

基于句法模式识别的射频指纹识别方法

陈彦君¹, 胡爱群^{1,2,3}

(1. 东南大学信息学院, 江苏 南京 210096; 2. 移动通信全国重点实验室, 江苏 南京 210096;

3. 紫金山实验室, 江苏 南京 211111)

摘要: 针对射频指纹特征提取中信道畸变与噪声干扰导致的数据维度高、分类鲁棒性差等问题, 提出基于句法模式识别的射频指纹识别方法。以 802.11 Wi-Fi 信号为对象, 通过提取射频指纹幅度谱的转折特征, 构建层次化句法模型以实现特征压缩。首先, 将幅度谱的局部形态归纳为峰、谷、上平肩、下平肩 4 类基元, 建立 3 层句法映射规则: 第 1 层提取单线段特征 (平直、上升、下降), 第 2 层组合相邻线段生成双线段特征 (如峰、谷), 第 3 层结合特征位置与电平值生成句法序列, 进一步提出子图融合算法, 通过嵌入、剪枝与拓扑重排, 融合同类设备的多样本句法模型, 最终将射频指纹特征压缩 60%~66% (从 50 维降至 30~33 维)。利用 10 个 ESP32-WROOM 32U 模组进行实验, 在信噪比 25 dB 下, 结合句法模型后, K-近邻 (K-nearest neighbor, KNN) 和朴素贝叶斯分类器的准确率分别提升 4.3% 和 5.0%, 开集识别的类内/类间错误率均低于 10.72%。实验结果表明, 所提方法在降低数据维度的同时仍能维持高判别性, 为射频指纹识别提供了一种新思路。

关键词: 射频指纹; Wi-Fi 信号; 句法模式识别; 物理层安全

中图分类号: TP393

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2025024

Radio frequency fingerprint identification method based on syntactic recognition

CHEN Yanjun¹, HU Aiqun^{1,2,3}

1. School of Information, Southeast University, Nanjing 210096, China

2. National Mobile Communication Research Laboratory, Nanjing 210096, China

3. Purple Mountain Laboratories, Nanjing 211111, China

Abstract: To address the difficulty in feature extraction due to channel characteristics and noise influence, a method based on syntactic pattern recognition was proposed. Using the 802.11 Wi-Fi signal as the object, a hierarchical syntactic model was constructed based on the folding features of the radio frequency fingerprint (RFF) amplitude spectrum to achieve feature compression. The local form of the amplitude spectrum was summarized into four

收稿日期: 2024-10-25; 修回日期: 2025-03-02

通信作者: 胡爱群, aqhu@seu.edu.cn

基金项目: 国家自然科学基金 (62171120)

Foundation Item: The National Natural Science Foundation of china(62171120)

引用格式: 陈彦君, 胡爱群. 基于句法模式识别的射频指纹识别方法[J]. 网络与信息安全学报, 2025, 11(2): 152-160.

Citation Format: CHEN Y J, HU A Q. Radio frequency fingerprint identification method based on syntactic recognition[J]. Chinese Journal of Network and Information Security, 2025, 11(2): 152-160.

elements: peak, valley, upper flat shoulder, and lower flat shoulder. Three layers of syntactic mapping rules were established. The first layer extracted single line segment features (straight, rise, fall), the second layer combined adjacent segments to generate double line segment features (e.g., peak and valley), and the third layer integrated characteristic position and level value. Additionally, a subgraph fusion algorithm was proposed. Through embedding, pruning, and topology rearrangement, various syntactic models of similar equipment were integrated. Ultimately, the RFF features were compressed by 60% to 66% (from 50 dimensions to 30-33 dimensions). Experiments using 10 ESP32-WROOM 32U modules demonstrate that with a signal-to-noise ratio of 25 dB, the accuracy of the K-nearest neighbor and naive Bayes classifiers is improved by 4.3% and 5.0%, respectively. Both the intra-class error rate and inter-class error rate of open set identification are below 10.72%. This method can reduce data dimensionality while maintaining high discriminability, offering a novel approach for RFF identification.

Keywords: radio frequency fingerprint, Wi-Fi signal, syntax recognition, physical layer safety

0 引言

无线通信网络已成为人类传输信息的主要方式，但无线网络的开放性使得通信系统时刻受到各种攻击的威胁，如假冒攻击、Sybil攻击、DDoS攻击等。为应对这些攻击，传统的加密方法有密钥分发、数字证书分发、MAC地址认证等^[1]。随着计算能力的日新月异，密钥加密总面临被更高算力机器暴力破解的风险^[2]；数字证书认证需要终端安装证书客户端，这需要终端具有较强的计算能力，不适合部署在物联网终端等小型终端上^[3]，且MAC地址极易被修改和伪造^[4]。

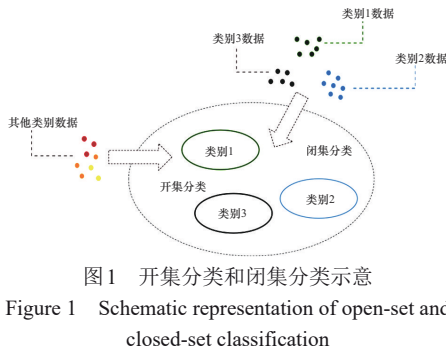
射频指纹是设备制造工艺的差异导致发射和接收的信号中自带的独一无二的设备信息，这已经被证明是一种可以有效应对超级计算机暴力破解的轻量级物理层安全技术^[5]。射频指纹从提取方法上分为瞬态指纹和稳态指纹。瞬态指纹如文献^[6]指出RFID信号接收机时钟晶振现象导致的电平持续时间在采样点上存在偏移，文献^[7]指出可以利用开关时信号的上下边沿特性提取指纹等。然而，瞬态部分的持续时间通常在纳秒或亚微秒水平^[8]，因此需要高速示波器或频谱分析仪等分析设备，所需成本较高。相比而言，提取稳态指纹对设备精度要求较低，所需成本也相应较低，所以近年来对射频指纹的研究重点逐渐偏向稳态指纹。例如，Peng等^[9]提出利用差分星座轨迹图区分Zigbee设备；Yu^[10]进一步针对Zigbee信号分析I/Q路信号频率偏移问题，提出指纹提取方法；Guo等^[11]利用LoRa信号的循环移位特性，提出一种基于卷积神经网络的射频指纹提取方案。以上方法提取的射频指纹数据维度大，且

指纹图形不规则，难以进一步提取有效特征进行压缩存储。Dong等^[12]利用最小二乘（least square, LS）信道估计获得上下行信道状态信息（channel state information, CSI），进而从802.11 Wi-Fi信号中提取射频指纹（仅包含64路复频率点），指纹数据维度相对较小，仍未解决如何进一步压缩不规则的指纹图形及提取射频指纹的有效特征的问题。

句法模式识别起源于Chomsky^[13]对于语句结构的划分，起初仅用于处理文本内容。Rao等^[14]首次将句法模式识别作为一种模式识别方法用于人体指纹识别。句法模式识别可以将指定的特征从原图像中抽象出来，但不同特征之间仍然保持类似于语法的线性联系，因此十分适合对无关信息较多的数据进一步地提取抽象模型。2013年，Pistori等^[15]首次将句法模式识别从较为简单的人体指纹图像分析向数据更为复杂的计算机视觉领域推广。在计算机视觉领域中，研究者会先利用句法模式识别对原图像抽象出句法模型，然后利用神经网络等深度学习方法^[16]或SVM等浅度学习方法^[17]对句法模型中的数据进行最终识别或决策。研究证明，经过句法模式识别处理过的图像，相比原图像的识别准确率往往会显著提高^[18]。在无线电信号处理领域，句法模式识别还被用于雷达信号识别。Visnevski等^[19]和Ou等^[20]将捕获到的多功能雷达信号片段抽象成句法模型，建立有限状态机，进而实现多功能雷达工作参数的推断；文献^[21]利用句法模式识别将多个载波工作参数相近的雷达区分开来，进而识别雷达脉冲。

在将射频指纹从设备信号中提取出来后，

需要进行射频指纹识别。射频指纹识别过程分为1对 N 的指纹分类（闭集识别/分类）问题和1对1的指纹识别（开集识别/分类）问题^[9]。对于一个分类问题，若输入数据原本所属集合均可在待匹配集合中找到，则为闭集分类问题；若待分类数据中还存在其他类别数据，则属于开集分类问题。开集分类和闭集分类示意如图1所示。射频指纹分类的实际应用场景中多为开集分类问题。



本文提出一种句法模式识别方法用以射频指纹分类，该方法在降低数据维度的同时仍能维持高差别性，为射频指纹识别提供了一种新思路。

1 句法模式识别方法

利用一对 ESP32 设备发送上下行 802.11 信号，利用 PPDU 帧结构中的长导码进行最小二乘信道估计获得 64 路上下行频域 CSI 信号，利用除法剔除多径特性后得到 50 路 802.11 Wi-Fi 射频指纹 $RFF_{1 \times 50}$ ，具体提取方法可参考文献[12]。接收机相同时，两个型号相同的 ESP32-WROOM 32U 芯片用该方法提取出射频指纹，两台设备的 RFF 曲线如图 2 所示，其中第一排是设备 1 的指纹，第二排是设备 2 的指纹。指纹体现在两个方面：一是整体趋势，设备 1 指纹中间隆起，设备 2 指纹较为平直；二是尽管高度可能不同，但同一设备的指纹总是会在指定位置的附近出现相同的峰、谷、肩等特征。因此，可以将峰、谷、肩等特征作为字母表，从射频指纹中提取句法模型，结合整体趋势分类，进而实现射频指纹的 1 对 1 识别认证。这样做可以仅考虑主要特征，排除不规则形状的影响，进一步提高稳定性。

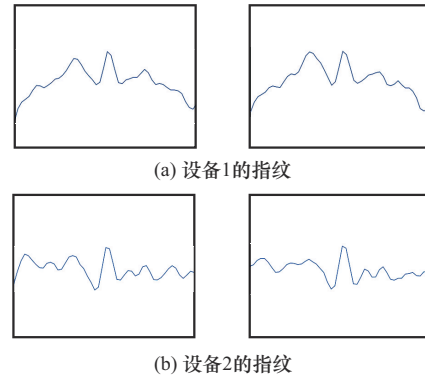


图2 两台设备的RFF曲线

Figure 2 The RFF curves from 2 devices

1.1 整体趋势定义

指纹图像有其固有的走势，可以依此来进行粗分类。利用矩形窗将原有的 RFF 分成 4 段小窗，记为 $[w_1, w_2, w_3, w_4]$ ，对各个小窗内求差分得 $[d_1, d_2, d_3, d_4]$ ，计算小窗内差分的均值，若均值大于 1，该小窗记为正，表示该小窗整体是上升趋势；若均值小于 -1，该小窗记为负，表示该小窗整体是下降趋势；若该小窗记为“0”，表示该小窗无明显上升下降趋势。如此可以获得指纹的整体趋势特征，记为 Type，整体趋势定义如表 1 所示。

表 1 整体趋势定义

Table 1 Overall trend definition

整体趋势	举例	差分描述	Type
峰型		先正后负	1
谷型		先负后正	2
平直型		非负非正，或无正负交替	3
扭曲型		多次正负交替	4

1.2 字母表生成方法

文献[13]通过标准语法定义正式语言。标准语法可定义为：

$$G = (V_N, \varepsilon, \Gamma, S) \quad (1)$$

其中， V_N 为字母表； ε 为符号，即字母的连接规则； Γ 为句法生成规则； S 为起始非终端，任何句法都以 S 起始。从信号中提取句法模型的关键是生成字母表 V_N 以及寻找句法生成规则 Γ 。

为将RFF映射至 V_N ，首先定义3层字母表：

$$\begin{aligned} V_0 &= \{a, b, c\} \\ V_1 &= \{d, e, f, g, h, r\} \\ V_2 &= \{u, v, w, x\} \end{aligned} \quad (2)$$

第一层映射定义为：

$$R_0: \text{RFF} \rightarrow [V_0, P_0] \quad (3)$$

其中， P_0 为字母表 V_0 中的元素在RFF中的对应下标。第一层映射规则如表2所示。

表2 第一层映射规则

Table 2 First-layer mapping rule

规则	描述	$[V_0, P_0]$
$\text{RFF}(i) = \text{RFF}(i+1)$	平直线段	$[a, i]$
$\text{RFF}(i) < \text{RFF}(i+1)$	上升线段	$[b, i]$
$\text{RFF}(i) > \text{RFF}(i+1)$	下降线段	$[c, i]$

单线段特征可以拼接成更复杂的双线段特征，记 $v_0, v_1 \in V_0, v_2 \in V_1$ ， p_0, p_1, p_2 分别是对应的RFF下标，则第二层映射 R_1 可描述为：

$$R_1: [(v_0, p_0); (v_1, p_1)] \rightarrow (v_2, p_2) \quad (4)$$

具体对应关系如下：

$$\begin{aligned} [(a, i); (b, i+1)] &\rightarrow (d, i+1) \\ [(a, i); (c, i+1)] &\rightarrow (e, i+1) \\ [(b, i); (a, i+1)] &\rightarrow (f, i+1) \\ [(b, i); (c, i+1)] &\rightarrow (g, i+1) \\ [(c, i); (a, i+1)] &\rightarrow (h, i+1) \\ [(c, i); (b, i+1)] &\rightarrow (r, i+1) \end{aligned} \quad (5)$$

记第三层映射为：

$$R_2: \{V_1, P_1\} \rightarrow \{V_2, P_{\text{begin}}, P_{\text{end}}\} \quad (6)$$

设 $v_2, v_3 \in V_1, v_4 \in V_2$ ， p_2, p_3 分别是 v_2, v_3 对应的RFF下标，且 $p_2 \leq p_3$ ，则映射 R_2 可描述为：

$$\begin{aligned} [(f, p_2); (e, p_3)] &\rightarrow (t, p_2, p_3) \\ [(h, p_2); (d, p_3)] &\rightarrow (u, p_2, p_3) \\ [(h, p_2); (e, p_3)] &\rightarrow (v, p_2, p_3) \\ [(f, p_2); (d, p_3)] &\rightarrow (w, p_2, p_3) \\ (g, p_2) &\rightarrow (t, p_2) \\ (r, p_2) &\rightarrow (u, p_2) \end{aligned} \quad (7)$$

为便于后续计算，给字母表 V_2 元素赋值电平 I ，字母表 V_2 定义如表3所示。

表3 字母表 V_2 定义

Table 3 Definition of alphabet V_2

V_2 元素	特征形状	I
t		+2
u		-2
v		-1
w		+1

于是 V_N 可表示为：

$$V_N = \{\text{Type}, V_2, I, P_{\text{begin}}, P_{\text{end}}\} \quad (8)$$

1.3 句法模型生成方法

设有 m 类训练样本 $\{\omega_1, \omega_2, \dots, \omega_m\}$ ，每类中含有 n 个样本，如 $\omega_i (i = 1, 2, \dots, m)$ 内的样本可记为 $L_i = \{x_{i1}, x_{i2}, \dots, x_{im}\}$ 。记第 i 类第 j 个训练样本的句法为 $G_{ij} (j = 1, 2, \dots, n)$ ，生成规则如下（由于 G_{ij} 的生成仅与 V_2 有关，这里用 V_2 代替 V_N ）：

$$\begin{aligned} \varepsilon &= \{S_0, S_1, S_2, S_3\} \\ S_0 &\rightarrow t \mid tS_1 \mid tS_2 \quad S_1 \rightarrow u \mid uS_0 \mid uS_3 \\ S_2 &\rightarrow v \mid vS_1 \mid vS_2 \quad S_3 \rightarrow w \mid wS_0 \mid wS_3 \end{aligned} \quad (9)$$

该生成规则考虑到函数的连续性，当检测到一个峰特征后，接下来仅有结束、下平肩、谷3种可能，而不会出现峰或上平肩，其他特征亦然。图2中设备1指纹的句法推断过程可以表示为：

$$\begin{aligned} x_{11}: \\ S_0 &\Rightarrow (t, i_t, p_{\text{begin}}, p_{\text{end}}) S_1 \\ &\Rightarrow (t, i_t, p_{\text{begin}}, p_{\text{end}}) (u, i_u, p_{\text{begin}}, p_{\text{end}}) S_3 \\ &\Rightarrow (t, i_t, p_{\text{begin}}, p_{\text{end}}) (u, i_u, p_{\text{begin}}, p_{\text{end}}) \\ &\quad (w, i_w, p_{\text{begin}}, p_{\text{end}}) S_0 \\ &\Rightarrow \dots \end{aligned} \quad (10)$$

在完成了单个样本的句法推断后，需要将同一类别全部训练样本的句法模型融合，得到类别 ω_i 的标准句法模型 Gr_i ，标准句法合成过程如图3所示。

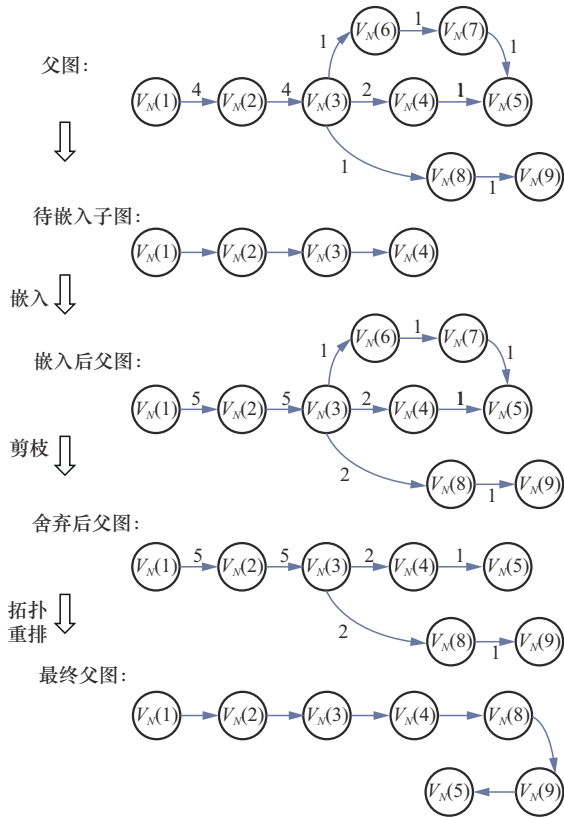


图3 标准句法合成过程

Figure 3 Procedure of standard syntactic synthesis

根据图论定义, 定义 Gr_i 为父图, 是融合后包含所有子图特征的全局结构, 定义句法集合 $\{G_{i1}, G_{i2}, \dots, G_{in}\}$ 为子图, 是单个样本的局部句法模型。通过子图嵌入、剪枝与父图节点重排, 可提取出表征同类设备的共性特征的父图, 具体步骤如算法 1 所示。

算法 1 子图融合算法

输入 句法集合 $\{G_{i1}, G_{i2}, \dots, G_{in}\}$

输出 融合后的父图 Gr_i

1) 初始化父图: 选取 G_{i1} 作为初始父图 Gr_i , 其余句法模型 $\{G_{i2}, \dots, G_{in}\}$ 视为子图集合 $G_{i,sub}$ 。

2) 子图嵌入: 遍历 $G_{i,sub}$, 若子图节点 V_i 和父图节点 V_j 的特征位置和电平值匹配 (误差阈值内), 则将 V_i 合并至父图, V_j 对应边权重+1。

3) 冗余边剪枝: 遍历 Gr_i , 若其中节点 V_j 的边权重低于阈值, 则进行剪枝, 保留高频特征路径。

4) 拓扑重排: 根据特征在射频指纹中的位置顺序, 对父节点进行头排序, 提取出标准句法模型。

假设有 m 个设备, 每个设备各 n 个 RFF 数据,

单个 RFF 数据处理后可以生成 $3 \times l$ 大小的句法模型, 模型可表示为 $G_{ij} = \{V_2, P_{begin}, P_{end}\}$, 现进行句法模型生成的算法复杂度分析。

对于单样本, 第 1 层映射: 遍历 n 维指纹数据差分计算, 时间复杂度为 $O(n)$, 空间复杂度为 $O(n)$; 第 2 层映射: 匹配相邻线段生成双线段特征, 时间复杂度为 $O(n)$, 空间复杂度为 $O\left(\frac{n}{2}\right)$; 第 3 层映射: 组合双线段生成句法模型, 时间复杂度为 $O\left(\frac{n}{2}\right)$, 空间复杂度为 $O(3l)$ 。综上, 单个样本映射过程的总时间复杂度为 $O\left(\frac{5}{2}n\right) \approx O(n)$, 空间复杂度为 $O(3l) \approx O(n)$ 。对于全体样本, 映射过程的总时间复杂度为 $O\left(\frac{5}{2}mn\right) \approx O(mn)$, 总空间复杂度为 $O(3lm) \approx O(mn)$ 。

1.4 传统分类器对句法模型的分类方法

现有来自 m 个设备各 n 个射频指纹提取出的句法模型 $\{G_{ij}\}_{m \times n}$ 与其对应所属的设备编号 $\{Label_{ij}\}_{m \times n}$, 其中 G_{ij} 可被视作一组仅含原射频指纹特征信息的向量, $Label_{ij}$ 取值为 $1 \sim n$ 中某一整数。

KNN 分类器是一种常见的机器学习算法, 可直接对提取出的射频指纹进行分类^[12]。本文首先利用句法模式识别对射频指纹数据进行处理, 生成句法模型后再利用 KNN 分类器进行分类判决。KNN 分类算法流程如图 4 所示, 训练集包含来自 m 个设备的各 n_1 个句法模型 $\{G_{ij}\}_{m \times n_1}^{train}$ 以及对应的类别标签 $\{Label_{ij}\}_{m \times n_1}^{train}$; 测试集仅包含来自 m 个设备的各 n_2 个句法模型 $\{G_{ij}\}_{m \times n_2}^{test}$, 其中:

$$\{G_{ij}\}_{m \times n} = \{G_{ij}\}_{m \times n_1}^{train} + \{G_{ij}\}_{m \times n_2}^{test} \quad (11)$$

在分类时, 首先计算测试数据与全部训练数据的距离, 在距离最小的 k 个训练样本中选择出现频率最高的对应类别作为最终分配判决的结果。

朴素贝叶斯分类器也是一种常见的可用于射频指纹分类的算法^[22], 其泛指所有基于贝叶斯公式推导出来的分类算法。

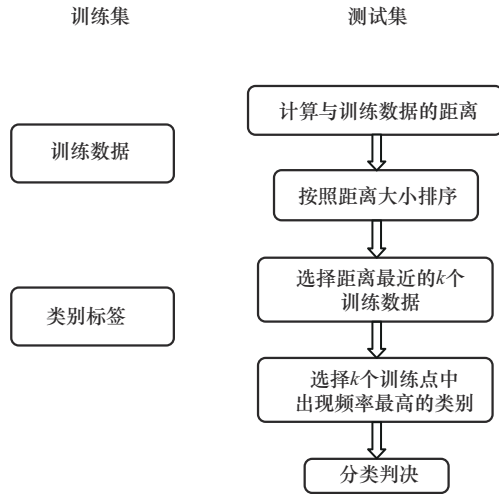


图4 KNN分类算法流程

Figure 4 The flow of KNN classification algorithm

对于测试集中的待分类数据 $\mathbf{G}_{ij}^{\text{test}}$ ，记 $P(\omega_i|\mathbf{G}_{ij}^{\text{test}})$ 为该数据点属于类别 ω_i 的概率，朴素贝叶斯分类器考虑做如下优化。

$$\text{Label}_{ij}^{\text{test}} = \arg \max_{\omega_i} P(\omega_i|\mathbf{G}_{ij}^{\text{test}}) \quad (12)$$

朴素贝叶斯分类器假设数据点 $\mathbf{G}_{ij}$ 的各个特征分量之间相互独立，于是根据贝叶斯公式可得：

$$P(\omega_i|\mathbf{G}_{ij}^{\text{test}}) = \frac{P(\omega_i) \prod_{k=1}^l P(\mathbf{G}_{ij}^{\text{test}}(k)|\omega_i)}{\prod_{k=1}^l P(\mathbf{G}_{ij}^{\text{test}}(k))} \quad (13)$$

其中，数据点 $\mathbf{G}_{ij}^{\text{test}}$ 是大小为 $1 \times l$ 的数据向量。可以发现，左式与右式分子成正比，于是贝叶斯分类器的目标函数可改写成：

$$\text{Label}_{ij}^{\text{test}} = \arg \max_{\omega_i} P(\omega_i) \prod_{k=1}^l P(\mathbf{G}_{ij}^{\text{test}}(k)|\omega_i) \quad (14)$$

其中， $P(\mathbf{G}_{ij}^{\text{test}}(k)|\omega_i)$ 表示在 ω_i 类别中一个数据向量内各分量的先验概率。不同朴素贝叶斯分类器的区别主要体现在对 $P(\omega_i)$ 的定义上，它是通过训练来给定的。

1.5 句法模式的识别方法

句法模式识别过程是对提取出的句法特征计算得分，最后根据得分门限判断是否合法。针对指纹图像，考虑总体特征和句法特征总得分计算式为：

$$\text{score}_{ij} = t_{ij} s_{ij} \quad (15)$$

其中，当样本 j 属于 ω_i 时， t_{ij} 值为 1，否则为 0；句法得分算法如算法 2 所示。

算法2 句法得分算法

输入 标准句法集合 Gr_i ，待计算指纹句法 $\mathbf{G}_{ij}$ ，检索范围 h

输出 得分 score_{ij}

1) $\text{score}_{ij} = 0$ 。

2) 遍历 Gr_i 中的 V_N ，在每个特征点附近 h 范围内寻找是否存在特征点。

3) 若相同位置下 $\mathbf{G}_{ij}$ 也存在特征点，检查 T_i 和 T_{ij} 。

if $(T_i = T_{ij})$: $\text{score}_{ij} += 5$ if $(T_i + T_{ij} = 0)$:
 $\text{score}_{ij} -= 5$

4) 将 score_{ij} 归一化。

th_i 是 ω_i 的判决门限，当 score_{ij} 超过门限时，即可判定为设备 ω_i 的合法设备。为确定各类的门限值，利用本文方法得到训练集的得分矩阵，记为：

$$\mathbf{S} = [\text{score}_{ij}]_{m \times n} \quad (16)$$

反映判决准确率的指标有误差率 (false accept rate, FAR) 和拒识率 (false rejection rate, FRR)，计算式为：

$$\begin{aligned} \text{FAR} &= \frac{\text{类内测试错误次数}}{\text{类内测试总次数}} \\ \text{FRR} &= \frac{\text{类间测试错误次数}}{\text{类间测试总次数}} \end{aligned} \quad (17)$$

分别计算不同 th_i 下各类指纹训练集的 ROC 曲线，FAR 和 FRR 曲线的交点即最终的判决门限，如图 5 所示。

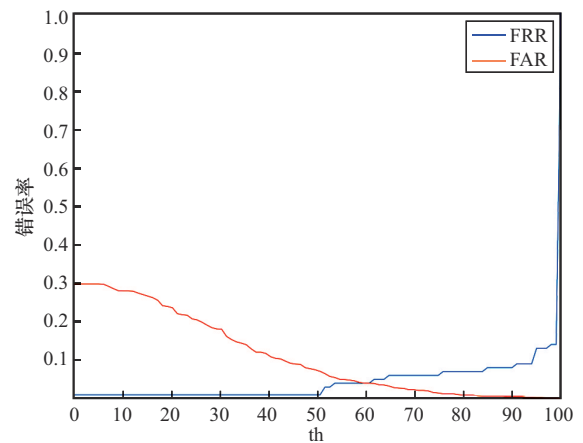


图5 某台设备的 ROC 曲线

Figure 5 The ROC curve for a certain device

2 实验验证

本节利用 1 个 ESP32-WROOM 32U 芯片作为 Wi-Fi 的 AP 端, 10 个 ESP32-WROOM 32U 芯片作为 Wi-Fi 的 STA 端, 在 SNR 为 25 dB 的静止状态下利用文献[12]所述上下行 CSI 回传方法采集 10 组 50 路载波的幅度谱作为 RFF 数据, 图 6 为某台设备的射频指纹幅度图。

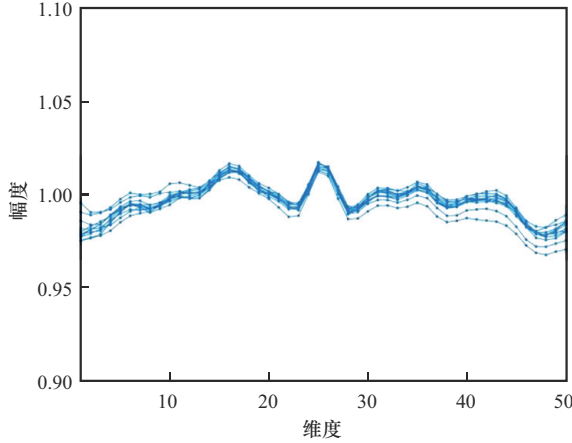


图 6 某台设备的射频指纹幅度图

Figure 6 The RFF amplitude figure of a certain device

将数据按照 3:7 分成训练集和测试集, 集合 $\omega_i (i = 1, 2, \dots, 10)$ 的训练集记为:

$$L_i = \{x_{i1}, x_{i2}, \dots, x_{iM}\} \quad (18)$$

测试集记为:

$$T_i = \{\sigma_{i1}, \sigma_{i2}, \dots, \sigma_{iN}\} \quad (19)$$

首先获得 1024×1024 的 RFF 灰度图像, 抹平肩、峰、谷的特征处出现的大量毛刺和阶梯状切分, 根据差分定位拐折点, 经过预处理后的射频指纹如图 7 所示。利用滑动窗逐一扫描过整个曲线, 利用上节所述方法分别识别整体和局部特征, 其中局部句法特征以电平的形式表现, 图 8 给出某台设备的射频指纹提取出的句法模型示意, 将原 50 维 RFF 数据用 10~11 个位于不同位置的句法特征进行描述, 而每个句法特征只记录起始点、特征宽度与电平值 3 个特征, 即将 50 维的指纹数据压缩至 30~33 维句法模型数据。除去个别样本的特征点有偏差, 样本整体的句法特征的一致性较好。

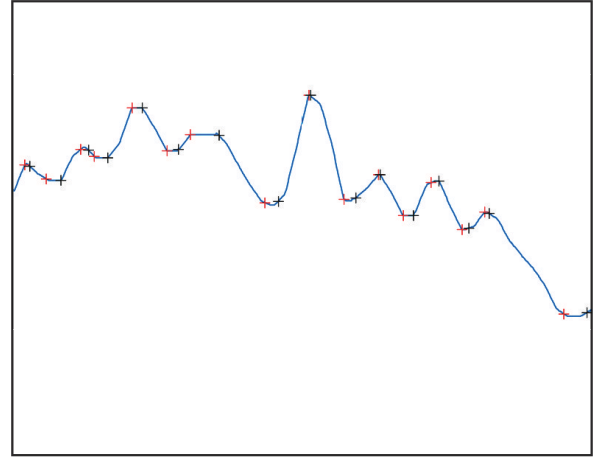


图 7 经过预处理后的射频指纹

Figure 7 The pre-processed RFF figure

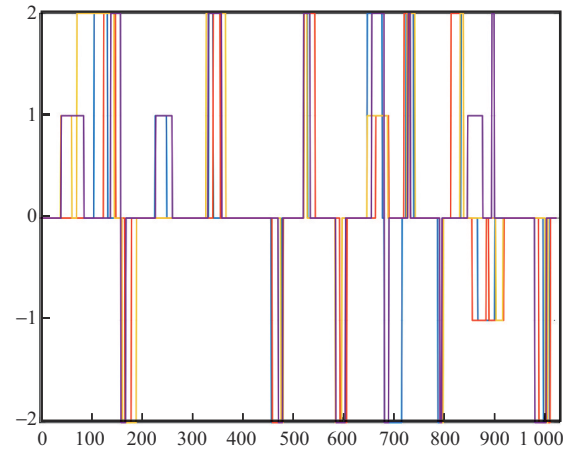


图 8 某台设备的射频指纹提取出的句法模型示意

Figure 8 Several syntactic features extracted by a certain device

下面对句法模式识别的分类准确率进行测试。

(1) 测试闭集分类准确率

针对 KNN 分类器, 首先初步比较不同 k 值下的整体分类准确率, 如图 9 所示, 整体准确率在 $k = 1$ 时最高, 因此可确定 k 值为 1; 针对朴素贝叶斯分类器, 以训练集内 ω_i 的频率作为 $P(\omega_i)$ 。表 4 分别给出单独使用 KNN 分类器、单独使用朴素贝叶斯分类器、KNN 分类器组合句法模型、朴素贝叶斯分类器组合句法模型的 10 组射频指纹分类准确率。由表 4 可知, 4 种方法均可有效区分不同设备的射频指纹, 而经过句法模式识别处理后的分类准确率分别提高 4.3% 和 5.0%。

(2) 测试开集分类准确率

在 L_i 中每组随机选取 10 个训练样本按照算

法1所述方法生成标准句法集合 G_i ，其余训练样本按照1.5节所述方法进行句法匹配，得到匹配分数，计算FRR与FAR并寻找ROC曲线交点处的th值作为各类数据的门限值，最终确定th值。

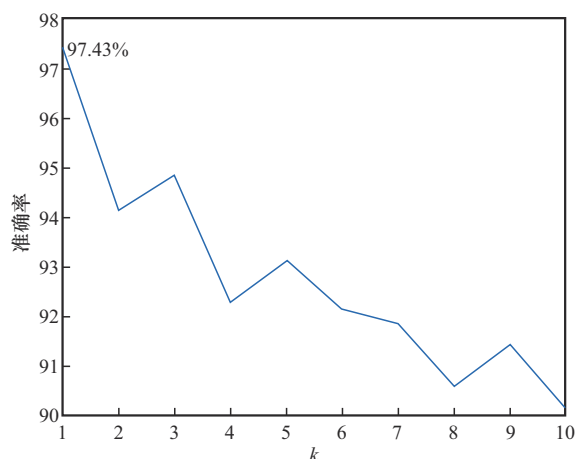


图9 k 值的选择与对应整体分类准确率

Figure 9 The k values in KNN with the corresponding overall classification accuracy

表4 4种方法对射频指纹分类的准确率

Table 4 Accuracy of 4 methods for RFF classification

类别	KNN 分类器	朴素 贝叶斯 分类器	KNN 分类器 + 句法模型	朴素贝叶斯 分类器 + 句法模型
ω_1	90.0%	90.0%	97.9%	97.9%
ω_2	100.0%	91.7%	96.9%	96.9%
ω_3	82.5%	100.0%	87.5%	90.0%
ω_4	87.5%	90.0%	95.0%	95.0%
ω_5	94.5%	99.0%	99.0%	99.0%
ω_6	88.9%	87.5%	97.9%	97.9%
ω_7	100.0%	90.0%	99.0%	98.0%
ω_8	87.5%	87.5%	92.5%	92.5%
ω_9	90.0%	80.0%	99.0%	99.0%
ω_{10}	100.0%	100.0%	99.0%	99.0%

利用测试集分别做类内测试和类间测试得到FAR和FRR，由表5可知，FRR最高为10.22%，FAR最高为10.72%。对比方法的性能如表6所示，经过句法处理后的射频指纹数据维度显著下降，但是分类准确率性能也相应有所降低。

表5 句法模式识别错误率

Table 5 Error rate of syntactic recognition

类别	th	FRR	FAR
ω_1	55	4.44%	8.67%
ω_2	80	4.44%	8.33%
ω_3	48	1.17%	6.17%
ω_4	52	6.67%	6.06%
ω_5	61	9.33%	10.72%
ω_6	75	6.67%	8.28%
ω_7	66	10.22%	8.11%
ω_8	69	7.78%	6.72%
ω_9	64	0	0.11%
ω_{10}	73	1.11%	10.39%

表6 对比方法的性能

Table 6 the performance of the Comparative methods

方法	最低分类 准确率	数据维度	计算复杂度
KNN	82.5%	50	$O(mn)$
本文(句法+KNN)	87.5%	30~33	$O(mn)$
文献[9](DCTF)	92.1%	64	$O(n^2)$
文献[11](CNN)	95.3%	128	$O(n^3)$
文献[12](LS-CSI)	89.5%	64	$O(n^2)$

3 结束语

射频指纹技术提供了一种轻量且可靠的保密方法，是未来信息安全的重要研究方向。但提取出的射频指纹图的不规律性，使得对射频指纹数据进行进一步的特征提取和压缩十分困难。因此，本文提出了一种基于句法模式识别的指纹识别方法，参考人体指纹的特征归类方法，将射频指纹图的拐折特征归纳为峰、谷、上平肩、下平肩4种，并根据在指纹图中的位置关系建立句法模型。为验证生成的句法模型的可区分性，在10台ESP32设备上进行了准确率测试，10台设备分别使用KNN分类器和朴素贝叶斯分类器对生成的模型进行分类，与仅使用KNN分类器和朴素贝叶斯分类器相比，分类准确率分别提升4.3%和5.0%，而直接进行开集分类错误率最高为10.72%。与现有其他方法相比，本文将KNN和句法识别方法结合，将数据维度降低40%以上，并且计算复杂度显著低于深度

学习方案。但是作为代价, 射频指纹分类准确率会略微降低。

参考文献:

- [1] YU J, HU A, ZHOU F, et al. Radio frequency fingerprint identification based on denoising autoencoders[J]. IEEE Transactions on Information Forensics and Security, 2019, 14(6): 1530-1542.
- [2] BASTOS D, SHACKLETON M, EL-MOUSSA F. Internet of things: a survey of technologies and security risks in smart home and city environments[C]//Living in the Internet of Things: Cybersecurity of the IoT. London, UK: Institution of Engineering and Technology, 2018.
- [3] JIAO L, WANG N, WANG P, et al. Physical layer key generation in 5G wireless networks[J]. IEEE Wireless Communications, 2019, 26(5): 48-54.
- [4] NGUYEN T, ZHENG G, HAN Z, et al. Device fingerprinting to enhance wireless security using nonparametric Bayesian method[C]//2011 Proceedings IEEE INFOCOM. 2011: 1404-1412.
- [5] PENG L, HU A, ZHANG J, et al. Design of a hybrid RF fingerprint extraction and device classification scheme[J]. IEEE Internet of Things Journal, 2019, 6(1): 349-360.
- [6] PAN Q, AN Z, ZHAO X, et al. The power of precision: high-resolution backscatter frequency drift in RFID identification[J]. IEEE Transactions on Mobile Computing, 2024, 23(8): 8370-8385.
- [7] DANEV B, ZANETTI D, CAPKUN S. On physical-layer identification of wireless devices[J]. ACM Computing Surveys, 2012, 45(1): Article 6.
- [8] YU J, HU A, LI G, et al. A robust RF fingerprinting approach using multisampling convolutional neural network[J]. IEEE Internet of Things Journal, 2019, 6(4): 6786-6799.
- [9] PENG L, HU A, JIANG Y, et al. A differential constellation trace figure based device identification method for ZigBee nodes[C]//2016 8th International Conference on Wireless Communications & Signal Processing (WCSP). 2016: 1-6.
- [10] 俞佳宝. 射频指纹提取与识别技术研究[D]. 南京: 东南大学, 2020. YU J B. Research on RF fingerprint extraction and identification technology[D]. Nanjing: Southeast University, 2020.
- [11] GUO Y, HU A, JIANG Y. Deep learning based RF fingerprint identification using cyclic shift characteristic[C]//2024 4th International Conference on Consumer Electronics and Computer Engineering (ICCECE). 2024: 369-374.
- [12] DONG B, HU A, YU J, et al. A robust radio frequency fingerprint extraction method based on channel reciprocity[C]//2024 IEEE Wireless Communications and Networking Conference (WCNC). 2024: 1-6.
- [13] CHOMSKY N. Three models for the description of language[J]. IEEE Transactions on Information Theory, 1956, 2(3): 113-124.
- [14] RAO K, BALCK K. Type classification of fingerprints: a syntactic approach[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1980, PAMI-2(3): 223-231.
- [15] PISTORI H, CALWAY A, FLACH P. A new strategy for applying grammatical inference to image classification problems[C]//2013 IEEE International Conference on Industrial Technology (ICIT). 2013: 1032-1037.
- [16] LI X, SONG X, WU T. AOGNets: compositional grammatical architectures for deep learning[EB].
- [17] ZARCHI M S, TAN R T, VAN-GEMEREN C, et al. Understanding image concepts using ISTOP model[J]. Pattern Recognition, 2016, 53: 174-183.
- [18] ASTOLFI G, REZENDE F, PORTO J, et al. Syntactic pattern recognition in computer vision: A systematic review[J]. ACM Computing Surveys, 2021, 54(3): Article 65.
- [19] VISNEVSKI N, KRISHNAMURTHY V, WANG A, et al. Syntactic modeling and signal processing of multifunction radars: a stochastic context-free grammar approach[J]. Proceedings of the IEEE, 2007, 95(5): 1000-1025.
- [20] OU J, CHEN Y, ZHAO F, et al. Novel method for radar word extraction in the syntactic model of multi-function radar[C]//2017 18th International Radar Symposium (IRS). 2017: 1-7.
- [21] 李莹, 陈艺廷, 李鸿. 基于句法模式识别的雷达识别方法研究[J]. 现代电子技术, 2014. LI Y, CHEN Y T, LI H. Research on radar recognition method based on syntactic pattern recognition[J]. Modern Electronic Technology, 2014, 37: 25-28.
- [22] 陈天舒. Wi-Fi 射频指纹提取与识别技术研究[D]. 南京: 东南大学, 2021. CHEN T S. Wi-Fi radio frequency fingerprint extraction and identification technology[D]. Nanjing: Southeast University, 2021.

作者简介



陈彦君 (2000-), 男, 江西九江人, 东南大学硕士生, 主要研究方向为物理层安全领域的射频指纹技术。



胡爱群 (1963-), 男, 江苏如皋人, 东南大学教授、博士生导师, 主要研究方向为无线物理层安全和移动通信内生安全理论与技术。