第 45 卷 第 11 期 计 算 机 工 程 2019 年 11 月
Vol. 45 No. 11 Computer Engineering November 2019

·安全技术· 文章编号：1000-3428(2019)11-0144-08 文献标志码：A 中图分类号：TP309

功耗分析攻击中机器学习模型选择研究

段晓毅，陈 东，高献伟，范晓红，周玉坤

（北京电子科技学院 电子信息工程系，北京 100070）

摘 要： 根据密码芯片功耗曲线的特性，对支持向量机、随机森林、K 最近邻、朴素贝叶斯 4 种机器学习算法进行分析研究，从中选择用于功耗分析攻击的最优算法。对于机器学习算法的数据选取问题，使用多组数量相同但组成元素不同的数据集的十折交叉验证结果进行模型选择，提高测试公平性及测试结果的泛化能力。为避免十折交叉验证过程中出现测试集误差不足以近似泛化误差的问题，采用 Friedman 检验及 Nemenyi 后续检验相结合的方法对 4 种机器学习算法进行评估，结果表明支持向量机是适用于功耗分析攻击的最优机器学习算法。

关键词： 机器学习；十折交叉验证；Friedman 检验；Nemenyi 后续检验；功耗分析攻击

开放科学（资源服务）标志码（OSID）：

中文引用格式： 段晓毅，陈东，高献伟，等. 功耗分析攻击中机器学习模型选择研究[J]. 计算机工程，2019，45(11)：144-151，158.

英文引用格式： DUAN Xiaoyi，CHEN Dong，GAO Xianwei，et al. Research on model selection of machine learning in power analysis attack[J]. Computer Engineering，2019，45(11)：144-151，158.

Research on Model Selection of Machine Learning in Power Analysis Attack

DUAN Xiaoyi，CHEN Dong，GAO Xianwei，FAN Xiaohong，ZHOU Yukun

（Department of Electronics and Information Engineering，Beijing Electronic Science and Technology Institute，Beijing 100070，China）

【Abstract】 According to the characteristics of the power curve of crypto chips，four kinds of machine learning algorithms including Support Vector Machine（SVM），Random Forest（RF），K-Nearest Neighbor（KNN） and Naive Bayes（NB） are analyzed and studied，and the optimal algorithm for power analysis attack is selected. For the data selection problem of the machine learning algorithm，the 10-fold cross-validation results of multiple sets of data sets with the same number but different constituent elements are used to select the model，which improves the test fairness and the generalization ability of the test results. In order to avoid the problem that the test set error is not enough to approximate the generalization error during the 10-fold cross-validation process，the four kinds of machine learning algorithms are evaluated by the combination of Friedman test and Nemenyi post-hoc test. The results show that the SVM is the optimal machine learning algorithm for power analysis attack.

【Key words】 machine learning；10-fold cross-validation；Friedman test；Nemenyi post-hoc test；power analysis attack

DOI： 10.19678/j.issn.1000-3428.0052538

0 概述

传统功耗分析攻击中的模板攻击是将攻击问题看作分类问题，而机器学习领域的各类算法也与模板攻击相似，同为将问题看作分类任务，这使得研究者将机器学习与功耗分析相结合实施更有效的攻击，通过使用机器学习能够在提高功耗分析攻击效率的同时降低攻击所需使用的功耗曲线条数。文献[1]将机器学习技术用于侧信道攻击中的功耗分析攻击，使用具有明显汉明重量泄漏的数据集，利用最小二乘支持向量机（Least Squares Support Vector Machine，LS-SVM），成功对高级密码标准（Advanced Encryption Standard，AES）的部分软件实施了攻击，并指出传统模板攻击其实是一个多分类问题，SVM 参数设置对分类性能有较大影响，但是功耗曲线的数量和抽样点的数量对结果的影响较小。文献[2]利用支持向量机（Support Vector Machine，SVM）分类算法对 8 位智能卡上运行的 DES 算法进行攻

基金项目： 国家自然科学基金（61701008）；中央高校基本科研业务费专项资金（2017LG05）。

作者简介： 段晓毅（1979—），男，讲师、博士，主研方向为信息安全；陈 东，硕士研究生；高献伟，教授；范晓红，讲师；周玉坤，副教授。

收稿日期： 2018-09-03 **修回日期：** 2018-11-01 **E-mail：** duanxiaoyi@besti.edu.cn

击。文献[3]基于汉明重量模型,利用带概率的多分类 SVM 将密钥分为 9 类。实验结果表明,在强噪声环境下,要达到同样的猜测熵时,SVM 攻击所需训练功耗曲线的数量需少于模板攻击,因此 SVM 攻击更具一般性。文献[4]利用无监督的机器学习方法,结合平均痕迹和主成分分析(Principal Component Analysis,PCA)提出一种可容忍分析和攻击痕迹之间存在一定差异的分类识别器。

但上述研究均仅针对 1 种算法实施功耗分析攻击,文献[5]对 SVM 和随机森林(Random Forest,RF)2 种算法的准确度进行对比,得出 SVM 优于 RF。本文对 SVM、RF、K 最近邻(K-Nearest Neighbor,KNN)、朴素贝叶斯(Naive Bayes,NB)算法进行分析研究,选择适合功耗攻击的最优算法。对于机器学习的数据选择问题,由于单一数据集一次测试的正确率具有偶然性,因此本文使用十折交叉验证的平均值代替一次测试的准确率,以确保测试的公平性。为避免一组数据的一次十折交叉验证的评估过程中存在测试集误差不足以近似泛化误差的问题,使用多组数量相同但组成元素不同的数据集的十折交叉验证结果进行模型选择,并采用 Friedman 检验以及 Nemenyi 后续检验对以上算法进行评估,从而选出更适合功耗分析攻击的机器学习算法。

1　机器学习算法

1.1　朴素贝叶斯分类器

朴素贝叶斯是基于贝叶斯公式的一种分类器,为避开基于贝叶斯公式估计后验概率时难以从有限样本中直接估计得到类条件概率 $P(x|c)$ 的问题,朴素贝叶斯采用“属性条件独立性假设”:对已知类别,假设所有属性相互独立。

基于该假设,贝叶斯公式可以重写为[6]:

$$P(c|x)=\frac{P(c)P(x|c)}{P(x)}=\frac{P(c)}{P(x)}\prod_{i=1}^{d}P(x_i|c) \tag{1}$$

其中,$P(c)$ 是先验概率,$P(x|c)$ 是样本 x 相对于类标记 c 的条件概率,$P(x)$ 是证据因子,用来对各类的后验概率之和进行归一化处理,d 为属性数目,x_i 为 x 在第 i 个属性上的取值。

由于 $P(x)$ 对于所有类别均相同,因此基于式(1)的朴素贝叶斯分类器的表达式为:

$$h_{\mathrm{nb}}(x)=\underset{c\in y}{\operatorname{argmax}}\ P(c)\prod_{i=1}^{d}P(x_i|c) \tag{2}$$

通过式(2)计算出最大的条件概率,预测未知样本类别。该算法源于古典数学理论,具有稳定的分类效率,适用于小规模数据,并且对数据缺失不敏感。

根据朴素贝叶斯的特点,被分类的数据要求有多个属性,根据数据携带的不同属性对其进行分类。在功耗分析攻击中,通过采集密码算法工作时的电压对其进行数据分析,每个电压都是它的一个属性,与朴素贝叶斯的要求一致,所以可将该算法运用于本文实验中。但由于密码算法工作时电压与电压之间不完全独立,而朴素贝叶斯的前提是假设独立,因此分类结果在一定程度上会受到影响。

1.2　K 最近邻算法

K 最近邻算法[7]是根据测试集最邻近的 K 个训练数据对测试集数据进行分类的算法。若这 K 个训练集数据的多数属于某个类,则将该测试数据分到这个类中。因此在使用 KNN 算法时 K 个“邻居”的选择会影响到分类结果。本文实验中使用欧式距离作为评估准则,选取距离最近的 K 个点对功耗数据进行分类。n 维欧式距离的计算公式如式(3)所示。

$$\rho(A,B)=\sqrt{[\sum(a[i]-b[i])^2]},i=1,2,\cdots,n \tag{3}$$

计算测试集数据与训练集所有数据的欧式距离,选取距离最小的 K 个数据,统计 K 个数据所在类别出现的频率,将频率最高的类别作为当前测试数据的预测分类。

KNN 算法优势在于其不需要进行数据训练,直接根据待测数据与已知数据的距离关系确定所属类别,但其在进行预测时需要将待测数据与所有已知数据进行距离计算,因此适用于小数据量的情况,数据量过大会消耗过多时间。

功耗分析攻击中能量迹的数量也是需要考虑的一个问题,应用尽可能少的数据得到尽可能好的预测结果。前期特征点选取等多种预处理方式都有助于减少数据量,因此 KNN 数据量较小,适用于功耗分析攻击。

1.3　随机森林算法

在处理实际问题时,需要根据具体情况改进传统算法,或按照一定的规则组合多种典型的机器学习算法,以满足实验结果对精确度的要求,因此本文将多类或多个学习器结合完成学习任务,称为集成学习。

随机森林属于将多个决策树按照投票法进行结合的一种集成学习算法[8],其性能优于单一决策树[9]的集成算法。随机森林算法结构如图 1 所示,首先,从数据集 $D=\{(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)\}$ 中进行有放回的采样,得到 m 个子数据集 $D=\{D_1,D_2,\cdots,D_m\}$;然后,利用 m 个子数据集构建 m 棵决策树,每个子决策树输出一个结果 $t_i=T_i(D_i)(i\in[1,m])$,其中,T_i 代表单个决策树算法,t_i 代表子决策树的输出结果;最后,通过对子决策树的判断结果进行多数投票法投票,即寻找矩阵 $\boldsymbol{t}=\{t_1,t_2,\cdots,t_i\}$ 的众数,得到随机森林的输出结果。

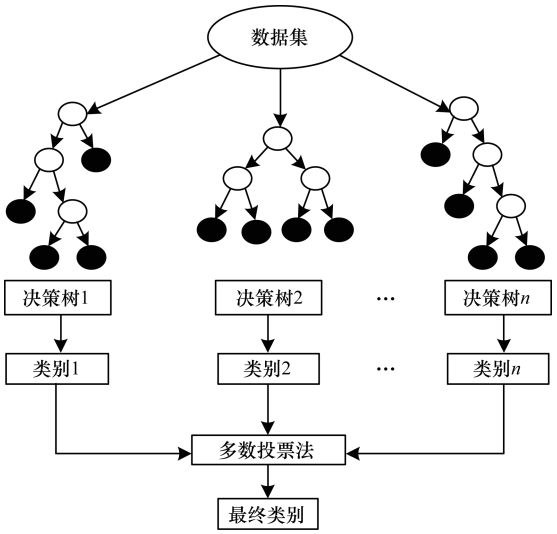


图 1 随机森林算法结构

随机森林作为一种能处理高维度数据的集成算法,可并行处理数据是其优点之一。由于算法会构建多棵决策树,因此对每个数据的属性要求不高,同样对于缺失数据也不敏感。但由于算法过程是个黑盒子,因此无法控制内部运行过程。功耗攻击中采集的曲线是针对同一密码算法,且电压与电压之间相差较小,相比于决策树,使用随机森林算法能够有效提高分类准确率。

1.4 支持向量机算法

支持向量机算法也是一种监督学习的机器学习算法,其关键在于寻找一个满足分类要求的最优超平面,如图 2 所示,其中 γ 是指样本到超平面的距离。

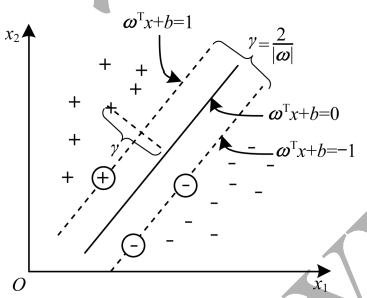


图 2 支持向量机算法原理

在样本空间中,划分超平面的方程如式(4)所示。

$$\boldsymbol{\omega}^{\mathrm{T}}x+b=0 \tag{4}$$

其中,$\boldsymbol{\omega}$ 为法向量,决定超平面的方向,b 是截距,决定超平面与原点的距离。为满足最优化分类的超平面,将式(5)作为支持向量机的基本模型。

$$\min_{\boldsymbol{\omega},b}\frac{1}{2}\|\boldsymbol{\omega}\|^2$$
$$\text{s.t. } y_i(\boldsymbol{\omega}^{\mathrm{T}}x_i+b)\geqslant 1,i=1,2,\cdots,m \tag{5}$$

对于线性可分问题,SVM 采用优化算法实现最大化分类间隔。当处理的问题是非线性问题时,SVM 会引入适当的核函数,将输入空间映射到高维空间,在高维空间,非线性问题被转化为线性可分问题,从而达到区分样本的目的[10]。

SVM 同样适用于高维度样本,与决策函数有关的只是少数的支持向量,同时支持向量的数目又决定了计算的复杂性,这在某种意义上避免了维数灾难,有利于抓住关键样本,剔除大量冗余样本,使得该算法具有较好的鲁棒性。

在功耗分析攻击中,会选择密码算法中的某个步骤作为攻击点,实际上只有少数几个与该攻击点有关的时间点,但由于噪声及算法连续性的影响,会采集几十个甚至几百个时间点的电压信息,利用 SVM 的优势,将其运用到功耗分析中,能够得到一个较好的分类结果。

2 特征点选择

本文使用差分功耗分析(Differential Power Analysis,DPA)国际学术大赛第四阶段(DPA Contest v4)的数据作为实验对象,DPA 国际学术大赛是由法国国家科学研究院、法国巴黎高等电信学校等单位联合主办,DPA Contest v4 分析对象为 ATMega163 芯片上的掩码类 AES-256 密码算法,共采集 100 000 条功耗曲线,每条功耗曲线包含 435 002 个采样点,其中所有的功耗曲线采用的密钥、明文、偏移量和掩码都是已知的。由于本文侧重于对模型选择进行研究,因此将掩码及偏移量作为已知量,使用 S 盒输出值的汉明重量作为标签,对这些数据实施基于机器学习的一阶功耗分析攻击。

能量迹曲线中包含 435 002 个采样点,而能够表现 S 盒汉明重量特征的点只占极少数,过多无关的冗余点会影响机器学习算法的性能,常见的特征点的选取有特征提取和特征选择 2 种方法。其中:特征提取是在原有特征的基础上产生新的特征集,例如主成分分析[11];特征提取是选取原数据集中的子集,例如通过相关性计算来选择评分最高的取样点。

本文使用功耗曲线的电压作为机器学习的输入数据,使用汉明重量作为输出标签。文献[12]中指出密码芯片的功耗电压与汉明重量成正比,因此本文选用皮尔森相关系数提取特征点,且由于实验中采样点数较多,使得提取后的电压呈线性关系,因此使用原始电压集的子集有效减少机器学习时的计算量。

设有 N 条能量迹,每条能量迹上有 M 个点,记能量迹上某一采样点电压为 $v_{(i,j)}$($i\in[1,N]$,$j\in[1,M]$),每条曲线的标签为 $H_j(j\in[1,N])$,根据皮尔逊相关系数公式计算电压值与标签的相关系数。

$$\rho_{X,Y}=corr(X,Y)=\frac{Cov(X,Y)}{\sigma(X)\sigma(Y)}=\frac{Cov(v_{i,j},h_j)}{\sigma(v_{i,j})\sigma(h_j)} \tag{6}$$

对相关系数的绝对值进行从大到小排序,选取

相关系数最大的 m 个点所对应时刻的电压值,形成新的数据矩阵。

通过 800 条能量迹计算能量迹中每点的功耗与汉明重量之间的相关系数值,皮尔森相关系数与样本点的关系如图 3 所示,相关系数最大值在样本点 101 589 的位置,最高为 0.868 736。

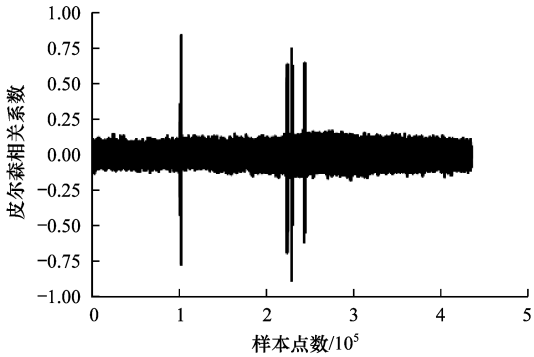


图 3　皮尔森相关系数与样本点的关系

对于机器学习分类算法来说,若采用图 3 中所有有关的点运行时间会过长,造成过拟合。图 4(a)给出从每条功耗曲线选取 0 个特征点到 100 个特征点时的分类准确率,可以看出,在每条曲线选取特征点数低于 30 个时,准确率呈上升状态,高于 50 个点时,支持向量机和 K 最近邻算法的准确率呈下降趋势,也就是过多或者过少的特征点都会降低模型性能。图 4(b)中截取了图 4(a)中 30 个 ~50 个特征点对应的准确率,可以看出此区间内各算法的正确率基本在同一水平线,因此本文实验中每条功耗曲线选取 40 个特征点。

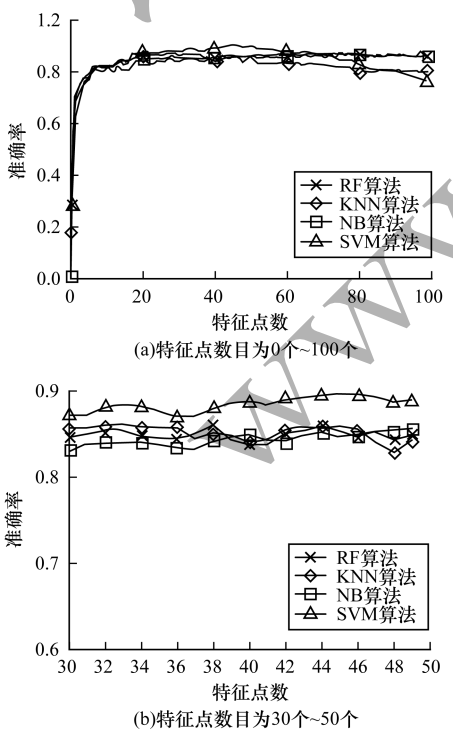


图 4　特征点数目对分类准确率的影响曲线

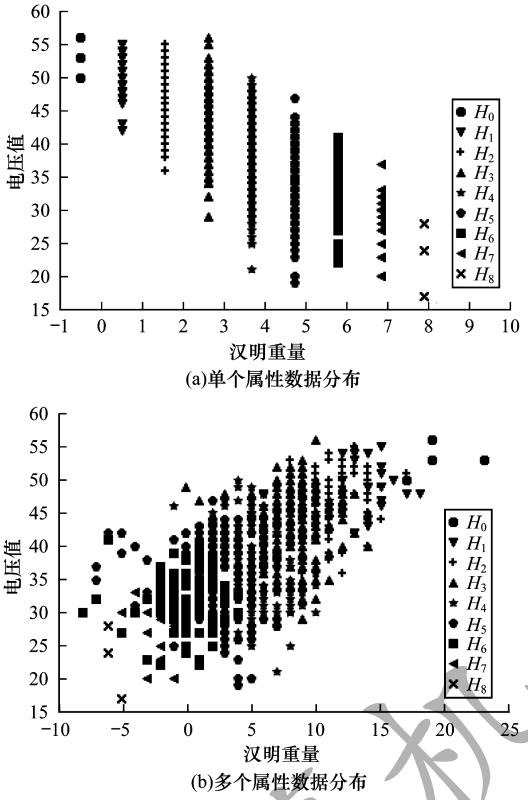
3　机器学习模型与数据分析

本文选用的 4 种机器学习算法分别为朴素贝叶斯、K 最近邻、随机森林以及支持向量机。这 4 种算法都属于监督学习算法,即通过部分被标记的样本数据训练模型,再用模型预测未知数据。从分类方式看可以分为两类:1)分类与样本间的距离有关,即将样本看成一个整体,根据样本间的距离给出待测数据的类别,包括 K 最近邻算法和支持向量机,其中,K 最近邻算法是计算出待测数据与训练样本间的欧氏距离,由 K 个最近邻确定该数据的类别,支持向量机是通过训练样本的距离确定合适的超平面,对待测数据进行分类;2)按照属性的特点进行分类,样本的特点即为属性,根据样本间属性的差异进行分类,包括随机森林和朴素贝叶斯,其中,随机森林是通过计算属性增益构成多棵决策树,再使用投票法对待测数据进行分类,朴素贝叶斯是通过以属性为依据的条件概率来计算后验概率,从而对待测数据进行分类。

在选择汉明重量作为标签时,每条曲线都属于并且仅属于汉明重量为 0 ~8(H_0 ~H_8)的类别中的一种,曲线上的电压就是它的属性。本文选出前 40 个相关系数最大的点,并按照相关系数从大到小的顺序排列。以 800 条数据为例,将相关系数最大的属性(电压值)作为纵轴,汉明重量作为横轴,如图 5(a)所示,可以看出,整体呈负相关性,但属性与属性间有重合部分,因此根据属性划分的类别结果会受到一定的影响。

图 5(a)仅能够看出不同类别间同一个属性的分布情况,无法看出功耗曲线间的位置关系,而 SVM 与 KNN 算法考虑的是整个特征点集合的相对空间位置,因此需要使用多个属性进行描述。本文实验共选取 40 个特征点,因此对于每一条功耗曲线来说,都可以看作是 40 维空间中的一个点。由图 4 可以看出,在属性个数为 2 时已经有正确分类的结果存在,所以可以选取任意两个属性(电压值)作为横坐标和纵坐标近似观察数据分布情况。如图 5(b)所示,相同形状的点代表相同的类别,反之亦然,可以看出数据分布从总体上看呈线性分布,相同类别的样本分布较集中。

由此可以得出,当使用与距离有关的机器学习算法时,结果准确率会高于根据属性划分类别的算法。由于针对本文实验数据时,K 最近邻算法和 SVM 算法效果更佳,而 K 最近算法容易受到数据不均衡的影响,本文实验中的数据中汉明重量为 0 和汉明重量为 8 的数据都只占少数,因此 SVM 算法更加适合本文实验中的功耗分析攻击。



(a)单个属性数据分布

(b)多个属性数据分布

图 5 属性数据分布

4 模型评估与选择

模型评估标准及选择方法决定了选择结果的合理性。本文中机器学习算法使用网格搜索的方法进行调参,保证各算法在同一数据集上的参数均为最优。在模型选择方面,根据“没有免费午餐”定理和“丑小鸭”定理,使用分类准确率作为评估准则,认为准确率越高模型性能越好,并基于该原则进行模型选择研究。

4.1 相关定理

文献[13]提出“没有免费的午餐”(No Free Lunch Theoren,NFL)定理。根据 NFL 可以得出,在实际背景不明确的情况下讨论机器学习算法,没有一种算法会比遍历搜索空间更优。也就是说,在比较算法优劣时,需要规定比较标准,如运行时间、准确率等。本文实验以机器学习十折交叉验证结果作为评估准则比较随机森林、K 最近邻、朴素贝叶斯以及支持向量机算法的性能优劣。

文献[14]提出“丑小鸭”定理。该定理指出世界上不存在分类的客观标准,一切分类的标准都是主观的。在进行机器学习模型选择时,不存在与问题无关的“优越”或“最好”的机器学习算法。即使是两个算法之间的“相似程度”也建立在某种假设的基础上。本文中以准确率作为评估准则,认为准确率越高模型性能越好,基于该选择标准选出更适合本文数据的基于机器学习的功耗分析攻击模型。

4.2 交叉验证

本文选用的机器学习算法都属于监督式机器学习算法,即通过一些已知类别的数据集训练模型,预测未知数据的类别,而实际生活中,待测数据的类别是未知的,因此不能评估已训练模型的优劣。因此将已知数据集进行划分,一部分作为训练集,另一部分作为验证集对模型进行评估,利用验证集误差近似泛化误差。本文中选用 k 折交叉验证(k_cv)[15]对模型进行评估,具体算法如下:

输入 训练集 $D=\{(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)\}$

输出 k-cv 下分类器的性能指标

1. 将训练集 D 分为大小相等的 k 份;
2. for i = 1,2,…,k do
3. 取第 i 份作为测试集;
4. for j = 1,2,…,k do
5. if i != j then
6. 将第 j 份加入到训练集中,作为训练集的一部分;
7. end if
8. end for
9. end for
10. for i = 1,2,…,k – 1 do
11. 训练第 i 个训练集,得到一个分类模型;
12. 使用该模型在对应的测试集中进行预测,计算并保存模型评估指标;
13. end for
14. 计算模型的平均性能;

4.3 Friedman 检验与 Nemenyi 后续检验

通常使用的假设检验方法是比较 2 种算法的性能优劣,例如交叉验证 t 检验,而有些假设检验仅适用于二分类,例如 McNemar 检验,结合实验本文需要比较的是 4 种算法的性能,且属于多分类问题(8 比特汉明重量共分为 9 类),因此使用 Friedman 检验与 Nemenyi 后续检验方法,选择更适合本文实验数据的机器学习模型。

Friedman 检验[6]属于统计假设检验,基本思想是首先建立一个假设检验问题,H_0:k 个算法间无显著差异,H_1:k 个算法间有显著差异。然后在同一组数据集上,根据测试结果(使用交叉验证的结果)对学习器性能进行排序,赋予序值 1,2,…,相同则平分序值,如表 1 所示。

表 1 算法比较序值

数据集	算法 A	算法 B	算法 C
D_1	1.000	2.000	3.000
D_2	1.000	2.500	2.500
D_3	1.000	2.000	3.000
D_4	1.000	2.000	3.000
平均序值	1.000	2.125	2.875

若学习器的性能相同,平均序值应该相同,且第 i 个算法的平均序值 r_i 服从正态分布 $N((k+1)/2,(k^2-1)/12)$,则可使用式(7)计算变量 τ_{χ^2}。

$$\tau_{\chi^2}=\frac{k-1}{k}\cdot\frac{12N}{k^2-1}\sum_{i=1}^{k}\left(r_i-\frac{k+1}{2}\right)^2=\frac{12N}{k(k+1)}\left(\sum_{i=1}^{k}r_i^2-\frac{k(k+1)^2}{4}\right)\tag{7}$$

其中,当 k 和 N 都较大时,r_i 服从自由度为 $k-1$ 的 χ^2 分布。若直接使用式(7)进行假设检验,则结果不够准确,因此对式(7)进行修改,得到变量 τ_F。

$$\tau_F=\frac{(N-1)\tau_{\chi^2}}{N(k-1)-\tau_{\chi^2}}\tag{8}$$

其中,τ_F 服从自由度为 $k-1$ 和 $(k-1)(N-1)$ 的 F 分布。表 2 给出了置信度为 95% 的情况下,F 检验常用的临界值,其中 k 表示算法种数。

表 2　F 检验常用的临界值

数据集个数 N	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$	$k=9$	$k=10$
4	10.128	5.143	3.863	3.259	2.901	2.661	2.488	2.355	2.250
5	7.709	4.459	3.490	3.007	2.711	2.508	2.359	2.244	2.153
8	5.591	3.739	3.072	2.714	2.485	2.324	2.203	2.109	2.032
10	5.117	3.555	2.960	2.634	2.422	2.272	2.159	2.070	1.998
15	4.600	3.340	2.827	2.537	2.346	2.209	2.104	2.022	1.955
20	4.381	3.245	2.766	2.492	2.310	2.179	2.079	2.000	1.935

若计算出的 τ_F 大于表 2 中对应的值,则接受“k 种算法间有显著差异”这一假设,然后使用 Nemenyi 后续检验找出合适的算法。Nemenyi 后续检验计算出平均序值差别的临界值域,表 3 是常用的 q_α 值,α 为显著性水平。根据式(9)计算临界值域 CD,若 2 种算法的平均序值差超出 CD,则能从置信度 $1-\alpha$ 中得出 2 种算法性能不同的结论。

$$CD=q_\alpha\cdot\sqrt{\frac{k(k+1)}{6N}}\tag{9}$$

表 3　Nemenyi 后续检验中常用的 q_α 值

α	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$	$k=9$	$k=10$
0.05	1.960	2.344	2.569	2.728	2.850	2.949	3.031	3.102	3.164
0.10	1.645	2.052	2.291	2.459	2.589	2.693	2.780	2.855	2.920

4.4　模型选择过程

大多数模型选择是通过一次测试集误差来近似泛化误差,但通过分析机器学习算法的特点,发现多次实验的预测结果可能会不一致,在正常情况下准确率会在一个范围内波动。本文选取的 4 种算法通过网格搜索进行参数调优后准确率都在 70% 以上,波动区间有所重合,并且在模型选择中存在随机抽取的过程,因此一次测试集误差不足以近似泛化误差。为弥补上述方法的局限性,本文通过十折交叉验证的均值代替一次测试集误差,即利用均值法削峰填谷的特点,去除过高和过低的准确率,使得结果能够更准确地反映模型的好坏。通过 10 次实验结果代替 1 次实验的结果,将原有的测试集 D 划分成 10 个大小相似的互斥子集,即:$D=D_1\cup D_2\cup\cdots\cup D_{10}$,$D_i\cap D_j=\varnothing(i\neq j)$,每次使用其中的 9 个子集作为训练集,剩下的 1 个子集作为验证集,进行 10 次训练和测试 $acc_1,acc_2,\cdots,acc_{10}$,得到 10 次结果的均值近似泛化误差,如式(10)所示。

$$Acc=\sum_{i=1}^{10}acc_i\tag{10}$$

根据实验中使用的功耗数据具有多样性的特点,测试数量越多,结果越近似泛化误差。以 800 条测试数据为例,将 800 条数据扩展为 3 200 条数据进行 1 次十折交叉验证,对于 SVM 和 KNN 来说,测试时间会远大于 4 次 800 条数据的测试时间。为此,本文使用 4 组数量相同但组成元素不同的数据集的十折交叉验证结果来评定模型的优劣,即 $\{Acc_1,Acc_2,Acc_3,Acc_4\}$,不仅节省了大量时间,而且对每组数据进行十折交叉验证,既满足了数据的多样性,又满足了多次实验近似泛化误差的条件。本文需要对 4 种机器学习算法进行择优,因此 4 个数据集的十折交叉验证结果可以表示为:

$$\begin{bmatrix}Acc_{\mathrm{rf1}} & Acc_{\mathrm{rf2}} & Acc_{\mathrm{rf3}} & Acc_{\mathrm{rf4}}\\ Acc_{\mathrm{nb1}} & Acc_{\mathrm{nb2}} & Acc_{\mathrm{nb3}} & Acc_{\mathrm{nb4}}\\ Acc_{\mathrm{knn1}} & Acc_{\mathrm{knn2}} & Acc_{\mathrm{knn3}} & Acc_{\mathrm{knn4}}\\ Acc_{\mathrm{svm1}} & Acc_{\mathrm{svm2}} & Acc_{\mathrm{svm3}} & Acc_{\mathrm{svm4}}\end{bmatrix}$$

假设每组有 n 条数据,上述方法共使用 $4\times n$ 条数据对模型进行 40 次评估,评估结果远比一次一组数据的准确率更具说服力。以十折交叉验证的结果为前提,对 4 种算法进行对比,选用 Friedman 检验与 Nemenyi 后续检验方法,选择更适合本文数据的机器学习模型。模型选择流程如图 6 所示。

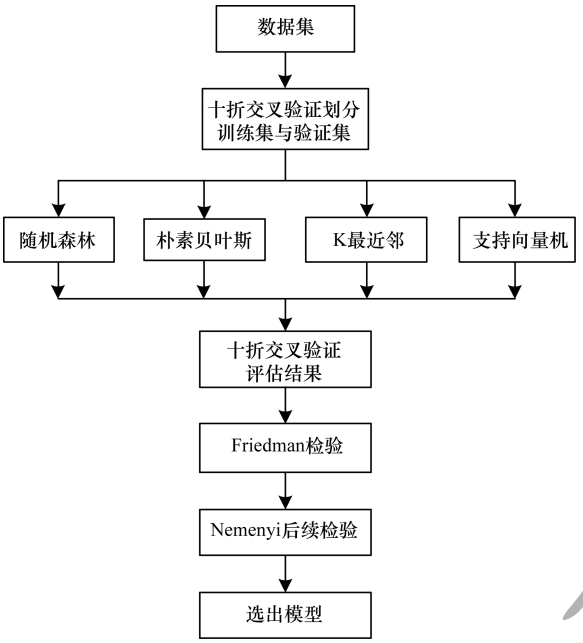


图 6 模型选择流程

5 实验验证

5.1 十折交叉验证

本文选用十折交叉验证方法对随机森林、K 最近邻、朴素贝叶斯及支持向量机 4 种算法进行评估，当使用 800 条数据(包含训练集和验证集)进行模型评估时，结果如图 7 所示。针对一组数据时，十折交叉验证的结果基本重合，无法辨认出模型优劣。同样使用 1 000 条数据进行测试，结果如图 8 所示，曲线间的重合度较 800 条时稍有减弱，但仍不明显，同时可以看出若不采用十折交叉验证，而仅采用一组数据的一次预测结果来近似泛化误差，则可能判定为决策树为最优模型。将数据条数增加到 2 000 条以及 5 000 条，结果分别如图 9、图 10 所示，朴素贝叶斯算法与其他 3 种算法的曲线没有重合之处。为更加严谨地判断模型的优劣，本文进行 Friedman 检验及 Nemenyi 后续检验。

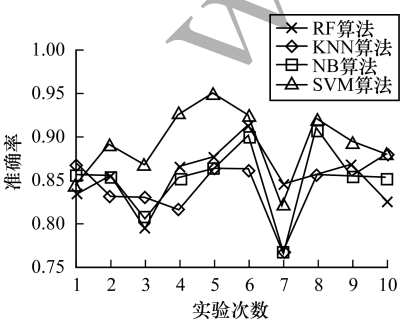


图 7 数据集为 800 条时的十折交叉验证准确率对比

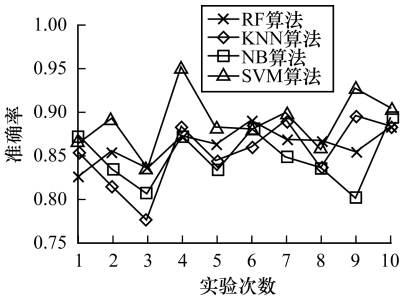


图 8 数据集为 1 000 条时的十折交叉验证准确率对比

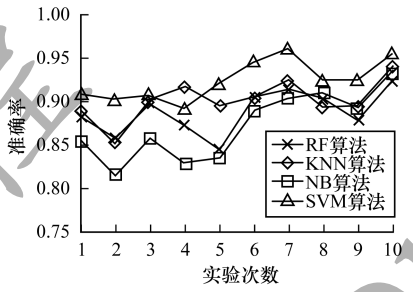


图 9 数据集为 2 000 条时的十折交叉验证准确率对比

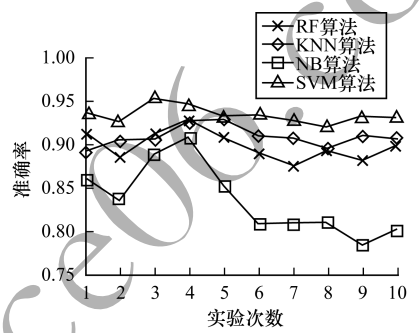


图 10 数据集为 5 000 条时的十折交叉验证准确率对比

5.2 Friedman 检验及 Nemenyi 后续检验结果

实验首先假设 H_0 表示 4 种算法间无明显差异，H_1 表示 4 种算法间有显著差异。然后计算出十折交叉验证的平均值，作为 Friedman 检验中各算法排序的依据。如表 4 所示，D_1、D_2、D_3 及 D_4 分别表示不同的数据集。按照准确率高低对其进行排序，计算每个算法的平均序值，如表 5 所示。

表 4 数据集为 800 条时十折交叉验证平均准确率 %

数据集	RF 算法	KNN 算法	NB 算法	SVM 算法
D_1	85.0	86.8	84.4	87.7
D_2	84.4	87.4	85.0	90.3
D_3	83.6	86.0	86.0	89.1
D_4	81.8	82.5	78.9	87.5

表 5 数据集为 800 条时算法比较序值

数据集	RF 算法	KNN 算法	NB 算法	SVM 算法
D_1	3.000	2.000	4.000	1.000
D_2	4.000	2.000	3.000	1.000
D_3	4.000	2.500	2.500	1.000
D_4	3.000	2.000	4.000	1.000
平均序值	3.500	2.125	3.375	1.000

通过式(8)计算出变量 $\tau_F = 10.230\ 8$,通过查表2中的临界值,当数据集种数为 4、算法个数为 4 时,临界值为 3.863,因此拒绝 H_0,接受 H_1。进而进行 Nemenyi 后续检验,使用式(9)计算出临界值域 $CD = 2.345$,如图 11 所示,每条线段的长度等于临界值域,SVM、朴素贝叶斯以及随机森林的平均序值之间的距离大于临界阈值,因此 SVM 优于朴素贝叶斯和随机森林。而 SVM 与 KNN 的平均序值之间的距离小于临界阈值,因此 SVM 与 KNN 之间无明显差异。但是从表 5 的序值可以看出 SVM 算法始终最优,比 KNN 算法更加稳定。

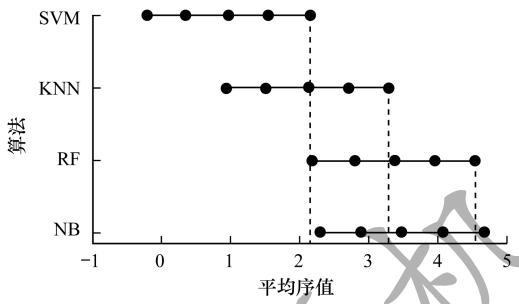


图 11　数据集为 800 条时的 Nemenyi 后续检验结果

将数据集条数扩展到 1 000 条、2 000 条及 5 000 条,按照上述步骤对 4 种机器学习算法进行评估,结果如图 12 ~ 图 14 所示。可以看出,在 4 种数据集情况下 SVM 始终优于朴素贝叶斯,与 K 最近邻没有明显差异,但 SVM 的平均序值小于 K 最近邻,从而表明其在实验中性能更加稳定。

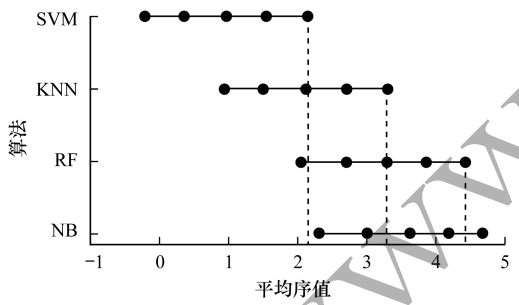


图 12　数据集为 1 000 条时的 Nemenyi 后续检验结果

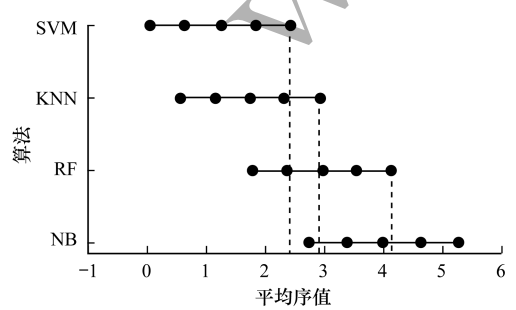


图 13　数据集为 2 000 条时的 Nemenyi 后续检验结果

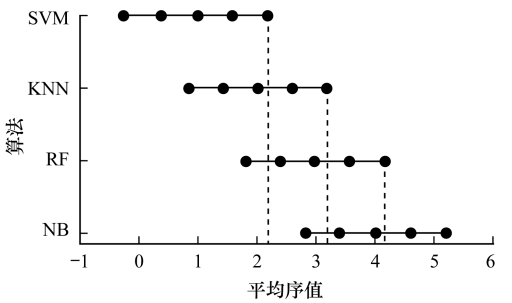


图 14　数据集为 5 000 条时的 Nemenyi 后续检验结果

6　结束语

本文根据密码芯片所采集的能量曲线数据特性,研究相关机器学习算法,并依据各算法特点,分析适合功耗分析攻击的机器学习算法。实验选用随机森林、朴素贝叶斯、K 最近邻以及支持向量机 4 种机器学习算法,结果表明按照相对距离进行划分的支持向量机算法是最适合功耗分析攻击的机器算法,Friedman 检验与 Nemenyi 后续检验结果也证明了支持向量机算法的性能更加稳定。然而,本文仅从机器学习算法分类准确率的角度进行性能评估,且在实验中发现数据具有不均衡特性,因此下一步将针对数据不均衡问题对算法进行改进,提高分类准确率。

参考文献

[1] HOSPODAR G, MINDER E D, GIERLICHS B, et al. Least squares support vector machines for side-channel analysis [C]//Proceedings of the 3rd International Conference on Control, Automation and Robotics. Washington D. C., USA: IEEE Press, 2011: 99-104.

[2] HE H, JAFFE J, ZOU Long. Side channel cryptanalysis using learning [EB/OL]. [2018-10-08]. http://cs229.stanford.edu/proj2012/HeJaffeZou-SideChannelCryptanalysisUsingMachineLearning.pdf.

[3] HEUSER A, ZOHNER M. Intelligent machine homicide [C]// Proceedings of the 3rd International Conference on Constructive Side-Channel Analysis and Secure Design. Washington D. C., USA: IEEE Press, 2012: 249-264.

[4] WHITNALL C, OSWALD E. Robust profiling for DPA-style attacks [C]//Proceedings of CHES'15. Berlin, Germany: Springer, 2015: 3-21.

[5] LERMAN L, MEDEIROS S F, BONTEMPI G, et al. A machine learning approach against a masked AES [J]. Journal of Cryptographic Engineering, 2015, 5(2): 123-139.

[6] 周志华. 机器学习[M]. 北京:清华大学出版社,2016.

(下转第 158 页)

(上接第 151 页)

[7] 李强,蒋静坪. 量子 K 最近邻算法[J]. 系统工程与电子技术,2008,30(5):940-943.
[8] 王奕森,夏树涛. 集成学习之随机森林算法综述[J]. 信息通信技术,2018,12(1):49-55.
[9] 张宇. 决策树分类及剪枝算法研究[D]. 哈尔滨:哈尔滨理工大学,2009.
[10] 刘方园,王水花,张煜东. 支持向量机模型与应用综述[J]. 计算机系统应用,2018,27(4):1-9.
[11] 赵蔷. 主成分分析方法综述[J]. 软件工程,2016,19(6):1-3.
[12] 罗鹏,冯登国,周永彬. 功耗分析攻击中的功耗与数据相关性模型[J]. 通信学报,2012,33(z1):276-281.
[13] WOLPERT D H. No free lunch theorems for search[EB/OL]. [2018-10-08]. https://core.ac.uk/display/24281656.
[14] WATANABE S. Knowingand guessing: a quantitative study of inference and information[M]. New York, USA:Wiley,1969.
[15] 范永东. 模型选择中的交叉验证方法综述[D]. 太原:山西大学,2013.

编辑 陆燕菲