

基于一维卷积神经网络的雌激素粉末拉曼光谱定性分类

赵 勇¹, 荣 康¹, 谈爱玲²

1. 燕山大学电气工程学院, 河北省测试计量技术及仪器重点实验室, 河北 秦皇岛 066004

2. 燕山大学信息科学与工程学院, 河北省特种光纤与光纤传感重点实验室, 河北 秦皇岛 066004

摘 要 拉曼光谱物质定性鉴别已被广泛应用于诸多行业和研究领域, 但传统拉曼光谱分析过程中的预处理主要依赖人为经验, 光谱特征提取虽然能够降低信号维度, 同时也会造成部分光谱信息损失。特性相近物质本身光谱相似度较高, 受到测量过程中环境干扰和分析过程中多种误差影响, 导致最终分类效果并不理想。针对此问题, 提出基于一维卷积神经网络(one-dimensional convolution neural network, 1D-CNN)的拉曼光谱定性分类方法。实验采集雌酮(Estrone)、雌二醇(Estradiol)、雌三醇(Estriol)三种不同雌性激素粉末的拉曼光谱, 设计随机平移、添加噪声和随机加权三种光谱数据增强方法, 构建数量充足的拉曼光谱数据库用于神经网络模型训练与测试; 基于拉曼光谱数据特点提出一维卷积神经网络分类模型, 将光谱预处理、特征提取和定性分类的全过程融为一体。通过大量仿真实验, 优化所提出的神经网络模型超参数和训练过程并测试分类效果, 从预处理对光谱分类结果的影响和模型抗干扰性能两个方面与多种传统拉曼光谱分类算法对比, 评价模型性能。实验结果表明, 本文提出的一维卷积神经网络模型可实现三类雌性激素粉末拉曼光谱快速准确分类, 分类正确率最高可达 98.26%, 分析过程中无需光谱预处理和特征提取步骤, 简化了光谱分析流程, 并能保留更多有效信息。同时, 当模拟测量噪声强度达到 60 dBW 时, 传统方法分类正确率均明显出现不同程度明显降低, 卷积神经网络模型依然能够取得 96.81% 的分类正确率, 说明相对传统拉曼光谱分类方法, 所提出方法受光谱测量噪声影响更小, 鲁棒性更强, 适用于分析更复杂现场测量的强噪声拉曼光谱信号。该研究结果表明深度学习方法在拉曼光谱的分析与处理领域具有很大的应用潜力和研究价值。

关键词 拉曼光谱; 深度学习; 一维卷积神经网络; 雌性激素; 定性分类

中图分类号: O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)12-3755-06

引 言

激光拉曼光谱(laser Raman spectroscopy)作为一种精准的物质检测技术, 具有数据采集简单、快速、灵敏度高等诸多优点^[1-2], 被称作物质的“指纹图谱”, 可有效定性分析多种形态物质, 已被广泛应用于工业、医学和生命科学等诸多领域^[3-4]。光谱处理与分析算法是实现拉曼光谱定性鉴别的关键, 分类算法直接影响物质检测准确性和效率。

传统拉曼光谱定性分析步骤比较繁琐, 分析过程多依赖于经验。首先需对光谱数据进行预处理, 如去荧光背景、平滑去噪和基线校正等。然后提取光谱中重要波峰信息或运用主成分分析等方法对光谱信号降维, 保留重要特征, 降低计算复杂性。最后运用机器学习方法分类, 常用的有随机森林、支持向量机等分类算法。拉曼光谱测量技术不断发展,

对光谱分析算法提出更高要求, 光谱分析趋向于更高效和自动化。

目前, 以深度学习(Deep Learning)为代表的复杂神经网络模型成为人工智能的重要研究领域, 其强大表征能力使计算机能够完成类似人脑的学习与判断, 在某些方面甚至具有超越人脑的决策能力。卷积神经网络(convolutional neural networks, CNN)是典型的前馈神经网络, 是深度学习算法的一个重要分支^[5-6]。凭借局部连接和权值共享两个重要特征, CNN 模型学习和分类能力远超传统神经网络, 具有广泛的适用性, 已在众多领域中展现出显著优势。Krizhevsky 等基于对图像分类算法的改进提出经典卷积神经网络结构 AlexNet^[7], 并凭借该模型在 2012 年 ImageNet 图像分类大赛中赢得第一名。近年来, 卷积神经网络结构更为复杂, 层次更深, VGGNet 具有 16 层网络结构^[8], GoogleNet 具有 22 层网络结构^[9], 而赢得 2015 年 ILSVRC 冠军的 ResNet 更是

收稿日期: 2019-05-08, 修订日期: 2019-09-20

基金项目: 国家自然科学基金项目(61575170), 国家重点研发计划项目(2017yfc1403800)和燕山大学博士基金项目(B779, B687)资助

作者简介: 赵 勇, 1978 年生, 燕山大学电气工程学院副教授 e-mail: zhaoyong@ysu.edu.cn

达到了 152 层的网络结构深度^[10]。

近年来,深度神经网络在光谱分析领域也得到研究与应用,Tuan 等提出深度学习、极限学习机-ELA 算法和可见、红外光谱联合建立煤矿分类模型,研究发现与 PCA 特征提取方法比较,CNN 可以更好地提取光谱数据特征^[11]。Liu 等提出一种使用深度卷积神经网络对多类别拉曼光谱数据分类的方法,在 RRUFF 矿物拉曼光谱数据库上取得了优异的分类效果^[12]。针对训练神经网络的数据库中样本常需要更新的问题,Liu 等提出使用 Siamese 神经网络的解决方案,该系统仅需极少量训练样本,可对未知物质拉曼光谱准确分类,同时将其动态添加至模型中^[13]。

本文将一维卷积神经网络模型应用于拉曼光谱分析,采集 3 种雌性激素粉末的拉曼光谱,设计拉曼光谱数据增强方

法,构建拉曼光谱数据库;针对拉曼光谱数据特点,提出适用的一维卷积神经网络分类模型,优化模型超参数并完成训练。从光谱预处理和抗干扰性能两个方面讨论所建立模型的优势和特点。

1 实验部分

实验样品为雌酮(Estrone)、雌二醇(Estradiol),雌三醇(Estriol)(纯度均大于 97.5%)三种雌性激素粉末,购自上海安谱实验科技股份有限公司,其化学结构式如图 1 所示。实验使用的拉曼光谱仪为 i-Raman[®] Plus,购自必达泰克光电科技(上海)有限公司。激光器激发光源 785nm,最大功率为 300 mW(可调),光谱范围 175~3 200 cm^{-1} ,光谱分辨率 4.5 cm^{-1} 。

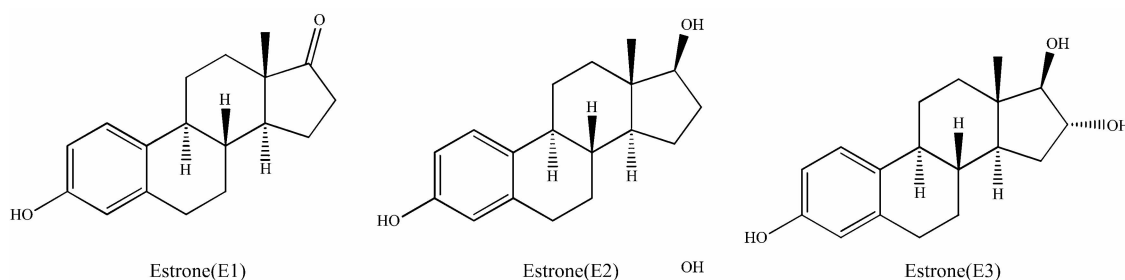


图 1 三类雌性激素化学结构

Fig. 1 The molecular structure of 3 classes of estrogen

拉曼光谱采集时,光谱仪功率设定为 30 mW,积分时间 10 s。为避免光照对测试结果的影响,保证采集环境一致性,测量全程在暗室内进行。光谱采集前,以相同的积分时间采集暗电流。每类准备 40 个样品,三类雌性激素样品平均拉曼光谱数据如图 2 所示。观察拉曼光谱数据,可知三种雌性激素样本均在 710 cm^{-1} 附近有相对强度最高的波峰,在 $900 \sim 1 500 \text{ cm}^{-1}$ 波段特征峰较为密集且相似,包含较丰富的物质结构信息。

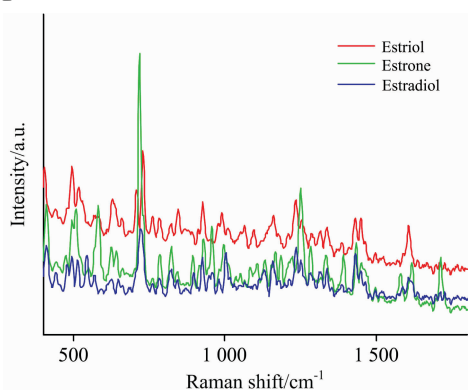


图 2 三类雌性激素粉末的平均拉曼光谱图

Fig. 2 Mean Raman spectra of 3 classes of estrogen powders

2 CNN 模型搭建与训练

2.1 拉曼光谱数据增强

卷积神经网络模型的优异性能是经过大量数据训练达到

的。深度神经网络模型训练过程由数据驱动,充足的训练数据可以使神经网络模型充分学习到数据类别内部特征和类别间的区别,增强模型的鲁棒性,尽可能避免过拟合现象。受拉曼光谱采集环境、人力和测量设备限制,一般很难大批量得到带有标签的拉曼光谱数据样本。

为充分训练卷积神经网络,提升网络泛化性能和鲁棒性,本文使用数据增强(Data Augmentation)的方式扩充样本数据,增强数据多样性。设计 3 种光谱数据增强方法,数据扩充步骤如下:

(1)在有限的拉曼位移移动范围内,左右随机平移光谱,将每类雌性激素拉曼光谱数据由 40 条扩充至 100 条;

(2)向拉曼光谱图像中添加强度不等的随机高斯白噪声,扩充每类光谱数据至 400 条;

(3)使用总和为 1 的随机比例系数,将数条同类别拉曼光谱数据线性叠加,扩充每类光谱数据至 2 000 条。

经数据增强,生成包含三类带标签的雌性激素拉曼光谱数据集共 6 000 条。数据集被按比例划分为三个部分:60% 数据作为训练集;10% 数据用作验证集,在反向传播训练过程中实时调整神经元权重参数;30% 数据作为测试集,用于测试已训练后的神经网络性能。

2.2 1D-CNN 雌性激素拉曼光谱定性分类模型

拉曼光谱信号是一维信号,参照经典卷积神经网络模型 LeNet-5,结合仿真实验调整 CNN 模型结构和参数,提出拉曼光谱数据分类的一维卷积神经网络模型,其结构如图 3 所示。输入层输入原始光谱,卷积层通常位于输入层之后,由一组参数可训练的滤波器组成,这些滤波器通常感知区域较

小, 也被称作卷积核。网络前向传播过程中, 每个卷积核都会在输入数据上按一定方向滑动, 对所覆盖区域执行卷积运算。一维信号卷积运算可由式(1)表示

$$y^j = f\left(\sum_i k^{ij} * x^i + b^j\right) \quad (1)$$

其中, * 表示卷积运算, y^j 为第 j 个输出特征图, x^i 为第 i 个输入特征图, k^{ij} 为本层卷积运算所使用的卷积核, b^j 为第 j 个特征图的偏置。

池化层位于卷积层后, 对卷积层运算生成的特征图采样, 也称为采样层。池化层运算并未减少特征图的个数, 而是减小每个特征图的维度, 缩减数据量, 可提升运算速度, 增强神经网络模型鲁棒性。本文采用最大池化法(Max-pooling), 以特定宽度采样窗口内最大值作为特征值, 其表达式如式(2)

$$f(X_k) = \max_{a \in X_k} \{a_1, \dots, a_s\} \quad (2)$$

式(2)是最大池化对信号降采样计算方法, 对卷积层运算所得的一个特征映射 Y_i , 将其划分成多个不重叠区域 $X_k, k=1, 2, \dots, K$, 区域大小为 $s \times 1$ 。

经过卷积层和池化层的运算, 神经网络有效提取并提炼原始数据中无法直接获取的局部和全局特征, 将特征输入全连接层分类。本文模型包含 2 个卷积层、2 个池化层和 2 个全连接层, 一维卷积核尺寸为 (10×1) 和 (5×1) 。每条拉曼光谱信号以一维张量形式输入至神经网络中, 记录了整条光谱由 $399.93 \sim 1803.27 \text{ cm}^{-1}$ 共 707 个拉曼偏移所对应的拉曼散射强度。卷积层使用线性修正单元变体 LeakyReLU 作为激活函数, 表达式如式(3)所示

$$f(x) = \begin{cases} x, & x > 0 \\ ax, & \text{otherwise} \end{cases} \quad (3)$$

为避免过拟合, 提高泛化性能, 在每层神经网络后加入批标准化层(batch normalization, BN), 在第一个全连接层后应用比例为 0.5 的随机失活(Dropout)。

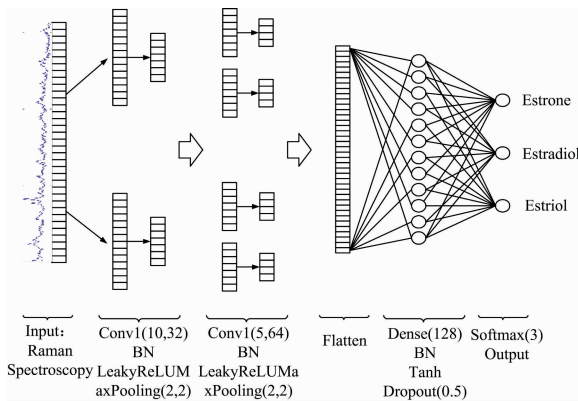


图 3 一维卷积神经网络拉曼光谱分类模型结构

Fig. 3 The structure of one-dimensional convolutional neural network Raman spectral classification model

2.3 神经网络模型训练

一维卷积神经网络模型在训练过程中, 可将网络模型整体视为一个复杂函数 F_{CNN} , 无需划分子问题, 求得全局最优解即可, 图 4 所示是反向传播算法训练神经网络过程。

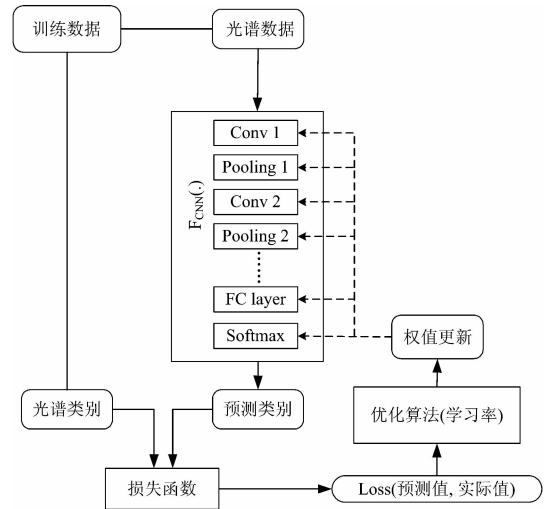


图 4 反向传播算法训练 CNN 模型流程图

Fig. 4 The flow chart of Backpropagation algorithm

卷积神经网络的训练和优化依赖于损失函数(Loss Function), 损失函数计算预测值和真实值之间的误差, 通过反向传播算法将误差从最后一层反向传播至网络各层并更新权值。更新后的参数继续参与训练, 循环往复, 直到损失函数数值达到最小, 即达到了最终训练的目的。本文 CNN 模型训练使用交叉熵损失函数, 表达式如(4)式所示

$$L(x_n, y_n) = -\frac{1}{N} \sum_{n=1}^N \sum_{k=1}^K y_{n,k} \ln p_{n,k} \quad (4)$$

其中, x_n 是训练光谱数据, $y_{n,k}$ 是第 n 个样本预测第 k 个数据的标签, $p_{n,k}$ 是第 n 个样本预测第 k 个标签值的概率, N 是总共的样本数, K 为总标签类数。

损失函数下降采用 Adam 优化算法, 参数设置为 $\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 10^{-8}, \eta = 0.001$ 。为实现 CNN 模型有序参数更新和快速收敛, 训练样本被分成多个批次(batch), 批处理样本数目(batch size)设置为 50, 训练过程中损失函数值下降曲线如图 5 所示。

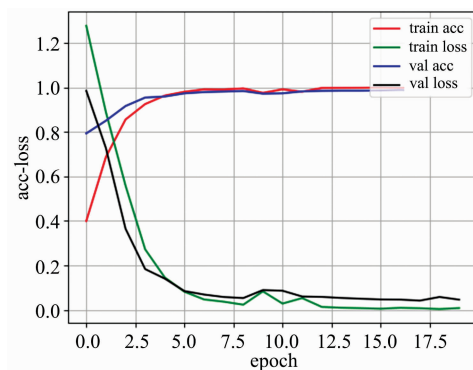


图 5 CNN 模型训练过程中损失值下降曲线

Fig. 5 The curve of decreasing loss value in the process of CNN model

1D-CNN 模型基于 Kears 框架, TensorFlow 后端, 通过 Python 计算机语言搭建。实验环境: Intel Core i7-7700HQ

CPU; Radeon Pro 560 GPU; 16GB 2133MHz 计算机内存。为实现数据优化和对比不同机器学习算法, 系统环境中配置 Numpy, Scipy, Sckit-learn 等 Python 运算库。

3 结果与讨论

3.1 不同分类方法准确性比较

为验证 1D-CNN 模型的拉曼光谱分类性能, 本文还使用 K 近邻(KNN)、梯度提升(Gradient boosting)、随机森林(Random forest)、支持向量机(SVM(linear) & SVM(rbf))、决策树(Decision tree)、卷积神经网络(CNN)共 7 种传统分类器对雌性激素光谱信号分类, 与本文 1D-CNN 模型对比。使用拉曼光谱传统分类方法时, 数据通常需要经过基线校正和平滑滤波预处理, 故本文以原始光谱数据、经基线校正和经平滑滤波处理后的 3 种拉曼光谱数据集为研究对象, 对比不同分类方法准确性的同时, 探究预处理对多种拉曼光谱分类方法效果的影响。

以上 7 种分类算法在每类光谱数据上分别实验 10 次, 记录其平均分类正确率。仿真实验遵循常规拉曼光谱处理与分析步骤规范, 光谱分类前经过主成分分析提取光谱有效信息, 各分类器均已经过参数调试和优化。通过 airPLS 算法对拉曼光谱信号做基线校正, 使用 Savitzky-Golay 滤波器对信号平滑处理, 实验结果见表 1。

表 1 不同算法的分类正确率

Table 1 The classification accuracy of different algorithms

分类算法	分类正确率/%		
	原始数据	基线校正	平滑滤波
KNN	78.34±0.83	90.94±1.54	87.27±0.21
Gradient boosting	78.28±1.32	92.17±0.64	92.78±0.19
Random forest	82.11±0.98	96.27±1.21	93.89±0.93
SVM(linear)	86.06±0.60	97.94±0.62	95.44±0.37
SVM(rbf)	83.89±1.31	96.39±0.95	93.11±0.50
Decision tree	81.44±0.88	89.72±0.70	92.07±1.01
CNN	98.29±0.12	96.54±0.21	95.83±0.30

对比不同分类算法的拉曼光谱数据分类正确率, 在原始拉曼光谱数据集上应用卷积神经网络模型分类效果最佳, 可达 98.29%。由表 1 可知, 拉曼光谱数据经过基线校正或平滑滤波处理后, 6 种传统分析方法的分类正确率均得到不同程度的提升。分类正确率提升最明显的是基线校正处理后的随机森林分类器, 由 82.11% 提高到 96.27%。

通过以上实验分析, 拉曼光谱的基线校正和平滑滤波预处理步骤均可显著提高传统光谱分类方法的分类正确率, 其中基线校正步骤对光谱分类效果的提升最为明显。对 CNN 模型而言, 基线校正和平滑滤波步骤反而使得其分类正确率有了小幅度降低, 可知在训练 CNN 模型时, 直接使用原始光谱数据即可。但若应用传统光谱分析方法, 则应首先对原始光谱数据执行基线校正和滤波处理。

3.2 光谱预处理对不同分类方法的影响

基线校正和平滑滤波都为光谱预处理的重要步骤, 为研

究预处理步骤对不同分类方法的影响, 分析不同分类模型对光谱预处理的依赖性。基于经过完整预处理(基线校正+平滑去噪)后的拉曼光谱数据和原始光谱数据, 测试 6 种传统定性分析方法和应用 CNN 模型分类正确率, 实验结果如图 6 所示。

经过完整拉曼光谱预处理步骤后, SVM(linear), SVM(rbf), Random forest 和 CNN 模型分类正确率分别为 96.89%, 98.21%, 96.27% 和 96.33%, 均可对三类不同雌性激素拉曼光谱数据准确区分, 实现有效鉴别。预处理步骤对 SVM 和 Random forest 方法的提升十分明显, 而一维卷积神经网络模型在原始光谱数据集上便取得了 98.3% 较高分类正确率, 表现出神经网络的强大学习和分类能力。

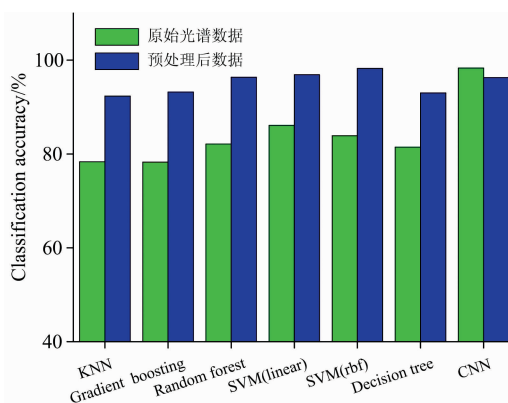


图 6 不同算法分类正确率对比

Fig. 6 Comparison of classification accuracy of different algorithms

通过以上实验结果分析, 传统拉曼光谱分析方法对光谱预处理步骤的依赖性较强。但拉曼光谱数据在预处理, 过程中必定会损失或改变部分有效信息, 致使 CNN 分类准确率出现小幅度下降。由此可见, 卷积神经网络可直接使用原始拉曼光谱数据训练模型和分类, 不再需要对光谱执行繁琐的预处理步骤。卷积神经网络可有效避免拉曼光谱预处理, 在节省时间成本和计算成本的同时实现准确分类。

3.3 噪声对不同分类方法的影响

拉曼光谱现场采集过程中通常存在较多不可控干扰, 包含环境中噪声、荧光、辐射等, 多种噪声干扰的存在对光谱分类模型提出了更高要求。因此当分类模型用于现场拉曼采集设备时, 需要具有更强的抗噪声干扰能力和鲁棒性, 以适用于更复杂的检测环境。为卷积神经网络模型的鲁棒性, 我们在原始光谱信号中加入不同梯度强度的高斯白噪声干扰, 使用 KNN, Gradient boosting, Random forest, SVM(linear), SVM(rbf), Decision tree 和本文 1D-CNN 分类方法, 对比不同强度噪声对多种模型分类准确性的影响。

拉曼光谱数据中添加的噪声强度分别为 20, 30, 40, 45, 50, 55 和 60 dBW。每组实验测试 3 次, 记录平均正确率, 将实验结果绘制折线图, 如图 7 所示。

从图 7 可以看出当原始拉曼信号中添加微弱噪声时, 各种分类方法都能达到较高的分类正确率。但随机噪声强度不

断增大时,各算法的分类正确率会受到不同程度影响,其中 KNN, Random forest, Decision tree 和 Gradient boosting 算法的分类正确率受噪声影响最为明显。当噪声强度超过 50 dBW 时, KNN 和 Decision tree 算法的分类正确率出现大幅度下降,降至 70% 甚至更低,当信号中噪声强度达到 55 dBW 时 Random forest 和 Gradient boosting 算法的分类正确率也降至 70% 左右。

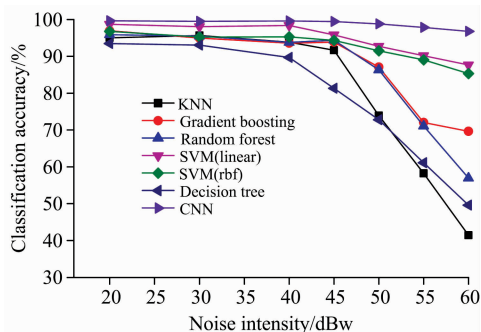


图 7 随机噪声对不同算法分类正确率的影响

Fig. 7 The influence of random noise on the classification accuracy of different algorithms

SVM(linear), SVM(rbf)算法和本文提出的 1D-CNN 模型受噪声的影响较小一些,卷积神经网络模型表现出最强的

抗噪声能力。添加 60 dBW 的随机噪声时,卷积神经网络模型依然能够学习到物质拉曼光谱信号的特征,取得 96.81% 的分类正确率。这充分表现出卷积神经网络模型对光谱特征的强大挖掘能力,在处理光谱分类任务时具有更强的鲁棒性,分类效果更加可靠。

4 结 论

提出了一维卷积神经网络拉曼光谱定性分类模型,实现 3 种雌性激素粉末拉曼光谱定性鉴别。设计 3 种拉曼光谱数据增强方法扩充数据量,以提升模型学习效果,构建充足的拉曼光谱数据集。搭建卷积神经网络拉曼光谱分类模型,确定最优结构参数和训练参数,完成 CNN 模型训练。与传统光谱分类方法相比,1D-CNN 模型可不经光谱数据预处理,实现 3 种雌性激素拉曼光谱数据准确分类。研究发现,CNN 模型能同时实现拉曼光谱信号分析过程中预处理、特征提取和分类,分析过程中无需人工干预,能在准确鉴别物质光谱数据的同时简化信号处理流程。通过在拉曼光谱信号中添加不同强度噪声,发现 CNN 模型具有更强抗噪声干扰性能,鲁棒性更高。可见,一维卷积神经网络是一种优秀的拉曼光谱定性分析方法,相比传统方法更为简便、快速、准确,可适应更复杂信号和环境,有明显的优势和重要的应用价值。

References

- [1] Schlücker S. *Angewandte Chemie International Edition*, 2014, 53(19): 4756.
- [2] Wang Alan, Kong Xianming. *Materials*, 2015, 8(6): 3024.
- [3] Tan Ailing, Zhao Yong, Sivashanmugan Kundan, et al. *Food Control*, 2019, 103: 111.
- [4] Aron Hakonen, Per Ola Andersson, Michael Stenbæk Schmidt, et al. *Analytica Chimica Acta*, 2015, 893: 1.
- [5] Sainath T N, Kingsbury B, Saon G, et al. *Neural Networks*, 2015, 64: 39.
- [6] Gu J, Wang Z, Kuen J, et al. *Computer Science*, 2015. arXiv: 1512. 07108.
- [7] Krizhevsky A, Sutskever I, Hinton G. *Advances in Neural Information Processing Systems*, 2012, 25(2): 1097.
- [8] Simonyan K, Zisserman A. *Computer Science*, 2015. arXiv 1409. 1556.
- [9] Szegedy C, Liu W, Jia Y, et al. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015. 1.
- [10] He K, Zhang X, Ren S, et al. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. 770.
- [11] LE Ba Tuan, XIAO Dong, MAO Ya-chun, et al(LE Ba Tuan, 肖冬, 毛亚纯, 等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2018, 38(7): 2107.
- [12] Liu J, Osadchy M, Ashton L, et al. *Analyst*, 2017, 142(21): 4067.
- [13] Liu J, Gibson S J, Mills J, et al. *Chemometrics and Intelligent Laboratory Systems*, 2019, 184: 175.

Qualitative Analysis Method for Raman Spectroscopy of Estrogen Based on One-Dimensional Convolutional Neural Network

ZHAO Yong¹, RONG Kang¹, TAN Ai-ling²

1. School of Electrical Engineering, Yanshan University, The Key Laboratory of Measurement Technology and Instrumentation of Hebei Province, Qinhuangdao 066004, China

2. School of Information and Science Engineering, Yanshan University, The Key Laboratory for Special Fiber and Fiber Sensor of Hebei Province, Qinhuangdao 066004, China

Abstract The qualitative identification of Raman spectroscopy has been widely used in many industries and research fields, but the preprocessing in traditional Raman spectroscopy analysis mainly relies on human experience. Although spectral feature extraction can reduce the signal dimension, it also causes partial spectral information loss. Materials with similar characteristics have high spectral similarity. In addition, due to the interference of measurement environment and the various errors in the analysis process, the final classification accuracy is not ideal enough. Aiming at these problems, this paper proposes a novel Raman spectral qualitative classification method based on one-dimensional convolution neural network. The Raman spectra of three different estrogen powders, estrone, estradiol and estriol, were collected, and three Raman spectral data augmentation methods were designed to construct Raman spectral database; a one-dimensional convolution neural network classification model for Raman spectral data was proposed, which integrated the whole process of spectral preprocessing, feature extraction and qualitative classification. The hyper-parameters and training process of the proposed classification model were optimized and the accuracy was tested by simulation experiments. Experimental results indicated that the 1D-CNN model can classify three similar estrogen powders Raman spectroscopy with the highest classification accuracy of 98.26%. No spectral preprocessing and feature extraction steps were required in the analysis process, which simplifies the spectral signal analysis process and can retain more vital information. In addition, when the noise intensity of the simulated measurement reached 60 dBW, the classification accuracy of the traditional methods decreased obviously in varying degrees, but the 1D-CNN model could still achieve 96.81% accuracy. Compared with traditional Raman spectral classification methods, the proposed method was less affected by the noise of measurement process and had stronger robustness, which was suitable for Raman spectral signals with strong noise measured in more complex environments. The results of this study show that deep learning method has great application potential in the field of analysis of Raman spectroscopy.

Keywords Raman spectroscopy; Deep learning; One-dimensional convolution neural network; Estrogen; Qualitative classification

(Received May 8, 2019; accepted Sep. 20, 2019)