

基于深度学习的抗菌药物耐药性分析研究

谢修娟^{1†}, 顾兵²

(1. 东南大学成贤学院 计算机工程系, 江苏 南京 210000;

2. 徐州医科大学 医学技术学院, 江苏 徐州 221004)

摘要:细菌耐药性的日益加剧,以及目前的耐药性检测方法周期长等问题,给临床第一时间准确用药带来极大的挑战和困境.为此,本文将探索深度学习技术在抗菌药物耐药性预测中的应用,提出一种融合注意力机制的双通道卷积神经网络模型,通过上下两个通道对建模后的送检数据做不同粒度的特征提取,每个通道经过卷积和池化后引入注意力机制,聚焦重要的特征信息,而后将两个通道的特征进行融合,从而完成分类输出.将模型在某三甲医院细菌药敏检测历史数据集上,与多种不同方法进行对比实验,结果表明,本文所提出方法在分类准确度 F 值指标中平均实现 20.35% 的提升,同时在小样本分类上表现出更好的效果.

关键词:深度学习;卷积神经网络;注意力机制;耐药性预测

中图分类号:TP301

文献标志码:A

Research on Antimicrobial Resistance Analysis Based on Deep Learning

XIE Xiujuan^{1†}, GU Bing²

(1. Department of Computer Engineering, Southeast University Chengxian College, Nanjing 210000, China;

2. College of Medical Technology, Xuzhou Medical University, Xuzhou 221004, China)

Abstract: The increasing drug resistance of bacteria, as well as the long cycle of current drug resistance testing methods, bring great challenges and difficulties to accurate drug use at the first time in clinic. Therefore, this paper will explore the application of deep learning technology in the prediction of antimicrobial resistance, and proposes a dual-channel convolution neural network model integrating attention mechanisms. Through the upper and lower channels, different granularity features are extracted from laboratory data after modeling. After convolution and pooling, an attention mechanism is introduced in each channel to focus on important feature information, and then the features of the two channels are fused to complete the classification output. The model is applied to the historical data set of bacterial drug sensitivity test in a tertiary hospital, and compared with other methods. The results show that the proposed method achieves an average improvement of 20.35% in F -value index of classification accuracy, and performs better in small sample classification.

Key words: deep learning; convolutional neural network; attention mechanism; drug resistance prediction

* 收稿日期:2020-09-22

基金项目:国家自然科学基金资助项目(81871734), National Natural Science Foundation of China(81871734);江苏省青蓝工程资助项目(202004), Qinglan Project of Jiangsu Province(202004)

作者简介:谢修娟(1981—),女,江苏高邮人,东南大学成贤学院副教授

† 通信联系人, E-mail:xiujuanx@126.com

在互联网技术高速发展的今天,医疗、教育、金融、电商等各个行业的数据以“爆炸”式速度在增长,如何从海量数据中,提取有价值的信息变得越来越重要.深度学习作为大数据分析的关键技术之一^[1],它将特征提取与分类器有机融合,能从大量输入信息中自动学习特征并完成分类,在图像识别、语音识别、文本分类及情感分析等诸多领域取得了较好的实践效果^[2-3].近年来,不少学者开始探索医疗卫生行业大数据的智能化处理,尝试运用深度学习技术处理医学大数据,主要集中于对医学图像、电子病历、基因组学的分析^[4],特别是医学影像图像,包括图像分类、图像分割、图像重建、目标检测等关键问题,如:Yasaka等^[5]使用深度学习技术对增强CT图像上肝脏占位性疾病进行包括典型的干细胞癌、不确定的肝肿块及肿块样病灶、血管瘤在内的五种病灶分类,平均AUC为0.92;Jiang等^[6]利用基于AlexNet的全卷积神经网络进行视网膜血管分割,在4个公开数据集上取得优异表现;You等^[7]提出一种基于半监督的图像恢复算法,将残差学习方式用于循环对抗生成网络,将低分辨率CT图像,准确恢复至高分辨率CT图像;Almasni等^[8]使用YOLO网络对乳腺肿块进行自动检测,定位准确率达99.7%,并能准确地区分良性和恶性病变.

由于医学数据存在多源异构性、公开数据集缺乏、高质量数据标注困难、对领域知识有一定要求等诸多因素,致使医学数据的处理与分析发展较慢.从已有统计数据来看,深度学习在检验医学中的研究比较少,主要集中在血细胞识别、体液分析、免疫分析、微生物领域、寄生虫领域和基因及蛋白组学领域^[9-15],而这些绝大多数还是属于图像处理应用.

随着抗菌药物耐药性的加剧,抗菌药物的不合理使用问题愈来愈严重,构成全球公共卫生的一大威胁^[16].目前,各大医院耐药性检测方法多采用传统的细菌培养方法,从鉴定细菌到完成药敏检测至少需要3~5d时间,这种传统检测方法延误了医生对患者的抗菌药物耐药性的知悉时机,给临床用药带来挑战和困境.大数据分析技术是否能为快速、准确地使用抗菌药物另辟蹊径?目前,对抗菌药物的数据分析处理研究极少,Nguyen等^[17]使用机器学习的XGBoost算法,构建非伤寒沙门氏菌基因组与15种抗生素间的MIC预测模型;梁治钢、王一敏^[18-19]先后使用机器学习的免疫遗传算法和深度学习的长短期记忆网络模型,基于电子病历数据,对抗菌药物的是否合理使用进行了探索性研究.

本文将探索深度学习技术在细菌药敏检测历史大数据中的应用,提出一种融合注意力机制的双通道卷积神经网络模型,充分挖掘不同深浅层次的药敏检测文本信息,并为不同重要程度的信息赋予不同的权重,构建检测样本与抗菌药物间的分类模型,而后利用此模型判断待测样本对抗菌药物的耐受程度,第一时间为医生临床用药提供最适合个体患者的决策支持,为抗菌药物的合理使用提供辅助作用.

1 所用技术

1.1 卷积神经网络

卷积神经网络作为当前最为流行的深度学习网络模型之一,近些年在计算机视觉领域取得了令人瞩目的成绩.它具有优异的特征自学习能力,从底层图像特征逐步映射到高层语义特征,常被用于分类和预测.卷积神经网络主要由输入层、卷积层、池化层和输出层四部分组成.

1)输入层:卷积神经网络的输入层一般是二维数据格式,可以是二维图像矩阵,也可以是二维文本矩阵,本文是将基于患者检测样本的二维文本数据作为输入.

2)卷积层:卷积层利用共享的卷积核与输入层或者前一个卷积层的局部感受野做卷积运算,再经过激活函数(如Sigmoid、tanh、Relu)做非线性运算,得到特征值,卷积运算如公式(1)所示.

$$c_i = f(\sum w, x_{i:i-h+1} + b) \quad (1)$$

式中: c_i 表示第*i*个特征值, f 为激活函数, w 为卷积核, h 为卷积核窗口滑动大小, $x_{i:i-h+1}$ 为从*x*(输入层或者前一个卷积层)中第*i*行至第*i-h+1*行提取到的局部文本矩阵, b 为偏置项.

输入层*X*经过卷积后得到特征向量*C*:

$$C = (c_1, c_2, c_3, \dots, c_{i-h+1}) \quad (2)$$

3)池化层:该层主要用于减少卷积得到的特征图维数,多采用公式(3)所示的最大池化法提取局部最优值.

$$c_{\max} = \max(c_i) \quad (3)$$

4)输出层:将经过卷积和池化后得到的特征值拉平,再进行全连接运算,此过程中为了防止过拟合,使用Dropout暂时屏蔽一些神经元,以增强模型的鲁棒性.最后,将Dropout操作后所得到的向量使用softmax函数进行分类,输出预测结果,本文需要预测出当前患者样本对应某抗菌药物的MIC(Minimal inhibitory concentration:最低抑菌浓度)分类值.

1.2 文本词向量处理

检测样本所对应的文本数据要用于深度学习, 第一步需要将文本数据数字化, 即将文本映射为 k 维的向量空间. 传统的做法是使用 One-hot 编码, 将文本中的最小单位数据(可以是数字、字或词等, 简称元素)表示成一个向量. 具体做法是对文本按照需求进行元素分割, 建立一个包含所有元素的元素库, 向量的维度与库大小相同, 向量中, 只有表示该元素的位置为 1, 其它为 0.

One-hot 方法有两个主要缺点, 一是当元素比较多时, 容易造成维度灾难; 二是元素之间的相关性无法体现. 鉴于这些问题, 提出了分布式表示法, 早期是使用统计学方法按照元素出现的次数将每个元素表示成一个实数向量, 通过向量之间的余弦/欧式距离来表示元素间的语义相似性. 后来, 又提出更为容易的基于神经网络的向量训练方法, Word2Vec^[20]是目前所有分布式向量训练中效果最好的一种框架.

2 模型构建

本文所提出的融合注意力机制的双通道卷积神经网络模型, 如图 1 所示. 对原始的检验数据经过预处理后, 进行向量化建模, 以此作为输入数据, 被分别送往相互独立且深度不同的上下两个通道的卷积神经网络中, 每个通道经过若干卷积和池化操作后, 引入注意力机制, 通过拟合多组权重向量来表征各特征分量的重要性, 再经过一个全连接层, 而后将两个通道的特征数据进行融合, 最后利用 softmax 函数

实现分类输出.

2.1 药敏检测数据建模

本文研究的数据源是细菌药敏检测数据, 每一条检测样本包含患者的年龄、性别、所属科室、送检日期、样本类型、细菌种类、报告日期等送检数据, 以及该患者对各个抗菌药物的最低抑菌浓度 (MIC) 检出结果值. 从送检数据中选取有可能影响抗菌药物耐受值的多个属性作为输入特征, 以检出结果 MIC 作为分类标签. 考虑到不同菌种对同一种抗菌药物的敏感性不同, 并且同一菌种在不同标本中对同一种抗菌药物的敏感值也不尽相同. 如: 细菌 aba 和细菌 aca 对抗菌药物 AMP 的最低抑菌浓度不同, 同时, 在不同检测样本中, 细菌 aba 对抗菌药物 AMP 会表现出来多个不同的 MIC 值. 因此, 有必要为每种抗菌药物建立一个分类模型, 根据输入的多特征送检数据, 输出其对当前抗菌药物的 MIC 分类值.

包含多特征的送检样本数据向量化表示, 是卷积神经网络进行特征提取和融合的重要前提. 本文借鉴文本问题中构建词向量模型的做法, 每一个属性特征值对应一个词, 使用一个向量来表示, 特征向量维度为所有特征值的个数, 由于本文最后筛选得到的药敏数据特征向量维度为 42, 显然不存在维度灾难问题, 同时各个特征间关联性比较弱, 故选用传统的 One-hot 法构建特征向量模型, 如: 性别男这个特征被表示为向量(1000...000). 而一条包含多个特征的送检样本数据则对应文本问题中的一条句子, 建模后为一个二维矩阵, 矩阵的每一行即是一个特征向量.

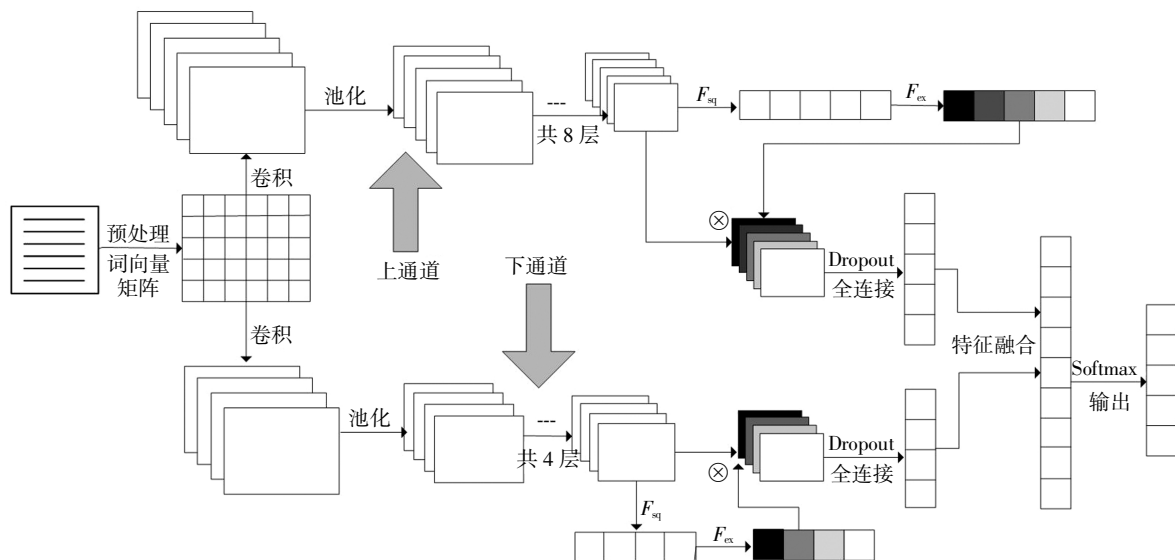


图 1 融合注意力机制的双通道卷积神经网络模型

Fig.1 Dual channel convolutional neural network model with attention mechanism

2.2 双通道卷积神经网络

卷积神经网络所提取的特征往往与网络深度有关,网络层次越深,越容易提取到表征整体的抽象特征;而网络层次越浅,则容易提取到表征局部区域的细节特征.鉴于此,本文设计了两种深度分别为6层和4层的双通道卷积神经网络,以提取深层和浅层不同抽象级别的特征,此外,考虑到检测样本的特征矩阵规模不大,故削弱了池化降维工作,上通道网络由5个不同卷积核大小和不同数量的卷积层、3个相同步长的最大池化层,以及1个全连接层组成,下通道网络由3个不同卷积核大小和不同数量的卷积层、1个最大池化层,以及1个全连接层组成.

上下两个通道,经过卷积、池化和全连接后,得到的特征向量分别记为 $c^{(1)}=(c_1, c_2, \dots, c_x)$, $c^{(2)}=(c_1, c_2, \dots, c_y)$,为便于后续处理,需要将两组特征向量进行合并,得到最终的特征向量记为 c ,可表示为:

$$c = c^{(1)} \oplus c^{(2)} \quad (4)$$

2.3 注意力机制

注意力机制,是通过给各个特征分量分配合适的权重,筛选出重要特征并且聚焦到这些信息上.它分成三个部分,分别为挤压(squeeze)、激励(excitation)和注意(attention),如图1所示.假设经过卷积和池化后的特征信息是一个包含 n 个特征向量的二维矩阵,记为 $U_{m \times n}$,进行挤压处理后,得到向量 $z=(z_1, z_2, \dots, z_n)$,挤压函数(F_{sq})的计算如公式(5)所示;随后,将向量 z 经过公式(6)的激励处理(F_{ex}),得到注意力权重向量 $a=(a_1, a_2, \dots, a_n)$;最后,将权重向量 a 与特征矩阵 U 相乘,如公式(7)所示,生成带注意力机制的特征输出 U' .

$$z = \frac{1}{m} \sum_{i=1}^m U_i \quad (5)$$

$$a = \text{softmax}(w_s \cdot \tanh(w_e \cdot z)) \quad (6)$$

$$U' = a \cdot U \quad (7)$$

式中: m 为特征向量维数, U_i 为第 i 个特征向量,softmax和tanh是激活函数, w_s 和 w_e 是可学习权重矩阵.

2.4 模型训练

采用softmax函数对双通道融合后的特征向量进行分类,输出结果类别的预测概率值:

$$P(k|x) = \frac{e^{x|k|}}{\sum_{i=1}^n e^{x|i|}} \quad (8)$$

式中: $P(k|x)$ 表示当前样本属于类别 k 的预测概率, n 是类别数, $x[k]$ 是网络前向传播得到的特征向量.

采用交叉熵损失函数,衡量预测概率与真实标签的差异,计算如公式(9)所示,以此进行模型的反向传播训练,目标是将函数的损失降低到最小.

$$\text{loss} = \sum_{i=1}^n q(k|x) \log(P(k|x)) \quad (9)$$

式中: $q(k|x)$ 表示当前样本对应类别 k 的实际编码向量,即真实标签.

3 实验及分析

3.1 数据集

本文数据来源为某三甲医院重症监护病房(ICU)2014年至2019年六年的细菌药敏检测历史数据,共计15128条.其中,细菌种类有156种,抗菌药物种类有53种,本文选取抗菌药物AMK的所有检测数据作为试点.实际生产数据比较复杂,必须经过预处理后才能用于实验,本文的数据预处理工作包括:数据过滤、异常值检测和规范化.第一,数据过滤:在检验领域专业人士的帮助下,对原始数据进行筛选,去除无关指标,挑选出有可能影响抗菌药物耐受值的5个属性特征作为输入,分别为患者性别、患者年龄、来源科室、标本类型和细菌种类;第二,异常值检测:根据专家自身检测经验,排查原始数据中存在的检测不合理的结果,同时,将缺失值较多以及记录数比较少的检测样本也纳入异常数据,直接删除它们.最终用于实验的数据集有1781条,其中,涉及到的细菌有aba、eco等5种,分类标签数有 ≤ 2 、 ≤ 4 等6个;第三,规范化:为方便后期的数据处理,对5个特征使用数值进行量化表示,以实现量化方式的统一.

3.2 评价指标

本文采用 F 度量值^[21]作为分类结果的评价标准.该方法同时兼顾了查准率(P)和查全率(R)两个指标, P 、 R 和 F 分别由式(10)~式(12)计算得到.

$$P = \text{TP}/(\text{TP} + \text{FP}) \quad (10)$$

$$R = \text{TP}/(\text{TP} + \text{FN}) \quad (11)$$

$$F = (2 \cdot P)/(P + R) \quad (12)$$

式中:TP为正类被划分为正类的样本数,FP为负类被划分为正类的样本数,FN为正类被划分为负类的样本数,TP+FP表示实际分类的样本数,TP+FN表示应有的样本数.

3.3 实验环境及参数设置

本文的实验环境:操作系统为Windows 10,处理

器为 Intel Core i3,GPU1060 6G, 内存 24G, 硬盘 256G, 编程语言为 Python 3.6, 深度学习框架为 Keras 2.1.2, 后端使用 Tensorflow 1.3.0.

为确保结果的客观性和有效性, 实验采取十折交叉验证法. 数据集被划分为 10 份, 1 份为测试集, 其余 9 份为训练集, 10 份样本随机轮流进行 10 次实验, 取平均值作为模型最终的结果值. 经过多次实验, 确定了一组最佳模型参数, 参数取值见表 1 所示.

表 1 关键参数设置

Tab.1 Key parameter setting

参数	参数值
最大迭代次数(epoch)	300
训练批次大小(batch size)	50
上通道卷积核个数	64, 128, 128, 128, 128
下通道卷积核个数	64, 128, 128
上通道卷积核大小	2, 3, 3, 3, 2
下通道卷积核大小	2, 3, 2
输出层 dropout 比率	0.5
优化器(optimizer)	Adam
学习率	0.000 5

3.4 实验结果与分析

3.4.1 对比方法介绍

为更好地对本文所提出的融合注意力机制的双通道卷积神经网络模型(CNN-Att-TChan)在药敏检测数据中的分类效果进行评估, 设计了如下对比实验:

1)单通道二维卷积神经网络模型(CNN-2D). 该模型基于传统的 CNN 网络, 卷积核为二维, 包含多层的卷积和池化操作.

2)带注意力机制的单通道二维卷积神经网络模型(CNN-2D-Att). 该模型在 CNN-2D 基础上, 增加了基于特征分量的注意力机制.

3)双通道二维卷积神经网络模型(CNN-2D-TChan). 在 CNN-2D 网络基础上, 设计了上下两个通道, 分别进行多次卷积和池化操作, 然后进行特征融合, 进而分类输出.

4)多通道一维有池化卷积神经网络模型(CNN-1D-MChan). 使用多个不同宽度, 但长度相同的一维卷积核从多个通道提取数据特征, 每一个卷积核长度等于药敏数据特征向量维度, 由于每个二维输入矩阵经过一次一维卷积后得到的已经是向量, 因此, 每个通道只进行一次卷积和池化, 其它操作与 CNN-

2D-TChan 类似.

5)多通道一维无池化卷积神经网络模型(CNN-1D-MChan-NoPooling). 在 4)中 CNN-1D-MChan 网络基础上, 去除池化操作, 每个通道经过一维卷积后, 直接进行向量融合, 而后进入全连接层.

6)经典深度学习模型^[22-25]. 除 1)-5)自构建的深度学习网络模型外, 还选用经典分类模型 ResNet-18、AlexNet-8、VGG-16, 进行对比实验.

7)机器学习模型. 选用机器学习分类方法中的随机森林算法(RF)和决策树算法(C4.5)作为代表, 进行对比实验. 为使实验结果更具有可比性, 采用与本文一致的 aba 等 5 种细菌和抗菌药物 AMK 作为输入, 为每个菌种和抗菌药物建立一对一的分类模型, 取 5 个模型的均值作为最终结果.

3.4.2 整体性能对比分析

不同模型的整体性能对比结果如图 2 和表 2 所示. 由于本次实验的数据属于结构化数据、数据量比较小、并且特征工程量不大等偶然因素, 机器学习不失为一种有效方法, 表现出了较好的分类性能. 但是实验过程中, RF 和 C4.5 两种机器学习方法也暴露

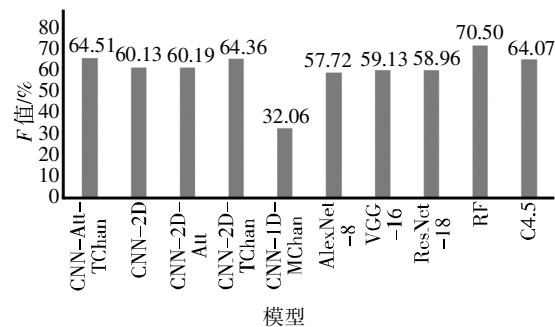


图 2 不同模型的 F 值比较

Fig.2 Comparison of F values of different models

表 2 不同模型的训练时长比较

Tab.2 Comparison of training time of different models

模型	训练时长/s
CNN-Att-TChan	103.52
CNN-2D	77.98
CNN-2D-Att	98.91
CNN-2D-TChan	97.81
AlexNet-8	173.63
VGG-16	200.54
ResNet-18	725.31
RF	60.23
C4.5	108.65

出一些问题,如:在面对训练集中的噪声数据时容易出现过拟合现象、随机森林法的 Bootstrap 自助采样做法会导致有些样本没有被训练到、不能表征特征之间的相互关联等,当训练数据量非常庞大时,模型性能会被弱化,相反此时深度学习模型却更具有优势。

本文主要探究深度学习模型在抗菌药物耐药性预测中的应用,从多组对比实验得到:

1) 分类准确度方面,本文所提出的 CNN-Att-TChan 模型与其它模型相比, F 值有一定提升,说明本文模型所引入的多通道等做法能有效提高分类准确度,如图 2 所示。第一,通道数影响。CNN-2D、CNN-2D-TChan、ResNet、Alexnet、VGG 的对比实验中,同是二维卷积,但是双通道网络表现出更好的分类效果,CNN-2D-TChan 的 F 值比 CNN-2D 约高出 7%,比经典模型平均约高出 10%,可见,增加通道数,能提取到更为丰富的融合高层次和低层次的语义特征,有利于模型的拟合训练;第二,网络层数影响。在 CNN-2D 和经典模型的对比实验中,同是单通道,同是二维卷积,但 CNN-2D 的 F 值比经典模型平均约高出 3%,这是因为本文实验经典模型的网络层数基本都比 CNN-2D 深,其中 ResNet-18 模型包含 17 个卷积层和 1 个全连接层,网络层数约 CNN-2D 的三倍,网络层次越高,越能提取到能表征局部和全局视图的复杂特征,这适合于复杂图像的认识与处理,而对于本文的药敏检测数据,虽然输入也可视为图像,但该图像矩阵的每一行仅有一个有效值,随着网络层数的深入,势必会引入更多的有效值以外的冗余信息,从而干扰模型的正常训练;第三,注意力机制影响。在 CNN-Att-TChan 与 CNN-2D-TChan 的对比实验中,前者比后者增加了一个注意力机制,从实验结果来看,分类准确度有一定提升,但提升幅度不是特别大,关于注意力机制的作用下文会详细叙述;第四,卷积核维度影响。在 CNN-2D、CNN-2D-TChan、CNN-1D-MChan 的一组实验中,发现一维卷积的分类准确度远远低于二维卷积,下文将进一步进行卷积核维度影响的对比实验。

2) 时间性能方面,Epochs 值为 50 时,各个模型的训练时长如表 2 所示,所提出模型的训练时长,比三个经典模型要短,但较网络层数相当的 CNN-2D 模型有所增加。实验结果表明,在二维卷积网络模型中,第一,添加注意力机制,训练时长约增长 26.84%,注意力机制是在“卷积池化后、全连接前”增加的一个操作,通过自学习给每个特征分量分配注

意力权重,必然会在时耗,但是由于注意力机制在各特征分量层面采取的是并行运算方式,因此对整个模型的时间效率不会有太大影响;第二,双通道较单通道,训练时长平均高出 25.43%,因为每增加一个通道,经过逐步卷积和池化后,势必会在全连接层产生“层内连乘、层间相加”的大量参数,极大地增加了计算量,从而影响网络的训练速度,可以说增加通道数是以降低模型执行效率为代价的。

3.4.3 单项性能对比分析

1) 注意力机制对比分析。在上述对比实验中发现,增加注意力机制,对分类准确度的提升虽然不是特别明显,但是被正确预测出来的分类数增多。CNN-2D、CNN-2D-TChan 和经典模型,在数次测试中最终预测正确的均只有“ ≤ 2 ”这一个分类,这可能是数据集不均衡原因造成,因为整个数据集有约 60%都属于“ ≤ 2 ”类别,大样本所带的信息量比小样本要高,使得分类器自然倾向于大样本数据。但是添加注意力机制后,发现有更多的其它小样本类别被正确预测出来,平均 100 次测试,有超过 50%的概率预测出其它小样本类别,这是因为注意力机制通过赋予高权重,能有机会聚焦到小样本特征。

为进一步验证,本文在一维卷积 CNN-1D-MChan 和 CNN-1D-MChan-NoPooling 模型中添加了本文所提出模型中的特征分量(通道域)注意力机制,辅以空间域注意力机制和混合注意力机制(空间+通道),进行对比实验,实验结果如表 3 所示。结果表明,注意力机制在一维卷积中的作用与在二维卷积中相当,增设注意力机制对于模型分类准确度影响不大,在一维卷积中甚至有微弱下降,但是能有效改善小样本的预测分类效果。此外,本文的通道域注意力机制,较空间域注意力机制和混合注意力机制,在预测准确度和小样本分类预测上的表现不分伯仲。

为进一步评估注意力机制对模型训练时间性能的影响,本文在一维卷积和二维卷积中分别进行了对比实验,图 3 为不同模型在不同的 Epochs 下完成一次迭代的训练时间变化曲线。由图 3 可见,两组对比实验中,添加注意力机制后训练时长均有一定程度的提升。添加了注意力机制后的一维卷积网络(CNN-1D-Att)与没有添加注意力机制的一维卷积网络(CNN-1D)训练时长提升幅度非常小,几乎可以忽略不计;而添加了注意力机制后的二维卷积网络(CNN-2D-Att)较没有添加注意力机制的二维卷积网络(CNN-2D),训练时长有一定差异,前者比后者

表 3 一维卷积中注意力机制对比实验

Tab.3 Contrast experiment of attention mechanism in one dimensional convolution

注意力机制	CNN-1D-MChan		CNN-1D-MChan-NoPooling	
	F 值	小样本分类出现次数(100 次测试)	F 值	小样本分类出现次数(100 次测试)
无注意力机制	32.06	0	40.85	3
通道注意力机制	30.84	54	38.95	64
空间注意力机制	31.92	58	33.14	53
混合注意力机制	31.17	62	35.31	61

平均高出 25.36%,如图,当 Epochs 为 300 时,训练时长多出约 42 s,但相对 CNN-2D 原始时长 440 s 来说,这个差值可以说比较小,故再次验证添加注意力机制对于模型训练时间整体性能影响不大。

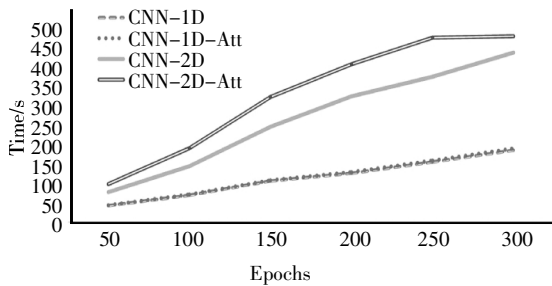


图 3 注意力机制对训练时间的影响

Fig.3 Effect of attention mechanism on training time

2)一维/二维卷积对比分析.考虑到本文特征向量采用类词向量建模的方法,输入矩阵中的每一行构成一个完整的药敏特征向量,因而有必要进行一维卷积核和二维卷积核的对比实验,一维卷积核的窗口长度等于样本特征向量的维数,使得卷积操作只在输入矩阵的一个方向上进行,以便提取一到多个完整的特征信息.设计了一组对比实验,分别是一维卷积模型(CNN-1D)、一维无池化卷积模型(CNN-1D-NoPooling)、普通二维卷积模型(CNN-2D),结果如图 4 和图 5 所示。

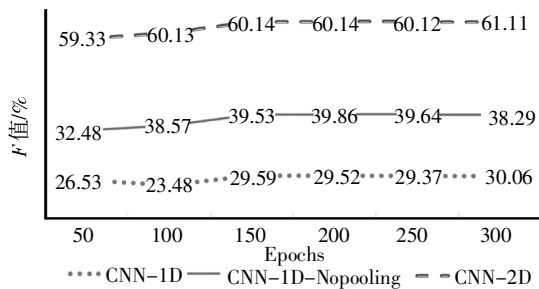


图 4 一维/二维卷积的 F 值比较

Fig.4 Comparison of F-values of one-dimensional and two-dimensional convolutions

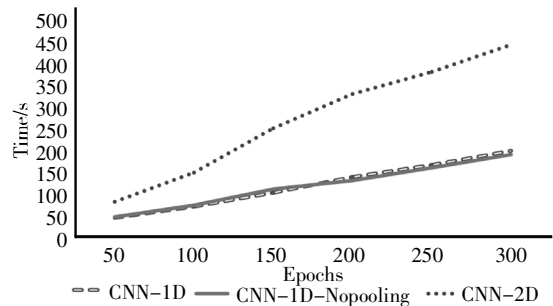


图 5 一维/二维卷积的训练时长比较

Fig.5 Comparison of training time of one-dimensional and two-dimensional convolution

图 4 结果表明,CNN-1D-NoPooling 的分类准确度略高于 CNN-1D,但它们都远远低于 CNN-2D,这进一步验证了前文一维卷积在分类准确度方面比二维卷积效果要差的结论,其原因主要是本文的一维卷积操作受输入矩阵规模局限,只有一层,只能提取低层次特征,虽然移除了池化层,不对特征做精简以保留尽可能多的特征,但由于一维卷积提取到的特征单一、泛化能力弱等问题,削弱了模型训练质量.此外,在时间复杂度影响方面,如图 5 所示,一维卷积的训练时长要比二维卷积少出约 60%,一方面是由于一维卷积的网络模型层数少,导致网络模型整体复杂度降低,另一方面是一维卷积过程中卷积核规模、输出特征规模变小致使单个卷积层时间复杂度下降。

4 结论

本文提出一种融合注意力机制的双通道卷积神经网络模型用于检验医学中抗菌药物的耐药性预测.该方法通过上下两个通道对建模后的送检数据矩阵做不同粒度的特征提取,每个通道经过卷积和池化后引入注意力机制,使模型聚焦重要的特征信息,而后将两个通道的特征进行融合,从而完成分类输出。

将模型在某三甲医院细菌药敏检测历史数据集上,与多种不同方法进行了对比实验,结果表明,本文所提出方法在整体分类准确度和小样本分类方面表现出更好的效果,具有一定的可行性和有效性。下一步工作,将对原始数据存在的细菌类型和类别不均衡问题作进一步的探索和研究,提高输入数据质量,从而进一步提升模型分类质量。

参考文献

- [1] 程学旗,靳小龙,王元卓,等.大数据系统和分析技术综述[J].软件学报,2014,25(9):1889—1908.
CHENG X Q, JIN X L, WANG Y Z, *et al.* Survey on big data system and analytic technology [J]. Journal of Software, 2014, 25 (9): 1889—1908. (In Chinese)
- [2] LECUN Y, BENGIO Y, HINTON G. Deep learning [J]. Nature, 2015, 521(7553):436—444.
- [3] 何炎祥,孙松涛,牛菲菲,等.用于微博情感分析的一种情感语义增强的深度学习模型[J].计算机学报,2017,40(4):773—790.
HE Y X, SUN S T, NIU F F, *et al.* A deep learning model enhanced with emotion semantics for microblog sentiment analysis [J]. Chinese Journal of Computers, 2017, 40(4):773—790. (In Chinese)
- [4] MIOTTO R, WANG F, WANG S, *et al.* Deep learning for healthcare: review, opportunities and challenges [J]. Briefings in Bioinformatics, 2018, 19(6):1236—1246.
- [5] YASAKA K, AKAI H, ABE O, *et al.* Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced CT: a preliminary study [J]. Radiology, 2018, 286(3):887—896.
- [6] JIANG Z X, ZHANG H, WANG Y, *et al.* Retinal blood vessel segmentation using fully convolutional network with transfer learning [J]. Computerized Medical Imaging and Graphics, 2018, 68:1—15.
- [7] YOU C Y, LI G, ZHANG Y, *et al.* CT super-resolution GAN constrained by the identical, residual, and cycle learning ensemble (GAN-CIRCLE) [J]. IEEE Transactions on Medical Imaging, 2020, 39:188—203.
- [8] ALMASNI M A, AL-ANTARI M A, PARK J M, *et al.* Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system [J]. Computer Methods and Programs in Biomedicine, 2018, 157:85—94.
- [9] 严虹,刘国焯,李砚,等.深度学习在检验医学中的研究与应用[J].中华检验医学杂志,2019,42(12):1063—1066.
YAN H, LIU G Y, LI Y, *et al.* Research and application of deep learning in laboratory medicine [J]. Chinese Journal of Laboratory Medicine, 2019, 42(12):1063—1066. (In Chinese)
- [10] SHAHIN A I, GUO Y, AMIN K M, *et al.* White blood cells identification system based on convolutional deep neural learning networks [J]. Comput Methods Programs Biomed, 2019, 168:69—80.
- [11] JAVADI S, MIRROSHANDEL S A. A novel deep learning method for automatic assessment of human sperm images [J]. Computers in Biology and Medicine, 2019, 109:182—194.
- [12] LI Y Q, MAHJOUBFAR A, CHEN C L, *et al.* Deep cytometry: deep learning with real-time inference in cell sorting and flow cytometry [J]. Scientific Reports, 2019, 9(1):11088.
- [13] JO Y, PARK S, JUNG J, *et al.* Holographic deep learning for rapid optical screening of anthrax spores [J]. Science Advances, 2017, 3(8):e1700606.
- [14] ZHANG Y B, KOYDEMIR H C, SHIMOGAWA M M, *et al.* Motility-based label-free detection of parasites in bodily fluids using holographic speckle analysis and deep learning [J]. Light: Science & Applications, 2018, 7:108.
- [15] 赵新元,秦伟捷,钱小红.深度学习在生物质谱及蛋白质组学中的应用[J].生物化学与生物物理进展,2018,45(12):1214—1223.
ZHAO X Y, QIN W J, QIAN X H. Application of deep learning in biological mass spectrometry and proteomics [J]. Progress in Biochemistry and Biophysics, 2018, 45(12):1214—1223. (In Chinese)
- [16] SHI J H, YAN Y, LINKS M G, *et al.* Antimicrobial resistance genetic factor identification from whole-genome sequence data using deep feature selection [J]. BMC Bioinformatics, 2019, 20(15):1—14.
- [17] NGUYEN M, LONG S W, MCDERMOTT P F, *et al.* Using machine learning to predict antimicrobial MICs and associated genomic features for nontyphoidal salmonella [J]. Journal of Clinical Microbiology, 2019, 57(2):e01260—18.
- [18] 王一敏,梁冶钢.基于免疫遗传算法的抗菌药物数据挖掘[J].计算机系统应用,2017,26(3):156—161.
WANG Y M, LIANG Z G. Data mining of antimicrobial drug based on immune genetic algorithm [J]. Computer System and Application, 2017, 26(3):156—161. (In Chinese)
- [19] 梁冶钢,王一敏.深度学习在电子病历抗菌药物使用方法分类中的应用[J].计算机系统应用,2019,28(8):71—77.
LIANG Z G, WANG Y M. Application of deep Learning in classification of antimicrobial using methods in electronic medical records [J]. Computer System and Application, 2019, 28 (8):71—77. (In Chinese)
- [20] MIKOLOV T, CHEN K, CORRADO G, *et al.* Efficient estimation of word representations in vector space [EB/OL]. arXiv:1301.3781, 2013—5—5.
- [21] 钟将,刘荣辉.一种改进的KNN文本分类[J].计算机工程与应用,2012,48(2):142—144.
ZHONG J, LIU R H. Improved KNN text categorization [J]. Computer Engineering and Applications, 2012, 48 (2):142—144. (In Chinese)
- [22] 陈红松,陈京九.基于ResNet和双向LSTM融合的物联网入侵检测分类模型构建与优化研究[J].湖南大学学报(自然科学版),2020,47(8):1—8.
CHEN H S, CHEN J J. Study on construction of IOT network intrusion detection classification model and optimization based on combination of ResNet and bidirectional LSTM network [J]. Journal of Hunan University (Natural Sciences), 2020, 47(8):1—8. (In Chinese)
- [23] ELNASIR S, SHAMSUDDIN S M. Proposed scheme for palm vein recognition based on linear discrimination analysis and nearest neighbour classifier [C]//2014 International Symposium on Biometrics and Security Technologies (ISBAST). Kuala Lumpur, Malaysia: IEEE, 2014:67—72.
- [24] KRIZHEVSKY A, SUTSKEVERI, HINTON G E. ImageNet classification with deep convolutional neural networks [C]//International Conference on Neural Information Processing Systems. New York: Curran Associates Inc, 2010:1097—1105.
- [25] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. 2014.