

基于标签识别的药品追溯算法设计与研究

霍 珊, 刘大维, 徐国成*

(长春中医药大学, 吉林 长春 130118)

摘要: 药品追溯过程中标签信息易出现擦碰导致模糊, 对追溯精准性造成影响, 提出一种基于标签识别的药品追溯算法。在 BP 神经网络的基础上对药品标签图像使用滴水算法进行分割, 运用 K-L 转换提取字符特征, 将自适应调节学习率及动态调整 S 型激励函数相结合对药品信息精确识别; 使用遗传算法按照种群特征来动态转换交叉概率和变异概率数值, 利用海明距离测度和适应度距离相融合手段, 保证药品信息的实时跟踪及有效查询。仿真结果表明, 吞吐量达到峰值 50tps, 液体几乎不会对标签的读取率产生影响, 药品追溯延迟时间在 10s 以下, 可实时追踪药品流通供应链信息, 有效解决药品使用安全问题。

关键词: 标签识别; 药品追溯; 神经网络; 遗传算法; 滴水算法

中图分类号: TP723 **文献标识码:** B

Design and Research of Drug Traceability Algorithm Based on Label Recognition

HUO Shan, LIU Da-wei, XU Guo-cheng*

(Changchun University of Chinese Medicine, Changchun Jilin 130118, China)

ABSTRACT: During the drug traceability, the label information is easy to be blurred, influencing the accuracy of traceability. Therefore, a drug traceability algorithm based on label identification was proposed. On the basis of BP neural network, the drop-fall algorithm was used to segment the drug label image, and K-L transformation was used to extract the features of characters. Moreover, the adaptive learning rate was combined with dynamic adjustment of S-incentive function to accurately identify drug information. According to population characteristics, the genetic algorithm was used to dynamically convert the crossover probability and mutation probability. Finally, Hemingway distance measure was integrated with fitness distance to ensure the real-time tracking and effective query of drug information. Simulation results show that the throughput reaches the peak value (50tps), and the liquid hardly affects the reading rate of label. The delay time of drug traceability is less than 10s, so that we can track the information of pharmaceutical distribution supply chain in real time and effectively solve the problem about safety in drug use.

KEYWORDS: Tag identification; Drug tracing; Neural network; Genetic algorithm; Drop-fall algorithm

1 引言

伴随经济全球化的迅猛发展, 药品供应链日益复杂, 安全问题也同样遭受巨大考验^[1]。据权威调查显示, 假冒伪劣药品销售量占全世界药品总销售的 10%。如果药品监管不当, 将会给人们身体健康带来严重威胁。由于部分药品生产厂商的制药技术水准不高, 各种恶性竞争频繁出现, 致使假

药贩卖现象不断发生^[2]。所以, 增强药品监督管理力度, 注重假冒伪劣药品的清除, 对于保障民众的用药安全具有重要意义。

由此, 本文提出一种基于标签识别的药品追溯算法。首先使用滴水算法将药品包装上的字符采取分割, 把分类后的字符信息使用 K-L 变换并提取最优特征, 以此准确识别出药品包装上的关键信息; 其次设计药品追溯管理系统, 将系统划分为硬件层、服务层、数据层及应用层, 辨别伪劣药品信息, 实现药品数据的在线交互; 最后运用遗传算法完成药品追溯, 让消费者可以准确查询到药品的流通信息, 保证自身

基金项目: 中国图书馆学会 2018 年阅读推广课题(YD2018B22)

收稿日期: 2020-02-18 修回日期: 2020-04-02

用药安全。与传统方法相比,本文方法的药品追溯效率更高,最大限度保证信息查询的及时性与有效性,可广泛应用于真实场景中。

2 基于 BP 神经网络的药品标签识别方法

本文将相机内拍摄的彩色药品包装图像转化为灰色图像,并利用滴水算法对标签字符进行分割^[3]。假设需要分割的包装图像是 P , 图像高度是 N , 宽度是 M , 构建一个坐标图, 将左上方作为原点 O , 把图像进行二值化处理, 这时分割路径 T 是由目前像素值之和及其临近的像素值决定的, 可将其描述为

$$T(x_{i+1}, y_{i+1}) = f(x_i, y_i, D_i) \quad i = 0, 1, \dots \quad (1)$$

其中 (x_{i+1}, y_{i+1}) 为目前像素坐标, (x_i, y_i) 为分割路径的初始点, D_i 为移动方向, 其数值取决于邻近像素 n_1, n_2, n_3, n_4, n_5 。具体表达式为

$$D_i = \begin{cases} 4 & \text{if } \sum_{j=1}^5 (1 - Z_j) D_j = 0 \text{ or } 15 \\ \max_{j=1}^5 (1 - Z_j) w_j & \text{else} \end{cases} \quad (2)$$

分割后形成的单字符图像通常尺寸各不相同。想要更加直观地提取药品包装内的有效信息, 就要利用双线性插值将图像采取归一化^[4], 把字符尺寸均设定为统一数值。再按照字符特点将其划分为 3 组, 实现字符的粗分类。

将获得的各个子类内的文字符号利用 K-L 转换完成特征提取, 再将各个类中的子类特征依次传输至 BP 神经网络中训练, 借此让药品包装上的文字符号性质的提取过程更加简便, 减少识别时间, 让识别准确率得到显著提升^[5]。具体的实现过程为:

首先找到正交矩阵 A , 令变换后的信号所对照的协方差矩阵为对角矩阵。针对平均值是 μ_x 的稳固随机矢量, 可将 C_x 协方差矩阵描述为

$$C_x = E[(x - \mu_x)(x - \mu_x)^T] \quad (3)$$

K-L 转换的详细过程为: 利用 λ 的 N 阶多项式得到矩阵 C_x 的特征值 λ_0 ; 因为 $C_x A_i = \lambda_i A_i$, 所以能够得到矩阵 C_x 的 N 个特征矢量 A_0 , 将特征矢量进行归一化处理, 让 $\langle A_i, A_i \rangle$ 组成矩阵 A , 完成信号 K-L 转换。

BP 神经网络是一种利用 BP 方法实现偏差纠正的多层前馈神经网络。在传统 BP 神经网络中, 权值的修正解析式为

$$\Delta w = -\eta \frac{\partial L}{\partial w} \quad (4)$$

式中 L 表示全局偏差函数。根据式(4)可以看出, 因为学习率 η 为固定数值, 继而对算法的平稳性造成一定的影响。如果学习率较大, 会致使网络产生振荡^[6]; 若学习率较小, 虽然确保了训练的平稳收敛性能, 但训练耗时过多, 收敛速率缓慢。因此本文使用自适应调整学习率算法, 其核心思想是在偏差缩小时, 增加学习率; 在偏差增多时, 降低学习率。其具

体表达式为

$$\eta_t = \eta_{t-1}(1 - \alpha \Delta L) \quad (5)$$

其中 η_t, η_{t-1} 依次表示学习率迭代的第 $t, t-1$ 次的值, ΔL 是偏差增量, $0 < \alpha < 1$ 是一个常量。

由式(5)可知, 在 $\Delta L > 0$ 的情况下, 偏差增多, 权值修正较多, 要降低 Δw 。如果减少 η , 那么 Δw 的值也随之降低, 反之 $\Delta L < 0$ 时, η 值会增大。通过此种手段, 可以在确保增加学习收敛速率的过程中, 不会产生因为 ΔL 过大让网络产生振荡, 维护了方法的平稳性。

因为传统 BP 神经网络算法通常选择 S 函数, 其导数是 $f(1-f)$ 。假设神经元的净输入为一个极端值, 那么输出结果将位于饱和位置范围, 导致得到的结果偏差较大, 此时获得的偏差曲线极有可能呈现不约束状态。由此, 本文将偏差元素 $f(1-f)$ 均扩大 $-\ln[f(1-f)]$ 倍, 以此对输出层及隐含层的偏差信号全部给予新值, 防止神经网络训练产生饱和状态。

把各个子类获得的特征维数当作输入层的神经元数目^[7], 将待识别的类别数量当作网络输出层神经元数目, 将隐含层节点数值表示成

$$s = \sqrt{q \times (v+1) + 1} \quad (6)$$

式中 q 是输入节点个数, v 是输出节点个数。

通过上述方法, 可以快速准确地识别出药品标签内的重要信息, 为药品追溯算法的平稳运行提供重要保证。

3 基于标签识别的药品追溯算法实现

利用基于遗传算法的药品追溯算法, 可以对药品标签信息实现的实时跟踪, 让消费者明确药品的真实来源, 保障其合法权益。

3.1 基于遗传算法的药品二维码识别

在本文方法中, 首先要构建基于遗传算法的药品二维码识别模型, 并对原药品二维码识别种群进行交叉及变异, 以此构成新的种群。在模型中, 交叉概率 P_c 和变异概率 P_m 的值均为程序初始值。针对变异概率 P_m 而言, 若 P_m 较小, 表明搜寻过程的速率较慢, 反之, 若 P_m 较大, 则新个体的生成速度较快, 但对种群内原本具备较优个体会构成一定损害^[9]。

因此, 提出一种可以自适应按照种群特点在算法计算中不定时得到 P_c 和 P_m 的遗传算法, 将自适应调整过程描述为

$$P_c = \begin{cases} P_{c1} - \frac{P_{c1} - P_{c2}}{f_{\max} - f_{\text{avg}}}(f' - f_{\text{avg}}), & f' \geq f_{\text{avg}} \\ P_{c1}, & f' < f_{\text{avg}} \end{cases} \quad (7)$$

$$P_m = \begin{cases} P_{m1} - \frac{P_{m1} - P_{m2}}{f_{\max} - f_{\text{avg}}}(f_{\max} - f), & f \geq f_{\text{avg}} \\ P_{m1}, & f < f_{\text{avg}} \end{cases} \quad (8)$$

其中 f_{avg} 表示每代群体的适应度均值, f_{\max} 表示群体内最大的适应度值, f 表示变异个体的适应度值, f' 表示交叉的适应度

值内最大个体。

从式(7)和式(8)中可以看出,在优化的过程中,如果种群的适应度较为分散时,公式就会自动减少 P_c 与 P_m 的值;如果群体的适应度值逐渐相同或接近局部最优时,则公式就会自动增加 P_c 与 P_m 值,以此对群体质量进行更新。

假设个体适应值不大于 f_{avg} ,则此种状况下的 P_c 、 P_m 值就相对较大,若个体的适应值均大于 f_{avg} 时,则该个体仅能得到较小的 P_c 、 P_m 值,以此保证该个体能够留存于新一代种群内。

适应能力共享函数可以精准分辨出搜寻空间在地理内的多个不同峰值,任意峰值均拥有一定比率数量的个体,比率大小与峰值高度间具备相当大的关联^[10]。想要实现此种分布形态,共享函数就要有效减少个体的目标适应度。使用 $niche$ 计数 m_i 除以适应度值 f_i 就会获得新的共享函数, $niche$ 计数 m_i 也能用于表示个体邻域的密集程度,具体表达为

$$m_i = \sum_{j \in Pop} sh[d[i, j]] \quad (9)$$

式中 $d[i, j]$ 为个体 i 、 j 的间距, $sh[d]$ 为共享函数, $sh[0] = 1$, $sh[d \geq \sigma_{share}] = 0$ 。

式(10)为一个典型的三角共享函数

$$sh(d) = \begin{cases} 1 - \frac{d}{\sigma_{share}}, & d \leq \sigma_{share} \\ 0, & d > \sigma_{share} \end{cases} \quad (10)$$

式中 σ_{share} 表示 $niche$ 的半径 r ,可根据当前状况对半径采取初始化设置,让它变成两个峰值个体间的最小距离。因为个体之间具备相近的 $niche$ 值,所以可在一个固定的 $niche$ 区域内对峰值进行约束,这样有助于全部种群的深层次收敛^[11]。

使用海明距离测量和适应度距离相融合的方法来表示 $niche$ 。若随机两个个体 x_i 和 x_j 之间的海明距离为 $d_1(x_i, x_j)$, 适应度距离为 $d_2(x_i, x_j)$, 就能获得一个全新的共享函数

$$sh(x_i, x_j) = \begin{cases} 1 - \frac{d_1(x_i, x_j)}{\sigma_1}, & d_1(x_i, x_j) < \sigma_1, d_2(x_i, x_j) \geq \sigma_2 \\ 1 - \frac{d_2(x_i, x_j)}{\sigma_2}, & d_1(x_i, x_j) \geq \sigma_1, d_2(x_i, x_j) < \sigma_2 \\ 1 - \frac{d_1(x_i, x_j) d_2(x_i, x_j)}{\sigma_1 \sigma_2}, & d_1(x_i, x_j) < \sigma_1, d_2(x_i, x_j) < \sigma_2 \\ 0, & \text{else} \end{cases} \quad (11)$$

其中 σ_1 和 σ_2 均为 $niche$ 的半径。

最终可得到新共享函数的个体适应度识别函数

$$f'(x_i) = \frac{f(x_i)}{\sum_{j=1}^M sh(x_i, x_j)} \quad (12)$$

式中 $f(x_i)$ 、 $f'(x_i)$ 依次为共享前及共享后的特体适应度识别函数。

3.2 药品信息的实时跟踪及有效查询

当识别药品包装上的二维码时,首先要将确定的参变量

转换为二维码的位置坐标,即对图像实施定位,然后对数据进行提取。只要对位置进行精准定位,就可以对码元信息进行读取^[12]。

进行二维码符号识别时,将二维码内的点 P 表示为

$$P = \{p(x_1, y_1), p(x_2, y_2), \dots, p(x_N, y_N)\} \quad (13)$$

式中 $p(x_i, y_i)$ 为与原点坐标板块相对照的灰度值。同理,针对旋转 θ 角度的二维码而言,可以预先将条码改正成序列 Q

$$Q = \{p'(x'_1, y'_1), p'(x'_2, y'_2), \dots, p'(x'_N, y'_N)\} \quad (14)$$

如果使用 R 代表图像内点序列和等待识别模式的匹配率,则其中包含 n_b 个点符合 $p(x'_j, y'_j)$

$$R = \frac{n_b}{N}, R \in [0, 1] \quad (15)$$

利用遗传算法可把问题变换成在四维空间(x_c, y_c, M, θ)内获得最大匹配率 R 的问题。假设遗传算法的种群个体为 $I_k(k = 1, 2, 3, \dots)$, 则将基因型 G_k 描述为

$$G_k = (x_{ck}, y_{ck}, M_k, \theta_k) \quad (16)$$

式中 x_{ck} 、 y_{ck} 、 M_k 、 θ_k 依次为 x_c 、 y_c 、 M 、 θ 的表现型,可将 G_k 当作四维空间内的某个点的表现型,该表现型的函数解析式为

$$H_k = \{h_{k1}(x'_1, y'_1), h_{k2}(x'_2, y'_2), \dots, h_{kN}(x'_N, y'_N)\} \quad (17)$$

按照图像的真实状况就可以把表现型的选择范围和映射关联表示出来。若药品包装上二维码图像的宽度为 w ,高度为 h ,则二维码的符号边长为 u 。按照收敛条件可知 $w \geq h \geq u$ 。由于拍摄图像时就能完全容纳二维码,且二维码在图像内的位置能够随意变换,拍摄角度和距离都具有一定的变动性,所以,利用二进制编码手段可以获得总长是 19 位的染色体。

使用式(18)中的匹配率可得到遗传算法中个体的适应度,适应度函数就是像素点相同数量和匹配总数量的比值。

$$f(I_k) = \frac{\sum_{j=1}^N f(x_{jk}, y_{jk})}{N} \quad (18)$$

使用遗传算法按照种群特征来动态转换交叉概率和变异概率数值,利用海明距离测度和适应度距离相融合手段,保证药品信息的实时跟踪及有效查询。

4 实验分析

4.1 实验环境

为了验证本文方法的优越性,将实验平台搭建于 MATLAB 7.0 软件上,并将本文方法与传统方法进行实验对比。测试环境如表 1 所示。

表 1 测试环境

配置项	描述	配置项	描述
服务器类型	物理机	操作系统	Ubuntu16.04LTS
内核版本	4.4.0-116-generic	CUP(GHz)	4* 2.0
内存(GB)	3	Peer 节点数量	8
Orderer 节点数量	1		

识别标签最好在外箱上的产品批号附近粘贴,方便信息管理。同时为避免监管码太靠底边无法识别,建议保持监管码底部与箱底边保证最小 30mm 的距离。如图 1 所示。

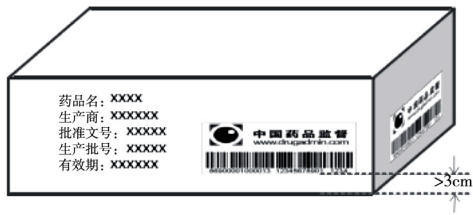


图1 药品标签贴放位置

4.2 实验结果分析

交易吞吐量 TPS,该指标反映的是区块链系统每秒能处理的交易数量,如果 TPS 每秒并发太低,很容易造成网络拥堵。方法吞吐率大小由标签数和搜索次数决定。

$$T = \frac{n}{n + m} \tag{19}$$

式中 n 表示标签数 m 表示搜索次数。

表2 吞吐量测试结果

测试轮数	吞吐量
1	27tps
2	36tps
3	46tps
4	42tps
5	50tps

由表 2 可以看出,在经过 5 轮测试后,吞吐量达到峰值 50tps,可保证整体交易成功率。

为了测量读取率,标签贴放方式如图,测量不同 DTW 距离下,标签的读取率。DTW = 0mm,表示标签直接贴附在盛满液体的容器表面;DTW = ∞,表示容器是空的,且标签直接贴附在容器壁上。



图2 标签贴放方式

当标签距离读写器天线较近时(即处在标签的稳定可读取时)随着标签与水面间距离的增大,标签的读取率都成平稳不变趋势,即液体几乎不会对标签的读取率产生影响。结果如表 3 所示。

表3 读取率结果分析

标签距离液体的距离	读取率
0	25
5mm	24.8
10mm	24.9
20mm	24.7
∞	25.1

图 3 为两种方法的药品信息追溯延迟时间对比图。从图 3 可知,本文方法的药品追溯延迟较少,保证了消费者查询药品详细数据的及时性。而传统方法药品追溯延迟时间较长,缺乏一定的有效性。

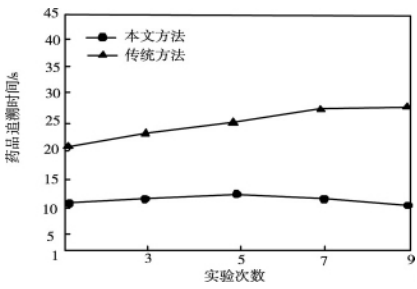


图3 药品信息追溯延迟对比图

5 结论

针对现阶段假冒伪劣药品流入市场难以分辨来源的严峻问题,本文提出一种基于标签识别的药品追溯算法。首先运用基于 BP 神经网络的药品标签识别方法,对药品包装上的关键信息进行精准识别;其次建立药品追溯管理系统,并采用遗传算法对药品流程的有效跟踪,明确药品的流通细节信息,真正实现药品规范化监管。

参考文献:

[1] 李勇. 多电子标签识别的 RFID 防碰撞方法[J]. 南京邮电大学学报(自然科学版), 2019, 39(4): 33-38.

[2] 李红伟, 陈佩斯, 梁志宏, 等. 智能滑套内低频段 RFID 标签识别率的研究[J]. 微波学报, 2018, 34(4): 48-54, 59.

[3] 冯霄婵, 杨璐瑶, 杨悦. 美国药品信息追溯与安全保障体系研究[J]. 中国药学杂志, 2019, 54(11): 934-940.

[4] 周丽金, 尚如霞, 詹长春. 社会共治视阈下药品安全监管体系优化研究[J]. 卫生经济研究, 2019, 36(6): 54-57.

[5] 魏玮. 基于无线传感网络的智慧药房药品配置仿真[J]. 计算机仿真, 2019, 36(9): 442-445.

[6] 李达, 韩冬桂, 沈程砚丹, 等. 土壤及标签贴附方式对 RFID 标签性能影响[J]. 电子技术应用, 2019, 45(6): 84-88.

[7] 郭鹏飞, 何昊葳, 付亚波, 等. 气敏类智能包装标签技术的研究进展[J]. 包装工程, 2018, 39(11): 13-18.

[8] 王可可, 陈志德, 徐健. 基于联盟区块链的农产品质量安全高效追溯体系[J]. 计算机应用, 2019, 39(8): 2438-2443.

- [9] 杜军龙,金俊平,周剑涛. 具备完整性追溯的系统数据容灾机制[J]. 计算机工程, 2019, 45(7): 170-175.
- [10] 周愉峰,李志,刘思峰. 基于随机 p-鲁棒优化的国家血液战略储备库选址-库存模型[J]. 中国管理科学, 2018, 26(10): 52-63.
- [11] 陈亮,张媛媛,刘韵婷. 基于改进的 LSTM 的药品温湿度预测方法[J]. 电子测量与仪器学报, 2019, 33(1): 106-112.
- [12] 阿仁宝力高,于晓杰,等. 医院药品验收药检单及冷链运输记录管理信息化[J]. 中国医院药学杂志, 2018, 38(11): 1222-1226.



[作者简介]

霍 珊(1981-),女(汉族),吉林长春人,硕士研究生,工程师,主要研究方向:数据分析、教育信息化。

刘大维(1980-),男(汉族),吉林长春人,硕士研究生,副研究员,主要研究方向:数据获取。

徐国成(1964-),男(汉族),吉林长春人,博士研究生,教授,主要研究方向:药物经济学。(通讯作者)

(上接第 281 页)

- [6] 闫腾飞,尚文利,赵剑明,等. 基于遗传算法优化的 OCSVM 双轮廓模型异常检测算法[J]. 计算机应用研究, 2019, 36(11): 1-6.
- [7] 张长勇,王立忠,步亚,等. 基于动力系统模型的多旋翼性能估算方法[J]. 计算机仿真, 2019, 35(5): 83-87.
- [8] 陈惠娟,冯月春,赵雪青. 利用 SSO 的自适应黑名单分组过滤器网络入侵检测方法[J]. 控制工程, 2018, 25(10): 1940-1945.
- [9] 陈冬阳,彭道刚,张浩,等. KPCA-IPSO-OCSVM 方法在工业控制系统入侵检测中的应用[J]. 中国科技论文, 2019, 14(3): 326-333.
- [10] 盛智勇,曾志强,曲洪权,等. 基于随机配置网络的光纤入侵信号识别算法[J]. 激光与光电子学进展, 2019, 56(14): 39-46.

- [11] 李春桃,李冬,齐欢. 基于 Bayes 网络的 IDS 告警分析模型研究[J]. 数学的实践与认识, 2018, 48(19): 100-104.

- [12] 夏景明,李冲,谈玲,等. 改进的随机森林分类器网络入侵检测方法[J]. 计算机工程与设计, 2019, 40(8): 2146-2150.



[作者简介]

钟小莉(1975-),女(汉族),陕西旬阳人,硕士,副教授,主要研究方向:图像处理。

谢旻旻(1976-),女(汉族),青海西宁人,硕士,副教授,研究方向:图像处理。

封希媛(1965-),女(汉族),青海西宁人,硕士,教授,研究方向:应用数学、算法。

(上接第 337 页)

显著提高了云资源调度的效率与成功率,同时提升了云资源节点抗击大量资源调度与负载均衡的能力。

参考文献:

- [1] Cao Jie, Zeng Guo-sun, Kuang Gui-juan, et al. An on-demand physical resource allocation method for cloud virtual machine to support random service requests[J]. Journal of Software, 2017, 28(2): 457-472.
- [2] 李成严,曹克翰,冯世祥,等. 不确定执行时间的云计算资源调度[J]. 哈尔滨理工大学学报, 2019, 24(1): 85-91.
- [3] Fan Wen-juan, Yang Shan-lin, Pei Jun. A novel two-stage model for cloud service trustworthiness evaluation[J]. Expert Systems, 2014, 31(2): 136-153.
- [4] Cao Jie, Zeng Guo-sun, Jiang Huo-wen, et al. Trust-aware dynamic level scheduling algorithm in cloud environment[J]. Journal on Communications, 2014, 35(11): 39-49.

- [5] 齐平,李龙澍,李学俊. 具有失效恢复机制的云资源调度算法[J]. 浙江大学学报(工学版), 2015, 49(12): 2305-2315.

- [6] Qi P, Li L S. A fault recovery-based scheduling algorithm for cloud service reliability[J]. Security and Communication Networks, 2015, 8(5): 703-714.

- [7] Zhou A, Wang S, Hsu C H, et al. Task rescheduling optimization to minimize network resource consumption[J]. Multimedia Tools and Applications, 2016, 75(20): 12901-12917.



[作者简介]

陈利军(1978-),男(汉族),河南新乡人,硕士,副教授,任职于河南经贸职业学院,主要研究方向:计算机应用。

王 畅(1975-),女(汉族),河南周口人,硕士,副教授,主要研究方向:网络安全。