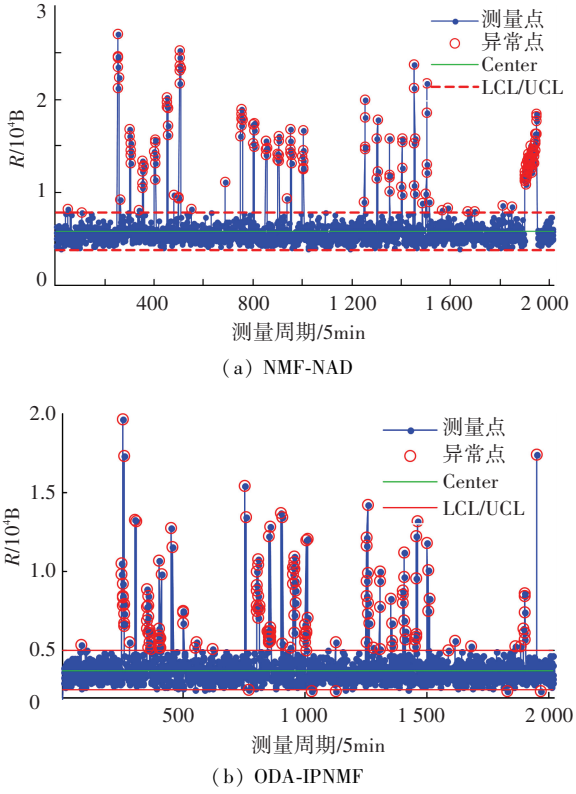

图 4　基于 Shewchart 的异常检测结果

表 2　异常检测结果　%

方法	检测率	误检率
MOADA-SVD	83.47	26.77
NMF-NAD	97.74	5.93
ODA-IPNMF	98.25	6.21

3.2　实测数据实验及其分析

3.2.1　实测数据集 Abilene

为验证、评价上述方法在实际网络中异常检测的效果及性能,本文采用来自于 Abilene 网络的流量矩阵数据,其描述如表 3 所示.上述数据集自身便含有网络异常,发现并确定异常是检测的关键.

表 3　Abilene 流量矩阵

数据集	时间段	间隔/min	测度	矩阵形式
数据集 1	2003.12.08~12.14	5	字节数	2 016×121

3.2.2　ODA-IPNMF 算法实验分析

按照上述实验流程对 Abilene 流量矩阵数据进行异常检测.其中 Dataset1 的 3 种方法的异常凸显结果见图 5.

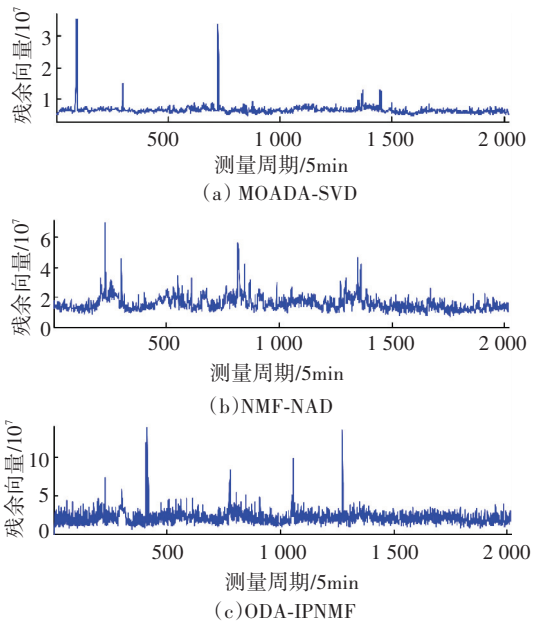

图 5　Dataset1 数据集异常凸显结果

实测网络数据的异常检测结果很好地验证了模拟数据的实验结论.MOADA-SVD 方法仅发现了 3 处网络异常,而其他两种方法则发现了包括上述异常在内的 5 至 6 处异常,表明基于 NMF 的异常检测方法(NMF-NAD 和 ODA-IPNMF)在异常检测性能上优于基于 PCA 的检测方法；本文提出的基于增量 NMF 的在线异常检测算法 ODA-IPNMF 与 NMF-NAD 相比,在异常检测性能上是相当的,但对于不同的异常表现出一些差异.

4　结　语

实验与分析表明基于 NMF 的异常子空间投影方法较基于 PCA 的方法更利于网络流量矩阵异常检测.然而现有基于 NMF 的异常检测方法只能处理离线数据,无法对网络流量进行在线检测.针对这一不足,本文给出了一种增量非负矩阵分解算法,并将其应用于网络异常检测中,并利用 Shewart 控制图分析残余流量,提出了一种基于 IPNMF 的在线全网络流量异常检测方法.该方法以增量的方式实时检测网络异常,所占用的存储空间以及计算资源要远远小于 NMF-NAD 算法.通过模拟数据以及实测网络数据实验表明本文所提方法具备良好的全网络异常流量检测效果.

参考文献

[1] BHUYAN M H, BHATTACHARYYA D K, KALITA J K. Network anomaly detection: methods, systems and tools[J]. Communications Surveys & Tutorials, IEEE, 2014, 16(1): 303-336.

[2] HUANG S Y, HUANG Y N. Network traffic anomaly detection based on growing hierarchical SOM[C]//2013 43rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN). Budapest: [s.n.], 2013: 1-2.

[3] LAKHINA A, CROVELLA M, DIOT C. Diagnosing network-wide traffic anomalies [C]//ACM SIGCOMM Computer Communication Review, [S.l.]: ACM, 2004, 34(4): 219-230.

[4] LAKHINA A, CROVELLA M, DIOT C. Mining anomalies using traffic feature distributions[C]//ACM SIGCOMM Computer Communication Review, [S.l.]: ACM, 2005, 35(4): 217-228.

[5] RINGBERG H, SOULE A, REXFORD J, et al. Sensitivity of PCA for traffic anomaly detection[C]//ACM SIGMETRICS Performance Evaluation Review, [S.l.]: ACM, 2007, 35(1): 109-120.

[6] 钱叶魁, 陈 鸣. 基于奇异值分解更新的多元在线异常检测方法[J]. 电子与信息学报, 2010, 32(10): 2404-2409.

[7] 钱叶魁, 陈 鸣, 叶立新, 等. 基于多尺度主成分分析的全网络异常检测方法[J]. 软件学报, 2012, 23(2): 361-377.

[8] NOVAKOV S, LUNG C H, LAMBADARIS I, et al. Studies in applying PCA and wavelet algorithms for network traffic anomaly detection [C]//High Performance Switching and Routing (HPSR), 2013 IEEE 14th International Conference on, [S.l.]: IEEE, 2013: 185-190.

[9] 魏祥麟, 陈鸣, 张国敏. NMF-NAD: 基于 NMF 的全网络流量异常检测方法[J]. 通信学报, 2012, 33(4): 54-61.

[10] VARDI Y M. Network tomography: estimating source-destination traffic intensities from link data[J]. Journal of the American Statistical Association, 1996, 91(433): 365-377.

[11] LEE D D, SEUNG H S. Learning the parts of objects by non-negative matrix factorization[J]. Nature, 1999, 401: 788-791.

[12] YANG Z, OJA E. Linear and nonlinear projective nonnegative matrix factorization[J]. IEEE Transaction on Neural Networks, 2010, 21(5): 734-749.

[13] DONOHO D, STODEEN V. When does non-negative matrix factorization give a correct decomposition into parts? [C]//Advances in Neural Information Processing Systems, 2004: 1141-1148.

[14] LAKHINA A, PAPAGIANNAKI K, CROVELLA M, et al. Structural analysis of network traffic flows[C]//SIGMETRICS, New York, NY, USA: [s. n.], 2004: 156-167.

(编辑　苗秀芝)