



## 基于机器学习的多源威胁情报质量评价方法

刘汉生<sup>1,2</sup>, 唐洪玉<sup>1</sup>, 薄明霞<sup>1</sup>, 牛剑锋<sup>1</sup>, 李天博<sup>1</sup>, 李玲晓<sup>1</sup>

(1. 中国电信股份有限公司上海研究院, 上海 200122;

2. 中国电信股份有限公司北京研究院, 北京 102209)

**摘要:** 在多源威胁情报收集过程中, 由于存在数据价值密度低、情报重复度高、失效时间快等问题, 情报中心难以对海量情报数据做出科学决策。针对上述问题, 提出一种基于机器学习的多源威胁情报质量评价方法。首先基于标准情报格式, 设计了一套多源情报数据标准化流程; 其次, 针对情报数据的特点, 分别从情报来源、情报内容、活跃周期、黑名单库匹配程度 4 个维度提取特征作为评估情报质量的依据; 然后针对提取的特征编码, 设计了一套基于深度神经网络算法和 Softmax 分类器的情报质量评价模型, 并利用反向误差传播算法最小化重构误差; 最后根据 2 000 条开源已标注样本数据, 利用  $K$  折交叉验证法对模型进行验证, 得到了平均 91.37% 的宏查准率和 84.89% 的宏查全率, 为多源威胁情报质量评估提供借鉴和参考。

**关键词:** 信息安全; 威胁情报; 质量评价; 深度神经网络

**中图分类号:** TP393

**文献标识码:** A

**doi:** 10.11959/j.issn.1000-0801.2020010

## A multi-source threat intelligence confidence value evaluation method based on machine learning

LIU Hansheng<sup>1,2</sup>, TANG Hongyu<sup>1</sup>, BO Mingxia<sup>1</sup>, NIU Jianfeng<sup>1</sup>, LI Tianbo<sup>1</sup>, LI Lingxiao<sup>1</sup>

1. Shanghai Research Institute of China Telecom Co., Ltd., Shanghai 200122, China

2. Beijing Research Institute of China Telecom Co., Ltd., Beijing 102209, China

**Abstract:** During the collection process of multi-source threat intelligence, it is very hard for the intelligence center to make a scientific decision to massive intelligence because the data value density is low, the intelligence repeatability is high, and the ineffective time is very short, etc. Based on those problems, a new multi-source threat intelligence confidence value evaluation method was put forward based on machine learning. First of all, according to the STIX intelligence standard format, a multi-source intelligence data standardization process was designed. Secondly, according to the characteristic of data, 14 characteristics were extracted from four dimensions of publishing time, source, intelligence content and blacklist matching degree to be the basis of determining the intelligence reliability. After getting the feature encoding, an intelligence confidence value evaluation model was designed based on deep neural network algorithm and Softmax classifier. Backward propagation algorithm was also used to minimize reconstruction error. Last but not least, according to the 2 000 open source marked sample data,  $k$ -fold cross-validation method was used to evaluate the model and get an average of 91.37% macro-P rate and 84.89% macro-R rate. It was a good reference for multi-source threat intelligence confidence evaluation.

**Key words:** information safety, threat intelligence, confidence evaluation, deep neural network



## 1 引言

随着信息技术的发展,网络安全形势日益严峻。为有效感知网络安全态势,建立以威胁情报为核心的安全防御体系逐渐成为主流选择<sup>[1]</sup>。目前在情报中心建设过程中,企业除收集内部情报数据外,还需通过开源数据爬取和外部厂商交换的方式获取外部情报。而在数据收集过程中,由于数据价值密度低、情报重复度高、失效时间快等问题,情报库中存在大量混淆和失效的错误情报。因此,有效地从海量数据中提取真实可信的威胁情报,对构建情报中心以及接下来的情报预警、威胁溯源具有重要影响。

由于威胁情报概念提出较晚,目前国内外对于威胁情报质量评估的研究屈指可数。2013年Dandurand等人<sup>[1]</sup>第一次提出威胁情报数据共享中存在数据质量验证问题,但未给出明确的解决方案。2016年Bloodgood等人设计了一种基于抽象数据模型的评判方法,包括MDS聚类、抽象、应用、应答4个阶段,主要依靠用户反馈和主数据依赖来判定数据质量,但实际应用中用户反馈数据较少且评估准确程度很难衡量,因此该方法具有一定的局限性。2017年刘春年等人<sup>[2]</sup>基于工程化思维与WRS方法论构建了应急信息可信度模型,为威胁情报质量评估模型提供参考。2018年贺雅琪<sup>[3]</sup>提出一种基于D-S证据理论的数据融合方法,针对数据不确定问题采用朴素贝叶斯和FCM算法进行判别,但缺乏在信息安全场景下对攻击时间、攻击类型等初始特征的分析。现阶段主要威胁情报公司,如微步在线、天际友盟等,其数据质量评价主要思路是从来源厂商、失效时间等角度进行评估以及参数及权重固定,评价体系主观性较强。

针对上述问题,本文为进一步提高威胁情报的可靠性,结合深度神经网络和情报数据特点,设计了一套数据清洗、特征提取、分类训练的数

据质量评估流程。并基于公开数据集对模型进行了验证,得到了平均91.37%的宏查准率和84.89%的宏查全率,为多源威胁情报质量评估的研究提供借鉴和参考。

## 2 威胁情报置信度模型研究

### 2.1 情报置信数据标准化

在情报数据格式方面,目前主要标准有CybOX、STIX、TAXII 3种<sup>[4]</sup>。CybOX(cyber observables eXpression)提供了一个标准化的威胁情报表达方法,主要用于规范描述在通信和网络操作中观察到事件的状态属性<sup>[5]</sup>。STIX(structured threat information eXpression)设计了一套结构化威胁信息表达式,通过威胁源、攻击动机、攻击手段、防御措施等特征表达威胁攻击细节<sup>[6]</sup>,是目前最普遍使用的威胁情报数据格式。TAXII(trusted automated eXchange of indicator information)则定义了数据传输共享的规范,在实施时可以跨组织、产品和服务边界共享网络威胁情报。目前主要基于CybOX进行情报描述,基于STIX进行格式定义,基于TAXII进行情报共享<sup>[7]</sup>。

目前企业或机构获取情报的方式有3种:第一种是已经标准化为STIX格式的威胁情报,通过共享或购买方式获取后可以直接入库<sup>[8]</sup>,例如微步在线、天际友盟等威胁情报厂商。第二种是企业从WAF(Web application firewall)、漏扫等设备中自产情报,情报格式一般为固定的syslog日志;其数据格式较为固定,维度也比较丰富,一般经过简单的数据处理就可以转换为STIX格式。第三种是开源情报共享平台,如badIPs、malware等,这类情报大多只提供恶意IP地址、域名、恶意文件Hash、恶意URL中的一类,单独来源难以形成标准化格式;其需要先通过调用第三方接口和数据挖掘的方法进行数据富化,然后再进行标准化处理。以上3种方式由于来源不同,获取的情报中存在很大的重复部分。因此,为了降低情报

重复率、增加情报数据可信度<sup>[9]</sup>, 本文通过比较攻击时间、攻击源、被攻击者等关键指标的相似程度, 进而对相似的情报进行数据归并, 整个情报标准化流程如图1所示。

本文情报聚合过程中的相似度计算主要首先从攻击时间、攻击者、被攻击者3个维度进行相似度计算。多个来源情报 $\{A_1, A_2, A_3, \dots, A_n\}$ , 其中任意两个来源间相似度的计算方式如式(1)所示,  $\text{Sim}_{\text{time}}$ 、 $\text{Sim}_{\text{attack}}$ 、 $\text{Sim}_{\text{defense}}$  分别代表时间相似度、攻击源相似度和被攻击源相似度。

$$\text{Sim}(A_i, A_j) = \text{Sim}_{\text{time}} \times \text{Sim}_{\text{attack}} \times \text{Sim}_{\text{defense}} \quad (1)$$

时间相似度计算方法如式(2)所示,  $\alpha$  为常参,  $t_i$ 、 $t_j$  分别代表两条情报的攻击发现时间的分钟值, 其将两条情报的时间差由 $[0, +\infty]$ 映射到 $[0, 1]$ 空间中。当两条情报发现时间完全一致时, 其时间相似度为1, 其间隔时间越长, 情报相似度越小。

$$\text{Sim}_{\text{time}} = 1 - \frac{\frac{e^{\frac{|t_i - t_j|}{\alpha}} - e^{-\frac{|t_i - t_j|}{\alpha}}}{e^{\frac{|t_i - t_j|}{\alpha}} + e^{-\frac{|t_i - t_j|}{\alpha}}}}{2} \quad (2)$$

攻击源是否相同是两条情报是否相似的最基本特征。本文仅从两条情报数据中是否有完全相同的攻击源进行比较, 其相似度计算如式(3)所示。当攻击源相同时, 攻击源相似度为1, 否则为0。

$$\text{Sim}_{\text{attack}} = \begin{cases} 1, S_i = S_j \\ 0, S_i \neq S_j \end{cases} \quad (3)$$

被攻击目标的相似度主要通过比较两条情报的攻击目标重叠程度, 考虑部分开源情报数据缺少被攻击目标数据, 设置了常参  $\theta$  在 $[0, 1]$ 作为调节, 具体计算方法如式(4)所示,  $D_1$  和  $D_2$  分别代表两条威胁情报的攻击目标集合, 分子和分母分别代表交集和并集的个数。

$$\text{Sim}_{\text{defense}} = \theta + (1 - \theta) \times \frac{D_1 \cap D_2}{D_1 \cup D_2} \quad (4)$$

## 2.2 情报特征编码

在第2.1节中利用数据标准化方法将情报归并, 并将其转换为标准结构化数据, 但是结构化数据仍然不能被计算机处理, 需要利用特征编码的方式将数据抽象表达出来。通过分析目前掌握

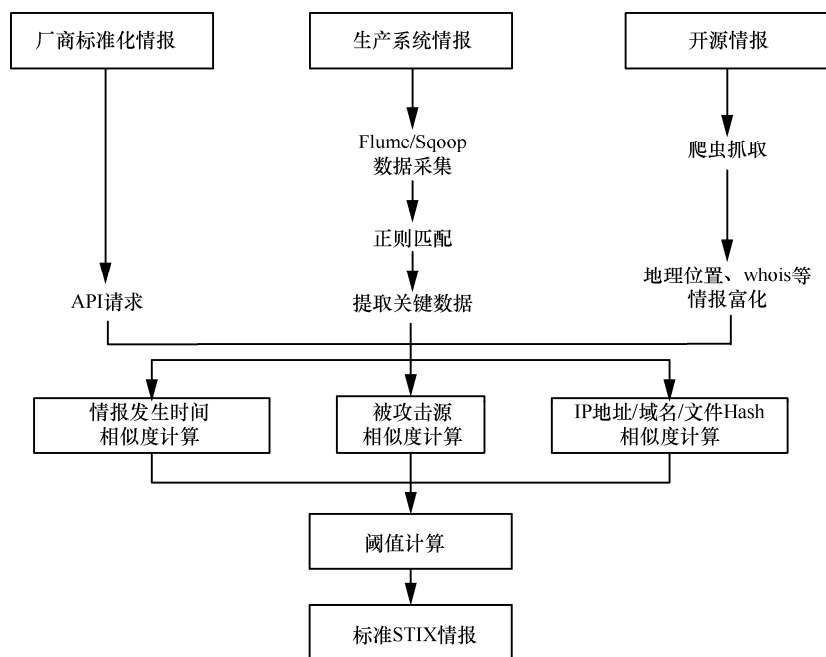


图1 情报数据标准化流程



的天际友盟、微步在线等威胁情报源来看,情报来源、发布时间、情报内容、黑名单库匹配程度是最直观的评价情报质量的特征,下文分别从这4个方面对情报数据进行特征编码。

情报的来源能在很大程度上反映出一条情报的可信程度,一般来说,多来源情报比单来源情报质量更高,知名威胁情报厂商、专业情报评估机构比个人情报数据质量更高<sup>[10]</sup>。由于情报来源之间的关联性较小,为了保证每个情报来源特征的独立性,本文对其采用 onehot 编码方式,将每个情报源映射到不同维度上。情报源 onehot 特征编码示意图如图2所示,对  $n$  个不同的情报源,对其顺序编号,分别将其映射到  $n$  维的 0、1 特征空间。

| 情报来源       | 编号  | onehot 编码 |
|------------|-----|-----------|
| badIPs     | 1   | 10000...0 |
| malware    | 2   | 01000...0 |
| 天际友盟       | 3   | 00100...0 |
| 微步在线       | 4   | 00010...0 |
| ⋮          | ⋮   | ⋮         |
| VirusTotal | $n$ | 00000...1 |

图2 情报源 onehot 特征编码示意图

由于威胁情报具有很强的时效性,发布时间是评价情报是否有效的重要特征指标。一般来说,当前时间离情报发布时间越近,其失效的可能性越小<sup>[11]</sup>。同时,由于一条情报可能会在不同时间段多次发布,通过记录其最近3次的发现时间,能够表征情报的波动趋势,有助于分析当前情报的可信程度。本方法情报基于时间特征的编码见表1。

表1 基于时间维度的特征编码

| 符号(特征)              | 描述           | 示例(将时间转换为毫秒值)     |
|---------------------|--------------|-------------------|
| $T_{\text{now}}$    | 当前时间         | 1 559 638 191 019 |
| $T_{\text{latest}}$ | 情报最近发布时间     | 1 559 638 132 222 |
| $T_{\text{sec}}$    | 情报发布时间2(空为0) | 1 559 638 131 115 |
| $T_{\text{thir}}$   | 情报发布时间3(空为0) | 0                 |

情报内容描述攻击者采取攻击行为的时间、地点和攻击手法<sup>[12]</sup>。有效分析威胁情报内容,有助于感知威胁态势,描绘出攻击发生的情境,为安全从业人员快速提供预警消息。基于内容维度的特征编码见表2,本方法基于内容维度提取了攻击类型、IP地址/域名经度、IP地址/域名纬度、恶意IP地址个数、恶意域名个数、恶意URL个数、恶意文件样本个数7个特征。其中攻击类型与上文情报源编码方式相同,将不同类型的攻击方式映射到不同维度。

表2 基于内容维度的特征编码

| 符号(特征)                 | 描述               | 示例           |
|------------------------|------------------|--------------|
| $T_{\text{attack}}$    | 攻击类型(可包含多种)      | 100000010000 |
| $N_{\text{IP地址}}$      | 情报中恶意IP地址个数      | 4            |
| $P_{\text{longitude}}$ | 情报中一个随机IP地址或域名经度 | 27.32        |
| $P_{\text{latitude}}$  | 情报中一个随机IP地址或域名纬度 | 106.37       |
| $N_{\text{domain}}$    | 情报中恶意域名个数        | 2            |
| $N_{\text{url}}$       | 情报中恶意URL个数       | 5            |
| $N_{\text{hash}}$      | 情报中恶意Hash个数      | 3            |

黑名单库是对历史恶意IP地址、恶意域名、恶意文件Hash、恶意URL的所有记录,虽然库中大部分数据可能已失效,但是情报内容中的IP地址、域名等数据与库中的重叠个数可以说明当前情报数据的可信程度。本方法基于黑名单库的编码见表3。

表3 基于黑名单库的特征编码

| 符号(特征)              | 描述            | 示例 |
|---------------------|---------------|----|
| $C_{\text{IP地址}}$   | 与黑名单库IP地址重叠个数 | 3  |
| $C_{\text{domain}}$ | 与黑名单库域名重叠个数   | 5  |
| $C_{\text{url}}$    | 与黑名单库URL重叠个数  | 1  |
| $C_{\text{hash}}$   | 与黑名单库Hash重叠个数 | 2  |

## 2.3 基于深度神经网络算法的情报评估模型研究

### 2.3.1 深度神经网络算法分析

对情报数据进行特征提取后,本文采用深度

神经网络算法训练质量分类模型。深度神经网络在网络结构上由输入层、输出层和多个全连接层 3 部分组成,如图 3 所示,输入层神经元负责接收特征输入,全连接层和输出层通过功能神经元对输入进行函数处理。函数处理与局部模型和感知机相同,由一个线性函数  $z = \sum_{i=1}^m w_i x_i + b$  ( $x_i$  为上一层神经元的输入,  $w_i$  为上一层神经元与当前神经元的连接权)与一个激活函数构成。

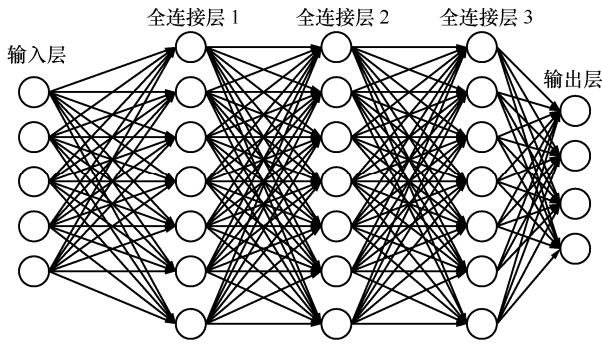


图3 深度神经网络结构示意图

为提高模型复杂度和训练准确率,深度神经网络有增加全连接层神经元数目和增加神经网络层数两种方法。但一般情况下,增加神经网络层数更加可靠有效。因为增加全连接层数不仅增加了拥有激活函数的神经元数目,还增加了激活函数嵌套层数,具有更强的特征表达和函数模拟能力。但是随着层数的增加,网络中无法使用单层的反向误差传播算法进行模型训练,因为在多层网络中误差逆传播时往往会出现梯度消失或梯度爆炸的情况,训练无法达到稳定收敛的状态。

为解决多层网络训练的问题,无监督逐层训练应运而生,其借鉴贪心算法的思想,每次只训练一层全连接层节点,训练过程中将上一层全连接层的输出作为本层输入,逐层预训练完成后,再利用 BP 等算法对网络进行有监督微调,提高模型的准确率。

### 2.3.2 基于深度神经网络算法的情报评估模型训练

本文是利用深度神经网络算法将威胁情报分为可信、不可信、无法判定三类,是一个多分类问题,具体实施流程如下。

首先,通过激活函数和随机化连接权值实现对深度神经网络的初始化,设定一个无标签输入情报数据集  $x$ ,利用第 2.2 节的特征编码方法得到编码矢量  $h_1$ 。然后将  $h_1$  作为输入,在  $[0,1]$  区间内取随机数设置权值,然后利用激活函数(式 5)得到第一层训练编码矢量  $X_1$  和对应连接权集合  $W_1$ 。然后再将第一层训练结果通过随机连接权和激活函数的方式得到第二层训练编码矢量  $X_2$  和对应的连接权集合  $W_2$ 。重复这个过程,直至得到第  $N$  层的特征向量  $X_n$  对应的连接权  $W_n$ 。

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (5)$$

然后,对深度神经网络添加分类器。通过上一步的无监督逐层网络训练完成了将输入特征映射到多维空间中,实现了对特征数据的筛选提取。但此时输出是一个多维向量,不能表征情报质量优劣。为了实现情报数据的分类,本文利用 Softmax 分类器作为深度神经网络的输出层。假设训练的数据集是  $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(n)}, y^{(n)})\}$ ,对于标签集合  $y^{(i)} \in \{1, 2, 3\}$  分别代表情报可信、不可信、无法确定 3 种类型,概率可以通过式(6)的假设函数来计算。

$$\begin{bmatrix} p(y^{(i)} = 1 | x^{(i)}; w) \\ p(y^{(i)} = 2 | x^{(i)}; w) \\ p(y^{(i)} = 3 | x^{(i)}; w) \end{bmatrix} = \frac{1}{\sum_{j=1}^3 e^{w_j^T(x^{(i)})}} \begin{bmatrix} e^{w_1^T(x^{(i)})} \\ e^{w_2^T(x^{(i)})} \\ e^{w_3^T(x^{(i)})} \end{bmatrix} \quad (6)$$

其中,  $w$  是分类器的模型参数,本文用最小化负对数似然函数作为分类器 loss, loss 的计算方法如式(7)所示。为了优化网络、提高分类精度,可以通过最小化 loss 调节顶层网络参数  $W_{n+1}$ 。



$$J_w(x_m) = -\frac{1}{M} \left[ \sum_{i=1}^n \sum_{j=1}^3 1\{y^{(i)} = j\} \ln \frac{e^{w_k^T x^{(i)}}}{\sum_{k=1}^3 e^{w_k^T x^{(i)}}} \right] \quad (7)$$

最后利用误差反向传播算法对网络进行微调。根据已标注的情报数据，利用样本标签与模型输出结果的误差，结合梯度下降优化算法进行有监督训练，网络参数 $\{W_{11}, W_{12}, \dots, W_{nm}\}$ 的更新过程如下：

$$W'_{ij} = W_{ij} - \eta \frac{\partial(J_w)}{\partial(W_{ij})} \quad (8)$$

其中， $\eta$  是神经网络的学习率， $W'_{ij}$ 是微调后更新的权值，通过反向传播算法不断微调网络权值，提高神经网络的训练精度。

### 3 情报置信模型评估验证

#### 3.1 验证实验设计

本文从天际友盟、badIPs 等 9 个威胁情报源中获取部分威胁情报数据，并利用开源接口(ipadjust.net、ipdb.me)富化 whois、地理位置等情报信息，然后利用第 2.1 节的方式对情报按照 STIX 格式进行数据标准化得到 2 000 条情报。之后两位具有 CISP(certified information security professional, 注册信息安全专业人员)认证的安全专家对 2 000 条情报数据进行了可信、不可信、无法确定 3 种结果的人工标记，得到可信情报 1 465 条、不可信情报 127 条、无法确定情报 408 条。

本文使用  $K$  折交叉验证的方法对模型进行验证，具体实施过程如图 4 所示，流程如下。

(1) 按照第 2.2 节的方法分别从情报来源、发布时间、情报内容、黑名单库匹配程度 4 个方面提取情报特征。由于本文中情报来源 9 个，攻击类型 15 种，所以编码后每条情报将由一个 38 维的特征向量表示。

(2) 将数据分成  $k$  份(本文测试  $k$  取 5)，每

次从中取一份作为测试集，其余  $k-1$  份为训练集，共进行  $k$  次，分别计算每次测试集的准确率和召回率。

(3) 为了预防后期网络陷入局部最优，采用无监督逐层贪婪预训练初始化网络参数，然后计算 Softmax 分类器 loss 值，采用反向误差传播算法更新网络权值，使得分类效果更优。

(4) 将测试集数据通过全连接层映射到多维可分空间中，并用 Softmax 计算分类结果。

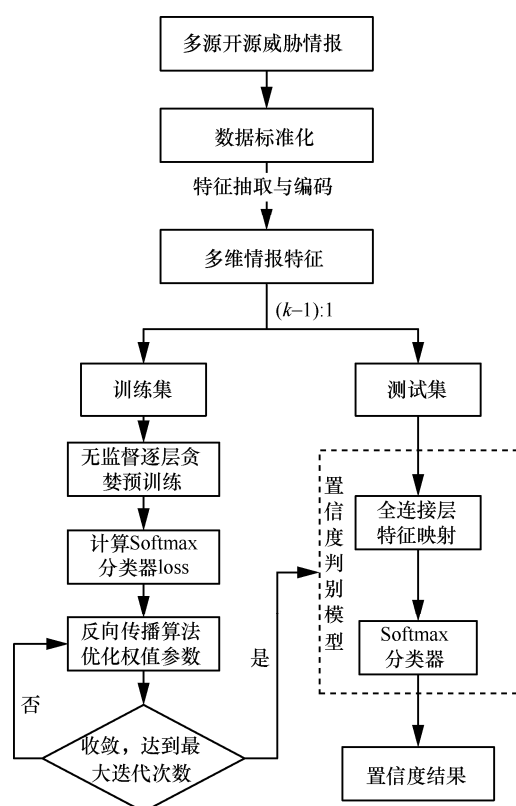


图4 情报质量评估模型验证流程

#### 3.2 实验结果分析

由于本文探讨的是一个情报多分类问题，对测试集中的情报数据，分别计算宏查准率(macroP)、宏查全率(macroR)和 macroF1 值作为评价算法的优劣的依据。macroP、macroR 和 macroF1 值的定义如式(9)~式(11)所示，其中  $P_i$ 、 $R_i$  分别代表每一类的查准率和查全率。本文采用  $K$  折交叉验证方法将情报均匀分成 5 份，训练集和测试集比例为

4:1, 实验将本文算法与 SVM、决策树算法进行对比, 对比结果见表 4。

$$\text{macro}P = \frac{1}{n} \sum_{i=1}^n P_i \quad (9)$$

$$\text{macro}R = \frac{1}{n} \sum_{i=1}^n R_i \quad (10)$$

$$\text{macro}F1 = \frac{2 \times \text{macro}P \times \text{macro}R}{\text{macro}P + \text{macro}R} \quad (11)$$

表 4 各种方法分类效果汇总

| 算法    | macroP | macroR | macroF1 |
|-------|--------|--------|---------|
| 本文算法  | 91.37% | 84.89% | 88.01%  |
| SVM   | 85.30% | 80.26% | 82.74%  |
| 决策树算法 | 75.71% | 68.34% | 71.83%  |

此次样本采用本文检测算法的宏查准率为 91.37%, 宏查全率为 84.89%, 综合评价为 88.01%。对照实验中的 SVM 算法与决策树算法采用的是 Sklean 库中自带的 SupportVectorsClassification 分类器和 DecisionTreeClassifier 分类器, 与本文提出的基于深度神经网络的情报质量评价方法比较, 本文算法在查准率和查全率均有较大提高, 说明本文提出的算法在情报数据质量评估能提供一定的参考。

## 4 结束语

本文提出了一种基于机器学习的多源威胁情报质量评价方法。首先通过情报标准化和相似度匹配算法对海量情报数据进行标准化预处理; 然后分别从发布时间、来源、情报内容、黑名单库匹配程度 4 个方面提取情报特征, 并利用 onehot 等特征编码方式对情报进行向量化; 最后结合深度神经网络和 Softmax 分类器训练优化情报评价模型。经过对天际友盟、badIPs 等 9 个威胁情报源收集到数据的实验论证, 本文提出的方法能对情报质量提供较为准确的评价。

## 参考文献:

- [1] DANDURAND L, SERRANO O S. Towards improved cyber security information sharing[Z]. 2013.
- [2] 刘春年, 张凌宇. 应急信息可信度研究范式的三维阐释与构建——基于工程化思维与 WSR 方法论[J]. 现代情报, 2017, 37(6): 24-30.
- [3] LIU C N, ZHANG L Y. Three dimensional interpretation and construction of emergency information credibility research paradigm\_based on engineering thinking and WSR methodology[J]. Journal of Modern Information, 2017, 37(6): 24-30.
- [4] 贺雅琪. 多源异构数据融合关键技术研究及其应用[D]. 成都: 电子科技大学, 2018.
- [5] HE Y Q. Research and applications on the key technology of multi-source heterogeneous data fusion[D]. Chengdu: University of Electronic Science and Technology of China, 2018.
- [6] 徐留杰, 翟江涛, 杨康, 等. 一种多源网络安全威胁情报采集与封装技术[J]. 网络安全技术与应用, 2018(10): 23-26.
- [7] XU L J, ZHAI J T, YANG K, et al. A multi-source network security threat information collection and packaging technology[J]. Network Security Technology & Application, 2018(10): 23-26.
- [8] MOHAISEN A, AL-IBRAHIM O, KAMHOUA C, et al. Rethinking information sharing for threat intelligence[Z]. 2017.
- [9] STHONNARD O, DACIER M. Actionable knowledge discovery for threats intelligence support using a multi-dimensional data mining methodology[Z]. 2008.
- [10] QAMAR S, ANWAR Z, RAHMAN M A, et al. Data-driven analytics for cyber-threat intelligence and information sharing[J]. Computers & Security, 2017(67): 35-58.
- [11] 侯艳芳, 王锦华. 基于自更新威胁情报库的大数据安全分析方法[J]. 电信科学, 2018, 34(3): 56-64.
- [12] HOU Y F, WANG J H. Big data security analysis method based on self-update threat intelligence database[J]. Telecommunications Science, 2018, 34(3): 56-64.
- [13] 李蕾. 网络空间中威胁情报可信度多维度分析模型研究[D]. 北京: 北京邮电大学, 2018.
- [14] LI L. Study on the multi-dimensional analysis model of threat intelligence credibility in cyberspace[D]. Beijing: Beijing University of Posts and Telecommunications, 2018.
- [15] BOU-HARB E, LUCIA W, FORTI N, et al. Cyber meets control: a novel federated approach for resilient cps leveraging real cyber threat intelligence[J]. IEEE Communications Magazine, 2017: 2-8.
- [16] SERKETZIS N, KATOS V, ILIOUDIS C, et al. Actionable threat intelligence for digital forensics readiness[J]. Information & Computer Security, 2019, 27(2).
- [17] 方滨兴. 定义网络空间安全[J]. 网络与信息安全学报, 2018, 4(1): 1-5.
- [18] FANG B X. Define cyberspace security[J]. Chinese Journal of Network and Information Security, 2018, 4(1): 1-5.



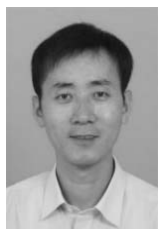
[作者简介]



**刘汉生**（1993- ），男，中国电信股份有限公司北京研究院新兴信息技术研究所网络AI研究中心工程师，主要研究方向为人工智能、威胁情报、网络智能化运维等。



**牛剑锋**（1993- ），男，中国电信股份有限公司上海研究院云安全研究所开发总监，主要研究方向为云安全、威胁情报等。



**唐洪玉**（1977- ），男，中国电信股份有限公司上海研究院云安全研究所副所长，主要研究方向为云安全、态势感知、威胁情报。



**李天博**（1988- ），男，中国电信股份有限公司上海研究院云安全研究所产品运维经理，主要研究方向为 Web 安全。



**薄明霞**（1978- ），女，博士，中国电信股份有限公司上海研究院云安全研究所高级工程师，主要研究方向为威胁情报、软件定义安全等。



**李玲晓**（1991- ），女，中国电信股份有限公司上海研究院云安全研究所工程师，主要研究方向为网站安全。