计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.190600173

一种基于时序性告警的新型聚类算法

邓甜甜[1,2]　熊荫乔[1,2]　何贤浩[2]
1 长沙学院电子信息与电气工程学院　长沙 410022
2 国防科技大学计算机学院　长沙 410073
(dtt@ccsu.edu.cn)

摘　要　云环境下,大规模集群设备将产生海量时序性的告警数据,实际应用中,运维人员通常利用这些告警数据来定位、排查、修复故障和错误,维持系统的正常运行。因此,如何将海量告警数据进行有效聚类,并挖掘告警中的关键信息,必将成为"云"能否持续稳定运行的核心问题。据此,文中提出了一种基于时序性告警的新型聚类算法。算法利用设定时间窗口内两两告警之间时间差的关系,构造告警之间新的关系矩阵,再利用 K-means 算法对关系矩阵中的列向量进行聚类,得到告警的聚类结果。实验结果表明,该算法能充分地将海量告警信息有效聚类。

关键词:告警;时序特征;数据挖掘;聚类
中图法分类号　TP274

Novel Clustering Algorithm Based on Timing-featured Alarms

DENG Tian-tian[1,2],XIONG Yin-qiao[1,2] and HE Xian-hao[2]
1 College of Electronic and Communication Engineering,Changsha University,Changsha 410022,China
2 College of Computer,National University of Defense and Technology,Changsha 410073,China

Abstract　In the cloud environment,large-scale cluster equipments will generate massive timing-featured alarms. In the practical application,operational personnel generally uses these alarms to locate,check and repair the faults and errors,and maintains the normal operation of the systems. So how to efficiently cluster the alarms and mine the key information will be core issues to keep continuous and stable operation of the cloud. Therefore,this paper proposes a novel clustering algorithm based on timing featured alarms. The algorithm constructs a new relation matrix by utilizing time difference between any two alarms in the given time window,then takes advantage of K-means algorithm to cluster the column vectors in the relation matrix,to get the cluster result of alarms. Experiment result shows that the algorithm can cluster massive alarms efficiently.

Keywords　Alarms,Timing feature,Data mining,Cluster

1　引言

云计算的快速发展,使得云环境中的设备越来越多,设备之间的拓扑结构越来越复杂,对设备的故障诊断和排查工作的难度也随之提高。一个告警的发生往往导致一系列告警的出现和不断迭代,进而导致系统出现海量告警,且不同原因产生的告警交织在一起,这些告警表面看来杂乱无章,使得维护人员在分析某一告警时,会受到其他大量与之无关的告警信息影响,从而无法准确地进行故障的定位、排查与修复。文献[1]指出,大型网站在恢复故障过程中,故障的定位和排查大约占据总时间的93%。由此可见,维护人员必须对网络设备产生的告警信息进行分析并对告警进行有效的聚类,才能实现高效的运维服务。云规模的扩大以及用户需求的增加,致使告警排查工作需要满足时效性,处理得不及时与不恰当会直接影响用户体验并给企业带来无法估量的损失。如何快速、准确地对告警信息进行聚类是工业界与学术界一直关注的问题[2]。

数据挖掘技术通常被称为数据的知识发现(KDD)。在目前的商业领域,数据挖掘技术都有应用场景,最广为人知的就是"啤酒与尿布"的案例。沃尔玛超市对顾客购物的交易数据进行分析,最终得出了"啤酒与尿布一起购买的可能性最大"这一结论。这个案例表明,日趋增加的数据迫使人们需要找到有效的方法来分析这些数据并找到有用的价值和规律,用以解决问题。文献[3]指出,数据挖掘算法在模式关联挖掘、分类、预测和聚类等方面有突出的表现,而设备故障的诊断和排查需要对大量的告警信息进行分析,因此数据挖掘技术正是解决这类问题的强有力工具。

目前,已经有许多学者采用不同的方式对告警数据进行挖掘。文献[4]提出了一种基于关联规则挖掘的 Apriori 算法,该算法在给定置信度的条件下,找到满足支持度限制下的所有关联规则。文献[5]提出了一种频繁模式挖掘,基于 FP-Tree 的数据结构使得数据能够压缩成树形结构并减少访问数据库的次数。文献[6]采用基于 WINEPI 算法的序列模式挖掘,该算法利用滑动窗口挖掘告警模式,计算频繁情节,最

基金项目:国家自然科学基金(61972058);湖南省自然科学基金(2020JJ5621);长沙市科技计划项目(ZD1601042,K1705031)
This work was supported by the National Natural Science Foundation of China(61972058),Natural Science Foundation of Hunan Province(2020JJ5621) and Science and Technology Planning Project of Changsha (ZD1601042,K1705031).
通信作者:熊荫乔(yq.xiong@ccsu.edu.cn)

后从情节中找到关联规则。文献[7]抽象语义关联关系，聚合相似的告警，进而判断多个告警是否由同一原因产生。为从告警泛洪中识别根源性问告警信息，刘冬生等[8]提出了一种结合元胞学习自动机(CLA)和决策树 ID3 的新告警关联聚类算法，但是该算法没有考虑时间窗问题。陈兴蜀等[9]提出了一种基于告警属性聚类的攻击场景关联规则挖掘算法，该算法首先基于误告警周期特性生成误告警过滤规则，然后基于告警属性相似性对告警信息进行聚类，最后利用 Apriori 频繁项挖掘算法生成攻击场景序列模式。通过真实网络环境实验，该算法能有效解决攻击链断裂问题，缓解误告警带来的影响。为重现多步攻击场景，樊迪等[10]提出了一种基于因果发现的攻击场景重现算法。该算法首先通过告警日志的 IP 属性构建序列集合，利用概率统计方法发现告警信息间的关联关系。在 DARPA2000 数据集上的验证表明，该方法能较好地识别多步攻击模式。针对告警关联分析中因果知识难以自动获取的问题，冯学伟等[11]利用马尔可夫链挖掘不同攻击类型的一步转移概率矩阵，对具有重复步骤的因果知识进行匹配融合。但这种方法仅考虑了告警信息中的 IP 属性，未考虑其他属性的影响。Ramaki 等[12-13]利用贝叶斯网络构建贝叶斯攻击图(BAG)，在此基础上生成关联规则，以预测攻击者的后续步骤。

大多数文献都是采用基于关联规则、频繁模式的挖掘算法，这些挖掘算法依赖于高支持度、高置信度条件[8]，并未考虑到在时间窗口内，各类告警发生的时间上存在新的关系。而文献[7]的方法需要预先定义好语义关系库，不具备时效性[9]。针对以上问题，本文提出了一种新型的聚类挖掘算法，该算法在设定时间窗口内，找出各个告警发生时间的时间差关系，并利用这种关系构造任一告警与其他告警之间新的关系向量，找出向量之间的相似度进行有效聚类，简化了运维过程的复杂性，并缩小了真实故障的排查范围。

2 相关问题描述

定义 1(告警类型) 告警类型表示不同异常导致设备产生不同告警信息的唯一标识。

定义 2(告警数据) 告警数据 D 表示为$\langle alarm,site,T\rangle$的三元组，$alarm$ 表示告警类型，$site$ 表示发告警的站点，T 表示发生告警的时间。以图 1 为例，时间轴 t 上有 4 个告警信息：

1)$\langle A,site_1,t_1\rangle$表示在时间 t_1，站点 $site_1$ 发生告警 A；

2)$\langle B,site_2,t_2\rangle$表示在时间 t_2，站点发生告警 B；

3)$\langle A,site_1,t_3\rangle$表示在时间 t_3，$site_1$ 站点再次发生告警 A；

4)$\langle C,site_3,t_4\rangle$表示在时间 t_4，$site_3$ 站点发生告警 C。

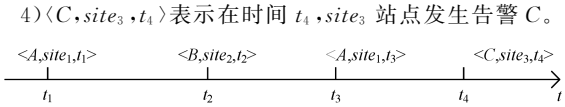

图 1 告警数据示例

Fig. 1 Example for alarms

定义 3(告警聚类) 告警聚类 $\Omega=\{\Omega_1,\Omega_2,\cdots,\Omega_k\}$表示警数据分为 k 簇，$\Omega_i=\{alarm_i,\cdots,alarm_k\}$表示告警类型$\{alarm_i,\cdots,alarm_k\}$属于第 i 簇。

定义 4(时间窗口) 告警数据$\langle alarm,site,T\rangle$的时间窗口 $W=\langle T_{\text{win}},\langle alarm,site,T\rangle\rangle$表示以告警 $alarm$ 发生的时间 T 为中心，T_{win}为时间跨度的时间区间。其中 n 表示告警数据的大小，T_{start}表示告警开始的时间，T_{end}表示告警结束的时间，α 为时间窗口的系数。

$$T_{\text{win}}=\alpha\cdot(T_{\text{end}}-T_{\text{start}})/n \tag{1}$$

定义 5(告警 i 与告警 j 的时间差绝对值之和) T_{start}至 T_{end}时间内，针对某一个站点，计算所有告警 i 与以其为中心的时间窗口内告警 j 时间差的绝对值之和。以图 2 为例，告警$\langle j,site,t_1\rangle$、$\langle j,site,t_3\rangle$和$\langle j,site,t_4\rangle$与告警$\langle i,site,t_2\rangle$属于同一个站点，但只有告警$\langle j,site,t_1\rangle$、$\langle j,site,t_3\rangle$位于以告警$\langle i,site,t_2\rangle$为中心的时间窗口内，所以：

$$\Delta_{ij}=|t_2-t_1|+|t_2-t_3| \tag{2}$$

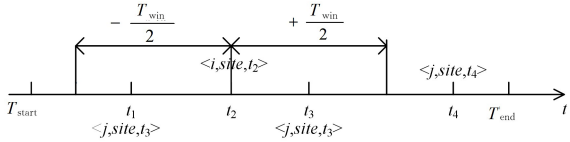

图 2 计算 Δ_{ij} 示例图

Fig. 2 Example for calculating Δ_{ij}

3 时序告警聚类挖掘算法

本算法流程分为 5 步：1)数据采集和存储；2)对不同站点进行分类；3)数据预处理；4)建立聚类分析模型；5)得到告警聚类结果。其基本框架如图 3 所示。

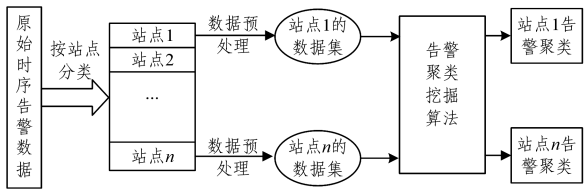

图 3 算法基本框架图

Fig. 3 Framework of our algorithm

3.1 数据采集和存储

云服务提供商网络中的设备，如基站、核心网服务器、传输网线路等，都通过设备厂商的 EMS 设备管理。EMS 探测到设备或线路故障后，会产生相应告警。所有 EMS 设备通过 Syslog 或者 SNMP 方式发送告警到 FM(Fault Management)平台探针，FM 探针统一采集告警，并解析告警内容，对告警名、告警发生时间、告警发生地点、告警摘要等信息进行标准化处理。之后，探针将告警送到集中的告警数据库存储。

3.2 对不同的站点进行分类

以每个站点的名称 $site$ 作为 key，获取不同站点内所有告警数据集$\langle D_1,D_2,\cdots,D_n\rangle$，每一个站点所有的告警数据集可作为该站点的 value，形成 key-value 的映射关系，即：

$$H_{site}=\langle D_1,D_2,\cdots,D_n\rangle \tag{3}$$

采用 key-value 的映射关系，在对不同站点内的告警数据进行聚类时，可以直接通过站点的名称取得该站点对应的告警数据集，避免多次访问数据库。其流程如 4 图所示。

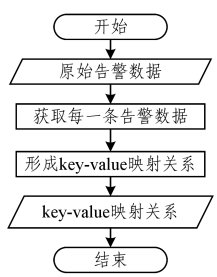

图 4 站点分类流程图

Fig. 4 Flowchart of sites' category

针对需要研究的站点，利用 key-value 的映射关系获取对应的告警数据集。利用需要研究的数据集，求出它的大小 n 及告警种类数 m，并估算出时间窗口 T_{win}。利用 Δ_{ij} 构造 $\Delta_{m\times m}$ 矩阵：

$$\Delta_{m\times m}=\begin{pmatrix}\Delta_{11} & \cdots & \Delta_{1m}\\ \vdots & & \vdots\\ \Delta_{m1} & \cdots & \Delta_{mn}\end{pmatrix} \tag{4}$$

并计算出 $\Delta_{m\times m}$ 矩阵的列向量间的协方差，形成协方差矩阵：

$$\sigma=\begin{pmatrix}cov(c_1,c_1) & \cdots & cov(c_1,c_m)\\ \vdots & & \vdots\\ cov(c_m,c_1) & \cdots & cov(c_m,c_m)\end{pmatrix} \tag{5}$$

其中，$\{c_1,\cdots,c_m\}$ 表示 $\Delta_{m\times m}$ 矩阵中的 m 个列向量，$cov(X,Y)=\frac{\sum_{i=1}^{m}(X_i-\overline{X})(Y_i-\overline{Y})}{m-1}$。协方差矩阵 σ_{ij} 表示 $\Delta_{m\times m}$ 矩阵中第 i 个列向量与第 j 个列向量的相关程度（正相关、负相关、零相关），则 σ 矩阵中第 i 个列向量可以视为告警 i 与其他告警的相关程度。

定义 6 告警 i 与告警 j 的距离为 σ 中第 i 个列向量与第 j 个列向量的距离：

$$d=|\sigma_{i1}-\sigma_{j1}|+\cdots+|\sigma_{im}-\sigma_{jm}| \tag{6}$$

当两个列向量中每个元素越相近时，则 d 越小，说明这两个列向量相似度越高。这是将时序告警进行聚类的基础。数据预处理流程如图 5 所示。

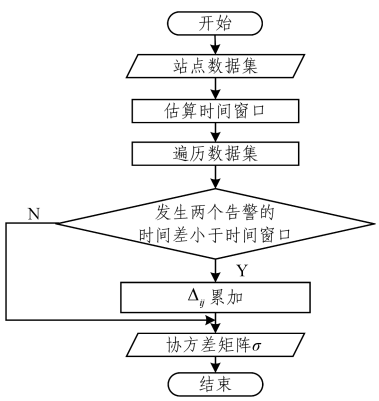

图 5 数据预处理流程图

Fig. 5 Flowchartof data preprocessing

3.3 建立聚类分析模型

将协方差矩阵 σ 的 m' 个非零列向量视为 m' 个样本点，并根据式(6)对这 m' 个样本点进行聚类分析，余下每个零向量单独作为一簇。聚类算法中最常用的是基于 K-Means 算法的聚类[10]：给定样本集 $D=\{x_1,\cdots,x_m\}$，K-Means 算法针对聚类所得簇划分 $C=\{C_1,\cdots,C_k\}$，最小化平方误差。

$$E=\sum_{i=1}^{k}\sum_{x\in C_i}\|x-\mu_i\|_2^2 \tag{7}$$

其中，$\mu_i=\frac{1}{|C_i|}\sum_{x\in C_i}x$ 是簇 C_i 的均值向量。K-Means 算法的关键在于以下 4 个步骤。

S_1：随机选择 k 个样本点，每个样本点作为一个簇的初始均值或中心；

S_2：对于其他样本点，根据其与各个簇中心的距离，将它分配给最相似的簇；

S_3：计算每个簇新的均值或中心；

S_4：重复 S_2 至 S_3，直至每个簇的中心不再变化。

上述算法需要用户事先给出精确的值，在一定程度上影响和限制其应用的合理性。本模型采用文献[11]中一种确定值的方法：构造代价函数，即一个好的聚类模型应该保证中心点之间的距离尽可能大，而同一簇内的点距中心点的距离尽可能小。据此，构造的代价函数为：

$$F(site,k)=\sum_{i=1}^{k}|\mu_i-\mu|+\sum_{i=1}^{k}\sum_{p\in C_i}|p-\mu_i| \tag{8}$$

其中，$site$ 表示站点名称，μ_i 表示第 i 簇的中心，μ 表示所有样本点的中心，C_i 表示属于第 i 个簇的点的集合，p 表示簇中的点，k 表示簇的个数。文献[11]指明，最佳的聚类数 k 应该满足：

$$\min_k\{F(site,k)\},k=1,2,\cdots,m \tag{9}$$

3.4 得到告警聚类结果

在建立模型的过程中，可以用：

$$R(site,k)=\{C_1,C_2,\cdots,C_k\} \tag{10}$$

来记录站点 $site$ 所划分的 k 个簇集合。针对每个研究的站点 $site$，只需找到满足 $F(site,k)$ 最小的 k 以及对应的 $R(site,k)$，即可找到该站点最佳的聚类数以及聚类结果。算法最终的流程如图 6 所示。

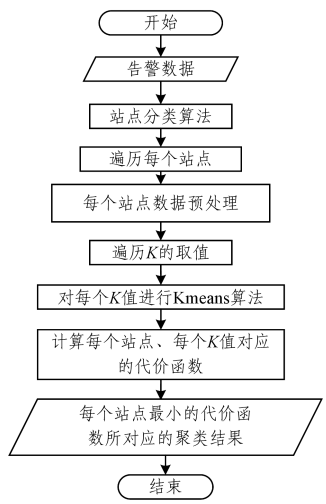

图 6 聚类分析流程图

Fig. 6 Flowchart of clustering analysis

4 实验测试及分析

在香港移动通讯有限公司（CSL）的无线传输网和无线核心网中，通过 EMS 检测到设备或线路故障，产生告警（包括基站、微波、核心网元设备告警等）。

利用 IBM Netcool FM 软件通过 Probe 探针采集告警并存储到 oracle 数据库。本实验数据采用了 oracle 数据库中 2013/3/1 0:00－2013/3/12 15:27 时间段内的 100 万条、544 种告警数据信息。实验语言为 Python3，实验平台为内存 8 GB，CPU 为 Core i5-2.3 GHz，操作系统为 Windows10 的个人计算机。

从数据库中获取不同数量的时序告警数据，分析数据量与站点种类数关系和数据量与 BMI 站点告警种类数的关系，得到了如图 7、图 8 所示的结果。随着时序告警数据量的增加，站点种类数和站点内告警种类数都有明显的上升趋势。这一规律表明，数据量越大，产生告警的原因涵盖得越广，算法的准确度越高。

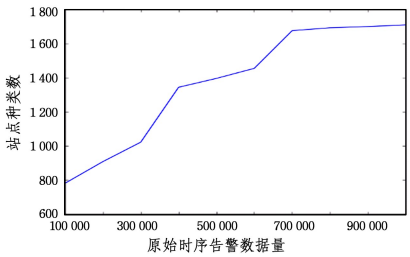

图 7 原始时序告警数据量与站点种类数关系图

Fig. 7 Diagram of relationship between amount of original alarm data for time series and number of sites'categories

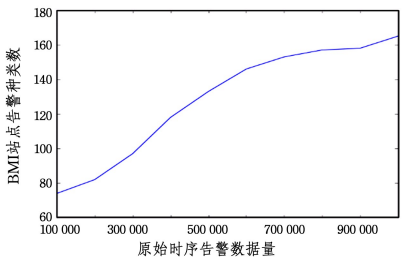

图 8 时间窗口系数与聚类率关系图

Fig. 8 Relationship between coefficient of time window and accuracy of clustering

在本实验中，时间窗口大小受到时间窗口系数 α 的影响。为了寻找最优的时间窗口大小，本实验选取不同的 α 值，并观察不同的 α 值所对应的聚类率 $Rate$[9]，其计算公式如下：

$$Rate=\left(1-\frac{K}{m}\right) \tag{11}$$

其中，K 表示簇数，m 表示告警种类数。设实验研究的站点为 HMK，α 的取值范围在 10 至 100，得到图 9 所示的时间窗口系数与聚类率的关系。

图 9 表明，时间窗口系数的增大导致聚类率上升，这是因为时间窗口系数的增大会使得时间窗口扩大，从而每一个时间窗口内告警数量增加，那么这些告警属于同一簇的可能性变大，最终的聚类率自然也会上升。但随着时间窗口的扩大，无关的告警可能聚成一类，导致聚类结果的准确度降低。文献[9]指出，为了找到最优时间窗口系数，在时间窗口系数与聚类率关系图中，选取斜率明显下降的点作为时间窗口系数的取值。因此，图 9 对应的最优时间窗口系数为 40 至 70。

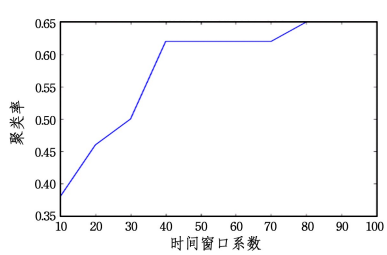

图 9 原始时序告警数据量与 BMI 站点告警种类数的关系

Fig. 9 Relationship between quantity of alarm data for original time series and number of categories for BMI sites

在时间窗系数为 40、研究站点为 HMK 的条件下，分析告警聚类数与代价函数的关系，得到如图 10 所示的结果。当聚类数较少或较多时，导致式(8)中 $\sum_{i=1}^{k}\sum_{p\in C_i}|p-\mu_i|$ 和 $\sum_{i=1}^{k}|\mu_i-\mu|$ 增大，最终使得代价函数较大。图 10 表明，当 HMK 站点的聚类数为 11 时，代价函数最小，并对应了表 1 的聚类结果。

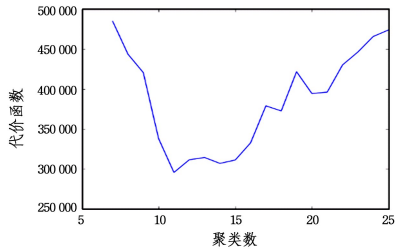

图 10 HMK 站点的聚类数与代价函数的关系图

Fig. 10 Relationship between number of clusters in sites of HMKand cost function

表 1 HMK 站点告警聚类结果

Table 1 Results of clustering of alarm data in sites of HMK

簇编号	告警类型
1	198092230，198096404，198092030，198092348，198092241，198092549-56
2	198092060，198093950，198093951
3	199083022
4	198092245，198093887，198092290，198093888，198092010，198096403
5	198092009，198092240，198092072，198096406
6	198096405
7	198092549-15
8	198092549-19
9	198092549-52
10	198099803
11	198092014

结合以往的告警摘要分析表 1 可得，簇 1 中，由于电源的问题而产生告警 198096404 和 198092549-56，使得硬件设备不能正常工作，产生告警 198092348，并最终导致软件的异常，产生告警 198092030；簇 2 中，因为信道出现异常，导致告警 198093950 和 198093951 被划分在一起。簇 7 中的告警 198092549-15 和簇 9 中的告警 198092549-52 都是由于变频器故障或传感器接受到烟雾而同时产生的，但该算法误认为这两类告警之间没有任何联系，因此将它们分到不同簇。实验结果表明，该算法能够将一些相同原因产生的告警划分在一起，对于维护人员进行故障定位和排查起到一定的指导作用，但该算法存在误判的问题。

结束语 文中针对告警信息的特点，提出了一种新型的告警聚类挖掘算法。其主要的思想是在时间窗口内，利用告警之间时间差的关系，构造出告警之间新的关系矩阵，并对关系矩阵中的列向量进行聚类分析，将不同原因产生的告警划分在一起。

本次实验只针对某一站点的告警数据进行分析，在未来的研究中，将通过结合多站点之间告警数据的联系进一步缩小故障的排查范围。在研究方法上，将采用图的方式建立硬件系统模型，对于有关联边相连或者联系紧密的硬件系统，计算它们所发出的告警是由同一原因产生的概率。与此同时，还可以扩展告警数据的属性，增加以往发生此类告警的摘要字段，对发出的告警做语义分析。

参 考 文 献

[1] KICIMAN E，FOX A. Detecting and localizing anomalous behavior to discover failures in component-based internet services [R]. Technical Report，Stanford，2004.

（下转第 473 页）

ry[J]. Neural Computation,1997,9(8):1735-1780.
[8] GERS F A,SCHMIDHUBER,JÜRGEN,et al. Learning to Forget:Continual Prediction with LSTM[J]. Neural Computation,2000,12(10):2451-2471.
[9] GRAVES A. Supervised Sequence Labelling with Recurrent Neural Networks[M]. Springer,2012.
[10] QIN H M,SUN X. Classifying Bug Reports into Bugs and Nonbugs Using LSTM[C]// The Tenth Asia-Pacifc Symposium on Internet-ware. 2018.
[11] HUANG Y M,JIANG Y,HASAN T,et al. A Topic BiLSTM Model for Sentiment Classification[C]// Innovation in Artificial Intelligence (ICIAI). 2018.
[12] NELSON D M Q , PEREIRA A C M , OLIVEIRA R A D. Stock market's price movement prediction with LSTM neural networks[C]// International Joint Conference on Neural Networks (IJCNN). 2017.
[13] LIN M,CHEN C X. Short-term prediction of stock market price based on GA optimization LSTM neurons[C]// International Conference on Deep Learning Technologies (ICDLT). 2018.
[14] SHIU J N,ZOU J Z,ZHANG J,et al. Research of stock price prediction based on dmd-lstm model [J/OL]. Application Research of Computers. https://doi. org/10. 19734/j. issn. 1001-3695. 2018. 08. 0657.
[15] CHEN J,LIU D X,WU D S. Stock index forecasting method based on feature selection and LSTM model[J]. Computer Engineering and Applications,2019,55(6):108-112.
[16] HO T K. Random decision forests [C]// International Conference on Document Analysis and Recognition. 1995:278-282.
[17] CHOLLET F. Keras [EB/OL]. https://github. com/fchollet/keras,2016.
[18] Keras Documentation[EB/OL]. https://keras. io.
[19] GRANGER C W J. Strategies for Modelling Nonlinear Time - Series Relationships[J]. Economic Record, 2010, 69(3): 233-238.

BAO Zhen-shan, born in 1965, is a member of China Computer Federation. His main research interests include machine learning and Financial technology.

ZHANG Wen-bo, born in 1980, Ph. D, lecturer, is a member of China Computer Federation. Her main research interests include heterogeneous computing and trust computing.

(上接第 443 页)

[2] 王肇刚. 基于网络拓扑约束的时序数据挖掘算法研究与应用[D]. 北京:北京邮电大学,2009.
[3] HAN J W,KAMBER M. 数据挖掘概念与技术(原书第 2 版)(计算机科学丛书)[M]. 北京:机械工业出版社,2008.
[4] AGRAWAI R. Mining association rules between sets of items in large databases[C]// Proceedings of the 1993 ACM SIGMOD Conference. Washington,D C,1993:207-216.
[5] HAN J,PEI J,YIN Y. Mining frequent patterns without candidate generation[C]// ACM SIGMOD International Conference on Management of Data. ACM,2000:1-12.
[6] HATONEN K. Knowledge discovery from telecommunication network alarm databases[C]// ICDE 96. New Orieans,1996:115-122.
[7] NING P,CUI Y,REEVES D S,et al. Techniques and tools for analyzing intrusion alerts[J]. ACM Transactions on Information and System Security(TISSEC),2004,7(2):274-318.
[8] 刘冬生,曾小荟,唐卫东,等. 一种新的告警关联聚类算法[J]. 计算机应用研究,2013,30(12):3786-3789,3793.
[9] 陈兴蜀,何涛,曾雪梅, 等. 基于告警属性聚类的攻击场景关联规则挖掘方法研究[J]. 工程科学与技术,2019,51(3):144-150.
[10] 樊迪,刘静,庄俊玺, 等. 基于因果知识发现的攻击场景重构研究[J]. 网络与信息安全学报,2017,3(4):58-68.
[11] 冯学伟,王东霞,黄敏桓, 等. 一种基于马尔可夫性质的因果知识挖掘方法[J]. 计算机研究与发展,2014,51(11):2493-2504.
[12] KHOSRAVI-FARMAD M,RAMAKI A A,BAFGHI A G. Risk-based Intrusion Response Management in IDS using Bayesian Decision Networks[C]// 2015 5th International Conference on Computer and Kknowledge Engineering(ICCKE). 2015:307-312.
[13] RAMAKI A A,RASOOLZADEGAN A ,BAFGHI A G. A Systematic Mapping Study on Intrusion Alert Analysis in Intrusion Detection Systems[J]. ACM Computing Surveys,2018,51(3):55.

DENG Tian-tian, Ph.D, senior engineer. Her research interests include big data analysis and open source ecology.

XIONG Yin-qiao, Ph.D. His research interests include privacy preserving, information security, and the Internet of Things.