基于评分卡—随机森林的云计算用户公共安全信誉模型研究

周胜利[1,2]，金苍宏[3]，吴礼发[1]，洪征[1]

（1. 陆军工程大学指挥控制工程学院，江苏 南京 210007；2. 浙江警察学院信息技术系，浙江 杭州 310053；
3. 浙江大学城市学院计算机与计算科学学院，浙江 杭州 310015）

摘 要：传统云计算用户信誉的研究主要集中在对用户操作行为信誉评估，较少涉及用户发布文本信息的安全管理，并且存在指标筛选欠准确、信誉评估结果缺乏科学验证等问题，难以满足监管部门的实际需求。针对以上问题，提出基于评分卡—随机森林的云计算用户公共安全信誉模型。首先，利用 Word2Vec 和卷积神经网络进行公共安全标签分类；其次，采用评分卡方法，筛选强相关性指标；最后，结合随机森林算法，建立云计算用户公共安全信誉模型。实验分析表明，所建立的模型能够解决云计算公共安全监管中用户信誉指标筛选不准确和信誉区分准确性低等问题，能够有效识别有害用户，提高云计算用户监管效率。

关键词：云计算安全；安全监管；评分卡；随机森林；卷积神经网络

中图分类号：TP393

文献标识码：A

doi: 10.11959/j.issn.1000-436x.2018085

Research on cloud computing users' public safety trust model based on scorecard-random forest

ZHOU Shengli[1,2], JIN Canghong[3], WU Lifa[1], HONG Zheng[1]

1. Institute of Command and Control Engineering, Army Engineering University of PLA, Nanjing 210007, China
2. Department of Information, Zhejiang Police College, Hangzhou 310053, China
3. School of Computer and Computing Science, Zhejiang University City College, Hangzhou 310015, China

Abstract: Traditional cloud computing trust models mainly focused on the calculation of the trust of users' behavior. In the process of classification and evaluation, there were some problems such as ignorance of content security and lack of trust division verification. Aiming to solve these problems, cloud computing users' public safety trust model based on scorecard-random forest was proposed. Firstly, the text was processed using Word2Vec in the data preprocessing stage. The convolution neural network (CNN) was used to extract the sentence features for user content tag classification. Then, scorecard method was used to filter the strong correlation index. Meanwhile, in order to establish the users' public safety trust evaluation model in cloud computing, a random forest method was applied. Experimental results show that the proposed users' public safety trust evaluation model outperforms the general trust evaluation model. The proposed model can effectively distinguish malicious users from normal users, and it can improve the efficiency of the cloud computing users management.

Key words: cloud computing security, security regulation, scorecard, random forest, convolution neural network

收稿日期：2017-11-30；**修回日期**：2018-04-18

通信作者：周胜利，zhoushengli@zjjcxy.cn

基金项目：国家重点研发计划基金资助项目（No.2017YFB0802900）；国家自然科学基金资助项目（No.U1509219）；杭州市科技发展计划基金资助项目（No.20162013A08）

Foundation Items: The National Key Research and Development Program of China (No.2017YFB0802900), The National Natural Science Foundation of China (No.U1509219), The Science & Technology Development Project of Hangzhou (No.20162013A08)

1 引言

随着云计算技术的迅速发展，云计算应用已经渗透到经济、政治、文化和国防等各个领域。一些著名的 IT 公司纷纷提供云计算服务[1]，如阿里巴巴的阿里云、腾讯公司的腾讯云、亚马逊的 EC2/S3[2]、谷歌的 AppEngine[3]、IBM 的蓝云[4]等。人们只要接入互联网，注册云计算访问账号并购买云服务，就能利用云服务进行工作学习[5]。一些有害用户利用云计算的开放性、便捷性以及监管部门云计算信息安全管理手段的滞后性，将低俗、诈骗、涉政、涉毒等有害信息放到云上并快速传播，造成极其恶劣的社会影响。因此，必须根据云计算信息安全监管的实际情况，研究科学的云计算用户公共安全信誉模型，对云计算用户传播信息进行分析，发现预警有害用户，打击云计算环境下的违法、违规行为。

传统的云计算用户安全管理主要是以识别恶意用户为目标，根据用户异常操作次数、云服务提供商历史反馈、交易时间、交易次数等因素，设置影响因素权重，建立云计算用户信誉模型，较少涉及云计算用户发布文本信息的安全管理，且缺乏对信誉区分准确性的评估论证，难以满足监管部门对云计算用户安全监管及打击云计算犯罪行为的需要。针对以上问题，本文提出基于评分卡—随机森林的云计算用户公共安全信誉模型（PST-SRF, cloud computing users' public safety trust model based on scorecard-random forest）。模型利用卷积神经网络分析云计算用户文本信息，同时基于评分卡—随机森林方法对云计算用户公共安全信誉相关指标进行筛选建模，识别有害用户，提高监管部门云计算安全管理效率。

2 研究现状

国内对云计算安全的监管主要是按照公安部云计算信息安全等级保护要求实行分级监管，主要涉及用户信誉评估、基础设施安全评估、违法用户识别、违法用户溯源分析以及犯罪打击等，尚未形成完善的云计算安全监管制度。目前，公安部正在开展云计算安全评估和认证工作，对于云计算用户信誉的监管还处于起步阶段。

学术界围绕云计算用户信誉模型开展了一系列的理论研究。文献[6]通过计算历史反馈因子、交易完成反馈因子的权重得到整体信誉，但未给出各因素权重的设定依据。文献[7]通过调整通信时间，提高信誉评估准确性。文献[8]对用户的信誉区间进行划分，并对信誉评估标准进行论证，但并未对用户信誉等级做区分准确性评估。文献[9]采用模糊层次分析法来量化云计算用户的行为，从直接信誉、推荐信誉、综合信誉、历史信誉这 4 个维度评估云计算用户信誉情况。由于评估结果只有可信与不可信 2 种，该方法无法实现对云计算用户信誉的量化评估。文献[10]构建了基于动态行为数据监控的信誉模型，利用数据挖掘与知识发现针对影响信誉的多个度测指标进行自适应的动态度量。文献[11]采用了基于改进的证据理论的云计算环境下海量用户行为信誉评估建模方法。该方法借鉴概率加权平均的原理来实现对云计算环境下海量用户行为信誉评估。其成本较低，但是建立的模型无法反映用户行为客观事实。针对上述海量用户信誉评估问题，文献[12]提出了一种基于改进动态博弈论算法的云计算用户行为信誉评估建模方法。该方法以博弈论为基础，利用不完全信息动态博弈对云计算环境下的海量用户进行分类，从不同角度将云计算环境下的用户行为可信关系量化，对用户信誉进行评估。

以上信誉模型的研究主要集中在用户操作行为，很少涉及用户的内容安全管理。在用户内容安全的研究上，传统的方法主要包括支持向量机[13]、朴素贝叶斯分类法[14]、随机森林、决策树法、*K*-最近邻法[15]等。以上方法主要采用浅层机器学习方法，缺乏对分词间逻辑关系的分析。为了提高浅层机器学习分类的精度，国内外学者引入深度学习。深度学习本质上是一种特征提取手段，能够较好地反映出文本信息的特征。文献[16]提出低维实数词向量表示的方法，利用不同词之间的相关性和依赖关系，有效降低了网络的深度。文献[17]介绍了一种高效的 Skip-Gram 模型，可以进行高质量的词向量特征训练，语义相似度效果非常好。文献[18]利用卷积神经网络进行分类，对网络参数进行微调，取得较好的分类效果。文献[19]针对英文文本提出基于 *K*-max 池化操作的动态卷积神经网络方法，提取文本特征向量，取得较好的英文分类效果。

云计算用户公共安全信誉评估本质上是分类问题研究，即把全部用户划分为合规用户和有害用户。文献[20]将线性回归用于信誉评分。文献[21]针对文献[20]信誉评分方法在处理交互效应和非线性因果关系的不足，采用逻辑回归进行信誉评分，

利用正则化对变量进行选择，能够解决共线性问题，实现简单快速，但当变量空间变大时，回归的性能下降明显并且容易存在欠拟合问题。文献[22,23]运用决策树方法进行信誉评分，能够实现自动选择变量，很好地处理缺失信息，防止过拟合，该方法的准确性较采用逻辑回归方法高，但是方法的泛化能力比较弱。

上述研究为云计算用户公共安全信誉评估提供了良好的理论基础，但传统文本分类模型对有害信息的识别效度仍需要评估；信誉量化模型基于结构化指标体系，对于文本等语义模糊且强度难以定义的指标仍需要研究；信誉模型的泛化能力、数据处理能力、分类准确性仍需要进一步研究。基于上述分析，结合云计算用户信息安全管理的需要，首先，利用 Word2Vec 和卷积神经网络进行公共安全标签分类；然后，采用评分卡方法，筛选强相关性指标；最后，结合随机森林算法，建立 PST-SRF 模型进行云计算有害用户的发现、预警。

3　PST-SRF 模型设计

3.1　PST-SRF 模型基本流程

PST-SRF 模型基本流程分 3 个阶段：预处理阶段、公共安全信誉评估阶段和预警监督阶段，如图 1 所示。

3.1.1　预处理阶段

预处理阶段主要进行用户发布文本内容的采集、自然语言处理和文本信息公共安全信誉标签处理，具体操作说明如下。

1) 爬取用户链接内容信息：通过云服务商提供爬虫工具以及日志采集机制，获得云计算用户链接跳转信息等。

2) 爬取用户内容信息：通过云服务提供商的文本信息分析系统，分析云计算用户各种内容信息，包括网页内容清洗、关键内容提取和非结构化信息保存等。

3) 用户信息源分析：将爬取的用户链接内容和用户内容数据进行降维处理，同时使用自然语言处理技术提取特征。

4) 用户公共安全标签：利用卷积神经网络对用户文本信息进行标签分类，按照预先设计的话题分类进行相关性分析。

3.1.2　公共安全信誉评估阶段

公共安全信誉评估阶段主要利用评分卡方法进行公共安全信誉指标分析、筛选，结合随机森林算法对用户公共安全信誉进行评估。具体步骤如下。

1) 指标体系分析：主要通过评分卡方法，利用 WOE（weight of evidence）和 IV（information value）指数评估指标对信誉值的影响，筛选云计算用户公共安全信誉相关性强的指标[24]。

2) 评估模型构建：对指标体系中各个特征利用随机森林进行建模、分值转换。

3) 模型评估：对模型进行评估，主要考察其准确率和召回率，对模型参数和算法进行优化调整，达到最佳的评估效果。

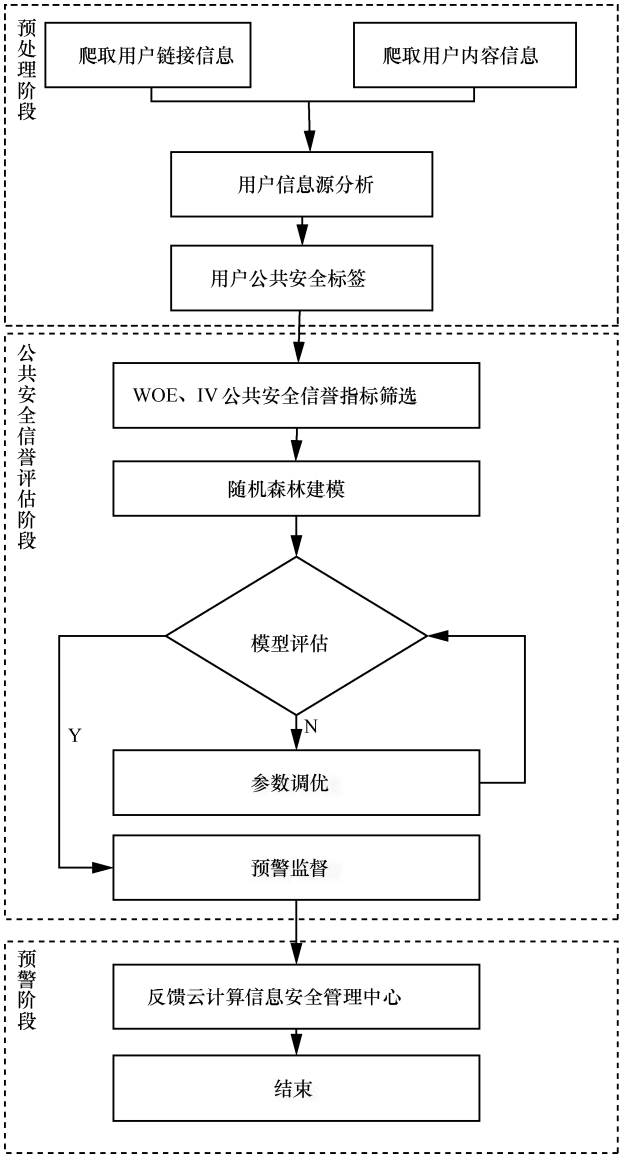


图 1　PST-SRF 模型工作流程

3.1.3　预警监督阶段

该阶段是信息的整体处理和反馈阶段，云服务

提供商将预警监督的数据反馈给云计算信息安全管理中心，便于预测、发现有害用户，及时做出判断。

3.2 公共安全信誉等级分配策略

云计算用户公共安全信誉等级分配策略是公共安全信誉评估的重要参考依据。公共安全信誉等级越低的用户，其访问行为的危险级别越高。用户公共安全信誉分成优、良、中、差 4 个级别，分别对应用户行为危险级别为正常、低危、中危、高危，如表 1 所示。

表 1 公共安全信誉级别与对应的用户可疑级别

用户公共安全信誉级别	用户行为危险级别
优	正常
良	低危
中	中危
差	高危

4 PST-SRF 模型实现

本节将从用户公共安全标签文本处理、指标体系构建、随机森林特征选择算法这 3 个方面介绍 PST-SRF 模型，其具体实现步骤如下。

4.1 用户公共安全标签文本处理

用户公共安全标签文本处理主要包括文本分词相关处理、词向量计算以及利用卷积神经网络进行文本向量分类。

4.1.1 文本分词相关处理

对云计算用户所发言论进行文本分词相关处理，包括分词、词性标注、实体识别等。在分词后，保留名词、动词，去掉相关的形容词、副词和停用词等。

4.1.2 词向量计算

对分词后获得的名词、动词进行向量计算，包括 3 个步骤。

1) 使用 mini-batch 方法对文本内容进行分组，并对文本缺少的分组，使用特殊字符进行补全。

2) 使用词向量计算并扩展相关单词的关联词，使用 Word2Vec 框架把相关单词变成词向量，词向量可计算词之间的距离关系，得到相近的词。

3) 使用 skip-gram 方法进行推测，该方法能根据目标词输出其周围最相关的词。对于词向量中的任意一个单词 w，通过 skip_window 的参数 n，形成一系列的二元组(Context(w), w)，生成 Huffman 树，生成的 Huffman 树节点根据词频进行构建，如图 2 所示。

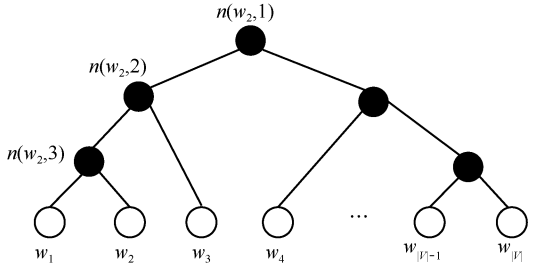


图 2 分词向量处理

4.1.3 文本向量分类

利用卷积神经网络进行文本向量分类。向量的卷积是文本特征向量进行高层次特征提取的过程，与卷积窗口大小、学习率、卷积步长以及正则化系数有关。云环境的文本经过数据预处理后，根据网络文本的长度，本文设置卷积窗口为 5，设置卷积步长为 1（经过实验，卷积窗口设置为 5 效果最好）。文本卷积式为

$$\boldsymbol{W}=\begin{bmatrix} X_{11} & X_{12} & X_{13} & \cdots & X_{1N} \\ X_{21} & X_{22} & X_{23} & \cdots & X_{2N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_{K1} & X_{K2} & X_{K3} & \cdots & X_{KN} \end{bmatrix} \tag{1}$$

$$\boldsymbol{G}=\begin{bmatrix} A_1 & 0 & 0 & 0 & 0 \\ A_2 & A_1 & 0 & 0 & 0 \\ \vdots & A_2 & A_1 & 0 & 0 \\ A_{\text{win}} & \vdots & A_2 & A_1 & 0 \\ 0 & A_{\text{win}} & \vdots & A_2 & A_1 \\ \vdots & \vdots & 0 & \vdots & A_2 \\ 0 & 0 & \vdots & A_{\text{win}} & \vdots \\ 0 & 0 & 0 & 0 & A_{\text{win}} \end{bmatrix} \tag{2}$$

$$\boldsymbol{H}=\boldsymbol{W}\boldsymbol{G}=\begin{bmatrix} H_{11} & H_{12} & H_{13} & \cdots & H_{1Q} \\ H_{21} & H_{22} & H_{23} & \cdots & H_{2Q} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ H_{K1} & H_{K2} & H_{K3} & \cdots & H_{KQ} \end{bmatrix} \tag{3}$$

其中，矩阵 $\boldsymbol{W}$ 表示输入的文本向量，$\boldsymbol{X}_{ij}$ 代表第 i 个文本的词特征向量，$\boldsymbol{G}$ 代表卷积核函数矩阵。$\boldsymbol{G}$ 大小受输入的文本数 K 和卷积的窗口 win 共同影响。矩阵 $\boldsymbol{W}$ 与矩阵 $\boldsymbol{G}$ 输出的结果矩阵为 $\boldsymbol{H}$。其中，$\boldsymbol{H}_{ij}$ 代表第 i 个文本通过 j 次卷积得到的向量。

卷积神经网络的最终目标是将相关词划分到几大类安全标签类别中，需要优化的参数主要包括 2 个部分：词向量和网络参数。本文记词向量为 $\boldsymbol{E}$，卷积操作的参数为 $\hat{W}$，分类器的参数为 W_e，记

$\boldsymbol{W}=\left((s_1,y_1),(s_2,y_2),\cdots,(s_{|q|},y_{|q|})\right)$，其中，$\boldsymbol{s}_i$ 表示对应句子的类别标签，$|\theta|$ 表示训练集样本个数。$p=(y_i|\boldsymbol{s}_i,\theta)$ 表示已知参数 θ 时将句子 $\boldsymbol{s}_i$ 的标签分为 y_i 的概率，则优化目标为 $L=\sum_{i=1}^{|\theta|}\ln p=(y_i|\boldsymbol{s}_i,\theta)+\lambda\theta_2^2$，其中，$\lambda$ 为正向参数。

4.2　指标体系构建

IV 是信息价值或者信息量，*WOE* 为证据权重，是对原始自变量的一种编码形式。

变量 *WOE* 编码，首先把变量进行分箱处理，然后对第 *i* 组 *WOE* 值进行计算。计算式为

$$WOE_i=\ln\left(\frac{py_i}{pn_i}\right)=\ln\left(\frac{\frac{\#y_i}{\#y_T}}{\frac{\#n_i}{\#n_T}}\right) \tag{4}$$

其中，py_i 是分组中符合条件用户（对应的是有害用户）占所有样本中所有相应用户的比例，pn_i 是分组中正常用户占样本中所有正常用户的比例，$\#y_i$ 是分组中有害用户的数量，$\#n_i$ 是分组中正常用户的数量，$\#y_T$ 是样本中所有有害用户的数量，$\#n_T$ 是样本中所有正常用户的数量。*WOE* 表示分组中有害用户占所有有害用户的比例和当前分组中正常用户占所有正常用户的比例的差异。

对式(4)做变换得

$$WOE_i=\ln\left(\frac{py_i}{pn_i}\right)=\ln\left(\frac{\frac{\#y_i}{\#y_T}}{\frac{\#n_i}{\#n_T}}\right)=\ln\left(\frac{\frac{\#y_i}{\#n_i}}{\frac{\#y_T}{\#n_T}}\right) \tag{5}$$

变换后，WOE_i 表示当前组有害用户与正常用户的比值和所有样本中这个比值的差异。这种差异是用这 2 个比值的比值再取对数来表示的。WOE_i 越大，这种差异越大，这个分组里的样本响应的可能性就越大；WOE_i 越小，差异越小，这个分组里的样本响应的可能性就越小。

基于前面的分析，可得 *IV* 的计算式为

$$IV_i=(py_i-pn_i)WOE_i=\left(\frac{\#y_i}{\#y_T}-\frac{\#n_i}{\#n_T}\right)\ln\left(\frac{\frac{\#y_i}{\#y_T}}{\frac{\#n_i}{\#n_T}}\right) \tag{6}$$

通过变量各分组的 *IV* 值可以计算整个变量的 *IV* 值，计算式为

$$IV=\sum_i^n IV_i \tag{7}$$

其中，*n* 为变量分组个数。

通过 *WOE* 和 *IV* 筛选后，保留下来的指标可以用作训练指标模型所需的特征。在训练之前，对每个样本指标进行 *WOE* 的转换。

IV 值的选取根据 *IV* 值衡量标准决定指标是否保留[24]，如表 2 所示。

表 1　　***IV* 值衡量标准**

IV 值	影响效果
< 0.02	对预测变量几乎无影响，不保留
0.02~0.1	对预测变量有一定影响，可保留
0.1~0.3	对预测变量有较大影响，可保留
0.3~0.5	对预测变量有很大影响，可保留
> 0.5	不现实，不保留

4.3　随机森林特征选择算法

通过 *WOE* 和 *IV* 筛选后，保留下来的指标可以作为训练指标模型所需的特征，在训练模型之前，首先要对每个样本指标进行 *WOE* 转换，然后针对该子类评分模型，利用随机森林特征选择（RFFS, radom forest feature selection）算法提取有效特征，具体算法如算法 1 所示。

MacAcc 表示最大分类正确率，*FGSet* 表示筛选后的特征集合，*LMaxAcc* 表示局部最大分类正确率，*LMeanAcc* 表示局部平均正确率。

算法 1 中步骤 1)和步骤 2)表示数据前处理步骤，步骤 3)~步骤 12)为特征选择最大子概率算法，步骤 13)和步骤 14)为从局部最优概率中选择出全局最优的正确率，并且提取相对应的特征。

算法 1　随机森林特征选择算法

输入　*WOE* 和 *IV* 选择为 0.1~0.5 之间的特征，生成数据集 *S*，共有 *N* 个

输出　验证集上的最大分类正确率 *MaxAcc* 和对应的特征集合 *FGSet*

初始化

1) 读入训练数据集 *S*

2) 设置 *MaxAcc* = 0

特征选择最大子概率算法（循环 *N*−2 次）

3) 将 *S* 随机 10 等分

4) 设置局部最大分类准确率 *LMaxAcc*=0 和平均分类准确率 *LMeanAcc* =0

5) *LAcc*[1:10] = 0

6) for (*i* in 1:10)

7) 在 *S* 上构建 *RandomForest* 分类器

8) 在测试集合上执行分类

9) 比较分类结果和观测值，计算 *LAcc*

10) $LMeanAcc = LMeanAcc + LAcc\frac{[i]}{10}$

11) *LMaxAcc* = 最大的 *LAcc*[*i*]

12) 对特征按重要性排序 *FGSet*

输出结果

13) 输出最大的 *LMaxAcc* 为 *MaxAcc*

14) 输出分类准确率最高 *MaxAcc* 的特征集合 *FGSet*

5 实验评估分析

通过实验分析，验证文本标签分类的正确性、模型指标筛选处理的合理性、随机森林参数调整的合理性以及评估模型的准确性。

在阿里云 ODPS（open data processing service）平台进行实验测试。该平台由 10 台 S10 机器组成，每台机器的配置为 32 核 64 GB 内存 2 TB 硬盘，其中包含 3 台管控集群和 10 台计算集群，管控集群用于对任务进行分发和管理，计算集群负责各机器学习分布式计算任务的运行。

5.1 用户文本标签分类实验

实验的目的是评估云计算用户文本标签分类的准确性。实验所使用的数据为某政法云平台的公共安全分类数据集。本文选取了其中包含政治人物、司法相关、公众人物等的实验数据。对每个类别分别随机选取其中的 70%用于训练、30%用于测试。数据预处理阶段使用 Word2Vec 完成文本的简单处理，包括分词、词性标注等。在预处理阶段会设置句子分词后最长词数，对未达到最长分词个数的分词单元统一使用规定符号补全。然后，利用卷积神经网络提取指定任务的句子特征。为了防止过拟合和提升模型的预测准确率，使用 dropout 算法进行优化。

实验主要通过准确率（accuracy）、灵敏度（sensitive）、特异度（specificity）、精确率（precision）、召回率（recall）、*F*1 值等指标进行评估[24,25]。评估指标如表 3 所示。

表 3 中，*P* 表示正样本数量；*N* 表示负样本数量；*TP* 表示被正确分类的正样本；*FP* 表示被错误地标记为正样本的负样本；*FN* 表示被错误地标记为负样本的正样本；*TN* 表示被正确分类的负样本。

表 4 的斜对角线体现了真实标签与预测标签的一致性。表 4 中 A~R 分别代表 18 种敏感话题。模型对话题 D、F、J、K、L、N 等标签预测准确率在 90%以下，其余各个类别的预测准确率在 90%以上，满足实际应用需求。

表 5 各项评估指标均在 90%以上，说明模型具有很好的分类效果，满足公共安全信誉评估应用需求。表 5 中 A~R 含义同表 4。

5.2 评分卡指标筛选实验

实验的目的是为了说明模型指标选择的科学性。实验数据为从某政法云平台中选取的 6 万个用户内容行为数据，包括 1 万个黑样本数据和 5 万个白样本数据。对各数据字段进行属性扩充后，最终进行模块化存储。

将黑样本和白样本分别按照 7:3 的比例（机器学习常用的比例）进行分割再组合，形成训练样本和测试样本。

表 3 模型评估指标

评估指标	计算式	说明
准确率	$accuracy = \frac{TP+TN}{P+N}$	表示模型在测试集上的准确性，即模型正确分类的样本数所占比例
灵敏度	$sensitive = \frac{TP}{P}$	正确判断的正样本的百分比
特异度	$specificity = \frac{TN}{N}$	正确判断的负样本的比例
精确率	$precision = \frac{TP}{TP+FP}$	被标记为正样本的样本实际为正样本的比例
召回率	$recall = \frac{TP}{TP+FN} = \frac{TP}{P}$	是一个完全性度量，同灵敏度
*F*1 值	$F1 = \frac{2precision \cdot recall}{precision + recall}$	

表 2　　**混淆矩阵值**

敏感话题	预测值																	
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
A	97	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0
B	0	93	0	0	1	0	0	1	0	0	0	0	1	0	0	1	0	0
C	0	0	96	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0
D	0	0	0	71	0	0	0	0	0	0	0	0	0	0	1	0	0	0
E	0	0	0	0	96	0	0	0	0	0	0	1	0	0	0	0	1	0
F	0	0	2	1	0	78	0	0	0	0	0	0	0	0	1	0	0	4
G	1	0	0	0	1	0	96	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	97	0	0	0	0	0	0	0	0	1	0
I	0	1	0	1	0	0	0	0	93	0	7	0	0	1	1	0	0	0
J	0	0	0	0	0	0	0	0	9	87	0	1	0	0	0	0	0	0
K	0	1	0	0	0	0	1	9	0	0	88	1	0	0	0	0	1	0
L	0	0	0	0	0	0	1	1	0	0	0	83	0	1	1	0	0	0
M	0	0	1	0	0	1	0	0	0	3	1	0	90	0	0	0	0	0
N	1	0	0	0	0	0	1	0	0	0	20	0	0	78	0	1	0	0
O	0	0	0	0	0	1	0	1	0	0	0	0	1	0	95	0	0	0
P	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	105	0	0
Q	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	1	97	0
R	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	96

表 5　　**混淆矩阵评估指标**

敏感话题	*TP*	*TN*	*FP*	*FN*	准确率	灵敏度	特异度	精确率	*F*1
A	97	1 634	3	2	0.997 11	0.979 79	0.998 16	0.970 00	0.974 87
B	93	1 637	2	4	0.996 54	0.958 76	0.998 77	0.978 94	0.968 75
C	96	1 634	4	2	0.996 54	0.979 59	0.997 55	0.960 00	0.969 69
D	71	1 662	2	1	0.998 27	0.986 11	0.998 79	0.972 60	0.979 31
E	96	1 634	4	2	0.996 54	0.979 59	0.997 55	0.960 00	0.969 69
F	78	1 648	2	8	0.994 23	0.906 97	0.998 78	0.975 00	0.939 75
G	96	1 634	4	2	0.996 54	0.979 59	0.997 55	0.960 00	0.969 69
H	97	1 625	13	1	0.991 93	0.989 79	0.992 06	0.881 81	0.932 69
I	93	1 623	9	11	0.988 47	0.894 23	0.994 48	0.911 76	0.902 91
J	87	1 636	3	10	0.992 51	0.896 90	0.998 16	0.966 66	0.930 48
K	88	1 605	30	13	0.975 23	0.871 28	0.981 65	0.745 76	0.803 65
L	83	1 646	3	4	0.995 96	0.954 03	0.998 18	0.965 11	0.959 53
M	90	1 636	4	6	0.994 23	0.937 50	0.997 56	0.957 44	0.947 36
N	78	1 633	2	23	0.985 59	0.772 27	0.998 77	0.975 00	0.861 87
O	95	1 633	5	3	0.995 39	0.969 38	0.996 94	0.950 00	0.959 59
P	105	1 625	3	3	0.996 54	0.972 22	0.998 15	0.972 22	0.972 22
Q	97	1 633	3	3	0.996 54	0.970 00	0.998 16	0.970 00	0.970 00
R	96	1 634	4	2	0.996 54	0.979 59	0.997 55	0.960 00	0.969 69

表 6 cnt_good 指标 *WOE* 分析结果

区间	*WOE*	总数	正例数	负例数	区间占比	正例占比	负例占比
(−∞,1]	−1.399	111	36	75	14.07%	6.91%	27.99%
(1, 3]	−1.054	52	21	31	6.59%	4.03%	11.57%
(3, 8]	−0.25	93	56	37	11.79%	10.75%	13.81%
(76, 109]	−0.207	80	49	31	10.14%	9.4%	11.57%
(8, 12]	0.104	60	41	19	7.6%	7.87%	7.09%
(109, 194]	0.207	78	55	23	9.89%	10.56%	8.58%
(43, 76]	0.207	78	55	23	9.89%	10.56%	8.58%
(12, 26]	0.945	84	70	14	10.65%	13.44%	5.22%
(26, 43]	1.445	74	66	8	9.38%	12.67%	2.99%
(194, +∞)	1.666	79	72	7	10.01%	13.82%	2.61%

表 6 表示 cnt_good 指标的 *WOE* 值及其代表性分析结果。表 6 中横坐标表示正面句子条数分箱区间，纵坐标 *WOE* 表示不同区间对应的值。图 3 进一步分析 cnt_good 指标的 *WOE* 值变化情况和分箱效果。图 3 中，*WOE* 值越大，反向指标越小；*WOE* 值越小，且中间无跳点，则说明该分箱效果较好。另外，正向指标的 *WOE* 值正斜率越大，反向指标的负斜率越大，则说明指标区分能力越好，即对评分的影响越大。

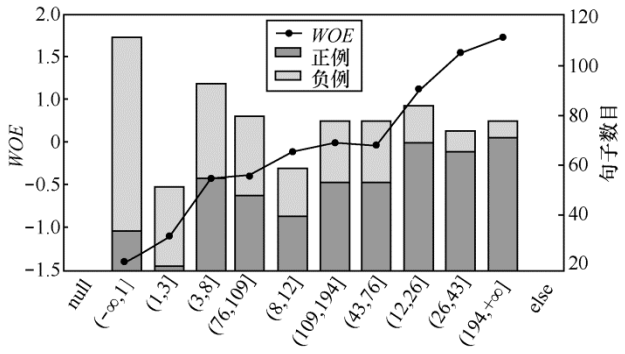


图 3 cnt_good 指标 *WOE* 分析结果

通过指标分析，可以筛选出符合统计学规律的指标作为模型训练的候选特征，然后通过多重共线性分析对候选特征进行相关性分析，删除一些相关性强的指标，减少模型特征的维度。

筛选的最后一步是通过 *IV* 分析来确定单一指标对模型预测能力的贡献，*IV* 表示每个指标所包含的信息量，相当于是指标 *WOE* 值的加权，其值的大小决定了自变量对于目标的影响程度，*IV* 的实验分析结果如表 7 所示。

5.3 公共安全信誉模型 *ROC*、*AUC* 混淆矩阵实验评估

经过数据分布情况分析，不断进行模型迭代，最终得到 22 个重要特征，树的棵数为 200，每棵树的最大深度为 5。通过 *ROC*、*AUC* 等指标评估模型效果如图 4 所示。

表 7 指标 *IV* 值

指标名称	中文解释	类型	分箱方法	*IV*
bad_total	负面占比	double	等频	0.363
middle_total	中性占比	double	等频	0.446
good_total	正面占比	double	等频	1.454
cnt_bad	负面句子数	bigint	等频	0.363
sum_sentence	句子总数	bigint	等频	0.700
cnt_middle	中性句子数	bigint	等频	0.521
cnt_good	正面句子数	bigint	等频	1.225

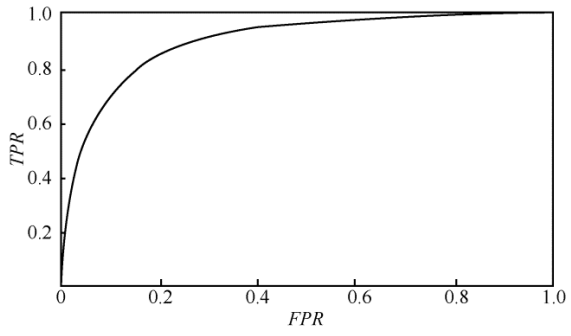


图 4 *ROC* 曲线

图 4 中纵轴表示 *TPR*，即实际正样本中被预测成正样本的比例；横轴表示 *FPR*，即实际负样本中被预测成正样本的比例。其中，4 个坐标点分别表示如下。

1) 坐标(0,0)表示实际正样本中，被预测成正样本的比例为 0，即所有预测都错误；而在实际负样本中，被预测成正样本的比例也为 0，即所有负样本都预测正确。

2) 坐标(0,1)表示实际正样本中，被预测成正样本的比例为 1，即所有正例都预测正确；同时，实际负例也被预测正确。

3) 坐标(1,1)表示实际正样本中，被预测成正样本的比例为 1，即所有正样本都预测正确；但实际负样本中，被预测成正样本的比例也为 1，所有负样本预测错误。

4) 坐标(1,0)表示实际正样本中，被预测成正样本的比例为 0，即所有正样本都预测错误；且实际负样本中，被预测成正样本的比例为 1，即所有负样本都预测错误。

可见，第 2)种情况效果是最好的，而第 1)和第 3)种情况是 2 个极端，第 4)种情况效果最差。因此，对于 *ROC* 曲线，越是靠近左上角，效果则越好。

为了对 *ROC* 曲线中的效果进行数值量化，引入了 *AUC* 指标，该指标表示的是 *ROC* 曲线下方的面积。模型效果越好，*ROC* 曲线越逼近左上角，*AUC* 值越靠近 1.0。本模型的 *AUC*=0.9，达到了预期效果。

在本文的模型中，以概率 0.5 作为预测为正样本和负样本的切分点，得到的混淆矩阵效果如表 8 所示。表 8 中预测为正样本（1 表示正样本，0 表示负样本）的数目为 1 843 个，正确数目为 1 340 个，错误数目为 503 个，正确率 72.7%，召回率 47.4%，*F*1 为 57.4%，满足实际应用需求。

6　结束语

本文所提 PST-SRF 模型对云计算用户的文本内容进行公共安全信誉评估。模型利用词向量结合卷积神经网络方法，对用户内容安全进行分类打标签，并结合评分卡方法中 *WOE* 和 *IV* 指标筛选方法，筛选云计算用户公共安全相关指标，之后根据数据离散情况、随机森林中树的个数、属性个数、树的深度，采用 bootstrap 对样本进行子采样，降低子模型之间的关联度，建立公共安全信誉评估模型。实验评估结果表明，PST-SRF 模型对用户公共安全信誉具有较好的区分效果，能够有效识别有害用户，提高云计算用户信息安全管理效率。

表 8　　**本文混淆矩阵**

分类值	正确个数/个	错误个数/个	总计/个	正确率	召回率	*F*1 指标
1	1 340	503	1 843	72.7%	47.4%	57.4%
0	14 430	1 490	15 920	90.6%	96.6%	93.5%

参考文献：

[1] 刘楠, 魏进武, 刘露.大数据交换信息链[J].电信科学, 2016, 32(10): 130-136.
LIU N, WEI J W, LIU L. Big data exchange based on information chain[J].Telecommunications Science, 2016, 32(10): 130-136.

[2] 周维, 路劲, 周可人, 等.基于并发跳表的云数据处理双层索引架构研究[J].计算机研究与发展, 2015, 52(7): 1531-1545.
ZHOU W, LU J, ZHOU K R, et al. Concurrent skiplist based double-layer index framework for cloud data processing[J].Journal of Computer Research and Devlopment, 2015, 52(7): 1531-1545.

[3] 张常有, 邵立向, 李文清, 等.云服务的自组织机制及性能分析[J].中国通信, 2012, 9(6):135-144.
ZHANG C Y, SHAO L X, LI W Q, et al. Self organizing mechanism for cloud services and performance analysis[J].China Communications, 2012, 9(6):135-144.

[4] 陈康, 郑纬民.云计算:系统实例与研究现状[J].软件学报, 2009, 20(5): 1337-1348.
CHEN K, ZHENG W M. Cloud computing: system instances and current research[J]. Journal of Software, 2009, 20(5):1337-1348.

[5] 王国峰, 刘川意, 潘鹤中, 等.云计算模式内部威胁综述[J].计算机学报, 2017, 40(2): 296-316.
WANG G F, LIU C Y, PAN H Z, et al. Survey on insider threats to cloud computing[J].Chinese Journal of Computers, 2017, 40(2): 296-316.

[6] 李丙戌, 吴礼发, 周振吉, 等. 基于信任的云计算身份管理模型设计与实现[J]. 计算机科学, 2014, 41 (10): 144-148.
LI B X, WU L F, ZHOU Z J, et al. Design and implementation of trust-based identity management model for cloud computing[J]. Computer Science, 2014, 41(10): 144-148.

[7] TIAN L Q, LIN C, YANG N. Evaluation of user behavior trust in cloud computing[C]//2010 International Conference on Computer Application and System Modeling (ICCASM). 2010: 567-572.

[8] 苏锷, 李凤华, 史国振. 基于行为的多级访问控制模型[J]. 计算机研究与发展, 2014, 51(7) : 1604-1613.
SU M, LI F H, SHI G Z. Action-based multi-level access control model[J]. Computer Research and Development, 2014, 51(7): 1604-1613.

[9] 周茜, 于炯. 云计算下基于信任的防御系统模型[J]. 计算机应用

2011, 31(6) : 1531-1535.
ZHOU Q, YU J. Defense system model based on trust for cloud computing[J]. Computer Application, 2011, 31(6) : 1531-1535.

[10] LI X Y, GUI X L, MAO Q, et al. Adaptive dynamic trust measurement and prediction model based on behavior monitoring: adaptive dynamic trust measurement and prediction model based on behavior monitoring[J]. Chinese Journal of Computers, 2009, 32(4): 664-674.

[11] 杨家兴. 复杂网络安全态势评估模型仿真分析[J]. 计算机仿真, 2013, 30(8): 289-292.
YANG J X. Simulation analysis of complex network security situation assessment model[J]. Computer Simulation, 2013, 30(8):289-292.

[12] 毛建景, 张凯萍. 云计算环境下海量用户行为信任评估模型[J]. 计算机仿真, 2016, 33(3): 385-388.
MAO J J, ZHANG K P. Behavior trust evaluation model for massive users under cloud computing environment[J]. Computer Simulation, 2016, 33(3): 385-388.

[13] 丁世飞, 张健, 张谢锴, 等. 多分类孪生支持向量机研究进展[J]. 软件学报, 2018, 29(1): 89-108.
DING S F, ZHANG J, ZHANG X K, et al. Survey on multi class twin support vector machines[J]. Journal of Software, 2018, 29(1): 89-108.

[14] WANG S C, XU G L, DU R J. Restricted Bayesian classification networks[J]. Science China(Information Sciences), 2013, 56(7): 210-224.

[15] LAM H K, EKONG U, LIU H B, et al. A study of neural-network-based classifiers for material classification[J]. Neurocomputing, 2014: 144.

[16] HINTON G E. Learning distributed representations of concepts[C]// The 8th Annual Conference of the Cognitive Science Society. 1986: 1-12.

[17] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[J]. Computer Science, 2013: 1-12.

[18] KIM Y. Convolutional neural networks for sentence classification[C]// The 2014 Conference on Empirical Methods in Natural Language Processing. 2014:1746-1751.

[19] STILO G, VELARDI P. Efficient temporal mining of micro-blog texts and its application to event discovery[J]. Data Mining and Knowledge Discovery, 2016, 30(2): 372-402.

[20] KENNETH A C, MICHAEL E S. Debit, credit, or cash: survey evidence on gasoline purchases[J]. Journal of Economics and Business, 1999, 51(5): 409-421.

[21] WIGINTON J C. A. note on the comparison of logit and discriminant models of consumer credit behavior[J].Journal of Financial and Quantitative Analysis,1980,15:757-770.

[22] MAKOWSKI P. Credit scoring branches out[J]. Credit World, 1985, 75:30-37.

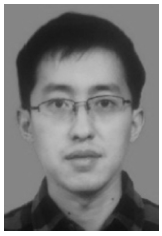
[23] CARTER C, CATLETT J. Assessing credit card application using machine learning[J].IEEE Expert Magazine, 1987, 2(3):71-79.

[24] WANG G, HAO J, MA J, et al, A comparative assessment of ensemble learning for credit scoring[J]. Expert Systems with Applications , 2011, 38(1) :223-230.

[25] GARETH J, DANIELA W, TREVOR H, et al. Anintroduction to statistical learning with application in R[M]//An Introduction to Satistical Learning. 2013: 78-129.

[作者简介]

周胜利（1982-），男，浙江苍南人，陆军工程大学博士生，主要研究方向为云计算安全。



金苍宏（1982-），男，浙江绍兴人，博士，浙江大学城市学院讲师，主要研究方向为机器学习、云计算。

吴礼发（1968-），男，湖北蕲春人，博士，陆军工程大学教授，主要研究方向为大数据安全。

洪征（1979-），男，江西南昌人，博士，陆军工程大学副教授，主要研究方向为网络安全。